https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Leveraging reforecasts for flood estimation with long continuous simulation: a proof-of-concept study

Daniel Viviroli[1], Martin Jury[2], Maria Staudinger[1], Martina Kauzlaric[3,4], Heimo Truhez[2], and Douglas Maraun[2]

[1]Department of Geography, University of Zurich, Zurich, Switzerland
[2]Wegener Center for Climate and Global Change, University of Graz, Graz, Austria
[3]Institute of Geography, University of Bern, Bern, Switzerland
[4]Oeschger Centre for Climate Change Research, University of Bern, Bern,Switzerland

Correspondence: Daniel Viviroli (daniel.viviroli@geo.uzh.ch)

Abstract. Flood estimation is critical for risk assessment, but traditional methods are often constrained by the limited length of observation data. This study explores the potential of reforecasts (RFs) to enhance flood estimation through use in long continuous simulation (CS) with a hydrological model. As a proof of concept, we processed individual RFs from the vast database of the European Center for Medium-Range Weather Forecasts (ECMWF) with bias correction, stochastic downscaling and disaggregation with analogs to finally obtain mean areal precipitation and mean areal temperature for a set of test catchments in Switzerland. We subsequently concatenated these RFs into a time series of close to 10 000 years length and used them in long CS to derive flood return levels. Results demonstrate the potential of RFs as a complementary tool in flood estimation, providing insights into extreme event magnitudes and frequencies. Moreover, RFs can provide a relevant alternative view on exceptionally high extremes when compared to flood estimates derived from using other inputs to long CS, such as those generated by a stochastic weather generator. Limitations apply to catchments smaller than approximately 500 km^2, where the stochastic downscaling becomes increasingly inadequate, especially for resolving convective events. There, dynamical downscaling would be more appropriate, but was not feasible with the data currently available.

1 Introduction

Rare to very rare floods, associated with return periods of 1000 to 10 000 years, can cause severe human and economic damage. However, their estimation is limited by the comparatively short streamflow records available and the unknown representativity of disaster-rich and disaster-poor periods within these records for present conditions (Redmond et al., 2010; Schmocker-Fackel and Naef, 2010a, b; Zeder and Fischer, 2024). At the upper end of the flood magnitude scale, the estimation of very rare or even unprecedented floods – and of weather hazards more broadly, see Kelder et al. (2025) – is explored through various approaches. These include transforming estimates of possible maximum precipitation (PMP) into estimates of possible maximum floods (PMF) (World Meteorological Organization, 2009), employing weather generators for simulations with hydrological models (Lamb et al., 2016), using large data sets of observed floods and pooling them into regions (Bertola et al., 2023), and developing storylines based on extremely rare observed events (e.g., de Bruijn et al., 2023).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Regarding meteorological extremes more generally, reforecasts (RF) have been used to explore yet unobserved extremes that would be difficult to anticipate based on historical records. Even though the primary purpose of reforecasts is skill assessment and calibration of forecasts (Robertson and Vitart, 2019), the large archive of model realisations under current atmospheric conditions that these ensemble prediction systems provide (Kelder et al., 2022) is also an interesting possibility to explore meteorologically plausible extreme events beyond the range of observations. RFs often have lead times of several weeks. That is, while starting from observed weather, they simulate the atmospheric state beyond the time horizon over which the atmosphere is well predictable and, therefore, substantially deviate from the weather which has actually manifested. Thus, RFs simulate plausible but non-realized weather states, including extreme events that have not yet been observed (Thompson et al., 2017). In this sense, RFs have been used to estimate, for example, extremes of sea surges (van den Brink et al., 2004), precipitation (Thompson et al., 2017; Kelder et al., 2020), temperature (Thompson et al., 2019; Kay et al., 2020) and wind (Osinski et al., 2016), and to map compound events (Hillier and Dixon, 2020). In connection with estimation of extreme floods, Brunner and Slater (2022) pooled RF ensemble members of river discharge from the European Flood Awareness System (EFAS) to examine the frequency and intensity of extreme floods (up to 200 years return period) at the daily scale across numerous catchments in Central Europe. They demonstrated that, compared to observation-derived flood estimates, this approach yields estimates with considerably narrower uncertainty bounds for extreme events occurring less than twice a century. Similarly, Klehmet et al. (2024) investigated the robustness of frequency estimates for both precipitation and streamflow extremes (up to 500 years return period), using synthetic time series derived from pooled meteorological seasonal reforecasts SEAS5 (Johnson et al., 2019). These time series were then used to generate hydrological reforecasts with the E-HYPE process-based model (Hundecha et al., 2016) for the pan-European domain. By generating 100 samples with replacement for several time series length, they showed that the relative interquartile range of the estimated 100-year return levels – used to quantify sampling uncertainty – decreased markedly as the time series length increased, dropping from over 100% with 50-year time series to below 10% with 500-year time series. Extending on these methods, Ganapathy et al. (2024) introduced an additional step in the workflow to account for different flood type clusters, modifying the approach by applying a mixing distribution after pooling. In their study, they used RF data from the Global Flood Awareness System (GLOFAS) and applied the UNprecendented Simulated Extreme ENsemble (UNSEEN) method (Kelder et al., 2020, 2022) for pooling, focusing on two catchments in Germany, with return periods of up to 200 years. Their results indicate that explicitly considering different flood types considerably alters the 100-year return level estimate ($\geq$15%) and reduces the Root Mean Square Error (RMSE) compared to the conventional approach, which assumes a single distribution. The application of the UNSEEN method using RFs extends beyond flood frequency estimation; it has also been shown to provide plausible unprecedented events, supporting the development of flood storylines under varying initial conditions (Kay et al., 2024).

Here, we explore the feasibility of utilizing extensive RF data on precipitation and temperature within a long continuous simulation (CS) framework, leveraging RFs with a total length of close to 10 000 years obtained from the vast database of the European Center for Medium-Range Weather Forecasts (ECMWF). Our proof of concept combines these data with key advantages of CS, enabling the estimation of return periods considerably larger than previous RF-related studies. In particular, this approach avoids assumptions about antecedent conditions and their spatial patterns while enabling a physically coherent

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

representation of the full spatial-temporal development of floods over larger areas (Lamb et al., 2016). For comparability with an existing long CS approach in the application domain that utilizes weather generator (WGEN) input, we employed the hydrometeorological simulation chain introduced by Viviroli et al. (2022) that serves as the foundation for the projects EXAR (hazard information for extreme flood events on the rivers Aare and Rhine) (Andres et al., 2021) and EXCH (Extreme Floods in Switzerland) (Viviroli et al., 2025). Both EXAR and EXCH aim at providing methodologically and spatially coherent flood hazard assessments, and the project results have been examined comprehensively as far as feasible from the data available on rare events (Andres et al., 2021; Viviroli et al., 2022, 2025).

To arrive at the spatial and temporal scales necessary for hydrological modelling, bias adjustment, downscaling and disaggregation of the relatively coarse RFs were necessary. To ensure comparability with the EXAR/EXCH framework, some methodological choices had to be carried over, most notably the use of a lumped catchment model combined with hydrological routing where necessary, and the interpolation of mean catchment meteorological inputs from point values. While specifics of this procedure could certainly be altered or fine-tuned, this framework offers the possibility to juxtapose RF-based (from physically consistent and plausible weather conditions) and WGEN-based (from a stochastic multi-site framework) flood estimates and provides valuable context for assessing very rare floods.

Departing from a broad scale range of approximately 20 to 3000 km^2, we explored the feasibility of both a statistical and a dynamical downscaling approach, and discovered that only the statistical approach – suitable for larger scales – was feasible with the data publicly available. The main parts of this paper therefore focus on this approach, while dynamical downscaling – which would be more suitable for small scales than the statistical approach – is briefly discussed in Sect. 5.3.

2 Data

2.1 Test catchments

For our proof of concept, we examined a total of 20 test catchments (Table 1, Fig. 1) from across Switzerland. These catchments represent important climate regions and runoff regime characteristic of a region with complex topography. Note that three catchments in each of the Aare river basin (Kander at Hondrich, Aare at Thun, Aare at Bern), the Thur river basin (Thur at Alt St. Johann, Thur at Jonschwil, Thur at Andelfingen) and the Maggia river basin (Riale di Calneggia at Cavergno, Isorno at Mosogno, Maggia at Locarno) are nested. The sites on the Aare, Drance de Bagnes, Maggia, Saltina and Sarine rivers are impacted by hydropower, and the Thun and Bern sites on the Aare River are influenced by the regulated lakes Brienz and Thun.

2.2 Reforecast data

Our work is based on the ensemble RFs provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) that allows considering plausible yet not actually occurred weather extremes (Owens and Hewson, 2018). For the statistical modelling, we used precipitation and near surface air temperature as input. The data

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Table 1. Test catchments examined in this proof-of-concept study, with area (A), mean catchment elevation (E) and runoff regime type (Weingartner and Aschwanden, 1992). The catchments are grouped by climate region (Bundesamt für Umwelt BAFU, 2022) and listed in order of decreasing area.

Climate region	River and site	ID	A [km^2]	E [m a.s.l.]	Runoff regime type
Northern Alps	Aare @ Bern *†	AarBrn	2941	1584	nivo-glaciaire
	Aare @ Thun *†	AarThu	2459	1739	nivo-glaciaire
	Sarine @ Broc *	SarBro	636	1501	nival de transition
	Thur @ Jonschwil	ThuJon	493	1026	nivo-pluvial préalpin
	Kander @ Hondrich	KanHon	491	1846	nivo-glaciaire
	Kleine Emme @ Emmen	KEmEmm	478	1058	nivo-pluvial préalpin
	Minster @ Euthal, Rüti	MinEut	59.1	1352	nival de transition
	Thur @ Alt St. Johann	ThuASJ	42.3	1489	nival alpin
	Allenbach @ Adelboden	AllAde	28.8	1858	nival alpin
Western Central Alps	Drance de Bagnes @ Le Châble *	DdBLeC	254	2601	b-glaciaire
	Lonza @ Blatten	LonBla	77.4	2615	a-glaciaire
	Saltina @ Brig *	SalBrg	76.5	2014	b-glacio-nival
	Goneri @ Oberwald	GonOwd	38.5	2375	b-glacio-nival
Eastern Plateau	Thur @ Andelfingen	ThuAnd	1702	773	pluvial supérieur
Eastern Central Alps	Rein da Sumvitg @ Sumvitg	RdSSum	21.8	2445	b-glacio-nival
Engadin	Inn @ S-Chanf	InnSCh	616	2460	b-glacio-nival
Southern Alps and Ticino	Maggia @ Locarno *	MagLcn	927	1534	nivo-pluvial méridional
	Isorno @ Mosogno	IsoMsg	125	1570	nivo-pluvial méridional
	Riale di Calneggia @ Cavergno	RdCCav	23.9	1986	nival méridional
	Krummbach @ Klusmatten	KruKlu	19.4	2265	nival méridional

* Flows at these sites are impacted by hydropower operations (see Margot et al., 1992)

† Flows at these sites are impacted by regulated natural lakes

were available on a 0.25° regular grid (~19×28 km^2 in the study domain) until forecast day 15, and on a 0.5° regular grid (~38×56 km^2) thereafter. The temporal resolution of the data was 6 hours.

ECMWF's IFS has undergone several changes since its inception in 2008 (ECMWF, 2024a, b, c). As detailed in Tab. 2, the ECMWF RFs (control and perturbed) are

- initialised weekly 2008/03–2015/05, and bi-weekly from 2015/05 onwards,
- initialised for the leading 18 years 2008/03–2012/06, and for the leading 20 years from 2012/06 onwards,

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

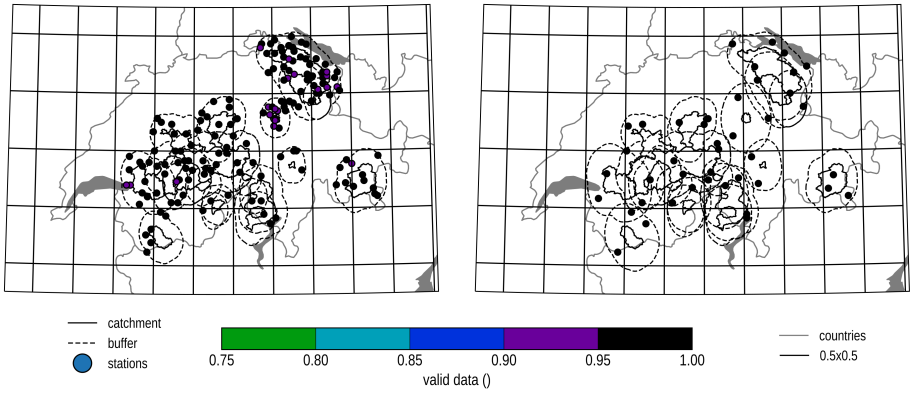

Figure 1. Study region with catchments (solid black lines) and meteorological stations (dots), also showing temporal coverage of station data for precipitation 1961–2019 (left) and temperature 1991–2019 (right). The 0.5° grid (~38×56 km^2) shown is equivalent to that of the reforecasts used. Dashed black lines indicate the buffer zone around the catchment used in the statistical downscaling.

- provide 4 ensemble members 2008/03–2015/05, and 10 ensemble members from 2015/05 onwards, and
- provide forecasts for a period covering 31 days (2008/03–2015/05), and 46 days from 2015/05 onwards.

In addition, the IFS underwent regular updates over the years, often multiple times per year (Tab. 2). To benefit from as much data as possible and derive a large collection of yearly time series for hydrological modelling, data from the different IFS cycles initialised within the same year were not treated differently.

However, a detailed analysis regarding inconsistencies and drifts revealed two issues:

1. The precipitation data show a spike on the transition day between medium-range ensemble (ENS) RFs (days 1–15) and the extended range RFs (days 16–46) (see Fig. S1 in the Supplement, left).
2. The temperature data for RFs initialised in 2015 have a strong drift over the forecast period (see Fig. S1 in the Supplement, right).

To address the first issue, we used only data from day 16 onward, corresponding to the extended range RFs. This also ensures a high degree of independence among individual ensemble members, as discarding the first 10 days alone would have been already sufficient for our application domain (Mahlstein et al., 2019). To resolve the second issue, we excluded the year 2015 entirely from the analysis. Moreover, we did not include the year 2008, as it does not provide a full year of data. As a result, data with a total length of more than 10 000 years were available for analysis and processing.

2.3 Meteorological data

We employ a statistical model to relate the gridded RF data to the point scale. As a meteorological reference for grid-scale precipitation, we used RhiresD, a spatial precipitation analysis available from MeteoSwiss (2021a), which provides daily

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Table 2. Reforecasts by initialisation year, showing number of initialisation per week, number of years reforecasted, number of members, the available forecast length (from day 16 onwards) and the respective IGS cycles for each year (ECMWF, 2024a, b, c). Note that years 2008 and 2015 were not used in this study.

year	initialisation frequency	n years before	members	forecast length	IFS cycles
2008 *	weekly	18	1+4	16	32r3V; 33r1; 35r1/33r2; 35r2
2009	weekly	18	1+4	16	35r2; 35r3; 36r1
2010	weekly	18	1+4	16	36r1; 36r2; 36r4; 37r2
2011	weekly	18	1+4	16	37r2; 37r3; 38r2
2012	weekly	18 / 20	1+4	16	38r2
2013	weekly	20	1+4	16	38r1; 38r2; 40r1
2014	weekly	20	1+4	16	40r1
2015 *	weekly / twice a week	20	1+4 / 1+10	16 / 30	40r1; 41r1; 41r2
2016	twice a week	20	1+10	30	41r2; 43r1
2017	twice a week	20	1+10	30	43r1; 43r3
2018	twice a week	20	1+10	30	45r1
2019	twice a week	20	1+10	30	45r1; 46r1
2020	twice a week	20	1+10	30	46r1; 47r1

* Not used in this study

precipitation data from 1961 onwards at a 1 km raster resolution. Additionally, we evaluated the hourly MeteoSwiss product CombiPrecip (Sideris et al., 2014), which combines weather radar fields with point precipitation measurements from 2005 onward, also at a 1 km resolution. However, we used CombiPrecip only for disaggregating daily precipitation (Sect. 3.1.3) because the data, despite MeteoSwiss' elaborate processing (Germann et al., 2022), are noticeably affected by shielding of the radar beam by terrain. Moreover, CombiPrecip data are only available from 2005 onward, which is too short for bias adjustment. We also examined two further gridded observation products but discarded them due to their limitations in the present context: E-OBS (Cornes et al., 2018), which has a low station density; and WFDE5 (Cucchi et al., 2020), which matches the grid point resolution of the coarse RFs (0.5°) but is misaligned by half a grid point.

As a meteorological reference for grid-scale temperature, we considered a spatial analysis of mean, minimum and maximum temperature available from MeteoSwiss (2021b), which provides daily data from 1961 onwards at a 1 km raster resolution. However, we opted to directly correlate the RF temperature data to the station scale and did not use this gridded analysis.

As a meteorological reference at the point scale for both precipitation and temperature, we used station data obtained from MeteoSwiss (2025, retrieved 01.05.2021). The temporal and spatial coverage varied between the available daily and the hourly gauging networks, and for consistency with the approach used in the projects EXAR/EXCH, we used hourly station data aggregated to the daily scale for both the stochastic downscaling of precipitation and the quantile mapping of temperature (see Sect. 3.1).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

2.4 Hydrological data

The continuous hydrological data encompassed hourly discharge records for the test catchments listed in Tab. 1. These data were available for a maximum period from 1974–2021. However, most time series were shorter, with a median length of 30 years and a range of 6–47 years. In addition, records of annual maximum floods (AMFs) were available for most of these stations. These records refer to instantaneous peak flow and date back even further, with a median length of 63 years and a range of 23–118 years. Extrapolations of the observed AMFs via the generalized extreme value (GEV) distribution were also available for these stations (Baumgartner et al., 2013). Maximum flood data from all over Switzerland were available from Kienzler and Scherrer (2018), covering measurements and historical reconstructions at 740 sites in total.

2.5 Hydrometeorological scenarios based on weather generator

For comparison, 300 000 years of hourly simulation results from a hydrometeorological model chain using the Generator of Weather EXtremes (GWEX) WGEN were available. GWEX is a multi-site, two-part stochastic weather generator for precipitation and temperature, based on the structure proposed by Wilks (1998). It is designed to reproduce the statistical behaviour of weather events at various temporal and spatial resolutions, with a focus on extremes. Since comparatively long events are relevant in the main EXAR/EXCH framework, GWEX first generates 3-day precipitation amounts. These amounts are then disaggregated into daily and ultimately hourly values using meteorological analogues. Details on GWEX are found in Evin et al. (2018, 2019), while its application to flood estimation is discussed in Viviroli et al. (2022) and Mas et al. (2023).

3 Methods

As mentioned in the introduction, some methodological choices from the EXAR/EXCH framework (Andres et al., 2021; Viviroli et al., 2022) were retained to facilitate comparisons. A lumped catchment model – combined with hydrological routing where necessary – was run using mean catchment precipitation and temperature inputs interpolated from point values. The raw RF data were bias adjusted, stochastically downscaled to multiple meteorological stations, and disaggregated to hourly temporal resolution. Subsequently, mean catchment precipitation and temperature were interpolated and adjusted to mean catchment elevation (Sect. 3.2).

3.1 Statistical postprocessing of reforecasts

The processing of the RF data involved multiple steps (Fig. 2), which were applied to all test catchments (Sect. 2.1).

As a model-derived product, the RF data are subject to biases, particularly at longer lead times. The statistical modelling chain served three main purposes: adjustment of model biases, downscaling to the point scale to multiple meteorological stations, and performing temporal disaggregation.

Bias adjustment is often calibrated between climate model output and station data as a simple downscaling method that both adjusts model biases and downscales to the point-scale station data. This approach has been shown to cause substantial

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

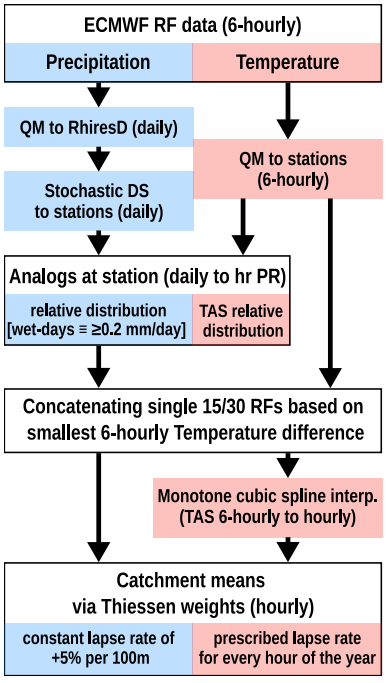

Figure 2. Workflow of statistical postprocessing, beginning with ECMWF raw reforecast (RF) data and proceeding through bias adjustment and stochastic downscaling (DS), application of analogs, concatenation, and interpolation to mean catchment values. PR is precipitation, TAS is temperature.

artefacts when applied to precipitation (Maraun et al., 2017). We therefore followed the conceptual approach by Volosciuk et al. (2017) and first bias-adjusted the gridded RF precipitation data at its native grid and subsequently statistically downscaled the adjusted data. The bias adjustment was carried out using quantile mapping against the gridded observations of RhiresD (Sect. 3.1.1). Subsequently, a multi-site stochastic downscaling method was applied to the adjusted RFs to generate spatially coherent precipitation time series at multiple station locations (Switanek et al., 2022). This model is based on a truncated and transformed multivariate Gaussian distribution (Sect. 3.1.2). To derive the hourly precipitation time series, daily precipitation values were disaggregated via analogs (see Sect. 3.1.3). Two-meter air temperature (TAS) was bias adjusted directly to the station locations using daily data, and applying the correction value to the 6-hourly RF data. Hourly values were subsequently built, after concatenating single 15/30 day long RFs via a cubic spline interpolation of 6-hourly temperatures. Time series of mean catchment precipitation and temperature were finally derived via Thiessen interpolation, using a constant adjustment factor for precipitation and a prescribed lapse rate for temperature as described in Viviroli et al. (2022).

3.1.1 Bias adjustment

In a first step, the RFs were corrected via basic quantile mapping (QM) (Jakob Themeßl et al., 2011), a common state-of-the-art bias adjustment method. We applied quantile mapping to the RF data separately for each month. For precipitation, we used

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

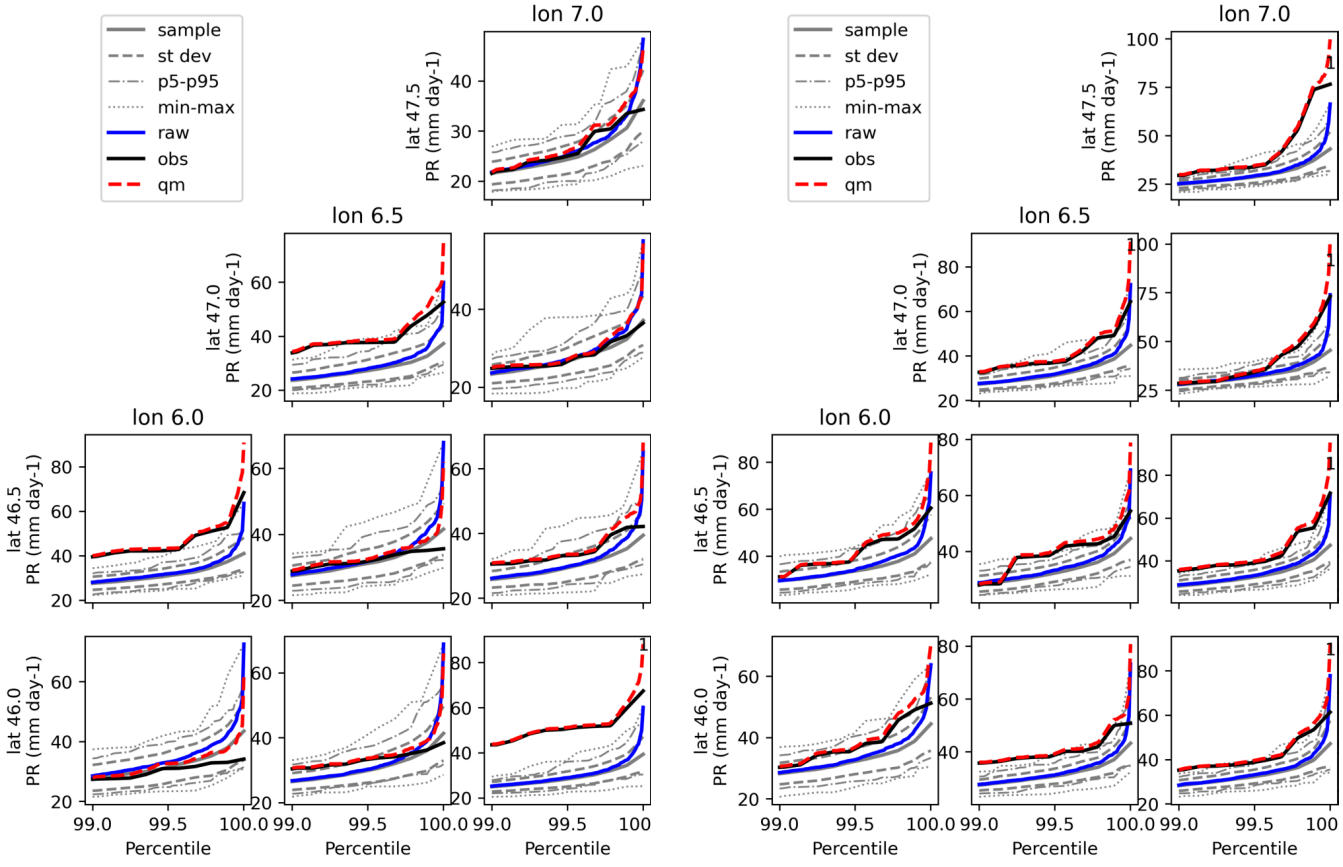

Figure 3. Daily precipitation (mm day^{-1}) starting from the 99th percentile for January (left) and August (right). Shown are RhiresD ('obs', dashed black line), ECMWF raw reforecast data ('raw', blue line), bias adjusted reforecast data ('qm', red dashed line), and the transfer function on which the correction is based ('sample', gray solid line) for selected grid points over Switzerland.

RhiresD as a reference, with daily data from the period 1991–2020. For temperature, we used station data located within a ~10–20 km buffer zone around the catchment as a reference, with data from the period 1991–2020. Temperature has been bias corrected using daily data, and applying the correction value to the 6-hourly data. Since the bias structure for temperature and precipitation in the RFs varied by year of initialisation (see Sect. 2.2), the periods 2009–2014 and 2016–2020 were adjusted separately.

Due to the large sample of ensembles, RFs include extremes not yet observed, and a direct bias correction would result in the elimination of heavy precipitation events when fitting to the distribution of the observational data. This would negate the key advantage of the RFs, which is their ability to provide insight into unseen extremes. Therefore, a transfer function was constructed by estimating the RF data distribution using non-overlapping samples of the same length as the observational data. Simply put, the raw RF data were corrected by the quantile differences between the observations and the sample mean to derive the corrected RF data (Fig. 3).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

3.1.2 Stochastic downscaling

For stochastic downscaling, we adapted code of Switanek et al. (2022) to handle missing station data. We then applied this stochastic downscaling using RhiresD and daily station data for calibration over the period 1961—2019 for each month and for each of the test catchments separately. We did not use stations that had more than 25% of missing data during this period. Using 6-hourly instead of daily station data for calibration yielded unrealistic values for some sites and was discarded. In addition to stations within a catchment, all stations that were located within a ~10 km buffer zone around the catchment were used. We applied the adjustment of the mean bias to the simulated data; however, the quantile-based multiplicative bias adjustment was omitted as it led to poorer results in a split sample test (see Switanek et al., 2022). We also tested the impact of extending the range of grid points included as predictors, but found no discernible difference between including grid points of up to 0.75° or 1.25° distance from the nearest stations. Therefore, we used grid points with a distance of up to 0.75° for the final downscaling.

3.1.3 Analogs and disaggregation

The bias-corrected and downscaled daily precipitation data had to be disaggregated to the hourly scale necessary for hydrological modelling. For this, the analog method was used as it is relatively easy to implement, and generally produces spatially and temporally coherent results. To have a large pool of reference data to draw the analogs from, we used daily and hourly station data (point values) as well as hourly CombiPrecip data (values on 1×1 km^2 raster).

First, stations within each single catchment and its vicinity were selected. Then, for each station, days with both daily and hourly data available over the period 1981–2019 were identified. For days where only daily station data were available, disaggregation was performed using daily fractions of hourly CombiPrecip values from the 2005–2019 period. Analogs were finally drawn for every month separately using a moving window of ± 1 month. Further, the data were separated into dry and wet days (threshold of 0.2 mm day^{-1} for daily catchment precipitation interpolated via Thiessen weights). For dry days (<0.2 mm day^{-1}), the hourly precipitation data were set to zero. For wet days, the daily precipitation data were disaggregated to hourly values by using relative hourly contributions of analogs at the station scale. As reference, all days within plus or minus one month of the respective modelled day were used. The best analog day was determined by calculating the RMSE at the station locations of both precipitation and mean daily temperature to all reference days, and selecting the analog day that showed the smallest mean precipitation and temperature RMSE.

3.1.4 Concatenation

The single RFs, which depending on their initialisation year have a length of 15 or 30 days, had to be concatenated into yearly time series for use in continuous hydrological simulations. A length of 360 days was assumed for each year, consisting of 24 and 12 single RFs, respectively. For each year, drawing of individual RFs was restricted to the same initialisation year, minimizing the mixing of RFs from different IFS cycles. To reproduce the annual cycle, we selected single RFs based on the dates that roughly represent a year. For instance, for forecasts before 2015 (initialised weekly, length 15 days), we built years using every second initialisation week, which resulted in two sets of years with the same initialisation dates. This approach introduces a

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

slight offset with the ongoing year, due to the difference of 15-day forecast length and 14-day gap between initialisation dates. Therefore, single initialisation dates had to be discarded to keep this offset reasonably small. With this approach, a total 9920 single 360-day years were obtained. The selection of the single RFs building a year was based on minimizing the temperature difference at the intersection at the catchment scale. Using raw 6-hourly temperature enabled us to minimize the temperature difference at the same time step. Grid points were weighted according to the station locations and their Thiessen weights. At the start of building years for each initialisation year, succeeding RFs can be drawn from a large pool (reforecasted years ensemble members). With each built year, this pool diminishes resulting in larger temperature differences to succeeding RFs. Mean differences over all date combinations stayed well below 1 °C for about three-quarters of the selected reforecasted years ensemble members selection steps, and increased to about 5 °C for the last built year (see Fig. S2 in the Supplement, left). The mean overall maximum temperature differences of the single years and selection steps increases from approximately 0.1 °C to about 14 °C for the last built year (see Fig. S2 in the Supplement, right). The differences shown are for temperature at 6-hourly resolution. In the final hourly time series, derived from 6-hourly temperature station data via a monotone cubic spline interpolation, differences are around one-sixth of the values shown. Fig. S3 in the Supplement shows one of the extreme gap years with an extreme temperature jump. The behaviour of the hydrological simulations in relation to the stitching points is examined in Sect. 4.2.

3.2 Continuous hydrological simulation and routing

The complete concatenated RF precipitation and temperature data were finally used as input for the bucket-type catchment model HBV (Hydrologiska Byråns Vattenbalansavdelning model, see Seibert and Vis, 2012). HBV was calibrated on hourly data spanning 1983–2019 and run at an hourly time-step, employing the non-linear response function of Lindström et al. (1997) to enhance the representation of flood behaviour. To achieve the time series of mean areal precipitation and mean areal temperature required for hydrological modelling in the given long CS framework, the RF values downscaled to locations of meteorological stations were interpolated via Thiessen polygons. Based on the difference of Thiessen weighted mean station elevation to mean catchment elevation, a constant adjustment factor with a linear increase of 5% for every 100 m was used for precipitation (see e.g., Farinotti et al., 2012; Ménégoz et al., 2020; Ruelland, 2020; Viviroli et al., 2022). For temperature, calendar lapse rates estimated from data 1981–2019 for each hour and each day of the year (and differentiated by six large river basins) were used (see Mas et al., 2023).

Observations of precipitation and temperature 1931–2019 were interpolated similarly to mean areal values and used as a basis for a control run with HBV.

In catchments strongly affected by hydropower operations, lake retention, lake regulation, bank overflow or floodplain retention, hydrological routing was applied using RS Minerve (García Hernández et al., 2020) to account for these effects.

For further details on the hydrological simulation set-up we refer to Viviroli et al. (2022).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

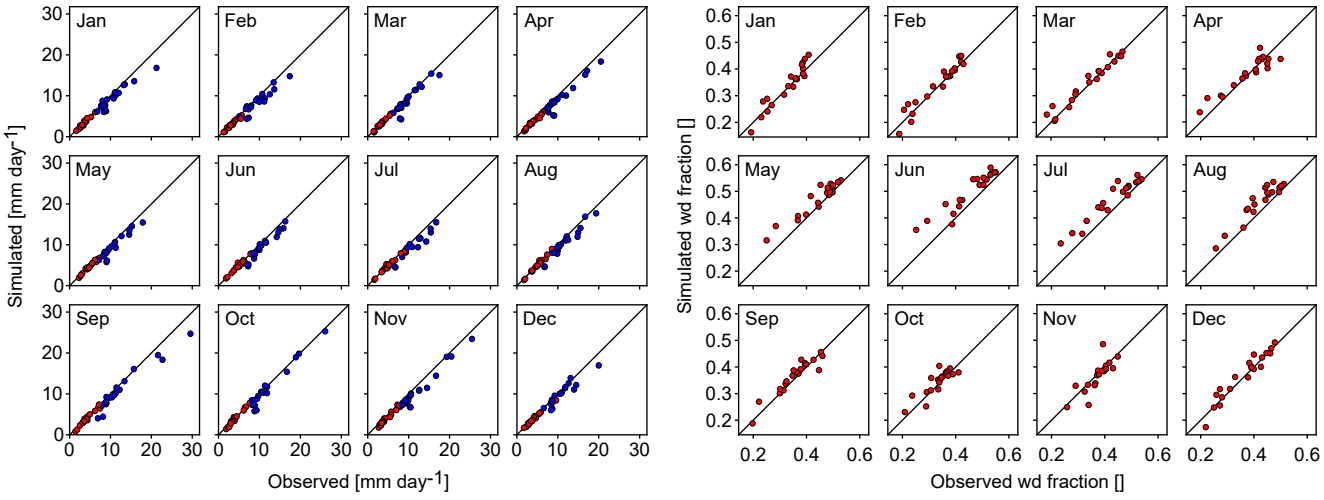

Figure 4. Left: Mean simulated and observed daily precipitation per month over all catchments for all day mean (red circles) and wet day mean (>1.0 mm day^{-1}, blue circles). Right: Mean simulated and observed wet day frequency per month (expressed as fraction of days per month) over all catchments (red circles). Station observations are for the period 1991–2019.

4 Results

4.1 Stochastic downscaling

Fig. 4 shows simulated and observed mean daily catchment precipitation (left) and mean wet day frequency (right). Note that there is a mismatch between the station observations used in processing the RFs and the data shown at point scale: The station observations are based on all available observed days over 1991–2019, which roughly coincides with the period used for the quantile mapping (1991–2020), while the stochastic downscaling was calibrated using grid data and station observations over the period 1961–2019. The stochastic downscaling reproduced these mean characteristics well. Fig. 5 shows simulations and observations at single station locations (circles) and in the Thiessen weighted mean (crosses) for the catchment of the Aare River at Bern. Also for the single stations, the mean characteristics are reproduced well. There is a tendency for a small underestimation of mean precipitation for single stations and months, and some deviations from observed wet-day frequency, which might be explained by the different data periods used for the bias adjustment and for calibrating the stochastic downscaling, as well as associated sampling differences of large scale modes of variability such as the Atlantic Multidecadal Oscillation (AMO) (see also Sect. 5.5).

The following evaluation of the downscaling results refers to the catchment scale. A clear and concise evaluation is hindered by two aspects. First, the bias adjustment of temperature and precipitation was done using reference data spanning the periods 1991–2019 and 1991–2020, respectively, while the stochastic downscaling was trained using grid and station data over the period 1961–2019, and the analogs were built using station data over 1981–2019. We here use the periods 1991–2019 and

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

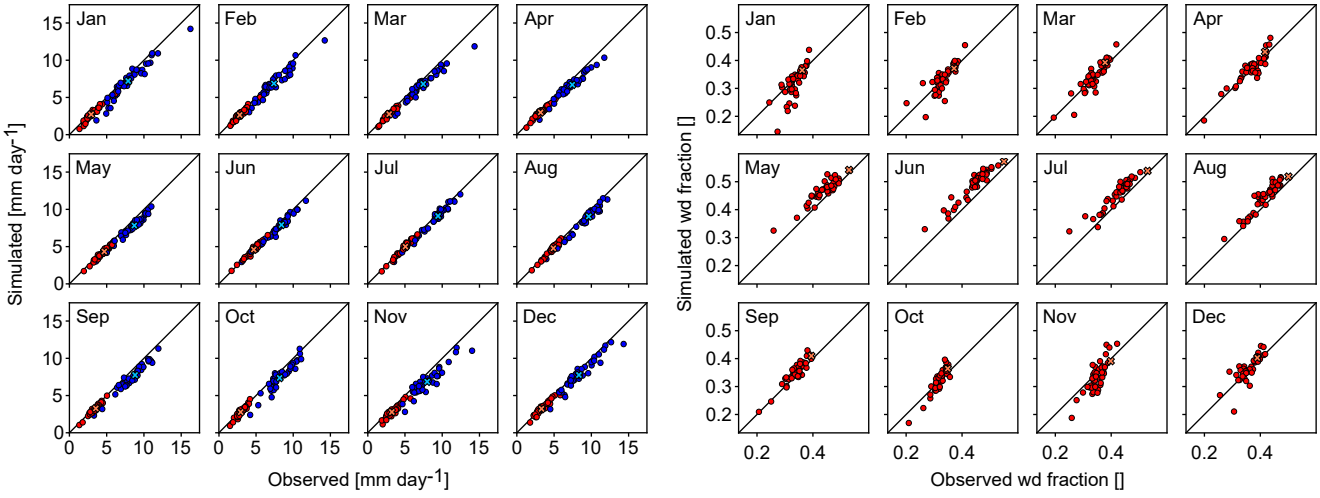

Figure 5. Same as Fig. 4 but for single stations within the Aare River catchment upstream of Bern (AarBrn). The crosses denote mean catchment all day and wet day precipitation and wet day frequency.

1981–2019 as reference. Second, the resulting 360-day calendar was built from the single RFs, which leads to small shifts in some of the seasonal statistics presented.

For monthly mean precipitation (Fig. 6, first column; for all results see Figs. S4–S7 in the Supplement), good agreement was found between observed and downscaled RF values for large and medium-sized catchments. Good agreement was also found for several small and very small catchments, while others – in particular Lonza, Saltina and Krummbach – showed notable differences for single months, with the mean observed annual cycle being outside the interquartile range (25th to 75th percentile) of the RFs.

In terms of annual 90th percentile daily precipitation (Fig. 6, second column), good agreement was found between the observations and the RFs for large and medium-sized catchments. Specifically, the values for the single observed years lie within the distribution of the RFs. While the same is true for many of the small and very small catchments, observed values for single years are larger than all downscaled values for the Lonza, Saltina and Krummbach river catchments, and smaller than all downscaled values for the Rein da Sumvitg river catchment.

Annual maximum 1-day and 3-day precipitation (Fig. 6, third and fourth column) show good agreement for the 29 years of observations used for large and medium-sized catchments, with the RFs falling within the confidence intervals of the observations. For 1-day maxima, the RFs show higher return levels for return periods greater than 10 years over the Thur river catchment at Jonschwil and at Andelfingen. For 3-day maxima, no such disagreement is noted. For 3-day maxima in the Inn and Kleine Emme river catchments, RFs indicate a smaller return value than observed above the 10-year return period. There is also a good agreement for several of the small catchments, however, over the Isorno river catchment, both 1- and 3-day

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

maxima deviate strongly from the observed return levels. The largest differences are found over the very small catchments starting already at a return period of 1 year.

The analog method was used to disaggregate daily precipitation. Results over the catchments are mixed, also within catchments of the same scale range. For many of the large and medium-sized catchments, results indicate a relatively good overlap with the intensity-duration-frequency (IDF) curves of the observations, in particular for the 2-year return level, with the exception of the Sarine, Inn and Minster river catchments (see Figs. S8 and S9 in the Supplement). For small and very small catchments, the downscaled RFs show overly strong precipitation intensities in particular for short durations (see Figs. S10 and S11 in the Supplement).

IDF curves for 1- to 7-day precipitation (Figs. S12–S15 in the Supplement) show good agreement between the downscaled RFs and the observations for the 2-year return level for all large and medium-sized catchments, and the small catchments Allenbach, Drance de Bagnes and Thur at Alt St. Johann. The remaining catchments show larger differences. For some of the catchments showing a good performance, there are indications that higher return level precipitation episodes spanning several days are underrepresented (e.g., the 25-year return level over the Aare at Thun), whereas they are represented well for other catchments (e.g., the Sarine river catchment).

Fig. 7 shows density plots based on annual and seasonal mean as well as daily mean temperature and precipitation for the large Aare River at Bern and the very small Krummbach River catchments. The observed seasonal and annual mean values (red crosses, rows 1, 2, 4 and 5) are within the density cloud of the RFs, except for single extreme years and seasons. For very small catchments, precipitation means tend to be closer towards the edge of the point cloud. Seasons in rows 1 and 4 were built using the RF forecasted days, while seasons in rows 2 and 5 are based on the 360-day calendar of the final time series. Not surprisingly, the concatenation of the single RFs to a 360-day calendar artificially increases seasonal variability, in particular for the transitional seasons. At the same time, observed extreme years indicated in the seasonal means are not depicted by the RFs, which is likely due to the opportunistic concatenation of the single RFs to yearly time series. Also for the remaining catchments there is overall a good agreement between RFs and observations (see Figs. S16–S19 in the Supplement). However, there are some indications of the RFs being slightly warmer than the observations in spring in the catchments of Thur (at Jonschwil and at Alt St. Johann), Maggia and Isorno.

4.2 Hydrological validation

To validate the suitability of the bias-corrected and downscaled RFs for hydrological simulations, we calculated flow duration curves (FDCs) from the RF-based simulations. To achieve these, we sampled daily values over the length of the observations (or, if shorter, the length of the control run) from each block of 1000 years and calculated the exceedance probabilities separately. For plotting, we calculated the 95% confidence intervals for each exceedance probability over the flow values and show these together with the observations and the control run. Comparing the FDCs from RF, control run and observations for selected sites (Fig. S20 in the Supplement), we find very similar behaviour for the Aare River at Bern, the Thur River at Andelfingen, and the Minster River. However, for the Maggia River and especially the Inn River, RFs and control run show higher values than observations in the upper flow ranges. In the case of the Maggia River, control run and RF-based simulation

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

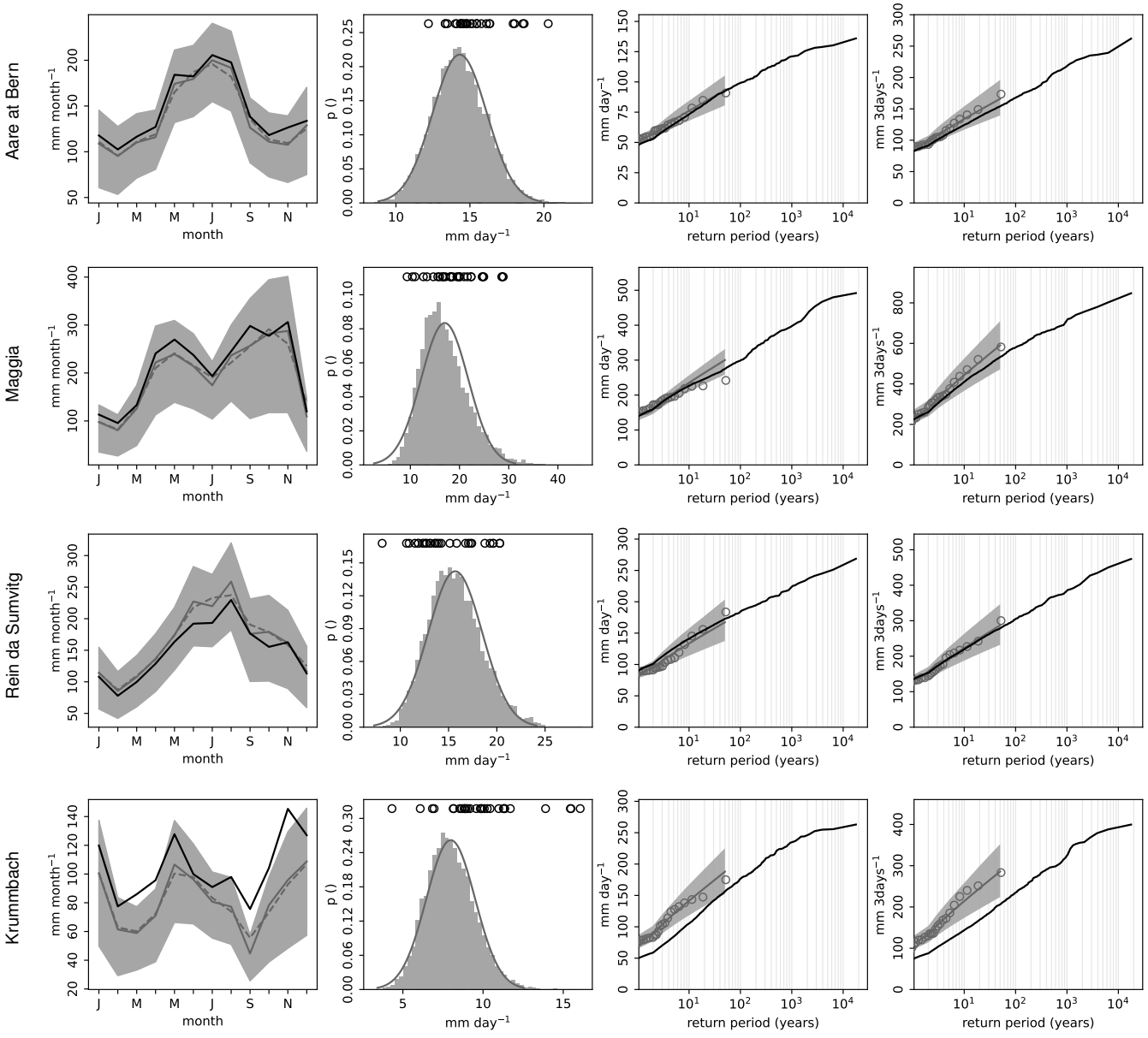

Figure 6. Evaluation of reforecast (RF) precipitation for selected catchments across different scale ranges: Aare at Bern (large), Maggia (medium), Rein da Sumvitg (small) and Krummbach (very small). From left to right: Annual cycle of mean monthly precipitation for observations (1991–2019, black) and RFs (grey; interquartile range (25^{th}–75^{th} percentile) in shading; mean based on 360-day calendar in dashed line); histogram of annual daily 90^{th} percentile precipitation for RFs, and observations (black circles); return levels for annual maximum daily and maximum 3-day precipitation for RFs (black) and observation (grey circles) using the Gringorten plotting position. For observations, return levels have additionally been fitted using a Gumbel distribution (using a GEV distribution resulted in overly broad confidence intervals) and maximum likelihood estimates, confidence intervals are based on 5000 bootstrap samples (mean in dark grey, 2.5^{th}–97.5^{th} percentile in grey shading). Results for all test catchments are shown in Figs. S4–S7.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

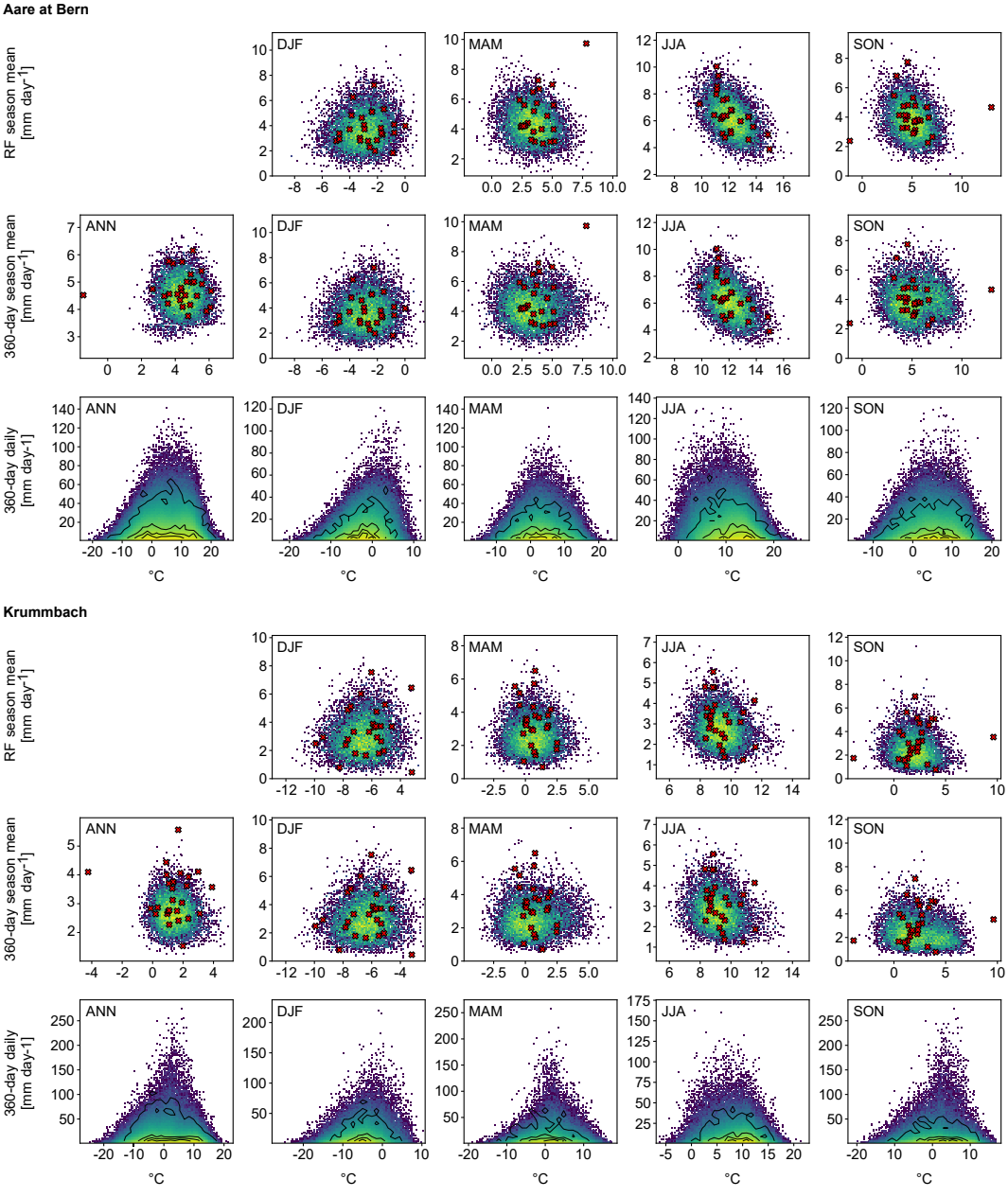

Figure 7. Density plots for the Aare River at Bern (top) and the Krummbach River (bottom), showing annual and seasonal precipitation and temperature based on seasonal means and the RF calendar dates (first row each), on seasonal means and the final 360-day calendar (second row each), and as based on daily means and the final 360-day calendar (third row each, days with a precipitation sum of $\geq$1 mm). Colours depict RF values, ranging from blue for low density to yellow for high density. Observations over the period 1991–2019 are denoted by red crosses (first and second row each) and contour lines (from low to high density in the innermost contours, third row each).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

agree well to very well. The discrepancy to observations is partly or entirely due to the absence of the large flood events of 1981 and 1982 in the records, as the station was destroyed by a massive flood in 1978 and did not resume operation until in 1995 (Näf-Huber et al., 2021). In the Saltina River, control run and observations agree well, but RFs show lower values overall, particularly in the upper flow ranges. This suggests limitations in the RF input for reproducing large floods in this small catchment.

We also examined the seasonality of AMF occurrence in the RF-based hydrological simulations and compared it with the seasonality in control run and observed AMFs at selected sites (Fig. S21 in the Supplement). The months with frequent AMF occurrence are generally consistent, with differences typically no greater than one month. However, discrepancies are noted in the strength of AMF seasonality in some cases. For example, in the Inn River, RF-based AMFs are more evidently distributed across the months of frequent occurrence and do not show the pronounced peak in observed AMF occurrence centered in June, which is well captured in the control run. For the interpretation of all sites, it should be noted that the time periods and their durations for AMF observations, control run and RFs do not fully align.

As individual RF segments were concatenated to obtain a continuous hourly time series (see Sect. 3.1.4), we verified whether the resulting AMFs do not show exceptional behaviour by comparing the time of occurrence and magnitude of the AMF with the stitching dates. For the smaller catchments, the densest occurrence of the AMFs is within a relatively narrow time window, occurring markedly after the stitching point, typically at least 100 days later. For the larger catchments, the AMFs are more widely distributed throughout the entire year following the stitching. Fig. S22 in the Supplement shows results for a small and a larger sub-catchment from each large river basin. In conclusion, the simulated AMFs appear to be unrelated in their magnitude and timing to the stitching points.

4.3 Flood estimation results

Exceedance curves based on RF inputs are shown for selected examples in Fig. 8. Looking at large catchments (Fig. 8, top row), very good agreement between RF-based AMFs and observed AMFs is found for the Aare River at Thun. Note that the observed time series is not stationary at this site (Bundesamt für Umwelt (BAFU), 2020) because a flood relief tunnel was taken into operation in 2009, and regulation rules for the upstream lakes Thun and Brienz were altered. The CS modelling chain used in this study depicts the current state, and comparison to observations of the period 2009–2021 is most appropriate, even though the statistics of observed floods from this period show an exceptionally wide confidence interval due to the short record length. Further downstream at Bern (not shown), the return levels based on RF simulations are higher in comparison to those of observations and GWEX, presumably due to overestimation of floods from smaller tributaries that join the Aare River. For the Thur River at Andelfingen, return levels are higher than expected from observations for control run, RF-based simulation and GWEX-based simulation for return periods greater than 5–10 years. It is difficult to decide whether this points at floods that could be systematically higher than expected, or whether the model tends to overestimate AMFs in this case. For the large Maggia River catchment, which exhibits very challenging meteorological conditions, the RF-based flood exceedance curve agrees remarkably well with that of the observations. The highest AMFs simulated from RFs, with estimated return periods

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

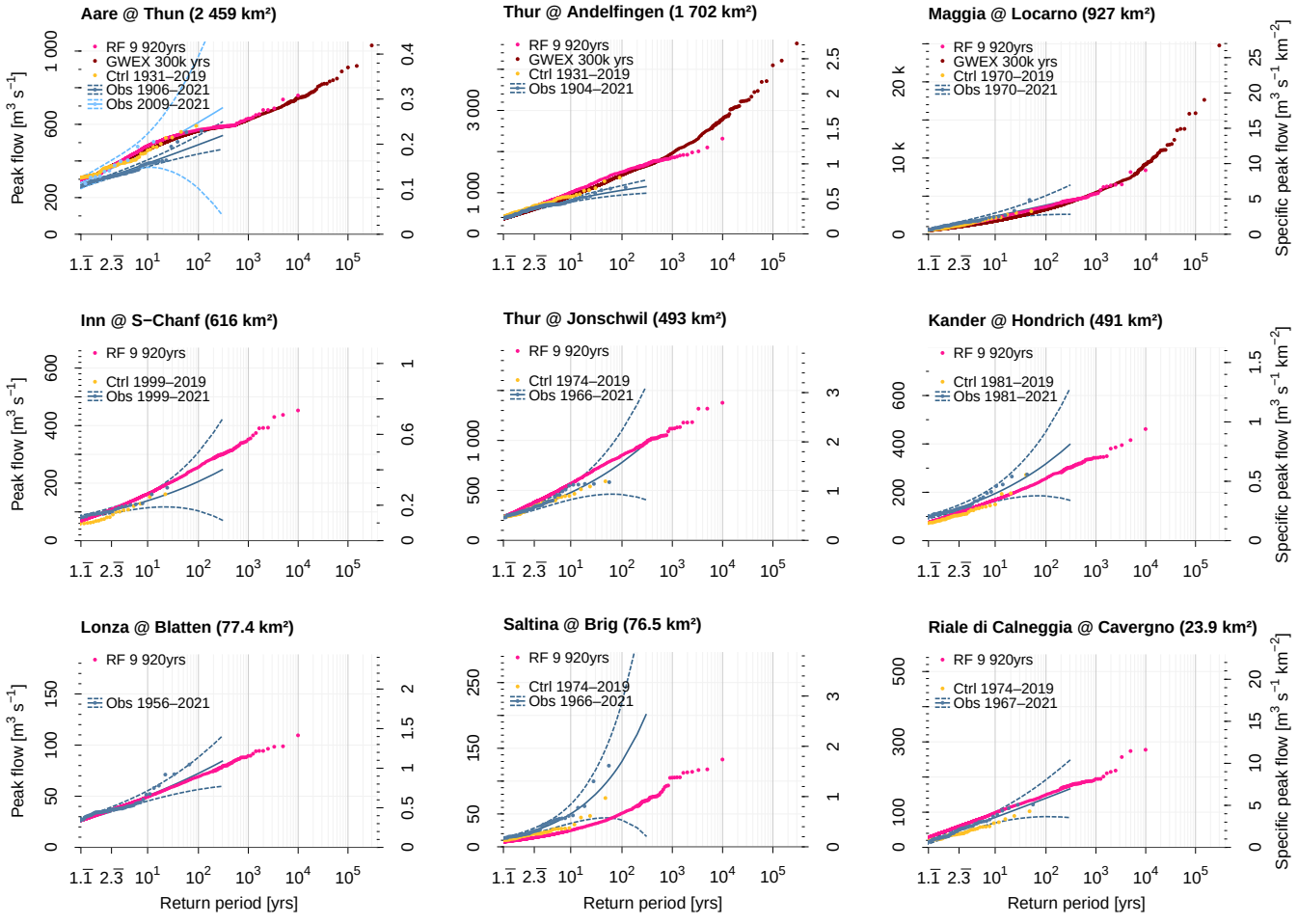

Figure 8. Exceedance curves from hydrological simulations using reforecast (RF) input. The top row shows three large catchments, for which long continuous simulations with input from the weather generator GWEX are available for comparison. The middle row shows three medium-sized catchments, the bottom row three small catchments. 'RF' refers to annual maximum floods (AMFs) from the long continuous simulation (CS) based on reforecasts, 'GWEX' to AMFs from the long CS based on weather generator input (only top row), 'Ctrl' to AMFs from a control run with observed weather, and 'Obs' are observed AMFs and a corresponding extrapolation.

between 1000 and 10 000 years, reach values of approximately 5300–8400 m^3s^{-1}. The corresponding return levels estimated from GWEX are approximately 5500–9150 m^3s^{-1}, which is higher, but does not indicate a fundamental disagreement.

For medium-sized catchments (Fig. 8, middle row), we note curves lower (Kander River) or higher (Inn River, Thur River at Jonschwil) than observed exceedance curves.

For small catchments (Fig. 8, bottom row), results matching well with observations are possible, such as for the Riale di Calneggia and Lonza rivers. But as expected, marked underestimation can also occur, like for the Saltina River, where the highest RF-based AMF is only slightly higher than the highest observed AMF in the 55 year long records.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Table 3. Precipitation gauges used for juxtaposition with reforecasts and weather generator (GWEX), with coordinates (x, y) and elevation (z).

Code	Name	x [m]	y [m]	z [m a.s.l.]
BOS	Bosco-Gurin	680 879	130 027	1486
CEV	Cevio	689 688	130 565	417
BEP	Belp	605 140	193 805	515
FRF	Frauenfeld	709 480	270 170	393
GRH	Grimsel Hospiz	668 583	158 215	1980
INT	Interlaken	633 019	169 093	577
LTB	Lauterbrunnen	635 851	160 291	815
MSG	Mosogno	692 803	117 050	771
SAE	Säntis	744 200	234 920	2502
URN	Urnäsch	739 205	241 765	825

5 Discussion

5.1 Comparison of reforecasts and weather generator precipitation

In the following, RF precipitation maxima are juxtaposed to the fundamentally different approach of constructing precipitation scenarios with the stochastic multi-site weather generator GWEX (see Sect. 2.5). Three catchments were selected for comparison as they well represent different conditions regarding climatology, station density and station representativeness: The Aare River at Bern, the Thur River at Andelfingen, and the Maggia River. These were evaluated in terms of annual peaks of areal precipitation, and three to four meteorological stations were selected in addition per river basin (Tab. 3) with focus on different station elevations.

To achieve a consistent comparison, a time series of 9900 years was taken each from RF and GWEX, and divided into 99 blocks with a length of 100 years each for computing confidence intervals. Observations (OBS) with a maximum length of 90 years in the period 1930–2019 were furthermore available for juxtaposition, with confidence intervals computed from a parametric bootstrap General Extreme Value (GEV) distribution fitted with L-Moments. The following discussion focuses on return periods that can be reasonably estimated from OBS, approximately 150 years as per the Gringorten (1963) plotting position.

It should be noted that GWEX was parameterized using OBS data from the same period 1930–2019, whereas the RF data cover the period 1991–2020 and were bias corrected using RhiresD over this period. Instead, the stochastic downscaling was calibrated using the full available RhiresD data 1961–2019 and station data 1961–2019. Caution is therefore warranted when comparing OBS, RF and GWEX, especially considering that the Atlantic Multidecadal Oscillation (AMO) index went through different phases between 1930 and 2019, indicating a relevant impact of internal climate variability (see also Sect. 5.5).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

The latest heavy precipitation statistics by MeteoSwiss (Fukutome et al., 2018, retrieved 27.03.2023) are shown for further reference. These use the entire measurement series available up to 2022 at each station, roughly compatible with the block length chosen for RF and GWEX.

At-station maximum 1-day and 3-day precipitation sums (Fig. 9, columns 1–4) generally show good to very good agreement between OBS, RF and GWEX for all three catchments considered. The largest disagreement is noted at the station Mosogno (MSG) in the Ticino River basin, where RF values are lower than OBS and GWEX. At the level of mean catchment precipitation, agreement is also high in all three river basins (Fig. 9, right). GWEX confidence intervals tend to exceed those of the RFs in the upper range of maximum 1-day and 3-day precipitation sums for return periods larger than 50 to 100 years at both station and river basin level. This also means that GWEX in general reaches higher precipitation extremes in the 9900 years of data analysed as compared to the RFs. Given the differences between how the RF and GWEX time series are composed, it appears not possible to draw firm conclusions beyond this statement. Also, it does not appear justified to make comparisons to the range found for observations since their confidence intervals are derived from a maximum of 90 years of record only. Both for 1-day and 3-day maxima, it is not possible to determine whether RFs or GWEX is closer to reality as regards high extremes.

5.2 Envelope curves for floods

Fig. 10 shows an evaluation of the ten highest RF-based flood estimates in the context maximum flood (Qmax) data from across Switzerland (Kienzler and Scherrer, 2018). Two envelope curves based on these data are shown. Their slope was derived from a log-log linear regression, after which the intercept was adjusted to either encompass all data or 95% of the data, respectively. Additionally, envelope curves for European (Bertola et al., 2023) and global (Herschy, 2003) maximum floods are shown.

Overall, the magnitude and estimated frequency of RF-based floods appear plausible, although they tend to be lower at smaller scales than expected from Qmax data and envelope curves. This is evident in catchments where storms increasingly dominate the generation of AMFs. Such catchments would require dynamical downscaling of RF, rather than stochastic downscaling, which was not feasible with the available data (see Sect. 5.3). In contrast, RF-based values for the Ticino region – specifically the Isorno River (IsoMsg, 124.7 km^2) and the Maggia River (MagLcn, 926.9 km^2) – are notably higher than in other regions. This is expected given the region's climatological characteristics, which favour heavy precipitation events, as well as the crystalline bedrock, which promotes a rapid and strong flood response. The comparatively low values for the Aare River at Thun (AarThu, 2459 km^2) and at Bern (AarBrn, 2965 km^2) can be attributed to the marked attenuation of floods by Lake Brienz and Lake Thun further upstream, which is represented in the simulation chain but not in the envelope curve.

The highest ten GWEX-derived values from the full 300 000-year simulations often reach considerably higher magnitudes than those derived from RFs (note the logarithmic y-axis). This is not surprising given the substantially longer duration of the weather generator scenarios. However, when considering only 9920 years of GWEX weather scenarios – matching the length of the RFs – the magnitudes align more closely. Notably, the exceptionally high RF results for the Ticino region are consistent with GWEX results.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

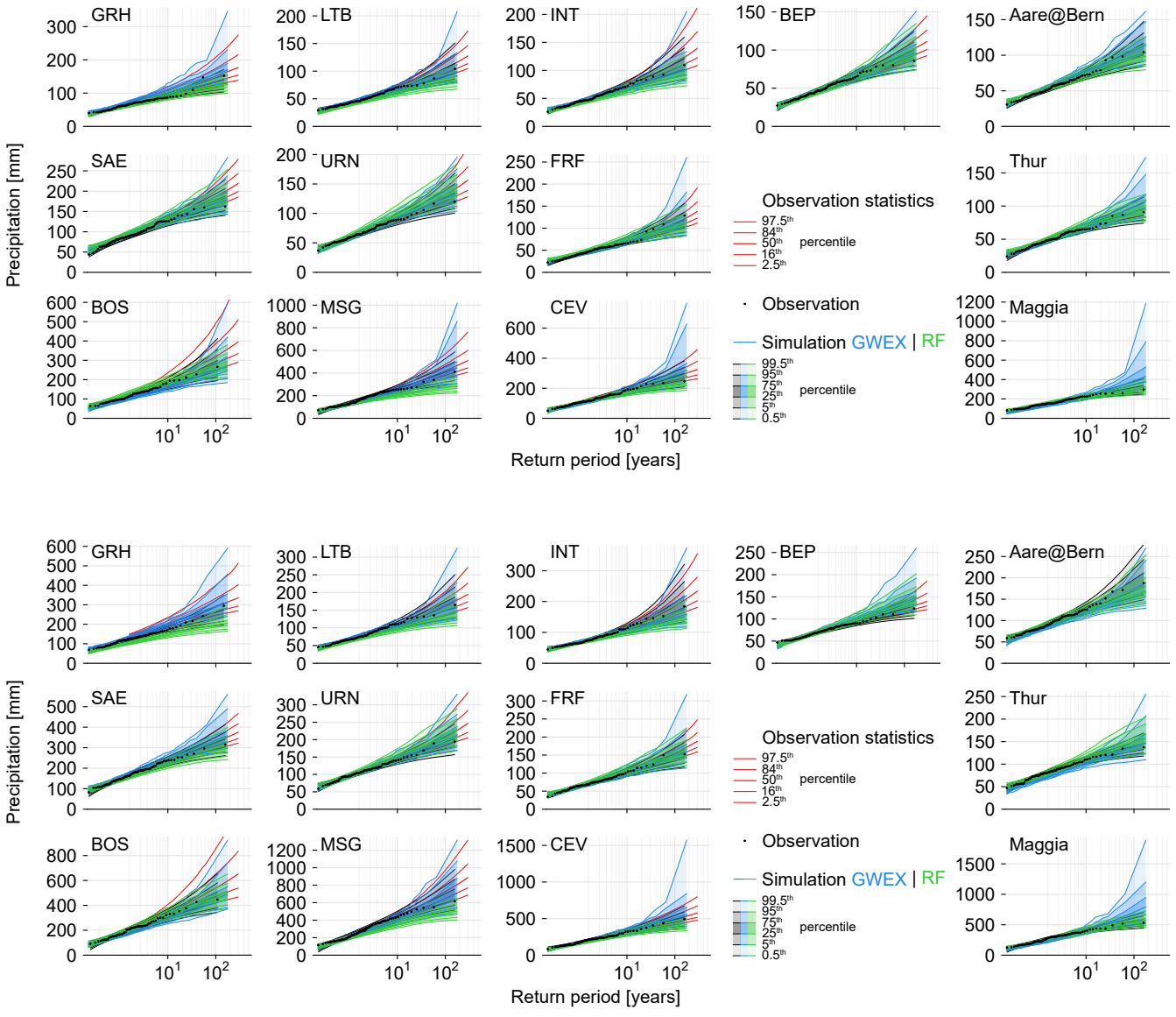

Figure 9. Comparison of maximum 1-day (top) and 3-day (bottom) precipitation sums from reforecasts (RF), weather generator (GWEX) and observations (OBS). Columns 1–4 show selected precipitation gauging stations (see Tab. 3) within the full catchment shown in column 5.

A comparison within the context of large river basins (Fig. S23 in the Supplement) shows that the RF-based flood estimates align well with the regional patterns and are often positioned near the regional 95th percentile envelope. Values that are noticeably lower can be explained by lake retention in the case of the Aare River at Thun and Bern, as noted above, and by a combination of slightly different climatological regime and small catchment area in the case of the Krummbach (KruKlu),

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

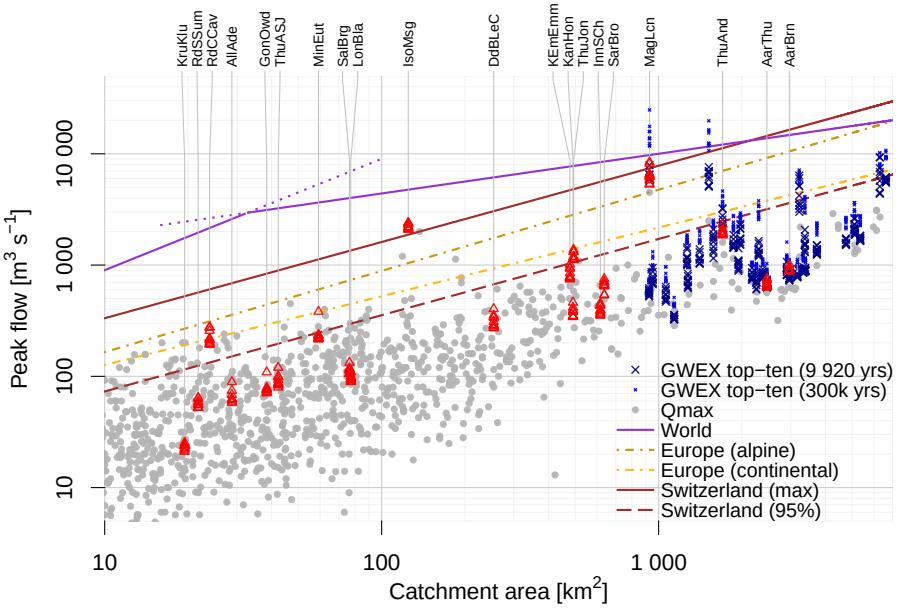

Figure 10. Highest ten annual maximum floods simulated using reforecast (RF) and weather generator (GWEX) input. Maximum observed floods in Switzerland as per Kienzler and Scherrer (2018) (gray dots) are shown for context, along with corresponding envelope curves (the solid line refers to all data, the dashed line to the 95th percentile). Also shown are envelope curves for maximum floods in Europe (Bertola et al., 2023) (two regions are relevant: alpine, applies to AarBrn, AarThu, KEmEmm, MinEut, SarBro, ThuAnd, ThuJon; continental, applies to the remaining catchments) and worldwide (Herschy, 2003). For catchment IDs see Table 1, for analyses separately for large river basins see Fig. S23.

which is the smallest test catchment considered. Compared to other large river basins, flood estimates for sites not influenced by lake attenuation fall somewhat more noticeably below the regional 95th percentile envelope in the Aare River basin.

5.3 Feasibility of dynamical downscaling

Statistical and dynamical downscaling each have their own pros and cons, and should not be understood as competing approaches. Statistical downscaling is comparatively cheap but involves a trade-off between model complexity and realism, which among other things has an impact on spatio-temporal consistency, replication of extremes and temporal disaggregation. In contrast, dynamical downscaling is based on models that use fundamental laws of physics, providing physically and spatio-temporally consistent results. However, its high computational demand limits its applicability. Specifically, within the framework presented in this paper, dynamical downscaling of the entire set of nearly 10 000 years of RF data is not feasible, requiring the selection of extreme events for dynamic downscaling.

We tested the feasibility of dynamically downscaling the ECMWF RFs to the convection-permitting scale (~3 km grid size), which would improve the representation of short local extreme precipitation events and be valuable for small catchments. To model rapid regional weather changes, as occurring during small-scale extreme events, the raw data (in our case RFs) should

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

be available at high temporal resolution. Typically, these data (e.g., reanalyses or data from global climate models) are stored every 6 or 3 hours and have to be interpolated to the required temporal resolution. The ECMWF RFs we used were stored at 6-hour resolution for the first two forecast weeks, but only at 12-hour resolution for the time beyond. This poses the risk that short-term extreme events, lasting no more than a few hours, will not be captured. This risk emerges due to the methodology of dynamical downscaling: The limited-area regional model needs atmospheric driving data at each model time step (20 seconds) along its lateral boundaries. These driving data are interpolated linearly in time from the available RF data. Hence, if the RF data are available only on a 12-hour resolution, processes beyond this resolution are hardly represented, although the regional model is to a certain extent capable of reintroducing sub-daily processes.

We therefore compared whether downscaling applied to 12-hourly input data is capable of simulating local extreme events that essentially correspond to those from downscaled hourly data. Since the RF data are only available at a coarse temporal resolution, ERA5 reanalysis data (Hersbach et al., 2020) with a grid width of 25 km and a temporal resolution of 1 hour were used as alternative input data. The ECMWF forecast model underlying ERA5 is very similar to that used for the RFs. The CCLM model (Rockel et al., 2008; Schättler et al., 2021) was used for the dynamic downscaling.

The two extreme events of 8 July 1996 and 23 June 2021 were considered. While the simulation results looked realistic at intervals of 1 hour, there was a marked reduction in the simulated extreme precipitation as well as spatial distortions in the precipitation field at 12-hour intervals. Due to the reduced simulated precipitation, there is a risk that extreme events (which are not based exclusively on slow processes) will no longer be represented as such.

In summary, it was not possible to identify the relevant extreme events for dynamical downscaling because the link between the scales was not pronounced enough, especially with the limited 12-hour resolution available. For successful selection, the ECMWF RF data would need to be available at a higher temporal and spatial resolution to simulate relatively small-scale extreme precipitation events (~500 km^2 or smaller). While dynamic downscaling would work for large-scale extreme events, no additional benefit is expected compared to stochastic downscaling. In the future, dynamic downscaling would be highly relevant if RF data became available at a higher temporal resolution.

5.4 Flood return levels

Some caution is warranted when interpreting the estimated return periods of AMFs derived from the RF-based long CS: Although the individual RF ensemble members are assumed independent after discarding days 1 to 15, their ability to accurately represent the flood-relevant meteorological behaviour across different regions cannot be examined systematically with the test catchments used in this proof of concept. The exceedance curves for the larger test catchments studied, however, indicate a generally plausible range when compared to observed AMFs and their extrapolation. However, it should be borne in mind that the observed AMFs are also subject to uncertainty (Westerberg et al., 2020), and that the observed AMFs available are instantaneous peak values, as opposed to the hourly simulated values.

An important question to which RF-based model runs can contribute is how the exceedance curve behaves as the return period increases. Notably, the observed AMFs sometimes appear to level off, whereas GWEX-derived AMFs typically do not. Interestingly, the RF-derived AMFs also generally show no levelling off or a tendency towards a plateau value. Moreover,

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

the agreement between RF and GWEX-based flood estimates in the tail is remarkably high where comparison can be made (Fig. 8), particularly also in the case of the Maggia River, which shows an exceptionally heavy tail. This general agreement is noteworthy because the RF offer a more physically based approach to precipitation scenarios, whereas the GWEX weather generator employs a stochastic approach without explicit physical boundary conditions. This suggests that indeed the AMF observations – paired with extrapolation using a GEV distribution – might underestimate the potential of catchments to generate increasingly higher floods within the return periods considered here. This interpretation is preliminary, as the range of test catchments examined in this study is limited, and the return periods covered extend only to 1000 to 10 000 years.

5.5 Limitations

Some methods, such as Thiessen interpolation of point values, precipitation adjustment factor and temperature lapse rates, were adopted to enable long CSs within the EXAR and EXCH project frameworks (Viviroli et al., 2022) (Sect. 3.2). Along with statistical RF downscaling, these methods have limitations for smaller catchments, which however were not the focus of EXAR and EXCH but here serve to explore the RF-CS approach across a broad range of scales from approximately 20 to 3000 km^2. Addressing some of the associated complexities – such as through dynamical downscaling, targeting mean areal precipitation and temperature, or even a spatially distributed modelling approach – will be time-consuming and require further research.

Further limitations are related to the resolution of the RF data. Initially, we planned to use both the ENS Model Climate (day 1–15) and the Extended Range Model Climate (day 16–45) for the statistical downscaling. However, due to inconsistencies in the data and concerns about the independence of short forecasting times from observed weather, only the Extended Range Model Climate data were used (see Sect. 2.2). In consequence, the statistically downscaled time series were only 9920 years long. On the other hand, ENS Model climate data are available at 0.25° resolution, whereas the longer forecasts are available on a coarser 0.5° grid, limiting the applied approach in terms of catchment size.

Since its introduction, the ECMWF IFS and hence the RFs have undergone continuous updates (see Sect. 2.2). When building the yearly time series required for the hydrological modelling, we were drawing from RFs initialised in the same year. For many initialisation years, this resulted in concatenation of RFs from different IFS cycles. Certainly, over the 12 initialisation years of RFs, considerable model evolution has occurred. However, while the extent of updates between successive IFS cycles varies, the updates are gradual, often introducing only minor changes. The resulting yearly time series are hence, at most, based on slightly different IFS model versions.

While results of the statistical downscaling approach match well with observations for medium and large catchments, the stochastic downscaling is not well suited for very small catchments, where only a few stations are scattered over few or even only one grid-point. The Extended Range Model Climate RFs used have a spatial resolution of 0.5° (~38×56 = 2128 km^2 in the study domain), whereas some of the catchments examined cover an area of less than 50 km^2, and only 2–3 stations were used in the calibration of the stochastic downscaling. We noted a systematically poor performance for these catchments. The smaller number of stations, together with the coarse resolution of the RFs, impedes estimating reliable statistical relationships between the large and small scale during the calibration of the stochastic downscaling.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

For the temporal disaggregation, we have experimented with several procedures to draw the best analogs. Additionally, we initially planned to disaggregate temperature using the analog method, alike precipitation. We tried using mean daily temperatures and absolute values of temperature as well as differences between minimum and maximum temperatures at the station level, all of which led to very strong temperature jumps when concatenating the single RFs to the yearly time series. To smooth those steep changes from one hour to the next, we pragmatically interpolated between the bias corrected 6-hourly temperature values. For precipitation, the applied analogs performed satisfactorily for large and medium sized catchments, but showed excessive precipitation intensities in particular for short durations and in small to very small catchments.

The concatenation of individual RFs to yearly time series furthermore disregarded other potentially more important aspects. Future efforts could seek to additionally incorporate variables accounting for the dynamic state of the atmosphere such as geopotential height, and aim at sampling inter-annual variability which might require the reuse of single RFs when building the annual time series. However, comparison of observations and RFs on a seasonal and annual scale indicated that the time series produced do not contain years with values overly above or below observed precipitation and temperature.

The RFs have been introduced in March 2008, reforecasting weather occurring up to 18 years prior. Counting full years only, the RFs used here span the period 1991–2019 (excluding 2015, see Sect. 2.2), with fewer data in the beginning and end of that period and more in the middle. Counting the number of RF days per month and year revealed that the Atlantic Multidecadal Oscillation (AMO) has been predominantly in a positive phase over the RF period (see Trenberth et al., 2023), in particular for the middle of the period for which most RF were available (see Fig. S24 in the Supplement). This results in an oversampling of positive AMO phases in the RFs processed here.

With ongoing climate change, mean precipitation over the Alps is projected to increase during winter and decrease during summer. Virtually all state-of-the art global and regional climate model ensembles indicate an increase in extreme precipitation over the Alpine region, except for summer, where only high resolution global climate models and regional climate models indicate an increase (Ritzhaupt and Maraun, 2023). The RF simulation period spans the period 1991–2019, which was also used for the bias adjustment. The RFs are thus representative for the current state of the climate, whereas aspects of non-stationarity are not accounted for. Given the dominant role of precipitation for the subsequent hydrological simulations (Kritidou et al., 2025), any resultant flood estimates should be re-assessed within a time-frame of approximately 10 years.

6 Conclusions

We have demonstrated the feasibility of processing extensive RF data to generate long continuous inputs for a hydrological model, with statistical downscaling proving to be a suitable approach. Good meteorological and hydrological results were achieved for large and medium-sized catchments (larger than approximately 500 km^2), while limitations were identified for small and very small catchments. In addition to contributing information on rare floods, the resulting RF-based flood estimates provide valuable insights for verifying the magnitude of flood estimates derived from other continuous simulation studies, such as those using weather generator inputs. In our study, we found that the stochastic weather generator employed as an

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

alternative is not likely to underestimate possible but unobserved floods. Feasibility is also ensured in terms of computational cost, as processing for the largest catchments was run within a few hours in the current setting using daily data as a reference.

In the future, ways to apply the stochastic downscaling to higher resolved temporal data should be explored. In our proof-of-concept study, one obstacle in this regard was data availability, in particular that of gridded observations at high temporal resolution. Although the hourly CombiPrecip dataset currently provides data only from 2005 onward, the gap to create a sufficiently long data record for use as a predictor dataset within the stochastic downscaling framework is expected to close over time. Also regarding station data, ongoing measurements will improve the data basis. In the meantime, disaggregated data could be generated using other sophisticated approaches over the applied analogs, such as stochastic disaggregation (Kossieris et al., 2018). As a way forward for achieving better results in small and very small catchments, the possibilities of dynamical downscaling should be further explored, on the condition that RF data become available at higher temporal and spatial resolution.

Other possible avenues to derive local information from the RFs include the use of emulators (see Maraun and Widmann, 2018, for an overview). Doury et al. (2023, 2024) proposed a regional climate model (RCM) emulator, combining the strengths of both empirical statistical downscaling methods and dynamical downscaling via neural networks. The emulator learns the complex spatial structure and daily variability simulated by the RCMs, particularly how the RCM refines the low-resolution climate patterns, and could be applied to the RFs data directly. Paired with recent efforts in convection permitting simulations such as the CORDEX FPS on convective phenomena over Europe and the Mediterranean (Ban et al., 2021) this approach could proof very useful.

Code and data availability. Code used in this project and resulting data can be obtained from the first author upon reasonable request.

Author contributions. Conceptualization: DM; funding acquisition: DV; methodology: DM, MJ, DV; investigation: MJ, DV, MS, MK, HT; visualization: MJ, DV; original draft preparation: DV, MJ; review and editing: all authors.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. We thank the ECMWF for producing and making available the vast database of RFs that made this study possible. We also thank MeteoSwiss, the Federal Office for the Environment FOEN, as well as the cantons of St. Gallen and Ticino for providing hydrometeorological data. Heimo Truhetz gratefully acknowledges the computational resources granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at the Julich Supercomputing Centre (JSC) through grant JJSC39 and by the Vienna Scientific Cluster (VSC) through grant 71193.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Financial support. This research was funded by the Federal Office for the Environment FOEN and the Swiss Federal Office of Energy SFOE as a part of the project “Extreme Floods in Switzerland” (EXCH).

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

References

Andres, N., Steeb, N., Badoux, A., and Hegg, C., eds.: Grundlagen Extremhochwasser Aare: Hauptbericht Projekt EXAR. Methodik und Resultate, vol. 104 of *WSL Berichte*, Swiss Federal Institute for Forest Snow and Landscape Research WSL, Birmensdorf, https://www.wsl.ch/de/publikationen/extremhochwasser-an-der-aare-hauptbericht-projekt-exar-methodik-und-resultate/, 2021.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., Bastin, S., Belušić, D., Berthou, S., Brisson, E., Cardoso, R. M., Chan, S. C., Christensen, O. B., Fernández, J., Fita, L., Frisius, T., Gašparac, G., Giorgi, F., Goergen, K., Haugen, J. E., Hodnebrog, Ø., Kartsios, S., Katragkou, E., Kendon, E. J., Keuler, K., Lavin-Gullon, A., Lenderink, G., Leutwyler, D., Lorenz, T., Maraun, D., Mercogliano, P., Milovac, J., Panitz, H.-J., Raffa, M., Remedio, A. R., Schär, C., Soares, P. M. M., Srnec, L., Steensen, B. M., Stocchi, P., Tölle, M. H., Truhetz, H., Vergara-Temprado, J., de Vries, H., Warrach-Sagi, K., Wulfmeyer, V., and Zander, M. J.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation, Climate Dynamics, 57, 275–302, https://doi.org/10.1007/s00382-021-05708-w, 2021.

Baumgartner, E., Boldi, M.-O., Kan, C., and Schick, S.: Hochwasserstatistik am BAFU – Diskussion eines neuen Methodensets, wasser, energie, luft, 105, 103–110, 2013.

Bertola, M., Blöschl, G., Bohac, M., Borga, M., Castellarin, A., Chirico, G. B., Claps, P., Dallan, E., Danilovich, I., Ganora, D., Gorbachova, L., Ledvinka, O., Mavrova-Guirguinova, M., Montanari, A., Ovcharuk, V., Viglione, A., Volpi, E., Arheimer, B., Aronica, G. T., Bonacci, O., Čanjevac, I., Csik, A., Frolova, N., Gnandt, B., Gribovszki, Z., Gül, A., Günther, K., Guse, B., Hannaford, J., Harrigan, S., Kireeva, M., Kohnová, S., Komma, J., Kriauciuniene, J., Kronvang, B., Lawrence, D., Lüdtke, S., Mediero, L., Merz, B., Molnar, P., Murphy, C., Oskoruš, D., Osuch, M., Parajka, J., Pfister, L., Radevski, I., Sauquet, E., Schröter, K., Šraj, M., Szolgay, J., Turner, S., Valent, P., Veijalainen, N., Ward, P. J., Willems, P., and Zivkovic, N.: Megafloods in Europe can be anticipated from observations in hydrologically similar catchments, Nature Geoscience, 16, 982–988, https://doi.org/10.1038/s41561-023-01300-5, 2023.

Brunner, M. I. and Slater, L. J.: Extreme floods in Europe: going beyond observations using reforecast ensemble pooling, Hydrology and Earth System Sciences, 26, 469–482, https://doi.org/10.5194/hess-26-469-2022, 2022.

Bundesamt für Umwelt (BAFU): Hochwasserstatistik Stationsbericht Aare - Thun, https://www.hydrodaten.admin.ch/documents/Hochwasserstatistikberichte/2030_hq_Bericht.pdf, 2020.

Bundesamt für Umwelt BAFU: Die biogeografischen Regionen der Schweiz, 2022.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth System Science Data, 12, 2097–2120, https://doi.org/10.5194/essd-12-2097-2020, 2020.

de Bruijn, K., van den Hurk, B., Slager, K., Rongen, G., Hegnauer, M., and van Heeringen, K. J.: Storylines of the impacts in the Netherlands of alternative realizations of the Western Europe July 2021 floods, Journal of Coastal and Riverine Flood Risk, 2, https://doi.org/10.59490/jcrfr.2023.0008, 2023.

Doury, A., Somot, S., Gadat, S., Ribes, A., and Corre, L.: Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach, Climate Dynamics, 60, 1751–1779, https://doi.org/10.1007/s00382-022-06343-9, 2023.

Doury, A., Somot, S., and Gadat, S.: On the suitability of a convolutional neural network based RCM-emulator for fine spatio-temporal precipitation, Climate Dynamics, 62, 8587–8613, https://doi.org/10.1007/s00382-024-07350-8, 2024.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

ECMWF: Integrated Forecasting System, https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model, 2024a.

ECMWF: Changes to the forecasting system, https://confluence.ecmwf.int/display/FCST/Changes+to+the+forecasting+system, 2024b.

ECMWF: IFS cycle upgrades pre 2015, https://confluence.ecmwf.int/display/FCST/IFS+cycle+upgrades+pre+2015, 2024c.

Evin, G., Favre, A.-C., and Hingray, B.: Stochastic generation of multi-site daily precipitation focusing on extreme events, Hydrology and Earth System Sciences, 22, 655–672, https://doi.org/10.5194/hess-22-655-2018, 2018.

Evin, G., Favre, A.-C., and Hingray, B.: Stochastic generators of multi-site daily temperature: comparison of performances in various applications, Theoretical and Applied Climatology, 135, 811–824, https://doi.org/10.1007/s00704-018-2404-x, 2019.

Farinotti, D., Usselmann, S., Huss, M., Bauder, A., and Funk, M.: Runoff evolution in the Swiss Alps: projections for selected high-alpine catchments based on ENSEMBLES scenarios, Hydrological Processes, 26, 1909–1924, https://doi.org/10.1002/hyp.8276, 2012.

Fukutome, S., Schindler, A., and Capobianco, A.: MeteoSwiss extreme value analyses: User manual and documentation, 2018.

Ganapathy, A., Hannah, D. M., and Agarwal, A.: Improved estimation of extreme floods with data pooling and mixed probability distribution, Journal of Hydrology, 629, 130 633, https://doi.org/10.1016/j.jhydrol.2024.130633, 2024.

García Hernández, J., Foehn, A., Fluixá-Sanmartín, J., Roquier, B., Brauchli, T., Paredes Arquiola, J., and de Cesare, G.: RS MINERVE – Technical manual, v2.25, https://crealp.ch/wp-content/uploads/2021/09/rsminerve_technical_manual_v2.25.pdf, 2020.

Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B.: Weather Radar in Complex Orography, Remote Sensing, 14, 503, https://doi.org/10.3390/rs14030503, 2022.

Gringorten, I. I.: A plotting rule for extreme probability paper, Global Biogeochemical Cycles, 68, 813–814, https://doi.org/10.1029/JZ068i003p00813, 1963.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., de Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Herschy, R. W.: World Catalogue of Maximum Observed Floods, vol. 284 of *IAHS publication*, IAHS Press, Wallingford, ISBN 1-901502-47-3, 2003.

Hillier, J. K. and Dixon, R. S.: Seasonal impact-based mapping of compound hazards, Environmental Research Letters, 15, 114 013, https://doi.org/10.1088/1748-9326/abbc3d, 2020.

Hundecha, Y., Arheimer, B., Donnelly, C., and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, Journal of Hydrology: Regional Studies, 6, 90–111, https://doi.org/10.1016/j.ejrh.2016.04.002, 2016.

Jakob Themeßl, M., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31, 1530–1544, https://doi.org/10.1002/joc.2168, 2011.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019, 2019.

Kay, A. L., Dunstone, N., Kay, G., Bell, V. A., and Hannaford, J.: Demonstrating the use of UNSEEN climate data for hydrological applications: case studies for extreme floods and droughts in England, Natural Hazards and Earth System Sciences, 24, 2953–2970, https://doi.org/10.5194/nhess-24-2953-2024, 2024.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Kay, G., Dunstone, N., Smith, D., Dunbar, T., Eade, R., and Scaife, A.: Current likelihood and dynamics of hot summers in the UK, Environmental Research Letters, 15, 094 099, https://doi.org/10.1088/1748-9326/abab32, 2020.

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., Bohlinger, P., Ferranti, L., and Nipen, T.: Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes, npj Climate and Atmospheric Science, 3, https://doi.org/10.1038/s41612-020-00149-4, 2020.

Kelder, T., Marjoribanks, T. I., Slater, L. J., Prudhomme, C., Wilby, R. L., Wagemann, J., and Dunstone, N.: An open workflow to gain insights about low–likelihood high–impact weather events from initialized predictions, Meteorological Applications, 29, https://doi.org/10.1002/met.2065, 2022.

Kelder, T., Heinrich, D., Klok, L., Thompson, V., Goulart, H. M. D., Hawkins, E., Slater, L. J., Suarez-Gutierrez, L., Wilby, R. L., Coughlan de Perez, E., Stephens, E. M., Burt, S., van den Hurk, B., de Vries, H., van der Wiel, K., Schipper, E. L. F., Carmona Baéz, A., van Bueren, E., and Fischer, E. M.: How to stop being surprised by unprecedented weather, Nature Communications, 16, 2382, https://doi.org/10.1038/s41467-025-57450-0, 2025.

Kienzler, P. M. and Scherrer, S.: Verzeichnis grosser Hochwasserabflüsse in Schweizerischen Einzugsgebieten: Auswertung und graphische Aufbereitung, Reinach, https://www.bafu.admin.ch/bafu/de/home/themen/wasser/zustand/wasser--methoden/grosse_hochwasserabflusse-2.html, 2018.

Klehmet, K., Berg, P., Bozhinova, D., Crochemore, L., Du, Y., Pechlivanidis, I., Photiadou, C., and Yang, W.: Robustness of hydrometeorological extremes in surrogated seasonal forecasts, International Journal of Climatology, 44, 1725–1738, https://doi.org/10.1002/joc.8407, 2024.

Kossieris, P., Makropoulos, C., Onof, C., and Koutsoyiannis, D.: A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures, Journal of Hydrology, 556, 980–992, https://doi.org/10.1016/j.jhydrol.2016.07.015, 2018.

Kritidou, E., Kauzlaric, M., Staudinger, M., Evin, G., Hingray, B., Vis, M., Seibert, J., and Viviroli, D.: Impact of different weather generator scenarios on extreme flood estimates in Switzerland, Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-024-02843-8, 2025.

Lamb, R., Faulkner, D., Wass, P., and Cameron, D.: Have applications of continuous rainfall-runoff simulation realised the vision for process-based flood frequency analysis?, Hydrological Processes, 30, 2463–2481, https://doi.org/10.1002/hyp.10882, 2016.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, Journal of Hydrology, 201, 272–288, https://doi.org/10.1016/S0022-1694(97)00041-3, 1997.

Mahlstein, I., Bhend, J., Spirig, C., and Martius, O.: Developing an Automated Medium-Range Flood Awareness System for Switzerland Based on Probabilistic Forecasts of Integrated Water Vapor Fluxes, Weather and Forecasting, 34, 1759–1776, https://doi.org/10.1175/WAF-D-18-0189.1, 2019.

Maraun, D. and Widmann, M.: Statistical downscaling and bias correction for climate research, Cambridge University Press, Cambridge, ISBN 9781107066052, https://doi.org/10.1017/9781107588783, 2018.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, 764–773, https://doi.org/10.1038/NCLIMATE3418, 2017.

Margot, A., Schädler, B., Sigg, R., and Weingartner, R.: Influence on Rivers by Water Power Stations (≥300kW) and the Lake Control, vol. 5.3 of *Hydrologischer Atlas der Schweiz*, Bundesamt für Umwelt BAFU, Bern, 1992.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Mas, A., Evin, G., and Hingray, B.: Simulation of MAT and MAP scenarios using the weather generators GWEX and SCAMP: EXCH Final working report – Large catchments, 2023.

Ménégoz, M., Valla, E., Jourdain, N. C., Blanchet, J., Beaumet, J., Wilhelm, B., Gallée, H., Fettweis, X., Morin, S., and Anquetin, S.: Contrasting seasonal changes in total and intense precipitation in the European Alps from 1903 to 2010, Hydrology and Earth System Sciences, 24, 5355–5377, https://doi.org/10.5194/hess-24-5355-2020, 2020.

MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Precipitation (final analysis): RhiresD, https://www.meteoswiss.admin.ch/dam/jcr:4f51f0f1-0fe3-48b5-9de0-15666327e63c/ProdDoc_RhiresD.pdf, 2021a.

MeteoSwiss: Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD, https://www.meteoswiss.admin.ch/dam/jcr:818a4d17-cb0c-4e8b-92c6-1a1bdf5348b7/ProdDoc_TabsD.pdf, 2021b.

MeteoSwiss: MeteoSwiss – Open Data, https://github.com/MeteoSwiss/opendata/blob/main/README.md, 2025.

Näf-Huber, D., Scherrer, S., and Kienzler, P.: Le grandi piene della Maggia (Ct. Ticino). Studio principale. Derivazione delle portate di piena per diversi periodi di ritorno lungo la Maggia, Reinach, 2021.

Osinski, R., Lorenz, P., Kruschke, T., Voigt, M., Ulbrich, U., Leckebusch, G. C., Faust, E., Hofherr, T., and Majewski, D.: An approach to build an event set of European windstorms based on ECMWF EPS, Natural Hazards and Earth System Sciences, 16, 255–268, https://doi.org/10.5194/nhess-16-255-2016, 2016.

Owens, R. and Hewson, T.: ECMWF Forecast User Guide. Technical report, https://www.ecmwf.int/node/16559, 2018.

Redmond, K. T., Enzel, Y., House, P. K., and Biondi, F.: Climate Variability and Flood Frequency at Decadal to Millennial Time Scales, in: Ancient floods, modern hazards, edited by House, P. K., Water Science and Application, pp. 21–45, American Geophysical Union, Washington, DC, ISBN 9781118665701, https://doi.org/10.1029/WS005p0021, 2010.

Ritzhaupt, N. and Maraun, D.: Consistency of Seasonal Mean and Extreme Precipitation Projections Over Europe Across a Range of Climate Model Ensembles, Journal of Geophysical Research: Atmospheres, 128, e2022JD037 845, https://doi.org/10.1029/2022JD037845, 2023.

Robertson, A. W. and Vitart, F., eds.: Sub-Seasonal to Seasonal Prediction, Elsevier, ISBN 9780128117149, https://doi.org/10.1016/C2016-0-01594-2, 2019.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteorologische Zeitschrift, 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, 2008.

Ruelland, D.: Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models?, Hydrology and Earth System Sciences, 24, 2609–2632, https://doi.org/10.5194/hess-24-2609-2020, 2020.

Schättler, U., Doms, G., and Schraff, C.: COSMO-Model Version 6.00: A Description of the Nonhydrostatic Regional COSMO-Model - Part VII: User's Guide, https://doi.org/10.5676/DWD_PUB/NWV/COSMO-DOC_6.00_VII, 2021.

Schmocker-Fackel, P. and Naef, F.: Changes in flood frequencies in Switzerland since 1500, Hydrology and Earth System Sciences, 14, 1581–1594, https://doi.org/10.5194/hess-14-1581-2010, 2010a.

Schmocker-Fackel, P. and Naef, F.: More frequent flooding? Changes in flood frequency in Switzerland since 1850, Journal of Hydrology, 381, 1–8, https://doi.org/10.1016/j.jhydrol.2009.09.022, 2010b.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrology and Earth System Sciences, 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, 2012.

Sideris, I. V., Gabella, M., Erdin, R., and Germann, U.: Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland, Quarterly Journal of the Royal Meteorological Society, 140, 1097–1111, https://doi.org/10.1002/qj.2188, 2014.

https://doi.org/10.5194/egusphere-2025-1920
Preprint. Discussion started: 26 June 2025
© Author(s) 2025. CC BY 4.0 License.

Switanek, M., Maraun, D., and Bevacqua, E.: Stochastic downscaling of gridded precipitation to spatially coherent subgrid precipitation fields using a transformed Gaussian model, International Journal of Climatology, 42, 6126–6147, https://doi.org/10.1002/joc.7581, 2022.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E.: High risk of unprecedented UK rainfall in the current climate, Nature Communications, 8, 107, https://doi.org/10.1038/s41467-017-00275-3, 2017.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Hardiman, S. C., Ren, H.-L., Lu, B., and Belcher, S. E.: Risk and dynamics of unprecedented hot months in South East China, Climate Dynamics, 52, 2585–2596, https://doi.org/10.1007/s00382-018-4281-5, 2019.

Trenberth, K., Zhang, R., and National Center for Atmospheric Research Staff: The Climate Data Guide: Atlantic Multi-decadal Oscillation (AMO) and Atlantic Multidecadal Variability (AMV), https://climatedataguide.ucar.edu/climate-data/atlantic-multi-decadal-oscillation-amo, 2023.

van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., and Burgers, G.: Improving 10^4-year surge level estimates using data of the ECMWF seasonal prediction system, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL020610, 2004.

Viviroli, D., Sikorska-Senoner, A. E., Evin, G., Staudinger, M., Kauzlaric, M., Chardon, J., Favre, A.-C., Hingray, B., Nicolet, G., Raynaud, D., Seibert, J., Weingartner, R., and Whealton, C.: Comprehensive space-time hydrometeorological simulations for estimating very rare floods at multiple sites in a large river basin, Natural Hazards and Earth System Sciences, 22, 2891–2920, https://doi.org/10.5194/nhess-22-2891-2022, 2022.

Viviroli, D., Staudinger, M., and Kauzlaric, M.: Extreme Floods in Switzerland: Hydrological scenarios for large catchments: Project report for FOEN and SFOE, 2025.

Volosciuk, C., Maraun, D., Vrac, M., and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrology and Earth System Sciences, 21, 1693–1719, https://doi.org/10.5194/hess-21-1693-2017, 2017.

Weingartner, R. and Aschwanden, H.: Discharge Regime – the Basis for the Estimation of Average Flows, in: Hydrological Atlas of Switzerland, p. Plate 5.2, Federal Office for the Environment FOEN, Bern, 1992.

Westerberg, I. K., Sikorska-Senoner, A. E., Viviroli, D., Vis, M., and Seibert, J.: Hydrological model calibration with uncertain discharge data, Hydrological Sciences Journal, pp. 2441–2456, https://doi.org/10.1080/02626667.2020.1735638, 2020.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, Journal of Hydrology, 210, 178–191, https://doi.org/10.1016/S0022-1694(98)00186-3, 1998.

World Meteorological Organization: Manual on Estimation of Probable Maximum Precipitation (PMP), vol. 1045 of *WMO Publ.*, World Meteorological Organization, ISBN 978-92-63-11045-9, 2009.

Zeder, J. and Fischer, E. M.: Decadal to centennial extreme precipitation disaster gaps — Long-term variability and implications for extreme value modelling, Weather and Climate Extremes, 43, 100 636, https://doi.org/10.1016/j.wace.2023.100636, 2024.