Supplement

Possible Geographical Predictors

The following geographical predictors were used in predictor selection on a 1-km grid basis:

- elevation
- exposure (based on a 3, 5, 11, 21 or 51 km smoothed topography)
- aspect
- slope
- latitude
- longitude
- distance to Alpine main ridge (assumed to be the highest grid cell, only in north-south direction)
- surface roughness, based on the CORINE dataset (European Environment Agency, 2019).

In the end, elevation and exposure on a 5 km and 51 km basis were selected based on an untuned random forest and a concurvity algorithm.

Predictor Maps

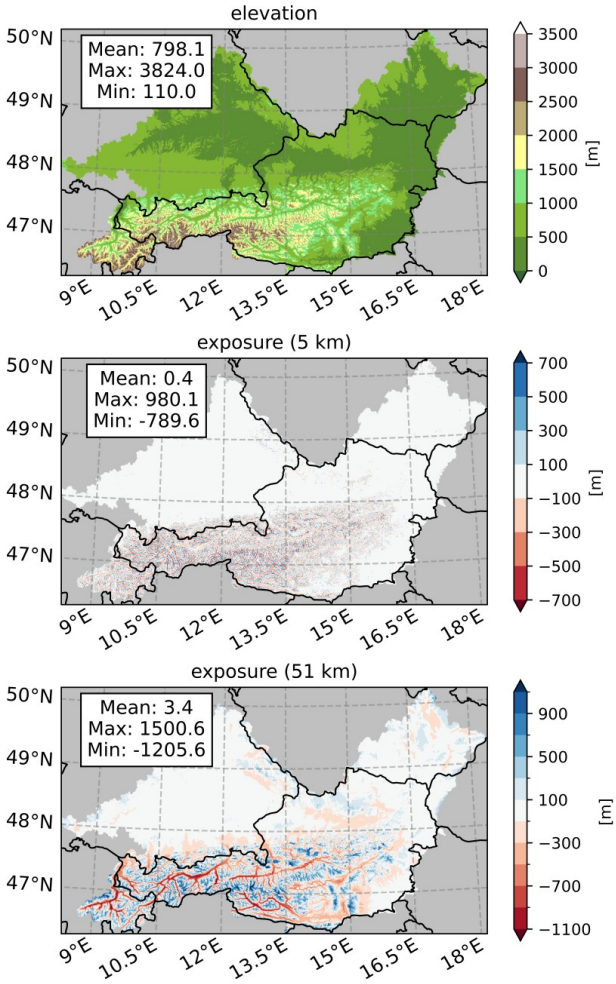

Figure S1. Selected predictors for GAM training, all color-coded and given in m. Mean, maximum and minimum levels are given in the text box.

GAM Smooth Functions

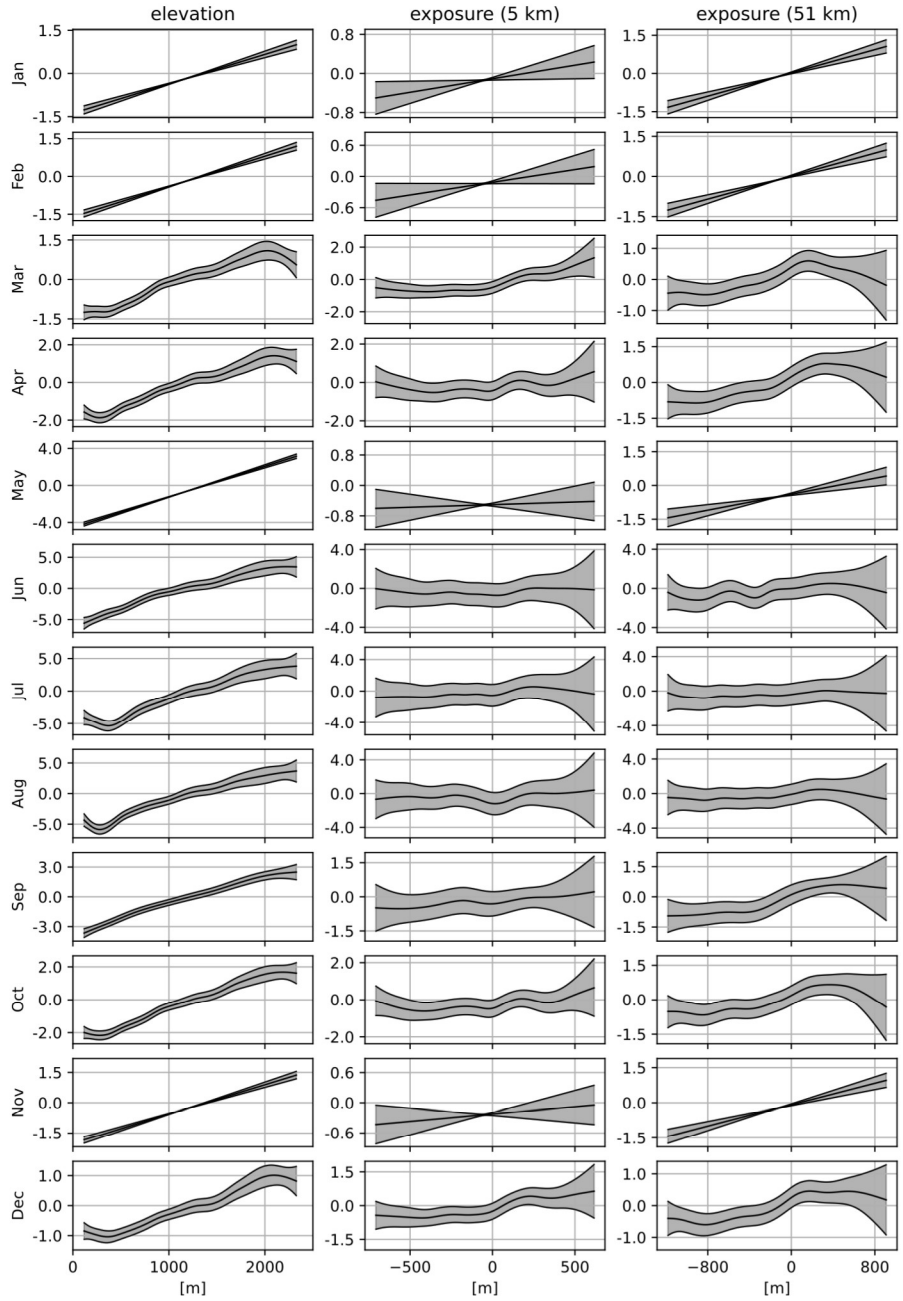

Figure S2. GAM smooth functions for every month (rows) and for the three chosen features, all in m. The shaded areas indicate the 95 % confidence interval.

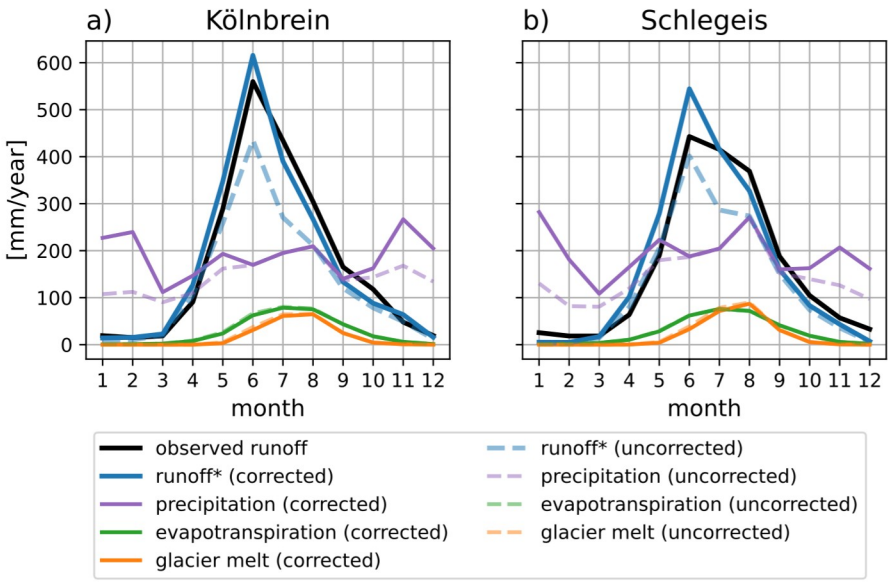

Figure S3. COSERO simulation results, as well as observed runoff, displayed as average monthly values for the high-alpine study regions a) Kölnbrein and b) Schlegeis. Solid lines refer to the simulation results using undercatch corrected data, while dashed lines refer to results using uncorrected precipitation input. *The simulated runoff is only considered for days with valid runoff observations.

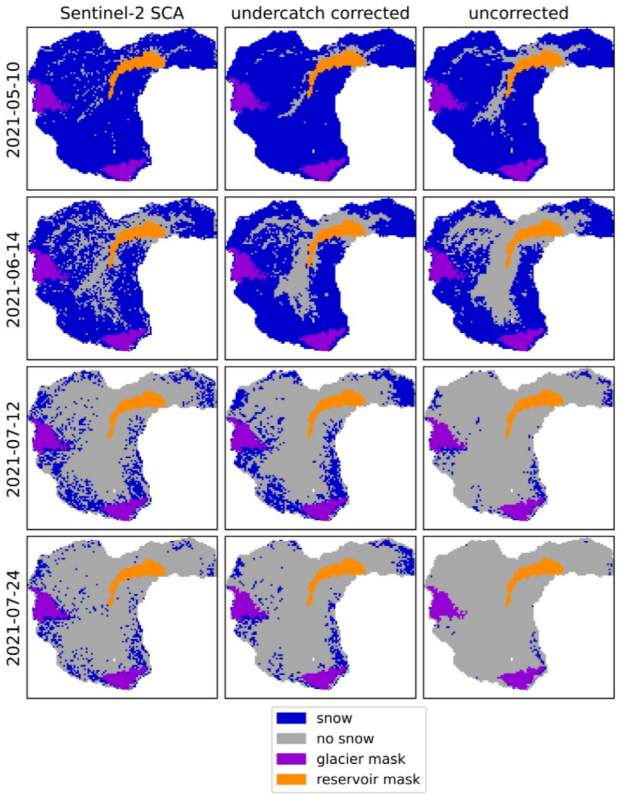

Figure S4. Comparison of SCA of Sentinel-2 data (first column) compared to the snow representation of the undercatch corrected (second column) and uncorrected (third column) Alpine3D-simulations for four dates (rows) in 2021 with satellite image coverage with cloud cover less than 5 %. Shown are grid cells with snow (blue), no snow (gray) and a mask for the glacier (violet) and reservoir area (orange).

Table 11. Performance metrics for the four snow-melt dates in 2021, obtained by comparing Alpine3D simulations with Sentinel-2 data on a grid cell basis. Accuracy is the correct attribution of true positives (TP) and true negatives (TN) in relation to the full dataset. Precision is the relation of TP to the sum of all positives (true and false). Recall is the relation of TP to the sum of TP and false negatives (FN). The numbers for positives and negatives refer to the number of grid cells in the catchment (excluding glacier and reservoir grid cells).

metric	undercatch corrected	uncorrected
accuracy	0.885	0.847
precision	0.869	0.945
recall	0.906	0.738
True positives (snow)	8652	7045
True negatives (no snow)	8242	9135
False positives (melt too late)	1304	411
False negatives (melt too early)	894	2501

RGI Glacier IDs in the Schlegeis catchment

Table 12. RGI Glacier IDs.

RGI ID	area [km^2]	catchment
RGI60-11.00476	2.0625	Schlegeis
RGI60-11.00434	1.3125	Schlegeis
RGI60-11.00459	2.25	Schlegeis
RGI60-11.00469	2.8125	Schlegeis
RGI60-11.00311	3.8125	Schlegeis
RGI60-11.00524	2.6875	Schlegeis
RGI60-11.00415	3.6875	Schlegeis

COSERO Calibration Performance Metrics

Figure S5 shows the Kling-Gupta Efficiency (KGE) of the Austrian-wide COSERO simulations, similar to Fig. 15. The KGE is defined as

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}\,, \tag{11}$$

where r is the Pearson correlation coefficient, $\alpha = \frac{\sigma_{\mathrm{sim}}}{\sigma_{\mathrm{obs}}}$ is the fraction of the simulated and observed variance, and $\beta = \frac{\overline{Q}_{\mathrm{sim}}}{\overline{Q}_{\mathrm{obs}}}$ is the mean runoff bias (Gupta et al., 2009).

Figure S6 shows the Nash–Sutcliffe Efficiency (NSE) of the Austrian-wide COSERO simulations, in the same structure as Fig. 15. The NSE is defined as

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}(Q_{\mathrm{obs}}^{t} - Q_{\mathrm{sim}}^{t})^2}{\sum_{t=1}^{T}(Q_{\mathrm{obs}}^{t} - \overline{Q}_{\mathrm{obs}})^2}\,, \tag{12}$$

where Q_{i}^{t} are the observed and simulated runoff, respectively, at time t and $\bar{Q}_{\mathrm{obs}}$ is the mean observed discharge (Nash and Sutcliffe, 1970).

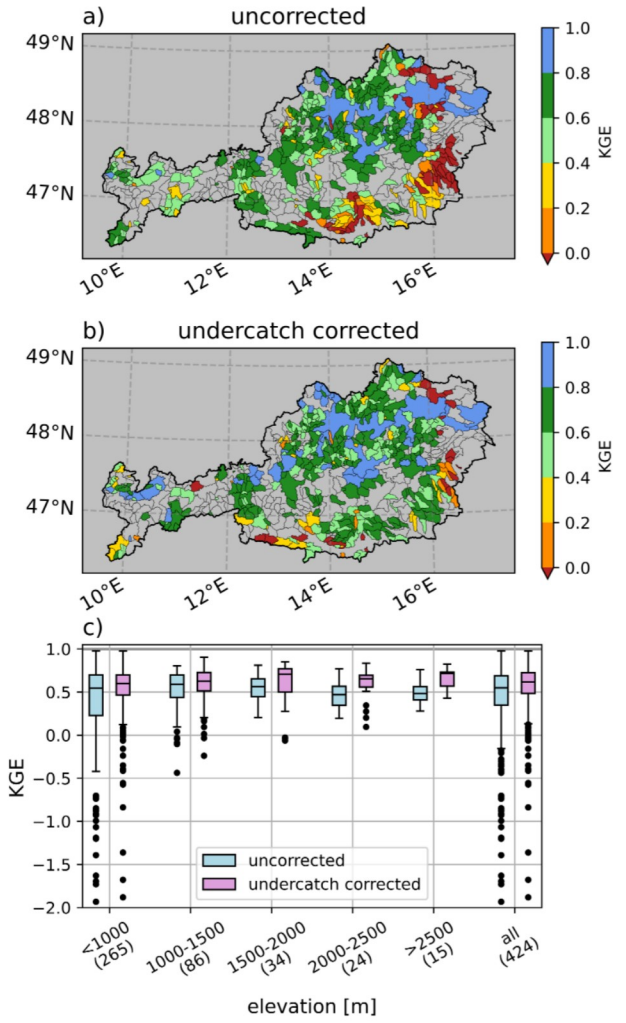

Figure S5. Kling-Gupta Efficiency (KGE) for the calibrated subcatchments of the Austrian-wide COSERO model, for simulations using a) uncorrected and b) undercatch corrected precipitation, as well as c) separated into elevation bands. The black outline in panel a) and b) shows the subcatchment borders. Numbers in parentheses in panel c) indicate the number of catchments within each elevation band. Whiskers represent 1.5 times the interquartile range, and outliers are shown as dots. The y-axis in panel c) is capped at -2 to improve legibility.

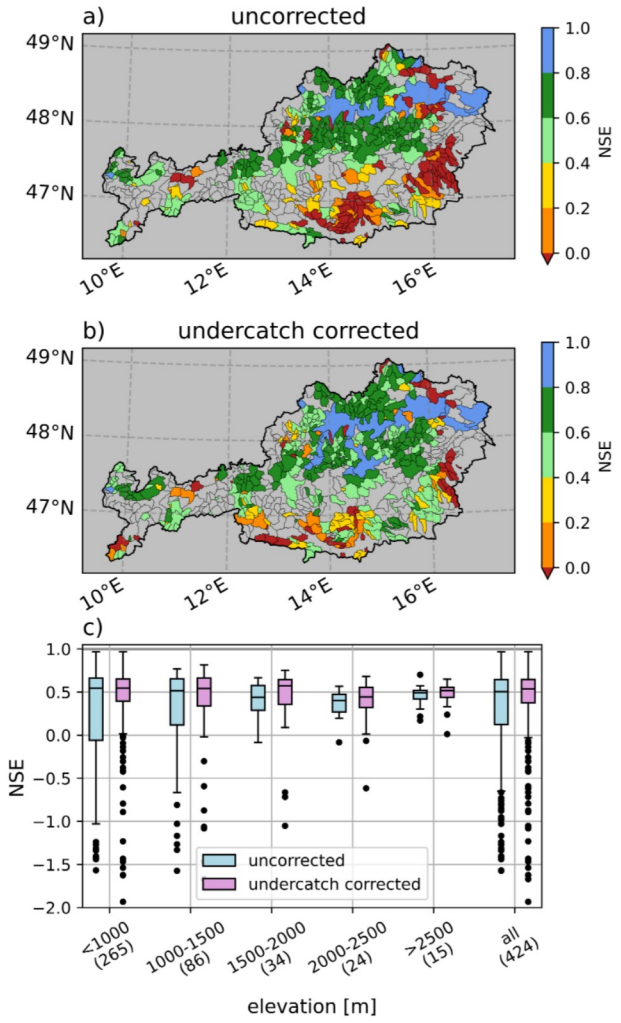

Figure S6. Nash-Sutcliffe Efficiency (NSE) for the calibrated subcatchments of the Austrian-wide COSERO model, for simulations using a) uncorrected and b) undercatch corrected precipitation, as well as c) separated into elevation bands. The black outline in panel a) and b) shows the subcatchment borders. Numbers in parentheses in panel c) indicate the number of catchments within each elevation band. Whiskers represent 1.5 times the interquartile range, and outliers are shown as dots. The y-axis in panel c) is capped at -2 to improve legibility.

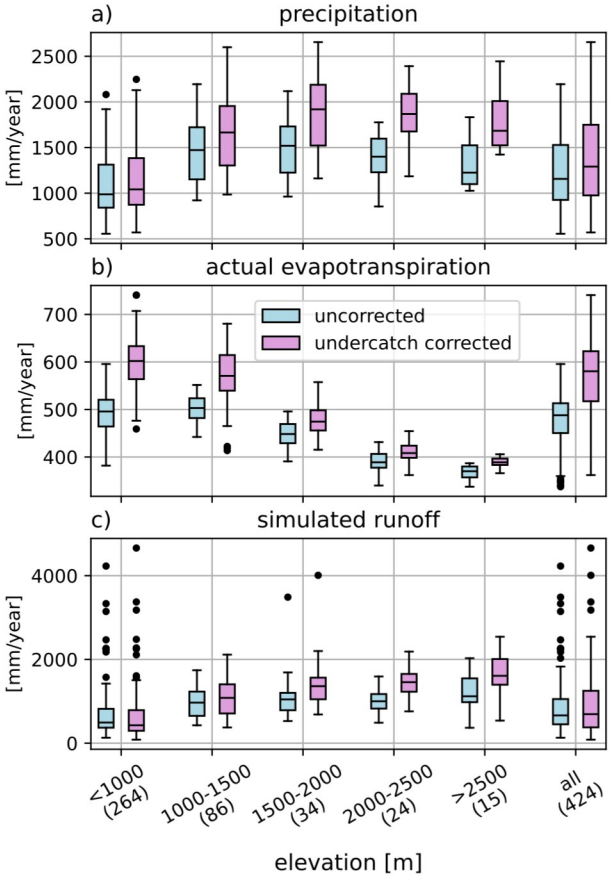

Figure S7. COSERO water balance components in [mm/year] of the calibrated subcatchments (n=424) in the Austrian-wide model. Displayed is the a) precipitation and b) actual evapotranspiration, separated into elevation bands. Numbers in parentheses in panel c) indicate the number of catchments within each elevation band. Whiskers represent 1.5 times the interquartile range, and outliers are shown as dots. Outliers are displayed as dots. As parts of some catchments (e.g. Upper Danube, Inn) are not explicitly modeled, runoff expressed in millimeters and normalized by the modeled area can exceed precipitation sums. The Danube runoff was excluded from panel c) for legibility.

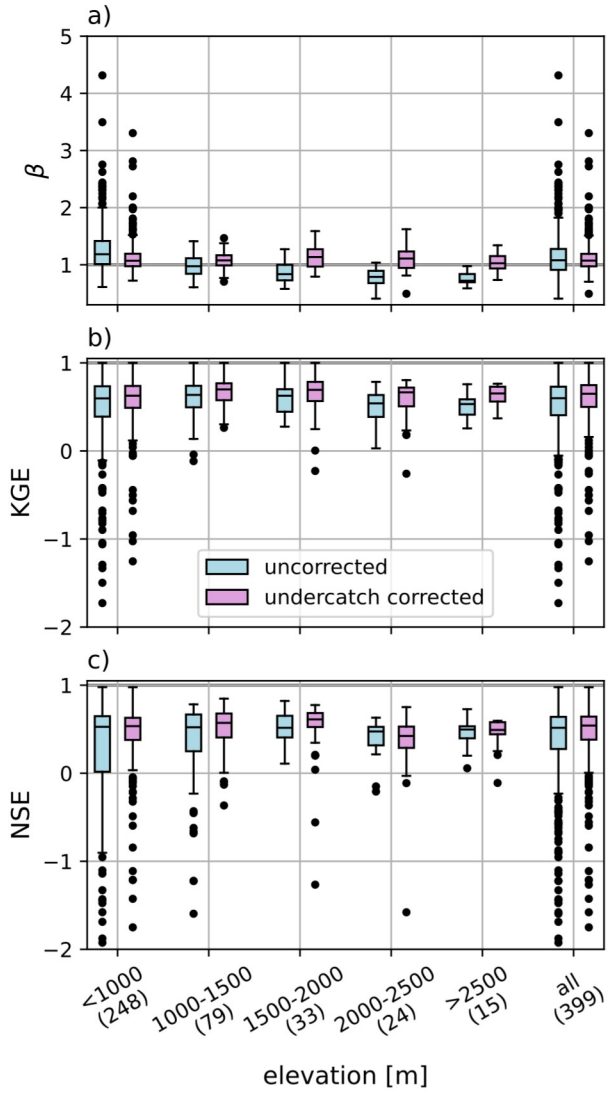

Figure S8. Performance metrics of calibrated subcatchments* the validation period (1980-2000) for the Austrian-wide COSERO model. Shown are a) the bias β, b) the Kling-Gupta Efficiency (KGE) and c) the Nash-Sutcliffe Efficiency (NSE), evaluated using uncorrected and undercatch-corrected precipitation input for different elevation bands. Numbers in parentheses indicate the number of catchments within each elevation band. Whiskers represent 1.5 times the interquartile range, and outliers are shown as dots. The y-axis in panels b) and c is capped at -2 to improve legibility. *For 25 catchments, runoff data are unavailable during the validation period; therefore, the total number of data points is lower than in the calibration analysis (399 vs. 424, respectively).