Food and Agriculture Organization
of the United Nations

FAO
AQUASTAT
Reports

Evaporation from artificial lakes and reservoirs

Evaporation from artificial lakes and reservoirs

Prepared by: **Amit Kohli** and **Karen Frenken**
AQUASTAT Programme, FAO [1]
July 2015

---

1. Introduction

Dams and their associated reservoirs provide many services, including water storage, flow regulation, navigation, hydropower, in-stream and off-stream uses, flood protection, amongst others. However, these artificial lakes and reservoirs evaporate more water than the natural surface water flow before the dam was built, because dams generally increase the surface area of the body of water. This means that more water is exposed to air and direct sunlight, thus increasing evaporation. This "lost" water is referred to as consumed, because it is removed from the system. In some cases, this water consumption can be quite substantial.

AQUASTAT gathers country-level information about dams and reservoirs (see Section 2), but information about evaporation from artificial lakes and reservoirs is rarely reported. Due to its importance, AQUASTAT has estimated the evaporation for all artificial lakes and reservoirs that are available in the geo-referenced dam database it maintains. This exercise is a very rough estimation, the limitations of which are documented in Section 3 and Section 5, and it thus should be considered as an 'order of magnitude' study. As always, AQUASTAT welcomes feedback which would help improving the information provided.

2. Inputs

The analysis was performed using the following inputs:

A. AQUASTAT geo-referenced database of dams: a database spanning 157 countries, 14 200+ dams, 6 700 km^3 total capacity. This database holds the following information:

- Country
- Name of dam
- Alternate dam name
- ISO alpha- 3
- Administrative unit
- Nearest city
- River
- Major basin
- Sub-basin
- Completed/operational since
- Dam height (m)
- Reservoir capacity (million m^3)
- Reservoir area (km^2)
- Sedimentation (latest known) (%)

---

[1] The AQUASTAT Programme of FAO collects, analyzes and disseminates information on water resources, water uses, and agricultural water management. Water usage time series per country can be observed in the AQUASTAT database. Questions and comments can be directed to aquastat@fao.org.

The authors would like to thank **Michael Kao** and **Alex Matrunich** for helpful discussions during this analysis.

- Purpose: irrigation, water supply, flood control, hydroelectricity (MW), navigation, recreation, pollution control, livestock rearing, other
- Decimal degree latitude
- Decimal degree longitude
- National reference(s)
- Other reference(s)
- Comments

B. Global map of monthly reference evapotranspiration - 10 arc minutes: Grid with estimated reference evapotranspiration per month with a spatial resolution of 10 arc minutes.

C. Dams within Open Street Maps (OSM): See Section 4

D. dams within Wikipedia: See Section 4

3. Methodology

This section describes the methodology followed in order to estimate the evaporation in depth for each artificial lake and reservoir behind a dam. These values then then were converted to volume and summed for all dams within each country to generate a national level evaporation value. In order to approximate the evaporation from each dam, the ET_O for each dam reservoir was gathered from the map of Reference Evapotranspiration, using the coordinates of each dam (see Section 5).

The evaporation of each dam reservoir was calculated by using Equation 56 of FAO (2000) as follows:

$$ET_c = K_c \times ET_O \quad \textbf{[Equation 1]}$$

Where:

ET_c = Crop evapotranspiration in depth (mm/day)
K_c = Crop coefficient (-)
ET_O = Reference evapotranspiration (mm/day)

For open water, assuming $K_c = 1.00$ (see Section 5), Equation 1 simplifies to:

$$ET_c = ET_O \quad \textbf{[Equation 2]}$$

Considering that there is no transpiration element to this analysis, Equation 2 becomes:

$$E_d = ET_O \quad \textbf{[Equation 3]}$$

Where (also converting the period from day to year):

E_d = Dam reservoir evaporation in depth (mm/year)
ET_O = Reference evapotranspiration (mm/year)

And in volumetric units and converting units:

$$E_v = ET_O \times A \times 10^{-6} \quad \textbf{[Equation 4]}$$

Where:

E_v = Dam reservoir evaporation in volume (km^3/year) [2]
ET_O = Reference evapotranspiration in depth (specific to each dam) (mm/year)
A = Reservoir surface area (specific to each dam) (km^2)

---

[2] 1 km^3 = 1 000 million m^3 = 1 000 000 000 m^3

Thus, the evaporation in volume was calculated for each dam by multiplying the surface area of each dam's reservoir by the ET_O for that geographic area. While this approach is intended to be used for crops and not open water, it does provide a way to quickly estimate the evaporation for the whole world with minimum data inputs. Of course, where accurate evaporation estimations like open pan measurements are available, it is advisable to use these methods instead of this equation. However, this information is only available for a miniscule fraction of dams. The error introduced by the use of this method is discussed in greater detail in Section 5.

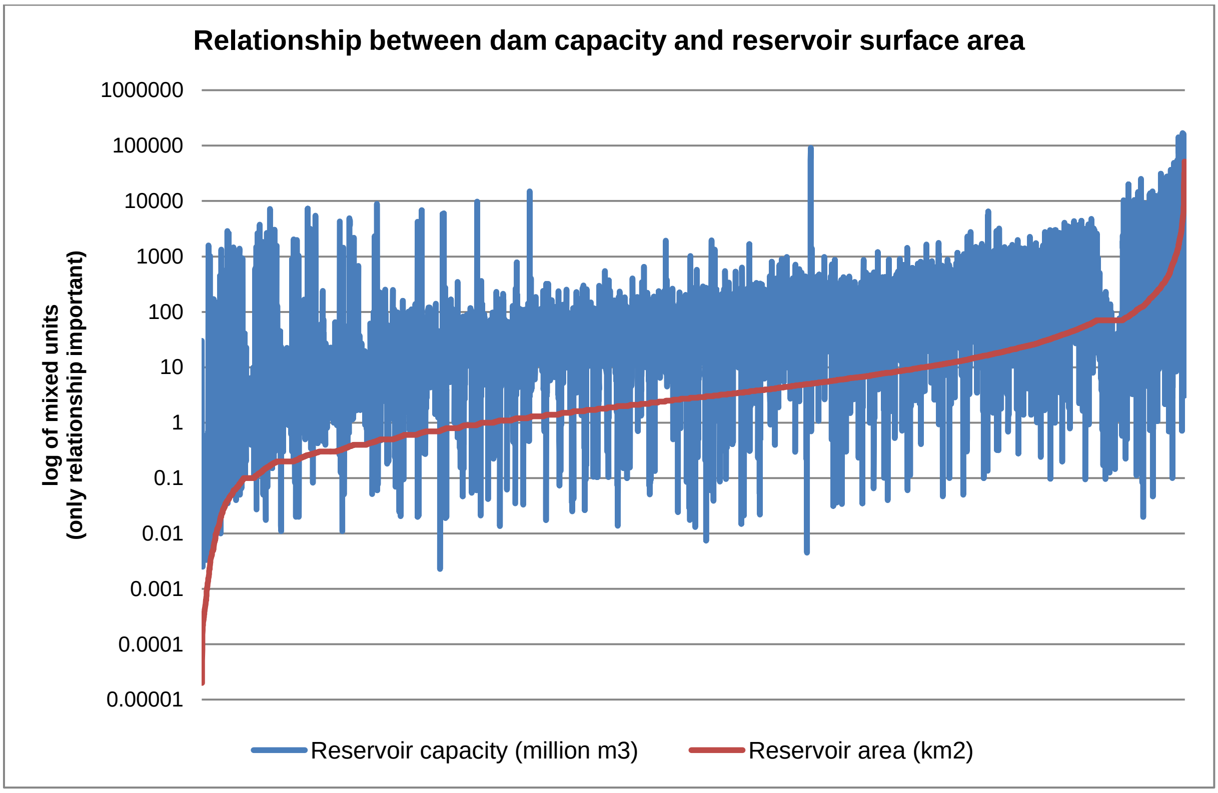
Therefore, what is required to calculate the evaporation from each dam are the coordinates (for the ET_O), and the surface area (A). This presented a challenge for AQUASTAT because of the 14 200+ dams held in the AQUASTAT database: (i) only 13 589 dams have data on capacity; (ii) of those, only 9 691 dams have data on surface areas; (iii) of those, only 8 091 dams have coordinates. Considering the big omissions, these gaps had to be filled.

To fill the missing coordinates, other geographic information was used to approximate these dam locations, using the following methodology:

- If the city and the state and the country are known: use that location
    - If the state and the country are known: use the center-point of the state
        - Failing this, the country center-point was used

The missing reservoir surface areas were much more complicated to fill. Given the vast variety of dam shapes and sizes, we were unable to find any meaningful proportional relationship. A quick depiction of the irregular relationship is displayed below in Figure 1. Consider that this chart shows the log of the relationships, which should make clear the poor predictability of the area given the capacity.

FIGURE 1



The following methodology, developed by GRanD (Global Reservoirs and Dams database) was also attempted:

$$V = 0.678\,(A\cdot h)^{0.9229} \quad \textbf{[Equation 5]}$$

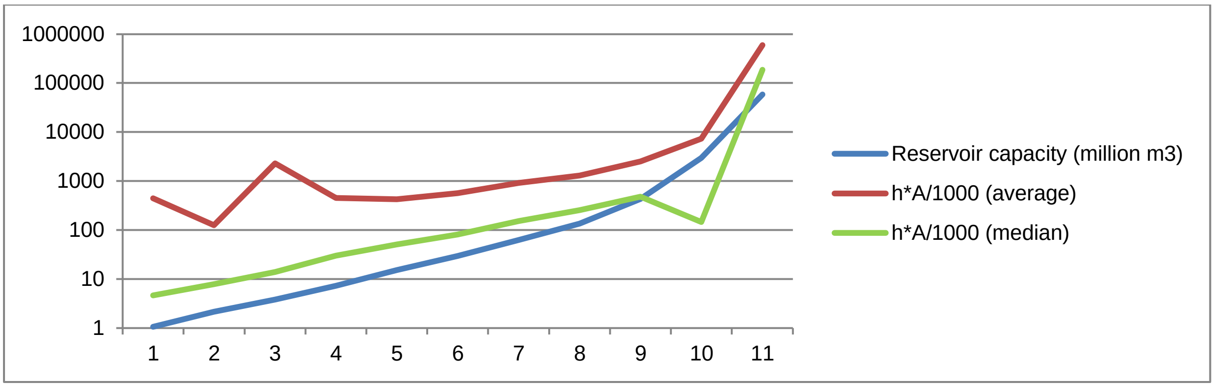
$$V = 30.684\,A^{0.9578} \quad \textbf{[Equation 6]}$$

Where: V = Reservoir volume (10^6 m^3)
A = Reservoir area (km^2)
H = Dam height (m)

These equations were developed by GRanD in order to estimate missing reservoir volumes if both area and dam height were available, or if only the reservoir area was available. They arrived at the coefficients by statistical regression analysis of the GRanD database (5 824 dams). Equation 5 was more accurate (R^2 = 0.92 for Equation 5 vs R^2 = 0.80 for Equation 6). For our purposes, we reversed these equations in order to calculate A.

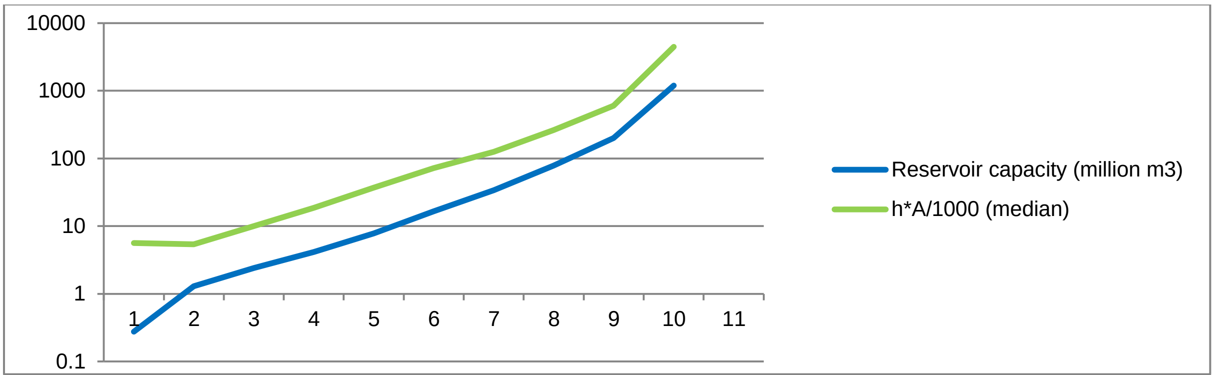
Unfortunately, for our analysis, this approach did not yield sufficiently useful results due to the fact that we have almost three times more dams and they include very small dams as well. Even after the regression coefficients were recalculated, the correlation was low. We decided to split the dams into ten categories based on their capacity. Figure 2 shows the average and median by category to check the correspondence at different intervals. As can be seen, the median is a bit closer, but fits unevenly at different categories, including a very large non-linearity. Therefore this model was rejected.

FIGURE 2



The dataset was cleaned a bit by removing dams where we knew the reported data for capacity, area, and height was impossible. Also removing all dams that didn't have information for ALL (V,h,A) in order to only use the best fit possible, improved the correspondence Figure 3).

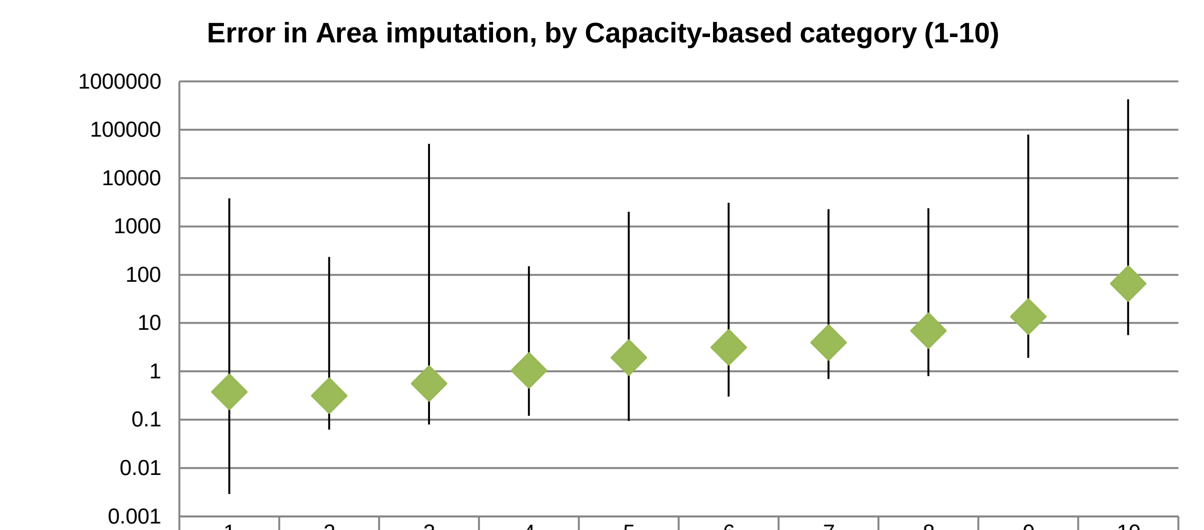
FIGURE 3



The method ultimately opted for was imputation of results based on the median of each category. For our needs it doesn't matter that the area be accurate for every dam, we just want to minimize the error caused by guessing an area for any given dam capacity. Therefore, imputation reduces the risk of guessing a wildly inaccurate surface area.

Figure 4 shows the correlation between REAL capacities and modelled capacities using the GRanD approach and imputation. In order to simplify the representation, the dams were first broken up into categories based on their dam capacity (in an effort to minimize the variability).

FIGURE 4



	1	2	3	4	5	6	7	8	9	10
High	3820	235.07	51110	150	2000	3091	2272	2375	79271	426009
Low	0.0029	0.062	0.08	0.12	0.09356	0.3	0.7	0.8	1.9	5.6
◆ Used value	0.372	0.31	0.5565	1.04668	1.927	3.125	3.925	6.89	13.503	65.55

Once the area and reference evapotranspiration for each dam were available, the evaporation was calculated using Equation 4.

The evaporations for all dams were aggregated into a global total that was **865 km^3/year**. However, this assumes that all dams are filled to capacity, which is clearly not the case. AQUASTAT decided to assume that all dams are half-full. Since surface areas of dams correlate almost exactly with dam capacities, we reduced this full capacity number by 50 percent. An additional 10 percent reduction was administered in order to account for the fact that rivers had also SOME evaporative loss before the dam was constructed. Therefore,

Actual Evaporation = 0.4 x Maximum Evaporation **[Equation 7]**

Therefore, AQUASTAT has calculated that the actual evaporation from artificial lakes and reservoirs is 346 km^3/year.

This value for each country and each year (taking into consideration the year a dam became operational as per our dams database) has been introduced in the database. While it might seem silly to worry about the minutia of details and then at the end to simply take a fraction of the total, readers and users of the data are reminded that this exercise was made to determine how water is being evaporated from reservoirs for every country in the world. Whether this analysis is overestimated or underestimated is less relevant than to treat all countries equally, which means keeping the data quality as high as possible for as long as possible.

4. Data fullness

In this exercise we take all known dams per country and aggregate the evaporations from their reservoirs. This introduces a large reporting bias because the more dams we know about, the higher the evaporation will be (see Section 5). In order to minimize the reporting bias, AQUASTAT attempted to see if there were any other dams that had capacity, surface area and coordinates information, in addition to the ones available in its database.

In order to cast the net wide, dams on Wikipedia and Open Street Map were consulted. This is how the information was gathered:

- Dams within Open Street Maps (OSM)
  - Harvested dams by using Overpass Turbo (http://overpass-turbo.eu/), searching for Nodes, Ways, and Relations containing the word "dam" within each bounding box. Bounding boxes were iteratively drawn in order to minimize the number of queries, while not returning an error[3]. This set was reduced by only keeping:
    - Nodes that had a name (in any language), and coordinates
    - Ways that had a name (in any language), and contained a Node with coordinates. For the sake of this exercise, the coordinates of the first Node were assigned to the Way. Taking the centroid of all Nodes would be more correct, but added complication with no meaningful benefit.
  - This final set of "useful dams" consisted of 58 668 dams with the following metadata:
    - Name
    - Decimal degree latitude
    - Decimal degree longitude
- Dams within Wikipedia
  - Harvested by downloading all wikipedia titles from dbpedia.org (accessed in June 2015). Since the data on dbpedia was not in sync with (and older than) the data on wikipedia, the list of dbpedia links was modified to be compatible with wikipedia, and all wikipedia had their "blue box" information scraped. There were several false-positives as well as inconsistent content within the blue boxes. This set having a lot of inconsistent data, it was cleaned.
  - The final set of "useful dams" consisted of 5759 dams with 32 variables (most sparsely filled). The important variables were:
    - Total capacity
    - Surface area
    - Decimal degree latitude
    - Decimal degree longitude
  - The total capacity and surface area were reported in a wide variety of units, so conversion had to be performed to ensure the data all used the same units.

---

[3] A better way to harvest this information was later identified: In R, create a loop for each latitude (360) drawing a thin bounding box spanning all longitudes, and run iteratively a script like this:

```
query <- 'http://overpass.osm.rambler.ru/cgi/interpreter?data=[out:json][timeout:2500];
          way["waterway"="dam"][~"^nam[e|e:en|e:fr|e:es]"~"."];out;'
response <- httr::GET(query)
httr::content(response)
```

The above is just for Ways. The same would have to be done for Nodes.

The dams of AQUASTAT, Open Street Map and Wikipedia have been combined into a massive Google Earth.kml file to examine better the geographic correlation. This file is available here: http://www.fao.org/nr/water/aquastat/dams/all.merged.kml (warning 200+ Mb file). In it:

- ✓ Blue LEFT arrow is an OSM dam
- ✓ Red DOWN arrow is a Wikipedia dam
- ✓ Green UP arrow is an AQUASTAT dam, the coordinates of which have been verified
- ✓ Green DIAGONAL arrow is an AQUASTAT dam, the coordinates of which have been approximated using the methodology in Section 3.

The OSM and Wikipedia dams were used to supplement the information on AQUASTAT dams and also to identify dams that AQUASTAT did not have. Matching dams proved quite challenging due to the fact that all three sources can have quality problems, misspellings, different spellings, mistranslations, etc. String matching was attempted after removing accents and white space, which was only moderately successful. String matching algorithms based on string distance (for example, the Levenstein algorithm) were used, but even when cutting the threshold of distance down to 1 (the minimum string distance that's not equal strings... for example "cat" and "hat"), the matches were not meaningful. For example:

```
696                      Kodna Dam/Koda Dam  1
697                     Kerit Dam/Kurit Dam  1
698                       Vail Dam/Vaal Dam  1
699                     Tanda Dam/Tansa Dam  1
700  Allal Al Fassi Dam/Allal al Fassi Dam  1
701                 Karpura Dam/Karpara Dam  1
702                       Kodo Dam/Koda Dam  1
703                   Morris Dam/Norris Dam  1
704                  Baisha Dam/Baishan Dam  1
705                Hassan II Dam/Hassan I Dam  1
706              Mohammed V Dam/Mohamed V Dam  1
707                     Altus Dam/Almus Dam  1
708               Hamahara Dam/Tamahara Dam  1
709                     Isaka Dam/Isawa Dam  1
710                     Ikawa Dam/Isawa Dam  1
711                       Kori Dam/Bori Dam  1
712                        Smir Dam/Sir Dam  1
713                     Tansi Dam/Tansa Dam  1
714               Liujiaxia Dam/Lijiaxia Dam  1
715                       Nina Dam/Sina Dam  1
716                     Kebir Dam/Kebar Dam  1
```

This is due to the fact that there is a tremendous amount of dams, but most dams have relatively short names, therefore it is almost more likely to find an identically named dam from somewhere in the world than finding a misspelled dam. In the end, exact matches and matches based on search engine hit correlation were used to identify dams in other datasets. Unfortunately, AQUASTAT was unable to improve ANY area and/or coordinates information from these other two data sources. Just a few new dams were identified and added to the AQUASTAT dams database.

One suggestion that was not attempted that could have helped was to try to match dams based on cluster analysis of the coordinates. This would only have helped identify useful Areas, since we would be assuming that the coordinates were correct. In the end, after consulting the dam distribution, it did not seem worthwhile to perform this analysis.

5. Caveats

This analysis is riddled with wild approximations. Readers are cautioned to review this section carefully.

a. Regarding the use of ET_O and K_c to estimate evaporation (Equation 1), in FAO (2000) the K_c values for open water are as follows:

	K_c ini	K_c mid	K_c end
Open Water, < 2 m depth or in subhumid climates or tropics	1.05	1.05	
Open Water, > 5 m depth, clear of turbidity, temperate climate	0.65*	1.25*	

* *These K_c values are for deep water in temperate latitudes where large temperature changes in the water body occur during the year, and initial and peak period evaporation is low as radiation energy is absorbed into the deep water body. During fall and winter periods (K_c end), heat is released from the water body that increases the evaporation above that for grass. Therefore, K_c mid corresponds to the period when the water body is gaining thermal energy and K_c end when releasing thermal energy. These K_c values should be used with caution.*

The data for dams that was readily available to AQUASTAT did not include water temperature, air temperature, evaporation data, or dam geometry. Lacking this specific information, and in light of the fact that a detailed ET_O global map was available, AQUASTAT decided to use Equation 1, where $K_c = 1.00$ was assumed for simplicity. This would introduce an error in the deep portion of the reservoir of -35 percent and +25 percent depending on the season. For the shallow portion of reservoirs, the evaporation would be underestimated by 5 percent. Given the rough nature of this analysis, and comparing obtained results to published data, this error was considered acceptable, although the reader's and data user's attention is brought to these limitations.

b. The more dams are analyzed, the more evaporation from artificial lakes and reservoirs will be identified. This represents a source of error since countries for which information is scarce will inevitably have a low evaporation reported in the AQUASTAT database. This also means that since different information on dams is available in different countries, there are necessarily differences in final results directly attributable to the input file used. While this shortcoming is regrettable, AQUASTAT has done what is possible to supplement the data with dam information available online.

c. The TOTAL evaporation assumes that all dams operate at the design capacity, and therefore, the surface area of the water body within each reservoir is assumed to be filled to the design capacity. This assumption may not be valid as dams often operate at a capacity less than the design capacity. Unfortunately, statistics regarding actual capacity are not available for all dams (and also vary from year to year). Therefore it was not possible to find a consistent methodology through which to determine how much the evaporation is overestimated in each country. This is the single most important parameter in determining the global total, and future collaborators are encouraged to identify a better way to predict the percent fullness of each dam.

d. This analysis assumes that dams began operating at design capacity on the "completed on" year. This is a source of error since, depending on the dam, it may take several years to fill the reservoir to the design capacity volume. This time-lag is present for all countries and therefore applies equal error throughout every country.

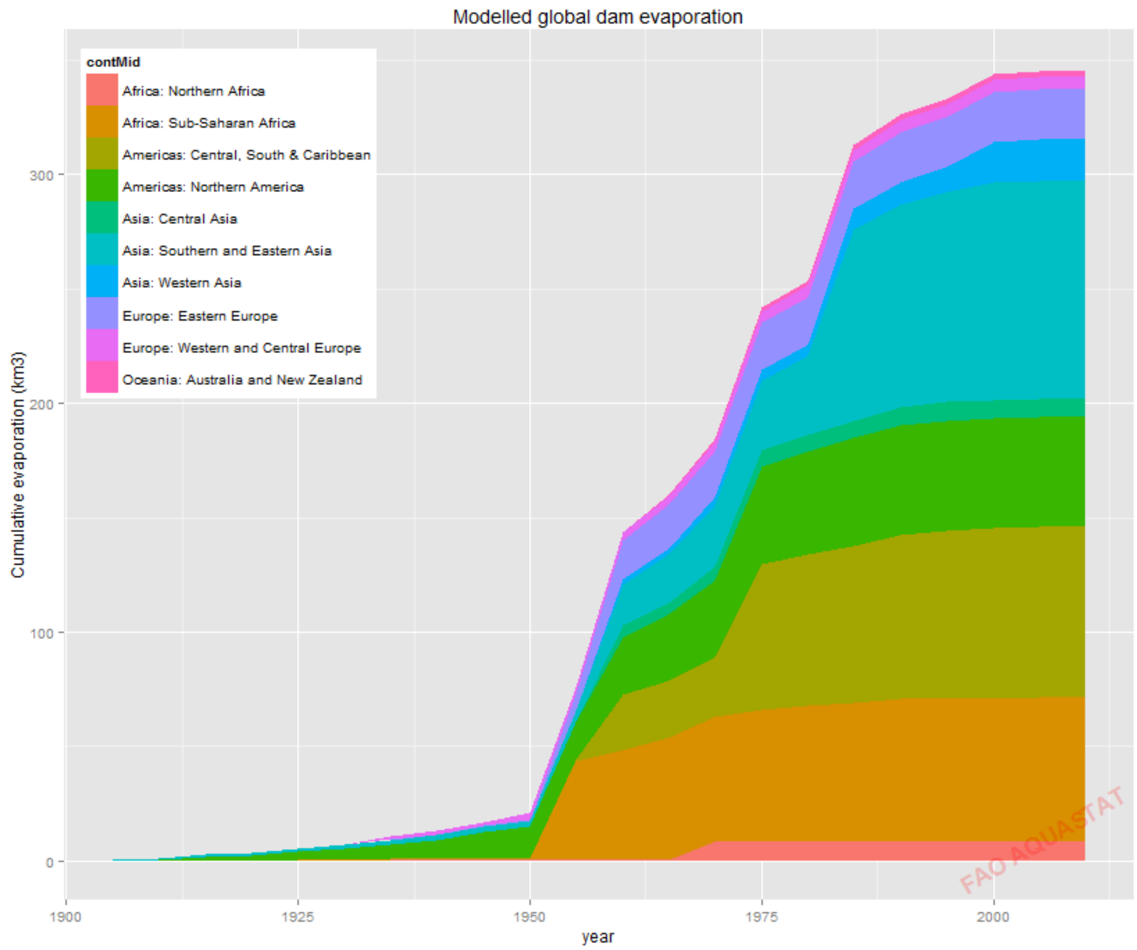
e. It is assumed for the sake of this analysis that the entire surface area of the reservoirs is anthropogenic, but that is not true. Rivers do have a surface area, and in some cases the surface area is not insubstantial. For the purposes of this analysis, we reduced 10 percent from the total evaporation for each country in order to account for this.

f. Another important caveat is that the data quality for this dataset, even in AQUASTAT which always performs a rigid quality control, is not ideal. For OSM and Wikipedia, obviously without a central authority verifying this data, the quality of the information can be suspect and should be not be just taken for granted but considered with extreme caution.

6. Conclusion

AQUASTAT has undertaken a FIRST STEP towards identifying the evaporative losses from artificial dams and reservoirs. This analysis and the output data are only intended to provide a general idea of the evaporation in each country.

The following Figure shows a time-series of evaporative losses by major AQUASTAT region, made possible by this analysis:



7. References

FAO. 2015. *AQUASTAT geo-referenced dam database*. Accessed on 15/06/2015.

FAO. 2004. *Global Map Of Monthly Reference Evapotranspiration - 10 Arc Minutes*.

FAO. 2000. Crop evapotranspiration: Guidelines for computing crop water requirements. By: R. Allen, L. Pereira, D. Raes, M. Smith, *Irrigation and Drainage Paper 56*. FAO, Rome.

GRanD. 2011. *Global Dams and Reservoirs (GRanD) database – Technical documentation*.