Applied Machine Learning 2024 - Glacier ice thickness via ML/DL - handbook

Niccolò Maffezzoli

May 19, 2024

Overview

On Earth there are more than 200,000 glaciers. The thickness distribution of ice within the glaciers has been measured only in few thousands of them and very sparsely, due to the difficulty in performing measurements in such harsh environments. Therefore, ice thickness is typically obtained from physics based models that have been developed over the last few decades.

In this project you attempt to take a data driven approach and use the world-wise available ice thickness measurements to create a predictive model for ice thickness.

Training dataset

The released dataset is a numerical dataset consisting of ice thickness measurements and a set of informative features which is naturally well suited for a ML/DL regression model. However, you may want to explore other/you own additional data, and/or pursue a different modeling approach.

- You can see the glacier outlines here: `https://www.glims.org/maps/glims` (flag RGI green outlines)
- Millan model (ith_m feature) visualization: `https://ige-vis.univ-grenoble-alpes.fr/glaciers/`

The dataset is released in its original version and in 3 spatially averaged versions (per glacier), in which all features (and the target) in each glacier have been gridded to a 20x20, 50x50 or 100x100 lat-lon pixel grid. The gridded datasets have the additional constraint to use only data more recent than 2005.
You may consider the dataset of your choice.
The ground truth:

- The dataset contains ice thickness measurements taken on glaciers and ice caps (not ice sheets). The target variable (ice thickness) is taken from the Glacier Ice Thickness Dataset (GlaThiDa 3.1.0), while the other features have been extracted from other products.

Along with the ice thickness target variable, in the dataset you also find the modeling solutions of Millan et et (2022) and Farinotti et al. (2019) interpolated locally. These are the only two modeling efforts existing for all glaciers world-wide.

Features

Some features are local (change within the glacier), while some are per-glacier constants, i.e. characterize the glacier as a whole and do not change within the glacier outline. Below the list features contained in the dataset. Note that not all of them are useful.

Features from **GlaThiDa**. See Welty et al. (2020) and here for a complete description of the features.

- **GlaThiDa_ID**: survey identifier. Many measurements will share the same ID.
- **POLITICAL_UNIT**: country code.
- **GLACIER_NAME**: glacier name.
- **SURVEY_DATE**: survey date.
- **PROFILE_ID**: field related to change wrt GlaThiDa v1.
- **POINT_ID**: point identifier.
- **POINT_LAT**: measurement latitude coordinate.
- **POINT_LON**: measurement longitude coordinate.
- **ELEVATION**: measurement elevation.
- **THICKNESS**: ice thickness measurement from GlaThiDa (see here). It's the target. Unit: m. The dataset is known to contain some weird zero measurements, which you may/may not want to remove (see `https://gitlab.com/wgms/glathida/-/issues/25`).
- **THICKNESS UNCERTAINTY**: estimated random error of THICKNESS (m).
- **DATA_FLAG**: erroneous data will have non-Nan here.
- **REMARKS**: any other important information about the survey not included elsewhere.

Features from other sources:

- **RGI** (from 1 to 19): Randolph Glacier Inventory region, version 6. See `https://nsidc.org/sites/nsidc.org/files/technical-references/RGI_Tech_Report_V6.0.pdf`, page 65.
- **RGIId**: RGI glacier ID. E.g. $RGI60-11.01450$: RGI(version60)-11(region 11).xxxxx
- **Area**: area of glacier. Unit: km^2
- **Zmin, Zmax, Zmed**: glacier minimum, maximum and mean glacier elevations. From OGGM.
- **Slope**: mean glacier slope. From OGGM.
- **Lmax**: glacier maximum length. See `https://essd.copernicus.org/articles/14/3889/2022/essd-14-3889-2022.pdf`. From OGGM.
- **Form**: Glacier type. {'0': 'Glacier', '1': 'Ice cap', '2': 'Perennial snowfield', '3': 'Seasonal snowfield', '9': 'Not assigned' }. See `https://docs.oggm.org/en/v1.1/_modules/oggm/utils/_workflow.html`
- **TermType**: Terminus type. {'0': 'Land-terminating', '1': 'Marine-terminating', '2': 'Lake-terminating', '3': 'Dry calving', '4': 'Regenerated', '5': 'Shelf-terminating', '9': 'Not assigned', }. See `https://docs.oggm.org/en/v1.1/_modules/oggm/utils/_workflow.html`.
- **Aspect**: glacier mean aspect. Int number from 0 to 360. Bad values: -9. See definition.
- **dmdtda_hugo**: glacier mean mass balance from Hugonnet et al. (2021). Unit: kg/m^2yr.

- **elevation**: measurement elevation interpolated from theTanDEM-X Edited Digital Elevation Model. This is possibly more accurate than 'ELEVATION'.
- **slope_lat(lon)_xxx**: elevation gradients along latitude and longitude axes. "gfxxx" reflect the width (in meters) of a gaussian filter used to smooth the elevation map before taking the derivatives. Various widths are used. "gfa" is an adaptive width which is function of the glacier area.
- **curv_50, curv_300, curv_gfa**: curvature of the elevation map (see xrspatial.curvature). 50, 300 and gfa are the same as in the slope (width of gaussian filter).
- **aspect_50, aspect_300, aspect_gfa'**: aspect. See xrspatial.aspect.
- **smb**: mass balance from a mix of different methods. Unit: kg/m^2yr.
- **vx, vy, ...**: surface velocity of the ice along x (longitude) and y (latitude) axes. The pedex reflect the width of the gaussian filter used to smooth the velocity fields. Unit: m/yr
- **dvx_dx, dvx_dy, dvy_dx, dvy_vy**: velocity derivatives.
- **dist_from_border_km_geom**: closest distance of the point to any glacier border. Unit: km.
- **ith_m**: ice thickness predicted by Millan et al. (2022). Unit: m.
- **ith_f**: ice thickness predicted by Farinotti et al. (2019). Unit: m.

On top of the features deposited in the training dataset, the following additional features can be added: $elevation - z_{min}$, $||s|| = \sqrt{slope_{lat}^2 + slope_{lon}^2}$, $||s_{50}||$, $||s_{100}||$, ... and similarly the modules of the smoothed velocity. You may also want to consider informing the model by using the rgi region, or the latitude, or the surface temperature (the colder it is, the thicker the ice ? Surely that can be true in some regions). Different types of satellite products are available from space missions, e.g. Sentinel. **Hint**: while the slope along the latitude and longitude axes is not really important, the module of the slope is very important (all physics-based models include the slope - the higher the slope the more shallow is the ice, since it flows away due to gravity).

Useful packages

For plotting purposes I'd recommend the following python packages:

- Geopandas, xarray, rasterio, rioxarray
- OGGM package: this is in itself a glacier model, but contains a lot of useful products as well (in fact some training features have been imported from OGGM). You can for example import all RGI glacier outlines.

Notes

- If you deal with satellite products, mind the concept of different projections, and the fact that not all projections preserve distance.