esa climate change initiative

European Space Agency

Climate-Space X-ECV

Karakoram Anomaly

Fit for Purpose Report (F4PR)

Prepared by:	X-ECV consortium
Contract:	4000145271/24/I-LR
Name:	X-ECV_KA_D2.2_F4PR
Version:	0.5
Date:	02.02.2026

Contact:
Frank Paul
Department of Geography
University of Zurich
frank.paul@geo.uzh.ch

Technical Officer:
Anna Maria Trofaier
ESA Climate Office

University of Zurich[UZH] · University of Oslo · ETH zürich · enveo · University of Bristol · DWD · CLS Collecte Localisation Satellites · GAMMA REMOTE SENSING

Document status sheet

Version	Date	Changes	Approval
0.1	12.05.2025	Initial draft	
0.2	06.06.2025	Extended draft	
0.3	25.09.2025	Consortium input included	
0.4	11.11.2025	Comments from TO P. Fisher included	
0.5	02.02.2026	Final version with edits in Section 3.8	

The work described in this document was done under ESA contract 4000145271/24/I-LR (Climate-Space X-ECV: Karakoram Anomaly). Responsibility for the contents resides with the authors who prepared it.

Author team:
Frank Paul (GIUZ), Thomas Nagler, Gabriele Schwaizer, Nico Mölg (ENVEO), Andreas Kääb, Désirée Treichler (GUIO), Lander van Tricht, Matthias Huss (ETHZ), Tazio Strozzi (Gamma), Fabien Maussion (UB), Rainer Hollmann (DWD), Christophe Fatras (CLS)

Glaciers_cci+ Technical Officer at ESA:
Anna Maria Trofaier

Related documents

Acronym	Acronym	Name	Version	Date
[RD1]	EID	Deliverable 1.2: ECV Inventory Database	0.2	30.03.2025

Table of Contents

Purpose

This is the Fit for Purpose Report (F4PR) of the X-ECV Karakoram Anomaly project. It presents the datasets that have been selected for processing along with the modifications required to make them useful for modelling and analysis. The focus is on the spatio-temporal subsetting and harmonization of datasets that differ widely in character (e.g. vector/raster, coverage, resolution, format/projection). We start here with the overall set-up and datasets as required for initial modelling and analysis, but also present the foreseen later additions.

1. Introduction

To investigate the Karakoram Anomaly as described in the project proposal, we need to handle and use datasets of a very diverse nature, i.e. with different spatial and temporal coverage/resolution, completeness and format (see EID [RD1]). These differences need to be harmonized to (1) use them in the mass balance model (MBM) and (2) perform time series analysis. As a third requirement we need to prepare and analyse them for different regions. For example, a more sophisticated mass balance model (MBM) will first be applied in a smaller test region before the more generalized MBM will be applied to all glaciers in the Karakoram. This also requires having the related subsets of glacier outlines, digital elevation models (DEMs) and other meteorological and glaciological data available. We will further compare trends of climatic variables inside and outside the Karakoram. We have thus created a set of sub-regions for the related model testing and analysis.

For (1) and (2) we also need to know which datasets should be used (in the case several are available), which datasets are required (e.g. DEMs of different spatial resolution), which parameters have to be extracted (e.g. temperature at the surface, precipitation amounts, cloud cover) and how to handle data gaps (in space and time). Depending on the application, individual datasets might require re-projection, re-gridding and vector-raster or file format conversion. This document describes the selected datasets and the foreseen transformations.

As an overall structure, we first present the range of spatial and temporal scales along with the selected sub-regions. Afterwards, the selected (mostly static) glaciological datasets and the selection of the (highly variable) meteorological datasets is presented. At the end, we summarize the selection of datasets and their modifications in view of the two main applications, i.e. forcing the mass balance model and trend analysis.

2. Temporal and spatial scales

2.1 Temporal scales

Datasets come along at a range of temporal scales, i.e. they can be an average over a minute or hour up to a year or decade. While point measurements (e.g. climate station data) refer to a specific minute or hour, reanalysis data are available every 3h or as a 1-day average and can be further aggregated to monthly, seasonal or annual means for climatic analysis. Mass balance measurements are usually performed at a specific day, but might refer to two half or a full hydrological year and geodetic mass balances are usually multi-year to decadal averages. Table 2.1 provides a condensed overview of the scales. Apart from the temporal scale the datasets refer to also the temporal frequency can differ. For example, snow cover data derived from a satellite sensor has a precise date and time, but might only be available every two to four weeks. Finally, the datasets have differences in temporal range. For example, climate or reanalysis data might be available from 1950 to 2025 (e.g. ERA5-L), whereas geodetic mass balances (derived from ASTER) are available from 2000 to 2019 and Sentinel-2 images from 2016 to 2025. So, depending on the specific requirements, a complete time series has to be created from a diverse set of input datasets that might require a different processing. The latter might also require restricting the sampling to specific sub-sets (which are available for all datasets). As a general rule, for the MBM we focus on the temporal range 2000 to 2019 as critical datasets such as geodetic mass balances are available over this period. For the analysis of temporal trends the time period might be longer (e.g. from the 1970s to 2025).

For some datasets it could also be beneficial to mix datasets with differing temporal (and often also spatial scales). For example, near daily MODIS snow cover products might be combined with weekly to monthly snow cover maps to improve spatial and/or temporal extrapolation. As a general rule, the climate data input for the MBM needs to be at the mean daily scale whereas glacier outlines or DEMs will likely only be available from one point in time.

Table 2.1: Overview of the typical temporal scales of selected datasets (mb: mass balance).

ID	Type	Dataset	Description
T0	Minute	Field measurement (incl. meteo station)	Precision of field measurements
T1	hour	Reanalysis data (ERA5-Land)	Frequency of reanalysis & data field measurements
T2	Day	Mean value (meteo data, velocity)	Meteo aggregation for mb model, satellite image date
T3	Month	Mean value (meteo data, mb gradient)	Meteo variability, frequency of snow cover data
T4	Season	Mean value (meteo trends)	For meteo trend analysis
T5	Year	Mean value (meteo), annual mb	For meteo & glaciologic trend analysis
T6	Decade	Geodetoc mb, surge period	For mb model calibration, glacier characterization

2.2 Spatial scales

The spatial scale of the input datasets ranges from the point (e.g. climate station), to a glacier or test region (for MBM development), to the entire Karakoram (main focus of this project), to neighbouring regions (e.g. for comparison of trends), High Mountains Asia (climatologies, lakes) and the world (re-analysis data, geodetic mass balance). Hence, apart from the selection of the variables and time periods, spatial sub-setting is required before downloading related datasets. This mostly applies to global scale reanalysis data and regional scale mass balance

data or glacier outlines. Thereby, not all input dataset are extracted to the same sub-regions and – depending on the application – several datasets have to be extracted for different sub-regions. Data at the point scale will stay at this scale, but might be extrapolated vertically.

2.3 Sub-regions

For the calculations, comparisons and data analysis we have selected sub-regions at various spatial scales. Apart from point scale climate station data, we list the regions in Table 2.2 along with their extent and purpose. Sub-regions SR1 to SR4 are shown in Fig. 2.1.

Table 2.2: Overview of the selected sub-regions.

ID	Type	Application	Size (km)	Projection	Creation
SR1	Test region	Testing of the 3D (raster) MBM	31*37	UTM43	RGI7, Manual
SR2	Karakoram vector	Perimeter outline(1), glacier selection	305 *425	UTM43	Literature
SR3	Karakoram raster	15 tiles (1° by 1°) covering Karakoram	333*450	geogr.	Fishnet
SR4	DEM	Comparison with climate model DEM	393*500	UTM43	Manual
SR5	HMA sub-regions	Comparison of climate trends	variable(3)	geogr.	Fishnet
SR6	HMA Tibet	Coverage of lakes(2)	c. 2000*3400	geogr.	Literature

(1) from Bhambri et al. (2022), (2) from Treichler et al. (2019), (3) details are provided below.

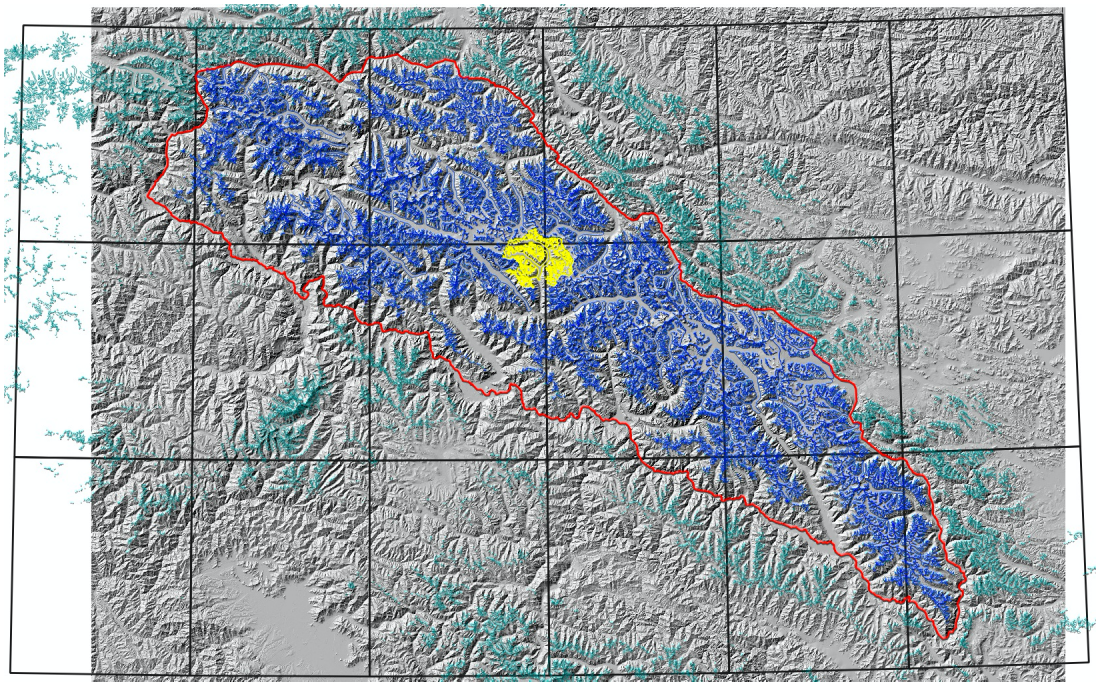

Fig. 2.1: Graphical representation of four sub-regions as listed in Table 2.2. SR1: yellow, SR2: red, SR3: black, SR4: grey hillshade of the DEM. Glaciers inside SR2 are shown in blue, glaciers outside are in darker green.

For analysis and comparison of climatic trends we have selected the following sub-regions:
(1) Karakoram: 74-79E, 34-37N (5° x 3°), see raster in Fig. 2.2;
(2) Hindu Kush: 71-74E, 35-37N (3° x 2°);
(3) Pamir: 72-76E, 37-40N (4° x 3°);
(4) Tien Shan: 78 (or 80)-82E, 41-43N (4° or 2° x 2°);
(5) Western Kunlun Shan: 79-83E, 34-37N (4° x 3°);
(6) Western Himalaya: 76-78E, 32-34N (3° x 2°).

3. Dataset selection and modification

3.1 Glacier outlines

For the study region Karakoram we have glacier outlines from (a) RGI 6.0 and (b) RGI 7.0. The outlines from RGI 6.0 served as the base for several glacier-related products (e.g. hypsometry, ice thickness, debris distribution and thickness, elevation changes) and glacier models such as GloGEM have converted the information from these products to a vector database, i.e. they are available per RGI 6.0 glacier-ID (e.g. geodetic mass balance) or glacier-specific elevation bands (e.g. debris cover). However, they suffer from partly mapped seasonal snow (i.e. they are too large in the upper regions) and some surge-type glaciers are connected to larger trunk glaciers and can thus not be investigated independently. RGI 7.0 is more precise but also suffers from connected surge-type glaciers. Moreover, several missing or wrongly placed ice divides need to be corrected / added. Accordingly, glacier IDs (and areas) change and the glaciological information required for GloGEM has to be recalculated for all glaciers.

As this is a considerable effort, we have decided to use RGI 6.0 outlines and the available glacier-specific meta-information as a starting point for GloGEM as well as the calculations with the raster-based 3D model. In a second step, we will determine results from the 3D model for the test region Panmah also for RGI 7.0 outlines and compare them to the results using RGI 6.0. Depending on the magnitude and location of the differences, we will decide whether or not GloGEM should be updated to glacier extents from RGI 7.0 (with new ice divides).

In a third step we intend to create minimum and maximum extents of selected surge-type glaciers and determine how their mass balance changes due to the extent change. In combination with a DEM (see Section 3.2) we will more closely examine glacier-specific parameters and determine how these change with the change of the DEM. Here we have to consider that outlines of RGI 7.0 are mapped from Landsat data acquired from 1995 to 2005 (i.e. West: 1995–1998, Central: 1999–2001, East: 2002–2005), i.e. for glaciers with strong changes in extent over time they better match to the SRTM DEM than to the Copernicus global DEM.

3.2 Debris extent and thickness

Supra-glacial debris plays a critical role in shaping glacier response to climate change (e.g. Anderson and Anderson, 2016). When debris is exposed at the surface (i.e. not snow covered), ablation is either amplified or suppressed depending on debris thickness. A thin debris layer enhances melt, whereas thicker debris provides insulation and reduces ablation. Within GloGEM, this is achieved by applying a scaling factor in combination with a glacier-specific debris parameter. Evolution of debris can be considered in the model as described by Compagno et al. (2022), but due to the short time period considered and the much larger changes in debris cover location due to surges, we do not consider a parameterized temporal evolution. Initial debris extent and thickness are derived from the global dataset of Rounce et al. (2021). These inputs are pre-processed for the flow-line model to provide, per elevation band, both the debris fraction and thickness distribution. Both datasets will thus not be modified for the initial modelling, but the impact of turning the debris on and off will be analysed once meaningful modelling results emerge.

3.3 Surge inventories

A number of surge inventories exist for the Karakoram specifically, or inventories that contain this region (e.g. Bhambri et al., 2017; Guillet et al., 2022; Guo et al., 2022; Kääb et al., 2023; Ke et al., 2022; Vale et al., 2021). Moreover, several studies exist that are not explicitly surge inventories, but describe glacier surges and thus contribute to related inventories and our understanding of surges (e.g. Copland et al., 2011; Gardelle et al., 2013; Hugonnet et al., 2021; Paul et al., 2022; Rankl and Braun, 2016). Most of this surge information has been compiled into RGI 7.0 with its four-step classification (0, 1, 2, 3). Still, it is important to keep in mind that glacier surging is not strictly defined, and the detection of surging depends on the method used, e.g. based on elevation changes, velocity changes, advance rates, geomorphological indicators or satellite radar backscatter changes (Paul et al., 2022; Kääb et al., 2023).

While the different surge inventories mentioned above are not strictly comparable and the four class assignment in RGI 7.0 might not reflect all details, the RGI 7.0 should contain the widest collection of surge-type glaciers in the Karakoram as it will incorporate the variety of spatially overlapping studies using different methods and thus ensures a broad detection of glaciers that is largely independent of specific definitions and indicators.

The caveat with the RGI 7.0 assignment is its restriction to glacier complexes, i.e. if surging tributaries are connected to a larger trunk glacier at the date of outline creation, the entire glacier complex becomes surge-type and it is unclear which tributaries are indeed surging (see Fig. 3.1). To solve this problem, we will a) use a modified glacier inventory where the surge-type tributaries are separated from the trunk glaciers (see Section 3.1) and b) use the Bhambri et al. (2017) assignment of surge-type glaciers that has coordinates for the individual glaciers and is available as a shape file with point topology. Indeed, also this dataset is not perfect, i.e. it includes glaciers that are not surging and misses some that are surging, but for the purpose of identifying the individual surge-type glaciers it is our first choice.

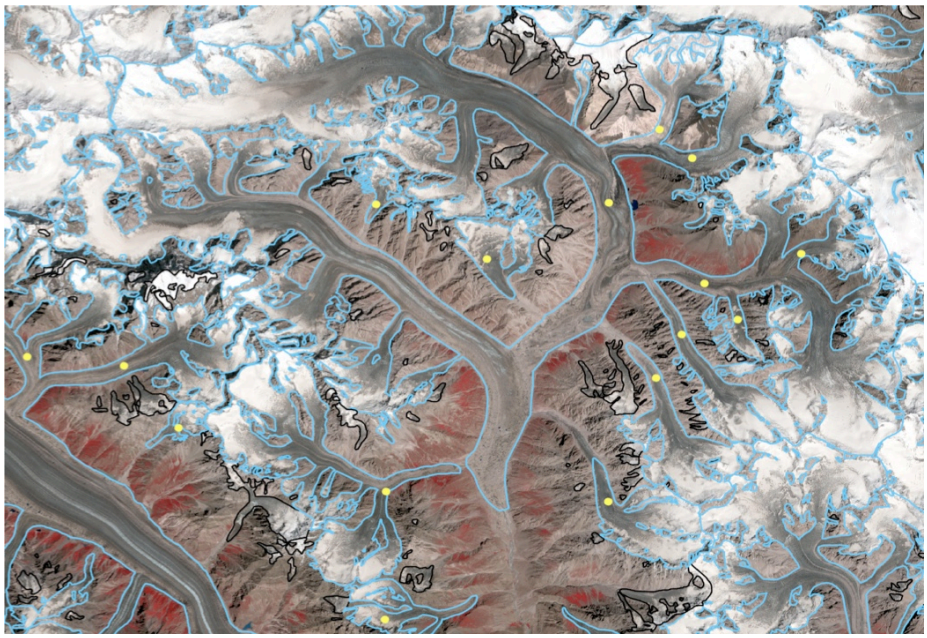

Fig. 3.1: Glaciers in the Panmah test region classified in RGI 7.0 as not surging (black) and with observed surges (blue). The yellow circles show the individual tributaries identified by Bhambri et al. (207) as surge-type.

3.4 Geodetic mass balance

To calibrate the mass balance model, glacier-specific geodetic mass balance datasets are required. This calibration involves adjusting model parameters and bias values so that simulated mass balances are consistent with observed geodetic estimates. Most regional- to global-scale glacier models (Zekollari et al., 2024) currently rely on the geodetic mass balances of Hugonnet et al. (2021), which cover 2000-2019 (also provided in 5-year intervals). These datasets are derived from spatially distributed elevation change maps, primarily based on multi-temporal ASTER imagery (Brun et al., 2017). Glacier-wide volume changes were calculated from aggregated multi-annual elevation differences and converted to mass changes per glacier using a fixed density of 850 kg m^3 (Huss, 2013). For comparative purposes, we will additionally use the geodetic mass balance estimates of Brun et al. (2019). Older geodetic mass balances, reconstructed from DEMs dating back to the 1970s-1980s (Bolch et al., 2017; Zhou et al., 2017), will primarily support long-term time series analyses.

As a starting point, the Hugonnet et al. (2021) dataset will be used as currently implemented in both MBMs, i.e. further modification is not required. However, we will recalculate glacier specific mean values also for the modified RGI 7.0 extents, probably using a revised and extended version of the geodetic mass balance dataset (Béraud et al. 2024).

3.5 Digital elevation models

Two Digital Elevation Models (DEMs) are available for a more detailed analysis:

(1) The SRTM DEM v4.1 (Jarvis et al., 2008) (Fig. 3.2, left) and
(2) the Copernicus Global DEM GLO30 (Fig. 3.2, right).

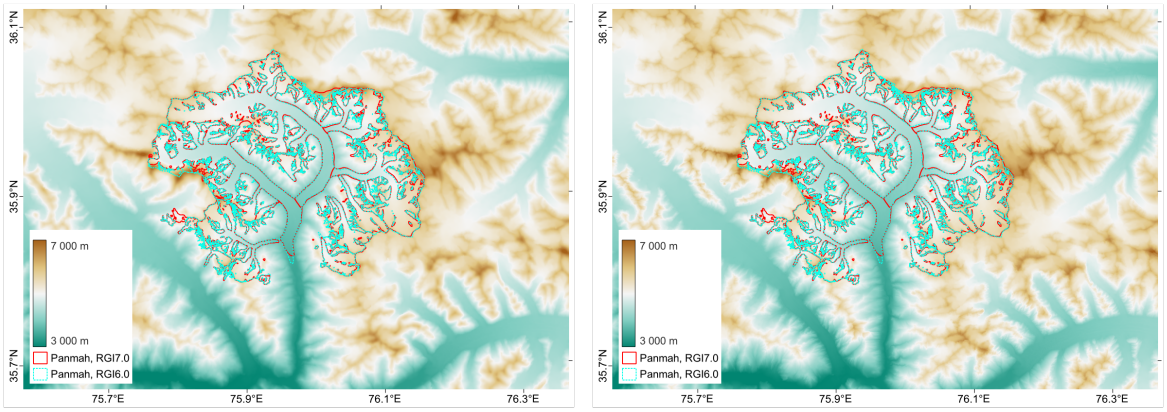

Figure 3.2: Orthometric heights from DEMs overlaid with glacier outlines from RGI6.0 and RGI7.0 for the test region Panmah. Left: SRTM DEM v4.1; Right: Copernicus GLO30 DEM.

The SRTM DEM v4.1 with a pixel spacing of about 90 m x 90 m is based on satellite acquisitions from February 2000 and has been used together with RI 6.0 to determine all glacier-specific information (e.g. elevation bands) for the mass balance models. The Copernicus GLO30 DEM with a pixel spacing of about 30 m x 30 m is based on satellite data acquired around the year 2015 and should be much more precise than the void-filled SRTM DEM. However, in terms of temporal consistency with glacier extents from RGI 6.0 and 7.0, the SRTM DEM v4.1 is closer.

In addition to the differences in the horizontal grids and resolutions, there are major differences in the vertical information between the two DEMs. To assess them, heights from both DEMs are prepared as orthometric heights in the UTM43/WGS84 map projection (EPSG: 32643). The difference between the Copernicus GLO30 DEM and the SRTM DEM v4.1 is calculated by subtracting one DEM from the other. For the central and eastern parts of the Karakoram, the resulting height difference reaches maximum values of more than 1000 m, which is likely also related to artefacts in the SRTM DEM. A subset of differences in the orthometric heights in these two DEMs over the test region Panmah is shown in Fig. 3.3. The impact of differences in the DEMs and the glacier outlines from the different RGI versions on the usage in and results of the glacier model will be investigated.

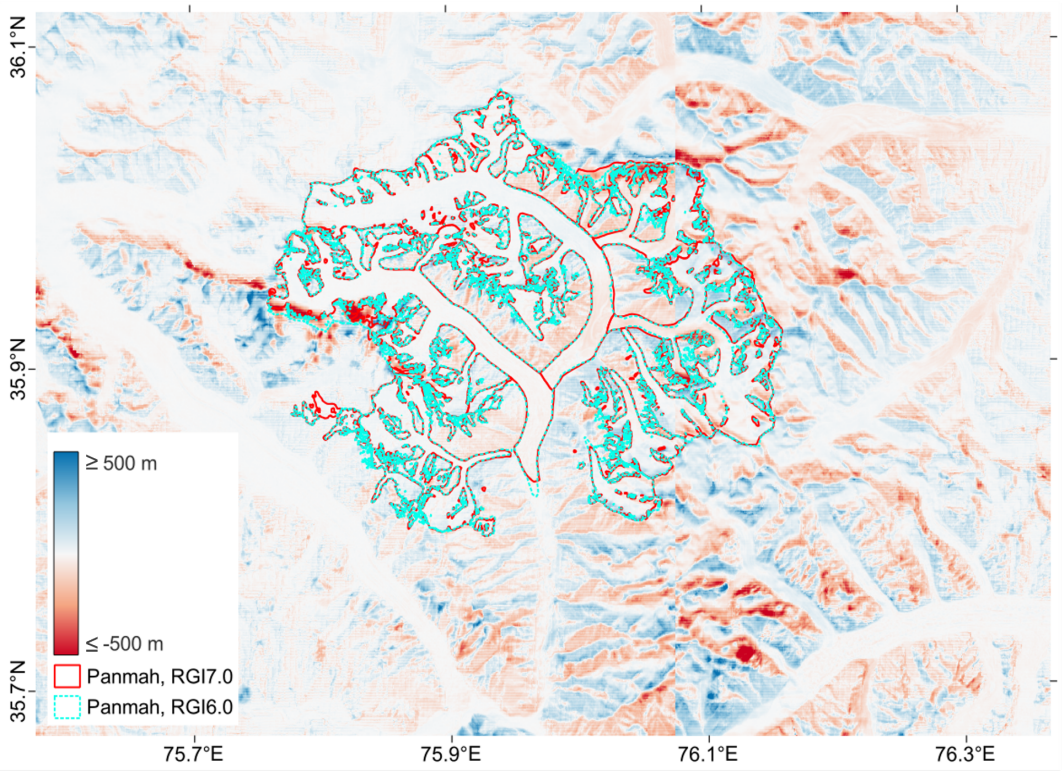

Figure 3.3: Orthometric height difference map from the Copernicus GLO30 DEM minus the SRTM DEM v4.1 over the test region covering the Panmah glacier. Maximum height differences over the central and east Karakoram region are more than 1000 m.

3.6 Snow cover from Landsat and Sentinel-2

The snow cover products derived in Glaciers_cci from Landsat at 30 m and Sentinel-2 imagery at 10 m resolution are available for the periods 1990–2023 and 2016–2023, respectively. The datasets provide a pixel-based glacier surface classification during the melt season, separating snow, snow free glacier ice, snow free debris cover, clouds over glaciers, and non-glacier areas (Fig. 3.4). For each classified pixel, the associated uncertainty information is provided as a flag, indicating low, medium or high uncertainty. There are some temporal gaps caused by persistent cloud cover at the satellite acquisition times during the melt season, in particular for the years 2001–2007 and 2012–2015. Moreover, Landsat-7 data have only been used for the period without the striping (1999-2002) and also for largely cloud-free scenes many glaciers have data gaps due to local cloud cover. Accordingly, a lower number of days and glaciers will have calibration data for the period before Sentinel-2.

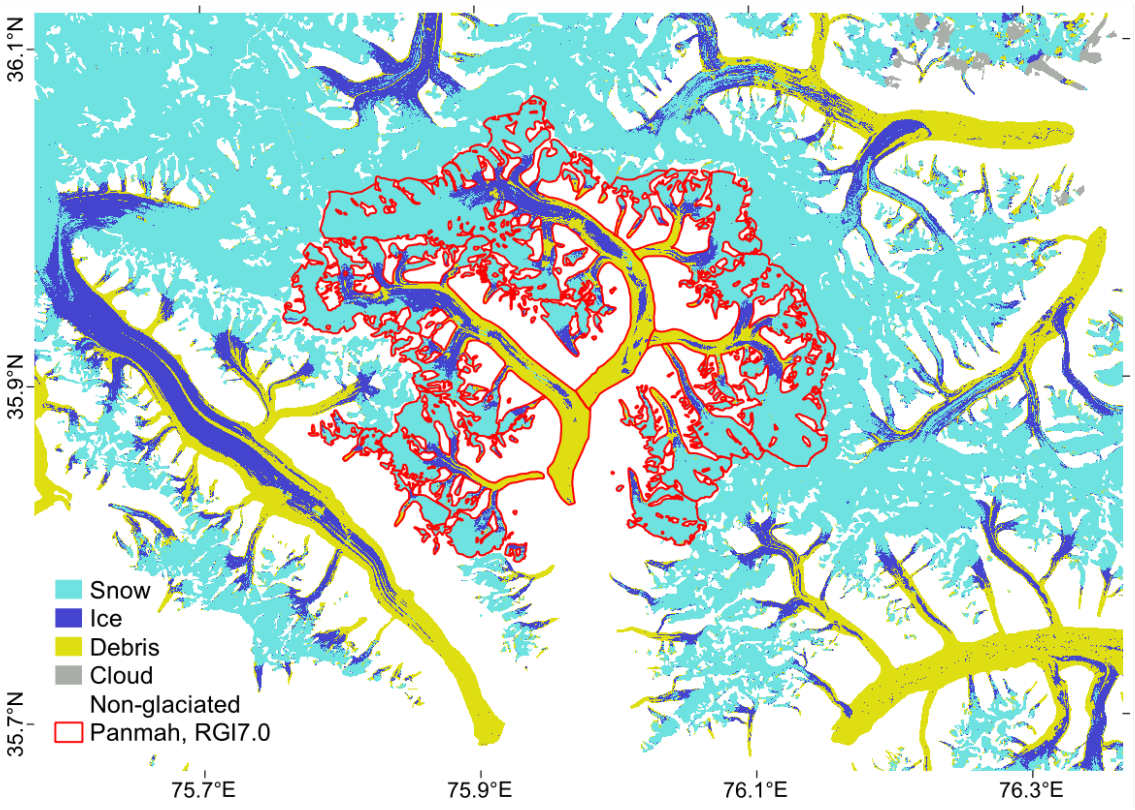

Figure 3.4: Glacier surface classification from Sentinel-2 imagery of 4 September 2022. Red outlined is the selected test region, Panmah glacier.

As the GloGEM model requires snow line elevations per glacier (i.e. vector data), the raster-based snow cover maps have to be combined with a DEM (SRTM) and glacier outlines (RGI 6.0). A common method to derive snow line elevations is to calculate the percentage of snow cover per elevation bin of the respective glacier and then determine the mean elevation of the lowest three intervals with at least 50% snow cover (e.g. Rastner et al. 2019). Uncertainties in the resulting elevation are due to a wrong classification, patchy snow cover, a lower snow cover limit that stretches over a large elevation range and the vertical reference system of the DEM (e.g. ellipsoid or geoid). Snow line elevations will thus be derived from the same DEM as used for the MBM. Apart from using this information for the MBM, the derived values will also be analysed statistically (e.g. spatio-temporal variability and trends).

3.7 Reanalysis data

3.7.1 Available datasets

The Karakoram area and surroundings are notoriously difficult to represent with climate reanalysis due to the lack of in-situ measurements to constrain the physical model. Also satellite data products are only assimilated to a limited degree in this area due to lower data quality/reliability caused by the complex topography. The different products solve this challenge in different ways and differ in how they represent relevant physical processes (e.g., related to snow cover). Currently, none of the available products can be considered superior or correct, and different considerations must be taken depending on the application.

Within this project, we distinguish between two applications of reanalysis data: (1) as input forcing data for the glacier mass balance model (MBM), and (2) to analyse regional changes and patterns in a climate context, considering different factors/ECVs included in this project. For purpose (1) we need to consider that errors in the dataset (e.g. precipitation amounts) will be compensated to some extent by the tuning of the MBM (e.g. with a precipitation correction factor). Differences among the different datasets are thus not very important. For (2) the analysis will focus on trends of climate data (temperature, precipitation, cloud cover) rather than absolute values, as trends are in general more robust. However, previous studies have shown that also here differences exist and we will analyse how these manifest in the study region.

The following overview presents a short description of the relevant reanalysis datasets:
(a) ERA5 – the most recent version of ECMWF's reanalysis, which is considered a very good global reanalysis and usually the first choice as input data for physical models.
(b) ERA5-Land – a dynamically downscaled version of ERA5 at higher resolution (ca. 9 km). The ERA5 products are generally found to overestimate precipitation in many parts of HMA (Fig. 3.5; Orsolini et al., 2019), but to underestimate high-altitude precipitation (Fig. 3.7; Liu and Margulis 2019). They do not assimilate station data over the Tibetan Plateau and also don't assimilate the snow satellite product IMS in the HMA region. A detailed description of ERA5-Land is provided by Muñoz-Sabater et al. (2021).
(c) ERA Interim – an older ECMWF reanalysis, which has been found to provide more realistic precipitation/snow values over the Tibetan Plateau area than the ERA products (Fig. 3.5), possibly because it did assimilate IMS snow cover data in HMA (Orsolini et al., 2019). Previous research, however, shows that this data does not represent decadal precipitation trends that could explain lake level changes observed in HMA (Treichler et al. 2019, Fig. 3.6). The dataset is no longer publicly available as it has been replaced by ERA5.
(d) JRA-55 - The Japanese reanalysis has been found to provide more realistic snow estimates in the region (Orsolini et al. 2019, Fig. 3.5). It differs from ERA5 through its interpolation-based snow depth analysis and assimilation of station data and satellite information (SSM/I, SSMIS) over the Tibetan Plateau.
(e) HAR2 - The High Asia Refined Analysis version 2 (HAR v2) seeks to combine the qualities of ERA5 with better snow representation of JRA-55. It is generated by dynamical downscaling using WRF (see below) of ERA5 data, but using snow data from JRA-55. It has a spatial resolution of 10 km, comparable to ERA5-Land.
(f) MERRA-2 – The NASA Modern-Era Retrospective analysis for Research and Applications, Version 2, performs an online correction for precipitation using data from the NOAA Climate Prediction Center. Compared to other reanalysis products in the HMA region it was found to represent more realistic snow volumes (Orsolini et al. 2019, Fig 3.5), high-altitude precipitation in parts of HMA (Liu and Margulis 2019, Fig. 3.7), and precipitation patterns that could explain lake level changes (Treichler et al., 2019, Fig. 3.6).

The ERA5 and MERRA-2 reanalysis datasets can be accessed programmatically via APIs, and they exist at different temporal resolutions/timesteps (e.g., monthly, daily, sub-daily at regular intervals). Also HAR2 and JRA-55 are available from data servers, only the ERA-Interim data is no longer publicly distributed (as it has been replaced by ERA5). For the purpose of this research, we primarily need parameters calculated at the surface level, and for the analysis of circulation patterns and clouds possibly also at certain atmospheric pressure levels. All data products provide the necessary variables, but internal variable names differ.

Due to its high spatial resolution and the already existing assimilation scheme in the MBM, we have decided to use ERA5-Land as a starting point for the modelling. For the analysis of trends we will consider ERA-5 Land and MERRA-2, possibly also HAR-2.

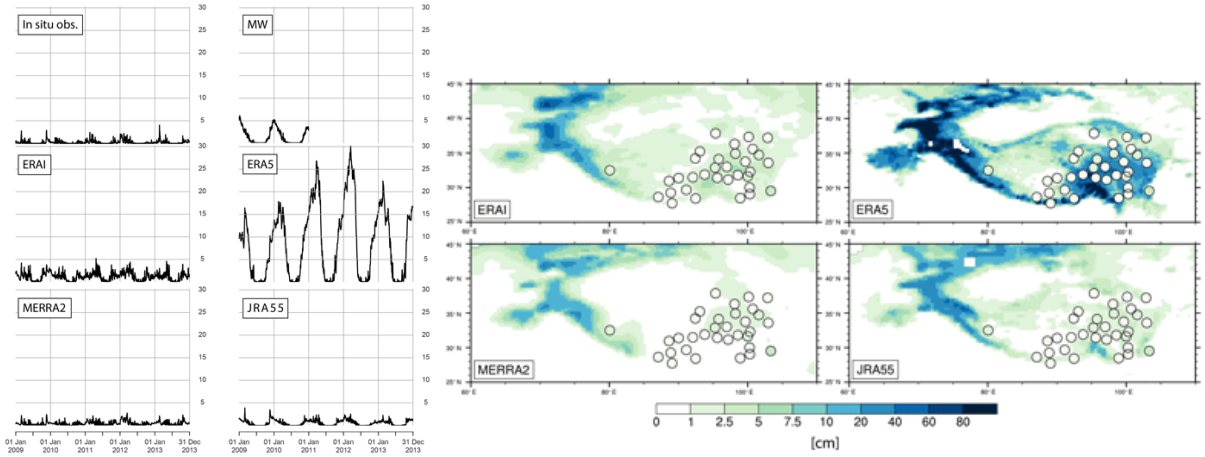

Figure 3.5: Snow depth (in cm) on the Tibetan Plateau. Left: average of 33 stations for the period 2009-2013, and for satellite microwave (MW) and four reanalysis products in the same locations. Right: 5-year mean of January snow depths for the four reanalysis products; station locations indicated with white circles. Source: Orsolini et al. (2019).

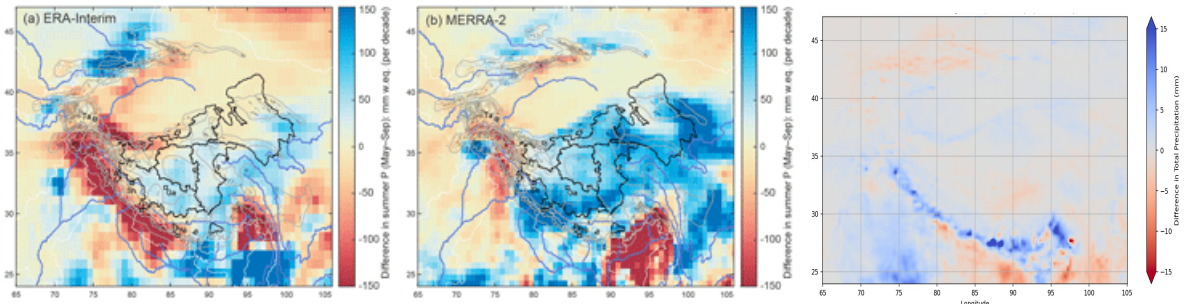

*Figure 3.6: Difference between decadal averages of summer precipitation (May-September) in 2000–2009 and 1990–1999 for **(a)** ERA-Interim, **(b)** MERRA-2, and c) ERA5-Land. Blue indicates increased precipitation for the later decade. Source: a) and b) from Treichler et al. (2019).*

3.7.2 Use as model input data

We will use the global glacier model GloGEM (Huss and Hock, 2015) to model glacier mass balance over the entire Karakoram and a more detailed and raster based 3D model for the test region Panmah. Both models can handle ERA5-Land data as an input for the climatic forcing. The MBM GloGEM requires temperature and precipitation time series as input data for each glacier, provided by the reanalysis datasets. The coarse spatial resolution and simplified representation of mountain topography requires downscaling of reanalysis products to the correct location and elevation. This can be achieved by application of elevation lapse rates and bias correction to match local conditions (Fig. 3.7), through statistical models (e.g., TopoPyScale, Fiddes et al., 2022), or nested physical models operating at higher spatial resolution (e.g., the Weather Research and Forecasting model WRF, as done for the reanalysis product HAR2).

The GloGEM workflow includes a downscaling and bias-correction (delta-method) routine to retrieve the necessary forcing data for the mass balance modelling, which are temperature and precipitation at the surface at daily temporal resolution. As all reanalysis datasets require bias correction to reproduce correct magnitudes of the selected parameter, the choice of reanalysis is of secondary importance. In other words, corrections and model calibration to make reanalysis products suitable for the modelling exercise are part of the workflow.

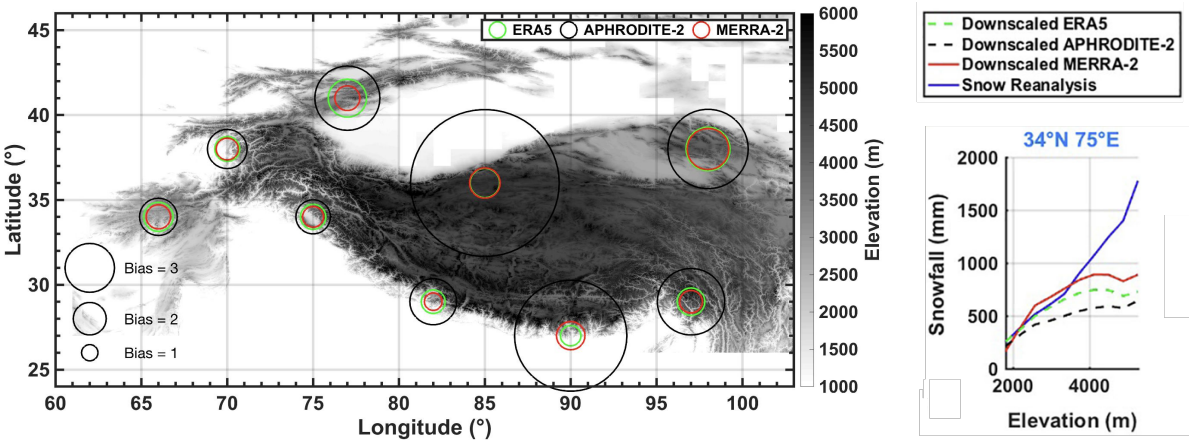

Fig. 3.7: Bias for annual total snowfall (water years 2001–2015) for the reanalysis products ERA5 and MERRA-2, and the gridded precipitation product APHRODITE. Left: spatial distribution of the bias (a larger circle means greater bias, i.e. snowfall is more heavily underestimated in that dataset), right: elevational distribution of snowfall for the upper Indus basin in the Karakoram region (corresponding circle at 34N, 75E), where a snow reanalysis based on satellite data (assumed more correct) indicates considerable elevation dependence of the bias. The pattern is similar for the other western locations/circles, but less severe for the eastern locations, in particular for the MERRA-2 data (not shown). Source: modified from Liu and Margulis (2019).

ERA5 data has been shown to be a suitable reanalysis dataset for this purpose and is therefore the dataset of choice for the modelling work in the project. Should progress later in the project point to issues due to this choice it would be possible to replace ERA5-L with a different reanalysis product in the glacier modelling workflow, and re-run the entire analysis with different input data sources. However, other factors and modelling choices are considered to have greater influence on the results and their uncertainty than the choice of forcing data. Hence, a sensitivity study for different reanalysis datasets will not be performed initially (e.g. Aguayo et al., 2024), but considered if first results of the glacier modelling or the analyses of the datasets show deficiencies of the ERA5-L reanalysis product – for example, in representing important spatio-temporal patterns or changes thereof.

Bulk download and duplication of reanalysis data onto the X-ECV project database is not foreseen due to the integrated workflow and accessibility of the reanalysis data. Rather, relevant (corrected) model input data and parameters will be stored in the project database and documented in a publication after completion of model runs alongside results.

3.7.3 Analysis of climate patterns and changes

Measurements and reanalysis data show that the HMA region suffers from increasing temperatures just as other mountain regions, but the development of other climate variables relevant

for the cryosphere and hydrosphere is less clear. Large parts of HMA are very cold due to the high altitude, and due to surrounding orographic barriers also extremely dry, with minimal annual precipitation – but potentially high evaporation/evapotranspiration and sublimation (of snow/ice). In addition to temperature (at different elevations) and precipitation (in various forms), we will analyse cloud cover and parameters related to water vapour fluxes between the surface and atmosphere (and within the atmosphere), and water retention in the soil or in the form of snow. It is also possible that other parameters that are less directly related to glacier mass balance (or other observed changes) could provide useful hints – about underlying model physics or processes (or missing relevant physics/processes).

Given the differences between, and varying performance of reanalysis data in the area, we need to analyse the Karakoram region in its greater spatial context, rather than just the mountain range on its own. Previous research has shown that representation of precipitation and snow differs greatly between the reanalysis products in the HMA region (snow volumes: Fig. 3.5, Orsolini et al., 2019, summer precipitation changes: Fig. 3.6, Treichler et al., 2019) – but these analyses are of small values with high uncertainty. Therefore, aggregation to monthly, seasonal, annual and multi-annual datasets is required to detect patterns and trends that are relevant for the observed (and modelled) changes in the climate variables. Consequently, the dataset required for such analyses is therefore different to the very localized, dense time-series needed as mass balance model input. Aggregation in space is another way to decrease uncertainties, and we therefore aim at using sub-regions as defined in Section 2.3. However, spatial patterns may not match these sub-regions well, and we lack knowledge to define sub-regions in a better way – these are therefore considered indicative rather than fixed for the purpose of spatial analyses.

In summary, the analysis of reanalysis data has explorative character, which might be hampered by strict limitations to only certain variables, resolutions or regions at this stage of the project. We will therefore use an agile approach with reusable and adaptable scripts that take advantage of the ready availability of the reanalysis products through data providers and APIs, rather than mirroring slices of these datasets in the project database. At temporally aggregated timescales, and for our limited geographical areas, file sizes of datasets are small and they can easily be re-downloaded on demand, whereby changes to the temporal/spatial domain and variable list are unproblematic.

3.8 Other datasets

Time-series of cloud cover data from Cloud_cci will be used as an input for the MBM and statistically analysed (e.g. monthly/seasonal means and trends for different regions). Apart from the required spatio-temporal sub-setting, the further modifications of the dataset (e.g. format transformations, outlier handling, gap filling) are yet unclear and will be reported in a later document. The extension of the lake extent and lake level datasets will be performed independently from the other investigations. We here aim at creating the longest possible time series for selected lakes by combining multiple datasets. These will be contrasted in the later analysis with the observed trends in mass balance and precipitation. For glacier flow velocities we will analyses long-term time-series of a larger glacier sample (e.g. to identify differences among glaciers) as well as dense time-series for individual glaciers in the case of surging. For this purpose, available datasets (e.g. from ITS_LIVE and Glaciers_cci) will be analysed along with datasets created by the project.

4. Summary of dataset selection and pre-processing

Getting the required datasets fit for purpose requires selecting them (when several are available) and modifying them in a way that they can be used for these applications. These are (A) forcing a mass balance model and (B) analysing characteristics and trends of the datasets. For (A) and (B) we need (1) meteorological and (2) glacier-related datasets which both vary in space and have a variable degree of temporal variability ranging from quasi static (e.g. glacier outlines, DEM) to hourly changes (e.g. reanalysis data). Moreover, all datasets come in different flavours (e.g. file formats, projections, spatial resolution) that have to be harmonized to some degree for the analysis. As the two MBM to be used (GloGEM and the 3D model) have already the required interfaces for assimilation of existing datasets implemented, we start with the datasets the two models are using and do thus not need to modify them in this regard. We need, however, to define the regions and periods of interest. For (A) this will be (i) a smaller test region in the central Karakoram (Panmah) and (ii) the Karakoram itself, whereas for (B) we will analyse the Karakoram and (iii) selected mountain regions in HMA (see Section 2.3). As a temporal period we have selected 2000-2019 for (A) and 1950-2025 for (B), i.e. the periods with available geodetic mass balances for calibration and reanalysis data, respectively. For the MBM and topographic analysis we will work in UTM43N projection with WGS84 datum, for the analysis of climate data in the different sub-regions we will use the original (geographic) projection and select the cells covering the study region.

As a starting point for the MBM we will apply both GloGEM and the 3D model only in the Panmah test region with their current set-up. This means we use the SRTM DEM (from 2000) as a reference for elevations and glacier hypsometry, glacier outlines from RGI 6.0, the debris extent and debris thickness data as prepared for RGI 6.0 and ERA5-Land reanalysis data (temperature T and precipitation P). Whereas the GloGEM model will run with mean daily averages of T and P on a per glacier basis, the 3D model will use hourly values of T and P, include a correction for solar radiation and works raster based, i.e. all values are determined for each cell of the underlying DEM, for SRTM at a spatial resolution of 90 m. The differences in the modelled mass balances when using the modified glacier outlines from RGI 7.0 and the Copernicus DEM will be determined in the test region with the 3D model in a second step. As the model is raster based, the DEM can be exchanged easily and different outlines can be used in the post-processing step to determine glacier-specific values by digital intersection. Further modifications of the input data (e.g. considering cloud cover or a distributed map of ice albedo) will be investigated in a third step with the 3D model.

The new element of the MBM is the use of satellite-derived snow cover maps (created in Glaciers_cci) that will be used for calibration (i.e. adjusting the precipitation correction factor in the model). As the both models work with glacier-specific elevations of the snow line, a method to extract this information has to be applied. Further datasets will be derived and analysed in the course of the project.

References

Aguayo, R., Maussion, F., Schuster, L., Schaefer, M., Caro, A., Schmitt, P., Mackay, J., Ultee, L., Leon-Muñoz, J., and Aguayo, M. (2024): Unravelling the sources of uncertainty in glacier runoff projections in the Patagonian Andes (40–56° S), The Cryosphere, 18, 5383–5406; doi.org/10.5194/tc-18-5383-2024

Anderson, L.S. and Anderson, R.S. (2016): Modeling debris-covered glaciers: response to steady debris deposition, The Cryosphere, 10, 1105–1124; doi.org/10.5194/tc-10-1105-2016

Bhambri, R., Hewitt, K., Kawishwar, P. and Pratap, B. (2017). Surge-type and surge-modified glaciers in the Karakoram. Scientific Reports, 7(1); doi.org/10.1038/s41598-017-15473-8.

Béraud, L., F. Brun, A. Dehecq, L. Charrier and R. Hugonnet (2024): An improved dataset of ASTER elevation time series in High Mountain Asia to study surge dynamics. EGU General Assembly 2024, Abstract EGU24-19324.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., and Treichler, D. (2017): A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nature Geosci, 10, 668–673; doi.org/10.1038/ngeo2999

Brun, F., Wagnon, P., Berthier, E., Jomelli, V., Maharjan, S.B., Shrestha, F., and Kraaijenbrink, P.D.A. (2019): Heterogeneous Influence of Glacier Morphology on the Mass Balance Variability in High Mountain Asia, Journal of Geophysical Research: Earth Surface, 124, 1331–1345; doi.org/10.1029/2018JF004838

Bolch, T., Pieczonka, T., Mukherjee, K., and Shea, J. (2017): Brief communication: Glaciers in the Hunza catchment (Karakoram) have been nearly in balance since the 1970s, The Cryosphere, 11, 531–539; doi.org/10.5194/tc-11-531-2017

Compagno, L., Huss, M., Miles, E. S., McCarthy, M. J., Zekollari, H., Dehecq, A., Pellicciotti, F., and Farinotti, D. (2022): Modelling supraglacial debris-cover evolution from the single-glacier to the regional scale: an application to High Mountain Asia, The Cryosphere, 16, 1697–1718; doi.org/10.5194/tc-16-1697-2022

Copland, L., Sylvestre, T., Bishop, M. P., Shroder, J. F., Seong, Y. B., Owen, L. A., Bush, A., and Kamp, U. (2011): Expanded and recently increased glacier surging in the Karakoram, Arct. Antarct. Alp. Res., 43, 503–516

Fiddes, J., Aalstad, K., and Lehning, M.: TopoCLIM (2022): rapid topography-based downscaling of regional climate model output in complex terrain v1.1, Geosci. Model Dev., 15, 1753–1768; doi.org/10.5194/gmd-15-1753-2022

Gardelle, J., Berthier, E., Arnaud, Y. and Kaab, A. (2013): Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999-2011. The Cryosphere, 7 (6), 1885-1886.

Guillet, G., King, O., Lv, M., Ghuffar, S., Benn, D., Quincey, D., and Bolch, T. (2022): A regionally resolved inventory of High Mountain Asia surge-type glaciers, derived from a multi-factor remote sensing approach. The Cryosphere, 16 (2), 603-623.

Guo, L., Li, J., Dehecq, A., Li, Z., Li, X. and Zhu, J. (2022): A new inventory of High Mountain Asia surge type glaciers derived from multiple elevation datasets since the 1970s. Earth System Science Data, 15 (7), 2841-2861.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., and Kääb, A. (2021): Accelerated global glacier mass

loss in the early twenty-first century, Nature, 592, 726–731; doi.org/10.1038/s41586-021-03436-z

Huss, M. (2013): Density assumptions for converting geodetic glacier volume change to mass change. Cryosphere, 7, 877–887; doi.org/10.5194/tc-7-877-2013

Huss, M. and Hock, R. (2015): A new model for global glacier change and sea-level rise, Front. Earth Sci., 3; doi.org/10.3389/feart.2015.00054

Kääb, A., Bazilova, V., Leclercq, P.W., Mannerfelt, E.S. and Strozzi, T. (2023): Global clustering of recent glacier surges from radar backscatter data, 2017–2022. Journal of Glaciology, 69 (277), 1515-1523.

Ke, L.H., Zhang, J.S., Fan, C.Y., Zhou, J.J. and Song, C.Q. (2022): Large-scale monitoring of glacier surges by integrating high-temporal- and -spatial-resolution satellite observations: a case study in the Karakoram. Remote Sensing 14(18), Artnr. 4668. doi.org/10.3390/rs14184668

Liu, Y. and Margulis, S.A. (2019): Deriving bias and uncertainty in MERRA-2 snowfall precipitation over High Mountain Asia. Frontiers in Earth Science, 7, 280. doi.org/10.3389/feart.2019.00280

Muñoz-Sabater, J., et al. (2021): ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data, 13, 4349–4383; doi.org/10.5194/essd-13-4349-2021

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W., Senan, R. and Arduini, G. (2019): Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations. The Cryosphere, 13, 2221–2239; doi.org/10.5194/tc-13-2221-2019

Paul, F., Piermattei, L., Treichler, D., Gilbert, L., Girod, L., Kääb, A., Libert, L., Nagler, T. Strozzi, T. and Wuite, J. (2022): Three different glacier surges at a spot: What satellites observe and what not. The Cryosphere, 16, 2505-2526.

Rankl, M. and Braun, M. (2016): Glacier elevation and mass changes over the central Karakoram region estimated from TanDEM-X and SRTM/X-SAR digital elevation models. Annals of Glaciology 57 (71), 273-281.

Rounce, D. R., Hock, R., McNabb, R. W., Millan, R., Sommer, C., Braun, M. H., Malz, P., Maussion, F., Mouginot, J., Seehaus, T. C., and Shean, D. E. (2021): Distributed Global Debris Thickness Estimates Reveal Debris Significantly Impacts Glacier Mass Balance, Geophysical Research Letters, 48, e2020GL091311; doi.org/10.1029/2020GL091311

Treichler, D., Kääb, A., Salzmann, N., and Xu, C.-Y. (2019): Recent glacier and lake changes in High Mountain Asia and their relation to precipitation changes. Cryosphere, 13, 2977-3005.

Vale, A.B., Arnold, N.S., Rees, W.G. and Lea, J.M. (2021): Remote detection of surge-related glacier terminus change across High Mountain Asia. Remote Sensing, 13 (7), 1309; doi.org/10.3390/rs13071309.

Zekollari, H., Huss, M., Schuster, L., Maussion, F., Rounce, D. R., Aguayo, R., Champollion, N., Compagno, L., Hugonnet, R., Marzeion, B., Mojtabavi, S., and Farinotti, D. (2024): Twenty-first century global glacier evolution under CMIP6 scenarios and the role of glacier-specific observations, The Cryosphere, 18, 5045–5066; doi.org/10.5194/tc-18-5045-2024

Zhou, Y., Li, Z., and Li, J. (2017): Slight glacier mass loss in the Karakoram region during the 1970s to 2000 revealed by KH-9 images and SRTM DEM, Journal of Glaciology, 63, 331–342; doi.org/10.1017/jog.2016.142, 2017.

Acronyms

ASTER	Advanced Spaceborne Thermal Emission and Reflection radiometer
DEM	Digital Elevation Model
ECV	Essential Climate Variable
ERA	European Reanalysis
GloGEM	Global Glacier Evolution Model
HMA	High Mountain Asia
MBM	Mass Balance Model
MODIS	Moderate Resolution Imaging Spectrometer
RGI	Randolph Glacier Inventory
SR	Sub-region
SRTM	Shuttle Radar Topography Mission
UTM	Universal Transverse Mercator
WRF	Weather Research and Forecasting