The analogue method for precipitation prediction: finding better analogue situations at a sub-daily time step

Pascal Horton[1,2], Charles Obled[3], and Michel Jaboyedoff[1]

[1]Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland
[2]Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland
[3]Laboratoire d'étude des Transferts en Hydrologie et Environnement (LTHE), Université de Grenoble-Alpes, Grenoble, France

Correspondence to: Pascal Horton (pascal.horton@alumnil.unil.ch)

Abstract. Analogue methods (AMs) predict local weather variables (predictands) such as precipitation by means of a statistical relationship with predictors at a synoptic scale. The analogy is generally assessed on gradients of geopotential heights first to sample days with a similar atmospheric circulation. Other predictors such as moisture variables can also be added in a successive level of analogy.

5   The search for candidate situations similar to a given target day is usually undertaken by comparing the state of the atmosphere at fixed hours of the day for both the target day and the candidate analogues. This is a consequence of using standard daily precipitation time series, which are available over longer periods than sub-daily data. However, it is unlikely for the best analogy to occur at the exact same hour for the target and candidate situations. A better analogue situation may be found with a time shift of several hours since a better fit can occur at different times of the day. In order to assess the potential for finding

10   better analogues at a different hour, a moving time window (MTW) has been introduced.

The MTW resulted in a better analogy in terms of the atmospheric circulation and showed improved values of the analogy criterion on the entire distribution of the extracted analogue dates. The improvement was found to increase with the analogue rank owing to an accumulation of better analogues in the selection. A seasonal effect has also been identified, with larger improvements shown in winter than in summer. This may be attributed to stronger diurnal cycles in summer that favour

15   predictors taken at the same hour for the target and analogue days.

The impact of the MTW on the precipitation prediction skill has been assessed by means of a sub-daily precipitation series transformed into moving 24 h totals at 12-h, 6-h, and 3-h time steps. The prediction skill was improved by the MTW, as was the reliability of the prediction. Moreover, the improvements were greater for days with heavy precipitation, which are generally related to more dynamic atmospheric situations in which the timing is more specific and fewer records are available in the

20   meteorological archive.

The improvements of the analogy criterion and the performance scores on precipitation were both found to be higher for MTWs with a smaller time step of 3 h. A 3-h MTW provides eight times more candidate situations even though they are not fully independent. Since the MTW provides additional situations to the pool of possible analogues, it can be considered as an inflation of the meteorological archive. Because this technique is very simple and easily applicable, it should be considered for

25   several applications in different contexts, such as operational forecasting or climate-related studies.

1 Introduction

The analogue method (AM) is a statistical downscaling method based on the hypothesis that two relatively similar synoptic situations may produce similar local effects (Lorenz, 1956, 1969). Other statistical downscaling methods for climate studies can be found in Maraun et al. (2010). The AM consists of finding past situations that are similar to the target day of interest in terms of atmospheric circulation or other synoptic predictors. Referred to as the predictand, the local weather variables of interest which were observed at the analogue dates are used to construct a probabilistic prediction for the target day (Duband, 1970; Zorita and Storch, 1999). Multiple variations of this method have been reported (Ben Daoud et al., 2016). AMs are used for operational precipitation forecasting in the context of weather forecasting, flood forecasting, and hydropower production (e.g., Bontron and Obled, 2005; Hamill and Whitaker, 2006; Desaint et al., 2008; García Hernández et al., 2009; Bliefernicht, 2010; Marty, 2010; Marty et al., 2012; Horton et al., 2012; Obled, 2014; Hamill et al., 2015; Ben Daoud et al., 2016), as well as for precipitation downscaling from a climate perspective (e.g. Radanovics et al., 2013; Chardon et al., 2014; Dayon et al., 2015). AMs can provide multivariate predictions which are physically consistent (Raynaud et al., 2016). Their spatial transferability is analysed in Chardon et al. (2014) and Radanovics et al. (2013), and their temporal transferability has been discussed in recent studies for past or future climates (Dayon et al., 2015; Caillouet et al., 2016).

The method requires two different archives. The first is a meteorological archive describing the state of the atmosphere at a synoptic scale, such as reanalysis datasets, at different hours of the day. The second is an archive of the target variable to be predicted, which is usually standard daily precipitation totals from 06:00 UTC to 06:00 UTC the following day either at a target station or integrated over a target catchment. Obviously, the period to be used is limited to the smallest period common to the two archives.

Ruosteenoja (1988) and Van Den Dool (1994) analysed the influence of the length of the meteorological archive on the quality of the analogy. They highlighted a three-way relationship among the quality of the analogy, the archive length, and the size of the spatial domain (or degrees of freedom) and determined that errors increase with a larger domain but decrease with a longer archive. For that reason, limited spatial windows are always considered when searching for analogues, and the archive length is maximised.

Therefore, owing to the availability of long precipitation archives at a daily time step that have no equivalent at a finer resolution, AMs are usually implemented on a daily basis. Given the cumulative aspect of the predictand, the corresponding meteorological situation is characterized by several predictors taken at different but fixed hours of the day, and the analogue situations are assessed by comparing these same predictors at the same fixed hours of the day. However, it can be expected that the best analogy of the synoptic situations does not occur systematically at the same time of the day and that better candidates can be found by shifting to different hours. With this assumption, a moving time window (MTW) was introduced to allow the search for candidates at different hours of the day. Previous tests have showed the benefit, in terms of analogy criterion values, of searching for analogue synoptic situations at a finer time step, although such research did not assess the impact on the prediction skill of a specific predictand (Finet et al., 2008).

Other possibilities exist for increasing the prediction skill of the AMs. A classical approach is to add new predictors or new successive levels of analogy (e.g. Horton, 2012; Ben Daoud et al., 2016; Caillouet et al., 2016). AMs can also be combined with other methods (e.g. Chardon et al., 2014). Another possibility is to use a global optimization technique, such as genetic algorithms, in order to better optimize the method and to add new parameters (Horton et al., 2017). However, the MTW technique is not in competition with other approaches and can be combined with them.

The outline of this paper is as follows. Section 2 presents the context of the study as well as the data and methods, including the proposed MTW technique. The benefits of introducing an MTW are assessed first regarding the improvement of the analogue date selection (Sect. 3.1) and then in terms of precipitation prediction skill (Sect. 3.2). Finally, the results are discussed in Sect. 4, and the conclusions are found in Sect. 5. A list of the acronyms and their definitions is provided in appendix A.

2 Data and methods

2.1 Study area and data

The study area is the upper Rhône catchment in Switzerland. The precipitation time series were obtained from six automatic weather stations, Ulrichen, Zermatt, Visp, Montana, Sion, and Aigle (Fig. 1), that are subject to various meteorological influences (Horton et al., 2012). The data were available at an hourly time step for 29 years (1982–2010) and were also obtained at a standard daily time step of 06:00 UTC to 06:00 UTC the following day for 50 years (1961–2010). Owing to the low density of weather stations with high temporal resolution and long archives, no spatially aggregated rainfall was processed.

Synoptic-scale variables used as predictors were extracted from two of the most recent global reanalysis datasets: the European Center for Medium Range Weather Forecasting 20th century reanalysis (ERA-20C) (Poli et al., 2016), with 3-h temporal resolution and a spatial resolution of 1°, and Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), from the National Aeronautics and Space Administration (NASA), with a 6-h temporal resolution and a spatial resolution of $0.625° \times 0.5°$. MERRA-2 is an update for the first MERRA reanalysis (Rienecker et al., 2011). The study was originally performed on the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis 1 dataset (Kalnay et al., 1996); the conclusions were similar. ERA-20C is built by assimilating only surface observations and is thus available for a long period (1900–2010). On the contrary, MERRA-2 is built by assimilating observations at higher levels and from more sensors including satellite data, but it is more limited in time (1980–present). It is of interest to assess the relevance of an MTW with these two datasets because they represent different types of products. The specific advantages of ERA-20C are that it allows for testing an MTW with a 3-h time step, and it covers a long period. On the contrary, MERRA-2 has a higher spatial resolution and can be expected to be more accurate at higher levels of the atmosphere. The variables extracted from these datasets were geopotential heights at 500 hPa (Z500) and 1000 hPa (Z1000), the total precipitable water (TPW), and the relative humidity at 850 hPa (RH850).

2.2 The considered analogue methods

The first considered AM is based on the analogy of atmospheric circulation only (Table 1; Obled et al., 2002; Bontron and Obled, 2005). Before searching for analogue situations for a target day, seasonal stratification is conducted through preselection of the potential candidates for analogy. The search is restricted to analogue days within a four-month window centred on the target date for every year of the archive. The similarity of the atmospheric circulation of the target date with every day of the preselection is assessed by processing the Teweles and Wobus (1954) score (S1), which is a comparison of gradients of geopotential heights over a particular spatial window and at certain hours:

$$
S_1 = 100 \frac{\sum\limits_{i}^{m} |\Delta \hat{z}_i - \Delta z_i|}{\sum\limits_{i}^{m} max(|\Delta \hat{z}_i|, |\Delta z_i|)},
\tag{1}
$$

where $\Delta \hat{z}_i$ is the forecast geopotential height difference between the ith pair of adjacent points from the grid of the target situation, Δz_i is the corresponding observed geopotential height difference in the candidate situation, and m is the number of pairs of adjacent points in the grid. The differences are processed separately in both directions. With smaller S1 values, there is greater similarity in the pressure fields.

The predictor variables extracted from reanalysis datasets are considered at different hours of the day. Based on Bontron and Obled (2005), geopotential heights are compared at 1000 hPa (Z1000) at 12:00 UTC and at 500 hPa (Z500) at 24:00 UTC. The time of the day at which the predictors are selected has a significant influence (Bontron, 2004).

Then, N_1 dates with the lowest values of S1 are considered as analogues to the target day, where N_1 needs to be calibrated (Sect. 2.3). Finally, the daily observed precipitation amount of the corresponding dates provide the empirical conditional distribution considered as the probabilistic prediction for the target day. This method is referred to as 2Z.

The second reference method (2Z-2MI, Table 2; Bontron and Obled, 2005) adds a subsequent level of analogy with moisture variables, which are compared by means of the root–mean–square error (RMSE) criteria:

$$
C_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{v}_i - v_i)^2},
\tag{2}
$$

where $\hat{v}_i$ is the ith predictor value from the grid of the target situation, v_i is the corresponding observed value in the candidate situation, and n is the number of points in the grid.

The additional predictor is a moisture index composed of the product of the total precipitable water (TPW) with the relative humidity at 850 hPa (RH850) at 12:00 UTC and 24:00 UTC (Bontron, 2004). When adding a second level of analogy, N_2 dates are subsampled from the N_1 analogues on the atmospheric circulation, resulting in a smaller number of analogue situations. Moreover, when a second level of analogy is added, a higher number of N_1 analogues is kept on the first level.

More complex AMs exist with additional predictors (e.g. Horton, 2012; Ben Daoud et al., 2016; Caillouet et al., 2016). The MTW can also be applied to these; however, it is easier to interpret the impact of the MTW by using more basic methods.

2.3 Calibration of the analogue method

AMs rely on parameters that need to be defined for every level of analogy. Here, the selection of the meteorological variables used as predictors, as well as their corresponding pressure level and hour of the day, were considered to be identical to those used in methods 2Z (Table 1) and 2Z-2MI (Table 2) described above. The parameters calibrated in this study are listed below.

- Spatial windows, which are the domains in which the predictors are compared. A spatial window is specific to each level of analogy; thus, its extent differs among circulation and moisture variables.

- The optimal number of analogue situations for every level of analogy.

The semi-automatic sequential procedure discussed by Bontron (2004) was used to calibrate the AM (Radanovics et al., 2013; Ben Daoud et al., 2016). The analogy levels (e.g. the atmospheric circulation or moisture index) are calibrated sequentially. The procedure, as defined by Bontron (2004), consists of the following steps:

1. For every level of analogy

    (a) Identification of the most skilled unitary cell of the predictor data over a large domain. Every point or cell of the full domain is jointly assessed on the predictors of the current level of analogy.

    (b) From this most skilled point, the spatial window is expanded by successive iterations in the direction of greater performance gain. The spatial window increases until no improvement is reached.

    (c) The number of analogue situations N_1, which was initially set to an arbitrary value, is then reconsidered and optimised for the current level of analogy.

2. A new level of analogy can then be added based on other variables such as the moisture index at chosen pressure levels and hours of the day. The number of analogues for the next level of analogy, N_2, is initiated at a chosen value. The procedure starts again from step 1 (calibration of the spatial window and the number of analogues) for the new level. The parameters calibrated on the previous analogy levels are fixed and do not change.

3. Finally, the numbers of analogues N_1 and N_2 for the different levels of analogy are reassessed. This is performed iteratively by varying the number of analogues of each level in a systematic manner.

The calibration is performed in successive steps and aims at minimizing the objective function (CRPS, Eq. 3). Except for the number of analogues, previously calibrated parameters are generally not reassessed. More advanced techniques such as using genetic algorithms (Horton et al., 2017) exist but are beyond the scope of the present study.

The 29-year hourly precipitation dataset was divided into a calibration period (CP) and a validation period (VP) in order to assess the robustness of the proposed improvements on independent data. The selection of the VP was evenly distributed over

the entire series (Ben Daoud, 2010) to reduce potential biases related to trends linked to climate change or to evolution in the measurement techniques. Thus, 1 out of every 5 years was selected for validation, which represents a total of 6 years for the VP and 23 years for the CP.

The continuous ranked probability score (CRPS, Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000) is often employed to assess the performance of AMs (e.g. Bontron, 2004; Bontron and Obled, 2005; Ben Daoud et al., 2008; Horton et al., 2012; Marty et al., 2012; Radanovics et al., 2013; Chardon et al., 2014; Junk et al., 2015; Ben Daoud et al., 2016; Caillouet et al., 2016). This allows for evaluation of the predicted cumulative distribution functions $F(y)$ of the precipitation values y from analogue situations compared with the observed value y^0; a better prediction has a smaller score. This score is defined as

$$S_{\mathrm{CRPS}} = \int\limits_{-\infty}^{+\infty} \left[F(y) - \mathrm{H}(y - y^0) \right]^2 \mathrm{d}y, \tag{3}$$

where $\mathrm{H}(y - y^0)$ is the Heaviside function, which is null when $y - y^0 < 0$ and has the value of 1 otherwise.

2.4 The moving time window approach

The MTW technique is used to find better analogue situations at different hours of the day rather than comparing the predictors at the same fixed hours. The target situation (the day to predict) is still the same as that in the conventional approach: a daily precipitation total between 06:00 and 30:00 UTC characterised by predictors, which in this case are Z1000 and Z500 at 12:00 and 24:00 UTC, respectively. The difference is that candidate situations are considered not only at the same time (12:00 and 24:00 UTC), but also at other hours by allowing a time shift. Therefore, instead of looking for analogues at a 24-h time step, they are sought at the time step matching the predictor temporal resolution (Fig. 2). The ERA-20C dataset was used here to test an MTW with 12-h, 6-h, or 3-h time steps, which are referred to as 12-h MTW, 6-h MTW, and 3-hr MTW, respectively. Thus, the candidates are two (12-h MTW), four (6-h MTW), or eight (3-h MTW) times as many as in the conventional approach even though they are not fully independent.

The target situations and their corresponding observed precipitation values do not change. The predictions are asessed on the exact same precipitation time series for both the MTW and the conventional approach so that the performance scores can be directly compared. In order to assess the benefit of searching for analogue situations at a sub-daily time step for quantitative precipitation prediction, an appropriate precipitation archive is required. On the basis of the hourly time series (Sect. 2.1), 24-h totals were processed at time steps matching the MTW by building moving 24-h totals on the period 1982–2010, for example starting at 0, 3, ..., 21 h UTC instead of 6 h UTC as for the standard archive.

3 Results

The considered AMs were applied to both ERA-20C and MERRA-2 datasets. Most of the results are presented for the ERA-20C dataset because it allows for testing a 3-h MTW, although similar conclusions can be found by using MERRA-2. The impact of an MTW approach was assessed on the two bases given below.

- Original parameters: The AMs were initially calibrated for each station based on the conventional fixed 24-h approach. Then, the MTW was introduced into the AMs, but the parameters (spatial windows and number of analogues) were not reassessed. The analogy was then identical, except for the fact that the MTW had more candidates at disposal.

- Recalibrated: The parameters of the AMs were reassessed after the introduction of the MTW at different time steps. Indeed, one can assume that the introduction of the MTW might change the optimum value of some parameters (different spatial windows and number of analogues). The calibration (Sect. 2.3) was then reprocessed to adapt the AM parameters to the new information available through the MTW. The main difference was the increase in the optimal number of analogues to retain, which is discussed in Sect. 4.2.

The MTW is expected to affect both the selection of analogue dates and the prediction skill. Both are analysed separately hereafter. The results are presented for the Ulrichen station, but were very similar for all other stations.

3.1 Influence of the MTW on the selection of analogue dates

First, the impact of the MTW on the selection of analogue dates was analysed independently of precipitation data. This was conducted for the period 1982–2010, although a longer period could have been used because no sub-daily precipitation series is required at this stage. The original parameters were used when assessing the impact of the MTW; no recalibration of the parameters was conducted.

3.1.1 Analogy of the atmospheric circulation

When searching for analogues in the first level of analogy, such as on the geopotential heights in the 2Z method (Sect. 2.2), up to eight times as many candidates were provided with the MTW than before even though they are not fully independent (Sect. 2.4). Figure 3 shows the changes in the distributions, for target dates over the CP, of the analogy criterion (S1) for the 1st, 15th, and 30th analogue rank due to the introduction of an MTW with a 12-h, 6-h, or 3-h time step. The shapes of the distributions of the conventional approach and the MTWs were found to be similar. However, the values of the analogy criterion were gradually reduced (shifted to the left) with smaller MTW time steps and were therefore better. The circulation analogy was regularly improved with the doubling of the MTW time step. For analogues with higher ranks (e.g. 15th or 30th), the difference between the two distributions was larger than that for the first rank, which means that the improvement increased with the rank of the analogue.

The improvements of the analogy with the rank of the analogues are summarised in Fig. 4, which shows the quantiles of the S1 criterion for the conventional method and the MTWs at different time steps (top) and those of the relative reduction

indicating improvement owing to the MTWs (bottom). This confirms that the improvement constantly increased from the first to the last analogue (Fig. 4 bottom). This can be explained by the accumulation of better analogues in the selection, whereby the new, better situations push the previous analogues to higher ranks. The regularity of the improvements brought by decreasing time steps is also evident. The introduction of the MTW enables better analogue situations to be found in the first level of

5   analogy, resulting in the selection of days with better similarity in the atmospheric circulation. This improvement was superior for MTWs with a smaller time step, which obviously allows better matching.

The impact of the MTW on the analogy criterion has been analysed per precipitation class, for the target day, for 6-h and 3-h MTWs. The results are summarised in Fig. 5 by the median reduction of S1 for the days with precipitation between two thresholds. With a reduction in the number of cases per class, the curves were not as smooth as those in the previous analyses.

10   It is nevertheless clear that the improvements were larger for days with higher precipitation for both the 6-h and 3-h MTW.

The effect of the MTW on the S1 criterion was decomposed per season and is presented in Fig. 6. The differences among the seasons were substantial, with greater improvements shown for winter than summer. Analysis of the selected hours for the geopotential height predictor was then performed for the MTW with different time steps (Fig. 7). The new choice of the temporal window in winter was more balanced between the different hours of the day, showing more regular repartition, for all

15   MTW time steps. This indicates a change in selection of a greater portion of the analogue dates for winter. On the contrary, the days in summer showed a preference for the initial temporal window (Z500, 24 h and Z1000, 12 h). This is likely attributed to the more pronounced diurnal effects, which reduced the potential for improvement of the criteria. These effects are in phase with the daily cycle, and good analogues were essentially found for the same hours. When a 3-h MTW was used, the time step following the initial temporal window (Z500, 27 h and Z1000, 15 h) was selected almost as often as the initial values for all

20   seasons. This temporal window can be considered as relatively similar in terms of the solar influence. The other seasons were between these two extremes, which is consistent with their respective improvements in S1.

3.1.2   Moisture analogy

When adding the second level of analogy of the 2Z-2MI method (Table 2), the number of candidate situations did not increase when using the original parameters because they were conditioned by the N_1 previously selected analogues; however, their

25   dates changed. In contrast to the AM on the atmospheric circulation only, both a reduction and an increase in the RMSE analogy criterion values were possible with the MTW compared with the conventional fixed approach. The results of the second level of analogy, presented in Fig. 8, showed no improvement in RMSE values regardless of the MTW time step. Unlike the first level of analogy, the relative changes in the RMSE values were distributed relatively symmetrically around zero.

These criterion values were not improved by the MTW because the number of candidates (N_1) was not higher. The use

30   of an MTW did not increase the sample size in this case because the second level of analogy consists in subsampling in the dates provided by the first level. However, this result of a globally null improvement in the RMSE values does not mean that the 2Z-2MI method cannot be improved by the MTW. After selection of the analogue dates for the synoptic circulation, the new candidate dates did not provide better analogues in terms of moisture. However, the selected dates changed in the first

level of analogy and in the final selection; thus, the distributions of precipitation values were different, which can impact the performance of the precipitation prediction.

3.2 Impact of the MTW on the precipitation prediction

The changes in the performance score (CRPS) of the precipitation prediction are provided in Fig. 9 and 10 for the 2Z and 5 2Z-2MI methods, respectively. The AMs were recalibrated on the CP and were assessed on the independent VP. The score was processed for each station and both reanalysis datasets. For ERA-20C, the 12-h MTW, 6-h MTW, and 3-h MTW could be assessed, whereas only the two first could be processed with the MERRA-2 dataset.

The MTW did improve the precipitation prediction because the CRPS globally decreased for all stations, both for the CP and the VP. The VP showed greater variability, which is likely related to its shorter period of six years. The prediction skill for the 10 CP was almost always improved further by reducing the time step of the MTW, but not of the same magnitude. For ERA-20C, the magnitude of the improvement of 2Z for the CP was relatively reduced when using a 3-h MTW instead of a 6-h MTW, and the variability for the VP increased. However, as shown in Fig. 4, the selection of analogue dates was improved relatively equally for every reduction of the MTW time step. A possible reason is that the ERA-20C dataset was built by assimilating only surface observations, whereas the AMs rely on the geopotential height at 500 hPa. Thus, the timing of the atmospheric 15 circulation at higher pressure levels might not be completely in phase with the actual perturbation systems which caused the observed precipitations. The MTW with the MERRA-2 dataset and the 2Z method also showed a certain slope break after the 12-h MTW that could not be explained and was less important for the 2Z-2MI method. However, the 12-h MTW resulted in important variability for the 2Z-2MI method and the MERRA-2 dataset.

The introduction of an MTW was previously found to show greater improvement of the S1 criterion for days with higher 20 precipitation values (Sect. 3.1.1). The impact of the MTW was then assessed in terms of changes in prediction performance per precipitation class for the target day. Figure 11 synthesises these differences for the Ulrichen station by using the original parameters with the MTW. After the introduction of the MTW, the performance score was generally further improved with reduced CRPS for days with higher precipitation than for non-rainy days and small precipitation values. When recalibrating the parameters using the MTW, the spread of the changes in CRPS tended to decrease, although a slight loss in performance 25 was noted for the most rainy days (not shown).

Reliability diagrams were also created for the VP for both 2Z and 2Z-2MI methods and for the exceedance of all days percentiles 80 % (P80, 4 mm), 90 % (P90, 9.5 mm), and 95 % (P95, 17.5 mm) at the Ulrichen station (Fig. 12). These diagrams plot the observed frequency against the predicted probability of a binary event, which in this case is threshold exceedance. For perfectly calibrated predictions, the curve should be along the diagonal. The VP contains only six years (Sect. 2); therefore, 30 higher variability is present for higher thresholds, and the curves are smoother and closer to the diagonals for the CP, which is longer (not shown). Figure 12 shows that the MTW improved the reliability for both 2Z and 2Z-2MI methods and for all thresholds because the curves moved towards the diagonal with decreasing MTW time steps. For P80, the 3-h MTW predictions fit very well the observed frequency. When considering higher thresholds, the 2Z-2MI method was generally better than 2Z. The case of the P95 threshold showed a conditional dry bias associated with larger forecast probabilities for the considered

AMs; however, this was substantially reduced by the 3-h MTW. This conditional dry bias means that the AM did not predict the event with a probability equal to 1 for 2Z. Obtaining a predicted probability of exceedance of 1 would mean that all analogue dates for a certain target date would be above the threshold. The number of analogue dates being 30 in this case, it is less likely to have 30 values above the threshold when this one increases. This issue has been addressed by Marty (2010) who proposed a

5 bias correction method for the AM. However, as shown in Figure 12, the introduction of an MTW, especially with a 3-h time step, improves this mismatch to some degree.

4 Discussion

4.1 Better prediction of heavy precipitation

Both the analogy criteria (Sect. 3.1.1) and the performance scores (Sect. 3.2) were improved to a greater extent for days with

10 heavier precipitation. This is likely attributed to the fact that higher precipitation events are a consequence of atmospheric conditions with greater dynamics such as weather disturbances, which have well-marked temporal evolution. Indeed, the position of the driving elements, such as the low-pressure centres and the fronts, change significantly during the course of a day. These situations are less numerous than anticyclonic situations, which makes it less likely to find very good analogues at the same time of the day. We can, therefore, expect to more significantly improve these situations with greater dynamics when introduc-

15 ing an MTW because better matches to the target situation may be found. In contrast, days with low dynamics in atmospheric circulation, such as anticyclonic situations, will not be radically improved by the introduction of the MTW.

The MTW improved the prediction for days with heavier precipitation. Therefore, it should improve the prediction of extremes owing to selection of better analogue situations and to addition of possible new extreme precipitation values resulting from moving 24 h totals with a certain time shift. However, even though the distribution of analogue precipitation values should

20 move towards the targeted extreme to provide better prediction, the MTW itself does not enable prediction of extreme events that were not yet observed and are therefore not present in the archive. However, the extremes in AMs can be modelled by extrapolation of a truncated exponential or gamma distribution fitted to the analogue values (Obled et al., 2002). Another possible approach is to combine AMs with other methods (e.g. Chardon et al., 2014). From this perspective, the MTW might improve the prediction of extremes because it improves the distribution of precipitation values for days with higher precipitation, on

25 which post-treatment techniques rely. However, this topic is beyond the scope of the present study.

4.2 The relationship between the MTW time step and the number of analogues

When recalibrating the AMs with MTWs of different time steps, the optimal number of analogues changed for both 2Z and 2Z-2MI methods. N_1 and N_2, from the first and the second level of analogy, tended to significantly increase with a reduced MTW time step (Fig. 13). When using ERA-20C, all optimal numbers of analogues for the different analogy levels tended

30 to double when using a 6-h MTW instead of the classic approach and to even triple when using a 3-h MTW. These higher

numbers of analogues were objectively chosen by using the calibration procedure (Sect. 2.3) in order to increase the prediction skill of the methods.

As shown in Fig. 4, the improvement of the S1 criterion increases along with the rank of the analogue, which shows an accumulation of better analogue situations in the distributions. Widening the selection of analogues appears to be beneficial for keeping some with increased rank because they appear to be relevant in predicting the precipitation values. Thus, the number of good analogues was globally increased.

A higher number of analogues generally implies a poorer analogy with an archive of fixed length. Indeed, when the choice of the predictors or the parameters improve, leading to a better prediction, the optimal number of analogue situations tends to decrease. However, when the length of the archive increases, the optimal number of analogues also tends to increase for better performance up to a certain threshold (Bontron, 2004; Hamill et al., 2006). The observed increase in the number of analogue situations with the MTW resulting in better performance skills for the given methods can therefore be considered as an inflation of the archive.

4.3 Inflation of the archive

Because sub-daily precipitation time series are often available over a shorter period than daily, a reduction of the overall archive length at disposal is necessary. This usually has a negative consequence on the skill of precipitation prediction. The role of the archive length was assessed on the ERA-20C dataset and is presented in Fig. 14 for both 2Z and 2Z-2MI methods using the conventional approach of fixed 24-h windows. As expected, the performance of the method globally increased with the length of the archive. The best performances were found with a 44-year archive, which corresponds to the period 1961–2010 minus the 6 years of independent VP. This addition of 21 years of archive to the original 23 years resulted in an improvement of both 2Z and 2Z-2MI methods that was of the same order of magnitude as the introduction of a 6-h MTW (Fig. 9 and 10). The 3-h MTW resulted in similar performance for the 2Z method and slightly superior improvements for the 2Z-2MI method.

As previously stated, the MTW can be considered as an inflation of the archive because it provides additional situations to the pool of possible analogues. With the introduction of the MTW, the performance loss related to an eventual reduction of the archive length to meet the length of the sub-daily precipitation archive was indeed compensated. A 6-h MTW brought gains of the same magnitude similar to an almost double archive length. However, it should be noted that despite the number of candidate situations, which in this case is four times the original number, the gains appear to be lower than those for a quadrupled archive length. The likely reason is that an actual longer archive would contain more atmospheric situations that might have been observed less frequently during a shorter period. Moreover, rather strong serial correlations between successive sub-daily circulation patterns are expected.

Many reanalysis datasets begin in the 80s owing to availability of satellite data and are thus more accurate than reanalysis datasets based on conventional data only; however, they provide shorter archives. When using such an archive, the MTW approach shows a certain potential for enriching the pool of possible analogues. Moreover, in a transient climate, the eventual nonstationarity of the link between predictors and precipitation might discard the relevance of analogues from the distant past

and can increase the relevance of using a more recent and shorter archive rather than a long one. In such cases, the archive inflation brought by the MTW is also relevant.

One can question the interest of using moving daily totals when, for example, 6-h precipitation series can be predicted instead. However, the 6-h time series generated by the AM might not accurately represent the intra-daily precipitation distribution or variability. In addition, sometimes a resolution finer than the daily time step is not needed, and another disaggregation technique may be used afterwards. Finally, when using a reconstructed precipitation archive, the errors in intra-daily precipitation distributions have a smaller impact on 24-h totals than on 6-h totals.

4.4 Reconstruction of a long precipitation archive suitable for the MTW

As discussed in Sect. 4.3, the skills of AMs improved with the length of the archive. It would then be even more profitable to apply the MTW to the longest archive possible rather than being limited to the period in which sub-daily precipitation data are available. Therefore, the idea is to reconstruct longer archives of moving 24-h totals from existing standard daily precipitation series. For this purpose, disaggregation techniques can be used. In this study, two simple disaggregation approaches of the daily precipitation time series were assessed.

The first technique is a proportional distribution in which the observed daily precipitations were considered to be constant during the measurement period of 06:00 h UTC to 06:00 h UTC the following day. The proportional parts of the original daily time series were allocated into a new moving average of 24 h totals (e.g. a 00:00 – 24:00 h UTC total would be made of ¼ of the standard precipitation of day one and ¾ of day two).

The second approach was used to obtain data that are closer to the chronology of the actual precipitation by relying on informative proxy variables during the reconstruction procedure. Data from NCEP/NCAR reanalysis 1 were used for this purpose. Precipitable water and relative humidity at 1000 hPa, 925 hPa, or 850 hPa were assessed at the four points surrounding the catchment. Precipitation time series from the reanalysis were not considered due to the presence of zeros, which may not always match the prediction; this can lead to an undefined temporal repartition. Time lapses from -12 h to +12 h between both series were introduced to consider the significant distance separating the weather stations and the reanalysis grid point. The best proxy variable, precipitable water, was identified through correlation analyses on non-zero values with the 6-h precipitation time series. When the best proxy was selected, its temporal profile was used to disaggregate the daily precipitation time series.

These two simple methods did not result in valuable outputs (results not shown). Indeed, the performance improvement brought by the MTW was lost due to the poor quality of the reconstructed precipitation archives. A slight improvement was obtained for the second method which relied on a proxy variable rather than the proportional distribution method. However, this improvement was relatively small, and most of the benefit of the MTW was lost. Another reanalysis dataset with more accurate moisture variables could produce better proxies.

These attempts to transpose the MTW on the (usually) longest meteorological archive highlight the importance of the actual rainfall chronology. The MTW is beneficial if the precipitation series are close to the observed values. In the absence of a precipitation series with accurate sub-daily chronology, the introduction of an MTW does not improve the precipitation prediction.

5 Conclusions

The AMs are most often based on a daily time step owing to the availability of long precipitation archives. However, it is unlikely that two analogue synoptic situations which evolve relatively quickly would match optimally at the same hours of the day. It is probable that better matches may be found at a different hour, which can change the selection of the analogue dates.

As shown in previous research (Finet et al., 2008), the introduction of an MTW enables better analogue situations in terms of atmospheric circulation to be obtained. Using recent reanalysis datasets, MTWs with 12-h, 6-h, and 3-h time steps were assessed in the present study. Improvement in the S1 criterion values increased with the rank of the analogue owing to the accumulation of better analogues within the predicted distributions. Moreover, the improvement in the S1 criterion increased further for smaller MTW time steps.

A seasonal effect was highlighted such that the MTW was more profitable for winter than for summer. A likely reason is that the diurnal cycle has a greater effect in summer than in winter, which results in better analogues at the same time of the day. The preference for the same hours in summer has been demonstrated, whereas a large part of the analogue situations were selected at a different time in winter.

The MTW improved the precipitation prediction, with the CRPS showing a global decrease for all stations for both the CP and VP. The prediction skill for the CP was almost always improved further by reducing the time step of the MTW, but the magnitude differed. The results for the VP showed the same global trends but had more variability. Moreover, the reliability of the considered AMs improved for the prediction of different threshold exceedances, especially with a 3-h MTW.

Improvements of the analogy criterion and the performance scores were both found to be higher for MTWs with a smaller time step of 3 h. A 3-h MTW provided eight times more candidate situations, although they were not fully independent. Because the MTW provides additional situations to the pool of possible analogues, it can be considered as an inflation of the archive, which can be interesting in several applications.

The improvement in the circulation analogy and the change in performance of the precipitation prediction was found to be more important for days with heavier precipitation. These days are generally related to more dynamic atmospheric situations and are less frequent in the archive. These situations have more specific circulation patterns that are evolving more rapidly. Therefore, an MTW was found to be of particular interest in this type of situation and is beneficial for the prediction of heavier precipitation events.

The importance of the quality of the precipitation archive was also demonstrated because simplistic reconstructions of sub-daily time series led to a loss of all the improvement brought by the MTW. The precipitation prediction improved only when the precipitation chronology was close to the accurate values. Attempts to reconstruct longer time series based on simplistic proportional distributions or by using meteorological variables from the NCEP/NCAR reanalysis 1 dataset as a proxy were not successful. Other reanalysis datasets with more accurate moisture variables could show better performance.

The use of the MTW relies partly on the availability of long precipitation series at a sub-daily time step and with high accuracy. The precipitation data archives of high temporal resolution have increased over time. Other possible sources of such archives is the establishment of precipitation reanalysis at a regional scale in addition to the use of reanalysis-driven regional

climate models or limited area models over a long period. Even though outputs from these models might be biased or not accurate enough, information regarding the timing of the precipitation events could be useful in disaggregating the station time series.

The introduction of the MTW improved the selection of synoptic analogues. Regardless of its impact on the prediction skill
5 for precipitation or the availability of a predictand time series with a sub-daily time step, this improvement has the potential for application to long meteorological archives. For example, when processing forecasts for a target day showing synoptic similarity with known situations from the past related to extreme weather, it can be used even if no precipitation archive is available. Thus, it is worth knowing whether the situation at hand had such an analogue in the past. Another possible application is quality assessment of the selection of analogues on a shorter period, where sub-daily precipitation data are available. Indeed,
10 if the selection of analogues with the MTW on a long period for a specific target day differs from the selection for the shorter period, a sub-optimal forecast could be identified. Finally, some other predictands might not need sub-daily total values but point observations such as hail or extreme wind gusts, which make them easier to use with the MTW.

The technique is very simple and easily applicable. Therefore, it should be considered for several applications in different contexts, such as operational forecasting or climate-related studies.

15 **Appendix A: Acronyms**

2Z   Name of the analogue method of atmospheric circulation

2Z-2MI Name of the analogue method composed of a first level on atmospheric circulation and a second level on a moisture index

AM   Analogue method

20 CP   Calibration period

CRPS   Continuous ranked probability score

ERA-20C European Center for Medium Range Weather Forecasting 20th century reanalysis

MERRA-2 Modern-Era Retrospective Analysis for Research and Applications, Version 2

MTW   Moving time window

25 NCAR   National Center for Atmospheric Research

NCEP   National Center for Environmental Prediction

RH850   Relative humidity at 850 hPa

RMSE   Root–mean–square error

S1   Teweles and Wobus (1954) score

TPW   Total precipitable water

VP   Validation period

Z1000   Geopotential height at 1000 hPa

5   Z500   Geopotential height at 500 hPa

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

Acknowledgements. The authors thank the Swiss Federal Office for Environment (FOEV), the Roads and Water Courses Service, Energy and Water Power Service of the Wallis Canton, and the Water, Land, and Sanitation Service of the Vaud Canton which financed the MINERVE (Modélisation des Intempéries de Nature Extrême des Rivières Valaisannes et de leurs Effets) project which started this research. Thanks are
10   also extended to Dominique Bérod for his support.

The fruitful collaboration with the Laboratoire d'Etude des Transferts en Hydrologie et Environnement of the Grenoble Institute of Technology (G-INP) was made possible by the Herbette Foundation.

NCEP reanalysis data were obtained from the Web site of the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at http://www.esrl.noaa.gov/psd/. Precipitation time series were provided by MeteoSwiss.
15   The authors would also like to acknowledge the work of anonymous reviewers who contributed to improving this manuscript.

References

Ben Daoud, A.: Améliorations et développements d'une méthode de prévision probabiliste des pluies par analogie., Ph.D. thesis, Université de Grenoble, 2010.

Ben Daoud, A., Sauquet, E., Lang, M., Obled, C., and Bontron, G.: La prévision des précipitations par recherche d'analogues: état de l'art et perspectives, in: Colloque SHF-191e CST - «Prévisions hydrométéorologiques», Lyon, 18-19 novembre 2008, 6, 2008.

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

Bliefernicht, J.: Probability forecasts of daily areal precipitation for small river basins, Ph.D. thesis, Universität Stuttgart, 2010.

Bontron, G.: 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., Ph.D. thesis, Institut National Polytechnique de Grenoble, 2004.

Bontron, G. and Obled, C.: L'adaptation probabiliste des prévisions météorologiques pour la prévision hydrologique, La Houille Blanche, 1, 23–28, doi:10.1051/lhb:200501002, 2005.

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

Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, Clim. Past, 12, 635–662, doi:10.5194/cp-12-635-2016, 2016.

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

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

Desaint, B., Nogues, P., Perret, C., and Garçon, R.: La prévision hydrométéorologique opérationnelle: l'expérience d'Electricité de France, in: Colloque SHF-191e CST - «Prévisions hydrométéorologiques», Lyon, 18-19 novembre 2008, p. 8 p., doi:10.1051/lhb/2009054, 2008.

Duband, D.: Reconnaissance dynamique de la forme des situations météorologiques. Application à la prévision quantitative des précipitations., Ph.D. thesis, Thèse de 3ème cycle de la faculté des sciences de Paris., 1970.

Finet, T., Marty, R., Zin, I., and Obled, C.: Developing and transferring the ANALOG approach for PQPF's from French OHMCV catchments to quick responding catchments in Venetia and Romania., Tech. rep., LTHE & CNRS, Grenoble, France, 2008.

García Hernández, J., Horton, P., Tobin, C., and Boillat, J.: MINERVE 2010: Prévision hydrométéorologique et gestion des crues sur le Rhône alpin., Wasser Energie Luft – Eau Energie Air, 4, 297–302, 2009.

Hamill, T. and Whitaker, J.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application, Mon. Weather Rev., 134, 3209–3229, doi:10.1175/mwr3237.1, 2006.

Hamill, T. M., Whitaker, J. S., and Mullen, S. L.: Reforecasts: An Important Dataset for Improving Weather Predictions, B. Am. Meteorol. Soc., 87, 33–46, doi:10.1175/BAMS-87-1-33, 2006.

Hamill, T. M., Scheuerer, M., and Bates, G. T.: Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses, Mon. Weather Rev., 143, 3300–3309, doi:10.1175/MWR-D-15-0004.1, 2015.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:dotcrp>2.0.co;2, 2000.

Horton, P.: Améliorations et optimisation globale de la méthode des analogues pour la prévision statistique des précipitations. Développement d'un outil de prévision et application opérationnelle au bassin du Rhône à l'amont du Léman, Thèse de doctorat, Université de Lausanne, https://tel.archives-ouvertes.fr/tel-01441762, 2012.

Horton, P., Jaboyedoff, M., Metzger, R., Obled, C., and Marty, R.: Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method, Nat. Hazard Earth Sys., 12, 777–784, doi:10.5194/nhess-12-777-2012, 2012.

Horton, P., Jaboyedoff, M., and Obled, C.: Global Optimization of an Analog Method by Means of Genetic Algorithms, Mon. Weather Rev., 145, 1275–1294, doi:10.1175/MWR-D-16-0093.1, 2017.

Junk, C., Delle Monache, L., Alessandrini, S., Cervone, G., and von Bremen, L.: Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble, Meteorol. Z., 24, 361–379, doi:10.1127/metz/2015/0659, 2015.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., and Woollen, J.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2, 1996.

Lorenz, E.: Empirical orthogonal functions and statistical weather prediction, Tech. rep., Massachusetts Institute of Technology, Department of Meteorology, Massachusetts Institute of Technology, Dept. of Meteorology, 1956.

Lorenz, E.: 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, 1969.

Maraun, D., Wetterhall, F., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, 1–34, doi:10.1029/2009RG000314, 2010.

Marty, R.: Prévision hydrologique d'ensemble adaptée aux bassins à crue rapide. Elaboration de prévisions probabilistes de précipitations à 12 et 24 h. Désagrégation horaire conditionnelle pour la modélisation hydrologique. Application à des bassins de la région Cév, Ph.D. thesis, Université de Grenoble, 2010.

Marty, R., Zin, I., Obled, C., Bontron, G., and Djerboua, A.: Toward real-time daily PQPF by an analog sorting approach: Application to flash-flood catchments, J. Appl. Meteorol. Clim., 51, 505–520, doi:10.1175/JAMC-D-11-011.1, 2012.

Matheson, J. and Winkler, R.: Scoring rules for continuous probability distributions, Manage. Sci., 22, 1087–1096, doi:10.1287/mnsc.22.10.1087, 1976.

Obled, C.: Daniel Duband - cinquante ans de contributions scientifiques à l'hydrologie (1962-2011), La Houille Blanche, pp. 55–68, doi:10.1051/lhb/2014017, 2014.

Obled, C., Bontron, G., and Garçon, R.: Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach, Atmos. Res., 63, 303–324, doi:10.1016/S0169-8095(02)00038-8, 2002.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J. N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An atmospheric reanalysis of the twentieth century, J. Climate, 29, 4083–4097, doi:10.1175/JCLI-D-15-0556.1, 2016.

Radanovics, S., Vidal, J.-P., Sauquet, E., Ben Daoud, A., and Bontron, G.: Optimising predictor domains for spatially coherent precipitation downscaling, Hydrol. Earth Syst. Sc., 17, 4189–4208, doi:10.5194/hess-17-4189-2013, 2013.

Raynaud, D., Hingray, B., Zin, I., Anquetin, S., Debionne, S., and Vautard, R.: Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb, Int. J. Climatol., doi:10.1002/joc.4844, 2016.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Clim., 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

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

Teweles, S. and Wobus, H. B.: Verification of prognostic charts, B. Am. Meteorol. Soc., 35, 455–463, 1954.

Van Den Dool, H. M.: Searching for analogues, how long must we wait?, Tellus, 46A, 314–324, 1994.

Zorita, E. and Storch, H. V.: The analog method as a simple statistical downscaling technique: comparison with more complicated methods, J. Climate, 12, 2474–2489, doi:10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2, 1999.
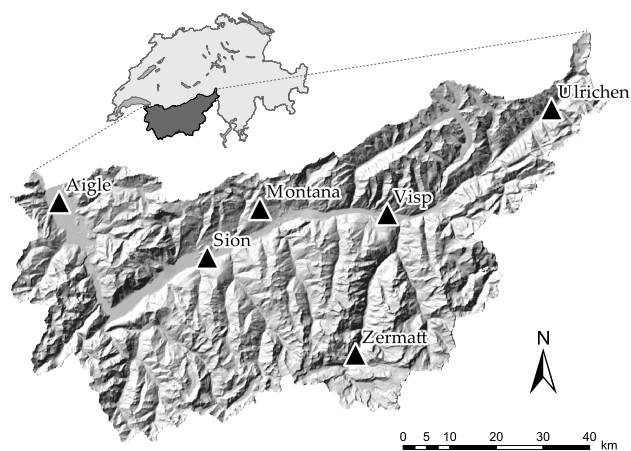
Figure 1. Position of the six weather stations of interest (Aigle, Montana, Sion, Ulrichen, Visp, and Zermatt) in the upper Rhône catchment in Switzerland.
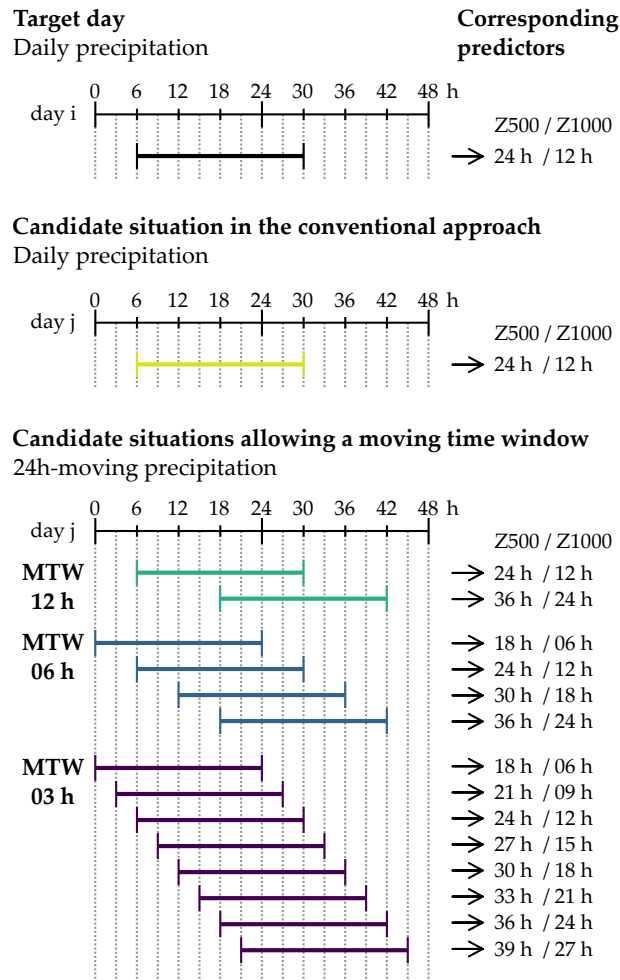
Figure 2. Illustration of the principle of a moving time window (MTW). The target situation is the same for the conventional approach and the MTW, although the candidate situations with the MTW are two, four, or eight times as many. The horizontal bars represent the 24h precipitation totals; their associated predictors are listed on the right-hand side.
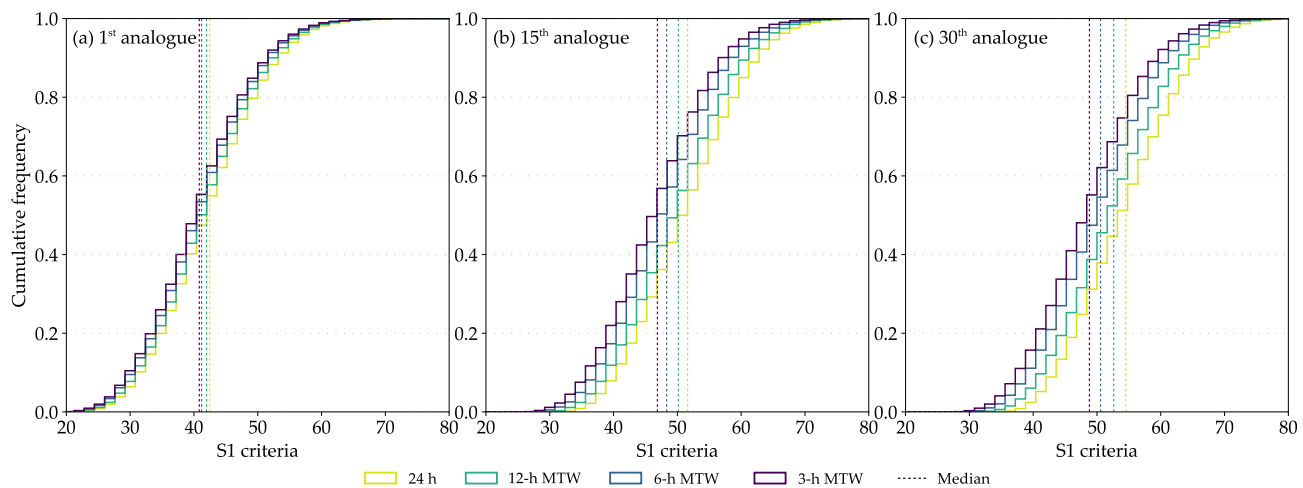
Figure 3. Changes in S1 criterion distributions owing to the introduction of an MTW with a 12-h, 6-h, or 3-h time step. Distributions are provided for (a) the 1st, (b) 15th, and (c) 30th analogue ranks for all days of the CP at the Ulrichen station.
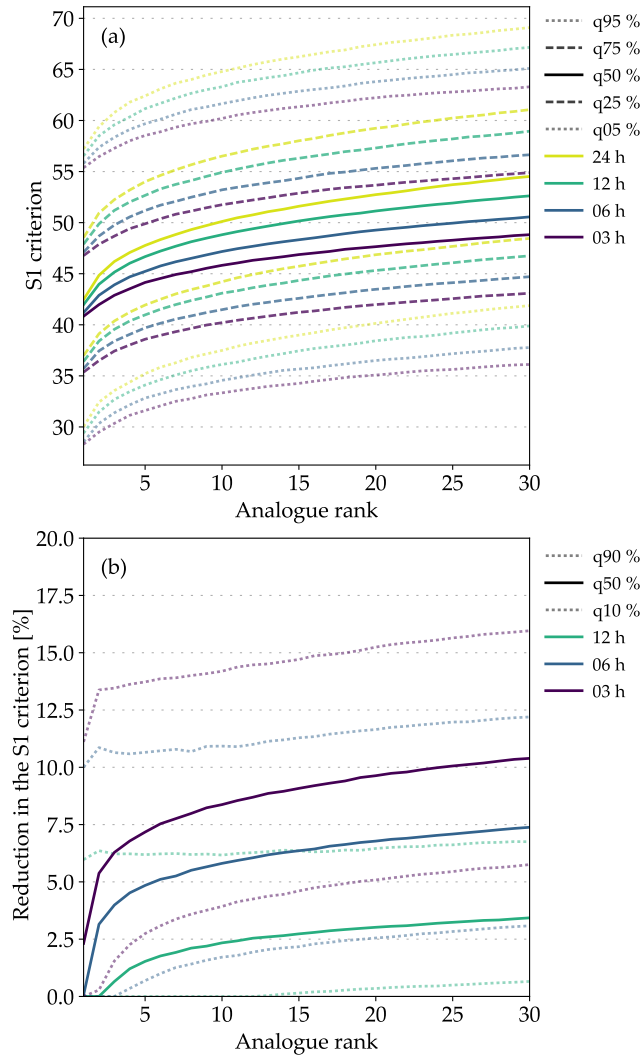
Figure 4. Synthesis of the changes in the S1 criterion owing to the MTW with 12-h, 6-h, and 3-h time steps for the Ulrichen station depending on the rank of the analogue. (a) Quantiles of the S1 distributions corresponding to the MTW with different time steps; 24-h is the conventional approach without MTW. (b) Quantiles of the relative improvements of the S1 criterion when using the MTW compared with the conventional approach.
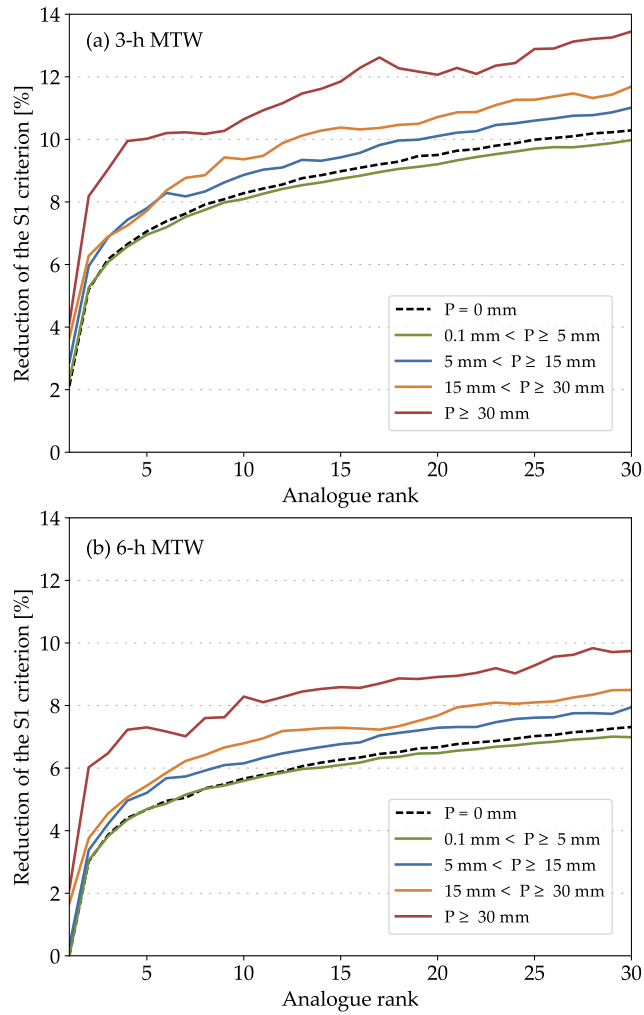
Figure 5. Distribution of the median improvements of the S1 criterion owing to the (a) 3-h MTW and (b) 6-h MTW, depending the on daily precipitation thresholds at the Ulrichen station.
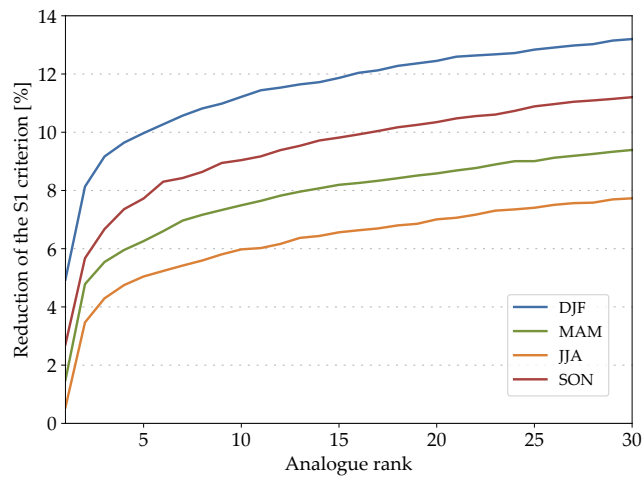
Figure 6. Seasonal effect on the median reduction of the S1 criterion for the Ulrichen station owing to the MTW. DJF: winter; MAM: spring; JJA: summer; SON: fall.
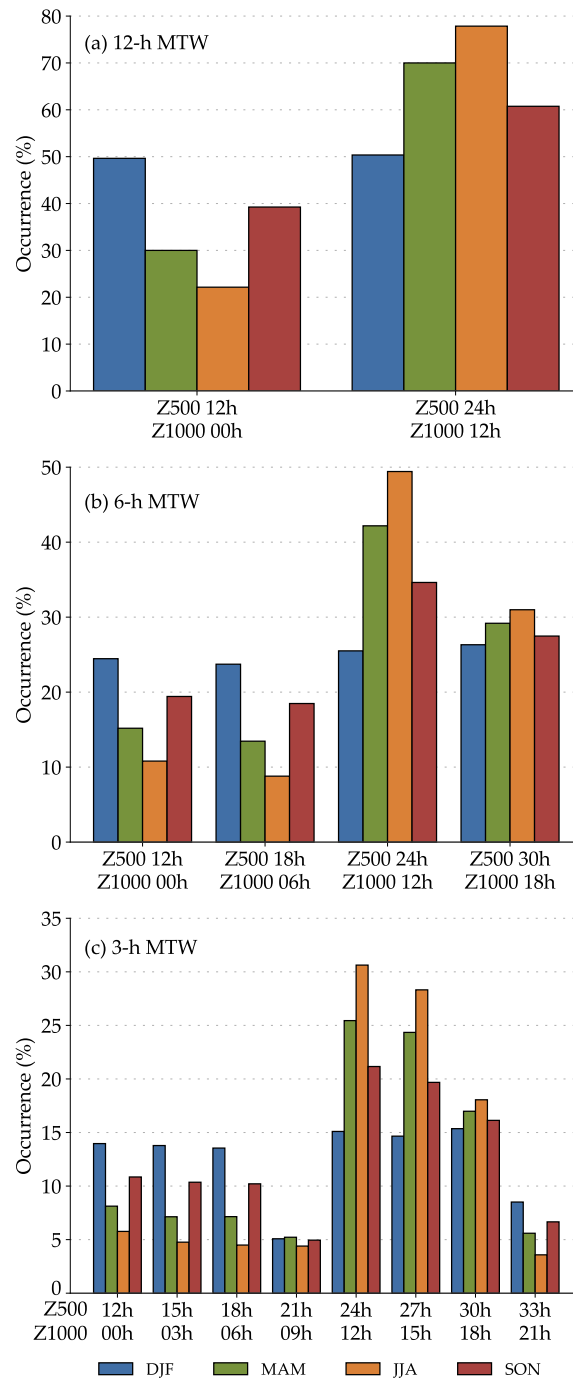
Figure 7. Distribution of the predictor hours on the selected analogue dates for the Ulrichen station when using a (a) 12-h MTW, (b) 6-h MTW, and (c) 3-h MTW, depending on the season.
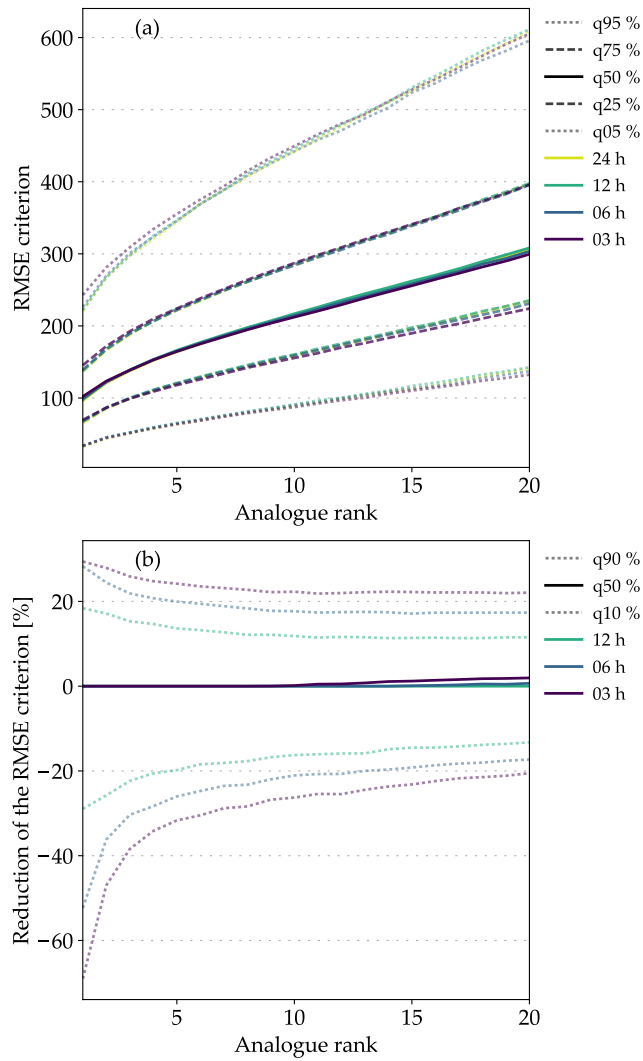
Figure 8. Synthesis of the changes in the RMSE criterion owing to the MTW for the Ulrichen station, depending on the rank of the analogue. (a) Quantiles of the RMSE distributions with and without the MTW. (b) Quantiles of the relative improvements of the RMSE criterion when using the MTW.
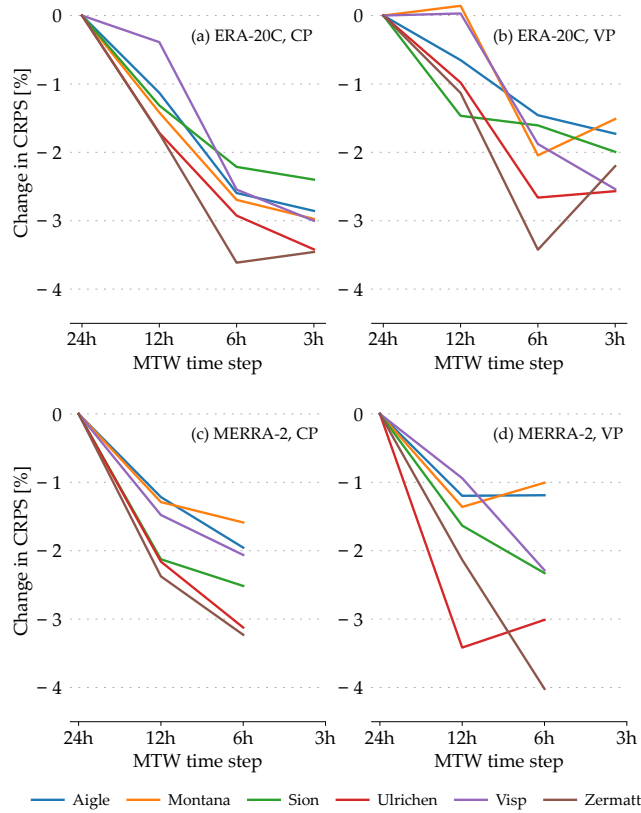
Figure 9. Changes in performance score (CRPS) of the 2Z method at the different stations for MTW of varying time steps relative to the conventional approach of 24 h. Results are provided for ERA-20C for the (a) CP and (b) VP as well as for MERRA-2 for the (c) CP and (d) VP. Lower CRPS values indicate better performance.
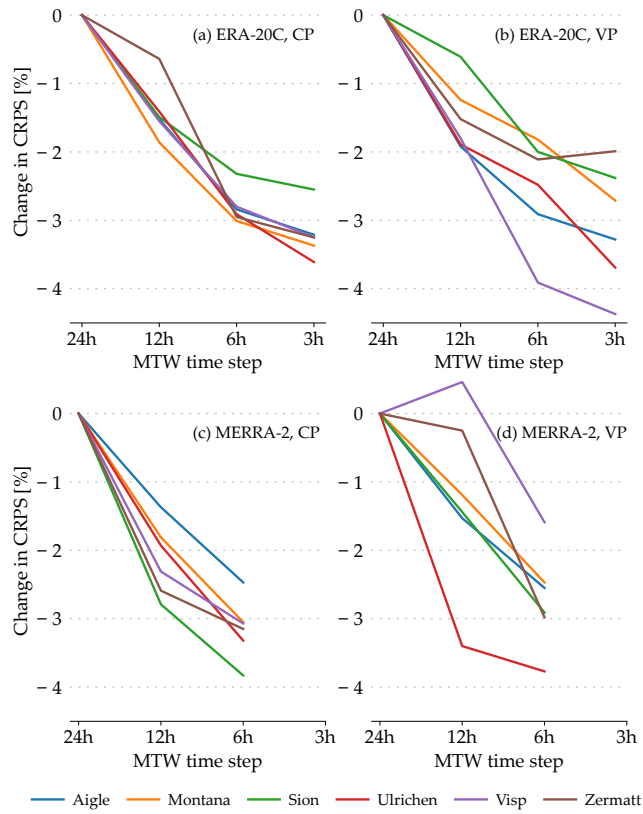
Figure 10. Same as Figure 9 but for the 2Z-2MI method.
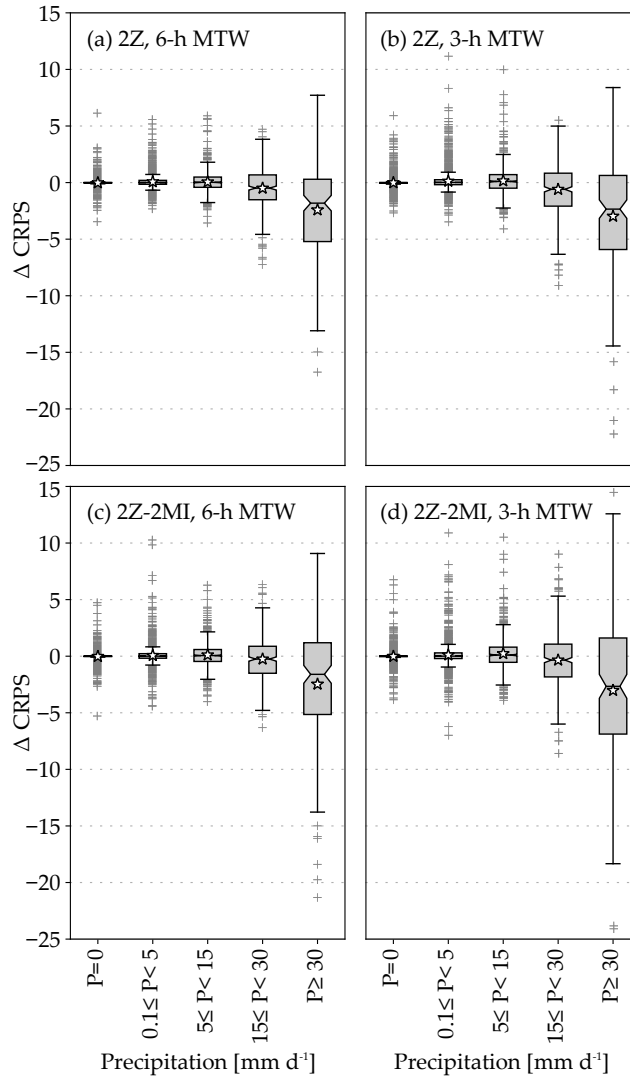
Figure 11. Differences in the CRPS values at the Ulrichen station owing to the introduction of the MTW as a function of different daily precipitation thresholds for the target date. The results are provided for the 2Z method (top) and the 2Z-2MI method (bottom) with a 6-h MTW (left) and a 3-h MTW (right). Stars represent averages. Lower CRPS values indicate better improvements.
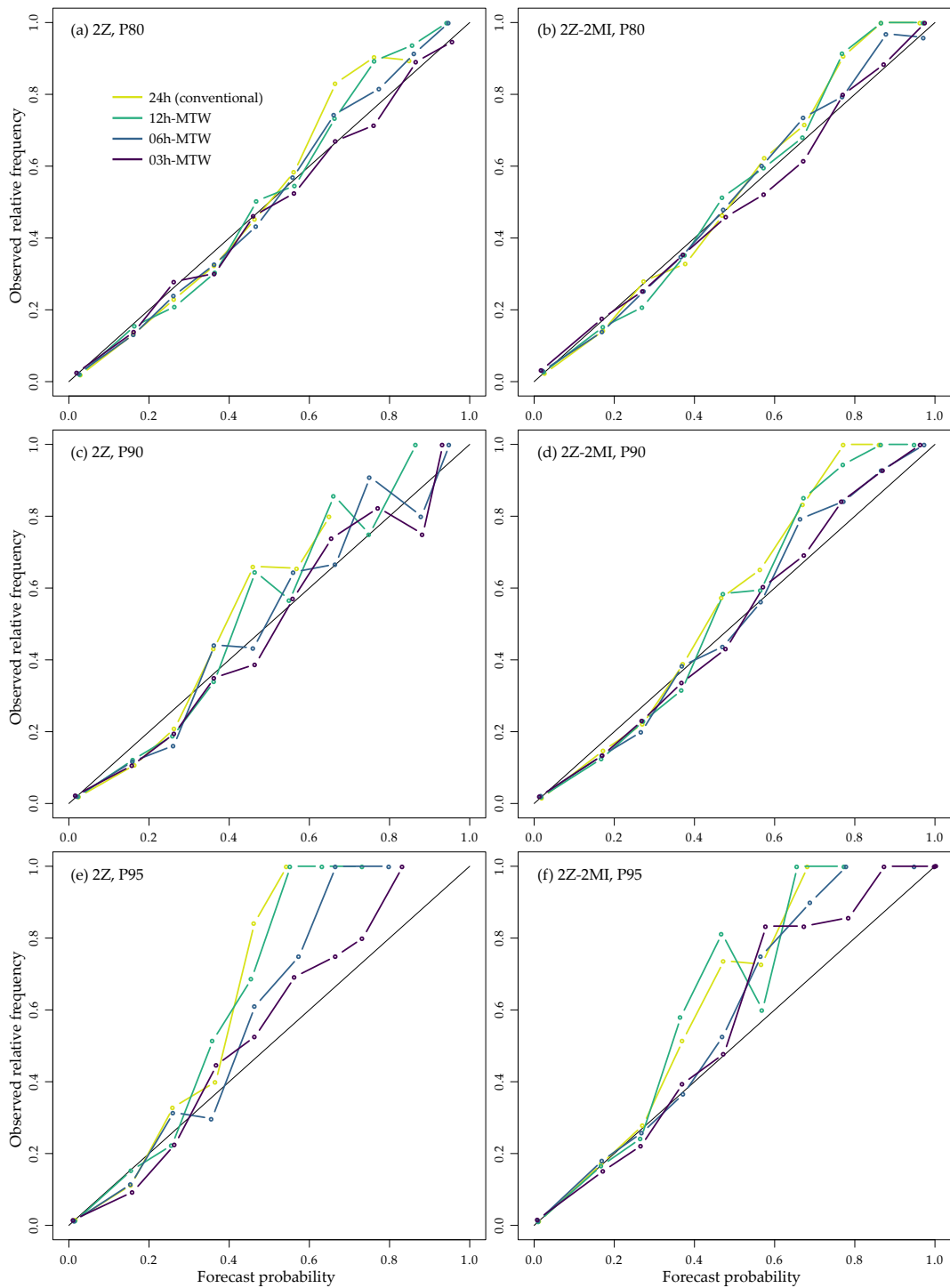
Figure 12. Reliability diagrams for the 2Z (left) and 2Z-2MI (right) methods and the prediction of the exceedance of percentiles 80 % (top), 90 % (middle), and 95 % (bottom) at the Ulrichen station for the VP. The conventional approach of 24 h is provided as well as the 12-h, 6-h, and 3-h MTWs.
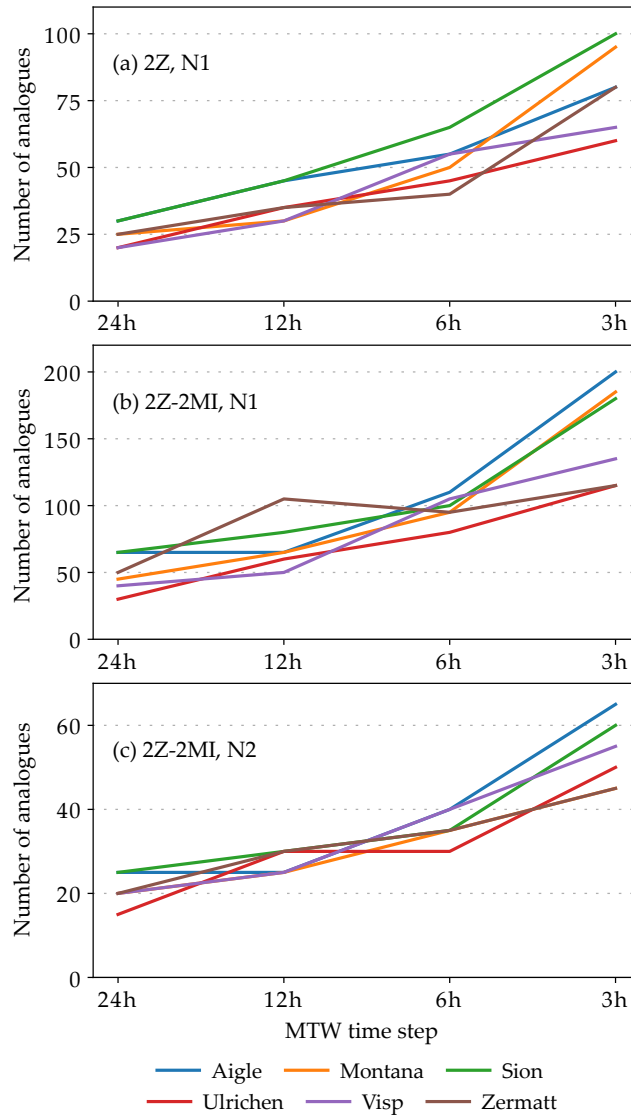
Figure 13. Optimal number of analogues of the first and second levels of analogy, N_1 and N_2, respectively, for the (a) 2Z and (b), (c) 2Z-2MI methods after recalibration using the MTW.
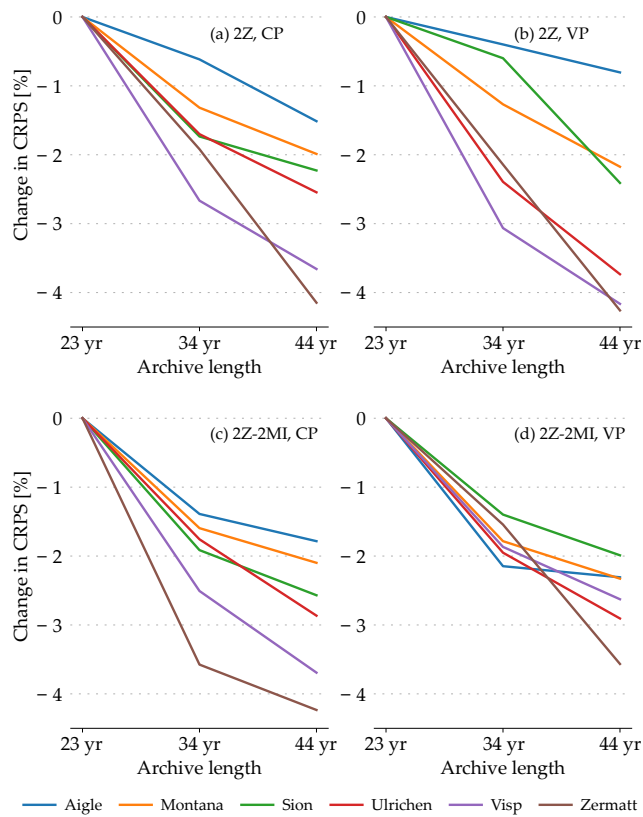
Figure 14. Change in performance score (CRPS) of the 2Z (top) and 2Z-2MI (bottom) methods at the different stations for the CP (left) and VP (right) for an increasing archive length with the conventional approach of 24-h window. Lower CRPS scores indicate better results.

Table 1. Parameters for the reference method on the atmospheric circulation (2Z). The first column shows the level of analogy (0 for preselection). The following columns shown the meteorological variable and its hour of observation, the criterion used for the current level of analogy, and the number of analogues.

Level	Variable	Hour	Criterion	Nb
0	± 60 days around the target date			
1	Z1000	12 h	S1	N_1
	Z500	24 h		

Table 2. Parameters of the reference method with moisture variables (2Z-2MI). The conventions are the same as those in Table 1

Level	Variable	Hour	Criterion	Nb
0	± 60 days around the target date			
1	Z1000	12 h	S1	N_1
	Z500	24 h		
2	TPW * RH850	12 h	RMSE	N_2
	TPW * RH850	24 h		