LEM Data-Driven Calibration

Calibration of a Landscape Evolution Model in an Alpine Region: on the value of in-situ observations and transferability

Rayane Meghezi

Course Number: GEO4-1520
ECTs: 30
Student Number: 9852514
Programme: Earth Structure and Dynamics
Track: Physics of the Solid Earth and Planet
1st Supervisor: Dr. Rens van Beek
2nd Supervisor: Dr. Cedric Thieulot

Universiteit Utrecht

Utrecht University
The Netherlands
1st October 2024

Abstract

Landscape evolution models (LEMs) are physics-based simulations that evaluate the influence of geomorphic processes on regional landscape development over longer time periods. These processes are represented in a simplified manner, and their parameterization cannot always be readily linked to real-world quantities. This makes their calibration and application challenging, especially as the calibrated parameter values are imposed to the entire model domain and extrapolated in time. This study introduces a data-driven calibration using historical observations from different stations in the Swiss Alps, aiming to automate and streamline the calibration process.

Using a Sobol sequence method allowed comprehensive exploration of the parameter space, identifying behavioral parameter sets that were then tested for their transferability across different spatial and temporal scales. Calibrating parameters on smaller catchments significantly reduced computational demands, streamlining the calibration process and making it more practical. Results from temporal transferability tests indicated consistent performance in discharge simulations, although sediment transport remained highly sensitive and variable over extended periods. Spatial transferability showed promising potential, suggesting that parameter sets calibrated on smaller catchments could be effectively applied to larger areas.

However, the study also highlighted significant challenges, particularly in simulating hydrological extremes and accurately capturing sediment transport processes. The use of a yearly timestep limited the model's ability to reflect seasonal dynamics, and the current model setup lacked sufficient data and complexity to properly simulate key processes such as glacier dynamics and temperature effects. These limitations underscore the need for further refinement in both data inputs and model structure to enhance the accuracy and robustness of LEMs.

Contents

List of Figures

List of Tables

1 Introduction

Landscape evolution models (LEMs) are essential for understanding how geomorphological processes, such as erosion, sediment transport, and deposition, shape the Earth's surface over long timescales. These models simulate the complex interactions between landscape dynamics and environmental factors, making them valuable tools for reconstructing past landscapes and forecasting future changes. However, LEMs often rely on simplified representations of processes and proportionality parameters that may not directly correspond to measurable real-world quantities (Reinhardt et al., 2010). This simplification is necessary due to practical and computational limitations, but it also introduces challenges in accurately calibrating and validating these models (Willgoose et al., 1991).

Calibration and validation of LEMs are crucial steps to ensure that the models realistically represent natural landscapes. LEM calibration methods involve adjusting parameters iteratively by comparing model outputs with observed landscape features (van der Beek and Bishop, 2003). However, due to the complexity of the Earth's system and the high-dimensional nature of parameter spaces, calibration is challenging and requires robust methods (Hancock et al., 2011). Key calibration approaches include:

- **Manual Calibration:** Parameters are manually adjusted by experts based on their domain knowledge and comparisons between model outputs and observed data. This iterative process involves fine-tuning until the model closely aligns with the real-world observations.
- **Automated Calibration:** This method uses optimization algorithms to adjust parameters, minimizing the difference between model outputs and observed data. It reduces subjectivity and labor by systematically exploring parameter space, often employing techniques like gradient descent and Bayesian optimization.
- **Parameter Space Exploration:** Involves navigating the multidimensional space of all possible parameter combinations. Techniques like Latin Hypercube Sampling (LHS) and Sobol sequences enhance coverage, while hybrid methods combine broad exploration with precise optimization to avoid pitfalls like local minima.
- **Data Mining:** Utilizes statistical and computational methods to extract patterns and correlations from large datasets. In calibration, it helps identify optimal parameter sets that accurately reflect observed real-world phenomena by analyzing and validating extensive data.

For model validation, the Kling-Gupta Efficiency (KGE) is used to compare simulated data with observations, providing an estimation of model performance by considering correlation, variability, and bias between simulated and observed data. KGE values are calculated for discharge and suspended sediment concentration (SSC) at a specified outlet node corresponding to the real-world location of the hydrological stations. This approach allows for a comprehensive assessment of how well the model replicates key hydrological and sediment transport processes.

Despite these approaches, the calibration and validation processes are complicated by the scale-dependency of parameters, making them difficult to transfer between different spatial or temporal contexts. Parameters effective in one catchment or time resolution may not perform well when applied elsewhere, raising questions about the robustness and generalizability of LEMs.

This study addresses these challenges by exploring the viability and utility of calibrating LEMs using data from specific stations. This restricts the calibration period to the timeframe covered

by the dataset, in this case, 49 years. A key aspect of this approach is evaluating how well calibrated parameters perform when simulations are extended beyond the calibration period, and whether they can accurately predict landscape changes over longer timescales. This is tested by running simulations extended in time of the best-performing models from the same catchment and comparing variations in discharge and sediment transport.

Additionally, the study investigates whether calibrating LEMs on smaller catchments, to save computational time, allows the derived parameters to be successfully applied to larger areas. This raises the question of whether LEM parameters can be transferred in space, offering a practical solution to the computational challenges of LEM calibration.

However, the calibration framework used in this study introduces certain limitations. The model simulates only sediment transport linked to river channels and does not include processes such as glacial dynamics, which are significant in alpine environments. The model also uses annual averages and yearly timesteps, thus missing seasonal variations and restricting the analysis to inter-annual changes. Furthermore, with only one station per catchment, the calibration lacks detailed insights into within-catchment variability, potentially reducing the robustness of the calibrated parameters.

The chosen study area is the region southeast of Interlaken, Switzerland, characterized by varied topography, hydrological complexity, and the availability of high-resolution data. This area includes three primary catchments: Aare, Lütschine, and Lonza. They feature diverse characteristics such as differences in altitude, glacier coverage, and catchment size. The detailed hydrological data available, including discharge and suspended sediment concentration measurements from multiple stations, make it an ideal location to test our calibration approach. By utilizing high-resolution digital elevation models and climatic data, we can simulate a range of geomorphic and hydrological conditions to assess the generalizability of the calibrated LEM parameters.

In summary, this study aims to contribute to the ongoing discussion about the transferability of LEM parameters across different spatial and temporal contexts. By examining whether calibration on limited datasets can yield reliable predictions when extended and whether parameters calibrated on smaller catchments can be effectively applied to larger regions, we hope to advance the practical use of LEMs in geomorphology and landscape management.

2 Methods

2.1 Model Description

2.1.1 Modelling Framework

LEM utilizes numerical methods to simulate the topographic changes. While non-exhaustive, the following process can be simulated:

1. The erosion and weathering effect caused by water, wind and ice on rock and soil.
2. The transport and deposition of sediment by river, glacier, shallow-water flow over the landscape.
3. The impact of tectonic movements on the landscape, including uplift and subsidence.

The predicting framework can be used to do landscape reconstruction or to forecast forward in time, both of which can be used on to gain an understanding of the processes that lead to the development of patterns in the landscape, conduct hazard assessment or evaluate the impact of climate change (Simpson, 2017; Temme et al., 2017).

The modelling framework in use for this project is Landlab 2.7.0, an open-source Python library designed for the creation and simulation of landscape (Hobley et al., 2017; Barnhart et al., 2020; Hutton et al., 2020). It provides tools and components for carrying out various types of numerical simulations in the geosciences, in particular on earth surface processes such as hillslope and fluvial geomorphology, hydrology, weathering and vegetation.

Landlab uses a grid data structure and support rectangular, hexagonal and Voronoi grid. Fields in Landlab are arrays of data associated with a specific grid element (nodes, cells or links). They can store various physical quantities such as elevation, temperature, soil moisture, etc. Each field is identified by a name and is associated with one of the grid elements. While there is a standard list of names for each field, it is possible to create custom one, as they work like any other variable in python.

Landlab's design feature modular components that describe specific physical processes or algorithms. These components interact with Fields and can be combined to build models with varying degree of complexity. This is a great feature of Landlab as it allow user to simulate environmental processes by simply integrating pre-built component but also allow the user to design their own. In this project, we will use a set of pre-built components to calculate the flow of water with a D8 algorithm, simulate the erosion using the stream power law and compute the discharge and suspended sediment at different point of the grid.

The selection of Landlab for this research is based on four key considerations. First, its component-based architecture facilitates customization and allows for varying degrees of complexity to suit specific modeling requirements. Second, the high-resolution Digital Elevation Models (DEMs) available for our study area can be seamlessly integrated into Landlab. Third, Landlab offers extensive documentation, tutorials, and examples, which greatly ease the learning curve and support for model development. Lastly, Landlab is built in Python, enabling easy integration with other scientific libraries such as NumPy (Harris et al., 2020) and Pandas (pandas development team, 2020), and it aligns with the author's programming expertise.

2.1.2 Model Structure

The overall model structure is presented in Figure 1, with inputs shown in blue and outputs in red. The landscape evolution model (LEM) in this study, implemented using the Landlab library, simulates discharge and suspended sediment concentration (SSC) based on key environmental inputs: a digital elevation model (DEM), temperature, precipitation, and soil type. The DEM defines the topography of the grid, enabling the application of a lapse rate correction to adjust precipitation data according to elevation.

Runoff is generated from precipitation using a spatially varying runoff coefficient, which is influenced by temperature and slope, the latter derived from the terrain. Water flow direction is determined from elevation differences between neighboring grid cells, guiding runoff and allowing the estimation of discharge at each node by multiplying runoff by the cell area.

Discharge, in combination with soil type, is then used to calculate the erosion rate. Erosion deposition is subsequently computed based on both erosion rate and discharge. The final SSC is determined by integrating the effects of erosion and deposition processes. These interactions and their mechanics will be explored in more details in the following sections.

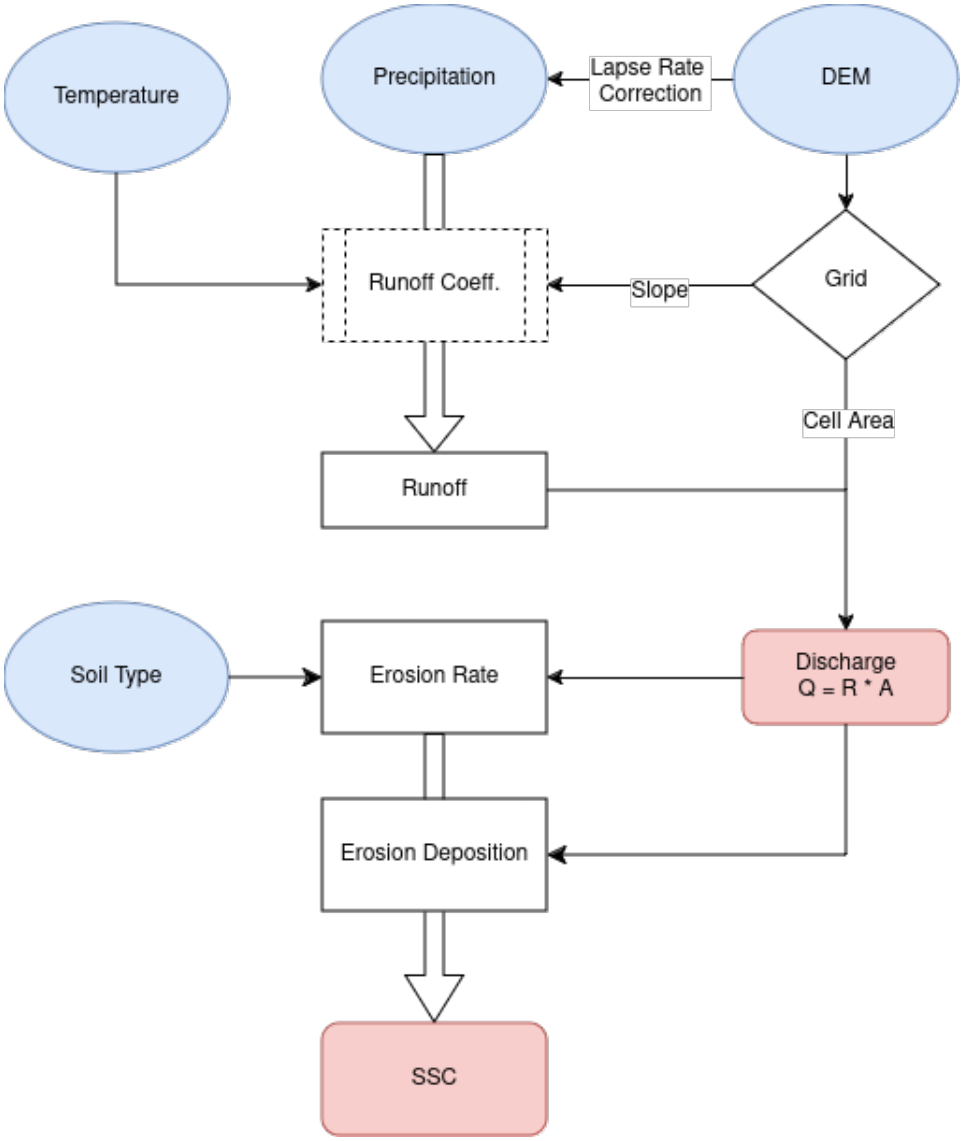


Figure 1: Flowchart of the model structure showing blue input data (Temperature, Precipitation, DEM, Soil Type) and red output data (Discharge, SSC).

2.1.3 Grid Setup and Resolution

A grid system organizes spatial data into a structured format where nodes act as the data collection points, links describe the connections between these points, and cells define the areas around nodes. Therefore, when a specific calculation depends on the area, it will generally use the cell, and when it depends on a point value, like the elevation, it will use the node.

In this study, we employed a regular rectangular grid within Landlab to represent the spatial domain of our LEM. Selecting the appropriate grid type and resolution is crucial, as it directly influences the accuracy of model outputs, the computational efficiency of simulations, and the complexity of implementation. We chose a rectangular grid because it is straightforward to implement, simplifying the development process. Moreover, flow accumulation is an essential and computationally demanding part of the hydrological modelling. These calculations run more efficiently when the cells have consistent arrangement.

Selecting an appropriate grid resolution was critical to ensure that the model accurately represented the landscape while remaining computationally feasible. We conducted preliminary tests using grid resolutions of 10m, 20m, 50m, and 100m to evaluate the model's performance across different scales. These tests assessed runtime, Kling-Gupta Efficiency (KGE), hydrological loss, and station placement accuracy in terms of elevation and catchment area. Hydrological loss was calculated as the difference between the total runoff generated within a catchment and the discharge observed at the corresponding hydrological station.

2.1.4 Hydrological Modeling

Hydrological modeling is a central component of the Landscape Evolution Model (LEM), involving the conversion of precipitation into runoff, determination of flow directions across the terrain, and calculation of discharge at each grid cell. This section details the processes and equations used to simulate these hydrological components.

Runoff Generation
Runoff generation transform precipitation into surface runoff. In the model, annual precipitation data are assigned to each grid cell after applying an elevation-based correction to account for the change in precipitation due to the altitude. The corrected precipitation is calculated as:

$$P_{corrected} = P_{mean} * (Z_{node} - Z_{mean}) * L_p^z \tag{1}$$

Where:

- $P_{corrected}$ is the corrected precipitation $[m/yr]$
- P_{mean} is the upscaled value computed using bilinear interpolation [m/yr][m/yr]. This interpolation is performed to match the resolution of the model grid, utilizing the SciPy library (version 1.13.1) (Jones et al., 2001)
- Z_{node} is the elevation of each grid node in the DEM $[m]$
- Z_{mean}, represents the average elevation at each data point within the precipitation dataset [m][m]. This is computed by applying a uniform filter, where the filter size corresponds to the ratio between the precipitation cell size and the grid size of the Landscape Evolution Model (LEM), effectively representing how many LEM cells fit within a single precipitation cell. The calculation is performed using the SciPy library (version 1.13.1) (Jones et al., 2001)

- L_p^z is the precipitation lapse rate $[m^{-1}]$, which varies every 1000m up to 3000m, with a fixed rate above 3000m. These lapse rates are calibration parameters. See Table 3 for the value in use

The amount of precipitation that becomes runoff is determined by a spatially variable runoff coefficient, which accounts for local factors such as slope and temperature that influence infiltration and evaporation. The runoff at each grid cell is calculated using the following formula:

$$R = P * R_{ratio} \tag{2}$$

The runoff coefficient, constraint between 0 and 1, is calculated with the following formula:

$$R_{\text{ratio}} = \min\left(1, \max\left(0, R_{\text{ref}} \times S_{\text{ratio}} \times T_{\text{ratio}}\right)\right) \tag{3}$$

Where:

- R_{ref} is the reference runoff coefficient $[-]$, a calibration parameter representing the maximum runoff potential before modifiers.
- S_{ratio} is the slope ratio $[-]$
- T_{ratio} is the temperature ratio $[-]$

The slope ratio S_{ratio} is given by:

$$S_{\text{ratio}} = S_\alpha + \left(\frac{S - S_{\text{min}}}{S_{\text{max}} - S_{\text{min}}}\right)^{S_\beta} \tag{4}$$

Where:

- S is the local slope at the grid cell $[-]$
- S_{min} and S_{max} are the minimum and maximum slopes in the grid $[-]$
- S_α is a small positive constant ensuring S_{ratio} is never zero, set at 0.01
- S_β is the slope power exponent $[-]$, a calibration parameter controlling the sensitivity to slope changes

The temperature ratio T_{ratio} is defined as:

$$T_{\text{ratio}} = \left(\frac{T_{\text{mean}} - T_{\text{min}}}{T_{\text{max}} - T_{\text{min}}}\right)^{T_\beta} \tag{5}$$

Where:

- T_{mean} is the mean annual temperature at each grid cell $[°C]$
- T_{min} and T_{max} are the minimum and maximum annual temperature at each grid cell $[°C]$
- T_β is the temperature power exponent $[-]$, a calibration parameter controlling sensitivity to temperature changes

The slope ratio represents the reduced water infiltration, as steeper slopes lead to faster runoff and reduced vegetation cover, while the temperature ratio capture the effect of evaporation.

Both coefficients are integrated as a reducing factor on the R_{ref}, which serves as a base reference value.

Flow Direction Calculation
After generating runoff, determining the flow direction at each grid cell is the next step for simulating streamflow across the landscape. The model use the D8 (eight-direction) algorithm via the *FlowDirectorD8* component in Landlab. The D8 algorithm assigns flow from each cell to one of its eight neighboring cells (north, northeast, east, southeast, south, southwest, west, northwest) based on the steepest downhill gradient (O'Callaghan and Mark, 1984). This method simply and efficiently models the path water takes under gravity on a rectangular grid.

To address depressions or pits where water could become trapped, the *DepressionFinderAndRouter* component is used. It identifies such areas and adjusts flow directions to route water around or through depressions, ensuring continuous downstream flow (Tucker et al., 2001). Without this adjustment, flow accumulation could be artificially halted, leading to errors in discharge calculations.

Discharge Calculation
Now that the flow direction is established, discharge Q at each grid cell is calculated using the *FlowAccumulator* component. This algorithm accumulates runoff from all upstream cells based on flow directions, effectively simulating water accumulation as it flows downstream.

The discharge is calculated as:

$$Q = \sum_i R * A_i \tag{6}$$

Where:

- Q is the discharge at each grid cell $[m^3/yr]$
- R is the runoff at each grid cell $[m/yr]$
- $\sum_i A_i$ is the accumulated upstream drainage area at each grid cell $[m^2]$

Station Placement
We use the Kling-Gupta Efficiency (KGE) to compare simulated data with observations. It provides an estimation of model performance by considering correlation, variability, and bias between the simulated versus observed data. KGE values are calculated for discharge and suspended sediment concentration (SSC) at a specified outlet node corresponding to the real-world location of the hydrological stations.

The placement of these outlets involved finding the nearest node, from the station's coordinates, with a significant discharge. Initially, the distances from each grid node to the provided station coordinates are calculated to determine their respective proximity. For each station, the script iterates through the sorted nodes by distance, selecting the first node where the discharge surpasses a threshold. This node is designated as the station location, and its coordinates and ID are stored. The threshold has been set at 5% of the maximum discharge, this has been tested for all the catchments at different resolution. To verify the accurate placement of the stations, we compared the catchment area and altitude of the station and the LEM's node defined as the outlet. The error margin for both measurements are within 5 to 8% depending on the resolution, with an exception for station 2200 that require a resolution of at least 50m. This difference is due to the proximity of station 2200 to the junction of two streams, approximately 200m, which can lead to significantly overestimating the discharge received at the outlet node. As an additional verification, the LEM generate Figure 2 to show the real-world location of the

station, in yellow (visible only when a mismatch is present), versus where they are placed, in red.

In an effort to reduce the error and improve the robustness, we attempted an approach that relied directly on the flow path. This was accomplished by recording all the nodes that belong to each catchment and place the station accordingly. While it achieved more consistent results and removed the need for a threshold parameter, it was computationally costly, adding about 10 to 20% additional runtime. This option is available but was not use in the generation of the results due to the computational cost.

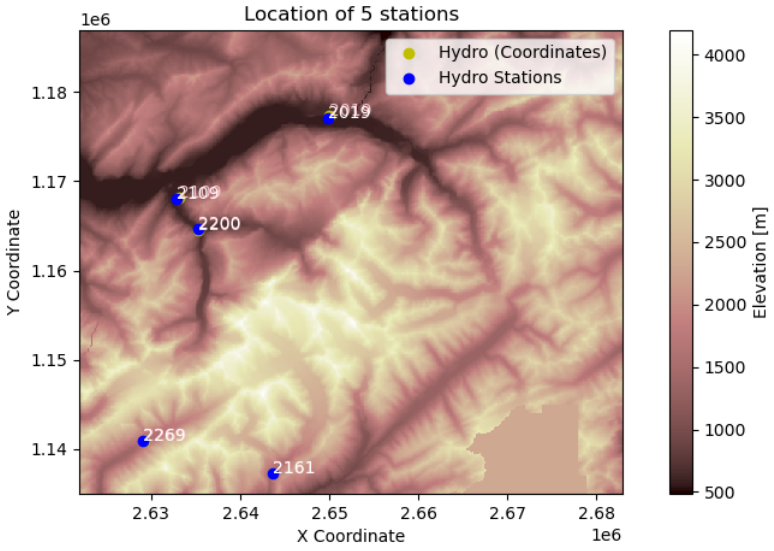


Figure 2: Station position in the grid superposed with their real world location

2.1.5 Erosion Transport

The *ErosionDeposition* component in Landlab models the processes of erosion and deposition in a landscape evolution context. This component is based on the equations presented by Davy and Lague (2009), which balances the complexity of fluvial processes, manageable parameters and simplicity in model formulation (Barnhart et al., 2019).

Erosion Rate Calculation
Erosion is calculated using a stream power law model, which considers both the topographic slope and the water discharge per unit width:

$$\dot{E} = Kq^m S^n - \dot{E}_c \tag{7}$$

where:

- $\dot{E}$ is the erosion rate $[m/yr]$
- K is an erosion efficiency factor with a variable unit, dependent on the value of m and n parameters
- q is the water discharge per unit width $[m^2/yr]$
- m and n are dimensionless power exponents use to control the influence of Q and S $[-]$
- S is the topographic slope $[-]$

- $\dot{E}_c$ is an erosion threshold, representing the critical erosion below which no erosion occurs $[m/s]$

The stream power law approach accounts for the ability of flowing water, within a river channel, to detach and transport sediment. We also incorporate a soil suitability map to approximate the distribution of different geological units, allowing the erosion efficiency factor K to vary spatially according to the underlying geology. A more detailed explanation of this approach is provided in Section 2.2.1, under the description of the Soil Suitability dataset.

Deposition Rate Calculation
Deposition occurs when the transported sediment settles due to reduced flow or increased resistance. The deposition rate is determined by the following equation:

$$D = F\frac{q_s v_s}{q} \tag{8}$$

Where:

- D is the deposition rate $[m/yr]$
- F is the deposition fraction $[-]$, representing the portion of sediment that is transformed into "fines" and are no longer available for deposition (considered removed from the mass balance system)
- q_s is the sediment flux $[m^2/yr]$
- v_s is the settling velocity $[m/yr]$
- q is the discharge per unit width $[m^2/yr]$

The equation shows that the deposition rate is proportional to sediment flux and settling velocity, but inversely related to discharge. Higher discharge keeps sediment in suspension, while lower discharge or higher settling velocity increases deposition. Additionally, v_s influences the response type: values less than 1 lead to a detachment-limited response, and values greater than 1 result in a transport-limited response.

Suspended Sediment Concentration (SSC) calculations
The *ErosionDeposition* component calculates the net sediment flux at each node by taking the difference between the erosion and deposition rates. The sediment outflux is the sediment that leaves the node and can be converted into Suspended Sediment Concentration (SSC) using the following formula:

$$SSC = q_s^{out} * \rho/q \tag{9}$$

With:

- SSC is the suspended sediment concentration $[kg/m^3]$
- q_s^{out} is the sediment outflux from the node $[m^2/yr]$
- ρ is the estimated rock density, set at 2000 $[kg/m^3]$
- Q is the discharge per unit width $[m^2/yr]$

This calculation allows us to estimate the suspended sediment concentration, linking the sediment transport processes modeled within Landlab to real-world measurable values of sediment load.

2.2 Data

2.2.1 Input Data

Digital Elevation Model (DEM)
The digital elevation model (DEM) used in this study is the swissALTI3D, provided by the Swiss Federal Office of Topography (SwissTopo) (SwissTopo, 2024). This high resolution model covers the entirety of Switzerland and Liechtenstein with elevation data free from vegetation and man-made structures, updated on a 6-year cycle, in LV95+ coordinates system (CRS) and with a 2m resolution on a regular grid. To assemble the model, the source includes laser measurements (LiDAR) and aerial imagery. The altitude accuracy vary from 50cm below 200m, up to 3m above 200m.

The data processing for geospatial raster images involves six operations to prepare the data for the integration into the LEM:

1. Obtain a CSV file containing download link for each raster (TIFF format) by indicating the desired coordinates range on the swisstopo website [1].

2. Each raster is downloaded into a directory (3000 map for the project area).

3. Each file is then downsample following a scale factor, factor 10, 20, 25 and 50 were used in the project, corresponding to 20m to 100m resolution. This is accomplished using the GDAL 3.8.5 library(GDAL/OGR contributors, 2020). This step reduces the file size and making it more manageable for subsequent calculations.

4. Using the Warp function of GDAL, the raster files are merged together. This is essential for creating a continuous spatial dataset from individual raster tiles.

5. Using PCRaster 4.4.1, we create a new DEM where each pit along a flow are modifying the elevation (Karssenberg et al., 2010). This is not necessary since Landlab reroute the flow around each pit but greatly help with the computation, since reducing the number of pit reduce the amount of computation necessary at each time step.

6. GDAL translate is used to convert the TIFF file (from step 4 and 5) to ASCII. Landlab is able to handle both format, but ASCII were easier to access and less computationally intensive to access. The ASCII is then meshed and serve as a base for the grid size and resolution.

We prepared a total of four maps for the simulation. One map covering the entire area and one map for each individual catchment: Aare, Lutschine and Lonza. These maps were prepared at different resolutions, ranging from 10 to 100m grid size.

Precipitation
The precipitation data used in this study is part of the Alpine Precipitation Grid Dataset (EURO4M-APGD)[2], provided by the Federal Office of Meteorology and Climatology (MeteoSwiss) (Isotta et al., 2014; MeteoSwiss, 2019). The dataset cover the European Alps and the adjacent regions (4.8E to 17.5E, 43N to 49N) over a 49 years period, from 1971 to 2019. It contains the daily precipitation (rainfall and snow water equivalent) sums in millimeter on a 5x5km grid using the ETRS89 CRS. These measurements are based on high-resolution rain-gauge from 8500 stations interpolated spatially with an effective resolution of approximately

[1] SwissAlti3D URL: https://www.swisstopo.admin.ch/en/height-model-swissalti3d
[2] EURO4M-APGD: https://www.meteoswiss.admin.ch/climate/the-climate-of-switzerland/spatial-climate-analyses/alpine-precipitation.html

10-20km. The interpolation was done by linear regression using regionally varying precipitation-topography relationships estimates using the PRISM algorithm (Schwarb, 2000).

The files are in monthly NetCDF format for the area covered, and two steps are required to prepare them for use in the simulation. After loading them using xarray 2024.3.0 (Hoyer and Hamman, 2017), we need to convert the CRS to LV95+, using pyproj 3.6.1, and slice them at the coordinates of interest, one per simulated area. Due to the difference in grid size, a 5 km margin has been taken in all directions to ensure that the grid is fully covered. The second step is to sum up each daily value to obtain the annual precipitation. The results are saved in a new NetCDF file containing the annual precipitation.

Given the high effective resolution, the data should be interpreted as area average for our application. Additionally, the difference in resolution and coordinate system between the DEM and the precipitation dataset make revising not as straightforward. To achieve this during the simulation, we need to take into account the local elevation during the resizing. The formula is explained in Section 2.1.4.

Temperature
The temperature dataset is provided by the Federal Office of Meteorology and Climatology (MeteoSwiss)[3] (Begert et al., 2003). It covers Switzerland and span from 1961 to 2023 and contained the mean, maximum and minimum temperature in Celsius in LV95+ CRS with a grid of 1x1km. The measurements are based on approximately 80 stations, taken two meters above the ground, and derived using daily values calculated using an automatic 10-minute measurement (to represent both daytime and nighttime conditions). All the measurements are assembled using a homogeneous time series to maintain consistency and interpolated using a regionally variable topography-temperature relationship.

The files are in yearly NetCDF format and were prepared for the simulation by extracting the data for the region covered by each DEM in a Numpy's compressed array format (NPZ), resulting in three files that contain the time series that overlap with the precipitation data (1971-2019): the yearly average, maximum and minimum. Numpy 1.26.4 was used for this operation (Harris et al., 2020).

A Lapse rate was considered in an earlier iteration, but the dataset is already taking into account the difference in elevation at a 1km resolution, which should be enough for the need of this study.

Soil Suitability
The soil suitability map utilize in this study is provided by the Office Federal of Agriculture (OFAG)[4] and the Office Federal of Statistics (OFAG, 2020). The map was originally created in 1980 and updated in 2000. It contains 144 different classes of soil types, based on their suitability for agricultural and forestry use, distributed over 11,000 polygons with a scale of 1:200,000. It considers various natural conditions such as soil depth, permeability and hydrological regime. The map units are coded with a letter and number, representing geological and geomorphological features, and store in a shape file.

In this project, we use this map to obtain an approximation of the different geological units, the following classification was of interest which were separated in four categories depending

[3]Temperature dataset URL: https://www.meteoswiss.admin.ch/climate/the-climate-of-switzerland/spatial-climate-analyses.html
[4]Soil Suitability URL: https://www.blw.admin.ch/blw/fr/home/politik/datenmanagement/geografisches-informationssystem-gis/bodeneignungskarte.html

on their estimated potential of SSC production (heuristic):

- High potential of SSC production:
  - Molasse partially covered by moraine (Ticino area)
  - Sandy Molasse
  - Flysh
  - Glacier and Névé
- Medium-High potential of SSC production:
  - Pudding, conglomerate
  - Schist (Bündnerschiefer)
- Medium potential of SSC production:
  - Limestone
- Low potential of SSC production:
  - Hard crystalline rock: Granite, Orthogneiss
  - Softer crystalline rock: Paragneiss

Each category were assigned a different erodibility factor K, allowing to approximate and calibrate the SSC production of each catchment and fit the observed dataset more closely.

2.2.2 Calibration Data

Hydrological Station
This study utilizes historical hydrological data from the *Federal Office for the Environment* (FOEN)[5]. Two types of measurements were used in this project: water discharge and suspended sediment concentration, collected from five stations in the region of Interlaken. Table 1 present information about each station and their corresponding dataset.

The dataset was prepared by regrouping all station data into CSV files, translating them in English (UTF-8), and inspected for inconsistencies and misalignment.

A more detailed overview of each catchment and station is available in Annex 5.

Study Area
This hydrological dataset were obtained in an area south-east of Interlaken, in Switzerland, characterized by varied topography and hydrological patterns. The high-resolution digital elevation model (DEM), precipitation records and freely available hydrological data make this region an ideal place to test the effectiveness of our method. The total study area cover an area of 3000 km^2 and include the three catchment with time series for the SSC: Aare (2019), Lutschine (2161) and Lonza (2269).

[5]Hydrological Dataset: https://www.bafu.admin.ch/bafu/en/home/topics/water/state/data/obtaining-monitoring-data-on-the-topic-of-water/hydrological-data-service-for-watercourses-and-lakes.html

Table 1: Hydrological Stations Information

Station 2019	**Brienzwiler**			**Station 2161**	**Blatten bei Naters**		
Stream	Aare			**Stream**	Massa		
Altitude	574 m a.s.l.			**Altitude**	1446 m a.s.l.		
Catchment Size	555 km^2			**Catchment Size**	196 km^2		
Glacier Area	15.5%			**Glacier Area**	56.5%		
Coordinates	2'649'942 / 1'177'374			**Coordinates**	2'643'694 / 1'137'291		
Discharge Data	1905-2022			**Discharge Data**	1931-2022		
SSC Data	1964-1991, 2020-2023			**SSC Data**	-		
Discharge [m^3/s]	**Avg**	**Max**	**Min**	**Discharge [m^3/s]**	**Avg**	**Max**	**Min**
	35	45	27		14	20	9.5
Station 2109	**Gsteig**			**Station 2200**	**Zweilutschinen**		
Stream	Lutschine			**Stream**	Weisse Lutschine		
Altitude	586 m a.s.l.			**Altitude**	652 m a.s.l.		
Catchment Size	381 km^2			**Catchment Size**	165 km^2		
Glacier Area	13.5%			**Glacier Area**	13.1%		
Coordinates	2'633'140 / 1'168'191			**Coordinates**	2'635'309 / 1'164'547		
Discharge Data	1908-2022			**Discharge Data**	1964-2023		
SSC Data	1964-1988, 2012-2023			**SSC Data**	-		
Discharge [m^3/s]	**Avg**	**Max**	**Min**	**Discharge [m^3/s]**	**Avg**	**Max**	**Min**
	19	23	15		10	7.9	6.4

Station 2269	**Blatten**		
Stream	Lonza		
Altitude	1523 m a.s.l.		
Catchment Size	77 km^2		
Glacier Area	24.7%		
Coordinates	2'629'128 / 1'140'919		
Discharge Data	1956-2022		
SSC Data	1966-1999		
Discharge [m^3/s]	**Avg**	**Max**	**Min**
	4.7	6.0	3.5

The largest catchment is Aare with a size of 555 km^2, high discharge and a relatively low glacier area and elevation. Lutschine catchment is relatively similar, albeit smaller, with an area of 381 km^2 and lower discharge. Lonza is the third catchment, situated 1000 m above the two others, significantly smaller (77 km^2) and covered by one fourth of its surface by glacier. We added two more catchments: one sub-catchment of Lutschine, using station Weiss Lutschine (2200). And a catchment that capture the opposite hillside of Lonza, using station Massa (2161). This station was introduced at a later stage to help the investigation on some of the challenges we faced in the calibration of station Lonza (2269). Aare and Lutschine are recharching the reservoir of Lake Brienz, while Lonza is flowing into the Rhone.

Figure 3 provides an overview of the area, including the station location and the glacier thickness. The addition of different catchment will provide comparison point, allowing to test the effect of resolution and parameters, and the viability of the calibration of an area using only key catchment.

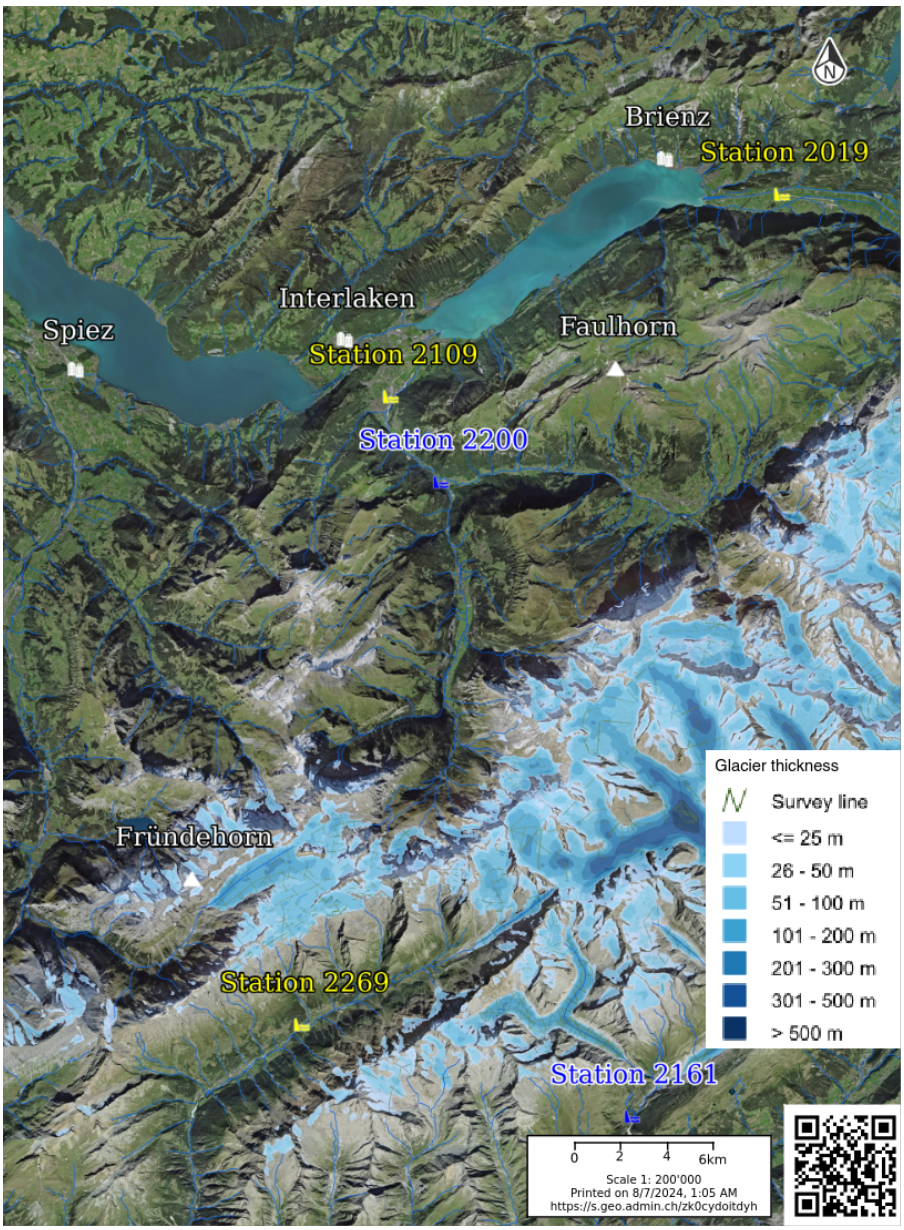


Figure 3: Overview of the simulated area. Stations with both Q and SSC observations are marked in yellow, while stations with only Q observations are marked in blue.

2.3 Model Calibration and Validation

2.3.1 Objective Function

The objective function used for model validation and calibration is the Kling-Gupta Efficiency (KGE), which evaluates model performance by combining correlation, variability, and bias. The KGE is calculated as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$$

Where:

1. r is the correlation coefficient between the observed and simulated data, indicated how well the two datasets are linearly related
2. $\alpha = \frac{\sigma_{sim}}{\sigma_{obs}}$ is the variability ratio, where σ are the standard deviations
3. $\beta = \frac{\mu_{sim}}{\mu_{obs}}$ is the bias ratio

KGE range from $-\infty$ and 1, where a value of 1 indicates a perfect match between simulated and observed data. A KGE value of $1 - \sqrt{2}$ correspond to model performance equivalent to random guessing.

The KGE is computed for both discharge and suspended sediment concentration (SSC) at various stations within the catchments. To aggregate KGE across multiple stations, each station's KGE is calculated individually, and the overall KGE score is derived by taking a weighted average of station-specific KGEs. The weights are based on the number of data point at each station, ensuring that stations with more data points have a greater influence on the overall score.

Additionally, we introduced a model accuracy metric to ease interpretation. By setting a 50% accuracy at $KGE = 1 - \sqrt{(2)}$ and 100% accuracy to $KGE = 1$, the conversion is done using the following formula:

$$\text{Model Accuracy } (\%) = KGE * 35.355 + 64.645$$

2.3.2 Calibration Process

The calibration process aims to identify the parameters controlling both hydrological and sediment processes. It is divided into three steps: preliminary tests, hydrological calibration, and sediment transport calibration. All calibration parameters are listed in Table 2. The initial parameter range was selected manually through a trial-and-error approach, guided by the typical ranges reported in the literature for Landscape Evolution Models (LEMs) in alpine environments.

Sampling Method
To explore the multidimensional parameter space, we employ a Sobol sequence, a quasi-random sequence often used in numerical simulations. Unlike purely random sampling, which can leave gaps in the parameters space, the Sobol sequences generate points that are more uniformly distributed across the entire range for each parameter. This uniform coverage ensures that all regions of the parameters space are explored more evenly, reducing the likelihood of missing important parameter combinations.

Each new point in the Sobol sequence is calculated using binary fractions and bitwise operations, ensuring that new points fill gaps left by previous ones. The method works best when the total number of runs is a power of 2, as this fully utilizes the binary structure. This allows for more evenly distributed points across the sample space.

Preliminary Tests
Before proceeding with the full calibration, preliminary tests were conducted to identify the optimal model resolution that balances accuracy and computational efficiency. We tested multiple resolutions, ranging from 10m to 100m, evaluating the model based on runtime, Kling-Gupta Efficiency (KGE), hydric loss, and station placement accuracy (elevation and catchment area). Hydric loss was calculated as the difference between the total runoff of the catchment and the discharge at the station.

Hydrological Calibration
The initial calibration was performed using KGE(Q) as the objective function. We ran 128 simulations at 50m resolution using a Sobol sequence to thoroughly explore the parameter space. From these results, we selected two parameter sets for each catchment to serve as the basis for the next calibration, focused on the Sediment Transport parameters.

Sediment Transport Calibration
Among the remaining parameters, the discharge and slope power coefficients (m and n) and the erodibility coefficient (K) are significant unknowns. The calibration of m and n is particularly sensitive, posing challenges when included in a Sobol sequence for sensitivity analysis. Many simulations resulted in excessively high erosion, often leading to errors that halted the calibration process. Therefore, instead of calibrating all three parameters simultaneously, we chose to fix m and n while leaving K unconstrained.

To guide the selection of m and n, we follow two rules: (1) n should range between 2/3 and 7/3, as proposed by Whipple et al. (2000), based on the mechanics of river incision into bedrock; (2) the ratio m/n should lie between 0.4 and 0.6. This approach is grounded in theoretical and empirical findings that link these values to the physics of stream incision. By fixing m and n within these empirically and theoretically supported ranges, we aim to balance model robustness with physical realism, allowing K to capture the spatial variability in erosion potential (Harel et al., 2016). Based on these considerations, we selected $m = 1.2$ and $n = 2$ for the main calibrations. To explore the impact of different values for these two parameters, we also tested two additional combinations: $m = 1.3, n = 2.3$ and $m = 1.1, n = 2.3$.

The final step in determining the optimal parameter sets involved calibrating the remaining parameters: the erosion threshold E_c, the effective settling velocity (V), the fraction of eroded material converted into fines (F), and the four erodibility coefficients (K). We conducted 128 simulations for each of the two hydrological calibrations using Sobol sequences at a 50m resolution.

Calibration Session	Label	Low Range	High Range	Unit	Equation
Preliminary	Precipitation Lapse Rate 1	0.001	0.005	*C/m	1
	Precipitation Lapse Rate 2	0.0001	0.005	*C/m	1
	Precipitation Lapse Rate 3	0.0001	0.005	*C/m	1
	Precipitation Lapse Rate 4	0.0001	0.005	*C/m	1
	m	–	–	–	7
	n	–	–	–	7
Hydrological	Runoff Ratio Ref.	0.7	0.95	–	3
	Slope Power Exponent	0.1	0.9	–	4
	Temp Power Exponent	0.1	0.9	–	5
Sedim. Transport	K High	1.00E-10	1.00E-06	var	7
	K MedHigh	1.00E-10	1.00E-06	var	7
	K Med	1.00E-10	1.00E-06	var	7
	K Low	1.00E-10	1.00E-06	var	7
	E_c	0	0.7	m/s	7
	v	0.1	9.5	m/s	8
	F	0	0.9	–	8

Table 2: The table contains all the calibration parameters, sorted according to the order in which they were calibrated. First are the preliminary tests, followed by the hydrological calibration, and then the sediment transport calibration.

2.3.3 Validation Approach

A full model validation, involving performance testing against entirely independent datasets, was not conducted in this study. Instead, validation was approached by performing sensitivity analysis and evaluating the spatial and temporal transferability of the calibrated parameters.

The model's spatial transferability was evaluated by applying parameter sets calibrated on one catchment (e.g., Aare) to other catchments (e.g., Lutschine and Lonza). This approach tests

whether parameters optimized for one area can successfully simulate discharge and suspended sediment concentration in a different geographic context. The results showed that while discharge simulations transfer relatively well between catchments, sediment transport simulations (SSC) were much more sensitive to local geomorphic conditions, resulting in lower transferability for SSC across catchments.

The temporal transferability of the model was evaluated by extending the simulation period beyond the calibration timeframe. Due to the limitations of our precipitation and temperature datasets, which span only a 49-year period, we needed a method to simulate longer timescales without introducing new data or altering the calibration. To achieve this, we looped the 49-year precipitation and temperature records, effectively repeating the datasets to cover a total simulation period of 1000 years.

This approach maintains consistency with the calibration period and ensures that the model operates under the same climatic conditions as during the calibration phase. It allows us to assess the model's performance over long timescales without introducing additional variables or uncertainties that new data might bring. However, this method introduces an artificial cyclicity into the model inputs, as the same 49-year climatic patterns are repeated throughout the simulation. This cyclicity does not account for natural climatic variability and trends that occur over centennial or millennial timescales, such as long-term climate change or rare extreme events.

The artificial repetition of climatic data is a limitation that must be considered when interpreting the long-term simulation results. It may affect the model's ability to capture long-term geomorphic processes and sediment transport dynamics accurately. The lack of natural variability could lead to an underestimation of extreme events' impact on landscape evolution, which are critical drivers of geomorphic change (Pazzaglia, 2003). Despite this limitation, the looping method provides a practical means to evaluate the model's stability and performance over extended periods. It offers insights into its temporal robustness within the constraints of the available data.

2.4 Previous Consideration for Calibration through Machine Learning

Initially, the aim was to use machine learning techniques to identify the patterns between the data and model parameters, potentially skipping or limiting the optimization steps of LEMs. However, a few challenges were encountered.

At the simplest implementation, we were trying to get a pattern between the validation metric of the LEM and its parameters. However, actual field data is necessary in the training to obtain any correlation that could explain the model parameters. One challenge is the difference in temporal and spatial resolution between the data. The training feature is in raster format that are spread in time and space, for example the precipitation, the discharge and the elevation. While the target model parameters are singular value that are fixed across the 50 years, and not spatially distributed (for most parameters). This dimension mismatch, both between the features and the target and within the target data themselves, made the design of the training dataset challenging. A potential solution would have been to reduce the feature dimension by transforming them in mean, standard deviation, etc. However, it would lead to a significant lost of information, so an alternative approach was required.

Two alternatives approaches were considered:

1. The use of more advance technique, such as convolution neural network algorithm (CNN) that are specifically designed to handle raster or images data. An interesting method, but much more involved to implement and calibrate. Moreover, ones can start questioning on the need and application of such a pipeline. Potentially, we could use the CNNs to directly predict the output of the LEMs instead of predicting the LEMs parameters, effectively using the CNNs as a generalized landscape simulation model. This is out of the scope of this project, but further research on that front could be considered.

2. On the other hand, the current LEM implementation is largely automated, run in parallel and allow changing area with relative ease. An adaptable simulation able to generate numerous parameters permutations and where the implementation of different calibration is straightforward, due to how the code is structured.

Focusing on the strength of this code instead, we started looking for statistical or computational approaches able to identify patterns in numerous parameters permutation and able to determine the best combination. This led us to the current methodology described in this report.

3 Results

In this section, we present the results of this project, aiming to cover all the possibilities in a well conceived manner. First, we will do preliminary tests to define an optimal resolution that provide a balance between accuracy and computational efficiency.

The second part present the calibration on the three selected catchments, selecting multiple "best model". Third, using the "best models", we will investigate the parameter's sets transferability both in time (running for 1000 of years) and space (using parameters set between catchment and to the whole area), parameters sensitivity (which parameters are the most important) and further tests on the resolution (how are the output changing with different resolutions). Lastly, we will take test the overall calibration process by investigating how long it takes with this method to reach good performing models.

3.1 Preliminary Tests

The preliminary tests were conducted to identify the optimal resolution that balances model accuracy and computational efficiency. For these tests, we used default parameters (Table 5), the precipitation dataset, and a fixed runoff coefficient of 0.6. The results indicated that while finer resolutions significantly increased computational time, the model's performance in terms of KGE and hydric loss showed minimal variation across different resolutions. However, at coarser resolutions (e.g., 100m), inaccuracies in station placement led to significant errors in catchment area estimation.

Table 4 shows the performance of the model at the three main stations: Station 2269 in the Lonza catchment, and Stations 2109 and 2200 in the Lütschine catchment. As expected, runtime increases significantly with finer resolutions. For example, in the Lonza catchment, the runtime jumps from 3.38 seconds at 100m to 240.83 seconds at 10m resolution.

Most evaluation metrics show minimal variation between different resolutions, indicating stable model performance across scales. However, a notable exception occurs at 100m resolution for Station 2200, where catchment area errors lead to a significant overestimation, doubling the catchment area. Further investigation shows that station 2200 is close to a junction, and incorrect positioning has led to the inclusion of an additional sub-catchment.

The 50 m resolution offers a balanced compromise, providing an acceptable runtime, allowing hundreds of simulations to be run in around 30 minutes, while retaining sufficient accuracy for proper station placement.

After selecting the resolution, we conducted tests to estimate the precipitation lapse rate at different altitudes. We ran 128 simulations with a fixed runoff coefficient of 0.6 and selected the model that achieved the highest KGE(Q). The final results are presented in Table 3.

Label	Calibration Results	Unit	Altitude Range
Precipitation Lapse Rate 1	0.00371	*C/m	0-1000m
Precipitation Lapse Rate 2	0.00155	*C/m	1000-2000m
Precipitation Lapse Rate 3	0.00167	*C/m	2000-3000m
Precipitation Lapse Rate 4	0.00081	*C/m	+3000m

Table 3: Calibration results for the precipitation lapse rate.

Catchment	Runtime [s]	Station Z Error [%]	Catchment Area Error [%]	Hydric Loss [m³/s]	KGE(Q)	KGE(SSC)	std(ssc)	mean(ssc)	std(Q)	mean(Q)
Lonza 10m	240.83	0.1716	-0.4249	4.9798E-05	0.1469	-0.2087	0.0055	0.2137	0.4769	3.3586
Lonza 20m	41.22	0.0357	-0.3512	-1.8851E-07	0.1472	-0.2054	0.0056	0.2126	0.4760	3.3515
Lonza 50m	8.54	-0.0970	-1.2403	5.2798E-08	0.1451	-0.2079	0.0059	0.2114	0.4813	3.3892
Lonza 100m	3.38	-0.2941	-1.3896	1.7958E-07	0.1457	-0.2073	0.0064	0.2125	0.4797	3.3788
Lu-2109 20m	1043.22	0.2395	0.1400	-4.5893E-04	0.6170	-0.4342	0.0089	0.3415	1.8717	1.8717
Lu-2109 50m	55.16	0.3660	0.0499	-6.2782E-04	0.6192	-0.4426	0.0094	0.3432	1.8658	16.6746
Lu-2109 100m	9.94	0.1983	5.5252	-2.7972E-04	0.6410	-0.4368	0.0101	0.3504	1.7970	15.9588
Lu-2200 20m	1043.22	0.0756	0.1438	-4.5893E-04	0.5548				0.8434	7.4486
Lu-2200 50m	55.16	0.1067	0.0551	-3.6664E-05	0.5573				0.8381	7.4019
Lu-2200 100m	9.94	0.0706	-109.0848	4.9298E-07	-0.8191				1.7481	15.4645

Table 4: Resolution Test Results Table.

3.2 Parameters Sensitivity Analysis

The correlation matrix in Figure 4 provides an overview of the relationships between the calibration parameters and the KGE for both discharge and suspended sediment concentration (SSC). It's important to note that the KGE for SSC remains negative across all simulations, indicating potential inaccuracies in the model's handling of sediment transport. Therefore, any conclusions regarding parameters related to sediment transport should be approached with caution. Furthermore, the interaction between sediment transport and discharge complicates the isolation of each parameter's influence (e.g., high erosion can lead to increased discharge).

The sensitivity analysis was challenging due to the number of parameters and their interactions. This complexity is evident in the correlation matrix, where parameters known to influence sediment transport appear to have higher correlations with discharge. For example, the parameter v (settling velocity) shows a correlation of -0.4 with discharge, and F (deposition fraction) correlates at 0.3 with discharge. Conversely, the slope exponent, which primarily affects discharge, has a correlation of 0.21 with KGE(SSC). As a result, the sensitivity analysis largely relied on manually testing each parameter through a trial-and-error process. Due to time constraints, the depth of this analysis was limited.

Several key observations were made, notably that the parameters K(high) and K(med) exerted the greatest influence across all catchments. This suggests that these regions are predominantly characterized by certain soil types, or that these soils contribute most significantly to sediment production.

Additionally, the parameters m and n (from the stream power law equation) had to be fixed, as slight variations in these values required vastly different sets of parameters for sediment transport. Interestingly, despite the challenges in adjusting m and n, the best model results

in Table 6 show that the variations attempted with different m and n values led to relatively similar accuracy ratings (e.g. models 2.Lu and 3.Lu). This suggests that while m and n have a large influence, their influence on the overall model performance might not drastically change the accuracy when within certain ranges. Further analysis is required to better understand this relationship.

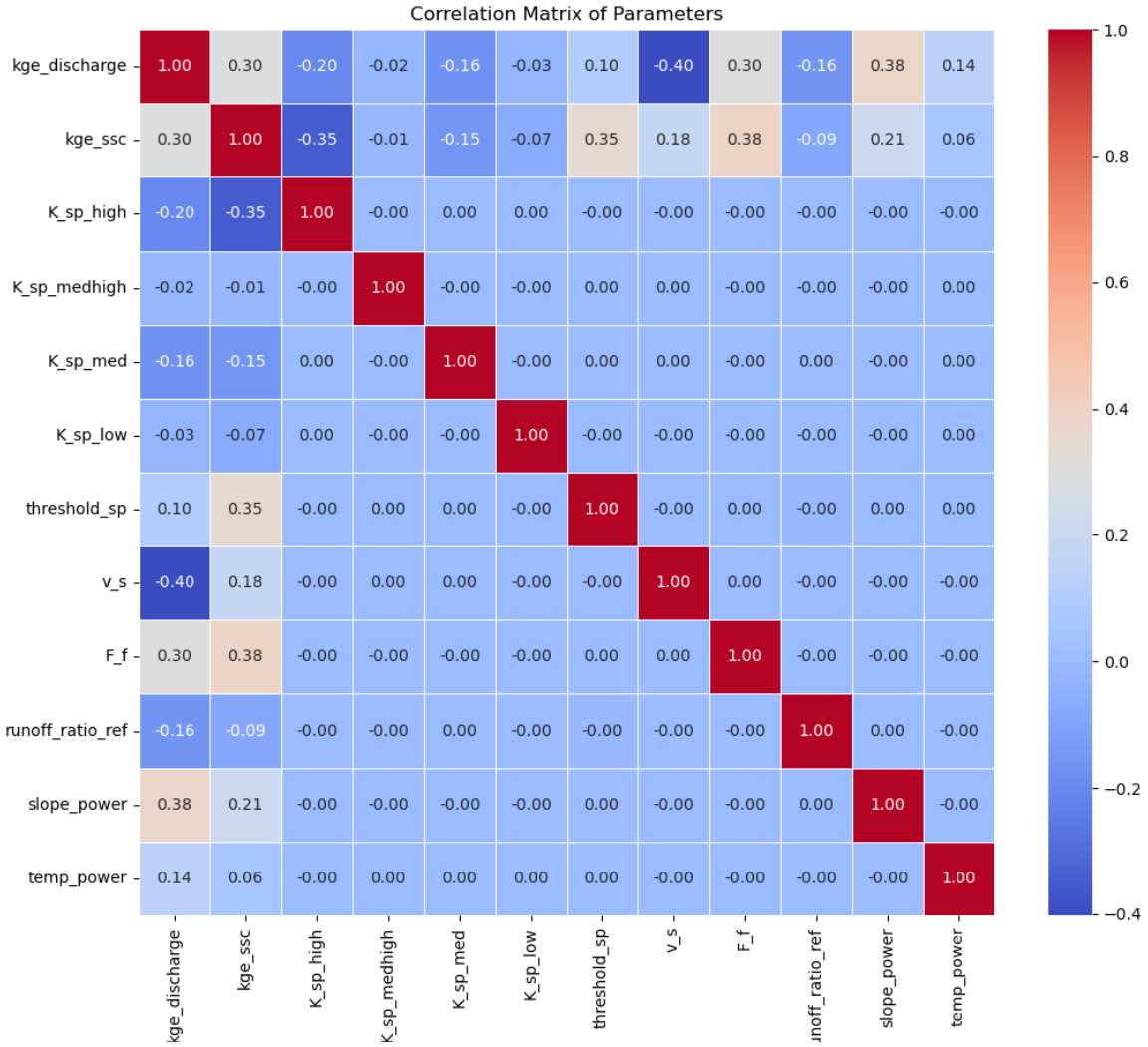


Figure 4: Correlation Matrix on all calibration parameters (except m and n) for Lütschine catchment using 1224 simulations with Sobol sequence

3.3 Calibration of Selected Catchments

The calibration process focused on the three selected catchments, aiming to identify the parameters sets that best matched the observed data. The "best models" were identified based on KGE scores for discharge and suspended sediments.

3.3.1 Hydrological Calibration

The initial calibration was performed using KGE(Q) as the objective function. We ran 128 simulations at 50m resolution using a Sobol sequence to thoroughly explore the parameter space. The results are detailed in the Annex: Table 10, Table 11, and Table 12. From these

results, we selected two parameter sets for each catchment to serve as the basis for the next calibration, focused on the Stream Power Law.

For each catchment, the first selected parameter set is the one with the highest KGE(Q) value, representing the best model performance. The second set was chosen to include a runoff coefficient that differs noticeably from the first, while ensuring the KGE(Q) remains high. This approach allows us to have two parameter sets with distinct characteristics, providing a wider range of model behaviors for the following analysis. The selected parameter sets for each catchment are as follows:

- **Aare Catchment:** Model 05 and Model 99
- **Lonza Catchment:** Model 215 and Model 125
- **Lütschine Catchment:** Model 35 and Model 93

3.3.2 Sediment Transport Calibration

We conducted 128 simulations for each of the two hydrological calibrations using Sobol sequences at a 50 m resolution. Detailed results are provided in the Annex (Table 13, 14, 15, 16, 17, 18, 19, 20). From each table, we selected one parameter set with the highest KGE(SSC) and another set with a distinctly different erodibility factor K.

Params Sets	Model ID	runoff_ratio_ref	slope_power	temp_power	m	n	K(high)	K(highmed)	K(med)	K(low)	E_c	V	F
Default	default	0.6000	0.0000	0.0000	1.2	1	9.0000E-09	9.0000E-09	9.0000E-09	9.0000E-09	0.0000	1.0000	0.0000
1.A-1-1	model_32	0.9410	0.0310	0.4014	1.2	2	9.8443E-10	7.0757E-10	9.7203E-10	3.9654E-09	0.5153	4.6522	0.4712
1.A-1-2	model_45	0.9410	0.0310	0.4014	1.2	2	3.6455E-09	2.2390E-09	3.3978E-09	3.5656E-09	0.5873	8.0171	0.7773
1.A-2-1	model_57	0.7100	0.0270	0.1277	1.2	2	4.8448E-10	7.0506E-10	8.8443E-10	3.7858E-09	0.4617	3.8709	0.3595
1.A-2-2	model_52	0.7100	0.0270	0.1277	1.2	2	3.4171E-09	3.2413E-09	3.5451E-09	3.6789E-09	0.6340	6.9320	0.7609
1.Lu-1s-1-1	model_77	0.9466	0.1986	0.1070	1.2	2	2.8659E-09	1.8791E-09	1.7399E-09	2.7510E-09	0.5323	4.2505	0.3466
1.Lu-1s-1-2	model_79	0.9466	0.1986	0.1070	1.2	2	1.3179E-09	5.0100E-10	3.0376E-10	3.6393E-09	0.1227	8.6728	0.2254
1.Lu-1s-2-1	model_32	0.7781	0.1269	0.1171	1.2	2	3.0456E-09	5.8693E-09	7.2862E-10	3.2291E-09	0.5975	1.3111	0.1237
1.Lu-1s-2-2	model_24	0.7781	0.1269	0.1171	1.2	2	2.0238E-09	4.5152E-09	1.2144E-09	1.9057E-10	0.2248	3.7389	0.2116
1.Lu-2s-1-1	model_08	0.9466	0.1986	0.1070	1.2	2	1.3565E-09	1.9307E-09	2.0814E-10	1.3447E-10	0.0518	0.7864	0.1528
1.Lu-2s-1-2	model_56	0.9466	0.1986	0.1070	1.2	2	1.4792E-09	1.6066E-09	1.2825E-09	5.4809E-10	0.2155	0.1683	0.3876
1.Lu-2s-2-1	model_61	0.7781	0.1269	0.1171	1.2	2	1.2198E-09	8.3950E-10	1.6809E-09	1.7245E-09	0.1585	0.3658	0.1589
1.Lu-2s-2-2	model_102	0.7781	0.1269	0.1171	1.2	2	1.9410E-09	1.7916E-09	9.8964E-10	2.8689E-10	0.0015	8.6885	0.7591
1.Lo-1-1	model_53	0.9595	0.1047	0.3338	1.2	2	1.3590E-09	2.3523E-08	2.6708E-08	9.5651E-09	0.4513	8.1119	0.7809
1.Lo-1-2	model_18	0.9595	0.1047	0.3338	1.2	2	5.1118E-09	8.6842E-10	8.3274E-09	3.3319E-08	0.4455	2.5213	0.5677
1.Lo-2-1	model_103	0.7534	0.1038	0.1064	1.2	2	2.7087E-08	3.9483E-08	6.8863E-09	9.1900E-09	0.6324	8.0099	0.7824
1.Lo-2-2	model_137	0.7534	0.1038	0.1064	1.2	2	5.4862E-10	3.9243E-08	2.7369E-09	3.0548E-08	0.3674	0.2937	0.3987
2.Lu-1s-1-1	model_15	0.9466	0.1986	0.1070	1.3	2.3	7.6075E-10	6.1548E-10	5.4567E-10	8.0814E-10	0.6786	7.4347	0.4915
2.Lu-2s-1-1	model_67	0.9466	0.1986	0.1070	1.3	2.3	3.2115E-10	3.6941E-10	1.6356E-10	1.9897E-10	0.3016	0.2247	0.1413
3.Lu-1s-1-1	model_10	0.9466	0.1986	0.1070	1.1	2.3	1.6446E-08	8.7026E-09	3.0776E-09	2.3506E-08	0.4758	0.8311	0.3995
3.Lu-2s-1-1	model_78	0.9466	0.1986	0.1070	1.1	2.3	7.1195E-09	2.0871E-10	2.9843E-09	6.8270E-09	0.2640	1.4247	0.1491

Table 5: Best Model Parameter Sets. Each parameter set is denoted by "X.Y-Z-W," where X represents the fixed m/n calibration number, Y indicates the catchment name (e.g., 'A' for Aare), Z corresponds to the hydrological calibration model number, and W represents the stream power law calibration. Each catchment has four primary calibrations, with additional variations included for different m/n values.

3.3.3 Best Models Results

This calibration process resulted in twenty "best models": four parameter sets per catchment, four additional sets for Lütschine calibrated using two stations and four sets accounting for variations in m and n. These parameter sets are summarized in Table 5.

Table 6 summarizes the evaluation of the best models based on the Kling-Gupta Efficiency (KGE) for discharge (Q) and suspended sediment concentration (SSC). The KGE(Q) values generally indicate good performance across most parameter sets, particularly those calibrated on the Lutschine catchment (e.g., 1.Lu-1s-1-1 and 1.Lu-1s-1-2), which consistently achieve the

Params Sets	KGE(Q)	KGE(SSC)	MA(Q)	MA(SSC)
1.A-1-1	0.61834	-0.24839	86.50638	55.86327
1.A-1-2	0.61827	-0.28091	86.50377	54.71328
1.A-2-1	0.61687	-0.27069	86.45428	55.07471
1.A-2-2	0.61670	-0.28459	86.44847	54.58334
1.Lu-1s-1-1	0.65356	-0.45042	87.75153	48.72043
1.Lu-1s-1-2	0.65355	-0.46153	87.75138	48.32754
1.Lu-1s-2-1	0.64836	-0.44992	87.56791	48.73821
1.Lu-1s-2-2	0.64836	-0.46069	87.56794	48.35748
1.Lu-2s-1-1	0.65355	-0.48965	87.75134	47.33325
1.Lu-2s-1-2	0.65354	-0.52791	87.75097	45.98067
1.Lu-2s-2-1	0.64836	-0.45916	87.56761	48.41153
1.Lu-2s-2-2	0.64846	-0.52030	87.57135	46.24983
1.Lo-1-1	0.15427	0.17962	70.09908	70.99553
1.Lo-1-2	0.15426	0.13143	70.09883	69.29165
1.Lo-2-1	0.15342	0.17330	70.06899	70.77209
1.Lo-2-2	0.15348	0.16798	70.07128	70.58396
2.Lu-1s-1-1	0.65356	-0.44731	87.75156	48.83027
2.Lu-2s-1-1	0.65353	-0.51877	87.75063	46.30405
3.Lu-1s-1-1	0.65355	-0.49547	87.75142	47.12756
3.Lu-2s-1-1	0.65353	-0.57595	87.75052	44.28232

Table 6: Table showing the evaluation of the best model. For each parameter set, the KGE for discharge and SSC is provided, along with the Model Accuracy [%]

highest KGE(Q) values around 0.65. The Aare catchment sets (e.g., 1.A-1-1, 1.A-1-2) also perform well, with KGE(Q) values slightly above 0.61, confirming their suitability for modeling discharge. In contrast, the performance for SSC, represented by KGE(SSC), is notably weaker across all parameter sets, with most values being negative, indicating poor model performance in simulating sediment transport. Parameter sets calibrated on the Lutschine catchment, although showing relatively better discharge performance, perform poorly for SSC, with values often below -0.4, and some as low as -0.57595. This trend highlights the difficulty of accurately modeling sediment dynamics across catchments, suggesting that the sediment processes are more complex and sensitive to calibration. Interestingly, the models calibrated on the Lonza catchment (e.g., 1.Lo-1-1, 1.Lo-1-2) show a contrasting pattern: while their KGE(Q) values are much lower, around 0.15, they achieve the highest KGE(SSC) values among all sets, with some positive values up to 0.17962. This suggests that the Lonza-specific parameter sets are better tailored to capture sediment transport processes, likely due to the distinct hydrological and geomorphological characteristics of the Lonza catchment, which includes higher altitudes and glacial influences.

The discharge and SSC time series for the Aare catchment, using model 1.A-1-1, are presented on Figure 5, with a corresponding KGE(Q) of 0.62. The observed discharge is represented by solid blue circles, while the simulated discharge is shown with blue dashed lines marked by 'x'. Additionally, the bar plot in the background illustrates the mean annual precipitation in the catchment, which is plotted on the secondary y-axis to provide context for discharge variations. Both simulated and observed discharge values exhibit a similar interannual trend, indicating that the model capture the general variability. However, they are noticeable discrepancies in magnitude, for example in 2003, where the simulated tend to underestimate. While both observed and simulated discharges show an expected response to precipitation, the observed

discharge tends to peak higher in drought years compared to the simulated discharge. This might indicate limitations in the model's capacity to capture the hydrological processes, and probably a lack of seasonal variability due to the yearly timestep.

Table 7 present the results of the Lonza catchment, using model 1.Lo-1-1, with a corresponding KGE(Q) of 0.15. The comparison between simulated and observed discharges reveals a general pattern of variability that aligns with precipitation trends, suggesting that the model captures some hydrological fluctuations. However, the model struggles to consistently reproduce observed peaks and troughs, showing weaker performance compared to other catchments. This discrepancy indicates that the Lonza catchment involves additional processes not adequately represented by the model, likely due to its higher altitude and significant glacial coverage, which introduce complex dynamics not fully captured by the current parameterization.

Table 6 and Table 8 contain the time series for the suspended sediment concentration for the Aare and Lonza catchment. The simulated SSC captures some inter-annual variability of the observed SSC, indicating the model's ability to reflect general sediment transport patterns. However, alignment between simulated and observed values is inconsistent, especially during pronounced peaks in observed SSC. Significant discrepancies occur during high SSC events, such as those around 1980, 1990, and 2010, where observed values show sharp peaks that the model fails to replicate in magnitude and timing. These peaks indicate intense sediment transport periods that the model does not accurately capture. The simulated SSC is generally more stable and subdued compared to the observed data. Suggesting an underestimation of variability and extremes in sediment dynamics, potentially due to missing seasonal events like heavy rainfall or snowmelt. During periods of lower observed SSC, the simulated values align somewhat better but still show noticeable differences. It could indicate that while the model represents base-level sediment transport reasonably well, it struggles with high-magnitude events.

3.4 Parameter Transferability and Sensitivity Analysis

To evaluate the robustness of the best-performing models, we tested the transferability of the parameters in terms of space, by using parameters set on different catchment which they were calibrated for and on the whole area, and temporal by extending the simulation period to 1000 years.

3.4.1 Spatial Transferability

The transferability of parameter sets across different catchments and their performance on the entire study area are shown in Table 7 for discharge and Table 8 for SSC. The results demonstrate the significant challenges of applying parameters calibrated in one catchment to another, especially when extended to the entire area. Notably, many simulations failed due to excessive erosion, particularly for models calibrated on the Lonza catchment, where none achieved successful outcomes.

Table 7 presents the Kling-Gupta Efficiency (KGE) values for discharge (Q) at various stations. The simulations used parameter sets originally calibrated on individual catchments. For instance, the parameter set 1.A-1-1-50m, calibrated on the Aare catchment, performed moderately well across most stations, with KGE values of 0.61886 at Aare 2019 and 0.56078 at Lutschine 2109. Even on the Lonza catchment, its performance was relatively acceptable, with values close to those of the models calibrated directly on Lonza, which barely exceeded 0.15.

The parameter set 1.Lu-1s-2-1-50m, calibrated on the Lutschine catchment, showed varied performance across the stations but generally performed worse than 1.A-1-1-50m except in its

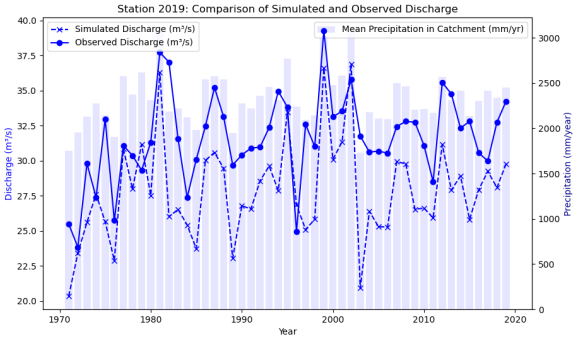


Figure 5: Time series of the observed and simulated discharge on Aare using model 1.A-1-1

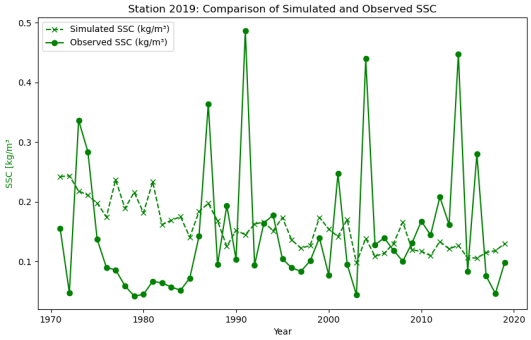


Figure 6: Time series of the observed and simulated SSC on Aare using model 1.A-1-1

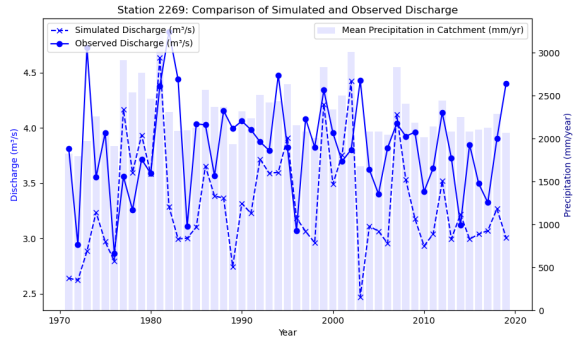


Figure 7: Time series of the observed and simulated discharge on Lonza using model 1.Lo-1-1

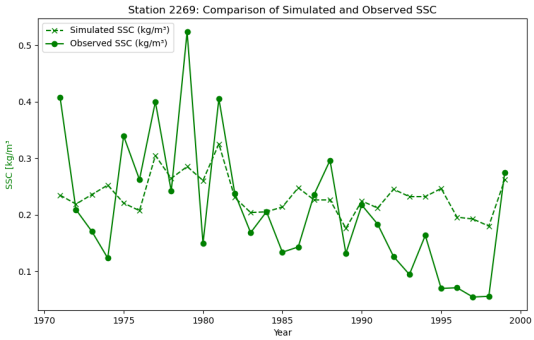


Figure 8: Time series of the observed and simulated SSC on Lonza using model 1.Lo-1-1

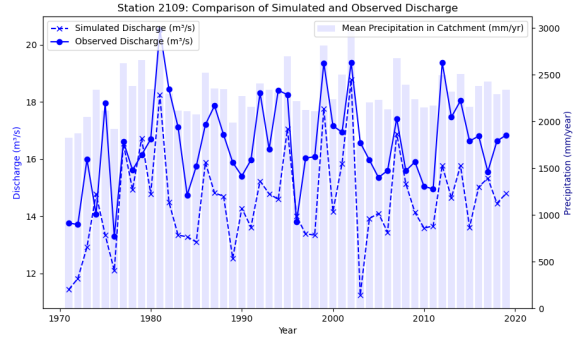


Figure 9: Time series of the observed and simulated discharge on Lütschine using model 1.Lu-1s-1-1

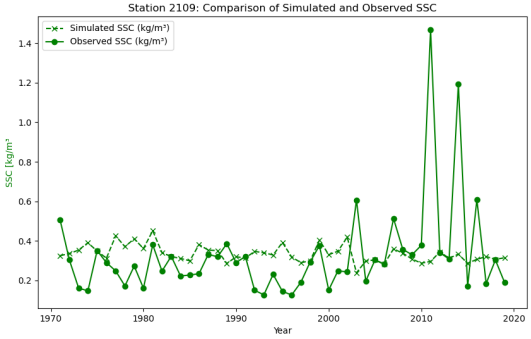


Figure 10: Time series of the observed and simulated SSC on Lütschine using model 1.Lu-1s-1-1

own catchment. When applied to the entire area, this set resulted in a negative overall KGE (-0.10943). Overall, the Aare catchment appears to be a suitable candidate for the calibration of the entire area, possibly because it is the largest catchment, covering 20% of the total area.

Table 8 highlights the KGE values for sediment concentration (SSC) at the same stations, where the SSC dataset was available. Both parameter sets performed poorly, demonstrating their inability to represent sediment transport accurately in catchments where they were not calibrated. This suggests that sediment transport processes are particularly sensitive and require

careful and dedicated calibration. These findings underscore the need for catchment-specific parameterization to improve model reliability and performance across varied landscapes.

These results underscore the difficulty of finding a parameter set that performs consistently across diverse catchments.

3.4.2 Consistency over extended runtime

To evaluate the consistency of model performance over extended periods, we conducted four simulations on the Lonza catchment using the "best models". This approach allows us to evaluate the variability in discharge and suspended sediment concentration (SSC) among models with different calibrations. The results are shown in Figure 11 for discharge and Figure 12 for SSC, where the differences in discharge and SSC were calculated using 1.Lo-1-1 as the reference model. The discharge and SSC values from the other three models were subtracted from the reference, and these differences are plotted over time. Due to the limitations of our precipitation and temperature datasets, which span only 49 years, we extended the simulations to 1000 years by looping these datasets. While this approach maintains consistency with the calibration data, it introduces an artificial cyclicity that must be considered when interpreting the long-term simulation results.

Examining the discharge results in Figure 11, models with identical hydrological calibrations, such as 1.Lo-2-2 and 1.Lo-2-2, demonstrate similar behavior, which is consistent with expectations, and show a limited influence of the Stream Power Law parameters on the discharge. During the first 400 years, the models with the same hydrological parameters produce near identical discharge outputs. However, minor variations begin to emerge after the second half, with a gradual divergence among models. It might be due to the poor calibration of the uplift, however negligeable difference were shown with an uplift between 0.1 and 0.5 mm. The deviation between all the models reach a difference of 0.35 m^3/s, representing approximately a divergence of 10% compare to the reference model, which can be acceptable over a long period.

In contrast, the SSC results depicted in Figure 12 reveal a slightly different pattern. Despite the discharge divergence observed after 400 years, the SSC differences do not display the same increasing trend over time. This is a bit surprising as it indicates that SSC, a process influenced by sediment dynamics and complex interactions between hydrology and geomorphology, does not directly mirror the discharge variability in these simulations.

The models 1.Lo-1-1 and 1.Lo-2-2 remain closely aligned throughout the simulation period, as do models 1.Lo-1-2 and 1.Lo-2-1, effectively forming two pairs. However, upon comparing their parameters, no clear trends emerge that fully explain these groupings. In addition, a significant divergence in SSC is observed early in the simulation, particularly for Lonza-1-2, which reaches a difference of over 0.1 kg/m^3, representing a variation exceeding 50% compared to the reference model. This finding could indicate that even minor parameter adjustments can lead to substantial differences in SSC predictions.

To further investigate the model's behavior over the extended simulation, we compared the changes in elevation after 1000 years. The results showed minimal differences, with a drop of only 0.2m in mean elevation and standard deviation. This minimal change is unrealistic, as significant landscape evolution would be expected over such a timescale. The underestimation of erosion is likely due to the model's inability to represent extreme events and accurately simulate sediment transport processes, leading to insufficient cumulative erosion over time.

These findings suggest that while the model maintains some consistency in discharge predic-

tions over extended periods, its performance in simulating SSC and landscape evolution is limited. The artificial repetition of climatic data and the lack of representation of extreme events contribute to the model's shortcomings in capturing long-term geomorphic changes.

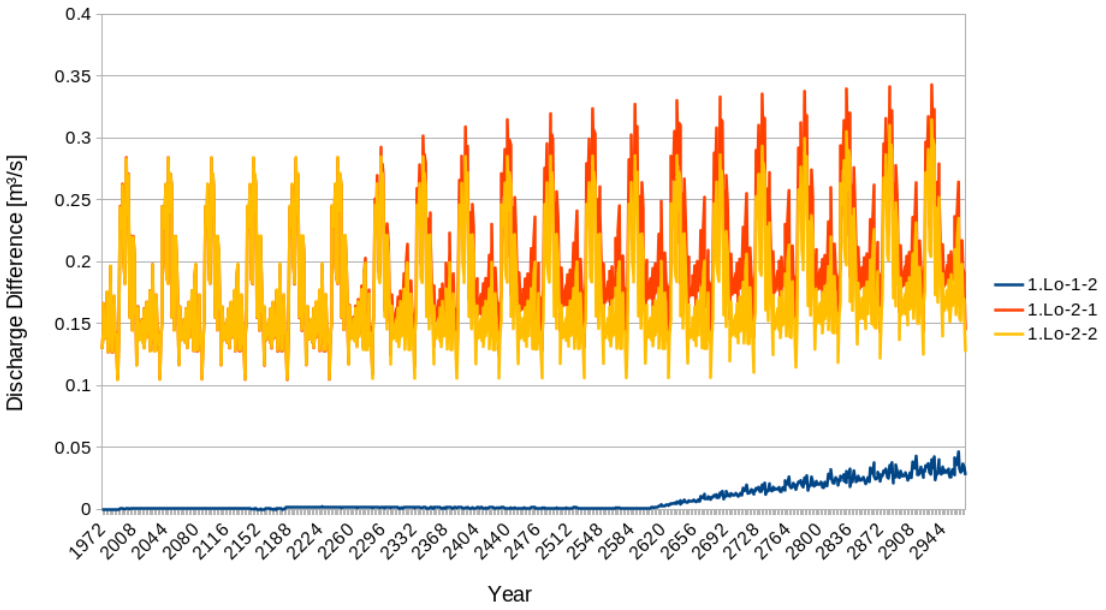


Figure 11: Time series of the discharge differences between model 1.Lo-1-1 and three other calibrated models (1.Lo-1-2, 1.Lo-2-1, 1.Lo-2-2) on the Lonza catchment.

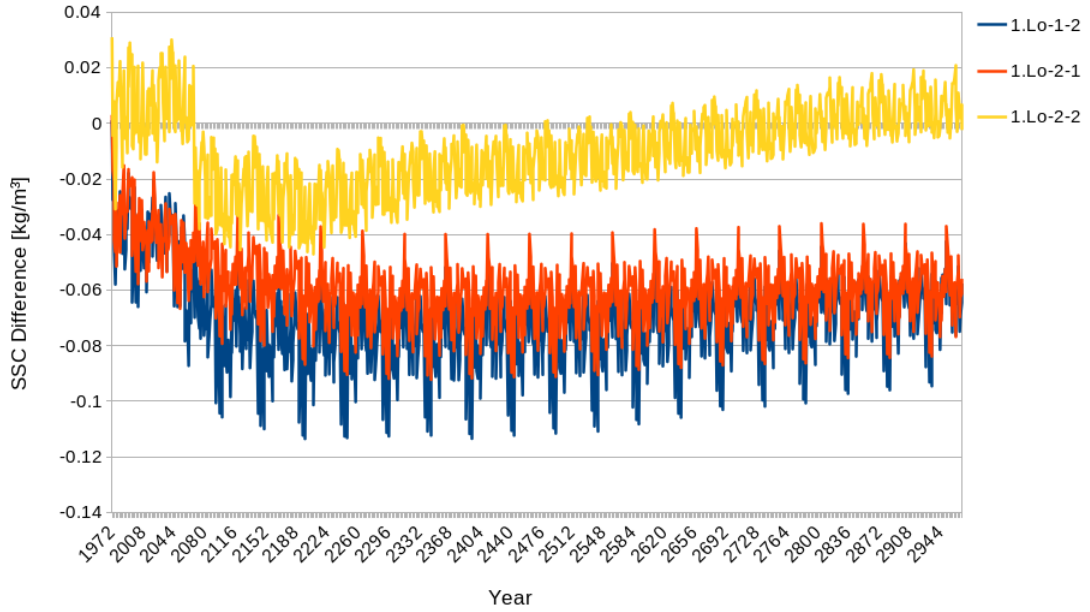


Figure 12: Time series of the SSC differences between model 1.Lo-1-1 and three other calibrated models (1.Lo-1-2, 1.Lo-2-1, 1.Lo-2-2) on the Lonza catchment.

3.5 Calibration Process Evaluation

This section evaluates the efficiency of the calibration process, focusing on the computational time and resources required to achieve well-performing models. The evaluation highlights the feasibility and scalability of the calibration process when extending it from individual catchments to the entire study area.

Parameter Set	Aare 2019	Lutschine 2109	Lutschine (sub) 2200	Lonza 2269	Area
1.A-1-1-50m	0.61886	0.56078	0.53790	0.12206	0.35890
1.Lu-1s-2-1-50m	0.46485	0.61863	0.53619	0.04586	-0.10943

Table 7: KGE(Q) at stations within the study area. The simulations were performed using parameter sets calibrated on a single catchment, with a 50m resolution and a 1000-year extended runtime. Precipitation and temperature data were applied in a 49-year loop. The overall KGE for the area was calculated as the mean, weighted by the number of data points at each station

Parameter Set	Aare 2019	Lutschine 2109	Lonza 2269	Area
1.A-1-1-50m	-0.28940	-0.71972	-0.55792	-0.51675
1.Lu-1s-2-1-50m	-1.27342	-0.45781	-0.52929	-0.78882

Table 8: KGE(SSC) at stations within the study area. The simulations were performed using parameter sets calibrated on a single catchment, with a 50m resolution and a 1000-year extended runtime. Precipitation and temperature data were applied in a 49-year loop.

Table 9 presents a comparison of the computational efficiency of sequential and parallel calibration approaches for two different Digital Elevation Models (DEMs): the entire study area ("Area") and the Aare catchment. The evaluation is based on the time required for a single simulation run, the total calibration time using sequential runs, and the improvements achieved by parallelizing the calibration process.

The table shows that calibrating the entire study area is significantly more demanding than calibrating the Aare catchment. Specifically, a single simulation run for the entire study area takes 4234.56 seconds (approximately 1.18 hours), while the Aare catchment requires only 42.14 seconds. This disparity illustrates the exponential increase in computational demand with the size of the DEM.

Sequential calibration of the entire study area requires approximately 300 hours, whereas calibrating the Aare catchment takes only 3 hours. To reduce to acceptable runtime, a parallel calibration approach was implemented. For the entire study area, the parallel calibration was tested with a capacity of 2 simultaneous simulations, reducing the calibration time from 300 hours to 166 hours. This reduction, while significant, still indicates high computational demands due to the large area size and limited parallel capacity under the available hardware (21 GB of RAM).

In contrast, parallel calibration of the Aare catchment allowed for 6 simultaneous simulations, reducing significantly the calibration time from 3 hours to just 40 minutes. This demonstrates the substantial gains in computational efficiency achievable through parallel processing.

Overall, these results highlight the advantages of calibrating the Landscape Evolution Model (LEM) using a single catchment rather than the entire study area. However, the effectiveness of this method heavily depends on the computational resources available and the specific size of the catchment selected. As we pointed out in the previous section, it is essential to select the catchment area carefully, as this has a direct impact on the transferability and performance of calibrated parameters.

DEM	DEM Size [km²]	Single Run	Serie Calibration Time	Parallel Model Capacity	Parallel Calibration Time
Area	3000	4234.56 s	300 h	2 simulations	166h
Aare	555	42.14 s	3 h	6 simulations	40min

Table 9: Computational time evaluation for the calibration process, comparing sequential and parallel calibration duration for the entire study area and the Aare catchment at 50m resolution. The parallel capacity was tested using 21 GB of RAM.

4 Discussion

This study explored the calibration of Landscape Evolution Models (LEMs), focusing on the transferability of calibrated parameters across different spatial and temporal scales. By testing whether parameters calibrated on smaller catchments could be applied to larger areas and their runtime extended beyond the calibration period, we aimed to understand model generalizability.

4.1 Resolution Effects

An unexpected outcome emerged regarding model resolution: there was minimal variation in results across different resolutions, despite our expectation of significant degradation of accuracy as resolution decreased. Measures such as hydric loss, changes in Kling-Gupta Efficiency (KGE), and statistical metrics like mean and standard deviation in time series were negligible. The main differences were linked to station placement. For instance, Station 2200 is close to the junction of two streams, within 200 meters. An error in station placement at low resolutions could double the estimated catchment size.

This finding contrasts with previous studies that have highlighted the importance of spatial resolution in LEMs. Willgoose (2005) discussed how grid resolution can affect the representation of topographic features and processes in landscape evolution models. However, in our study, the lack of significant variation suggests that the model may not be sensitive to the scales tested or that the processes governing SSC are not adequately captured at these resolutions.

Wohl (2018) highlighted the complexity of channel initiation and headwater processes, which are sensitive to spatial resolution and may not be well-represented in coarser models. The poor performance of SSC prediction might explain this lack of sensitivity to resolution. Further investigation is required, particularly after improving the sediment transport predictions.

4.2 Parameter Sensitivity and Calibration Challenges

Our sensitivity analysis revealed that the stream power law parameters m and n significantly influence sediment transport predictions. Small changes of 0.1 in these parameters led to substantial variations in model outcomes. In some cases, the suspended sediment concentration (SSC) became almost constant; in others, excessive erosion caused the model to fail due to numerical errors like dividing by zero.

This high sensitivity aligns with findings from Harel et al. (2016), who highlighted the importance of accurately estimating m and n for reliable model predictions. They suggest fixing these parameters before proceeding with the calibration of others. The challenge in calibrating m and n underscores the complexity of geomorphological processes and the difficulty of capturing them accurately in models.

Interestingly, when the SSC remained constant, the mass flux still varied. This indicates that the concentration was directly linked to the discharge, as SSC is obtained by dividing the mass

flux by the discharge. The interdependence of parameters like m and n and their impact on model stability make calibration complex.

Throughout the study, we realized the difficulty of calibrating LEMs. Initially, we aimed to automate the process using machine learning and specialized algorithms. However, the sensitivity of parameters, particularly m and n, made the process complex. Slight variations significantly altered model behavior, changing the initial range of parameters K by factors of 10 or 100. Many simulations resulted in excessive erosion, causing the models to fail.

This experience mirrors the challenges discussed by Beven and Binley (1992), who highlighted the limitations of automatic calibration methods in complex, non-linear models. Similarly, Boyle et al. (2000) suggested combining manual and automatic methods to improve calibration outcomes, using expert judgment to navigate sensitive parameter spaces.

Time constraints and underestimation of calibration complexity led us to simplify the methodology. We adopted the Sobol sequence, which, while simplistic, offers transparency and allows for manual adjustments. The number of parameters and their interdependence made calibration challenging. With 12 total parameters, automatic calibration was only partially achieved. Balancing model complexity to capture essential processes while keeping it interpretable was difficult. Coulthard (2001) noted that increasing model complexity often comes at the cost of increased calibration difficulty and computational demand.

4.3 Model Performance

4.3.1 Discharge Predictions

The best-performing models showed consistent discharge simulations across different catchments. Large catchments like the Aare exhibited fewer problems, suggesting that key hydrological processes are being captured adequately. Duan et al. (2003) emphasized the importance of capturing essential hydrological dynamics for reliable model performance.

In contrast, the Lonza catchment showed a sharp decline in the KGE for discharge simulations, likely due to its large glacier coverage. Unlike precipitation-driven catchments, the Lonza's hydrology is more influenced by glacial melt, which is highly sensitive to temperature changes. Our model, using annual averages, failed to capture important seasonal temperature variations. Karger et al. (2023) emphasized the need for high-resolution climate data, including seasonal temperature fluctuations, when modeling this type of catchment. The absence of these fluctuations, combined with the annual time step, likely contributed to the reduced model performance. Glaciers act as reservoirs, releasing water in response to temperature changes, which our model could not fully reflect.

4.3.2 Sediment Transport Predictions

The models had difficulty representing seasonal processes and glacier-related dynamics. They struggled to capture extreme events and inter-annual variability, mainly due to limited model complexity and temporal resolution. The use of an annual time step and lack of seasonal data hindered accurate simulation, leading to discharge prediction errors that were even more pronounced for sediment transport dynamics.

Pazzaglia (2003) noted that modeling sediment transport accurately requires considering short-term events and seasonal variations. In our study, although the Lonza catchment had the best SSC prediction results, it performed poorly in simulating discharge. As the smallest catchment, the Lonza may be more vulnerable to errors in precipitation estimates and runoff representation,

especially since spatial precipitation inaccuracies would have a more noticeable impact. The high KGE(SSC) observed for the Lonza catchment could be explained by the buffering effect of glaciers, which smooths out extreme events in both discharge and SSC. In other catchments, the failure to represent extreme events led to poor SSC predictions, while the stability of the Lonza catchment helped the model perform better for SSC.

This suggests that the model's temporal resolution limitations have a greater impact on catchments where seasonal and extreme events are key drivers of hydrology and sediment transport.

4.4 Model Calibration and Transferability

4.4.1 Spatial Transferability

The calibration process revealed consistent performance in discharge simulations across different catchments. Key hydrological parameters exhibited a degree of spatial transferability. This finding suggests that models calibrated on smaller, less computationally intensive catchments can be effectively applied to larger and more complex areas without significant loss of accuracy. This aligns with previous studies by Hancock et al. (2011) and Willgoose et al. (1991), which emphasize the challenges and potentials in parameter transferability.

However, the limited success in transferring sediment transport parameters points to the complex and localized nature of geomorphic processes. This supports the findings of Davy and Lague (2009) and Tucker and Hancock (2010), who noted the sensitivity of sediment transport to catchment-specific factors. To confirm these finding, further testing of sediment transport spatial transferability will be necessary once within-catchment predictions are improved.

Parameter transfer was most successful for the Aare catchment. Aare is the largest catchment (555 km^2) with a mean elevation of 2135m, covering both high and low altitudes, making it the most representative of the entire area. In contrast, the Lonza catchment is smaller (196 km^2) with a higher mean elevation of 2937m and over 50% glacier cover. The Lutschine catchment, at 381 km^2 with a mean elevation of 2050m and 13% glacier cover, is closer to Aare in size but has lower altitudes and less glacier cover, making it less representative of the area.

This difference in representativeness was reflected in the results, where the Aare calibration performed better across all catchments except Lutschine. Lutschine struggled, particularly in simulating the Lonza station, where it barely produced positive KGE results for the discharge, likely due to having the lowest glacier cover. Aare's diverse characteristics make it a better candidate for parameter transfer, contributing to its overall success. This highlights the potential for spatial transferability, but emphasizes the importance of carefully selecting the catchment.

4.4.2 Temporal Transferability

We extended our simulations up to 1,000 years to evaluate the long-term robustness of the models. The hydrological model maintained consistent performance over this period, with only minor deviations appearing after the initial 400 years. The maximum deviation in discharge reached approximately 0.35,m^3/s, representing about a 10% difference compared to the reference model. This level of divergence is acceptable over such an extended period and suggests that the hydrological components of the model exhibit long-term stability. This finding aligns with broader literature, where LEMs effectively capture long-term hydrological trends (Whipple et al., 2000).

In contrast, the sediment transport model showed significant variability over the extended timescale. Substantial divergences in SSC predictions occurred early in the simulations. Mi-

nor parameter adjustments led to variations exceeding 50% compared to the reference model. This instability indicates that sediment transport processes are highly sensitive to parameter variations and may not be reliably predicted over long periods without more precise calibration. Since the model already struggles with SSC predictions in the short term, assessing its robustness over extended periods is challenging.

These findings support the conclusions of Simpson (2017) and Temme et al. (2017), who emphasize the inherent challenges in simulating sediment transport over extended periods due to the stochastic nature of sediment supply and climate variability. In our simulations, the necessity to repeatedly use a limited 49-year climatic dataset introduces an artificial cyclicity that complicates interpretation. This cyclicity fails to capture the gradual changes in temperature and precipitation expected over these timescales, such as those resulting from global warming and other natural climatic cycles.

Furthermore, the lack of correlation between discharge variability and SSC differences emphasizes the complexity of sediment dynamics. While discharge predictions remained consistent, SSC predictions did not display the same stability or direct relationship with discharge changes. This suggests that factors beyond hydrological inputs, such as sediment availability and channel morphology, might play significant roles in sediment transport over long timescales.

Improving model accuracy for sediment transport on shorter timescales is essential. Without reliable short-term predictions, extending the model to longer periods may not yield meaningful insights. Re-evaluating the model's performance after enhancing its short-term accuracy could determine if the approach is viable for long-term sediment transport predictions.

Additionally, the minimal changes observed in elevation after 1,000 years, a drop of only 0.2m in mean elevation, are unrealistic. Significant landscape evolution would be expected over such a timescale. This underestimation of erosion underscores the model's limitations in representing long-term geomorphic changes. It highlights the importance of incorporating processes that account for extreme events and climate variability, which drive much of the landscape evolution.

Overall, while the hydrological model components are relatively robust over time, the sediment transport model requires further refinement. Addressing model accuracy for sediment transport by incorporating more detailed climate variability, extreme event representation, and enhanced process simulations is necessary. Re-evaluating the transferability of sediment transport parameters with these improvements could determine if this approach becomes viable for long-term landscape evolution predictions.

4.5 Limitations

Our study has several limitations affecting the accuracy and applicability of the results.

First, while transferring hydrological parameters across catchments reduces computational effort, sediment transport parameters proved less transferable. This necessitates localized calibration to predict sediment dynamics accurately. Regions with diverse geomorphic conditions require detailed, catchment-specific calibration. As Pazzaglia (2003) emphasized, sediment transport is highly sensitive to local factors, and general parameters may miss critical details.

Second, the model overlooks important processes like landslides, floodplain erosion, and bank dynamics. These processes significantly contribute to sediment production, especially during extreme events not captured on an annual scale. Including such processes is crucial for improving the model's predictive capability, as highlighted by Temme et al. (2017).

Third, the coarse resolution of input data introduces uncertainties, especially in smaller catchments. Low-resolution models can misrepresent flow paths and catchment areas, leading to inaccuracies in discharge simulations. For example, errors in station placement caused by low resolution can result in incorrect catchment size estimates and discharge values. Adopting a different method for station placement or adjusting the station location at each time step could help mitigate these issues.

Fourth, the precipitation data we used have limitations. Interpolated precipitation can be highly uncertain, especially in alpine regions where precipitation varies greatly over short distances. In the Lonza catchment, underestimation of precipitation made it impossible to simulate the observed discharge accurately. Adjustments like increasing precipitation by 30% helped, but did not fully resolve the issue.

Fifth, the model assumes glaciers contribute negligibly to runoff, which may not be accurate in alpine settings with significant glacier coverage. This assumption affects the simulation of discharge and sediment transport, as glaciers can act as important water reservoirs.

Lastly, the limited number of available monitoring stations restricts our ability to calibrate the model effectively. More stations within catchments would provide better data to refine model calibration and validation. The lack of stations makes it difficult to capture the spatial variability of hydrological and sediment processes.

These limitations suggest that while our approach reduces computational demands, careful consideration of local conditions and data quality is essential. Future studies should aim for higher-resolution data, include additional geomorphic processes, and gather more extensive field measurements to improve model accuracy.

4.6 Broader Implications

The ability to transfer hydrological parameters across catchments reduces the computational burden of calibrating models for large or complex regions. This has significant implications for practical applications of LEMs in geomorphology and landscape management.

However, the limited transferability of sediment transport parameters suggests that localized calibration is necessary to predict sediment dynamics accurately. Regions with highly variable geomorphic conditions require more detailed, catchment-specific calibration. As Pazzaglia (2003) emphasized, sediment transport processes are highly sensitive to local factors, and generalized parameters may not capture critical nuances.

4.7 Recommendations for Future Research

Improving the sediment transport simulation is essential. Including additional data or more accurate datasets could enhance model performance. For instance, incorporating groundwater data, precise glacier coverage, and more accurate precipitation and temperature records could refine the model. High-resolution climate datasets like those described by Karger et al. (2023) could improve the representation of climatic inputs.

Adding processes like soil infiltration, overflow, glacial dynamics, and river incision calculations could better represent hydrological and geomorphic processes. However, these additions would increase model complexity and require careful consideration. Willgoose (2005) noted that while adding complexity can improve model realism, it also demands more detailed data and increases calibration challenges.

Moreover, increasing the number of observation stations per catchment would help capture the gradient of discharge and provide a better understanding of material transport across the landscape. Enhanced spatial data could reduce uncertainty and improve model calibration, as suggested by Temme et al. (2017).

5 Conclusion

This study aimed to improve the calibration of Landscape Evolution Models (LEMs) in alpine regions using data-driven methods. The use of in-situ observations and a comprehensive exploration of parameter space allowed for more robust calibration results. The findings demonstrate that calibrating LEMs for smaller catchments can significantly reduce computational demands while maintaining acceptable performance levels, particularly for discharge simulations.

Spatial transferability tests revealed that parameters calibrated on smaller catchments performed well when applied to larger areas. However, sediment transport predictions remained highly sensitive to local conditions, indicating that further refinement is needed for this aspect of the model. The study also highlighted the limitations of using a yearly timestep, which restricts the model's ability to capture seasonal variations and extreme events.

While the results show promise for enhancing LEM calibration, challenges such as hydrological extremes and accurately simulating sediment transport processes remain. Future work should focus on improving data inputs, such as incorporating glacier dynamics and increasing temporal resolution, to provide more accurate predictions over longer timescales.

In summary, this research contributes to the understanding of LEM calibration and parameter transferability, offering practical insights for future applications in landscape management and geomorphological studies. However, continued refinement of both model structure and data inputs is essential to fully realize the potential of LEMs. By addressing the limitations identified, we can move toward more robust and generalizable LEMs, as envisioned by van der Beek (2013).

References

Barnhart, K. R., Glade, R. C., Shobe, C. M., and Tucker, G. E. (2019). Terrainbento 1.0: A Python package for multi-model analysis in long-term drainage basin evolution. *Geoscientific Model Development*, 12(4):1267–1297.

Barnhart, K. R., Hutton, E. W., Tucker, G. E., Gasparini, N. M., Istanbulluoglu, E., Hobley, D. E., Lyons, N. J., Mouchene, M., Nudurupati, S. S., Adams, J. M., et al. (2020). Landlab v2. 0: A software package for Earth surface dynamics. *Earth Surface Dynamics*, 8(2):379–397.

Begert, M., Seiz, G., Schlegel, T., Musa, M., Baudraz, G., and Moesch, M. (2003). *Homogenisierung Vom Klimareihen Der Schweiz Und Bestimmung Der Normwerte 1961–1990. Schlussbericht Des Projekts NORM90*, volume 67. MeteoSwiss.

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3):279–298.

Boyle, D. P., Gupta, H. V., and Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12):3663–3674.

Coulthard, T. J. (2001). Landscape evolution models: A software review. *Hydrol. Process.*

Davy, P. and Lague, D. (2009). Fluvial erosion/transport equation of landscape evolution models revisited. *Journal of Geophysical Research: Earth Surface*, 114(F3).

Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R. (2003). *Calibration of Watershed Models*, volume 6. John Wiley & Sons.

GDAL/OGR contributors (2020). *GDAL/OGR Geospatial Data Abstraction Software Library.*

Gillies, S. et al. (2013). Rasterio: Geospatial raster I/O for Python programmers.

Hancock, G. R., Coulthard, T. J., Martinez, C., and Kalma, J. D. (2011). An evaluation of landscape evolution models to simulate decadal and centennial scale soil erosion in grassland catchments. *Journal of Hydrology*, 398(3):171–183.

Harel, M. A., Mudd, S. M., and Attal, M. (2016). Global analysis of the stream power law parameters based on worldwide 10Be denudation rates. *Geomorphology*, 268:184–196.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hobley, D. E. J., Adams, J. M., Nudurupati, S. S., Hutton, E. W. H., Gasparini, N. M., Istanbulluoglu, E., and Tucker, G. E. (2017). Creative computing with Landlab: An open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics. *Earth Surface Dynamics*, 5(1):21–46.

Hoyer, S. and Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1).

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Hutton, E., Barnhart, K., Hobley, D., Tucker, G., Nudurupati, S., Adams, J., Gasparini, N., Shobe, C., Strauch, R., Knuth, J., Mouchene, M., Lyons, N., Litwin, D., Glade, R., Giuseppecipolla95, Manaster, A., Abby, L., Thyng, K., and Rengers, F. (2020). Landlab.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G. (2014). The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34(5):1657–1675.

Jones, E., Oliphant, T., Peterson, P., et al. (2001/). SciPy: Open source scientific tools for Python.

Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., Conrad, O., Zimmermann, N. E., and Frieler, K. (2023). CHELSA-W5E5: Daily 1 km meteorological forcing data for climate impact studies. *Earth System Science Data*, 15(6):2445–2464.

Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K., and Bierkens, M. F. P. (2010). A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software*, 25(4):489–502.

MeteoSwiss (2019). EURO4M-APGD.

O'Callaghan, J. F. and Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 28(3):323–344.

OFAG (2020). Carte des aptitudes des sols de la Suisse. https://www.blw.admin.ch.

pandas development team, T. (2020). Pandas-dev/pandas: Pandas. Zenodo.

Pazzaglia, F. J. (2003). Landscape evolution models. In *Developments in Quaternary Sciences*, volume 1 of *The Quaternary Period in the United States*, pages 247–274. Elsevier.

Reinhardt, L., Jerolmack, D., Cardinale, B. J., Vanacker, V., and Wright, J. (2010). Dynamic interactions of life and its landscape: Feedbacks at the interface of geomorphology and ecology. *Earth Surface Processes and Landforms*, 35(1):78–101.

Schwarb, M. (2000). *The Alpine Precipitation Climate: Evaluation of a High-Resolution Analysis Scheme Using Comprehensive Rain-Gauge Data*. PhD thesis, ETH Zurich.

Simpson, G. (2017). *Practical Finite Element Modeling in Earth Science Using Matlab*. John Wiley & Sons, Ltd, Chichester, UK.

SwissTopo (2024). swissALTI3D.

Temme, A., Armitage, J., Attal, M., van Gorp, W., Coulthard, T., and Schoorl, J. (2017). Developing, choosing and using landscape evolution models to inform field-based landscape reconstruction studies. *Earth Surface Processes and Landforms*, 42(13):2167–2183.

Tucker, G., Lancaster, S., Gasparini, N., and Bras, R. (2001). The Channel-Hillslope Integrated Landscape Development Model (CHILD). In Harmon, R. S. and Doe, W. W., editors, *Landscape Erosion and Evolution Modeling*, pages 349–388. Springer US, Boston, MA.

Tucker, G. E. and Hancock, G. R. (2010). Modelling landscape evolution. *Earth Surface Processes and Landforms*, 35(1):28–50.

van der Beek, P. (2013). Modelling Landscape Evolution. In *Environmental Modelling*, chapter 19, pages 309–331. John Wiley & Sons, Ltd.

van der Beek, P. and Bishop, P. (2003). Cenozoic river profile development in the Upper Lachlan catchment (SE Australia) as a test of quantitative fluvial incision models. *Journal of Geophysical Research: Solid Earth*, 108(B6).

Whipple, K. X., Hancock, G. S., and Anderson, R. S. (2000). River incision into bedrock: Mechanics and relative efficacy of plucking, abrasion, and cavitation. *Geological Society of America Bulletin*, 112(3):490–503.

Willgoose, G. (2005). Mathematical Modeling of Whole Landscape Evolution. *Annual Review of Earth and Planetary Sciences*, 33(Volume 33, 2005):443–459.

Willgoose, G., Bras, R. L., and Rodriguez-Iturbe, I. (1991). A coupled channel network growth and hillslope evolution model: 1. Theory. *Water Resources Research*, 27(7):1671–1684.

Wohl, E. (2018). The challenges of channel heads. *Earth-Science Reviews*, 185:649–664.

Appendix

Calibration Results

Hydrological

model_name	kge_discharge	runoff_ratio_ref	slope_power	temp_power
model_05	0.61824	0.94104	0.03099	0.40138
model_53	0.61781	0.92572	0.10463	0.19547
model_99	0.61665	0.71000	0.02696	0.12769
model_39	0.61648	0.80261	0.03469	0.25469
model_89	0.61442	0.89421	0.08328	0.24683
model_113	0.61342	0.86882	0.11203	0.03135
model_71	0.61050	0.79456	0.08722	0.11160
model_37	0.60717	0.94561	0.19290	0.01741
model_95	0.60581	0.84334	0.01164	0.43921
model_45	0.60094	0.96586	0.05966	0.46705

Table 10: Aare Catchment - 10 best models for hydrological calibration, from 128 simulations

model_name	kge_discharge	runoff_ratio_ref	slope_power	temp_power
model_215	0.15416	0.95948	0.10473	0.33381
model_99	0.15382	0.89026	0.11951	0.23650
model_03	0.15375	0.89629	0.15574	0.13092
model_183	0.15363	0.97158	0.17007	0.22013
model_125	0.15349	0.75339	0.10379	0.10643
model_111	0.15230	0.95032	0.18493	0.18442
model_71	0.15170	0.98231	0.16183	0.27125
model_147	0.15124	0.90106	0.11129	0.19336
model_201	0.15031	0.82444	0.12046	0.19592
model_15	0.15006	0.94274	0.13991	0.28956

Table 11: Lonza Catchment - 10 best models for hydrological calibration, from 256 simulations

model_name	kge_discharge	runoff_ratio_ref	slope_power	temp_power
model_35	0.61057	0.94662	0.19862	0.10700
model_47	0.60504	0.85630	0.12002	0.25428
model_93	0.60432	0.77812	0.12691	0.11708
model_11	0.60317	0.96589	0.16507	0.26748
model_87	0.60229	0.89142	0.10616	0.35648
model_67	0.60045	0.93635	0.14440	0.30236
model_117	0.59392	0.83591	0.11174	0.28473
model_49	0.59009	0.73361	0.10237	0.16614
model_83	0.58983	0.92489	0.20006	0.14623
model_59	0.58613	0.97374	0.13015	0.43242

Table 12: Lütschine Catchment - 10 best models for hydrological calibration, from 128 simulations

Stream Power Law

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_32	0.61834	-0.24839	9.84425E-10	7.07566E-10	9.72029E-10	3.96539E-09	0.51527	4.65219	0.47121
model_69	0.61833	-0.26796	3.27767E-09	3.35832E-09	1.41188E-09	3.04318E-09	0.55498	3.12706	0.70942
model_45	0.61827	-0.28091	3.64555E-09	2.23900E-09	3.39784E-09	3.56556E-09	0.58728	8.01712	0.77727
model_20	0.61832	-0.28199	1.49165E-09	1.36891E-09	1.83015E-09	1.55802E-09	0.48857	3.60711	0.38107
model_120	0.61831	-0.28426	4.60052E-10	1.18187E-09	1.22600E-09	3.82150E-09	0.59870	8.89239	0.10641
model_62	0.61834	-0.28998	2.65783E-09	8.04853E-10	1.90465E-09	3.01359E-09	0.60980	1.82423	0.68169
model_92	0.61832	-0.29089	1.77337E-09	1.56648E-09	1.51331E-09	2.72343E-09	0.42626	3.96314	0.57039
model_03	0.61829	-0.29569	1.57498E-09	2.97398E-09	1.70723E-09	2.20575E-09	0.56603	5.53303	0.48490
model_118	0.61837	-0.29635	2.31518E-09	5.46336E-10	1.48831E-09	1.26929E-09	0.67349	0.99087	0.46661
model_52	0.61821	-0.30113	1.40975E-09	9.80028E-10	3.30879E-09	2.27082E-09	0.37140	8.79667	0.54140

Table 13: Aare Model 5 - 10 best models for stream power calibration, from 128 simulation

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_57	0.61687	-0.27069	4.84482E-10	7.05063E-10	8.84429E-10	3.78585E-09	0.46165	3.87086	0.35951
model_92	0.61683	-0.27187	3.52176E-09	2.55220E-09	1.17797E-09	3.32637E-09	0.67217	3.44295	0.73156
model_45	0.61671	-0.27281	1.73581E-09	9.24983E-10	3.68031E-09	2.92912E-09	0.40604	5.51879	0.64419
model_52	0.61670	-0.28459	3.41707E-09	3.24133E-09	3.54507E-09	3.67893E-09	0.63396	6.93201	0.76093
model_97	0.61682	-0.28506	9.27363E-10	1.67743E-09	1.39238E-09	3.95225E-09	0.64278	5.43558	0.17814
model_39	0.61690	-0.28514	2.25008E-09	7.52172E-10	1.82530E-09	2.31670E-09	0.64882	1.34342	0.52605
model_81	0.61682	-0.28696	5.89107E-10	1.21891E-09	2.40971E-09	3.15960E-09	0.51098	5.08121	0.33258
model_26	0.61672	-0.28919	1.50101E-09	3.72119E-09	1.61338E-09	3.00334E-09	0.66589	8.75079	0.38535
model_13	0.61684	-0.29541	1.56471E-09	1.94381E-09	1.88880E-09	1.86545E-09	0.45538	3.22259	0.44707
model_19	0.61690	-0.29811	3.82075E-09	1.87475E-09	9.43048E-10	3.98150E-10	0.62100	0.77549	0.66351

Table 14: Aare Model 99 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_53	0.15427	0.17962	1.35895E-09	2.35227E-08	2.67081E-08	9.56509E-09	0.45135	8.11187	0.78088
model_197	0.15428	0.13815	4.93449E-09	2.44745E-08	3.06668E-08	2.83982E-08	0.43214	1.82944	0.88153
model_18	0.15426	0.13143	5.11181E-09	8.68420E-10	8.32739E-09	3.33192E-08	0.44546	2.52126	0.56765
model_147	0.15427	0.13100	2.09085E-08	2.55668E-08	2.95788E-08	5.45771E-09	0.55479	4.24063	0.87933
model_77	0.15426	0.13080	6.84368E-09	2.40037E-08	8.64131E-10	3.60953E-08	0.64415	9.27641	0.31252
model_117	0.15426	0.12359	2.17275E-09	3.87768E-08	8.18658E-09	2.59793E-08	0.34559	5.55039	0.40786
model_226	0.15426	0.10749	8.70184E-09	2.40534E-09	1.41859E-08	1.44533E-08	0.42097	6.74084	0.66976
model_230	0.15437	0.10480	1.79856E-08	1.60626E-08	2.13564E-08	3.70298E-08	0.66754	5.10633	0.86205
model_251	0.15400	0.09552	2.97151E-08	3.45242E-08	1.35798E-08	3.74568E-09	0.58028	4.92438	0.83578
model_28	0.15415	0.09151	2.27907E-08	1.90131E-08	6.64283E-09	1.17663E-08	0.37664	8.94412	0.85491

Table 15: Lonza Model 35 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_103	0.15342	0.17330	2.70869E-08	3.94828E-08	6.88631E-09	9.18998E-09	0.63242	8.00992	0.78235
model_137	0.15348	0.16798	5.48615E-10	3.92433E-08	2.73693E-09	3.05477E-08	0.36738	0.29367	0.39869
model_54	0.15348	0.16714	1.01522E-09	4.61550E-09	1.91630E-09	2.83520E-08	0.44062	8.23406	0.13102
model_190	0.15350	0.15890	1.00406E-08	4.11029E-09	2.36815E-08	3.11657E-08	0.67773	3.11962	0.75004
model_182	0.15348	0.15636	1.04575E-09	1.06457E-08	1.15434E-08	2.29694E-08	0.29978	1.61264	0.63576
model_151	0.15348	0.14387	2.82375E-08	3.85799E-08	1.91701E-08	2.38349E-08	0.59116	3.62719	0.87649
model_109	0.15348	0.13797	1.11516E-08	3.55493E-08	2.94362E-08	4.60621E-09	0.35906	0.70484	0.88165
model_78	0.15350	0.13517	6.08502E-09	3.21224E-09	3.10557E-08	1.58877E-08	0.54858	8.52134	0.85668
model_138	0.15358	0.13391	1.69302E-08	1.17605E-08	3.90016E-08	1.54286E-08	0.69080	3.97238	0.84469
model_25	0.15348	0.11683	3.29280E-09	3.24541E-08	5.86902E-09	2.41418E-09	0.17239	4.53491	0.18604

Table 16: Lonza Model 125 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_11	0.10755	-0.30059	1.58671E-09	3.02050E-09	4.05387E-09	2.16033E-09	0.25632	7.84809	0.41424
model_45	0.10747	-0.30696	4.95859E-09	6.13238E-10	6.45232E-09	2.98711E-09	0.51967	8.11562	0.71768
model_33	0.10737	-0.33304	5.51796E-09	1.17259E-09	4.10203E-09	4.07181E-09	0.17236	6.26558	0.76099
model_32	0.64836	-0.44992	3.04560E-09	5.86931E-09	7.28621E-10	3.22911E-09	0.59746	1.31109	0.12370
model_26	0.64838	-0.45653	3.67075E-09	6.74297E-09	2.17252E-09	3.08074E-09	0.65577	6.04912	0.42746
model_24	0.64836	-0.46069	2.02379E-09	4.51522E-09	1.21438E-09	1.90566E-10	0.22475	3.73887	0.21161
model_120	0.64837	-0.46462	2.06487E-09	3.87621E-09	2.85248E-09	3.38013E-09	0.49545	9.33568	0.37578
model_55	0.64837	-0.46595	2.41066E-09	2.73792E-09	1.40240E-10	4.21551E-09	0.47363	7.12670	0.29845
model_20	0.64836	-0.47048	1.66679E-09	4.17144E-09	2.43286E-09	6.86835E-09	0.57751	1.22856	0.26752
model_116	0.64837	-0.47068	1.54611E-09	5.08250E-09	7.95282E-10	3.60256E-09	0.14275	5.02544	0.34399

Table 17: Lütschine Model 35 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_25	0.14962	-0.35586	3.15576E-09	6.87954E-10	1.61964E-09	4.05889E-10	0.25454	6.14816	0.23349
model_77	0.65356	-0.45042	2.86589E-09	1.87911E-09	1.73992E-09	2.75097E-09	0.53228	4.25046	0.34663
model_106	0.65355	-0.45470	2.50704E-09	2.69490E-09	2.68316E-09	8.09067E-10	0.54487	5.75331	0.10105
model_23	0.65355	-0.46043	1.67320E-09	2.03001E-09	1.96586E-09	3.61232E-09	0.33161	3.91433	0.36590
model_21	0.65355	-0.46097	2.49534E-09	3.46259E-10	5.37948E-10	2.55068E-09	0.41341	9.48402	0.26203
model_79	0.65355	-0.46153	1.31791E-09	5.00998E-10	3.03761E-10	3.63932E-09	0.12270	8.67284	0.22545
model_54	0.65356	-0.46154	3.31598E-09	2.11378E-09	1.22692E-09	3.68455E-09	0.59670	1.75008	0.28360
model_17	0.65357	-0.46341	3.88716E-09	1.48172E-09	2.23008E-09	1.74170E-09	0.61205	7.29902	0.46404
model_48	0.65354	-0.46428	1.26985E-09	2.54019E-09	2.90794E-09	1.65422E-09	0.34090	5.35536	0.18644
model_37	0.65355	-0.46704	2.25807E-09	8.03582E-10	3.20970E-09	2.18644E-09	0.57087	8.56499	0.35450

Table 18: Lütschine Model 93 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_59	0.14966	-0.30382	8.79937E-10	8.52418E-10	2.72774E-10	2.96181E-10	0.66221	8.45008	0.31249
model_00	0.14962	-0.30909	6.80882E-10	5.23844E-10	6.98468E-10	5.55066E-10	0.33130	8.60722	0.55480
model_15	0.65356	-0.44731	7.60749E-10	6.15476E-10	5.45669E-10	8.08137E-10	0.67856	7.43472	0.49147
model_46	0.65355	-0.45081	3.77615E-10	4.31093E-10	3.54325E-10	7.90552E-10	0.28203	6.40959	0.29516
model_53	0.65354	-0.45975	3.14324E-10	6.69880E-10	2.35301E-10	8.48617E-10	0.56849	2.41997	0.16860
model_24	0.65355	-0.46512	5.68051E-10	3.72861E-10	4.39610E-10	4.75279E-10	0.38185	1.75275	0.33299
model_47	0.65356	-0.46624	7.22490E-10	7.75465E-10	5.91973E-10	4.91299E-10	0.60135	3.33094	0.51471
model_13	0.65354	-0.47238	1.77515E-10	7.95064E-10	1.91105E-10	6.38330E-10	0.08387	3.83601	0.27191
model_05	0.65354	-0.47274	2.38004E-10	6.04651E-10	5.96175E-10	9.77411E-10	0.51207	1.34058	0.37598
model_52	0.65358	-0.47623	7.85788E-10	3.11237E-10	9.08089E-10	5.49360E-10	0.31497	7.26111	0.64749

Table 19: Lütschine Model 35 m/n variation 1 - 10 best models for stream power calibration, from 128 simulations

model_name	kge_discharge	kge_ssc	K_sp_high	K_sp_medhigh	K_sp_med	K_sp_low	E_c	v_s	F_f
model_10	0.65355	-0.49547	1.64460E-08	8.70259E-09	3.07764E-09	2.35061E-08	0.47576	0.83106	0.39955
model_41	0.65353	-0.50136	1.45914E-09	2.24808E-08	1.58487E-08	3.74974E-08	0.68097	5.64001	0.35150
model_17	0.65355	-0.50318	8.84775E-09	3.06112E-08	1.39045E-08	2.07894E-08	0.26641	8.54335	0.43719
model_03	0.65362	-0.50758	3.39628E-08	3.68430E-08	1.56829E-08	1.81654E-08	0.51538	8.31593	0.76252
model_07	0.65357	-0.50896	2.06389E-08	2.63459E-08	2.62757E-08	2.24950E-08	0.65326	9.36110	0.69246
model_34	0.65356	-0.51389	1.16039E-08	6.13888E-09	1.33859E-08	1.95508E-08	0.23832	4.88474	0.56372
model_13	0.65357	-0.52025	1.46290E-08	2.97076E-08	1.78546E-08	2.80795E-08	0.40335	2.65658	0.50045
model_05	0.65363	-0.52447	1.80357E-08	3.46692E-08	3.61259E-08	3.89616E-08	0.22231	1.44727	0.85726
model_61	0.65356	-0.52510	1.06277E-08	2.27911E-08	3.41932E-08	2.28709E-08	0.55007	3.31793	0.73065
model_59	0.65381	-0.52800	3.65489E-08	2.62678E-08	1.48095E-08	3.94260E-09	0.09874	6.44588	0.84878

Table 20: Lütschine Model 35 m/n variation 2 - 10 best models for stream power calibration, from 128 simulations

Project Scripts

The scripts used in this project are available on GitHub for review and reproducibility. This repository includes the source code with four main script for the user to make use of:

1. config/config.yml: to set the base parameters and setting of the LEM.
2. data_processing_run.py: to download and process the DEM in .tif format and the precipitation data.
3. lem_run.py: to run multiple simulation of the LEMs in parallel.
4. data_analysis_run.py: to conduct the data analysis on the model's output and calibrate the right set of parameters and range of value.

You can access them in the following link:
https://github.com/barcazama/UU-MScThesis-landscape-calibration.

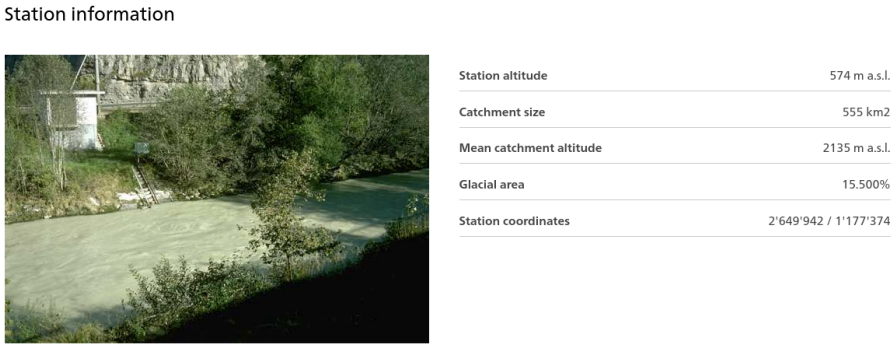
Relevant Library

- Pandas 2.2.2 (pandas development team, 2020)
- Rasterio 1.3.10 (Gillies et al., 2013)
- Matplotlib 3.8.4 (Hunter, 2007)

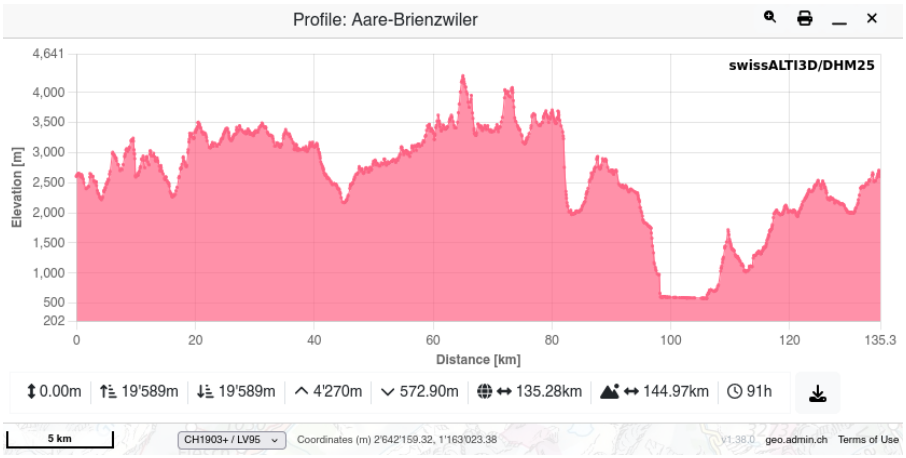
Hydrological Station - Location

Station 2019: Additional Information

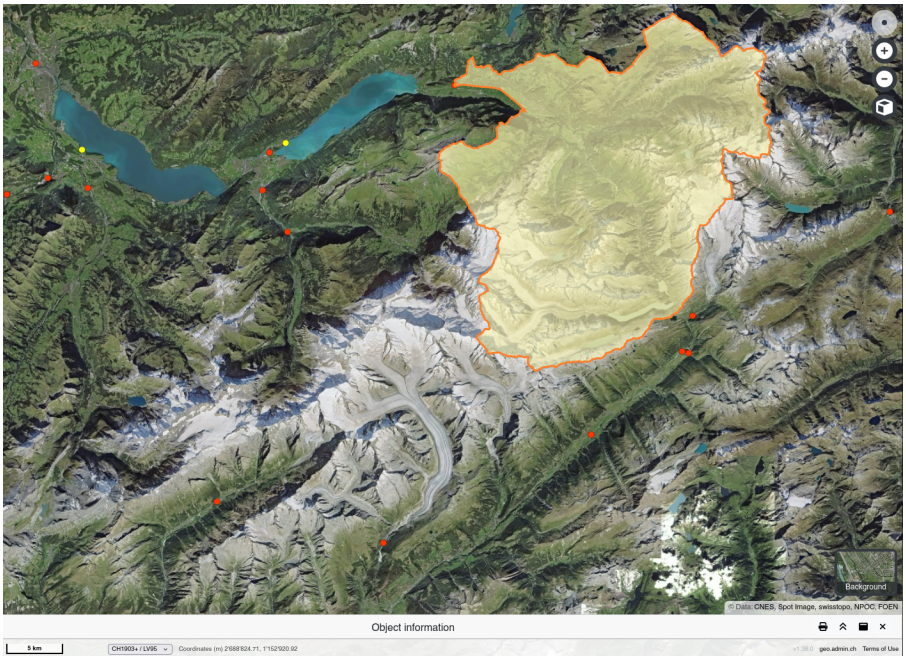
(a) Station Information



(b) Catchment Profile



(c) Catchment Map



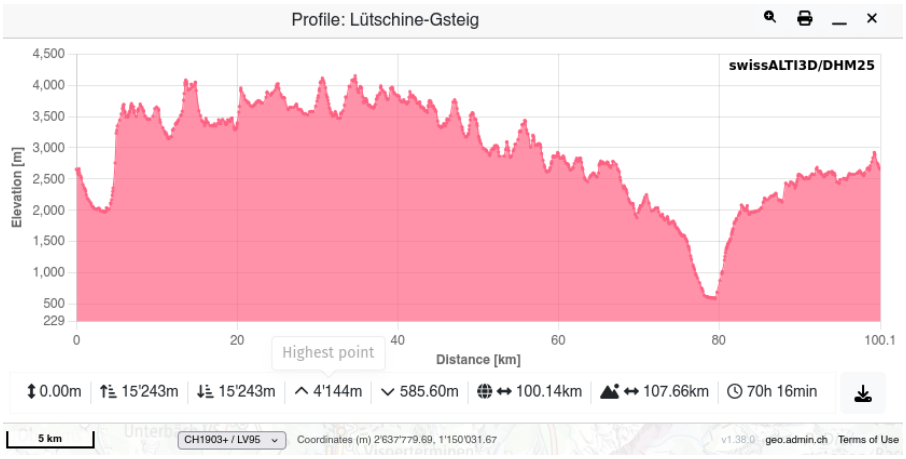
Station 2109: Additional Information

(a) Station Information

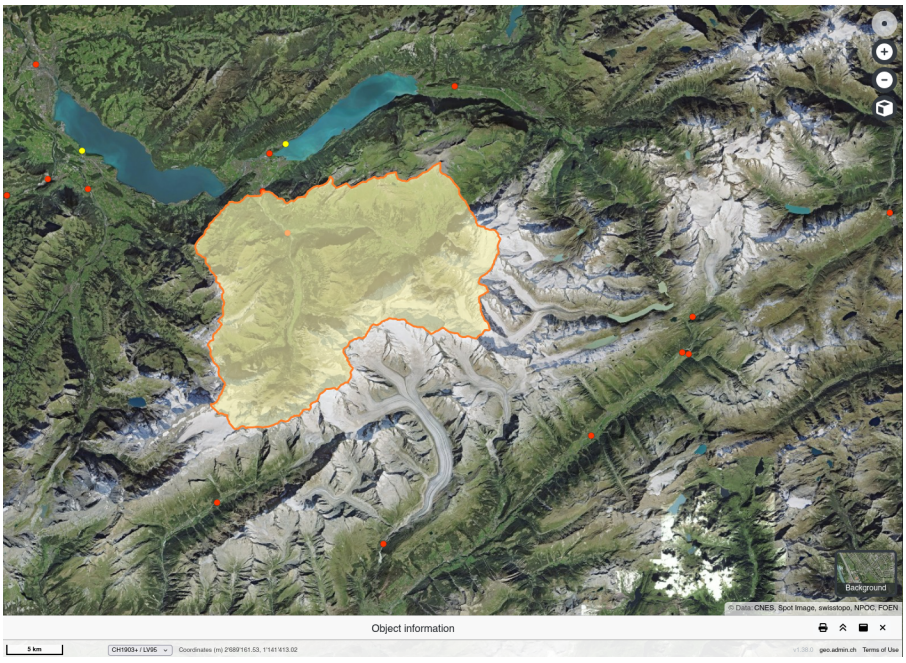
Station information

Station altitude	586 m a.s.l.
Catchment size	381 km2
Mean catchment altitude	2050 m a.s.l.
Glacial area	13.500%
Station coordinates	2'633'140 / 1'168'191

(b) Catchment Profile



(c) Catchment Map



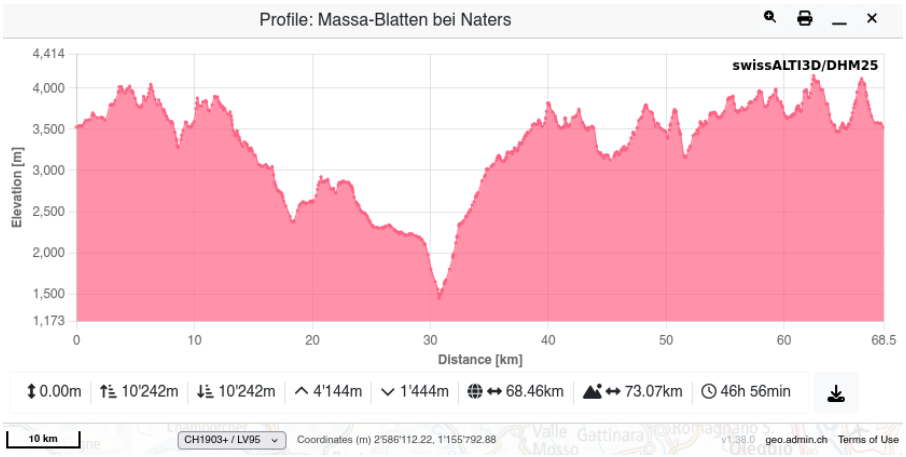
Station 2161: Additional Information

(a) Station Information

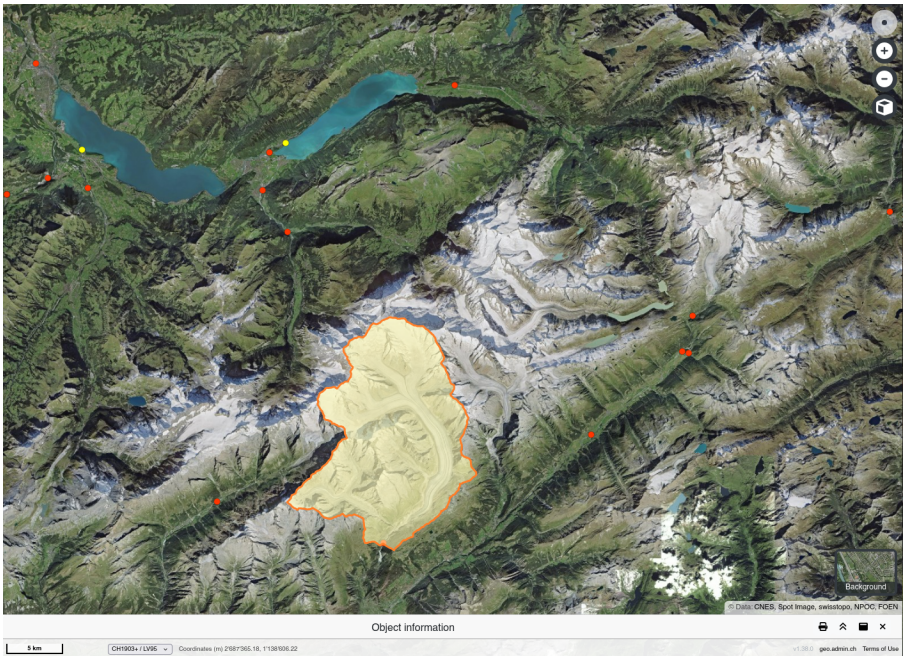
Station information

Station altitude	1446 m a.s.l.
Catchment size	196 km2
Mean catchment altitude	2937 m a.s.l.
Glacial area	56.500%
Station coordinates	2'643'694 / 1'137'291

(b) Catchment Profile

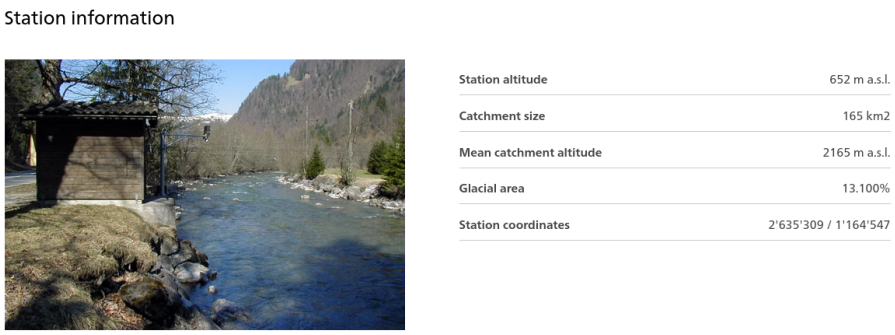


(c) Catchment Map

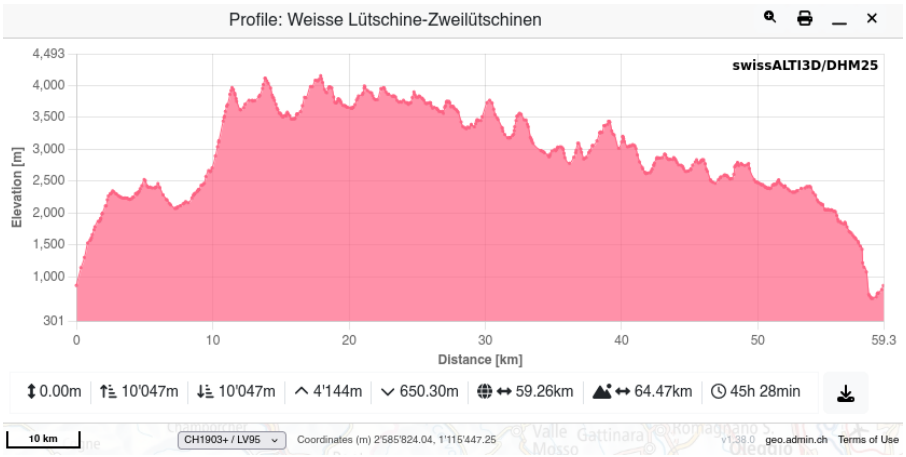


Station 2200: Additional Information

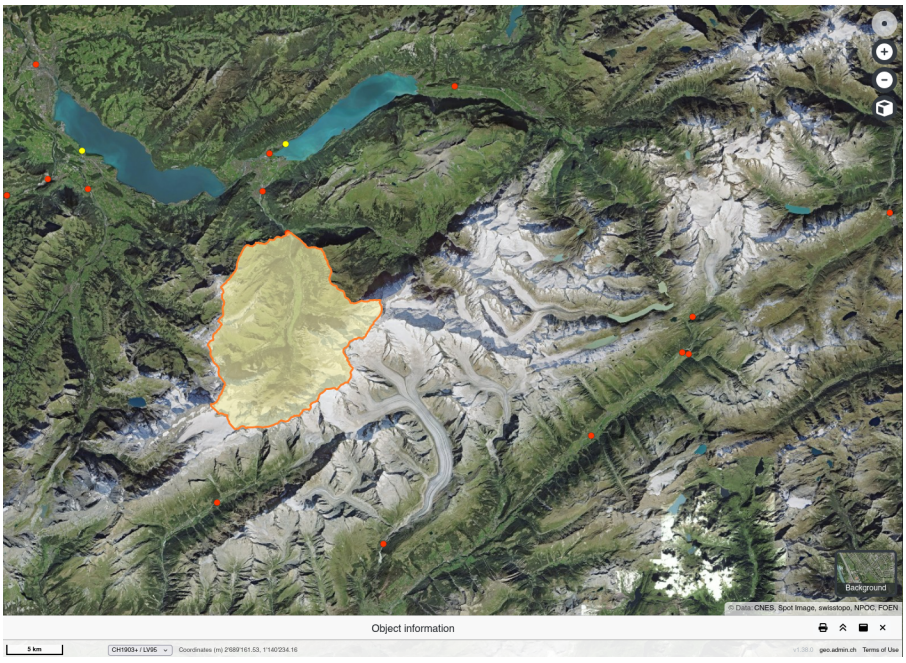
(a) Station Information



(b) Catchment Profile



(c) Catchment Map



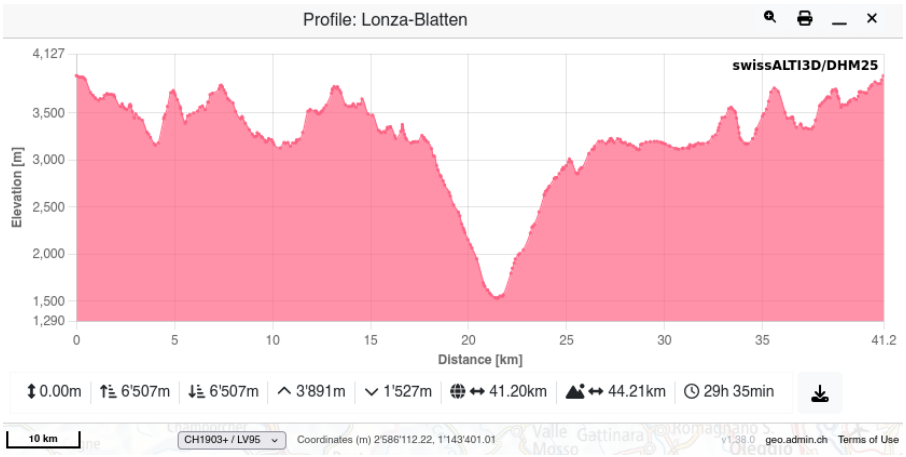
Station 2269: Additional Information

(a) Station Information

Station information

Station altitude	1523 m a.s.l.
Catchment size	77 km2
Mean catchment altitude	2624 m a.s.l.
Glacial area	24.700%
Station coordinates	2'629'128 / 1'140'919

(b) Catchment Profile



(c) Catchment Map