

Replication Notes¹

February 2023

This file contains a description of all codes in the GDAT_code folder. These codes are used to generate dam catchment areas and analyze dam-induced trends in global water coverage. All .ipynb files require Jupyter Notebook to run, and all .do files require STATA.

- **01_correct_gdat_coordinate.ipynb** creates a new GDAT database with dam coordinates modified to fit HydroRivers lines. The database merges the original GDAT database with several continent-scale interpolated databases (see code 02). The merged GDAT database contains all dam attributes and revised interpolated coordinates, with erroneous interpolations removed.
- **02_snap_gdat_river.ipynb** generates interpolated GDAT coordinates in which the dam locations are snapped to the closest HydroRivers line. This would allow subsequent analyses on watersheds to detect the presence of a dam when calculating watershed characteristics using flow-direction data.
- **03_pcraster_dam_catchment.ipynb** calculates the catchment areas of each dam using software from PCRaster (<https://pcraster.geo.uu.nl/>). The code uses elevation data to calculate the local drainage direction of each pixel, then uses these directions to generate polygons for the upstream area captured by each dam (the dam's catchment area). These catchment-area polygons are saved as GDAT catchment area shapefiles.
- **04_training_completion_year.ipynb** generates a "training" dataset used for the completion-year imputation procedure (see code 11). The code subsets GDAT dams that have completion years within the Global Surface Water temporal scope (1984-2018 at the time of analysis). The code then uses Google Earth Engine to calculate the count of each type of water-cover pixel in the catchment areas of these dams for each year from 1984 to 2018. This creates a panel dataset with pixel counts for each year and each catchment area.
- **05_testAZ_completion_year.ipynb** and **05_testVG_completion_year.ipynb** both generate "testing" datasets for the completion-year imputation procedure (see code 12). The code operates in the same way as code 04, but uses GDAT dams that do not have completion-year data.
- **10_GDAT_Descriptives.do** generates all summary-statistics tables. This includes descriptive statistics on the distribution of the main purposes of dams, total dam counts by continents, completion years, etc. The code also generates charts that compare GDAT statistics with those of GRanD and AQUASTAT.
- **11_GDAT_imputeCompletionYear_train.do** uses data generated from code 04 to impute the completion years of dams based on water-coverage trends. The code uses the structural breaks method to calculate the completion year of each dam. Because the "training" dataset are dams that have completion-year data, results of this code can be compared with actual data to check for accuracy.
- **12_GDAT_imputeCompletionYear_test.do** uses data generated from both codes 05 to impute the completion years of dams based on water-coverage trends. The code follows the same procedure as code 11. Results from this code serve as the imputed completion years for dams without completion-year data.
- **13_GDAT_calculateChange.do** calculates summary statistics of pixel values before and after the completion year for each dam (actual year for "train," imputed year for "test"), and generates figures to visualize the changes in water coverage.

¹ Zhang, Alice Tianbo, and Vincent Xinyi Gu. 2023. "Global Dam Tracker: A Database of More than 35,000 Dams with Location, Catchment, and Attribute Information." *Sci Data*. Corresponding Author: Alice Tianbo Zhang (alice.tianbo.zhang@gmail.com).