

Evaluation of global glacio-hydrological model
coupling for runoff prediction:
PCR-GLOBWB 2 and GloGEM

Pau Wiersma

February 11, 2021

Thesis committee: Dr. Markus Hrachowitz (Chair)
Dr. Ir. Rolf Hut
MSc Jerom Aerts
Dr. Harry Zekollari



Abstract

Global hydrological models (GHMs) have become an increasingly valuable tool in a range of global impact studies related to water resources. However, glacier parameterization is often overly simplistic or non-existent in GHMs. The representation of glacier dynamics and evolution, including related products such as glacier runoff, can be improved by relying on dedicated global glacier models (GGMs). In this study we test the hypothesis that coupling a GGM to a GHM can lead to increased GHM predictive skills and decreased GHM uncertainty through better glacier parameterization. To this end, the GGM GloGEM is coupled with the GHM PCR-GLOBWB 2 within the eWaterCycle II framework. For the years 2001-2012, the coupled model is evaluated against the uncoupled benchmark in 25 large ($>50,000 \text{ km}^2$) glacierized basins. Across all basins, the coupled model produces higher runoff throughout the melt season. In July and August, it ranges between 100.07% and 352% of the mean monthly benchmark runoff in lowly and highly glaciated basins respectively. The difference can primarily be explained by the inability of PCR-GLOBWB 2 to simulate snow redistribution and glacier retreat, causing an underestimation of glacier runoff. The coupled model better reproduces basin runoff observations primarily in highly glaciated basins, i.e. where the coupling has the most impact. This study underlines the importance of glacier representation in GHMs and demonstrates the potential of coupling a GHM with a GGM for better glacier representation and runoff predictions in glaciated basins.

1 Introduction

1.9 billion people worldwide rely on glacial meltwater as part of their water resources (Immerzeel et al., 2020). Particularly in regions prone to drought, glaciers can act as a crucial multiannual buffer (Pritchard, 2019, Biemans et al., 2019). Huss (2011) showed that in a dry year, even a glacier area fraction of as low as 0.06% can contribute up to 9% of the total basin discharge. However, with the rapid retreat of glaciers as a consequence of climate change, their contribution to runoff is expected to change in the future. On the intra-annual scale, peak runoff will occur earlier in summer, while on the interannual scale glacier runoff will increase until a peak is reached due to increased melt, after which a steady decline is expected due to the reducing glacier area (Jansson et al., 2003, Huss and Hock, 2018). In many basins throughout the world, the peak water already lies in the past (Huss and Hock, 2018) and the shift from a glacial to a nival-pluvial regime is underway. This will likely not only impact the water supply of millions of people, but also lead to increased natural hazards, hydro-political tension (Immerzeel et al., 2020) and instability of many glacier-influenced ecosystems (Cauvy-Fraunié and Dangles, 2019).

To account for the contribution of glaciers to runoff, many hydrological models in glacierized catchments include glacier parameterization schemes to form glacio-hydrological models (van Tiel et al., 2020). These models have been applied both at a local (e.g. Huss et al. (2008), Ragettli et al. (2016)) as well as at a regional scale (e.g. Farinotti et al. (2012), Frans et al. (2018)). Another approach involves the use of glacier geometry evolution estimates of an independent glacier model as forcing to a hydrological model, which has likewise been applied on a local (Laurent et al., 2020) and regional (Brunner et al., 2019) scale.

On a global scale, however, the integration of glacier processes in hydrological modeling is still lacking. While global hydrological models (GHMs) have gained relevance in recent years and have been used to study many different global issues including water scarcity (Wada et al., 2013, Haddeland et al., 2014, Schewe et al., 2014), drought severity (Van Huijgevoort et al., 2014), flood hazard (Dankers et al., 2014) and groundwater depletion (Wada et al., 2010, Gleeson et al., 2012), they are reported to have an overly simplistic description of glacier dynamics (Van Dijk et al., 2014, Cáceres et al., 2020) and to mostly treat glaciers as non-glaciated land (Cáceres et al., 2020). The complex and dynamic contribution of glacier runoff to basin runoff is therefore unlikely to be captured by GHMs.

Models that simulate glacier evolution on a global scale do exist in the form of global glacier models (GGMs) (Hock et al., 2019, Marzeion et al., 2020). These models combine a surface mass balance model with a glacier geometry change model, which ranges in complexity from a volume-area-length scaling model (Marzeion et al., 2012, Radić and Hock, 2014) to a mass-conserving retreat parameterization (Huss and Hock, 2015) to a prognostic ice dynamics model (Zekollari et al., 2019, Maussion et al., 2019). Although most GGMs are developed with the goal of simulating the mass balance and evolution of glaciers, some also produce glacier runoff as a model output (Hirabayashi et al., 2010, Radić and Hock, 2014, Huss and Hock, 2018), thus making them suitable for coupling with GHMs.

Several studies have been carried out on the interface between hydrology and glaciology on a global scale. Kaser et al. (2010) compared glacier runoff with the mean upstream precipitation at several elevations to estimate the contribution of glacier runoff along the course of a multitude of large glaciated rivers. To a similar purpose, Schaner et al. (2012) used a land surface hydrological model combined with an energy-balance model. Huss and Hock (2018) compared the runoff of a GGM to monthly average basin runoff observations to assess the changing contribution to basin-scale runoff and the timing of intra- and inter-annual peak water. Finally, Cáceres et al. (2020) coupled a GGM with a GHM to assess the joint contribution of glacial and non-glacial water storage anomalies to ocean mass change. While these studies give valuable insights into the general contribution of glacier runoff to basin runoff and sea-level rise, no study so far has investigated whether the coupling of hydrological models and glacier models on a global scale can lead to better daily runoff predictions. This knowledge gap was likewise pointed out by Radić and Hock (2014).

In this study, we test the hypothesis that the coupling of a GGM with a GHM can lead to an increase in runoff prediction quality and a decrease in model uncertainty in glacierized basins, considering the poor glacier representation in GHMs as compared to GGMs. To this end, the GHM PCR-GLOBWB 2 (Sutanudjaja et al., 2018) is coupled with the GGM GloGEM (Huss and Hock, 2015). We evaluate the results of the coupled model over 25 large glacierized basins through a comparison with the uncoupled GHM, which serves as the benchmark. In the evaluation we aim to identify general differences in behavior between the two models, as well as which model best reproduces the observed basin runoff. Finally, by using the eWaterCycle II framework (Hut et al., 2018) to run the GHM and perform the coupling we intend to increase the (re)producibility of this study.

2 Models and data

2.1 Global hydrological model

We used PCR-GLOBWB 2 (Sutanudjaja et al., 2018), the latest version of PCR-GLOBWB (PCRaster Global Water Balance) (Van Beek et al., 2011), as the global hydrological model. It covers all continents, except Greenland and Antarctica, with a 5 arcmin grid cell resolution, which is comparable to other state-of-the-art GHMs (Sood and Smakhtin, 2015, Bierkens et al., 2015). All processes are calculated on a daily time step, except for the hydrodynamic river routing, which uses sub-daily time stepping. The model fully integrates both withdrawal by and return flow from irrigation, livestock, industry and households. The four standard land cover types are tall natural vegetation, short natural vegetation, non-paddy irrigated crops and paddy irrigated crops, but there is an option to include custom land cover types. The snow module of PCR-GLOBWB is based on the HBV snow module (Bergstrom, 1995) and considers accumulation, melt according to a degree day factor, and refreezing. No redistribution or geometry change of the snowpack is considered, and glacier processes are not accounted for. For latitudes up to 60 degrees the Digital Elevation Model (DEM) of HydroSHEDS (Lehner et al., 2008) is used, while for latitudes over 60 degrees the lower resolution HYDRO1K DEM of USGS is used (Verdin and Greenlee, 1998). For the sake of consistency with the GGM, ERA-Interim Reanalysis temperature and precipitation data are used as forcing (Dee et al., 2011). It should be emphasized that PCR-GLOBWB 2 is generally not calibrated (Sutanudjaja et al., 2018), so this practice is followed here. Rather, mostly the initial parameterization of Van Beek et al. (2011) is used.

2.2 Global glacier model

The Global Glacier Evolution Model (GloGEM) was initially developed by Huss and Hock (2015). It calculates the mass balance and the glacier geometry changes for each glacier individually at a monthly resolution. The data used in this study is from a follow-up study focusing on the glacier runoff (Huss and Hock, 2018). For this study GloGEM was run for all glaciers in the 56 glacierized drainage basins with an area of more than 50.000 km², a glacier area of more than 30 km² and an ice cover of more than 0.01%. The glacier outline data were drawn from the Randolph Glacier Inventory (RGI) v4.0 (Pfeffer et al., 2014). The glacier runoff is defined as the total amount of water originating from the glacierized area drawn from the Randolph Glacier Inventory, even when a part of the initially glaciated area becomes ice-free throughout the simulation. Calibration was done by Huss and Hock (2015) using glacier mass balances derived from in-situ and gravimetry measurements and observed glacier area changes, under the assumption that glacier melt is the dominant process for glacier runoff. The model was run for the period 1980-2100. For the period 1980-2012 the monthly near-surface air temperature and precipitation data were drawn from the ERA-Interim Reanalysis (Dee et al., 2011), while for the period 2013-2100 three different emission scenarios (RCP2.6, RCP4.5 and RCP8.5) from 14 GCMs of the CMIP5 framework (Taylor et al., 2012) were used.

Given the scope of this study, we only used a limited temporal range of the runoff data provided by Huss and Hock (2018). The upper bound is the last year in which ERA-Interim is used to force the model, since the climate projections of CMIP5 are not suitable for an evaluation study, as they are not constrained by observations. The lower bound is dictated by the dates of the first RGI measurements, since they define the area from which the glacier runoff originates in GloGEM (Huss and Hock, 2018). These first measurements are mainly from 1999-2000 following the application of new satellite imagery (Pfeffer et al., 2014). The considered date range thus becomes October (April) 2000 to September (March) 2012 for the Northern (Southern) hemisphere. This corresponds to the hydrological years 2001-2012.

2.3 Glacier runoff preprocessing

To match the spatial and temporal resolution of PCR-GLOBWB 2, the individual glacier runoff data was first rasterized from the individual RGI extents to 5 arcmin grid cells, and consequently resampled from monthly to daily resolution. The resampling was done with a weight function using the ERA-Interim surface temperature data. Since melt is determined by the daily maximum temperature instead of the daily mean temperature, only days with a daily mean temperature below -5°C were excluded from the weighting. A resampling to diurnal resolution was not possible given the daily time step of PCR-GLOBWB 2, despite glaciers exhibiting a strong day-night cycle. The following weight function was applied:

$$w_D = \frac{1 + \alpha * \frac{T_D - \overline{T_{T>-5}}}{\overline{T_{T>-5}}}}{N_{T>-5}} * (T_D > -5)$$

$$R_D = w_D * R_M$$

In which w_D is the weight given to a particular day, T_D is the mean daily temperature at surface, $\overline{T_{T>-5}}$ is the average of all mean daily temperatures above -5 °C in the considered month and $N_{T>-5}$ is the number of days with a mean daily temperature above -5 °C. α is a weighting factor that was set to 20 after calibration on the runoff of the Great Aletsch glacier (BAFU) (see Appendix B). Finally, R_D and R_M are the daily and monthly grid cell glacier runoff respectively. Since it is linear in nature and the sum of the weights always equals $N_{T>-5}$, this weight function is mass-conserving.

2.4 Model framework

eWaterCycle II (Hut et al., 2018) is a global-scale hydrological modeling platform that aims to improve the accessibility and reproducibility of hydrological models. In eWaterCycle II, hydrological models are run in containers and communicate with the central experiment that runs in a Jupyter Notebook. Communication with hydrological models is made independent of the language the model is written in through GRPC4BMI (van den Oord et al., 2019) and BMI (Hutton et al., 2020). Additionally, the ESMValTool (Eyring et al., 2016) implementation in eWaterCycle II allows for smooth preprocessing and high compatibility of forcing data. We made use of eWaterCycle II to obtain the adequate ERA-Interim forcing data, to access PCR-GLOBWB 2 and to couple it with GloGEM. All coding used to setup this study will be made available to allow for easy reproduction or continuation.

2.5 Basin runoff observations

As the source for runoff observations to evaluate against, we used the Global Runoff Data Centre (GRDC) (GRDC, 2016) for all basins except for the Rhone, where we used observations from the French national hydrological service (Hydrobanque, 2020). Out of the 56 basins used by Huss and Hock (2018), 30 are present in the GRDC database with more than 5 years of daily runoff observations between 2000 and 2012. If a basin contained more than one measuring station in the GRDC database the most upstream station was chosen that still included all the glacier runoff (e.g. for the Rhine a gauging station at Basel was chosen instead of the most downstream station at Lobith). The most upstream point that included all glacier runoff, hereafter called glacier sink, was found using HydroSHEDS (Lehner et al., 2008). If the only available station was upstream of the glacier sink we excluded the glaciers that would not drain into that station from our analysis. Appendix A contains more information on the GRDC station numbers and on the available years. Detailed maps of all basins including glacier coverage, gauging station location and glacier sink location are presented in Appendix H. While the GRDC database does contain stations along the Rhone in Switzerland, the glacier sink is near the river mouth at Beaucaire. Therefore, observations at Beaucaire from the Hydro banque were used as an alternative.

Of the 30 resulting basins, 5 were discarded for analysis for various reasons (see Appendix E). The remaining 25 large-scale glacierized basins are mostly concentrated in North-West America and Europe (see Figure 1). Unfortunately, runoff data from rivers originating in the Himalayas are very scarce, despite many of them being some of the world’s most important and vulnerable glacier-fed river basins (Immerzeel et al., 2020). (Seasonally) arid regions are likewise underrepresented, the only exceptions being the Rhone and the Negro (respectively Cfb/Csa and Csb/Bsk on the Köppen-Geiger climate classification scale (Kottek et al., 2006)).

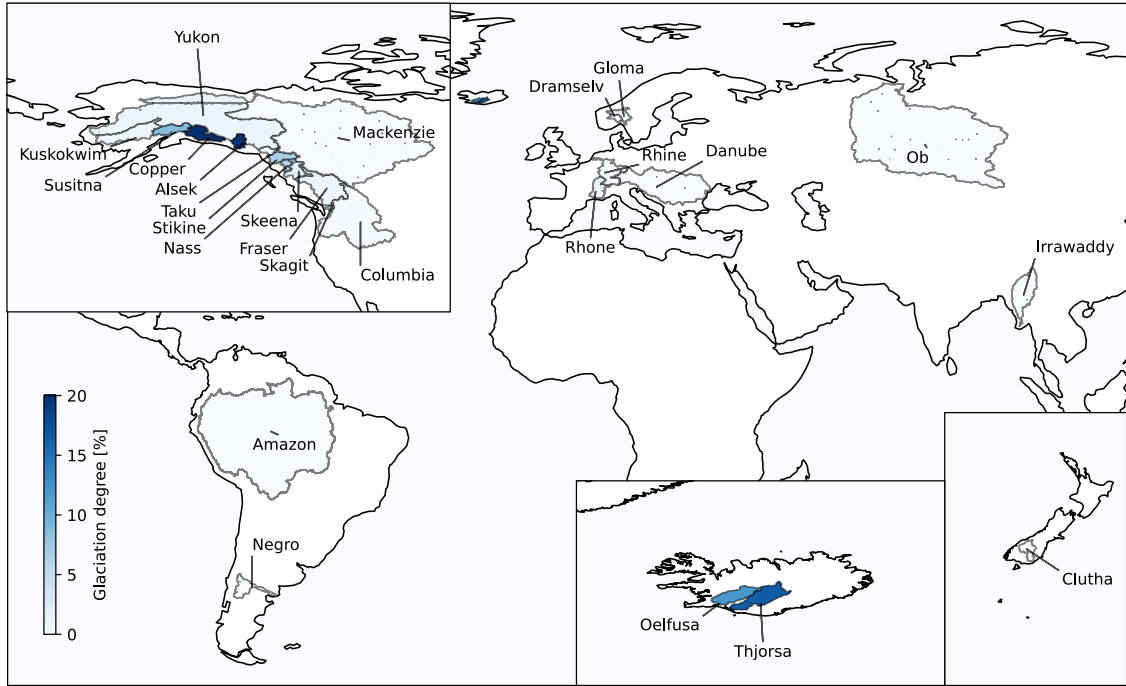


Figure 1: The 25 large-scale (>50.000 km²) glacierized basins for which sufficient runoff observations are found. The color represents the percentage of basin area covered by glaciers.

3 Methods

3.1 Model coupling

Within the context of this study, with ‘coupling’ simply the insertion of the GloGEM output into PCR-GLOBWB 2 is meant. Several situations can be thought of for which further coupling between a glacier model and a hydrological model could be applied, such as surging glaciers damming upstream rivers (Sevestre and Benn, 2015) or the flow of subglacial groundwater (Vincent et al., 2019), but these are not considered in this study.

Simply adding the GloGEM output to PCR-GLOBWB 2 would result in double counting, since PCR-GLOBWB 2 already calculates its own runoff for glacier-covered areas. Ideally, the specific glacier-covered land in PCR-GLOBWB 2 would be replaced by GloGEM, but such a landcover class is not included. Short natural vegetation is practically the only landcover class present in high mountain grid cells, given that glaciers do not tend to extend below the tree line, and can therefore be held accountable for the output of glacier-covered areas in PCR-GLOBWB 2. Thus, to prevent double counting, the glacier fraction of each grid cell is calculated with the RGI glacier extents and subsequently subtracted from the short natural vegetation landcover class, thereby restraining the landcover classes from adding up to 1:

$$(f_{short\ natural\ veg.} - f_{glacier}) + f_{tall\ natural\ veg.} + f_{paddy\ crop} + f_{non-paddy\ crop} = 1 - f_{glacier}$$

Effectively, this causes PCR-GLOBWB 2 to omit any calculations on the glacier-covered area, without having to adjust the source code or the forcing and without having to create a new landcover class. This is beneficial regarding the reproducibility of this approach with other models than the ones used in this study. The only additional adjustment that had to be made was the disabling of the PCR-GLOBWB 2 setting that ensures the landcover classes to add up to one. The option of subtracting the glacier fraction from each landcover class proportionally, rather than only from the short natural vegetation landcover, was tested, but produced only minimal difference in results and was consequently discarded. It should be re-emphasized that this methodology replaces an arbitrary fraction of the PCR-GLOBWB 2 grid cells that does not directly represent the glacier-covered area.

As for the coupling itself, the GloGEM glacier runoff is added to the PCR-GLOBWB 2 variable called *channel_storage* for each time step, which is equivalent to a direct routing into the stream. This is a heavy simplification, since groundwater infiltration both under the glacier and after the glacier terminus is possible

and even likely with certain rock types (Vincent et al., 2019). Among several examples, Vincent and Hart (2017) estimated the subglacial groundwater recharge to be 85% of the glacial melt at the Skalafell glacier in Iceland, while Castellazzi et al. (2019) finds an increased total water storage in the Canadian Rocky Mountains despite a decreased total ice mass, indicating a large glacial groundwater recharge that could delay the transfer to rivers by tens to hundreds of years. Still, according to Vincent et al. (2019), “most of the glaciological and/or hydrological models and studies of present systems do not take into account groundwater dynamics”. Nevertheless, given the large scope of this study and the lacking research on this topic (Vincent et al., 2019), we will ignore glacial groundwater recharge, in line with other studies (Marzeion et al., 2020, Cáceres et al., 2020).

3.2 Spilling prevention

As a consequence of the basin boundary rasterization, a considerable part of the glaciers ended up in grid cells that were at the risk of being routed into adjacent basins, causing the runoff of these glacier to be ‘spilled’. To neutralize this spilling, the runoff of these glaciers was transferred to neighboring grid cells that were certain to drain in basin in question. This spilling prevention was only applied to the coupled model and not to the benchmark, effectively leading to a difference in total basin area.

3.3 Model settings

Three different model settings were run (see Table 1). The benchmark model is the default PCR-GLOBWB 2 run to which the coupled model will be compared. The bare model is an auxiliary model setting that does not include the GloGEM coupling but does apply the adjusted landcover ratios discussed in section 3.1. Finally, the coupled model applies both the adjusted landcover fractions and the GloGEM coupling. In theory, the difference between the benchmark and the bare model gives the routed PCR-GLOBWB 2 runoff for glacier-covered areas, while the difference between the coupled and the bare model gives the routed GloGEM runoff. We assume the bare model to not include any glacier runoff. To find the accurate variable states to initialize the models with, a year with the climatological average of the period 1990-1999 was run 50 times to serve as a spin-up. This approach is similar to the spin-up process used by Sutanudjaja et al. (2018).

Model setting	Adjusted landcover	GloGEM coupling
<i>Benchmark</i>		
<i>Bare</i>	x	
<i>Coupled</i>	x	x

Table 1: The three different model settings used in this study. The difference between the bare model and the coupled model is assumed to give the routed GloGEM runoff.

3.4 Evaluation

3.4.1 Benchmark comparison

To identify differences in basin runoff between the coupled model and the benchmark as a function of the time of the year, we apply the following normalized difference metric over all basins:

$$Normalized\ difference_{t=d} = \frac{\overline{(Q_{Coupled} - Q_{Benchmark})_{t=d}}}{q_{99}(Q_{Coupled} - Q_{Benchmark})}$$

in which d is the calendar day, Q is the basin runoff and q_{99} is the 99th quantile of the difference taken over the whole time range. The normalization was applied with the 99th quantile instead of the maximum difference to avoid the influence of extreme maxima. With this metric, a positive value indicates a higher average discharge of the coupled model on that particular calendar day, and vice versa. Given the highly seasonal pattern of glacier melt seasonal patterns in the difference are expected.

Additionally, to quantify not only the nature of the difference between the coupled model and the benchmark but also its relative magnitude, the ratio among them is calculated by dividing the coupled model runoff by the benchmark runoff and taking the average for each month.

3.4.2 Evaluation against observations

In analyzing whether the coupled model performs better than the benchmark in reproducing the observed basin runoff, it is important to consider the maximum possible performance difference (Seibert et al., 2018). Since in

this study the difference between the two models can only be attributed to a difference in glacier representation, we assume the maximum possible performance difference to occur when one model accurately simulates the true glacier runoff and the other model simulates a glacier runoff of zero. The true glacier runoff is then assumed to be the maximum among the PCR-GLOBWB 2 glacier runoff and the GloGEM glacier runoff, calculated through their respective difference with the bare model (see section 3.3). The performance difference between the coupled model and the benchmark can then be expressed relative to the maximum possible performance difference as follows:

$$RRD = \frac{RMSE(Q_{Obs}, Q_{Coupled}) - RMSE(Q_{Obs}, Q_{Benchmark})}{RMSE(Q_{Bare}, \max(Q_{Coupled}, Q_{Benchmark}))}$$

in which RRD stands for the relative RMSE difference and $RMSE$ entails the root of the mean squared error. With the RRD, a positive sign indicates whether the coupled model performs better compared to the benchmark and vice versa, while the value indicates the fraction of the difference to the maximum possible difference. The RRD is therefore always between -1 and 1. A discussion on the metric choice is presented in Appendix D.

4 Results

4.1 Hydrographs

In line with the expectations, the difference between the benchmark and the coupled model is most pronounced in the melt season and in highly glaciated basins (see Figure 2, and Appendix G for full hydrographs for all basins). In highly glaciated basins, the glacier representation is a major factor of uncertainty in the summer months whereas in lowly glaciated basins hardly any difference can be observed.

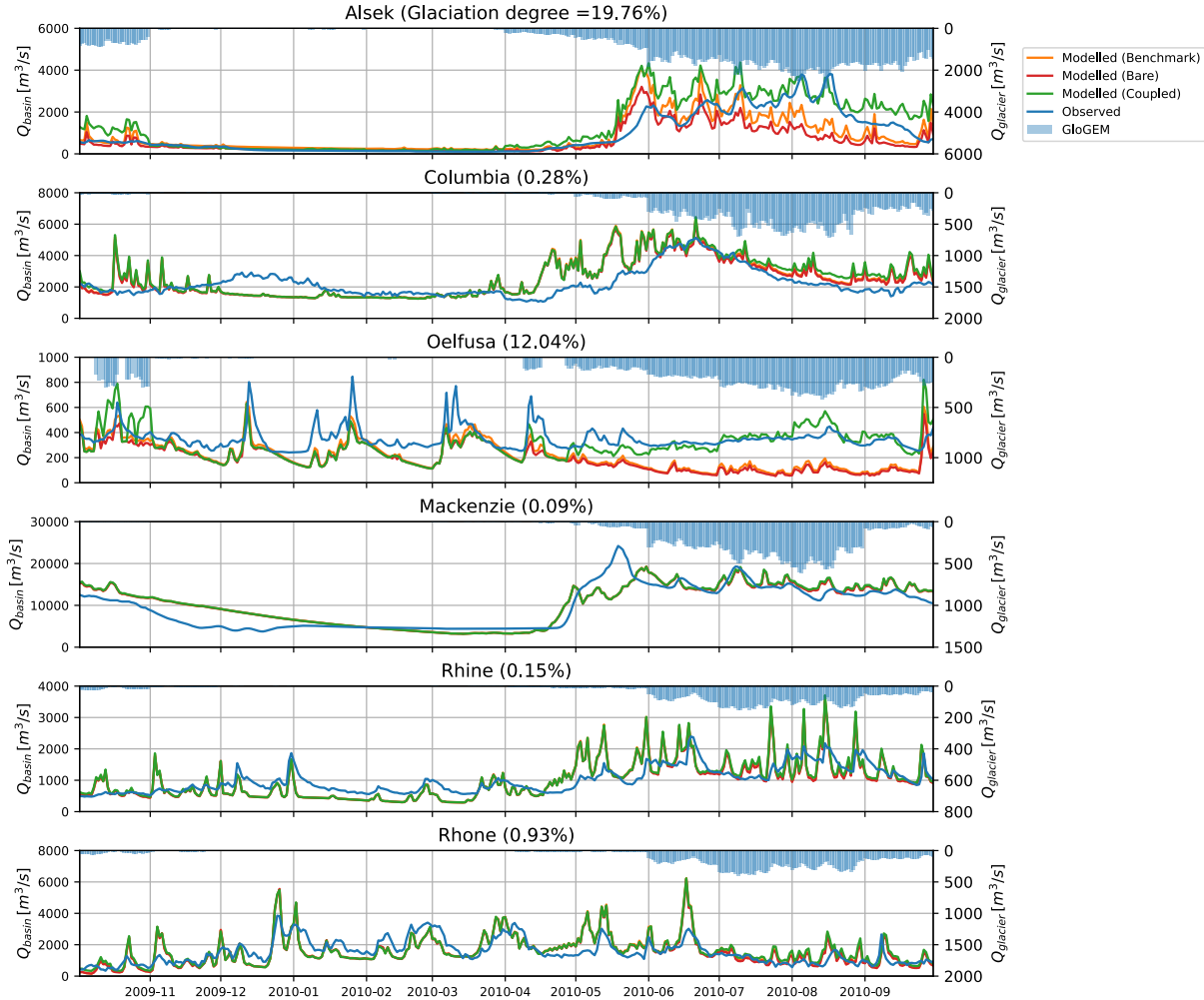


Figure 2: Hydrographs of a representative selection of the 25 basins for a representative year. The left y-axis represents the runoff at the selected gauging station, while the right y-axis represents the GloGEM total basin glacier runoff. Notice the different extents per basin on the y-axes. The glacier area fraction is presented with the basin name.

4.2 Benchmark comparison

The normalized difference between the coupled model and the benchmark indicates that the coupled model produces higher basin runoffs during the melt season, especially in July/January, August/February and September/March in the northern/southern hemisphere. (see Figure 3a). Only a few lowly glaciated basins deviate from this overall pattern (e.g. Amazon, Ob and Negro). However, there are some basins that exhibit lower basin runoffs for the coupled model in May/November, and to a lesser degree in October/April.

While the general pattern in difference is shared by nearly all basins, the impact this difference has on the total simulated runoff is greater in highly glaciated basins (see Figure 3b). In the Amazon, the coupled model runoff at the peak of the melt season (July/January and August/February) is 100.07% of the benchmark runoff, while in the Oelfusa it amounts to 352.27%.

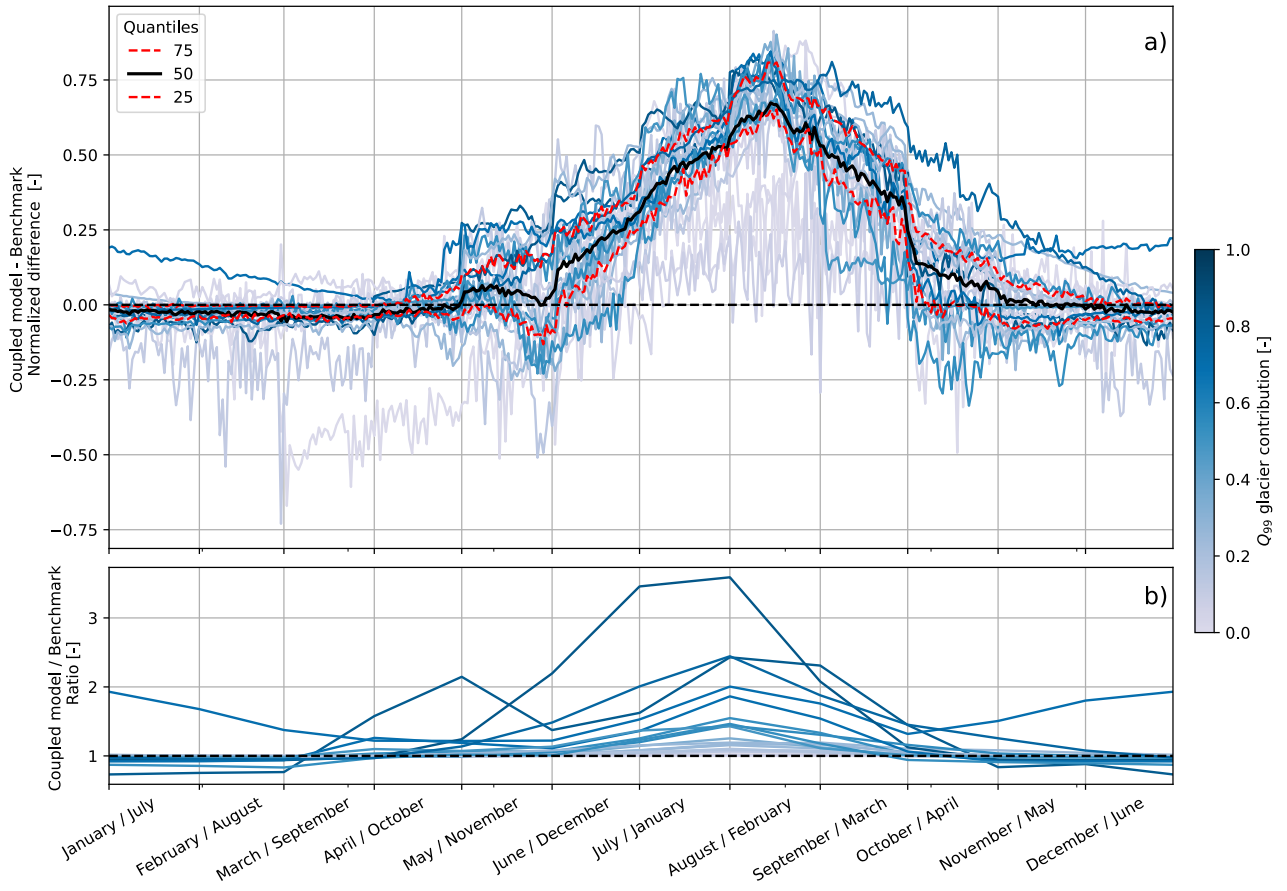


Figure 3: Mean differences between the coupled model and the benchmark for all 25 basins. The color represents the 99th quantile of the contribution of the routed GloGEM glacier runoff to the coupled model runoff. The months on the x-axis are given for both the northern and southern hemisphere. **a)** The difference normalized using the 99th quantile of the difference over the whole time range. The mean is computed for each calendar day. **b)** The ratio of the coupled model over the benchmark. The mean is computed for each month.

4.3 Evaluation against observations

The coupled model better reproduces the observations in highly glaciated basins than in lowly glaciated basins, particularly at the peak of the melt season (July/January and August/February in the northern/southern hemisphere) (see Figure 4). This is reflected by the coefficient of determination (R^2). Most positive scores (14 out of 25 basins) are found in May/November, when the coupled model produces lower basin runoffs than the benchmark in some basins, while the least positive scores (6 out of 25 basins) are found in September/March.

An independent performance evaluation of the benchmark is presented in Appendix C. The results of three other metrics taken over the whole time range are presented in Appendix D.

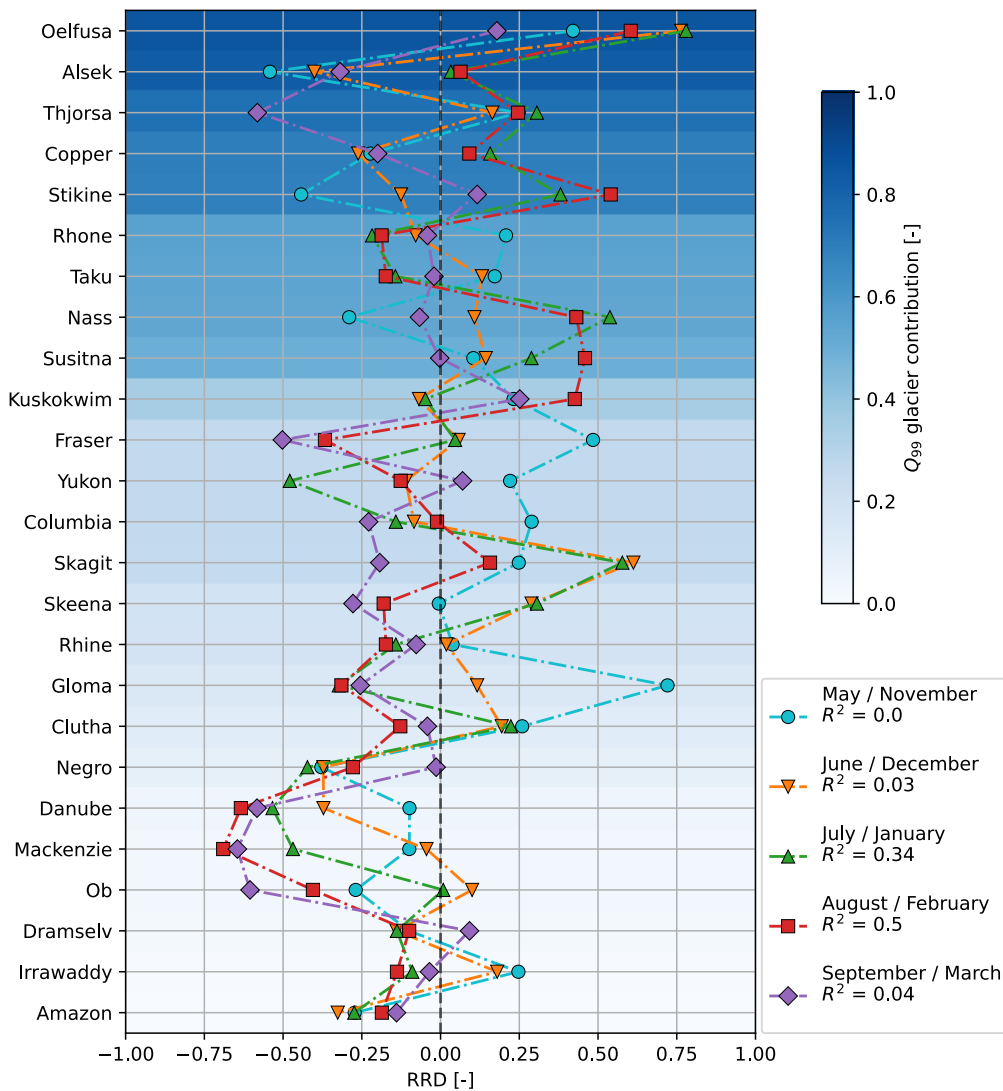


Figure 4: The relative RMSE difference (RRD) over all 25 basins over the melt season. The RMSE difference is calculated relative to the routed glacier runoff, which embodies the maximum possible RMSE difference. Positive scores indicates an improvement of the coupled model over the benchmark and vice versa. The RRD always lies between -1 and 1. The basins are sorted based on the 99th quantile of the contribution of the routed GloGEM glacier runoff to the coupled model runoff. In the legend, the coefficient of determination (R^2) represents the correlation with the glacier contribution, and the months are given for both the northern and southern hemisphere.

5 Discussion

5.1 Benchmark comparison

5.1.1 Overall difference

Since the coupling only involves the glacier-covered area, the difference between the coupled model and the benchmark can only be attributed to the different representation of glaciers. Several explanations were identified for this difference and have been quantified for four representative basins (see Figure 5). Firstly, the lack of a snow redistribution parameterization in PCR-GLOBWB 2 leads to the formation of ‘snow towers’ (Freudiger et al., 2017). Snowfall at high elevations is known to be redistributed through wind and avalanches, as well as through gravitational glacier flow, towards lower elevations, where melt is more likely to occur. Not accounting for these processes can lead to multiannual accumulation of snow at high elevations where temperatures rarely drop below melting point. This phenomenon is also acknowledged by Sutanudjaja et al. (2018) for polar

regions. Out of the 25 basins, 17 simulate significant snow towers (see Appendix F). Meanwhile, GloGEM account for glacier flow through a geometry change module (Huss et al., 2010) and acknowledge the importance of snow redistribution onto glaciers by taking a precipitation correction factor, c_{prec} , as the primary calibration parameter.

Secondly, PCR-GLOBWB 2 does not include estimates of the initial glacier volumes and is therefore unable to capture the additional melt caused by their retreat. Glaciers worldwide have been losing mass (Zemp et al., 2019, Wouters et al., 2019) and will continue to lose mass (Marzeion et al., 2020) as they gradually adapt to changing climate conditions (Zekollari et al., 2020). Currently, many glaciers experience a peak in mass loss and therefore in glacier runoff (Huss and Hock, 2018). Without initial glacier volume estimates, no adaptation to global warming is needed and therefore no additional inter-annual glacier runoff will be simulated by PCR-GLOBWB 2. It should be mentioned that the lack of glacier and ice processes was already mentioned by Sutanudjaja et al. (2018) after observing a poor correlation of the simulated total water storage with gravimetry measurements in Alaskan and Icelandic basins.

Thirdly, while the spilling prevention may have helped in accurately routing all GloGEM glacier runoff in the coupled model, no similar measures were taken for the benchmark. Effectively this leads to a greater basin area for the coupled model and consequently to a greater basin runoff. This effect is greater in basins where a large portion of the glaciers is located at the basin boundary.

Finally, while the above-mentioned c_{prec} helps in accounting for snow redistribution, with a c_{prec} larger than one and without a counter-correction factor in PCR-GLOBWB 2 there is a risk of ending up with an excess in grid cell precipitation. Without snow towers, this could in turn cause an overestimation of runoff, albeit with a certain time lag. An additional reason through which c_{prec} could cause an overestimation is suggested by Zekollari et al. (2019). By calibrating against regional glacier mass balance estimates (Gardner et al., 2013), rather than against individual glacier mass balance estimates, GloGEM potentially underestimates the net mass loss of large glaciers and overestimates the mass loss of small glaciers. Being the primary calibration parameter, GloGEM is likely to apply a c_{prec} of greater than one to match the mass balance estimates in large glaciers, which in turn could potentially lead to an increased sustained runoff on the long term.

Other sources of difference between the coupled model and the benchmark remain unquantified. Particularly in the Alsek and Oelfusa basins large gaps are left unaccounted for (see Figure 5). Evaporation/sublimation and groundwater recharge calculations are included in PCR-GLOBWB 2 and not by GloGEM, but while they could (temporarily) account for part of the mass imbalance, their effect is estimated to be small.

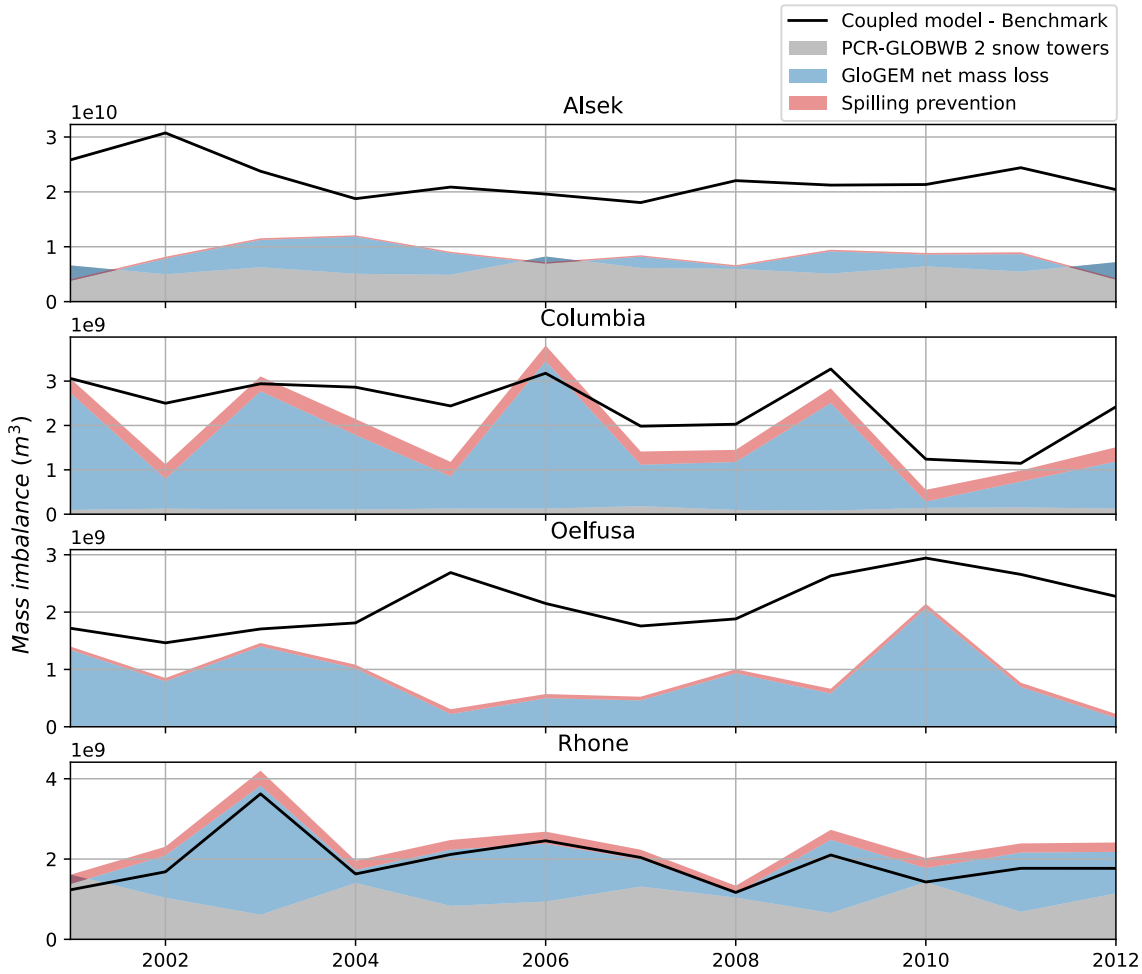


Figure 5: Factors determining the annual mass balance difference between the coupled model and the benchmark. The black line represents the annual sum of the difference between the coupled model and benchmark basin runoff. The annual increase in snow water equivalent modeled by PCR-GLOBWB 2 following the lack of snow redistribution parameterization is depicted in grey. The annual net mass loss from retreating glaciers as modeled by GloGEM is presented in blue. The annual glacier runoff that would have spilled into neighboring basins as a consequence of basin boundary rasterization is shown in red. This contributes to the imbalance as the spilling prevention is not applied to the benchmark. Other factors explaining the imbalance are left unquantified.

5.1.2 Late spring difference

While throughout most of the melt season the coupled model produces higher runoff, in the beginning of the melt season this difference is of opposite sign in many basins. These basins only partially overlap with the basins in which no snow towers were found, and equally no correlation with geographical location or climate was discovered. We hypothesize this effect to be the result of the limited horizontal and vertical spatial resolution of the temperature forcing in PCR-GLOBWB (Van Beek et al., 2011). Mountainous regions are characterized by steep horizontal and vertical temperature gradients, causing snow and glacier melt and accumulation to be highly spatially dependent. If due to an insufficient spatial resolution a model fails to capture these gradients there is a high chance of the melt being simulated too suddenly (Sexstone et al., 2020, Immerzeel et al., 2014). PCR-GLOBWB 2 does facilitate a temperature downscaling from the 45 arcmin of ERA-Interim to the 5 arcmin model resolution using lapse rates from the CRU CL 2.0 climatology (New et al., 2002) to better account for snow dynamics, but this is arguably still too coarse for the gradients present at glaciated mountain areas. Additionally, Van Beek et al. (2011) hypothesize part of the melt timing error of PCR-GLOB to be a consequence of the use of a constant melt rate and threshold temperature in the snow module. However, it should be mentioned that the high spatial resolution needed in mountainous areas is still rather unfeasible for models that are designed to operate on a global scale, and therefore a certain degree of simplification will always be needed.

GloGEM does not downscale ERA-Interim data horizontally, but it does apply a set of twelve constant monthly temperature lapse rates to all glacier elevation bands (Huss and Hock, 2015). Additionally, a shift in the air temperature series is applied as a potential final calibration parameter, which also increases the likelihood of agreement with local temperature gradients. Thus, particularly in the high temperature gradients around the highest elevations, GloGEM is likely to ensure a more gradual melt process along late spring. While in many basins PCR-GLOBWB 2 simulates a stronger onset of the melt season than the observations (e.g. Columbia, Susitna, Copper), it should be mentioned that for some basins the opposite is the case (e.g. Kuskokwim and Mackenzie) (see Appendix G).

5.2 Evaluation against observations

Using the RRD to evaluate whether the coupled model better reproduces the observed basin runoff, the results for highly glaciated basins are more meaningful than for lowly glaciated basins. In lowly glaciated basins, the glacier parameterization can only have a limited influence on the total runoff simulated by PCR-GLOBWB 2. Since the coupling of GloGEM generally leads to higher runoff values, the RRD will therefore simply be negative when PCR-GLOBWB 2 overestimates the basin runoff, and vice versa. In many basins PCR-GLOBWB 2 mostly overestimates the basin runoff (e.g. Danube, Ob, Irrawaddy), causing the RRD to be negative even in the hypothetical case that the glacier runoff is simulated perfectly with the coupled model. On the other hand, in highly glaciated basins glacier runoff is an important part of the basin runoff and therefore the quality of the glacier parameterization is a decisive factor. The RRD value can better reflect this quality and is more meaningful due to the higher glacier contribution.

The majority of RRD values for highly glaciated basins are positive. If we consider the threshold of highly glaciated basins to be at a value of 0.5 for the 99th quantile of glacier contribution to streamflow (i.e. Susitna), then the RRD is positive for 5 out of 9 values in May and June, 7 out of 9 in July and August and 2 out of 9 in September (northern hemisphere). Thus, particularly at the peak of the melt season (July and August) the coupled model performs better than the benchmark. The lesser performance in September can partially be explained by PCR-GLOBWB 2 reproducing the observations more accurately, causing the addition of GloGEM to lead to an overestimation (e.g. Thjorsa, Alsek). The highest scores over all metrics are obtained by the basin with the highest maximum glacier contribution, the Oelfusa in Iceland, where PCR-GLOBWB 2 heavily underestimates summer basin runoff.

Considering the greater meaning and higher scores of the RRD in highly glaciated basins, we can conclude that there is a high likelihood that the coupled model provides a better representation of glacier runoff than the benchmark. While in this study the coupling does not lead to better results for lowly glaciated basins, it is probable that the glacier parameterization has in fact improved in these basins, but that this is not visible in the results.

5.3 Weighting factor sensitivity

The resampling from monthly to daily resolution of the GloGEM output is important to accurately capture the daily fluctuations of glacier runoff needed for daily runoff prediction. In the weighting function the weighting factor α was chosen after a simple calibration on the Aletsch glacier runoff observations (Appendix B). Over a series of 8 evenly spaced values between 0 and 40, a value of 20 gave the lowest RMSE. To measure the impact of the weighting factor on the basin scale, a sensitivity analysis is performed on the same basins as in section 5.1.1. The coupled model runs are repeated with weighting factors of 10 and 30. The resulting RRD values are consequently compared with the reference case ($\alpha=20$). The sensitivity analysis shows that on the basin scale a weighting factor of 30 leads to better results in 3 out of 4 basins (see Table 2). Additionally, it ensures a smoother glacier runoff transition between months where jumps are sometimes visible in the reference model (e.g. Rhone). Nonetheless, the maximum mean differences in RRD are -0.009 for $\alpha=10$ and +0.0176 for $\alpha=30$. The sensitivity of the weighting factor is therefore limited compared to other factors of the glacier parameterization, such as snow redistribution and mass balance calculations.

α	Alsek	Columbia	Oelfusa	Rhone
10	-0.00905	-0.00144	-0.00587	0.00015
30	0.01364	0.00247	0.01759	-0.00232

Table 2: Difference in RRD between the $\alpha=10$ and $\alpha=30$ cases with the reference case ($\alpha=20$).

6 Conclusion

In this study, we coupled the global hydrological model PCR-GLOBWB 2 with the global glacier model GloGEM to investigate whether their coupling can lead to better runoff predictions in glaciated basins. The coupling was done within eWaterCycle II by adding the rasterized and resampled GloGEM glacier runoff to the channel storage of PCR-GLOBWB 2 grid cells. To avoid double counting, in each grid cell a fraction equal to the glaciation degree was subtracted from the grassland landcover type. Both the uncoupled benchmark and the coupled model were run for 25 large ($>50.000 \text{ km}^2$) glaciated basins across multiple continents during the hydrological years 2001-2012. The results were evaluated mutually and against GRDC runoff observations. The main outcomes are the following:

- The coupled model generally produces higher runoff across all basins. In summer, it ranges between 100.07% and 352% of the mean monthly benchmark runoff in lowly and highly glaciated basins respectively. This difference can mainly be attributed to an underestimation by the benchmark, which simulates the formation of permanent ‘snow towers’ and is unable to capture the additional melt induced by the retreat of glaciers worldwide.
- In some basins the coupled model produces lower runoff than the benchmark in late spring, when the benchmark is likely to simulate a more abrupt onset of the melt season due to a limited spatial resolution.
- While in the evaluation against basin runoff observations the coupling does not lead to better runoff predictions across all basins, it does lead to majoritarily positive results in highly glaciated basins, where the coupling has the most impact.

Taken together, these outcomes suggest that the coupling of a GHM and a GGM can lead to a better glacier parameterization, and therefore a high likelihood of increased runoff prediction quality and decreased model uncertainty in glaciated basins. This study also underlines the importance of glacier representation in highly glaciated basins. Furthermore, it validates the feasibility of eWaterCycle II as a hydrological model framework.

Given the increased viability of GHMs but their limited glacier representation, there is a large potential in the research covering the coupling of GHMs and GGMs. To further test this approach, future studies could apply ensembles of GHMs and/or GGMs, include more basins, particularly around the Himalayas, or perform a joint calibration to improve model parameters and reduce GHM equifinality. To facilitate this future work, the availability and applicability of GGMs would be of great importance. Ultimately, the coupling of GHMs and GGMs could lead to a better understanding of the impact that climate change will have in glaciated basins.

References

- BAFU. Hydrologische Daten und Vorhersagen. URL <https://www.hydrodaten.admin.ch/de>.
- S. Bergstrom. The HBV model. *Computer models of watershed hydrology*, 1995.
- H. Biemans, C. Siderius, A. F. Lutz, S. Nepal, B. Ahmad, T. Hassan, W. von Bloh, R. R. Wijngaard, P. Wester, A. B. Shrestha, and W. W. Immerzeel. Importance of snow and glacier meltwater for agriculture on the Indo-Gangetic Plain. *Nature Sustainability*, 2(7):594–601, 2019. ISSN 23989629. 10.1038/s41893-019-0305-3. URL <http://dx.doi.org/10.1038/s41893-019-0305-3>.
- M. F. Bierkens, V. A. Bell, P. Burek, N. Chaney, L. E. Condon, C. H. David, A. de Roo, P. Döll, N. Drost, J. S. Famiglietti, M. Flörke, D. J. Gochis, P. Houser, R. Hut, J. Keune, S. Kollet, R. M. Maxwell, J. T. Reager, L. Samaniego, E. Sudicky, E. H. Sutanudjaja, N. van de Giesen, H. Winsemius, and E. F. Wood. Hyper-resolution global hydrological modelling: What is next?: "Everywhere and locally relevant" M. F. P. Bierkens et al. Invited Commentary. *Hydrological Processes*, 29(2):310–320, 2015. ISSN 10991085. 10.1002/hyp.10391.
- M. I. Brunner, D. Farinotti, H. Zekollari, M. Huss, and M. Zappa. Future shifts in extreme flow regimes in Alpine regions. *Hydrology and Earth System Sciences*, 2019. ISSN 16077938. 10.5194/hess-23-4471-2019.
- D. Cáceres, B. Marzeion, J. H. Malles, B. Gutknecht, H. Müller Schmied, and P. Döll. Assessing global water mass transfers from continents to oceans over the period 1948–2016. *Hydrology and Earth System Sciences Discussions*, (February):1–37, 2020. ISSN 1027-5606. 10.5194/hess-2019-664.

- P. Castellazzi, D. Burgess, A. Rivera, J. Huang, L. Longuevergne, and M. N. Demuth. Glacial Melt and Potential Impacts on Water Resources in the Canadian Rocky Mountains. *Water Resources Research*, 55(12): 10191–10217, 2019. ISSN 19447973. 10.1029/2018WR024295.
- S. Cauvy-Fraunié and O. Dangles. A global synthesis of biodiversity responses to glacier retreat. *Nature Ecology & Evolution*, 3(12):1675–1685, 2019.
- R. Dankers, N. W. Arnell, D. B. Clark, P. D. Falloon, B. M. Fekete, S. N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada, and D. Wisser. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):3257–3261, 2014. ISSN 00278424. 10.1073/pnas.1302078110.
- D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, and F. Vitart. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656): 553–597, 2011. ISSN 00359009. 10.1002/qj.828.
- V. Eyring, M. Righi, A. Lauer, M. Evaldsson, S. Wenzel, C. Jones, A. Anav, O. Andrews, I. Cionni, E. L. Davin, C. Deser, C. Ehbrecht, P. Friedlingstein, P. Gleckler, K. D. Gottschaldt, S. Hagemann, M. Jukes, S. Kindermann, J. Krasting, D. Kunert, R. Levine, A. Loew, J. Mäkelä, G. Martin, E. Mason, A. S. Phillips, S. Read, C. Rio, R. Roehrig, D. Senftleben, A. Sterl, L. H. Van Ulft, J. Walton, S. Wang, and K. D. Williams. ESMValTool (v1.0)-a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, 9(5):1747–1802, 2016. ISSN 19919603. 10.5194/gmd-9-1747-2016.
- D. Farinotti, S. Usselman, M. Huss, A. Bauder, and M. Funk. Runoff evolution in the Swiss Alps: Projections for selected high-alpine catchments based on ENSEMBLES scenarios. *Hydrological Processes*, 26(13):1909–1924, 2012. ISSN 08856087. 10.1002/hyp.8276.
- C. Frans, E. Istanbuloglu, D. P. Lettenmaier, A. G. Fountain, and J. Riedel. Glacier Recession and the Response of Summer Streamflow in the Pacific Northwest United States, 1960–2099. *Water Resources Research*, 54(9):6202–6225, 2018. ISSN 19447973. 10.1029/2017WR021764.
- D. Freudiger, I. Kohn, J. Seibert, K. Stahl, and M. Weiler. Snow redistribution for the hydrological modeling of alpine catchments. 4(October):1–16, 2017. 10.1002/wat2.1232.
- A. S. Gardner, G. Moholdt, J. G. Cogley, B. Wouters, A. A. Arendt, J. Wahr, E. Berthier, R. Hock, W. T. Pfeffer, G. Kaser, S. R. Ligtenberg, T. Bolch, M. J. Sharp, J. O. Hagen, M. R. Van Den Broeke, and F. Paul. A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. *Science*, 340(6134):852–857, 2013. ISSN 10959203. 10.1126/science.1234532.
- T. Gleeson, Y. Wada, M. F. P. Bierkens, and L. P. H. Van Beek. Water balance of global aquifers revealed by groundwater footprint. *Nature*, 488(7410):197–200, 2012.
- GRDC. The Global Runoff Data Centre, 56068 Koblenz, Germany, 2016. URL www.bafg.de/GRDC.
- I. Haddeland, J. Heinke, H. Biemans, S. Eisner, M. Flörke, N. Hanasaki, M. Konzmann, F. Ludwig, Y. Masaki, J. Schewe, T. Stacke, Z. D. Tessler, Y. Wada, and D. Wisser. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):3251–3256, 2014. ISSN 00278424. 10.1073/pnas.1222475110.
- Y. Hirabayashi, P. Döll, and S. Kanae. Global-scale modeling of glacier mass balances for water resources assessments: Glacier mass changes between 1948 and 2006. *Journal of Hydrology*, 390(3-4):245–256, 2010. ISSN 00221694. 10.1016/j.jhydrol.2010.07.001. URL <http://dx.doi.org/10.1016/j.jhydrol.2010.07.001>.
- R. Hock, A. Bliss, B. E. Marzeion, R. H. Giesen, Y. Hirabayashi, M. Huss, V. Radic, and A. B. Slangen. GlacierMIP-A model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology*, 65(251):453–467, 2019. ISSN 00221430. 10.1017/jog.2019.22.
- M. Huss. Present and future contribution of glacier storage change to runoff from macroscale drainage basins in Europe. *Water Resources Research*, 47(7):1–14, 2011. ISSN 00431397. 10.1029/2010WR010299.

- M. Huss and R. Hock. A new model for global glacier change and sea-level rise. *Frontiers in Earth Science*, 3 (September):1–22, 2015. ISSN 22966463. 10.3389/feart.2015.00054.
- M. Huss and R. Hock. Global-scale hydrological response to future glacier mass loss. *Nature Climate Change*, 8 (2):135–140, 2018. ISSN 17586798. 10.1038/s41558-017-0049-x. URL <http://dx.doi.org/10.1038/s41558-017-0049-x>.
- M. Huss, D. Farinotti, A. Bauder, and M. Funk. Modelling runoff from highly glacierized alpine drainage basins in a changing climate. *Hydrological processes*, 22(19):3888–3902, 2008.
- M. Huss, G. Jouvett, D. Farinotti, and A. Bauder. Future high-mountain hydrology: A new parameterization of glacier retreat. *Hydrology and Earth System Sciences*, 14(5):815–829, 2010. ISSN 10275606. 10.5194/hess-14-815-2010.
- R. Hut, N. Drost, W. Van Hage, and N. Van De Giesen. eWaterCycle II. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, page 379. IEEE, 2018.
- E. Hutton, M. Piper, and G. Tucker. The Basic Model Interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software*, 5(51):2317, 2020. ISSN 2475-9066. 10.21105/joss.02317.
- Hydrobanque. Hydrobanque, 2020. URL <http://hydro.eaufrance.fr/>.
- W. W. Immerzeel, L. Petersen, S. Raetzli, and F. Pellicciotti. The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. *Water Resources Research*, 50(3):2212–2226, 2014.
- W. W. Immerzeel, A. F. Lutz, M. Andrade, A. Bahl, H. Biemans, T. Bolch, S. Hyde, S. Brumby, B. J. Davies, A. C. Elmore, A. Emmer, M. Feng, A. Fernández, U. Haritashya, J. S. Kargel, M. Koppes, P. D. Kraaijenbrink, A. V. Kulkarni, P. A. Mayewski, S. Nepal, P. Pacheco, T. H. Painter, F. Pellicciotti, H. Rajaram, S. Rupper, A. Sinisalo, A. B. Shrestha, D. Viviroli, Y. Wada, C. Xiao, T. Yao, and J. E. Baillie. Importance and vulnerability of the world’s water towers. *Nature*, 577(7790):364–369, 2020. ISSN 14764687. 10.1038/s41586-019-1822-y. URL <http://dx.doi.org/10.1038/s41586-019-1822-y>.
- P. Jansson, R. Hock, and T. Schneider. The concept of glacier storage: A review. *Journal of Hydrology*, 282 (1-4):116–129, 2003. ISSN 00221694. 10.1016/S0022-1694(03)00258-0.
- G. Kaser, M. Großhauser, and B. Marzeion. Contribution potential of glaciers to water availability in different climate regimes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47):20223–20227, 2010. ISSN 10916490. 10.1073/pnas.1008162107.
- M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263, 2006.
- L. Laurent, J. F. Buoncristiani, B. Pohl, H. Zekollari, D. Farinotti, M. Huss, J. L. Mugnier, and J. Pergaud. The impact of climate change and glacier mass loss on the hydrology in the Mont-Blanc massif. *Scientific Reports*, 10(1):1–11, 2020. ISSN 20452322. 10.1038/s41598-020-67379-7. URL <https://doi.org/10.1038/s41598-020-67379-7>.
- B. Lehner, K. Verdin, and A. Jarvis. New global hydrography derived from spaceborne elevation data. *Eos*, 89 (10):93–94, 2008. ISSN 00963941. 10.1029/2008EO100001.
- B. Marzeion, A. H. Jarosch, and M. Hofer. Past and future sea-level change from the surface mass balance of glaciers. *Cryosphere*, 6(6):1295–1322, 2012. ISSN 19940416. 10.5194/tc-6-1295-2012.
- B. Marzeion, R. Hock, B. Anderson, A. Bliss, N. Champollion, K. Fujita, M. Huss, W. W. Immerzeel, P. Kraaijenbrink, J. H. Malles, F. Maussion, V. Radić, D. R. Rounce, A. Sakai, S. Shannon, R. van de Wal, and H. Zekollari. Partitioning the Uncertainty of Ensemble Projections of Global Glacier Mass Change. *Earth’s Future*, 8(7):1–25, 2020. ISSN 23284277. 10.1029/2019EF001470.
- F. Maussion, A. Butenko, N. Champollion, M. Dusch, J. Eis, K. Fourteau, P. Gregor, A. H. Jarosch, J. Landmann, F. Oesterle, B. Recinos, T. Rothenpieler, A. Vlug, C. T. Wild, and B. Marzeion. The Open Global Glacier Model (OGGM) v1.1. *Geoscientific Model Development*, 12(3):909–931, 2019. ISSN 19919603. 10.5194/gmd-12-909-2019.

- J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.
- M. New, D. Lister, M. Hulme, and I. Makin. A high-resolution data set of surface climate over global land areas. *Climate Research*, 21(1):1–25, 2002. ISSN 0936577X. 10.3354/cr021001.
- W. T. Pfeffer, A. A. Arendt, A. Bliss, T. Bolch, J. G. Cogley, A. S. Gardner, J. O. Hagen, R. Hock, G. Kaser, C. Kienholz, E. S. Miles, G. Moholdt, N. Mölg, F. Paul, V. Radić, P. Rastner, B. H. Raup, J. Rich, M. J. Sharp, L. M. Andreassen, S. Bajracharya, N. E. Barrand, M. J. Beedle, E. Berthier, R. Bhambri, I. Brown, D. O. Burgess, E. W. Burgess, F. Cawkwell, T. Chinn, L. Copland, N. J. Cullen, B. Davies, H. De Angelis, A. G. Fountain, H. Frey, B. A. Giffen, N. F. Glasser, S. D. Gurney, W. Hagg, D. K. Hall, U. K. Haritashya, G. Hartmann, S. Herreid, I. Howat, H. Jiskoot, T. E. Khromova, A. Klein, J. Kohler, M. König, D. Kriegel, S. Kutuzov, I. Lavrentiev, R. Le Bris, X. Li, W. F. Manley, C. Mayer, B. Menounos, A. Mercer, P. Mool, A. Negrete, G. Nosenko, C. Nuth, A. Osmonov, R. Pettersson, A. Racoviteanu, R. Ranzi, M. A. Sarikaya, C. Schneider, O. Sigurdsson, P. Sirguey, C. R. Stokes, R. Wheate, G. J. Wolken, L. Z. Wu, and F. R. Wyatt. The randolph glacier inventory: A globally complete inventory of glaciers. *Journal of Glaciology*, 60(221): 537–552, 2014. ISSN 00221430. 10.3189/2014JoG13J176.
- H. D. Pritchard. Asia’s shrinking glaciers protect large populations from drought stress. *Nature*, 569(7758): 649–654, 2019. ISSN 14764687. 10.1038/s41586-019-1240-1. URL <http://dx.doi.org/10.1038/s41586-019-1240-1>.
- V. Radić and R. Hock. Glaciers in the Earth’s Hydrological Cycle: Assessments of Glacier Mass and Runoff Changes on Global and Regional Scales. *Surveys in Geophysics*, 35(3):813–837, 2014. ISSN 01693298. 10.1007/s10712-013-9262-y.
- S. Ragettli, W. W. Immerzeel, and F. Pellicciotti. Contrasting climate change impact on river flows from high-altitude catchments in the Himalayan and Andes Mountains. *Proceedings of the National Academy of Sciences of the United States of America*, 113(33):9222–9227, 2016. ISSN 10916490. 10.1073/pnas.1606526113.
- B. Schaeffli and H. V. Gupta. Do Nash values have value? *Hydrological Processes*, 21(ARTICLE):2075–2080, 2007.
- N. Schaner, N. Voisin, B. Nijssen, and D. P. Lettenmaier. The contribution of glacier melt to streamflow. *Environmental Research Letters*, 7(3), 2012. ISSN 17489326. 10.1088/1748-9326/7/3/034029.
- J. Schewe, J. Heinke, D. Gerten, I. Haddeland, N. W. Arnell, D. B. Clark, R. Dankers, S. Eisner, B. M. Fekete, F. J. Colón-González, S. N. Gosling, H. Kim, X. Liu, Y. Masaki, F. T. Portmann, Y. Satoh, T. Stacke, Q. Tang, Y. Wada, D. Wisser, T. Albrecht, K. Frieler, F. Piontek, L. Warszawski, and P. Kabat. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):3245–3250, 2014. ISSN 00278424. 10.1073/pnas.1222460110.
- J. Seibert, M. J. Vis, E. Lewis, and H. J. van Meerveld. Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, 32(8):1120–1125, 2018. ISSN 10991085. 10.1002/hyp.11476.
- H. Sevestre and D. I. Benn. Climatic and geometric controls on the global distribution of surge-type glaciers: implications for a unifying model of surging. *Journal of Glaciology*, 61(228):646–662, 2015.
- G. A. Sextstone, J. M. Driscoll, L. E. Hay, J. C. Hammond, and T. B. Barnhart. Runoff sensitivity to snow depletion curve representation within a continental scale hydrologic model. *Hydrological Processes*, 34(11): 2365–2380, 2020. ISSN 10991085. 10.1002/hyp.13735.
- A. Sood and V. Smakhtin. Revue des modèles hydrologiques globaux. *Hydrological Sciences Journal*, 60 (4):549–565, 2015. ISSN 21503435. 10.1080/02626667.2014.950580. URL <http://dx.doi.org/10.1080/02626667.2014.950580>.
- E. H. Sutanudjaja, R. Van Beek, N. Wanders, Y. Wada, J. H. Bosmans, N. Drost, R. J. Van Der Ent, I. E. De Graaf, J. M. Hoch, K. De Jong, D. Karssenber, P. López López, S. Peřkenteiner, O. Schmitz, M. W. Straatsma, E. Vannamettee, D. Wisser, and M. F. Bierkens. PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, 11(6):2429–2453, 2018. ISSN 19919603. 10.5194/gmd-11-2429-2018.
- K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498, 2012. ISSN 00030007. 10.1175/BAMS-D-11-00094.1.

- L. P. Van Beek, Y. Wada, and M. F. Bierkens. Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, 47(7), 2011. ISSN 00431397. 10.1029/2010WR009791.
- G. van den Oord, S. Verhoeven, I. Pelupessy, J. Aerts, M. de Vos, B. Weel, M. van Meersbergen, R. van Haren, Y. Dzigan, B. van Werkhoven, et al. Grpc4bmi: Running earth system models as remote services. In *Geophysical Research Abstracts*, volume 21, 2019.
- A. I. Van Dijk, L. J. Renzullo, Y. Wada, and P. Tregoning. A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrology and Earth System Sciences*, 18(8):2955–2973, 2014. ISSN 16077938. 10.5194/hess-18-2955-2014.
- M. H. Van Huijgevoort, H. A. Van Lanen, A. J. Teuling, and R. Uijlenhoet. Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge. *Journal of Hydrology*, 512:421–434, 2014. ISSN 00221694. 10.1016/j.jhydrol.2014.02.060. URL <http://dx.doi.org/10.1016/j.jhydrol.2014.02.060>.
- M. van Tiel, K. Stahl, D. Freudiger, and J. Seibert. Glacio-hydrological model calibration and evaluation. *Wiley Interdisciplinary Reviews: Water*, 7(6), 2020. ISSN 20491948. 10.1002/wat2.1483.
- K. L. Verdin and S. K. Greenlee. HYDRO1k documentation. *Sioux Falls, ND, US Geological Survey, EROS Data Center*, <http://edcdaac.usgs.gov/gtopo30/hydro/readme.html>, 1998.
- A. Vincent and J. Hart. Under the glacier, the groundwater—the case of Skálafell area, Iceland. *EGUGA*, page 2322, 2017.
- A. Vincent, S. Violette, and G. Aðalgeirsdóttir. Groundwater in catchments headed by temperate glaciers: A review. *Earth-Science Reviews*, 188(June 2018):59–76, 2019. ISSN 00128252. 10.1016/j.earscirev.2018.10.017. URL <https://doi.org/10.1016/j.earscirev.2018.10.017>.
- Y. Wada, L. P. H. Van Beek, C. M. Van Kempen, J. W. T. M. Reckman, S. Vasak, and M. F. P. Bierkens. Global depletion of groundwater resources. *Geophysical research letters*, 37(20), 2010.
- Y. Wada, D. Wisser, S. Eisner, M. Flörke, D. Gerten, I. Haddeland, N. Hanasaki, Y. Masaki, F. T. Portmann, T. Stacke, Z. Tessler, and J. Schewe. Multimodel projections and uncertainties of irrigation water demand under climate change. *Geophysical Research Letters*, 40(17):4626–4632, 2013. ISSN 00948276. 10.1002/grl.50686.
- B. Wouters, A. S. Gardner, and G. Moholdt. Global glacier mass loss during the grace satellite mission (2002–2016). *Frontiers in earth science*, 7:96, 2019.
- H. Zekollari, M. Huss, and D. Farinotti. Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble. *Cryosphere*, 13(4):1125–1146, 2019. ISSN 19940424. 10.5194/tc-13-1125-2019.
- H. Zekollari, M. Huss, and D. Farinotti. On the Imbalance and Response Time of Glaciers in the European Alps. *Geophysical Research Letters*, 47(2):1–9, 2020. ISSN 19448007. 10.1029/2019GL085578.
- M. Zemp, M. Huss, E. Thibert, N. Eckert, R. McNabb, J. Huber, M. Barandun, H. Machguth, S. U. Nussbaumer, I. Gärtner-Roer, L. Thomson, F. Paul, F. Maussion, S. Kutuzov, and J. G. Cogley. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature*, 568(7752):382–386, 2019. ISSN 14764687. 10.1038/s41586-019-1071-0.

Appendix A Basin information

Basin name	Center lon.	Center lat.	Glaciation degree (%)	GRDC station no.	GRDC station name	Obs. start	Obs. End
Alsek	-137	60	19.76	4102050	Near Yakutat	2000	2012
Amazon	-64	-6	0.03	3629001	Obidos - Linigrafo	2000	2007
Clutha	169	-45	0.31	5868050	Clyde	2000	2008
Columbia	-116	46	0.28	4215210	International Boundary (Canada)	2000	2012
Copper	-143	61	20.01	4102710	Million Dollar Bridge Near Cordova, Ak.	2003	2011
Danube	18	46	0.05	6742201	Bazias	2000	2007
Dramselv	9	61	0.19	6731310	Dovikfoss	2000	2012
Fraser	-122	52	1.04	4207900	Hope	2000	2012
Gloma	11	61	0.63	6731403	Solbergfoss	2000	2012
Irrawaddy	96	23	0.02	2260400	Katha	2000	2009
Joekulsa	-16	65	15.03	6401702	Grimsstadir	2000	2012
Kalixaelven	22	67	0.22	6233850	Raektfors	2000	2012
Kuskokwim	-156	61	0.87	4102100	Crooked Creek, Alas.	2000	2012
Lule	18	67	0.98	6233750	Bodens Krv (+ Vattenverk, Trangfors)	2002	2011
Mackenzie	-120	61	0.09	4208025	Arctic Red River	2000	2012
Nass	-129	56	6.3	4206100	Above Shumal Creek	2000	2012
Negro	-68	-39	0.05	3275990	Primera Angostura	2000	2012
Nelson	-101	51	0.03	4213711	Long Spruce Generating Station	2000	2012
Ob	75	55	0.03	2912600	Salekhard	2000	2009
Oelfusa	-21	64	12.04	6401090	Selfoss	2000	2012
Rhine	7	49	0.15	6935051	Basel, Rheinhalle	2000	2012
Rhone	5	45	0.93	-	Beaucaire/Tarrascon	2000	2012
Santa_Cruz	-73	-50	9.89	3276800	Charles Fuhr	2000	2012
Skagit	-121	49	2	4145080	Near Mount Vernon, Wa	2000	2012
Skeena	-127	55	1.73	4206250	Usk	2000	2012
Stikine	-131	57	6.78	4204900	Near Wrangell	2000	2012
Susitna	-149	62	8.7	4102820	Gold Creek, Ak	2000	2012
Taku	-132	58	0.41	4202601	Near Juneau	2000	2012
Thjorsa	-19	64	16.63	6401120	Krokur	2000	2012
Yukon	-144	65	1.15	4103200	Pilot Station, Ak	2000	2012

Table S1: Information on each basin and their corresponding observations. For the Rhone, data from the Hydrobanque was used instead of the GRDC as an exception. Only basins with more than 5 years of observation records between 2000 and 2012 were selected. The Colorado (Argentina) complied with these requirements but its GRDC observations were defected. See Appendix E for the five basins in this list assumed to be invalid for analysis.

Appendix B Aletsch glacier runoff calibration

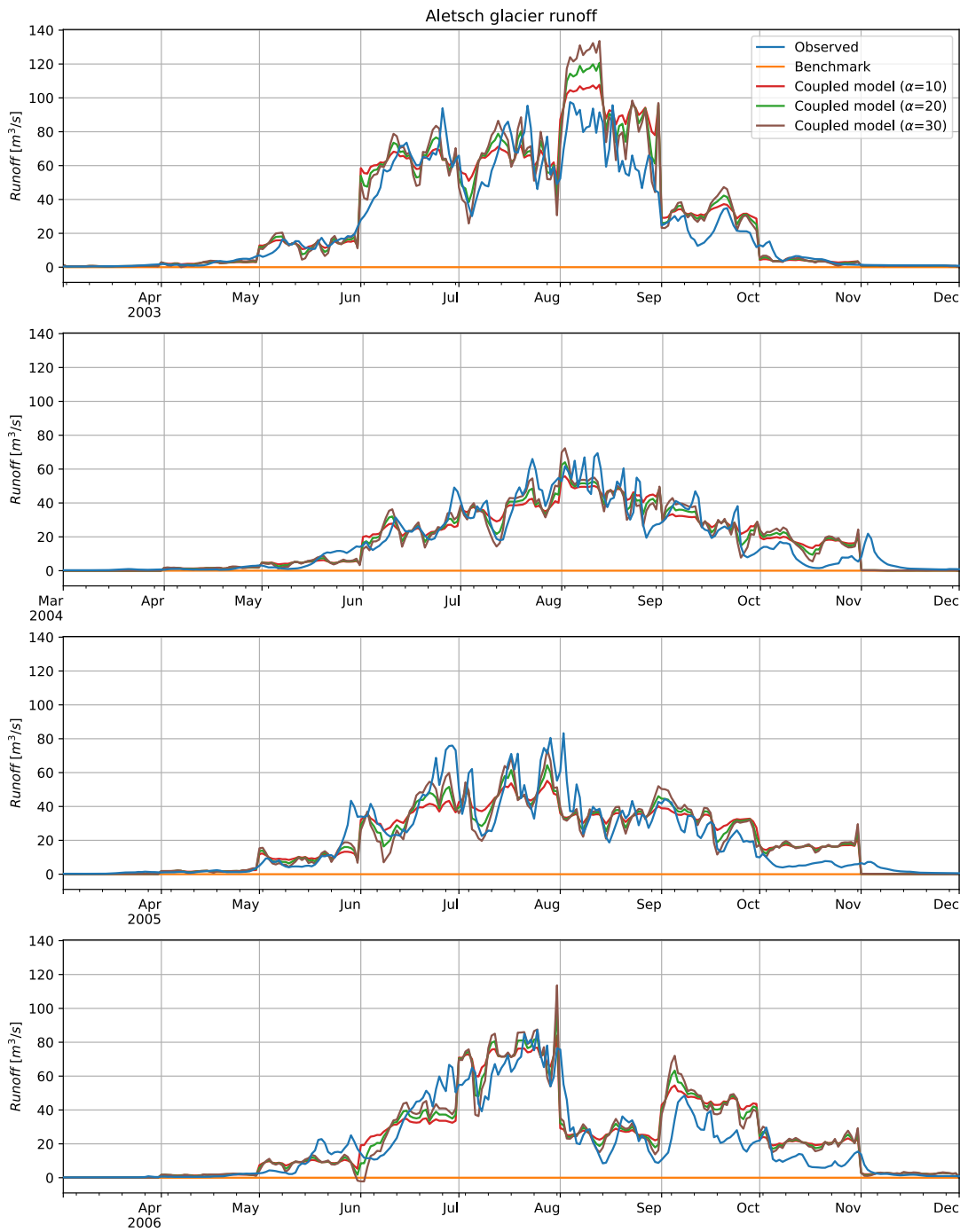


Figure S1: To determine the optimal weighting factor α to use in the weighting function (section 2.3), a simple calibration was performed on the runoff downstream of the Aletsch glacier (BAFU). An α -value of zero gives a monthly step-wise function, while a higher α -value leads to a higher sensitivity to daily temperature. An α -value of 20 produces the lowest RMSE over the 10 years considered and was therefore selected, although the sensitivity analysis suggested an α -value of 30 to possibly be better on the basin scale. The benchmark simulates a runoff of zero, which can likely be attributed to the formation of snow towers.

Appendix C Benchmark independent evaluation

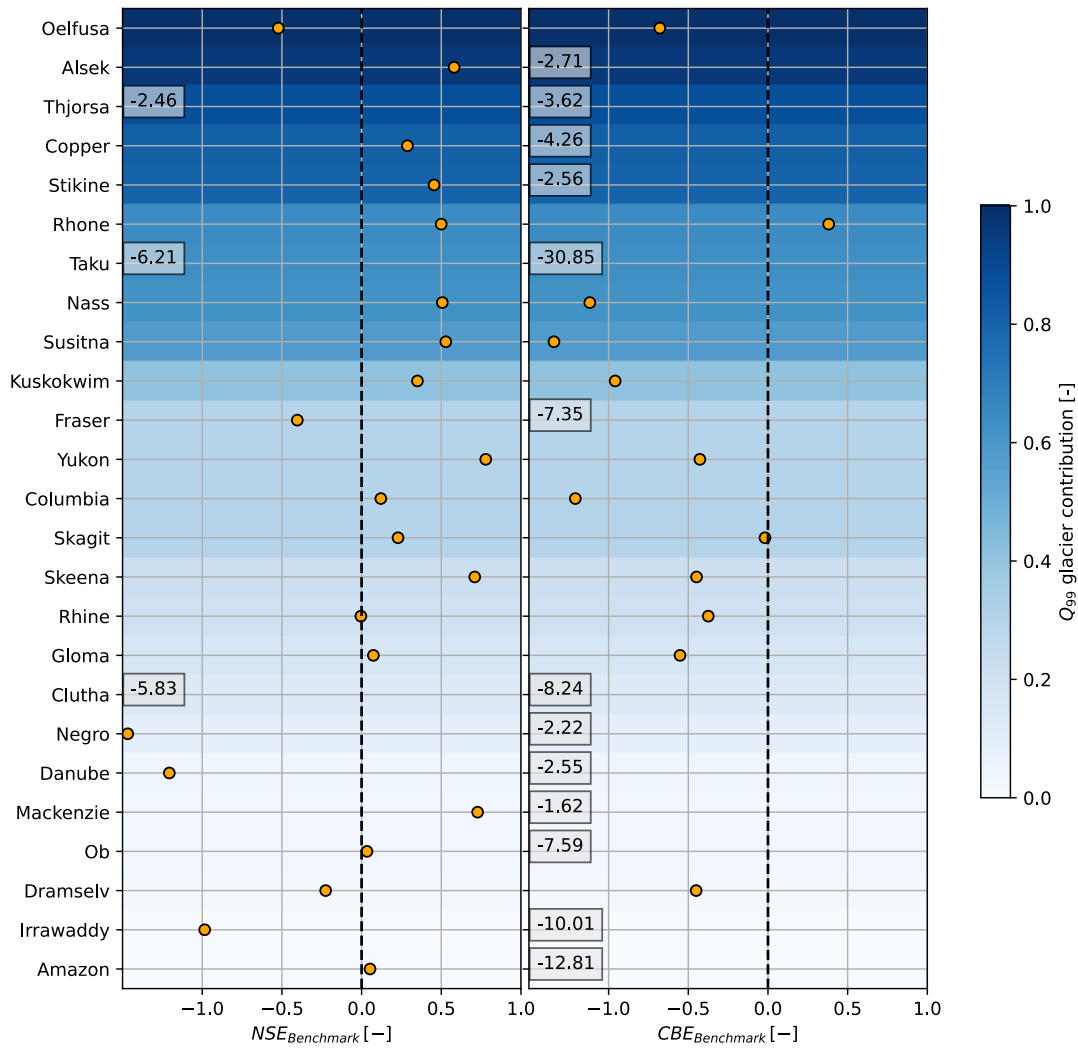


Figure S2: Evaluation of the benchmark (standard PCR-GLOBWB 2) against observations. The performance is expressed both in the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe (1970)) and calendar day benchmark efficiency (CBE, Schaeffli and Gupta (2007)). In seasonal runoff regimes the mean flow is a poor benchmark, resulting in high NSE-values. The CBE is more suitable for seasonal regimes and was therefore included in the evaluation. While in many basins the benchmark performs better than the mean flow, it only performs better than the mean flow of each calendar day in the Rhone.

Appendix D Overall metrics

In this appendix three other metrics are applied on each of the 25 basins over the whole time range for an overall evaluation. Additionally, an explanation is provided on the choice of the relative RMSE difference over these overall metrics.

Overall metrics

Since no significant runoff timing difference is involved between the benchmark and the coupled model, we only consider metrics evaluating the value differences. Firstly, a benchmark efficiency (BE) is applied as follows:

$$BE = 1 - \frac{\sum_{t=1}^N (Q_{Obs} - Q_{Coupled})^2}{\sum_{t=1}^N (Q_{Obs} - Q_{Benchmark})^2}$$

where Q_{obs} is the observed basin runoff as reported in the GRDC and N is the number of data points. With this metric, a value of 1 indicates perfect correlation with the observations and a value of 0 or lower indicates equal or worse performance compared to the benchmark respectively. This benchmark efficiency is similar to the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and to the benchmark efficiency defined by Schaeffli and Gupta (2007), but while those metrics need an artificial benchmark to compare the model against, the benchmark in this study is already present. Additionally, while its use would facilitate comparison with other studies given its widespread use in the hydrological modeling community, the NSE is a poor metric choice in highly seasonal flow regimes (Schaeffli and Gupta, 2007, van Tiel et al., 2020).

Secondly and thirdly, the flow-duration benchmark efficiency (FDBE) and total flow benchmark efficiency (TFBE) are applied analogous to the above-defined BE:

$$FDBE = 1 - \frac{\sum_{t=1}^N (FD_{Obs} - FD_{Coupled})^2}{\sum_{t=1}^N (FD_{Obs} - FD_{Benchmark})^2}$$
$$TFBE = 1 - \frac{|\sum_{t=1}^N Q_{Obs} - \sum_{t=1}^N Q_{Coupled}|}{|\sum_{t=1}^N Q_{Obs} - \sum_{t=1}^N Q_{Benchmark}|}$$

in which FD represents the flow-duration curve. These metrics assess the ability of the model to reproduce the flow regime and the total basin runoff of the observations as compared to the benchmark.

Results

For all three BE's a low sensitivity can be observed in lowly glaciated basins, owed to the limited influence the glacial runoff has on the total runoff, flow-duration curve and total flow respectively. The opposite is true for highly glaciated basins (see Figure S3). For both the BE and the FDBE, around half of the basins score positively, with seemingly no correlation to glaciation degree. For TFBE only a small number of basins score positively, since PCR-GLOBWB 2 often already overestimates the basin runoff in many cases and the additional basin runoff in the coupled model only exacerbates this.

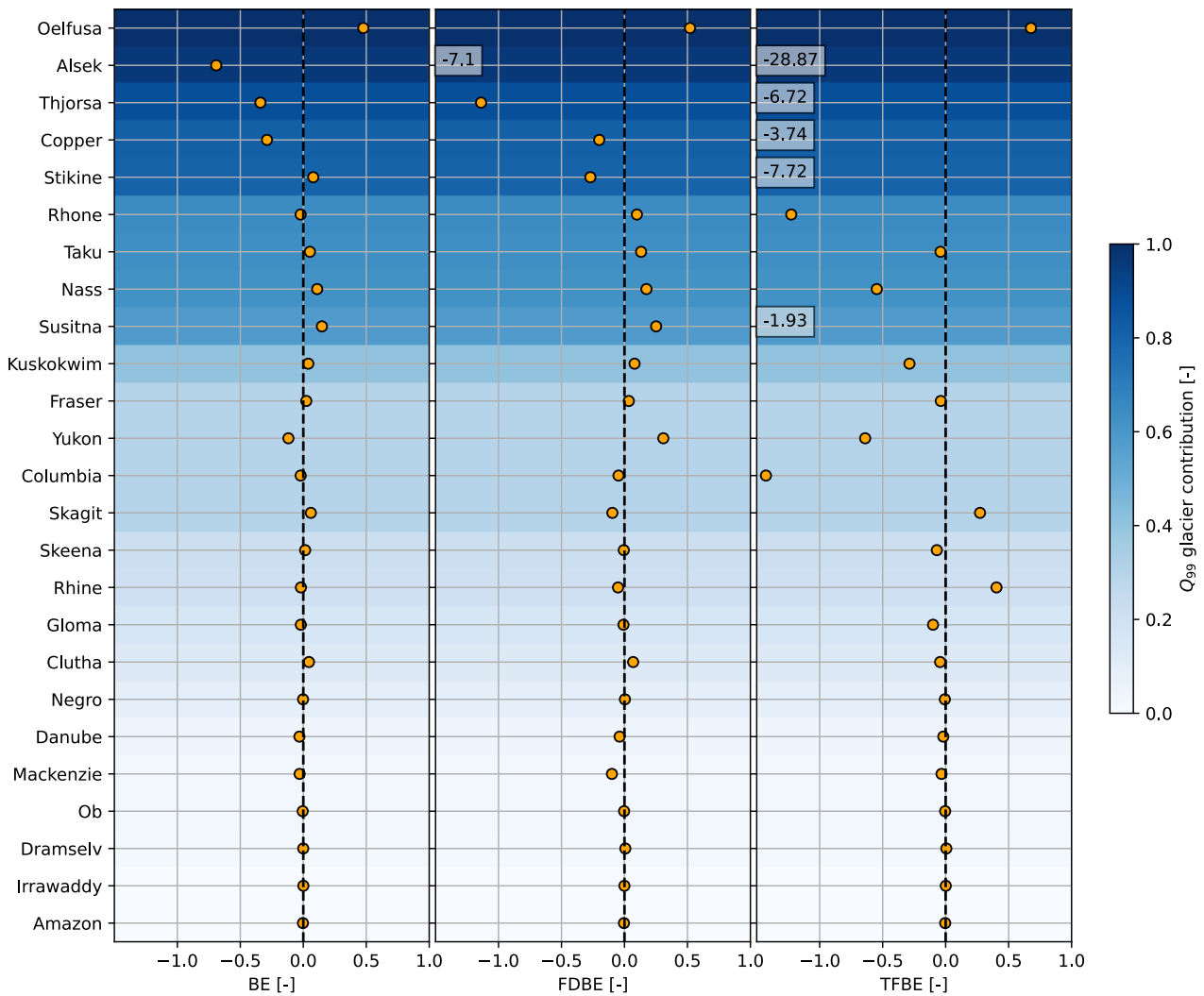


Figure S3: Results of the three overall metrics: the benchmark efficiency (BE), the flow-duration benchmark efficiency (FD BE) and the total flow benchmark efficiency (TF BE). The basins are sorted based on the 99th quantile of the contribution of the routed GloGEM glacier runoff to the coupled model runoff. A value of 1 indicates perfect correlation of the coupled model with the observations, while a negative value indicates a lesser performance compared to the benchmark.

Metric choice

While the above-mentioned metrics provide a good overall evaluation, their capacity of interpreting the data is limited for three reasons. Firstly, since the melt simulated by the models is highly seasonal, the difference between the models is also likely to be highly seasonal. It is therefore worth looking at the average monthly performance of the coupled model as compared to the benchmark, instead of only at the entire time range. Secondly, since the mean basin runoff and the fraction of glacial meltwater to basin runoff are different for each basin, an absolute error metric such as the BE does not allow for a fair comparison between the coupled model and the benchmark and between the different basins. In a lowly glaciated basins such as the Amazon, an additional error caused by the coupling of GloGEM will be only a fraction of the total error and will cause the BE-score to deviate only minimally from zero, and vice versa for highly glaciated basins such as the Oelfusa. Finally, the BE fails to express the performance change (+/-) between the coupled model and the benchmark relative to the maximum possible performance change (Seibert et al., 2018). In other words, the BE misses out on the fact that the same error decrement is worth more on a day with little melt than on a day with a high melt rate. The relative RMSE difference (RRD) introduced in section 3.4 meets these three criteria and was therefore chosen for this particular study. To avoid the introduction of an entirely new metric, the RMSE was used as a basis and the calculation was kept simple.

Appendix E Discarded basins

Although in total 30 basins with runoff observations data were found, 5 of them proved to be unsuited for further analysis. Firstly, the flow of the Kalix was routed upstream towards a bifurcation of the Torne that in reality forks into the Kalix. This is likely caused by the reduced DEM quality above the polar circle (Lehner et al., 2008). Secondly, the Santa Cruz, the Lule and the Nelson contain lakes upstream of the GRDC station, making any meaningful evaluation of daily streamflow impossible. Finally, the Joekulsa is simulated to contain an endless reservoir just downstream of its glaciers which fills up during the summer and drains slowly over the course of the winter and spring. The cause of this misrepresentation likely has to do with the routing module settings of PCR-GLOBWB. It should be noted that in other basins, such as the Copper and Skeena basins, this problem might be partly present as well but to a much lesser degree and they are therefore not excluded from analysis.

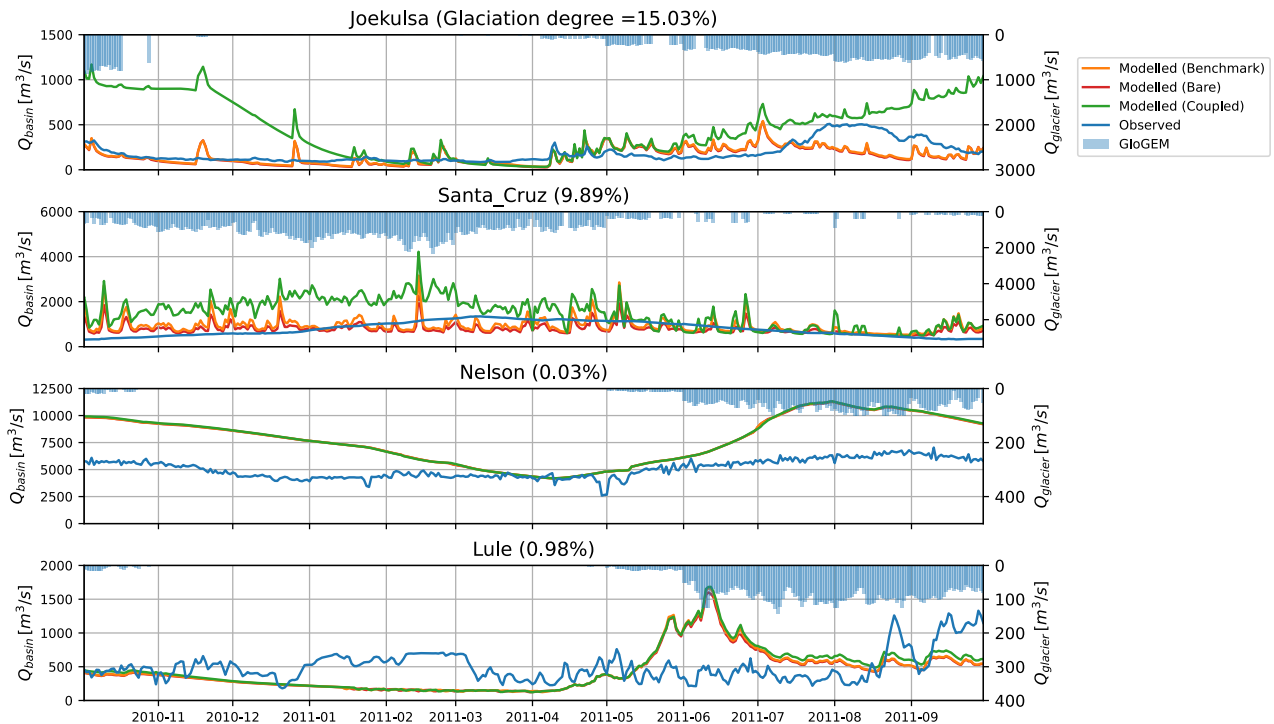


Figure S4: Four of the five discarded rivers: the Joekulsa due to routing problems, the remaining three due to large lakes in the river course. The Kalixa is not shown as the modelled discharge is simply zero, since all discharge is routed into a neighboring river.

Appendix F PCR-GLOBWB 2 SWE evolution

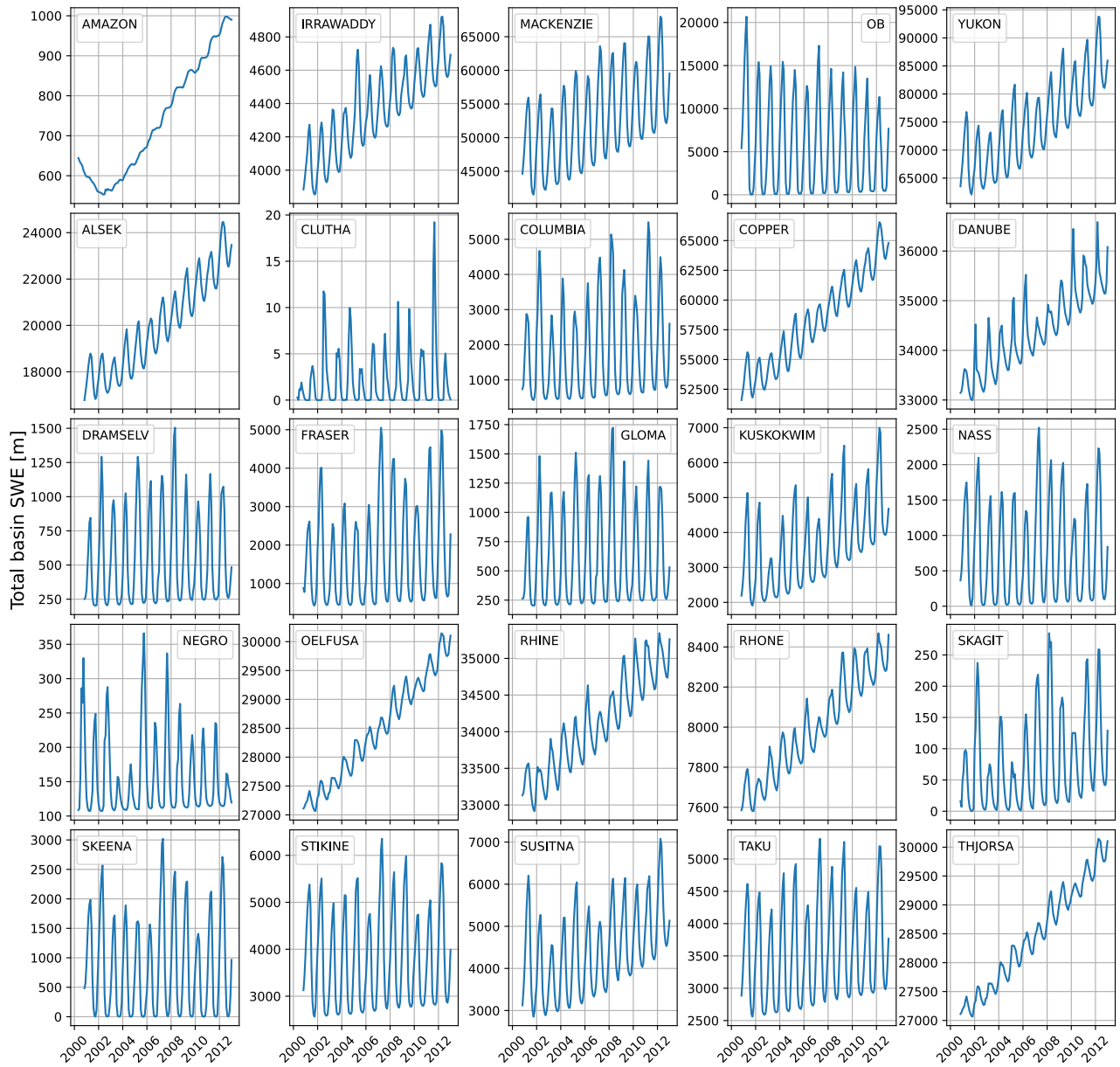
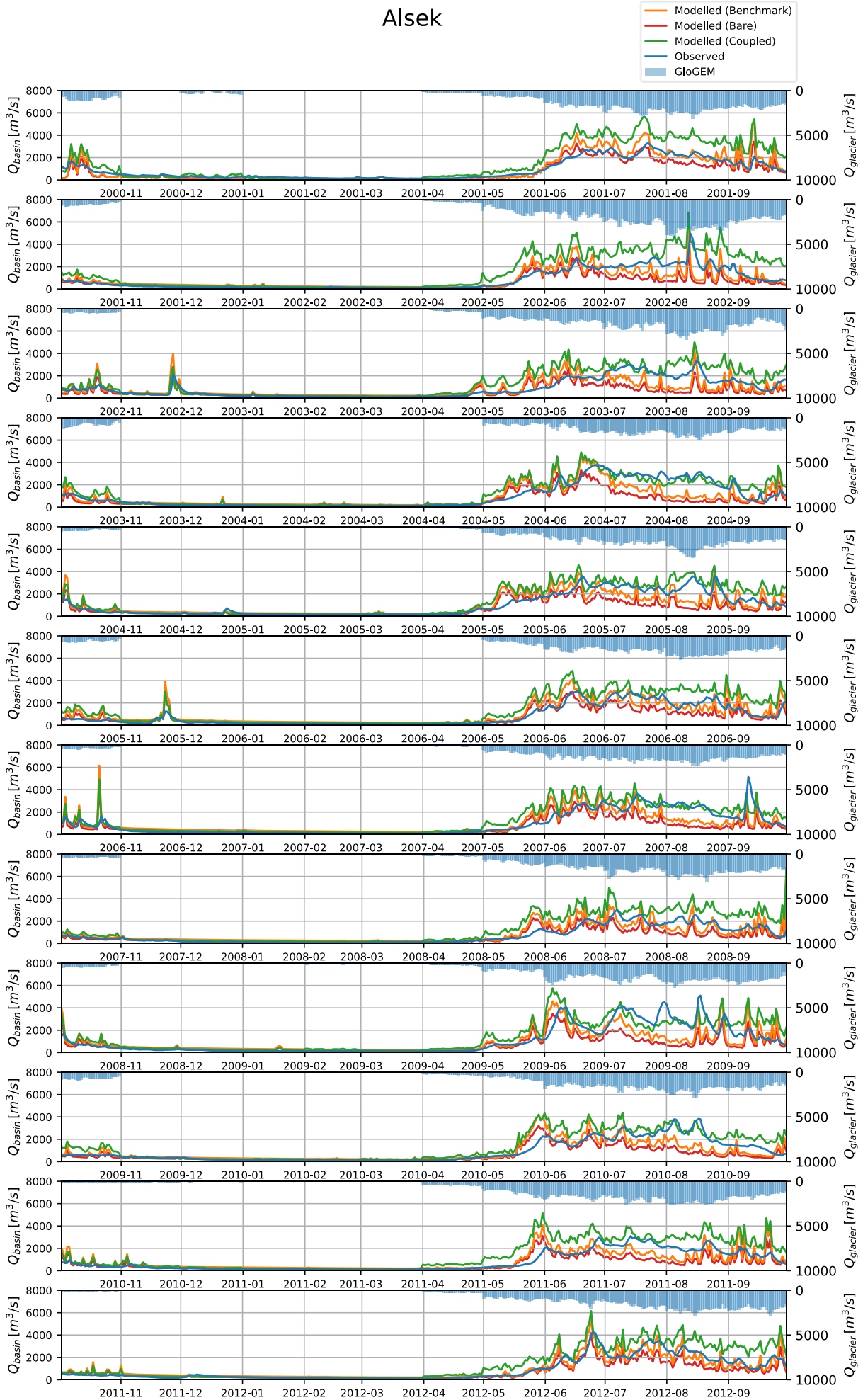


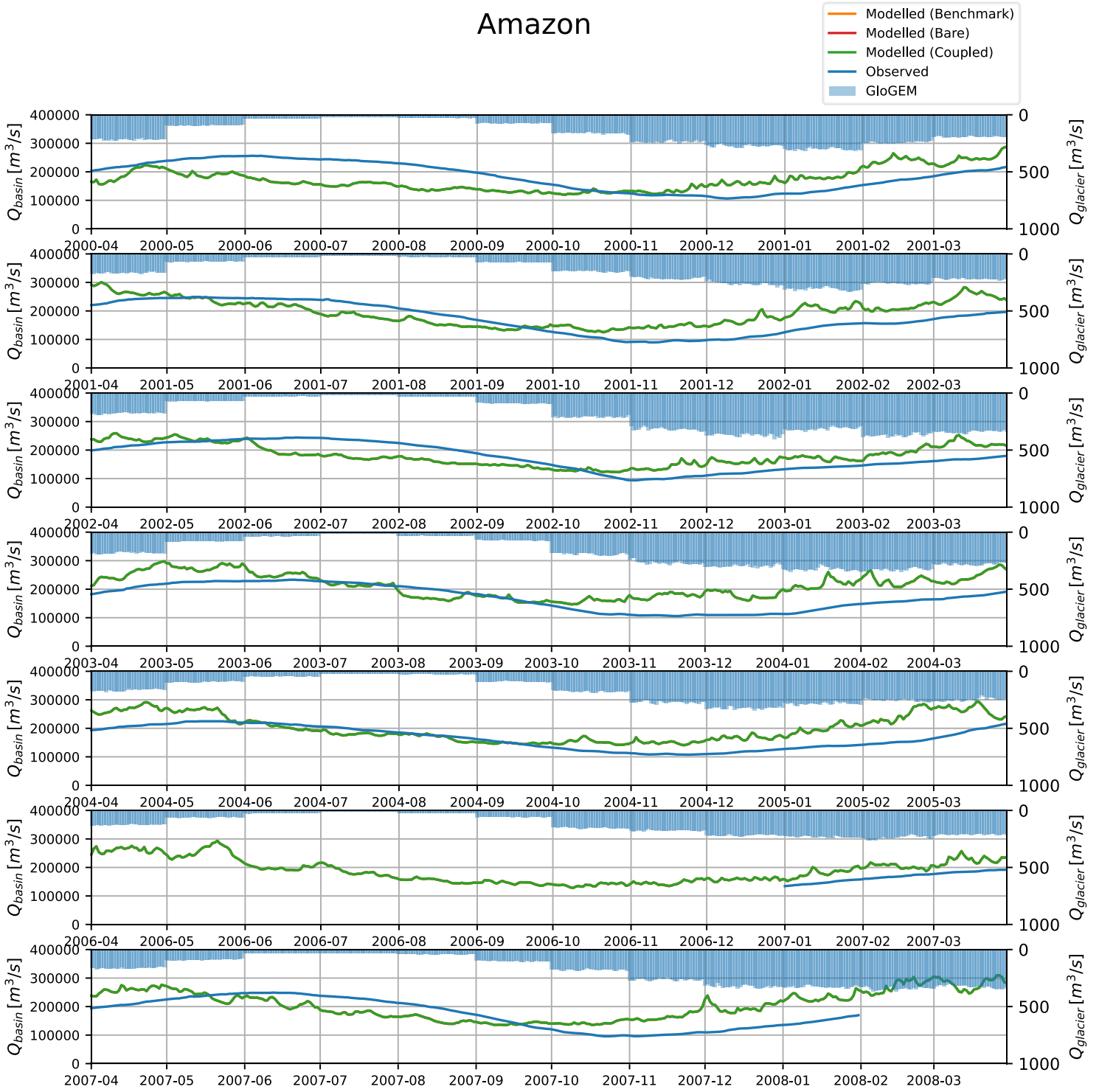
Figure S5: Basin-wide snow water equivalent (SWE) evolution. The multi-year accumulation of 'snow towers' due to a lacking snow redistribution representation in PCR-GLOBWB 2 is present in all basins with an increasing annual trend (16/25).

Appendix G Hydrographs

Alek

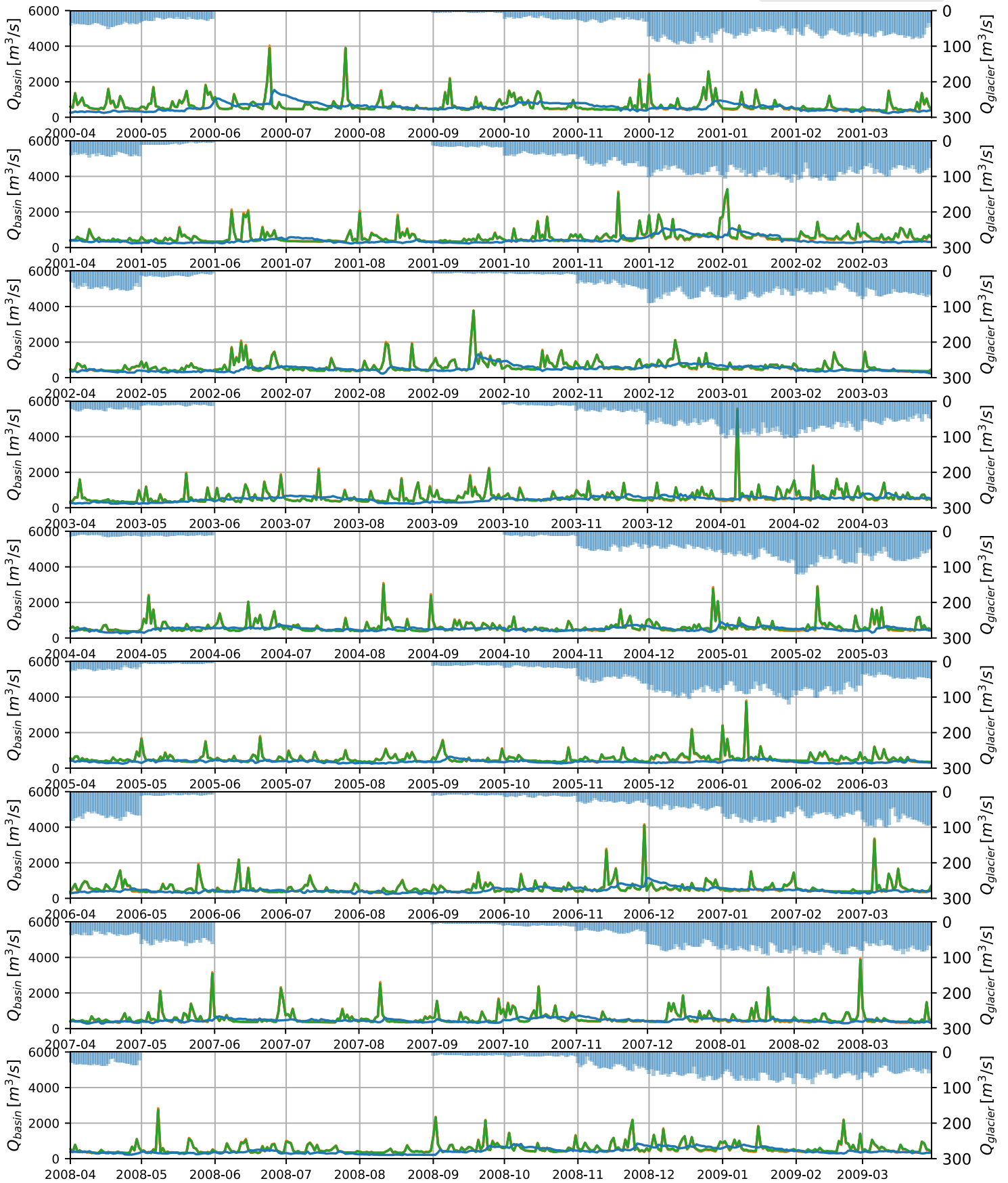


Amazon

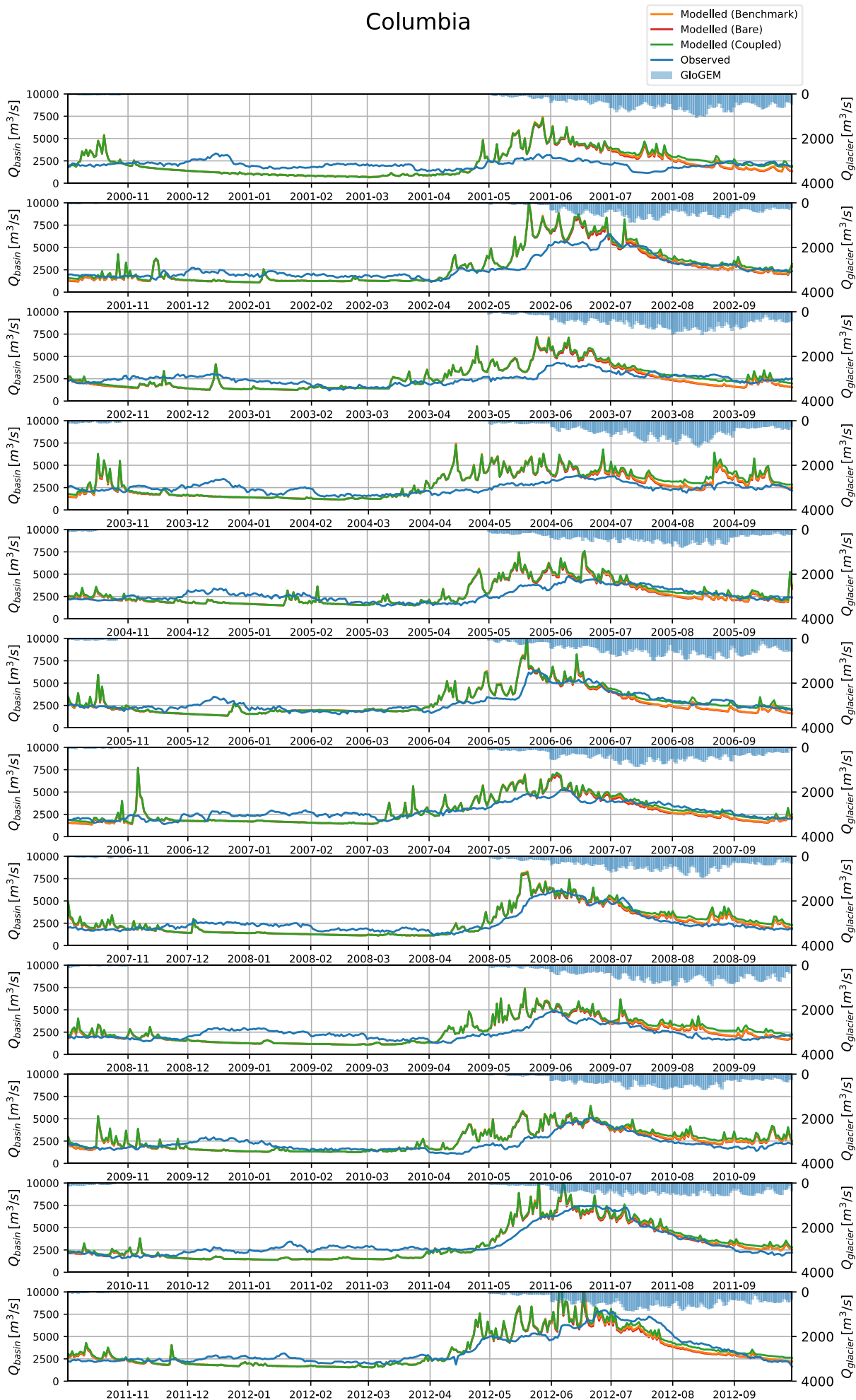


Clutha

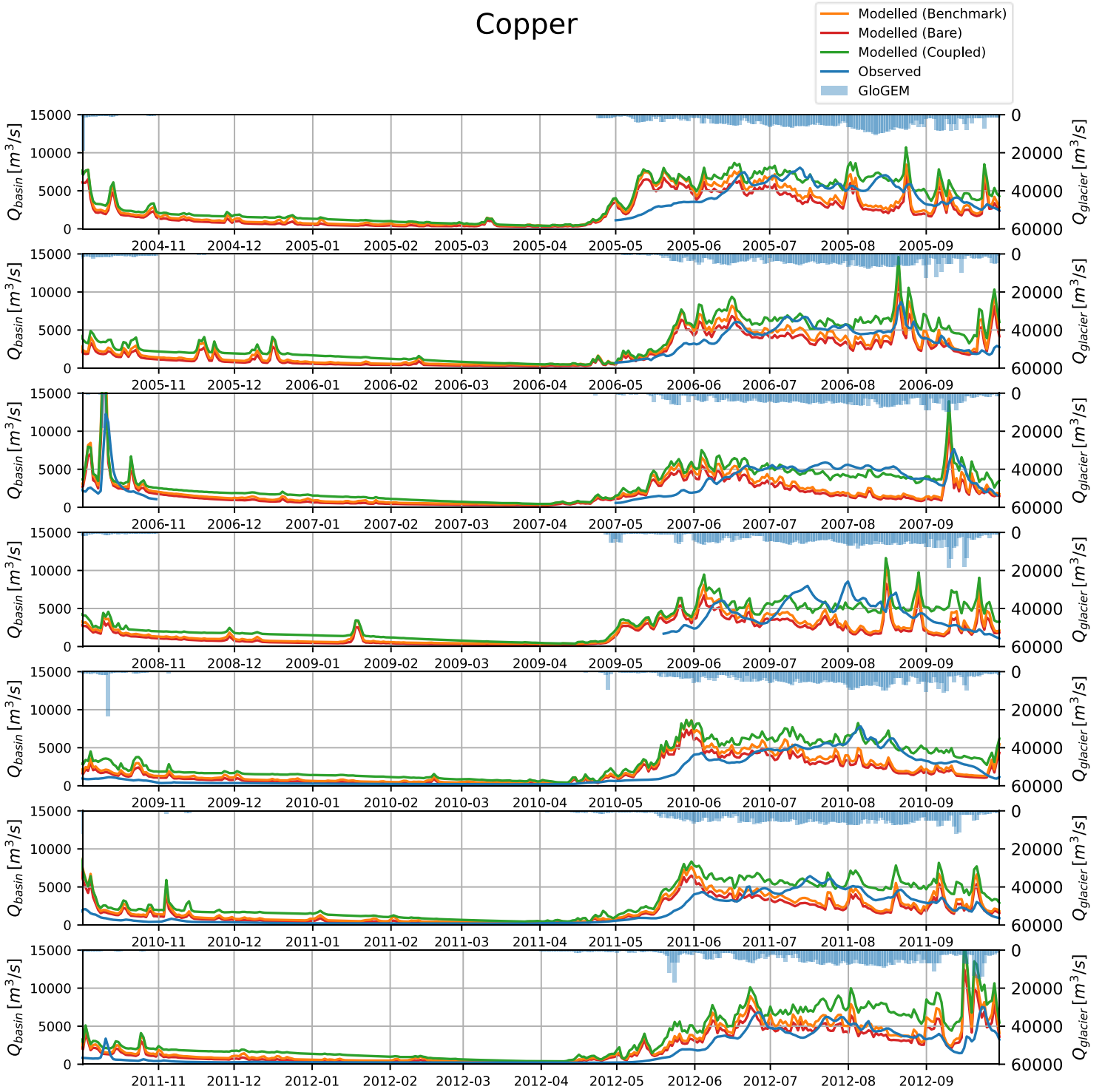
- Modelled (Benchmark)
- Modelled (Bare)
- Modelled (Coupled)
- Observed
- GloGEM



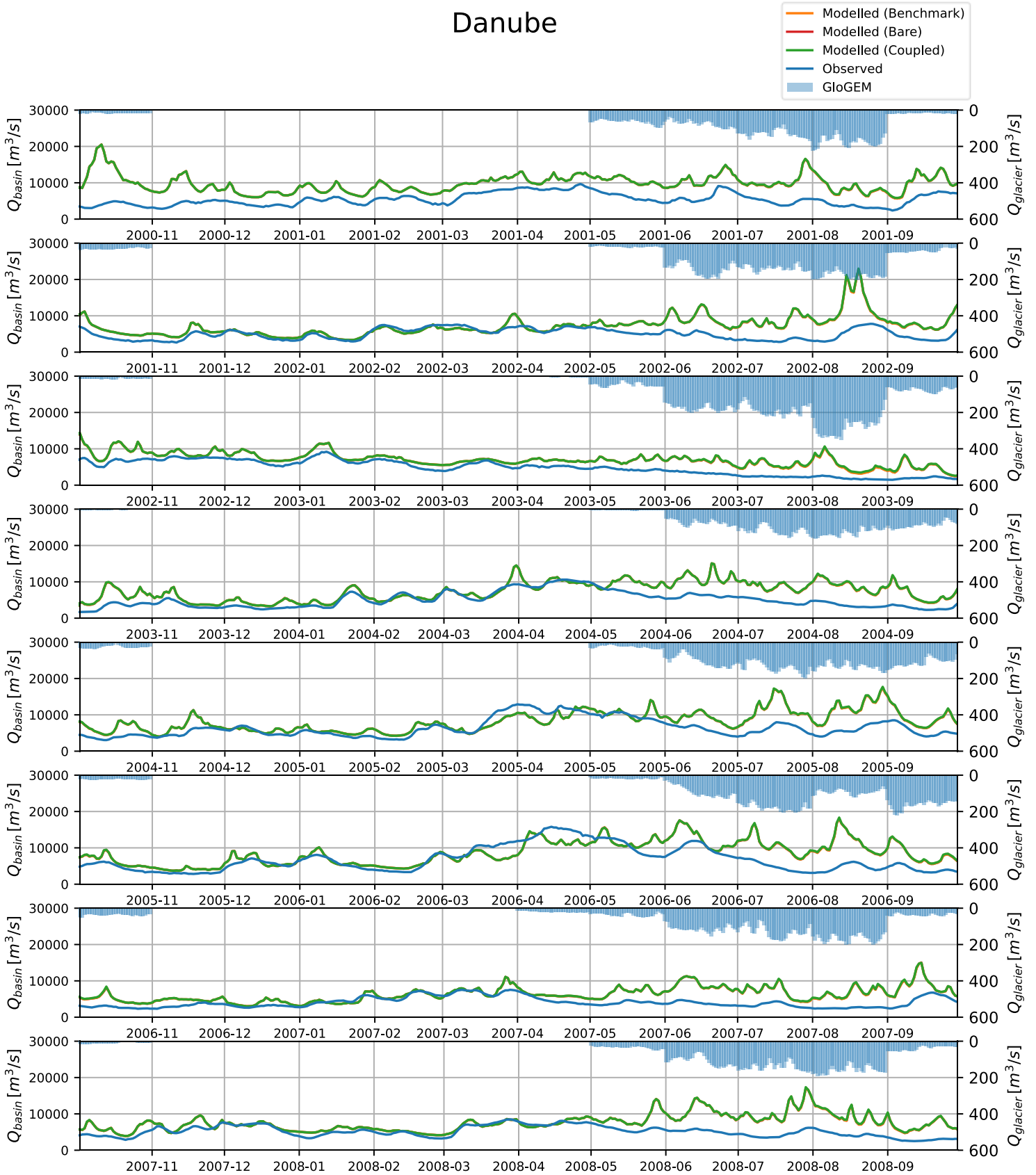
Columbia



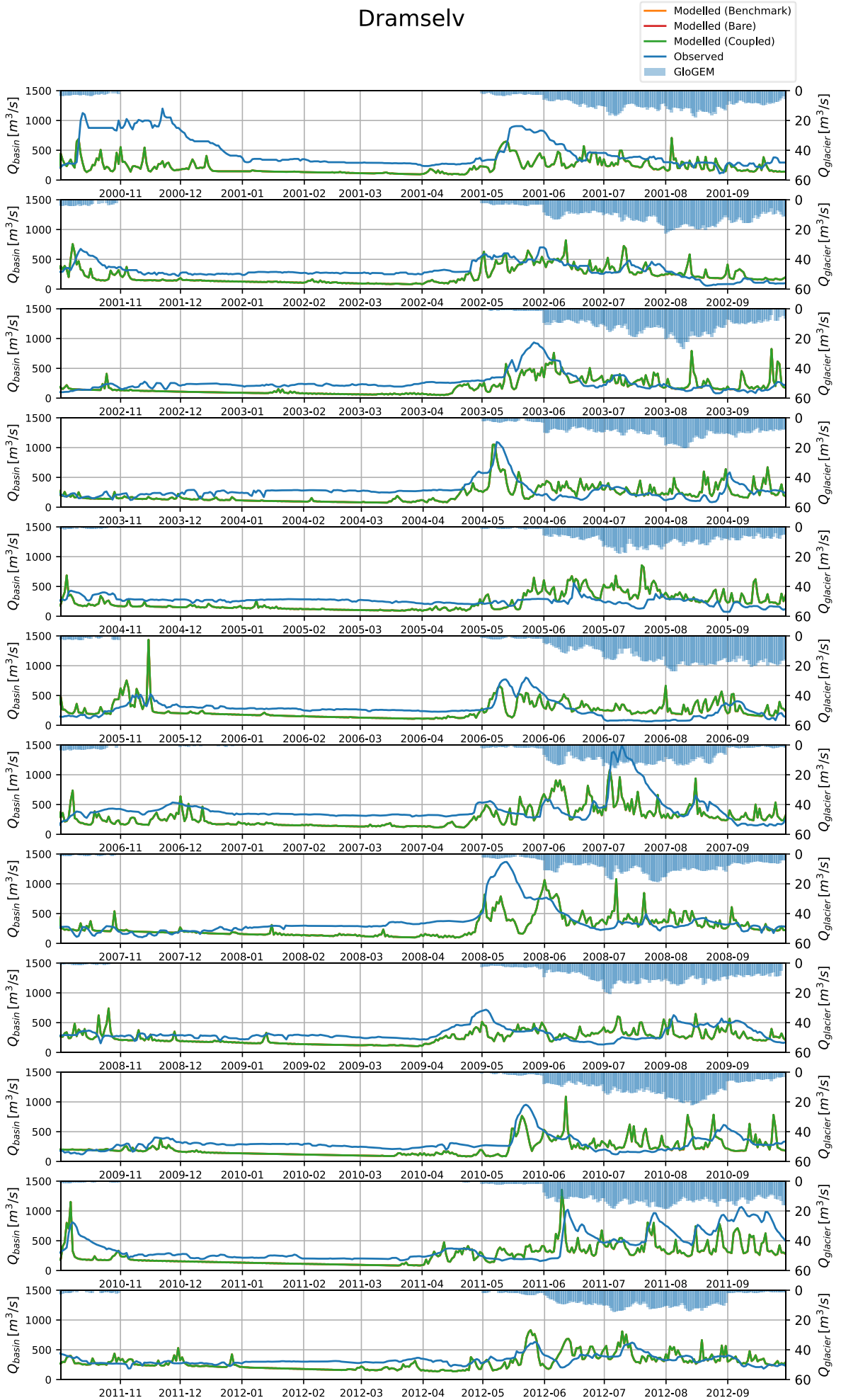
Copper



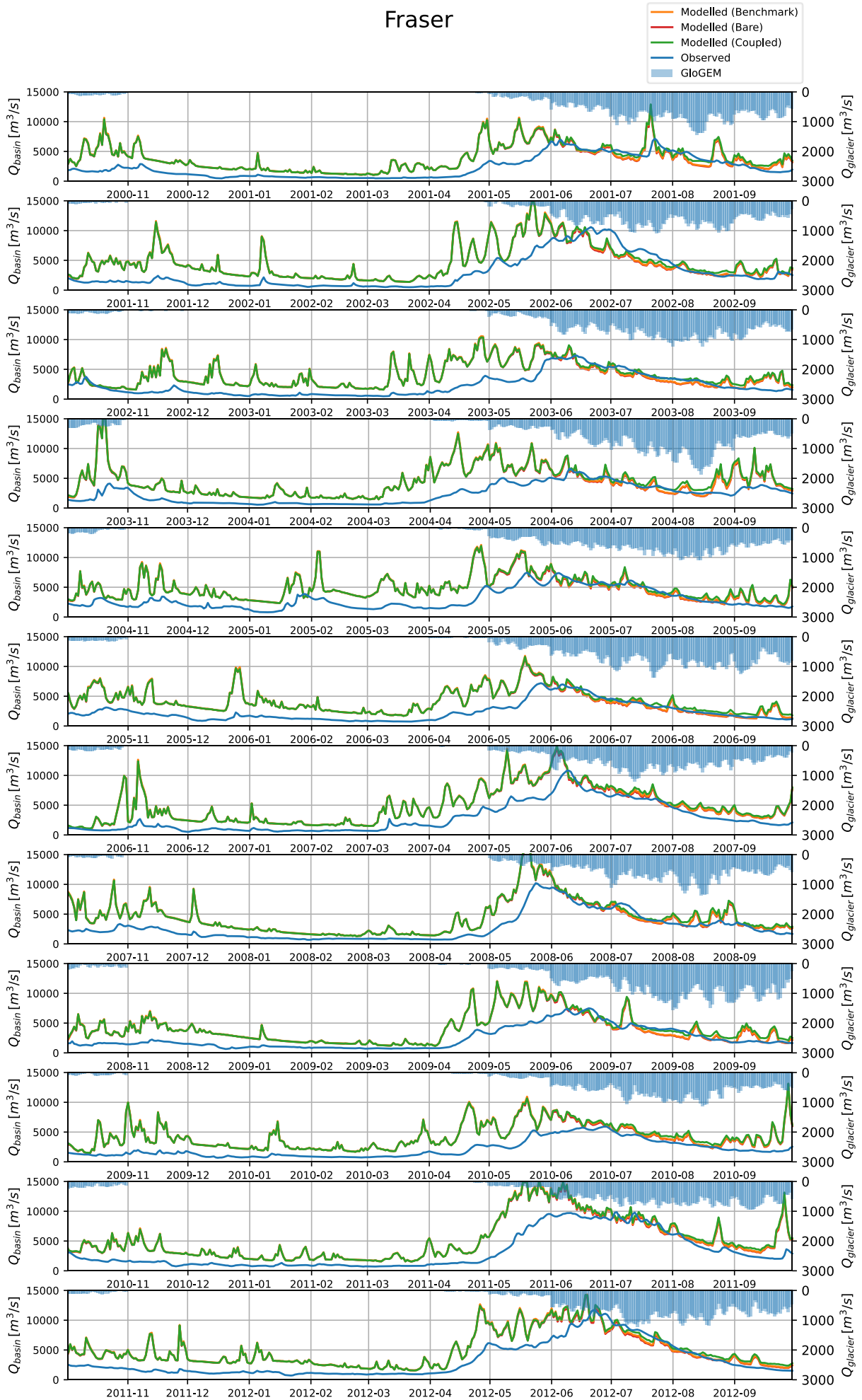
Danube



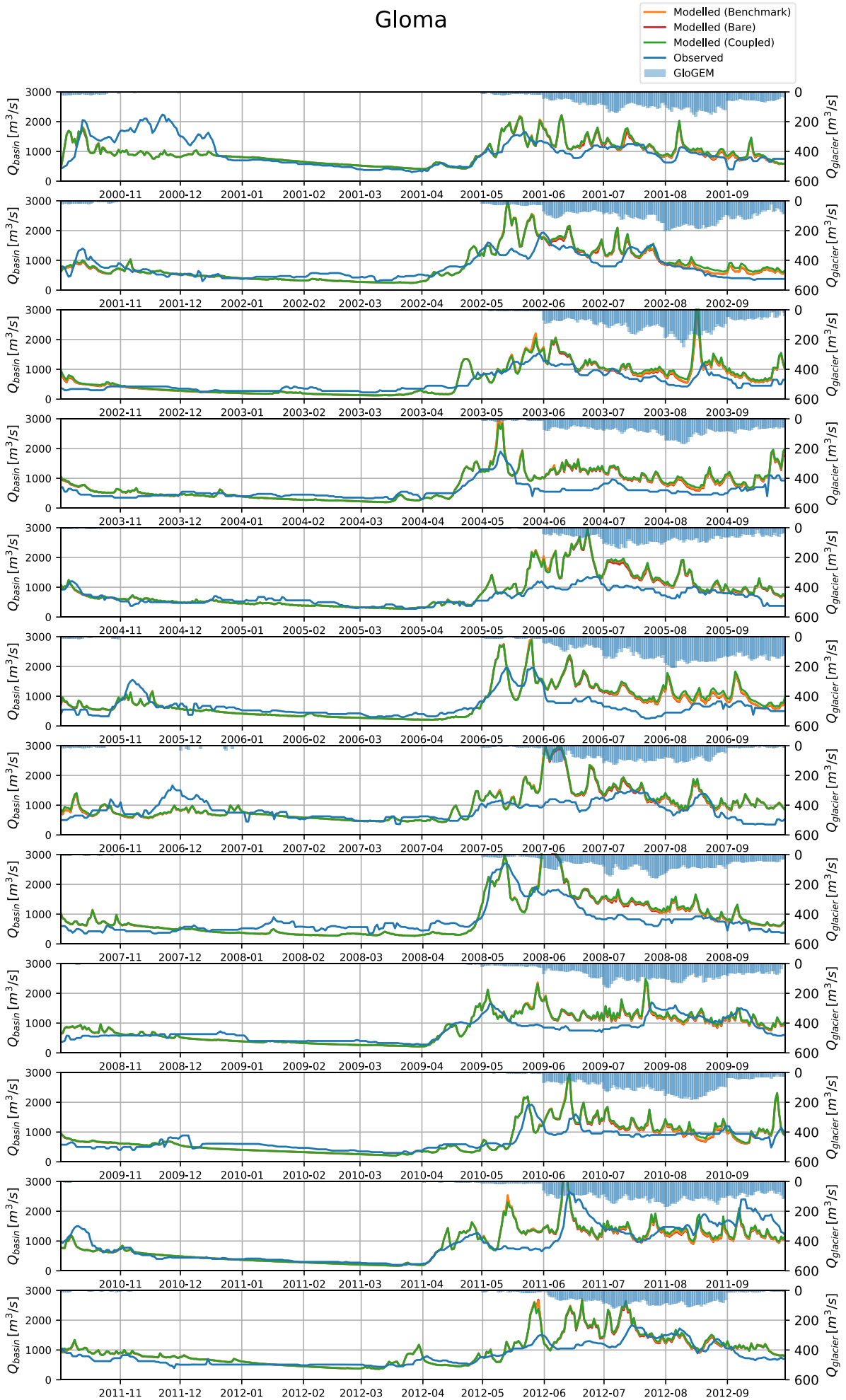
Dramselv



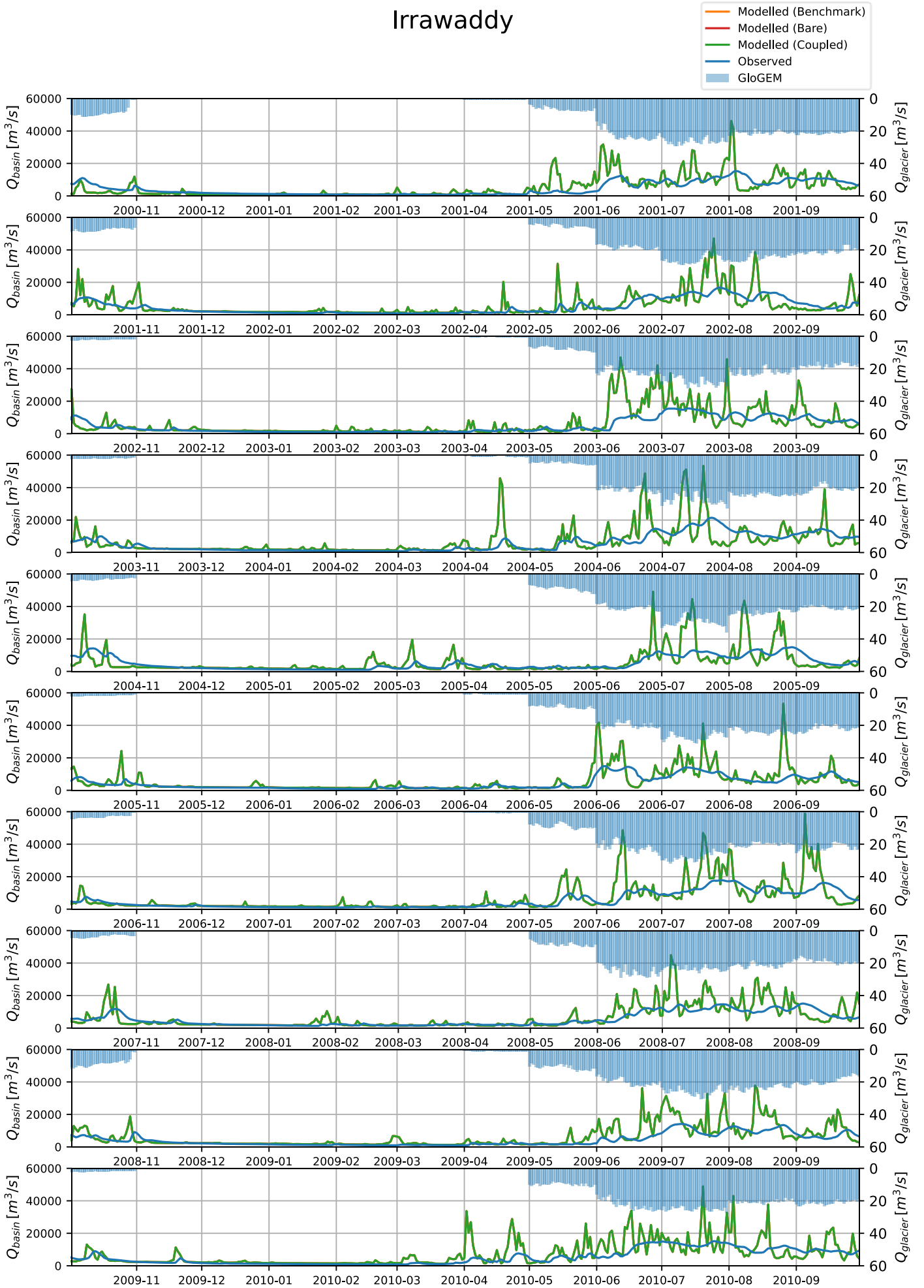
Fraser



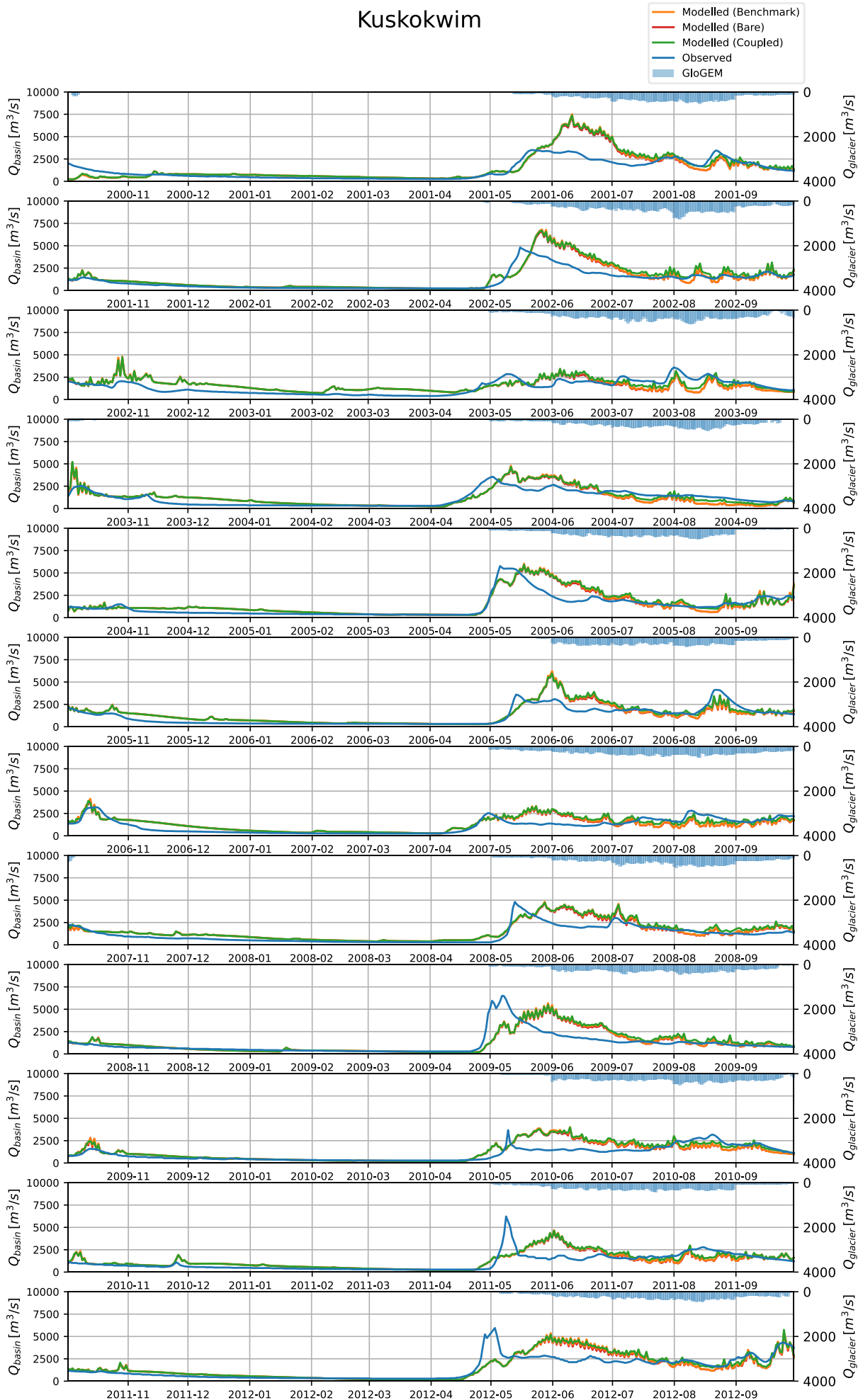
Gloma



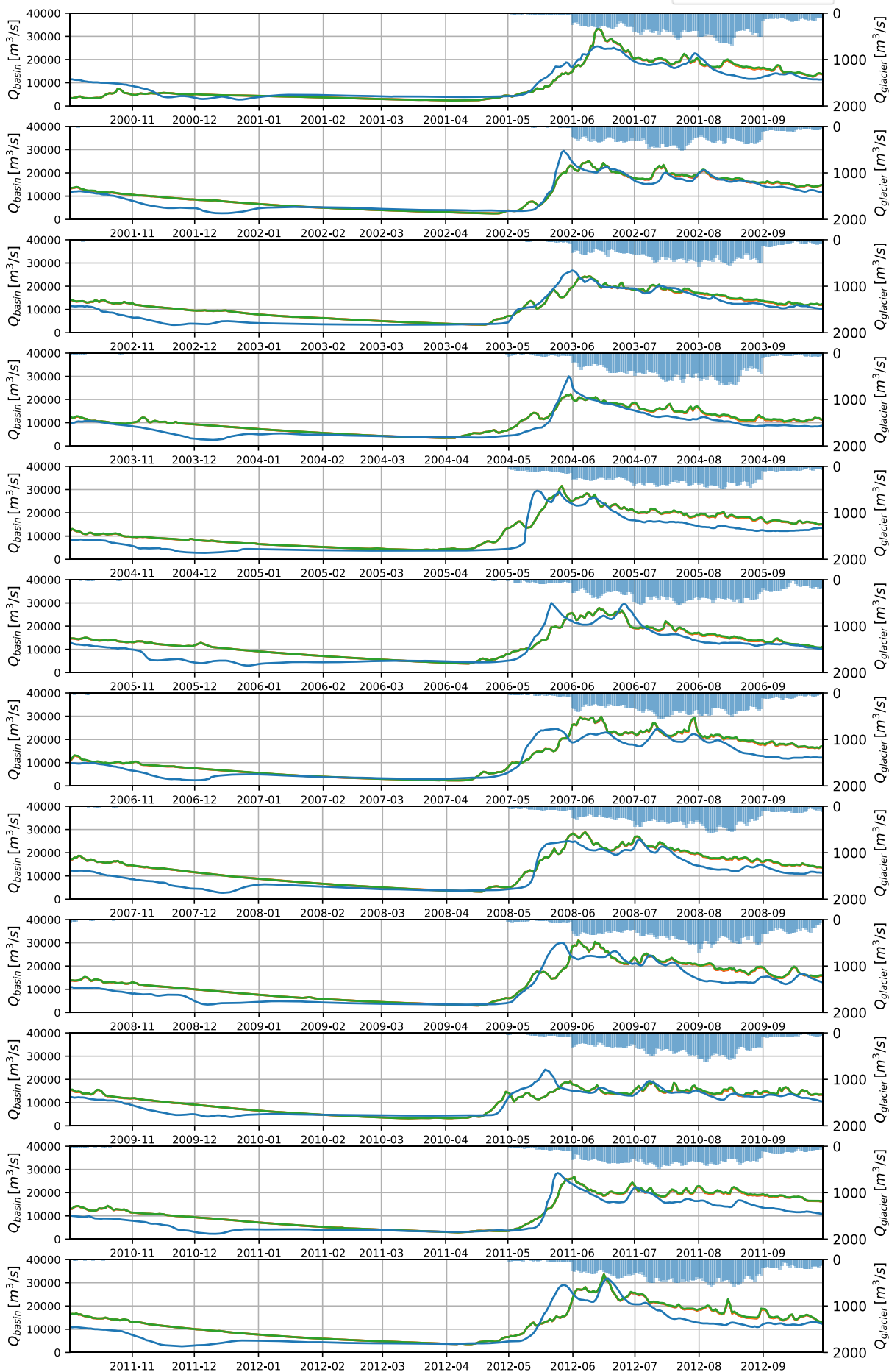
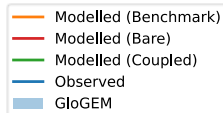
Irrawaddy



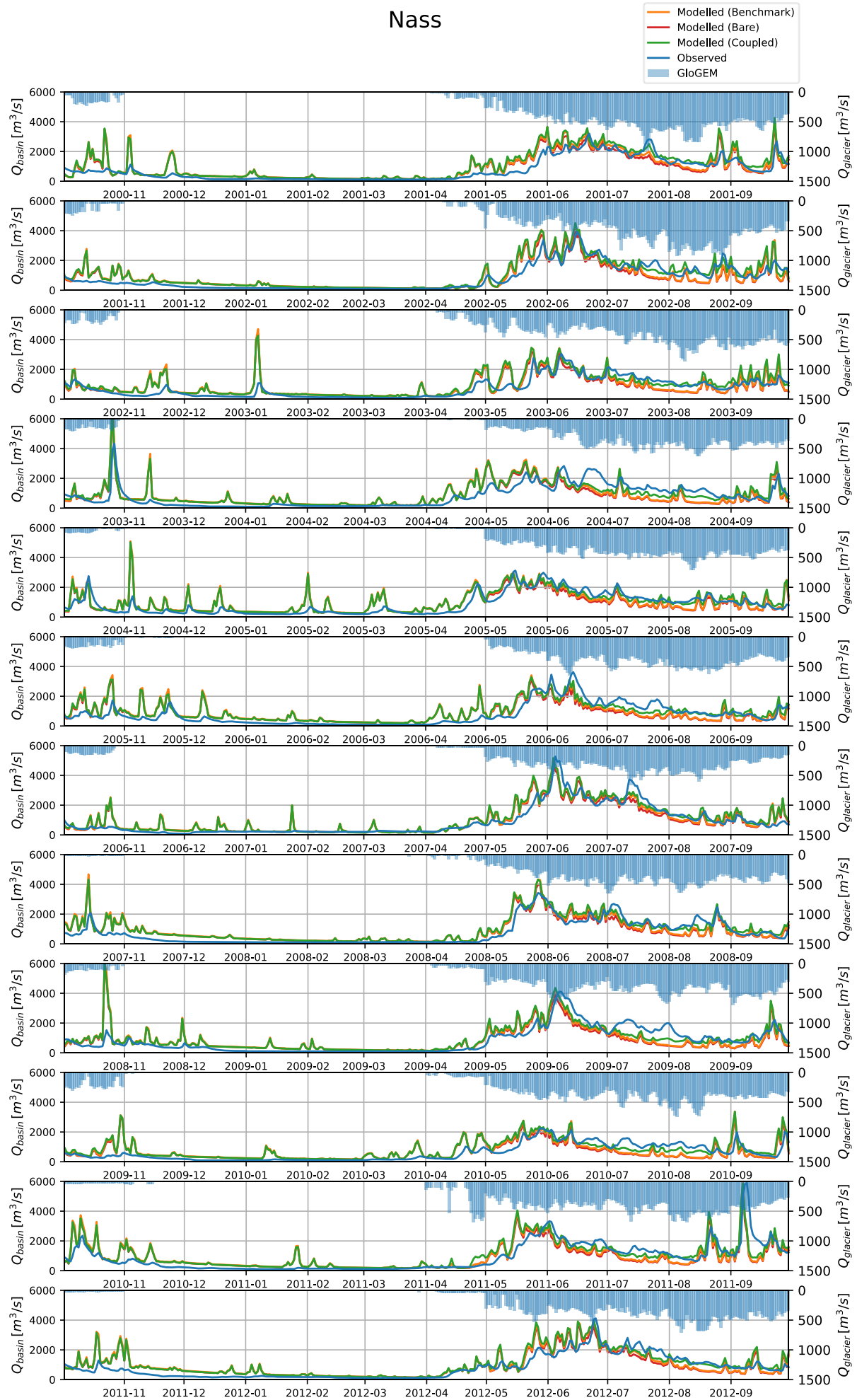
Kuskokwim



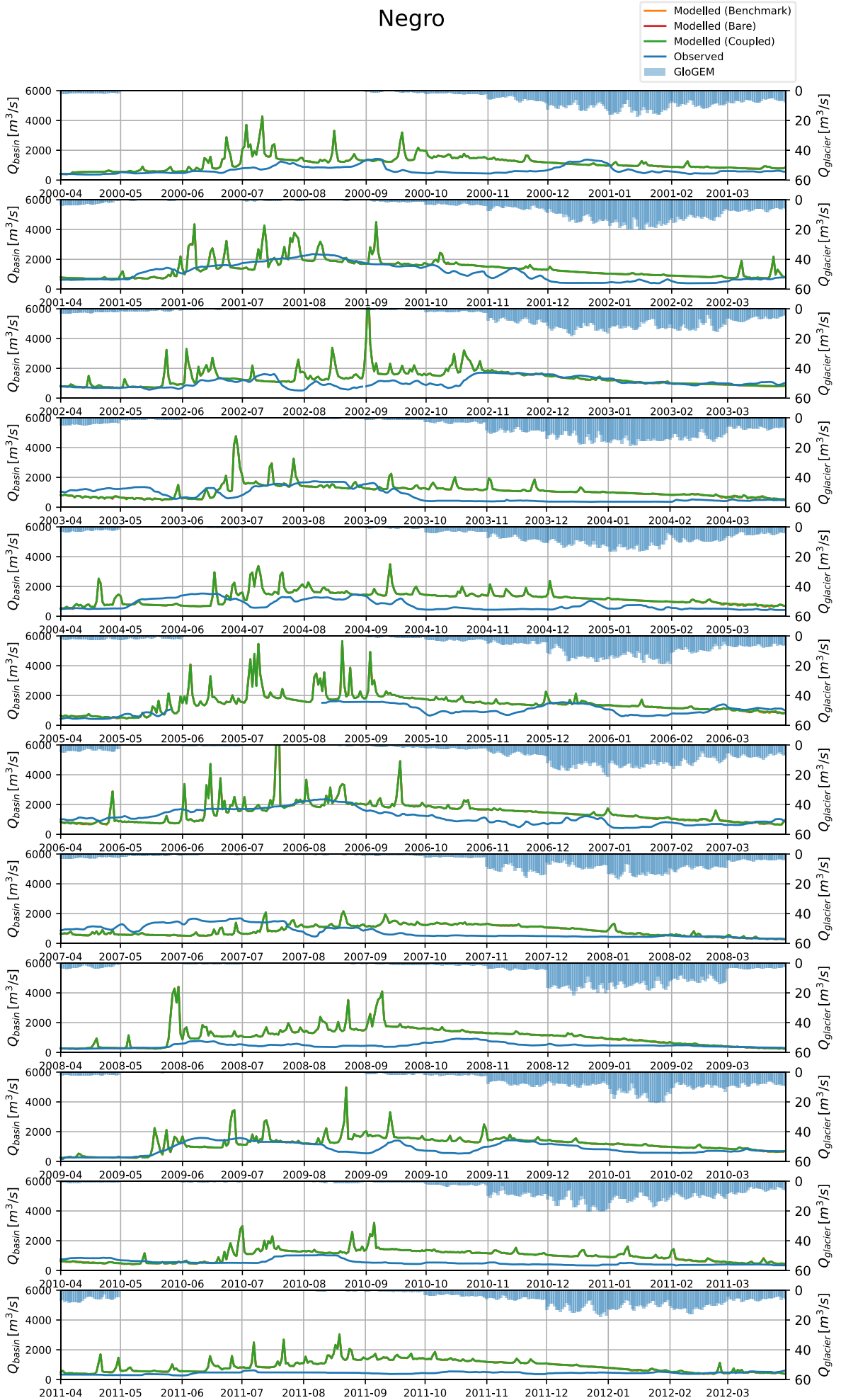
Mackenzie



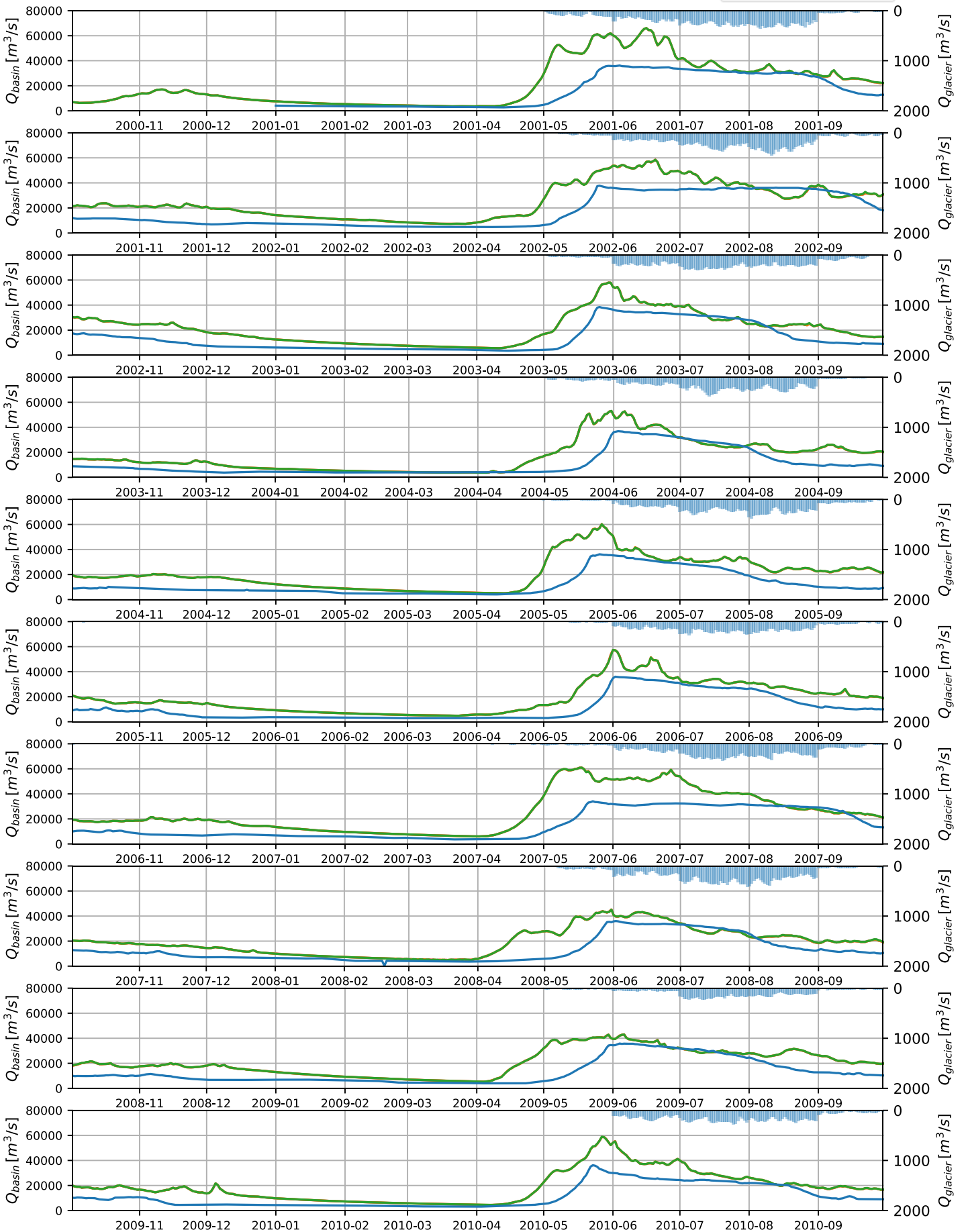
Nass



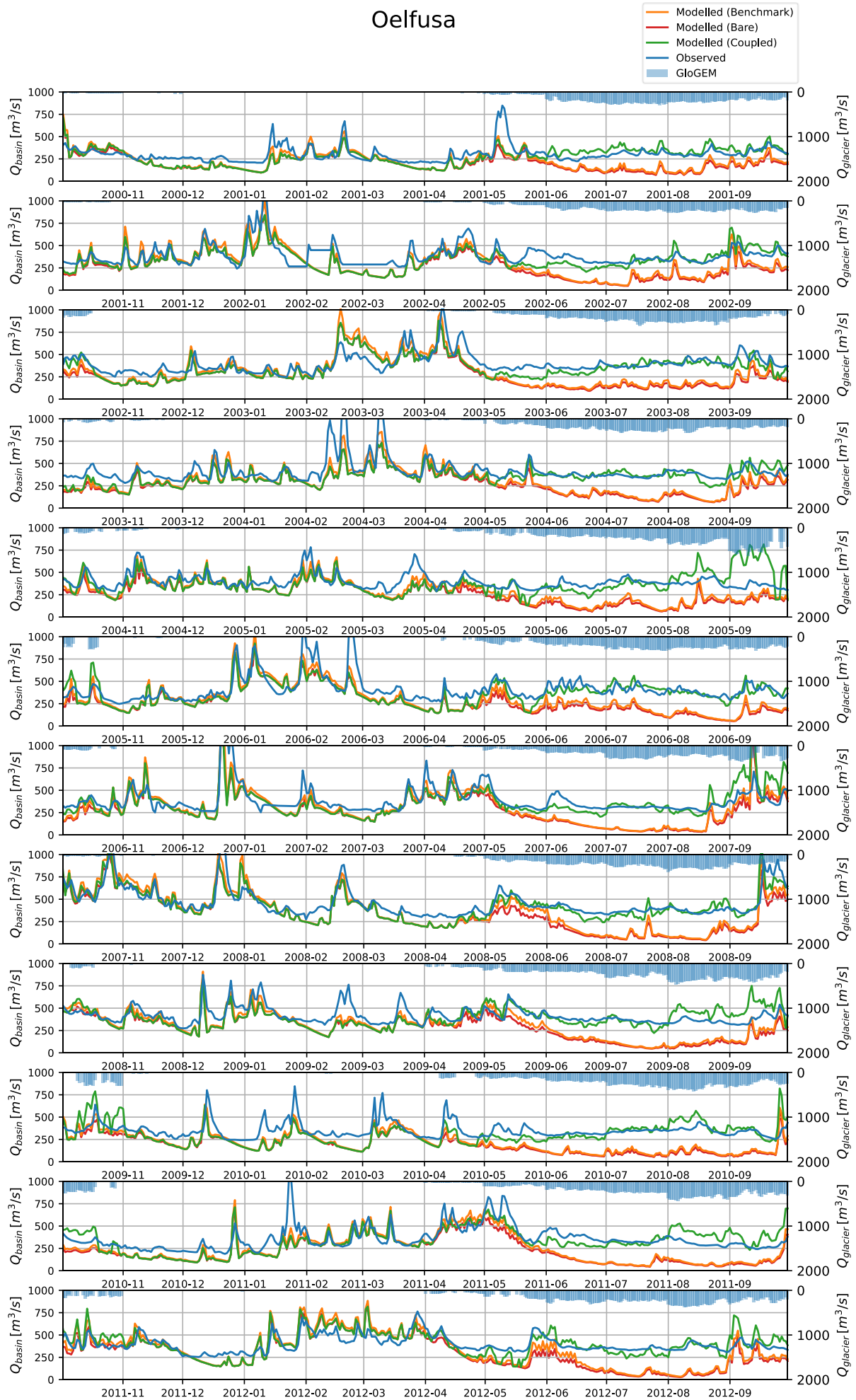
Negro



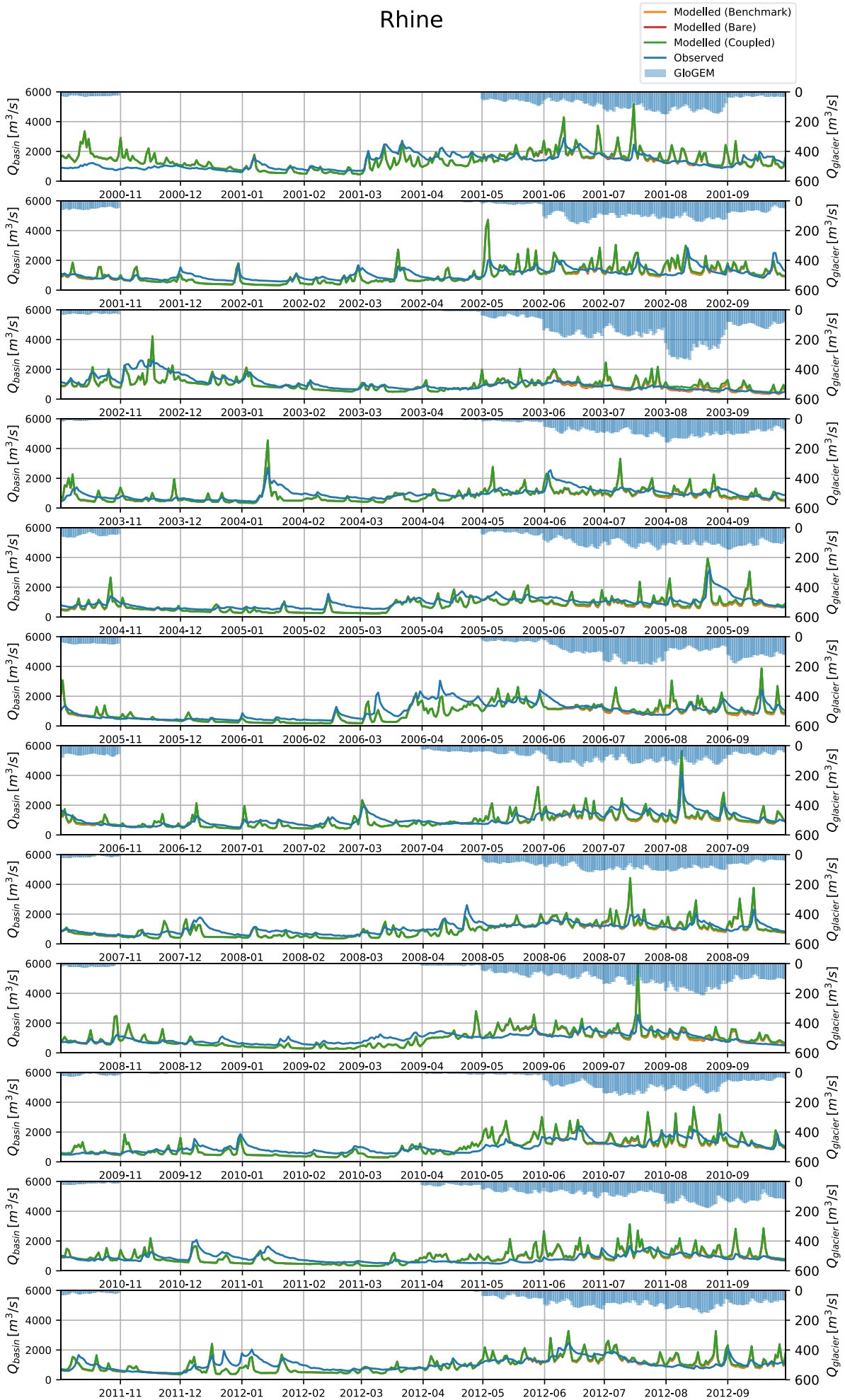
Ob



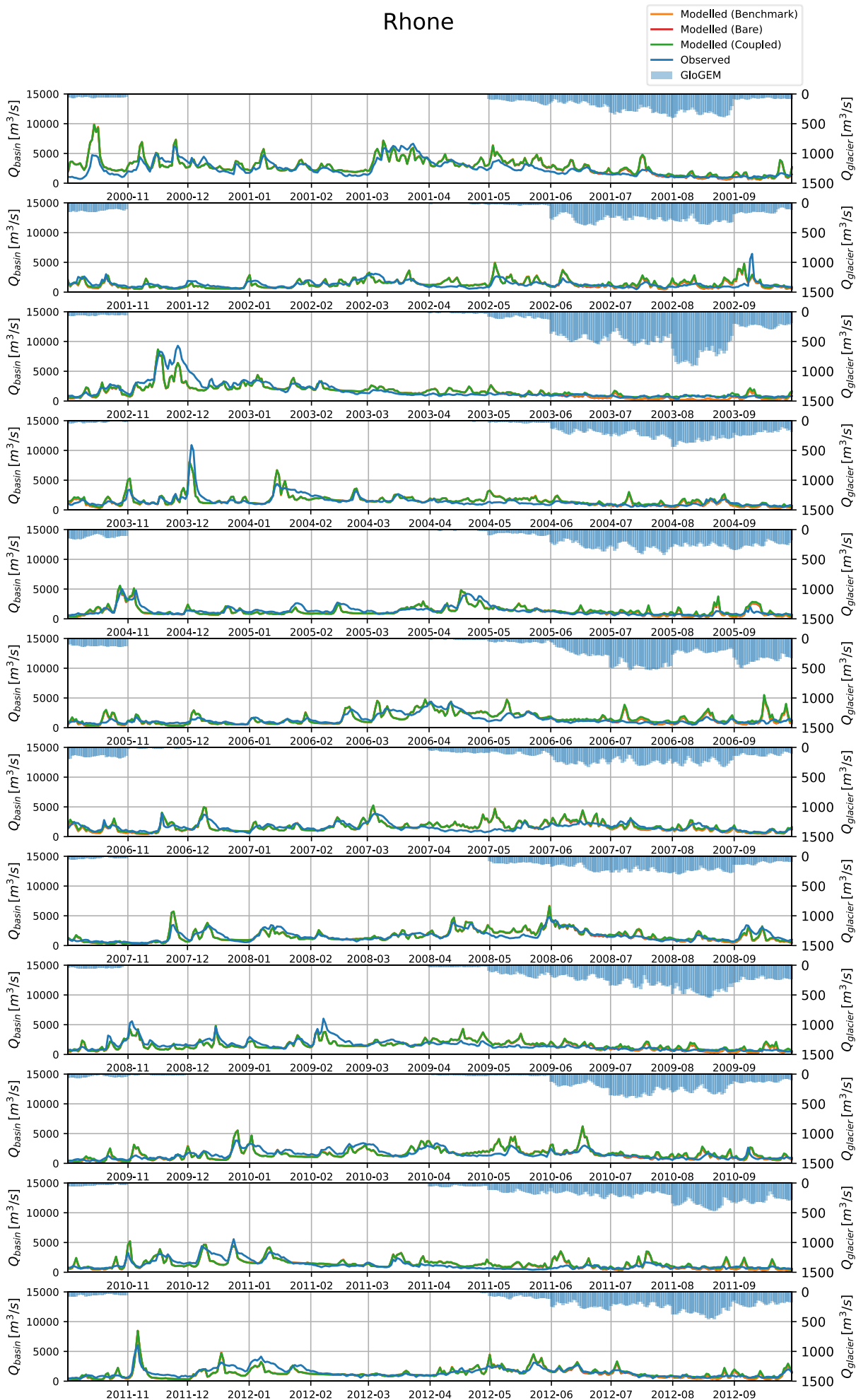
Oelfusa



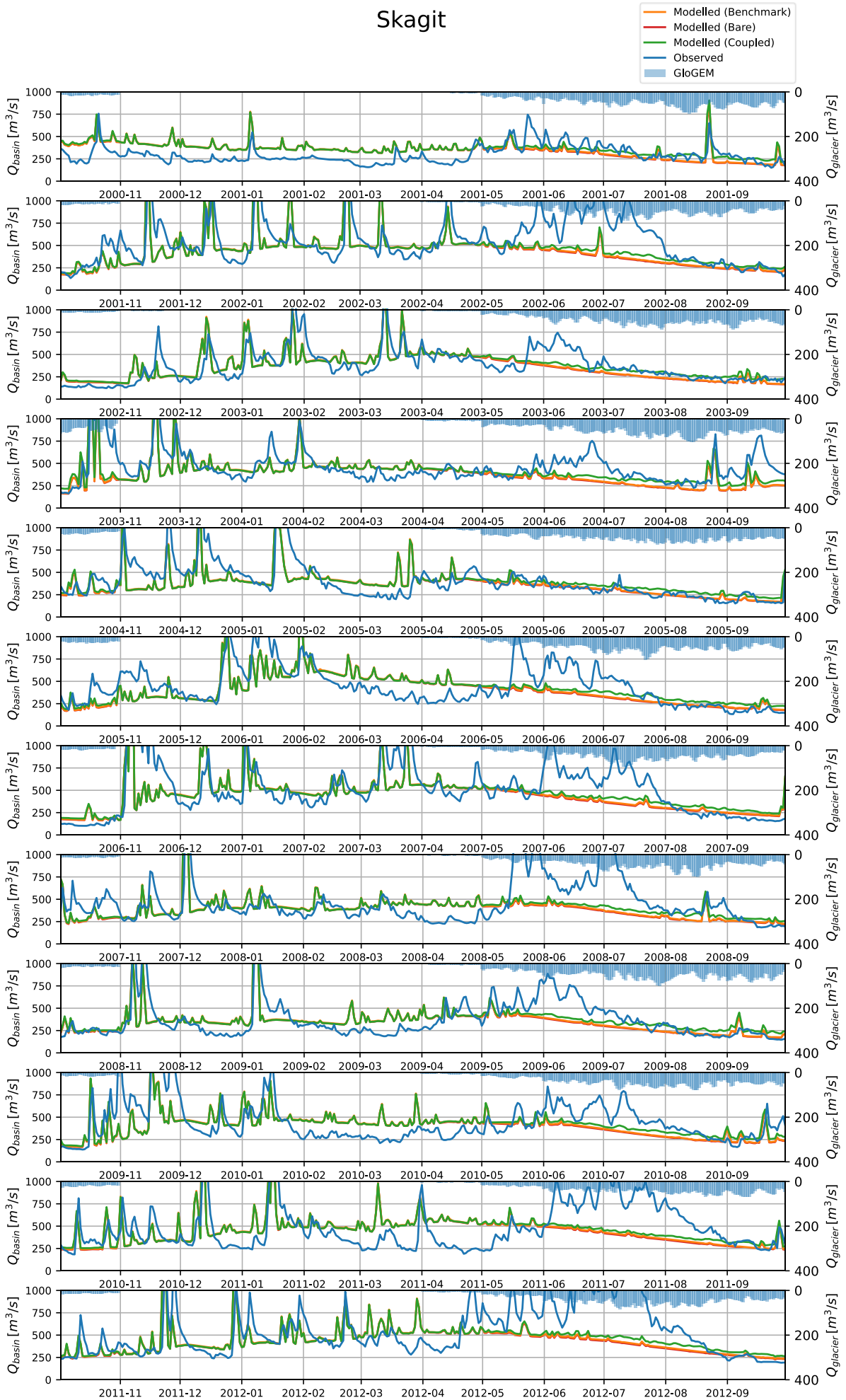
Rhine



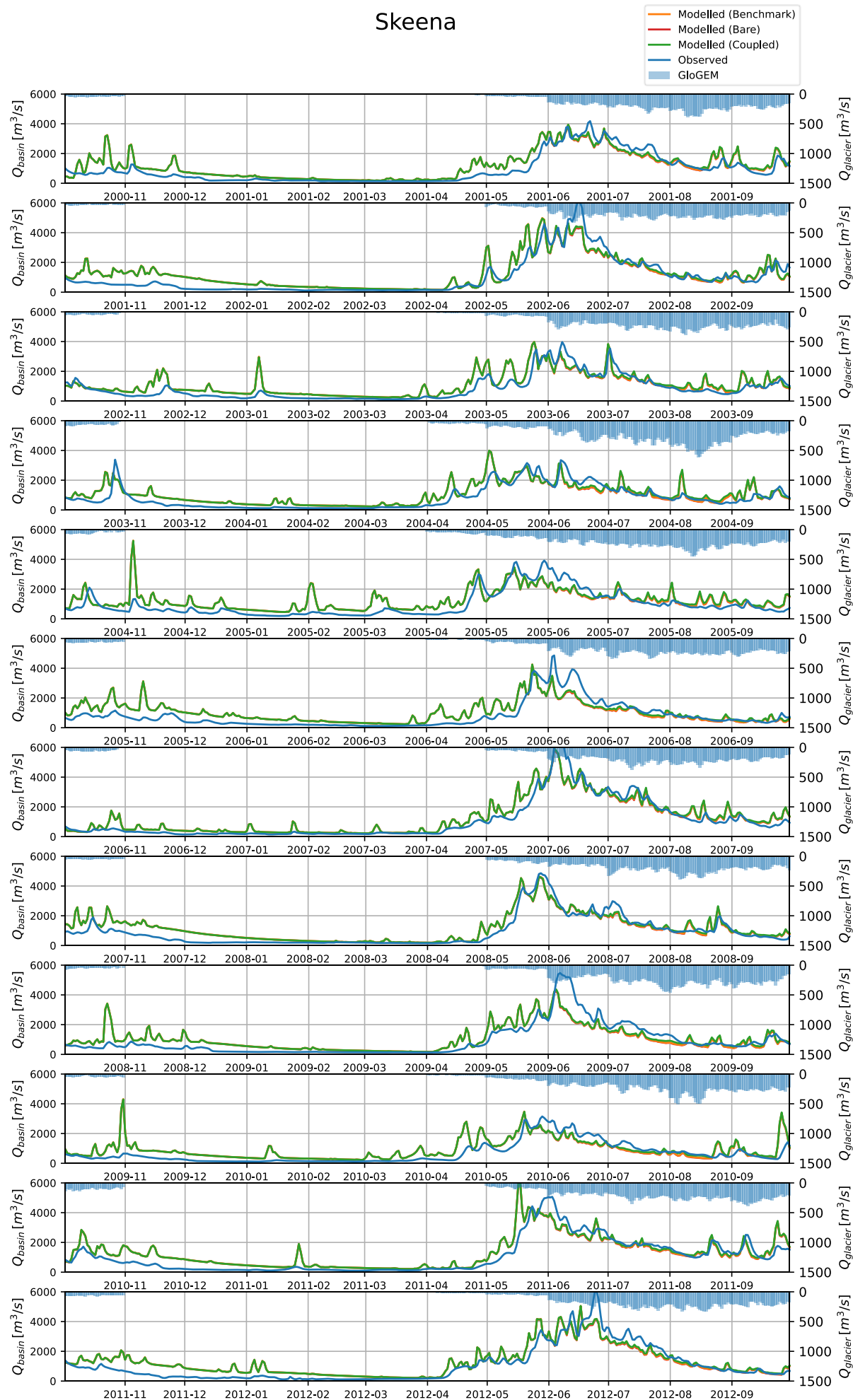
Rhone



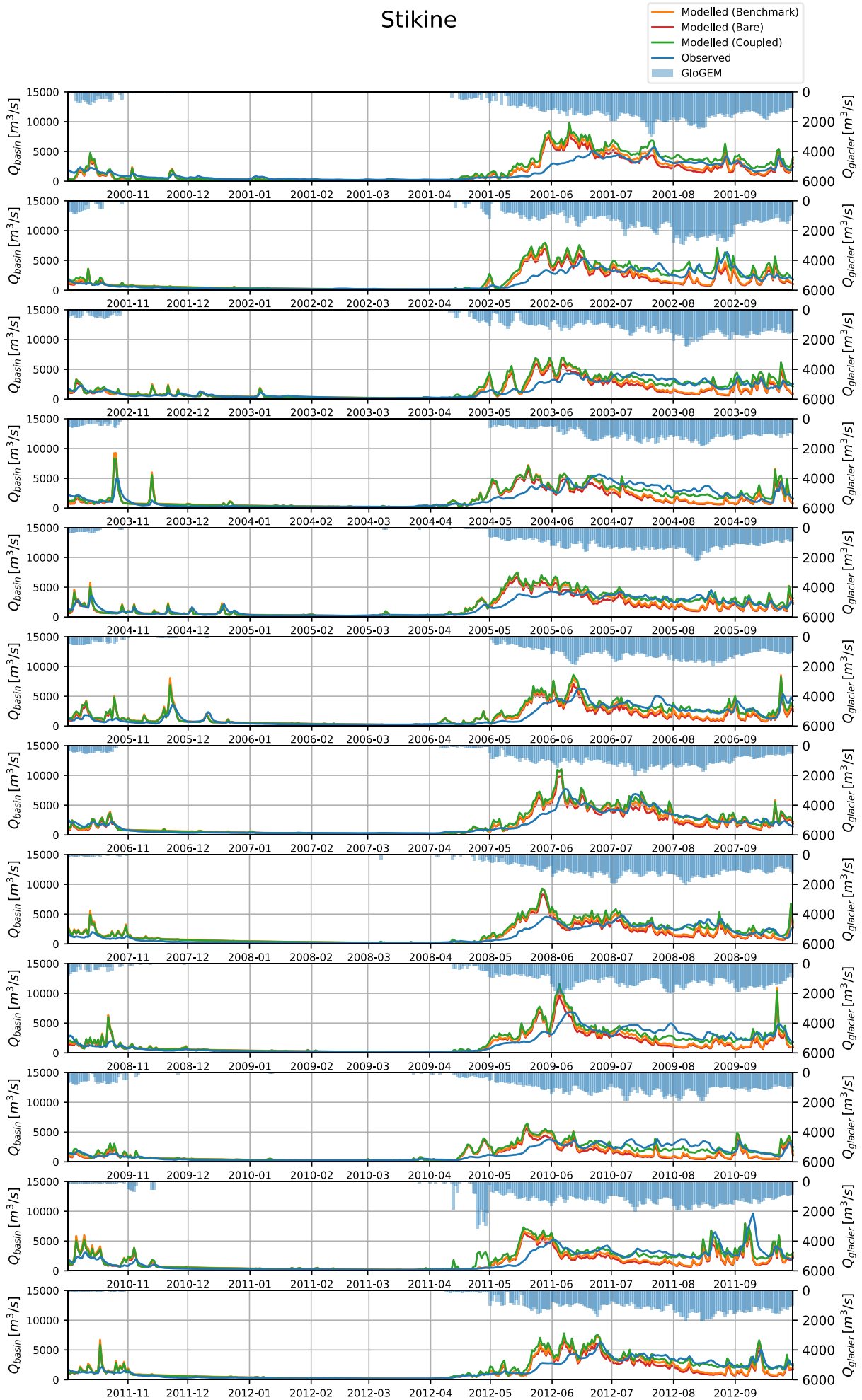
Skagit



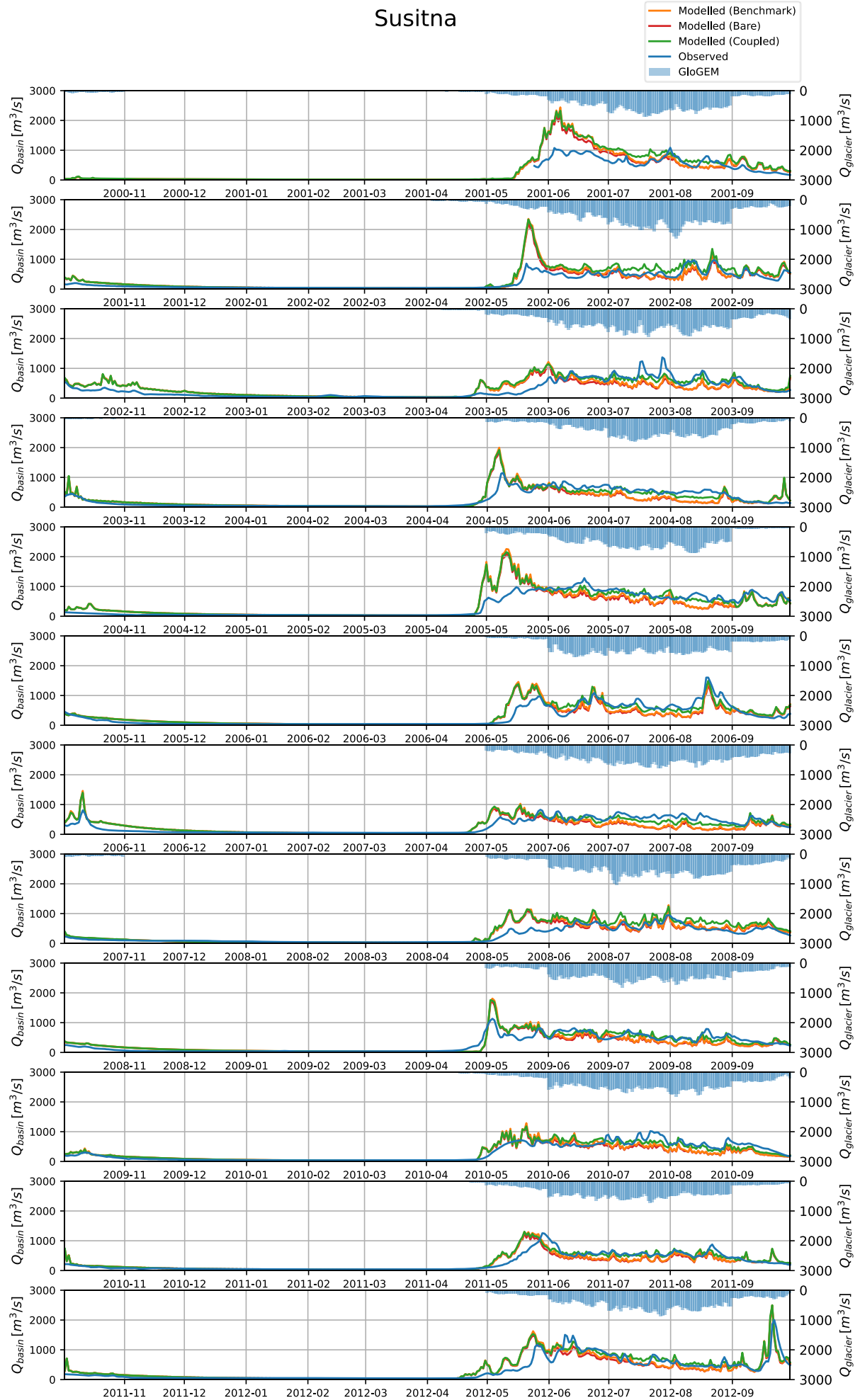
Skeena



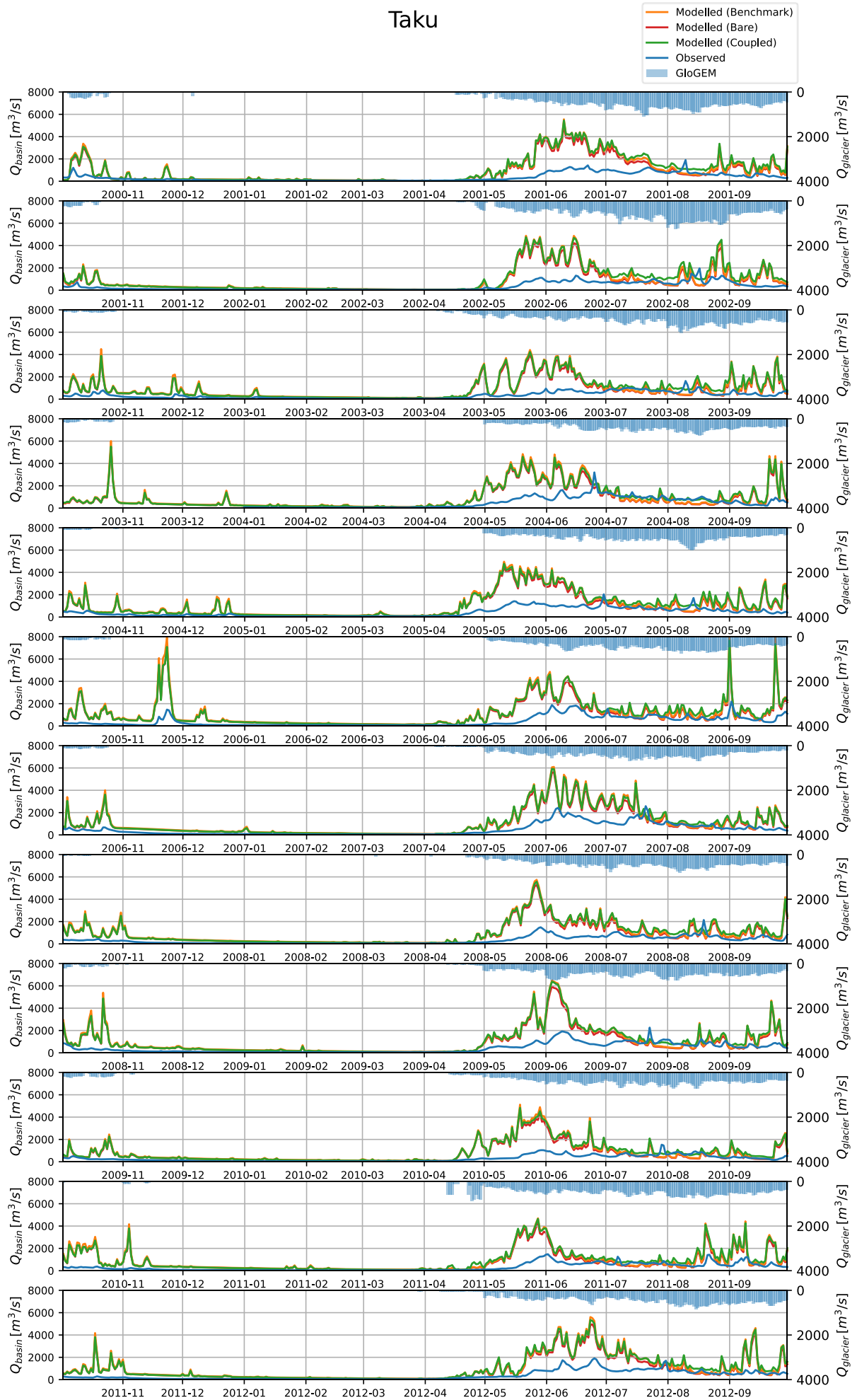
Stikine



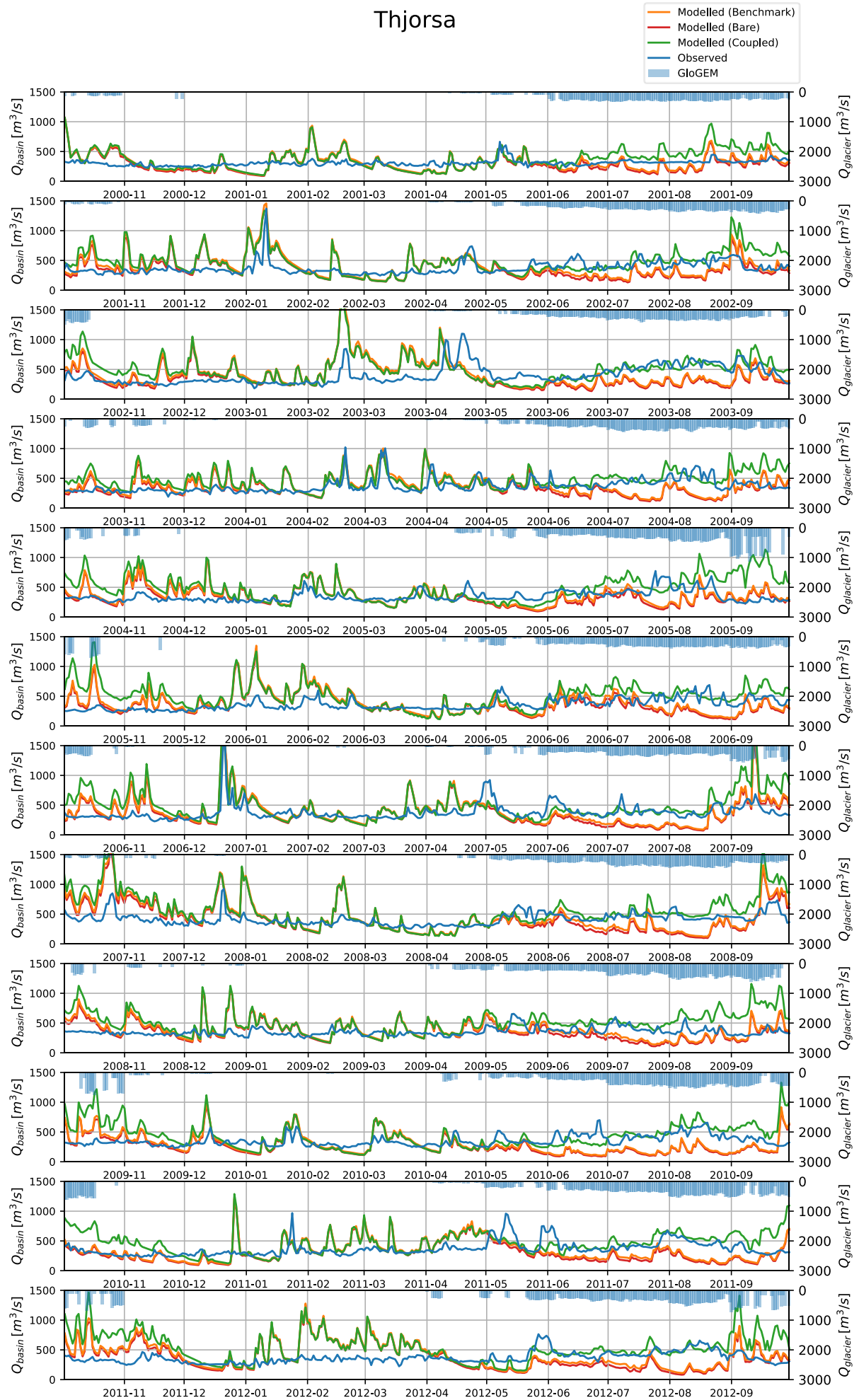
Susitna



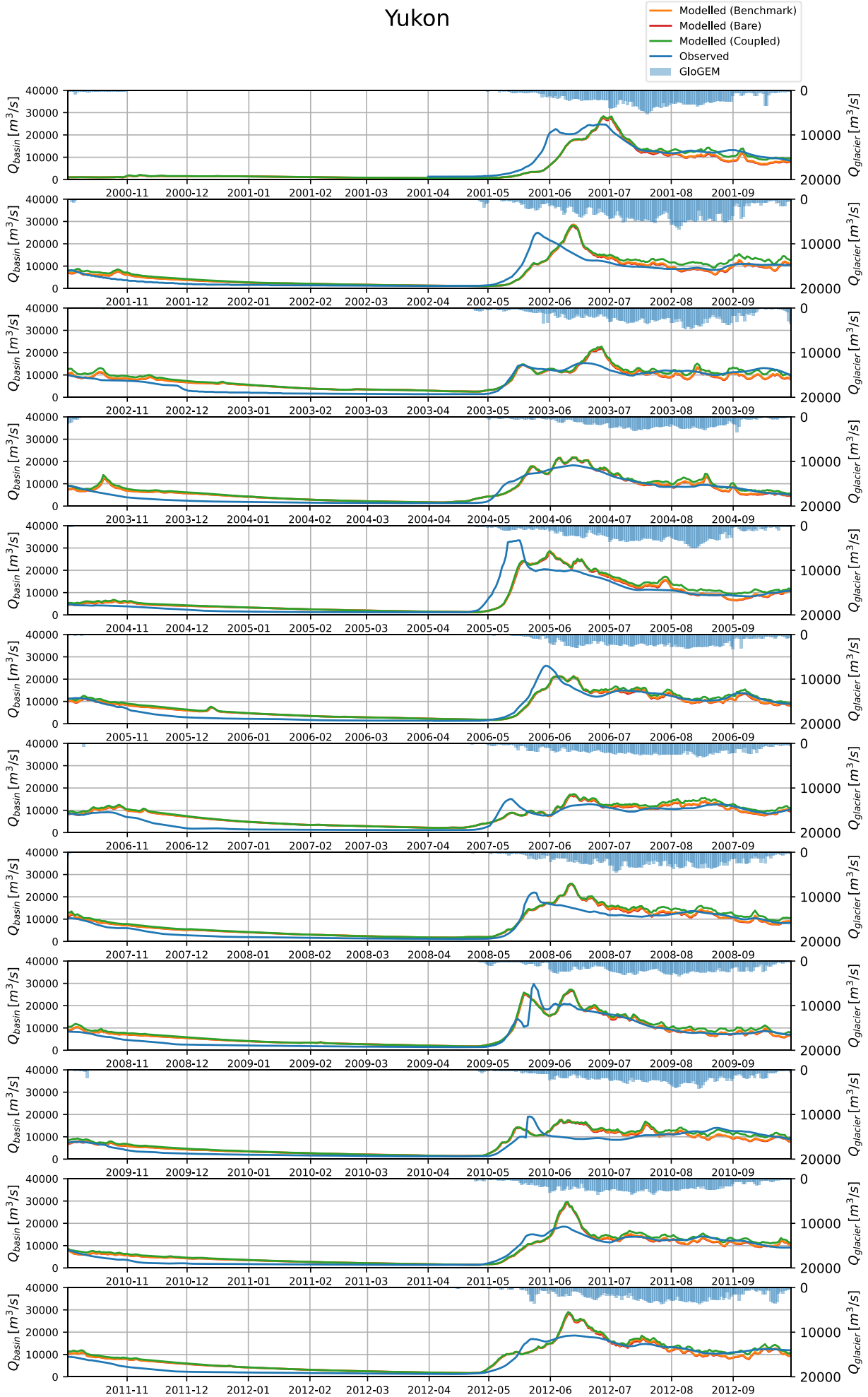
Taku



Thjorsa



Yukon



Appendix H Basin maps

