%\documentclass[5p]{elsarticle}

\documentclass[review]{elsarticle}

\usepackage{lineno,hyperref}

\modulolinenumbers[5]

\journal{Journal of Hydrology}

\usepackage{multirow}

\usepackage{gensymb}

%TODO: read again

%TODO: Elsevier english editing

%TODO: relative improvement of the crpss...

%TODO: homogenize hours: 6 a.m. or 06 h ??

%%%%%%%%%%%%%%%%%%%%%%%

%% Elsevier bibliography styles

%%%%%%%%%%%%%%%%%%%%%%%

%% To change the style, put a % in front of the second line of the current style and

%% remove the % from the second line of the style you would like to use.

%%%%%%%%%%%%%%%%%%%%%%%

%% Numbered

%\bibliographystyle{model1-num-names}

%% Numbered without titles

%\bibliographystyle{model1a-num-names}

%% Harvard

\bibliographystyle{model2-names}\biboptions{authoryear}

%% Vancouver numbered

%\usepackage{numcompress}\bibliographystyle{model3-num-names}

%% Vancouver name/year

%\usepackage{numcompress}\bibliographystyle{model4-names}\biboptions{authoryear}

%% APA style

%\bibliographystyle{model5-names}\biboptions{authoryear}

%% AMA style

%\usepackage{numcompress}\bibliographystyle{model6-num-names}

%% `Elsevier LaTeX' style

%\bibliographystyle{elsarticle-num}

%%%%%%%%%%%%%%%%%%%%%%%

\begin{document}

\begin{frontmatter}

\title{Using Genetic Algorithms to Optimize the Analogue Method for Precipitation Prediction in the Swiss Alps}

%% Group authors per affiliation:

\author[unil,unibe,terranum]{Pascal Horton\corref{mycorrespondingauthor}}

\cortext[mycorrespondingauthor]{Corresponding author}

\ead{pascal.horton@alumnil.unil.ch}

\author[unil]{Michel Jaboyedoff}

\author[lthe]{Charles Obled}

\address[unil]{University of Lausanne, Institute of Earth Sciences, Lausanne, Switzerland}

\address[unibe]{University of Bern, Oeschger Centre for Climate Change Research, Institute of Geography, Bern, Switzerland}

\address[lthe]{Universit\'{e} de Grenoble-Alpes, LTHE, Grenoble, France}

\begin{abstract}

Analogue methods provide a statistical precipitation prediction based on synoptic predictors supplied by general circulation models or numerical weather prediction models. The method samples a selection of days in the archives that are similar to the target day to be predicted, and consider their set of corresponding observed precipitation (the predictand) as the conditional distribution for the target day. The relationship between the predictors and predictands relies on some parameters that characterize how and where the similarity between two atmospheric situations is defined.

This relationship is usually established by a semi-automatic sequential procedure that has strong limitations: (i) it cannot automatically choose the pressure levels and temporal windows for a given meteorological variable, (ii) it cannot handle dependencies between parameters, and (iii) it cannot easily handle new degrees of freedom. In this work, a global optimization approach relying on genetic algorithms was able to optimize all parameters jointly and automatically. It allowed consideration of parameter inter-dependencies, and objective selection of some parameters that were manually selected beforehand, which, among other things, obviates the need to assess a large number of combinations of pressure levels and temporal windows of predictor variables.

The global optimization was applied to some variants of the analogue method for the Rh\^{o}ne catchment in the Swiss Alps. The performance scores increased compared to reference methods, especially for days with high precipitation totals. The resulting parameters were found to be relevant and coherent between the different subregions of the catchment. Moreover, they were obtained automatically and objectively, which reduces the effort that needs to be invested in exploration attempts when adapting the method to a new region or for a new predictand. In addition, the approach allowed for new degrees of freedom, such as a possible weighting between pressure levels, and non-overlapping spatial windows.

\end{abstract}

\begin{keyword}

precipitation prediction\sep

analogue method\sep

optimization\sep

genetic algorithms\sep

Alpine climate

\end{keyword}

\end{frontmatter}

\linenumbers

**\section{Introduction}**

**\label{sec:intro}**

The analogue method (AM) is a downscaling technique based on the idea expressed by \citet{Lorenz1956, Lorenz1969} that similar situations in terms of atmospheric circulation are likely to lead to similar local weather \citep{Duband1970}. It uses predictor variables describing the synoptic atmospheric circulation in order to predict local-scale predictands of interest. It is often used to predict daily precipitation, either in an operational forecasting context \citep[e.g.][]{Guilbaud1997, Bontron2005, Hamill2006, Bliefernicht2010, Marty2012, Horton2012, Horton2016a, Hamill2015, BenDaoud2016} or a climate downscaling context \citep[e.g.][]{Radanovics2013, Chardon2014, Dayon2015, Raynaud2016b}. Other predictands are also considered, such as precipitation radar images \citep{Panziera2011,Foresti2015a}, temperature \citep{Radinovic1975, Woodcock1980, Kruizinga1983, DelleMonache2013, Caillouet2016, Raynaud2016b}, wind \citep{Gordon1987, DelleMonache2013, DelleMonache2011, Vanvyve2015, Alessandrini2015, Junk2015, Junk2015c}, solar power \citep{Alessandrini2015a, Bessa2015}, snow avalanches \citep{Obled1980, Bolognesi1993}, and radiation \citep{Bois1981, Raynaud2016b}.

In real-time forecasting, it is used mainly by practitioners, notably hydropower companies or flood forecasting services, that need to anticipate water yields or issue early flood warnings several days in advance. The classical forecasting chain consists of using limited area models (e.g. AROME, or COSMO) forced by global NWP (numerical weather prediction) models with a lower resolution. However, their use requires very important processing capacities, and the resulting forecast still presents large uncertainties and biases. Although these outputs are essential, they can be supplemented by other sources of forecasts providing useful information. In contrast to local NWP models, AMs can transform at low cost the synoptic-scale information provided by the global NWP model into precipitation forecasts, by using the natural local behaviour in response to synoptic-scale influences stored in the archive of observed precipitation. Running an AM approach is fast enough that it can search for analogues for each day, up to ten days ahead, eventually for the different traces of an ensemble forecast and/or those issued by different NWP models (e.g. NOAA-GFS or ECMWF-IFS), in a matter of minutes.

In climate studies, AMs are used to downscale the outputs of a general circulation model (GCM) or regional climate model (RCM) simulation runs \citep{Dayon2015} or to reconstruct past weather conditions \citep{Caillouet2016}. In future climate studies, RCMs are often used to dynamically downscale precipitation to a local scale. However, even though the relevance of RCMs' outputs increases, a bias correction of the outputs is often still required, particularly in complex terrain. Moreover, their application is computer-intensive, which makes it difficult to cover all combinations of climate scenarios and GCMs. Therefore, the idea is to bypass the small-scale simulations and to go from the large-scale situation to the end variables such as precipitation by statistical downscaling \citep{Maraun2010}.

Beyond being computationally inexpensive, another big advantage of AMs is that they create realistic precipitation patterns for a region, provided that the analogue dates are the same, since they are based on observed situations with consistent spatial distribution \citep{Radanovics2013, Chardon2014}. For the same reason, they can also provide multivariate predictions that are physically consistent \citep{Raynaud2016b}.

The method can be designed with multiple successive subsampling steps, or analogy levels, each of them relying on different meteorological variables. A certain number of parameters define the relationship between predictors and predictands, such as the choice of the predictor variable, its pressure level and temporal window to consider, the spatial domain to use for the comparison, as well as the analogy criterion itself, and finally, the number of analogue situations to keep at each subsampling level. These parameters are usually calibrated by means of a semi-automatic sequential procedure \citep{Bontron2004, Radanovics2013}, i.e. by optimizing each single parameter, one at a time, in an arbitrarily chosen order, with no or little reassessment. This sequential approach therefore has strong limitations: (i) it cannot automatically choose the optimal pressure levels and the temporal windows for a given meteorological variable, (ii) it cannot handle dependencies between the parameters within a level of analogy, and even less between them, and (iii) it cannot easily handle new degrees of freedom, such as a possible weighting between the pressure levels. Thus, even if the processing involved is relatively fast, the sequential approach requires laborious assessments of predictor combinations (variables, pressure levels, temporal windows), and presents a high risk of ending in a local optimum because of subjective initial choices and lack of consideration of parameter inter-dependencies. Other calibration methods exist for specific applications, such as radar images \citep{Panziera2011, Foresti2015a}.

Aiming to overcome these limitations, a global optimization by genetic algorithms (GAs) was introduced. An intensive assessment resulted in recommendations to parametrize GAs in order to optimize AMs successfully \citep{Horton2016}. The present paper is based on these recommendations, and applies them to precipitation prediction for the upper Rh\^{o}ne catchment in the Swiss Alps, using AMs of varying complexity. It aims at illustrating the relevance of a fully automatic, objective, and global, optimization technique for AMs. The applications are indeed numerous, as AMs have to be adapted to every new location they are applied, or to any new predictand they should predict.

The data, AMs, and optimization techniques (sequential and GAs) are presented in Section \ref{sec:data\_methods}. The results are first given for the optimization of the analogy of atmospheric circulation only (Section \ref{sec:optim\_circul}), before being extended to a method adding a second level of analogy on moisture variables (Section \ref{sec:optim\_moisture}). General discussions (Section \ref{sec:discussion}) and conclusions (Section \ref{sec:conclusions}) follow.

**\section{Data and methods}**

**\label{sec:data\_methods}**

**\subsection{Case study description}**

**\label{sec:case\_study}**

The study area is the alpine upper Rh\^{o}ne catchment in Switzerland (Fig.\ \ref{fig:map}). The altitude ranges from 372 to 4634~m.a.s.l.\ and the area is 5524~km$^{2}$. This region is the target of the MINERVE (Mod\'{e}lisation des Intemp\'{e}ries de Nature Extr\^{e}me sur les Rivi\`{e}res Valaisannes et de leurs Effets) project aimed at real-time flood management on the upper Rh\^{o}ne catchment \citep{GarciaHernandez2009b}. Even though the region is rather small, the meteorological influences related to extreme weather conditions vary substantially within it \citep[see][]{Horton2012}. Indeed, a high spatial variability of precipitation climatology exists, which is due to the complex orography of the region, and the mix of various meteorological influences. Based on different climatological analyses, the precipitation gauge stations in the catchment were clustered in ten subregions (Fig.\ \ref{fig:map}):

\begin{enumerate}

\item Swiss Chablais

\item Trient Valley

\item West Bernese Alps

\item Lower Rhone Valley

\item Southern valleys

\item Southern ridges

\item Upper Rhone Valley

\item Southeast ridges

\item East Bernese Alps

\item Conches Valley

\end{enumerate}

**\subsection{Data}**

**\label{sec:data}**

AMs rely on two types of data: predictors, which are atmospheric variables describing the state of the atmosphere at a synoptic scale, and the predictand, which is the local weather variable one wants to predict.

Predictors are generally extracted from reanalysis datasets. The NCEP-NCAR reanalysis I \citep[6-hourly, 17 pressure levels at a resolution of 2.5\degree, see][]{Kalnay1996} was used here, but it could have been any other reanalysis dataset.

The predictand (which is to be predicted) is here the daily precipitation (6~a.m. to 6~a.m. the next day) measured at the MeteoSwiss network stations, for the period 1961--2008. The time series from every available gauge station were averaged over the ten subregions (Fig.\ \ref{fig:map}), which were approximately 500~km$^{2}$ each, in order to smooth local effects \citep{Obled2002, Marty2012}. This helps account for local variability, mainly when convective processes are involved, which slightly increases the prediction skill.

It must be stressed that the predictand here is a temporally cumulated variable, compared to the meteorological predictors, which may be considered instantaneous. Depending on the duration of the accumulation period (here 24~h, but could have been 6~h, 12~h, or more than 24~h), the choice of predictors will vary.

The 48-yr precipitation dataset was divided into a calibration period (CP) and a validation period (VP). Using data independent of the CP to validate the results is very important in order to assess the robustness of the proposed improvements and to avoid over-parametrization of the method.

In order to reduce potential biases related to trends linked to climate change or to the evolution in measurement techniques, the selection of the VP was evenly distributed over the entire series \citep{BenDaoud2010}. Thus, one out of every six years was selected for validation, which represents a total of 8 years for the VP and 40 for the CP. This choice of sequence was made in order to have similar statistical characteristics between the CP and VP.

**\subsection{Analogue method}**

**\label{sec:references}**

Multiple variations of the analogue method exist, most of which are not detailed here \cite[see][for a more comprehensive listing]{BenDaoud2016}. However, there are mainly two parameterizations that are most often used for precipitation prediction and that are considered as reference: one that relies on an analogy of the atmospheric circulation, and another that adds a second level of analogy on moisture variables \citep{Obled2002, Bontron2005, Marty2012}.

The method based on the analogy of synoptic circulation consists of the following steps (Table \ref{table:params\_R1}): the similarity of the atmospheric circulation of a target date with every day of the archive is assessed by processing the S1 criterion \citep[Eq.\ \ref{eq:S1}, ][]{Teweles1954, Drosdowsky2003}, which is a comparison of gradients, over a certain spatial window:

\begin{equation}

\**label{eq:S1}**

S1=100 \frac {\displaystyle \sum\_{i} \vert \Delta\hat{z}\_{i} - \Delta z\_{i} \vert}

{\displaystyle \sum\_{i} max\left\lbrace \vert \Delta\hat{z}\_{i} \vert , \vert \Delta z\_{i} \vert \right\rbrace }

\end{equation}

where $\Delta \hat{z}\_{i}$ is the difference in geopotential height between the \textit{i}-th pair of adjacent points of gridded data describing the target situation, and $\Delta z\_{i}$ is the corresponding observed geopotential height difference in the candidate situation. The differences are processed separately in both North and East directions over the selected spatial domain. The smaller the S1 values, the more similar the pressure fields.

\citet{Bontron2005} show that the geopotential height at 500~hPa (Z500) and 1000~hPa (Z1000) are the best first predictors of the NCEP/NCAR reanalysis I dataset, and that the S1 criterion performs better than scores based on absolute distances. The reason for such better results is that the S1 criterion allows comparison of the circulation patterns, by means of the gradients, rather than the absolute value of the geopotential height, which better represent the flow direction. To cope with seasonal effects, candidate dates are extracted within a period of four months centred around the target date, for every year of the archive. This method using two geopotential heights is named 2Z.

The $N\_{1}$ dates with the lowest values of S1 are considered as analogues to the target day. The number of analogues, $N\_{1}$, is a parameter to calibrate. Then, the daily observed precipitation amount for the $N\_{1}$ resulting dates provide the empirical conditional distribution, considered as the probabilistic prediction for the target day.

The other most well-known parametrization adds a second level of analogy on the moisture variables (method 2Z-2MI, Table \ref{table:params\_R2}). The predictor that \citet{Bontron2004} found optimal for France is a moisture index made of the product of the total precipitable water (TPW) with the relative humidity at 850~hPa (RH850). \cite{Horton2012a} confirms that this index is also better for the Swiss Alps than any other variable from the NCEP/NCAR reanalysis I considered independently. When adding a second level of analogy, $N\_{2}$ dates are subsampled within the $N\_{1}$ analogues of the atmospheric circulation, to end up with a smaller number of analogue situations. When this second level of analogy is added, a higher number of analogues $N\_{1}$ is kept on the first level. Moisture fields are not as well-predicted by NWP models as pressure variables. This implies that the 2Z-2MI method, when used in real-time forecasting, is very dependent on the skill of the NWP model in predicting moisture fields, and thus its use is often restricted to the first lead times.

**\subsection{Performance assessment}**

**\label{sec:score}**

The performance assessment in the present context consists of verifying the prediction of an ensemble probabilistic technique. The set of precipitation values collected with each analogue can be considered as a sample drawn from the conditional distribution associated with the current circulation. The score that is most often used to assess an AM performance is the CRPS \citep[Continuous Ranked Probability Score,][]{Brown1974, Matheson1976, Hersbach2000}. It allows evaluating the predicted cumulative distribution functions $F(y)$, for example, of the precipitation values $y$ from analogue situations, compared to the observed value $y^{0}$. The better the prediction, the smaller the score. The mean CRPS of a prediction series of length $n$ can be written as:

\begin{equation}

\**label{eq:CRPS}**

CRPS = \frac{1}{n} \sum\_{i=1}^{n} \left( \int\_{-\infty}^{+\infty} \left[ F\_{i}(y)-H\_{i}(y-y\_{i}^{0})\right]^{2} dy \right)

\end{equation}

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 mean CRPS is averaged on the calibration, respectively, the validation periods, on all days.

In order to compare the value of the score relative to a reference, one often considers its skill score expression, and uses the climatological distribution of precipitation from the entire archive as the reference. The CRPSS (Continuous Ranked Probability Skill Score) is thus defined as follows:

\begin{equation}

\**label{eq:CRPSS}**

CRPSS = \frac{CRPS-CRPS\_{r}}{CRPS\_{p}-CRPS\_{r}} = 1-\frac{CRPS}{CRPS\_{r}}

\end{equation}

where $CRPS\_{r}$ is the CRPS value for the reference and $CRPS\_{p}$ would be the one for a perfect prediction (which implies $CRPS\_{p}~=~0$). A better prediction is characterized by an increase in CRPSS.

Note, however, that the choice of reference does not matter so much when assessing potential improvements of the method, since we consider more its relative increase or decrease rather than the CRPSS absolute value.

**\subsection{Sequential calibration}**

**\label{sec:sequential}**

AMs are usually calibrated by a semi-automatic sequential procedure, as elaborated by \citet{Bontron2004} \cite[see also ][]{Radanovics2013, BenDaoud2016}. The calibration technique optimizes the spatial windows in which the predictors are compared and the number of analogues for every level of analogy, by maximizing the performance score (CRPSS). However, the different analogy levels are calibrated sequentially, and the meteorological variables, pressure levels, and temporal windows are chosen manually. The procedure, as defined by \citet{Bontron2004}, consists of the following steps:

\begin{enumerate}

\item Manual selection of the following parameters:

\begin{enumerate}

\item Meteorological variable

\item Pressure level

\item Temporal window (hour of the day)

\item Initial analogue numbers

\end{enumerate}

\item For every level of analogy:

\begin{enumerate}

\item Identification, for the analogy level considered, of the most skilled unitary cell of all predictors jointly, over a large domain, by a full scanning of the grid.

\item From this most skilled cell, the spatial window is expanded by successive iterations in the direction of greater performance gain until no improvement is reached.

\item The number of analogue situations $N\_{1}$ is then reconsidered and optimized for the current level of analogy.

\end{enumerate}

\item A new level of analogy can then be added, based on other variables (such as the moisture index) with the chosen pressure levels, temporal windows, and initial number of analogues $N\_{2}$. The procedure starts again from step 2 (calibration of the spatial window and the number of analogues) for the new level. The parameters calibrated on the previous analogy levels are fixed and do not change.

\item Finally, the numbers of analogues $N\_{1}$ and $N\_{2}$ for the different levels of analogy are reassessed by systematic increments.

\end{enumerate}

The calibration is done in successive steps with a limited number of parameters. Previously calibrated parameters are generally not reassessed (except for the number of analogues).

This procedure was used to calibrate the methods that were here considered as references to further assess the ability of genetic algorithms to outperform the classic approach.

**\subsection{Genetic algorithms}**

**\label{sec:gas}**

Genetic algorithms (GAs) were developed by \citet{Holland1992b} and \citet{Goldberg1989}. They are part of the Evolutionary Algorithms \citep{Back1993b, Schwefel1993}, which were inspired by some mechanisms in biological evolution, such as reproduction, genetic mutations, chromosomal crossovers, and natural selection. GAs seek the global optimum on a complex surface, theoretically without restriction, which is of interest for AMs, which are characterized by a complex high-dimensional error function having multiple local optima. Practically, GAs allow rapidly approaching satisfactory solutions, but they are not guaranteed to provide the optimum solution \citep{Zitzler2004a}. It is indeed mainly a matter of time. When the optimizer gets closer to the global optimum, any new improvement takes more time to appear (see for example the slow-down of the improvements over generations in Figure \ref{fig:evolution}), and the final adjustment of the parameters can be very time consuming \citep{Back1993a}. For problems that require a significant amount of time to evaluate the objective function, as in the case of AMs (because it needs to make a prediction for every day of the CP), the number of generations has to be limited in order to ensure a reasonable processing time. Thus, different acceptable solutions can result from one or more optimization runs \citep{Holland1992b}. This is both a strength and a weakness of GAs: they are very good at exploring complex parameter spaces in order to identify the most promising areas, but they will not necessarily always find the best solution with the optimal values for all parameters \citep{Holland1992b}.

The optimizations here were performed based on the recommended GA parametrization for AMs as described in \citet{Horton2016}. As the optimization is mostly sensitive to the mutation operator (that randomly changes some values in the parameter sets), parallel optimizations are considered with variants of this operator, according to \citet{Horton2016}:

\begin{itemize}

\item 3x non-uniform mutation \citep{Michalewicz1996} with varying parameters

\item 1x multi-scale mutation \citep{Horton2016}

\item 2x chromosome of adaptive search radius \citep{Horton2016}.

\end{itemize}

A population size of 500 individuals (i.e. parameter sets of the AM to be detailed below) was considered, and the optimization was stopped when the best individual (with the highest CRPSS performance score) did not evolve for 20 generations (cycles of optimization).

**\section{Optimization of circulation analogy}**

**\label{sec:optim\_circul}**

The analogy of the atmospheric circulation was optimized for the ten subregions (Section \ref{sec:case\_study}) independently. We started from the simplest AM, and increased the complexity in order to identify the degrees of freedom that are of particular interest. Thus, the tested parametrization evolved iteratively in complexity. The detailed results of the intermediate stages are not provided in this paper \citep[see][for the details]{Horton2012a}.

The reference method for the analogy of the atmospheric circulation (2Z, Table \ref{table:params\_R1}), based on Z500 and Z1000, was first considered. The optimizer had to choose simultaneously the number of analogues, both spatial windows with no overlapping constraint (i.e. they can differ from one pressure level to another), as well as the temporal windows (hours of observation of the geopotential), which cannot be achieved with the sequential calibration technique. The performance score (CRPSS) was slightly improved, with these limited degrees of freedom, relative to the 2Z reference method calibrated with the sequential procedure. Some tests showed that most of the gains were due to the non-overlapping spatial windows. This demonstrated that the optimizer was able to obtain relevant parameters for a simple method.

Then, an additional degree of freedom was provided to the GAs by letting them choose the pressure levels along with the other parameters (analogue numbers, spatial and temporal windows), which is also a non-automated process in the sequential calibration. This degree of freedom increased the optimization time, and might decrease the number of simulations that converge to a single solution. However, most solutions were very close in terms of the performance score, which was further improved. The selected pressure levels were Z500 or Z700 for the upper level, and Z925 or Z1000 (most often) for the lower level.

Parallel analyses showed that the analogy of circulation is incomplete, and that geopotential heights still contain relevant information that can improve the statistical relationship. Therefore, a third, followed by a fourth circulation predictor were added (still only geopotential heights). There was no constraint on the predictors, so that the same pressure level could be selected more than once. Further improvements were found in the performance score, both for the CP and the VP, confirming that this additional information was beneficial for the quality of the prediction.

Finally, a weighting of the analogy criteria values per pressure level was proposed, again optimized by GAs. The weighting operates in the combination of the S1 criteria processed on every level, which were previously averaged with equal weights. The role of this new degree of freedom is to give more weight to the levels with greater predictive capacity, and to consider the differences in the geopotential height variability with altitude.

The number of circulation predictors (still only geopotential heights) was then successively increased up to ten, considering the weighting of the analogy criteria values. The addition of circulation predictors globally improved the prediction skill (for both the CP and the VP) only up to four predictors (Figure \ref{fig:figure\_nb\_levels}). Afterwards, the score on the VP was more variable, eventually even showing a decrease, which revealed an over-parametrization of the method, and thus a lack of robustness. After four predictors, the score for the CP did not increase substantially, and even presented a local decrease due to increasing difficulty for the optimizer to converge. Selecting four circulation predictors (geopotential heights) was considered optimal for this case study, since the gain in CRPSS was significant, and the model remained relatively simple. It cannot be ruled out that another number would prevail in a region other than the upper Rh\^{o}ne catchment, under other meteorological conditions, or with another reanalysis dataset.

**\subsection{Which parameters are optimized?}**

The chosen method for the atmospheric circulation analogy, based on four circulation predictors (geopotential heights), and which is named 4Zo (o for optimized), was based on the following degrees of freedom:

\begin{itemize}

\setlength\itemsep{-4px}

\item selection of pressure levels (4 degrees)

\item temporal windows (4 degrees)

\item spatial windows (4 x 4 degrees)

\item weights (4 degrees)

\item number of analogues (1 degree).

\end{itemize}

This adds up to 29 degrees of freedom that were optimized simultaneously.

**\subsection{Results for 4Zo method}**

The resulting optimized parameters for 4Zo vary from one subregion to another. The optimized spatial windows are given for every subregion in Figure \ref{fig:spatial\_windows\_4Zo}, and the selected pressure levels in Table~\ref{table:levels\_GA\_4Zo}.

The resulting CRPSS scores are provided in Figure \ref{fig:figure\_crpss\_4Zo} and were on average 35.8\% for the CP and 35.5\% for the VP, compared to 31.1\% and 32.3\%, respectively, for the reference method 2Z on the atmospheric circulation (optimized by the sequential procedure). The score was also calculated for three precipitation thresholds: P \(\geq\) 1~mm, P \(\geq\) 0.1\(\cdot\)P10, and P \(\geq\) 0.5\(\cdot\)P10, P10 being the daily precipitation with a 10~year return period (Table \ref{table:scores\_thresholds\_4Zo}). The gain in score increased with the precipitation threshold: the relative improvement of the CRPSS was, on average, for the different thresholds, 13.3\%, 15.4\%, and 29.1\% for the CP and 7.9\%, 11.1\%, and 34.5\% for the VP. The optimization thus improved the prediction even more for days with significant precipitation than for the usual days.

To assess the parameters’ cross-compatibility and the spatial coherence of the resulting parameters, those optimized for one subregion were applied to the others. The resulting losses or gains of the CRPSS are displayed in Figure \ref{fig:crossing\_4Zo}.

**\subsection{Analysis}**

The automatic selections of pressure levels (Table \ref{table:levels\_GA\_4Zo}) and temporal windows (not shown) for the analogy of circulation showed a great homogeneity and were spatially consistent. First of all, the level Z1000 was always selected twice (the first time at 6 or 12~h, and the second always at 30~h) and Z700 was selected once for every subregion (always at 24~h). The level that varied from one subregion to another, albeit in a spatially consistent way, was the upper level (always at 12~h), which was Z300 for the north-west part of the catchment, Z500 for most of the other subregions, and Z600 for the Conches Valley. The optimizer thus provided consistent selections of pressure levels and temporal windows. The automatic selection of pressure levels is a big advantage in favour of global optimization.

The resulting spatial windows (Figure \ref{fig:spatial\_windows\_4Zo}) may look very diverse first, but there are significant similarities for subregions located within the same vicinity. The first four subregions were characterized by a large spatial window on the upper level, whereas it was smaller elsewhere. For most subregions, the second level (Z700) was compared on thin and longitudinally extended spatial windows. The third level (Z1000 at 6 or 12~h) also had longitudinally extended domains, but were slightly larger. The last one (Z1000 at 30~h) had rather large and squared windows. Subregions number 5 (southern valleys) and 6 (southern ridges) had exactly the same spatial windows, which suggests that they behave in a similar way and thus could have been merged. This similarity is a good sign for the accuracy of the optimized parameters.

The performance scores showed non-negligible improvements for both the CP and VP (Figure \ref{fig:figure\_crpss\_4Zo}) compared to the 2Z reference method optimized by the sequential procedure. Even more interestingly, the results for higher precipitation thresholds (Table \ref{table:scores\_thresholds\_4Zo}) showed the largest improvements. This is of particular interest in the framework of flood forecasting. The further improvement of days with higher precipitation totals is likely related to the fact that larger values contribute more to the CRPS score, which means that better predicting these days results in significant increase in the global performance score.

The analysis of the parameters’ cross-compatibility showed that the parameters were obviously optimal on the CP for the subregion for which they were optimized (Figure \ref{fig:crossing\_4Zo} top). However, the losses in CRPSS when exchanging the parameters were not of the same magnitude among the different subregions. Indeed, the Upper Rhone Valley (7) and, moreover, the southeast ridges (8) seemed to behave significantly differently. These two regions have different climatic properties than the others, as they are particularly sensitive to southerly flows. Indeed, almost all heavy precipitation events occurred under a southerly regime, such as in the Liguria, Piedmont, and Aosta regions in Italy, whereas the other subregions of the catchment had extreme events mainly under a westerly regime \citep{Horton2012}. Thus, as the performance score is significantly influenced by heavy precipitation values, the parameters for the different subregions are likely optimized to better predict these days. It can then be expected that the optimal parameters differ between these two subregions and the others. This points at the importance of taking into account leading meteorological influences during precipitation station clustering, which are not always best represented by geographical distance.

Globally, the same cross-compatibility structure could be observed for the VP (Figure \ref{fig:crossing\_4Zo} bottom), but in this case, minor improvements were occasionally observed when crossing the parameters, because of the presence of other events in the VP that might be better predicted by a different parameter set. The relatively small differences in scores between parameterizations indicated that even though the parameters might differ significantly, the performance might not be drastically affected. Even a change in the pressure level did not mean a radical drop in the score value. A different parametrization may lead to a distinct selection of analogue days, and thus to an improvement of the prediction under certain weather conditions at the expense of others.

**\section{Optimization of analogy with moisture information}**

**\label{sec:optim\_moisture}**

It is known that moisture variables as a second level of analogy do provide improvements to the method (section \ref{sec:references}). The moisture index, which is a combination of the relative humidity and precipitable water, has thus also to be optimized. In order to do so, a constraint on the optimizer had to be introduced, so as to select the same temporal window (time of observation) for both variables.

Two methods were assessed: one with two moisture predictors (moisture index on two pressure levels or at two different hours), named 4Zo-2MIo, and one with four moisture predictors, named 4Zo-4MIo. When introducing two predictors for the moisture analogy, the number of degrees of freedom increased to 42, and to 54 with four predictors. However, there was no substantial difference in the performance scores between both 4Zo-2MIo and 4Zo-4MIo methods, which suggests that considering four moisture predictors is not necessary. For this reason, only the results of 4Zo-2MIo are presented.

The optimization was processed on both levels of analogy simultaneously. This implies that the analogy of the atmospheric circulation could change because of the new moisture information.

**\subsection{Results for 4Zo-2MIo method}**

As seen previously, the optimized parameters differed from one subregion to another, but to an even greater extent. The resulting spatial windows are displayed in Figure \ref{fig:spatial\_windows\_4Zo-2MIo} for 4Zo-2MIo, along with the selected pressure levels for both the circulation and moisture analogy (Table \ref{table:levels\_GA\_4Zo\_2MIo}).

The CRPSS scores of the optimized 4Zo-2MIo method are provided in Figure \ref{fig:figure\_crpss\_4Zo-2HIo} and amounted on average to 40\% (CP) and 40.3\% (VP), compared to 35.2\% (CP) and 36.2\% (VP) for the reference method 2Z-2MI on the moisture analogy optimized with the sequential procedure. The parameters’ cross-compatibilities are shown in Figure \ref{fig:crossing\_4Zo-2MIo}. As for 4Zo, the 4Zo-2MIo method presented larger improvements in the prediction of significant rainfall (Table \ref{table:scores\_thresholds\_4Zo-2MIo}).

**\subsection{Analysis}**

When optimizing a method consisting of two levels of analogy, the introduction of moisture variables in the second level has an influence on the parameter values of the first level. This means that the two levels of analogy bring complementary information, and are thus not independent. This is first visible in the number $N\_{1}$ of analogues to be selected on the first level, and in the selection of the pressure levels for the circulation analogy. If the change in the optimal value of $N\_{1}$ was already known, a change in the optimal pressure levels for the circulation analogy has never been identified before.

As for the sequential procedure, the optimal value of $N\_{1}$ increased when adding a second level of analogy (Figure \ref{fig:figure\_nb\_analogs}). One can also see that the optimal number of analogues $N\_{2}$ for the second level of analogy of 4Zo-2MIo was slightly inferior to $N\_{1}$ from 4Zo, but very close. There is a globally common tendency between the optimal analogue number values of both methods: $N\_{1}$ of the 4Zo method, and $N\_{1}$ and $N\_{2}$ of 4Zo-2MIo tend to be higher or lower together for a given region.

The optimal final numbers of analogues did not vary much: $23 \leq N\_{1} \leq 33$ for 4Zo and $21 \leq N\_{2} \leq 28$ for 4Zo-2MIo. However, the optimal number of the $N\_{1}$ analogues of the first level of 4Zo-2MIo varied to a greater extent: $48 \leq N\_{1} \leq 84$. In this latter method, it may be problematic to consider a fixed and unique value for all regions.

As for the pressure levels, Z1000, which was previously systematically selected twice (Table \ref{table:levels\_GA\_4Zo}) was here less often chosen (once or even not at all) for 4Zo-2MIo (Table \ref{table:levels\_GA\_4Zo\_2MIo}). There was indeed a vertical shift in the previously selected Z1000 for higher levels that was even slightly stronger with four moisture predictors than with two. This change is likely due to the fact that when considering only the circulation analogy, the method tried to take into account information that can serve as a proxy for moisture assessment, whereas it did not need it with the moisture index. This can only be assessed by a global optimization technique that can work jointly on both levels of analogy.

The selected pressure levels for the analogy of the moisture index were strongly centred around 700~hPa and 600~hPa. No other value was selected when considering two moisture predictors (Table \ref{table:levels\_GA\_4Zo\_2MIo}). It was sometimes more efficient, in terms of prediction performance, to consider the moisture at 700~hPa twice, but at different hours, rather than selecting another pressure level. Besides, the optimizer never chose the same pressure level at the same hour for any variable, even though it was allowed to do so. The selected pressure levels for the moisture analogy differed from the reference method (Tables \ref{table:params\_R2} and \ref{table:levels\_GA\_4Zo\_2MIo}, last row).

The selection of temporal windows for atmospheric circulation was similar to the preceding optimization (in order of increasing pressure: 12~h, 24/30~h, 12~h, 30~h), but sometimes with some variability. When it comes to the moisture analogy, there was a clear tendency to select 12~h and 24~h. However, it must be remembered that this holds for our predictand, the accumulated precipitation over 06--30~h~UTC, and that it is expected to differ if the temporal window changes (e.g. 00--24~h~UTC, or another accumulation duration).

The optimized spatial windows for the atmospheric circulation have also changed (Figure \ref{fig:spatial\_windows\_4Zo-2MIo}). The very large domains on the upper level of the first four subregions were not present anymore, and more variability could be observed. The selected points for the moisture analogy were always located near the catchment, including at least one of the nearest points from the reanalysis dataset, and the spatial windows were relatively small. Thus, for this case study, there is no need to look for distant moisture information, and the search could be reduced to a smaller domain.

The CRPSS scores were improved by considering the moisture information (Figure \ref{fig:figure\_crpss\_4Zo-2HIo} to be compared with Figure \ref{fig:figure\_crpss\_4Zo}). The optimized method also performed significantly better than the 2Z-2MI reference method optimized by the sequential procedure. When it comes to improvements for days with precipitation above the three thresholds (P\(\geq\)1~mm, P \(\geq\) 0.1\(\cdot\)P10, and P \(\geq\) 0.5\(\cdot\)P10), the conclusion is the same as before, that is, a significant improvement in the prediction compared to the reference method, mainly for heavy rainfall.

The analysis of the parameters’ cross-compatibility (Figure \ref{fig:crossing\_4Zo-2MIo}) was also very similar to the one of the circulation analogy only. The same pattern could be observed, with a drop of performance for the subregions submitted to different meteorological influences. However, the losses in performance were globally more important than before, suggesting that more complex methods with moisture variables are less transposable to another subregion (consistent with the observations of \citet{Chardon2014}), even though both were located within the same grid cell of the reanalysis dataset. Moisture fields have greater variability than pressure fields, and thus a change in the spatial windows can have a greater impact on the method performance. Indeed, the two regions with the lowest cross-compatibility with the others were the upper Rhone Valley (7) and the southeast ridges (8), which had similar optimal pressure levels and temporal windows to other regions, but had rather different spatial windows on the moisture predictor.

Predictors based on moisture variables do significantly increase the prediction skill, and are thus recommended, as long as they are reliable. In real-time forecasting, their reliability depends on the lead time: for lead times superior to 3--4 days, the uncertainties related to moisture variables from NWP models become fairly high, which reduces the relevance of methods relying on this information. In climate downscaling studies, it mainly depends on the coherence of the climatologies between the archive and the GCM model outputs. One should, however, not establish an AM with moisture variables for too large a region, as the transferability is reduced \citep[see][for alternative approaches]{Chardon2014}.

**\section{Discussion}**

**\label{sec:discussion}**

The optimization of the AM by means of GAs has been undertaken in successive stages by releasing progressively new degrees of freedom. This approach allowed us to differentiate the contributions to performance gains, as well as to identify possible over-parametrization. The main improvements obtained in the present case study are due to the following elements:

\begin{itemize}

\item Using four pressure levels for the circulation analogy seemed to be an optimal number for the studied region, length of archive available, and target predictand considered. Beyond four, the validation score was more variable, revealing a loss in robustness due to over-parametrization.

\item The automatic and joint optimization of all parameters: the analogue numbers, selection of pressure levels and temporal windows, and spatial windows. These parameters are highly interdependent, so one needs to optimize them jointly in order to identify optimal combinations. Indeed, there is a strong interdependence between space and time in the atmospheric circulation, so that, e.g. the spatial window should move upstream in the main atmospheric flow for earlier temporal windows.

\item The introduction of distinct spatial windows between pressure levels. The synoptic circulation is characterized by features with very different scales depending on the height, and important information for predicting precipitation is not necessarily located in the same area from one level to another.

\item The weighting of the analogy criteria between different pressure levels. This can be influenced by the variability of the geopotential height with altitude, or the levels of significance in regards to the meteorological processes specific to a region. There is a trend in the weighting of circulation predictors to decrease with the increase in pressure, as one can see in Figure \ref{fig:levels\_weights\_average} for the three optimized methods. However, the values remained approximately equal. This may not be the most influencing factor, and we may suggest removing it first when trying to reduce the degrees of freedom.

\item The joint optimization of the circulation and moisture analogy levels, which are usually calibrated successively. We were able to demonstrate that there is a dependency between the analogy levels, and that in order to approach the optimal parameters, one must consider them jointly.

\end{itemize}

GAs have proved very useful to optimize complex variants of the AM, and to assess new degrees of freedom that were not available thus far. However, it can be dangerous to add too many parameters to optimize. Indeed, the optimizer will probably use them to successfully improve the calibration score, so the validation control remains very important in order to determine if one is actually improving the method, or if it is being over-parametrized. Moreover, it might not always be desirable to increase the degrees of freedom, and some constraints (e.g. same weighting of the analogy criteria between different pressure levels) can be justified. However, one should first assess the consequence of a constraint before establishing it. In this sense, even though not all degrees of freedom are useful, GAs allow us to assess their influence. Finally, GAs could be used to identify, among other things, the best pairs of spatial and temporal windows, in order to later create a simpler regional method.

The convergence of parallel optimizations decreased when the optimization method became more and more complex. The optimizer did not always converge to the exact global optimum, but to its surroundings. This is related to the fact that the optimization slows down when it gets closer to the global optimum, and that one has to stop it before the end, because of the required processing time (see Figure \ref{fig:evolution} as an example). The resulting parameters might sometimes present non-negligible differences, even though the score is similar. Through Monte-Carlo analyses of the parameter space properties of the AM, \citet{Horton2012a} showed that some parameters of the method have a wide range of acceptable values. The spatial windows, for example, can be larger than the optimal size without much impact on the score, while they cannot be smaller \citep[see also][]{Bontron2004}. We also observed that the selection of pressure level is not a parameter as discrete as we would have thought, and that choosing another level may have reduced the consequences on the performance. This is particularly true for higher pressure levels, but can be more critical for lower layers. It was thus interesting to sometimes obtain several sets of near-optimal parameters, but with some nuances, in order to get an idea of the sensitivity of the parameters for a given region, and to compare the score for the VP. In this regard, a cross-validation technique may be advisable. However, as solutions identified at different regions of the parameter space might provide sufficiently good performance, an ensemble of these could be used, instead of a unique solution. These could account for the parameters’ uncertainty in the AM and increases in the sample size contributing to the empirical distribution of precipitation values. An approach that can also be recommended is to first explore a wide range of the parameter space with some optimizations, and to narrow it according to the results for more targeted optimizations that are likely to go faster and to perform better.

We tried to optimize the length of the preselection period (i.e. the seasonal stratification, which was a 4-month window leading up to the present) jointly with the other parameters, but no improvement was observed. Optimizing the moisture flux, which is composed of the moisture index multiplied by the wind flux, was also assessed. However, the results were not better than when considering the moisture index alone. This may be related to the fact that the optimizer tries to provide the best analogy of the atmospheric circulation in the first place, which makes the wind information less relevant in the second level of analogy.

As has been observed, methods with higher complexity that integrate moisture predictors are less transposable than simpler ones. It was also noticed in another unpublished work, that it is by far better to optimize for two subregions jointly than to optimize on one and to apply its parametrization to the other. Finally, the discretization in subregions is an important process and should be handled with care. Indeed, the geographical distance is not always the leading factor to define a subregion. For example, the southeast ridges subregion does not behave like its surroundings, and differs in its parametrization because of different leading meteorological influences.

GAs are relatively heavy on processing and require an IT infrastructure capable of performing thousands of hours of calculations. However, they automatically optimize all parameters of the AM, which is not possible with sequential calibration. Therefore, much human time, previously required to successively assess numerous combinations of parameters (particularly the selection of pressure levels and temporal windows),is saved. The ability to jointly optimize all parameters is important given the strong dependencies between them and between the levels of analogy.

Furthermore, AMs optimized with GAs showed an improvement in predictions for days with heavier precipitation, including extremes. Even though no new extreme value was added to the existing precipitation archive, the distribution of analogue precipitation values for a target situation can move towards the targeted extreme by sampling better candidate situations. Then, the subset of precipitation values collected on the analogue dates can be considered as a sample of the conditional distribution of precipitation associated with this situation. A truncated exponential or a gamma distribution model can be fit and extrapolated to extreme values not contained in the sample or even in the whole precipitation archive \citep{Obled2002}. Another possible approach is to combine AMs with other methods \citep[e.g.][]{Chardon2014}.

**\section{Conclusions and perspectives}**

**\label{sec:conclusions}**

The parameters resulting from the optimization by GAs were very consistent in terms of the selection of pressure levels and temporal and spatial windows. There were clear trends or even identical results for subregions under similar meteorological influences, which confirm that the optimized parameters were coherent, despite an eventual first impression of great variability in the spatial windows. When adding moisture variables, the results showed a higher variability, but remained highly acceptable and coherent.

Strong dependencies between the parameters of the AM were observed. Thus, the sequential calibration, which optimizes the parameters successively, may not lead to the optimal combination. Moreover, it contains several manual systematic assessments, such as the selection of pressure levels and temporal windows. GAs, however, can automatically select pressure levels and temporal windows, which can save a considerable amount of human time. A great advantage of a global optimization is its ability to approach or reach optimal parameter values when they are considered jointly.

A dependence in the selected parameters between the circulation analogy and moisture variables was identified. When the two analogy levels are considered together, the optimal parameters of the circulation analogy changed. This complexity can only be exploited in a suitable manner by global optimization methods.

For the present case study, there seemed to be an optimum number of pressure levels to consider for the circulation analogy, which is four, before losing consistency of the real gains. The circulation analogy was improved by introducing a weighting between pressure levels, and considering independent spatial windows between pressure levels.

GAs provided parameterizations of AMs that exceeded the performance of sequential calibration. In addition, it has been observed that the prediction for days with strong precipitation were improved to a greater extent, which is clearly interesting in the context of flood forecasting.

This work is by no means exhaustive, and is meant to open a door to new explorations of AMs with GAs. It is even possible to let the optimizer chose the meteorological variable to be used as a predictor, as well as the analogy criteria, which is the topic of work in progress. Moreover, the AM has been explored for decades for precipitation prediction, but not as intensively for other predictands. A global optimizer, such as a GA, can speed up this assessment significantly.

**\section\*{Acknowledgments}**

We thank Hamid Hussain-Khan of the University of Lausanne for his help and availability, and for the intensive use of the cluster he administers. We also thank Renaud Marty for fruitful collaboration over the years, and Dominique B\'{e}rod for his support.

We thank the Roads and Water courses Service, the Energy and Water Power Service of the Wallis Canton, the Water, Land and Sanitation Service of the Vaud Canton, and the Swiss Federal Office for Environment (FOEV) for financing the MINERVE (Mod\'{e}lisation des Intemp\'{e}ries de Nature Extr\^{e}me des Rivi\`{e}res Valaisannes et de leurs Effets) project which started this research. Fruitful collaboration with the Laboratoire d'Etude des Transferts en Hydrologie et Environnement of the Grenoble Institute of Technology (G-INP) was made possible thanks to the Herbette Foundation. NCEP reanalysis data were provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, via their website at http://www.esrl.noaa.gov/psd/. The precipitation time series were provided by MeteoSwiss.

The authors would also like to acknowledge the work of Loris Foresti and another anonymous reviewer, which contributed to improving this paper.

**\section\*{References}**

%\bibliography{references}

\**bibliography{../\_refs/4\_articles-2016\_JoH\_optimizations}**

\clearpage

%% FIGURES

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig01.pdf}}

\caption{Location of alpine Rh\^{o}ne catchment in Switzerland. (source: Swisstopo)}

**\label{fig:map}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig02.pdf}}

\caption{Performance score (CRPSS) for CP and VP for three stations (1 - Swiss Chablais; 5 - Southern valleys; 8 - Southeast ridges) when varying the number of geopotential height levels available to the optimizer.}

**\label{fig:figure\_nb\_levels}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.4cm]{figures/fig03.pdf}}

\caption{Optimized spatial windows for 4Zo method (analogy of atmospheric circulation on four pressure levels).}

**\label{fig:spatial\_windows\_4Zo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig04.pdf}}

\caption{Performance score (CRPSS) of reference methods 2Z (Table \ref{table:params\_R1}) and optimized 4Zo method for CP and VP for every subregion.}

**\label{fig:figure\_crpss\_4Zo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.4cm]{figures/fig05.pdf}}

\caption{Losses or gains (in \%) in CRPSS from applying optimized parameters for the series in columns to those in rows. Method 4Zo, calibration and validation periods.}

**\label{fig:crossing\_4Zo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.4cm]{figures/fig06.pdf}}

\caption{Optimized spatial windows for 4Zo-2MIo method (analogy of atmospheric circulation on four pressure levels and analogy on the moisture index on two pressure levels).}

**\label{fig:spatial\_windows\_4Zo-2MIo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig07.pdf}}

\caption{Performance score (CRPSS) of reference methods 2Z-2MI (Table \ref{table:params\_R2}) and optimized 4Zo-2MIo method for CP and VP for every subregion.}

**\label{fig:figure\_crpss\_4Zo-2HIo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig08.pdf}}

\caption{Optimal numbers of analogues for different regions and the two methods, resulting from optimization. Method 4Zo is made of a single level of analogy with $N\_{1}$ analogues, whereas 4Zo-2MIo has two levels of analogy with $N\_{1}$ and $N\_{2}$ analogues.}

**\label{fig:figure\_nb\_analogs}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.4cm]{figures/fig09.pdf}}

\caption{Losses or gains (in \%) of the CRPSS from applying optimized parameters for the series in columns to those in rows. Method 4Zo-2MIo, calibration and validation periods.}\*\*\*\*\*

**\label{fig:crossing\_4Zo-2MIo}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=8.8cm]{figures/fig10.pdf}}

\caption{Distribution of optimal weights for predictors of first level of analogy (geopotential heights) of (blue) 4Zo, (red) 4Zo-2MIo, and (green) 4Zo-4MIo methods. Results are for the ten subregions. Geopotential heights are sorted by increasing pressure and hour.}

**\label{fig:levels\_weights\_average}**

\end{figure}

\begin{figure}[t]

\centerline{**\includegraphics**[width=7.5cm]{figures/fig11.pdf}}

\caption{Example of evolution of performance score of the best individual over eight independent optimizations.}

**\label{fig:evolution}**

\end{figure}

\clearpage

%% TABLES

\begin{table}[t]

\caption{Parameters of reference method on atmospheric circulation (2Z). First column is level of analogy (0 for preselection); subsequent columns list meteorological variable, its hour of observation within the target day (temporal window), criterion used for current level of analogy, and number of analogues.}

\footnotesize

\begin{center}

\begin{tabular}{ccccc}

\hline

Level & Variable & Hour & Criterion & Nb \\

\hline

0 & \multicolumn{4}{l}{$\pm 60$ days around the target date} \\

\hline

\multirow{2}{\*}{1} & Z1000 & 12~h & \multirow{2}{\*}{S1} & \multirow{2}{\*}{50} \\

& Z500 & 24~h & & \\

\hline

\end{tabular}

\end{center}

**\label{table:params\_R1}**

\end{table}

\begin{table}[t]

\caption{Parameters of reference method with moisture variables (2Z-2MI). Conventions are the same as in Table \ref{table:params\_R1}}

\footnotesize

\begin{center}

\begin{tabular}{ccccc}

\hline

Level & Variable & Hour & Criterion & Nb \\

\hline

0 & \multicolumn{4}{l}{$\pm 60$ days around the target date} \\

\hline

\multirow{2}{\*}{1} & Z1000 & 12~h & \multirow{2}{\*}{S1} & \multirow{2}{\*}{70} \\

& Z500 & 24~h & & \\

\hline

\multirow{2}{\*}{2} & TPW \* RH850 & 12~h & \multirow{2}{\*}{RMSE} & \multirow{2}{\*}{30} \\

& TPW \* RH850 & 24~h & & \\

\hline

\end{tabular}

\end{center}

**\label{table:params\_R2}**

\end{table}

\begin{table}[t]

\caption{Pressure levels ($\sim$) automatically selected for 4Zo method for different subregions (ID). R represents 2Z reference method (Table \ref{table:params\_R1}).}

\footnotesize

\begin{center}

\begin{tabular}{ccccccccc}

\hline ID & 300 & 400 & 500 & 600 & 700 & 850 & 925 & 1000 \\

\hline

1 & $\sim$ & & & & $\sim$ & & & $\sim \sim$ \\

2 & $\sim$ & & & & $\sim$ & & & $\sim \sim$ \\

3 & $\sim$ & & & & $\sim$ & & & $\sim \sim$ \\

4 & $\sim$ & & & & $\sim$ & & & $\sim \sim$ \\

5 & & & $\sim$ & & $\sim$ & & & $\sim \sim$ \\

6 & & & $\sim$ & & $\sim$ & & & $\sim \sim$ \\

7 & & & $\sim$ & & $\sim$ & & & $\sim \sim$ \\

8 & & & $\sim$ & & $\sim$ & & & $\sim \sim$ \\

9 & & & $\sim$ & & $\sim$ & & & $\sim \sim$ \\

10 & & & & $\sim$ & $\sim$ & & & $\sim \sim$ \\

\hline

R & & & $\sim$ & & & & & $\sim$ \\

\hline

\end{tabular}

\end{center}

**\label{table:levels\_GA\_4Zo}**

\end{table}

\begin{table}[t]

\caption{Relative improvement (\%) in CRPSS for different precipitation thresholds for optimized 4Zo method, compared to reference method.}

\footnotesize

\begin{center}

\begin{tabular}{ccccccc}

\hline

ID & \multicolumn{2}{c}{P\(\geq\)1 mm} & \multicolumn{2}{c}{P\(\geq\)0.1\(\cdot\)P10} & \multicolumn{2}{c}{P\(\geq\)0.5\(\cdot\)P10} \\

& CP & VP & CP & VP & CP & VP \\

\hline

1 & 10.2 & 9.4 & 8.5 & 7.9 & 17.0 & 14.2 \\

2 & 9.9 & 3.4 & 10.2 & 7.3 & 19.3 & 13.7 \\

3 & 13.3 & 10.5 & 13.3 & 10.9 & 19.7 & 9.7 \\

4 & 11.0 & 7.4 & 12.9 & 10.0 & 23.2 & 23.8 \\

5 & 8.6 & 4.2 & 10.9 & 6.2 & 25.2 & 23.8 \\

6 & 10.5 & 5.1 & 11.1 & 7.1 & 21.2 & 41.1 \\

7 & 24.3 & 12.4 & 33.1 & 26.0 & 71.2 & 104.3 \\

8 & 19.0 & 12.7 & 26.2 & 19.2 & 39.4 & 34.9 \\

9 & 12.4 & 6.8 & 13.8 & 9.9 & 24.9 & 48.1 \\

10 & 13.6 & 6.8 & 14.4 & 6.9 & 29.9 & 31.5 \\

\hline

av. & 13.3 & 7.9 & 15.4 & 11.1 & 29.1 & 34.5 \\

\hline

\end{tabular}

\end{center}

**\label{table:scores\_thresholds\_4Zo}**

\end{table}

\begin{table}[t]

\caption{Atmospheric levels automatically selected for analogy of atmospheric circulation ($\sim$) and moisture analogy ($\bullet$) of 4Zo-2MIo method, for different subregions (ID). R represents 2Z-2MI reference method (Table \ref{table:params\_R2})}

\footnotesize

\begin{center}

\begin{tabular}{ccccccccc}

\hline ID & 300 & 400 & 500 & 600 & 700 & 850 & 925 & 1000 \\

\hline

1 & $\sim$ & & $\sim$ & & $\bullet \bullet$ & $\sim$ & & $\sim$ \\

2 & $\sim$ & & & & $\sim \bullet \bullet$ & $\sim$ & & $\sim$ \\

3 & $\sim$ & & & & $\sim \bullet \bullet$ & $\sim$ & $\sim$ & \\

4 & & & $\sim$ & $\bullet$ & $\sim \bullet$ & $\sim$ & & $\sim$ \\

5 & & $\sim$ & & & $\sim \bullet \bullet$ & & $\sim \sim$ & \\

6 & & $\sim$ & & $\bullet$ & $\sim \bullet$ & $\sim$ & & $\sim$ \\

7 & & $\sim$ & & $\bullet$ & $\sim \bullet$ & $\sim$ & & $\sim$ \\

8 & & & $\sim$ & $\bullet$ & $\sim \bullet$ & & $\sim \sim$ & \\

9 & & $\sim$ & & $\bullet$ & $\sim \bullet$ & $\sim$ & $\sim$ & \\

10 & & $\sim$ & & $\bullet$ & $\sim \bullet$ & $\sim$ & & $\sim$ \\

\hline

R & & & $\sim$ & & & $\bullet \bullet$ & & $\sim$ \\

\hline

\end{tabular}

\end{center}

**\label{table:levels\_GA\_4Zo\_2MIo}**

\end{table}

\begin{table}[t]

\caption{Relative improvement (\%) in CRPSS for different precipitations thresholds for optimized 4Zo-2MIo method, compared to reference method.}

\footnotesize

\begin{center}

\begin{tabular}{ccccccc}

\hline

ID & \multicolumn{2}{c}{P\(\geq\)1 mm} & \multicolumn{2}{c}{P\(\geq\)0.1\(\cdot\)P10} & \multicolumn{2}{c}{P\(\geq\)0.5\(\cdot\)P10} \\

& CP & VP & CP & VP & CP & VP \\

\hline

1 & 12.6 & 9.3 & 12.4 & 9.7 & 15.8 & 11.0 \\

2 & 10.4 & 7.7 & 11.2 & 10.5 & 18.9 & 16.6 \\

3 & 14.5 & 11.6 & 14.1 & 11.4 & 18.7 & 14.6 \\

4 & 11.4 & 9.4 & 11.5 & 11.6 & 14.9 & 22.7 \\

5 & 11.8 & 8.0 & 12.2 & 8.9 & 12.0 & 12.8 \\

6 & 11.3 & 7.1 & 11.2 & 8.0 & 15.3 & 29.1 \\

7 & 20.5 & 15.5 & 25.2 & 24.0 & 43.0 & 79.5 \\

8 & 19.3 & 15.7 & 23.1 & 18.6 & 25.2 & 31.7 \\

9 & 17.0 & 15.4 & 17.4 & 16.5 & 23.7 & 39.4 \\

10 & 12.9 & 9.6 & 13.8 & 11.1 & 28.5 & 32.1 \\

\hline

av. & 14.2 & 10.9 & 15.2 & 13.0 & 21.6 & 28.9 \\

\hline

\end{tabular}

\end{center}

**\label{table:scores\_thresholds\_4Zo-2MIo}**

\end{table}

\end{document}