The impact of reservoirs on hydrological model performance in Brazilian catchments

S.C.H. van Langen

March 2021

MSc thesis
Hydrology and Quantitative Water Management Group
Wageningen University

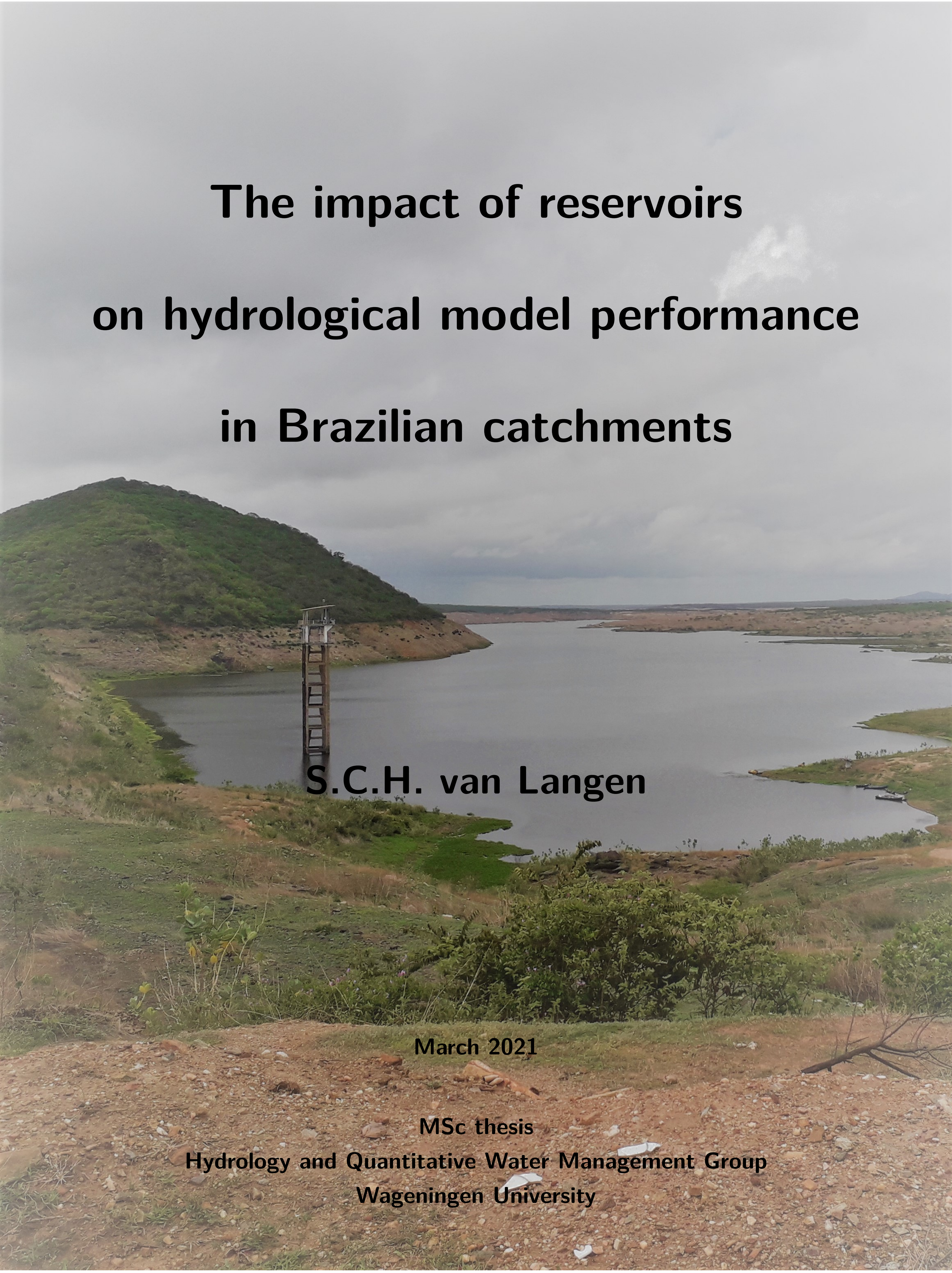


Photo of the Banabuiú reservoir in the northeast of Brazil taken by the author (2020)

Abstract

Many reservoirs have been constructed in Brazil during the past decades. These reservoirs are often not included in hydrological models, while they have a large impact on catchment hydrology. This study aimed to investigate the effects of including reservoirs in two hydrological models on model performance, measured by the Kling Gupta Efficiency (KGE). 403 catchments across Brazil were modeled using the HBV and GR4J models. Two scenarios, with and without reservoir, were simulated and compared to each other. For the HBV model, a significant increase in model performance was found when the reservoirs were included in the model, but overall performance was poor. The mean KGE increased from 0.21 to 0.40 when reservoirs were added. The GR4J model, on the other hand, showed better overall performance, but without the improvement when including reservoirs. Here, the mean KGEs were 0.57 without and 0.56 with reservoirs. In the catchments with the largest/most reservoirs, the HBV reservoir scenario outperformed both GR4J scenarios. The results are promising, because they show that model performance can increase when the reservoirs are included. Better model performance can still be obtained with a smaller spatial scale or other methods of including reservoirs, which might require more data.

Contents

1 | Introduction

Catchments around the world are being modeled for a variety of purposes like improving water management, forecasting hydrological extremes and understanding hydrological processes in the catchment (Blöschl, 2006). Many different hydrological models are available. The suitable model to use depends on local circumstances in the catchment and the aim of the modeling exercise (Holländer et al., 2014).

Models are simplifications of reality and therefore always come with uncertainties. To be able to make predictions, it is important to limit these uncertainties. This can be done in several ways, for example by using the right model that can simulate the important processes, calibrating the model to local circumstances and providing enough data with a good quality as input for the model. However, some scholars also consider it important that models do not become needlessly complex and over-parameterized (Perrin et al., 2001; Seibert et al., 2019; Whittaker et al., 2010). This could especially cause problems when data availability is limited. On the other hand, if enough detailed data is available, more complex models can have a higher fidelity. A high model fidelity means that simulations give a more realistic representation of processes in the real world and thus the results correspond well with reality (high model performance), or as Kirchner (2006, p.1) states, to "get the right answers for the right reasons".

The most important natural processes are already included in hydrological models and efforts are made to increase their realism (Clark et al., 2011). However, there is something else going on in many catchments that is often not included in hydrological models. People are interfering with natural hydrological systems, for example by abstracting water from various sources and building dams. De Graaf et al. (2019) found that unsustainable groundwater pumping for irrigation purposes is occurring around the world, depleting this water resource quickly. Ye et al. (2003) studied the Lena river catchment in the Arctic, where a hydropower dam was constructed. Because of this dam, the seasonality of the streamflow changed, with increasing low-flows and decreasing high-flows. These are two examples of how people are changing the natural hydrological system. Human interference in catchments can thus cause significant changes in streamflow (Van Loon et al., 2019; Wada et al., 2017; Wanders and Wada, 2015; Woo et al., 2008). This development causes an increasing interest in including such processes in hydrological studies. New concepts have been introduced like socio-hydrology (Sivapalan et al., 2012) and water science in the antropocene (Savenije et al., 2014; Van Loon et al., 2016). Furthermore, there is an increasing interest in incorporating this human interference into hydrological models to increase model fidelity. This is not an easy task, since there are many challenges, including (but not limited to) how to incorporate human influences in models and data availability regarding water management (Wada et al., 2017; Zhou et al., 2016). Because of these challenges, improved model realism does not always lead to improved model performance (DelSole and Shukla, 2010).

Over the past decades, many dams have been constructed in Brazil (Cavalcante et al., 2020; Souza Filho, 2009). These dams can be used for flow regulation, providing water during dry periods in the semi-arid northeast of Brazil (Braga et al., 2012). Another purpose of many dams in the rest of the country is hydro-power production, which is an important energy source for Brazil (Braga et al., 2012). Several studies have found that these dams have a significant impact on downstream hydrology (e.g. Almeida et al., 2020; Cavalcante et al., 2020; Dantas et al., 2020; Fantin-Cruz et al., 2015; Souza Filho, 2009). Therefore, this process seems important to be included in hydrological modeling of Brazilian catchments.

The new CAMELS-BR (Catchment Attributes and MEteorology for Large-sample Studies - Brazil) data set, introduced by Chagas et al. (2020), contains information on total reservoir capacity in Brazilian catchments next to other relevant data used in hydrological modeling. According to Wanders and Wada (2015), this human factor has an important impact on hydrology. Consequently, this data set provides a great opportunity to investigate how reservoirs impact hydrological model performance in a large-scale modeling exercise.

The aim of this study was to investigate the effect of increasing hydrological model realism by including reservoirs on model performance across catchments in Brazil. To reach this aim, 403 Brazilian catchments were modelled with two commonly used hydrological models. The model performance was compared between two scenarios, one with and one without reservoirs. This made it possible to study the effect of including reservoirs on model performance for different model structures across a variety of catchments with different characteristics.

2 Methods

2.1 Study area and data

The study area consists of 403 catchments across Brazil, as shown in Figure 2.1. These 403 catchments were selected from the CAMELS-BR data set (Chagas et al., 2020), only including catchments with reservoirs. This was done by selecting the catchments with a total reservoir capacity that is not equal to zero. The white parts in Figure 2.1 were either not included in the data set or excluded because there is no reservoir there. The colors of the catchments show the total reservoir capacity relative to annual streamflow per catchment. Brazil is an interesting study area to investigate reservoirs, because of the large number of them (thousands, although the exact number is unknown (Mulligan et al., 2020)). Therefore, reservoirs are likely to have a large impact on the hydrological system. They are used for the purpose of water availability in the dry season or hydro-power production (Braga et al., 2012). The large size of Brazil allows this study to consider a great variety in catchments in for example size, climate, topography and land use. Elevations of the land surface of Brazil vary from around sea level in the northwest to approximately 2000 m above sea level in the southeast. The land is covered mostly by forests in the northwest of Brazil, while cropland and shrubs are most common in the rest of the country. The northwest region receives most precipitation, up to 3000 mm/year. The least precipitation is found in the semi-arid northeast with 400-800 mm/year. In the south, the annual precipitation is around 1000-2000 mm. Average annual temperatures are high in general, ranging from 20°C in the south to 30°C in the north (FAO, 2021). Because of these large variations, the findings of this study are widely applicable.

The CAMELS-BR data set provides catchment properties and daily forcing data from 1980 to 2018. However, the streamflow time-series are smaller for some catchments. Data was used from 1990 to 2008, which made it possible to include all catchments with reservoirs. This period was still long enough for proper calibration and validation of the model. It should be noted that in the CAMELS-BR data set, only the total reservoir capacity of the catchment is given. This is only one fixed value, instead of a time series of reservoir volumes, for example. This study also aimed to assess whether this information is enough for including the reservoirs in a hydrological model or more information should be gathered for that purpose. Next to the reservoir capacities, the CAMELS-BR data about consumptive water use are available and may be included in the modeling structures as an extra outflow of water. However, this outflow was small compared to other outflows. Its influence on the model performance was negligible and therefore, it was not included in this study.

The other data that were necessary for modeling the catchments included time-series with a daily timescale of precipitation (P), potential evaporation (PE, which also includes transpiration), minimum and maximum temperatures (T_{min} and T_{max}) and other relevant catchment characteristics (e.g. soil, land use and topography). The CAMELS-BR data set contains different types of data for precipitation (Climate Prediction Center (CPC), Multi-Source Weighted-Ensemble Precipitation (MSWEP) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)). These data are all similar, but with different collection methods as well as spatial and temporal scales. CHIRPS precipitation have been used in this study. These data have the highest spatial

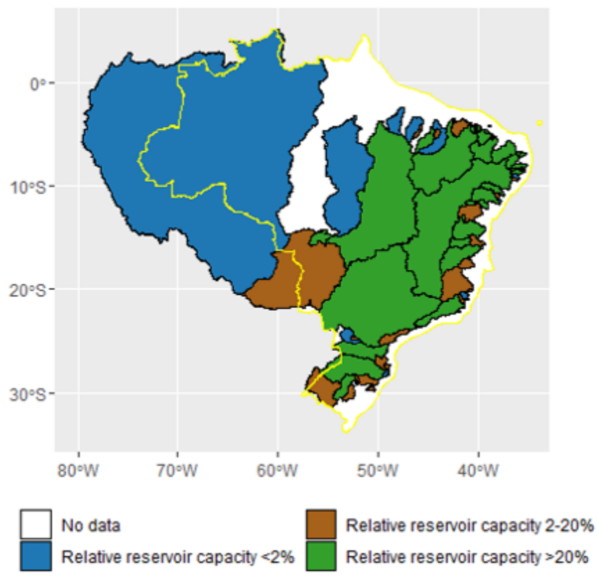


Figure 2.1: Catchments in the study area, with different colors showing the reservoir capacity as percentage of the total annual streamflow. The boundary of Brazil is shown in yellow.

resolution (0.05°) and are based on a combination of gauge data and remote sensing (Chagas et al., 2020). Therefore, it is assumed that these data are most realistic. Luo et al. (2019) also show that CHIRPS precipitation data perform well.

2.2 Modeling

In this section, the modeling methods for this study are explained and an overview is given in Figure 2.2. The two selected models are described, followed by the modeling scenarios as well as the calibration and validation methods and data analysis.

To reach the objective of this study, two hydrological modeling structures were compared, using the RAVEN modular modeling framework (Craig et al., 2020). This is a flexible framework, which allows many different algorithms to be used for different parts of the water cycle as well as usage of a variety of routing mechanisms. Several hydrological modeling structures can be reproduced nearly exact: UBCWM (Quick, 1995), HBV-EC model (Bergström, 1995), HMETS (Martel et al., 2017), MOHYSE (Fortin and Turcotte, 2006), and GR4J (Perrin et al., 2003). This framework was chosen because it includes some modules that allow modeling of human interference. It can thus be adapted easily to include reservoirs.

2.2.1 Models

The two models that were selected in this study are the RAVEN interpretations of HBV-EC and GR4J. HBV-EC has a slightly more complex model structure than GR4J, but both are relatively simple and widely used. These models have been used in many previous studies with good performance (e.g. Engeland and Hisdal, 2009; Payan et al., 2008; Unduche et al., 2018). The structures of both models as implemented in RAVEN are shown in Figure A.1 and A.2. An overview of both models is given in Table 2.1.

HBV-EC is a Canadian version of the HBV (Hydrologiska Byråns Vattenbalansavdelning) model (Bergström, 1995; Lindström et al., 1997) and is referred to as the HBV model in this study. It is a partially distributed conceptual model with sixteen parameters, but in this study it was used as a lumped model, by using only one "sub"-catchment. GR4J is a somewhat simpler model, developed by Perrin et al. (2003). This is a four-parameter lumped conceptual rainfall-runoff model.

Table 2.1: Overview of the RAVEN interpretation of the HBV-EC and GR4J models

	GR4J	**HBV-EC**
Water inflow	rain + snow	rain + snow
Surface water	- Ponded water - Water flowing to catchment outlet - Reservoir	- Ponded water - Water flowing to catchment outlet - Reservoir
Soil	4 conceptual layers - Product store (top soil) - Temporary store - Routing store - Groundwater	- Top soil - Fast and slow reservoir from where baseflow originates
Snow	Simple balance between snow and ponded water	More complex snow balance with liquid snow that can refreeze between snow and ponded water.
Routing to outlet	Fixed 10% fast (through temporary soil store) and 90% slow runoff (through routing store)	Separated fast and slow runoff based on parameters
Water outflow	Evaporation from: - Soil - Reservoirs Catchment outlet Groundwater	Evaporation from: - Soil - Canopy - Reservoirs Catchment outlet
Number of parameters	16 (17 with reservoir)	6 (7 with reservoir)

However, the RAVEN emulation contains two additional parameters to add a snow routine to the model. The parameters are given in Table A.1 (HBV-EC) and A.2 (GR4J).

To run the models in RAVEN, five input files are needed. For this study, the initial conditions file (.rvc) was kept empty and instead a warm-up time of three years (1990-1992) was used. For the primary input file (.rvi), the readily available templates for HBV-EC and GR4J models were used (Craig et al., 2020). For GR4J, the process for open water evaporation was added in this file to account for evaporation from reservoirs. The hydrological response units (HRU)/basin definition file (.rvh) contains one "sub"-basin and one HRU per catchment or two HRUs when reservoirs are added, because the open water requires its own HRU. The time-series file (.rvt) contains time-series of observed streamflow, precipitation, minimum temperature, maximum temperature and snowfall. Snowfall was initially set at zero mm/d, but later recalculated by the models. The parameters

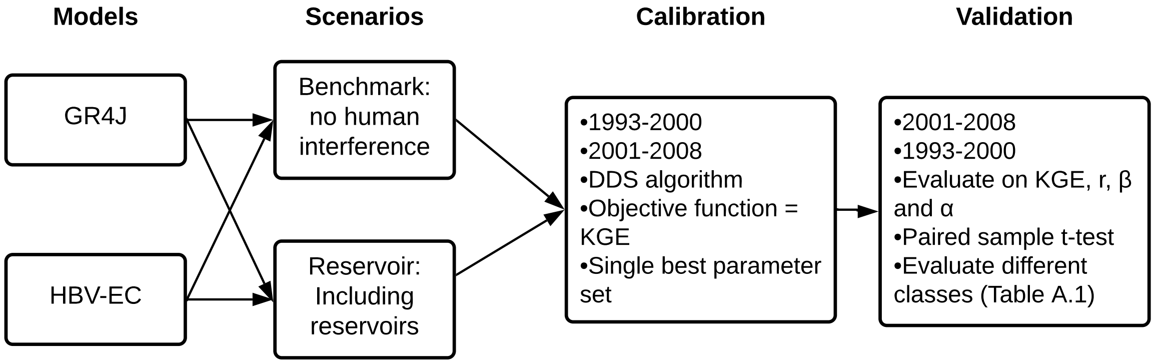


Figure 2.2: Overview of methods

file (.rvt) contains the model parameters. Part of the information in this file was taken from the catchment properties of the CAMELS-BR data or default values were used. The remaining parameters (shown in Table A.1 (HBV-EC) and A.2) (GR4J) were calibrated. Some assumptions and simplifications had to be made about the information in this file regarding vegetation and land use. Where possible, CAMELS-BR data or default values were used. If both were unavailable, values from the RAVEN tutorial files were used, as this was the best guess to apply for all catchments. This can have an impact on the performance of especially HBV-EC, while GR4J requires less information. It was assumed that this did not influence the results about changes in model performance when a reservoir is added to the model, since the same assumptions were made for all catchments. It can have an influence, however, on overall model performance, which was taken into account in the analysis of the results.

2.2.2 Scenarios

Two modeling scenarios were used in this study; with and without reservoirs. The benchmark model performance was assessed by running the model without reservoirs. Then reservoirs were included using the :Reservoir function in the hydrological response unit file. A lake-like reservoir was created, which required information about the weir coefficient (C, default 0.6), crest width (calibrated), maximum depth (h) and surface area (A). A and h can be calculated from the reservoir capacity (V) by reversing the equations given by Chagas et al. (2020):

$$V = 0.678 \times (Ah)^{0.9229} \tag{2.1}$$

$$V = 30.682 \times A^{0.9578} \tag{2.2}$$

In the same file, a separate lake HRU was created with the area of the reservoir to account for evaporation. Otherwise, evaporation from the reservoir would be assumed negligible in RAVEN (Craig et al., 2020). The reservoir is always placed automatically at the outlet of a subbasin. In this case, there is only one "subbasin", so the position of the reservoir is at the outlet of the catchment.

2.2.3 Calibration and cross-validation

Calibration was performed on streamflow at the catchment outlet using the model-independent, multi-algorithm optimization and calibration tool Ostrich (Matott, 2017). After a warm-up period of three years (1990-1992), the models were calibrated for the years 1993 to 2000 (8 years), which is the number of years recommended by Yapo et al. (1996). The rest of the data set (the years 2001 to 2008) was used for validation. Then cross-validation was performed where the calibration and validation periods were swapped. For calibration, the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007) was used and the objective function was the Kling-Gupta efficiency (*KGE*) (Gupta et al., 2009). Particle Swarm Optimization (PSO) was also tested as an alternative calibration algorithm, but this algorithm only found better results for one out of six calibration runs (the two modeling scenarios for three random catchments, selected to test the methods). The run time was over thirty minutes for three catchments compared to just a few minutes with DDS. The best parameters found through calibration were then used for validation. There are sixteen and six parameters that were calibrated for the HBV and GR4J model, respectively (Table A.1 and A.2). When the reservoir was added, an extra parameter was calibrated that represents the unknown crest width. The range for this parameter was 1-50 m. This extra parameter could be the reason for better model performance instead of the reservoir itself, which makes the results from this study more uncertain. However, Perrin et al. (2001) argue that more complex models may perform better in the calibration period, but not in the validation period, which would mean that the extra parameter does not influence the results. Nevertheless, this is taken into account in the analysis of the results.

Model performance was assessed using the *KGE*, which was also used for calibration. Its separate components were also assessed to determine which was the main cause for lower performances. These components are the linear correlation coefficient (r), bias (β) and variability (α) and are all optimal at 1, with r always being lower than (or equal to) 1, while α and β can also be higher. The components all have equal weights for the performance, as seen in the following equation (Gupta et al., 2009):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (2.3)$$

2.2.4 Model performance analysis

The changes in *KGE* between the scenarios were assessed with a paired samples t-test. This showed if including reservoirs increased model performance significantly. This was assessed for all 403 catchments and for classes of the catchments, with similar aridity (ar), seasonality (sea), asynchronicity (asy), land use (lu), catchment area (ca), total reservoir capacity (tc), total relative reservoir capacity (cap), latitude (lat) or longitude (lon), to see if these catchment characteristics influence the model performance A.3. For each catchment characteristic the data set was split into three classes, which in total contain all 403 catchments. To assess if there is a significant difference in model performance between the three classes of the same characteristic an ANOVA test was performed. The classes were all taken from the catchment properties in the CAMELS data set (Chagas et al., 2020). A more detailed description can be found in the documents of this data set. Relative reservoir capacity is called regulation degree by Chagas et al. (2020) and is the ratio of the total reservoir capacity compared to the annual streamflow. Next to the effect of using two different scenarios and multiple classes, the effect of model structure is analysed by performing this study with two different models.

3 Results and discussion

In this chapter, the results are shown and discussed for the HBV and GR4J models. The results are given in detail for the HBV model, starting with the general results and then some more specific results are discussed regarding the different classes and the different KGE components. These results are then compared to the results of the GR4J model. At the end of this chapter, the findings are compared to existing literature.

3.1 Calibration periods

The result analysis is focused on one of the two cross-validation (calibrated for 1993-2000 and validated for 2001-2008). In this section, the both cross-validations are discussed to explain this choice. Figure 3.1a shows the KGE of the validation period (KGE_{val}) corresponding to the different calibration periods. This shows that there are no large differences between the two. It should be noted, however, that there is a significant difference for the reservoir scenario, which gives better results for the first calibration period. The mean difference in KGE_{val} for all catchments is 0.03, which is considered small enough to neglect. Furthermore, the conclusions would not change if the second calibration period or both calibration periods were used. Therefore, it is clearer to only focus on one calibration period in the remainder of this chapter.

The KGE in the calibration period (KGE_{cal}) is significantly higher than the KGE_{val} (3.1b). The mean differences are 0.06 and 0.05 for the benchmark and reservoir scenarios, respectively. This could mean that the model is suffering from some over-parameterization, as also found by Orth et al. (2015) for the HBV-EC model. However, it also makes sense that the model performs somewhat better for the calibration period and the differences found here are assumed to be reasonable, since differences in this range are also found in other studies (e.g. Orth et al., 2015; Wallner et al., 2012).

3.2 HBV model performance

Figure 3.2 shows boxplots with the distribution of the KGE for the two simulation scenarios by the HBV model and the difference between them. The reservoir scenario has a significantly better performance than the benchmark scenario, with mean KGEs of 0.40 and 0.21 respectively. This makes the mean difference 0.19 (Table A.4). Despite this difference, a mean KGE of 0.40 is

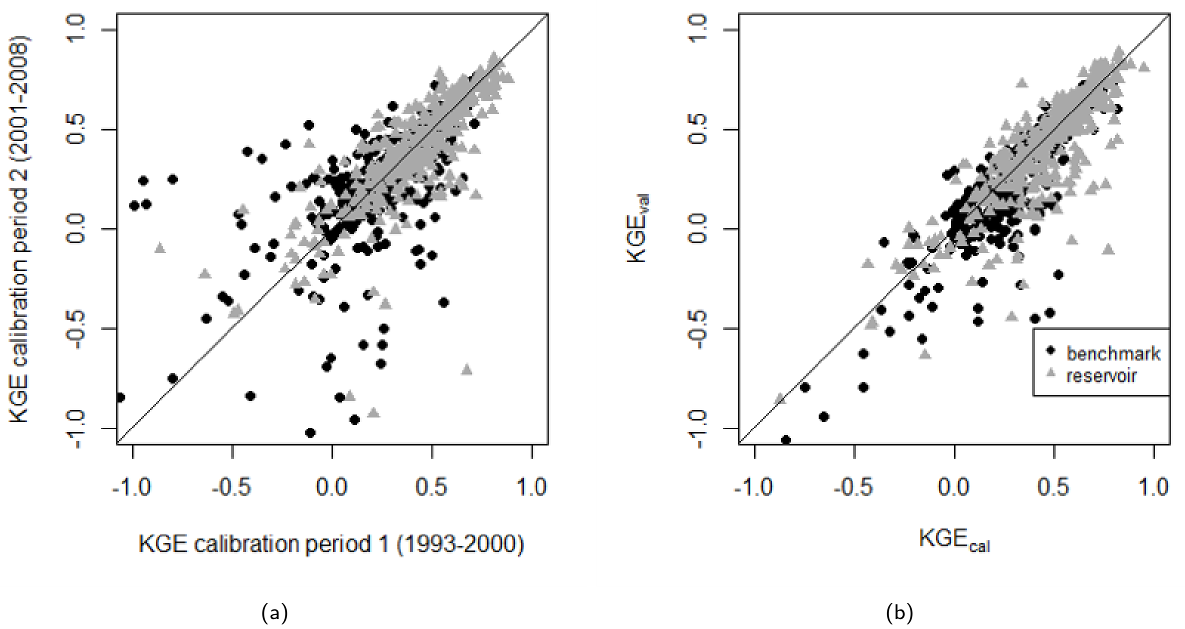


Figure 3.1: (a) KGE of the validation period corresponding to calibration periods 1 and 2 for the benchmark and reservoir scenario and (b) KGE_{cal} plotted against KGE_{val} for calibration period 1 of both scenarios

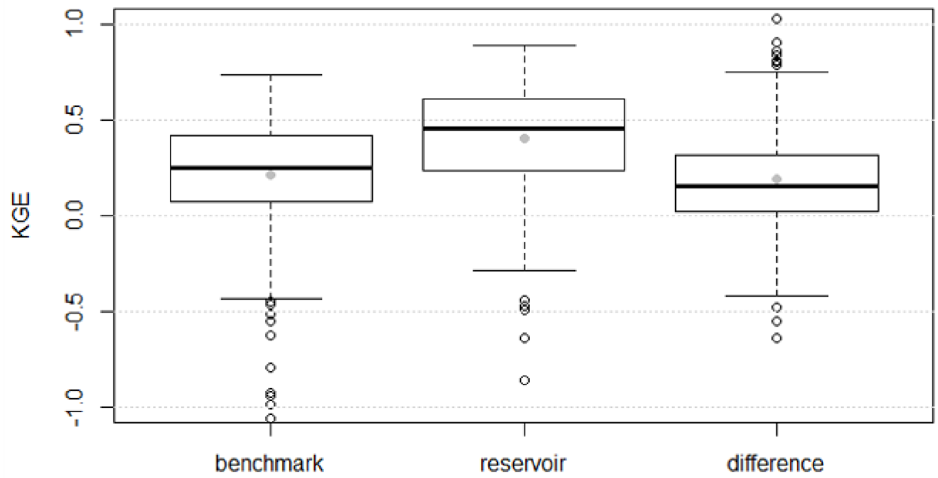


Figure 3.2: KGE of the validation period for the benchmark and reservoir scenario. On the right, the difference between the reservoir and benchmark scenario is shown. The gray dots represent the means of all values. (HBV)

still low and is not considered a good overall model performance (Pechlivanidis et al., 2014). However, there are catchments for which the KGE is higher, indicating a good model performance. The hydrographs of ten random catchments show that the simulated benchmark streamflow often had higher, narrower peaks and lower base-flows than the observed streamflow. Examples of hydrographs and the corresponding flow duration curves of two of these catchments are given in Figure A.3, which indeed indicate lower simulated low-flows more often than observed and the highest streamflow is also higher. Including a reservoir in the model solves this problem in some catchments and therefore improves the performance (Figure A.3b and d), but for most catchments the performance remains poor (Figure A.3a and c).

It is not the purpose of this study to optimize the model performance. Therefore, some choices were made that can negatively impact the KGE. Firstly, because of the large number of catchments, a relatively simple calibration period was chosen, which may not lead to the best parameters. It can require more time to obtain optimal parameters. Secondly, there was no focused study of one of the catchments, looking at it in full detail. The data was taken from the CAMELS data set without having a very close look at whether or not it is correct, because that would be too time-consuming considering the number of catchments. Thirdly, the data set itself was a limiting factor, because of limited details on vegetation (e.g. height/leaf area index) and reservoirs (only total and relative capacity were available).

HBV was found to overestimate low-flows by Unduche et al. (2018), but it works well for estimating peaks in their study. However, their study area, a Canadian Prairie catchment, was completely different than the catchments in this study and they focused on flood forecasting, making the peak flows the most important aspect. Engeland and Hisdal (2009) also report relatively poor performance of the HBV model for low-flows, but not necessarily an overestimation. Due to the variability of catchment characteristics in this study, it cannot be said that either low-flows or peak flows would have a larger influence on the model performance. However, it is certain that this study includes several catchments in the semi-arid region in the northeast of Brazil, where low-flows are common. This could therefore also be (part of) the reason for the relatively low model performance. In this study, the HBV model seems to underestimate low-flows, rather than overestimating them.The logarithm of the KGE quantifies the model performance with a focus on low-flows. This gives mean values of -1.81 for the benchmark scenario and -0.83 for the reservoir scenario. This indeed shows a bad model performance for low flows.

Despite the overall poor model performance, it is interesting to see that the performance increases significantly for the reservoir scenario. This shows that

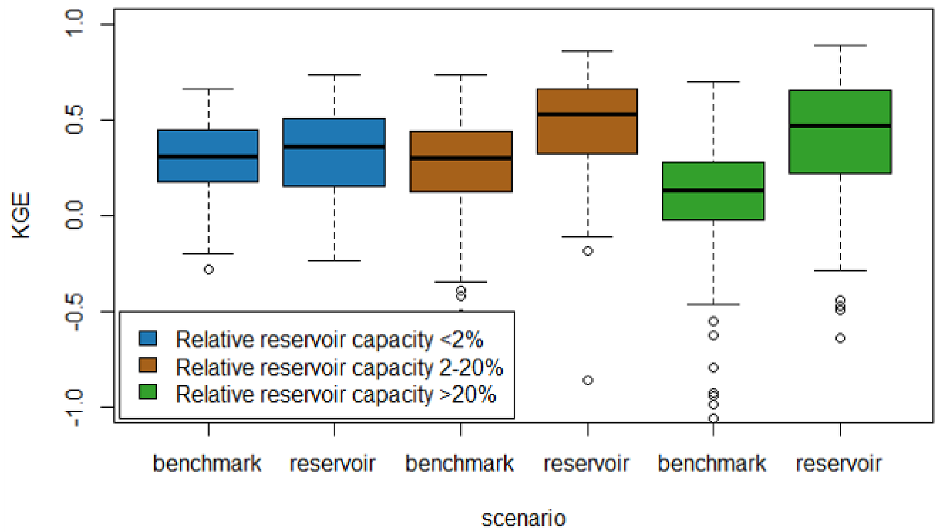


Figure 3.3: KGE of the different classes of reservoir capacity relative to streamflow (HBV)

it helps to add this information to the HBV model, even if not a lot of data about it is available.

3.2.1 Catchment Classes

The classes described in Section 2.2.4 and Table A.3 were investigated to see whether differences in model performance could be found for different classes based on several catchment characteristics. Most classes show the same general trend that the KGE was significantly higher for the reservoir scenario (Table A.4). The only classes that did not result in a significant improvement were the random sample class and the class with the smallest relative reservoir capacities. The random sample class was only used to check if the methods work for 3 random catchments and was not actually meant for analysis. It makes sense that the class with the smallest reservoir capacities does not show a significant improvement, since the difference that is made between the scenarios is adding the reservoir. Therefore, a smaller difference in performance is expected when there are relatively less reservoirs. This also shows that the improvement of the model performance is actually caused by adding the reservoir and not by adding the extra parameter in the reservoir scenario.

The largest increase in KGE between the scenarios is seen for the largest total reservoir capacity (tc3: 0.37) and the relative reservoir capacity (cap3: 0.33) (green cells in Table A.4). Figure 3.3 shows a boxplot of the three classes of relative reservoir capacity to visualize this. The benchmark scenario performance decreases with relatively larger reservoir capacities, while the reservoir scenario performance increases. However, for both total and relative reservoir capacity, the middle class (tc2 and cap2 in Table A.4) have a higher mean KGE in the reservoir scenario than the class with the largest (relative) reservoir capacity (tc3 and cap3 in Table A.4). This is likely due to other catchment characteristics, such as climate and land use, since the benchmark scenario also showed higher performance for these middle classes.

3.2.2 KGE components

All mean KGE components are better in the reservoir scenario than in the benchmark scenario. When the reservoir was included, the mean r increased from 0.57 to 0.67, mean α decreased from 1.22 to 1.01 and mean β increased from 0.53 to 0.65. This shows that all improved and the mean variability is almost perfect when reservoirs are included. The reason for the poor overall performance are thus mostly related to the correlation and bias, meaning the linear relation between the simulated and observed hydrographs and their means. Nevertheless, it should be noted that all KGE components range from very bad (near 0 or higher than 2) to (almost) perfect. This is only a general analysis and not true for every single catchment in this study. The values of β are below 1 for over 80% of the catchments for both scenarios, so in general, the simulated mean streamflow is underestimated. The values of α are mostly (slightly) above 1, which means that the standard deviation is in general overestimated in the simulations. However, since this is only the case for 64% and 55% of the catchments for the benchmark and reservoir scenarios, respectively, this difference is less clear. Nevertheless, it is in line with the description of the hydrographs in Section 3.2.

All KGE components for all of the relative reservoir capacity classes (not shown) have better mean values for the reservoir scenario, except for α of cap1, which is slightly better for the benchmark scenario. The mean value for β was best for cap3, while r and α had better values for cap2. This is in line with what was described in Section 3.2.1 about the overall higher performance of cap2. However, this analysis adds the information that still the mean bias is lower for cap3. All of these classes still have a large range of values for the KGE components.

3.3 Another model

The same analysis can also be done using other models. RAVEN is a useful tool for this because of its flexibility. Many existing models can be used and easily modified (Craig et al., 2020). For this study, one other model (GR4J) was used to see if this led to different results. In this section the results of this model are shown and compared to the results of the HBV model.

3.3.1 GR4J

The GR4J model results in different performance than the HBV model. The benchmark scenario performs significantly better than the reservoir scenario when all catchments are considered, with mean KGEs of 0.57 and 0.56, respectively (Table A.5). However, the difference of the mean KGE (-0.013) is small and the difference is not significant for the other calibration period (not shown). It is more interesting to consider the differences in model performance for the relative reservoir capacity classes (Figure 3.4, Table A.5). The red cells in Table A.5 show that the difference in mean KGE is lowest (highest negative difference) for the tc3 and cap3 classes, which are the classes with the absolutely and relatively largest total reservoir capacity. The difference in model performance between the classes of relative reservoir capacity is also significant, with the lowest performance for both scenarios for the relatively largest reservoir capacity (Figure 3.4). Although the reservoir scenario does not result in different performance for most scenarios and otherwise a (slightly) lower performance, it can still be considered important to include reservoirs in the model, because the overall performance is lower when with a (relatively) larger total reservoir capacity. However, the way in which the reservoirs are implemented for this study does not increase the

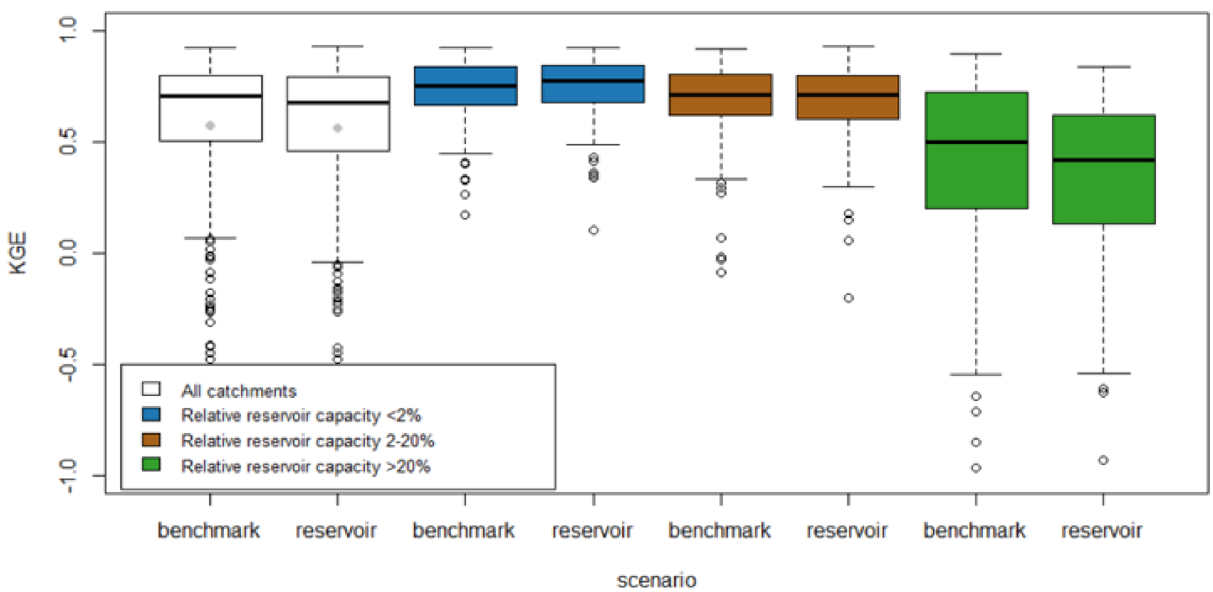
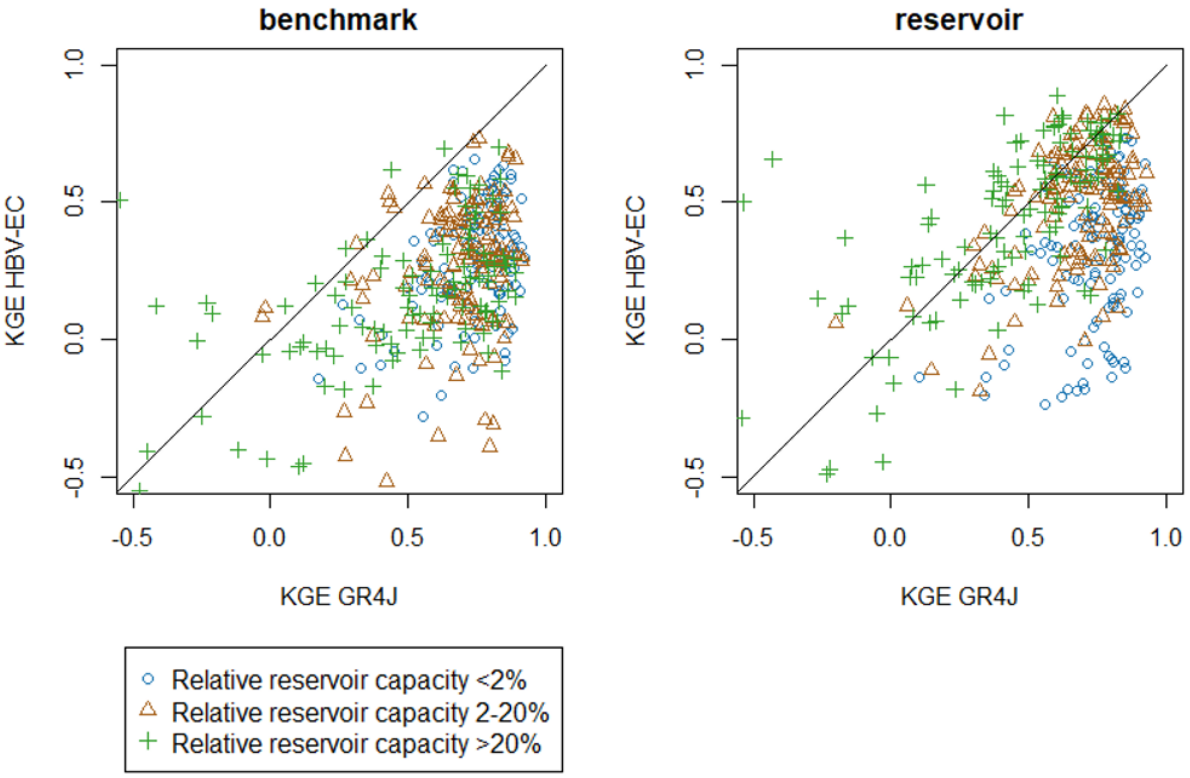


Figure 3.4: KGE of all catchments (white, with the grey dots representing the mean) and the different classes of reservoir capacity relative to streamflow (GR4J)



(a) (b)

Figure 3.5: KGE of the HBV model plotted against the KGE of the GR4J model for the benchmark (a) and reservoir scenario (b)

performance, but rather decreases it. Therefore, this way of including reservoirs does not work well for the GR4J model. The different components of the KGE were all significantly better for the benchmark scenario, but again with very small differences (smaller than 0.1).

3.3.2 Differences between HBV and GR4J

The differences between the performance of the two models can be observed by comparing Figures 3.2, 3.3 and 3.4. As an overview of the main differences between the results, Figure 3.5 shows the KGE for both models with different colors for the relative reservoir capacity classes. Overall, GR4J performs significantly better than HBV, both with and without reservoir. The difference is smaller for the reservoir scenario. For some classes the HBV reservoir scenario performance is better than the GR4J performance, but this is never significant. The most interesting result is found for the relative reservoir capacity classes again. For the reservoir scenario, the difference between the performance of the two models is largest for cap1, with GR4J performing better. However, the cap3 class shows one of the largest differences between the two models in favor of HBV. The mean KGE of this class is slightly (but not significantly) higher for the HBV than for GR4J. This is visualised in Figure 3.5b, where the points for cap3 lay around the 1:1 line. Although no clear conclusions can be drawn from this, it suggests that with a larger relative total reservoir capacity, the reservoir scenario of HBV might work better than GR4J. Possible reasons for these different results, overall and with the added reservoirs, are discussed below. Model structures, parameters and results of other studies that used these models are considered.

The main differences between the models are the structure and the number of parameters. HBV has a more complex model structure, including more processes. One of these processes is related to snow, but this is assumed to be negligible because of the low amounts of snowfall in the catchments. Next to that, canopy is included in the HBV model, which causes increased evaporation. The soil reservoirs are also represented differently in both models, but it is not clear how this would impact the simulation, since both have conceptual instead of physical soil layers. One of these layers for GR4J is the groundwater layer, which can be a source or sink of water. This helps to close the water balance, although it may not be physically correct. The more complex HBV model also has more parameters, 16 compared to 6 for GR4J. If the snow parameters are excluded these numbers are 7 and 4, respectively. It might be expected that a more complex model has a better performance, but this also depends on the availability of data. With lower data availability, less complex models are likely to perform better ((Grayson and Blöschl, 2001)). In this study, the data about canopy is limited, which could lead to lower performance for HBV. Nevertheless, the increase in information by including the reservoir may be handled better by this more complex model.

In other studies that compare these two models, but are not focused on reservoirs, varying results are found. Demirel et al. (2015) and Faiz et al. (2018) found that the performance of HBV is higher, but Piotrowski et al. (2017) found that it depends on the catchment. In this study, enough catchments are used that this catchment dependency should be negligible. Therefore, it seems more likely that the HBV model would perform better. However, in all of these studies, one or a limited number of catchments were studied. For this reason, they may have had more data available or were better able to estimate values if data was unavailable. Ayzel et al. (2020) found that GR4J had a better performance in their large-scale study. Therefore, the overall difference in performance between the two models can be attributed to data availability and the large number of catchments used in this study.

3.4 Synthesis

This study shows that it is important to include reservoirs in hydrological models. However, it is not straightforward to do so. The model performance can improve when reservoirs are included, but it remains poor in most of the catchments. Savenije et al. (2014) and Van Loon et al. (2016) have also identified the need to improve the understanding of complex interactions between people and water. People have had a huge impact on water systems over the past decades in many ways. Construction of reservoirs is only one aspect, but it is a good start to attempt to understand them. Reservoirs contain large amounts of water and are easily visible. It is easier to get data about reservoirs than for example groundwater. Nevertheless, Savenije et al. (2014) and Van Loon et al. (2016) also mention that new data should be collected to achieve better understanding of the human-water system. For reservoirs, information about operation rules would be useful. In this study, it was found that with limited data it is difficult to obtain good performance when modeling reservoirs.

Other studies have used similar data sets to simulate river flows without including reservoirs (Berghuijs et al., 2014; Valipour, 2015). The results of these studies could be reasonable, but this study suggests that the performance of the models could be poor for some catchments with large/many reservoirs. Berghuijs et al. (2014) disregarded results from some agricultural and more arid areas because of poorer model performance. These are areas where reservoirs can be expected. Including those in the model could have improved the performance. Hiep et al. (2018) studied one catchment without including upstream reservoirs in it. They also found that this probably caused an underestimation of low-flows. These studies strengthen the idea that reservoirs should be included in hydrological models.

Nevertheless, there have also been studies where reservoirs are included in hydrological models on various spatial scales. The scale used in this study is unique, because it is at the same time a small scale (catchment scale) and a large scale (because of the number of catchments). Other studies about reservoirs usually either use a global scale (Van Beek et al., 2011; Wanders and Wada, 2015), or a focus on one or a few catchment (Rougé et al., 2019; Turner et al., 2020). When using a global scale, processes are usually simplified more. This study shows that studying reservoirs in such a simplified way, does not result in great model performance. Therefore, the quality of the results of these global scale studies could be questionable. In smaller scale studies, reservoir operation can be modeled in more detail. Using more data to model a reservoir is beneficial for model performance (Turner et al., 2020).

3.5 Outlook

As mentioned before, there are different ways of implementing reservoirs in a hydrological model. For this study, the reservoirs are included as a lake-type reservoir using RAVEN (Craig et al., 2020). This is a very simple approach, requiring only the surface area and depth of the reservoir, the weir coefficient (default

0.6) and the crest width (calibrated). It is also possible to include a man-made reservoir in RAVEN, but this requires information about the relation between the reservoir stage and its discharge, volume and area. If even more information is known about reservoir management, this can also be added (e.g. maximum monthly storage or discharge). Therefore, RAVEN will also be a useful tool for more detailed studies. In RAVEN, the reservoir is always placed at the outlet of a subbasin. In this case this was the same as the outlet of the catchment, because of the lumped models used. Spatially distributed models can better account for the placement of a reservoir. However, Payan et al. (2008) introduced a different method of including reservoirs in a lumped model (GR4J), without accounting for the exact location with good results. This method does not add additional functions of parameters, but requires storage volumes as additional input data.

To be able to improve hydrological modeling of areas with a lot of reservoirs, more data is crucial. The CAMELS-BR data set is the first CAMELS data set that includes some form of reservoir data (Craig et al., 2020). This is not enough data to obtain great model performance, but it is a step in the right direction. Another potential source for reservoir data is the Global Reservoir and Dam (GRanD) database (Lehner et al., 2011). This is probably currently the most complete global database about reservoirs and can thus be an important factor in improving model performance. Furthermore, more detailed data could be obtained from local institutions. An increasing data availability would allow for more complex methods of implementing the reservoir. This can in turn improve model performance.

There are great opportunities and much ongoing research about the place of reservoirs in hydrological systems. However, there is also still a lot unknown. Future research will show better methods of including reservoirs in hydrological models and the data required to do so.

4 | Conclusions and recommendations

The aim of this study was to investigate the effect of including reservoirs in hydrological models on their performance across catchments in Brazil. This was done by including reservoirs in two lumped models (HBV-EC and GR4J) in a simplified way. Lake type reservoirs were implemented using the modular modeling framework RAVEN. Model performance was measured using the Kling Gupta Efficiency (KGE). These are the main findings of this study:

- It is possible to improve model performance by including reservoirs in the model structure. This is seen for the HBV-EC model which showed a significant improvement of model performance with the reservoir scenario. Adding the reservoir caused an increasing mean KGE from 0.21 to 0.40.
- The largest improvement of model performance occurred in the catchments with relatively the most/largest reservoirs. In these catchments, the benchmark performance was poor in both models (mean KGEs of 0.07 for HBV and 0.35 for GR4J), so improvement was also needed the most there. This shows the importance of including reservoirs in hydrological models and the promising improvement of model performance of HBV-EC, where the mean KGE increased to 0.40 for these catchments (For GR4J the KGE decreased to 0.31).
- The improvement of model performance also depends on the model structure. While improved model performance was found using the HBV-EC model, this cannot be concluded for GR4J. Overall performance was higher using GR4J, with a mean benchmark KGE of 0.57, but the performance decreased slightly to a mean KGE of 0.56 when reservoirs were added. This decrease was the worst with the largest/most reservoirs, with a difference in mean KGE of 0.07 between the scenarios. Therefore, HBV-EC seems more promising for modeling reservoirs at this scale.

For future research, it is recommended to focus on getting the best model performance possible for specific catchments. Different models could be compared or different ways of implementing the reservoir. Furthermore, it is strongly recommended to gather more data related to reservoirs in data sets like the CAMELS-BR data used in this study. This would allow for application of more sophisticated methods for reservoir modelling to improve model performance.

Acknowledgements

First of all, I would like to thank my supervisors, Tim van Emmerik, Chanoknun Wannasin (Mo) and Lieke Melsen. Even if Lieke had to end the supervision early, her input and enthusiasm were incredibly valuable and she suggested using RAVEN and CAMELS-BR, which inspired the subject of this thesis. Also Tim and Mo gave a lot of valuable feedback to make this thesis possible. I really liked working with all of them.

I would also like to thank everyone in the thesis-ring. It was nice to see everyone weekly even though we could not go to university to work there. It was also inspiring to read your pieces and I appreciate your feedback on mine.

Furthermore, I would like to thank my friends, housemates, family and boyfriend. It was nice to still see so many people (not at the same time obviously) during this pandemic and be able to get out of the house regularly. Thank you all for keeping me sane during this difficult time for everyone. I guess that also means I should thank the Dutch government for allowing this and not locking us up completely.

Bibliography

Almeida, R.M., Hamilton, S.K., Rosi, E.J., Barros, N., Doria, C.R., Flecker, A.S., Fleischmann, A.S., Reisinger, A.J., Roland, F., 2020. Hydropeaking operations of two run-of-river mega-dams alter downstream hydrology of the largest amazon tributary. Frontiers in Environmental Science 8, 1–11. doi:https://doi.org/10.3389/fenvs.2020.00120.

Ayzel, G., Kurochkina, L., Zhuravlev, S., 2020. The influence of regional hydrometric data incorporation on the accuracy of gridded reconstruction of monthly runoff. Hydrological Sciences Journal , 1–12doi:https://doi.org/10.1080/02626667.2020.1762886.

Beck, H., van Dijk, A., Roo, A., Miralles, D., McVicar, T., Schellekens, J., Bruijnzeel, L., 2016. Global-scale regionalization of hydrologic model parameters. Water Resources Research 52. doi:https://doi.org/10.1002/2015WR018247.

Berghuijs, W.R., Sivapalan, M., Woods, R.A., Savenije, H.H., 2014. Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. Water Resources Research 50, 5638–5661. doi:https://doi.org/10.1002/2014WR015692.

Bergström, S., 1995. The hbv model, in: Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, Colorado, pp. 443—-476.

Blöschl, G., 2006. Rainfall-runoff modeling of ungauged catchments, in: Anderson, M.G. (Ed.), Encyclopedia of Hydrological Sciences, John Wiley & Sons, Chichester. p. 2061–2080.

Braga, B.P.F., Filho, J.G.C.G., von Borstel Sugai, M.R., da Costa, S.V., Rodrigues, V., 2012. Impacts of sobradinho dam, brazil, in: Tortajada, C., Altinbilek, D., Biswas, A.K. (Eds.), Impacts of Large Dams: A Global Assessment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 153–170. doi:https://doi.org/10.1007/978-3-642-23571-9_7.

Carlyle-Moses, D.E., Gash, J.H.C., 2011. Rainfall Interception Loss by Forest Canopies. Springer Netherlands, Dordrecht. pp. 407–423. doi:10.1007/978-94-007-1363-5_20.

Cavalcante, G., Vieira, F., Campos, E., Brandini, N., Medeiros, P.R., 2020. Temporal streamflow reduction and impact on the salt dynamics of the são francisco river estuary and adjacent coastal zone (ne/brazil). Regional Studies in Marine Science 38, 1–11. doi:https://doi.org/10.1016/j.rsma.2020.101363.

Chagas, V.B., Chaffe, P.L., Addor, N., Fan, F.M., Fleischmann, A.S., Paiva, R.C., Siqueira, V.A., 2020. Camels-br: hydrometeorological time series and landscape attributes for 897 catchments in brazil. Earth System Science Data 12, 2075–2096. doi:https://doi.org/10.5194/essd-12-2075-2020.

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research 47. doi:https://doi.org/10.1029/2010WR009827.

Craig, J.R., 2020. Raven. URL: http://raven.uwaterloo.ca/.

Craig, J.R., Brown, G., Chlumsky, R., Jenkinson, R.W., Jost, G., Lee, K., Mai, J., Serrer, M., Sgro, N., Shafii, M., Snowdon, A.P., Tolson, B.A., 2020. Flexible watershed simulation with the raven hydrological modelling framework. Environmental Modelling & Software 129, 1–16. doi:https://doi.org/10.1016/j.envsoft.2020.104728.

Dantas, J.C., da Silva, R.M., Santos, C.A.G., 2020. Drought impacts, social organization and public policies in north-eastern brazil: a case study of the upper paraíba river basin. Environ Monit Assess 192, 765–785. doi:https://doi.org/10.1007/s10661-020-8219-0.

De Graaf, I.E., Gleeson, T., van Beek, L.R., Sutanudjaja, E.H., Bierkens, M.F., 2019. Environmental flow limits to global groundwater pumping. Nature 574, 90–94. doi:https://doi.org/10.1038/s41586-019-1594-4.

DelSole, T., Shukla, J., 2010. Model fidelity versus skill in seasonal forecasting. Journal of Climate 23, 4794–4806. doi:https://doi.org/10.1175/2010JCLI3164.1.

Demirel, M.C., Booij, M.J., Hoekstra, A.Y., 2015. The skill of seasonal ensemble low-flow forecasts in the moselle river for three different hydrological models. Hydrology and earth system sciences 19, 275–291. doi:https://doi.org/10.5194/hess-19-275-2015.

Engeland, K., Hisdal, H., 2009. A comparison of low flow estimates in ungauged catchments using regional regression and the hbv-model. Water Resources Management 23, 2567–2586. doi:https://doi.org/10.1007/s11269-008-9397-7.

Faiz, M.A., Liu, D., Fu, Q., Li, M., Baig, F., Tahir, A.A., Khan, M.I., Li, T., Cui, S., 2018. Performance evaluation of hydrological models using ensemble of general circulation models in the northeastern china. Journal of Hydrology 565, 599–613. doi:https://doi.org/10.1016/j.jhydrol.2018.08.057.

Fantin-Cruz, I., Pedrollo, O., Girard, P., Zeilhofer, P., Hamilton, S.K., 2015. Effects of a diversion hydropower facility on the hydrological regime of the correntes river, a tributary to the pantanal floodplain, brazil. Journal of Hydrology 531, 810–820. doi:https://doi.org/10.1016/j.jhydrol.2015.10.045.

FAO, 2021. Brazil - maps. URL: http://www.fao.org/countryprofiles/Maps/en/?iso3=BRA.

Fortin, V., Turcotte, R., 2006. Le modèle hydrologique mohyse. Note de cours pour SCA7420, Département des sciences de la terre et de l'atmosphere, Université du Québeca Montréal .

Grayson, R., Blöschl, G., 2001. Spatial modelling of catchment dynamics. Spatial patterns in catchment hydrology: observations and modelling , 51–81.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of hydrology 377, 80–91. doi:https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hiep, N.H., Luong, N.D., Nga, T.T.V., Hieu, B.T., Ha, U.T.T., Du Duong, B., Long, V.D., Hossain, F., Lee, H., 2018. Hydrological model using ground-and satellite-based data for river flow simulation towards supporting water resource management in the red river basin, vietnam. Journal of environmental management 217, 346–355. doi:https://doi.org/10.1016/j.jenvman.2018.03.100.

Holländer, H.M., Bormann, H., Blume, T., Buytaert, W., Chirico, G.B., Exbrayat, J.F., Gustafsson, D., Hölzel, H., Krauße, T., Kraft, P., et al., 2014. Impact of modellers' decisions on hydrological a priori predictions. Hydrology and Earth System Sciences 18, 2065–2085. doi:http://doi.org/10.5194/hess-18-2065-2014.

Huard, D., 2020. Calibrating the gr4j-cemaneige hydrological model using ostrich on the raven server. URL: https://pavics-sdi.readthedocs.io/projects/raven/en/test-rtd-build/notebooks/running_OSTRICH_gr4jcn.html.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42, 1–5. doi:https://doi.org/10.1029/2005WR004362.

Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., et al., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment 9, 494–502. doi:http://dx.doi.org/10.1890/100125.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed hbv-96 hydrological model. Journal of Hydrology 201, 272–288. doi:https://doi.org/10.1016/S0022-1694(97)00041-3.

Luo, X., Wu, W., He, D., Li, Y., Ji, X., 2019. Hydrological simulation using trmm and chirps precipitation estimates in the lower lancang-mekong river basin. Chinese Geographical Science 29, 13–25. doi:https://doi.org/10.1007/s11769-019-1014-6.

Mährlein, M., 2016. Streamflow response to forest fire and salvage harvesting in a snow dominated catchment:a model-based change detection approach. Master's thesis. Albert-Ludwigs-Universitat Freiburg.

Martel, J.L., Demeester, K., Brissette, F.P., Arsenault, R., Poulin, A., 2017. Hmet: a simple and efficient hydrology model for teaching hydrological modelling, flow forecasting and climate change impacts. The

International journal of engineering education 33, 1307–1316.

Matott, L., 2017. Ostrich: an optimization software tool, documentation and user's guide, version 17.12. 19. URL: http://www.civil.uwaterloo.ca/envmodelling/Ostrich.html.

Mulligan, M., van Soesbergen, A., Sáenz, L., 2020. Goodd, a global dataset of more than 38,000 georeferenced dams. Scientific Data 7, 1–8. doi:https://doi.org/10.1038/s41597-020-0362-5.

Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M., 2015. Does model performance improve with complexity? a case study with three hydrological models. Journal of Hydrology 523, 147–159. doi:https://doi.org/10.1016/j.jhydrol.2015.01.044.

Payan, J.L., Perrin, C., Andréassian, V., Michel, C., 2008. How can man-made water reservoirs be accounted for in a lumped rainfall-runoff model? Water Resources Research 44. doi:https://doi.org/10.1029/2007WR005971.

Pechlivanidis, I., Jackson, B., McMillan, H., Gupta, H., 2014. Use of an entropy-based metric in multiobjective calibration to improve model performance. Water Resources Research 50, 8066–8083. doi:https://doi.org/10.1002/2013WR014537.

Perrin, C., Michel, C., Andréassian, V., 2001. Does a large number of parameters enhance model performance? comparative assessment of common catchment model structures on 429 catchments. Journal of hydrology 242, 275–301. doi:https://doi.org/10.1016/S0022-1694(00)00393-0.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of hydrology 279, 275–289. doi:https://doi.org/10.1016/S0022-1694(03)00225-7.

Piotrowski, A.P., Napiorkowski, M.J., Napiorkowski, J.J., Osuch, M., Kundzewicz, Z.W., 2017. Are modern metaheuristics successful in calibrating simple conceptual rainfall–runoff models? Hydrological Sciences Journal 62, 606–625. doi:https://doi.org/10.1080/02626667.2016.1234712.

Quick, M., 1995. The ubc watershed model, in: Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, Colorado. chapter 8, pp. 233–280.

Rougé, C., Reed, P.M., Grogan, D.S., Zuidema, S., Prusevich, A., Glidden, S., Lamontagne, J.R., Lammers, R.B., 2019. Coordination and control: Limits in standard representations of multi-reservoir operations in hydrological modeling. Hydrology and Earth System Sciences Discussions , 1–37doi:https://doi.org/10.5194/hess-2019-589.

Savenije, H.H., Hoekstra, A.Y., van der Zaag, P., 2014. Evolving water science in the anthropocene. Hydrology and Earth System Sciences 18, 319–332. doi:https://doi.org/10.5194/hess-18-319-2014.

Seibert, J., Staudinger, M., van Meerveld, H.J.I., 2019. Validation and over-parameterization - experiences from hydrological modeling, in: Beisbart, C., Saam, N.J. (Eds.), Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives. Springer International Publishing, Cham, pp. 811–834. doi:https://doi.org/10.1007/978-3-319-70766-2_33.

Sivapalan, M., Savenije, H.H., Blöschl, G., et al., 2012. Socio-hydrology: A new science of people and water. Hydrol. Process 26, 1270–1276. doi:https://doi.org/10.1002/hyp.8426.

Souza Filho, E., 2009. Evaluation of the upper paraná river discharge controlled by reservoirs. Brazilian Journal of Biology 69, 707–716. doi:https://doi.org/10.1590/S1519-69842009000300024.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research 43. doi:https://doi.org/10.1029/2005WR004723.

Turner, S.W., Doering, K., Voisin, N., 2020. Data-driven reservoir simulation in a large-scale hydrological and water resource model. Water Resources Research 56, e2020WR027902. doi:https://doi.org/10.1029/2020WR027902.

Unduche, F., Tolossa, H., Senbeta, D., Zhu, E., 2018. Evaluation of four hydrological models for operational flood forecasting in a canadian prairie watershed. Hydrological Sciences Journal 63, 1133–1149. doi:https://doi.org/10.1080/02626667.2018.1474219.

Valipour, M., 2015. Long-term runoff study using sarima and arima models in the united states. Meteorological Applications 22, 592–598. doi:https://doi.org/10.1002/met.1491.

Van Beek, L., Wada, Y., Bierkens, M.F., 2011. Global monthly water stress: 1. water balance and water availability. Water Resources Research 47. doi:https://doi.org/10.1029/2010WR009791.

Van Loon, A.F., Gleeson, T., Clark, J., Van Dijk, A.I., Stahl, K., Hannaford, J., Di Baldassarre, G., Teuling, A.J., Tallaksen, L.M., Uijlenhoet, R., et al., 2016. Drought in the anthropocene. Nature Geoscience 9, 89–91.

Van Loon, A.F., Rangecroft, S., Coxon, G., Breña Naranjo, J.A., Van Ogtrop, F., Van Lanen, H.A., 2019. Using paired catchments to quantify the human influence on hydrological droughts. Hydrol. Earth Syst. Sci 23. doi:https://doi.org/10.5194/hess-23-1725-2019.

Wada, Y., Bierkens, M.F., De Roo, A., Dirmeyer, P.A., Famiglietti, J.S., Hanasaki, N., Konar, M., Liu, J., Schmied, H.M., Oki, T., et al., 2017. Human-water interface in hydrological modelling: current status and future directions. Hydrology and Earth System Sciences 21, 4169–4193. doi:https://doi.org/10.5194/hess-21-4169-2017.

Wallner, M., Haberlandt, U., Dietrich, J., 2012. Evaluation of different calibration strategies for large scale continuous hydrological modelling. Advances in Geosciences 31, 67–74. doi:https://doi.org/10.5194/adgeo-31-67-2012.

Wanders, N., Wada, Y., 2015. Human and climate impacts on the 21st century hydrological drought. Journal of Hydrology 526, 208–220. doi:https://doi.org/10.1016/j.jhydrol.2014.10.047.

Whittaker, G., Confesor, R., Di Luzio, M., Arnold, J., et al., 2010. Detection of overparameterization and overfitting in an automatic calibration of swat. Transactions of the ASABE 53, 1487–1499. doi:https://doi.org/10.13031/2013.34909.

Woo, M.k., Thorne, R., Szeto, K., Yang, D., 2008. Streamflow hydrology in the boreal region under the influences of climate and human interference. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 2249–2258. doi:https://doi.org/10.1098/rstb.2007.2197.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181, 23–48. doi:https://doi.org/10.1016/0022-1694(95)02918-4.

Ye, B., Yang, D., Kane, D.L., 2003. Changes in lena river streamflow hydrology: Human impacts versus natural variations. Water resources research 39. doi:https://doi.org/10.1029/2003WR001991.

Zhou, T., Nijssen, B., Gao, H., Lettenmaier, D.P., 2016. The contribution of reservoirs to global land surface water storage variations. Journal of Hydrometeorology 17, 309–325. doi:https://doi.org/10.1175/JHM-D-15-0002.1.

A Appendix

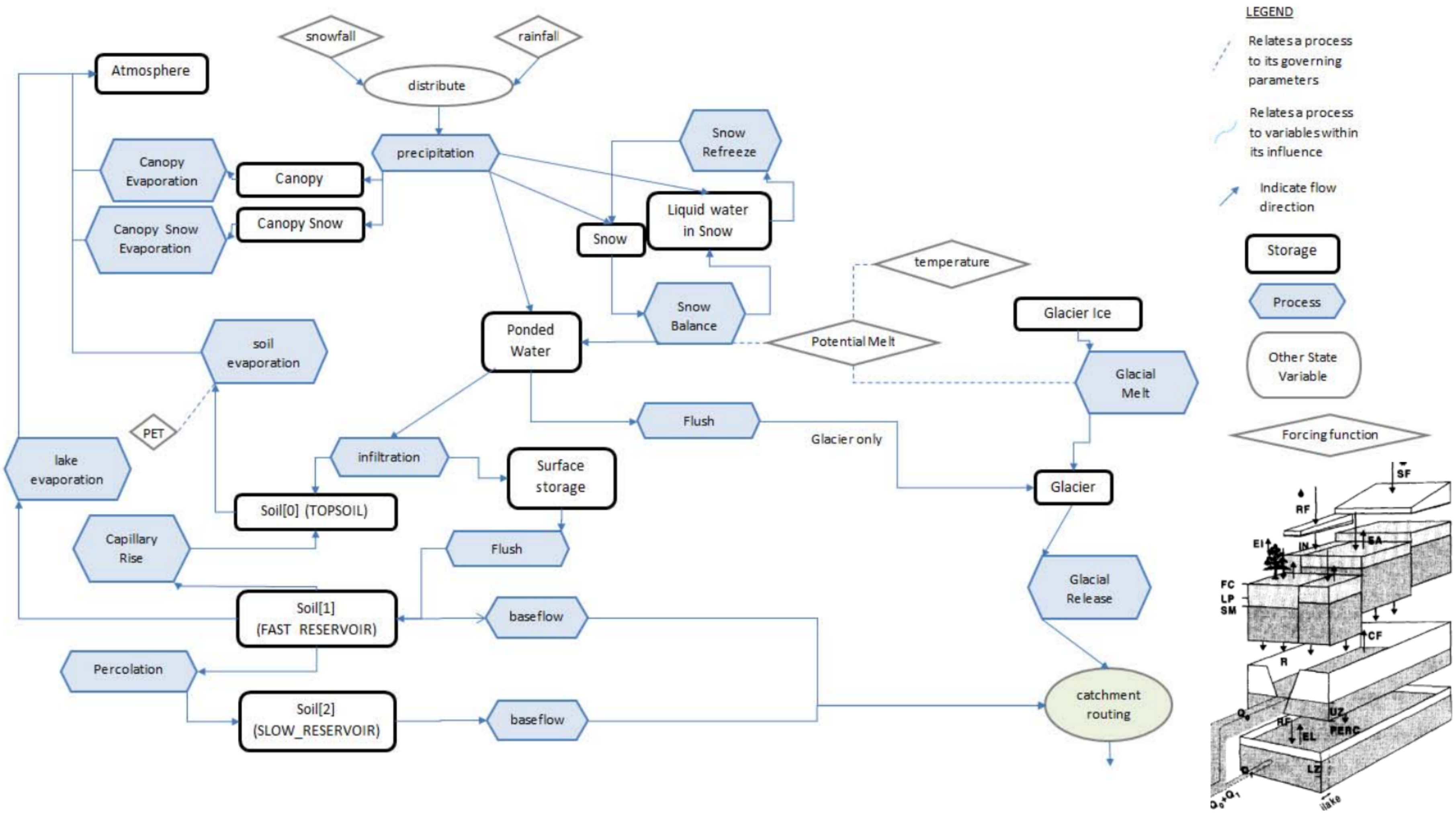


Figure A.1: Structure of the HBV-EC model in RAVEN ((Bergström, 1995; Craig, 2020)). This is one of the models used in this study and is shortly described in Section 2.2.1

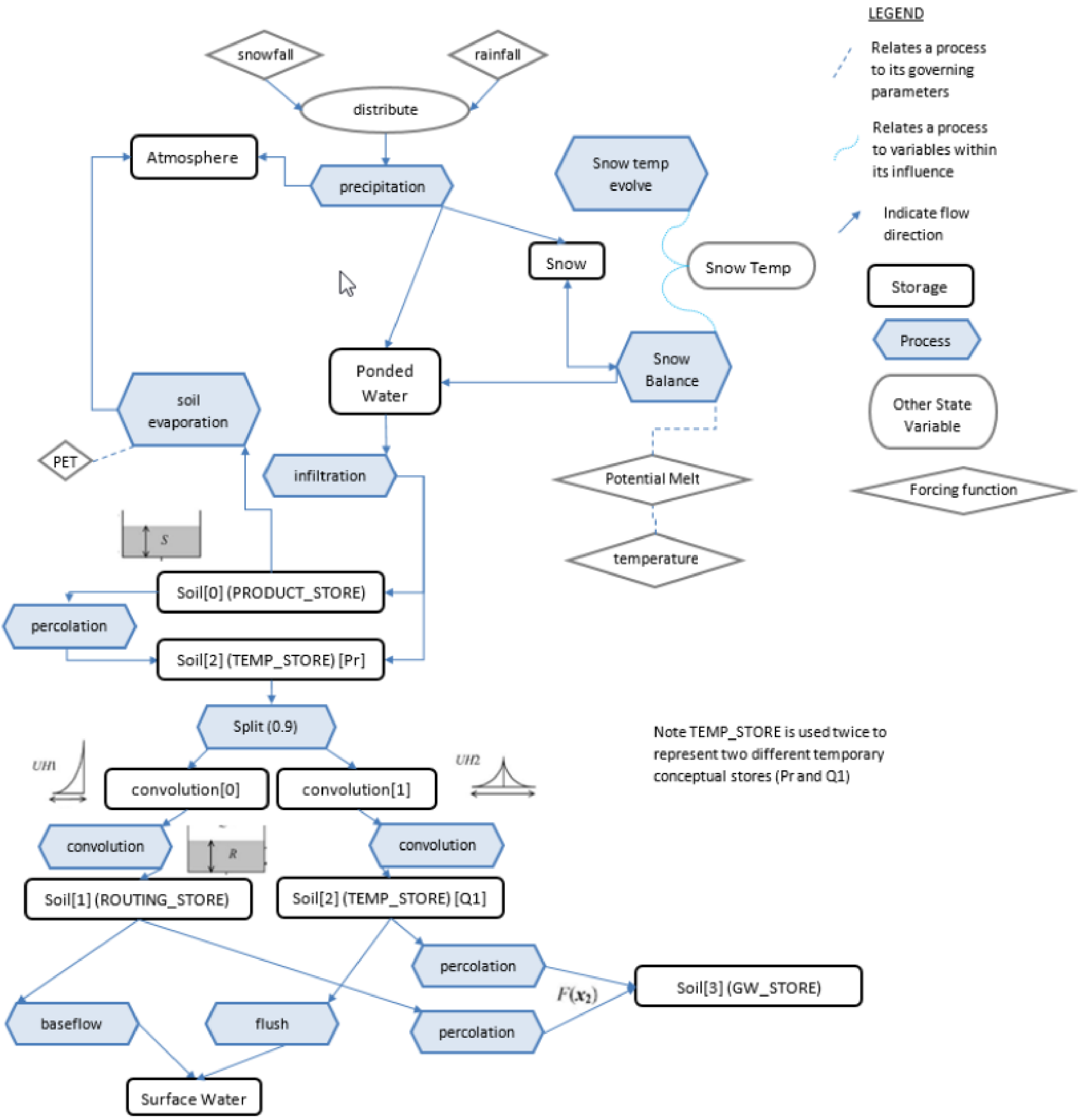


Figure A.2: Structure of the GR4J model in RAVEN((Craig, 2020; Perrin et al., 2003)). This is one of the models used in this study and is shortly described in Section 2.2.1

Table A.1: Parameters and ranges used for calibration of the HBV-EC model. Adapted from Mährlein (2016) with value ranges from, Beck et al. (2016), Carlyle-Moses and Gash (2011) and Craig et al. (2020). These are referred to in Sections 2.2.1 and 2.2.3

Parameter	**Description**	**Range**
TFr(ain)	Fraction of rainfall not lost by interception	0.7 - 1
TFs(now)	Fraction of snowfall not lost by interception	0.7 - 1
Tlapse	Temperature lapse rate	0 - 7
TT	Threshold temperature limit for snow/rain [°C]	-1 - 1
TTi	Temperature interval for mixture of snow and rain [°C]	0 - 4
Cmin	Minimum melt factor [mm/°C/d]	1.5 - 2.5
Cmax	Maximum melt factor [mm/°C/d]	3 - 4
MRF	Ratio between the melt factor in forest to open areas	0 - 1
CRFR	Melt factor for freezing of liquid water in snow	2 - 4
WHC	Macimum liquid water content of smow	0.04 - 0.07
AM	Aspect melt factor	0 - 1
FC	Field capacity [mm]	0 - 1
BETA	Exponent in soil drainage function	1 - 6
K1	Outflow coefficient fast reservoir	0.01 - 0.8
ALPHA	Exponent in outflow for fast reservoir	1 - 10
K2	Outflow coefficient for slow reservoir	0.001 - 0.15

Table A.2: Parameters and ranges used for calibration of the GR4J model, ranges from Huard (2020). These are referred to in Sections 2.2.1 and 2.2.3

Parameter	**Description**	**Range**
x1	Maximum soil moisture content (production store) [m]	0.01 - 2.5
x2	Water exchange coefficient with groundwater [mm/d]	-15 - 10
x3	Reference capacity of the routing store [mm]	10 - 700
x4	lag between rainfall and runoff [d]	0 - 7
x5	Melt factor [mm/d/°C]	1 - 30
x6	Air snow coefficient	0 - 1

Table A.3: Classes with a short description and the number of catchments in the class. Model performance was assessed for all of these different classes to assess the influence of different catchment characteristics on change of model performance between the benchmark and reservoir scenarios. A more detailed description can be found in the document that comes with the attributes of the CAMELS-BR data set (Chagas et al., 2020)

class	Description	Number of catchments
all	All 403 catchments	403
rand	Random sample	3
ar1	Aridity < 0.5	33
ar2	Aridity 0.5-1.0	262
ar3	Aridity > 1.0	108
sea1	Seasonality < 0	74
sea2	Seasonality 0-0.8	157
sea3	Seasonality > 0.8	172
asy1	Asynchronicity < 0.05	128
asy2	Asynchronicity 0.05-0.15	151
asy3	Asynchronicity > 0.15	124
lu1	Land use = Forest	151
lu2	Land use = Crops + Crop Mosaic	219
lu3	Land use = Shrub	33
ca1	Catchment area < 1000 km^2	32
ca2	Catchment area 1000-10000 km^2	172
ca3	Catchment area > 10000 km^2	199
tc1	Reservoir capacity < 100 hm^3	178
tc2	Reservoir capacity 100 - 1000 hm^3	129
tc3	Reservoir capacity > 1000 hm^3	96
cap1	Relative reservoir capacity < 2%	120
cap2	Relative reservoir capacity 2-20%	136
cap3	Relative reservoir capacity > 20%	147
lat1	latitude < -20	182
lat2	latitude -20 - -10	121
lat3	latitude > -10	100
lon1	longitude < -50	131
lon2	longitude -50 - -45	86
lon3	longitude > -45	186

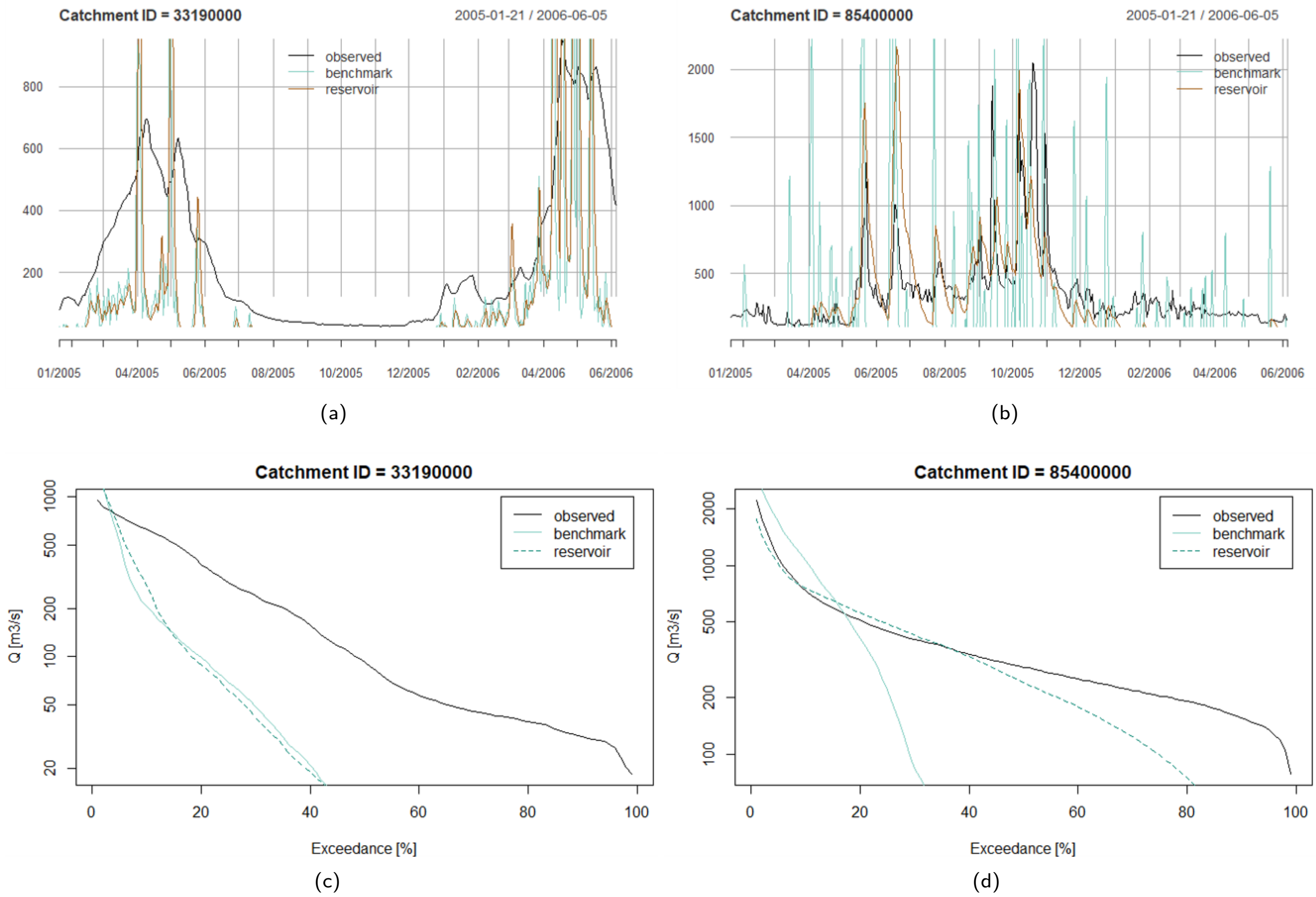


Figure A.3: Example hydrographs with streamflow in m^3/s on the y axis (a,b) and flow duration curves(c,d) of two catchments with the observed streamflow and the two scenarios simulated using the HBV-EC model of one catchment with relatively poor performance (a,c, KGE benchmark = 0.21, KGE reservoir = 0.29) and one with relatively good performance in the reservoir scenario (b,d, KGE benchmark = -0.05, KGE reservoir = 0.76). These figures visualise the description in Section 3.2

Table A.4: Mean KGE of all catchments and different classes for the two scenarios and the difference between them using the HBV-EC model. Significance: *: p = 0.01-0.05, **: p = 0.001-0.01, ***: p < 0.001. Green cells show the largest improvement of model performance and red cells the smallest improvement. These results are explained and discussed in Section 3.2

Class	Benchmark	Reservoir	Difference	Significance
all	0.209	0.401	0.192	***
rand	0.421	0.475	0.054	-
ar1	0.453	0.593	0.140	***
ar2	0.209	0.396	0.187	***
ar3	0.110	0.340	0.230	***
sea1	0.275	0.393	0.118	***
sea2	0.271	0.419	0.148	***
sea3	0.128	0.389	0.261	***
asy1	0.194	0.407	0.213	***
asy2	0.227	0.450	0.224	***
asy3	0.204	0.331	0.127	***
lu1	0.193	0.398	0.205	***
lu2	0.224	0.408	0.184	***
lu3	0.194	0.370	0.176	***
ca1	0.175	0.370	0.195	***
ca2	0.230	0.343	0.113	***
ca3	0.197	0.456	0.259	***
tc1	0.270	0.308	0.039	*
tc2	0.209	0.486	0.277	***
tc3	0.097	0.466	0.370	***
cap1	0.290	0.314	0.023	-
cap2	0.264	0.484	0.219	***
cap3	0.071	0.397	0.326	***
lat1	0.220	0.416	0.196	***
lat2	0.160	0.392	0.233	***
lat3	0.247	0.381	0.134	***
lon1	0.295	0.440	0.146	***
lon2	0.141	0.349	0.208	***
lon3	0.175	0.397	0.221	***

Table A.5: Mean KGE of all catchments and different classes for the two scenarios and the difference between them using the GR4J model. Significance: *: p = 0.01-0.05, **: p = 0.001-0.01, ***: p < 0.001. Green cells show the largest improvement of model performance and red cells show the largest decrease. These results are explained and discussed in Section 3.3.1

Class	Benchmark	Reservoir	Difference	Significance
all	0.573	0.560	-0.013	*
rand	0.464	0.488	0.024	-
ar1	0.735	0.715	-0.020	-
ar2	0.682	0.680	-0.002	-
ar3	0.234	0.195	-0.040	*
sea1	0.468	0.444	-0.025	-
sea2	0.631	0.620	-0.011	-
sea3	0.564	0.553	-0.011	-
asy1	0.654	0.664	0.010	-
asy2	0.560	0.535	-0.025	*
asy3	0.502	0.477	-0.025	*
lu1	0.618	0.617	-0.001	-
lu2	0.573	0.555	-0.019	*
lu3	0.324	0.281	-0.042	-
ca1	0.519	0.524	0.005	-
ca2	0.533	0.539	0.007	-
ca3	0.617	0.584	-0.034	***
tc1	0.638	0.636	-0.002	-
tc2	0.526	0.534	0.008	-
tc3	0.512	0.447	-0.065	***
cap1	0.730	0.742	0.012	-
cap2	0.660	0.659	-0.001	-
cap3	0.353	0.305	-0.048	***
lat1	0.645	0.646	0.001	-
lat2	0.513	0.489	-0.024	-
lat3	0.510	0.483	-0.027	*
lon1	0.737	0.733	-0.004	-
lon2	0.590	0.590	0.001	-
lon3	0.445	0.418	-0.027	*