Politecnico di Torino

Master's Degree Programme in Environmental and Land Engineering

A.A. 2023/2024
Graduation Period October 2024

Dams and Flood Reduction

A modelling study of the effects of dams on flood reduction in Germany

Supervisors:
Prof. Alberto Viglione
Prof. Dr. Ralf Merz

Candidate:
Serena Giammarino

Table of Content

List of Figures

Tables

Abstract

The main intent of this thesis is the evaluation, through modelling scenarios, of how large-scale water flows patterns in Germany would be modified if reservoirs were missing. This issue is not too popular among the scientific panorama, that, being quite divided on the question of dams bringing more advantages to people and economy or disadvantages to the environment and the ecosystems, often focus their studies on a single catchment or small geographical areas.

To evaluate the relationship between rivers discharge and the action of the dams, machine learning tools are exploited to run SALTO. SALTO is a distributed conceptual rainfall-runoff model on large scale that was first locally and then regionally calibrated with the PASS approach, an algorithm that works without any previous definition of the dominant catchment descriptors controlling regional patterns (Merz, Tarasova and Basso, 2020). The calibration function is defined by a weighted sum of Kling-Gupta efficiency and a metric that focuses on flood quantiles evaluated for the return period of 5 years. With this model it is possible to simulate the runoff both in normal conditions and in a scenario in which the action of the dams is ignored.

Analysis of model efficiencies after the local calibration show that three quarters out of the total catchments assume values higher than 0.75, suggesting good performances, and that to larger basins generally correspond lower results. Comparisons between the model efficiency after the regional calibration in normal conditions and when neglecting the action of the dams, show that the second one is in many cases higher than the first, and this reflects the higher complexity that the algorithm face when it considers basins.

By effectively plotting hydrographs for selected catchments, the difficulties of the model in reproducing the runoff arise. Nevertheless, their analysis, together with flow duration curves, highlight some peculiarities in the behaviours of the streamflow in those catchments containing dams, that distinguished themselves from the watersheds without barriers.

Anyway, the key point of the study is the analysis of flood quantiles (determined for return periods of 2, 5, 10, 20, 50, 100 and 500 years), based on runoff computed in three different cases: from observations, from simulations considering the effects of the dams and from simulations ignoring them. Two kinds of tests are carried out: the computation of relative differences between the first two abovementioned datasets, and the calculation of relative differences between the two distinct simulations. To reach some plausible conclusion, all the catchments for which the second analysis resulted to be a negative number (ambiguity of dams that increase the magnitude of the runoff) or a value greater than 10 (dams that would reduce the flux of tens, hundreds or even higher orders of magnitude) are neglected. With filtered results, it can be concluded that, for the majority of catchments, the model slightly underestimates the runoff with respect to the one relative to observations for smaller return periods up to 50 years; on the other hand, on average, dams make the runoff decrease of around 20% to 30%, according to the different return period.

Of course, when considering results of all catchments, many no-senses and absurdities arise, and this imprecise outcome could be imputed to impressive approximation in data or to inadequacies during the calibration procedure.

1 Introduction

1.1 Dams: a scientific overview

Dams are structures capable of capturing water and altering magnitude and timing of its flow downstream (Leroy Poff and Hart, 2002). Built in the whole globe since the last 5000 years, they faced a big push in the 20 years following the end of the Second World War, especially in the United States (Graf, 1999), and in Europe (Tianbo Zhang and Xinyi Gu, 2023), due to increasing attention in hydropower and water resources management. On the contrary, nowadays, their construction is dealing with a significant shrinkage. As a matter of fact, in the last decades the public attention is mainly thrown towards dams' consequences and effects on the surrounding nature and on the ecosystems, leading to the development of a new "Science of Dam Removal" (Leroy Poff and Hart, 2002) that carefully analyse the balance between benefits and costs of the barriers from both an economical and environmental point of view (Graf, 1999).

This necessity arose naturally when the dimensions of dams started to grow remarkably: from the quite small structures of the past, that reflected the limited preindustrial technical skills but at the same time managed to satisfy the minimal needs of the agrarian society of ancient ages, to the evolution during 19th and 20th centuries that led towards larger and larger dams. This made experts begin to investigate not only about their positive effects on small scale, often undeniable and self-evident, such as improvement in irrigation capability, progress in navigation skills, increase in protection from floods, and boost in recreation opportunities (Graf, 1999), but also about their influence on the territory on wider proportions.

When thinking about effects of dams on large scale, a big portion of the literature focuses on fluvial ecology and ecosystems. The majority of research is conducted on understanding how dams cause the interruption of river connectivity, that is the ability for water to flow freely through a system without anthropogenic intervention (Spinti, Condon and Zhang, 2023), since currently only 23% of river worldwide travels uninterrupted to the ocean (Boulange, Hanasaki, Yamazaki and Pokhrel, 2021). The consequences of this situation are primarily damages for migratory fishes (even if precautional measures such as fish ladders are being adopted to minimize their vulnerability) (Spinti, Condon and Zhang, 2023), and also an increment of the potential risk of pollution, with the formation of contaminated settlements, the reduction of dissolved oxygen content, and the interruption of movement of sediments that leads to debris deposition (Yi, Gao and Zhang, 2020). Other repercussions are associated with increasing temperatures (Leroy Poff and Hart, 2002), alterations in hydrological dynamics, and variation in the water flow, where the discharge amplitude in some specific cases was modified by up to three orders of magnitude (Chaudhari and Pokhrel, 2022). However, this last part is still not very popular in the scientific literature, and the existing studies mainly focus on one single river and never analyse the problem on a wider spatial domain.

Actually, the need of including the action of dams while modelling the runoff is fundamental, as recognized in the research of Boulange, Hanasaki, Yamazaki and Pokhrel (2021). This study regards the relationship between dams and floods and the damages that they will provoke to the exposed population in a near future. This issue has a strong association with climate change, well known to alter distribution, variability and intensity of precipitation events. The authors quantify how much the action of dams, capable of modifying the frequency, the duration and the timing of annual flooding events, could significantly affect the estimation of future population exposure to calamities, concluding that downstream of dams, floods occurred less frequently in the circumstance in which barriers are considered than when their action is neglected.

The study from Cipollini, Fiori and Volpi (2022) goes even further, focusing on the attenuation of flood events due to the superposition of the action of many dams. This question could be solved by treating all the reservoir along a main channel as a single equivalent one governed by a physically based index R. Furthermore, another research on what effectively reduces the downstream flood peak discharge when reservoirs are present is conducted by Volpi, Di Lazzaro, Bertola, Viglione and Fiori (2018). The authors conclude that there are three main factors controlling it: the spillway dimensions, the storage capacity and the reservoir position with respect to the river channel. Each one of these characteristics is quantified by a parameter (obtained as a ratio between specific attributes of the dam): respectively, an increase of the first and the second leads to a bigger attenuation of the degree of flood peak, whereas regarding the third parameter, a range of optimal position of the dam along the river exists to maximise the attenuation.

Just by reading this overview, is it easy to notice that the scientific panorama has very different opinions regarding the presence of dams and their advantages, according to their priority being the healthiness of the river and its ecosystems or the defence against floods. Keeping in mind that both issues are fundamental, this study will develop closer to the second direction, maintaining the focus on the present-day situation in Germany.

1.2 About this thesis

The central topic of this thesis is the relationship between the river discharge and the action of dams. More specifically, the key point is the evaluation, through modelling scenarios, of how large-scale water flows patterns in Germany would be modified if reservoirs were missing.

To obtain results, machine learning tools (especially the R language and its extensions) were used to run SALTO, a distributed conceptual rainfall-runoff model on large scale, that was first locally and then regionally calibrated with the PASS approach, developed by Merz, Tarasova and Basso (2020).

According to the authors, the peculiarity of this method is that it works without any a priori definition of the dominant catchment descriptors that control regional patterns. This allows to overcome a significant challenge, since the relationship between catchment descriptors and model parameters is often hidden in the data and not totally satisfying. Additionally, the dominant descriptors might be modified with the spatial scale of the model, thus failing to provide consistent results on the spatial variability of parameters.

From the results it was possible to realize and plot many different charts (using two software, R and QGIS), that allow to have graphical feedback of the computations, in order to easily look up for relationships in space and time, and draw some final conclusion on the matter.

2 Data

2.1 German hydrology in a nutshell

To understand better the hydrogeological processes analysed in this study, it is essential to locate geographically and climatically the area of interest, that is Germany. Situated in the centre of Europe, this country has generally a moderately humid and temperate climate (Karthe et al., 2017). More in detail, it is oceanic in the western and northern parts, and the temperature is usually moderate by the coast. Also, mild winters and cool summers are typical of these regions. The central and eastern areas of Germany on the other hand, have a more continental climate and higher temperature difference between the seasons. Moreover, on a wider perspective, variations of large-scale pressure systems such as the North Atlantic Oscillation (NAO) influences widely the climatic situation of the county (Karthe et al., 2017).

Regarding hydrological resources, Germany is considered to be quite rich thanks to its 188 km^3 of water available per year (Karthe et al., 2017). Out of the total, around one fifth is used for human purposes, mainly as a coolant for power stations (more than 60% of the overall), but also for other industrial functions, for household use, for irrigation and agriculture (Hirschfeld, Nilson and Keil, 2014).

Despite this positive picture, the distribution of water supplies is uneven between the *Länder*, mainly due to the different precipitation conditions: a lot of rain and snow falls in the Alps and central Uplands, whereas way less hits the flat North, that is also characterised by nearly total evaporation. As a matter of fact, although the average national precipitation is about 790 mm per year, this value varies broadly between higher mountain zones, that reach values of over 2000 mm/year, and the North German Plain, characterised by averages around 500 and 700 mm/year. Also, there is a decreasing trend in precipitation going eastward, thus reflecting the more and more continental climate that is faced when moving from west to east (Karthe et al., 2017).

The unequal distribution of water resources characterises the groundwater situation as well, with isolated and often unproductive aquifers in the South alpine region, and complex interactions between groundwater and surface water, interlaced with the presence of many lakes too, typical of the Northern lowlands (Karthe et al., 2017).

This bipartition regarding the whole hydrological panorama of Germany is tried to be levelled off by a very efficient extraction and distribution system that is able to satisfy the water demand of the entire nation (Karthe et al., 2017). However, these differences are challenging when there is the necessity of dealing with droughts and low flows in some regions of Germany, and face the risk of severe floods in others just a few hundred kilometres away.

In addition to this very peculiar situation caused by the different morphology at specific latitudes and altitudes, in last decades the situation has been accentuated by the aggressive action of climate change. As an example, last 10 years have been the warmest registered since 1881(Ionita, Scholz and Grosfeld, 2022), especially 2018 and 2019 (Ionita, Nagavciuc, Kumar and Rakovec, 2020), and the situation got even worst during 2022 and 2023[1]. However, the amount of precipitation has not

[1] Statista Research Department, Jan 3, 2024.

decreased, but has just changed its distribution: throughout the last two decades there have been large amounts of rain falling in always shorter intervals of time (that often resulted in huge damages to both the population and economy) commonly followed by long periods of really low or zero precipitation (Ionita, Scholz and Grosfeld, 2022).

The increase in flood risks could also be attributed to modifications in the hydromorphology and to the conversion of floodplains into agricultural land and urban areas, leading to a desiccation of important wetlands as well (Karthe et al., 2017). Furthermore, the presence of many forests, that occupy about 31% of the German surface, is a key factor regarding the hydrological cycle and the water retention, since they are able to delay the effects of heavy rain events and their runoff and to produce unpolluted groundwater zones (Karthe et al., 2017).

The last major aspect to take into consideration in order to have an exhaustive picture of the hydrological situation in Germany, especially when dams and their effects are involved, is water quality. Even though wastewater treatment has massively improved since the second half of the last century, such a densely populated country still has to face the high pollutant concentrations in water, mainly due to agricultural, industrial and mining activities (Karthe et al., 2017). Moreover, the nowadays challenge is embodied by new emerging pollutants and their transformation products, that are an even bigger concern for public health, because it is not clear how to face them (Karthe et al., 2017).

2.2 Inputs of the model

The hydrological status of Germany just described should be quantified by numerical data, in order to be handled by the model.

As a matter of fact, the code requires as input meteorological information collected in a gridded form: daily values of precipitation, air temperature, and potential evapotranspiration for every cell of dimensions 5x5km (Merz, Tarasova and Basso, 2020).

Precipitations belong to HYRAS dataset, in which values were calculated using the REGNIE method, that is a combination of multiple linear regression considering orographic conditions and inverse distance weighting (Rauthe et al., 2013). Air temperature comes from the open data server of German Weather Service (DWD) (Merz, Tarasova and Basso, 2020). Potential evapotranspiration has been computed through the Hargreaves Method (Hargreaves and Samani, 1985), based on an empirical relationship in which reference evapotranspiration was regressed with solar radiation and air temperature data[2].

Timeseries of the observed runoff are also needed. Overall, the period of activity of the model goes from November 1st, 1985, to October 31st, 2000, because for every cell there are data for different intervals of time, but at least values from these years are available (Merz, Tarasova and Basso, 2020).

The list of all the catchment descriptors (from the work of Merz, Tarasova and Basso, 2020) is displayed in the appendix.

[2] USACE Engineering Hydrologic Center.

3 The Model

3.1 Description of the model

To get results on the flood variations, machine learning tools have been used. The model performed is developed in the R Language: the code can be divided into three main sections, that are the local calibration, the regional calibration, and the execution of the model. This "nucleus" is then integrated by an introduction, a plot section, and a part for the calculation of flood quantiles. In chronological order:

- The introduction is needed to prepare the environment in which the computations will be executed, with the definition of working directories and the loading of libraries required as first step, followed by the reading of input data from specific directories, and the supply of two source files containing functions for model calibration and operation;
- The local calibration is fulfilled as a loop for every catchment. Its output are lumped parameters, that are constant values for all model elements within the same catchment (Merz, Tarasova and Basso, 2020);
- PASS approach is used as an alternative to the regional calibration to get a consistent parameterization of the distributed model (Merz, Tarasova and Basso, 2020). It is built upon the best results obtained from the local calibration;
- SALTO model is finally executed using calibrated parameters from PASS. It returns the simulated runoff under the two conditions and statistics about the model efficiency of every catchment;
- Results of the model can be plotted as time series;
- The last part is an analysis of flood quantiles according to a GEV distribution based on the maximum annual runoff data for selected return periods.

Outstanding phases are thereafter explained.

As already briefly mentioned, the interval of time with available data spreads for 15 years (from 1/11/1985 to 30/10/2000); however, the first 5 years (from 1/11/1985 to 30/10/1990) are just used as a warm-up period to allow the model to reach an equilibrium and work at its best. Hence, the calibration process works with data successive to this preparation term.

Moreover, further controls on precipitations, typically collected from 7 a.m. of one day to 7 a.m. of the next, should be carried out, to verify the accordance between different dataset that could possibly not refer to the exact same day.

Metadata file containing information on the spatial domain, divided into squared cells of length 5 km each, are read. Every cell gets as input the timeseries of precipitation, temperature, and evapotranspiration. Moreover, several information about catchment is read by the code, and among these there are also the areas conveying into the dams' basins that were previously computed on QGIS.

Then the local calibration takes place. First of all, a parallelization process is arranged to faster calculate 25 independent parameters sets (and their efficiencies) for every catchment. Some of them, according to their efficiency (that should be high) and the number of valid results obtained, will be later used as an input for the PASS approach (Merz et al., 2022).

The local calibration works thanks to two external functions and scripts required for the calibration process. They are the DDS (Dinamically Dimentioned Search) algorithm, and the SALTO_Model_function_reservoir, that is a function to run the SALTO model in Fortran.

The objective function to calibrate the model is given by:

$$(1)\ ME \leftarrow 0.6*ME.KGE + 0.4*(1 - me.fl).$$

ME.KGE is the sum of Kling - Gupta efficiency and Kling - Gupta computed on inverse streamflow efficiency, in which both terms have the same mathematical importance (Merz, Tarasova and Basso, 2020):

$$(2)\ ME.KGE \leftarrow 0.5*KGE(qobs,qsim,iwarmup) + 0.5*KGE(1/qobs,1/qsim,iwarmup)$$

with the Kling – Gupta efficiency KGE given by:

$$(3)\ KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2},$$

where r is the linear correlation between observations and simulations, α a measure of the flow variability error, and β a bias term (Knoben, Freer, Woods, 2019).

The second term, *me.fl*, is a metric that focuses on floods, and it is estimated by calculating the GEV distribution for the return period of 5 years for the observed and for the simulated streamflow (using the annual highest values). Then, the square roots of the ratios of these values are computed, and this corresponds to *me.fl*.

Actually, in the previous formula *(1)*, the addend is *1 - me.fl*. This is because a low value means that the ratios of simulations to observations are close to 1, thus the model is reproducing fairly the flood magnitudes; on the other hand, when the result is high, the ratios are differing from 1, and simulations are more imprecise.

The regional calibration procedure on the other hand, consists of finding other parameter sets to improve their similarities with the catchment attributes. To guarantee the outcome of more effective and accurate results, it is helpful to provide some catchments for the training of the model. They must meet some criteria, that are: at least two parameter sets computed, a 70% of efficiency reached, and an extension of less than 10000 km^2 found. This step allows reducing the risk of having non representative inputs or overfitted models, minimizing at the same time the number of iterations to obtain adequate results.

The PASS approach can be summarised in the following steps (Merz, Tarasova and Basso, 2020). First of all, all the catchments with low model performances resulted from the local calibration are removed. For all the others that provided good results, one out of the best lumped parameter set resulting from the local calibration is randomly selected. This random selection is repeated for each iteration of PASS. Afterwards, with the help of data mining tools, a regional relationship between catchment descriptors and each model parameter is built. The machine learning tool used in this case is the “Random Forest” technique, where every tree decides according to random data, and then all the decisions are combined together. This method is selected due to its capability of handle complex relationships without prior assumptions. Then, the whole procedure could be repeated until better results of model efficiencies are found. The next step is to apply the regional functional relationship obtained to predict parameters for each other element of the distributed model: this will ensure parameters consistency across the region.

During the process, PASS is executed with different sampling modalities: random and optimal. The first one is used to randomly select parameters according to the "try and error" technique, in order to explore many different combinations of basins, selection criteria and calibration settings to determine which is the best choice for the model. Later these results are iteratively improved, and the optimization modality is designated to ultimately select the best solutions.

Finally, the distributed model is run using the predicted regional parameters.

Introduced by Merz, Tarasova and Basso (2020), SALTO is a conceptual rainfall – runoff bucket-style model to keep track of the soil moisture. The idea behind it is the same that regulates other large scale hydrological models, such as the HBV (Hydrologiska Byråns Vattenbalansavdelning) model (Bergström, 1995) or the components of the SUPERFLEX modeling toolbox (Fenicia et al., 2011).

For the estimation of the input elements, that are precipitation, snowfall and snowmelt, the Degree-Day method with threshold temperatures is used. Output elements, that are actual evaporation, runoff and percolation, are calculated as a nonlinear function of the corresponding soil moisture state. Precipitation and snowmelt accumulate in the different layers of the soil, and then water is released as actual evaporation, as percolation into the lower layer, or as direct runoff that is conveyed to river and described by a nonlinear reservoir approach. Groundwater dynamics are reproduced by a slow reacting non-linear reservoir that holds water percolating from deepest layer of the soil. Is it possible to convert this amount of water either as river runoff or as part of the groundwater reservoir of the downstream model element. Hence, the river runoff is described by a non-linear reservoir approach, and it is sustained by the runoff coming from different soil layers, by the groundwater inflow and by the river runoff of upstream soil stratum. (Merz et al., 2022).

The list of the 21 parameters employed in the SALTO model is reported in Table 1, together with their description and with the minimum and maximum value that they can assume. Intensive parameters, that are the ones for which the minimum value is equal to the maximum value, were not calibrated.

Table 1: Parameters of SALTO.

PARAMETER NAME	DESCRIPTION	MINIMUM VALUE	MAXIMUM VALUE
TS	threshold temperature for rain/snow and melt [degC]	-2.0	2.0
D_TS	temperature range above/below TS in which precipitation is mix of snow and rain [degC]	0.0	3.0
DDF	degree day factor [mm/degC/timestep] for no rain	0.1	10.0
D_DDF	increase of degree day factor [DDF/mm rain] with rain for rain on snow events	0.1	2.0

SM_LAYER	number of soil layers (here PAR_MIN=PAR_MAX) (Parameters are numbered from top SM_MIN_1, SM_MIN_2, ...)	1.0	1.0
SM_MIN_1	min soil moisture storage [mm]	0.1	100.0
SM_MAX_1	max soil moisture storage [mm]	10.0	1000.0
BETA_RC_1	the non-linear parameter for runoff production [-]	0.1	8.0
BETA_AET_1	the non-linear parameter for evapotranspiration [-]	0.1	2.0
K_LF_1	storage coefficient for fast response [timestep]	0.1	10.0
BETA_LF_1	power coefficient for fast response [-]	0.1	5.0
PERCMAX_1	max percolation rate [mm/timestep]	0.1	10.0
BETA_PERC_1	the non-linear parameter for percolation [-]	0.1	2.0
K_VZ	storage coefficient for vadose zone [timestep]	0.1	10.0
BETA_VZ	the non-linear parameter for vadose zone percolation [-]	0.7	1.5
K_GW	storage coefficient for groundwater response [timestep]	10.0	400.0
BETA_GW	the non-linear parameter for groundwater response [-]	0.7	1.5
S_GW_MAX	max. groundwater storage [mm]	100.0	5000.0
BETA_DQ_GW	the non-linear parameter to separate between intercell GW flow and GW BF	0.5	2.0
K_RIVER	storage coefficient for river routing [timestep]	0.1	5.0
BETA_RIVER	the non-linear parameter for river routing [-]	0.1	30.0

The results of the model can be graphically reproduced. Charts show plots of precipitation [mm/day], actual evapotranspiration (AET) [mm/day] and snow water equivalent (SWE) [mm/day], soil moisture (SM) [%], groundwater storage (GW storage) [mm] and groundwater recharge (GWR) [mm/day], and runoff calculated with both observed data and simulation results [mm/day]. Examples could be found in the Appendix.

Lastly, flood quantiles were computed for return periods of 2, 5, 10, 20, 50, 100 and 500 years for all catchments with at least 5 years of available data. Annually, sets with highest values were fitted to a GEV (Generalised Extreme Values) distribution, that is the most appropriate one to describe flood frequency data (Singh, 1998). According to the extreme value theorem, the GEV is the limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. Thus, it can be used as an approximation to model the maxima of long and finite arrangements of random variables[3].

[3] NASA.gov

3.2 Model functioning

The hydrological model in R recalls some subroutines from a Fortran code to calculate the filling and emptying of reservoirs (RESERVOIR_OPERATION), the snow movement (SALTO_SNOW), the runoff in a soil layer (SALTO_SOILLAYER), the simulation of the travelling time between soil moisture layer and ground water storage (SALTO_VADOSEZONE), the groundwater storage (SALTO_GW) and the river routing (SALTO_riverrouting).

RESERVOIR_OPERATION is the subroutine that simulates the functioning of the basin considering all the parameters that affect the process, distinguishing between the ones that increase the level of water inside the reservoir, and the ones that decrease it.

First of all, the amount of water inside the basin is updated according to the ingoing quantity from rivers that enters the lake (RES_q_in), multiplied by a timestep, and to direct precipitation on the lake surface (RES_prec), times the lake area. These quantities are added to the already present amount of water.

Then, terms to be subtracted from the total volume are the water evaporating from the lake (RES_pet) and the one needed for irrigation or environmental purposes (RES_quse_target), which is subtracted from the reservoir before calculating the outgoing flux (RES_q_out). To estimate it, the following scenarios must be considered.

If the updated quantity of water overcomes the maximum capacity of the basin (RES_capacity), then the surplus is released to keep the volume at the maximum level.

If the new quantity of water is less than the maximum capacity, then the model has to decide how much water to release, according to the amount needed downstream (for agricultural use or environmental needing), and the one required inside the basin (as reserve or prevention from inundation).

As starting value, the water exiting the basin is set equal to the one entering it. Then a total error on the discharge is calculated, as the weighted sum of two differences (the weighting factor is the timestep): the actual discharge (RES_q_out) minus the target discharge (RES_q_out_target), and the amount of water resulting after the discharge (RES_volume – RES_q_out*RES_timestep) minus the target volume (RES_volume_target). From this error is possible to find a discharge value (RES_q_out) that minimizes both the differences between the actual discharge and the target ones, and the total amount of water and the target one. As a matter of fact, the subroutine looks for many possible discharge values and then choses the ones that minimizes the error function (pot_volerr_min), finding the best compromise between the two factors.

After having calculated the flux exiting from the basin, the subroutine updates the volume of the basin. If the volume of water after the discharge would be negative, then the outgoing flux is limited to the actual available amount.

To conclude, the subroutine keeps track of the volume of the basin, avoiding that it overcomes its capacity, and keeping balance between discharge and volume objectives; it decides how much water needs to be released and considers the effects of meteorological conditions on the lake.

SALTO_SNOW simulates snow melting through the Degree-Day method, and it describes melting events, snow and rain-on-snow precipitations. The assumption is that the snow melting rate is related to air temperature, as long as it is above a critical threshold, generally similar to the melting point of ice (Braithwaite, 2011).

Inputs of the code are the total amount of precipitation considering both snow and rain, air temperature, and the snow water equivalent (before and after the melting process). Expected outputs are the portion of fallen precipitation as rain, the quantity of snow melting during a certain time period, and the snow water equivalent at the end of the phenomenon.

To determine the ratio of snow and rain, the air temperature is compared with some thresholds, snowmod_tmin and snowmod_tmax, and the precipitation is considered to be entirely snow if temperature is below the first one, and entirely rain when it is above the second one, whereas between these two values a partitioning is computed. Quantity of snow obtained is added to the already accumulated one. Later, the correct daily melting factor is chosen (standard one in case of only snow, multiplied by a factor in the eventuality of rain accelerating the process). Finally, the melted snow is calculated, and it is subtracted from the stored one.

SALTO_SOILLAYER simulates the runoff inside a soil layer, taking into account the soil humidity, and the effective evapotranspiration and percolation downward. After the declaration of the variables, the soil moisture value (soilmod_sm) is updated by adding the infiltration (soilmod_infiltration) caused by precipitation, snow melting and percolation from the upper layer.

Subsequently, lateral discharge (soilmod_sm_q) is calculated through a non-linear approach. All the water accumulated in the layer is added to this quantity (soilmod_lf_s), and part of it is released as direct or lateral discharge (soilmod_lf_q).

Lastly, the effective evapotranspiration (soilmod_aet) is computed according to soil moisture and potential evapotranspiration of the layer (soilmod_pet_layer), and the percolation downward is determined through a non-linear approach.

While doing these calculations, humidity is always kept upon a minimum threshold (soilmod_SM_MIN), for the correct functioning of the model.

SALTO_VADOSEZONE simulates water behaviour (percolation and discharge) in the vadose zone, that is that region of the soil between the Earth's surface and the regional groundwater table[4]. After the variables' declaration, the quantity of water present in the vadose zone (vzmod_s) is updated by adding water percolated from the soil (vzmod_inf).

Next is the calculation of the vadose zone discharge (vzmod_grw), that is the amount of water going downward to feed groundwater, through a non-linear model. A check that this value is lower than the amount available in the vadose zone (vzmod_s) is necessary.

[4] P.A. Holden, N. Fierer, VADOSE ZONE | Microbial Ecology, Editor(s): Daniel Hillel, Encyclopedia of Soils in the Environment, Elsevier, 2005, Pages 216-224.

Finally, the quantity of water stored (vzmod_s) is updated by subtracting to the old value the computed discharge (vzmod_gwr).

SALTO_GW simulates what happens underground, hence the storage and the movement towards the rivers or inside the groundwater.

First of all, the amount of water stored is updated by adding the percolation from the vadose zone (gwmod_gwr) and the flux from the underground upper cell (gwmod_gw_in). Then, the underground flux is computed through a non-linear approach dependent on the current groundwater store (gwmod_gw_s). The flux is limited by the maximum amount of water available underground.

The portion of flux going towards the river (frac) is computed through a non-linear function according to the current underground water store and the maximum store capacity. If the water cannot flow laterally or downward, but only towards the river, then “frac” is set equal to 1.

According to this parameter, the amount of water towards the river (gwmod_gw2r) and the one toward the next underground cell (gwmod_gw2gw) can be calculated, and the underground store is updated by subtracting these quantities to the old value.

SALTO_riverrouting simulates how water moves along the river, computing the final river discharge, taking into account all the sources and the water that is already present. It uses a non-linear reservoir model to represent the river behaviour.

First of all, the river tank (rivermod_river_s) is updated by adding the discharge from the current cell (rivermod_q_catch) and the one from upstream cells (rivermod_q_river_in).

Then, the flow of the river is computed by using a non-linear function that depends on stored water in the river basin (rivermod_river_s). Simulated flow (rivermod_qsim), that must be lower than water available in the riverbed (rivermod_river_s), is subtracted from the stored one for the updated value.

3.3 Model efficiency

Model efficiency is a fundamental indicator to understand if it is possible to expect good results from the model.

When the local calibration is performed, an efficiency is associated to each of the 25 parameter sets found for every catchment. The highest value in every case is then selected to populate a table, that contains also other geolocation information about the basins and the rivers.

A spatial plot of ME is showed in the following figure. Colours divide the total amount of values in quartiles. Dark blue is more present in central and North areas of Germany than in the South, and this could probably be linked to the different morphologies of the two regions that makes it easier for the model to have more reliable results for plains and low mountains with respect to the steep Alpine zones.

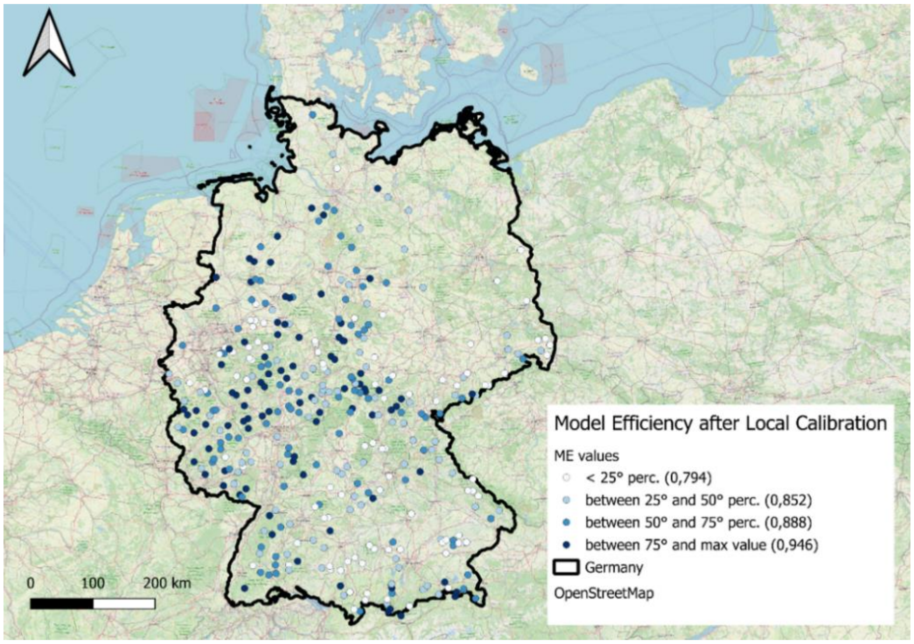

Figure 1: Spatial variation of the model efficiency (ME) after the local calibration.

To testify whether a relationship between model efficiencies and the extent of catchment areas exist, it is possible to examine Figure 4 (where the dotted line represents the mean value). In general, it can be noticed that for large catchments efficiencies are low. Smaller catchments on the other hand are mainly associated with higher numbers (big blue assemblage of dots in the top left corner), but a cloud of points with values below the average areas is quite noticeable as well. For an enhanced visualization of the colour's variation, it is better to consider the second graph, in which the few efficiencies below 0.5 are neglected.

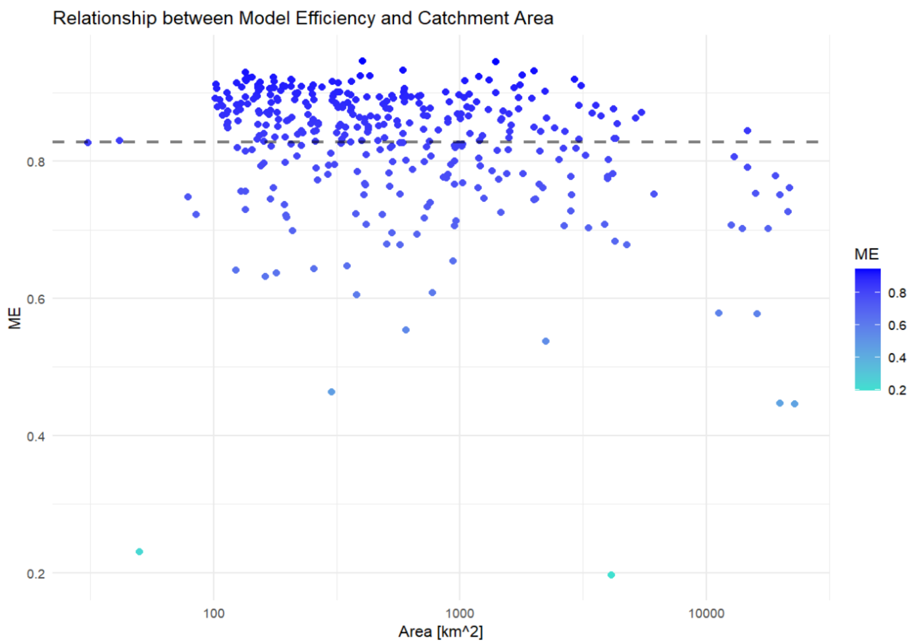

Figure 2: Relationship between catchment areas and model efficiency (ME), considering all cases

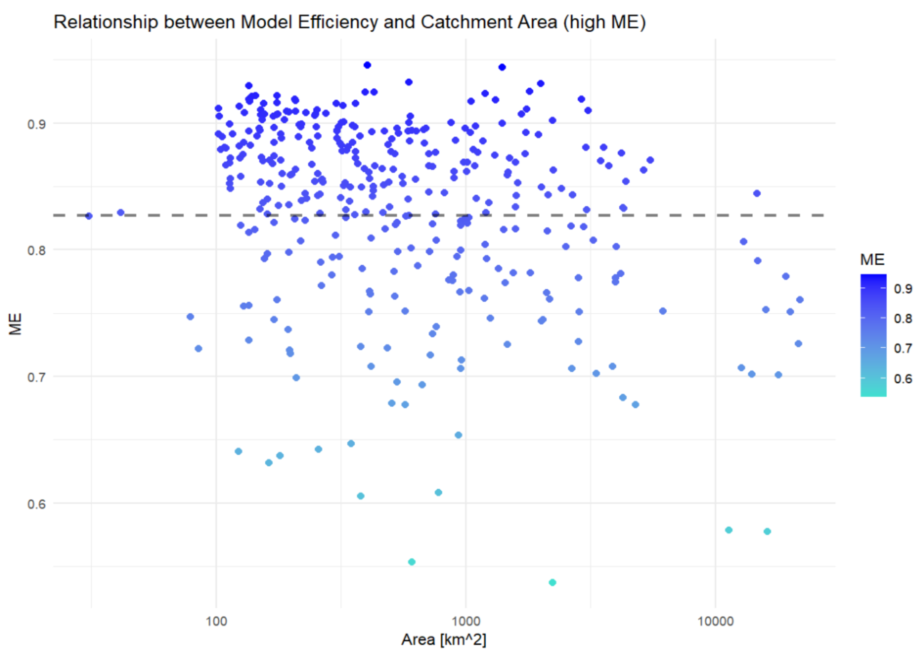

Figure 3: Relationship between catchment areas and model efficiency (ME), neglecting cases with efficiency lower than 0.5

Later, after PASS being performed, it is possible to analyse two new variables: MEreg, that is the efficiency of the model for every catchment after the regional calibration, and MEreg.wores, that is the same but in the case in which the effects of the dams are not considered. Every dot is coloured according to how much of the area of the catchment it represents is occupied by lakes formed by dams (Low less than 20%, Medium between 20% and 70%, High greater than 70%). As it can be seen from the scatterplot below (Figure 5), dots are mainly on the bisector (meaning that the efficiencies in the two cases are the same), and above it (hence the action of dams reduce the goodness of the model.

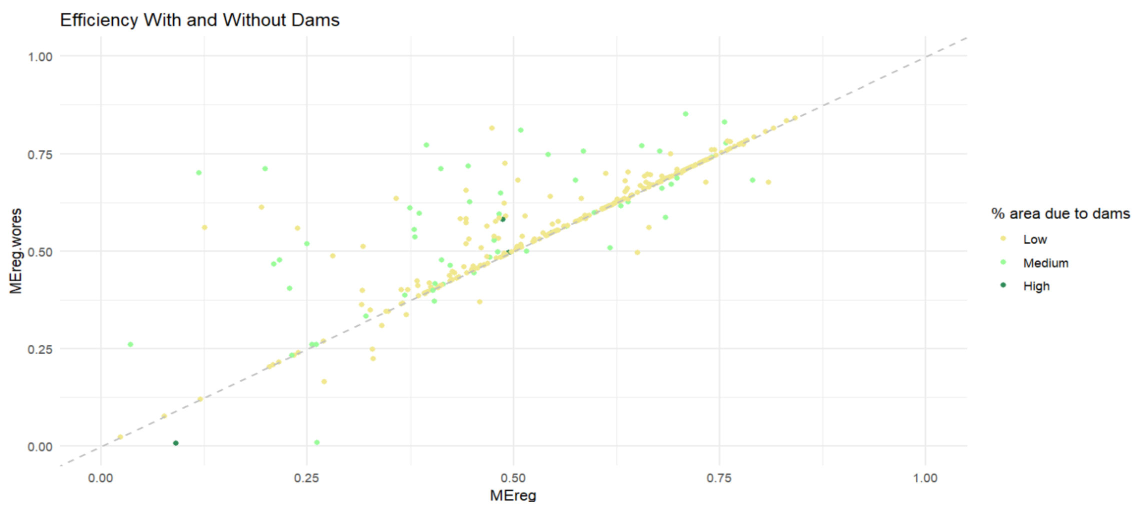

Figure 4: Scatterplot comparing the efficiency of the model for every dam after the regional calibration, in the case in which dams are considered vs dams are not considered.

The spatial distribution of the efficiencies after the regional calibration can be visualised in the following map. In this case darker dots seem to be more distributed on the whole territory, even if there is a higher concentration at middle latitudes. Furthermore, a comparison of these efficiencies with catchment areas leads to analogous results as before.

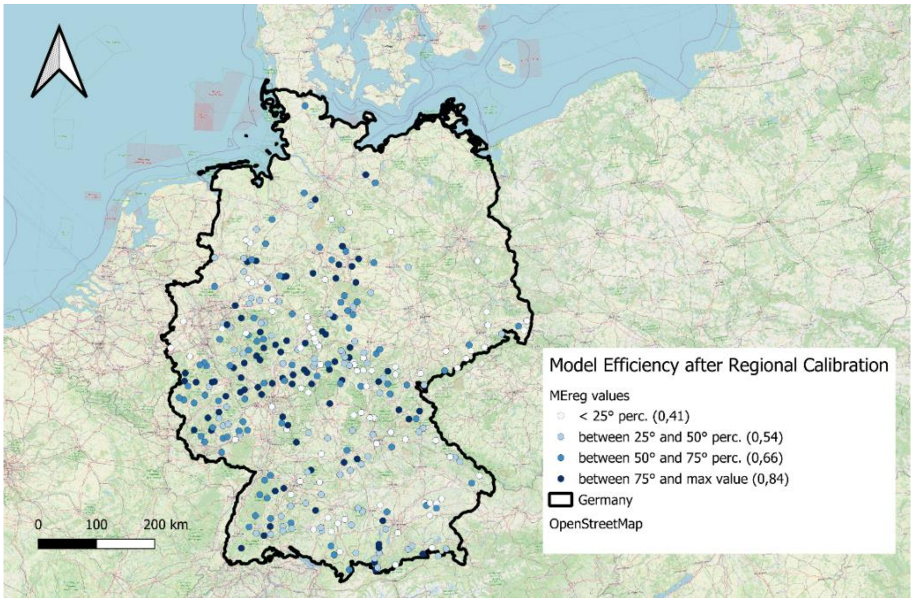

Figure 5: Spatial distribution of the model efficiency (MEreg) after regional calibration.

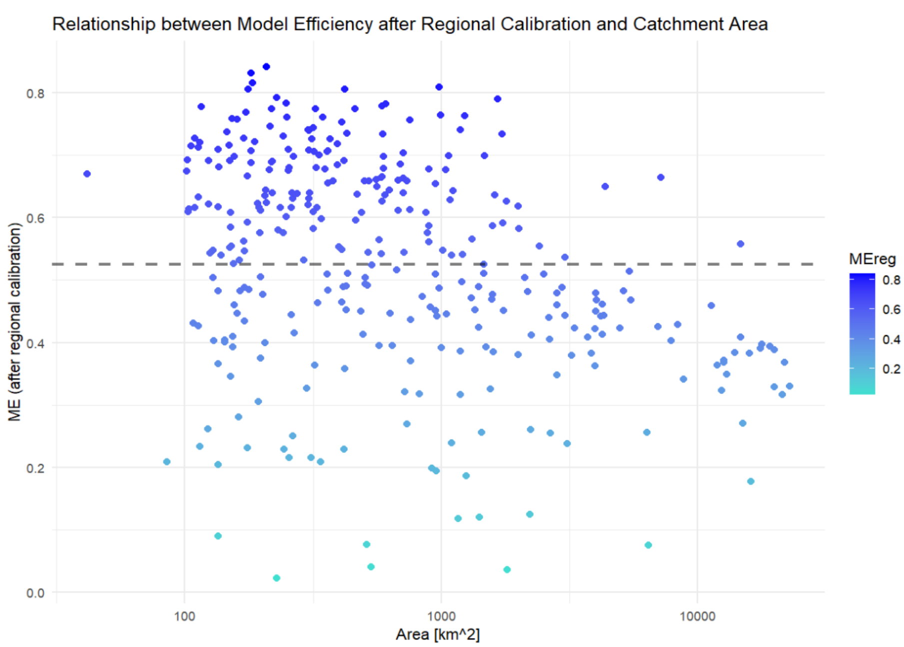

Figure 6: Relationship between catchment areas and model efficiency (MEreg)

It should also be noticed that among these efficiency values there were some peculiar cases that resulted to be negative. These instances were ignored in the plots.

For every catchment it is possible to compare ME and MEreg. As expected, values of the model efficiencies after the regional calibration are lower, because of the higher number of factors taken into account that led to an increasing complexity in the model. An analytical confirmation is given by Table 2.

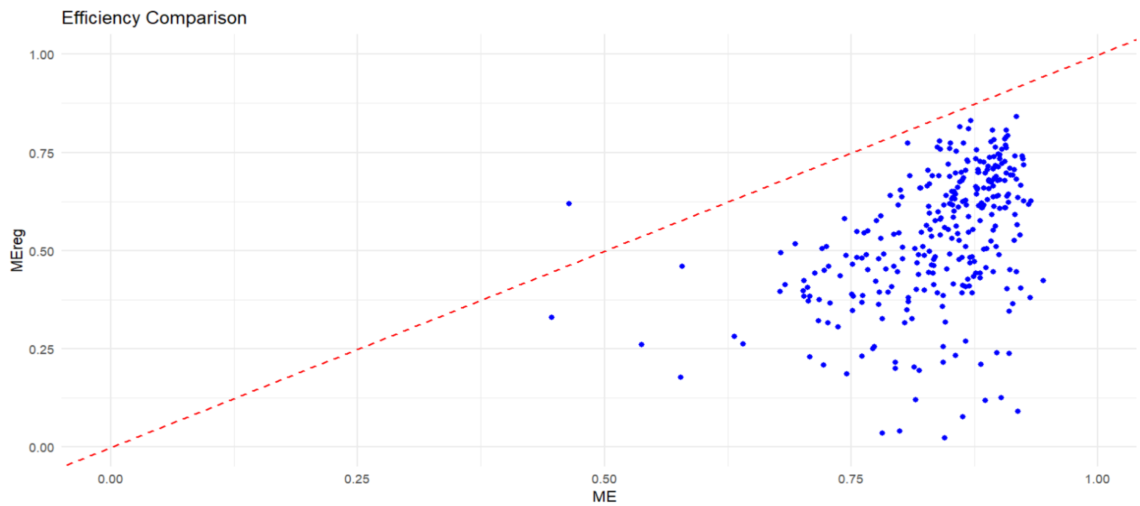

Figure 7:Comparison between model efficiency after local calibration (ME) and after regional calibration (MEreg) for every catchment.

Table 2: Statistics comparison between ME and MEreg.

	Statistic	ME	MEreg
1	Minimum	0.1971882	0.02329733
2	1st Quartile	0.7943287	0.40955030
3	Median	0.8522785	0.54138697
4	Mean	0.8273576	0.52556283
5	3rd Quartile	0.8881455	0.66426440
6	Maximum	0.9457848	0.84122940

The Cumulative Distribution Function (CDF) of ME and MEreg shows again that efficiencies after the local calibration are generally higher than efficiencies after the regional one. For example, half of the values assumed by ME are greater than around 0.8, whereas for MEreg this number goes down to less than 0.6.

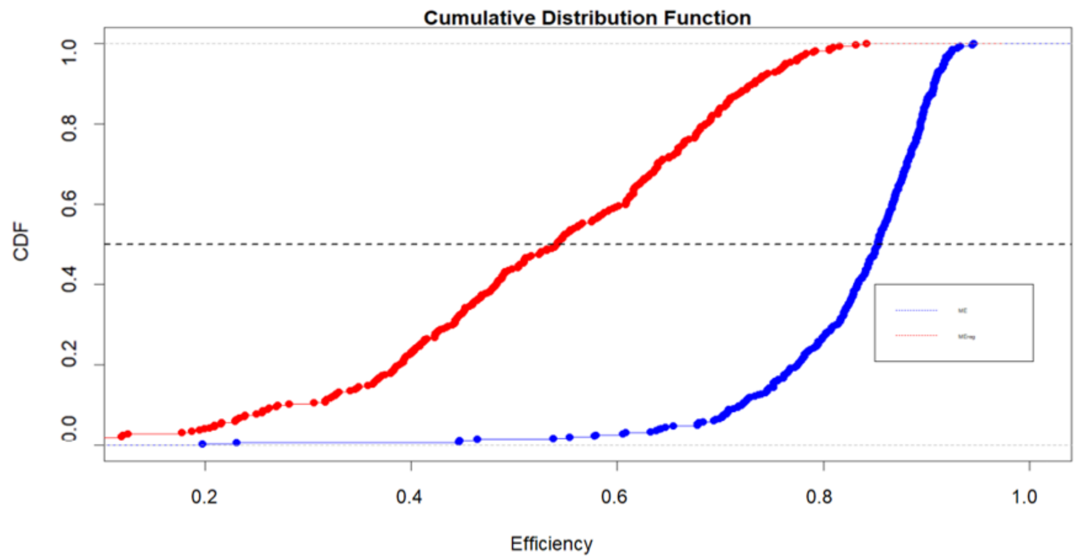

Figure 8: Cumulative Distribution Function (CDF) of the model efficiencies after local calibration (ME) and regional calibration (MEreg).

Before starting with the analysis, it is dutiful to say that some regions were not properly detected by the model, hence some approximations and imprecisions could arise. As instance, the North-East part of the country is nearly totally imperfect, because it is really flat and this leads to difficulties in reproducing the runoff: as a matter of fact, water flow here is likely not to follow a specific pattern, but it tends to spread laterally, and it extends onto the banks.

4 Results

4.1 QGIS

4.1.1 Inventory of dams

Among the inputs of the code, there are the drainage areas of the dams' basins; hence, the first step was calculating them one by one. The work of Speckhann, Kreibich and Merz (2020) was taken into account as guideline.

The authors developed an inventory filled with all the German dams, containing information on their name, date of construction (year in which building works began), the start of operation (year in which construction works ended), the German State in which they are situated, the name of the river where the dam is located or is close to, the dam height (in m, from base to top), the crest length (in m), the lake area at the full capacity (km^2), the lake volume at the full capacity (km^3), the purpose of the dam (which is its function), the type of the structure, the building characteristics, and the geographic coordinates in WGS84.

The "purpose of the dam" regards energy production, flood control, recreational use, water supply, industrial and agricultural water supply, fishing, transport or nature protection (Speckann, Kreibich and Merz, 2020).

The "dam type" could be dam, flood control, pump storage, water tank, upstream dam, forebay, polder, residual lake associated with mining, reservoir, storage basin, "Kulturwehr" (cultural, recreational), barrage or pond (Speckann, Kreibich and Merz, 2020).

The "building characteristics" refers to arch dam, embankment dam, homogeneous dam, gravity dam, flap weir, buttress dam, ring dam, segment with Fish belly flap, rockfill dam, residual lake associated with mining, rolling weir or zone dam (Speckann, Kreibich and Merz, 2020).

The authors stated that the collection of these data is the result of a meticulous work of research among federal agency reports, scientific publications, books, journals and web pages. Overall, the main source was the book "Talsperren in Deutschland" published in 2013 by the German Commission on Dams, that contains information on 340 structures. The second step was the localization of each one of the 530 barriers by using Google Earth. Once the name and the coordinates were known, the list was completed, when possible, with all the further information.

4.2 Drainage areas of dams

Building upon this inventory, drainage areas for the basins of the 530 dams were calculated in QGIS by using algorithms belonging to the SAGA tool: "Strahler Order", "Channel Network and Drainage Basins" and "Upslope Area".

The procedure began with the supply of two files:

- Digital Elevation Model (already filled), that is a raster file containing information of the "vertical datum" of the studied area with spatial resolution of 100m. The resolution being

pretty low contributed to several difficulties in the representation of water bodies, that were subject to massive approximations;
- Shape file, containing all the information of the dams that allowed to visualize them in the map.

Both these layers were reprojected in order to work with the EPSG:25832 coordinates, suitable for the German territory.

Then, the catchment area calculation started with the "Strahler Order" algorithm. The Strahler Order is a number used in hydrology as criterion to define the complexity of rivers by looking at the pecking order of tributaries. The QGIS algorithm requires the filled DEM as input, and provides a raster file showing the hierarchy of the hydrological net as output: values range from 1 to 10, where 1 corresponds to smaller size streams and 10 to the larger ones. These numbers need to be used as a threshold for what should be characterized as a river during the calibration process. This value is defined by the "Raster Calculation" tool, which provides a Boolean map assigning 1 to all the cells respecting the desired condition, and 0 to the others.

The second step is the performance of an algorithm called "Channel Network and Drainage Basins", used to delineate the stream. In input it needs the filled DEM, the calibrated threshold, and as outputs it will produce the flow direction, the drainage basins, and the channels.

The subsequent phase is the usage of the algorithm "Upslope Area". To function, it needs the UTM coordinates of the outlets (the dams' position), that can be find through the "Coordinate Capture" tool by placing the target on the "pixeled river", and the filled DEM. The "Deterministic 8" method is used, that is every raster cell is taken as the central starting point, and then the water flow is directed towards one of the eight adjacent directions.

The final stage consisted of converting the result, obtained as a raster layer, into a vector layer, to properly measure the area; this passage is performed by "Polygonise" algorithm.

As a first try, the Strahler Order was set equal to 5, and ultimately 226 catchment areas were calculated. It was impossible to define the others because, probably due to the low resolution of the DEM, three big problems were mainly faced:

- simulated rivers did not overlap with the real ones that were supposed to reproduce;
- water basins were wrongly represented by set of parallel lines;
- entire regions (for example nearly the whole North) were not covered at all by the model.

Hence, in a second time, a new threshold equal to 2 was set, and even though some of the previous issues were still encountered, the new representation was of course more detailed and allowed the calculation of 103 new values, for a total final of 329.

The remaining 201 areas were selected from a database obtained by processing in R the same DEM. Their reliability was further verified with some basic checks, such as looking for relationships of these catchment areas and dam's height or crest lengths, to look for linear dependences, or verify the compulsory condition that every catchment area must at least be wider than the respective lake area. Actually, this last condition was not respected for every dam, but values were considered as good anyway.

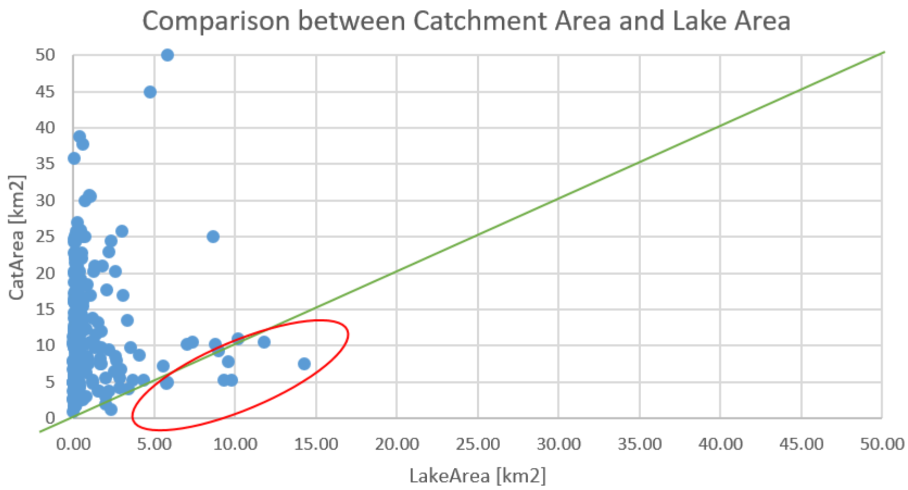

Figure 9: Comparison between the area of the catchment and the area of the lake for every dam. As is it highlighted by the red ellipse, in some cases there is the nonsense of the second being bigger than the first.

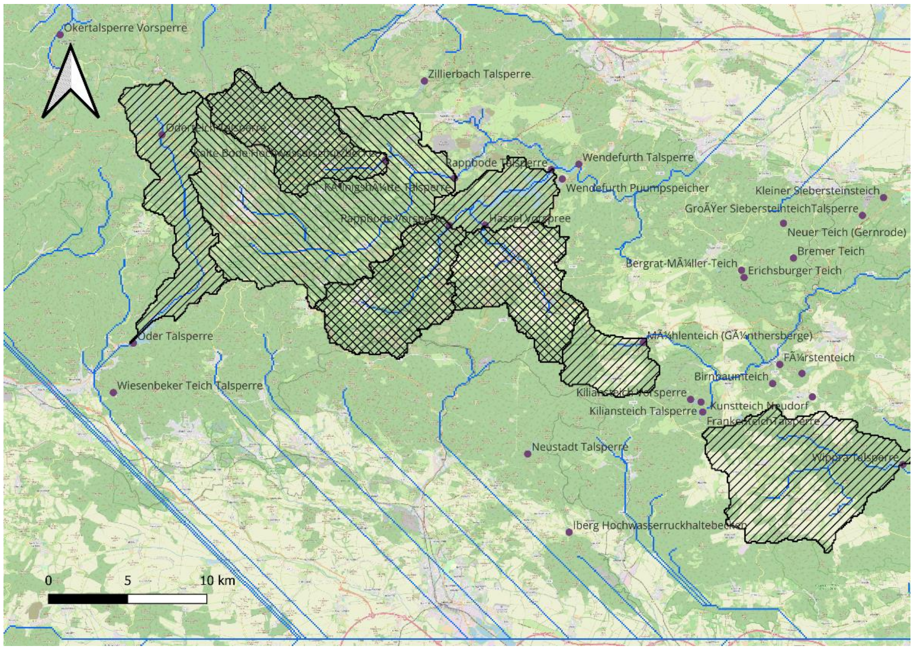

Figure 10: Calculation of catchment areas when the algorithm managed to reproduce the river. Parallel lines are an example of how often simulation could not be trusted. Purple dots represent dams.

4.2 Hydrograph and Flow Duration Curve

4.2.1 Sample catchments

To have an idea of how well the model responds to the presence of dams in the territory, it is useful to complete some conventional analysis, such as the evaluation of the evolution of the discharge and the interpretation of the flow duration curve.

Catchments chosen for the analysis are the one of Schwalm (gauging station at Utterschausen) in which there are dams, and the one of Efze (gauging station at Hebel) without them, and located upstream to the first one, in order to have a similar trend of the runoff in both cases.

To make this selection, it was necessary to look at a variable in the code that measures how much of a catchment area is actually related to the presence of dams: picking sites with high value means to look for catchments that could possibly be highly influenced by the barriers action. It should be noticed that, even though this number should represent a percentage, there were 21 different cases in which it exceeded 1: this might be due to a lack of precision in the geolocation of the dam that led to the consideration of the wrong territory. However, for the purpose of this thesis, these cases were ignored, and the ambiguities were not further explored into deeper details.

Chosen rivers are in the Hessen state. Schwalm is nearly 100 km long, it is part of the Weser fluvial system, and ultimately flows in the North Sea. Efze is less than 40 km long, and it is the most important affluent of Schwalm. Measurements are taken in gauging stations in the Schwalm-Eder district. Climatic conditions in these hilly lands are generally mild summers, and really cold, windy, and usually cloudy winters. During the year, temperature tipically ranges between some degrees below 0°C and 25°C, and it rarely goes below -10°C or overcomes 30°C.

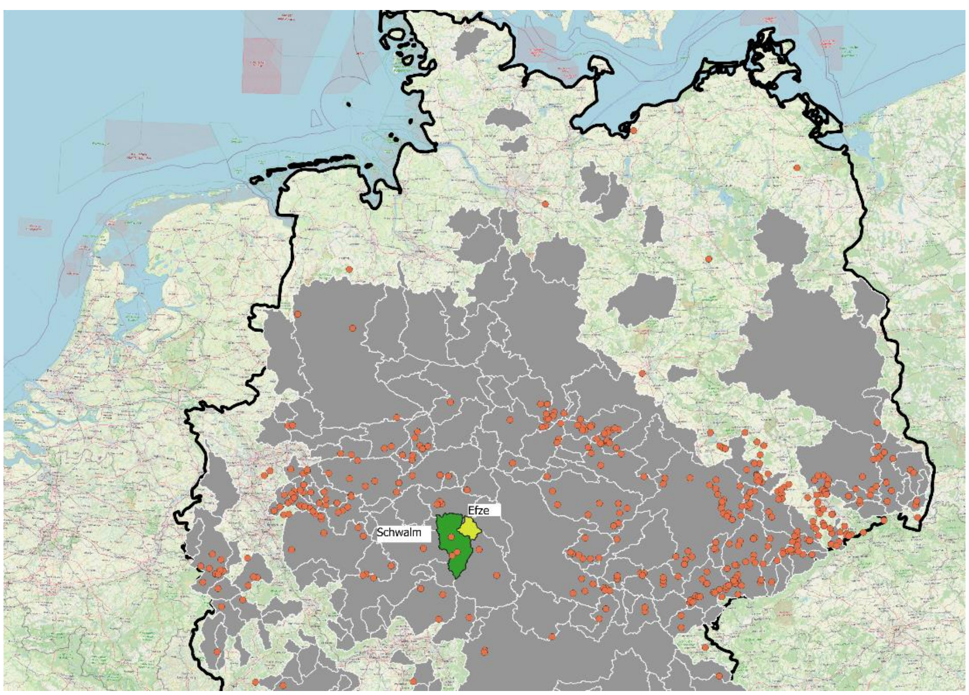

Figure 11: Position of Schwalm (dark green) and Efze (yellow) catchments. Orange dots represent dams. There are three dams inside the Schwalm catchment, and none of them inside the Efze catchment.

4.2.2 Evaluation of results

Hydrographs help make a comparison between the real runoff, calculated with data sampled at gauging stations, and the simulated one, computed by the model in the two cases of considering and ignoring dams.

The precision of the simulated runoff with respect to the measured one can show if the model works better with catchments that do not contain dams or if it is not excessively affected by their presence. What is expected is that the model cannot reproduce properly the discharge in periods in which dams are used to control the flow of the river, for example during drought periods.

Tipically, a hydrograph is characterized by the volume and the shape. The first information is determined by precipitation, generation of the runoff at the land surface, infiltration processes and evapotranspiration. The second one instead, is regulated by water movement in the river network and other flow paths on the land surface, in the unsaturated zone and in the groundwater. The main indication that hydrographs provide is the description of how the discharge develops chronologically (Blöshl, Sivapalan, Wagener, Viglione and Savenije, 2013).

Hereafter there are the timeseries for the two selected rivers, Schwalm and Efze.

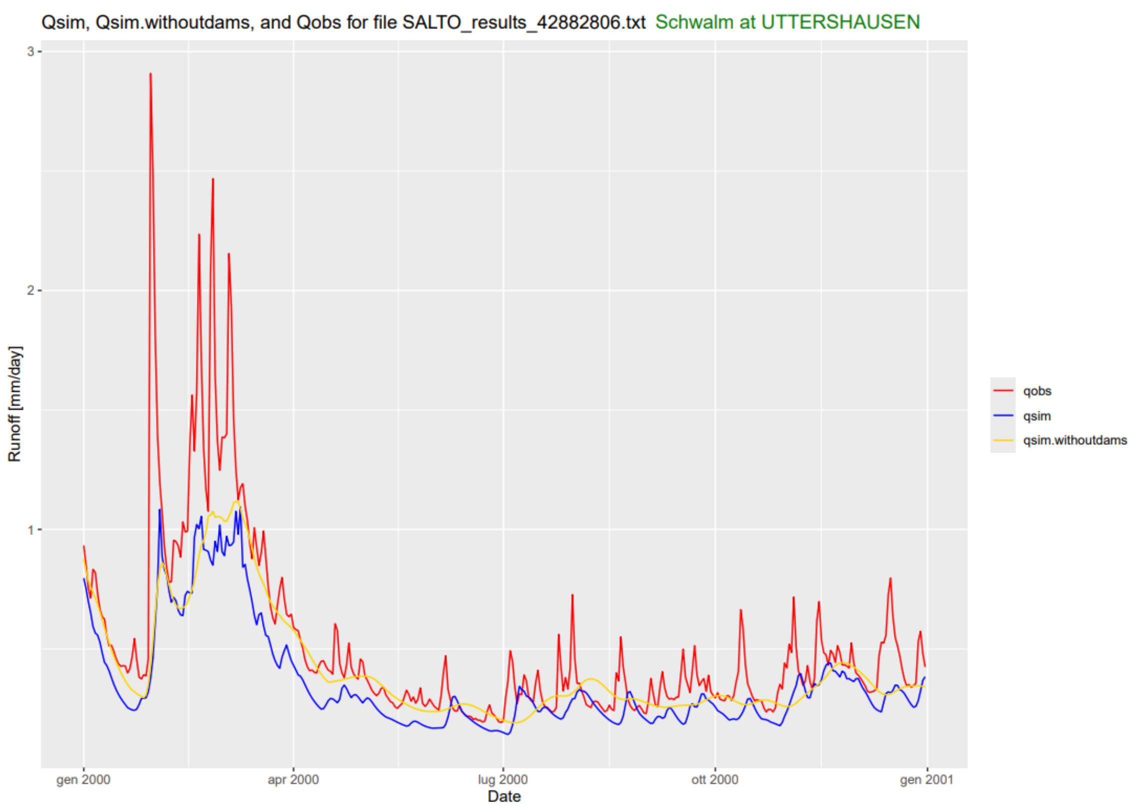

Figure 12: Hydrograph of Schwalm River, year 2000.

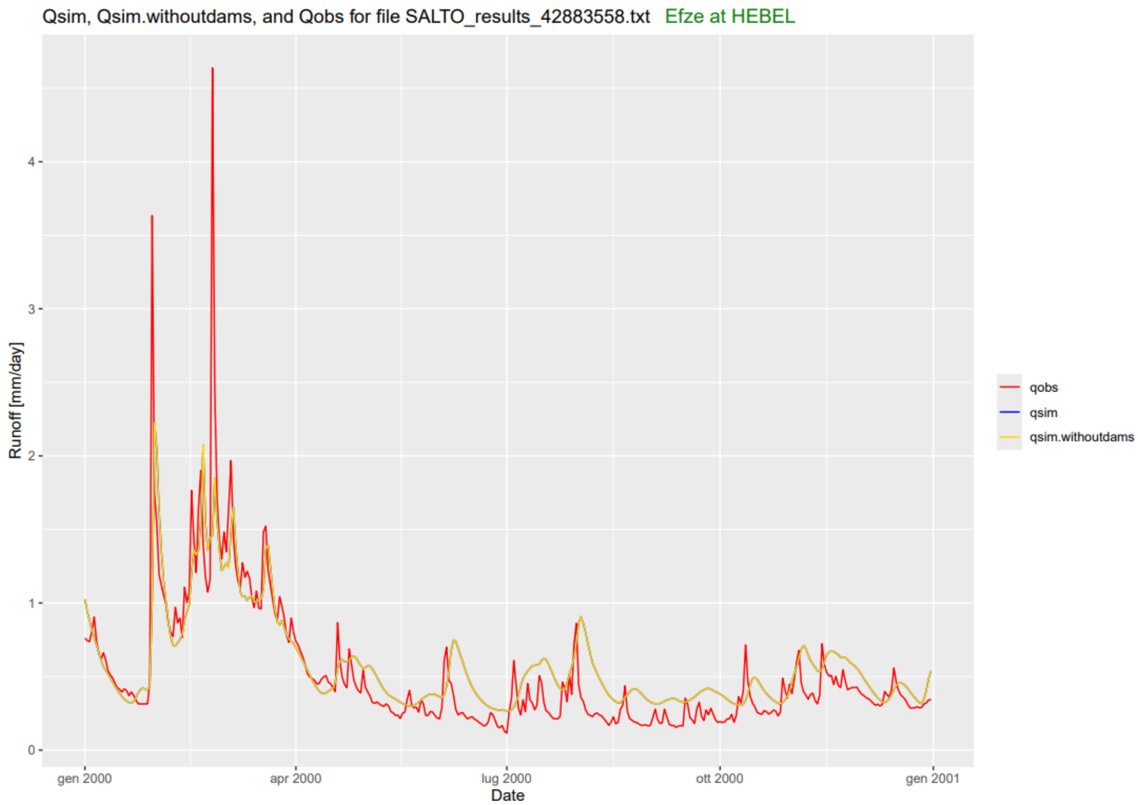

Figure 13: Hydrograph of Efze River, year 2000.

Observing red lines (observations) in the hydrographs, it can be noticed that the runoff dynamic for the two rivers is similar: higher values of the discharge are localized around February and March, then values are lower with small oscillations, probably caused by rainfall events. This situation is common in catchments mainly fed by snow melting at the beginning of spring.

By the comparison of the two cases, it is possible to notice that the second one assumes generally higher values, and this could make sense if the basin stores a part of water after precipitation events.

In both figures, it can be observed that the simulation has troubles in representing the high flow peaks. However, in the Schwalm case, the situation seems to be worst, and this could maybe be caused by the presence of multiple dams that make the algorithm more complex. No constant periods in the observations curve, possibly due to controlled release of water, are noticeable.

In Efze catchment, the two simulations coincide, because there are no dams in the territory. They are smoother with respect to the red curve, and this is a common thing in modelling because it is difficult to faithfully reproduce the strong dynamism of real phenomena.

To continue, Flow Duration Curves (FDC) could be considered as another indicator of the possible activities of dams, and their impact on the runoff.

By comparing the FDC of the catchment with the barriers and the one without, is it likely to expect that the runoff of the first one will assume a lower range of values with respect to the second one, due to the regulated flow at which it is subjected (holding water in the basin during floods, releasing it during low flows periods). Of course, dimensions of the rivers should be taken into account as well.

Generally speaking, that is neglecting dams for a while, the shape of the FDC is determined by many factors of different nature, such as climatic events, geological characteristics of catchments and human actions (Blöshl, Sivapalan, Wagener, Viglione and Savenije, 2013).

Yokoo and Sivapalan (2011) stated that a FDC is composed by three sections: the upper part (correspondent to high flows, for which the main control is the interaction of extreme rainfall events and fast runoff processes), the middle part (characterised by the mean runoff and its seasonality, for which the dominant force is the competition and seasonal interaction between available water, energy and storage), and the lower part (for which the leading regulation is the balance between deep drainage controlled by the geology and the evaporation of the riparian area) (Blöshl, Sivapalan, Wagener, Viglione and Savenije, 2013).

A steeper FDC is typical of catchments dominated by rapidly response near surface runoff dynamics, in opposition to catchments where discharge generation is managed primarily by slow processes, characterised by flat FDC. Of course, precipitation exerts the major climatic control on the runoff, but many other climatic forcing should be considered as well.

For instance, soil characteristics determine the partitioning of incoming precipitation into interception, infiltration and overland (fast) runoff, influencing the shape of the FDC. Thus, a less permeable soil, like clay, will correspond to a steeper FDC, whereas a more permeable one, like chalk, will result in a flatter curve.

The aridity of the climate, given by the annual potential evaporation over the precipitation, E_p/P, reflects the balance between energy and water availability, and it is another factor influencing the streamflow. This aspect is also linked to the spatial distribution of the vegetation cover and to its functioning, because it affects the amount and timing of evaporation, and hence of runoff.

A flat FDC could also results from a catchment connected with glaciers melting and snow processes, that contributes to discharge of the river especially during spring and summer. Moreover, the groundwater recharge could influence the runoff as well, contributing to make it more constant during low flow periods.

Coming back to Germany, the graphs of FDC for the selected catchments were plotted both in a linear and logarithmic way: the first one highlights better low flows, whereas the second is more suitable for high flows periods. However, in the following analysis the focus will be exclusively on the linear plots.

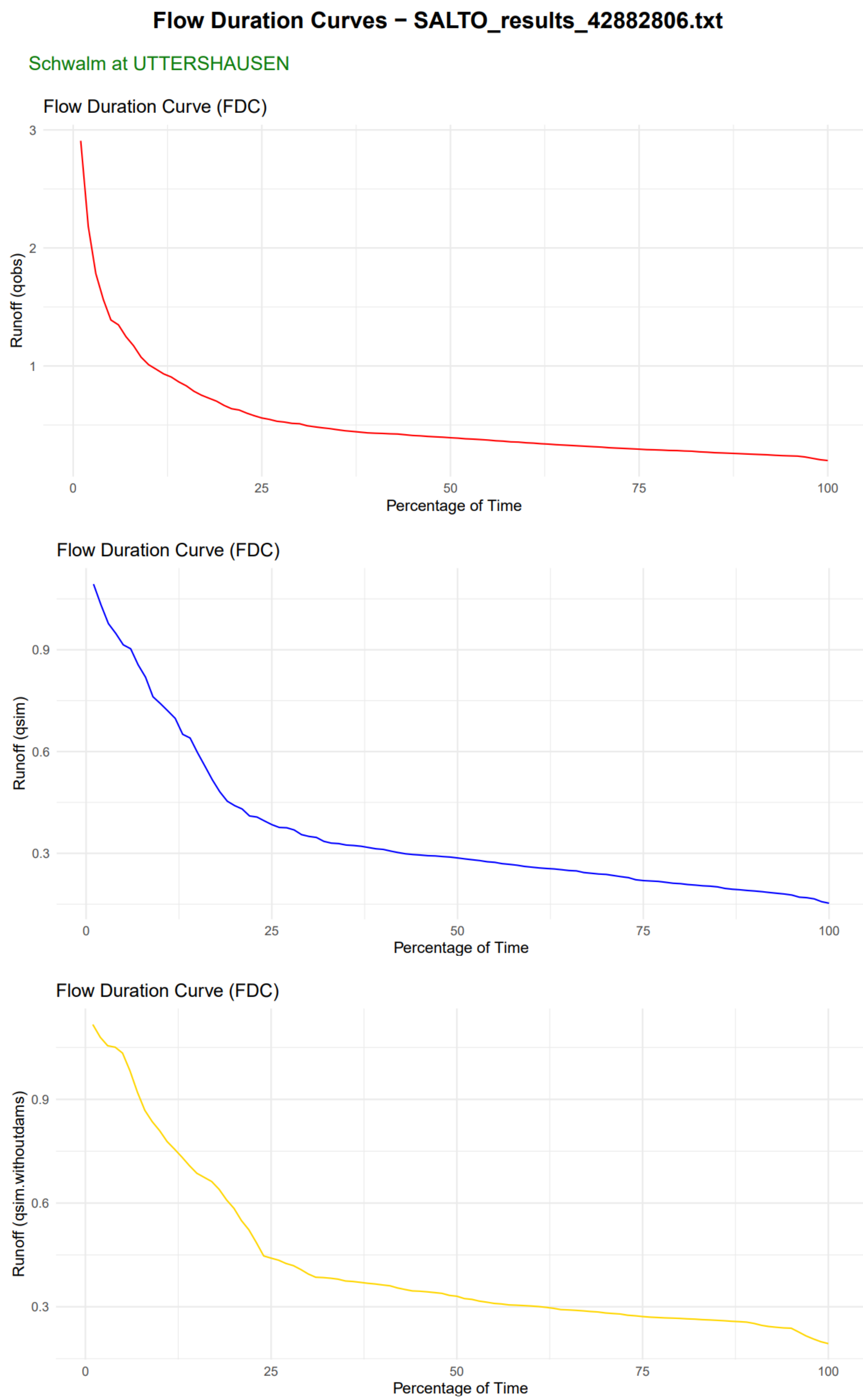

Figure 14: FDC of Schwalm River, year 2000.

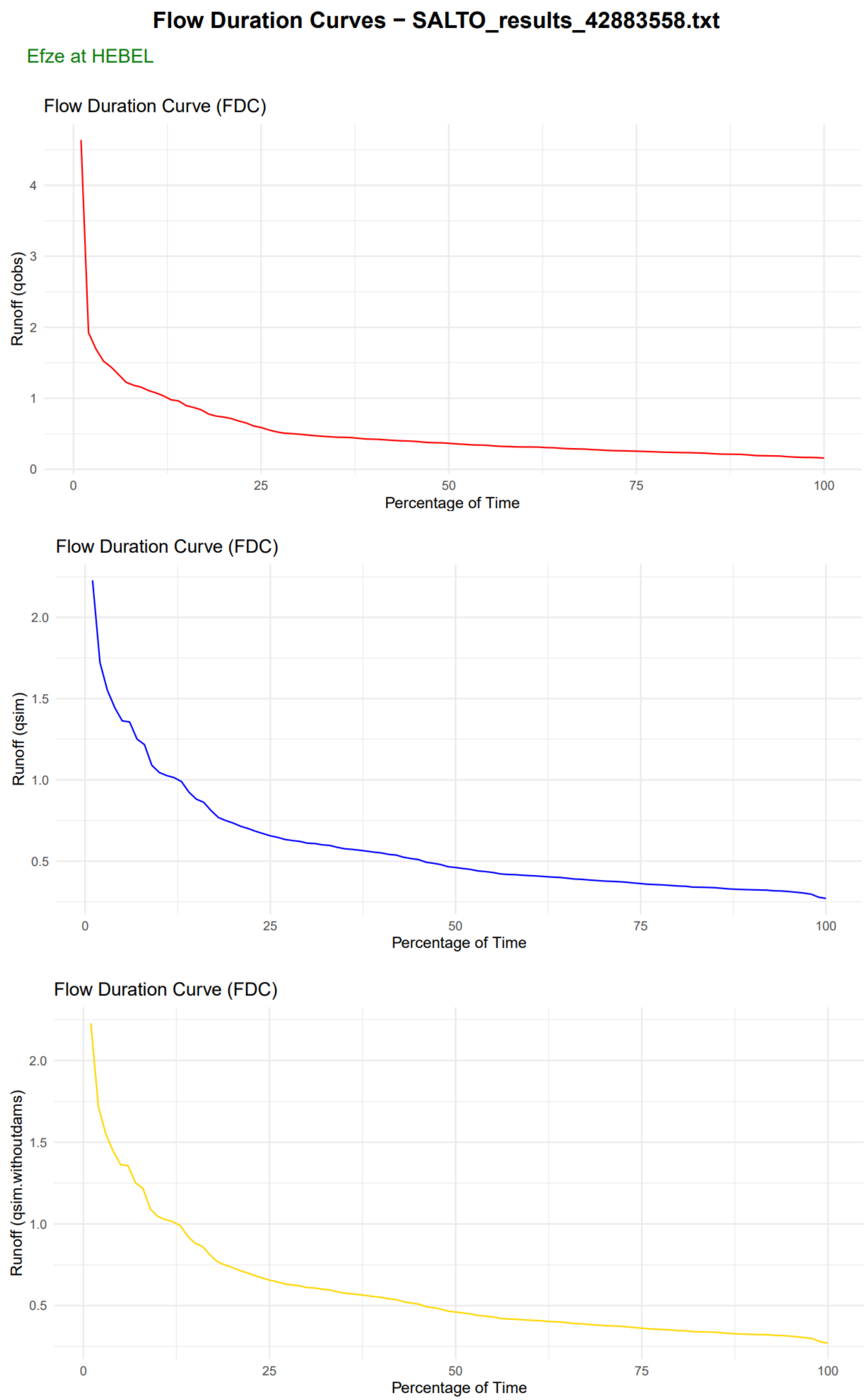

Figure 15: FDC of Efze River, year 2000.

The range of runoff values assumed by Schwalm is not too big, going from a maximum of approximately 3 mm/day to nearly no flow. The first part is not too steep, even if later, for three quarters of the time, the runoff is below 0,5 mm/day, decreasing nearly linearly. This could make think about a catchment in which discharge generation is managed primarily by snow phenomena. The simulation instead, ranges between 1mm/day and no water, meaning that it was not able to reproduce high flows. It also presents a slightly stepped structure. Comparing it with no-dam simulation, to verify the previous hypothesis, the opposite result is found (blue line assume a smaller range of value than the yellow one). As instance, in the dam case, null flow seems to be reached, and this is odd because one of the main purposes of reservoirs is to avoid this condition, to preserve the environmental flow and for human needing. However, this does not seem to be the case in the observation curve.

FDC of Efze is steeper at the beginning, and this could be and indicator of the fast response of the catchment, justified by a bigger altitude and steeper orographic characteristics of the locality. Then, for nearly three quarters of the time, the flow is below 0,5 mm/day, meaning that during the rest of the year the quantity of water is moderate.

Overall, looking also at the hydrographs and FDC of all the catchments, results are not so accurate, suggesting that perhaps the model has some imprecisions that do not allow to fully trust on its outcome.

4.3 Flood quantiles

4.3.1 Aim of the analysis

Flood quantiles are values of the runoff that are exceeded with a certain probability during a flood event. This probability of exceedance is given by 1/T, where T is the return period. In other words, the return period is the time that *on average* occurs between successive exceedances of a given threshold.

Analysing flood quantiles calculated from observed data, from simulated data, and from simulated data obtained in the situation in which the effect of the dams was neglected, it was possible to draw some conclusions on how they change, in relationship to different return periods and distinct geographical characteristics of Germany.

More in detail, two different kinds of evaluations were carried out, that are the calculation of relative differences between flood quantiles determined by observed data and the ones estimated by the model (considering the action of the dam), and the relative differences between flood quantiles simulations that omit the dam effects and the ones that take them into account. These two analyses are repeated for return periods T equal to 2, 5, 10, 20, 50, 100 and 500 years, to inspect the consequences of more and less frequent flood events.

4.3.2 Simulations vs observations

The following plots represent the difference between flood quantiles computed from simulations and observations, divided by observations ones. Comparing these two values is useful to have an idea of the efficiency of the model. Thus, when the ratio is equal or close to 0, it means that the result of the simulated value is really similar to the one obtained from the measurement, and hence the model worked properly. On the other hand, when numbers are high in absolute value, it is likely that the runoff calculated by the simulation was very different from the real one.

For an easier interpretation, the obtained results are divided into three different colours: the light blue stands for values smaller than -1 or higher than 1, the standard blue for -1 to -0.25 and for 0.25 to 1 intervals, and the dark blue for the central range.

In general, the majority of values span between -1 and -0.25, and 0.25 and 1, meaning that there is, respectively, a significant underestimation or overestimation of the model with regard to observed values, but however this does not represent a catastrophic error in the calibration. However, values reaching order of magnitude up to 10^4.

Catchments are as well divided into three categories, with a pale green representing the ones whose territory is occupied up to 20% by dams watersheds, a standard green for percentages ranging from 20% to 70%, and a dark green for basins that are highly influenced by barriers (from 70% to 100%). Only 5 items belong to this last category, whereas for around one quarter out of the total catchments considered, null results were computed.

Hereafter there are the maps for return periods of 2 and 50 years; all the others are in the Appendix.

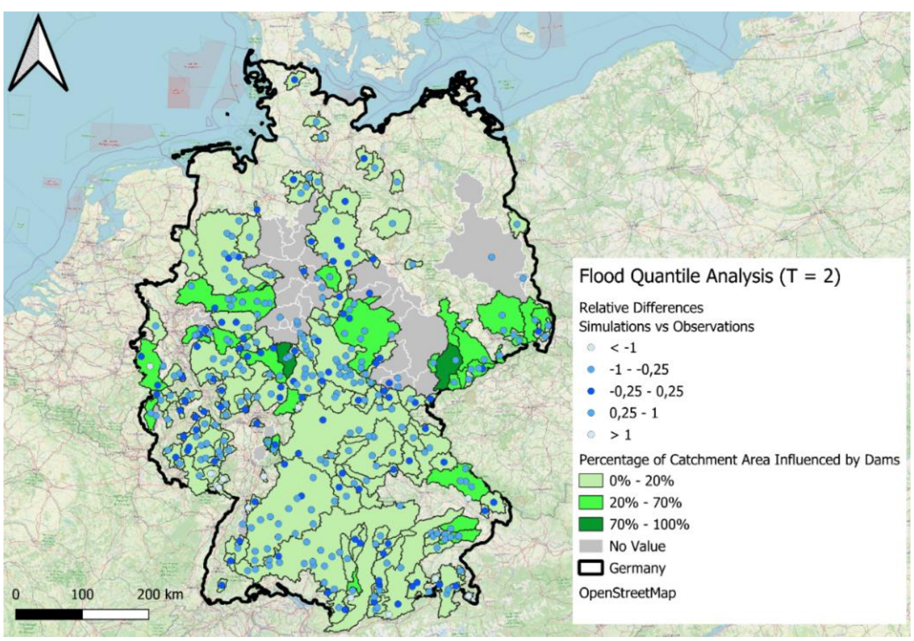

Figure 16: Relative differences of simulated and observed values for T = 2 years.

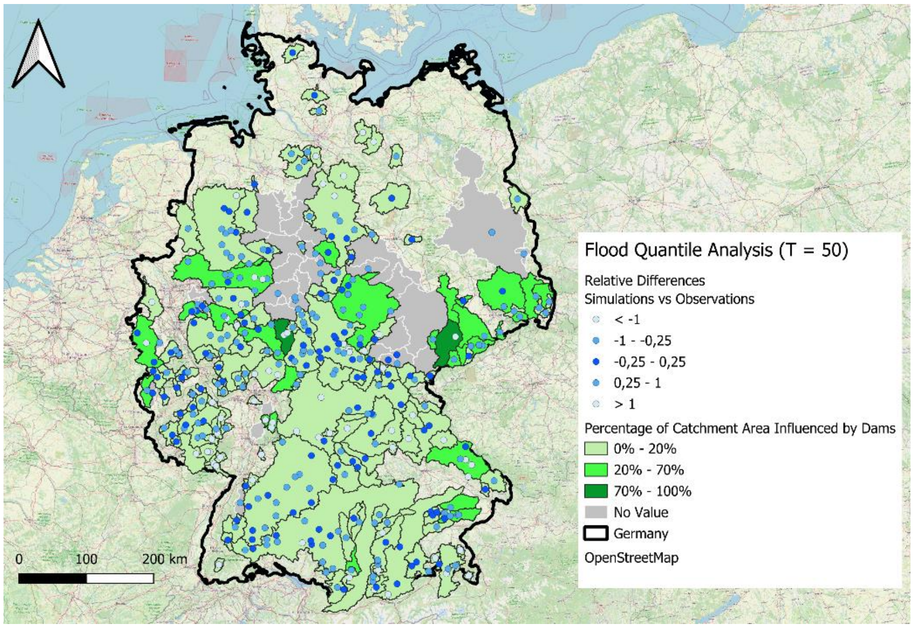

Figure 17: Relative differences of simulated and observed values for T = 50 years.

Comparing the maps obtained for all the different return periods, it seems that for central T, hence for medium frequent events, the flood quantiles computed with simulated data of runoff are more similar to the ones calculated with observations with respect to lower T. An explanation could be that the model responds better to high flows, because of the lower influence of daily and seasonal variation of the discharge, and also because of the littler subjection they have to meteorological phenomena.

Furthermore, it seems that, independently of the return period considered, less precise results are obtained for larger catchments, whereas many dark blue dots are associated to smaller ones. The explanation of this statement could depend on the relative more easiness with which the latter are represented in the model, that lead to greater accuracy. As a matter of fact, here climatic and geographic characteristics are more homogeneous, and just the simpler hydrological processes occur, whereas in bigger catchments also more complex phenomena, such as infiltration or sedimentation, play a key role and need to be taken in consideration, making it more difficult to be simulated. Another advantage of small catchments is the fast hydrological response to meteorological events. In addition, in large ones, it can happen that observations are aggregated on wide areas, leading to significant approximations in the simulations. This situation reflects what Figure 2 and Figure 6 showed earlier when comparing efficiencies and areas.

From a spatial point of view, it seems that more precise results are generally homogeneously located, even if, especially for lower return periods, the Southern area has a lower concentration of dark blue dots. This could be imputed to the morphology of that zone, characterised by high mountains difficult to reproduce in simulations due to more complex hydrological phenomenon happening there, and due to the fact that often is difficult to collect data in such inaccessible and steep places. However, for central return periods is it more difficult to visualize this trend.

Plotted maps are useful to visualize the spatial variation of the results, but it is difficult to say exactly which are the mean trends for the different return periods. For this purpose, it is possible to consult statistical indexes such as mean value, median, variance and standard deviation.

Table 3: Statistical indexes of the relative differences between "simulated" and "observed" flood quantiles.

	Mean	Median	Variance	StdDev
HQ2	−0.22697192	−0.34491691	4.178455e−01	0.6464097
HQ5	−0.13348741	−0.29932996	5.269238e−01	0.7258952
HQ10	0.01183959	−0.22371887	9.110674e−01	0.9544985
HQ20	0.33420402	−0.14975858	6.601026e+00	2.5692462
HQ50	1.88108649	−0.01798744	3.110586e+02	17.6368542
HQ100	6.94937466	0.09285991	6.507770e+03	80.6707487
HQ500	196.49585764	0.36211916	8.267738e+06	2875.3674899

By looking at the median, it is fair to conclude that in the majority of cases there are underestimations of the runoff for return periods of 2, 5, 10, 20 and 50 years, whereas a slight overestimation is linked to events of higher magnitude.

Regarding the mean instead, on average the model slightly underestimates the runoff computed with observed data for small return periods (T = 2, T = 5). By increasing the magnitude of the events, there is instead a slight overestimation.

Results for high return periods (T = 100 years and T = 500 years) are strongly influenced by wrong results and errors in the simulation. This can also be seen in the next tables and in the boxplot, in which it is clear that up to the third quartile numbers are still meaningful, hence there are some high results that heavily impact the mean.

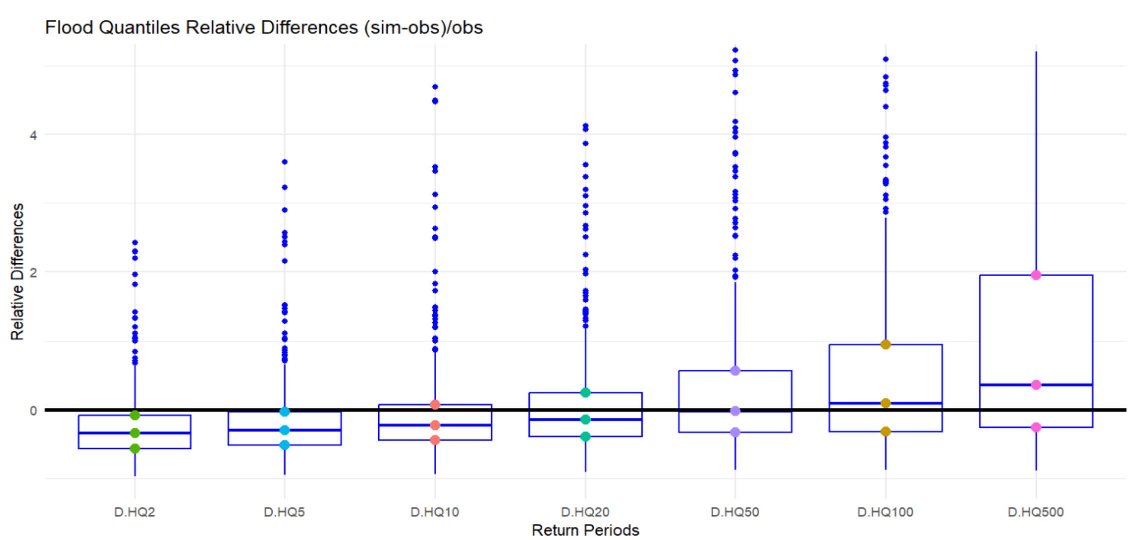

Figure 18: Box plots of the relative differences between "observations" and "simulations" for different return periods.

Table 4: Subdivision of the relative differences in quartiles.

	P25	P50	P75
D.HQ2	−0.5703817	−0.34491691	−0.08003154
D.HQ5	−0.5151747	−0.29932996	−0.03366806
D.HQ10	−0.4459188	−0.22371887	0.06826845
D.HQ20	−0.3932512	−0.14975858	0.24306727
D.HQ50	−0.3323656	−0.01798744	0.56332104
D.HQ100	−0.3219703	0.09285991	0.93976984
D.HQ500	−0.2604899	0.36211916	1.95547952

Table 5: Number of elements belonging to each class for the different return periods.

	RelDiff < -1	-1 < RelDiff < -0.25	-0.25 < RelDiff < 0.25	0.25 < RelDiff < 1	RelDiff > 1
D.HQ2	0	226	100	24	15
D.HQ5	0	198	115	34	18
D.HQ10	0	171	126	40	28
D.HQ20	0	147	130	46	42
D.HQ50	0	122	118	58	67
D.HQ100	0	107	108	62	88
D.HQ500	0	93	75	63	134

4.3.3 Neglecting vs considering dams

The comparison between the flood quantiles values produced with simulated value of the runoff including the action of the dam and ignoring their presence is expressed through the ratio between the difference of these two quantities at the numerator, over the second one at the denominator. This fraction establishes how much larger would the flow of a river be in the hypothesis in which no dam was built with respect to the real situation: it is an indicator to quantify the magnitude of their response.

Hence, only positive numbers are expected: reduced effects of the dams are associated to values lower than 1, whereas they result to be more and more significant when the ratio increases. In other words, the fraction being 2 means that the runoff of a river without dams is twice as the runoff of the same river where a dam works. When the result is 0 instead, no differences in the two cases are detected, meaning that the dam is not producing any effect (or the catchment does not have dams in its territory). Negative values are not awaited, if not as model imprecisions.

Theoretically, the forecasts are that the effects of dams should be stronger for middle return periods (usually 100 years)[5]: this corresponds to runoff that are not massive, intended to be contained in basins and later released when the discharge is smaller. Events corresponding to higher return

[5] JICA (Japan International Cooperation Agency) reports

periods are instead extreme, and maybe the dam is not big enough to store all that amount of water, so the results could be less visible. When designing the dam, a meticulous work of evaluation of the hydrologic risk is carried out to provide detailed information on the final dimensions of the structure and its relationships with return periods (Gebregiorgis and Hossain, 2012).

Also, it could be fair to think that the result changes according to the topography of the nation: mountainous areas are commonly characterised by smaller basins, where the action of the dam is likely to influence more the typical fast runoffs with respect to big lowlands and their wide watersheds, where the slow flowing water has more opportunities to disperse.

The last consideration is that probably effects are more visible in regions where there are many interconnected dams with respect to places where just one structure is present: their action can assemble, resulting in more powerful consequences.

The following plots illustrate which is the situation deriving from the simulation for every return period. Despite of the presuppositions, some negative results were obtained (in the maps in this Chapter they were ignored). In pink on the other hand, there are all the catchments for which the result of the ratio is equal to 0 (“Null”), in red the ones with positive values lower than 1 (“Light”) and in burgundy positive values higher than 1 (“Strong”). Actually, there were around 30 catchments that provided results of the order of magnitude up to 10^5, that were not considered while plotting the maps as well. More on these catchments can be read in Chapter 5.

Hereafter there are maps corresponding to T = 2 years and T = 50 years. The other ones can be found in the Appendix.

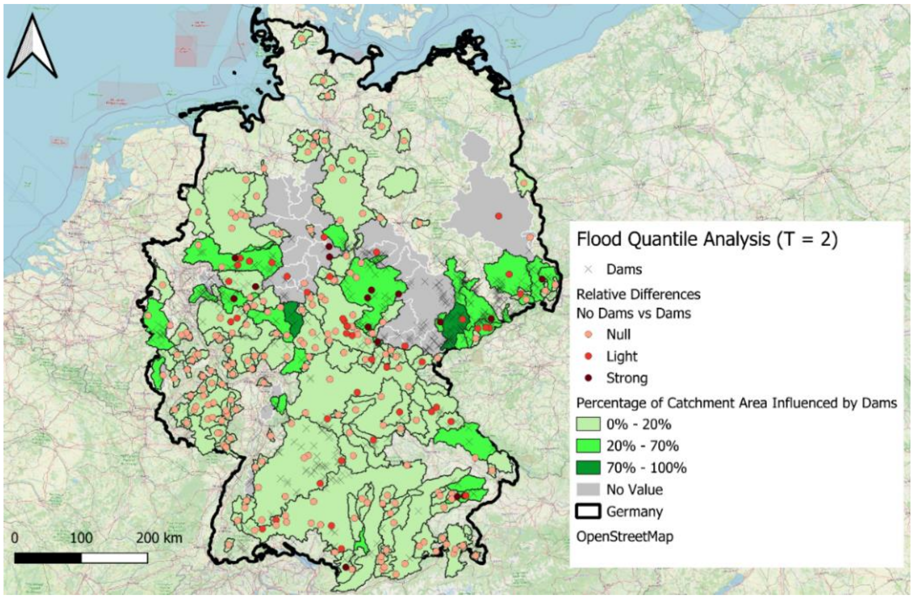

Figure 19: Relative differences of neglecting and considering the action of dams for T = 2 years.

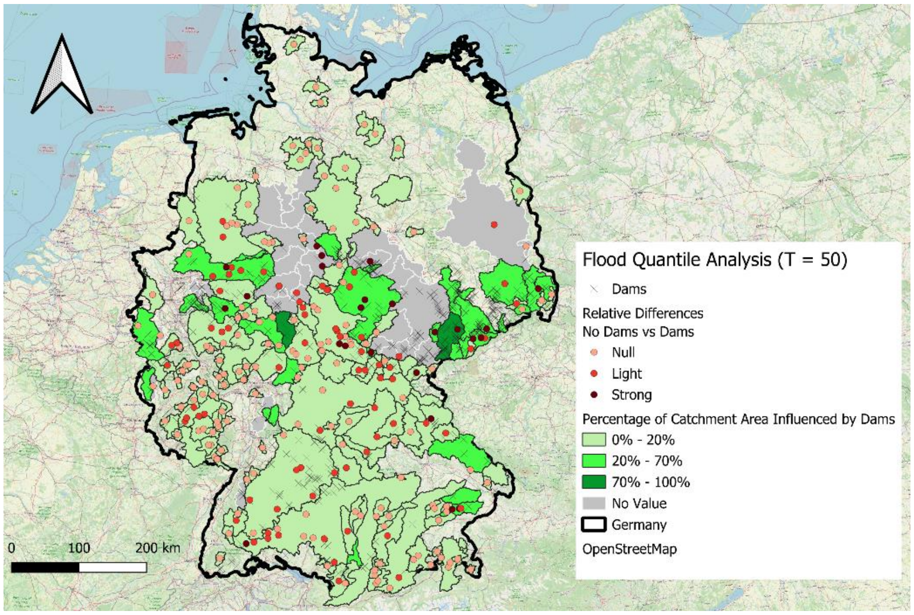

Figure 20: Relative differences of neglecting and considering the action of dams for T = 50 years.

By visually analysing the maps, it can be noticed that effectively, places that presents a high density of dams are subjected to stronger effects. Moreover, dark dots are mostly concentrated in the central part of Germany for lower return periods, whereas they are quite homogeneously distributed, covering also the South Alpine area, for higher T's.

To have a quantitative confirmation of visual impressions, the following tables and plots can be consulted.

Table 6: Statistical indexes of the relative differences between ignoring and considering dams "simulated" flood quantiles.

	Mean	Median	Variance	StdDev
HQ2	0.2177004	0	0.4956639	0.7040340
HQ5	0.2277778	0	0.4583957	0.6770493
HQ10	0.2362964	0	0.4301764	0.6558784
HQ20	0.2473291	0	0.4205398	0.6484904
HQ50	0.2675938	0	0.4547255	0.6743334
HQ100	0.2885690	0	0.5341228	0.7308370
HQ500	0.3638140	0	1.1129934	1.0549850

From the Table 6 it can be noticed that if dams were not present, an average increment of around 20% to 30% in the runoff would occur, depending on the return period. Higher boosts are for bigger return periods that correspond to more rare events.

Results meet the expectations that for central return periods, effects of dams are more visible: this is because smaller ones stand for events that are not so considerable, hence it is possible that, especially big dams, do not even feel the impact of these reduced magnitudes. Regarding the return period of 500 years, the effect results to be even bigger (average of 36%), because they are influenced by many "Strong" cases that, although being less than 10, bias the result.

Table 7: Number of elements belonging to each class for the different return periods.

	Reversed	Null	Light	Strong
D.HQ2.res	0	161	84	16
D.HQ5.res	0	161	84	16
D.HQ10.res	0	161	80	20
D.HQ20.res	0	161	77	23
D.HQ50.res	0	161	76	24
D.HQ100.res	0	161	73	27
D.HQ500.res	0	161	71	29

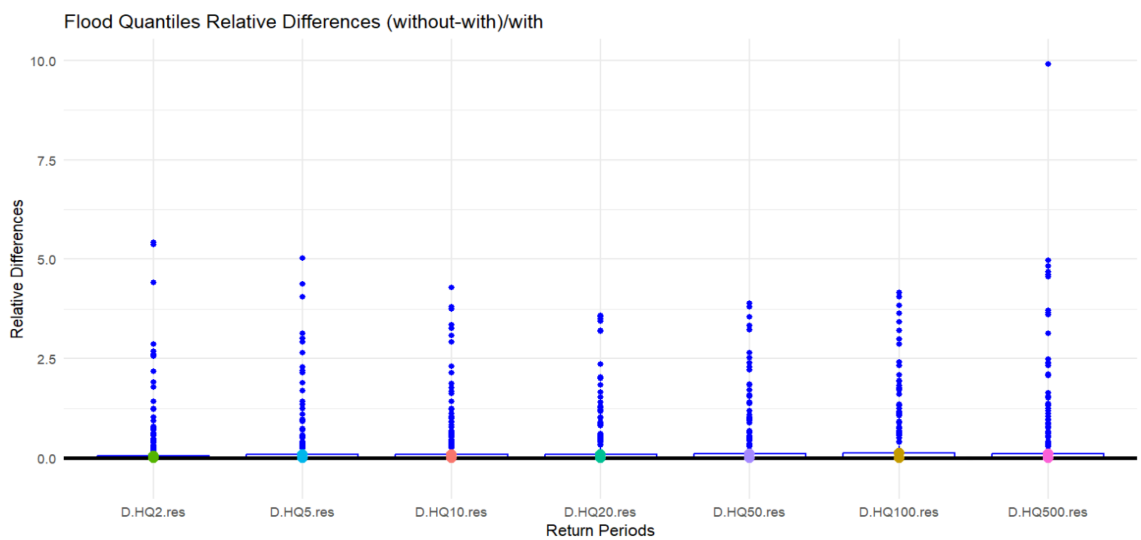

Figure 21: Box plots of the relative differences between neglecting and considering dams "simulations" for different return periods.

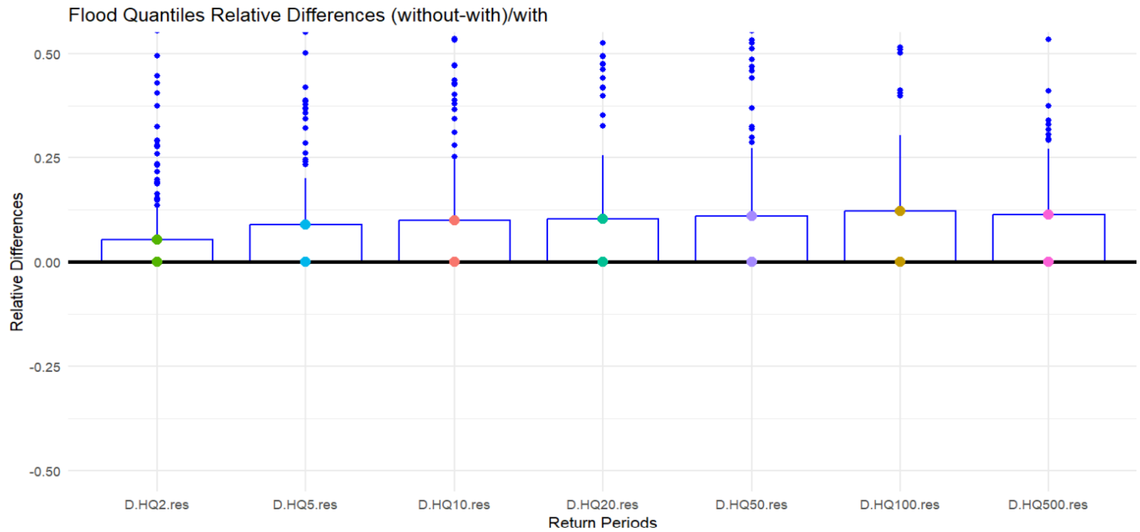

Figure 22: Box plots of the relative differences between neglecting and considering dam "simulations" for different return periods, values between 0 and 0,5.

Table 8: Subdivision of the relative differences in quartiles.

	P25	P50	P75
D.HQ2.res	0	0	0.05369434
D.HQ5.res	0	0	0.08951195
D.HQ10.res	0	0	0.10055169
D.HQ20.res	0	0	0.10314337
D.HQ50.res	0	0	0.10975336
D.HQ100.res	0	0	0.12255066
D.HQ500.res	0	0	0.11434776

5 Discussion

5.1 Interpretation of results

Results of the model highlighted a reduction of the flood quantiles due to the action of dams dependent on the return period, that on average ranges from 20% to 30%. To be more specific, it is possible to make some considerations on decrement situations, that is catchments with relative differences between flood quantiles computed from simulations of runoff neglecting and considering dams higher than 0. As instance, it could be interesting to investigate possible trends regarding the percentage of decrease and catchment area, basin area or basin volume.

To compare catchment areas, the mean surface values for every combination of increment and return period were computed, with results displayed in Table 9. Moreover, a three-dimensional plot that shows how increment percentages are related to the mean areas of catchments belonging to that category for every return period is plotted.

Table 9: Average of catchment areas (in Km^2) of catchments producing a certain difference in flood quantiles, according to distinct return periods.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	3098.014	3072.5782	3336.576	3146.230	3224.351	3809.2024	3942.773
0.2-0.4	4995.433	5008.3714	4974.000	6656.182	5865.067	2883.0867	1935.969
0.4-0.6	1487.087	1511.9350	1645.801	1368.678	1561.689	3689.3750	2637.333
0.6-0.8	2103.594	938.1117	1140.668	776.000	1915.362	1544.6683	3496.857
0.8-1	901.000	2850.1000	1327.333	2558.918	678.734	677.3333	2509.503
> 1	3405.358	3405.3575	3184.001	2877.131	3097.626	2837.2848	2612.679

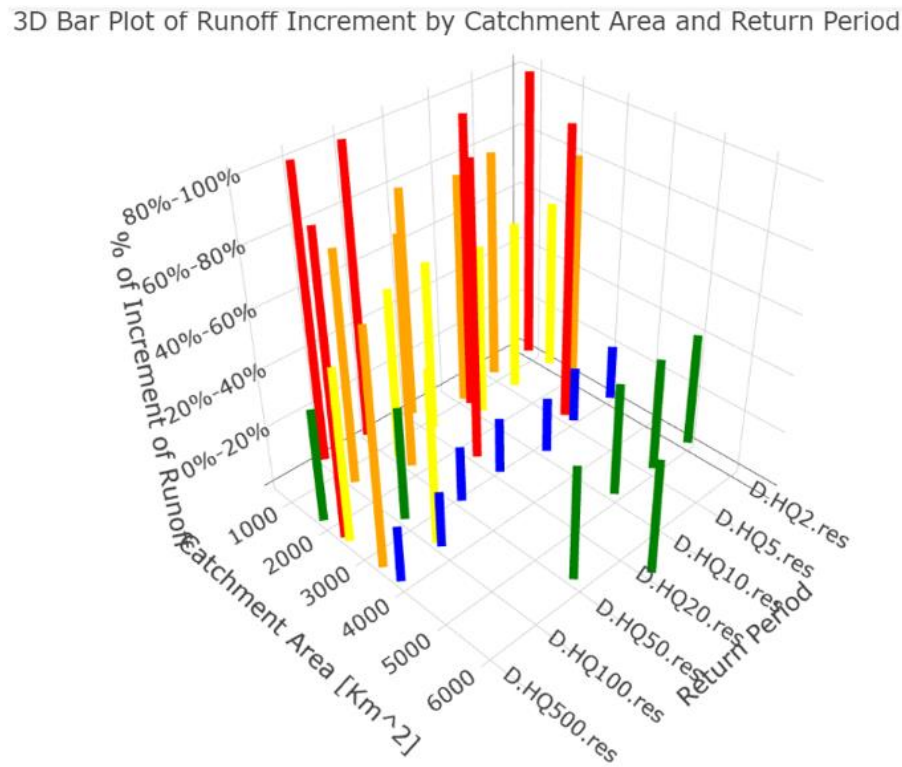

Figure 23: Visual representation of decrements according to the mean area of catchments producing them, for every return period. Colours vary according to the decrement percentage (blue < 20%, green between 20% and 40%, yellow between 40% and 60%, orange between 60% and 80%, red > 80%).

From the graph, it is evident that smaller catchment areas correspond to higher increments (more than 40%), independently from the return period considered. Medium catchment areas are related to smaller decrements (lower than 20%). Wider catchments are associated with medium increments (between 20% and 40%), even if the model is not so precise for such big territories.

This result aligns with expectations, because when the barrier of a little basin is neglected, a lot of water feeds the outgoing river of modest dimensions, causing a big difference in the runoff.

To put in relationship these results with the number of catchments belonging to each combination "increment – return period", is it possible to analyse Figure 24. Here, darker colours indicate that the majority of catchments produce small decrements. This is in line with results of analysis obtained in Chapter 4, stating that there is a very high number of catchments producing small variations, and a minority causing a strong decrement. As a matter of fact, Table 8 reports that the catchments up to the 75° percentile produces at most an increment of 12% when considering basins (related to the return period of 100 years). Table 10 shows the precise number of catchments belonging to each combination.

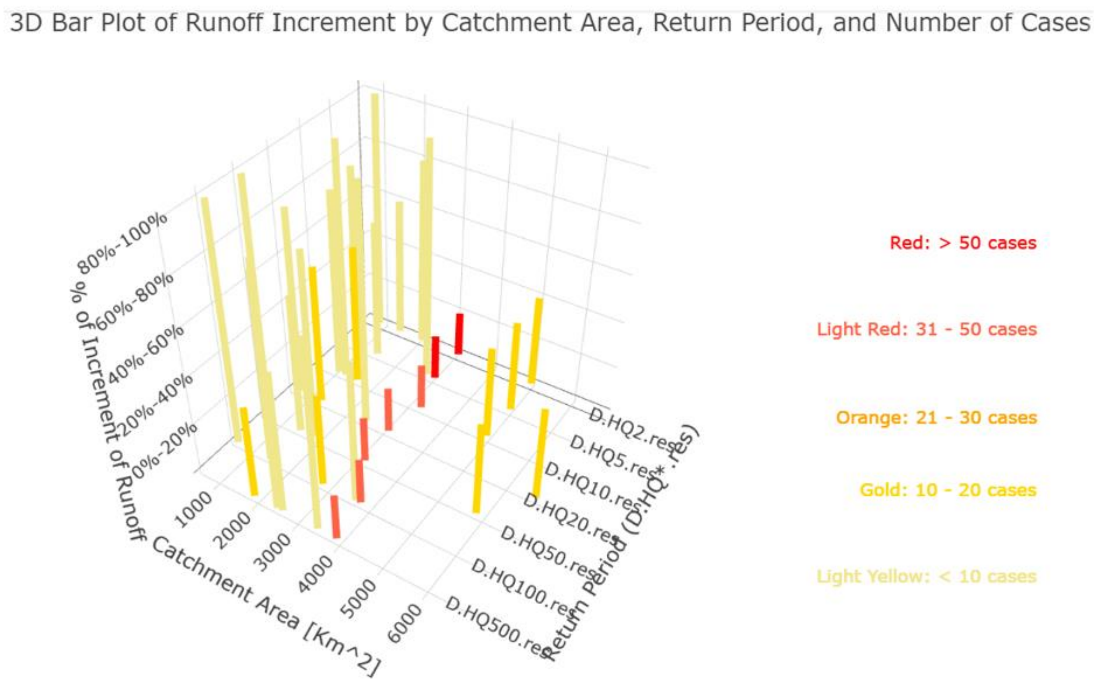

Figure 24: Visual representation of decrements according to the mean area of catchments producing them, for every return period. Colours vary according to the number of catchments for every category.

Table 10: Number of cases producing a percentage of decrement for every return period.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	58	55	50	47	45	41	44
0.2-0.4	12	14	12	11	12	15	13
0.4-0.6	7	6	11	13	9	8	3
0.6-0.8	5	6	4	2	5	6	7
0.8-1	2	3	3	4	5	3	4
> 1	16	16	20	23	24	27	29

To make a comparison between decrements of flood quantiles and areas and volumes of the basins, the same approach was used: computation of the mean of the variables, for every set of catchments belonging to each category.

In the code, it is possible to consult variables called sum.dams.area and sum.dams.vol, that for each catchment express, respectively, the sum of areas or volumes of every basin belonging to it. For example, a catchment having three dams will have a value of sum.dams.area correspondent to the sum of areas (and volumes) of the three lakes formed by the barriers; analogously, a catchment with no dams inside will have a null value of the two parameters.

It is self-evident to think that catchments producing an increment in the runoff when neglecting dams are the ones in which there are dams. However, a double check was made, and results confirm pretty much the triviality. Exceptions regarding small increments are accepted as imprecisions of the model or insignificant variations. Table 11 represent, for each category, the ratio between the number of catchments for which sum of areas (and volumes, since of course catchments without dams have null values for both variables) are higher than 0, over the total ones.

Table 11: For every category, ratio between the number of catchments having dams over total number of catchments producing a decrement.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	0.9827586	0.9818182	0.98	0.9787234	0.9777778	0.9756098	0.9772727
0.2-0.4	1.0000000	1.0000000	1.00	1.0000000	1.0000000	1.0000000	1.0000000
0.4-0.6	1.0000000	1.0000000	1.00	1.0000000	1.0000000	1.0000000	1.0000000
0.6-0.8	1.0000000	1.0000000	1.00	1.0000000	1.0000000	1.0000000	1.0000000
0.8-1	1.0000000	1.0000000	1.00	1.0000000	1.0000000	1.0000000	1.0000000
> 1	1.0000000	1.0000000	1.00	1.0000000	1.0000000	1.0000000	1.0000000

Next, comparison with average sum of areas and volumes of catchments producing a determined decrements could be performed.

Focusing on areas first, results are displayed in Table 12.

Table 12: Average of basin areas (in Km^2) of catchments producing a certain difference in flood quantiles, according to distinct return periods.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	2475.3368	2599.5936	2859.2418	2330.0238	2425.7296	2640.26883	5156.6811
0.2-0.4	100623.6898	86289.9315	10179.2442	14081.6239	12935.0632	8199.71587	327.6573
0.4-0.6	671.3800	528.3902	98965.2813	83771.3667	120704.9772	139917.32513	1341.1007
0.6-0.8	27181.0330	452.2275	291.7315	81.7745	549.8288	468.04850	159905.6559
0.8-1	834.6745	45469.1960	617.9913	384.4162	174.7576	73.04767	530.1448
> 1	101788.4739	101788.4739	88259.5777	76817.6349	73649.8854	65493.73226	60872.6408

In this case, there are not clear patterns defining the decrements according to sum of basin areas. For example, focusing on differences for return period of 2 years, catchments that on average are poorly influenced by dams (lower than 1000 Km^2) produce variations between 40% and 60%, and between 80% and 100% (hence medium-high variations). However, there is not a linear trend stating that when increasing the amount of territory covered by basin lakes, than a regular variation in the runoff arises. By looking at other return period, it is noticeable that every column present different values, but with no outstanding trend.

The same applies to volumes, as pictured in Table 13.

Table 13: Average of basin volumes (in Km^3) of catchments producing a certain difference in flood quantiles, according to distinct return periods.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	53345724	54237855	57369040	53454979	54738978	60042049	61744182
0.2-0.4	126885167	115545857	93739333	121125455	101025000	58048800	48628154
0.4-0.6	55553000	36398333	67157273	64786154	85947111	122131500	60348044
0.6-0.8	50624600	45042500	20956000	2377000	46982000	54628333	134261143
0.8-1	25745750	73596833	75940333	46950000	14966000	34260000	70062500
> 1	149155369	149155369	132647320	122482930	122409475	109503237	101607354

To be more accurate, the same analysis was repeated using as entries of the table the average values of ratio_area, a variable of the code pointing out how much of a catchment territory is occupied by dams, in terms of areas. Results are displayed in table 14, and an increase in the percentage of decrements while incrementing the area due to reservoir surfaces shows up.

Table 14: Average of percentage of catchment areas occupied by basin lakes of catchments producing a certain difference in flood quantiles, according to distinct return periods.

	D.HQ2.res	D.HQ5.res	D.HQ10.res	D.HQ20.res	D.HQ50.res	D.HQ100.res	D.HQ500.res
0.01-0.2	0.03477284	0.0344772	0.02847886	0.02654969	0.02549253	0.02389633	0.02950661
0.2-0.4	0.20003935	0.1777541	0.22748379	0.20953166	0.21570704	0.17417166	0.14100407
0.4-0.6	0.30759264	0.2694919	0.24479002	0.27110816	0.21340420	0.19907358	0.32830064
0.6-0.8	0.35516324	0.2875520	0.28074814	0.14346474	0.28894232	0.33007696	0.28049624
0.8-1	0.11080617	NaN	0.11856663	0.25137955	0.32091772	0.22860264	0.19355983
> 1	0.39005801	0.3900580	0.38087274	0.35336431	0.31965719	0.32301470	0.30078756

This trend means that the more a catchment territory is occupied by dams' lakes, the higher is the increment in the runoff when ignoring the barrier, for every return period.

5.2 Analysis of all catchments

Although results obtained in Chapter 4 could seem meaningful and promising, it is important to remember that all catchments related to a “strange” outcome of relative difference between simulations ignoring and considering dams were neglected in the Figures of the previous Chapter. For “strange” is it meant of course missing values NA (more than 100), but also negative numbers (around 60 cases), and catchments in which for at least one return period, the relative difference was higher than 10 (30 instances). Excluding this last category implicates not to consider all the situations in which an increment of more than 1000% resulted from the calculations (that is, runoff without barriers is 10 times higher than runoff with them). These unused values reach orders of magnitudes up to 10^7. The list of all ignored catchments is in the Appendix.

However, for a fairer analysis, plots considering all catchments are reported here. Maps corresponding to return periods of 2 and 50 years follows, whereas in the Appendix there are all the other cases.

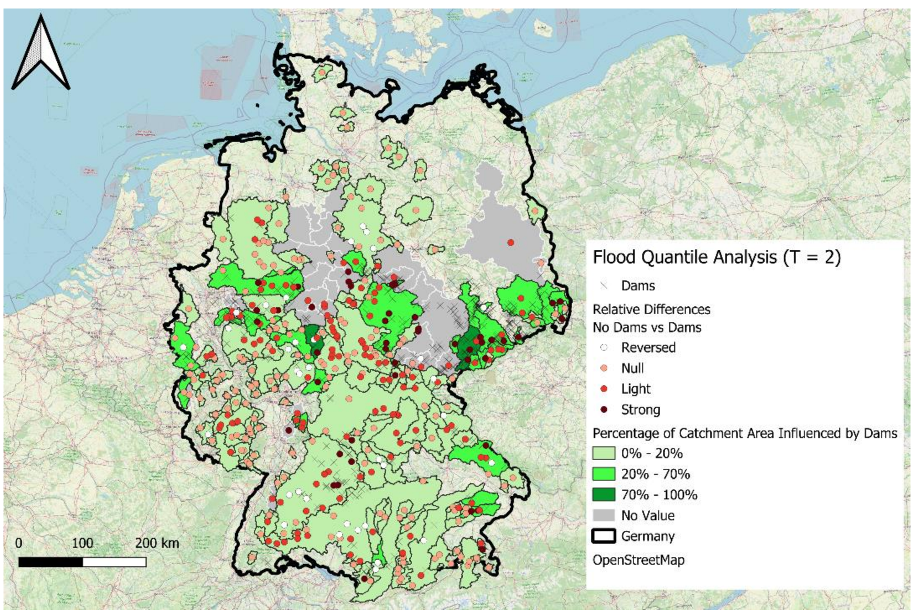

Figure 25: Relative differences of neglecting and considering the action of dams for T = 2 years.

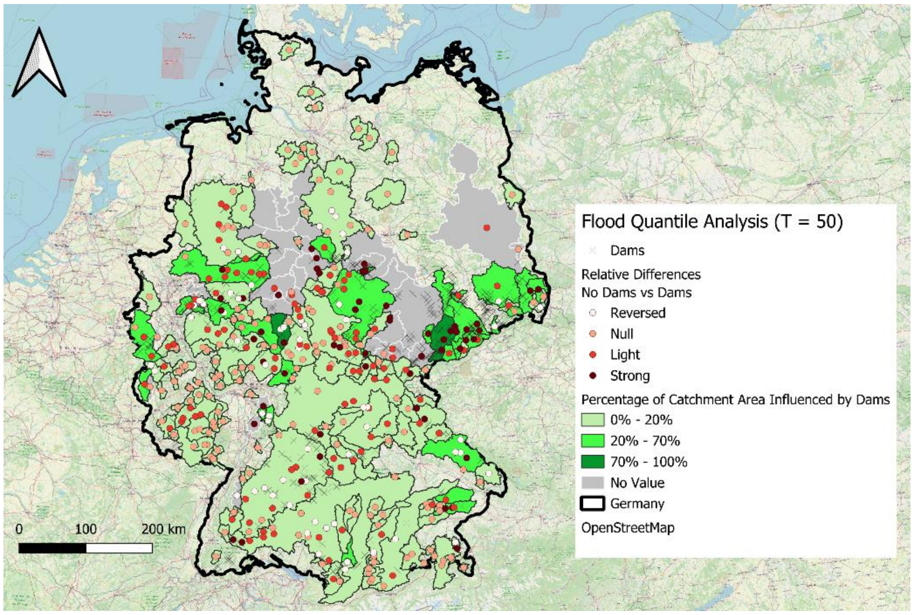

Figure 26: Relative differences of neglecting and considering the action of dams for T = 50 years.

In these examples, white dots correspond to negative results of relative differences (absurdity in which the action of dams increases the runoff). Burgundy dots are obviously more than in the previous case, because all the “Strong” instances were considered here.

Is it not easy to find a trend along the space or the return periods; however, it seems that “Strong” category dots are mainly concentrated at middle latitudes. For more precise considerations it is possible to consult the following tables.

Table 15: Statistical indexes of the relative differences between ignoring and considering dams “simulated” flood quantiles, all catchments.

	Mean	Median	Variance	StdDev
HQ2	1995.7725	0	5.996711ε+08	24488.183
HQ5	504.7536	0	2.980699ε+07	5459.577
HQ10	433.9382	0	1.702818ε+07	4126.521
HQ20	457.1861	0	1.514657ε+07	3891.859
HQ50	727.1085	0	3.542580ε+07	5951.957
HQ100	1979.6496	0	6.084838ε+08	24667.463
HQ500	138845.9510	0	7.339617ε+12	2709172.809

By looking at the Mean, Variance and Standard Deviation it is clear that there are results that completely bias the overall analysis, producing decrements in the runoff of hundreds or thousands of percentage points. These outputs seem to be completely out of scale.

However, the Median is always 0 because the majority of cases have a null relative difference result (no change in the runoff considering or ignoring the possible dams in the territory), as it is confirmed by the subdivision in quartiles below, and already discussed in Chapter 4.

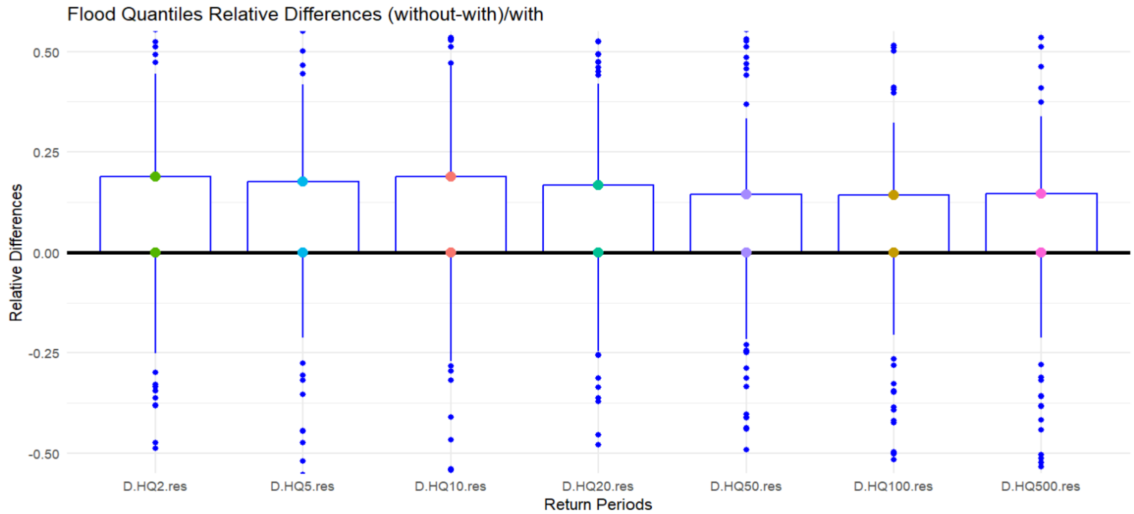

Figure 27: Box plots of the relative differences between neglecting and considering dam "simulations" for different return periods.

Table 16: Subdivision of relative differences in quartiles.

	P25	P50	P75
D.HQ2.res	0	0	0.1884689
D.HQ5.res	0	0	0.1768056
D.HQ10.res	0	0	0.1879580
D.HQ20.res	0	0	0.1682257
D.HQ50.res	0	0	0.1447792
D.HQ100.res	0	0	0.1422658
D.HQ500.res	0	0	0.1468630

Possible explanations for these results are imprecisions in the inputs, and also in the model structure. As a matter of fact, this algorithm had never been used previously for such a huge spatial domain. During the making of this thesis, some changes were brought to the code lines, resulting in improvements and always more promising outputs. This encourages to look for good conclusions for always a higher number of catchments.

6 Conclusion

This thesis aimed to evaluate flood quantiles in two different scenarios: one in which there are active dams in the catchment, that could potentially store water and release it in more convenient periods, and one in which, considering the same watershed, they are neglected.

To compute these results, SALTO model was exploited. Its calibration was performed in two steps: initially by finding lumped parameters, hence in every catchment each model element assumes a constant value (local calibration), in a second time by computing distributed parameters, that is for every model element within the same catchment, there are different values according to the cell of the grid they belong (regional calibration). This last phase is carried out through the PASS method, innovative because it does not need any a priori definition of the dominant catchment descriptors that control regional patterns. The calibration function used is a weighted sum of the Kling – Gupta metric and another one focusing on flood quantiles evaluated for return periods of 5 years, to examine better high flows.

Results of the calibrations were promising, since values of the efficiency of the model obtained for the majority of catchments were quite high. However, yet from the analysis of hydrographs and flow duration curves, many imprecisions arose, suggesting that the model might not work properly.

As a matter of fact, the simulation of relative differences between flood quantiles in the two compared cases produced unexpected (and sometimes absurd) results, with some of the catchments getting a negative number (i.e., dams increase the runoff), and others corresponding to a runoff decrease of thousands of percentage points.

Thus, by neglecting these “strange” outputs, the analysis of filtered results pointed out that for higher return periods the contribution of the reservoirs is more visible, with the variation of mean values ranging between 20% and 30%. Also, stronger effects are related to places in which a thicker concentration of barriers is present. Lastly, big variations in the runoff in the two cases of considering and neglecting dams are quite homogeneous on the whole territory for higher return periods, and mainly distributed at middle latitudes for smaller T’s. On the other hand, when all catchments are considered, the mean variation for all return periods assumes results of the order of 10^4.

In conclusion, the model seemed to have many imprecisions for the evaluation of flood quantiles in the two scenarios. This could certainly depend by inaccuracies of some data provided, such as the area of the catchments formed by the dams. As a matter of fact, many of these values are for sure far from the reality, because of the difficulties in their calculation derived from the low resolution of the DEM. Also, for a more accurate analysis, it would be convenient to look at how much nowadays every structure effectively work, that is if it is used to store water, to avoid low flow periods, or if it is not in service at all, because this can have an impact on the final evaluation of results.

The aim of studying the whole German territory all at once was challenging and stimulating, but the extended dimensions of the domain made arise many difficulties. Is it possible that better results could be obtained by reducing the sphere of interest to the single major basins constituting the country, that can be later be synthesized together.

References

- Bergström, S. (1995). *The HBV model*. In V. Singh (Ed.), Computer Models of Watershed Hydrology. Highlands Ranch, Colorado: Water Resour. Publ., pp. 443–476.
- Bloschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H. (2013). *Runoff Prediction in Ungauged Basins. Synthesis across Processes, Places and Scales.* Cambridge University Press.
- Boulange, J., Hanasaki, N., Yamazaki, D. et al. (2021). *Role of dams in reducing global flood exposure under climate change.* Nat Commun 12, 417.
- Braithwaite, R.J. (2011). Degree-Days. In: Singh, V.P., Singh, P., Haritashya, U.K. (eds) Encyclopedia of Snow, Ice and Glaciers. Encyclopedia of Earth Sciences Series. Springer, Dordrecht.
- Chaudhari, S., Pokhrel, Y. (2022). *Alteration of river flow and flood dynamics by existing and planned hydropower dams in the Amazon River basin*. Water Resources Research, 58.
- Cipollini, S., Fiori, A., & Volpi, E. (2022). *A new physically based index to quantify the impact of multiple reservoirs on flood frequency at the catchment scale based on the concept of equivalent reservoir*. Water Resources Research, 58, e2021WR031470.
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). *Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development.* Water Resources Research, 47, W11510.
- Gebregiorgis, Abebe & Hossain, Faisal. (2012). *Hydrological Risk Assessment of Old Dams: Case Study on Wilson Dam of Tennessee River Basin*. Journal of Hydrologic Engineering. 17. 10.1061/(ASCE)HE.1943-5584.0000410.
- Graf W.L. (1999). *Dam nation: A geographic census of American dams and their large-scale hydrologic impacts*. Water Resources Research, VOL. 35, NO. 4, 1305-1311.
- Hargreaves, G., Samani, Z. (1985). *Reference Crop Evapotranspiration From Temperature. Applied Engineering in Agriculture*. 1. 10.13031/2013.26773.
- Hirschfeld J., Nilson E., Keil F. (2014). *Everything Flows - A German Water Balance*. 10.13140/2.1.5063.8726.
- Ionita, M., Nagavciuc, V., Kumar, R. et al. (2020*). On the curious case of the recent decade, mid-spring precipitation deficit in central Europe*. npj Clim Atmos Sci 3, 49.
- Karthe, D., Chifflard, P., Cyffka, B. et al. (2017). *Water research in Germany: from the reconstruction of the Roman Rhine to a risk assessment for aquatic neophytes*. Environ Earth Sci 76, 549.
- Knoben, W. J. M., Freer, J. E., and Woods, R. A. (2019). *Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores.* Hydrol. Earth Syst. Sci., 23, 4323–4331.
- Leroy Poff N., Hart D.D. (2002). *How Dams Vary and Why It Matters for the Emerging Science of Dam Removal*. Bioscience, Vol. 52 No. 8, 659-668
- Merz, R., Tarasova, L., Basso, S. (2020). *Parameter's controls of distributed catchment models—How much information is in conventional catchment descriptors?* Water Resour. Res. 56 (2), e2019WR026008
- Merz, R., Miniussi, A., Basso, S., Petersen, K-J., Tarasova, L. (2022). *More Complex is Not Necessarily Better in Large-Scale Hydrological Modeling. A Model Complexity Experiment across the Contiguous United States.* American Meteorological Society, vol. 103: issue 8, E1947–E1967.
- Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., Gratzki, A. (2013). *A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS).* Meteorologische Zeitschrift. 22. 235-256.
- Singh, V.P. (1998). *Generalized Extreme Value Distribution. In: Entropy-Based Parameter Estimation in Hydrology.* Water Science and Technology Library, vol 30. Springer, Dordrecht.
- Speckhann, G., Kreibich, H., Merz, B. (2020). *Inventory of dams in Germany.* 10.5194/essd-2020-266.
- Spinti, R.A., Condon, L.E. & Zhang, J. (2023) *The evolution of dam induced river fragmentation in the United States.* Nat Commun 14, 3820.
- van der Kwast, H., Menke, K. (2022). *QGIS for Hydrological Applications - Second Edition Recipes for Catchment Hydrology and Water Management.*
- Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., & Fiori, A. (2018). Reservoir effects on flood peak discharge at the catchment scale. Water Resources Research, 54, 9623–9636. https://doi.org/10.1029/2018WR023866
- Yi Y., Gao Y., Zhang S. (2020). *The impact of dams on the river connectivity of the two largest river basins in China*. Authorea.
- Zhang, A. T., and Xinyi Gu, V. (2023). *Global Dam Tracker: A Database of More than 35,000 Dams with Location, Catchment, and Attribute Information*. Scientific Data, 10 (1): 111.

- Ionita M., Scholz P., Grosfeld K. (2022). *Is Germany drying out?* - https://www.reklim.de/en/archive-news/news-2022/is-germany-drying-out/
- JICA. Annex M. Planning and Design Guidelines for Delay ActionsDams, - https://openjicareport.jica.go.jp/pdf/11389715_15.pdf
- NASA - https://gmao.gsfc.nasa.gov/research/subseasonal/atlas/GEV-RV-html/GEV-RV-description.html
- Statista Research Department, Jan 3, 2024 - https://www.statista.com/statistics/1386617/highest-average-temperature-years-germany/

- The-German-American - https://www.the-german-american.com/germany/climate-of-germany/
- USACE Engineering Hydrologic Center - https://www.hec.usace.army.mil/confluence/hmsdocs/hmstrm/evaporation-and-transpiration/hargreaves-method
- Wikipedia. *Geography of Germany* - https://en.wikipedia.org/wiki/Geography_of_Germany
- Wikipedia. *Strahler Number* - https://en.wikipedia.org/wiki/Strahler_number
- *weatherspark.com* - https://it.weatherspark.com/y/61268/Condizioni-meteorologiche-medie-a-Homberg-Assia-Germania-tutto-l'anno

Appendix

I Other graphs

- Graphical output of the model for Schwalm and Efze at catchments.

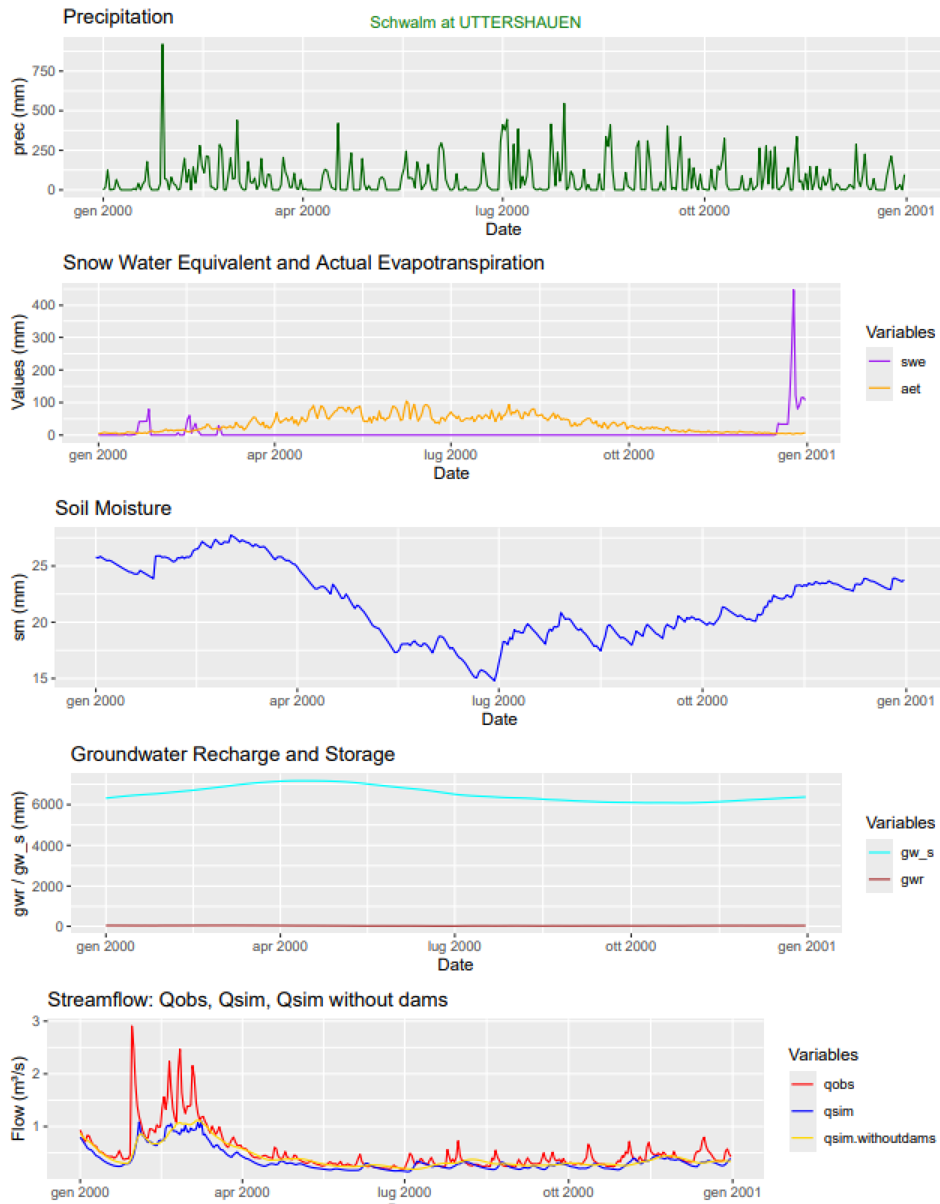

Figure 28: Precipitation [mm/day], AET [mm/day] & SWE [mm/day], SM [%], GW storage [mm] & GWR [mm/day], runoff of Schwalm River.

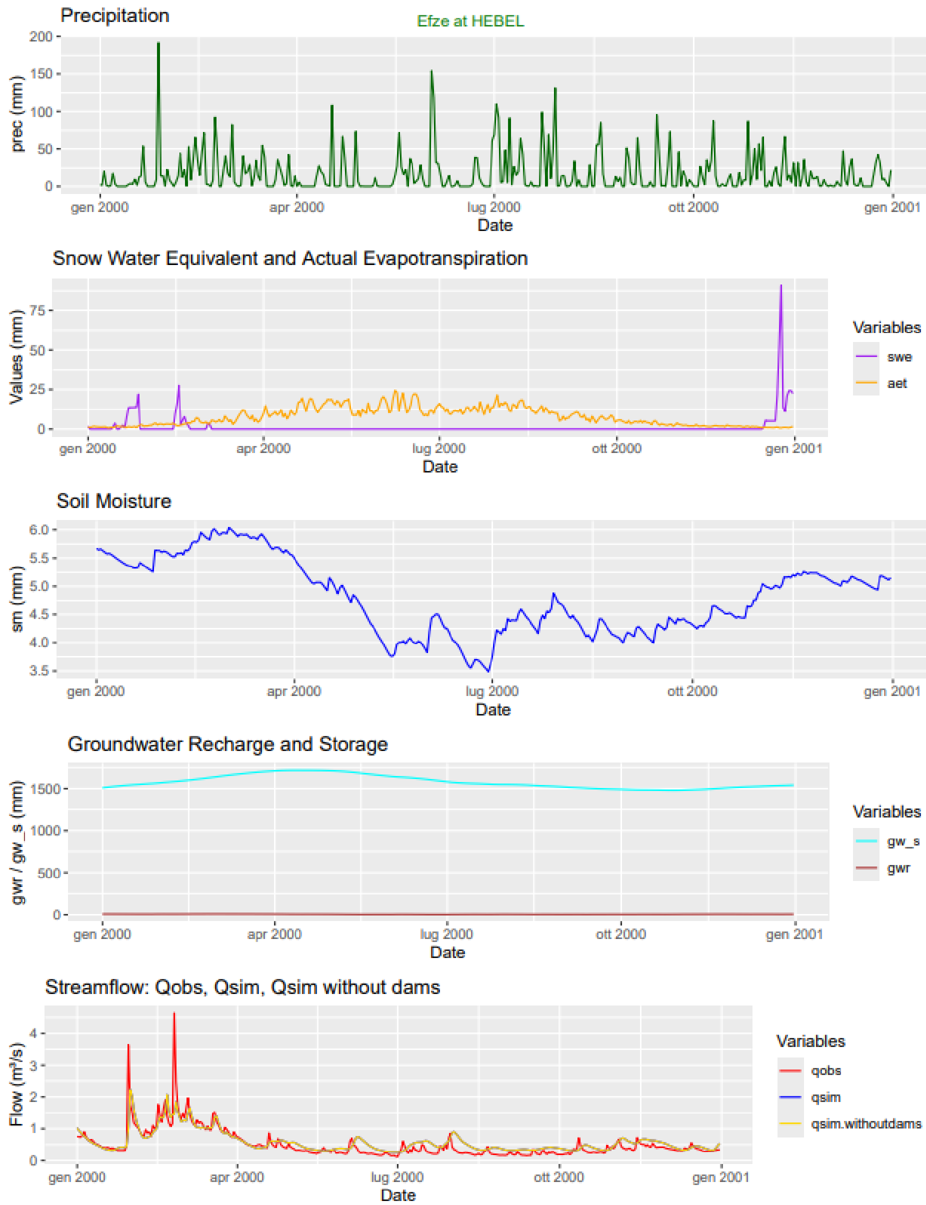

Figure 29: Precipitation [mm/day], AET [mm/day] & SWE [mm/day], SM [%], GW storage [mm] & GWR [mm/day], runoff of Efze River.

- Maps representing relative differences between flood quantiles obtained from simulations and observations of the runoff for T =2, 5, 10, 20, 50, 100, 500 years.

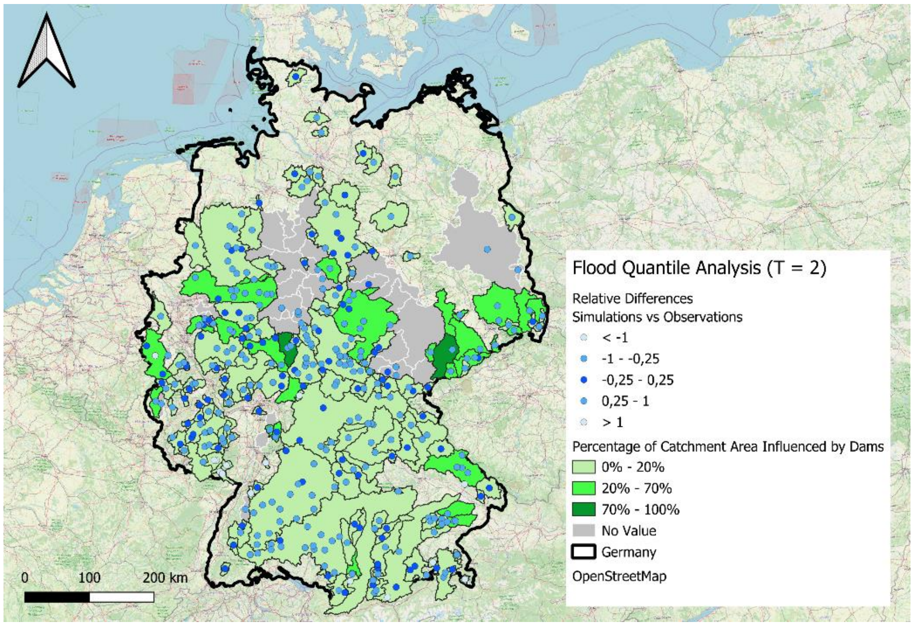

Figure 30: Relative differences of simulated and observed values for T = 2 years.

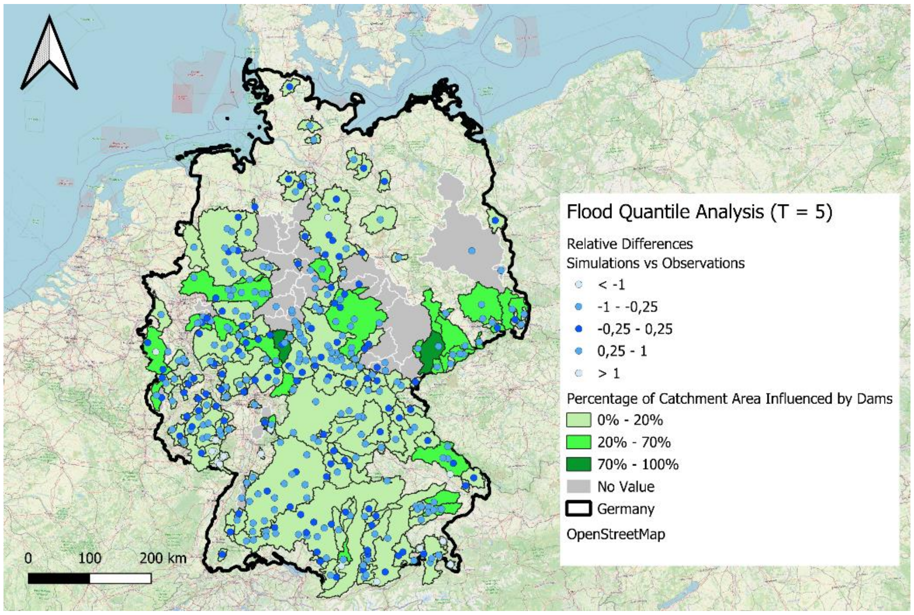

Figure 31: Relative differences of simulated and observed values for T = 5 years.

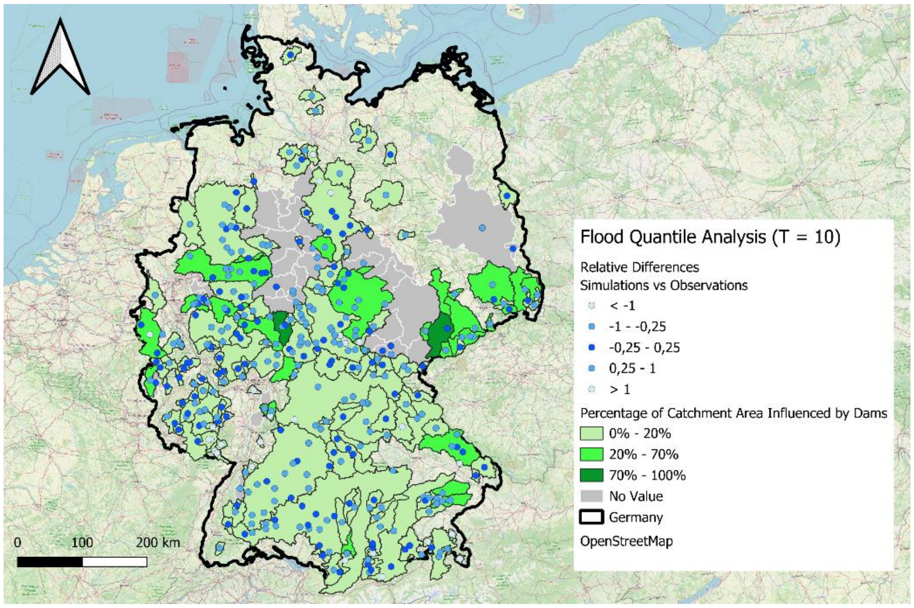

Figure 32: Relative differences of simulated and observed values for T = 10 years.

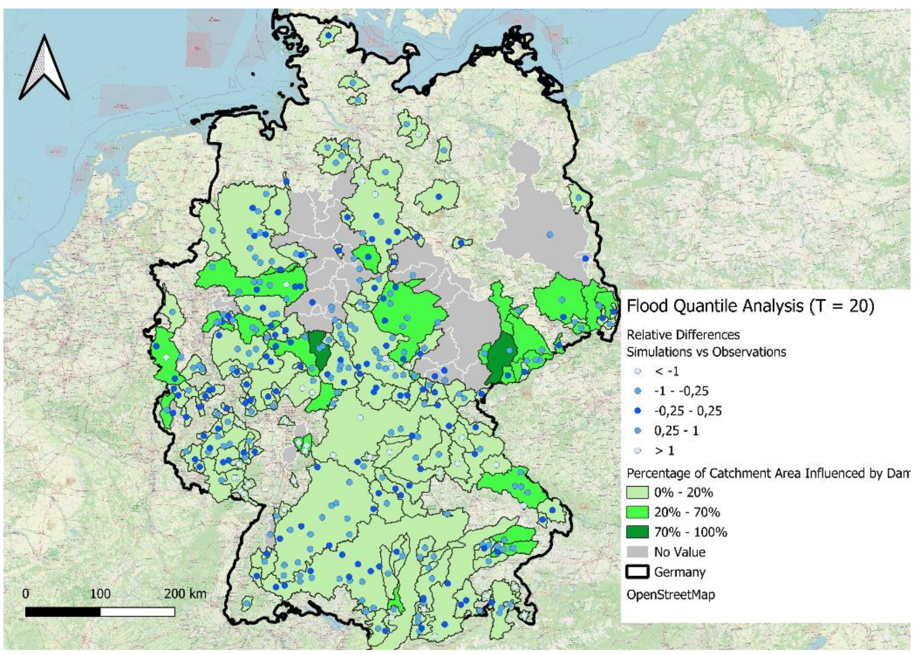

Figure 33: Relative differences of simulated and observed values for T = 20 years.

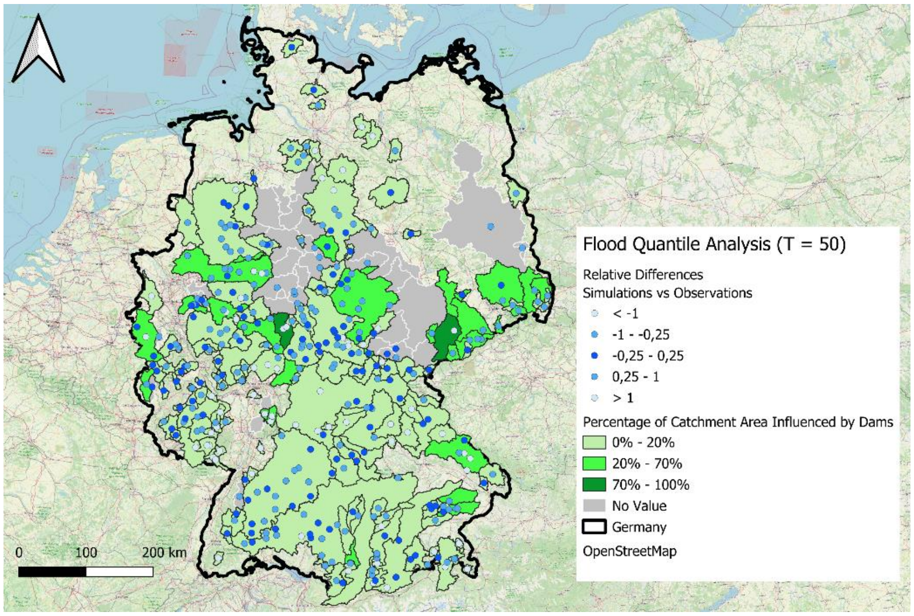

Figure 34: Relative differences of simulated and observed values for T = 50 years.

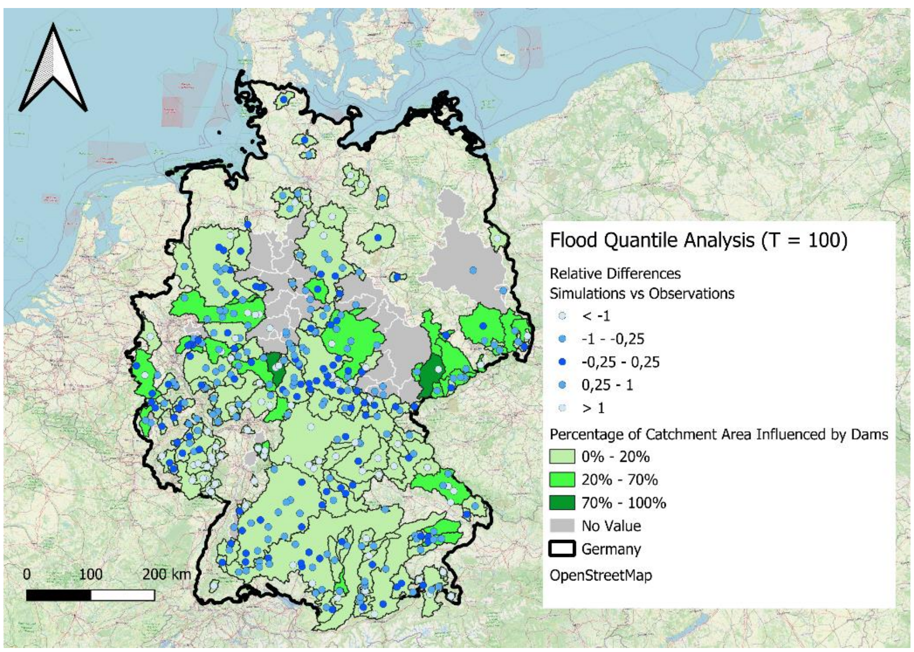

Figure 35: Relative differences of simulated and observed values for T = 100 years.

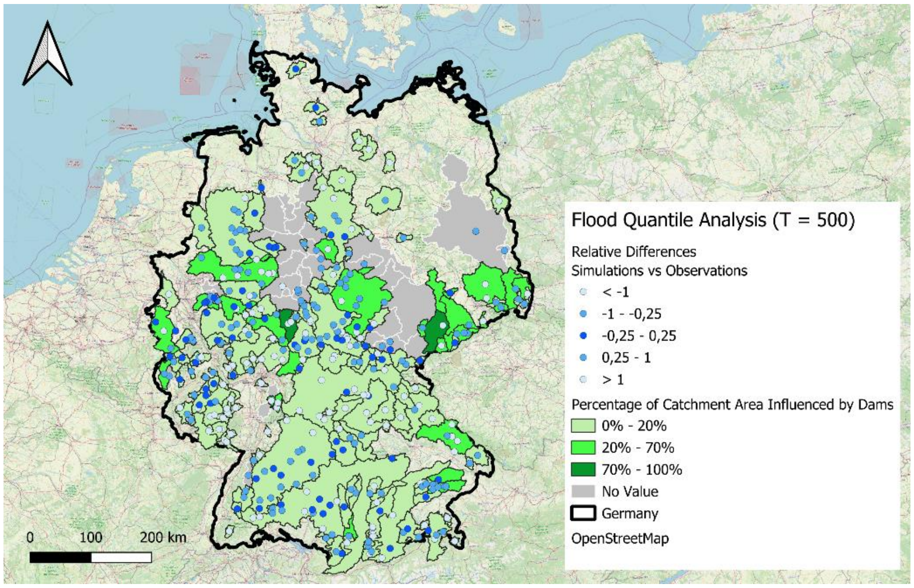

Figure 36: Relative differences of simulated and observed values for T = 500 years.

- Maps representing relative differences between flood quantiles obtained from simulations of the runoff while ignoring and considering the effect of the dams for T =2, 5, 10, 20, 50, 100, 500 years, filtered catchments only.

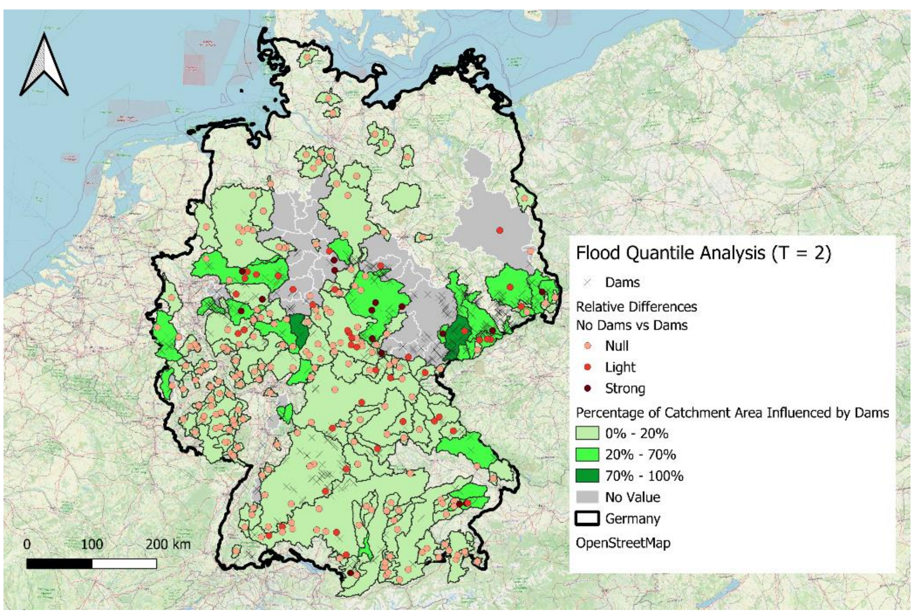

Figure 37: Relative differences of neglecting and considering the action of dams for T = 2 years.

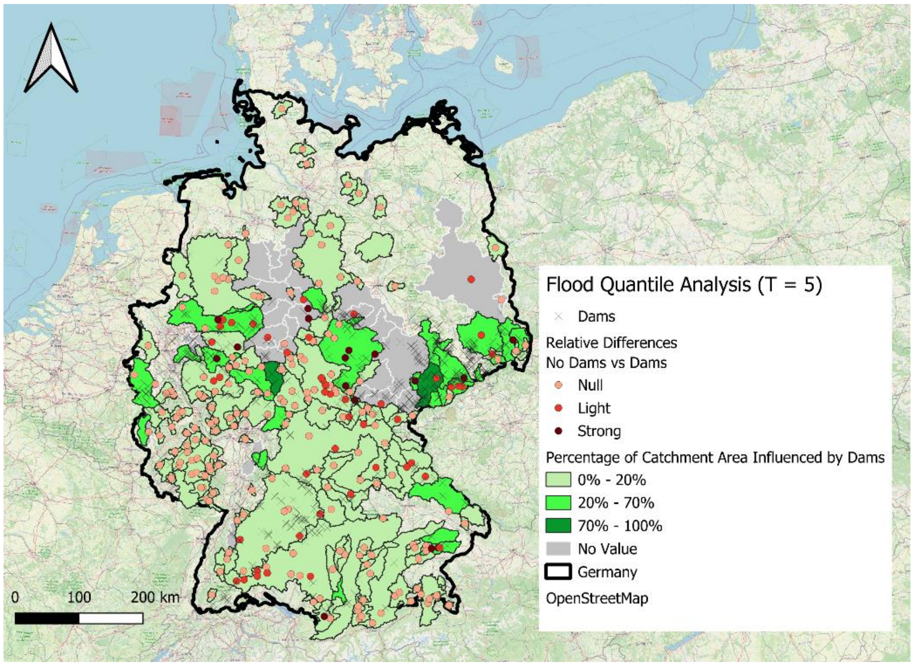

Figure 38: Relative differences of neglecting and considering the action of dams for T = 5 years.

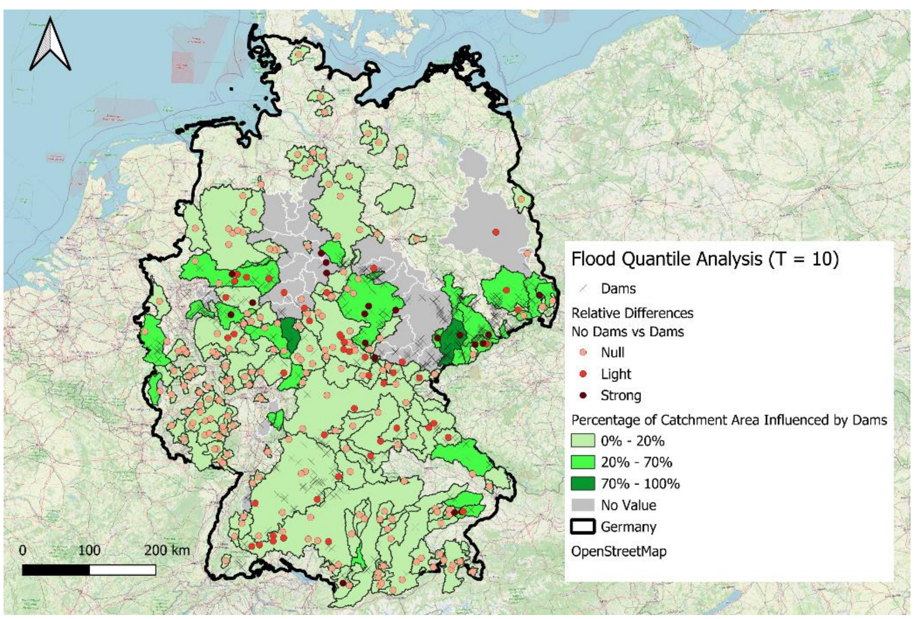

Figure 39: Relative differences of neglecting and considering the action of dams for T = 10 years.

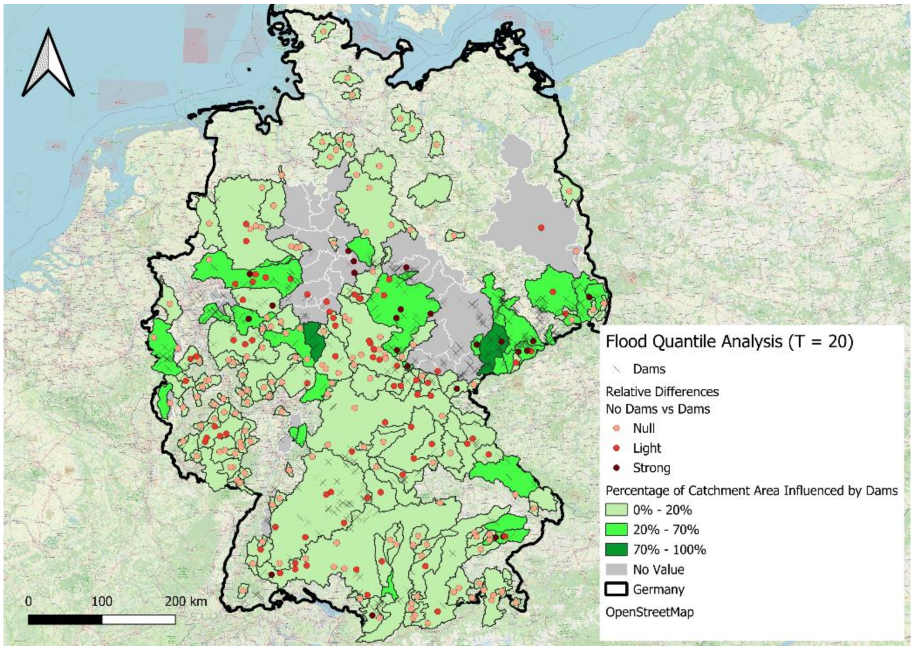

Figure 40: Relative differences of neglecting and considering the action of dams for T = 20 years.

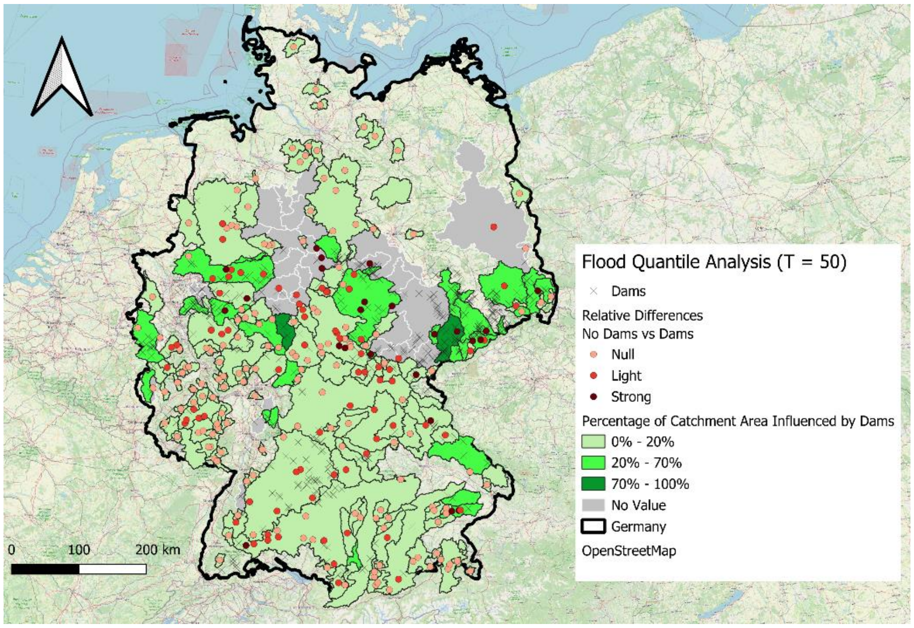

Figure 41: Relative differences of neglecting and considering the action of dams for T = 50 years.

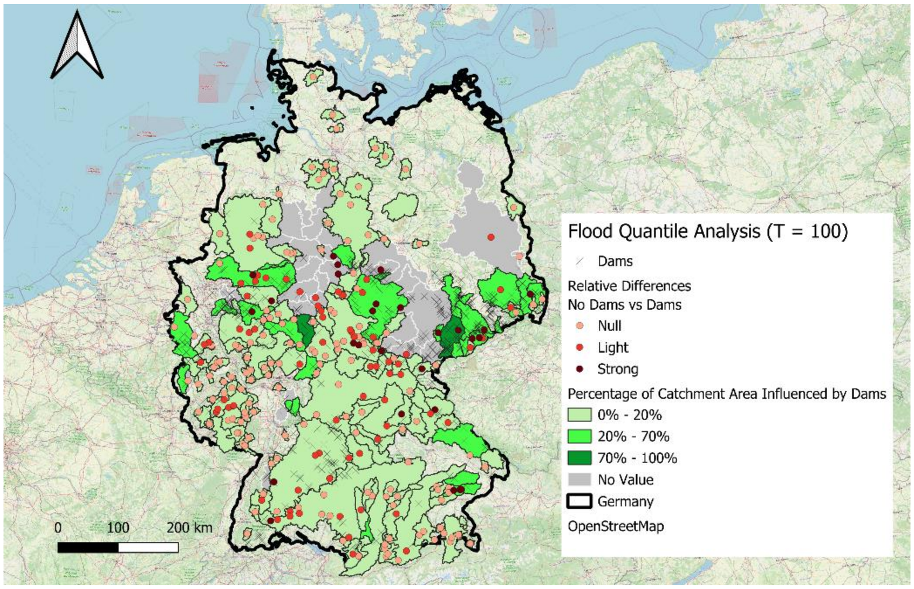

Figure 42: Relative differences of neglecting and considering the action of dams for T = 100 years.

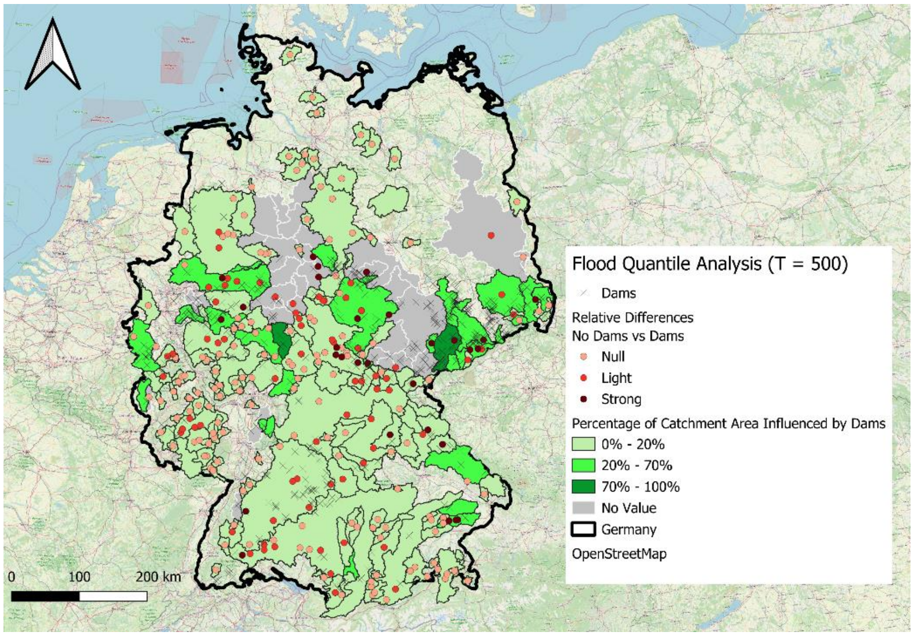

Figure 43 : Relative differences of neglecting and considering the action of dams for T = 500 years.

- Maps representing relative differences between flood quantiles obtained from simulations of the runoff while ignoring and considering the effect of the dams for T =2, 5, 10, 20, 50, 100, 500 years, all catchments.

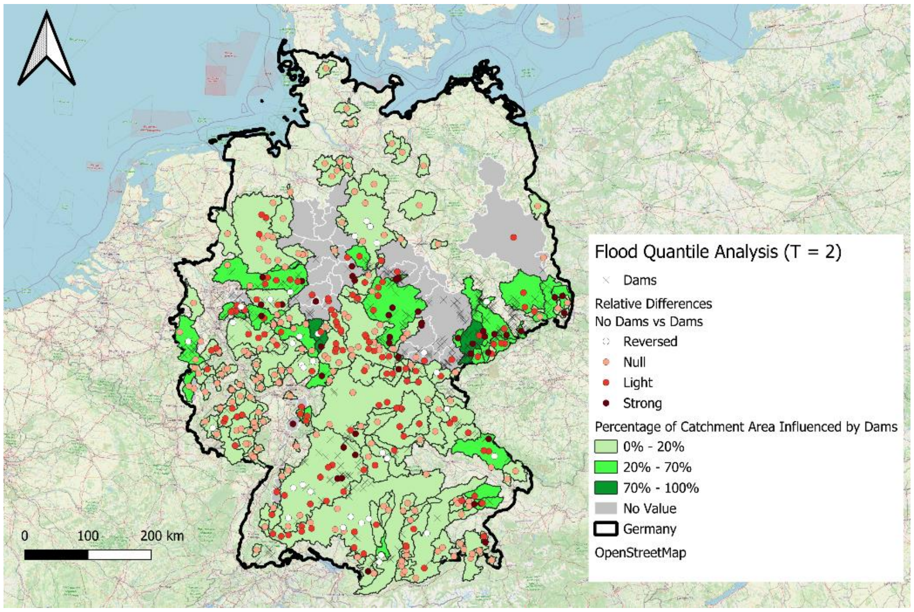

Figure 44: Relative differences of neglecting and considering the action of dams for T =2 years.

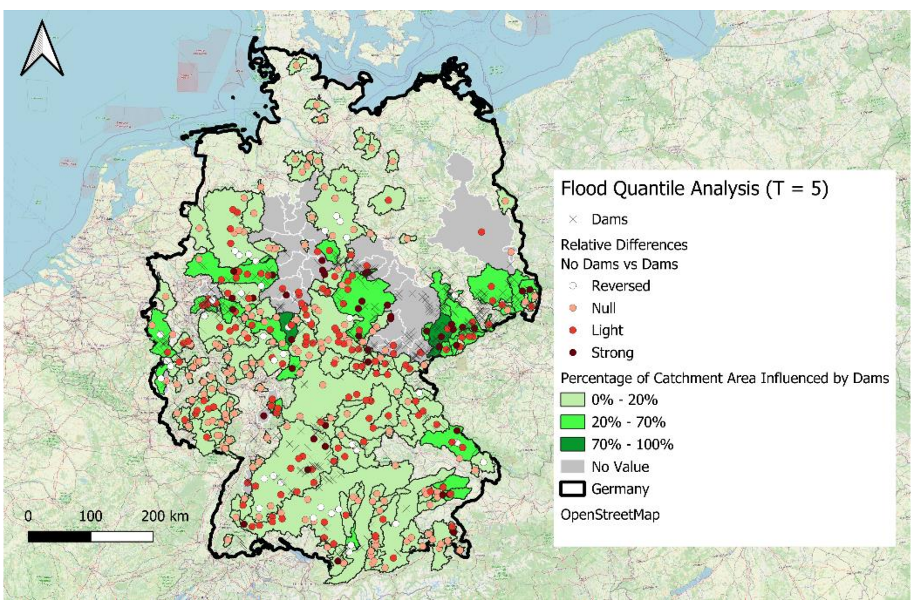

Figure 45: Relative differences of neglecting and considering the action of dams for T = 5 years.

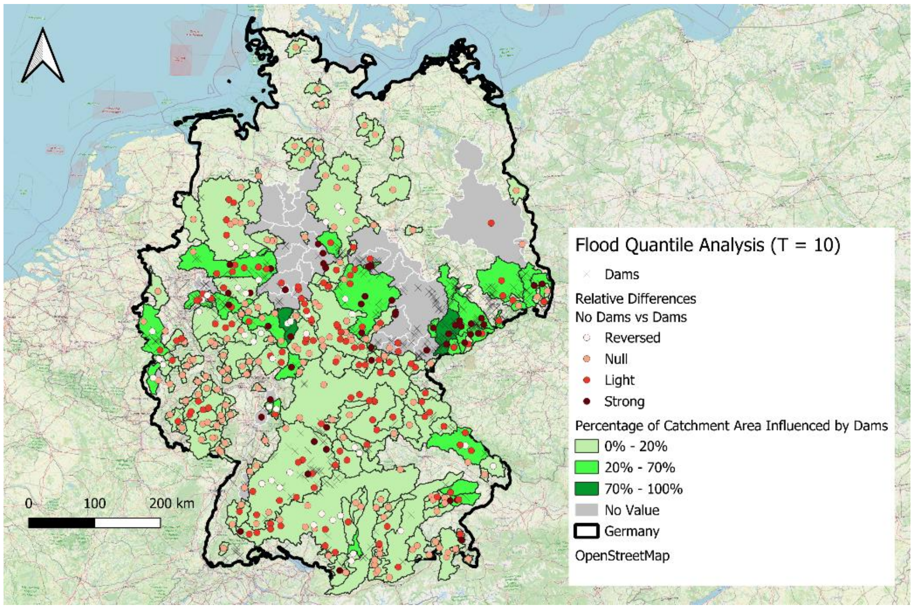

Figure 46: Relative differences of neglecting and considering the action of dams for T = 10 years.

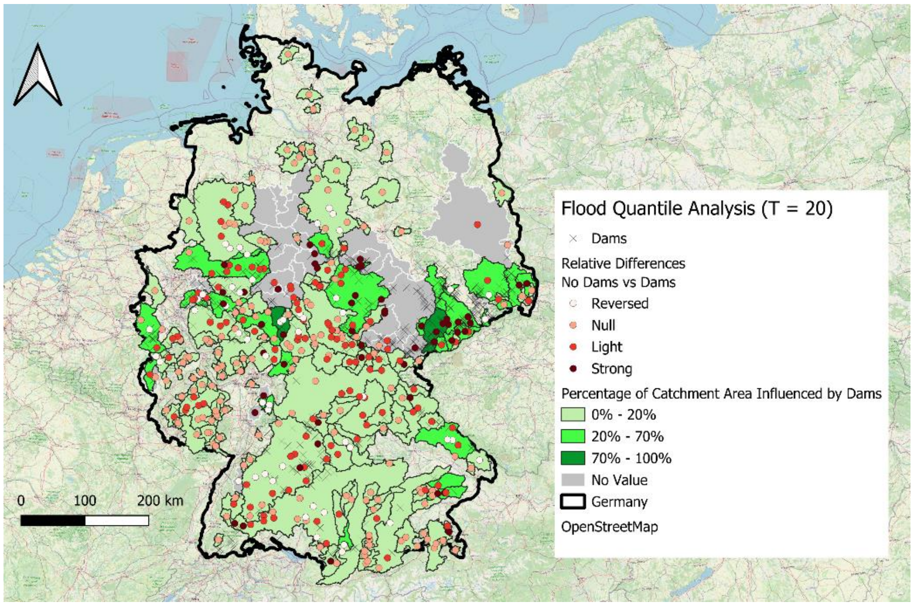

Figure 47: Relative differences of neglecting and considering the action of dams for T = 20 years.

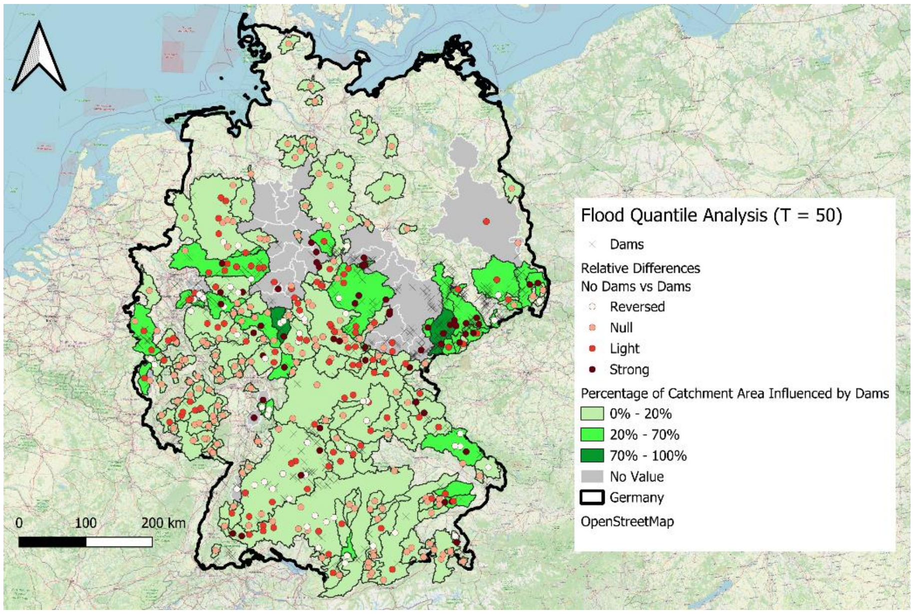

Figure 48: Relative differences of neglecting and considering the action of dams for T = 50 years.

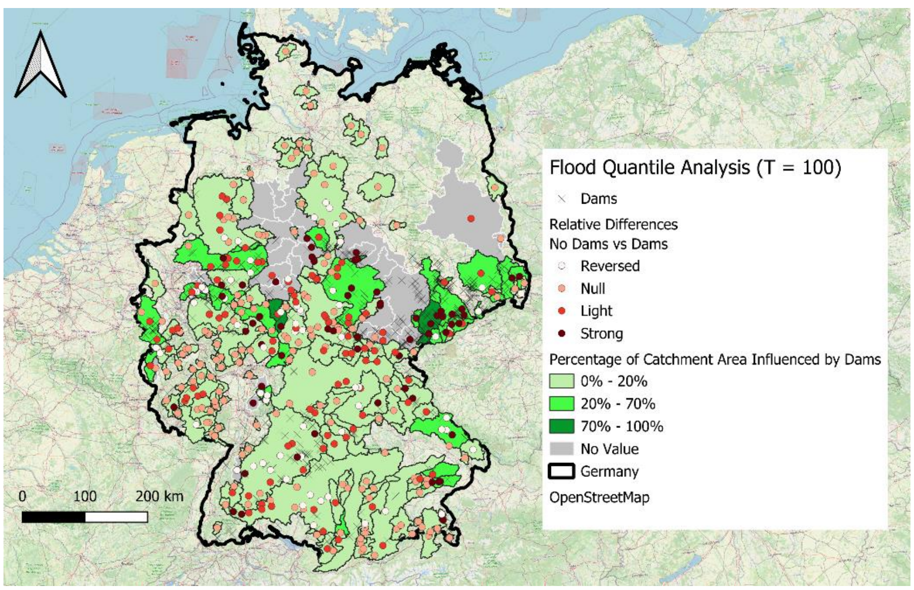

Figure 49: Relative differences of neglecting and considering the action of dams for T = 100 years.

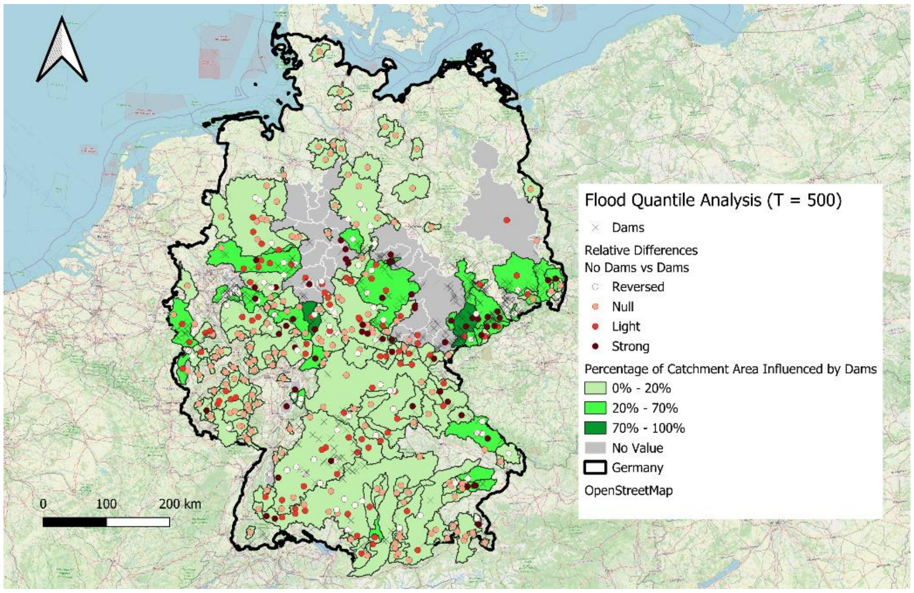

Figure 50: Relative differences of neglecting and considering the action of dams for T = 500 years.

- ID's of ignored catchments

Table 17: List of catchments producing relative differences of considering and neglecting the action of dams higher than 10.

	A	B	C	D
1	ID	River	Station	Period
2	18662000	Sur	Teisendorf	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
3	579605	Selke	SILBERHUETTE	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
4	42880458	Schwalm	ALSFELD	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
5	44480552	Erpe	EHRINGEN	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
6	1139	Breg	HAMMEREISENBACH	D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
7	579610	Selke	MEISDORF	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
8	24162206	Itz	COBURG	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
9	23942300	Weschnitz	Lorsch	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
10	24762653	Gersprenz	HARRESHAUSEN	D.HQ100.res, D.HQ500.res
11	567420	Zschopau	HOPFGARTEN	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
12	6335690	Rems	NEUSTADT	D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
13	566040	Freiberger Mulde	NOSSEN 1	D.HQ500.res
14	583200	Schwarzer Schöps	BOXBERG	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
15	24784259	Kinzig	HANAU	D.HQ500.res
16	477	Jagst	DOERZBACH	D.HQ2.res, D.HQ5.res, D.HQ10.res
17	562070	Zwickauer Mulde	ZWICKAU-POELBITZ	D.HQ2.res, D.HQ5.res
18	24850058	Nidda	ILBENSTADT	D.HQ500.res
19	576470	Weißer Elster	GREIZ	D.HQ50.res, D.HQ100.res, D.HQ500.res
20	25830056	Lahn	MARBURG	D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
21	14002305	Naab	UNTERKOEBLITZ	D.HQ100.res, D.HQ500.res
22	562115	Zwickauer Mulde	WECHSELBURG 1	D.HQ50.res, D.HQ100.res, D.HQ500.res
23	4885118	Leine	GREENE	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
24	566100	Zwickauer Mulde	ERLLN	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
25	25800200	Lahn	LEUN	D.HQ500.res
26	27695100	Ruhr	HATTINGEN	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
27	25800608	Lahn	KALKOFEN	D.HQ500.res
28	560021	Mulde	GOLZERN 1	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
29	560051	Vereinigte Mulde	BAD DUEBEN 1	D.HQ2.res, D.HQ5.res, D.HQ10.res
30	570810	Saale	HALLE TROTHA UP	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res
31	570910	Saale	BERNBURG UP	D.HQ2.res, D.HQ5.res, D.HQ10.res, D.HQ20.res, D.HQ50.res, D.HQ100.res, D.HQ500.res

II Catchment descriptors

The following table, from the work of Merz, Tarasova and Basso (2020), shows all the catchment descriptors of the model.

Table 18: Catchment descriptors.

Group	Label	Units	Description	Source and resolution of raster data or map scale for vector data
Climate	CL_MAP	mm	Long-term mean annual precipitation	REGNIE, DWD (*Rauthe et al.*, 2013), 1x1 km raster
	CL_MAT	°C	Long-term mean annual temperature aggregated from daily fields interpolated by external drift kriging (*Zink et al.*, 2017)	DWD, 8x8 km raster
	CL_PET.P	-	Aridity index (*Budyko,* 1974) as ratio of mean annual potential evaporation and mean annual precipitation	DWD, 8x8 km raster
	CL_P.sum2win	-	Ratio of long-term summer precipitation and winter precipitation	REGNIE, DWD (*Rauthe et al.*, 2013), 1x1 km raster
	CL_R50	mm/day	Long-term median maximum daily precipitation	REGNIE, DWD (*Rauthe et al.*, 2013), 1x1 km raster
	CL_R95	mm/day	Long-term 95th percentile of maximum daily precipitation	REGNIE, DWD (*Rauthe et al.*, 2013, 1x1 km raster)
	CL_dR.D2D	mm	Long-term mean difference of rainfall amount between two consecutive days	REGNIE, DWD (*Rauthe et al.*, 2013, 1x1 km raster)
	CL_DS.mean	days	Long-term mean duration of dry spells (minimum 1 wet day between dry spells)	REGNIE, DWD (*Rauthe et al.*, 2013), 1x1 km raster
	CL_DS.max	days	Mean of yearly maximum dry spells (minimum 1 wet day between dry spells)	REGNIE, DWD (*Rauthe et al.*, 2013), 1x1 km raster
Morphology & topography	MP_mean_dem	m asl	Mean elevation	DEM, SRTM, 30x30 m raster
	MP_cv_dem	-	Coefficient of variation of elevation in the catchment	DEM, SRTM, 30x30 m raster
	MP_mean_slope	%	Median slope	DEM, SRTM, 30x30 m raster
	MP_mean_aspect	°	Mean aspect	DEM, SRTM, 30x30 m raster
	MP_mean_TWI	-	Mean topographic wetness index (*Beven and Kirkby*, 1979) defined as ln(area/slope)	DEM, SRTM, 30x30 m raster
	MP_DD	km/km^2	Drainage density	DEM, SRTM, 30x30 m raster

Land use	LD_smallveg	%	Percent of the catchment covered with herbaceous, little or no vegetation, and open spaces	CORINE 2000, EEA, 100x100 m raster
	LD_agri	%	Percent of the catchment covered with agricultural areas	CORINE 2000, EEA, 100x100 m raster
	LD_wetland	%	Percent of the catchment covered with wetlands and lakes	CORINE 2000, EEA, 100x100 m raster
	LD_urban	%	Percent of the catchment covered with artificial surfaces	CORINE 2000, EEA, 100x100 m raster
	LD.forest	%	Percent of the catchment covered with various types of forests	CORINE 2000, EEA, 100x100 m raster
Soil physical and water properties	SOIL_mean_ subsoil_silt	%	Mean fraction of silt in subsoil (30-100 cm)	HWSD, 1x1 km raster
	SOIL_mean_ subsoil_sand	%	Mean fraction of sand in subsoil (30-100 cm)	HWSD, 1x1 km raster
	SOIL_mean_ subsoil_clay	%	Mean fraction of clay in subsoil (30-100 cm)	HWSD, 1x1 km raster
	SOIL_mean_subs oil_gravel	%	Mean fraction of gravel in subsoil (30-100 cm)	HWSD, 1x1 km raster
	SOIL_mean_ topsoil_silt	%	Mean fraction of silt in topsoil (0-30 cm)	HWSD, 1x1 km raster
	SOIL_mean_ topsoil_sand	%	Mean fraction of sand in topsoil (0-30cm)	HWSD, 1x1 km raster
	SOIL_mean_ topsoil_clay	%	Mean fraction of clay in topsoil (0-30 cm)	HWSD, 1x1 km raster
	SOIL_mean_ topsoil_gravel	%	Mean fraction of gravel in topsoil (0-30 cm)	HWSD, 1x1 km raster
	SOIL_mean_ subsoil_bulk_ density	kg/dm^3	SOTWIS bulk density (*van Engelen et al.*, 2005) of subsoil (30-100 cm)	HWSD, 1x1 km raster
	SOIL_mean_ topsoil_bulk_ density	kg/dm^3	SOTWIS bulk density (*van Engelen et al.*, 2005) of topsoil (0-30 cm)	HWSD, 1x1 km raster
	SOIL_mean_soil_ depth	dm	Mean soil depth	BÜK1000, BGR, 250x250 m raster
	SOIL_mean_FC	mm	Mean effective field capacity in rooting zone	BÜK1000, BGR, 250x250 m raster
	SOIL_awc_large	%	Percent of catchment with large (125-150 mm/m) available water content (*FAO,* 2006)	HWSD, 1x1 km raster
	SOIL_awc_med	%	Percent of catchment with medium (75-125 mm/m) available water content (*FAO,* 2006)	HWSD, 1x1 km raster
	SOIL_awc_small	%	Percent of catchment with small (15-75 mm/m) available water content (*FAO,* 2006)	HWSD, 1x1 km raster
	SOIL_ impermeable_ layer_toplayer	%	Percent of catchment with impermeable layer located within 80 cm of soil profile	HWSD, 1x1 km raster
	SOIL_ impermeable_ layer_ mediumlayer	%	Percent of catchment with impermeable layer located within 80-150 cm of soil profile	HWSD, 1x1 km raster
	SOIL_ impermeable_ layer_deeplayer	%	Percent of catchment with no impermeable layer located within 150 cm of soil profile	HWSD, 1x1 km raster
	SOIL_water_ regime_notwet	%	Percent of catchment with dominant annual average soil water regime class: not wet within 80 cm for over 3 months and not wet within 40 cm for over 1 month	HWSD, 1x1 km raster

	SOIL_water_ regime_mediumwet	%	Percent of catchment with dominant annual average soil water regime class: wet within 80 cm for 3 -6 months, but not wet within 40 cm for over 1 month	HWSD, 1x1 km raster
	SOIL_water_ regime_wet	%	Percent of catchment with dominant annual average soil water regime class: wet within 80 cm for over 6 months, but not wet within 40 cm for over 11 months	HWSD, 1x1 km raster
	SOIL_water_ regime_totalwet	%	Percent of catchment with dominant annual average soil water regime class: wet within 40 cm for over 11 months	HWSD, 1x1 km raster
	SOIL_topsoil_ texture_coarse	%	Percent of catchment with coarse topsoil texture	HWSD, 1x1 km raster
	SOIL_topsoil_ texture_medium	%	Percent of catchment with medium topsoil texture	HWSD, 1x1 km raster
	SOIL_topsoil_ texture_fine	%	Percent of catchment with fine topsoil texture	HWSD, 1x1 km raster
	SOIL_subsoil_ usda_clay	%	Percent of catchment with clay subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_subsoil_ usda_siltloam	%	Percent of catchment with silt and loam subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_subsoil_ usda_sand	%	Percent of catchment with sand subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_topsoil_ usda_clay	%	Percent of catchment with clay subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_topsoil_ usda_siltloam	%	Percent of catchment with silt and loam subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_topsoil_ usda_sand	%	Percent of catchment with sand subsoil according to USDA classification	HWSD, 1x1 km raster
	SOIL_drainage_ large	%	Percent of catchment belonging to “excessive” and “well” drainage class. Soil drainage classes are based on the guidelines from *FAO* (2006):	HWSD, 1x1 km raster
	SOIL_drainage_med	%	Percent of catchment belonging to “moderate” and “imperfect” drainage class. Soil drainage classes are based on the guidelines from *FAO* (2006):	HWSD, 1x1 km raster
	SOIL_drainage_ small	%	Percent of catchment belonging to “poor” and “very poor” drainage class. Soil drainage classes are based on the guidelines from *FAO* (2006):	HWSD, 1x1 km raster
Hydrogeology	HGEO_mean_ recharge	mm/a	Mean groundwater recharge from water balance for 1961-1990. The recharge is estimated as the difference between total precipitation, actual evapotranspiration and volume of direct runoff. Total runoff is separated to into direct and base flow by the empirical method of *Kille* (1970). Base flow index of 106 gauges is interpolated to the ungauged grids using multiple regression model. Slope, drainage density, land use, effective field capacity and depth of ground water table are chosen as explanatory variables (*Jankiewicz et al.,* 2003)	HAD, BGR, 1x1 km raster
	HGEO_kf_small	%	Percent of catchment with hydraulic conductivity lower than 10^{-5} m/s	HÜK200, BGR, map scale 1:200,000
	HGEO_kf_med	%	Percent of catchment with hydraulic conductivity 10^{-3}-10^{-5} m/s	HÜK200, BGR, map scale 1:200,000
	HGEO_kf_large	%	Percent of catchment with hydraulic conductivity higher than 10^{-3} m/s	HÜK200, BGR, map scale 1:200,000

	HGEO_gw_yield_small	%	Percent of catchment with yield less than 500 m^3/d. Measured yield of groundwater wells is regionalized using hydrogeological and geological information about aquifers.	HAD, BGR, map scale 1:200,000
	HGEO_gw_yield_med	%	Percent of catchment with yield 500-1300 m^3/d. Measured yield of groundwater wells is regionalized using hydrogeological and geological information about aquifers.	HAD, BGR, map scale 1:200,000
	HGEO_gw_yield_large	%	Percent of catchment with yield 1300-4000 m^3/d. Measured yield of groundwater wells is regionalized using hydrogeological and geological information about aquifers.	HAD, BGR, map scale 1:200,000
	HGEO_aquifer_type_aquitard	%	Percent of the catchment with aquitard	HÜK200, BGR, map scale 1:200,000
	HGEO_aquifer_type_porous	%	Percent of the catchment with porous aquifer	HÜK200, BGR, map scale 1:200,000
	HGEO_aquifer_type_lakes	%	Percent of the catchment covered by lakes	HÜK200, BGR, map scale 1:200,000
	HGEO_aquifer_type_fractured_karstic	%	Percent of the catchment with fractured and karst aquifer	HÜK200, BGR, map scale 1:200,000

III Code

```
# fq_observed is a dataframe where for every ID there is the value of the observed flood quantile calculated for every return period

# Load the data
file_path <- "C:\\Users\\seren\\Desktop\\uni\\POLI\\Magistrale\\Ambientale\\secondo anno\\tesi\\Halle\\writing\\luglio\\agosto\\cat_table_results.txt"
cat_table_results <- read.table(file_path, header = TRUE, sep = "", quote = "\"", stringsAsFactors = FALSE)

# Selecting columns for observed flood quantile
selected_columns_fq <- c("ID", "XCENT", "YCENT", "HQ2_obs", "HQ5_obs", "HQ10_obs", "HQ20_obs", "HQ50_obs", "HQ100_obs", "HQ500_obs")
fq_observed <- cat_table_results[selected_columns_fq]

# Selecting columns for relative differences
selected_columns_reldif <- c("ID", "XCENT", "YCENT", "D.HQ2", "D.HQ5", "D.HQ10", "D.HQ20", "D.HQ50", "D.HQ100", "D.HQ500")
reldif_simobs <- cat_table_results[selected_columns_reldif]

# Remove NA values
reldif_simobs_naomit <- na.omit(reldif_simobs)

# Save as CSV file for QGIS
reldif_simobs_csv <- "C:\\Users\\seren\\Desktop\\uni\\POLI\\Magistrale\\Ambientale\\secondo
anno\\tesi\\Halle\\writing\\luglio\\agosto\\results_aug\\reldif_simobs.csv"
write.csv(reldif_simobs_naomit, file = reldif_simobs_csv, row.names = FALSE)

# ***************************************************************************** #

# reldif_wowi contains relative differences (wo-wi)/wi

file_path <- "C:\\Users\\seren\\Desktop\\uni\\POLI\\Magistrale\\Ambientale\\secondo anno\\tesi\\Halle\\writing\\luglio\\agosto\\cat_table_results.txt"
cat_table_results <- read.table(file_path, header = TRUE, sep = "", quote = "\"", stringsAsFactors = FALSE)

# Selecting columns for relative differences (wo-wi)/wi
selected_columns_reldif_wowi <- c("ID", "XCENT", "YCENT", "D.HQ2.res", "D.HQ5.res", "D.HQ10.res", "D.HQ20.res", "D.HQ50.res", "D.HQ100.res",
"D.HQ500.res")
reldif_wowi <- cat_table_results[selected_columns_reldif_wowi]

# Remove NA values
reldif_wowi_naomit <- na.omit(reldif_wowi)

# Save as CSV file for QGIS
reldif_wowi_csv <- "C:\\Users\\seren\\Desktop\\uni\\POLI\\Magistrale\\Ambientale\\secondo
anno\\tesi\\Halle\\writing\\luglio\\agosto\\results_aug\\reldif_wowi.csv"
write.csv(reldif_wowi_naomit, file = reldif_wowi_csv, row.names = FALSE)

# statistics

# relative differences simulations and observations

hq_values <- c(2, 5, 10, 20, 50, 100, 500)
statistics_matrix_simobs <- matrix(NA, nrow = length(hq_values), ncol = 4)
colnames(statistics_matrix_simobs) <- c("Mean", "Median", "Variance", "StdDev")
rownames(statistics_matrix_simobs) <- paste0("HQ", hq_values)

for (i in seq_along(hq_values)) {
  hq <- hq_values[i]
  column_name <- paste0("D.HQ", hq)

  mean_value <- mean(reldif_simobs_naomit[[column_name]], na.rm = TRUE)
  median_value <- median(reldif_simobs_naomit[[column_name]], na.rm = TRUE)
  variance_value <- var(reldif_simobs_naomit[[column_name]], na.rm = TRUE)
  stdv_value <- sd(reldif_simobs_naomit[[column_name]], na.rm = TRUE)

  statistics_matrix_simobs[i, ] <- c(mean_value, median_value, variance_value, stdv_value)
}

# statistics relative differences (sim-obs)/obs
print(statistics_matrix_simobs)

# relative differences without and with

hq_values <- c(2, 5, 10, 20, 50, 100, 500)
statistics_matrix_wowi <- matrix(NA, nrow = length(hq_values), ncol = 4)
colnames(statistics_matrix_wowi) <- c("Mean", "Median", "Variance", "StdDev")
rownames(statistics_matrix_wowi) <- paste0("HQ", hq_values)
```

```
for (i in seq_along(hq_values)) {
  hq <- hq_values[i]
  column_name <- paste0("D.HQ", hq, ".res")

  mean_value <- mean(reldif_wowi_naomit[[column_name]], na.rm = TRUE)
  median_value <- median(reldif_wowi_naomit[[column_name]], na.rm = TRUE)
  variance_value <- var(reldif_wowi_naomit[[column_name]], na.rm = TRUE)
  stdv_value <- sd(reldif_wowi_naomit[[column_name]], na.rm = TRUE)

  statistics_matrix_wowi[i, ] <- c(mean_value, median_value, variance_value, stdv_value)
}
print(statistics_matrix_wowi)

# relative differences without and with only positive values

reldif_wowi_naomit_onlypos <- reldif_wowi_naomit[apply(reldif_wowi_naomit, 1, function(row) all(row >= 0)), ]

hq_values <- c(2, 5, 10, 20, 50, 100, 500)

statistics_matrix_wowi_onlypos <- matrix(NA, nrow = length(hq_values), ncol = 4)
colnames(statistics_matrix_wowi_onlypos) <- c("Mean", "Median", "Variance", "StdDev")
rownames(statistics_matrix_wowi_onlypos) <- paste0("HQ", hq_values)

for (i in seq_along(hq_values)) {
  hq <- hq_values[i]
  column_name <- paste0("D.HQ", hq, ".res")

  mean_value <- mean(reldif_wowi_naomit_onlypos[[column_name]], na.rm = TRUE)
  median_value <- median(reldif_wowi_naomit_onlypos[[column_name]], na.rm = TRUE)
  variance_value <- var(reldif_wowi_naomit_onlypos[[column_name]], na.rm = TRUE)
  stdv_value <- sd(reldif_wowi_naomit_onlypos[[column_name]], na.rm = TRUE)

  statistics_matrix_wowi_onlypos[i, ] <- c(mean_value, median_value, variance_value, stdv_value)
}
print(statistics_matrix_wowi_onlypos)

# Box plots
# relative differences simulations observations

reldif_simobs_naomit_bp <- reldif_simobs_naomit[, !(names(reldif_simobs_naomit) %in% c("ID", "XCENT", "YCENT"))]
percentiles <- apply(reldif_simobs_naomit_bp, 2, quantile, probs = c(0.25, 0.5, 0.75))


percentiles_df <- data.frame(t(percentiles))
colnames(percentiles_df) <- c("P25", "P50", "P75")

percentiles_df <- percentiles_df %>%
  mutate(percentile_type = rownames(percentiles_df))

percentiles_long <- melt(percentiles_df, id.vars = "percentile_type")
reldif_simobs_naomit_bp_long <- melt(reldif_simobs_naomit_bp, id.vars = NULL, variable.name = "Periodo_di_Ritorno", value.name = "Valore")

bp <- ggplot(reldif_simobs_naomit_bp_long, aes(x = Periodo_di_Ritorno, y = Valore)) +
  geom_boxplot(color = "blue") +
  geom_hline(yintercept = 0, color = "black", linewidth = 1.2) +
  geom_point(data = percentiles_long, aes(x = percentile_type, y = value, color = percentile_type), size = 3) +
  labs(title = "Flood Quantiles Relative Differences (sim-obs)/obs",
       y = "Relative Differences",
       x = "Return Periods") +
  coord_cartesian(ylim = c(-1, 5)) +
  theme_minimal() +
  theme(legend.position = "none")

# table of percentiles

percentiles_df_simobs <- percentiles_df
percentiles_df_simobs <- percentiles_df_simobs %>%
  select(-percentile_type)

# counting how many in each category

columns_of_interest <- c("D.HQ2", "D.HQ5", "D.HQ10", "D.HQ20", "D.HQ50", "D.HQ100", "D.HQ500")
count_matrix_simobs <- matrix(0, nrow = length(columns_of_interest), ncol = 5)
rownames(count_matrix_simobs) <- columns_of_interest
colnames(count_matrix_simobs) <- c("RelDiff < -1", "-1 < RelDiff < -0.25", "-0.25 < RelDiff < 0.25", "0.25 < RelDiff < 1", "RelDiff > 1")
```

```
for (col in columns_of_interest) {
  count_matrix_simobs[col, "RelDiff < -1"] <- sum(reldif_simobs_naomit_bp[[col]] <= -1, na.rm = TRUE)
  count_matrix_simobs[col, "-1 < RelDiff < -0.25"] <- sum(reldif_simobs_naomit_bp[[col]] > -1 & reldif_simobs_naomit_bp[[col]] <= -0.25, na.rm = TRUE)
  count_matrix_simobs[col, "-0.25 < RelDiff < 0.25"] <- sum(reldif_simobs_naomit_bp[[col]] > -0.25 & reldif_simobs_naomit_bp[[col]] <= 0.25, na.rm = TRUE)
  count_matrix_simobs[col, "0.25 < RelDiff < 1"] <- sum(reldif_simobs_naomit_bp[[col]] > 0.25 & reldif_simobs_naomit_bp[[col]] <= 1, na.rm = TRUE)
  count_matrix_simobs[col, "RelDiff > 1"] <- sum(reldif_simobs_naomit_bp[[col]] > 1, na.rm = TRUE)
}


# relative differences without with

reldif_wowi_naomit_bp <- reldif_wowi_naomit[, !(names(reldif_wowi_naomit) %in% c("ID", "XCENT", "YCENT"))]
percentiles <- apply(reldif_wowi_naomit_bp, 2, quantile, probs = c(0.25, 0.5, 0.75))


percentiles_df <- data.frame(t(percentiles))
colnames(percentiles_df) <- c("P25", "P50", "P75")

percentiles_df <- percentiles_df %>%
  mutate(percentile_type = rownames(percentiles_df))

percentiles_long <- melt(percentiles_df, id.vars = "percentile_type")
reldif_wowi_naomit_bp_long <- melt(reldif_wowi_naomit_bp, id.vars = NULL, variable.name = "Periodo_di_Ritorno", value.name = "Valore")

bp <- ggplot(reldif_wowi_naomit_bp_long, aes(x = Periodo_di_Ritorno, y = Valore)) +
  geom_boxplot(color = "blue") +
  geom_hline(yintercept = 0, color = "black", linewidth = 1.2) +
  geom_point(data = percentiles_long, aes(x = percentile_type, y = value, color = percentile_type), size = 3) +
  labs(title = "Flood Quantiles Relative Differences (without-with)/with",
       y = "Relative Differences",
       x = "Return Periods") +
  coord_cartesian(ylim = c(-0.5, 0.5)) +
  theme_minimal() +
  theme(legend.position = "none")

# table of percentiles

percentiles_df_wowi <- percentiles_df
percentiles_df_wowi <- percentiles_df_wowi %>%
  select(-percentile_type)

# counting how many in each category

columns_of_interest <- c("D.HQ2.res", "D.HQ5.res", "D.HQ10.res", "D.HQ20.res", "D.HQ50.res", "D.HQ100.res", "D.HQ500.res")

count_matrix_wowi <- matrix(0, nrow = length(columns_of_interest), ncol = 4)
rownames(count_matrix_wowi) <- columns_of_interest
colnames(count_matrix_wowi) <- c("Reversed", "Null", "Light", "Strong")

for (col in columns_of_interest) {
  count_matrix_wowi[col, "Reversed"] <- sum(reldif_wowi_naomit_bp[[col]] < 0, na.rm = TRUE)
  count_matrix_wowi[col, "Null"] <- sum(reldif_wowi_naomit_bp[[col]] == 0, na.rm = TRUE)
  count_matrix_wowi[col, "Light"] <- sum(reldif_wowi_naomit_bp[[col]] > 0 & reldif_wowi_naomit_bp[[col]] <= 1, na.rm = TRUE)
  count_matrix_wowi[col, "Strong"] <- sum(reldif_wowi_naomit_bp[[col]] > 1, na.rm = TRUE)
}

# Number of negative (Reversed), 0 (Null), between 0 and 1 (Light), >1 (Strong) catchments foe every return period
print(count_matrix_wowi)

# IDs che producono valori eccessivamente elevati

columns_of_interest <- c("D.HQ2.res", "D.HQ5.res", "D.HQ10.res", "D.HQ20.res", "D.HQ50.res", "D.HQ100.res", "D.HQ500.res")


id_list_wowi <- list()

count_matrix_wowi <- matrix(0, nrow = length(columns_of_interest), ncol = 5)
rownames(count_matrix_wowi) <- columns_of_interest
colnames(count_matrix_wowi) <- c("Reversed", "Null", "Light", "Strong", "GreaterThan10")

for (col in columns_of_interest) {
  reversed_ids <- reldif_wowi_naomit$ID[reldif_wowi_naomit[[col]] < 0 & !is.na(reldif_wowi_naomit[[col]])]
  null_ids <- reldif_wowi_naomit$ID[reldif_wowi_naomit[[col]] == 0 & !is.na(reldif_wowi_naomit[[col]])]
  light_ids <- reldif_wowi_naomit$ID[reldif_wowi_naomit[[col]] > 0 & reldif_wowi_naomit[[col]] <= 1 & !is.na(reldif_wowi_naomit[[col]])]
  strong_ids <- reldif_wowi_naomit$ID[reldif_wowi_naomit[[col]] > 1 & !is.na(reldif_wowi_naomit[[col]])]
```

```
  greater_than_10_ids <- reldif_wowi_naomit$ID[reldif_wowi_naomit[[col]] > 10 & !is.na(reldif_wowi_naomit[[col]])]

  count_matrix_wowi[col, "Reversed"] <- length(reversed_ids)
  count_matrix_wowi[col, "Null"] <- length(null_ids)
  count_matrix_wowi[col, "Light"] <- length(light_ids)
  count_matrix_wowi[col, "Strong"] <- length(strong_ids)
  count_matrix_wowi[col, "GreaterThan10"] <- length(greater_than_10_ids)

  id_list_wowi[[col]] <- list(
    Reversed = reversed_ids,
    Null = null_ids,
    Light = light_ids,
    Strong = strong_ids,
    GreaterThan10 = greater_than_10_ids
  )
}

result_list <- lapply(columns_of_interest, function(col) {
  data.frame(Period = col, IDs = id_list_wowi[[col]]$GreaterThan10)
})

result_df <- do.call(rbind, result_list)

final_table_grouped <- result_df %>%
  group_by(IDs) %>%
  summarise(Period = paste(unique(Period), collapse = ", "))

ids_to_remove <- final_table_grouped$IDs

reldif_simobs_naomit_formaps <- reldif_simobs_naomit[!reldif_simobs_naomit$ID %in% ids_to_remove, ]

reldif_wowi_naomit_formaps <- reldif_wowi_naomit[!reldif_wowi_naomit$ID %in% ids_to_remove, ]
# eliminazione anche dei valori negativi
reldif_wowi_naomit_formaps_onlypos <- reldif_wowi_naomit_formaps[rowSums(reldif_wowi_naomit_formaps < 0) == 0, ]

# salvataggio come csv per plottare su QGIS

# file_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/reldif_simobs_filtered.csv"
# write.csv(reldif_simobs_naomit_formaps, file = file_path, row.names = FALSE)
#
# file_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/reldif_wowi_filtered.csv"
# write.csv(reldif_wowi_naomit_formaps_onlypos, file = file_path, row.names = FALSE)


# creazione di un nuovo dataframe che ha una colonna con i valori della frazione sums.dams.area/AREA. Questi valori dovrebbero essere tutti tra 0 e 1, ma
alcuni sono maggiori di 1( probabilmente è un problema di geolocalizzazione delle dighe).

cat_table_ratio_info <- data.frame(
  ID = cat_table_results$ID,
  XCENT = cat_table_results$XCENT,
  YCENT = cat_table_results$YCENT,
  ratio_area = cat_table_results$sum.dams.area / cat_table_results$AREA
)

# rimozione valori maggiori di 1
cat_table_ratio_info_valid <- cat_table_ratio_info[cat_table_ratio_info$ratio_area <= 1, ]

# esportazione come file csv
# file_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo
anno/tesi/Halle/writing/luglio/agosto/results_aug/cat_table_ratio_info_valid.csv"
# write.csv(cat_table_ratio_info_valid, file = file_path, row.names = FALSE)

# per mappa QGIS: ovviamente sono stati eliminati molti catchments che avevano i valori di ratio > 1

# library(sf)

# # Carica lo shapefile catchment come dataframe
# catchment <- st_read("C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo
anno/tesi/Halle/writing/luglio/agosto/shp_files/catchments_RR2.shp")
# catchment <- rename(catchment, ID = GaugeID)
#
# catchment_area_ratio <- merge(catchment, cat_table_ratio_info_valid[, c("ID", "ratio_area")], by = "ID", all.x = TRUE)
# st_write(catchment_area_ratio, "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo
anno/tesi/Halle/writing/luglio/agosto/shp_files/to_serena/catchment_area_ratio_RR2.shp", append = TRUE)
```

```
library(plotly)
library(dplyr)

# Funzione per assegnare colori in base all'incremento percentuale del runoff (asse Z)
color_by_increment <- function(increment) {
  if (increment <= 0.2) return('blue')
  else if (increment <= 0.4) return('green')
  else if (increment <= 0.6) return('yellow')
  else if (increment <= 0.8) return('orange')
  else if (increment <= 1) return('red')
  else return('purple')  # Colore per incrementi > 100%
}

# Creare il grafico a barre 3D
fig <- plot_ly()

# Aggiungere le barre 3D al grafico
for (i in unique(mean_area_long$Classe)) {
  for (j in unique(mean_area_long$Intervallo)) {
    # Filtrare i dati per la combinazione corrente di Classe e Intervallo
    subset_data <- mean_area_long %>%
      filter(Classe == i, Intervallo == j)

    # Assegnare le variabili per ogni asse
    x_value <- as.character(i)  # L'asse X rappresenta D.HQ*.res
    y_value <- subset_data$MediaAree  # L'asse Y rappresenta le aree del bacino
    z_value <- as.numeric(gsub(".*-(.*)", "\\1", j))  # Estrae il valore numerico dalle percentuali

    # Verifica che ci siano valori validi
    if (!is.na(y_value) && length(y_value) > 0 && !is.na(z_value)) {
      # Usare il colore basato sull'incremento percentuale del runoff
      color <- color_by_increment(z_value)

      # Aggiungere la traccia per questa combinazione usando scatter3d per simulare le barre
      fig <- fig %>%
        add_trace(
          x = rep(x_value, 2),  # X: D.HQ*.res
          y = rep(y_value, 2),  # Y: Area del bacino
          z = c(0, z_value),    # Z: Incremento di runoff (%)
          type = 'scatter3d',
          mode = 'lines',
          line = list(color = color, width = 10)  # Simula la barra
        )
    }
  }
}

# Layout per il grafico con l'asse X per i periodi di ritorno, Y per le aree, e Z per le percentuali
fig <- fig %>%
  layout(scene = list(
    xaxis = list(title = 'Return Period'),
    yaxis = list(title = 'Catchment Area [Km^2]'),
    zaxis = list(
      title = '% of Increment of Runoff',
      tickvals = c(0.2, 0.4, 0.6, 0.8, 1, 1.2),
      ticktext = c('0%-20%', '20%-40%', '40%-60%', '60%-80%', '80%-100%', '> 100%')
    )
  ),
  title = "3D Bar Plot of Runoff Increment by Catchment Area and Return Period",
  showlegend = FALSE)  # Rimuove la legenda

# dataframe per grafici sulle efficienze

selected_columns <- c("ID", "XCENT", "YCENT", "AREA", "ME", "MEreg", "MEreg.wores", "sum.dams.area")
efficiencies_df <- cat_table_results[selected_columns]

efficiencies_df$ratio_area <- efficiencies_df$sum.dams.area / efficiencies_df$AREA

ME_df <- efficiencies_df[!is.na(efficiencies_df$ME), ]
MEreg_df <- efficiencies_df[!is.na(efficiencies_df$MEreg), ]

MEreg_df_onlypos <- MEreg_df[MEreg_df$MEreg > 0, ]
```

```
# library(dplyr)
ME_df_loweff <- ME_df[ME_df$ME < 0.5, ]
# ME_df_higheff <- anti_join(ME_df, ME_df_loweff, by = "ID")

reldif_simobs_naomit_higheff <- anti_join(reldif_simobs_naomit, ME_df_loweff, by = "ID")
reldif_wowi_naomit_higheff <- anti_join(reldif_wowi_naomit, ME_df_loweff, by = "ID")

MEreg_df_onlypos_loweff <- MEreg_df_onlypos[MEreg_df_onlypos$MEreg < 0.5, ]

reldif_simobs_naomit_higheff <- anti_join(reldif_simobs_naomit, MEreg_df_onlypos_loweff, by = "ID")
reldif_wowi_naomit_higheff <- anti_join(reldif_wowi_naomit, MEreg_df_onlypos_loweff, by = "ID")

# plot the variation of ME or MEreg and AREA

library(ggplot2)
library(gridExtra)

mean_ME <- mean(ME_df$ME)
mean_MEreg <- mean(MEreg_df_onlypos$MEreg)
mean_MEregwores <- mean(MEreg_df_onlypos$MEreg.wores)

# ME vs AREA

# ggplot(ME_df, aes(x = AREA, y = ME, color = ME)) +
#   geom_point(shape = 16, size = 2, alpha = 1) +
#   geom_hline(yintercept = mean_ME, linetype = "dashed", color = "black", size = 1, alpha = 0.5) +
#   labs(title = "Relationship between Model Efficiency and Catchment Area",
#        x = "Area [km^2]",
#        y = "ME") +
#   scale_color_gradient(low = "turquoise", high = "blue") +
#   scale_y_log10() +
#   theme_minimal()

ggplot(ME_df, aes(x = AREA, y = ME, color = ME)) +
  geom_point(shape = 16, size = 2, alpha = 1) +
  geom_hline(yintercept = mean_ME, linetype = "dashed", color = "black", size = 1, alpha = 0.5) +
  labs(title = "Relationship between Model Efficiency and Catchment Area",
       x = "Area [km^2]",
       y = "ME") +
  scale_color_gradient(low = "turquoise", high = "blue") +
  scale_x_log10() +
  theme_minimal()

ggplot(subset(ME_df, ME > 0.5), aes(x = AREA, y = ME, color = ME)) +
  geom_point(shape = 16, size = 2, alpha = 1) +
  geom_hline(yintercept = mean(ME_df$ME, na.rm = TRUE), linetype = "dashed", color = "black", size = 1, alpha = 0.5) +
  labs(title = "Relationship between Model Efficiency and Catchment Area (high ME)",
       x = "Area [km^2]",
       y = "ME") +
  scale_color_gradient(low = "turquoise", high = "blue") +
  scale_x_log10() +
  theme_minimal()

# MEreg vs AREA

ggplot(MEreg_df_onlypos, aes(x = AREA, y = MEreg, color = MEreg)) +
  geom_point(shape = 16, size = 2, alpha = 1) +
  geom_hline(yintercept = mean_MEreg, linetype = "dashed", color = "black", size = 1, alpha = 0.5) +
  labs(title = "Relationship between Model Efficiency after Regional Calibration and Catchment Area",
       x = "Area [km^2]",
       y = "ME (after regional calibration)") +
  scale_color_gradient(low = "turquoise", high = "blue") +
  scale_x_log10() +
  theme_minimal()

ME_MEreg_comparison <- MEreg_df_onlypos[!is.na(MEreg_df_onlypos$ME), ]

ggplot(ME_MEreg_comparison, aes(x = ME, y = MEreg)) +
  geom_point(color = "blue") +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
  labs(x = "ME", y = "MEreg") +
  ggtitle("Efficiency Comparison") +
  xlim(0, 1) +
  ylim(0, 1) +
  theme_minimal()
```

```
# confronto MEreg e MEreg.wores

MEreg_df_onlypos_valid <- MEreg_df_onlypos[MEreg_df_onlypos$ratio_area <= 1, ] # tolgo i casi in cui la frazione è maggiore di 1

ggplot(MEreg_df_onlypos_valid, aes(x = MEreg, y = MEreg.wores, color = ratio_area)) +
 geom_point() +
 geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
 labs(x = "MEreg", y = "MEreg.wores") +
 ggtitle("Efficiency With and Without Dams") +
 xlim(0, 1) +
 ylim(0, 1) +
 theme_minimal() +
 scale_color_gradient(low = "turquoise", high = "blue")  # in base a ratio_area in modo continuo

# dividendo in tre classi (low, medium e high) in base a ratio_area

ggplot(MEreg_df_onlypos_valid, aes(x = MEreg, y = MEreg.wores, color = ratio_area_class)) +
 geom_point() +
 geom_abline(intercept = 0, slope = 1, color = "grey", linetype = "dashed") +
 labs(x = "MEreg", y = "MEreg.wores", color = "% area due to dams") +
 ggtitle("Efficiency With and Without Dams") +
 xlim(0, 1) +
 ylim(0, 1) +
 theme_minimal() +
 scale_color_manual(values = c("Low" = "khaki", "Medium" = "palegreen", "High" = "seagreen"))

# CDF di ME e MEreg

par(mar = c(1, 1, 1, 1) + 0.1)
cdf_ME <- ecdf(ME_df$ME)
cdf_MEreg <- ecdf(MEreg_df_onlypos$MEreg)

plot(cdf_ME, main = "Cumulative Distribution Function", xlab = "Efficiency", ylab = "CDF", col = "blue")
lines(cdf_MEreg, col = "red")
abline(h = 0.5, col = "black", lty = 2, lwd = 1.5)
legend(x = 0.85, y = 0.4, legend = c("ME", "MEreg"), col = c("blue", "red"), lty = 3, cex = 0.35)

# trasformo i dataframe in file csv per poter plottare su QGIS
path_to_file <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/ME_df.csv"
write.csv(ME_df, file = path_to_file, row.names = FALSE)
path_to_file <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/MEreg_df.csv"
write.csv(MEreg_df_onlypos, file = path_to_file, row.names = FALSE)

# Flow Duration Curve
directory <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/PASS_flood_RM/SALTO_out"
file_names <- list.files(directory, pattern = "\\.txt$", full.names = TRUE)

create_fdc_plot <- function(data, file_name, log_scale = TRUE) {
 # Convert the 'date' column to Date class
 data$date <- as.Date(data$date, format = "%Y-%m-%d")

 # Extract data for the year 2000
 data_2000 <- subset(data, format(data$date, "%Y") == "2000")

 if (nrow(data_2000) == 0) {
  warning("No data available for the year 2000 in ", file_name)
  return(NULL)  # Return NULL if no data available
 }

 data_2000_qsim <- data_2000[order(data_2000$qsim, decreasing = TRUE), ]
 data_2000_qobs <- data_2000[order(data_2000$qobs, decreasing = TRUE), ]
 data_2000_qsim_withoutdams <- data_2000[order(data_2000$qsim.withoutdams, decreasing = TRUE), ]

 percentile <- seq(1, 100, by = 1)
 # Calculate the flow corresponding to each percentile for qsim
 flow_qsim <- quantile(data_2000_qsim$qsim, probs = percentile / 100, na.rm = TRUE)
 fdc_data_qsim <- data.frame(percentile = percentile, flow = rev(flow_qsim))

 # Calculate the flow corresponding to each percentile for qobs
 flow_qobs <- quantile(data_2000_qobs$qobs, probs = percentile / 100, na.rm = TRUE)
 fdc_data_qobs <- data.frame(percentile = percentile, flow = rev(flow_qobs))

 # Calculate the flow corresponding to each percentile for qsim.withoutdams
 flow_qsim_withoutdams <- quantile(data_2000_qsim_withoutdams$qsim.withoutdams, probs = percentile / 100, na.rm = TRUE)
 fdc_data_qsim_withoutdams <- data.frame(percentile = percentile, flow = rev(flow_qsim_withoutdams))
```

```
  # Determine y-axis scale
  y_scale <- if (log_scale) scale_y_continuous(trans = "log10") else scale_y_continuous()

  # Plot the Flow Duration Curves (FDC) for qsim, qobs, and qsim.withoutdams
  plot_qsim <- ggplot(fdc_data_qsim, aes(x = percentile, y = flow)) +
    geom_line(color = "blue") +
    labs(title = "Flow Duration Curve (FDC)", x = "Percentage of Time", y = "Runoff (qsim)") +
    scale_x_continuous() +
    y_scale +
    theme_minimal()

  plot_qobs <- ggplot(fdc_data_qobs, aes(x = percentile, y = flow)) +
    geom_line(color = "red") +
    labs(title = "Flow Duration Curve (FDC)", x = "Percentage of Time", y = "Runoff (qobs)") +
    scale_x_continuous() +
    y_scale +
    theme_minimal()

  plot_qsim_withoutdams <- ggplot(fdc_data_qsim_withoutdams, aes(x = percentile, y = flow)) +
    geom_line(color = "gold") +
    labs(title = "Flow Duration Curve (FDC)", x = "Percentage of Time", y = "Runoff (qsim.withoutdams)") +
    scale_x_continuous() +
    y_scale +
    theme_minimal()

  # Combine plots into a single PDF page with a title, arranged vertically
  combined_plot <- grid.arrange(
    grobs = list(
      textGrob(label = paste("Flow Duration Curves -", file_name),
               gp = gpar(fontsize = 16, fontface = "bold")),
      arrangeGrob(plot_qobs, plot_qsim, plot_qsim_withoutdams, ncol = 1)
    ),
    ncol = 1,
    heights = c(0.1, 0.9)
  )

  return(combined_plot)
}
# Create PDF for linear-scale plots in portrait format
pdf("C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/FDC_Lin_Official_2.pdf", width = 8.5,
height = 14)

# Loop through each file (reuse the same file_names and create_fdc_plot function)
for (file in file_names) {
  # Read the text file
  data <- read.table(file, header = TRUE)

  # Generate FDC plot
  plot <- create_fdc_plot(data, basename(file), log_scale = FALSE)

  if (!is.null(plot)) {
    # Print the plot to PDF
    print(plot)
  }
}

# Close the PDF device for linear-scale plots
dev.off(dev.prev())
```

 # comparison of timeseries qobs, qsim, qsim.withoutdams

```
dir_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/PASS_flood_RM/SALTO_out"
output_file_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo
anno/tesi/Halle/writing/luglio/agosto/results_aug/qobs_qsim_withwithout_plots_2000_orizz.pdf"

file_list <- list.files(path = dir_path, pattern = "*.txt", full.names = TRUE)

pdf(output_file_path, width = 11, height = 8.5, paper = "a4r")

for (file_path in file_list) {

data <- read.csv(file_path, sep = "", header = TRUE, stringsAsFactors = FALSE)

data_selected <- data %>%
select(date, qobs, qsim, qsim.withoutdams)
```

```
  data_selected$date <- as.Date(data_selected$date, format = "%Y-%m-%d")

  data_filtered <- data_selected %>%
    filter(date >= as.Date("2000-01-01") & date <= as.Date("2000-12-31"))

  # Create the plot
  plot <- ggplot(data_filtered, aes(x = date)) +
    geom_line(aes(y = qobs, color = "qobs")) +  # Plot qobs
    geom_line(aes(y = qsim, color = "qsim")) +
    geom_line(aes(y = qsim.withoutdams, color = "qsim.withoutdams")) +
    labs(title = paste("Qsim, Qsim.withoutdams, and Qobs for file", basename(file_path)),
         x = "Date",
         y = "Runoff [mm/day]") +
    scale_color_manual("",
                       breaks = c("qobs", "qsim", "qsim.withoutdams"),
                       values = c("qobs" = "red", "qsim" = "blue", "qsim.withoutdams" = "gold"))

  # Print the plot to the PDF device
  print(plot)
}

# Close the PDF device
dev.off(dev.prev())

# graphs of prec, swe aet, sm, gwr gw_s, qsim qobs qsim.withoutdams

dir_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/PASS_flood_RM/SALTO_out"
output_file_path <- "C:/Users/seren/Desktop/uni/POLI/Magistrale/Ambientale/secondo anno/tesi/Halle/writing/luglio/agosto/results_aug/total_plots.pdf"
file_list <- list.files(path = dir_path, pattern = "*.txt", full.names = TRUE)
pdf(output_file_path, width = 8.5, height = 11)  # Paper size A4 (vertical)

for (file_path in file_list) {

  data <- read.csv(file_path, sep = "", header = TRUE, stringsAsFactors = FALSE)
  data_selected <- data %>%
    select(date, prec, swe, aet, sm, gwr, gw_s, qobs, qsim, qsim.withoutdams)
  data_selected$date <- as.Date(data_selected$date, format = "%Y-%m-%d")
  data_filtered <- data_selected %>%
    filter(date >= as.Date("2000-01-01") & date <= as.Date("2000-12-31"))
  plot_prec <- ggplot(data_filtered, aes(x = date, y = prec)) +
    geom_line(color = "darkgreen") +
    labs(title = "Precipitation", x = "Date", y = "prec (mm)")
  plot_swe_aet <- ggplot(data_filtered, aes(x = date)) +
    geom_line(aes(y = swe, color = "swe")) +
    geom_line(aes(y = aet, color = "aet")) +
    labs(title = "Snow Water Equivalent and Actual Evapotranspiration", x = "Date", y = "Values (mm)") +
    scale_color_manual("Variables", breaks = c("swe", "aet"), values = c("swe" = "purple", "aet" = "orange"))

  plot_sm <- ggplot(data_filtered, aes(x = date, y = sm)) +
    geom_line(color = "blue") +
    labs(title = "Soil Moisture", x = "Date", y = "sm (mm)")

  plot_gwr_gws <- ggplot(data_filtered, aes(x = date)) +
    geom_line(aes(y = gwr, color = "gwr")) +
    geom_line(aes(y = gw_s, color = "gw_s")) +
    labs(title = "Groundwater Recharge and Storage", x = "Date", y = "gwr / gw_s (mm)") +
    scale_color_manual("Variables", values = c("gwr" = "brown", "gw_s" = "cyan"))

  plot_q <- ggplot(data_filtered, aes(x = date)) +
    geom_line(aes(y = qobs, color = "qobs")) +
    geom_line(aes(y = qsim, color = "qsim")) +
    geom_line(aes(y = qsim.withoutdams, color = "qsim.withoutdams")) +
    labs(title = "Streamflow: Qobs, Qsim, Qsim without dams", x = "Date", y = "Flow (m³/s)") +
    scale_color_manual("Variables", breaks = c("qobs", "qsim", "qsim.withoutdams"),
                       values = c("qobs" = "red", "qsim" = "blue", "qsim.withoutdams" = "gold"))

  # Combine the plots in a single vertical page (5 graphs per page)
  grid.arrange(plot_prec, plot_swe_aet, plot_sm, plot_gwr_gws, plot_q, ncol = 1)

}

# Close the PDF device
dev.off()
```