

These Python scripts streamline the key procedures for (1) georeferencing dam records in the World Register of Dams (WRD) of the International Commission on Large Dams (ICOLD) (<https://www.icold-cigb.org>) and then (2) extracting their associated reservoir polygons from existing water masks. They were used to produce the Georeferenced global Dam and Reservoir (GeoDAR) dataset (Wang et al., 2021, in review) and refer to Wang et al. (2021) for more details. The scripts are free for use or modify, and we kindly request users to cite the GeoDAR paper. The authors claim no responsibility or liability for any consequences related to the use, citation, or dissemination of these scripts. For any questions, please contact Jida Wang ([jjdawang@ksu.edu](mailto:jjdawang@ksu.edu) or [gdb Bruins@ucla.edu](mailto:gdb Bruins@ucla.edu)).

Script files:

`Georeferencing_functions.py`: This script contains a list of customized functions used for georeferencing ICOLD WRD. See their names, purposes, inputs, and outputs in the script file.

`Reverse_geocoding_register.py`: This module performs reverse geocoding for each record in a georeferenced regional register/inventory using the latitude and longitude coordinates provided in the register. The inventory for Brazil, Relatório de Segurança de Barragens (Dams Safety Report 2017, <http://www.snisb.gov.br/portal/snisb/relatorio-anual-de-seguranca-de-barragem/2017>), was given here as an example (see “CadastroRSB2017\_Portal\_SNISB(v4).xlsx” in ancillary files). Reverse geocoding converts spatial coordinates to a nominal address with consecutive administrative levels. In this module, reverse geocoding was implemented using the cloud-based service through Google Maps geocoding API (<http://developers.google.com/maps>). See more about reverse geocoding and Google Maps geocoding API at: <https://developers.google.com/maps/documentation/geocoding/overview#ReverseGeocoding>. The returned address components were then appended to the original regional register. An example of the output (using the Brazilian register) is given as “CadastroRSB2017\_Portal\_SNISB(v4)\_revgeo.xlsx”. The output was then used as an input of “Geomatching\_ICOLD.py” for geo-matching records in ICOLD WRD.

`Geomatching_ICOLD.py`: This module geo-matches (table-associates) the dam records in a georeferenced regional register (with latitude/longitude) with the ICOLD WRD records for the same region. The input regional register is the output of “Reverse\_geocoding\_register.py” and thus has been through reverse geocoding to extend the address components. An example of the reverse-geocoded regional register is given as “CadastroRSB2017\_Portal\_SNISB(v4)\_revgeo.xlsx” (also see the description of “Reverse\_geocoding\_register.py”). In the output, each geo-matched WRD record will be assigned the latitude and longitude coordinates as documented in the regional register. Meanwhile the QA level of each geo-matched WRD record will also be labeled.

`Geocoding_ICOLD.py`: This module performs forward (or regular) geocoding for each record in ICOLD WRD and output all geocoded results and their quality scenarios. Forward geocoding converts a nominal address to a pair of geographic coordinates (latitude and longitude), and meanwhile, the name and administrative divisions associated with the geographic coordinates are also returned. In this module, forward geocoding was implemented using the cloud-based geocoding service through Google Maps geocoding API. See more about forward geocoding at: <https://developers.google.com/maps/documentation/geocoding/overview>. In the output, one WRD

record may have multiple geocoding solutions. These solutions will be ranked by "Geocoding\_QA.py" and the best quality result will be selected and labeled with its corresponding QA level.

Geocoding\_QA.py: This module loops through all geocoding solutions for each ICOLD WRD record and rank them based on the corresponding QA levels. For each unique WRD record, the geocoding solution with the best possible rank is written to the output with the associated QA label. Note that the input file (geocoding solutions) is the output of "Geocoding\_ICOLD.py", except that it was then further expanded with more detailed address components by another round of reverse geocoding. The script for this round of reverse geocoding is not provided here but is similar to "Reverse\_geocoding\_register.py".

Reservoirs\_retrieval.py: This module loops through the dam points to retrieve their associated reservoir polygons from a given water mask (such as HydroLAKES). The retrieval process involves two consecutive rounds, with the first using a 500-m spatial tolerance and the second using a 1-km tolerance. Each round consists of another three iterations to progressively optimize reservoir-dam association. The output is one-to-one relationship (one dam point associated with one reservoir polygon). Although our goal is to retrieve reservoirs as thoroughly as possible, there is no guarantee that the reservoirs for all dam points can be assigned.

More detailed descriptions are provided in each script file.

#### Ancillary files:

Countries\_lookup.csv: a lookup table for country ISO 3166-2 codes. This file is a required input for "Geomatching\_ICOLD.py", "Geocoding\_ICOLD.py", and "Geocoding\_QA.py".

US\_states\_lookup.csv: a lookup table for the ISO 3166-2 codes for US states. This file is a required input for "Geocoding\_ICOLD.py" and "Geocoding\_QA.py".

Korea\_states\_lookup.csv: a lookup table for South Korean provinces. This file is a required input for "Geocoding\_ICOLD.py" and "Geocoding\_QA.py".

CadastroRSB2017\_Portal\_SNISB(v4).xlsx: the original dam register for Brazil. This register is Relatório de Segurança de Barragens (Dams Safety Report 2017), archived in Sistema Nacional de Informações sobre Segurança de Barragens (SNISB, Brazilian National Dam Safety Information System), accessed at <http://www.snisb.gov.br/portal/snisb/relatorio-anual-de-seguranca-de-barragem/2017>. This file is the input of "Reverse\_geocoding\_register.py".

CadastroRSB2017\_Portal\_SNISB(v4)\_revgeo.xlsx: reverse-geocoded register for Brazil. This file is the output of "Reverse\_geocoding\_register.py" and the input of "Geomatching\_ICOLD.py".

#### References:

Wang, J., Walter, B.A., Yao, F., Song, C., Ding, M., Maroof, M.A.S., Zhu, J., Fan, C., Xin, A., McAlister, J.M., Sikder, M.S., Sheng, Y., Allen, G.H., Crétau, J.-F., and Wada, Y., 2021. GeoDAR: Georeferenced global dam and reservoir database for bridging attributes and geolocations. Earth System Science Data, in review.