



Statistical reconstruction of daily precipitation and temperature fields in Switzerland back to 1864

Lucas Pfister^{1,2}, Stefan Brönnimann^{1,2}, Mikhaël Schwander³, Francesco Alessandro Isotta³, Pascal Horton^{1,2}, and Christian Rohr^{1,4}

¹Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

²Institute of Geography, University of Bern, Bern, Switzerland

³Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

⁴Institute of History, University of Bern, Bern, Switzerland

Correspondence: Lucas Pfister (lucas.pfister@giub.unibe.ch)

Received: 6 October 2019 – Discussion started: 5 November 2019

Revised: 6 February 2020 – Accepted: 7 February 2020 – Published: 2 April 2020

Abstract. Spatial information on past weather contributes to better understanding the processes behind day-to-day weather variability and to assessing the risks arising from weather extremes. For Switzerland, daily resolved spatial information on meteorological parameters is restricted to the period starting from 1961, whereas prior to that local station observations are the only source of daily long-term weather data. While attempts have been made to reconstruct spatial weather patterns for certain extreme events, the task of creating a continuous spatial weather reconstruction dataset for Switzerland has so far not been addressed. Here, we aim to reconstruct daily high-resolution precipitation and temperature fields for Switzerland back to 1864 with an analogue resampling method (ARM) using station data and a weather type classification. Analogue reconstructions are post-processed with an ensemble Kalman fitting (EnKF) approach and quantile mapping. Results suggest that the presented methods are suitable for daily precipitation and temperature reconstruction. Evaluation experiments reveal excellent skill for temperature and good skill for precipitation. As illustrated with the example of the avalanche winter of 1887/88, these weather reconstructions have great potential for various analyses of past weather and for climate impact modelling.

1 Introduction

Historical meteorological measurements are invaluable not just for studying climate variability, but also for long-term variability in weather, its extremes, and its relation to the large-scale circulation. Day-to-day weather data allow for the calculation of targeted indices (e.g. consecutive dry days, growing degree days), which are more useful than monthly climate data for assessing climate impacts. Moreover, daily data feed into current impact models and allow for the study of crop growth, water availability, and the impacts of droughts, floods, or avalanches numerically. However, historical station observations only capture local weather conditions. Most of the applications mentioned above require spatial fields of meteorological parameters.

For Switzerland a long-term, high-resolution, and time-consistent spatial dataset of precipitation and temperature starting in 1864 is available only with monthly resolution, as introduced recently by Isotta et al. (2019). A comparable daily dataset, however, which is needed to analyse past weather, covers a relatively short period starting in 1961 (MeteoSwiss, 2016a, b). Prior to 1961, observations from weather stations are the only sources that provide continuous information on daily weather. Today's dynamical and stochastic models offer new possibilities to make use of this sparse information and enable us to create spatial reconstructions of past weather. In recent years, several efforts have been made to create high-resolution temperature and precipitation reconstructions for historical extreme events in Switzerland using dynamical (Brugnara et al., 2017; Stucki

et al., 2018) and statistical (Flückiger et al., 2017) downscaling methods. While for Switzerland, the task of creating long-term, high-resolution daily spatial weather reconstructions has not been addressed so far, Caillouet et al. (2016, 2019) have presented a continuous dataset of daily precipitation and temperature fields for France starting in 1871 by statistically downscaling data from the 20CR reanalysis. This study aims to create such a dataset of long-term, high-resolution daily spatial reconstructions of precipitation and temperature for Switzerland by extending the currently available datasets backwards in time until 1864.

We use a statistical approach that has been applied in various research areas related to climate sciences: the so-called analogue method (Lorenz, 1969; Zorita and von Storch, 1999; Ben Daoud et al., 2016; Barnett and Preisendorfer, 1978; Kruijzinga and Murphy, 1983; Horton et al., 2012). In recent years, this method has also been introduced to local-scale weather reconstruction using historical station data (Flückiger et al., 2017; Rössler and Brönnimann, 2018). It is based on the assumption that over time similar spatial patterns of atmospheric states occur that produce similar local effects (Lorenz, 1969; Horton et al., 2017). The analogue approach makes use of this statistical relationship between large-scale and local weather or meteorological patterns, while one can be used to predict the other. To predict a certain atmospheric feature, e.g. precipitation and temperature fields for a given day of interest (predictand), the analogue method looks for the day with the most similar predictor values (best analogue) and takes the atmospheric feature from this (or multiple) best analogue day(s) as a prediction (Zorita et al., 1995). As it is basically a resampling of observed states of the atmosphere (spatial weather data) along the time axis to optimally fit certain predictors (Graham et al., 2007; Franke et al., 2011), the term analogue resampling method (ARM) is used in this paper.

In our approach, analogue reconstructions are further improved. Using techniques borrowed from data assimilation, reconstructed temperature fields are adjusted towards station measurements with a so-called ensemble Kalman fitting approach (Whitaker and Hamill, 2002; Franke et al., 2017) that is adapted to analogue reconstructions. The combination of the analogue method with a Kalman filter was tested, e.g. in Tandeo et al. (2014) and Lguensat et al. (2017), for Lorenz models and has proven to provide good results. Reconstructed precipitation data are bias-corrected using a quantile mapping method (Gudmundsson et al., 2012) by fitting reconstructed to observed precipitation distributions.

The result is a long-term, daily resolved spatial dataset of precipitation and temperature with a 2.2×2.2 km spatial resolution for the period of 1 January 1864–31 December 2017. Reconstructions are evaluated against gridded data from MeteoSwiss and against station observations. To demonstrate the potential of the reconstructions, we analyse the avalanche winter in 1887/88 by comparing reconstructions to previous studies and documentary data

(Vieli, 2017; Coaz, 1889). This paper accompanies the on-line publication of the reconstructed precipitation and temperature datasets at the open-source repository PANGAEA (<https://doi.org/10.1594/PANGAEA.907579>).

The paper is organised as follows: Sect. 2 provides an overview of the data used. Section 3 describes the methods of weather reconstruction and post-processing and presents the validation strategy as well as the measures applied for assessing the reconstructions. In Sect. 4, results from the validation of reconstructed and post-processed temperature and precipitation fields are presented and discussed before analysing the avalanche winter of 1887/88 in Sect. 5. Conclusions are drawn in Sect. 6.

2 Data

Statistical weather reconstruction methods require long-term and if possible homogeneous series of station measurements. In Switzerland, we can benefit from the network of MeteoSwiss going back to the year 1864 (Fülleman et al., 2011; Begert et al., 2005). All 68 meteorological stations used for reconstruction are part of the Swiss National Basic Climatological Network (Swiss NBCN), a network of long-term, continuous, and high-quality measurements used for climate monitoring (Begert et al., 2007; Begert, 2008). To ensure the consistency of the reconstructions over time, the set of meteorological stations and parameters used is ideally not changed over time. Therefore, only measurement series starting prior to 1901 and continuing until today with interruptions of no longer than 5 years were selected. One exception is the MeteoSwiss station of Col du Grand St-Bernard (GSB), for which data show a gap from 30 July 1925 to 31 December 1933. This station was included, as it lies within a data-scarce region and represents higher altitudes. In all cases, homogenised daily mean temperature and precipitation sums were used, as were daily mean pressure values at station height (QFE). An overview of measurement locations and variables, as well as their vertical distribution is given in Fig. 1.

In total we used 10 pressure, 25 temperature, and 67 precipitation series. The large number of precipitation measurements was chosen to account for the high spatial variability of this variable, while 10 stations are enough to cover surface pressure. Most stations are located at lower elevations and in valleys, while higher altitudes (hillsides, mountains) are under-represented. Two independent measurement series (yellow) from the Swiss Plateau and the Alpine region were used for validation: Schaffhausen (SHA, 438 m a.s.l.) with measurements from 1 January 1864–31 December 2017 and Grimsel Hospiz (GRH, 1980 m a.s.l.), covering the period from 1 January 1932–31 December 2017. Note that data from Schaffhausen are not homogenised.

Furthermore, the ARM requires spatial data, from which the best analogue reconstructions are drawn. Here we used

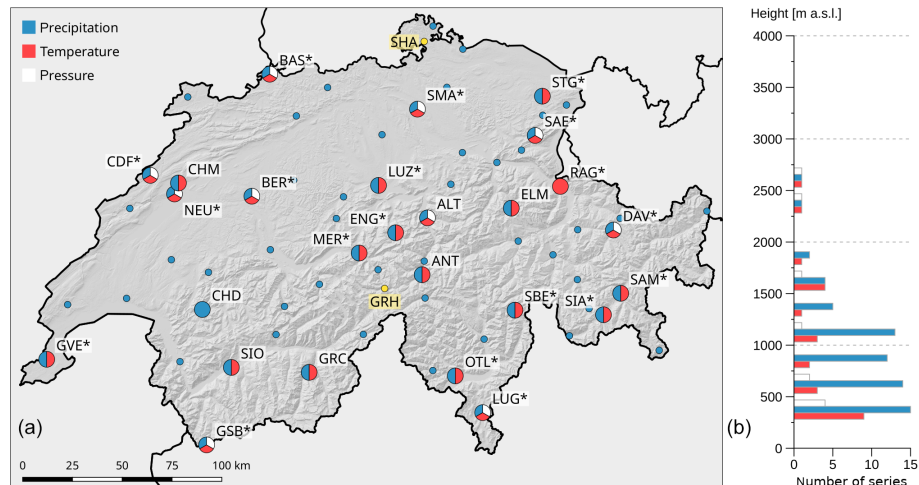


Figure 1. Station map. (a) Measured variables are indicated as colours. Labelled pie charts represent NBCN climate monitoring stations. Additional NBCN precipitation stations are indicated as small blue dots. Stations that were used for temperature assimilation are marked by an asterisk. Yellow dots represent series used for station-based validation. (b) The vertical distribution of measurement series is indicated by altitude class for each variable.

daily gridded precipitation and temperature data provided by MeteoSwiss (MeteoSwiss, 2016a, b) with a spatial resolution of 2.2 km, covering the period 1 January 1961–31 December 2017. Precipitation data (RhiresD) indicate accumulated precipitation (rainfall and snowfall water equivalent) from 06:00 to 06:00 UTC of the following day (MeteoSwiss, 2016a), spatially interpolated from daily precipitation sums measured at the MeteoSwiss high-resolution rain-gauge station network. Topographic effects and differences in station distribution are accounted for. Errors are estimated to be of the order of factor 1.7 for precipitation below the 20 % quantile (tendency towards overestimation) and 1.3 for precipitation above the 90 % quantile (tendency towards underestimation); they are higher in mountainous areas (MeteoSwiss, 2016a). A detailed description of this dataset and the methods to derive it can be found in MeteoSwiss (2016a) and Schwarb (2001).

Gridded temperature (TabsD) displays daily mean (00:00 to 00:00 UTC) air temperature measured in degrees Celsius ($^{\circ}\text{C}$) at 2 m above the ground (MeteoSwiss, 2016b). As homogenised station data were used for interpolation, errors resulting from changes in measurement location or instruments are corrected. Regional differences in vertical temperature gradients, as well as the effects of warm boundary layers and temperature inversions are taken into account. Standard errors in the TabsD dataset range from 0.6 to 1.1 $^{\circ}\text{C}$ in the Swiss Plateau (smaller in summer) and reach values of 4 $^{\circ}\text{C}$ in inner Alpine valleys in winter. For further information on the interpolation method and validation, the reader is referred to Frei (2014) and MeteoSwiss (2016b).

Furthermore, a daily weather type (WT) classification is used (Schwander et al., 2017), covering the period of interest. These WT reconstructions are based on the CAP9 clas-

sification used by MeteoSwiss that distinguishes nine different WTs for central Europe (Weusthoff, 2011), which show good skill at predicting daily weather, especially precipitation in the Alpine region (Schiemann and Frei, 2010). Merging two pairs of similar CAP9 WTs, Schwander et al. (2017) reconstructed WTs from 1763 to 2009 using homogenised instrumental measurement series from different locations in Europe. For each day, this dataset provides the probabilities of each CAP7 WT. WTs from 2010 onwards were calculated from the CAP9 data from MeteoSwiss (Weusthoff, 2011).

As argued by Schwander et al. (2017), reconstructed WTs are more reliable for winter than for summer months, as the underlying meteorological patterns are more pronounced during winter. For weather reconstruction, this property has to be taken into account.

3 Methods

3.1 The analogue resampling method

The application of the ARM in this paper is based on the work by Flückiger et al. (2017). The ARM requires two meteorological archives: data to predict the spatial fields and a record of the spatial data from which the reconstructions are drawn. To predict the spatial fields we used daily station observations, while the RhiresD and TabsD datasets for 1961–2017 from MeteoSwiss serve as records of spatial data (see Sect. 2). For a given day in the past, we screen the period for which we have spatial data (analogue pool) for the most similar day in terms of station data (best analogue). Precipitation and temperature fields from this day serve as an estimate for the day in the past. Predictands of the analogue method

are thus the unknown spatial fields, while station data and weather types serve as predictors.

Given a sufficiently large analogue pool, the ARM generally has the advantage of reproducing the mean and variance of the target, as well as naturally occurring spatial patterns (Zorita and von Storch, 1999). With input from both more coarsely resolved data, e.g. reanalyses or weather types, and local information (station data), it can make use of more data sources than the simple downscaling or interpolation of station observations alone.

In order to maintain the physical consistency of the reconstructions, further conditions are established.

1. The day of interest and possible analogue days are required to be of the same WT to ensure similar synoptic-scale atmospheric conditions, e.g. wind fields (Weusthoff, 2011). To account for the uncertainty in WT reconstructions, we did not restrict the analogue to the most probable WT but accepted additional WTs such that they cover the true WT with a combined probability of at least 95 % according to Schwander et al. (2017).
2. The day of interest and possible analogue days are required to be within the same season to account for seasonally different spatial patterns. The time window is set to ± 60 d centred at the target day.

Following these conditions, the best analogue is defined as a day within the analogue pool, with the same weather type, and within the same time window that shows the most similar values of certain meteorological variables from a defined set of stations to the day of interest.

Before the application of the ARM, station and gridded data are preprocessed. As observed variables have different scales, each measurement series is standardised. Furthermore, temperature data from weather stations and spatial fields are decomposed into a smoothed mean climatology and the respective anomalies. For each observation series and each cell of the gridded data, a smoothed mean seasonality curve is estimated by fitting the first two harmonics of temperature time series following Eq. (1) using linear regression.

$$S = c_0 + c_1 \sin\left(\frac{2\pi \text{doy}}{n_{\text{doy}}}\right) + c_2 \cos\left(\frac{2\pi \text{doy}}{n_{\text{doy}}}\right) + c_3 \sin\left(\frac{4\pi \text{doy}}{n_{\text{doy}}}\right) + c_4 \cos\left(\frac{4\pi \text{doy}}{n_{\text{doy}}}\right) \quad (1)$$

Here, *doy* denotes the day of year, n_{doy} is the number of days in a year, and c_0 , c_1 , c_2 , c_3 , and c_4 are parameters to estimate. After the calculation of the analogue reconstructions using temperature anomalies, mean climatology is then again added to the reconstructed temperature deviation fields to get absolute temperature data. This procedure slightly alters the characteristics of the ARM, as adding climatology and re-sampled temperature deviations not only resamples known temperature fields, but also creates new ones. An elimination

of the signal from climatic temperature changes over the last centuries did not improve the results and was not further pursued in this study.

With preprocessed data, the analogue method is applied. Following Horton et al. (2017), the root mean squared error (RMSE) is used as a measure of similarity (Eq. 2). This distance measure was chosen to reduce the levelling-out of station data resulting in decreased variance of reconstructions. As the RMSE is sensitive to large deviations, station data of the best analogues will closely follow observed ones, thus also reproducing local phenomena and extremes. To a certain extent, however, an underestimation of variance due to the levelling-out of station observations cannot be avoided, especially for days with less suitable analogues and for skewed distributions like precipitation.

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Here, \mathbf{x} and \mathbf{y} are vectors of observations from the day of interest and a day within the analogue pool, respectively; i denotes the different observations within this vector. Other measures of similarity like the Mahalanobis distance were not examined.

There are some limitations to the method, the most crucial one being the size of the analogue pool, which has to be large enough to provide reasonably matching analogues to a given atmospheric state (Zorita and von Storch, 1999). This is especially relevant for extreme events as they occur more rarely, and therefore less suitable analogues exist. Also, the coverage by observation series plays an important role, as stations might fail to capture local events like thunderstorms. Furthermore, the methodological setting described above, e.g. the coupling of precipitation and temperature reconstructions and the choice of the similarity measure, can have an impact on the capability of the ARM to correctly reproduce the variance and mean, as well as spatial distribution of the variables. As the coupling was established for reasons of physical consistency and the distance measure was chosen to maintain natural variability, these effects are not further assessed in Sect. 4. Finally, temporal consistency is not guaranteed.

3.2 Post-processing methods

The best analogue may not perfectly fit all observations. To further improve the temperature reconstructions, we borrow from data assimilation techniques (see, e.g. Daley, 1999; Kalnay, 2007). The method used here is based on the ensemble Kalman filter (Kalman, 1960; Evensen, 1994; Burgers et al., 1998), which is applied e.g. for data assimilation of ensemble forecasts from dynamical models. Here we use the best analogue in the same way as the forecast (termed background or first guess) and best n analogues as an ensemble. However, neither the analysis nor the covariance matrix

(see below) is passed on to the next time step. This simplification is called ensemble Kalman fitting (EnKF) (Bhend et al., 2012; Franke et al., 2017) or offline data assimilation (Matsikaris et al., 2015). The EnKF essentially minimises a least-squares problem (Franke et al., 2017). The state vector \mathbf{x} that minimises the following cost function J is optimal in the case of Gaussian errors:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T (\mathbf{P}^b)^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}]), \quad (3)$$

where \mathbf{x}^b is the first guess (background), in this case reconstructions from the best analogue. \mathbf{P}^b is the background error covariance matrix that in this particular case is estimated from an “ensemble” of the best n analogues, i.e. temperature fields from the n most similar days to the day of interest. Vector \mathbf{y} contains station observations, and operator H is used to extract the observations from the model space. \mathbf{R} is the error covariance matrix of $\mathbf{y} - H(\mathbf{x})$. With station observations and the ensemble of the best n analogues, a new estimation of temperature \mathbf{x}^b that is the best estimate for the true atmospheric state \mathbf{x} is calculated from Eq. (4):

$$\mathbf{x}^b = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b), \quad (4)$$

where \mathbf{x}^b denotes the updated state vector (analysis), \mathbf{x}^b and \mathbf{y} are as described above, and \mathbf{K} is the Kalman gain or innovation matrix calculated from the ensemble. In this and the following equations, \mathbf{H} describes the Jacobian matrix of $H(\mathbf{x})$ and extracts the values from the grid cell closest to the observation site of \mathbf{y} .

We use an implementation (Whitaker and Hamill, 2002; Bhend et al., 2012) in which each observation is assimilated sequentially (Eqs. 5a, 5b–7). The fitting procedure is into two steps: an update of the ensemble mean $\bar{\mathbf{x}}$ (Eq. 5a) and an update of the anomalies \mathbf{x}' with respect to the ensemble mean (Eq. 5b). Equations (6a) and (6b) depict the calculation of the Kalman gain \mathbf{K} for the ensemble mean and $\tilde{\mathbf{K}}$ for the anomalies.

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}(\bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}^b) \quad (5a)$$

$$\mathbf{x}'^a = \mathbf{x}'^b + \tilde{\mathbf{K}}(\mathbf{y}' - \mathbf{H}\mathbf{x}'^b) \quad \text{with: } \mathbf{y}' = 0 \quad (5b)$$

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (6a)$$

$$\tilde{\mathbf{K}} = \mathbf{P}^b \mathbf{H}^T \left[\left(\sqrt{\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}} \right)^{-1} \right]^T \times \left(\sqrt{\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} + \sqrt{\mathbf{R}}} \right)^{-1} \quad (6b)$$

Here, $\bar{\mathbf{x}}^a$ and $\bar{\mathbf{x}}^b$ denote the analysis and background of the ensemble mean and \mathbf{x}'^a and \mathbf{x}'^b the corresponding anomalies. \mathbf{P}^b and \mathbf{R} are the error covariance matrices as in Eq. (3). The observation error \mathbf{R} is roughly estimated to be 1 °C. The

background error covariance matrix \mathbf{P}^b is calculated from the best n analogues following Eq. (7), where i and j denote grid boxes and k the ensemble members.

$$\mathbf{P}_{i,j}^b = \frac{1}{n_{\text{ens}} - 1} \sum_{k=1}^{n_{\text{ens}}} (\mathbf{x}_{i,k}^b - \bar{\mathbf{x}}_k^b) (\mathbf{x}_{j,k}^b - \bar{\mathbf{x}}_k^b) \quad (7)$$

Following Whitaker and Hamill (2002), instead of the full error covariance matrix, the conversion $\mathbf{P}^b \mathbf{H}^T$ is calculated directly in order to save computational resources. Covariance matrices estimated from small samples may exhibit spurious covariances far away from the observation. Spatial localisation is often used to minimise these effects. In our case, the study areas is too small and the ensemble size sufficient such that localisation is not necessary (tests using a Gaussian weighting function did not show improvement).

For each day, the EnKF is applied to the analogue reconstructions using a selection of measurement series that exhibit an average monthly correlation with co-located data from TabsD above 0.975 (see Fig. 1). This is to avoid measurement series subject to local influences, which are not resolved by spatial data and would thus lead to erroneous assimilations. To account for a bias between local measured temperature at a weather station and spatially aggregated temperature values of the corresponding grid cell, station data are adjusted by subtracting the mean bias between the measurement and grid cell value from the TabsD dataset over the period 1961–2017 for each month. This procedure prevents systematic biases in fitted temperature fields. The ensemble size is set to the 50 best analogues.

Given the limitations of the ARM described in the previous section, precipitation reconstructions might also be affected by difficulties, e.g. in reproducing extreme events due to the limited availability of matching analogues, and as a consequence have biases in the mean and issues regarding areas that are sparsely covered by weather stations. Therefore, analogue reconstructions of precipitation are post-processed. Although attempts have been made to assimilate precipitation with the application of a Kalman filter (Lien et al., 2013, 2016), in this paper, a much simpler approach is used: quantile mapping (QM). This method of model output statistics aims to transform the cumulative distribution functions (CDFs) of modelled precipitation to match the CDFs of observed precipitation by finding a statistical transfer function h (Maraun et al., 2010; Maraun, 2013); i.e. it is mapping modelled to the observed distribution. As pointed out by Cannon (2018), this procedure is asynchronous; that is, it does not consider any chronological aspects of precipitation. In its simplest application, QM corrects the model bias according to observed precipitation values (Piani et al., 2010) and can be generally expressed by Eq. (8):

$$\mathbf{P}_o = h(\mathbf{P}_m), \quad (8)$$

where \mathbf{P}_o and \mathbf{P}_m are observed and modelled precipitation, respectively, and h the transfer function (Gudmundsson et al.,

2012). Based on the probability integral transform theorem (Angus, 1994), the transformation can be described as

$$\mathbf{P}_o = F_o^{-1}(F_m(\mathbf{P}_m)), \quad (9)$$

with F_m the CDF of modelled precipitation and F_o^{-1} the inverse CDF of the observed precipitation. To solve this equation, the distribution of the variable of interest has to be defined. In this paper, a parametric transformation using an exponential asymptotic function to estimate precipitation distribution was chosen following Gudmundsson et al. (2012). This parametric transformation is described by Eq. (10), where $\hat{\mathbf{P}}_o$ denotes the best estimate of \mathbf{P}_o and parameters a , b , x , and τ are to be determined.

$$\hat{\mathbf{P}}_o = (a + b\mathbf{P}_m) \left(1 - e^{-(\mathbf{P}_m - x)/\tau} \right) \quad (10)$$

The best prediction of parameters a , b , x , and τ is estimated by minimising the residual sum of squares for wet days (Gudmundsson et al., 2012). To define wet days, a threshold for $P > 0.1$ mm was set. Precipitation values beyond 0.1 mm were set to zero. Parametric transfer functions were calculated from all data within the calibration period 1 January 1961–31 December 2017 for each grid cell after Piani et al. (2010), with \mathbf{P}_m the values from the analogue reconstructions and \mathbf{P}_o the values from the RhiresD dataset. No discrimination between different seasons has been made. Based on the assumption that the transfer function derived from this period is robust, i.e. precipitation distribution is not subject to changes in time, these functions can then be extrapolated in time to transform the precipitation distributions of the reconstructed datasets back to 1864.

Note that the method as applied in this paper only corrects model bias. This simple application of QM was chosen to be able to extrapolate distribution correction in time, as more complex approaches would likely be less robust. To substantially improve e.g. dry–wet day discrimination or extreme values, other approaches have to be applied (Cannon et al., 2015).

3.3 Validation

The validation of precipitation and temperature reconstructions follows common measures and strategies used in the validation of field forecasts (Wilks, 2009; Jolliffe and Stephenson, 2012). If not indicated otherwise, validation measures and skill scores are computed on absolute values.

We use the Pearson correlation coefficient for temperature, while for non-Gaussian-distributed precipitation the Spearman correlation is calculated. Note that for temperature, correlation is computed on anomalies from mean seasonality (compare Sect. 3.1), so it reflects day-to-day variability rather than the seasonal cycle. Error magnitudes are indicated as root mean squared error (RMSE), as this measure is sensitive to larger errors. Furthermore, systematic biases between reconstructions and observations are evaluated.

Additionally, the mean squared error skill score (MSESS) or reduction of error statistic (RE value) is calculated for temperature reconstructions following Eq. (11), allowing us to analyse the skill of reconstructions compared to mean climatology in terms of the mean squared error:

$$\text{MSESS} = 1 - \frac{\sum_{i=1}^n (x_i^{\text{rec}} - x_i^{\text{ref}})^2}{\sum_{i=1}^n (x_i^0 - x_i^{\text{ref}})^2}, \quad (11)$$

with x^{rec} the reconstruction, x^0 a “no knowledge prediction” (in this case mean climatology), x^{ref} the reference data from TabsD, and i the time step (validation over time) or grid cell (validation over space). An MSESS value of 1 indicates a perfect reconstruction. With an MSESS of zero, the prediction skills of reconstruction and climatology are equal and values below zero denote a decline in skill compared to climatology (Jolliffe and Stephenson, 2012). Note that this measure punishes variance; i.e. a reconstruction with the correct variance but zero correlation will have an MSESS of -1 .

For precipitation reconstructions, another property of interest is the discrimination between wet and dry. For this purpose, the Brier score (BS) was calculated (Eq. 12) that compares the predicted probability of an event to observations (Wilks, 2009):

$$\text{BS} = \frac{1}{n} \sum_{i=1}^n (y_i - o_i)^2, \quad (12)$$

where y and o denote the probability of rain in reconstructions and observations, respectively, and i is as above. As reconstructions do not provide probabilities, y and o are binary with $1 = \text{rain}$ and $0 = \text{no rain}$ with a wet–dry threshold of 0.1 mm. The BS describes the percentage of time steps (or grid cells) wrongly assigned as wet or dry.

In a first part, a leave-one-out validation was performed on daily gridded data within the period 1 January 1961–31 December 2017. For each day, the best analogue day is calculated, excluding data from 5 d before and after the day of interest, as spatial patterns from neighbouring days can be similar. Precipitation and temperature reconstructions are then validated against the RhiresD and TabsD dataset, respectively. To analyse the full time span of the dataset, reconstructions are compared to station observations in a second part. For this purpose, two independent station series from Schaffhausen and Grimsel Hospiz (see Sect. 2) were used. Measurements were compared to reconstructions by extracting values from the corresponding grid cells without interpolation.

4 Results and discussion

As described in Sect. 3.3, a leave-one-out validation over the period 1961–2017 was performed, and reconstructions were

compared to the MeteoSwiss RhiresD and TabsD datasets, as well as station data. In this section, we will illustrate and discuss general results from the grid-based validation of precipitation and temperature reconstructions for 1961–2017. Of particular interest are seasonal differences and extreme events, for which we also evaluate the accuracy of reconstructions to reproduce spatial patterns. Furthermore, we compare reconstructed time series for Schaffhausen and Grimsel Hospiz to corresponding station observations.

4.1 Leave-one-out validation in time

Figure 2 shows results from the validation over time for analogue precipitation reconstructions (a–d) and quantile-mapped data (e–h) against RhiresD data. Depicted are the rank correlation (a, e), RMSE (b, f), mean bias (c, g), and Brier score (d, h). The Spearman correlation coefficient for analogue reconstructions is 0.79 on average and attains values from 0.62 to 0.86, with maximum values in central Switzerland (a). Quantile mapping does not change the ranks of precipitation distribution, therefore the two correlation maps are identical. Regarding the RMSE (e, f), an average error magnitude of less than 5 mm in the Swiss Plateau, as well as the inner-alpine valleys and large parts of the canton of Grisons can be observed. Errors are larger in mountainous areas and in Ticino, reaching values of 6–15 mm. Post-processed data (f) reveal a negligible increase in these errors in the range of 0.1–0.6 mm. Analogue reconstructions show a negative bias between 0.2 and 0.5 mm in the Swiss Plateau (c). The underestimation is more pronounced in mountainous regions and in Ticino, with values of 0.5–1.6 mm. Using the quantile mapping approach described in Sect. 3.2, this bias could be eliminated for the given time span (g). The Brier score indicates relatively high error rates in the discrimination between wet and dry days at individual locations, with values between 0.13 and 0.23 (d). Post-processed data reveal slightly negative changes in terms of Brier scores (h).

While rank correlation values show satisfying results, the bias and RMSE patterns of ARM reconstructions could possibly be explained by an underestimation of extreme and convective precipitation, which occur in the Alpine region and in southern Switzerland, especially in summer. While quantile-mapped data correct the bias, error values still remain large. We will elaborate on this issue below by looking at seasonal patterns and extremes. Another problem of the reconstructions is indicated by the Brier score: on average, 17% of days are wrongly assigned as wet or dry. This relatively high fraction is not improved with post-processing, as the quantile mapping approach used here is not designed to address this particular problem.

The validation of temperature reconstructions over time in Fig. 3 reveals a good correlation for unprocessed data, ranging between 0.76 and 0.95 with a mean of 0.91 (a). Correlation is slightly lower in Ticino and the southern valleys of Grisons. With ensemble Kalman fitting, the Pearson correla-

tion could be increased to values between 0.83 and 0.99 with a mean of 0.96, showing similar spatial patterns (e). The error (RMSE) could also be reduced with post-processing from 1.52 to 0.96 °C on average (b, f). In the Swiss Plateau, the error attains values below 1 °C, while in the Alpine region, in the Jura Mountains, and in southern Switzerland RMSE values up to 2.7 °C can be observed. Unprocessed reconstructions show a systematic overestimation of temperature in the Swiss Plateau, in the Rhone valley in Valais, and in the northern valleys of Ticino with values up to 0.06 °C (c). On the other hand, temperatures at higher altitudes and in southern Ticino are underestimated by 0.05 to 0.15 °C. Post-processed data (g) show less bias and a more balanced spatial pattern, with values ranging between –0.08 and 0.03 °C and a mean of –0.01 °C. The MESS compared to mean seasonality (d, h) is high all over Switzerland and could be increased from 0.83 to 0.93 on average using EnKF. The pattern follows correlation.

Overall, reconstructed temperature fields can be considered to very accurately reproduce the temporal evolution of the weather. Errors are relatively low, although in regions with sparse meteorological observations, larger errors are observed. Station coverage thus plays a crucial role for analogue reconstructions. The local field of larger errors in the western Jura near La Brévine might be explained by cold air pooling, which occurs frequently in this region during winter (Vitasse et al., 2017) and is not captured by any of the measurement series used for reconstruction. Bias patterns suggest that ARM reconstructions have problems correctly reproducing vertical temperature gradients. A major issue here could also be inversions. In the vertical distribution of the temperature stations used, higher altitudes are only sparsely covered (see Fig. 1), making the correct determination of vertical gradients and inversion heights difficult. Post-processed data seem to solve large parts of this problem, but this needs further investigation.

To find possible explanations behind the issues mentioned above and to gain more insight into the dataset, precipitation and temperature reconstructions are assessed in detail for differences between seasons and extremes.

Figure 4 depicts the rank correlation (a–d), RMSE (e–h), and averaged bias (i–l) over time of post-processed precipitation for each season. We see a relatively uniform correlation pattern over all seasons, with slightly higher values for summer (JJA) along the northern Prealps (c). Correlation values are lowest on the southern side of the Alps, especially in winter (a). Error values (e–h) show a similar spatial pattern throughout the year and are smallest from December to February (DJF) and slightly higher in spring (MAM) and autumn (SON); maximum values of the RMSE occur during the summer months and reach values of 8–15 mm in Ticino. The mean bias over time (i–l) still shows minor seasonal differences. In the Swiss Plateau and the Jura Mountains, the mean deviation is approximately zero. The largest differences occur in the Alpine region and in southern Switzerland, where

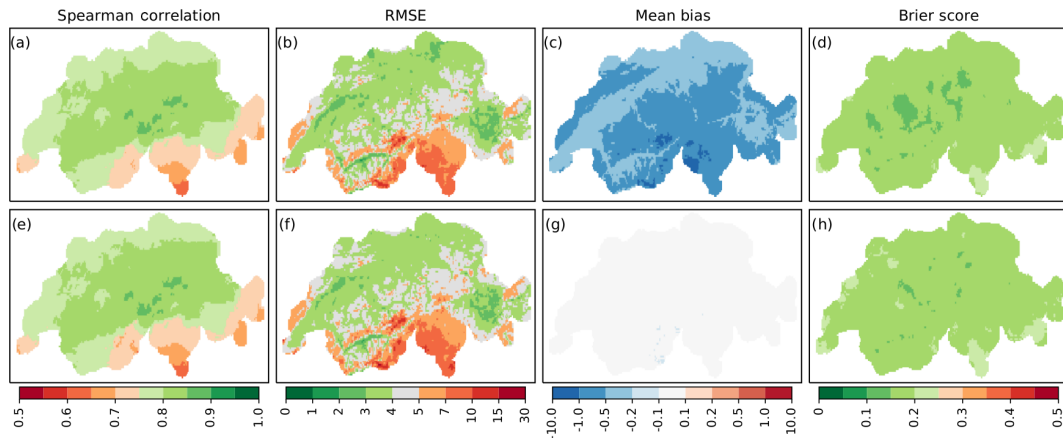


Figure 2. Validation over time of precipitation during 1961–2017 for analogue reconstructions (a–d) and quantile-mapped data (e–h). Shown are the Spearman correlation (a, e), RMSE (mm) (b, f), mean bias (mm) (c, g), and Brier score (d, h).

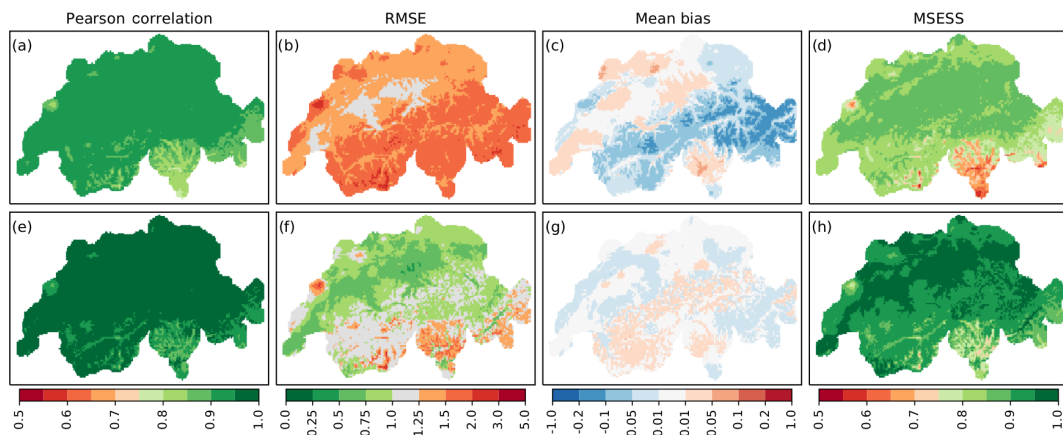


Figure 3. Validation of temperature over time during 1961–2017 for ARM (a–d) and EnKF (e–h) reconstructions. Shown are the Pearson correlation (a, e), RMSE (°C) (b, f), mean bias (°C) (c, g), and MSESS (d, h).

autumn and winter precipitation show a tendency towards overestimation, while in spring and summer post-processed precipitation fields exhibit a mostly negative bias in western Valais and the Gotthard region.

The pattern of RMSE, with higher values during the warmer periods of the year and maxima in summer, supports the previous assumption that reconstructions have problems reproducing intensive or convective precipitation. The latter, which are local-scale phenomena, may not be detected by measurement stations, making station coverage an important issue in terms of obtaining reliable reconstructions.

Analysing the same for temperature (Fig. 5), we see that Pearson correlation values (a–d) exhibit maximum correlation values in spring and summer, while in autumn and winter these values are slightly lower. The RMSE (e–h) is higher in winter than during the other seasons and reaches minima in the summer months. Maximum errors of up to 3 °C occur during winter in the Alpine region and the Jura Mountains (e). Overall, average bias (i–l) is only marginal for all sea-

sons, with values between -0.2 and 0.1 °C. Generally, vertical temperature gradients seem to be corrected by Kalman fitting. However, we can see higher values and a distinct spatial pattern related to topography in winter, as can also be seen in the RMSE (e). This indicates that inversions, which occur more frequently during this season, also remain a problem in post-processed reconstructions.

The above-mentioned issue with cold air pools in the western Jura seems to be confirmed by the seasonal error patterns, although larger errors for this region persist throughout the year. Interestingly, the mean bias shows an underestimation of winter temperature for this region.

4.2 Leave-one-out validation in space

Post-processed temperature and precipitation data were further assessed by quantiles of Swiss mean temperatures and precipitation during 1961–2017 calculated from the MeteoSwiss TabsD and RhiresD datasets to analyse the accuracy

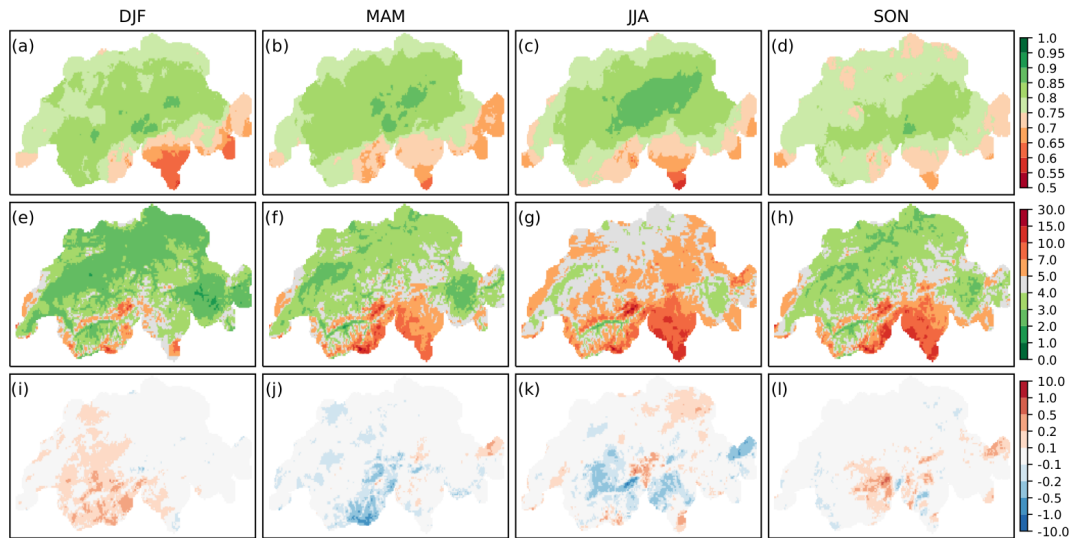


Figure 4. Validation of quantile-mapped precipitation over time by season. Shown are the Spearman correlation (a–d), RMSE (mm) (e–h), and bias (mm) (i–l).

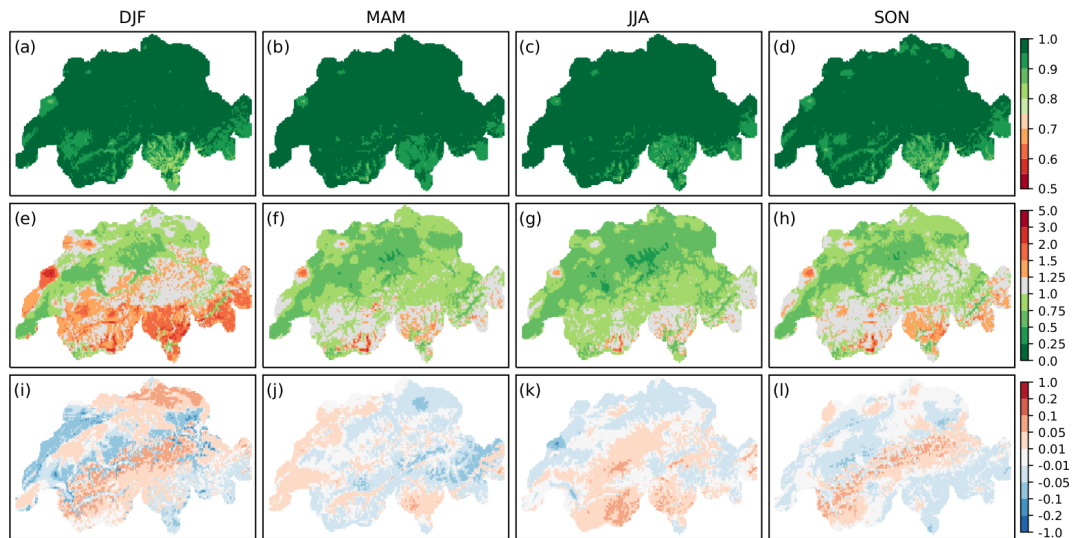


Figure 5. Results of validation over time of EnKF temperature reconstructions for each season. Shown are the Pearson correlation (a–d), RMSE (°C) (e–h), and bias (°C) (i–l).

of reconstructions in reproducing extremes (Fig. 6). Note that, as quantiles were calculated for average values over Switzerland, local extremes do not necessarily correspond to the highest or lowest quantiles for the whole area. In the following, results from validation over space are shown to analyse the capability of reconstruction methods to reproduce spatial patterns. For comparison, validation results from the analogue method are indicated in grey.

The spatial correlation of precipitation (Fig. 6, top) is relatively low for low to moderate precipitation events and increases with precipitation quantiles. Looking at the RMSE, the mean and the spread of errors also increase with increasing precipitation. While the RMSE shows a median of less

than 5 mm up to the 70 % quantile, for extreme precipitation events above the 95 % quantile errors attain values of 10–15 mm in the interquartile range. Compared to unprocessed data, a slight decrease in correlation and increase in the RMSE are visible. However, the bias is considerably improved. While, as argued before, analogue reconstructions indeed reveal a strong underestimation of extreme events, the median bias becomes approximately zero for all quantiles. Uncertainties, however, remain large. The Brier score reveals that for days with zero to low precipitation as well as for extreme events, the precipitation area is well represented in the reconstructions. The problems here lie in the correct reconstruction of precipitation areas for moderate events.

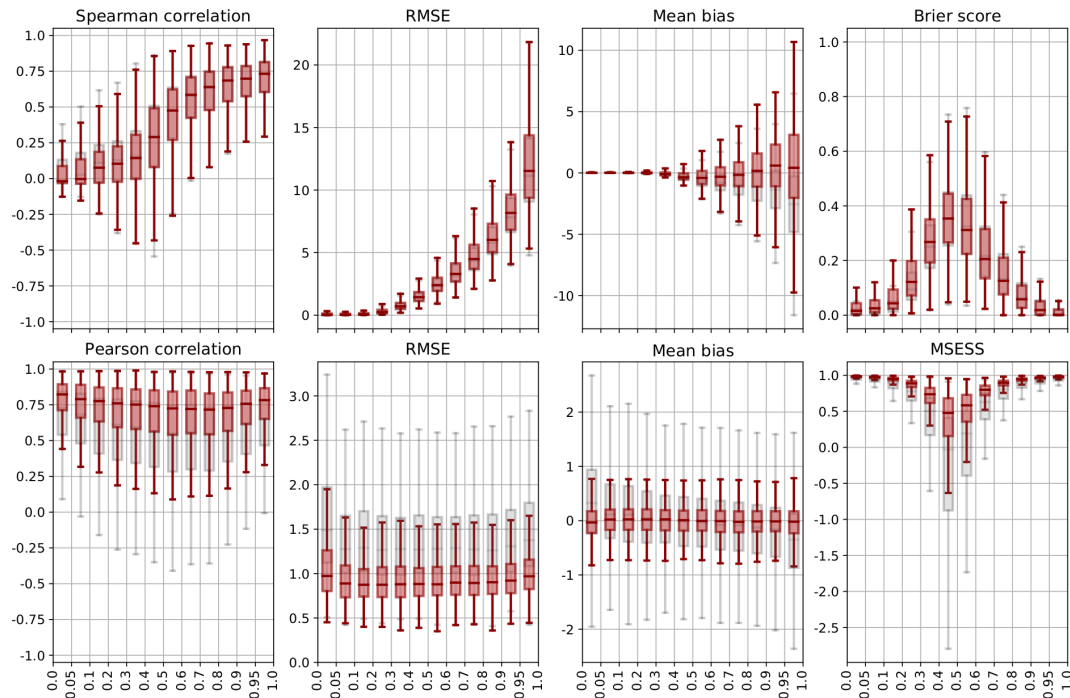


Figure 6. Validation over space of precipitation (top) and temperature (bottom) for ARM reconstructions (grey) and post-processed data (red), separated by quantile groups of spatial average precipitation and temperature, respectively. Shown are the Spearman (precipitation) and Pearson (temperature) correlation, RMSE, bias, Brier score (precipitation), and MESS (temperature). Boxes range from the first to the third quartile, and whiskers extend to 1.5 times the interquartile range outside the box.

Compared to unprocessed data, quantile mapping leads to a slightly better discrimination between wet and dry grid boxes for upper quantiles, while the BS becomes larger for lower quantiles.

From this, we can conclude that reconstructions provide accurate precipitation fields for low to moderate precipitation events. For the benefit of unbiased reconstructions, a slight decrease in correlation and an increase in the RMSE and BS have to be accepted. Extreme events, however, are underestimated by ARM reconstructions and also show large errors for post-processed data. As extreme events by definition occur more rarely, the number of suitable analogues is limited. As argued in Sect. 3, more station data and a bigger analogue pool would also lead to more accurate results for extremes. Different post-processing methods might help to improve reconstructions, especially regarding wet–dry discrimination and extremes.

The validation of temperature by spatial mean temperature quantiles (Fig. 6, bottom) shows a considerable improvement for post-processed data compared to analogue reconstructions. Correlation values exhibit slightly better correlations for extreme temperatures, while reconstructed fields for medium temperatures are less correlated with TabsD data. RMSE values are higher for upper and lower extreme values. In general, errors could be significantly reduced with Kalman fitting. The average bias reveals that while analogue recon-

structions tend to overestimate negative extreme values and underestimate extremely high temperatures, post-processed temperature data show a median of approximately zero for all quantiles. The bias pattern of the ARM can be explained as for precipitation by a limited number of suitable analogues for extreme events. Kalman fitting solves this problem. Furthermore, the spread of bias values is within ± 1 °C for 4 times the interquartile range. Post-processed temperature reconstructions are thus also accurate and precise for extreme temperatures. MESS values are better for upper and lower quantiles and show worse results around the median. As days around the median temperature are closer to average climatology, this pattern has to be expected. Nonetheless, the MESS of post-processed data is still within the area of natural variability (see Sect. 3.3).

4.3 Validation against independent observations

In Fig. 7, reconstructed precipitation and temperature are compared to station observations from Schaffhausen (left) and Grimsel Hospiz (right) over the full length of the respective series. Plotting reconstructed values against observations, we see a large spread of values for analogue reconstructions (grey) and for post-processed data (red). While for Schaffhausen, quantile-mapped precipitation exhibits a distribution closer to station observations in the Q–Q plot (centre) compared to ARM reconstructions, we can see a ten-

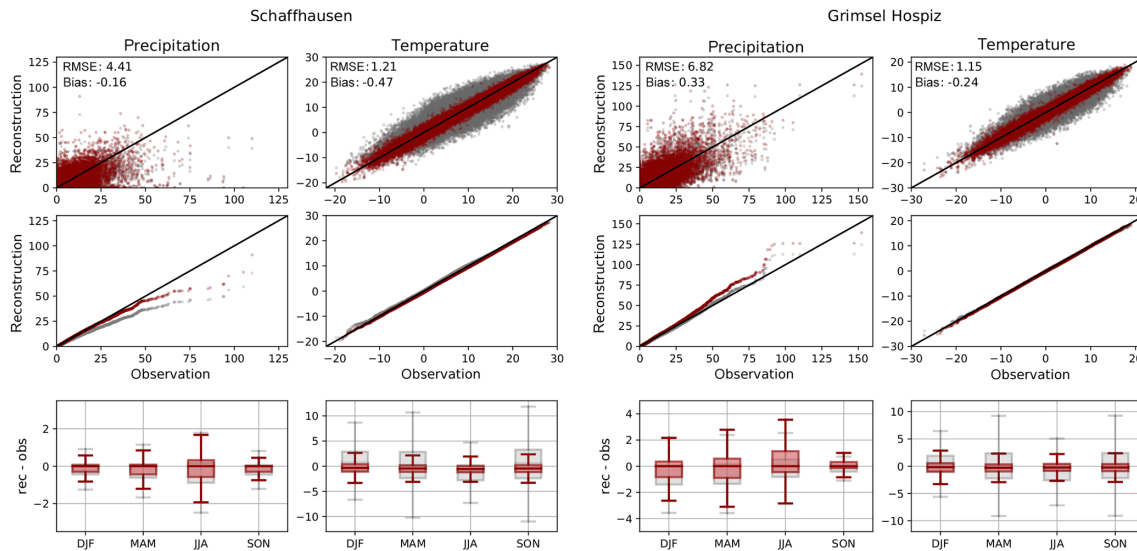


Figure 7. Comparison between reconstructions and station observations of precipitation (mm) and temperature ($^{\circ}\text{C}$) for Schaffhausen (left) and Grimsel Hospiz (right) with ARM reconstructions (grey) and post-processed data (red). Shown are observed vs. reconstructed values (top), quantile–quantile plots (centre), and box plots of the deviation between the reconstruction and observation by season (bottom).

endency towards overestimation in post-processed precipitation reconstructions for Grimsel Hospiz. The seasonal pattern (bottom) reveals the lowest differences in autumn, whereas the uncertainty is highest during summer, which is again in line with more frequent convective activity during the latter season as discussed before. The systematic underestimation of precipitation by the ARM is adjusted by post-processing for all seasons. The larger uncertainty in the Alpine region discussed before is also visible in station data for which the spread of bias values over all seasons (bottom) is larger for Grimsel Hospiz. This is also the case for the RhiresD gridded dataset used for reconstruction (see Sect. 2). The differences between reconstructions and station observations might at least partly be explained by the high spatial variability of precipitation; thus, spatially coarser gridded data can differ considerably from local measurements. For reasons of higher spatial variability, a less perfect fit has to be expected compared to temperature reconstructions. These are closer to observed values compared to precipitation and show a smaller spread of deviations. Kalman-fitted reconstructions are even more precise; not only is the spread of values reduced, but the tilt in distribution could also largely be corrected. Seasonal patterns of ARM reconstructions show larger deviations during spring and autumn. These seasonal differences are eliminated by EnKF. Reconstructed temperature fields thus accurately reproduce local temperature measurements, even for remote locations. With precipitation data, however, one has to be more careful when generating station series at individual locations.

4.4 The avalanche winter of 1887/88

The winter of 1887/88 was one of the most severe avalanche winters during the last 150 years, boosting efforts in avalanche prevention in Switzerland (SLF, 2000; Margreth, 2019). Intensive snowfall in February and March 1888 brought large snow masses to Switzerland, leading to 1094 disastrous avalanches, damaging 850 buildings, destroying over 1300 ha of forest, and burying 49 people under the snow (Coaz, 1889; Laternser and Pfister, 1997). Documentary data from Coaz (1889) provide a detailed description of this winter and a comprehensive survey of avalanche activity gathered by cantonal forestry offices. From a historical perspective, it has recently been assessed by Vieli (2017). However, quantitative data on the weather of this avalanche winter are restricted to sparse station observations so far. The gridded weather reconstructions presented in this paper can help to analyse the 1887/88 winter weather quantitatively, thus helping to better understand the weather patterns leading to such an event. For demonstration purposes, we performed some simple calculations of monthly averages, mean snow precipitation, and the zero-degree level that are summarised in Fig. 8.

Shown are post-processed precipitation reconstructions aggregated over 1 month (a–d) for winter 1887/88 and the recently published monthly precipitation reconstructions by MeteoSwiss (Isotta et al., 2019) (e–h). Both datasets reveal similar patterns of monthly precipitation, although regional differences occur. More deviations between the datasets can be observed in the amount of reconstructed precipitation. From both datasets, large precipitation sums can be determined in December 1887 on the northern flank of the Alps

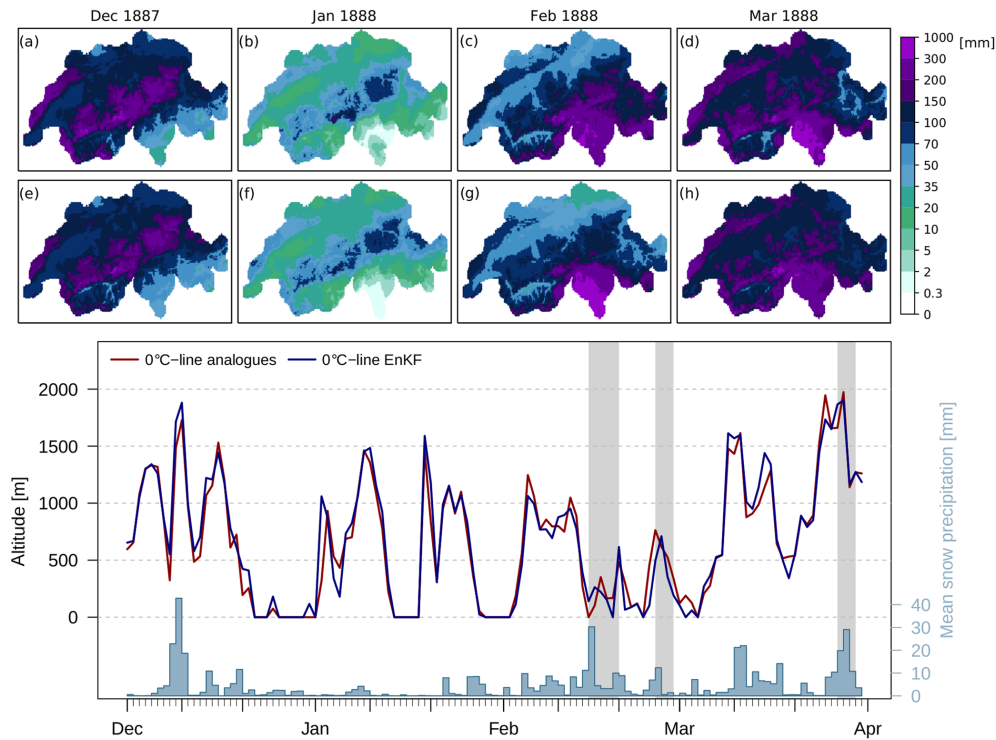


Figure 8. Avalanche winter 1887/88: monthly mean precipitation from December 1887 to March 1888 calculated from post-processed daily reconstructions (top, **a–d**) compared to monthly reconstructions from Isotta et al. (2019) (top, **e–h**). At the bottom, estimations of the zero-degree level are indicated for ARM (red lines) and EnKF (blue lines) reconstructions, as is average snow precipitation (blue bars) calculated from post-processed data. Grey shaded areas depict periods of increased avalanche activity as determined by Coaz (1889).

and in the Jura Mountains. January shows little precipitation, with the highest values in the north-eastern Alps. February and March show extreme precipitation values in Ticino and the Gotthard region, as well as in March over the remaining part of Switzerland.

Daily reconstructions allow for going more into detail. For example, we can calculate the development of the zero-degree level from gridded temperatures by taking the intercept of a linear regression between temperature and altitude (bottom of Fig. 8). Altitude data used here were aggregated from the Shuttle Radar Topography Mission (SRTM) 90 digital elevation dataset (Jarvis et al., 2008) to fit the resolution of reconstructions. Another value of interest is the intensity of snowfall precipitation. In Fig. 8 (bottom), the average snow precipitation per snowfall area is shown, calculated from post-processed reconstructions and assuming an estimated 1°C threshold (Jennings et al., 2018), below which precipitation falls as snow. Grey shaded areas depict periods of high avalanche activity as reported by Coaz (1889).

Analogue reconstructions (red) and assimilated temperature data (blue) show similar values. The extreme precipitation event in December 1887 coincides with a high altitude of the zero-degree level, thus leaving a snow-covered area restricted to higher altitudes. This event was followed by several cold episodes and low precipitation in January. The first

avalanche period was dominated by low temperatures and intensive precipitation. During the second avalanche period, the reconstructed zero-degree level rises to approximately 600 m a.s.l. with almost no snow precipitation. March shows two periods of high temperatures and intensive precipitation in the middle of the month and during the third avalanche period.

Reconstructed precipitation patterns and the development of temperature are in line with the findings from documentary data (Coaz, 1889) that report strong snowfall during December, a dry January, and intensive precipitation again during February and March, especially in the Southern Alps. While the first two avalanche periods are determined by low temperatures, Coaz describes a sharp rise in the zero-degree level to about 2000 m a.s.l. preceding the third period that led to a high number of wet avalanches, which can clearly be seen in the reconstructions. However, during the second period, reconstructions show relatively low snow precipitation values, contesting the high avalanche activity. Nonetheless, avalanches are not only triggered by intensive precipitation. For example, the intensive snowfall period in December 1887 and the first two in March 1888 are not accompanied by more frequent avalanches. To analyse this, other factors like temperature and wind as well as the composition of different snow layers also play an important role and have to

be assessed. A closer look at these periods would probably reveal more about the processes that triggered or prevented avalanches.

From precipitation and temperature reconstructions, new insight on the avalanche winter of 1887/88 can already be gained with simple methods. Using more sophisticated snow models that also take into account evaporation and snowmelt, high-resolution daily spatial data of the snow cover could be established that may be able to further explain avalanche activity. This is but one example of what the new daily reconstructions of temperature and precipitation could be used for. Analogue reconstructions have already been applied as input to numerical models, such as crop modelling (e.g. Flückiger et al., 2017) and hydrological modelling (e.g. Brönnimann et al., 2018), but the list could be extended. Many other phenomena, e.g. heatwaves and droughts, can be analysed spatially, and making use of the long time span changes in climate and extreme events over time could be investigated.

5 Conclusions

As shown in this paper, the analogue resampling method is a suitable and efficient approach for reconstructing daily precipitation and temperature fields from station observations. Using CAP7 weather types as a criterion for physical consistency and a set of observations from 68 weather stations, we present a long-term, physically consistent, high-resolution spatial dataset of these meteorological parameters for Switzerland since 1864. The datasets are published at the open-access repository PANGAEA (<https://doi.org/10.1594/PANGAEA.907579>; Pfister, 2019). Analogue reconstructions for temperature and precipitation show good results but experience difficulties in reproducing vertical temperature gradients and show a general negative bias for precipitation arising mainly from the underestimation of extreme events. Furthermore, analogue reconstructions reveal difficulties in correctly distinguishing between wet and dry days. On average, 17% of days were wrongly assigned. Temperature reconstructions could be considerably improved by assimilating station data using an ensemble Kalman fitting approach. Assimilated temperature fields show average error magnitudes of less than 1 °C and are nearly unbiased for the mean. The issue with vertical temperature gradients could be largely eliminated, although in winter some problems remain that could probably be attributed to difficulties of reconstructions in determining inversion heights. Precipitation data were post-processed with quantile mapping, adjusting the distributions of daily precipitation for each grid cell to obtain more accurate values. The mean bias could be successfully reduced, while a larger uncertainty for extreme events persists. However, error values show a slight increase in post-processed data. With the simple approach of quantile mapping presented in this paper, the

problem of wet and dry day discrimination could not be addressed.

There are some limitations to the analogue method, as the availability and coverage of station observations affect the accuracy of the results, especially for precipitation reconstruction. In regions with sparse information from weather stations, the uncertainty of reconstructions is larger. In Switzerland, this mostly regards mountainous areas. A second constraint is the comparatively small size of the analogue pool that is available for this application, which is especially relevant for extreme events as for such events less suitable analogues exist. To reconstruct extremes more accurately, notably for precipitation, a longer series of spatial data and a denser station network would be needed. An option to address the problem of small analogue pools as proposed by Van Den Dool (1994) is to construct more similar analogues through the linear combination of several possible analogue dates. With more sophisticated post-processing methods for precipitation, errors in wet and dry day discrimination could also be reduced. As mentioned, the analogue approach does not guarantee temporal consistency and therefore is not completely suitable to analyse trends. However, the dataset presented in this study complements the monthly reconstructions by Isotta et al. (2019), which were specifically designed for this purpose, very well.

The assessment of the avalanche winter of 1887/88 in Switzerland shows that the reconstructed development of temperature and precipitation corresponds well to documentary sources and to monthly reconstructions by Isotta et al. (2019). Possible applications of our daily high-resolution precipitation and temperature reconstructions range from crop modelling to the reconstruction of river runoff and the study of weather phenomena in the context of climate change.

Could daily reconstructions be extended even further back in time? For Switzerland, a recent survey brought to light a large amount of early instrumental data (Pfister et al., 2019). An extension of the dataset to the pre-industrial period is therefore envisaged, although larger measurement errors and less consistent measurement series make this endeavour rather challenging. The method should also be suitable to reconstruct daily meteorological fields for other regions of central and western Europe.

Data availability. Reconstructed daily precipitation and temperature datasets over the period 1 January 1864–31 December 2017 are published at the open-access repository PANGAEA (<https://doi.org/10.1594/PANGAEA.907579>; Pfister, 2019).

Author contributions. LP designed the study and performed the analysis with support from SB, FAI, and PH. MS provided weather type data and FAI the monthly reconstructions used in Sect. 4.4.

CH contributed to the analysis of the historical avalanche winter of 1887/88.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work has been supported by Swiss National Science Foundation project CHIMES (169676). Station and gridded data on meteorological parameters were obtained courtesy of MeteoSwiss. All methods were computed in R (R Core Team, 2017) using the quantile mapping package from Gudmundsson et al. (2012) and in Python. We would like to thank all researchers that assisted with advice and input on methodological matters.

Financial support. This research has been supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant no. 169676).

Review statement. This paper was edited by Hugues Goosse and reviewed by two anonymous referees.

References

- Angus, J. E.: The Probability Integral Transform and Related Results, *SIAM Rev.*, 36, 652–654, <https://doi.org/10.1137/1036146.1994>.
- Barnett, T. P. and Preisendorfer, R. W.: Multifield Analog Prediction of Short-Term Climate Fluctuations Using a Climate State vector, *J. Atmos. Sci.*, 35, 1771–1787, [https://doi.org/10.1175/1520-0469\(1978\)035<1771:MAPOST>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<1771:MAPOST>2.0.CO;2), 1978.
- Begert, M.: Die Repräsentativität der Stationen im Swiss National Basic Climatological Network (Swiss NBCN), *Arbeitsberichte der MeteoSchweiz*, 217, 2008.
- Begert, M., Schlegel, T., and Kirchhofer, W.: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000, *Int. J. Climatol.*, 25, 65–80, <https://doi.org/10.1002/joc.1118>, 2005.
- Begert, M., Seiz, G., Foppa, N., Schlegel, T., Appenzeller, C., and Müller, G.: Die Überführung der klimatologischen Referenzstationen der Schweiz in das Swiss National Basic Climatological Network (Swiss NBCN), *Arbeitsberichte der MeteoSchweiz*, 215, 2007.
- Ben Daoud, A., Sauquet, E., Bontron, G., Obled, C., and Lang, M.: Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin, *Atmos. Res.*, 169, 147–159, <https://doi.org/10.1016/j.atmosres.2015.09.015>, 2016.
- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Clim. Past*, 8, 963–976, <https://doi.org/10.5194/cp-8-963-2012>, 2012.
- Brugnara, Y., Brönnimann, S., Zamuriano, M., Schild, J., Rohr, C., and Segesser, D. M.: Reanalysis sheds light on 1916 avalanche disaster, *ECMWF Newsletter*, <https://doi.org/10.21957/h9b197.2017>.
- Brönnimann, S., Rohr, C., Stucki, P., Summermatter, S., Bandhauer, M., Barton, Y., Fischer, A., Froidevaux, P. A., Germann, U., Grosjean, M., Hupfer, F., Ingold, K., Isotta, F., Keiler, M., Martius, O., Messmer, M. B., Mülchi, R. I., Panziera, L., Pfister, L. M., Raible, C., Reist, T., Rössler, O. K., Röthlisberger, V. E., Scherrer, S. C., Weingartner, R., Zappa, M., Zimmermann, M., and Zischg, A. P.: 1868 – das Hochwasser, das die Schweiz veränderte. Ursachen, Folgen und Lehren für die Zukunft, [info:eu-repo/semantics/report](https://doi.org/10.4480/GB2018.G94.01), *Geographica Bernensia*, Bern, <https://doi.org/10.4480/GB2018.G94.01>, 2018.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, *Mon. Weather Rev.*, 126, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2), 1998.
- Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, *Clim. Past*, 12, 635–662, <https://doi.org/10.5194/cp-12-635-2016>, 2016.
- Caillouet, L., Vidal, J.-P., Sauquet, E., Graff, B., and Soubeyroux, J.-M.: SCOPE Climate: a 142-year daily high-resolution ensemble meteorological reconstruction dataset over France, *Earth Syst. Sci. Data*, 11, 241–260, <https://doi.org/10.5194/essd-11-241-2019>, 2019.
- Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, *Clim. Dynam.*, 50, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>, 2018.
- Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, *J. Climate*, 28, 6938–6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>, 2015.
- Coaz, J.: Der Lauinenschaden im schweizerischen Hochgebirge im Winter und Frühjahr 1887–88, *Stämpfli'sche Buchdruckerei*, Bern, 1889.
- Daley, R.: Atmospheric data analysis, vol. 2 of Cambridge atmospheric and space science series, Cambridge Univ. Press, Cambridge, 1. paperback ed., reprinted, transferred to digital print edn., 1999.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143, <https://doi.org/10.1029/94JC00572>, 1994.
- Flückiger, S., Brönnimann, S., Holzkämper, A., Fuhrer, J., Krämer, D., Pfister, C., and Rohr, C.: Simulating crop yield losses in Switzerland for historical and present Tambora climate scenarios, *Environ. Res. Lett.*, 12, 074026, <https://doi.org/10.1088/1748-9326/aa7246>, 2017.
- Franke, J., González-Rouco, J. F., Frank, D., and Graham, N. E.: 200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction analog method, *Clim. Dynam.*, 37, 133–150, <https://doi.org/10.1007/s00382-010-0802-6>, 2011.
- Franke, J., Brönnimann, S., Bhend, J., and Brugnara, Y.: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations, *Scientific data*, 4, 170076, <https://doi.org/10.1038/sdata.2017.76>, 2017.
- Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, *Int. J. Climatol.*, 34, 1585–1605, <https://doi.org/10.1002/joc.3786>, 2014.

- Fülleemann, C., Begert, M., and Croci-Maspoli, M., Brönnimann, S.: Digitalisieren und Homogenisieren von historischen Klimadaten des Swiss NBCN: Resultate aus DigiHom, Arbeitsberichte der MeteoSchweiz, 236, 2011.
- Graham, N. E., Hughes, M. K., Ammann, C. M., Cobb, K. M., Hoerling, M. P., Kennett, D. J., Kennett, J. P., Rein, B., Stott, L., Wigand, P. E., and Xu, T.: Tropical Pacific – mid-latitude teleconnections in medieval times, *Climatic Change*, 83, 241–285, <https://doi.org/10.1007/s10584-007-9239-2>, 2007.
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, *Hydrol. Earth Syst. Sci.*, 16, 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>, 2012.
- Horton, P., Jaboyedoff, M., Metzger, R., Obled, C., and Marty, R.: Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method, *Nat. Hazards Earth Syst. Sci.*, 12, 777–784, <https://doi.org/10.5194/nhess-12-777-2012>, 2012.
- Horton, P., Obled, C., and Jaboyedoff, M.: The analogue method for precipitation prediction: finding better analogue situations at a sub-daily time step, *Hydrol. Earth Syst. Sci.*, 21, 3307–3323, <https://doi.org/10.5194/hess-21-3307-2017>, 2017.
- Isotta, F. A., Begert, M., and Frei, C.: Long-Term Consistent Monthly Temperature and Precipitation Grid Data Sets for Switzerland Over the Past 150 Years, *J. Geophys. Res.-Atmos.*, 124, 3783–3799, <https://doi.org/10.1029/2018JD029910>, 2019.
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available at: <http://srtm.csi.cgiar.org/> (last access: 4 February 2020), 2008.
- Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere, *Nat. Commun.*, 9, 1148, <https://doi.org/10.1038/s41467-018-03629-7>, 2018.
- Jolliffe, I. T. and Stephenson, D. B. (Eds.): *Forecast verification: A practitioner's guide in atmospheric science*, Wiley-Blackwell, Chichester, UK, 2. ed. edn., <https://doi.org/10.1002/9781119960003>, 2012.
- Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *J. Basic Eng.-T. ASME*, 82, 35–45, <https://doi.org/10.1115/1.3662552>, 1960.
- Kalnay, E.: *Atmospheric modeling, data assimilation and predictability*, Cambridge Univ. Press, Cambridge, 4. print edn., 2007.
- Kruizinga, S. and Murphy, A. H.: Use of an Analogue Procedure to Formulate Objective Probabilistic Temperature Forecasts in The Netherlands, *Mon. Weather Rev.*, 111, 2244–2254, [https://doi.org/10.1175/1520-0493\(1983\)111<2244:UOAAPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<2244:UOAAPT>2.0.CO;2), 1983.
- Latenser, M. and Pfister, C.: Avalanches in Switzerland 1500–1990, in: *Rapid mass movement as a source of climatic evidence for the Holocene*, edited by: Matthews, J. A., 241–264, G. Fischer, Stuttgart and New York, 1997.
- Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R.: The Analog Data Assimilation, *Mon. Weather Rev.*, 145, 4093–4107, <https://doi.org/10.1175/MWR-D-16-0441.1>, 2017.
- Lien, G.-Y., Kalnay, E., and Miyoshi, T.: Effective assimilation of global precipitation: simulation experiments, *Tellus A*, 65, 19915, <https://doi.org/10.3402/tellusa.v65i0.19915>, 2013.
- Lien, G.-Y., Miyoshi, T., and Kalnay, E.: Assimilation of TRMM Multisatellite Precipitation Analysis with a Low-Resolution NCEP Global Forecast System, *Mon. Weather Rev.*, 144, 643–661, <https://doi.org/10.1175/MWR-D-15-0149.1>, 2016.
- Lorenz, E. N.: Atmospheric Predictability as Revealed by Naturally Occurring Analogues, *J. Atmos. Sci.*, 26, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2), 1969.
- Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, *J. Climate*, 26, 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>, 2013.
- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, 48, RG3003, <https://doi.org/10.1029/2009RG000314>, 2010.
- Margreth, S.: Lawinenwinter der letzten 150 Jahre: ihre Bedeutung für die Entwicklung des Lawinenschutzes, *WSL Berichte*, 78, 21–30, 2019.
- Matsikaris, A., Widmann, M., and Jungclaus, J.: On-line and off-line data assimilation in palaeoclimatology: a case study, *Clim. Past*, 11, 81–93, <https://doi.org/10.5194/cp-11-81-2015>, 2015.
- MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Precipitation (final analysis): RhiresD, available at: https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-niederschlag/doc/ProdDoc_RhiresD.pdf (last access: 4 February 2020), 2016a.
- MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD, available at: https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-temperatur/doc/ProdDoc_TabsD.pdf (last access: 4 February 2020), 2016b.
- Pfister, L.: Statistical Reconstruction of Daily Precipitation and Temperature Fields in Switzerland back to 1864, PANGAEA, <https://doi.org/10.1594/PANGAEA.907579>, dataset, 2019.
- Pfister, L., Hupfer, F., Brugnara, Y., Munz, L., Villiger, L., Meyer, L., Schwander, M., Isotta, F. A., Rohr, C., and Brönnimann, S.: Early instrumental meteorological measurements in Switzerland, *Clim. Past*, 15, 1345–1361, <https://doi.org/10.5194/cp-15-1345-2019>, 2019.
- Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, 99, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>, 2010.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <https://www.R-project.org/> (last access: 4 February 2020), 2017.
- Rössler, O. and Brönnimann, S.: The effect of the Tambora eruption on Swiss flood generation in 1816/1817, *Sci. Total Environ.*, 627,

- 1218–1227, <https://doi.org/10.1016/j.scitotenv.2018.01.254>, 2018.
- Schiemann, R. and Frei, C.: How to quantify the resolution of surface climate by circulation types: An example for Alpine precipitation, *Phys. Chem. Earth*, 35, 403–410, <https://doi.org/10.1016/j.pce.2009.09.005>, 2010.
- Schwander, M., Brönnimann, S., Delaygue, G., Rohrer, M., Auchmann, R., and Brugnara, Y.: Reconstruction of Central European daily weather types back to 1763, *Int. J. Climatol.*, 37, 30–44, <https://doi.org/10.1002/joc.4974>, 2017.
- Schwarb, M.: The alpine precipitation climate : Evaluation of a high-resolution analysis scheme using comprehensive rain-gauge data, vol. 80 of *Zürcher Klima-Schriften*, Verlag Institut für Klimaforschung, ETH Zürich, Zürich, 2001.
- SLF: Der Lawinenwinter 1999: Ereignisanalyse, SLF Bibliothek, Davos, 2000.
- Stucki, P., Bandhauer, M., Heikkilä, U., Rössler, O., Zappa, M., Pfister, L., Salvisberg, M., Froidevaux, P., Martius, O., Panziera, L., and Brönnimann, S.: Reconstruction and simulation of an extreme flood event in the Lago Maggiore catchment in 1868, *Nat. Hazards Earth Syst. Sci.*, 18, 2717–2739, <https://doi.org/10.5194/nhess-18-2717-2018>, 2018.
- Tandeo, P., Ailliot, P., Fablet, R., Ruiz, J., Rousseau, F., and Chapron, B.: The Analog Ensemble Kalman Filter and Smoother, 4th International Workshop on Climate Informatics, Boulder, USA, 2014.
- Van Den Dool, H.: Searching for analogues, how long must we wait?, *Tellus A*, 46, 314–324, <https://doi.org/10.3402/tellusa.v46i3.15481>, 1994.
- Vieli, I.: Wenn die Tochter der Hochalpe in ihre weissen Gewänder gehüllt zu Tal donnert, *Der Lawinenwinter 1887/88 im Berner Oberland*, <https://doi.org/10.7892/boris.107117>, 2017.
- Vitasse, Y., Klein, G., Kirchner, J. W., and Rebetez, M.: Intensity, frequency and spatial configuration of winter temperature inversions in the closed La Brevine valley, Switzerland, *Theor. Appl. Climatol.*, 130, 1073–1083, <https://doi.org/10.1007/s00704-016-1944-1>, 2017.
- Weusthoff, T.: Weather Type Classification at MeteoSwiss: Introduction of new automatic classification schemes, *Arbeitsberichte der MeteoSchweiz*, 235, 2011.
- Whitaker, J. S. and Hamill, T. M.: Ensemble Data Assimilation without Perturbed Observations, *Mon. Weather Rev.*, 130, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2), 2002.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, vol. 91 of International geophysics series, Elsevier, Amsterdam, 2. ed., 2009.
- Zorita, E. and von Storch, H.: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, *J. Climate*, 12, 2474–2489, [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2), 1999.
- Zorita, E., Hughes, J. P., Lettemaier, D. P., and von Storch, H.: Stochastic Characterization of Regional Circulation Patterns for Climate Model Diagnosis and Estimation of Local Precipitation, *J. Climate*, 8, 1023–1042, [https://doi.org/10.1175/1520-0442\(1995\)008<1023:SCORCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1023:SCORCP>2.0.CO;2), 1995.