University of Osnabrück

Project Report M. Sc.

Future Hydropower Reservoirs & Dams

Automated reservoir calculation for all available future hydropower dam locations worldwide

by:
Leonhard Urs Bürger
M. Sc. Environmental Systems and Resource Management
Matriculation Number: 965753

Supervisor:
Dr. Jürgen Berlekamp

February 2021

Contents

1 Motivation

Using the Future Hydropower Reservoirs and Dams Database (Zarfl et al. 2015) to estimate impacts of future dam build projects always came with missing information. Firstly the uncertainties concerning the from grey-sources collected data and secondly the derived reservoir volume and area information. The first attempts to derive those reservoir attributes and then their partial automatisation by Albers (2017) showed good results, but involved a modeller making crucial decisions in the process. Therefore the results were influenced by the modeller and not always consistent. Deriving the dam reservoirs by hand also required a lot of time and labour.
To enhance the process Pollmüller (2019) developed a new fully automated process to derive such reservoirs. It works well for dams in valleys or mountainous regions, but showed weaknesses in flat regions. Furthermore it shows occasional problems with some dams and is unable to process multiple dams in one run due to ArcGIS Python 2.7 RAM usage restrictions to 2 GB.
The goal for this work is therefore to generalize the automated process to a worldwide usage and to optimize RAM usage to derive as many dams as possible in one run.

2 Process

The process of the reservoir estimation is based on the work of Pollmüller (2019) with changes mainly focused on large-scale calculation, data availability and bug fixing.

2.1 Data

Therefore the data needed for the calculation is selected mainly due to accessibility and worldwide availability. The used data consists of a digital elevation model (DEM), a river network and the future hydropower dams.
The used hydropower dams are part of the Future Hydropower Reservoirs and Dams Database by Zarfl et al. (2015). It consists of roughly 3600 dams worldwide containing dams both planned and under construction. The database gives information about the potential dam position, height, predicted power output etc., but some information is not available for all dams. Additionally the quality of data is not always consistent and needs to be plausibility checked before usage (e.g. an unrealistically high power output).
To derive a realistic dam reservoir a 3 arc-second unconditioned digital elevation model is used. The used HydroSHED DEM (Lehner et al. 2008) is based on the near worldwide SRTM data gathered in early 2000. All landmasses between 60° north and 58° south are covered by the DEM in a grid format with a cellwidth of 92 m near the equator.
All further data is taken form the datasets published by Linke et al. (2019). On the one hand a worldwide 15 arc-second river network called RiverATLAS with additional information about discharge, evaporation etc. and on the other hand the BasinATLAS. It consists of watershed basins for large rivers and sub-rivers to easily split the world into hydrological regions. The BasinATLAS is not essential for the process, but speeds up the derivation by allowing for better preprocessing.

2.2 General process

The process is implemented as a Python-Toolbox using the Python 2.7 interface of ArcMap 10.6. It is only executable when the *Spatial Analyst*-Extension is licensed and activated.

The general reservoir derivation process is shown in figure 1. As the first step of the

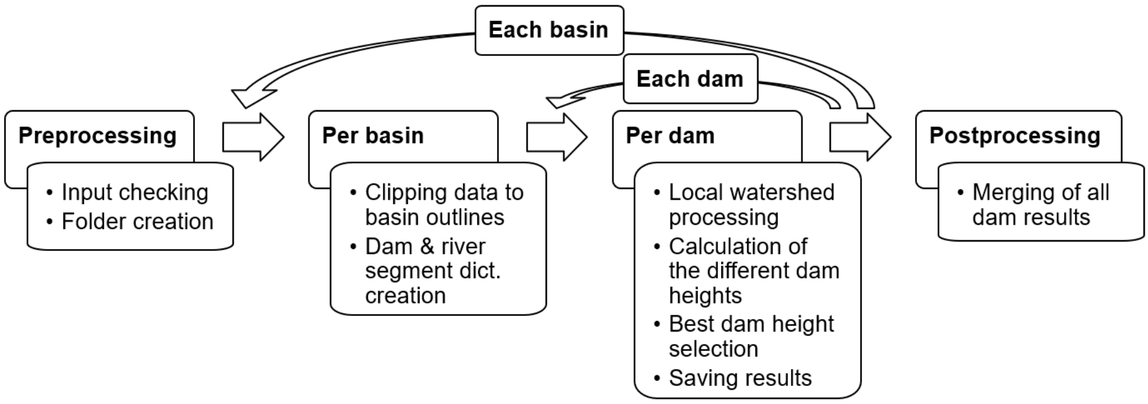

Figure 1: Diagram of the general reservoir derivation process.

preprocessing all inputs are read and converted into variables. Some inputs are checked for plausibility, for example that the dam calculation height step must be a positive value. If all inputs are given and plausible, the folder and geodatabase structure for the derivation process and the results is created.

The actual reservoir calculation works iteratively by first clipping all inputs to individual basin outlines and then deriving the reservoirs of each dam inside the basins. This intermediate step of using the basins as way to narrow the data to a certain area before the actual calculation is used mainly to reduce processing time and memory usage. For the iteration on the basin-level dictionaries are used, which contain all dams and river segments inside each basin and some of their attributes. Each dam inside those per-basin dictionaries is then individually processed. The processing is composed of the watershed estimation for the given dam, the calculation of all possible different dam wall heights and the selection of the best dam height. This process results in a predicted most likely reservoir with a certain volume, area and shape and the volumes and areas of all the other reservoirs for the less likely dam heights.

After all dam reservoirs inside one basin are derived the process is repeated for the next basin. When all calculations are completed all the data is merged into datasets containing the information of all reservoirs.

2.3 Process details

The general process only outlines the individual processing steps and their interaction. Therefore the details of each processing step are described in the following subsections. The process is described using the standard values, all values and properties written in *italic* are designed to be changeable in the initialization part of the script code.

2.3.1 Basin preparation

The first step for the basin preparation is to align the dam positions with the used river network. To achieve a consistency all dams are copied and their positions snapped to the closest river segments (max. *500 m*). If no river segment is close enough, the dams reservoir will not be calculated and the position needs to be reassessed by hand.
Next the dams are intersected with the basins and only the basins containing at least one dam are used for further calculations. For each remaining basin the input elevation model, river segments and snapped dam positions are clipped to the basin outline.
To simplify the reservoir derivation two easy to access key-value data structures (dictionaries) are created for each basin. One such dictionary is used for all dams inside the basin and filled with the ID of the closest river segment and its *maximum* monthly discharge. The other dictionary is filled with the river segments and the IDs of their upriver segments. The used river network normally uses the ID of the downriver segment, but the upriver segments are needed for the efficient watershed processing.

2.3.2 Local watershed processing

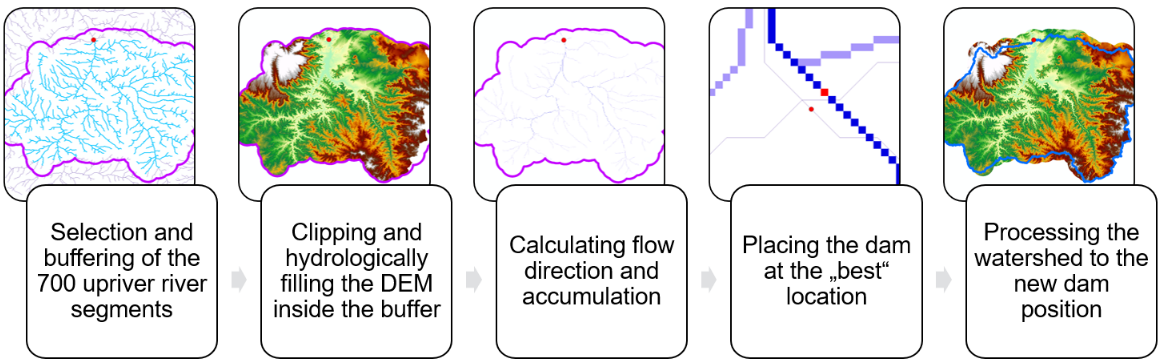

Figure 2: Diagram of the watershed derivation process.

The local watershed of a dam could reach in all directions and could have all kinds of different shapes. Using a fixed distance around the dam positions either results in to small watersheds or a time-consuming calculation, depending on the used distance. To bypass this problem a different strategy shown in figure 2 is used to derive the watershed.
The assumption for the strategy is that only the part of the watershed is needed, that may contain the dam reservoir. Thus the area of the watershed far away from the dam position is not needed for the local watershed. The strategy additionally uses the river network above the dam to further narrow the area before the calculation. To achieve

this, the river segment dictionary is used and the *700* topologically closest upriver segments are selected and buffered (*10 km*) to roughly limit the watershed. This rough watershed may not always be sufficient for very long and large reservoirs, but it spans up to a couple hundred kilometres depending on the upriver topography covering the dimensions of almost all reservoirs.

This rough watershed is then used to clip and fill the elevation model. The filling of the DEM is important to make sure, that no sinks and inconsistencies are left before the hydrological derivation of the watershed starts. This hydrological derivation begins with the calculation of the flow direction for surface water in each raster cell based on the slope of the surrounding cells. Those flow directions are then used to calculate the flow accumulation for each cell by adding up all cells whose water would flow into the cell. This process produces a local river network with a more accurate resolution (3 arc-seconds) compared to the used river network (15 arc-seconds) and can therefore be used to position the dam more accurately.

This accurate dam placement considers the flow accumulation raster in a 1 km square around the dam position. To identify the main river most likely hosting the dam in this square the biggest leap in flow accumulation values is used. All cells with values above the leap are considered part of the main river. If another tributary joins the main river in this square and adds at least 20 % of the main river's flow accumulation the dam will only be placed downriver of this confluence. The dam is then placed for the watershed calculation closest to its original position on the (remaining part of the) main river in the flow accumulation square.

This process of determining the main river and potential tributaries is used to ensure that the dam is actually placed on the main stream, but also avoiding to place the dam far downriver without intending to. This would happen using the 'normal' way of snapping the pour point for the watershed derivation. Because then the dam would be placed on the cell with the highest flow accumulation in a certain radius around its original position. This would result in the possibility to only snap to a tributary when using a small snapping radius. Additionally all dams would be placed downstream of their original position due to higher flow accumulation values there.

The position of the dam on the main river is then used as pour point to calculate the local watershed as a grid. It consists of all cells whose water flows through the dam position according to the previously established flow directions. The watershed may contain some cells that are only connected via corner with the rest. This is due to the flow directions having eight possible directions and each cell only having four connected neighbours. To make sure that the watershed has no such irregularities a one-way boundary clean algorithm is used. It slightly expands the watershed to make sure all cells part of the watershed are connected via edge and no isolated cells are left. This cleaned watershed is then used to create a hydrologically filled DEM of the watershed to later determine the reservoir shape.

In order to smooth the watershed shape, the watershed grid is converted into a simplified polygon with a less edgy appearance and a watershed line. This line is derived in a 20 km radius around the dam position to aid determining the dam profile. Both vectorized watershed variants are then projected to get accurate lengths and areas.

The watershed line around the dam position is then used to determine the elevations along the line in the unfilled DEM. If the line has larger indentations and notches the

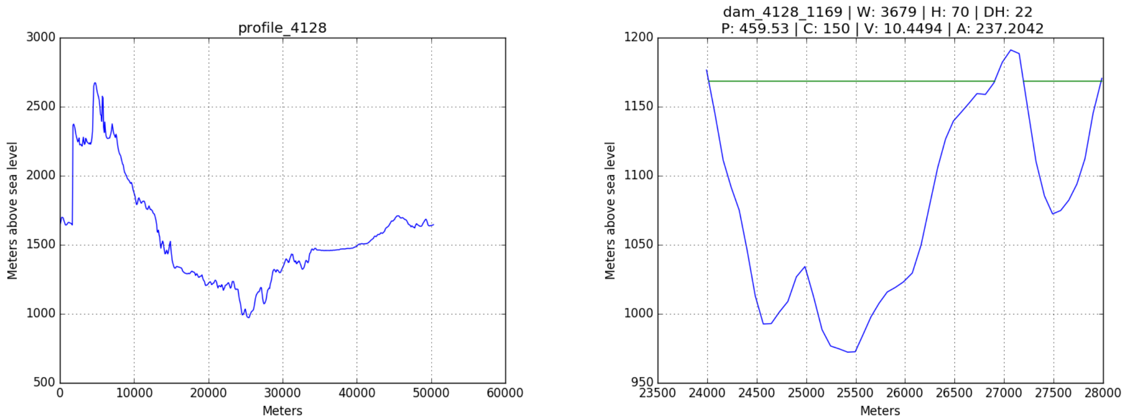

Figure 3: Visualisation of the elevation height profile along the local watershed outline around the predicted dam position in the center for the already build Tekezé Dam in Ethiopia (left). The right figure shows the derived dam height visualised by a green line in the enlarged elevation height profile.

process may result in different profiles each only covering a part of the watershed line. In this case the profile containing the lowest point is used as the potential dam wall profile. This dam profile is later used to calculate parameters to assess the viability of certain dam heights and to visualize the elevation height profile along the dam wall. One such height profile is shows in figure 3 in the left graph for a dam in Ethiopia.

2.3.3 Calculation of the different dam heights

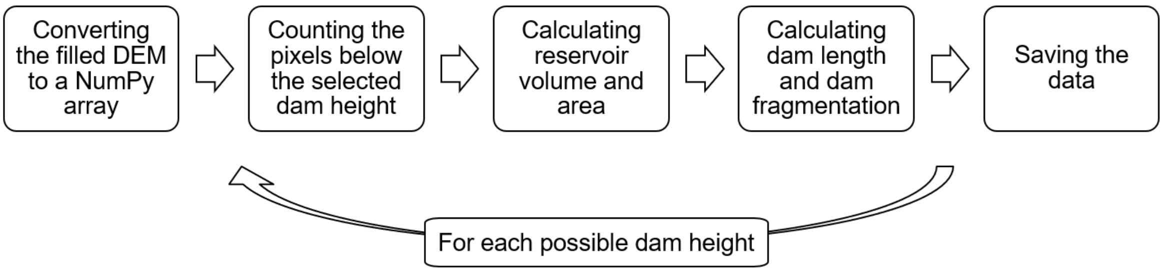

Figure 4: Diagram of the processing of the different dam heights.

To calculate the reservoir volumes and areas for the different possible dam heights the filled elevation model inside the derived local watershed is used. The used process is depicted in figure 4. Possible dam heights used are limited by the maximum dam height of *300 m* or the highest point in the surroundings of the dam and are evaluated every couple meters in height. The calculation is done in an array representation of the DEM to simplify and accelerate the calculation. Each cell in the DEM represents a 3 arc-second wide and tall area in the real world. But due to the earth curvature those 3 arc-second cells have different sizes on different latitudes ϕ (in radians) and therefore the values need to be length-corrected with the correction factor η using equation (1).

$$\eta = \frac{U}{360°} \cdot \cos(|\phi|) \cdot \alpha_{equator} \quad (1)$$

Additional parameters used are the earths diameter U (40030 km) and the departure constant $\alpha_{equator}$ ($\frac{1°}{111.319\,\mathrm{km}}$) representing the length of 1° latitude near the equator. This correction factor is then used to derive the different areas A and volumes V of the dam reservoir for all the different dam heights. To get those reservoir metrics all array elements $N_{<h}$ with elevation heights lower than the dam base elevation plus the currently calculated dam height h are counted and then calculated using equations (2) and (3). The additionally needed length L of a DEM cell at the equator is for the DEM used 92 m.

$$A = N_{<h} \cdot L^2 \cdot \eta \tag{2}$$

$$V = N_{<h} \cdot L^2 \cdot \eta \cdot h \tag{3}$$

For each dam height a rough estimation of the potential power output P_{el} (in kW) is also calculated using the *maximum annual* discharge Q_{max} (in $\mathrm{m^3/s}$) and the current dam height tested h (in m) as shown in equation (4).

$$P_{el} = c \cdot h \cdot Q_{max} \tag{4}$$

The constant c combines the efficiency of the turbine and generator ($\sim 85\,\%$), the water density ($1000\,\mathrm{kg/m^3}$), the gravitational acceleration ($9.81\,\mathrm{m/s^2}$) and the conversion from Watts to Kilowatts. Therefore the Value of c is roughly set to $8.5\,\mathrm{kN/m^3}$ (Pollmüller 2019).
Part of the later best dam selection is the fragmentation and estimated cost of the dam wall. To get those number the intersections between the dam height line and the elevation profile are used as seen in figure 3 on the right. The different fragment are then counted and the dam fragmentation $\xi_{longest}$ is calculated using the length $l_{frag_{longest}}$ of the longest dam fragment and the total dam length l_{total} an seen in equation 5.
A high fragmentation $\xi_{longest}$ indicates that the longest fragment only covers a small portion of the total dam length and therefore indicates a fragmented dam wall. This $\xi_{longest}$ is mainly used to easily grasp different dam fragmentations. For the actual dam selection the fragmentation value ξ_{above} is used which compares the combined lengths $l_{frag_{above_x}}$ of the watershed line parts x above the current dam height but in between dam wall parts (i.e. the "natural" part of the dam structure) with the total dam length l_{total} as defined in equation 6.

$$\xi_{longest} = 1 - \frac{l_{frag_{longest}}}{l_{total}} \tag{5}$$

$$\xi_{above} = \frac{\sum_i l_{frag_{above_i}}}{l_{total}} \tag{6}$$

To get an estimation of the dam wall construction costs C_{AUD} all fragments are viewed as parts of equal heights. Therefore the estimation assumes a wall being constructed for each fragment x with the length of the fragment l_{frag_x} and the maximum dam height needed in the fragment h_{frag_x} via equation 7. The resulting cost is in million AUD and should not be used as exact constructing cost for a dam project and will therefore only be used to compare the economic feasibility of different dam solutions (Petheram et al. 2017).

$$C_{AUD} = \sum_i 0.0039 \cdot (h_{frag_i})^{1.5681} \cdot (l_{frag_i})^{0.6148} \tag{7}$$

All dam and reservoir information generated for the different dam heights is saved as an CSV to ease further automated usage of the calculation results.

2.3.4 Best dam height selection

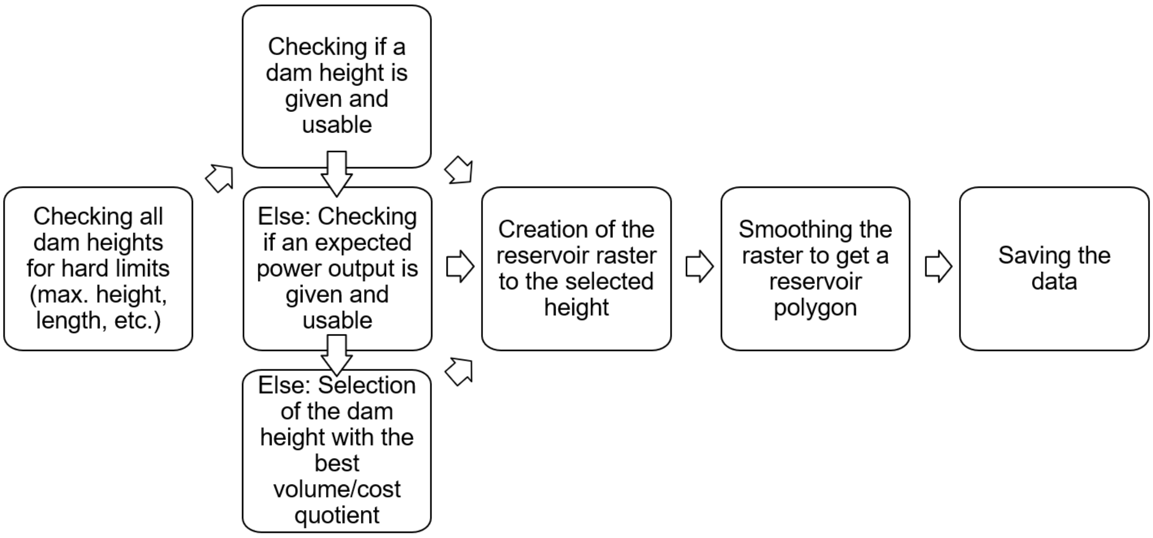

Figure 5: Diagram of the best dam selection process.

The selection of the best dam height out of all the calculated dam heights tries to asses the height via three different criteria as seen in figure 5. The first two criteria require additional data (i.e. listed dam height or power output) that may be present for a certain dam, while the last approach needs no additional information and just uses the best reservoir volume per estimated building cost quotient. Therefore the database needs to be checked first whether it contains a dam height or a listed power output for the certain dam. If it contains a height, the value can be used as it is and needs no further pre-calculations. A given power output P (in kW) on the other hand needs to be calculated into a dam height h_{kW} (in m) using the *maximum annual* discharge (in $\mathrm{m^3/s}$) and a rearranged version of equation (4).
Before the best height can be selected all possible dam heights are checked for general plausibility using the checks shown in table 1. If no possible dam heights remain after the check, the check is done again using soft filters. They only check the dam fragmentation ξ_{above} and the dam length l_{total} using a 30 % higher maximum length. When no heights remain using the soft filters no dam is considered viable and the process stops with no reservoir for the given dam.
The best dam selection then tries to use the listed dam height. If a height is given an the difference between the closest viable dam height and the given dam height is less than 50 % of the given height, the closest viable dam height is selected. Else the next criterion utilizing the power output derived dam height h_{kW} is applied in a similar way. If a power output is given and the difference between the closest viable dam height and the derived dam height is less than 50 % of the derived height, the closest viable dam height is selected. Dams with given or derived heights smaller than the minimum dam height ($5\,m$) are just assumed to be as high as the minimum dam height.

If both values are not listed in the database or differ to much from the dam heights the topography around the dam position allow, the third criterion is used. Here all viable dam heights are considered and the quotient of the reservoir volume V and the estimated cost C_{AUD} is calculated for each height and the height with the highest quotient (V/C_{AUD}) is selected.

Table 1: List of dam and reservoir values checked for each possible dam height before the best dam selection. All values shown can be edited in the initialization part of the script code.

dam height h	$\leq$	5 m
dam height h	$\geq$	300 m
dam length l_{total}	$\leq$	8,000 m
dam "area" $h \cdot l_{total}$	$\leq$	300,000 m^2
dam fragmentation ξ_{above}	$\leq$	20 %
reservoir area A	$\leq$	10,000 $\cdot 10^6 m^2$
reservoir volume V	$\leq$	250,000 $\cdot 10^6 m^3$

The selected dam height is then used to calculate the estimated reservoir shape using the filled elevation model. The resulting "boxy" raster is unrealistic for an actual reservoir. Therefore the raster boundry is first cleaned then resampled to a higher resolution to smooth the edges. To further polish the result a lowpas filter is used, the polygon is derived and then smoothed again. The resulting polygon now looks natural and still spans the area of the calculated reservoir raster.

2.4 Main changes

Compared to the estimation process by Pollmüller (2019) the main changes shown in table 2 are introduced to improve the calculation.

2.5 Remaining Problems

The reworkings in this work mainly targeted crashes, bugs and errors in the process while the general process remained unchanged.

2.5.1 Dam wall line

One major problem remaining in the process is the dam wall line following the local watershed outline. This causes derived dam wall lines to 'bend' downriver and therefore sometimes exceeding the thresholds listed in table 1. Here would a process building straight or concave dam wall lines proof favourable.

2.5.2 Best dam strategy

The decision of the best dam still highly relies on the given dam height or power output. Here a new strategy analysing the topography or the volume changes per additional wall

Table 2: Main changes between the old and the new version on the estimation script.

	old	**new**
Discharge	WaterGAP as grid	WaterGAP as part of the RiverATLAS
Watershed	Calculation in a certain square area around the dam position	Calculation in the area around the upriver stream segments
Watershed pour point	Highest flow accumulation in a proximity around the dam position	Main stream closest to the dam position
Elevation profile for the dam wall	Uses the longest connected profile around the watershed (may result in errors!)	Uses the lowest profile around just the dam position
Output CSV	Outputs a human readable CSV containing the results	Outputs a human and a machine readable CSV
RAM usage	Saves most working data in the RAM	Tries to save everything on the disc to save RAM
Best dam selection	Estimated drop height → cost/volume	given height → estimated drop height → cost/volume

height meter may prove advantageous because it would reflect the by-hand derivation process more closely which produced good results.

2.5.3 Crashes

The script and ArcGIS may crash during the derivation process with the last message in the log file being:
[Message]: Differs between 'high' and 'low' flow values and snaps the dam to the closest 'high' value
This crash happens with no error message or notice of any kind and is not reproducible. It may occur after the first couple dams or late into the process.
If it happens it is best to restart the computer and restart the process. It is very unlikely to crash again even when the same data and parameters are used.

3 Results & validation

The worldwide future reservoir estimation results are impossible to validate for dams not jet build. Therefore the results and the methods shortcomings are validated using already build hydropower dams.

3.1 Reservoir volume & area results

Using the given parameter values the script calculates 3570 of 3673 dams (97,2 %) successfully and adds an additional $3.35 \cdot 10^{12}\,\mathrm{m}^3$ of reservoir volume and $112 \cdot 10^{9}\,\mathrm{m}^2$ of reservoir area to the worldwide reservoirs. This low number of failed dams includes dams, that are located out of the DEMs borders, located too far away from streams (maybe pump storages) and dams for which no plausible dam wall could be derived. The additional reservoirs would increase the worldwide reservoir volume by 48.7 % compared to the GRanD reservoirs (see figure 6).

The mean volume is $937.8 \cdot 10^{6}\,\mathrm{m}^3$ (σ:$5402 \cdot 10^{6}\,\mathrm{m}^3$) and the mean area $31.5 \cdot 10^{6}\,\mathrm{m}^2$ (σ:$144.5 \cdot 10^{6}\,\mathrm{m}^2$). The large standard deviations σ compared to the mean values indicate a very heterogeneous set of dams with reservoirs sizes ranging from small to big.

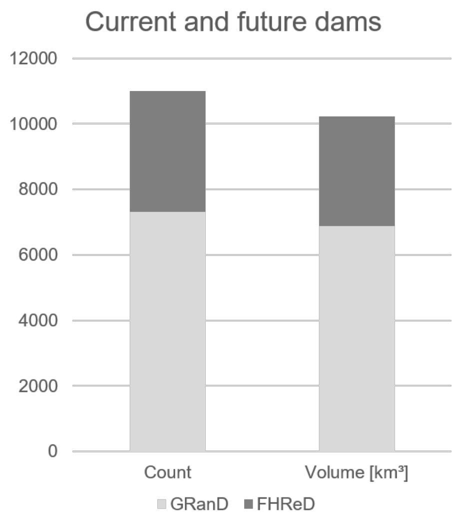

Figure 6: Bar graphs showing the count and volume of the GRanD reservoirs (Lehner et al. 2011) and the derived FHReD reservoirs (Zarfl et al. 2015).

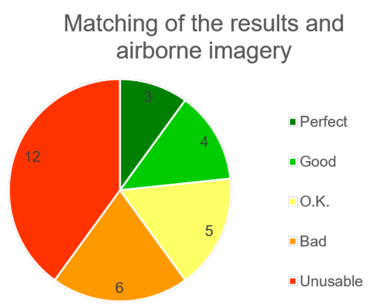

Figure 7: Pie graph showing the usability of the derived reservoirs compared to Google Earth imagery.

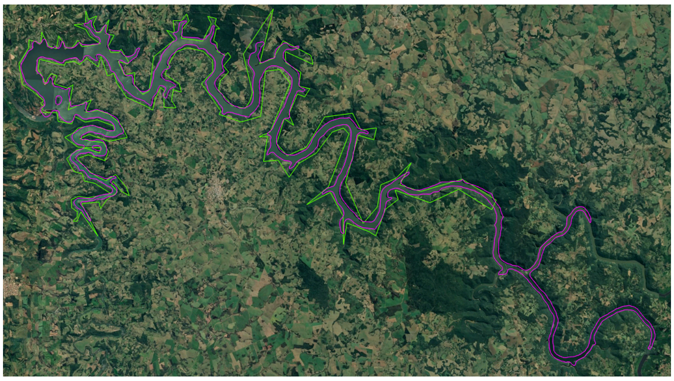

Figure 8: Comparison of the derived reservoir outline (pink; 'Perfect') and the GRanD reservoir outline (green; 'Good') using Google Earth imagery for the dam with GRanD ID 2446.

3.2 Airborne validation

Validating the results proofs difficult because no real 'ground truth' is available to compare the results to. Additionally future dams are not build jet and therefore not comparable to anything already present. To allow some validation 30 hydropower dams from the GRanD Database (Lehner et al. 2011) build since 2001 are selected. The dam height and power output are used to derive the reservoirs using the process and then compared to the GRanD reservoirs. The comparison shows large deviations in volume and area.
Using Google Earth airborne imagery of said dam reservoirs shows that the GRanD data can not be used as absolute ground truth for validation because their reservoir predictions also differ from the imagery. Therefore a validation classifying the reservoir outline predictions in usability classes is used. They range from 'Prefect' to 'Unusable' and are classified by hand and are shown in figure 7. Figure 8 to 10 depict some example classifications.
The first classification figure shows a very good derivation result where the result even outperforms the GRanD reservoir. Next next figure 9 shows a derived reservoir in the wrong valley due to a slightly off dam position in the used elevation model. This problem can be easily corrected in a real application of the process by relocating the dam in the right valley and recalculating the reservoir.
The last figure describes a reservoir that may still be in the filling phase. The GRanD reservoir on the one hand is too small and is likely taken from older imagery earlier in the filling phase. The derived reservoir on the other hand is way too big and therefore not usable to model the current reservoir. Such not fully filled reservoirs are especially hard to validate because it is impossible to predict the full extend of the reservoir from airborne imagery alone. For some dams and their reservoirs adjacent villages and other buildings can be used as upper bound for the reservoir extend. This can only be uses as a rough measure and villages may be relocated during dam construction. Therefore are reservoirs in the filling phase almost impossible to validate using only the available data and a validation can only be performed for the current water level.

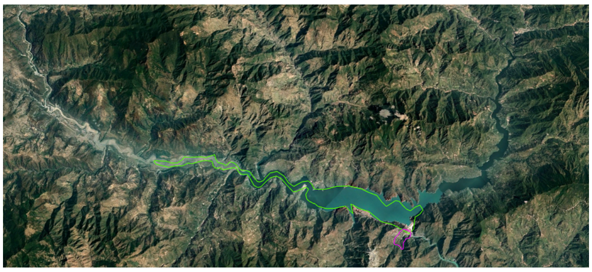

Figure 9: Comparison of the derived reservoir outline (pink; 'Unusable') and the GRanD reservoir outline (green; 'Bad') using Google Earth imagery for the dam with GRanD ID 4795.

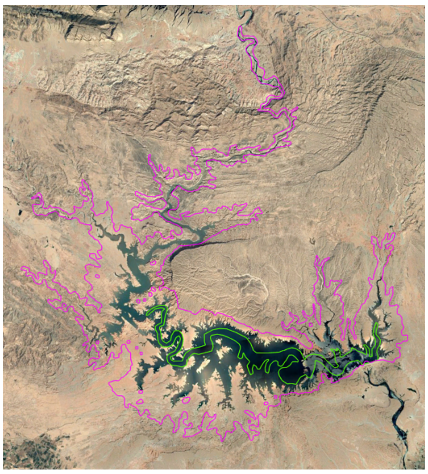

Figure 10: Comparison of the derived reservoir outline (pink; 'Unusable') and the GRanD reservoir outline (green; 'Bad') using Google Earth imagery for the dam with GRanD ID 4654.

3.3 DEM validation

The underlying DEM (Lehner et al. 2008) used for the estimation process shows problems in certain regards. On the one hand issues arise when the DEM is filled to make it hydrologically consistent to derive the watershed, which sometimes lifts the river bed in the filled DEM. This results in too high dam walls when using a given dam height. On the other hand are inconsistencies already present in the DEM data before filling. The figures 11 and 12 show two problems occurring in the DEM before or after filling. Out of the 30 derived dams for the validation most are in regions with no problems in the DEM and no problems due to hydrologically filling the DEM. For 6 of the dams the difference between the DEM and filled DEM is in the reservoir area way above 50 m resulting in derived reservoirs with false water levels. The overall deviation between the filled and unfilled DEM reaches from 700 m above to 390 m below the unfilled DEM. Those extreme values are only single pixels and most likely show errors in the original DEM.
Those problems mostly arise in areas with high altitude variations while in regions of equal elevation another problem is present. When a pixel has only pixels of the same altitude as neighbours no flow direction can be calculated which excludes the pixel from the watershed derivation. Because of this some areas which lie below the reservoir water level are not part of the reservoir due to no calculable flow direction. Nevertheless, the used DEM mostly produces accurate local elevations for the reservoir derivation.

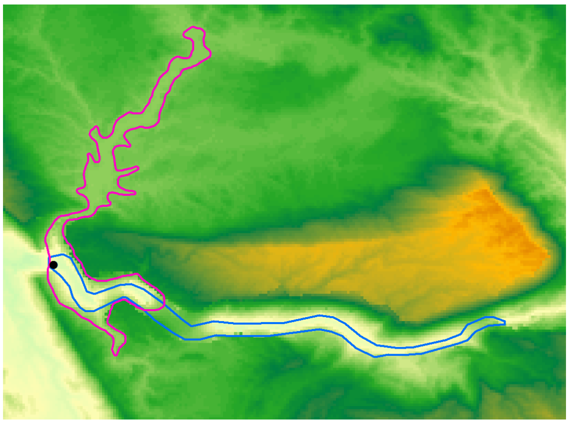

Figure 11: Example for a DEM problem where the river and GRanD reservoir (blue) is intercepted by a 300 m mound stopping the derived reservoir (pink) from reaching further in that direction. The DEM around the shown DAM 4656 reaches from high to low with colors orange-green-white.

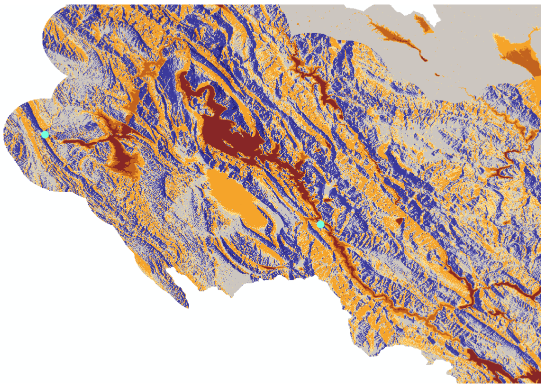

Figure 12: Example for a region where the DEM filling alters the elevation values significantly. Small local inconsistencies in the DEM can propagate to large differences in the filled DEM as shown here. Orange/brown pixels indicate the filled DEM to be more than 50/100 meters above the unfilled DEM while blue values show filled dam values more than 50 m below the DEM. The left dam in teal is the same dam as in figure 11 with GRanD ID 4656.

This small sample of dams is not enough to estimate the impact for all dams, therefore the difference between the filled DEM and the unfilled DEM is calculated for all used FHReD dam positions. The results of the analysis can be seen in figures 13 and 14. The difference is for most (56 %) dams smaller than ±10 m making the reservation estimation results quite viable. 13 % of the dams a located in areas with a difference between the filled and unfilled DEM of above ±50 m introducing large uncertainties to the estimation process and therefore the resulting reservoirs. Such high deviations only occur in mountainous region where the filling process has to alter the DEM more to accomplish a sink-free elevation model for the watershed derivation.

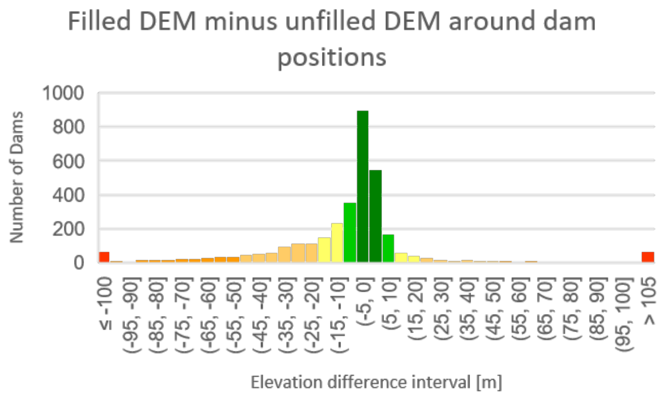

Figure 13: Bar graph of the differences between the filled and unfilled DEM on the worldwide dam positions grouped in 5 m intervals (only every other interval is labelled in the chart). A positive difference indicates that the filled DEM is higher than the unfilled DEM, a negative difference that it is lower.

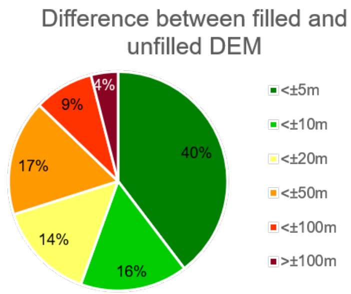

Figure 14: Pie chart aggregating the absolute differences shown in figure 13 as percentages of all dams.

4 Discussion & Outlook

The one main goal to generalize the process to other regions and minimize failed dam executions seems to be successful. The other goal to calculate more dams in one run was also successful, but could be further improved. The current state of the script can derive roughly 350 dams in one run, based on the number of basins in which the dams are located. More basins mean more precalculations, more RAM usage and therefore less dam calculable in one run.
Further improvements could be made using an x64 Version of ArcGIS (like ArcGIS Pro) and Python 3.x enabling the usage of more RAM. One of those ca. 350 dam runs take roughly 3-5 hours using the computers in the IUSF CIP-Pool and can easily be parallelized by using multiple PCs for different continents.
The process can and should be further improved by e.g. improving some of the problems mentioned in section 2.5 or other improvements described by Pollmüller (2019) in his master thesis.
The quality of the estimation could also be improved by improving the DEM problems covered in section 3.3. Here an elevation model with less noise and imperfections could highly benefit the process. This could be achieved either with a different DEM source or some preprocessing to minimize DEM filling problems.

Bibliography

Albers, A. S. (2017). A gis-based toolset for estimating reservoirs of future hydropower dams. Universität Osnabrück. Master Thesis.

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(9):494–502.

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10):93–94.

Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., et al. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, 6(1):1–15.

Petheram, C., Gallant, J., and Read, A. (2017). An automated and rapid method for identifying dam wall locations and estimating reservoir yield over large areas. *Environmental Modelling & Software*, 92:189–201.

Pollmüller, F. (2019). Automated gis-based estimation of future hydropower reservoirs. Universität Osnabrück. Master Thesis.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., and Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences*, 77(1):161–170.