Reply to reviewer 1

Thanks to the reviewer for his positive review and his constructive comments.

Comment 1:

P1 L1: I would not use the term "prediction" as, in my point of view, the AMs are not forecasting methods by themselves, but rather adaptation methods, which link predictand to predictors (as it is well explained by the author himself p.7, l6-25).

Reply: The first sentence has been changed, as well as other uses of the term "prediction".

Comment 2:

P2. L7-8: "(…) one describing the situation (…)" you should specify that these are 'historical situations' that will be compared to the situation at hand.

Reply: Thanks for the suggestion; this was added.

Comment 3:

P3 L23-13 : I do not agree to the terms "partially independent forecast". As express before, the AMs are not forecasting methods. The forecasting capacity is due to the NWP. The AMs are adapatation methods that can enhance the forecasting skill of the NWP.

Reply: I do agree with reviewer 1, and this was the meaning of "partially". However, to avoid confusion, this has been changed to "statistical adaptation".

Comment 4:

P4 L20-21 : these results were obtained considering daily rainfall (for shorter time-step, we may assume that we could use shorter archives).

Reply: Thanks for the suggestion; it was added to the manuscript.

Comment 5:

P7 L31 : Indicate that the CRPSS score used in Fig. 1 is explained section 3.6.2. Indicate also what is the reference forecast used to compute the CRPSS

Reply: This has been added to the caption of Fig. 1.

Comment 6:

P15 equation (2) : I think the subscript i of H must be removed.

Reply: Correct, thanks for identifying this.

Comment 7:

P17. Section 4 : All this section is very interesting. Is it possible to add the computing time requested by the Monte-Carlo simulation, the sequential calibration and the GAs calibration, for each case? The comparison of these computing times with the obtained skills might be quite interesting.

Reply: The following paragraph was added: "In terms of processing resources, all experiments were done under similar conditions, i.e. using 16 cpus on a Linux cluster. For 2Z, the sequential calibration

took 7 min (time is expressed as wall clock time), Monte Carlo took 12.9 h (50,000 evaluations), and GAs took 11.6 h on average (41,000 evaluations on average). For 2Z-2MI, the sequential calibration took 12.5 min, Monte Carlo took 16.8 h, and GAs took 20.4 h on average (61,000 evaluations on average). The computation time should be taken into account in the choice of a calibration strategy."

Reply to reviewer 2

Thanks to the reviewer for his thoughtful and detailed comments. All technical corrections were addressed and will not be discussed here.

Comment 1:

page 2 line 30-33: The literature cited here is very old. Are analogue methods still used for the cited purposes? If this is the case, please replace with more recent literature, if not, I think this paragraph could be omitted. If you decide to keep it, please marque as historical use.

Reply: This literature has been updated.

Comment 2:

page 3 line 1-5: Please mention the predictands evaluated in the mentioned project.

Reply: This has been added (for daily precipitation).

Comment 3:

page 5 line 18: "This preselection is now often implemented as a moving selection..." Please add a reference.

Reply: Some references were added.

Comment 4:

page 5 line 28 and page 19 line 13-14: I don't understand the hours UTC here. Especially 24h UTC, that would be rather 00h UTC. What is the reasoning behind taking values at a specific time of day? Does the choice depend on the longitude? Or are those hours meant to be forecast lead-times or time ranges? In this case "UTC" doesn't make any sense. Please clarify.

Reply: The following sentence was added in parenthesis to explain the selection of hours: ''reference time of the predictors as they are usually available at a 6-hrly temporal resolution or higher''. 24h UTC means here at 00h UTC but the next day. The reason being that daily precipitation is usually measured between 6h UTC and 6h UTC the next day, and so the centre of the accumulation period is 18h UTC. A couple of predictors at 12h UTC and 24h UTC is then centred on the accumulation period.

Comment 5:

page 6 line 10: You state that the moisture index MI "does not represent an actual physical quantity, but expresses the water content and the degree of saturation". To me the water content and the degree of saturation are physical quantities. Please clarify.

Reply: This sentence has been removed.

Comment 6:

page 6 line 21: What does "close in distance but too dissimilar in pattern" mean? Doesn't a distance in PCA space measure dissimilarity in the contribution of different patterns?

<u>Reply</u>: This sentence has been removed as it is out of the scope of the paper anyway.

Comment 7:

page 6 line 26: Similarly, "an analogy of the atmospheric circulation instead of a Euclidean distance" is not clear to me.

<u>Reply</u>: The sentence has been changed to: '' S1 allows for a comparison of the gradients and thus an analogy of the atmospheric circulation instead of considering the actual values at the grid points''.

Comment 8:

page 6 line 27: Isn't the RMSE the same as the Euclidean distance in this case?

<u>Reply</u>: Yes. The sentence is now: ''For other predictors, classic criteria representing Euclidean distances between grid point values are used: …''.

Comment 9:

page 7 line 16-25: The term "temporal extrapolation" is confusing in this paragraph. It makes me think of techniques like kinematic extrapolation which are used for example in a nowcasting context. I understood that in the first part of the paragraph you talk about analogy of temporal trajectories and their limitations. In the second part NWP forecasts are used on the synoptic scale, but they are based on the numerical resolution of dynamic equations and not extrapolation. Please revise.

<u>Reply</u>: The paragraph has been edited and is now: '' In one of the very first uses in operational forecasting, radiosonde observations were used as predictors to predict precipitation for the next two days. However, because of the chaotic nature of the atmosphere, two analog situations quickly diverge over time (Lorenz, 1969). Thus, the AM has strong limitations regarding the analogy of temporal trajectories (Bontron, 2004). Given the superior capability of numerical models for simulating the dynamic evolution of the atmosphere, their outputs are now used as predictors for the coming days. …''

Comment 10:

page 7 line 22: You mention precipitation and temperature as examples for predictands that are difficult to simulate for numerical models. I'm not sure that temperature is a very good example here, given the performance of modern weather forecast models. What do you think? Under which circumstances and for which temporal scales an analogue forecast of temperature typically performs better than a numerical model?

<u>Reply</u>: That was not intentional, but the result of consecutive editions. The sentence has been changed to ''… with a local predictand, especially precipitation, which is more difficult to simulate for numerical models.''

Comment 11:

page 12 line 23: "Different authors" which ones?

<u>Reply</u>: They are actually listed at the end of the sentence: (Djerboua, 2001; Bontron, 2004; Marty, 2010)

Comment 12:

page 16 line 25: I think it would be useful to specify which kinds of objective functions are minimized and which ones are maximized. For example error functions are minimized and skill scores are maximized.

Reply: Thanks for the suggestion. This has been added.

Comment 13:

page 18: At some point I got a bit lost between "calibration periods", "optimisation periods", "archive periods" and "validation periods". Please define calibration period vs. optimisation period. In which cases within sample skill is measured, and in which cases out of sample skill is measured? Is the archive length always the same? In line 29: "The contrary is expected for the later period" please explain why.

Reply: The section has been edited and a paragraph was added in the beginning: '' For all analyses, the archive period is 1981-2010 and the results are shown for the evaluation period (EP) 2001-2010. For methods requiring a calibration, the calibration period (CP) is 1981-2000.'' The sentence "The contrary is expected for the later period" has been removed as it is not of primary importance. A visual caption has also been added to the figure.

Comment 14:

page 19 line 10: Why is this expected? Please specify.

Reply: The sentence has be rephrased: '' The parameters calibrated for the EP perform better than the ones established on the CP and assessed on the EP, which can be expected.''

Comment 15:

figure 4: This figure is unclear to me. Especially the meaning of the connections with different line types, arrows and points. I didn't look at the code, so the figure might be useful in the software documentation or user manual, but I don't see the purpose of this figure within the paper.

Reply: The figure has been removed.

Comment 16:

figure 5: What are the numbers in the circles? Please add the information in the caption.

Reply: The caption has been completed: The values on the map represent the 90th percentile (as selected on the left panel) of the precipitation values from the analogs at the different stations and for the selected lead time. The colour is proportional to the selected return period (10 years here).

Comment 17:

figure 7 and 8: Incomplete caption. What are the crosses?

Reply: This has been added for figure 7: "The thick blue lines represent the 90th, 60th, and 20th percentiles for the given lead times. The thin blue lines represent the equivalent time series but from previous forecasts. The small grey crosses represent all analog values and the larger crosses highlight the 10 best analogs (with a colour gradient from red for the best to yellow for the 10th)."

For figure 8: "The blue line represent the full distribution provided by all analogs, the circles are the 90th, 60th, and 20th percentiles, and the crosses correspond to the distribution provided by the 10 best analogs (with a colour gradient from red for the best to yellow for the 10th). The vertical red line is here the precipitation value for a 10 year return period."

Comment 18:

figure 9 caption: What does "optimised directly" mean?

Reply: "directly" has been removed.

Comment 19:

figure 11, 12 and 13: The axis annotations, legends and crosses are very small. Please increase their size.

Reply: This has been fixed.

AtmoSwing: Analog Technique Model for Statistical Weather forecastING and downscalING (v2.1.0)

Pascal Horton[1,2,3]

[1]University of Bern, Oeschger Centre for Climate Change Research, Institute of Geography, Bern, Switzerland
[2]University of Lausanne, Institute of Earth Sciences, Lausanne, Switzerland
[3]Terranum SARL, Bussigny, Switzerland

Correspondence: Pascal Horton (pascal.horton@giub.unibe.ch)

Abstract. Analog methods (AMs) ~~allow for the prediction of local meteorological variables of interest(predictand) such as the daily precipitation, on the basis of synoptic variables (predictors)~~ use synoptic scale predictors to search in the past for similar days to a target day in order to infer the predictand of interest, such as daily precipitation. They can rely on outputs of numerical weather prediction (NWP) models in the context of operational forecasting or outputs of climate models in the context of climate impact studies. AMs require low computing capacity and have demonstrated a useful potential for application in several contexts.

AtmoSwing is an open source software written in C++ that implements AMs in a flexible ~~manner~~ way so that different variants can be handled dynamically. It comprises four tools: a Forecaster ~~that performs operational forecasts~~ for use in operational forecasting, a Viewer ~~for displaying~~ to display the results, a Downscaler for climate studies, and an Optimizer ~~for inferring~~ to establish the relationship between ~~the predictand~~ predictands and predictors.

The Forecaster handles every required processing internally, such as ~~operational predictor~~ NWP outputs downloading (when possible) and reading, grid interpolation, etc., without external scripts or file conversion. The processing of a forecast ~~is extremely low-intensive in terms of computing infrastructure~~ requires low computing efforts and can even run on a Raspberry Pi computer. It provides valuable results, as revealed by a three-year-long operational forecast in the Swiss Alps.

The Viewer displays the forecasts in an interactive GIS environment ~~. It contains several layers~~ with several levels of syntheses and details~~in order to provide~~. This allows providing a quick overview of the potential critical ~~events~~ situations in the upcoming days, as well as the possibility for the user to delve into the details of the forecasted predictand and criteria distributions.

The Downscaler allows ~~the use of~~ using AMs in a climatic context, either for climate reconstruction or for climate change impact studies. When used for future climate studies, it is necessary to pay close attention to the selected predictors, so that they contain the climate change signal.

The Optimizer implements different optimization techniques, such as a semi-automatic sequential approach, Monte–Carlo ~~simulation~~simulations, and a global optimization technique using genetic algorithms. ~~The process of inferring~~ Establishing a statistical relationship between predictors and ~~predictand is quite intensive in terms of processing~~predictands is computationally intensive because it requires numerous assessments over decades. To this end, the ~~Optimizer~~ code was highly optimized ~~in~~

~~terms of~~ for computing efficiency, is parallelized ~~over multiple threads~~(using multiple threads) and scales well on a Linux cluster. This procedure is only required to ~~infer~~ establish the statistical relationship, which can then be used ~~in~~ for forecasting or downscaling at a low computing cost.

1 Introduction

Approaches based on the concept of analogy are widespread in different domains of science and engineering. In hydrometeorology, it entails retrieving data on atmospheric conditions from the past that can be considered as similar to the situation at hand, with consequences that may be expected to be similar. The consequences can be local variables of interest such as the occurrence of fog, favourable conditions for avalanches, wind intensity, or the precipitation amount. The approach relies on the idea expressed by Lorenz (1956, 1969), that similar situations in terms of atmospheric circulation are likely to lead to similar local weather. ~~The approach requires~~ AMs require at least two concurrent archives: one ~~describing the situation through different variables called predictors, and another one~~ that provides the value of the local variable of interest called the predictand, and another one describing the past atmospheric situations through different variables called predictors, to which the situation at hand will be compared.

Usually, the predictand values could be derived by modelling the chain of processes linking the predictors to the predictand. The processes involved range from large-scale dynamical states of the atmosphere down to very small-scale microphysical processes. These require models that are extremely complex, data-intensive, and time-consuming. Conversely, given an appropriate set of predictor archives, a sufficient number of situations analogous to a target situation ~~could~~ can be identified so that reasonable values ~~will~~ can be obtained for the predictand, ~~at a reasonable coding and computing-time cost~~with low computing effort. This is particularly true for a specific predictand that is critical in hydrometeorological applications, namely, the precipitation amount over a given domain and time duration. ~~Incidentally, the forecast is proposed~~ The forecast provided by AMs is issued as a statistical distribution based on the ~~values assumed by the predictand in the set of analogs selected~~ observed predictand values from the selected analogs, unless only the single best analog is considered, which ~~may not prove to be the most efficient~~ usually results in a lower skill (Bontron and Obled, 2005).

Analog methods (AMs) are used in two different types of approaches (Rummukainen, 1997): perfect prognosis, for which the statistical relationship is calibrated ~~based on~~ using observed predictors, and model output statistics (MOS), for which the relationship is calibrated ~~against the~~ using outputs of a specific climate or numerical weather prediction (NWP) model. AMs are often used to predict daily precipitation, either in an operational forecasting context (e.g. Guilbaud, 1997; Bontron and Obled, 2005; Hamill and Whitaker, 2006; Bliefernicht, 2010; Marty et al., 2012; Horton et al., 2012; Hamill et al., 2015; Ben Daoud et al., 2016) or a climate downscaling context (e.g. Zorita and von Storch, 1999; Wetterhall, 2005; Wetterhall et al., 2007; Matulla et al., 2007; Radanovics et al., 2013; Chardon et al., 2014; Dayon et al., 2015; Raynaud et al., 2016). Other predictands are also considered, such as precipitation radar images (Panziera et al., 2011; Foresti et al., 2015), temperature (Radinovic, 1975; Woodcock, 1980; Kruizinga and Murphy, 1983; Delle Monache et al., 2013; Caillouet et al., 2016; Raynaud et al., 2016), wind (Gordon, 1987; Delle Monache et al., 2013, 2011; Vanvyve et al., 2015; Alessandrini et al., 2015b; Junk

et al., 2015b, a), solar radiation or power production (Alessandrini et al., 2015a; Bessa et al., 2015; Raynaud et al., 2016), snow avalanches (Obled and Good, 1980; Bolognesi, 1993), and the trajectory of tropical cyclones (Keenan and Woodcock, 1981; Sievers et al., 2000; Fraedrich et al., 2003). ~~Guilbaud (1997) performed a literature review on the use of the AM in long-term forecasting and identified operational applications for monthly forecasts in many countries, including Canada (?), Hungary~~

5   ~~(?), the Netherlands (?), and England (?), as well as seasonal forecasts: ?, ? and ?~~AMs are also used for seasonal forecast (Barnston et al., 1994; Xavier and Goswami, 2007; Charles et al., 2012; Wu et al., 2012; Shao and Li, 2013).

An AM was evaluated during the project STARDEX (*STAtistical and Regional dynamical Downscaling of EXtremes for European regions*, see Goodess, 2003; STARDEX, 2005). One of the goals of the project was to compare various downscaling methods to determine weather extremes, and the AM was selected as being among the most useful ~~based on several techniques~~

10   techniques for daily precipitation (Maheras et al., 2005; Schmidli et al., 2007). Bliefernicht (2010) obtained superior results with ~~the AM~~ AMs than downscaling methods based on weather typing.

The use of ~~the~~ AMs for operational forecasting of daily precipitation originates in the work of Duband (1970, 1974, 1981). They were then designed for operational forecasting at EDF (Electricité de France) in order to better manage water resources and flood risks. They ~~have been~~ are used mainly by practitioners, notably hydropower companies (Desaint et al., 2008; Ben

15   Daoud et al., 2009; Obled, 2014) or flood forecasting services in France and Switzerland (Marty, 2010; García Hernández et al., 2009; Horton et al., 2012). When comparing the results from AMs to an ensemble forecast, Marty (2010) found AMs to be better than the considered ensemble, particularly ~~in the case of~~ for strong precipitation. However, AMs should not be considered as a substitute for NWP models, but as a complement in order to obtain a fast ~~and partially independent forecast~~ statistical adaptation that is known to be accurate several days in advance. Therefore, they contribute to the analysis of potentially critical

20   situations in flood forecasting, for example, and are very useful in early warning.

Hamill and Whitaker (2006) used an analogy-based approach on the GFS reforecasts in order to correct systematic errors in the ensemble forecasts of temperature and precipitation. These biases could be corrected by taking into account the intrinsic local climatology ~~from~~ provided by the AM. Moreover, the under-dispersion of the ensemble forecast from the numerical model has also been corrected using analogs (Hamill and Whitaker, 2006). Correction of ensemble forecast under-dispersion

25   using AMs is also ~~utilized~~ used operationally at EDF (Électricité de France).

The present work does not introduce a new method, but ~~instead, a software called~~ a software named AtmoSwing that implements AMs in a versatile and efficient way. It is versatile in that it facilitates the building of AM structures in a dynamic way~~with XML files~~, and because the code is written with an object-oriented architecture. It is efficient because it is written in C++ and leverages parallel computing. AtmoSwing is made up of different modules targeted either for operational forecasting

30   (the Forecaster and the Viewer) or for climate impact studies (the Downscaler). Additionally, a module (the Optimizer) is available for calibrating the different parameters ~~required for the method, namely the Optimizer~~of the method. AtmoSwing is continuously evolving and has been used in Horton et al. (2012, 2017a, b, 2018) and Horton and Brönnimann (2018).

Some existing AMs designed for daily precipitation will first be described along with the required data (Sect. 2) and the software will then be presented (Sect. 3) together with the details of the modules: the Forecaster (Sect. 3.3), the Viewer (Sect.

35   3.4), the Downscaler (Sect. 3.5), and the Optimizer (Sect. 3.6). ~~The conclusion~~ Section 4 discusses the parameters space of

AMs through different calibration techniques and Sect. 5 provides a feedback from operational precipitation forecasting in the Swiss Alps. Some limitations of the AM are discussed in Sect. 6. The conclusions (Sect. 7) ~~includes~~ provide some additional perspectives for future developments of AtmoSwing.

2 Data and methods

2.1 Required data

AMs generally require three datasets: the historical predictand values, the historical predictor values for the same period and the predictors describing the target situation.

The predictand is often a daily or 6-hrly time series. ~~It can have a higher temporal resolution such as 6 hours, but not higher than the time step of the predictors. The~~ One of the most used predictand is the daily precipitation, which is usually averaged over subregions in order to smooth local effects (Obled et al., 2002; Marty et al., 2012). These time series ~~are frequently~~ can be normalized by the precipitation value for a ~~return period of 10 years (Djerboua, 2001). This normalization allows~~ certain return period (for example 10 years, Djerboua, 2001) to allow for an easier comparison between subregions subject to different precipitation regimes~~, and thus to better identify the most important contributions~~.

In the early days ~~when the method was first used~~ of AMs in operational forecasting, the predictors were based on radio sounding data. ~~Currently~~ Nowadays, the predictors' archive is often a global atmospheric reanalysis dataset, which provides gridded large-scale variables at 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 on the locations where few or no observations are available, including variables that are not directly observed (Gelaro et al., 2017). Even though reanalyses are considered as very accurate in a data-rich region such as Europe, they can have a non-negligible impact on the skill of the ~~prediction~~ AMs, that can be even higher than the choice of the predictor variables ~~Dayon et al. (2015); Horton and Brönnimann (2018)~~ (Dayon et al., 2015; Horton and Brönnimann, 2018). AtmoSwing can read ~~ten~~ eleven different reanalyses (Table 1), and others can be easily added ~~due~~ thanks to the encapsulation of the dataset characteristics in the objects. ~~Users can find recommendations~~ Recommendations for the selection of a reanalysis can be found in Horton and Brönnimann (2018). Other predictor archives can also be used such as Sea Surface Temperature (SST, Reynolds et al., 2007). Bontron (2004) proposed that the minimum length of the archive should be 30 years for the prediction of daily precipitation under usual conditions, and 40 years or more for ~~intense events.~~ heavy rainfall. For smaller time-steps, shorter archives can be used (Horton et al., 2017b).

The predictors' dataset that describes the target situation varies according to the application of the AM. For operational forecasting (Sect. 3.3) they are outputs of NWP models such as the European Centre for Medium–Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) or the National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS, Kanamitsu et al., 1991; Kanamitsu, 1989). For climate impact studies (Sect. 3.5), they are outputs of general circulation models (GCMs) or regional climate models (RCMs), such as the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2012) and EURO-CORDEX (Jacob et al., 2014).

4

2.2 Analog methods for daily precipitation

AtmoSwing does not rely on a single ~~variant~~ structure of the AM, but ~~instead~~ can implement different variants. A non-exhaustive selection of methods developed for different regions will be presented hereafter, focusing on the prediction of daily precipitation. ~~Some of these are more specific for a certain region and may not be relevant to others. In addition, some~~
5 ~~perform better depending on the lead time.~~

2.2.1 Characteristics of the AM

Definition of the analogy – The AM is based on the principle that two similar synoptic situations may produce similar local effects ~~(Lorenz, 1956)~~(Lorenz, 1956, 1969). The perfect analogy does not exist, but sufficiently similar situations leading to similar effects can be identified. ~~Thus, two states of the atmosphere that are alike are called analogs (Lorenz, 1969).~~ To be
10 relevant, this analogy must be selected by optimizing the following elements:

– The meteorological variables (predictors) must contain synoptic scale information with a direct or indirect dependency with the target predictand.

– The pressure ~~level, or height,~~ (or isentropic) levels at which the ~~predictor is~~ predictors are selected.

– The spatial ~~window is the domain~~ windows are the domains over which predictors are compared. ~~The ideal size of this~~
15 ~~area is that which maximizes the useful information and minimizes noise.~~

– The temporal ~~window is the hour(s)~~ windows are the hours of the day ~~for~~ at which the predictors are considered when the time step of the predictors is smaller than the one of the predictand.

– The analogy ~~criterion required to compare the variables on the chosen spatial and temporal windows is a distance measure~~ criteria are distance measures used to rank ~~observed~~ past situations according to their degree of similarity with the target
20 situation.

– ~~Eventual~~ Possible weights between the predictors (e.g., Horton et al., 2017b; Junk et al., 2015b).

– The ~~optimal~~ number of analog situations N_i ~~for the~~ to retain for the analogy level i~~which is the best compromise to take~~ ~~into account local variability and maximize useful synoptic information~~.

Seasonal preselection – Lorenz (1969) restricted the search for analog situations to the same period of the year to cope with
25 seasonal effects. This preselection is now often implemented as a moving selection of ±60 days centred around the target date, for every year of the archive ~~(Table 2)~~(Table 2, Bontron, 2004; Marty et al., 2012; Horton et al., 2012; Ben Daoud et al., 2016). Alternatively, the candidate dates can be selected based on similar air temperature at the nearest grid point ~~(Table 2 Ben Daoud et al., 2016)~~ (Table 2, Ben Daoud et al., 2016).

Analogy of atmospheric circulation – A conditioning by variables describing the atmospheric circulation is present in a
30 vast majority of AMs. The geopotential field (Z) is often used as a predictor since Lorenz (1969), who based the analogy

on the levels 200, 500, and 850 hPa. Several pressure levels were later assessed by means of various criteria for the analogy based on the geopotential field (Duband, 1970, 1974, 1981; Guilbaud, 1997). It was ~~determined~~ found to be important to calculate the analogy for multiple pressure levels and different temporal windows (~~time of observation~~reference time of the predictors as they are usually available at a 6-hrly temporal resolution or higher) instead of a unique selection (Guilbaud and

5 Obled, 1998; Obled et al., 2002). Bontron (2004) showed that the choice of the temporal window ~~plays a greater significance compared to~~ can be more important than the choice of the atmospheric level for ~~the performance of the AM for daily precipitations~~daily precipitation (usually measured between 6 h UTC and 6 h UTC the following day). He concluded that the coupled geopotential heights at 1000 hPa (Z1000) at 12 h UTC and 500 hPa (Z500) at 24 h UTC provided the best performance ~~(for a subset of the NCEP/NCAR Reanalysis I; Kalnay et al., 1996; Kistler et al., 2001)~~ (for a subset of the NCEP/NCAR Reanalysis I – N

10 for the investigated regions in France (Table 2). The analogy ~~for~~ of the atmospheric circulation proposed by Bontron (2004) is still used operationally at the time of writing. Marty (2010) tested other temporal windows for intraday application on the basis of a more comprehensive reanalysis dataset and proposed to change the hours of observation to 06 h UTC and 18 h UTC. Horton et al. (2018) showed that a selection of four combinations of pressure levels and temporal windows instead of two for the geopotential height improves the ~~prediction skill (PC-4Z~~skill of the method (4Z, Table 2). The pressure levels and temporal

15 windows were automatically selected by genetic algorithms for the upper ~~Rhone~~ Rhône catchment in Switzerland.

Additional levels of analogy – Additional levels of analogy are subsequent steps that subsample a lower number of analog situations from the antecedent level of analogy, based on other variables. A second level of analogy was first introduced by Mandon (1985) ~~, Vallée (1986), and Gibergans-Báguena and Llasat (2007),~~ and Vallée (1986) based on wind, moisture variables, ~~stability indexes,~~ or temperature. Gibergans-Báguena and Llasat (2007) used the same kind of variables along with

20 stability indexes. After a systematic assessment of the variables provided by NR-1, Bontron (2004) noted that a moisture index (MI) based on the product of the relative humidity at 850 hPa (RH850) and the total precipitable water (TPW) ~~gave the best performance~~ provided the best skill (Table 2). ~~This index does not represent an actual physical quantity, but expresses the water content of the air column and its degree of saturation.~~ Marty (2010) selected the MI at 925 hPa instead of 850 hPa and also considered the moisture flux (MF) at 700 or 925 hPa (Table 2). The MF is the product of the MI with the wind intensity. Horton

25 et al. (2018) determined that the MI at 600 and 700 hPa were more useful than MF after the circulation analogy was applied to the four atmospheric levels (Table 2). Ben Daoud et al. (2016) also reconsidered the parameters of the MI and ended up with both 925 hPa and 700 hPa levels (Table 2). Subsequently, they added an additional level of analogy between the circulation and the moisture analogy (Table 2) based on the vertical velocity at 850 hPa (W850). This AM, termed "SANDHY" for Stepwise Analogue Downscaling method for Hydrology (Ben Daoud et al., 2016; Caillouet et al., 2016), was primarily developed for

30 large and relatively flat/lowland catchments in France (Saône, Seine).

Analogy criteria – In early applications of AMs, the geopotential height was condensed using principal component analysis (PCA) and the selection of analog situations was performed according to a Euclidean distance in the space of the PCA~~, which was eventually combined with a correlation criterion in order to remove days that are close in distance but too dissimilar in pattern~~. Guilbaud (1997) stopped using PCA to work directly with the raw data interpolated on grids, which resulted in an

35 improvement. In the case of ~~the~~ variables that describe atmospheric circulation, the Teweles–Wobus (S1) criterion (Eq. (1),

Teweles and Wobus, 1954; Drosdowsky and Zhang, 2003) was identified as the most suited criteria based on different studies (Wilson and Yacowar, 1980; Woodcock, 1980; Guilbaud and Obled, 1998; Bontron, 2004). S1 allows for a comparison of the gradients and thus an analogy of the atmospheric circulation instead of ~~a Euclidean distance~~considering the actual values at the grid points. For other predictors, ~~the~~ classic criteria representing ~~absolute distances~~ Euclidean distances between grid point

5   values are used: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), the latter being used most often.

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

where $\Delta \hat{z}_i$ is the gradient between the ith pair of adjacent points from the geopotential field of the forecasted target situation, and Δz_i is the corresponding observed geopotential gradient in the candidate situation. The differences are processed separately in both directions. The smaller the S1 values, the more similar the pressure fields. AtmoSwing allows processing real gradients

10   by taking into account the ~~horizontal~~ actual distance between points, or simple height differences by ignoring the horizontal distance. Under the latitudes of central Europe, the impact of neglecting the horizontal distance is small (not shown), but it can become more important at higher latitudes.

Other parameters – The predictors are compared ~~for~~ on a defined spatial window, which must be optimized to maximize the useful information and minimize noise. The spatial window is usually considered unique for all predictors of a level of

15   analogy. Using genetic algorithms, Horton et al. (2018) introduced different spatial windows between the pressure levels, which increased ~~performance~~the skill. Additionally, a weighting between the predictors was also successfully added instead of a simple equal-weights averaging. The number of analogs to select at each level of analogy should be optimized to be the best trade-off between taking into account local variability and maximizing useful synoptic information. It depends on the predictor dataset, the size of the spatial window and the length of the archive ~~Ruosteenoja (1988); Van Den Dool (1994)~~

20   (Ruosteenoja, 1988; Van Den Dool, 1994).

Probabilistic forecast – After the last level of analogy, the observed values of the predictand of interest (~~for example, daily precipitation amount~~here daily precipitation amounts) of the N_i resulting dates provide the empirical conditional distribution considered as the probabilistic forecast for the target day. The empirical frequencies are processed for every ~~value of the predictand~~ predictand value after classification, based on the Gringorten parameters (for a Gumbel or exponential law; see

25   Gringorten, 1963) and a probabilistic model can eventually be fitted (e.g. Gamma function, Obled et al., 2002). The forecast is finally often synthesized according to percentiles 20, 60 and 90 % (Guilbaud, 1997; Guilbaud and Obled, 1998).

Use in operational forecasting – In one of the very first uses in operational forecasting, ~~the forecast was performed based only on radiosonde observations and was temporally extrapolated to the two following~~ radiosonde observations were used as predictors to predict precipitation for the next two days. However, because of the chaotic nature of the atmosphere, two analog

30   situations quickly diverge over time (Lorenz, 1969). Thus, the AM has strong limitations regarding the ~~temporal extrapolation in operational forecasting~~ analogy of temporal trajectories (Bontron, 2004). Given the superior capability of numerical models for simulating the dynamic evolution of the atmosphere, their outputs are now used as ~~temporal extrapolation of the synoptic~~

~~variables~~predictors for the coming days. The search for analogy thus aims to connect the forecasted synoptic situation with a local predictand~~(temperature, precipitation, etc.),~~ especially precipitation, which is more difficult to simulate for numerical models. When using AMs in operational forecasting, it should be noted that some variables such as moisture or vertical velocity might not be accurately predicted after a lead time of a few days due to higher uncertainties. Predictors describing the

5   atmospheric circulation are generally considered to be more reliable.

2.2.2   Regional characteristics

The optimal predictors vary from one region to another, along with the leading atmospheric processes. Thus~~a unique version of the AM valid for any place on earth cannot be obtained. However~~, the method needs to be adapted to the local conditions, available data, and to the size of the ~~catchment~~ region of interest. ~~Thus, there will always be local adaptations to be made for use~~

10   ~~in a new region.~~ Even for two locations that are close to each other but subject to different critical atmospheric conditions, the selection of the best predictors can vary. This is illustrated in Fig. 1 for two subregions of the Rhône catchment in Switzerland. For both ~~regions~~subregions, all variables of NR-1 were assessed by optimizing the spatial window and the number of analogs for each one of them using the sequential calibration tool implemented in AtmoSwing (Sect. 3.6.4). The main similarities in the selection of the best predictor from ~~the NCEP/NCAR reanalysis I~~ NR-1 at both locations are: (1) the variables describing

15   the atmospheric circulation (pressure fields or geopotential heights) perform best, and (2) they are better when compared with the S1 criteria (asterisk in Fig. 1) instead of the RMSE. The main difference is that the pressure fields better explain the precipitation when they are considered close to the ground for the Chablais region, and at a higher altitude for the South-east crests. This is driven by the elevation of the stations and by the main atmospheric ~~influences related to precipitation events~~ drivers related to the precipitation at these locations.

20   The choice of the best predictors is likely to vary from one reanalysis dataset to another. This comprehensive comparison was not repeated with other datasets, because a selection of the best predictors using genetic algorithms would be less cumbersome (Sect. 3.6.5).

2.2.3   Method nomenclature

Variants of the AMs are numerous and it is not always easy to reference them in a short and descriptive way. In AtmoSwing,

25   a basic nomenclature is used (Fig. 2) in order to express the structure into a simple identifier. This cannot describe all the parameters of the AM, but quickly illustrates the structure of the ~~implementation~~method. This is particularly useful when working with a global optimization method, where nothing is fixed but the structure of the AM. This nomenclature has been used in Horton et al. (2017a, b, 2018) and Horton and Brönnimann (2018).

The naming contains different blocs (separated by a hyphen) for the various levels of analogy. It starts with the specification

30   of the preselection (P; can be omitted when comparing AMs with the same preselection approaches), which can be one of two types:

  – PC: calendar period (± 60 days around the target date)

8

- PT: based on air temperature (Ben Daoud, 2010)

Then, the following levels of analogy are listed, which may start with an optional A (for analogy). For every level of analogy, the number of variables used (combination of atmospheric levels and time of observation) is first provided, and then the short name of the variable is given (according e.g. to ECMWF conventions; in upper case), for example:

5
- Z: geopotential (circulation)

- TPW: total precipitable water

- RH: relative humidity

- V: wind velocity

- W: vertical velocity

10
- MI: moisture index (TPW * RH)

- MF: moisture flux (V * TPW * RH)

In order to keep the identifier simple, no value of atmospheric level or time of observation is specified. Moreover, the analogy criterion is not specified and is supposed to be S1 for Z and RMSE for the other variables. If anything changes from these conventions, it can be noted as a flag. The flag (lower case) can also provide other information, such as the optimization
15 method:

- sc : sequential calibration (can be omitted as considered as default, see Sect. 3.6)

- go (or just "o"): global optimization (by means of genetic algorithms for example)

This nomenclature can be adapted to specific needs or simplified for better readability (e.g. by removing the specification of the preselection). Examples can be found in Table 2.

20 ## 3 AtmoSwing

AtmoSwing is made of 4 main modules that are standalone, but do share a common code basis: the Forecaster for operational forecasting, the Viewer for displaying the forecast in a GIS environment, the Downscaler for climate applications, and the Optimizer that is used to ~~infer~~ establish the statistical relationship that defines the analogy for a given predictand ~~time series~~.
Separating the Forecaster and the Viewer allows for automation of the forecast on a server and the local display of the re-
25 sults. The Forecaster, the Downscaler and the Optimiser can be used either with a graphical user interface or a command-line interface.

3.1 Technical aspects

The code is written in object-oriented C++ and relies on the wxWidgets (Smart et al., 2006) library to provide a cross-platform native experience to users. CMake is used to build AtmoSwing ~~in MS~~ under Windows, Linux~~/ Unix~~, or Mac ~~(macOS)~~OSX. Developments have been partly performed using a test-driven development (TDD) approach. Continuous integration has been set up (on Travis CI and AppVeyor) so that a collection of more than 600 tests can be evaluated on the three operating systems every time new code is pushed to the server, to prevent regressions. Every analogy criterion, ~~prediction~~ performance score, searching and sorting functions, data manipulation, etc., ~~were~~ are tested. Some tests specific to the AM rely on the results of another analog sorting software developed at the Université Grenoble Alpes. They ensure that the results of AtmoSwing are exactly equivalent to this model, given the same parameters and data. The source code is under version control (Git) and is open source (on GitHub, www.atmoswing.org, Horton, 2018a). The GitHub organization page (https://github.com/atmoswing) also contains toolboxes to work with the outputs of AtmoSwing in R (Horton and Burkart, 2018) or Python (Horton, 2018b).

Although processing an analog ~~prediction~~ adaptation for a given target date is fast, numerous hindcasts over periods of several decades must be performed for calibration, which may become very time-consuming. Thus, great effort has been focused on minimizing the processing time using profiling tools. Firstly, all identified redundancies in the processing ~~are~~ were removed. Then, when searching for a certain date or data, the search ~~initially occurs~~ starts in the region where it is likely to be found instead of exploring an entire array. Similar data are not loaded twice, but instead shared pointers are used. Several other improvements allow reducing the computing time, for example the use of the quicksort method (Hoare, 1962) to sort the date vectors according to the analogy criterion. Different implementation variants were tested in order to select the most efficient approach: for example, when storing analog dates according to their criterion value, it is faster to insert them in a fixed-size array instead of storing them all and subsequently sorting the array. When using the S1 criteria, the gradients are pre-processed on the predictor data, so that they are only processed once. AtmoSwing also uses the linear algebra library Eigen 3 (Guennebaud et al., 2010) for calculations on vectors and matrices, which result in time-saving. Multi-threading is also implemented so that the search for analog situations in the archive is distributed among the available threads.

A user interface allows for the creation of the predictand database in the NetCDF format from text files. During the process, Gumbel adjustments are automatically calculated for precipitation data to determine the values corresponding to different return periods. The time series are normalized using a selected return period (default 10 years) and their square root can be processed. The final database file contains both the raw and the normalized series, as well as characteristics of the gauging stations and some metadata.

3.2 Modular approach and implementation

AtmoSwing's great strength is that it is designed to process the analog method in a modular fashion. The structure of the AM (number of analogy levels, number of predictors) is built dynamically (Fig. 3), and nothing is fixed a priori. The software then successively performs as many analogy levels as the user specifies, using all the predictors indicated. Each level of analogy results in an object containing target dates, analog dates, values of the analogy criteria, values of the predictand (at the final

stage), and other data. This object can be saved as a NetCDF file and/or can be injected into a new analogy level. The whole structure of the AM is defined through an XML file. Even the time step of the method (6 or 24 hours for example) is a dynamic parameter.

Each implementation of the AM (see Sect. 2.2) may enter this scheme, even if it consists of pre-processed variables (e.g.

5   moisture index). Various pre-processing functions are implemented as the calculation of the moisture index or flux, multiplication operations, or calculation of the gradients. The user can dynamically specify the pre-processing method and the predictors to use in the XML file.

This modular approach is implemented through object-oriented programming, as a direct consequence of polymorphism. This allows, for example, processing of a predictor object as a single interface to entities representing any reanalysis dataset.

10   Similarly, the criterion can be of different types, as well as the score for calibrating. ~~Figure ?? illustrates the main classes or objects involved in the core of the analog method processing in AtmoSwing in a simplified way.~~ The different types of objects that are instantiated are defined in the XML parameters file. Thus, there is a single implementation of the analog method capable of interacting with different ~~object types~~ types of objects in various contexts (calibration, forecasting, downscaling).

3.3 AtmoSwing Forecaster

15   The Forecaster module allows processing of operational forecasts. The software can be compiled with a graphical user interface (GUI), or without it to be used on a headless server through a command line interface (CLI). Processing a forecast requires very low computing capabilities and can be performed on a low-end computer. It successfully runs on a Raspberry Pi 3 (Model B).

To this day, the software can use the outputs of IFS or GFS (see Sect. 2.1). When ~~using GFS~~possible, it first downloads

20   the ~~predictor describing the target situation. It then linearly~~ relevant model outputs and interpolates the gridded data to match the resolution of the archive. The analogs dates are next extracted according to the selected AM variant and the predictand data are associated with the corresponding dates. The results are finally saved in auto-describing NetCDF files. If requested, a synthetic XML file is generated for easier integration on a web platform, for example. Every step of the forecast, from predictor downloading (when possible) to the final results, is performed in the software (and controlled through configuration), without

25   the use of external scripts (e.g. for data conversion).

Both the GUI and the CLI facilitate the processing of a forecast based on the most recent NWP outputs, or for a given date or period. When there is no new predictor data available, the forecast is not processed and computing resources are not ~~utilized~~consumed. The recommended use is thus to set up an automatic task on a server to trigger the forecast every 30 minutes. This would for example provide four forecasts a day.

30   Before being used in operational forecasting, the AMs were calibrated in a perfect prognosis framework, usually using a reanalysis dataset (Sect. 3.6). However, this does not take into account the uncertainty related to the forecast of the target situation by ~~Numerical Weather Prediction~~ NWP models. One might be willing to take into account this uncertainty, which increases with the lead time. A solution is to increase the number of analog situations with the lead time, which should be

11

optimized for every lead time on a forecast archive or a reforecast dataset (Thevenot, 2004). This technique is available in AtmoSwing, as the number of analogs can be specified for every lead time.

A meteorological variable that was proven as a good predictor in the perfect prognosis framework may eventually be poorly predicted by the selected NWP beyond a certain lead time. It should then be dropped after this lead time. For example, when using moisture variables for the second level of analogy, Thevenot (2004) showed that beyond a lead time of three days the AM with two levels did not perform better than the one with a single level of analogy. Datasets of reforecast from the selected NWP models allow assessing these aspects for different lead times.

3.4 AtmoSwing Viewer

AtmoSwing Viewer allows for the display of the files produced by the Forecaster in an interactive GIS environment (Fig. 4) ~~.~~ ~~It provides~~ with several levels of synthesis~~of the forecasts~~. It first provides an overview of possible alerts using colour codes on the lead time switcher (upper right in the GUI, see Fig. 4) that represent the worst case scenarios, or in the alarm panel (on the left side of the GUI). The alarm panel allows for a synthesis of the highest forecasted values for the different AMs and the different lead times. By default, the colours are expressed relative to the 10 year return period, for the 90th percentile (which can be changed in the preferences). This highest level of synthesis allows for quick identification of potentially critical situations in the days ahead.

Then, the user can explore the forecasts in more details, starting from the provided map (Fig. 4). The map displays the forecast of the selected AM variant (selected in the upper left panel) and the selected lead time (upper right). During the forecast, one AM might have parameters that differ by subregions, such as the number of analogs or the spatial windows. The Viewer automatically gathers the similar AM types and provides a composite view of the optimal forecasts per subregion. The user can, however, choose to display the results associated with a single parameter set for the entire region (by opening the tree view and selecting a child element), which provides a homogeneous set of analog dates. A display of all lead times on a single map is possible based on a symbolic representation on a circular band with a box for every lead time (Fig. 5). The number of boxes is adjusted to the number of lead times. This representation offers a global spatiotemporal visualization for a chosen AM.

Colour scales in the map can be adjusted by choosing (on the left part of the GUI) the predictand reference (raw value or ratio to different return periods) and the quantile of the distribution. Using a ratio to a certain return period eases the interpretation of the expected precipitation given that reference ~~volumes~~ values can drastically differ from one location to another, particularly in mountainous regions. All information relative to a rain gauge station (or catchment), such as its location, its name, or the values of different return periods, are stored in the forecast files to be displayed for end users who do not have the predictand database.

By clicking on a station on the map (or by selection from a dropdown list on the left), a new window appears with a plot of the forecasted time series (Fig. 6). By default, the plot contains the usual three considered percentiles (90th, 60th, and 20th), along with the 10 best analogs (crosses) with a colour code from yellow (tenth) to red (first). The 10-year return period value is also displayed ~~to set the perspective of the forecast~~as a red line. The user can choose to hide any data or to display

supplementary information (all analogs, all 10th percentiles, or all return periods) in the left panel. Traces of previous forecasts are also automatically loaded and displayed to provide information on the consistency of the forecasts.

The user can then delve into further details and display the predictand cumulative distribution for a given lead time (Fig. 7). This can inform if there is a shift between the distribution of all analogs versus the 10 best. Such a shift warns of a risk of
5 under/overestimation when considering the full distribution, particularly for high precipitation amounts. Indeed, the number of extreme precipitation events in the archive is limited and they are thus likely to be underrepresented in the selected analog dates. Different authors have shown that if the 60th percentile is best to forecast the occurrence and the amount of precipitation for common situations, the 90th percentile is a better indicator for strong to extreme events (Djerboua, 2001; Bontron, 2004; Marty, 2010). It is, therefore, necessary to pay close attention when the 90th percentile reaches high values, as this may be
10 indicative of possible extreme precipitations due the presence of several analog dates with high precipitation amounts in the distribution (Djerboua, 2001).

The distribution of the analogy criteria (not shown) can also be displayed to identify eventual discontinuities in the criteria values. Finally, one can display ~~a grid containing~~ the analog dates with the corresponding predictand and criteria values in an interactive spreadsheet (not shown).
15 AtmoSwing Viewer relies on workspaces defined in XML files to specify the path to the forecast directories and the GIS layers. It is thus easy to switch from a forecast for a region to another. Many GIS formats are supported thanks to GDAL (Geospatial Data Abstraction Library, GDAL Development Team, 2014). A user can have as many layers as desired and can control their display properties (colour, transparency).

3.5 AtmoSwing Downscaler

20 The Downscaler module is the last addition to AtmoSwing. Its purpose is to downscale either climate model outputs for climate impact studies or reanalyses for climate reconstruction of the past.

The Downscaler is able to read outputs of general circulation models (GCMs) or regional climate models (RCMs), such as the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2012) and EURO-CORDEX (Jacob et al., 2014), and can be extended to other datasets. CMIP5 and EURO-CORDEX are distributed in the NetCDF format, but present a great
25 variety of time steps, temporal references, spatial resolution, and file structures. A complete redesign of the management of the predictor data was necessary to provide the flexibility required to account for this variety. The Downscaler is thus able to parse these datasets original files by exploiting the self-descriptive capacity of NetCDF files.

The use of AMs in the context of future climate is rather new. Not all AMs can be used for this purpose, because some predictors might not capture the climate change signal well and the preservation of the relationship between predictors and
30 predictands must prevail. However, several authors have demonstrated the transferability of some AMs for future climate (Dayon et al., 2015, 2018; Raynaud, 2016; Turco et al., 2017). The transferability of an AM must be assessed before it is used in such a context.

AMs have also been used to perform climate reconstruction of the past (Caillouet et al., 2016, 2017; Bonnet et al., 2017). Such applications allow, for example, hydrological modelling of flood events in periods where no meteorological data are available, or analysis of past severe droughts.

3.6 AtmoSwing Optimizer

AtmoSwing Optimizer is a single tool that integrates different optimization methods, presented in Sect. 3.6.3 to 3.6.5. Its purpose is to ~~infer~~ establish the statistical relationship between the predictors and a predictand. The calibration framework is detailed in Sect. 3.6.1 and the implemented skill scores are listed in Sect. 3.6.2.

3.6.1 Calibration framework

The calibration of the AM is usually performed in a perfect prognosis (Klein et al., 1959) framework (Bontron, 2004; Ben Daoud, 2010). Perfect prognosis uses observed or reanalyzed data to calibrate the relationship between predictors and predictands, as opposed to the MOS approach that establishes the relationship based on model outputs. As a result, perfect prognosis yields relationships that are as close as possible to the natural links between predictors and predictands. However, there are no perfect models and even reanalysis data may contain ~~bias~~ biases that cannot be ignored (Dayon et al., 2015; Horton and Brönnimann, 2018). Thus, the considered predictors should be as robust as possible, i.e., they should have minimal ~~dependence~~ dependency on the model. With MOS approaches, reforecasts can be used to establish the relationship between predictors and predictands, provided that the archive is long enough. However, the calibration procedure must be performed every time a new version is available in order to reduce the bias (Wilson and Vallée, 2002).

A statistical relationship is ~~inferred~~ established with a trial and error approach by processing a forecast for every day of a calibration period. A certain number of days close to the target date are excluded to consider only independent candidate days. Validating the parameters of AMs on an independent validation period is very important to avoid over-parametrization and to ensure that the statistical relationship is valid for another period. In order to account for climate change and the evolution of measuring techniques, it is recommended that a noncontinuous period for validation should be used, distributed over the entire archive (Ben Daoud, 2010; Horton and Brönnimann, 2018). AtmoSwing's users can thus specify a list of the years to set apart for the validation that are removed from possible candidate situations. At the end of the optimization, the validation score is processed automatically.

3.6.2 Implemented performance scores

Multiple scores are implemented in AtmoSwing Optimizer and are listed hereafter. Details are only provided for the CRPS (Continuous Ranked Probability Score, Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000), which is most often used.

Discrete deterministic predictions - These are, for example, deterministic predictions of threshold exceedances. The continuous probabilistic nature of an ensemble of analogs can be transformed into a discrete prediction by considering a fixed

percentile from the distribution, which is compared to a threshold exceedance of the predictand. On the basis of a contingency table (Wilks, 2006), multiple scores can be processed with AtmoSwing:

- Proportion correct (Finley, 1884)

- Threat Score (Gilbert, 1884)

- Bias

- False Alarm Ratio

- Hit Rate or Probability of Detection

- False Alarm Rate

- Heidke Skill Score (Heidke, 1926)

- Peirce Skill Score (Peirce, 1884)

- Gilbert Skill Score or Equitable Threat Score (Gilbert, 1884)

Continuous deterministic predictions - These types of predictions must be evaluated using distance measures. For AMs, the provided distribution is summarized by a chosen percentile, which is compared to the predictand value. Available scores are as follows:

- Mean Absolute Error

- Root Mean Squared Error

Discrete probabilistic predictions – ~~In this instance~~Here, the probability of occurrence or the probability of belonging to a certain category is considered. The implemented scores are as follows:

- Brier Score (Brier, 1950)

- ROC diagram (Relative Operating Characteristic or Receiver Operating Characteristic, Mason, 1982)

- RPS (Ranked Probability Score, Epstein, 1969)

- SEEPS (Stable Equitable Error in Probability Space, Rodwell et al., 2010, 2011)

Continuous probabilistic predictions - These types of predictions are issued in the form of the expected statistical distribution for a variable, which needs to be compared to an observed value. This is the situation encountered when using multiple analogs from AMs.

Most assessment of AMs performance use the CRPS (Continuous Ranked Probability Score, Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000). It allows for evaluation of the predicted cumulative distribution functions $F(y)$, for example, the precipitation values y associated with the analog situations, compared to the single observed value y^0 for a day i:

$$CRPS_i = \int\limits_0^{+\infty} \left[F_i(y) - H_i(y - y_i^0) \right]^2 dy \tag{2}$$

5 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. This score is now commonly used for the evaluation of continuous variable prediction systems (Casati et al., 2008; Marty, 2010). It can be decomposed into several indicators also implemented into AtmoSwing Optimizer, such as: reliability – resolution / uncertainty (Hersbach, 2000), or sharpness – accuracy (Bontron, 2004).

Its skill score expression is often used, with the climatological distribution of precipitation as the reference. The CRPSS
10 (*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.

Finally, the rank diagram (Talagrand et al., 1997) and its accuracy as defined by Candille and Talagrand (2005) are also
15 available.

3.6.3 The sequential calibration

The calibration procedure that we call "sequential" or "classic" was elaborated upon by Bontron (2004) (see also Radanovics et al., 2013; Ben Daoud et al., 2016). It is a semi-automatic procedure that optimizes the spatial windows in which the predictors are compared and the number of analogs for every level of analogy. The different analogy levels (e.g. the atmospheric
20 circulation or moisture index) are calibrated sequentially. The procedure consists of the following steps (Bontron, 2004):

1. Manual selection of the following parameters:

    (a) Meteorological variable

    (b) Pressure level

    (c) Temporal window (hour of the day)

25   (d) Number of analogs

2. For every level of analogy:

16

(a) Identification of the most skilled unitary cell (~~1 point for moisture variables and 4 for~~ four points for the geopotential height when using the S1 criteria and one point otherwise) of the predictor data over a large domain. Every point or cell of the full domain is ~~jointly~~ assessed based on the predictors of the current level of analogy.

(b) From this most skilled cell, the spatial window is expanded by successive iterations in the direction of the largest performance gain until no further improvement is possible.

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

3. 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 analogs for the next level of analogy, N_2, is initiated at a chosen value. The~~ procedure starts again from step 2 (calibration of the spatial window and the number of analogs) for the new level. The parameters calibrated for the previous analogy levels are fixed and do not change.

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

The calibration is performed in successive steps for a limited number of parameters with the aim of minimizing ~~/maximizing the chosen objective function~~error functions or maximizing skill scores. Except for the number of analogs, previously calibrated parameters are generally not reassessed. The benefit of this method is that it is relatively fast, it provides acceptable results, and it has low computing requirements.

Small improvements were incorporated into this method in AtmoSwing Optimizer, then termed as "classic+", by allowing the spatial windows to perform other moves, such as: (1) increase in 2 simultaneous directions, (2) decrease in 1 or 2 simultaneous directions, (3) expansion or contraction (in every direction), (4) shift of the window (without resizing) in 8 directions (including diagonals), and finally (5) all the moves described above, but with a factor of 2, 3, or more. For example, an increase by two grid points in one (or more) direction is assessed. This allows skipping one size that may not be optimal. These supplementary steps often result in spatial windows that are slightly different. The performance gain is rather marginal for reanalyses with a low resolution such as NR-1, but might be more consistent for reanalyses with higher resolutions due to the presence of more local minima.

3.6.4 Variables exploration

The sequential calibration can also be used to explore the variables of a dataset. A list of variables, pressure levels, and temporal windows can be provided and all combinations are assessed through the classic(+) calibration. This functionality facilitates a comparison between the different variables of a dataset while considering the effect of the pressure level and the temporal window. Using this approach, only one variable is assessed at a time, but multiple levels of analogy are possible. Figure 1 results from such an analysis of the NR-1 reanalysis.

3.6.5 Global optimization

The sequential calibration has strong limitations: (i) it cannot automatically choose the pressure levels and temporal windows (hour of the day) for a given meteorological variable, (ii) it cannot handle dependencies between parameters, and (iii) it cannot easily handle new degrees of freedom. For this reason, genetic algorithms (GAs) were implemented in AtmoSwing Optimizer to perform a global optimization of AMs. This allows for optimization of all parameters jointly in a fully automatic and objective way. The method is described in Horton et al. (2017a) and an application is provided in Horton et al. (2018).

3.6.6 Monte–Carlo simulations

A Monte–Carlo analysis is also implemented in AtmoSwing. The procedure performs thousands of assessments of random parameters within given ranges. This method is not efficient for finding the best parameters set, but facilitates a better understanding of the sensitivity of the parameters. Its relevance is however limited for AMs with multiple levels of analogy and variables. Indeed, for methods with a high number of parameters with wide authorized value ranges, the probability is too low to obtain an acceptable configuration, and thus the resulting response surface might not be representative of the actual distribution of optimal values (See examples in Sect. 4).

4 Parameters space of AMs

An analysis of the parameters resulting from Monte–Carlo simulations, the sequential calibration, and GAs was performed for the Binn station in Switzerland (Fig. 1) with ERA-INT (Table 1). ~~Binn is characterized by high precipitation totals and heavy rainfall in this region , and on several occasions was~~ High precipitation in the Binn region were responsible for large damages downstream ~~. For this reason, it is~~ on several occasions, making it a station of particular interest. The results for this station cannot be generalized to all stations, but similar conclusions can be drawn for other locations. Moreover, the parameters space at a single station is expected to be ~~significantly more irregular~~ less regular than averaged regional ~~precipitations.~~

~~The analysis was first performed for the 2Z method (Table 2) for the period 2001–2010 for the target, and~~ precipitation. For all analyses, the archive period is 1981–2010 ~~as the archive. A relatively short period of 10 years was chosen~~ and the results are shown for the evaluation period (EP) 2001–2010. For methods requiring a calibration, the calibration period (CP) is 1981–2000. These periods were chosen relatively short to allow for 50,000 Monte–Carlo simulations. ~~The plots in Fig. 8 are truncated at the best 25th percentile~~

The analysis was first performed for the 2Z method (Table 2). The Monte–Carlo simulations (Fig. 8) show that the spatial window for Z needs to cover a certain region, but can be larger than a critical size. The extent of the spatial window can thus be substantially different without significantly affecting the ~~performance~~ skill score. This was also observed by Bontron (2004), who noted that "*performance slowly decreases if we consider a window that is slightly too large, while using windows that are too small results in strong performance loss*". The dilution of the relevant synoptic information therefore does not necessarily have a significant negative impact on ~~performance~~ the skill, while ignoring some of this information leads to stronger losses of

~~performance~~skill. The issue of equifinality related to the spatial windows is discussed in Radanovics et al. (2013). The station is usually contained within the optimal spatial domain, provided the predictors are ~~considered on~~ selected for the same day as the predictand. The optimal number of analogs is relatively well defined, although the selection of more analog candidates is possible without a strong penalty in terms of ~~performance~~skill.

5   The results of the sequential calibration are also illustrated in Fig. 8 (with squares). The calibration was first performed for the ~~period 1981–2000~~ CP and applied to ~~2001–2010~~ EP (blue squares)~~but also calibrated directly for the 2001–2010 period~~, but was also achieved directly on the EP (red squares). Here, the parameters established on ~~a different calibration period~~ the CP rather than the EP provided slightly better results when assessed on the EP. This is due to the limitations of the sequential calibration that can easily be trapped in a local optima. Indeed, the resulting spatial windows are small in this case, and the

10   algorithm stops as soon as an increase of the domain does not improve the score. This might not be an issue with a low-resolution reanalysis such as NR-1 (2.5°; Table 1), but this might become more of an issue with higher resolutions, such as ERA-INT used here (0.75°), because local minima are more frequent. In this case, the classic+ approach (Sect 3.6.3) might be relevant, but is more time consuming. The Monte-Carlo analysis yielded some better parameter sets than the sequential calibration, due to the constraint on the latter to have the same spatial window for both pressure levels.

15   Fourteen optimizations by GAs were performed for the same setup (seven optimizations using the ~~2001–2010 period as validation~~ classical CP/EP setting (blue triangles) and seven optimizations using ~~this period~~ the EP as the calibration period directly (red triangles)). The optimization with GAs was given the same degrees of freedom as the Monte–Carlo simulations, so no weighting of the pressure levels was considered (as in Horton et al., 2018). Thus, the parameters optimized for the ~~2001–2010 period~~ EP (red) could have been found randomly using the Monte–Carlo simulations. However, this did not oc-

20   cur due to the low probability of obtaining this combination. GAs also result in more skilful parameters than the sequential calibration. When optimized for the ~~2001–2010 period~~ EP (red triangles), the parameters yielded results that outperform the optimization for the ~~1981–2000 period~~ CP (blue triangles) ~~. However, the contrary is expected for the later period~~when assessed on the EP, as it can be expected. Most optimizations converge to a narrow range of values, supposedly, the global optimum for the respective period. The main difference compared to the sequential calibration is that the spatial windows are substantially

25   larger, mainly for Z500, and they differ between pressure levels.

Monte–Carlo simulations were also performed for 2Z-2MI (Table 2) for the same periods and for the same station. Figure 9 shows that the Monte–Carlo simulations could not properly use the moisture variables of the second level of analogy. The boxplots for the second level of analogy show an indifference of the location and the size of the spatial windows, which is demonstrated to be wrong ~~, based on the other optimization methods~~by the sequential calibration and the GAs. Moreover, the

30   achieved CRPS here based on random parameters is not better than the ~~case~~ method without the moisture variables (Fig. 8). Additionally, the number of analogs corresponding to the best CRPS values are similar between the two levels of analogy, which means that the second level is simply discarded. There are too many parameters with acceptable ranges that are too narrow to obtain meaningful parameters randomly. Monte–Carlo simulations with uniform probability laws is not suited for even moderately complex AMs.

The sequential calibration results in small spatial windows, especially for moisture variables. The differences with the 2Z methods for the first level of analogy are due to the different initial number of analogs, which has an influence on the choice of the spatial windows. The ~~optimized parameters for the 2001–2010 period~~ parameters calibrated for the EP perform better than the ones established ~~based on the other calibration period~~on the CP and assessed on the EP, which can be expected.

The results of the GAs show more variability than previously, which is likely due to a higher difficulty related to the larger number of parameters that have to be optimized, and to the presence of potentially more correlated information. The choice of the spatial windows for the moisture index at 12 h UTC is similar between the different optimization techniques and is a small line of zonally extended points. The chosen spatial windows by GAs for the moisture index at 24 h UTC is surprisingly large. This is likely due to the search of the GAs for additional information at a more distant location due to highly correlated data between 12 h UTC and 24 h UTC at the same 850 hPa level. The lack of convergence for this second spatial window means that the use of this variable is likely not optimal, and it would probably add more information considered at another pressure level, which was shown in Horton et al. (2018). The analysis of the convergence of multiple GA optimizations can thus be useful in interpreting the results and in identifying potentially suboptimal structures or variables.

The former results present a relatively noisy signal for the different optimization methods or the Monte–Carlo simulations. This may be due (1) to the fact that we consider a station's time series instead of regional ones ~~, which could decrease some~~ (related to variability from small-scale patterns in the precipitation fields), and (2) because we consider a short period for calibration. Despite the high number of simulations, Monte–Carlo simulations with a uniform probability law are not appropriate for even moderately complex AMs. It is likely that using a Gaussian probability law centered on the station (for the spatial windows) would be more appropriate.

In terms of processing resources, all experiments were done under similar conditions, i.e. using 16 cpus on a Linux cluster. For 2Z, the sequential calibration took 7 min (time is expressed as wall clock time), Monte Carlo took 12.9 h (50,000 evaluations), and GAs took 11.6 h on average (41,000 evaluations on average). For 2Z-2MI, the sequential calibration took 12.5 min, Monte Carlo took 16.8 h, and GAs took 20.4 h on average (61,000 evaluations on average). The computation time should be taken into account in the choice of a calibration strategy.

5   Feedback from operational forecast

AtmoSwing Forecaster has been issuing operational forecasts since 2012 for the upper Rhône catchment in Switzerland (Fig. 1) in the context of a flood management project (García Hernández et al., 2009). First, the 2Z and 2Z-2MI methods were implemented using NR-1 as the archive and GFS outputs to describe the target situation (Horton, 2012). Two more recent methods that were optimized by genetic algorithms (Horton et al., 2018) were also implemented since 2016. These methods were found to provide better results both in the perfect prognosis context and in the operational forecast. The results of the forecasts are provided for the Binn station (as in Sect. 4) for the 4Zo method with a lead time of three days (forecast issued three days before the target day; Fig. 10) and the 4Zo-2MIo method with a lead time of one day (Fig. 11). For both methods and both lead times, the forecast obtained by analogs is satisfactory with observations falling within the distribution provided by

the analogs. Moreover, when the analog distribution reached high values, it often matched with the observed high precipitation values. As discussed in Sect. 3.4, high precipitation events are better described by the 90th percentile than the center of the distribution. The distributions provided by the analogs are quite large but nevertheless provide useful information.

Figure 12 shows how the most significant event (Oct./Nov. 2018) is predicted by the four implemented methods at different lead times. The two older methods (2Z and 2Z-2MI) do not forecast the main peak as well as the optimized ones (4Zo and 4Zo-2MIo). The forecast with a lead time of seven days show that high precipitation amounts can be expected, but the timing is not well defined as the four daily forecasts show high variability of the timing of the occurrence of the peaks (illustrated by the four forecasted 90th percentiles in Fig. 12). The timing and amplitude of the event are relatively well captured by the 4Zo method with a lead time of four days. For the same lead time, adding moisture data (4Zo-2MIo) is not informative as the distributions are wider, and thus the occurrence of the peaks is more uncertain. Globally, moisture data was not very informative for this event. This might be related to the use of NR-1 as the archive, which has a very coarse resolution and was shown to not perform as well as other reanalyses (Horton and Brönnimann, 2018). It is likely that another dataset would be more accurate, and would be recommended for operational use.

6   Limitations of AMs

Although AMs were found to be relevant for several applications, they have some limitations which must be considered. The first is their lower performance for summer compared to winter when using standard predictors (Bliefernicht, 2010). The relationship between synoptic predictors and local rainfall is lower in the summer, due to convective precipitations that present higher spatial variability and depend on other parameters. The variables that describe the synoptic circulation are indeed not able to predict the location of thunderstorm cells. This was also observed by Ben Daoud (2010), who set up a specific model for the summer months (June 15 to September 15).

Another limitation is the need for a long archive of the predictand variable, for example, precipitation. ~~Without several decades of data, AMs cannot be used.~~ An alternative for regions without long archives of station measurements can be using satellite-derived precipitation. Long predictor archives are also required, which is easily satisfied with reanalyses. These may not be perfect in terms of homogeneity, but several can be considered to be of sufficient quality ~~Horton and Brönnimann (2018)~~ (Horton and Brönnimann, 2018). Moreover, reanalysis data are available all around the world, which represent a great potential for AMs.

~~Then, there is also the issue that extreme~~ Extreme events may be under-represented in the considered sample of analog situations. Indeed, in a limited weather archive, events with high return periods are not frequent, which can introduce a bias in the prediction. There are however techniques to correct for this bias (see Marty, 2010). In order to produce new extremes, postprocessing of the distribution of analogs might be necessary, for example, by using a scaling based on a predictor variable.

It is also legitimate to raise the question of the relevance of an approach based on archives of past situations in the context of climate change. Changes in circulation frequencies and the persistence of certain weather types (Hewitson and Crane, 1996) can be accounted for by AMs that contain predictors that characterize atmospheric circulation. Thus, if the archive of weather

events is long enough, it is reasonable to assume that a large part of future events is already represented, even those whose frequency will change under different climatic conditions (Wetterhall, 2005). Changes in moisture and temperature variables must be accounted for to correctly capture the climate change signals. Dayon et al. (2015) has demonstrated the transferability of certain AMs to future climate conditions.

7   Conclusions

AMs are cost-effective techniques for downscaling local meteorological variables from large-scale predictors. They are used in the context of operational forecasting for flood management or hydropower production, or in a climatic context for climate change impact modelling or reconstruction of past meteorological conditions.

AtmoSwing is a suite of tools that facilitate processing of multiple AM structures in a flexible and efficient way. It consists of four software: the Forecaster for operational forecasting, the Viewer for displaying the Forecaster outputs, the Downscaler for applying AMs in a climatic context, and the Optimizer to ~~infer~~ establish the relationship between predictors and predictands. AtmoSwing is written in C++, is open source, and has been extensively tested.

Processing operational forecasts with AtmoSwing requires very low computing infrastructure (implementation is possible on a Raspberry Pi 3) yet it can yield useful information, such as early warning for high precipitation events in the case of an application to flood forecasting. Valuable results were obtained in a three-year-long operational forecast in the Swiss Alps. With the global availability of reanalyses, it can be applied to any region with a relatively long predictand time series. The predictors and the structure of the method can be adapted to the local meteorological processes and controlled through xml files. The connection with open access NWP models such as GFS is integrated into AtmoSwing and requires no prior processing. The Forecaster can be installed on a computer or a headless server and run automatically to issue a forecast as soon as new NWP outputs are available. The Viewer offers a user-friendly display of the forecasts, with different levels of synthesis and details. It first provides an overview of potentially critical situations (possibility of high precipitation at a station for a certain lead time) but also allows plotting of the details of the distributions provided by the selected analog dates.

The Downscaler allows the AMs to be used in a climatic context, either for climate reconstruction or for climate change impact studies. When used for future climate analysis, the user must pay close attention to the selected predictors, so that they are able to represent the climate change signal. This is a relatively new field of application of AMs, which was proven to be of interest.

The Optimizer implements different optimization techniques, such as the sequential approach, a Monte–Carlo simulation, and a global optimization technique. ~~Inferring~~ Establishing the statistical relationship between predictors and predictand is quite intensive in terms of processing, as it requires numerous assessment over decades. To this end, the Optimizer has been highly optimized in terms of computing efficiency and is parallelized over multiple threads. It scales well on a Linux cluster. This procedure is only required to ~~infer~~ establish the statistical relationship, which can then be used in forecasting or downscaling at a low computing cost.

22

One possible key improvement to AtmoSwing Forecaster is a multi-models approach that relies on outputs from multiple global NWP models to better take into account the uncertainty of the NWP forecasts. Similarly, Thevenot (2004) demonstrated the benefit of using ensembles from global NWP as input for the method. The implementation consists of combining the selected analog days associated with each of the members. The forecast on the ensemble was found to be more accurate than the deterministic control for a lead time of four days and more (Thevenot, 2004).

AtmoSwing aims to facilitate implementation of AMs with different types of structure and various predictors while being computationally efficient with low computing requirements. It can be applied to different contexts, ~~allowing them to be~~ such as operational forecasting or climate impact studies. It is open source and will hopefully save future users some development time.

References

Alessandrini, S., Delle Monache, L., Sperati, S., and Cervone, G.: An analog ensemble for short-term probabilistic solar power forecast, Appl. Energy, 157, 95–110, https://doi.org/10.1016/j.apenergy.2015.08.011, 2015a.

Alessandrini, S., Delle Monache, L., Sperati, S., and Nissen, J. N.: A novel application of an analog ensemble for short-term wind power forecasting, Renew. Energy, 76, 768–781, https://doi.org/10.1016/j.renene.2014.11.061, 2015b.

Barnston, A. G., van den Dool, H. M., Rodenhuis, D. R., Ropelewski, C. R., Kousky, V. E., O'Lenic, E. A., Livezey, R. E., Zebiak, S. E., Cane, M. A., Barnett, T. P., Graham, N. E., Ji, M., Leetmaa, A., Barnston, A. G., van den Dool, H. M., Zebiak, S. E., Barnett, T. P., Ji, M., Rodenhuis, D. R., Cane, M. A., Leetmaa, A., Graham, N. E., Ropelewski, C. R., Kousky, V. E., O'Lenic, E. A., and Livezey, R. E.: Long-Lead Seasonal Forecasts—Where Do We Stand?, Bull. Am. Meteorol. Soc., 75, 2097–2114, https://doi.org/10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2, 1994.

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, La Houille Blanche, pp. 60–65, https://doi.org/10.1051/lhb/2009079, 2009.

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

Bessa, R., Trindade, A., Silva, C. S., and Miranda, V.: Probabilistic solar power forecasting in smart grids using distributed information, Int. J. Electr. Power Energy Syst., 72, 16–23, https://doi.org/10.1016/j.ijepes.2015.02.006, 2015.

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

Bolognesi, R.: Premiers développements d'un modèle hybride pour le diagnostic spatial des risques d'avalanches, La Houille Blanche, 8, 551–553, https://doi.org/10.1051/lhb/1993045, 1993.

Bonnet, R., Boé, J., Dayon, G., and Martin, E.: 20th-Century Hydro-Meteorological Reconstructions to Study the Multidecadal Variations of the Water Cycle Over France, Water Resour. Res., 53, 1–17, https://doi.org/10.1002/2017WR020596, http://doi.wiley.com/10.1002/2017WR020596, 2017.

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, https://doi.org/10.1051/lhb:200501002, 2005.

Bradley, A. A. and Schwartz, S. S.: Summary Verification Measures and Their Interpretation for Ensemble Forecasts, Mon. Weather Rev., 139, 3075–3089, https://doi.org/10.1175/2010MWR3305.1, 2011.

Brier, G.: Verification of forecasts expressed in terms of probability, Mon. Weather Rev., 78, 1–3, http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1950)078{%}3C0001:VOFEIT{%}3E2.0.CO;2, 1950.

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, https://doi.org/10.5194/cp-12-635-2016, 2016.

Caillouet, L., Vidal, J. P., Sauquet, E., Devers, A., and Graff, B.: Ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871, Hydrol. Earth Syst. Sci., 21, 2923–2951, https://doi.org/10.5194/hess-21-2923-2017, 2017.

Candille, G. and Talagrand, O.: Evaluation of probabilistic prediction systems for a scalar variable, Q. J. R. Meteorol. Soc., 131, 2131–2150, https://doi.org/10.1256/qj.04.71, http://doi.wiley.com/10.1256/qj.04.71, 2005.

5 Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., and Mason, S.: Forecast verification: current status and future directions, Meteorol. Appl., 15, 3–18, https://doi.org/10.1002/met.52, http://doi.wiley.com/10.1002/met.52, 2008.

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. Clim., 27, 5056–5074, https://doi.org/10.1175/JCLI-D-13-00464.1, 2014.

10 Charles, A., Timbal, B., Fernandez, E., and Hendon, H.: Analog Downscaling of Seasonal Rainfall Forecasts in the Murray Darling Basin, Mon. Weather Rev., 141, 1099–1117, https://doi.org/10.1175/mwr-d-12-00098.1, 2012.

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, https://doi.org/10.1002/2014JD022236, 2015.

Dayon, G., Boé, J., Martin, É., and Gailhard, J.: Impacts of climate change on the hydrological cycle over France and associated uncertainties, 15 Comptes Rendus - Geosci., 350, 141–153, https://doi.org/10.1016/j.crte.2018.03.001, 2018.

Delle Monache, L., Nipen, T., Liu, Y., Roux, G., and Stull, R.: Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions, Mon. Weather Rev., 139, 3554–3570, https://doi.org/10.1175/2011MWR3653.1, 2011.

Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., and Searight, K.: Probabilistic Weather Prediction with an Analog Ensemble., Mon. Weather Rev., 141, 3498–3516, https://doi.org/10.1175/MWR-D-12-00281.1, 2013.

20 Desaint, B., Nogues, P., Perret, C., and Garçon, R.: La prévision hydrométéorologique opérationnelle: l'expérience d'Electricité de France, in: Colloq. SHF-191e CST - «Prévisions hydrométéorologiques», Lyon, 18-19 novembre 2008, Tableau 1, p. 8 p., https://doi.org/10.1051/lhb/2009054, 2008.

Djerboua, A.: Prédétermination des pluies et crues extrêmes dans les Alpes franco-italiennes - Prévision quantitative des pluies journalières par la méthode des Analogues, Ph.D. thesis, Institut National Polytechnique de Grenoble, 2001.

25 Drosdowsky, W. and Zhang, H.: Verification of Spatial Fields, in: Forecast Verif. a Pract. Guid. Atmos. Sci., edited by Jolliffe, I. T. and Stephenson, D. B., chap. 6, pp. 121–136, Wiley, 2003.

Duband, D.: Reconnaissance dynamique de la forme des situations météorologiques. Application à la prévision quantitative des précipitations., Ph.D. thesis, Faculté des sciences de Paris, 1970.

Duband, D.: Reconnaissance dynamique de la forme des situations météorologiques, application à la prévision numérique des précipitations 30 journalières, in: Congrès la Société Hydrotechnique Fr. XIIIèmes journées I'Hydraulique, Paris, 1974.

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

Epstein, E.: A scoring system for probability forecasts of ranked categories, J. Appl. Meteorol., 8, 985–987, http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1969)008{%}3C0985:ASSFPF{%}3E2.0.CO{%}3B2, 1969.

Finley, J.: Tornado prediction, Am. Meteorol. J., 1, 85–88, 1884.

35 Foresti, L., Panziera, L., Mandapaka, P. V., Germann, U., and Seed, A.: Retrieval of analogue radar images for ensemble nowcasting of orographic rainfall, Meteorol. Appl., 22, 141–155, https://doi.org/10.1002/met.1416, 2015.

Fraedrich, K., Raible, C. C., and Sielmann, F.: Analog Ensemble Forecasts of Tropical Cyclone Tracks in the Australian Region, Weather Forecast., 18, 3–11, https://doi.org/10.1175/1520-0434(2003)018<0003:AEFOTC>2.0.CO;2, 2003.

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

GDAL Development Team: GDAL - Geospatial Data Abstraction Library, Version 1.11.0, www.gdal.org, 2014.

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

Gibergans-Báguena, J. and Llasat, M.: Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia, Atmos. Res., 86, 173–193, https://doi.org/10.1016/j.atmosres.2007.04.002, http://linkinghub.elsevier.com/retrieve/pii/S0169809507000695, 2007.

Gilbert, G. K.: Finley's tornado predictions, Am. Meteorol. J., 1, 166–172, 1884.

Goodess, C.: Statistical and regional dynamical downscaling of extremes for European regions: STARDEX, Eur. Geophys. union Inf. Newsl., 6, 25–29, http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Statistical+and+regional+dynamical+downscaling+of+extremes+for+European+regions:+STARDEX{#}0, 2003.

Gordon, N. D.: Statistical very short-range forecasting via analogues, in: Proc. Symp. mesoscale Anal. Forecast., pp. 307–312, ESA SP-282, Vancouver, Canada, 1987.

Gringorten, I. I.: A plotting rule for extreme probability paper, J. Geophys. Res., 68, 813–814, https://doi.org/10.1029/JZ068i003p00813, http://doi.wiley.com/10.1029/JZ068i003p00813, 1963.

Guennebaud, G., Jacob, B., and Others: Eigen v3, http://eigen.tuxfamily.org, 2010.

Guilbaud, S.: Prévision quantitative des précipitations journalières par une méthode statistico-dynamique des recherche d'analogues - Application à des bassins du pourtour méditerranéen, Ph.D. thesis, Institut National Polytechnique de Grenoble, 1997.

Guilbaud, S. and Obled, C.: Prévision quantitative des précipitations journalières par une technique de recherche de journées antérieures analogues: optimisation du critère d'analogie, Comptes Rendus l'Académie des Sci. Ser. II, A-Earth Planet. Sci., 327, 181–188, https://doi.org/10.1016/s1251-8050(98)80006-2, http://www.sciencedirect.com/science/article/pii/S1251805098800062, 1998.

Hamill, T. and Whitaker, J.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application, Mon. Weather Rev., 134, 3209–3229, https://doi.org/10.1175/mwr3237.1, 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, https://doi.org/10.1175/MWR-D-15-0004.1, 2015.

Heidke, P.: Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst, Geogr. Ann., 8, 301–349, http://www.jstor.org/stable/519729, 1926.

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

Hewitson, B. and Crane, R.: Climate downscaling: techniques and application, Clim. Res., 7, 85–95, http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Climate+downscaling+:+techniques+and+application{#}0, 1996.

Hoare, C. A. R.: Quicksort, Comput. J., 5, 10–16, 1962.

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.: AtmoSwing v2.1.0, https://doi.org/10.5281/zenodo.3208134, 2018a.

Horton, P.: AtmoSwing Python tools, https://doi.org/10.5281/zenodo.1495057, 2018b.

Horton, P. and Brönnimann, S.: Impact of global atmospheric reanalyses on statistical precipitation downscaling, Clim. Dyn., 0, 0, https://doi.org/10.1007/s00382-018-4442-6, 2018.

Horton, P. and Burkart, K.: AtmoSwing R tools, https://doi.org/10.5281/zenodo.1305098, 2018.

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

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

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

Horton, P., Jaboyedoff, M., and Obled, C.: Using genetic algorithms to optimize the analogue method for precipitation prediction in the Swiss Alps, J. Hydrol., 556, 1220–1231, https://doi.org/10.1016/j.jhydrol.2017.04.017, 2018.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J. F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: New high-resolution climate change projections for European impact research, Reg. Environ. Chang., 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Junk, C., Delle Monache, L., and Alessandrini, S.: Analog-based Ensemble Model Output Statistics, Mon. Weather Rev., 143, 2909–2917, https://doi.org/10.1175/MWR-D-15-0095.1, 2015a.

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. Zeitschrift, 24, 361–379, https://doi.org/10.1127/metz/2015/0659, 2015b.

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, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kanamitsu, M.: Description of the NMC global data assimilation and forecast system, Weather Forecast., 4, 335–342, 1989.

Kanamitsu, M., Alpert, J., Campana, K., Caplan, P., Deaven, D., Iredell, M., Katz, B., Pan, H., Sela, J., and White, G.: Recent changes implemented into the global forecast system at NMC, Weather Forecast., 6, 425–435, http://journals.ametsoc.org/doi/abs/10.1175/1520-0434(1991)006{%}3C0425:RCIITG{%}3E2.0.CO{%}3B2, 1991.

Keenan, T. D. and Woodcock, F.: Objective Tropical Cyclone Movement Forecasts Using Synoptic and Track Analogue Information, Tech. rep., Department of Science and Technology, Bureau of Meteorology, 1981.

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., and Fiorino, M.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bull. Am. Meteorol. Soc., 82, 247–267, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

Klein, W., Lewis, B., and Enger, I.: Objective prediction of five-day mean temperatures during winter, J. Meteorol., 16, 672–682, https://doi.org/10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0469(1959) 016{%}3C0672:OPOFDM{%}3E2.0.CO;2, 1959.

Kruizinga, S. and Murphy, A.: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands, Mon. Weather Rev., 111, 2244–2254, https://doi.org/10.1175/1520-0493(1983)111<2244:uoaapt>2.0.co;2, 1983.

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, https://doi.org/10.1175/1520-0469(1969)26<636:aparbn>2.0.co;2, 1969.

Maheras, P., Anagnostopoulou, C., and Tolika, K.: Contribution to D12 (AUTH, APRA-SMR, FIC and UEA) Downscaling of extreme indices in Greece, Tech. rep., STARDEX project, http://www.cru.uea.ac.uk/projects/stardex/deliverables/D12/D12{_}regional{_}Greece. pdf, 2005.

Mandon, S.: Comparaison d'épisodes pluvieux intenses sur le Sud-Est de la France et de situations analogues au sens de la circulation générale. Recherche de variables discriminantes., Tech. rep., Rapport de stage de fin d'études de l'Ecole de la Météorologie Nationale, 1985.

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. Climatol., 51, 505–520, https://doi.org/10.1175/JAMC-D-11-011.1, 2012.

Mason, I.: A model for assessment of weather forecasts, Aust. Meteorol. Mag., 30, 291–303, http://www.nssl.noaa.gov/users/brooks/ public{_}html/feda/papers/mason82.pdf, 1982.

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

Matulla, C., Zhang, X., Wang, X. L., Wang, J., Zorita, E., Wagner, S., and von Storch, H.: Influence of similarity measures on the performance of the analog method for downscaling daily precipitation, Clim. Dyn., 30, 133–144, https://doi.org/10.1007/s00382-007-0277-2, http: //link.springer.com/10.1007/s00382-007-0277-2, 2007.

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

Obled, C. and Good, W.: Recent developments of avalanche forecasting by discriminant analysis techniques: a methodological review and some applications to the Parsenn area (Davos, Switzerland), J. Glaciol., 25, 315–346, https://doi.org/10.3198/1980JoG25-92-315-346, 1980.

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, https://doi.org/10.1016/S0169-8095(02)00038-8, 2002.
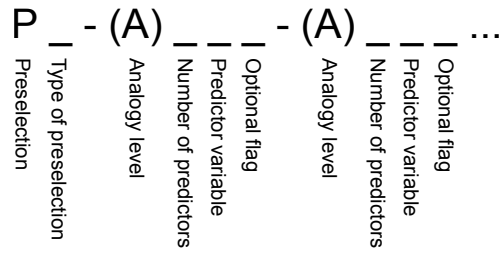
Panziera, L., Germann, U., Gabella, M., and Mandapaka, P. V.: NORA-Nowcasting of Orographic Rainfall by means of Analogues, Q. J. R. Meteorol. Soc., 137, 2106–2123, https://doi.org/10.1002/qj.878, http://doi.wiley.com/10.1002/qj.878, 2011.

Peirce, C.: The numerical measure of the success of predictions, Science, 4, 453–454, 1884.

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

Radinovic, D.: An analogue method for weather forecasting using the 500/1000 mb relative topography, Mon. Weather Rev., 103, 639–649, https://doi.org/10.1175/1520-0493(1975)103<0639:aamfwf>2.0.co;2, 1975.

Raynaud, D.: Hydroclimatic variability and the integration of renewable energy in Europe, Ph.D. thesis, Université Grenoble Alpes, 2016.

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., https://doi.org/10.1002/joc.4844, http://doi.wiley.com/10.1002/joc.4844, 2016.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Clim., 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1824.1, 2007.

Rodwell, M., Haiden, T., and Richardson, D. S.: Developments in precipitation verification, ECMWF Newsl., 128, 12–16, http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Developments+in+precipitation+verification{#}1, 2011.

Rodwell, M. J., Richardson, D. S., Hewson, T. D., and Haiden, T.: A new equitable score suitable for verifying precipitation in numerical weather prediction, Q. J. R. Meteorol. Soc., 136, 1344–1363, https://doi.org/10.1002/qj.656, http://doi.wiley.com/10.1002/qj.656, 2010.

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

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.

Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundecha, Y., Ribalaygua, J., and Schmith, T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, J. Geophys. Res., 112, D04 105, https://doi.org/10.1029/2005JD007026, http://doi.wiley.com/10.1029/2005JD007026, 2007.

Shao, Q. and Li, M.: An improved statistical analogue downscaling procedure for seasonal precipitation forecast, Stoch. Environ. Res. Risk Assess., 27, 819–830, https://doi.org/10.1007/s00477-012-0610-0, 2013.

Sievers, O., Fraedrich, K., and Raible, C. C.: Self-Adapting Analog Ensemble Predictions of Tropical Cyclone Tracks, Weather Forecast., 15, 623–629, https://doi.org/10.1175/1520-0434(2000)015<0623:SAAEPO>2.0.CO;2, 2000.

Smart, J., Hock, K., and Csomor, S.: Cross-Platform GUI Programming with wxWidgets, Prentice Hall, 2006.

STARDEX: STARDEX, Downscaling climate extremes, Tech. rep., STARDEX consortium, 2005.

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of probabilistic prediction systems, in: Proc. ECMWF Work. Predict., pp. 1–25, 1997.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

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

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

Turco, M., Llasat, M. C., Herrera, S., and Gutiérrez, J. M.: Bias correction and downscaling of future RCM precipitation projections using a MOS-analog technique, J. Geophys. Res., 122, 2631–2648, https://doi.org/10.1002/2016JD025724, 2017.

Vallée, J. L.: Précipitations sur le Sud-Ouest du Massif Central et l'Est des Pyrénées. Optimisation du modèle EDF/DTG de prévision par recherche d'analogues., Tech. rep., Rapport de stage de fin d'études de l'Ecole de la Météorologie Nationale, 1986.

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

Vanvyve, E., Delle Monache, L., Monaghan, A. J., and Pinto, J. O.: Wind resource estimates with an analog ensemble approach, Renew. Energy, 74, 761–773, https://doi.org/10.1016/j.renene.2014.08.060, 2015.

Wetterhall, F.: Statistical Downscaling of Precipitation from Large-scale Atmospheric Circulation: Comparison of Methods and Climate Regions, Ph.D. thesis, Uppsala University, 2005.

Wetterhall, F., Halldin, S., and Xu, C.-y.: Seasonality properties of four statistical-downscaling methods in central Sweden, Theor. Appl. Climatol., 87, 123–137, https://doi.org/10.1007/s00704-005-0223-3, http://link.springer.com/10.1007/s00704-005-0223-3, 2007.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: An Introduction. 2nd edition, Elsevier, 2006.

Wilson, L. and Vallée, M.: The Canadian updateable model output statistics (UMOS) system: Design and development tests, Weather Forecast., 17, 206–222, https://doi.org/10.1175/1520-0434(2002)017<0206:TCUMOS>2.0.CO;2, 2002.

Wilson, L. J. and Yacowar, N.: Statistical weather element forecasting in the Canadian Weather Service, in: Proc. WMO Symp. Probabilistic Stat. Methods Weather Forecast., pp. 401–406, Nice, France, 1980.

Woodcock, F.: On the use of analogues to improve regression forecasts, Mon. Weather Rev., 108, 292–297, https://doi.org/10.1175/1520-0493(1980)108<0292:otuoat>2.0.co;2, 1980.

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

Xavier, P. K. and Goswami, B. N.: An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability, Mon. Weather Rev., 135, 4149–4160, https://doi.org/10.1175/2007MWR1854.1, 2007.

Zorita, E. and von Storch, H.: The analog method as a simple statistical downscaling technique: comparison with more complicated methods, J. Clim., 12, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2, 1999.
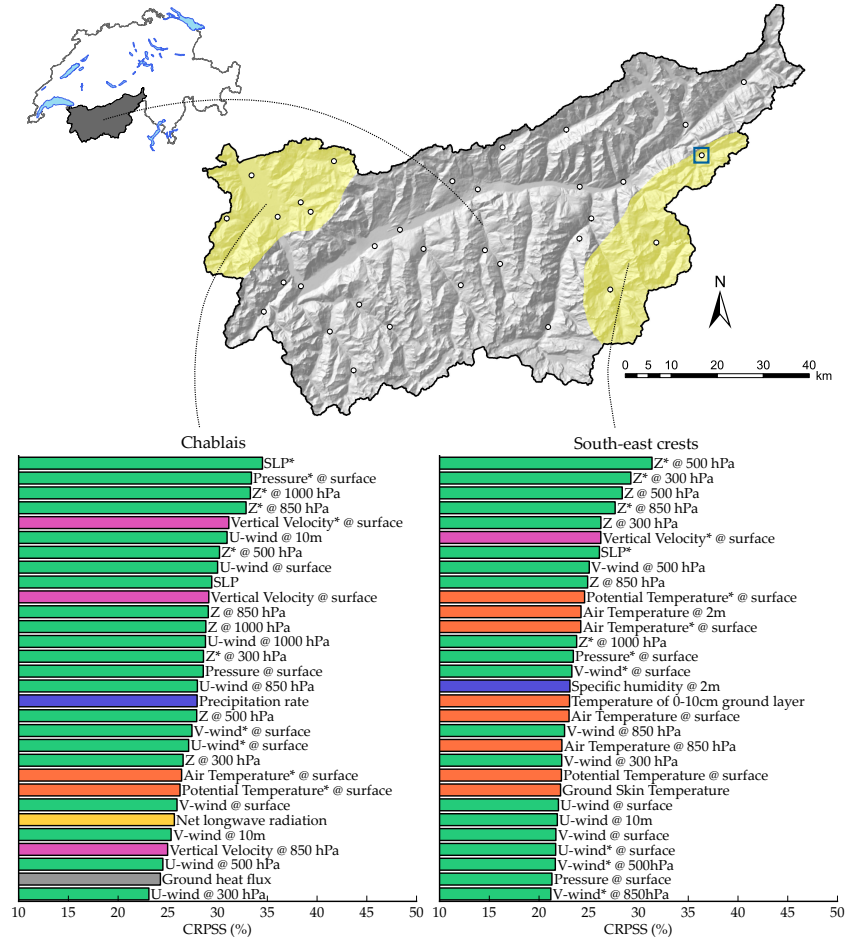
Figure 1. Performance score (CRPSS; Eq. 3; the reference being the climatological precipitation distribution) of the 30 best variables from the ~~NCEP/NCAR reanalysis~~ NR-1 dataset, when considered separately (no combination), for the Chablais region and the southeast ridges in the upper Rhône catchment in Switzerland. The analogy criterion is S1 when ~~there is~~ an asterisk is present next to the variable name, and RMSE otherwise. Colour illustrates the variable type: green = atmospheric circulation, blue = moisture, orange = temperature, yellow = radiation, purple = vertical velocity, and gray = other. SLP stands for sea level pressure and Z for geopotential height. The blue square indicates the Binn station~~, which is analyzed in more detail later on~~.
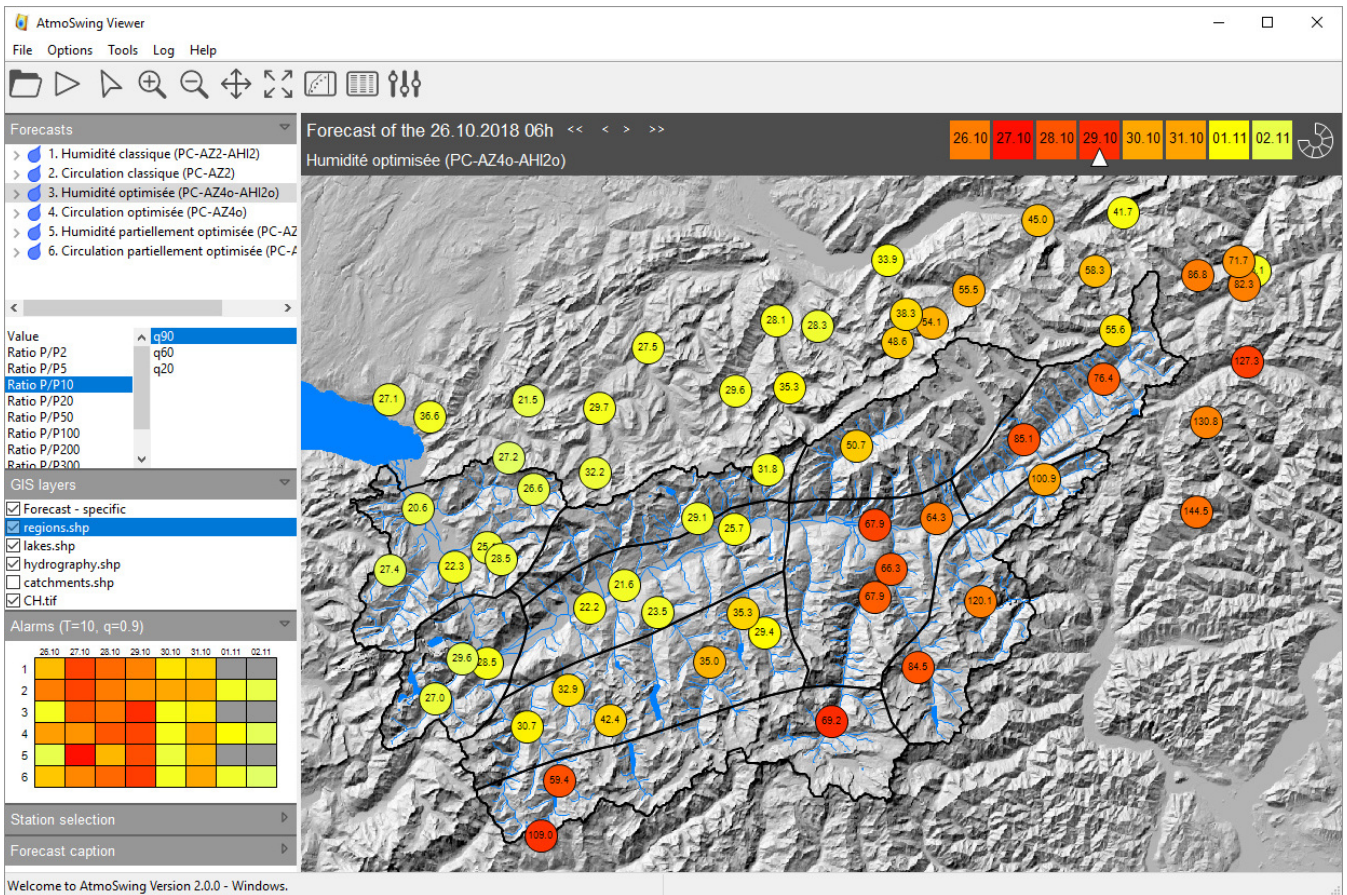
P _ - (A) _ _ _ - (A) _ _ _ ...

Preselection
Type of preselection
Analogy level
Number of predictors
Predictor variable
Optional flag
Analogy level
Number of predictors
Predictor variable
Optional flag

Figure 2. Proposed ~~structure for naming~~ nomenclature to describe the ~~parameterizations of the~~ AM structure.
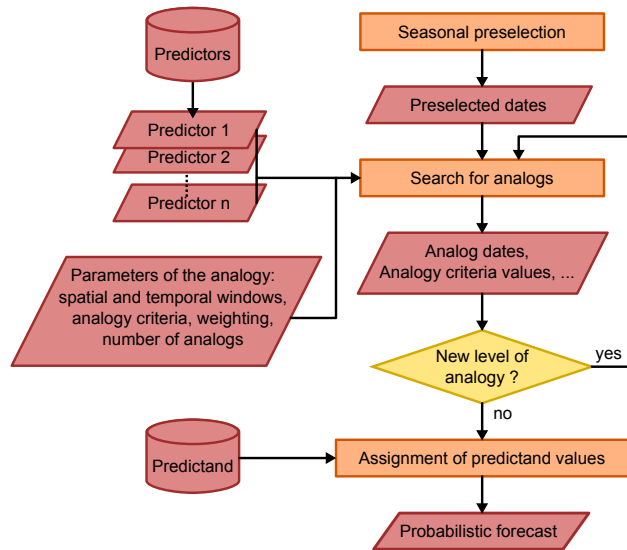

**Figure 3.** Simplified flowchart of the AM implementation in AtmoSwing.

Figure 4. ~~Simplified illustration~~ Graphical user interface of the ~~main classes or objects involved in the core of the AM processing in AtmoSwing.~~ Viewer module (Elevation data from ~~The~~ ~~processor class interacts with parent classes that can represent different entities~~ Shuttle Radar Topography Mission - SRTM, ~~such as different reanalysis datasets, predictand, criteria, scores,~~ and ~~in different contexts: calibration, forecasting, and downscaling~~ hydrological network from SwissTopo). The ~~items in green are only available in~~ values on the ~~Optimizer,~~ map ~~represent~~ the ~~ones in blue, in~~ 90th percentile (as selected on the ~~Forecaster, and~~ left panel) of the ~~ones in Orange, in~~ precipitation values from the ~~Downscaler. The area represents~~ analog samples at the ~~spatial window~~ different stations and ~~for~~ the selected lead time~~array a list of candidate dates (from preselection or previous analogy levels)~~. The ~~links~~ colour is proportional to the ~~parameters illustrate the dynamic definition of the different types by the parameters from the XML file~~ selected return period (10 years here).

~~Graphical user interface of the Viewer module (Elevation data from The Shuttle Radar Topography Mission (SRTM), and hydrological network from SwissTopo).~~

33
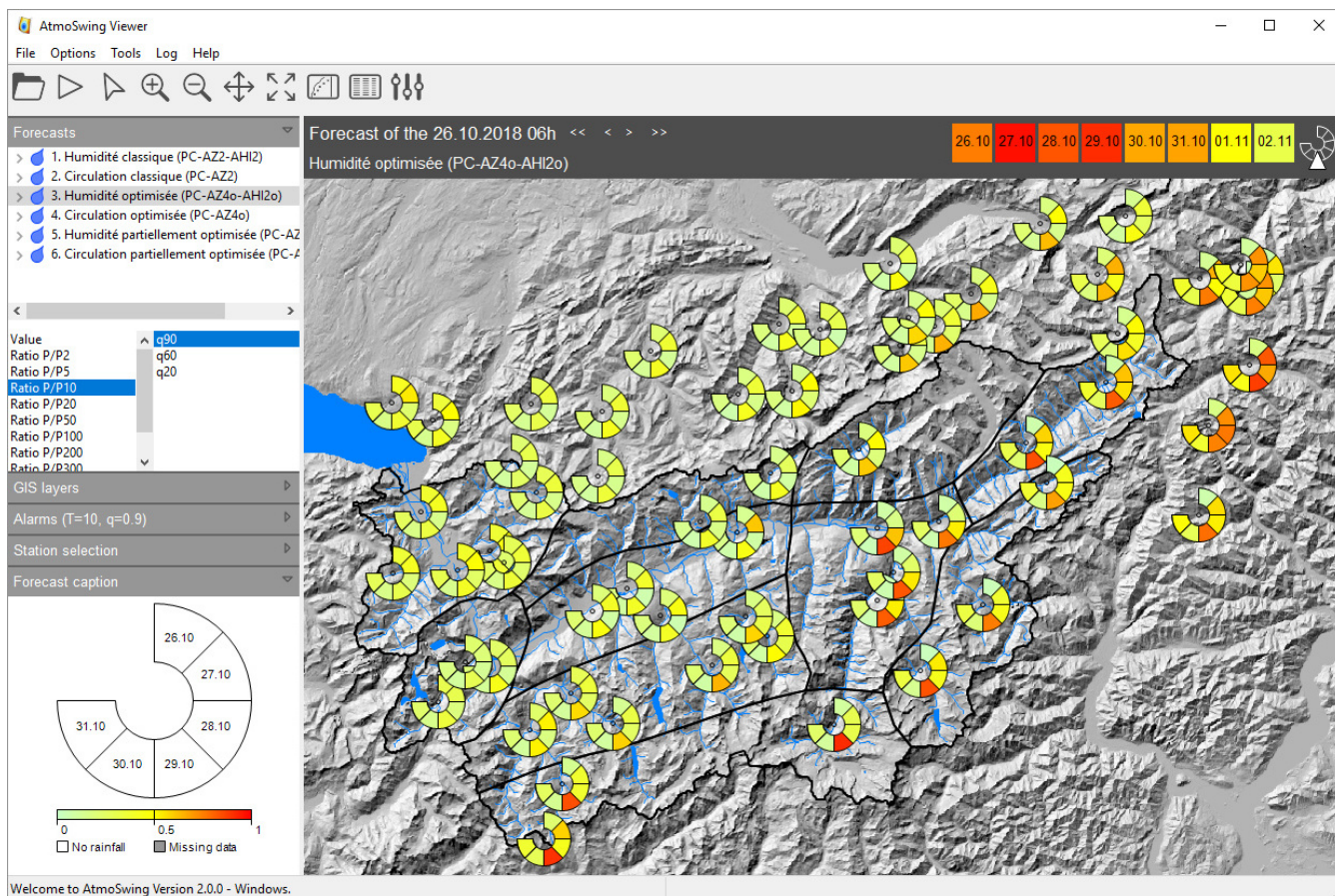
Figure 5. ~~Visualization of~~ Same as Fig. 4 but for multiple lead times~~on the map (Elevation data from the SRTM, and hydrological network from SwissTopo)~~.
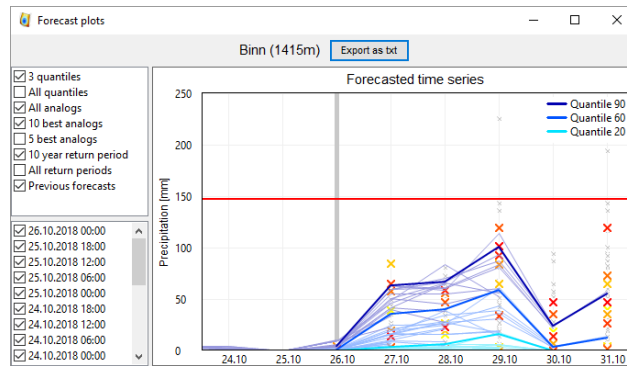
Figure 6. Visualization of the forecasted time series for an event at the Binn station (Fig. 1) in October 2018. The thick blue lines represent the 90th, 60th, and 20th percentiles for the given lead times. The thin blue lines represent the equivalent time series but from previous forecasts. The small grey crosses represent all analog values and the larger crosses highlight the 10 best analogs (with a colour gradient from red for the best to yellow for the 10th).
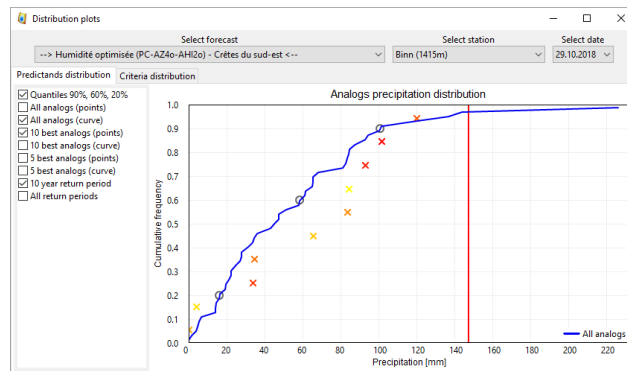

Figure 7. Visualization of the forecasted precipitation distribution for a given lead time for an event at the Binn station (Fig. 1) in October 2018. The blue line represent the full distribution provided by all analogs, the circles are the 90th, 60th, and 20th percentiles, and the crosses correspond to the distribution provided by the 10 best analogs (with a colour gradient from red for the best to yellow for the 10th). The vertical red line is here the precipitation value for a 10 year return period.
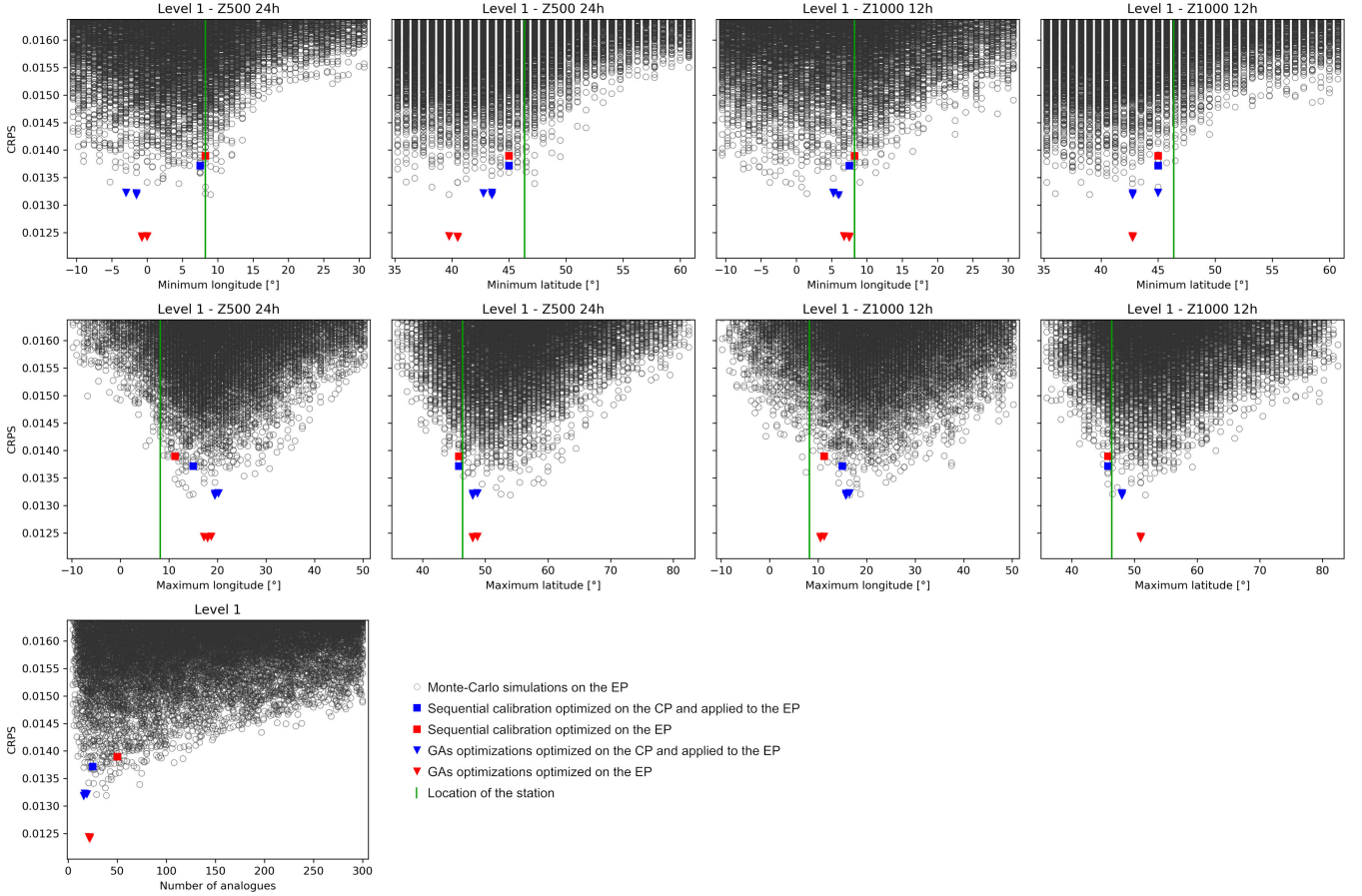
Figure 8. Example of parameter values for 2Z (Table 2) for the precipitation at the Binn station (Fig. 1) on the ~~period 2001–2010~~EP. The parameters are the extent (min/max longitude/latitude) of the spatial windows for the geopotential height at 500 and 1000 hPa, and the number of analogs. The green vertical bar in the plots represents the location of the station. The circles represent random parameters from the Monte–Carlo analysis. The plots are truncated at the 25th best percentiles for 50,000 realizations. Squares are the results of the sequential calibration and triangles result from genetic algorithms. Markers in blue represent parameters optimized for the ~~period 1981–2000~~ CP and applied to ~~2001–2010~~EP. Markers in red represent parameters optimized ~~directly~~ for the ~~period 2001–2010~~EP.
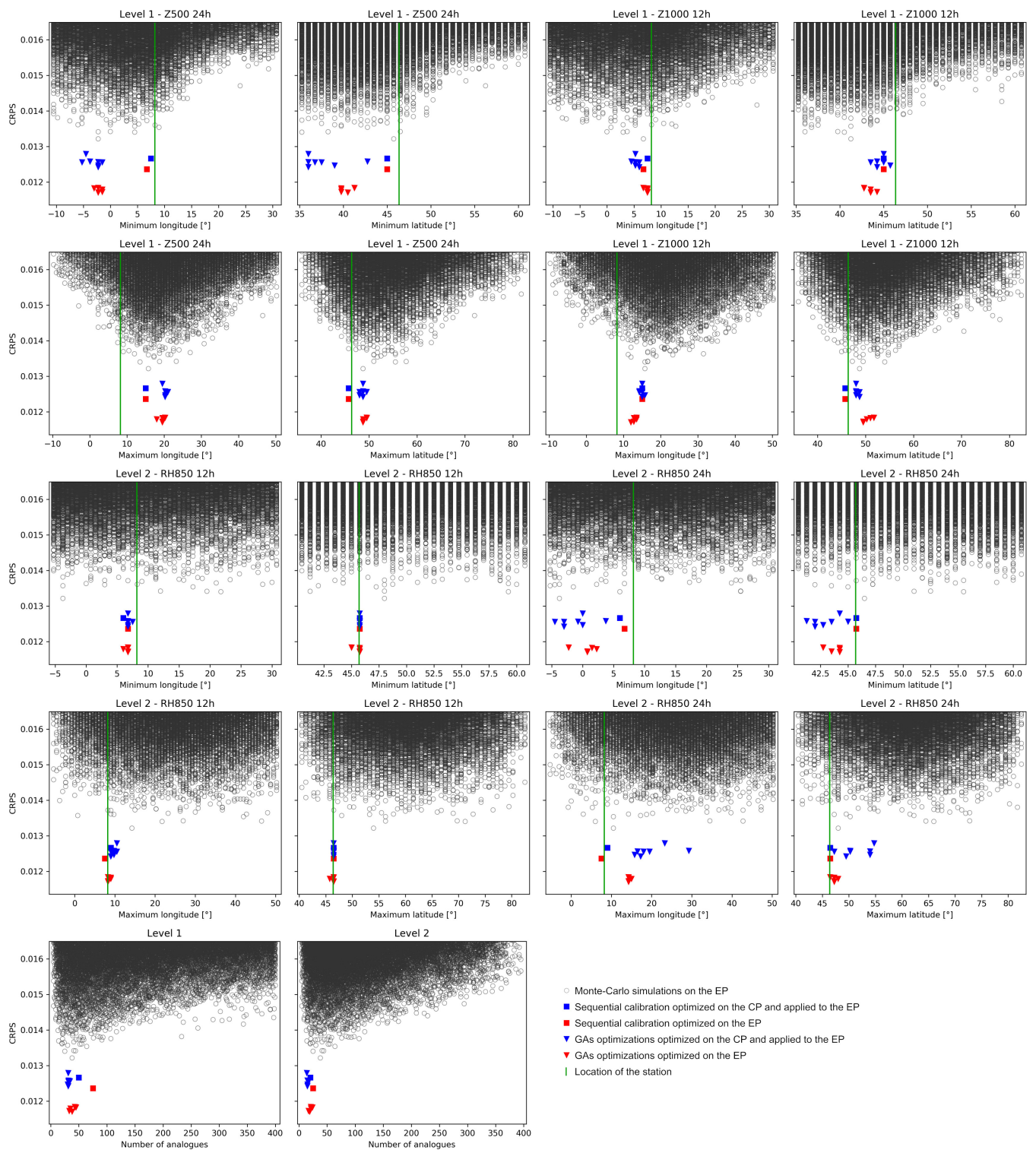
Figure 9. Same as Fig. 8 but for 2Z-2MI (Table 2). Results are shown for both levels of analogy (geopotential height and moisture index).
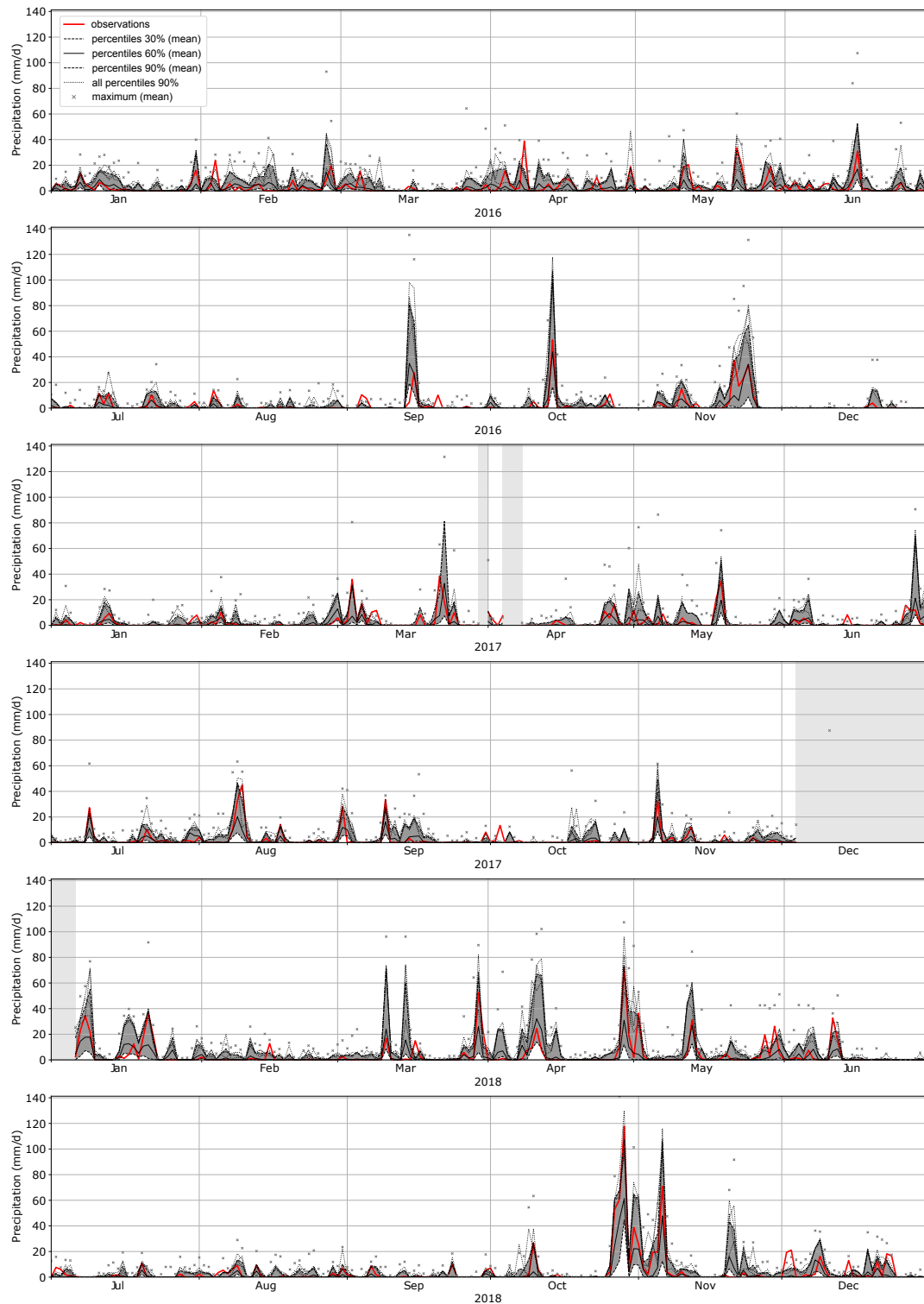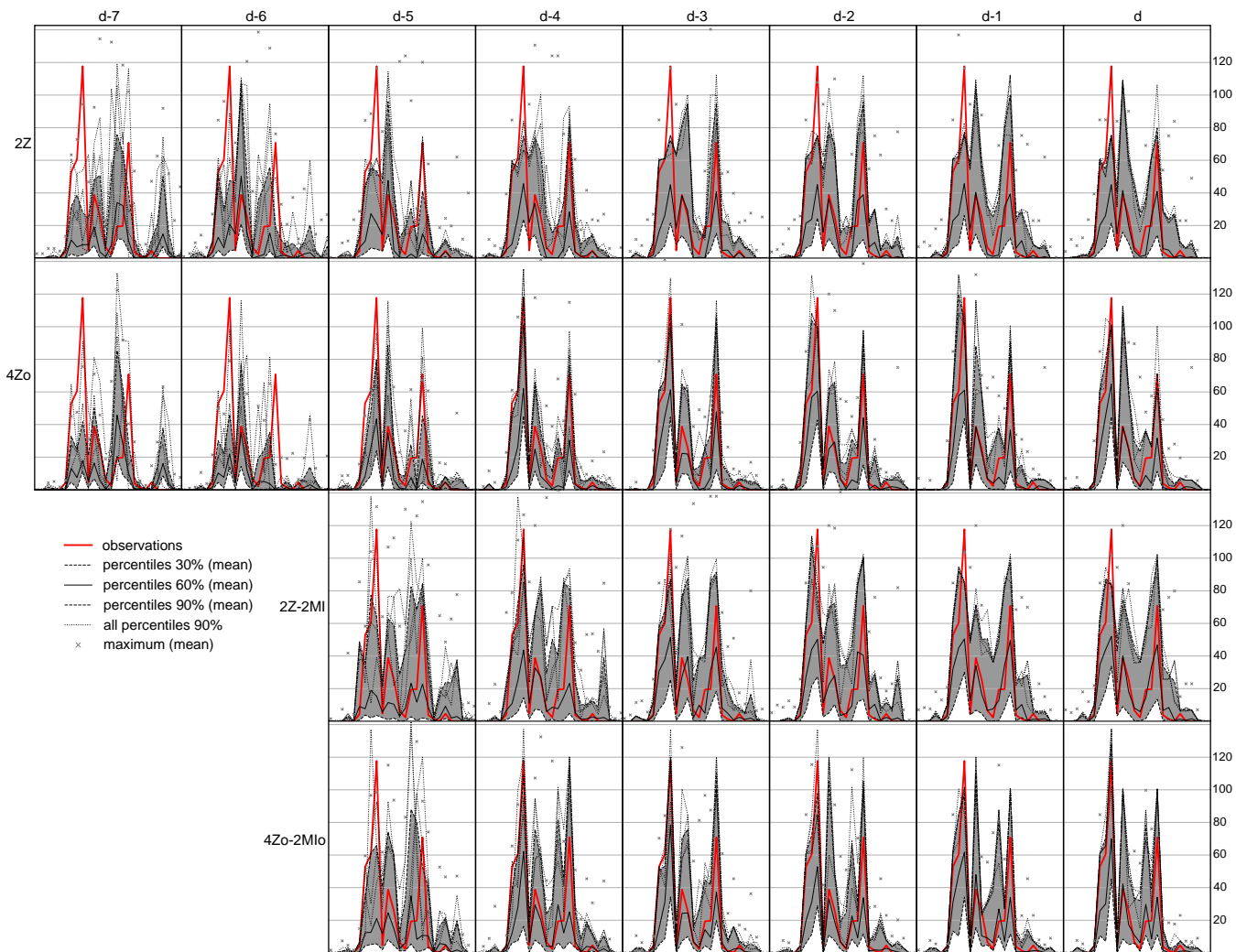
Figure 10. Forecasts for the Binn station (Fig. 1) over the period 2016–2018 obtained using the 4Zo method (Table 2) with a lead time of three days. The distributions provided by the analog values are summarized by the 90th, 60th, and 30th percentiles, as well as the maximum (crosses), all of them averaged over the four daily forecasts. Additionally, the four 90th percentiles were also plotted to show the consistency / variability between the four daily forecasts. The shaded areas correspond to forecasts downtime.
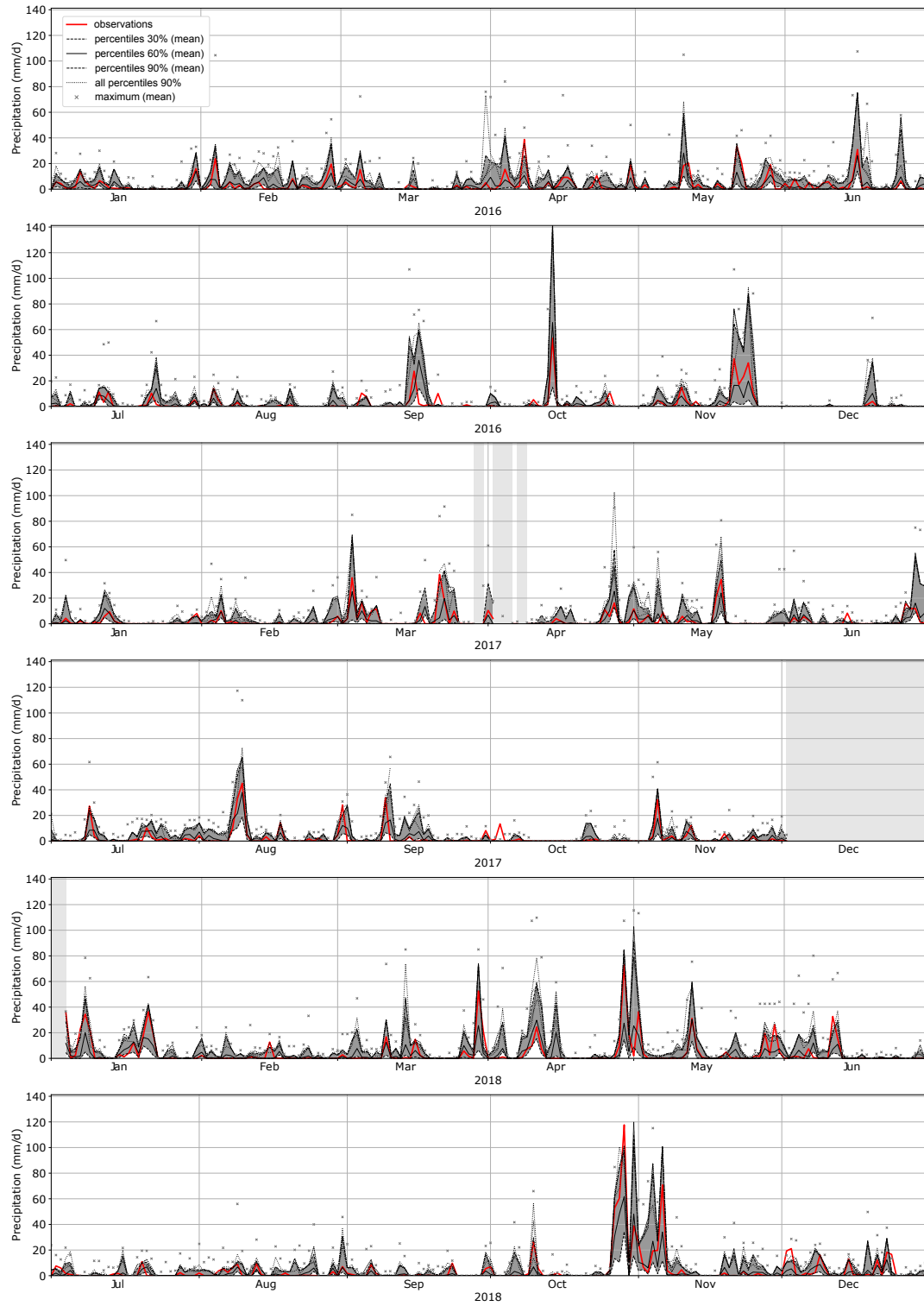
Figure 11. Same as Fig. ~~11~~ 10 but for the 4Zo-2MIo method (Table 2) with a lead time of one day.

Figure 12. Forecasts for the Oct./Nov. 2018 event at the Binn station (Fig. 1) for 2Z, 4Zo, 2Z-2MI and 4Zo-2MIo for lead times from seven to zero days prior to the target day.

Table 1. Reanalysis datasets that can be read by AtmoSwing.

Name	Institution	Period of record	Output resolution	Model resolution	Model ~~vintage~~generation	Type of input
NR-1	NCEP, NCAR	1948 – present	2.5°x 2.5°	T62 ($\sim$1.88°), L28	1995	full
NR-2	NCEP, DOE	1979 – present	2.5°x 2.5°	T62 ($\sim$1.88°), L28	2001	full
ERA-INT	ECMWF	1979 – present	0.75°x 0.75°	TL255 ($\sim$0.70°), L60	2006	full
20CR-2c	NOAA-CIRES	1851 – 2014	2°x 2°	T62 ($\sim$1.88°), L28	2008	surface
CFSR	NCEP	1979 – present	0.5°x 0.5°	T382 ($\sim$0.31°), L64	2009	full
JRA-55	JMA	1958 – present	1.25°x 1.25°	TL319 ($\sim$0.36°), L60	2009	full
JRA-55C	JMA	1958 – 2015	1.25°x 1.25°	TL319 ($\sim$0.36°), L60	2009	conventional
ERA-20C	ECMWF	1900 – 2010	1°x 1°	TL159 ($\sim$1.13°), L91	2012	surface
MERRA-2	NASA GMAO	1980 – present	0.625°x 0.5°	0.625°x 0.5°, L72	2014	full
CERA-20C	ECMWF	1901 – 2010	1°x 1°	T159 ($\sim$1.13°), L91	2016	surface
ERA5	ECMWF	1979 – present	0.25°x 0.25°	TL639 ($\sim$0.28°), L137	2016	full

Table 2. Some existing analog methods, 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. N1, N2 and N3 are the number of analogs to select at each level of analogy. The meteorological variables are: 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, MF – moisture flux, which is the product of MI with the wind intensity. The analogy criterion is S1 for Z and RMSE for the other variables.

Type	P0	L1	N1	L2	N2	L3	N3	Reference
~~PC-2Z~~2Z	PC	Z1000@12h Z500@24h	50					Bontron 2004
~~PC-4Z~~4Z	PC	Z1000@06h Z1000@30h Z700@24h Z500@12h	~27					Horton et al. 2018
~~PC-2Z-2MI~~2Z-2MI	PC	Z1000@12h Z500@24h	70	MI850@12h MI850@24h	30			Bontron 2004
~~PC-2Z-2MI~~2Z-2MI	PC	Z1000@06h Z500@18h	75	MI925@06h MI925@18h	30			Marty 2010
~~PC-2Z-2MF~~2Z-2MF	PC	Z1000@06h Z500@18h	60	MF700@06h[†] MF700@18h	25			Marty 2010
~~PC-4Z-2MI~~4Z-2MI	PC	Z1000@30h Z850@12h Z700@24h Z400@12h	~63	MI700@24h MI600@12h	~24			Horton et al. 2018
PT-2Z-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	70	MI925@12h MI925@24h MI700@12h MI700@24h	25			Ben Daoud et al. 2016
PT-2Z-10MI	T925@36h T600@12h	Z1000@12h Z500@24h	70	MI925@06-30h MI700@06-30h	25			Ben Daoud 2010
PT-2Z-4W-4MI	T925@36h T600@12h	Z1000@12h Z500@24h	170	W850@06h W850@12h W850@18h W850@24h	70	MI925@12h MI925@24h MI700@12h MI700@24h	25	Ben Daoud et al. 2016

† or MF925@06h+18h as an alternative