Noname manuscript No.
(will be inserted by the editor)

Impact of global atmospheric reanalyses on statistical precipitation downscaling

Pascal Horton · Stefan Brönnimann

Received: date / Accepted: date

Abstract Statistical downscaling based on a perfect prognosis approach often relies on global reanalyses to infer the statistical relationship between synoptic predictors and the local variable of interest, here daily precipitation. Nowadays, many reanalyses are available and their impact on the downscaled variable is not often considered. The present work assessed the impact of ten reanalyses on the performance of seven variants of analogue methods for statistical precipitation downscaling at 301 stations in Switzerland. Even though the study location is in a data-rich region, significant differences were found between reanalyses and their impact on the performance of the method can be even higher than the choice of the predictor variables. There was no single overall winner, but a selection of recommended reanalyses resulting in higher skill scores depending on the considered predictor variables. The impact of the output spatial resolution was assessed for different types of variables. Output resolutions below one degree were found to be often of low to no interest. Reanalyses with longer archives allow the pool of potential analogues to be increased, resulting in better performance. However, when adding variables affected by errors in a more distant past, the skill score decreased again. The use of multiple members from two reanalyses was also tested over a recent and a past period. The benefit of using members to increase the performance by better incorporating the uncertainties was found to be limited, and even problematic with methods using multiple analogy levels.

Keywords Reanalyses · Precipitation · Statistical downscaling · Analogue method

P. Horton
University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland
Tel.: +41 31 631 80 21
E-mail: pascal.horton@giub.unibe.ch
ORCID: 0000-0003-0466-0359

S. Brönniman
University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland

1 Introduction

Statistical downscaling is widely used to bridge the resolution gap between climate model outputs and impact models, and to bias-correct them, but also to bypass some physical parameterizations. Some of these methods rely on empirical statistical relationships between large-scale atmospheric variables and local variables of interest. Following the classification of Rummukainen (1997), which was also used in Maraun et al (2010), there are basically two types of approaches: perfect prognosis, for which the relationship is calibrated between large-scale and local-scale observations, and model output statistics, for which the relationship is calibrated against the outputs of a specific global or regional climate model and local-scale observations. Here we investigate an approach of the former type to downscale precipitation in Switzerland. Statistical downscaling is of particular interest for precipitation, due to the difficulty for numerical models to accurately simulate all the processes involved.

Perfect prognosis approaches rely on large-scale observations. Global atmospheric reanalyses are useful to fulfill this role, as they provide gridded large-scale variables that are available for any location in the world. Reanalyses are produced using a single version of a data assimilation system coupled with a forecast model constrained to follow observations over a long period. They provide multivariate outputs that are physically consistent, which contain information in locations where few or no observations are available, also for variables that are not directly observed (Gelaro et al, 2017). Their accuracy depends on both the quality of the model physics and that of the analysis process, and thus indirectly on the quantity and quality of the assimilated observations (Dee et al, 2011). The homogeneity of a reanalysis in time is a challenge due to significant changes in observing systems. The onset of satellite observations drastically changed the amount of data available, particularly for regions with sparse conventional observation networks. The assimilation of a temporally variable amount of observations is likely to lead to inhomogeneities in the reanalysis. For this reason, some reanalyses are limited to the satellite era, and others do not use satellite observations at all. Because of these discontinuities in the available observations, some variables from the reanalyses, such as precipitation and evaporation are to be used with great caution (Kobayashi et al, 2015).

The present work focuses on the analogue method (AM), which is a statistical downscaling technique that relies on the hypothesis that similar synoptic situations are likely to result in similar local effects, plus a certain variability that is not explained by the considered predictors (Lorenz, 1969). The local variable of interest, here, is daily precipitation. Different versions of AMs exist, relying on various predictors considered over domains of variable size. However, they generally contain predictors characterizing atmospheric circulation, considered over domains of width/length of about 5 to 20° depending on the method and the reanalysis. In order to take into account the unexplained variability, several analogue days are usually selected and their observed precipitation values are used to provide an empirical conditional distribution that is the statistical prediction for the considered target date.

In one of the first AM versions, the predictors were extracted from radiosounding data (Duband, 1981), which involved heavy pre-treatment to get a complete and homogeneous dataset that could be used. Other authors worked with rather short, local analysis from forecast models (for example Kruizinga and Murphy, 1983; Van den Dool, 1989). The release of the first reanalysis (NCEP/NCAR Reanalysis I, NR-1 – Kalnay et al, 1996; Kistler et al, 2001) greatly simplified the implementation of the AM, and made available potential new predictor variables, which increased the popularity of the method (Timbal et al, 2008).

Timbal et al (2003) and Bontron (2004) were the first authors to use NR-1 in the AM. NR-1, and its updated version NCEP/DOE Reanalysis 2 (NR-2 – Kanamitsu et al, 2002), remained popular for a long time and were often used until recently in AMs (Wetterhall et al, 2005; Gangopadhyay et al, 2005; Altava-Ortiz et al, 2006; Barrera et al, 2007; Cannon, 2007; Matulla et al, 2007; Bliefernicht and Bárdossy, 2007; Maurer and Hidalgo, 2008; Wu et al, 2012; Marty et al, 2012; Teng et al, 2012; Horton et al, 2012; Yiou et al, 2014). The first European long reanalysis ERA-40 (Uppala et al, 2005) then became popular within the European community (Willems and Vrac, 2011; Themessl et al, 2011; Ben Daoud et al, 2011; Turco et al, 2011; Franke et al, 2011; Pascual et al, 2012; Schenk and Zorita, 2012; Ribalaygua et al, 2013; Osca et al, 2013; Radanovics et al, 2013; Martín et al, 2014; Chardon et al, 2014; Ben Daoud et al, 2016). Ben Daoud et al (2009) analyzed the impact of choosing NR-1 or ERA-40 in the AM developed by Bontron (2004) and found no significant difference for the predictors considered. The more recent ERA-Interim (ERA-INT, Dee et al, 2011) was used by Raynaud et al (2016), and MERRA (Rienecker et al, 2011) was used by Vanvyve et al (2015). Several recent reanalysis products have not yet been used in AMs.

In almost all of these works, a single reanalysis was used. The choice is likely to be primarily driven by the ease of access and the availability of some datasets in research units, along with the code required to read them. Indeed, it might not be considered as a priority to use the latest reanalysis available if the benefit for AMs is unknown, as it requires effort to acquire ever larger datasets and to adapt code to read them. Moreover, they are often considered as rather equivalent for a data-rich region, such as Europe.

AMs are also used to reconstruct weather conditions for the more distant past, such as the entire Twentieth Century. Then, reanalyses spanning this period are required, such as the ECMWF twentieth century reanalyses (ERA-20C or CERA-20C – Poli et al, 2016; Laloyaux et al, 2016) or the Twentieth Century Reanalysis (20CR – Compo et al, 2011) produced by NOAA (for example, Kuentz et al, 2015; Caillouet et al, 2016; Brigode et al, 2016; Bonnet et al, 2017).

To our knowledge, Dayon et al (2015) made the most comprehensive comparison of the reanalyses in the AM so far. They compared NR-1, MERRA, ERA-INT and 20CR in terms of inter-annual correlations and biases and noted that the choice of the reanalysis is a non-negligible source of uncertainty, and that it can even impact the performance of the method to a greater extent than the choice of the predictors. They concluded that "the substantial differences in downscaling results associated with reanalyses [...] suggests that the role of reanalyses should not be underestimated when evaluating the statistical downscaling method". The choice of the predictors was also found to vary from one reanalysis to another, in a way that the optimization of the method is likely to be reanalysis dependent and that using a single reanalysis might introduce a lack of robustness (Dayon et al, 2015). Reanalyses were also found to impact other statistical downscaling methods (e.g. Koukidis and Berg, 2009).

The present work aims at assessing the impact of most of the currently available reanalyses on the performance of the AM. Ten reanalyses were compared for seven AMs at 301 stations in Switzerland (Sect. 3). Additionally, the role of spatial resolution (Sect. 4.1), the length of the archive (Sect. 4.2), and the use of different members from ensemble datasets (Sect. 4.3) were investigated. The discussion and conclusion (Sect. 5) provide some guidelines for the use of these reanalyses in AMs.

2 Data and methods

2.1 Reanalysis datasets

Different types of reanalyses exist, primarily characterized by their observational inputs. Fujiwara et al (2017) define three classes: "surface-input" reanalyses that assimilate surface data only, "conventional-input" reanalyses that additionally assimilate upper-air conventional data, and "full-input" reanalyses that additionally assimilate satellite data.

The global atmospheric reanalyses under evaluation are briefly described hereafter, providing first the full and conventional-input datasets (1–6), and then the surface-input ones (7–9). Some of their characteristics are provided in Table 1. The period common to all datasets is 1981–2010. The predictors are considered at a 6-hr time step in the present work, even though some products have higher temporal resolutions.

2.1.1 NCEP Reanalysis I

The NCEP/NCAR Reanalysis I (NR-1 – Kalnay et al, 1996; Kistler et al, 2001) was the first global reanalysis. It was done with a forecast model frozen at the state-of-the-art of 1995 and is a full-input dataset. Upper-air observations were found to have a much larger influence on the analysis than the surface observations (Kistler et al, 2001). The data assimilation system is a 3D variational technique (3D-Var). The model resolution is T62 (about 210 km) with 28 sigma levels. All major physical processes are parameterized. The period of coverage initially started in 1957, before being extended back to 1948. Kalnay et al (1996) were aware that assimilating all the available data at a given time would have an impact on the climate of the reanalysis due to changes in the observing system, but the choice was made for accuracy over stability of the climate. A comparison of two sets of analyses made with and without the use of satellite data showed that even without satellite data, almost 100% of the daily variance of the geopotential height was explained in the Northern Hemisphere (NH) extra-tropics (Kalnay et al, 1996). Lower correlation values were found in other regions of the globe, particularly in the Southern Hemisphere (SH), where the uncertainty is much higher due to the lack of rawinsonde data. However, RMS of the analysis increments (the differences between the forecast and the analysis) at 500 hPa showed large differences between a data-poor year (1958) and a data-rich year (1996), and the climate before and after 1979 differ significantly due to the use of satellite data (Kistler et al, 2001).

2.1.2 NCEP Reanalysis II

The NCEP/DOE Reanalysis 2 (NR-2 – Kanamitsu et al, 2002) is a follow-on to NR-1 that aims to correct some identified problems. However, these issues have consequences for a limited number of applications. NR-2 also relies on updated versions of the assimilation system and the forecast model, with improvements to the model physics. Changes in parameterizations have improved the precipitation estimate, but may have caused deterioration of other variables (Kistler et al, 2001; Kanamitsu et al, 2002). Geopotential heights only exhibit minor differences when compared to those of NR-1. The model and the outputs have the same spatial and temporal resolution as NR-1, and, mostly, the same observational data were assimilated. The dataset starts in 1979.

2.1.3 ERA-Interim

ERA-Interim (ERA-INT – Dee et al, 2011) is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and covers the period from 1979 onwards. It replaced ERA-40 (Uppala et al, 2005), which replaced ERA-15 (Gibson et al, 1997), reanalyses of 45 and 15 years respectively. ERA-INT aims to address problems in data assimilation of ERA-40.

ERA-INT uses a 4D variational technique (4D-Var) with sequential data assimilation in 12-hourly analysis cycles. 4D-Var is expected to make a more effective use of observations (Dee et al, 2011). ERA-INT also relies on several bias and error correction techniques that were introduced after ERA-40, in order to minimize inconsistencies between observations of different types.

The forecast model uses a hybrid sigma-pressure vertical coordinate on 60 layers and has a T255 horizontal resolution (about 79 km) and a 30 min time step. Orographic effects and convection schemes, among others, have been improved since ERA-40.

2.1.4 Climate Forecast System Reanalysis

The Climate Forecast System Reanalysis (CFSR – Saha et al, 2010) is provided by NCEP. The model resolution has increased significantly since NR-1 and NR-2: horizontal resolution of T382 (about 38 km) and 64 levels on sigma-pressure hybrid vertical coordinates. Both the forecast model and the assimilation were improved, and a coupling to the ocean, as well as a sea-ice model, were introduced. New parameterizations were used, resulting in more realistic moisture prediction and mountain blocking representation, among others (Saha et al, 2010). Temperature and moisture are also better adjusted to match the observed radiances.

CFSR was the first to use the historical tropical storm locations to avoid distortion of the circulation by the mismatch of guess and observed locations. The assimilation scheme relies on the 3D-Var technique, but with a certain consideration of the time aspect by using time tendencies of state variables. The analysis system used in CFSR for the atmosphere is similar to the one used by MERRA (Rienecker et al, 2011), with nearly the same input data. The period covered is from 1979 onwards, but with a plan to extend it back to 1947 or earlier (Saha et al, 2010).

2.1.5 Japanese 55-year Reanalysis

The Japanese 55-year Reanalysis (JRA-55 – Kobayashi et al, 2015; Harada et al, 2016) is produced by the Japan Meteorological Agency (JMA). It starts in 1958, which makes it the first reanalysis applying 4D-Var to this period. The forecast model used has a TL319 spectral resolution (about 60 km) and 60 levels in the vertical. JRA-55 shows substantial improvements compared to JRA-25 (Onogi et al, 2007), the first Japanese product. The observations used consist of those archived by JMA and those used in ERA-40 (Uppala et al, 2005). Tropical cyclones data are also assimilated, and they are well represented compared to other reanalyses (Harada et al, 2016). JRA-55 is sensitive to changes in the observing networks for some characteristics, but far less than JRA-25 was, which is probably related to improvements in the forecast model providing greater physical consistency of the JRA-55 product (Kobayashi et al, 2015).

JMA also released JRA-55 Conventional (JRA-55C – Kobayashi et al, 2014), a version of the reanalysis based on the assimilation of only conventional data, including upper air observations, without any satellite observation. The dataset is thus more homogeneous as it is unaffected by changes in satellite observing systems, even though the temporally variable number of observations may also have an impact. JRA-55C starts in 1972; the full 55-year reanalysis is obtained by using outputs from JRA-55 prior to 1972.

Globally, the anomaly of geopotential height is highly correlated between both datasets, except where conventional observations are sparse, especially for high latitude areas of the SH (Kobayashi et al, 2014).

2.1.6 MERRA-2

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2 – Gelaro et al, 2017) is an improvement of the first MERRA reanalysis (Rienecker et al, 2011) produced by NASA's Global Modeling and Assimilation Office (GMAO). One of its objectives is to improve the hydrological cycle represented in reanalysis products, primarily by providing improvement in precipitation and water vapor climatology. An important improvement in MERRA-2 over MERRA is that it shows a reduction of biases and imbalances in the water cycle, and a reduction of discontinuities in precipitation related to changes in the observing system (Gelaro et al, 2017). The forecast model has also improved both in its dynamical core and its physical parameterizations.

A peculiarity of MERRA-2 compared to the other reanalyses considered in the present work is that it uses a finite-volume dynamical core with a cubed-sphere horizontal discretization rather than a spectral model. The model grid has a relatively uniform resolution of 0.5° x 0.625° with 72 levels in the vertical.

2.1.7 NOAA-CIRES 20th Century Reanalysis

The Twentieth Century Reanalysis version 2c (20CR-2c – Compo et al, 2011) produced by NOAA starts in 1851. Unlike the other reanalyses, it only assimilates surface pressure data and relies on observed monthly sea-surface temperature and sea-ice distributions as boundary conditions. The omission of upper-air and satellite observations aims at increasing the homogeneity of the reanalysis over the whole period. The consequence is that the dataset is not the best estimate for more recent periods compared to other reanalyses (Poli and National Center for Atmospheric Research Staff, 2017).

The assimilation technique used is an Ensemble Kalman Filter (EnKF) that allows time-variant observational uncertainty related to the evolution of the measuring networks to be taken into account. The forecast model used is the NCEP Global Forecast System (GFS) with a T62 horizontal resolution and 28 vertical hybrid sigma-pressure levels. The reanalysis contains 56 members and an ensemble mean. As expected, the ensemble uncertainty varies with the time-changing observation network, i.e., it decreases over time. The outputs are available with a 2° resolution on 24 pressure levels (for the ensemble mean – fewer levels are publicly available for the individual members).

Although 20CR-2c only relies on surface data, it shows relevant information for the state of the atmosphere at higher levels, such as the 500 hPa geopotential height and the 850 hPa air temperature (Compo et al, 2011).

2.1.8 ECMWF 20th Century Reanalysis

The ECMWF twentieth century reanalysis (ERA-20C – Poli et al, 2016) starts in 1900. Unlike 20CR-2c, it is single-member. Additionally to surface pressure, ERA-20C also assimilates marine wind observations. It is forced by sea surface temperature, sea ice cover, atmospheric composition changes, and solar forcing. The forecast model used is the ECMWFs Integrated Forecast System (IFS) with a time step of 30 min, a T159 resolution (approximately 125 km), and 91 levels. The assimilation technique is 4D-Var on a 24 h window, which is also able to account for spatially and temporally varying errors in the model and the observations. A previously produced 10-member ensemble was used to derive these errors estimates.

2.1.9 ECMWF Coupled 20th Century Reanalysis

The ECMWF coupled twentieth century reanalysis (CERA-20C) is an update of ERA-20C, with an additional coupling to the ocean and a more recent version of the IFS model (Laloyaux et al, 2018). It provides 10 members and spans the period 1901–2010. The additional assimilated data are ocean temperature and salinity profiles. The coupled data assimilation system is able to accommodate feedback between the ocean and atmosphere in the forecast, as well as the analysis step through an additional iteration to account for the update of each component (Laloyaux et al, 2016), which ensures physical consistency between the upper ocean and the lower atmosphere. Changes in atmospheric temperature occur near the ocean surface, but there is no impact for the upper atmosphere. The coupled system has shown a neutral impact for the geopotential height or wind speeds (Laloyaux et al, 2016).

2.2 Precipitation dataset

The predictands – variables to be predicted – considered here were daily precipitation totals (06:00 h UTC to 06:00 h UTC the following day) at 301 weather stations of the MeteoSwiss network in Switzerland (Fig. 1). All stations with a good data record over the period 1981–2010 were considered. Often, applications of AMs use gridded precipitation or catchment-scale aggregated series, but any data manipulation was avoided here to obviate undesired interference with the sensitivity analysis. Precipitation data were also not transformed by a square root, as they are in some other studies (see e.g. Bontron, 2004). Thirty stations – those with longer time series – were selected for additional analyses (Sect. 4). Out of these 30 stations, 20 start in 1881 or earlier, four in 1882, two in 1883, two in 1884, and the last two in 1886 and 1887.

The 30-year precipitation dataset was divided into a calibration period (CP) and an independent validation period (VP). In order to reduce the impact of potential inhomogeneities in the time series, the selection of the VP was evenly distributed over the entire series (as in Ben Daoud, 2010). A total of 6 years was considered for the VP by selecting 1 year out of every 5 (explicitly: 1985, 1990, 1995, 2000, 2005, 2010). The archive period (AP), where the analogue dates are being retrieved, is the same as the CP for most of the study, except in Sect. 4.2. The VP is also excluded from the AP (days from the VP were never used as candidate situations for the selection of analogues), as well as a period of ±30 days around the target date. Unless stated otherwise, all results are presented for the VP; results on the CP were similar.

2.3 Considered analogue methods

Different variants of the AM were considered in the present work (Table 2). These methods have varying degrees of complexity and comprise one or more subsequent levels of analogy with predictor variables of different kinds. The first method developed with NR-1 by Bontron (2004) is based on the analogy of synoptic circulation on the geopotential height at two pressure levels (Z1000 at +12 h and Z500 at +24 h) and is known in this work as 2Z.

The 2Z method consists of the following steps: firstly, to cope with seasonal effects, candidate dates are extracted from the AP within a period of four months centered around the target date, for every year of the archive (PC: preselection on calendar basis in Table 2). Then, the analogy of the atmospheric circulation of a target date with every day from the preselection set (excluding a period of ±30 days around the target date along with the VP) is assessed by processing the S1 criterion (Eq. 1, Teweles and Wobus, 1954; Brown et al, 2012), which is a comparison of gradients, over a defined spatial window (the domain on which the predictors are compared). S1 is processed on each level and the average is then considered, here with the same weights.

$$S1 = 100 \frac{\sum_i |\Delta \hat{z}_i - \Delta z_i|}{\sum_i max\left\{|\Delta \hat{z}_i|, |\Delta z_i|\right\}} \tag{1}$$

where $\Delta \hat{z}_i$ is the difference in geopotential height between the i-th pair of adjacent points of gridded data describing the target situation, and Δz_i is the corresponding observed geopotential height difference in the candidate situation. The smaller the values S1 are, the more similar the pressure fields. This criterion, being processed on gradients, is insensitive to biases in the considered predictors, as long as the circulation is correctly represented.

The N_1 dates, where N_1 is a parameter to be calibrated, with the lowest values of S1 are considered as analogues to the target date. Then, the daily observed precipitation values of the N_1 selected dates provide the empirical conditional distribution, considered as the probabilistic prediction for the target date.

A variation of the former method, but based on the mean sea level pressure (2SLP), rather than the geopotential height, was also assessed in this work. The S1 criterion was also used to quantify the analogy between the pressure fields. SLP was used in AMs by Zorita and von Storch (1999), Timbal and McAvaney (2001) and Martín et al (2014), amongst others.

Another method relying only on atmospheric circulation has also been considered. It uses the geopotential height on four combinations of pressure levels and temporal windows (4Z, Table 2) at levels that were automatically selected by genetic algorithms for the upper Rhone catchment in Switzerland (Horton et al, 2017a). The 4Z method was shown to outperform 2Z by exploiting more information from the geopotential height and by taking advantage of additional degrees of freedom, such as different spatial windows between the pressure levels and the introduction of a weighting between them. However, due to the high number of reanalyses and stations considered in this work, it was not possible to use genetic algorithms in order to optimize the method. Thus, the 4Z method considered here is a simplification of the results from Horton et al (2017a), and only the selection of the optimal pressure levels and temporal windows were considered (Z1000 at +06 h and +30 h, Z700 at +24 h, and Z500 at +12 h), and used for all stations. Such simplifications of the parameters resulted in a decrease of the performance score, which, however, was still superior to that of 2Z.

The other methods considered hereafter add a second, or more, subsequent level(s) of analogy after the analogy of the atmospheric circulation, in a stepwise manner.

The next method adds a second level of analogy with moisture variables (method 2Z-2MI, Table 2), using a moisture index (MI), which is the product of the total precipitable water (TPW) and the relative humidity at 850 hPa (RH850) (Bontron, 2004). When adding a second level of analogy, N_2 dates are subsampled from the N_1 analogues of the atmospheric circulation, to end up with a smaller number of analogue situations. When this second level of analogy is added, a higher number of analogues N_1 is kept at the first level.

Similar to the 4Z method, the 4Z-2MI is a simplification of the methods optimized by genetic algorithms in Horton et al (2017a). It consists of a first level of analogy on the geopotential height at four pressure levels (Z1000 at +30 h, Z850 at +12 h, Z700 at +24 h, and Z400 at +12 h), different from 4Z, followed by the moisture index (MI) at two pressure levels (MI700 at +24:00 h and MI600 at +12 h).

To constrain the seasonal effect, Ben Daoud et al (2016) replaced the calendar preselection (± 60 days around the target date) by a preselection based on similarity of air temperature (T925 at +36 h and T600 at +12 h, at the nearest grid point). It allows a more dynamic screening of similar situations in terms of air masses as the seasonal signal is also present in the temperature data. The undesired mixing of spring and autumn situations is discussed in Caillouet et al (2016). The number of preselected dates (N_0) is equivalent to the number of days selected with the calendar approach, and thus depends on the archive size. In this method, named PT-2Z-4MI, the analogy of the atmospheric circulation is the same as in the 2Z method, but the moisture analogy is different (MI925 and MI700 at +12 h and 24 h).

Subsequently, Ben Daoud et al (2016) introduced an additional level of analogy between the circulation and the moisture analogy (PT-2Z-4W-4MI, Table 2), based on the vertical velocity at 850 hPa (W850). This AM, named "SANDHY" for Stepwise Analogue Downscaling method for Hydrology (Ben Daoud et al, 2016; Caillouet et al, 2016), was primarily developed for large and relatively flat/lowland catchments in France (Saône, Seine) and is the most complex method considered in this work. It has also been applied to the whole France territory by Radanovics et al (2013) with ERA-40 and by Caillouet et al (2016) with 20CR-V2b.

Precipitation variables from reanalyses are generally not considered as predictors, as they strongly depend on the model physics (Rienecker et al, 2011) and have significant biases, which would make them not interchangeable with the outputs of another model. Dayon et al (2015) assessed the relevance of using precipitation from four reanalyses as predictors and finally rejected precipitation as a predictor due to strong biases in the downscaled series.

2.4 Calibration of the AMs

The parameters (specific to each level of analogy) that were calibrated here for every station, method, and reanalysis, are: (1) the spatial windows, which are the domains on which the predictors are compared, and (2) the optimal number of analogues to select.

The semi-automatic sequential procedure developed by Bontron (2004) was used to calibrate the AM. The procedure is described in Horton et al (2017b) and is similar to the work of Radanovics et al (2013) and Ben Daoud et al (2016). It was implemented in the open source AtmoSwing-optimizer software v1.5.0 (www.atmoswing.org, Horton, 2017), which was used to perform the calibrations and the analyses.

When calibrating the method, the CRPS (Continuous Ranked Probability Score, Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000) is often used as the objective function. It allows evaluating the predicted cumulative distribution functions $F(y)$, here of the precipitation values y associated with the analogue situations, compared to the single observed value y^0 for a day i:

$$CRPS_i = \int_0^{+\infty} \left[F_i(y) - H_i(y - y_i^0)\right]^2 dy \tag{2}$$

where $H(y - y_i^0)$ is the Heaviside function that is null when $y - y_i^0 < 0$, and has the value 1 otherwise; the better the prediction, the lower the score.

Its skill score expression is often used, with the climatological distribution of precipitation as the reference. However, the choice of a reference is not important when comparing performances. The CRPSS (Continuous Ranked Probability Skill Score) is thus defined as follows (Bradley and Schwartz, 2011):

$$CRPSS = 1 - \frac{\overline{CRPS}}{\overline{CRPS_{clim}}} \tag{3}$$

where $CRPS_{clim}$ is the CRPS value for the climatological distribution. A better prediction is characterized by an increase in CRPSS.

All AMs were calibrated for every reanalysis and station, which resulted in a total of 21,070 calibrations being processed on a HPC cluster at the University of Bern. For every combination, the spatial windows and the number of analogues of each analogy level were calibrated for each station in order to be optimal. These optimized parameters are the focus of another coming article and were published as datasets (see Sect. Data availability).

3 Impact of the reanalysis

The results of the reanalyses comparison are shown for the VP (independent validation period, Sect. 2.2). The reanalyses were used at their original spatial resolution and thus differ from one another, the impact of which is analyzed in Sect. 4.1.

The results of 20CR-2c are shown here for the ensemble mean only (see Sect. 4.3 for the impact of using multiple members). The same analyses were performed on a single member (the first one), but no significant difference was observed. The single-member was slightly less skillful than the ensemble mean, but to a negligible extent (not shown).

One has to keep in mind that biases in the variables might not affect the performance of the AM, as long as they are constant over time and the prediction methods are used in a perfect prognosis framework. For example, a constant bias in the values of Z will not alter the selection of analogues, whereas a bias in the circulation frequency will affect the performance.

3.1 Impact on the skill

The CRPSS of all considered AMs and reanalyses are shown in Fig. 2. Globally, the skill tends to increase with the complexity of the AM. The first two methods based on two circulation predictors, 2SLP and 2Z, were equivalent, except for MERRA-2, where SLP showed a higher predictive skill than Z. Then, there was a systematic increase of the skill from 2Z, 4Z, 2Z-2MI, up to 4Z-2MI. Finally, the respective performance of 4Z-2MI, PT-2Z-4MI and PT-2Z-4W-4MI varied from one reanalysis to another. The spread was relatively similar between reanalyses.

As Dayon et al (2015) also observed for inter-annual correlations, the reanalysis had an impact on the skill of the AM that was sometimes larger than the choice of predictors, and is thus a non-negligible source of uncertainty. The impact of the reanalysis was isolated in Fig. 3 by processing the difference in CRPSS for one reanalysis compared to the mean performance on all reanalyses, per station and per method. The variability is reduced because the climatological differences between the stations were mostly removed. Except for 2SLP, there is a tendency for the impact of the reanalysis to increase with the complexity of the method. This is particularly visible for ERA-INT, JRA-55, JRA-55C and 20CR-2c. The boxplot spread cannot be interpreted in Fig. 3, as it is more akin to the average performance of all the reanalyses.

In general, modern full-input or conventional-input reanalyses, including ERA-INT, CFSR, JRA-55, JRA-55C, and MERRA-2, performed better than older ones (NR-1 and NR-2) and the surface-input ones (20CR-2c, ERA-20C, and CERA-20C) for this region of the globe (i.e., Switzerland), independently of the assimilation technique or the availability of high resolution outputs.

The first two reanalyses NR-1 and NR-2 were mostly slightly below the average. ERA-INT generally performed well, except for 2SLP, where it showed lower skills for several stations. However, the addition of more levels of the geopotential height or moisture variables made it a skillful dataset (from 2Z-2MI on). CFSR was always in the best reanalyses, except when vertical velocity was used, which decreased slightly its performance. The two Japanese reanalyses JRA-55 and JRA-55C performed equally well, despite the fact that JRA-55C does not assimilate satellite observations. MERRA-2 was also part of the top selection, and its SLP was found to be particularly skillful compared to other reanalyses. 2SLP with MERRA-2 was found to perform even better than using four levels of the geopotential height. 20CR-2c systematically resulted in lower performances, and its relative skill significantly decreased for more complex methods. ERA-20C, which is also a surface-input dataset, had an average impact. It did perform slightly better than NR-1 and NR-2, and largely better than 20CR-2c, but not as well as the full-input reanalyses. CERA-20C performed similarly to ERA-20C.

The impact of the reanalysis was then investigated by considering precipitation thresholds for the target date (not shown). The same tendencies could generally be observed for all thresholds considered, with a nuance: MERRA-2's remarkably high skill score for 2SLP was first related to days with precipitation, of any intensity.

Daily correlations were processed between the median or the mean precipitation from the selected analogues and the observations. The results were similar to Fig. 2 (same relative differences) and are thus not presented. The inter-annual Pearson correlation coefficient was processed in the same way (Figure 4, based on the mean precipitation), but both the CP and the VP were included to increase the sample size. There is only a slightly increasing trend in the correlation coefficient with the complexity of the method, but most of the differences are between reanalyses, with a growing impact for more complex methods. When using moisture variables, ERA-INT, MERRA-2, and CERA-20C were slightly superior to the others. Although it is usually advisable to assess different properties of the optimized methods, one has to remember that these were not optimized for this metric specifically. Thus, optimizing the methods in terms of inter-annual correlation might have resulted in a totally different picture. In conclusion, this analysis should not be used as a basis for selecting a dataset over another for a work relying on inter-annual correlations, but it shows that there are non-negligible differences between datasets in terms of annual volumes that should also be considered.

An analysis of variance (ANOVA) emphasized highly significant impact of the reanalyses on the skill. A Tukey Honest Significant Differences test showed highly significant differences of the skill between all pairs of reanalyses, except between NR-1 and NR-2 and between JRA-55, JRA-55C, ERA-INT, and CFSR. The NR-1 – NR-2 and JRA-55 – JRA-55C pairs are produced with the same model and are very similar products. The contribution to variance was thus processed after removing NR-2 and JRA-55C from the analysis to work with a setting more respectful of the independence assumption. As the three datasets JRA-55, ERA-INT, and CFSR might also not be independent, the same analysis was performed again by additionally removing JRA-55 and CFSR (contribution provided in parentheses hereafter). In order to remove the influence of the different climatic conditions at each station, the mean skill score per station (for all methods and reanalyses) was subtracted before processing the variance decomposition. The contribution to variance of the skill score was finally 63.8% (60.2%) for the methods, 20.4% (23.6%) for the reanalyses, 3.7% (4.3%) for the interaction between methods and reanalyses, and 12.0% (11.9%) for the residuals. An analysis with linear mixed-effects models was also performed and provided similar results.

The impact of the reanalyses on the biases was assessed for the first analogue. Considering only the first analogue is not recommended when using the results of the AM for hydrological modelling for example, but it was considered reasonable for the purpose of comparing reanalyses. A better way would be to use an approach such as the Schaake Shuffle (Clark et al, 2004) that reorders the ensemble members (here the analogue dates) in order to restore consistency in the spatio-temporal variability. Figure 5 shows that the biases seem to depend on both the method and the reanalysis. In terms of methods, 2SLP induced a dry bias for most reanalyses, as well as PT-2Z-4W-4MI, while PT-2Z-4MI resulted in a wet bias for most reanalyses. The bias related to 2Z-2MI and 4Z-2MI was generally more contained within a relative 5% range for most of the reanalyses. It can be noted that the reanalyses showing the larger bias in PT-2Z-4W-4MI were the ones with a higher CRPSS. The bias of PT-2Z-4W-4MI is due to a selection of too many dry days (Caillouet et al, 2016), which is addressed in Caillouet et al (2017). MERRA-2 is often showing a slightly stronger dry bias. NR-1, NR-2 and 20CR-2c were generally on the higher (wetter) part of the ensemble of reanalyses, which was to their advantage when the others showed a dry bias, but which was detrimental when the ensemble was more balanced. Putting this in perspective with the known dry bias of PT-2Z-4W-4MI, it is likely that a wet bias related to these three reanalyses by chance compensated the dry bias of these methods. It should then not be an argument to consider them as superior to the others.

3.2 Spatial patterns

The 301 precipitation stations are located at different elevations and are subject to various meteorological influences. In order to analyze spatial patterns of the methods/reanalyses relationships, maps of the best methods per reanalysis are presented in Fig. 6. The selection of an optimal method was not systematic for all stations, but some spatial patterns appeared, depending on the local climate. The three most complex methods (4Z-2MI, PT-2Z-4MI, and PT-2Z-4W-4MI) were almost always selected. The PT-2Z-4MI and PT-

2Z-4W-4MI methods were developed for the context of large and relatively flat/lowland catchments, and 4Z-2MI in the context of the upper Rhone catchment in Switzerland. There is a tendency in these maps for the methods to be selected as optimal in their original context, respectively in relatively flat plains or an Alpine environment. Indeed, the use of a variable, such as vertical velocity, at a relatively low resolution may still make sense in large plains as an uplift/subsidence index, but may be less relevant in narrow alpine valleys. The variability between the maps is probably related to the predictive skill of the variables from the different reanalyses. Overall, vertical velocity seems to be sub-optimal in 20CR-2c, but preferable in JRA-55(C) and ERA-20C.

A similar figure shows the best reanalyses for each method (Fig. 7). One has to keep in mind that only the best reanalysis is shown, but others might provide almost similar results. For 2SLP, MERRA-2 was the best reanalysis for almost all stations. The other methods did not show a single best reanalysis, but a selection of about 3–4 datasets. This selection was not completely random, as some spatial patterns could be identified. For the south-eastern part of Switzerland, MERRA-2 was often selected as the best option for different methods. Both methods based on the geopotential height only (2Z and 4Z) showed some clusters of CFSR, JRA-55(C), ERA-INT and CERA-20C, with a more defined pattern for 4Z with JRA-55(C) and the European products on the Plateau and CFSR in the reliefs. When moisture was added (from 2Z-2MI on), ERA-INT was more present as the first choice, particularly for the western and northern part of Switzerland. The presence of four variables from the geopotential heights in 4Z-2MI gave the advantage to CFSR over ERA-INT. The preselection on the temperature in PT-2Z-4MI introduced a cluster of CFSR in the eastern central part of Switzerland, while ERA-INT and MERRA-2 dominated the rest. CFSR was far less selected when vertical velocity was introduced in PT-2Z-4W-4MI, while JRA-55(C) appeared as one of the favorites along with ERA-INT and MERRA-2. In conclusion, the choice of the reanalysis and the AM should take into account the context of the area of interest.

3.3 Selection of the analogue dates

The use of a particular reanalysis in preference to another has an influence on the selection of the analogue dates. These were compared between reanalyses for all stations and all AMs. Figure 8 shows the percentage of identical analogue dates, per target date, selected when using the reanalyses in columns that were also found when using the reanalyses in rows for the different AMs. The values were averaged over time for all stations on the VP (same results on the CP). The different spatial resolutions are likely to play a role in the differences of selected analogue dates. Additionally, the spatial windows on which the predictors were compared might differ from one reanalysis to another (as the methods were calibrated for all stations and all reanalyses independently), which could potentially also play a role.

As expected, more complex AMs showed lower percentages of identical analogue days between the reanalyses. Indeed, higher correspondence is expected for circulation variables than moisture variables, which are more model-dependent. Reanalyses that are relatively similar, such as NR-1 and NR-2 or JRA-55 and JRA-55C, showed the highest percentage of shared dates. Higher similarities were also observed between CERA-20C and ERA-20C for methods based on circulation, but not, significantly, for more complex methods. This suggests that at least humidity variables are substantially different between CERA-20C and ERA-20C. The selection based on ERA-INT, JRA-55 and JRA-55C had globally the highest correspondence to the other reanalyses.

20CR-2c differed the most from other reanalyses for most methods. This difference in the selection of analogue days led to lower performance of the methods (Fig. 3). Another noticeable difference is for MERRA-2 and the 2SLP method; in this case, this departure led to better performance scores (Fig. 3).

The same analysis has been performed for different precipitation thresholds (days with precipitation $> 0.1\ mm$, and the 95^{th} and 99^{th} percentiles of rainy days) and the results are provided in the supplementary material. Overall, the patterns remained akin, but the percentage of similar days was slightly superior for days with high precipitation.

4 Assessing the characteristics of reanalyses

4.1 On the spatial resolution

The different reanalyses are characterized by various grid resolutions. Obviously, higher model resolutions usually allow for better modelling precipitation. What is not so clear, however, is the influence of the output grid resolution within the AM. In order to assess its impact on the methods performance, reanalyses with higher resolution were degraded to increasingly lower resolutions. This was performed simply by skipping points, which provided reduced resolution as factors of the original one. No more-advanced techniques, such as spectral transformations, were considered. For each resolution, the parameters of the AMs were calibrated again, independently for every method, reanalysis, and station, and were thus optimal for a given configuration.

The impact of the degradation in resolution is presented in Fig. 9 for six AMs and a selection of 30 stations (orange points in Fig. 1). No significant impact on the skill of the methods was found between a resolution of about 1° and higher resolutions, at least for the geopotential height. MERRA-2's SLP might benefit a bit more from high resolution than others. The geopotential at 500 hPa presents a half-autocorrelation distance of about 1000 km for equivalent latitudes (Thiébaux, 1985), so future increases in output resolutions should not bring substantial improvements to the circulation analogy. Higher model resolutions might however allow for better representation of orographic effects and complex processes, and thus improve the variables' accuracy.

Beyond 1°, the decrease in performance was systematic, but not of the same magnitude for every reanalysis and method. As expected, methods relying on Z were less sensitive to the resolution than the ones with moisture variables that have a smaller autocorrelation distance. 2SLP was more sensitive to the resolution than 2Z and 4Z, as it relies on a single level, which is, besides, more variable than the geopotential height at higher levels. When geopotential heights are considered alone, even a reduction of the resolution to 2–3° had limited impact. The most complex method, PT-2Z-4VV-4MI, was globally the most sensitive to the resolution as it relies on more local information.

4.2 On the archives length

All previous comparisons were performed with the period 1981–2010 as AP. However, some reanalyses have the important added value of covering longer periods. Longer reanalyses have mainly two benefits: they allow the investigation of periods in the past, for example to reconstruct the meteorological conditions related to a flood event, and they enrich the pool of potential analogue situations, primarily for less frequent situations; the second aspect was the focus of the present analysis. Ruosteenoja (1988) and Van Den Dool (1994) have shown that a longer archive improves the quality of the meteorological analogy.

Different AMs were recalibrated on the same CP as before and assessed on the same VP (Sect. 2.2), but with an increasing AP, which constituted the pool of potential analogue situations, by adding dates (by blocks of 10 to 20 yrs) farther in the past back to 1881 (for 20CR-2c). The influence of the archive's length on the VP is presented in Fig. 10 for five AMs and the NR-1, JRA-55, CERA-20C, and 20CR-2c reanalyses, on the 30 stations with longer precipitation series available (Fig. 1). Note that some precipitation data are missing for a couple of stations prior to 1887 (Sect. 2.2), but this does not seem to impact the analysis.

As expected, there was an overall improvement in skill with archives longer than the 24 years from the CP. The gain of longer archives for AMs based on the atmospheric circulation only (2Z and 4Z, Fig. 10 panel a) was generally superior to other methods with multiple levels of analogy. Figure 10 also shows that the improvement did not increase constantly with the archive's size, and a decrease of the performance even appeared for some reanalyses and methods. NR-1 showed a discontinuity in performance when adding moisture variables from the period 1961–1971, and CERA-20C showed a decrease for different methods from about 1941 backwards.

With perfect predictor and predictand (precipitation) archives, the prediction skill of the different methods would only increase thanks to the enrichment of the pool of potential analogues, up to a certain point where it might flatten out. A decrease in performance can be explained by the presence of less good analogues that degrade the prediction. The presence of less good analogues can be due to (a) the non-preservation of the relationship between predictors and predictands over time, (b) errors in the precipitation archives, or (c) inhomogeneities or errors in the early years of the reanalyses. It is obvious that the quality of precipitation measurement is not constant over time, and that the climate system presents trends on that period. However, if these were the main reasons, a break in performance would have appeared at the same time for all reanalyses and methods, as they all rely on the same predictand time series. The presence of breaks at different years that are reanalysis– and variable–dependent would suggest that the variability in the predictors' quality is likely the causative factor. It must be noted here that differences in improvements between reanalyses in Fig. 10 do not represent differences in quality between datasets, as these improvements must be interpreted relatively to the baseline performance of the reanalyses.

NR-1 is known to have significant differences between climates before and after the introduction of satellite data (Kistler et al, 2001), which might explain these discontinuities. CERA-20C and 20CR-2c are more homogeneous in terms of the type of observations that are assimilated, but the number of observations fluctuates over time, resulting in higher variability for the early years. Thus, for periods where measurements were scarce, the models were less constrained to observations and predictors such as moisture, temperature and vertical velocity are more uncertain. First guess errors or ensemble spreads from a given reanalysis might be used to motivate the choice of an acceptable archive period.

4.3 On the use of ensemble members

As discussed in the previous section, the reanalyses spanning the 20^{th} century are more uncertain for the early part of the period. In order to take this uncertainty into account, CERA-20C and 20CR-2c provide 10 and 56 members respectively. These ensemble datasets can be used in the AM by looking for similar days in every member. Both the target and the candidate situations are thus extracted from the same member. Two options are possible for merging the selected analogues: (a) by keeping all analogue dates including the duplicates, or (b) by removing duplicates. For both options, the optimal number of analogues needs to be reassessed. If the data from the different members were perfectly identical, the optimal number of analogues of the first approach would be m times higher than the selection from a single member, m being the number of members considered. On the contrary, the number of analogues would not change for the second approach. Both approaches were assessed here for the 2Z (Fig. 11) and 2Z-2MI (Fig. 12) methods, due to the availability of the variables in 20CR-2c's ensemble dataset. As the spread is lower for recent periods than in the past (Compo et al, 2011), two periods were assessed: the original 1981–2010 period with its VP (Sect. 2.2) and an earlier period 1901–1930 (with the following validation years: 1905, 1910, 1915, 1920, 1925, 1930). There might be other benefits in using members, such as a better consideration of the uncertainty when working on the distant past. However, their impact was only assessed here in terms of performance.

The introduction of members slightly improved the performance of the 2Z method, but typically only when keeping duplicate dates (Fig. 11 a and b). Indeed, the exclusion of duplicate dates led to minor or no improvement. The likely reason is that the recurring analogues are probably the best ones, and allowing duplicates gives them more weight, otherwise their importance decreases within a growing selection of analogues. Unsurprisingly, the benefit of using members was also higher for the early period 1901–1930 (Fig. 11 right), where larger uncertainties are present. In most situations, the additional gain in performance brought by new members flattened out relatively rapidly. Indeed, when using 20CR-2c, the increase in skill after 5 members was marginal, which was also the case with CERA-20C in the more recent (1981–2010) period. Using all 56 members of 20CR-2c was very costly in terms of processing time and provided no improvement to the performance.

The results of the 2Z-2MI method (Fig. 12) led to the same conclusions in terms of higher gains when allowing duplicates and also for the earlier (1901–1930) period. However, a major difference was that after having reached an optimal number of members (4–5), the performance did not flatten out, but decreased below the score based on a single member. This behavior was investigated and a peculiar characteristic of the number of analogues was found. The number of analogues was optimized for each level of analogy when adding new members, by assessing multiple combinations, so that they were optimal for the provided predictors. Here, the optimal number of analogues tended to be equal for both levels after addition of some members, which means that the subsampling of the second level of analogy (on moisture) was discarded. This behavior did not happen when real data from the past was added (Sect. 4.2). The uncertainty between the members is not of the same magnitude for the different variables. A likely hypothesis is that because moisture variables are more uncertain, their related number of analogues grew faster than for Z, but were limited by the selection of the first level of analogy. Great caution is therefore advised when using AMs with multiple analogy levels on ensemble reanalyses.

5 Discussion and conclusion

Some constraints might drive the choice of a certain reanalysis over another, for example when working on earlier periods. However, when the period of interest falls within the satellite era, one has to choose one reanalysis from among all the existing reanalyses. The choice is often motivated by either ease of access (availability of the dataset at the institution), ease of use (availability of code to read it), or by the preference for the local provider (such as ECMWF for Europe). This choice has a non-negligible impact, which was quantified in this work.

Although compared in a recent period over a data-rich region, the tested reanalyses resulted in large differences in terms of performance of the AMs. The impact of the reanalyses was sometimes found to be even larger than the choice of the method and its related predictors, in accordance with Dayon et al (2015). An analysis of variance emphasized highly significant impact of the reanalysis on the skill, with a contribution to variance (of the skill score) above 20%. There was no single overall winner, but different alternatives that provided similar performances.

The impact on the skill of AMs is not a direct assessment of the quality of the reanalysis, but it characterizes an indirect impact on the quality of the relationship between predictors and the precipitation, which makes it complex to interpret. However, given the results obtained, it seems manifest that there is indeed a link between the quality of a reanalysis and its impact on the skill of the AMs.

Figure 13 synthesizes the suggested choice of reanalyses for different periods and variables, providing the preferred reanalyses and their alternatives. These suggestions are specific for the use of AMs optimized, in terms of CRPSS, for daily precipitation in Switzerland or possibly similar contexts. The temporal homogeneity of the reanalyses was not fully assessed here, and users should consider this aspect depending on the application. The different reanalyses are discussed hereafter.

NR-1 and NR-2 were the first reanalyses available and were used until recently. Despite their age, and the progress made in terms of data assimilation and numerical modelling since their introduction, they still provide valuable outputs. However, they systematically performed slightly below average, and are thus of less interest than other options. Even though NR-1 starts in 1948, which is prior to many reanalyses, there are better alternatives, and we do not recommend using it exclusively any more.

ERA-INT is often the default choice in Europe nowadays for various applications. It was found to be amongst the best performing reanalyses, particularly for moisture variables, but it might not be the best choice for SLP.

The new NCEP reanalysis, CFSR, systematically surpassed its predecessors NR-1 and NR-2. It was in the top selection except for the vertical velocity (W), where it did not perform as well as other options.

The two Japanese reanalyses, JRA-55 and JRA-55C, are less well-known, but they result in remarkably good performances overall and are systematically a first choice or alternative selection (Fig. 13). A striking element is the similar performance of both reanalyses, despite the fact that JRA-55C only assimilates conventional observations. It is probably due to the good coverage of upper-air observations in Europe (C. Kobayashi, pers. comm., November 29, 2017). JRA-55C is the recommended reanalysis when the working period starts prior to the satellite era (from 1958 onward), as it is expected to be more homogeneous than JRA-55 due to its use of conventional-only data.

MERRA-2 showed good overall performance for all methods, both at a daily time step and for annual correlations. It showed a particularly striking performance with SLP, which was as skillful as using four levels of the geopotential height. MERRA-2 differs from other reanalyses in that it includes changes in atmospheric mass due to evaporation and precipitation in order to conserve atmospheric dry mass (Gelaro et al, 2017). This characteristic is likely to impact areas with strong precipitation events and may be related to the observed difference in skill (M. Suarez, pers. comm., January 25, 2018).

20CR-2c is the only reanalysis so far that provides data for the second half of the 19^{th} century, which makes it a valuable asset. However, it is not the best estimate for more recent periods (Poli and National Center for Atmospheric Research Staff, 2017), and its performance for daily precipitation was systematically and substantially inferior to that of other reanalyses. Although it sometimes showed inter-annual correlations at the same level as other reanalyses, its overall lower performance at a daily time step disqualifies it as an option for periods other than the distant past. Its lower performance in the AM was also raised by Dayon et al (2015), particularly when local predictors are included. It can be at least partly explained by the fact that 20CR-2c assimilates less data compared than other reanalyses. Additionally, 20CR-2c exhibits fewer westerlies and more easterlies over Western Europe than other reanalyses (Rohrer et al, 2018). Nevertheless, it is noteworthy to mention all the informative outputs generated over such a long period on the basis of so few assimilated data.

ERA-20C assimilates marine wind observations in addition to the data included in 20CR-2c, and the model is also forced by more data for its boundary conditions. This, along with a different model and assimilation technique, resulted in higher skills than 20CR-2c within the AM. However, ERA-20C did not compete at a daily time step for more recent periods with other reanalyses that assimilate more observations. CERA-20C has an additional coupling to the ocean and is processed with a more recent version of the IFS forecast model. This resulted in relatively equivalent skills at a daily time step, but higher inter-annual correlations; thus CERA-20C should be chosen over ERA-20C.

Switzerland is a small country, but with high contrasts in terms of climate, with regions sensitive to different meteorological situations, as well as a wide range of elevations. The choice of the best method or the best dataset was found to depend on the context of the station, with spatial patterns emphasizing the different climatic regions. The choice of the variables also had a strong impact on the selection of the best reanalysis, with MERRA-2 being the best for SLP; CFSR and JRA-55(C) along with MERRA-2 were often selected for Z; ERA-INT was more often selected when moisture variables were considered; JRA-55(C), ERA-INT and MERRA-2 were most often chosen for the most complex method with W.

The biases seemed to depend on both the method and the reanalysis. 2SLP induced a dry bias for most reanalyses, as well as PT-2Z-4W-4MI, while PT-2Z-4MI resulted in a wet bias for most reanalyses. The bias related to 2Z-2MI and 4Z-2MI was generally more contained within a relative 5% range for most of the reanalyses. NR-1, NR-2 and 20CR-2c generally resulted in wetter predictions, and MERRA-2 in dryer ones. The bias can be crucial depending on the use of the downscaled precipitation, and should then be considered in the choice of the method and the reanalysis. It can also be corrected in a postprocessing stage.

The percentage of similar analogue days between reanalysis decreased with the complexity of the method. Similar reanalyses showed a higher percentage of shared analogue dates. This percentage increased slightly for days with high precipitation. This is likely due to more defined circulation patterns associated with e.g. cyclonic circulations, and to the fact that these situations are less frequent, which increases the probability to select the same analogue dates. However, as the numbers do not drastically differ, most of the difference in the selection of analogue dates in the all-days analysis is not only related to situations with a less defined atmospheric circulation, such as an anticyclonic condition protecting Europe from disturbances. On the other hand, some of these similar analogue dates are driven by similarities between products, in terms of input data or concepts of numerical modelling, rather than being "perfect analogue" situations. There could also be differences between stations or seasons, which were not investigated. All analogue dates were published as datasets (see Sect. Data availability) in order to allow the community for further analyses.

The differences in skill between reanalyses did not depend so much on the assimilation technique (at least between 3D-Var and 4D-Var), but rather on the assimilated data and on the forecast model. Although higher spatial resolutions in the forecast models are likely to result in better reanalyses, higher output resolutions were not found to contribute to the differences in skill between reanalyses (Sect. 4.1).

Longer archives are commonly considered to improve the analogy by providing more candidate analogues. However, as shown in Sect. 4.2, it is not always the case when adding years from a more distant past as one should consider the temporal homogeneity of the archive and the reliability of the variables considered in earlier years. First guess errors or ensemble spreads from a given reanalysis might be used to influence the choice of an acceptable archive period. As expected, the geopotential height showed a greater robustness over time than moisture variables.

Some reanalyses provide multiple members, which is an added value for many applications. However, no substantial improvement of the skill was found when using ensemble reanalyses in the AM, at least for recent periods. Moreover, using multiple members in AMs with multiple levels of analogy might even reduce the performance of the method, possibly due to mismatches between the uncertainties of the variables under consideration. Thus, we recommend not using ensembles in the AM for present periods and to use them with great caution for past periods. When using AMs in operational forecasting, the use of forecast ensembles to characterize the target date is, however, valuable, due to greater uncertainties being related to the unknown evolution of the meteorological situation (Thevenot, 2004).

Hopefully, the present work can help drive a decision about the future use of reanalyses in AMs. The assessment focused on Switzerland only, but it can be expected that the results will be transferable to other data-rich regions, at least in Western Europe. Indeed, Switzerland has a rich climate with multiple meteorological influences, and the trends of the influence of the reanalyses were consistent from one climatic region to another even though one dataset might be just superior to others for specific regions. Moreover, the spatial quality of a reanalysis is closely related to the number of assimilated observations, which are relatively dense over Western Europe. For use of AMs in a different context, for example in a data-poor region of the SH, similar comparative work can be undertaken. The present work can still, however, help reduce the number of reanalyses considered.

When looking for analogues in a reanalysis to target situations described by NWP or climate model outputs, certain precautions must be taken to account for different model climates and biases (Scaife et al, 2010; Cattiaux et al, 2013). Additionally, it would be preferable to use several reanalyses as an ensemble rather than a single product. The most recent products of different institutions should be considered by default for this kind of approach.

The choice of some predictors common to most AMs from the literature was based on the first reanalysis dataset, NR-1, and new methods are often built on these foundations by adding complexity. However, the new reanalyses provide new or improved variables. Assessing systematically most variables from different products, and combination of these variables, would be cumbersome. In the continuity of this work, an automatic selection of variables from different reanalyses will be explored by means of genetic algorithms in order to extract potential new variables of interest or a combination of these.

Data availability

All calculations were performed with the open source AtmoSwing software v1.5.0 (Horton, 2017). The resulting files were processed using AtmoSwing R-toolbox v1.2.0 (Horton, 2018k).

The resulting analogue dates for every combination of station, dataset, and analogue method were published. Along with these, different files are also available: the parameter files used in AtmoSwing for the calibration, the resulting calibrated parameters, and files listing all assessed parameter sets. The datasets are available for each reanalysis: NR-1 (Horton, 2018i), NR-2 (Horton, 2018j), ERA-INT (Horton, 2018e), CFSR (Horton, 2018c), JRA-55 (Horton, 2018f), JRA-55C (Horton, 2018g), MERRA-2 (Horton, 2018h), 20CR-2c (Horton, 2018a), ERA-20C (Horton, 2018d), and CERA-20C (Horton, 2018b). Additional data can be obtained by contacting the authors.

Acknowledgements The authors want to thank M. Rohrer and P. Laloyaux for their valuable inputs and C. Obled for correcting the manuscript. Special thanks are due to M. Suarez, L. Takacs, R. Gelaro, and M. G. Bosilovich, from the Global Modeling and Assimilation Office, NASA, for their help in investigating the differences in SLP between MERRA-2 and the other reanalyses. Stefan Brönnimann acknowledges funding from the FP7 project ERA-CLIM2.

Precipitation time series were provided by MeteoSwiss. The NCEP/NCAR, NCEP/DOE, and 20CR-2c were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at http://www.esrl.noaa.gov/psd/. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. The CFSR, and JRA-55 were obtained from the CISL Research Data Archive (http://rda.ucar.edu/) at NCAR in Boulder, Colorado, and the NCAR is supported by grants from the National Science Foundation. The Climate Forecast System Reanalysis (CFSR) project is carried out by the Environmental Modeling Center (EMC), National Centers for Environmental Prediction (NCEP). The Japanese 55-year Reanalysis (JRA-55) project is carried out by the Japan Meteorological Agency (JMA). The MERRA-2 was obtained from the Goddard Earth Sciences Data and Information Services Center, Greenbelt, Maryland, from their website at http://disc.sci.gsfc.nasa.gov/mdisc. The ERA-interim, ERA-20C, and CERA-20C were obtained from the ECMWF Data Server at http://apps.ecmwf.int/datasets/.

Calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern.

The authors are grateful to the editor Susanna Corti and two anonymous reviewers for their invaluable and constructive comments that helped improve the manuscript.

References

Altava-Ortiz V, Barrera A, Llasat MC, Prat MA, Gibergans-Báguena J, Barnolas M (2006) Application of the MM5 and the analogous method to heavy rainfall event, the case of 1618 October 2003 in Catalonia (NE Spain). Adv Geosci 7:313–319, DOI 10.5194/adgeo-7-313-2006

Barrera A, Altava-Ortiz V, Llasat MC, Barnolas M (2007) Heavy rain prediction using deterministic and probabilistic models the flash flood cases of 11 13 October 2005 in Catalonia (NE Spain). Adv Geosci 12:121–126, DOI 10.5194/adgeo-12-121-2007

Ben Daoud A (2010) Améliorations et développements d'une méthode de prévision probabiliste des pluies par analogie. PhD thesis, Université de Grenoble

Ben Daoud A, Sauquet E, Lang M, Obled C, Bontron G (2009) La prévision des précipitations par recherche d'analogues : état de l'art et perspectives. La Houille Blanche (6):60–65, DOI 10.1051/lhb/2009079

Ben Daoud A, Sauquet E, Lang M, Bontron G, Obled C (2011) Precipitation forecasting through an analog sorting technique: A comparative study. Adv Geosci 29(1997):103–107, DOI 10.5194/adgeo-29-103-2011

Ben Daoud A, Sauquet E, Bontron G, Obled C, Lang M (2016) Daily quantitative precipitation forecasts based on the analogue method: improvements and application to a French large river basin. Atmos Res 169:147–159, DOI 10.1016/j.atmosres.2015.09.015

Bliefernicht J, Bárdossy A (2007) Probabilistic forecast of daily areal precipitation focusing on extreme events. Nat Hazards Earth Syst Sci 7(2):263–269, DOI 10.5194/nhess-7-263-2007

Bonnet R, Boé J, Dayon G, Martin E (2017) 20th-Century Hydro-Meteorological Reconstructions to Study the Multidecadal Variations of the Water Cycle Over France. Water Resour Res 53:1–17, DOI 10.1002/2017WR020596

Bontron G (2004) Prévision quantitative des précipitations: Adaptation probabiliste par recherche d'analogues. Utilisation des Réanalyses NCEP/NCAR et application aux précipitations du Sud-Est de la France. PhD thesis, Institut National Polytechnique de Grenoble

Bradley AA, Schwartz SS (2011) Summary Verification Measures and Their Interpretation for Ensemble Forecasts. Mon Wea Rev 139(9):3075–3089, DOI 10.1175/2010MWR3305.1

Brigode P, Brissette F, Nicault A, Perreault L, Kuentz A, Mathevet T, Gailhard J (2016) Streamflow variability over the 18812011 period in northern Québec: comparison of hydrological reconstructions based on tree rings and geopotential height field reanalysis. Climate Past 12:1785–1804, DOI 10.5194/cp-2016-5

Brown BG, Gilleland E, Ebert EE (2012) Forecasts of Spatial Fields. In: Jolliffe IT, Stephenson DB (eds) Forecast Verif., Wiley, chap 6, pp 95–117, DOI 10.1002/9781119960003.ch6, URL http://doi.wiley.com/10.1002/9781119960003.ch6

Brown T (1974) Admissible Scoring Systems for Continuous Distributions. Tech. rep., URL http://eric.ed.gov/?id=ED135799

Caillouet L, Vidal JP, Sauquet E, Graff B (2016) Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France. Climate Past 12(3):635–662, DOI 10.5194/cp-12-635-2016

Caillouet L, Vidal JP, Sauquet E, Devers A, Graff B (2017) Ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871. Hydrol Earth Syst Sci 21(6):2923–2951, DOI 10.5194/hess-21-2923-2017

Cannon AJ (2007) Nonlinear analog predictor analysis: A coupled neural network/analog model for climate downscaling. Neural Networks 20(4):444–453, DOI 10.1016/j.neunet.2007.04.002

Cattiaux J, Douville H, Peings Y (2013) European temperatures in CMIP5: Origins of present-day biases and future uncertainties. Clim Dyn 41(11-12):2889–2907, DOI 10.1007/s00382-013-1731-y

Chardon J, Hingray B, Favre AC, Autin P, Gailhard J, Zin I, Obled C (2014) Spatial Similarity and Transferability of Analog Dates for Precipitation Downscaling over France. J Climate 27(13):5056–5074, DOI 10.1175/JCLI-D-13-00464.1

Clark M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R (2004) The Schaake Shuffle: A Method for Reconstructing SpaceTime Variability in Forecasted Precipitation and Temperature Fields. J Hydrometeor 5(1):243–262, DOI 10.1175/1525-7541(2004)005¡0243:TSSAMF¿2.0.CO;2, arXiv:1011.1669v3

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant aN, Groisman PY, Jones PD, Kruk MC, Kruger aC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The Twentieth Century Reanalysis Project. Quart J Roy Meteor Soc 137(654):1–28, DOI 10.1002/qj.776, arXiv:1011.1669v3

Dayon G, Boé J, Martin E (2015) Transferability in the future climate of a statistical downscaling method for precipitation in France. J Geophys Res Atmos 120(3):1023–1043, DOI 10.1002/2014JD022236

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kallberg P, Köhler M, Matricardi M, Mcnally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart J Roy Meteor Soc 137(656):553–597, DOI 10.1002/qj.828

Duband D (1981) Prévision spatiale des hauteurs de précipitations journalières. La Houille Blanche 7-8:497–512

Franke J, Gonzalez-Rouco JF, Frank D, Graham NE (2011) 200 years of European temperature variability: Insights from and tests of the proxy surrogate reconstruction analog method. Climate Dyn 37(1):133–150, DOI 10.1007/s00382-010-0802-6

Fujiwara M, Wright JS, Manney GL, Gray LJ, Anstey J, Birner T, Davis S, Gerber EP, Lynn Harvey V, Hegglin MI, Homeyer CR, Knox JA, Krüger K, Lambert A, Long CS, Martineau P, Molod A, Monge-Sanz BM, Santee ML, Tegtmeier S, Chabrillat S, Tan DG, Jackson DR, Polavarapu S, Compo GP, Dragani R, Ebisuzaki W, Harada Y, Kobayashi C, McCarty W, Onogi K, Pawson S, Simmons A, Wargan K, Whitaker JS, Zou CZ (2017) Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. Atmos Chem Phys 17(2):1417–1452, DOI 10.5194/acp-17-1417-2017

Gangopadhyay S, Clark M, Rajagopalan B (2005) Statistical downscaling using K -nearest neighbors. Water Resour Res 41(2):W02,024, DOI 10.1029/2004WR003444

Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, Randles CA, Darmenov A, Bosilovich MG, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva AM, Gu W, Kim GK, Koster R, Lucchesi R, Merkova D, Nielsen JE, Partyka G, Pawson S, Putman W, Rienecker M, Schubert SD, Sienkiewicz M, Zhao B (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Climate 30(14):5419–5454, DOI 10.1175/JCLI-D-16-0758.1

Gibson JK, Kallberg P, Uppala S, Hernandez A, Nomura A, Serrano E (1997) ERA description. ECMWF Re-Analysis Project Report Series 1. Tech. rep., ECMWF, Reading, UK

Harada Y, Kamahori H, Kobayashi C, Endo H, Kobayashi S, Ota Y (2016) The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability. J Meteor Soc Japan 94(3):269–302, DOI 10.2151/jmsj. 2016-015 J-STAGE

Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. Wea Forecasting 15(5):559–570, DOI 10.1175/1520-0434(2000)015¡0559:dotcrp¿2.0.co;2

Horton P (2017) AtmoSwing v1.5.0. DOI 10.5281/zenodo.821559

Horton P (2018a) Analog dates for 20CR-v2c at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/8yw7r57sck.1, Mendeley Data, v1

Horton P (2018b) Analog dates for CERA-20C at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/vr6chcpfsf.2, Mendeley Data, v2

Horton P (2018c) Analog dates for CFSR at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/w7k5g94khj.2, Mendeley Data, v2

Horton P (2018d) Analog dates for ERA-20C at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/9ny7jp9gxg.2, Mendeley Data, v2

Horton P (2018e) Analog dates for ERA-interim at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/b3xfgk42h8.2, Mendeley Data, v2

Horton P (2018f) Analog dates for JRA-55 at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/zxyb7p825m.2, Mendeley Data, v2

Horton P (2018g) Analog dates for JRA-55C at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/sw8gr593rr.2, Mendeley Data, v2

Horton P (2018h) Analog dates for MERRA-2 at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/5sbnwcsh6p.2, Mendeley Data, v2

Horton P (2018i) Analog dates for NCEP-R1 at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/5rcwmfkg3v.1, Mendeley Data, v1

Horton P (2018j) Analog dates for NCEP-R2 at 301 precipitation stations in Switzerland (1981-2010). DOI 10.17632/vmdpg7ckpf.1, Mendeley Data, v1

Horton P (2018k) AtmoSwing R-toolbox v1.2.0. DOI 10.5281/zenodo.1305098

Horton P, Jaboyedoff M, Metzger R, Obled C, Marty R (2012) 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, DOI 10.5194/nhess-12-777-2012

Horton P, Jaboyedoff M, Obled C (2017a) Using genetic algorithms to optimize the analogue method for precipitation prediction in the Swiss Alps. J Hydrol in press, DOI 10.1016/j.jhydrol.2017.04.017

Horton P, Obled C, Jaboyedoff M (2017b) The analogue method for precipitation prediction: finding better analogue situations at a sub-daily time step. Hydrol Earth Syst Sci 21:3307–3323, DOI 10.5194/hess-21-3307-2017

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77(3):437–471, DOI 10.1175/1520-0477(1996)077¡0437:TNYRP¿2.0.CO;2, arXiv:1011.1669v3

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL (2002) NCEPDOE AMIP-II Reanalysis (R-2). Bull Amer Meteor Soc 83(11):1631–1643, DOI 10.1175/BAMS-83-11-1631

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, Van Den Dool H, Jenne R, Fiorino M (2001) The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bull Amer Meteor Soc 82(2):247–267, DOI 10.1175/1520-0477(2001)082¡0247:TNNYRM¿2.3.CO;2

Kobayashi C, Endo H, Ota Y, Kobayashi S, Onoda H, Harada Y, Onogi K, Kamahori H (2014) Preliminary Results of the JRA-55C, an Atmospheric Reanalysis Assimilating Conventional Observations Only. Sola 10(0):78–82, DOI 10.2151/sola.2014-016

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K (2015) The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J Meteor Soc Japan Ser II 93(1):5–48, DOI 10.2151/jmsj.2015-001

Koukidis E, Berg A (2009) Sensitivity of the Statistical DownScaling Model (SDSM) to reanalysis products. Atmos-Ocean 47(1):1–18, DOI 10.3137/AO924.2009

Kruizinga S, Murphy A (1983) Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. Mon Wea Rev 111(11):2244–2254, DOI 10.1175/1520-0493(1983)111¡2244:uoaapt¿2.0.co;2

Kuentz A, Mathevet T, Gailhard J, Hingray B (2015) Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: The ANATEM model. Hydrol Earth Syst Sci 19(6):2717–2736, DOI 10.5194/hess-19-2717-2015

Laloyaux P, Balmaseda M, Dee D, Mogensen K, Janssen P (2016) A coupled data assimilation system for climate reanalysis. Quart J Roy Meteor Soc 142(694):65–78, DOI 10.1002/qj.2629

Laloyaux P, de Boisseson E, Balmaseda M, Bidlot JR, Broennimann S, Buizza R, Dalhgren P, Dee D, Haimberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D (2018) CERA-20C: A Coupled Reanalysis of the Twentieth Century. J Adv Model Earth Syst pp 1172–1195, DOI 10.1029/2018MS001273

Lorenz E (1969) Atmospheric predictability as revealed by naturally occurring analogues. J Atmos Sci 26:636–646, DOI 10.1175/1520-0469(1969)26¡636:aparbn¿2.0.co;2

Maraun D, Wetterhall F, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user. Rev Geophys 48(RG3003):1–34, DOI 10.1029/2009RG000314

Martín ML, Valero F, Pascual A, Sanz J, Frias L (2014) Analysis of wind power productions by means of an analog model. Atmos Res 143:238–249, DOI 10.1016/j.atmosres.2014.02.012

Marty R, Zin I, Obled C, Bontron G, Djerboua A (2012) Toward real-time daily PQPF by an analog sorting approach: Application to flash-flood catchments.

J Appl Meteor Climatol 51(3):505–520, DOI 10.1175/JAMC-D-11-011.1
Matheson J, Winkler R (1976) Scoring rules for continuous probability distributions. Manage Sci 22(10):1087–1096, DOI 10.1287/mnsc.22.10.1087
Matulla C, Zhang X, Wang XL, Wang J, Zorita E, Wagner S, von Storch H (2007) Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. Climate Dyn 30(2-3):133–144, DOI 10.1007/s00382-007-0277-2
Maurer EP, Hidalgo HG (2008) Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. Hydrol Earth Syst Sci 12(2):551–563, DOI 10.5194/hess-12-551-2008
Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R (2007) The JRA-25 Reanalysis. J Meteor Soc Japan 85(3):369–432, DOI 10.2151/jmsj.85.369
Osca J, Romero R, Alonso S (2013) Precipitation projections for Spain by means of a weather typing statistical method. Glob Planet Change 109:46–63, DOI 10.1016/j.gloplacha.2013.08.001
Pascual A, Valero F, Martín ML, Morata A, Luna MY (2012) Probabilistic and deterministic results of the ANPAF analog model for Spanish wind field estimations. Atmos Res 108:39–56, DOI 10.1016/j.atmosres.2012.01.011
Poli P, National Center for Atmospheric Research Staff (2017) The Climate Data Guide: ERA-20C: ECMWF's atmospheric reanalysis of the 20th century (and comparisons with NOAA's 20CR). URL https://climatedataguide.ucar.edu/climate-data/era-20c-ecmwfs-atmospheric-reanalysis-20th-century-and-comparisons-noaas-20cr
Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut JN, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. J Climate 29(11):4083–4097, DOI 10.1175/JCLI-D-15-0556.1
Radanovics S, Vidal JP, Sauquet E, Ben Daoud A, Bontron G (2013) Optimising predictor domains for spatially coherent precipitation downscaling. Hydrol Earth Syst Sci 17(10):4189–4208, DOI 10.5194/hess-17-4189-2013
Raynaud D, Hingray B, Zin I, Anquetin S, Debionne S, Vautard R (2016) Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb. Int J Climatol DOI 10.1002/joc.4844
Ribalaygua J, Torres L, Pórtoles J, Monjo R, Gaitán E, Pino MR (2013) Description and validation of a two-step analogue/regression downscaling method. Theor Appl Climatol 114(1-2):253–269, DOI 10.1007/s00704-013-0836-x
Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim GK, Bloom S, Chen J, Collins D, Conaty A, Da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. J Climate 24(14):3624–

3648, DOI 10.1175/JCLI-D-11-00015.1

Rohrer M, Brönnimann S, Martius O, Raible CC, Wild M, Compo GP (2018) Representation of extratropical cyclones, blocking anticyclones, and Alpine circulation types in multiple reanalyses and model simulations. J Clim 31(8):3009–3031, DOI 10.1175/JCLI-D-17-0350.1

Rummukainen M (1997) Methods for statistical downscaling of GCM simulations. Tech. Rep. 80, URL http://agris.fao.org/agris-search/search/display.do?f=2012/OV/OV201205379005379.xml;SE19970167745

Ruosteenoja K (1988) Factors affecting the occurrence and lifetime of 500 mb height analogues: a study based on a large amount of data. Mon Wea Rev 116:368–376

Saha S, Moorthi S, Pan HL, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou YT, Chuang HY, Juang HMH, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Van Den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm JK, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou CZ, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M (2010) The NCEP climate forecast system reanalysis. Bull Amer Meteor Soc 91(8):1015–1057, DOI 10.1175/2010BAMS3001.1

Scaife AA, Woollings T, Knight J, Martin G, Hinton T (2010) Atmospheric blocking and mean biases in climate models. J Clim 23(23):6143–6152, DOI 10.1175/2010JCLI3728.1

Schenk F, Zorita E (2012) Reconstruction of high resolution atmospheric fields for Northern Europe using analog-upscaling. Climate Past 8(5):1681–1703, DOI 10.5194/cp-8-1681-2012

Teng J, Chiew FHS, Timbal B, Wang Y, Vaze J, Wang B (2012) Assessment of an analogue downscaling method for modelling climate change impacts on runoff. J Hydrol 472-473:111–125, DOI 10.1016/j.jhydrol.2012.09.024

Teweles S, Wobus HB (1954) Verification of prognostic charts. Bull Amer Meteor Soc 35:455–463

Themessl MJ, Gobiet A, Leuprecht A (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int J Climatol 31(10):1530–1544, DOI 10.1002/joc.2168

Thevenot N (2004) Prévision quantitative des précipitations: Adaptation par une méthode d'Analogie de la prévision d'ensemble du CEPMMT. Aspects opérationnels. PhD thesis, INP Grenoble

Thiébaux HJ (1985) On approximations to geopotential and windfield correlation structures. Tellus A 37 A(2):126–131, DOI 10.1111/j.1600-0870.1985.tb00275.x

Timbal B, McAvaney BJ (2001) An Analogue based method to downscale surface air temperature: Application for Australia. Climate Dyn 17:947–963, DOI 10.1007/s003820100156

Timbal B, Dufour A, Mcavaney B (2003) An estimate of future climate change for western France using a statistical downscaling technique. Climate Dyn 20:807–823, DOI 10.1007/s00382-002-0298-9

Timbal B, Li Z, Fernandez E (2008) The Bureau of Meteorology Statistical Downscaling Model Graphical User Interface: user manual and software documentation. Tech. Rep. 004

Turco M, Quintana-Seguí P, Llasat MC, Herrera S, Gutiérrez JM (2011) Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain. J Geophys Res Atmos 116(18):1–14, DOI 10.1029/2011JD016166

Uppala SM, Kallberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, Mcnally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. Quart J Roy Meteor Soc 131(612):2961–3012, DOI 10.1256/qj.04.176

Van den Dool H (1989) A new look at weather forecasting through analogues. Mon Wea Rev 117:2230–2247

Van Den Dool HM (1994) Searching for analogues, how long must we wait? Tellus 46A(3):314–324

Vanvyve E, Delle Monache L, Monaghan AJ, Pinto JO (2015) Wind resource estimates with an analog ensemble approach. Renew Energy 74:761–773, DOI 10.1016/j.renene.2014.08.060

Wetterhall F, Halldin S, Xu Cy (2005) Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 306(1-4):174–190, DOI 10.1016/j.jhydrol.2004.09.008

Willems P, Vrac M (2011) Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. J Hydrol 402(3-4):193–205, DOI 10.1016/j.jhydrol.2011.02.030

Wu W, Liu Y, Ge M, Rostkier-Edelstein D, Descombes G, Kunin P, Warner T, Swerdlin S, Givati A, Hopson T, Yates D (2012) Statistical downscaling of climate forecast system seasonal predictions for the Southeastern Mediterranean. Atmos Res 118:346–356, DOI 10.1016/j.atmosres.2012.07.019

Yiou P, Boichu M, Vautard R, Vrac M, Jourdain S, Garnier E, Fluteau F, Menut L (2014) Ensemble meteorological reconstruction using circulation analogues of 1781-1785. Climate Past 10(2):797–809, DOI 10.5194/cp-10-797-2014

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Climate 12(8):2474–2489, DOI 10.1175/1520-0442(1999)012¡2474:TAMAAS¿2.0.CO;2

Table 1 Assessed reanalysis datasets with their respective properties, sorted by type and model age.

Name	Institution	Period of record	Output resolution	Model resolution	Model vintage	Type of input	Assimilation technique
NR-1	NCEP, NCAR	1948 – present	2.5°x 2.5°	T62 (∼1.88°), L28	1995	full	3D-Var
NR-2	NCEP, DOE	1979 – present	2.5°x 2.5°	T62 (∼1.88°), L28	2001	full	3D-Var
ERA-INT	ECMWF	1979 – present	0.75°x 0.75°	TL255 (∼0.70°), L60	2006	full	4D-Var
CFSR	NCEP	1979 – present	0.5°x 0.5°	T382 (∼0.31°), L64	2009	full	3D-Var
JRA-55	JMA	1958 – present	1.25°x 1.25°	TL319 (∼0.36°), L60	2009	full	4D-Var
JRA-55C	JMA	1958 – 2015	1.25°x 1.25°	TL319 (∼0.36°), L60	2009	conventional	4D-Var
MERRA-2	NASA GMAO	1980 – present	0.625°x 0.5°	0.625°x 0.5°, L72	2014	full	3D-Var
20CR-2c	NOAA-CIRES	1851 – 2014	2°x 2°	T62 (∼1.88°), L28	2008	surface	EnKF
ERA-20C	ECMWF	1900 – 2010	1°x 1°	TL159 (∼1.13°), L91	2012	surface	4D-Var
CERA-20C	ECMWF	1901 – 2010	1°x 1°	T159 (∼1.13°), L91	2016	surface	4D-Var

Table 2 Analogue methods considered in the study, listed by increasing complexity. P0 is the preselection (PC: on calendar basis, that is ±60 days around the target date), L1, L2 and L3 are the subsequent levels of analogy. The meteorological variables are: SLP – mean sea level pressure, Z – geopotential height, T – air temperature, W – vertical velocity, MI – moisture index, which is the product of the relative humidity at the given pressure level and the total water column. The analogy criterion is S1 for SLP and Z and RMSE for the other variables.

Method	P0	L1	L2	L3	Reference
2SLP	PC	SLP@12h SLP@24h			
2Z	PC	Z1000@12h Z500@24h			Bontron 2004
4Z	PC	Z1000@06h Z1000@30h Z700@24h Z500@12h			Horton et al 2017a
2Z-2MI	PC	Z1000@12h Z500@24h	MI850@12+24h		Bontron 2004
4Z-2MI	PC	Z1000@30h Z850@12h Z700@24h Z400@12h	MI700@24h MI600@12h		Horton et al 2017a
PT-2Z-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	MI925@12+24h MI700@12+24h		Ben Daoud et al 2016
PT-2Z-4W-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	W850@06-24h	MI925@12+24h MI700@12+24h	Ben Daoud et al 2016

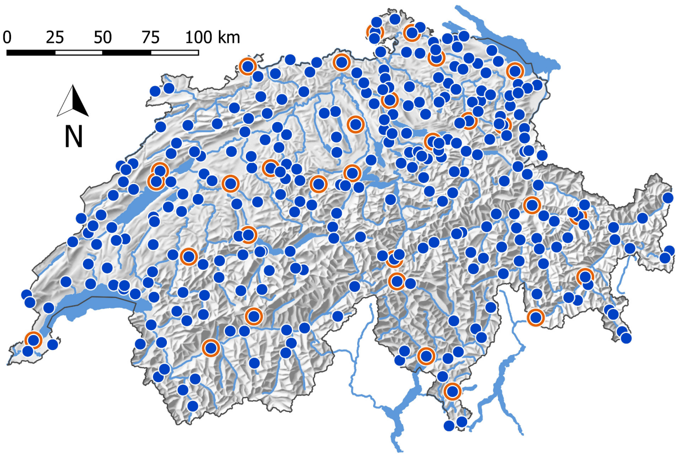

Fig. 1 Map of the 301 precipitation stations with good data coverage of the period 1981–2010 (blue dots), and the 30 stations with long archives (orange). Background map: © SwissTopo.

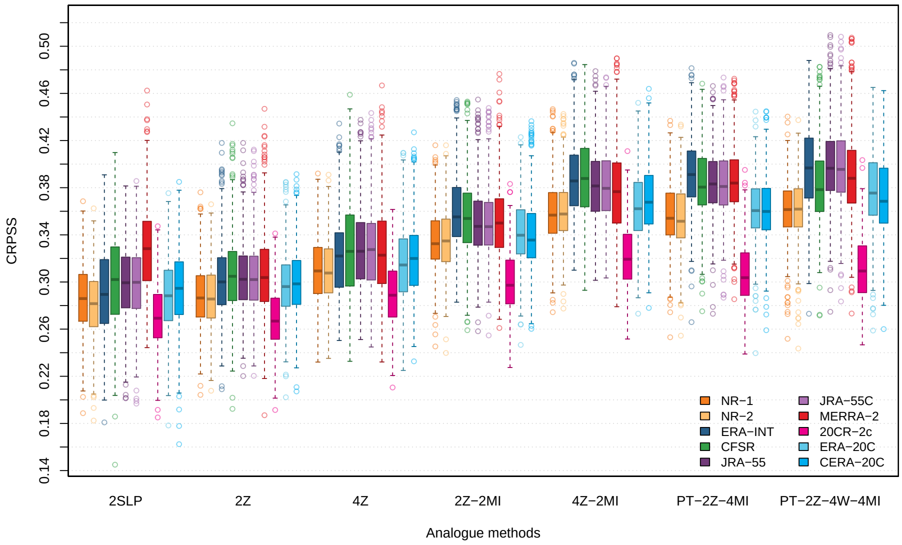

Fig. 2 CRPSS for all stations, and for all considered AMs and reanalysis datasets on the VP. A higher CRPSS means better performance. The parameters of the AMs were calibrated for every station, every dataset, and every method. The boxes show the 25th, 50th, and 75th percentiles. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range.

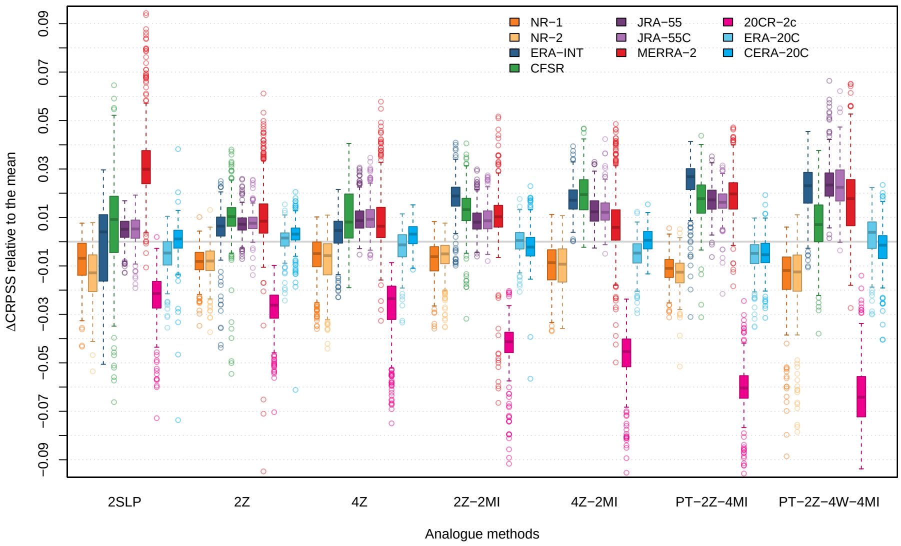

Fig. 3 Impact of the reanalysis dataset on performance, isolated by processing the improvement in CRPSS for one dataset compared to the mean performance on all datasets, per station and per method. Note that the methods cannot be compared here, only the datasets. Same conventions as Fig. 2.

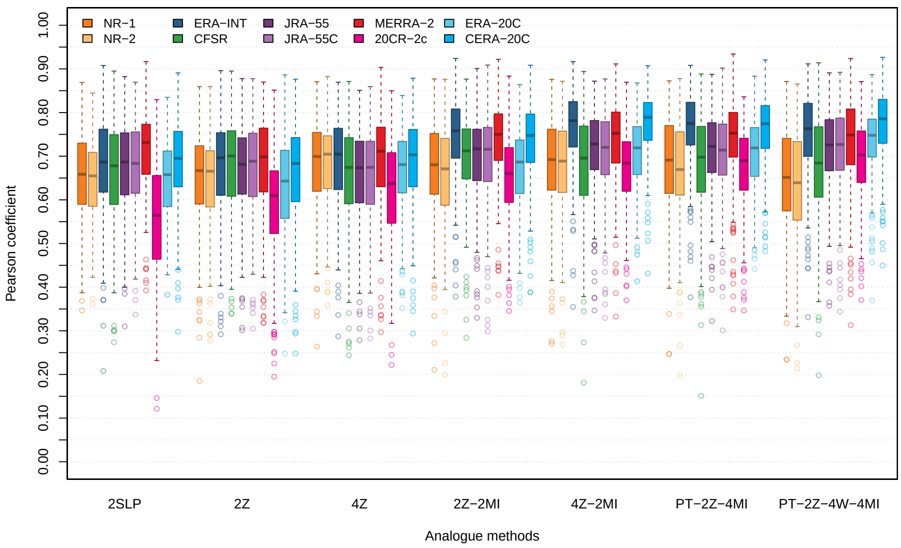

Fig. 4 Inter-annual correlation between the mean precipitation from the selected analogues and the observations for all stations and for all considered AMs and reanalysis datasets on both the CP and the VP. Same conventions as Fig. 2.

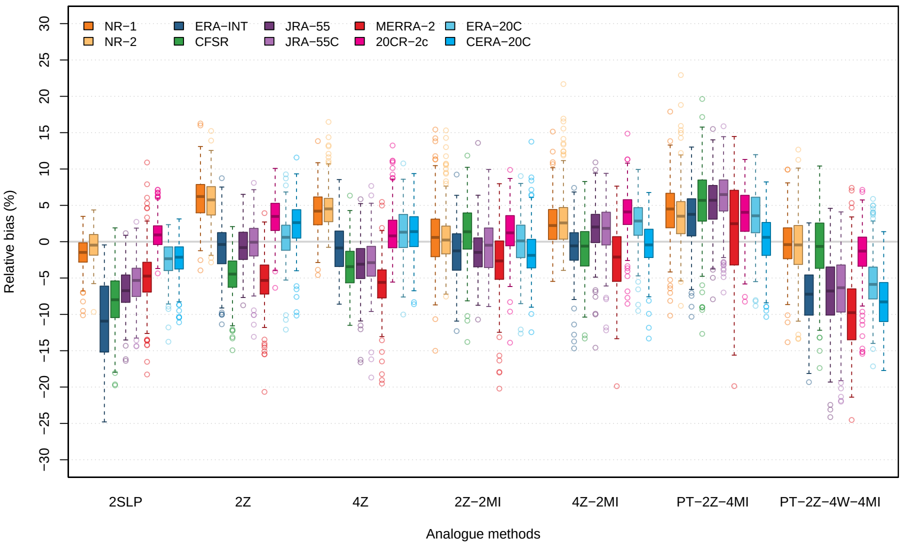

Fig. 5 Same as Fig. 4, but for relative biases.

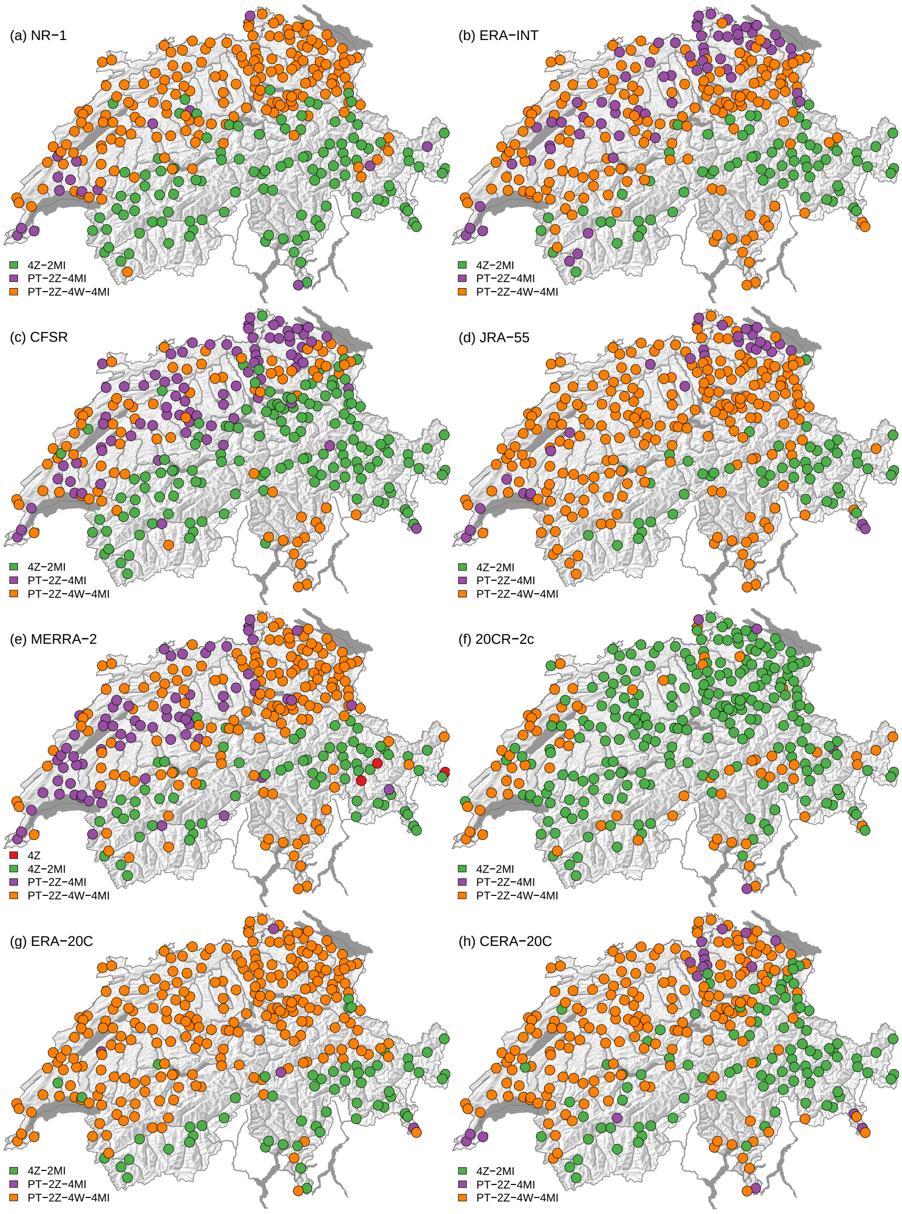

Fig. 6 Best method per station for the different datasets. NR-2 and JRA-55C are not shown as they are similar to NR-1 and JRA-55 respectively. Background map: © Swisstopo.

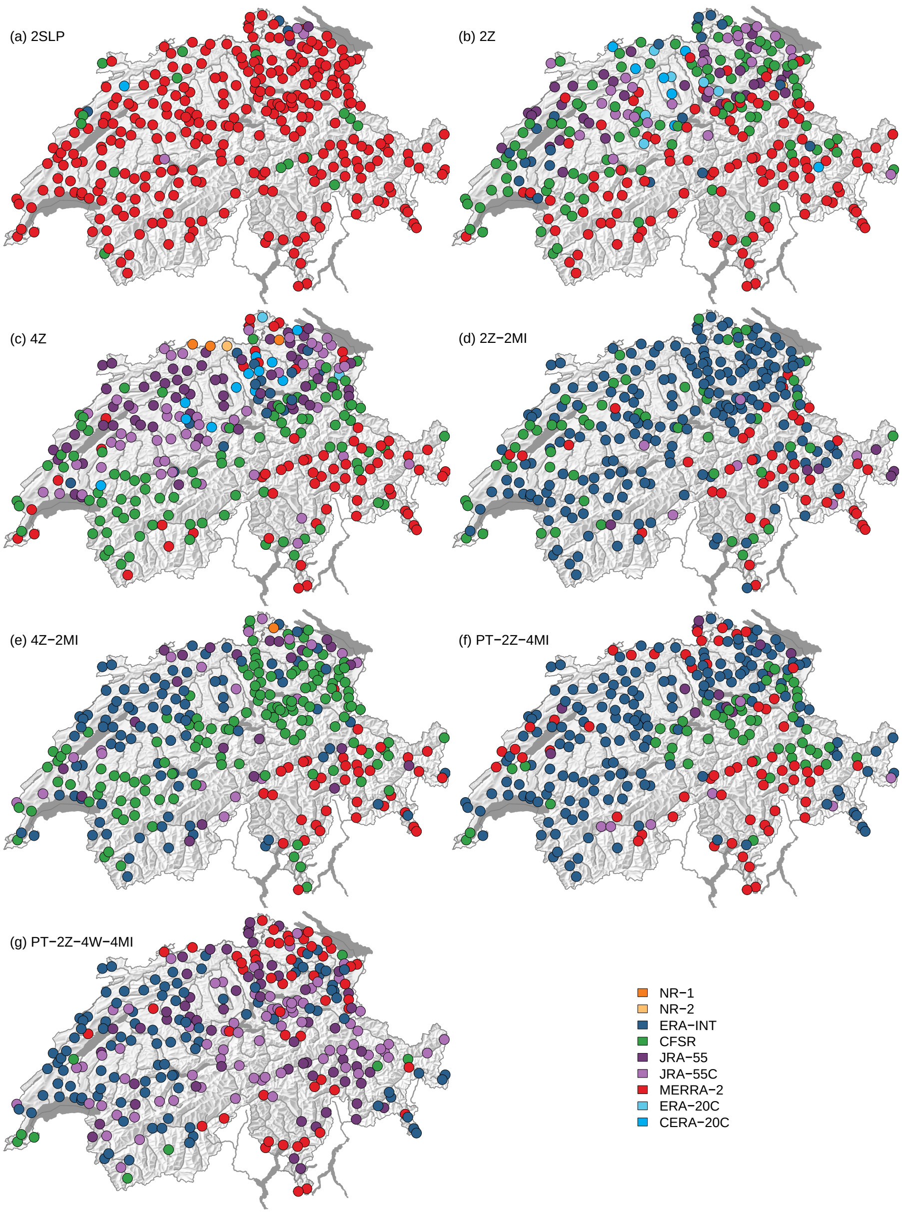

Fig. 7 Best reanalysis per station for the different methods. Background map: © Swisstopo.

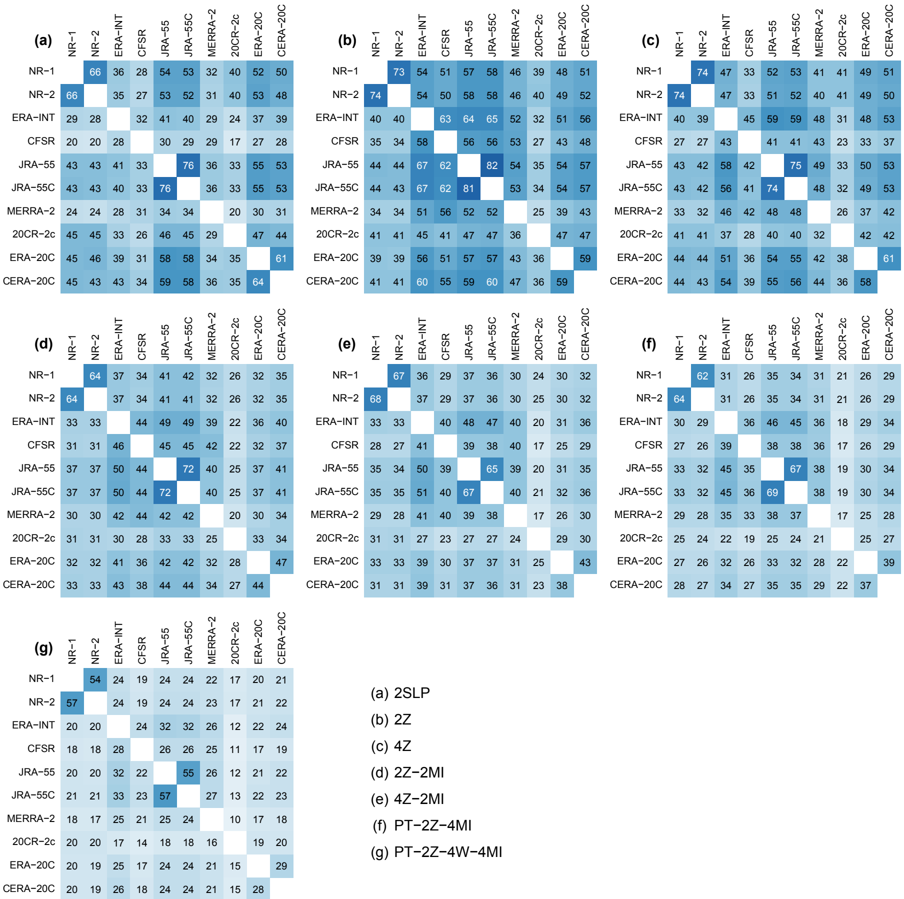

Fig. 8 Percentage of identical analogue dates selected when using the reanalysis datasets in columns that are also found when using the datasets in rows for different AMs. The values are averaged for all stations on the VP.

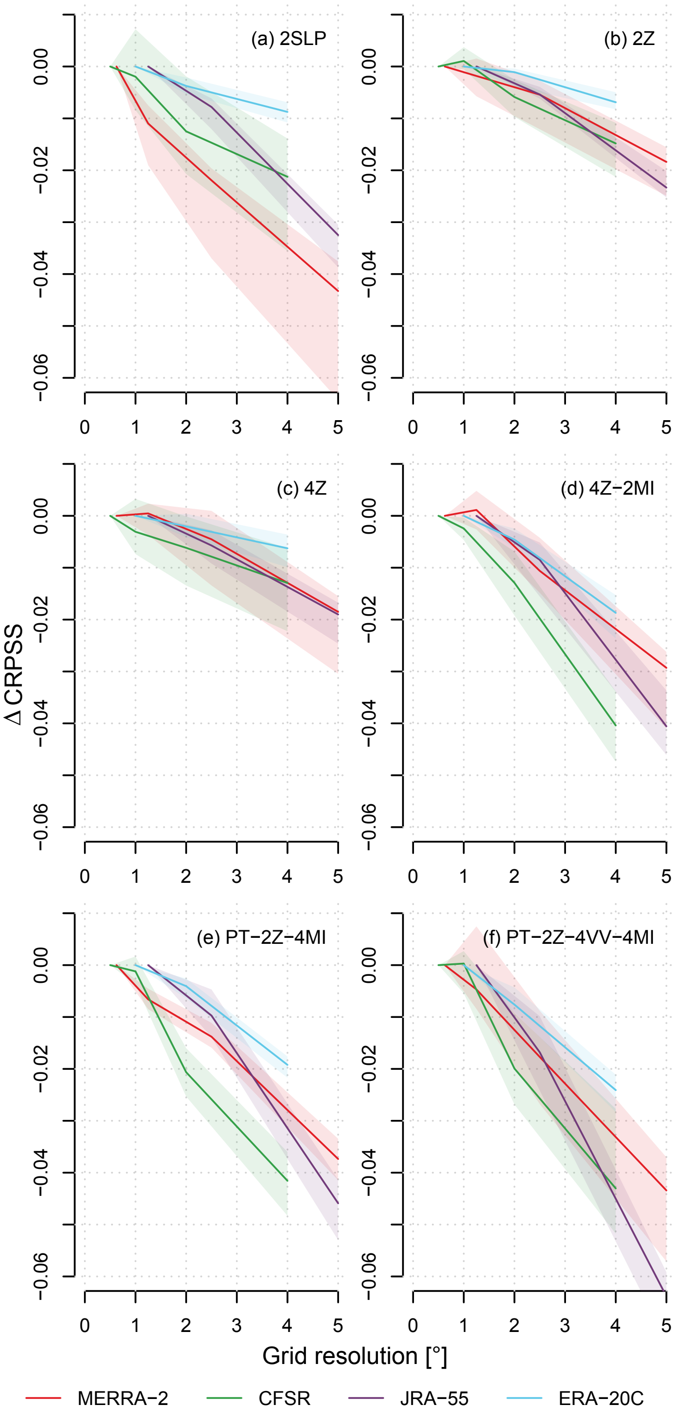

Fig. 9 Impact (difference in CRPSS) of a decrease in grid resolution (degrees) for different datasets and AMs on the CP. The line represents the median and the shaded area represents the first and the third quartiles (on 30 stations).

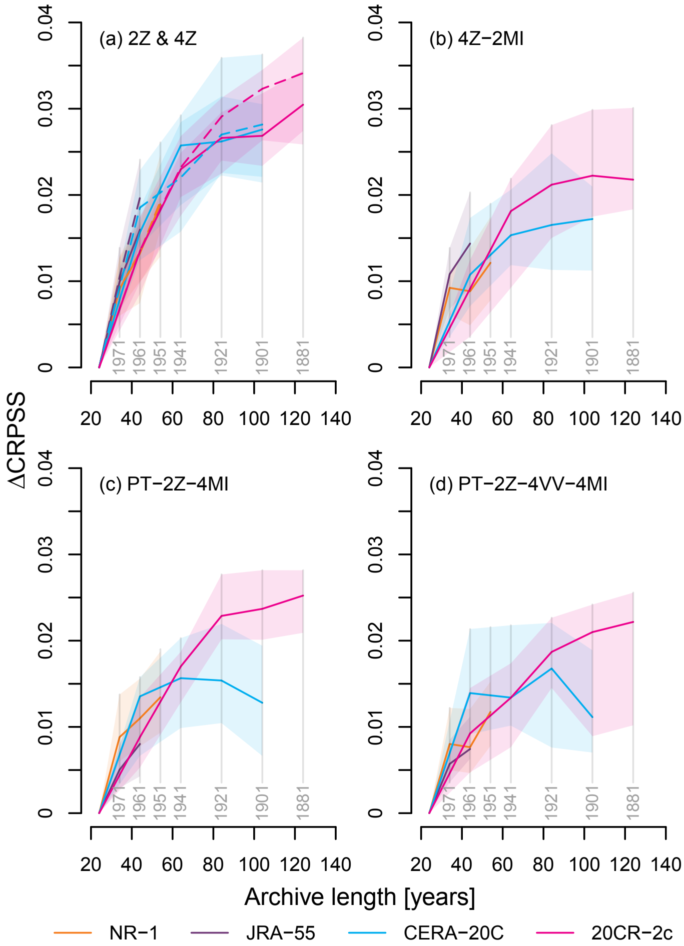

Fig. 10 Impact (difference in CRPSS) on the VP of an increase in the archive length (years) for different datasets and AMs. Results for the 4Z method (shown by the dashed lines) are displayed along with the 2Z method. The line represents the median and the shaded area represents the first and the third quartiles (on 30 stations).

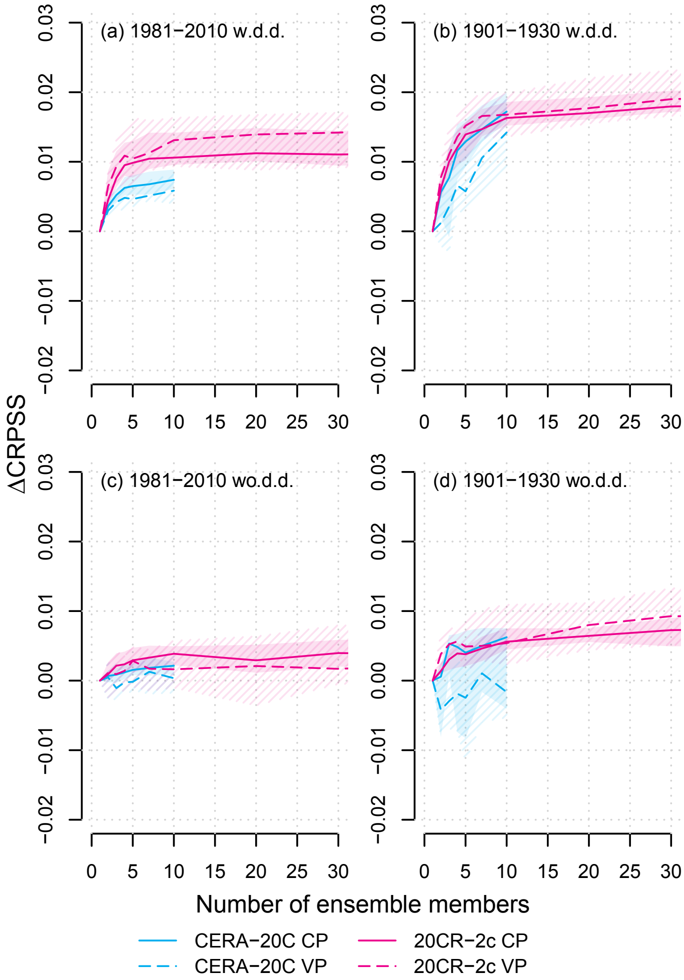

Fig. 11 Impact (difference in CRPSS) of an increase in the number of ensemble members used for the 2Z method, and for CERA-20C and 20CR-2c datasets. The results are provided for two periods: (a, c) 1981–2010 and (b, d) 1901–1930. Two approaches were assessed: (a, b) the first allowing duplicate analogue dates ("w.d.d.") and (c, d) the second without duplicate analogue dates ("wo.d.d."). The line represents the median and the shaded area represents the first and the third quartiles (on 30 stations). The dashed line and striped area correspond to results on the VP. All 56 members of 20CR-2c were assessed and the tendencies continue, but the plots are split at 30 members.

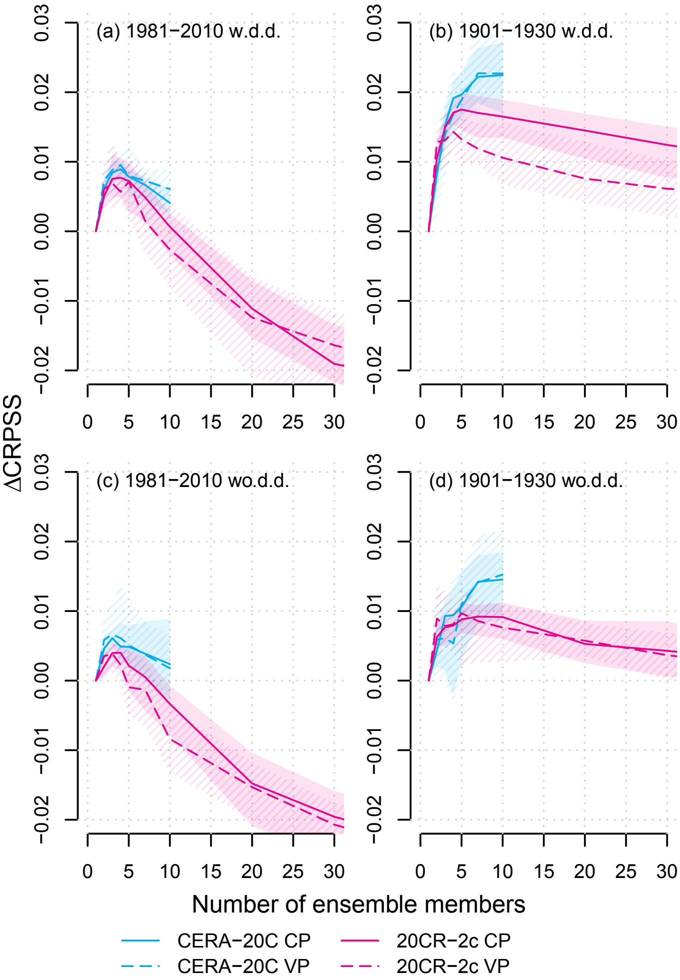

Fig. 12 Same as Fig. 11 but for the 2Z-2MI method.

	1851 - ...	1900 - ...	1958 - ...	1980 - ...	2010 - ...
SLP	20CR-2c	CERA-20C	JRA-55C CERA-20C	MERRA-2	MERRA-2
		ERA-20C	ERA-20C	CFSR JRA-55[C] CERA-20C	CFSR JRA-55
Z	20CR-2c	CERA-20C	JRA-55C	CFSR JRA-55[C] MERRA-2	CFSR JRA-55 MERRA-2
		ERA-20C	CERA-20C	ERA-INT CERA-20C	ERA-INT
M	20CR-2c	CERA-20C	JRA-55C	ERA-INT MERRA-2	ERA-INT MERRA-2
		ERA-20C	CERA-20C	CFSR JRA-55[C]	CFSR JRA-55
T	20CR-2c	CERA-20C	JRA-55C	ERA-INT CFSR MERRA-2 JRA-55[C]	ERA-INT CFSR MERRA-2 JRA-55
		ERA-20C	CERA-20C		
W	20CR-2c	CERA-20C	JRA-55C	JRA-55[C] ERA-INT MERRA-2	JRA-55 ERA-INT MERRA-2
		ERA-20C	CERA-20C	CFSR	CFSR

Fig. 13 Synthesis table of the recommended reanalyses to use in AMs for different periods and variables. This recommendation applies to Europe and eventually other data-rich regions of the world. The darker shaded area represents the first choice and the lighter shaded area represents alternatives. When a reanalysis is not mentioned, it is either not available or not recommended.