

## Response to Reviewer #1

The manuscript „Using genetic algorithms to optimize the Analogue Method for precipitation downscaling in the Swiss Alps" by Horton et al. addresses the problem of precipitation prediction by analogue methods (AM). The main goal of the paper is to present and test a new optimization method based on Genetic Algorithms (GA) which has in my opinion two main advantages to the traditional sequential procedure: (a) it optimizes all predictors jointly and independently; and (b) it allows for a free selection of the spatial integration window (and time) and can give weights to different pressure levels. The paper shows improvement compared to a reference method optimized by the sequential procedure. The analysis is conducted over the Rhone Basin in Switzerland divided into 10 different regions. Predictors come from NCEP-NCAR reanalysis, predicted precipitation comes from Meteoswiss stations averaged over 500 km<sup>2</sup> grids.

My main impression is that this is solid work for people interested in improving AM for predicting precipitation. As far as I can tell methodologically the paper is correct and in my opinion the results are well explained. But I do have a concern about the interest for a general hydrological audience and suggest that the authors in their revision try to make the paper shorter and more focussed on the message and expand some discussion on the actual usefulness and application of the method to precipitation prediction. Also a clearer statement on the novelty of this paper compared to the recommendations of Horton et al. (2016b) is necessary (line 40-41). The paper can be published in JH after revision. My main questions/issues which the authors may wish to consider and/or expand on are listed below.

Thank you for your positive feedback.

We removed many figures and tables and some secondary analyses in order to focus on the main message. We also added applications examples for hydrology mainly in the introduction. Please note that this paper is part of the special issue "Precipitation measurement and modeling: uncertainty, variability, observations, ensemble simulation and downscaling", and so the focus on precipitation is justifiable in our view.

Horton et al. (2016b) focused on the parametrization of GAs in order to optimize AMs successfully, not on their application on more elaborated AMs. This has been clarified.

1. Problem of non-uniqueness. It is true that the global optimization method provides more freedom in the combination of parameters (predictors) and therefore performance. However this increase in the degrees of freedom is to me not necessarily a win-win situation. As the authors point out several different combinations of parameters may now lead to similar model performance, especially in the two analogue version. I do not see a constraint on parameters as necessarily negative. I have a feeling the authors could engage with this question more in their work.

We agree and added some sentences in the discussion (section 5): l. 575-580

2. The use for prediction. I am not very familiar with applications of AM for prediction of precipitation, but as I understood it from the paper only observed precipitation amounts and patterns can be reproduced (from the historical archive), even though N1 analogues can be selected and a "probabilistic forecast" from them derived. Do these N1 cases cover the entire uncertainty, including high precipitation (extremes) which have not been observed? In other words, what is the advantage of these methods against stochastic predictive models conditioned on similar (circulation and moisture) predictors? Do the authors have an opinion on this?

The following has been added in the discussion about extremes: l. 635-645

The following has been added in the introduction about advantages of the AMs: l. 46-51

3. RCM is not an option? Although I agree that GCM precipitation performance is poor, in all fairness I think it should be mentioned in the Introduction (lines 11-20) that limited area RCMs are an alternative at space and time scales that are getting pretty close to station data, or rather to the spatial averaging scale for precipitation used in this study.

Some sentences were added in the introduction (l. 37-45).

4. Methodology. I appreciate the approach in Chapter 5 starting from the two fixed pressure level reference and then releasing constraints, i.e. adding degrees of freedom, and eventually a second analogue (moisture). Although I agree that the added level of performance is not incredible, there is measurable improvement in the calibration. There is also consistency in the results, in that some pressure levels, etc., are consistently chosen over others. Finally it was concluded that 4 predictors are optimal because more dropped the performance in validation and the authors considered the model then over-parameterized. I would like to read some physical explanation of this feature in the paper connected to the fields of pressure and moisture and precipitation formation in the Alps. Why should it be that including more parameters, if they are responsible for rain formation, becomes eventually counterproductive?

About the optimal number of 4 predictors: the text was maybe not so clear, so it has been reworked (l. 319-334) and illustrated in figure 2. The general term “predictor” might be ambiguous here, because only geopotential heights were considered at this stage. So, there is no moisture predictors at this stage, only atmospheric circulation data. As shown in figure 2, with 4 geopotential heights, most of the predictive information seem to be integrated. Adding other thermodynamic predictors will increase the prediction skill, as shown with the moisture index.

5. Performance dependent on precipitation depth. The authors present that the performance of their GA-AM method is best for days with heavy rain. I am wondering what is the reason for this, especially considering that I would also expect the predictand derived from station observations averaged over 500 km<sup>2</sup> grids to be most uncertain then. Meaning similar pressure/moisture analogues may lead to more different precipitation on the ground during heavy rain which is concentrated in space, but this would not reflect in the grid-averaged precipitation because the spatial density of gauges is not high enough to capture the large spatial variability all of the time.

As mentioned to reviewer #2, larger precipitation values contribute more to the performance score. Thus, GAs will optimize these days more thoroughly than the sequential calibration as they are more powerful and can handle supplementary degrees of freedom. The following sentence was added in section 3.3: l. 390-393

6. Cross-compatibility. This analysis (lines 263-265, 431-438) is really interesting. Is there any spatial pattern evident in the performance drop in space that the authors find physically explainable? There is an attempt starting on line 308 which I do not fully follow.

Yes, it is related to different climate properties of the subregions, mainly for heavy precipitation events. We developed a bit further this point in section 3.3 (l. 394-410) and we hope it is now more understandable.

7. Please explain in a sentence what accuracy and sharpness are (line 295) so that readers don't have to go to Bontron (2004).

For the sake of conciseness, the analysis of the CRPS decomposition into sharpness and accuracy has been removed as it was not extensively analyzed

8. Adding moisture analogue. The performance after adding the moisture analogue improved further, although 2 parameters were sufficient for this analogue. The transferability however starts to be a problem. This is an expected result because I expect atmospheric moisture to vary more significantly in space (one of many reasons?). It is however not super clear to me what the authors recommend: should moisture be used as a predictor or not?

A paragraph was added with recommendations (last one of section 4.2, l. 525-534). Another part was added to discuss the spatial variability of moisture variables (as the main difference between the 2 less transferable subregions with other regions is the spatial windows on which the moisture is considered).

9. Optimization window. The preselection period of 4 months was also tested, without much effect on the results. I understood this was  $\pm 2$  months around the target date. Would the results improve if seasons are instead considered, for example the window is fixed for selected summer months to capture convective events, etc, regardless of the target date in that season?

Maybe the sentence was not clear. We use the 4-month preselection window anyway. It was the optimization of the length of this window that was unsuccessful. The 4-month preselection window was introduced some decades ago as an improvement of a fixed seasonal preselection. It is expected to be better than a fixed seasonal window.

10. Finally, I find the paper very long and cumbersome to read. It has 17 figures and 9 tables. I wonder if some of the figures/tables can be passed to Supplementary materials to help the reader focus on the main messages. This is only my subjective opinion.

The number of figures was decreased to 10 and the number of tables to 6. We removed the following elements:

- The analysis of the CRPS decomposition into sharpness and accuracy, as it was not extensively analyzed, along with Fig. 4 and 10.
- Tables 3 and 7 (containing the resulting parameters for the Chablais subregion) were removed as the same information can be found in figures for all subregions.
- Table 4 was removed (values of the CRPSS score) and figure 3 and 8 were changed in order to represent the values of the CRPSS score instead of the relative difference.
- Figure 9 (CRPSS of the 4Zo-4Mio method) was removed as the results are very similar to those of the 4Zo-2Mio method, and thus not so interesting.
- Figures 5 and 6 were merged into a single figure, as well as figures 13 and 14.
- Figure 12 (relationship between the different number of analogues) was removed as it is a bit redundant with figure 11.
- Figure 15 (Optimized weighting for the pressure levels of the 4Zo method) was removed and standard deviations were added to figure 16 in order to show the variability.

11. Some editorial issues

> abstract: parameter inter-dependencies, not parameters inter-dependencies (check also elsewhere in text)

Corrected, thanks.

> line 38: presents

Corrected, thanks.

> typo in the subscript in Equation (1)

We didn't find any mistake here... The score name is written S1.

> line 138: Figure 17 where?

Some clarifications were added

> line 173: what are left side valleys?

This was changed for "southern valleys"

## Response to Reviewer #2

### Summary

The authors present a methodology to optimize the parameters of an analogue-based precipitation downscaling system using genetic algorithms (GA). The GA is not only an optimization technique but allows discovering parameter inter-dependencies and possibly give a better understanding of the dynamics that lead to high precipitation accumulations in Canton Valais, Switzerland.

The paper is well written and a pleasure to read. I believe that the use of genetic algorithms within analogue-based forecasting techniques is an interesting idea. In fact, it increases the objectivity of current "rule-of-thumb" decisions that are done to drive the selection of analogue situations. Consequently, I recommend the publication of the paper after having addressed the remarks that I list hereafter.

Thank you for your very nice feedback.

### Major comments

Page 4, Line 66 - I would rather put the equation just after mentioning the Teweless-Wobus criterion S1.

We changed that, thanks.

Page 5, line 97 - Here I would also mention that the skill of analogue forecasts that include as predictor variable the moisture index depends on the skill of the NWP model in predicting moisture fields (when used in real-time).

Good point. We added a note on that.

Page 5, line 8-9 - Does the AM perform well also when looking for analogues for a single rain gauge? What is the consequence of computing a local average given the high spatial variability and intermittency of precipitation, e.g. for convective cases? There is no need to do analysis to answer this question

There is a slight increase in performance due to the smoothing of local variability when working on local averages. The difference is however rather low. A comment has been added.

Page 6, line 122 - Is the climatological distribution of precipitation over a single day sufficiently stable as reference to account for seasonality? Have you tried to include a temporal smoothing or pool the

data over days before and after the given day? A harder reference to beat could be the Eulerian persistence forecast (the precipitation observed on the previous day).

The climatological distribution of precipitation usually considered is the one from the entire archive. It could be discussed that a climatological distribution built on the  $\pm 2$  months' window (thus seasonal) might be more relevant. However, most applications use the distribution of the full archive. This does not play a major role in this paper, as we are interested in the improvement relatively to reference methods.

Page 9, line 200-202 - What is the overlapping constraint? The expression "what the sequential calibration cannot do" is not clear to me.

No overlapping constraint of the spatial windows means that they can differ from one pressure level to another (this has been specified in the article).

"what the sequential calibration cannot do" was changed for "which cannot be achieved with the sequential calibration technique".

Page 10, line 226 - It would be very interesting to show a plot with the CP and VP error as a function of number of predictors to illustrate that the VP error reaches an optimum around 4 predictors while the CP error keeps decreasing for increasing number of predictors (overfitting).

The assessment of the optimal number of predictors has been performed again on 3 subregions in order to consider the weighting between the predictors in that process. Consequently, this has slightly changed the results (not anymore a clear decrease on the VP), but it has not changed the conclusion that 4 seems to be an optimal number of predictors. A figure (Fig. 2) has been added to illustrate this aspect.

Page 13, line 316 - You could add that there are multiple local optima in very different regions of the parameter space that provide sufficiently good performance. Instead of using only one single optimal solution for the selection of analogues, you could use an ensemble of optimal solutions. This way you could both account for the parameter uncertainty of the analogue technique and increase the number of samples contributing to the empirical distribution of precipitation at the rain gauge (ensemble size). This could be considered for future studies.

This is a good idea. This point was added in the discussion section (l.597-606).

Page 14, lines 351- 356 - When optimizing an error function depending on precipitation totals, the large precipitation values (and errors) will contribute more to the total error. Thus, using GA allows to minimize the forecast error in particular for days with high precipitation accumulations. Therefore, it is quite reasonable that you beat the reference method, which has no optimization of an error function.

There are two different elements here:

- Indeed, larger precipitation values contribute more to the error function. The following sentence was thus added in section 3.3: l. 390-393
- However, the reference methods are calibrated by means of the sequential procedure, which also aims at reducing the error function. GAs can reduce the errors to a greater extent than the sequential procedure thanks to more efficient techniques and more degrees of freedom. We realized that the establishment of the references is not clear enough and thus we added a section on the sequential calibration (2.5).

Page 17, line 435-438 - Could the over-parametrization of the regions be due to the larger spatial variability of moisture fields? Pressure fields are known to be smoother and could be expected to generalize more to close regions than moisture fields.

It is likely to play a role. Indeed, the main difference with other regions is the spatial windows on which the moisture is considered. We added a comment on that in the end of section 4.2 (l. 518-524).

Page 17, line 450 - It would be interesting to mention that there is an interdependence between the location (or size) of the spatial window and the temporal window. In fact, if we follow Taylor's hypothesis, space and time could be easily related if we consider a moving precipitation system (or other) that has no significant growth and decay processes. More we go backwards in time more we have to move upstream the analogy window.

We now mention this point: l. 548-551

Page 20, line 536 - I wonder whether it would be useful to compute and show a correlation matrix between the different predictors.

We consider this idea interesting, but not feasible, because of the unlimited possible combinations. Indeed, the predictors are considered at different pressure levels, temporal and spatial windows.

Figures and Tables - The number of figures and tables in the paper is quite high, but I do not know which ones could be removed, perhaps those that are not discussed in detail in text or that are giving redundant conclusions.

We removed 7 figures and 3 tables. See the last comment to referee #1 about what was removed.

#### **Minor comments**

Abstract, line 2 - "provided by global models" is a bit too general. I would rather use general circulation models or numerical weather prediction models.

Corrected, thanks.

Abstract, par 2, line 2 - "strong limitations". You could complete the sentence by listing a couple of them.

Some information were added.

Page 2, line 9 - "Other predictands are also often considered". Here I would also add which ones, e.g. ...

Examples and references were added.

Page 2, Line 15 - "get down" → resolve, compute, forecast. I would use a more appropriate term.

Corrected, thanks.

Page 2, Line 21 - "made" → "designed"?

Corrected, thanks.

Page 3, Line 26 - "criterion itself" or "criteria themselves"

Corrected, thanks.

Page 3, lines 27-29 - Here you could also mention that ad-hoc techniques for the selection of predictors were also used by Panziera et al. (2011) and Foresti et al. (2015) for ensemble radar rainfall nowcasting. The GA technique could also be adapted for these applications.

A sentence referring to this work was added at the end of the paragraph.

Page 3, Line 30 - I would find a better term for “reconsidering”

Corrected, thanks.

Page 3, Line 31 - “pressure levels” → “optimal pressure levels”

Corrected, thanks.

Page 3, Line 44 - “on precipitation predicting” → “for precipitation prediction”

Corrected, thanks.

Page 4, Line 74 - “of the geopotential height”. I would add “, which represent better the upper level flow direction”

Added to the text, thanks.

Page 5, Line 81 - “both North and East directions”

Added to the text, thanks.

Page 5, line 102 - “Predictors are generally extracted from reanalysis datasets”

Added to the text, thanks.

Page 6, line 111 - It would be interesting to mention that you are trying to verify the performance of an ensemble-probabilistic forecast technique.

A sentence has been added.

Page 7, line 133 - “complex surface”. You could add that you are trying to find the global optimum of a complex high-dimensional error function having multiple local optima.

A sentence has been added.

Page 8, line 165 - Here you could add that the high spatial variability of precipitation is due to complex orography.

Added to the text, thanks.

Page 9, line 196 - “are not provided in this paper”

Added to the text, thanks.

Page 9, line 204 - “respectively, w.r.t. the reference method based on Z500 and Z1000”

Added to the text, thanks.

Page 9, 205 - “tremendous” → “very significant”, “large”

Corrected, thanks.

Page 9, line 209 - “other parameters (...)”

Information added.

Page 9, line 211 - “and may” → “but may”

We meant “and”, because both points are negative consequences.

Page 10, line 225 - “but always more to a smaller extent” could be rephrased

Corrected, thanks.

Page 10, line 229 - “another region than Valais” to clarify that it is not another region within your domain.

Clarification added.

Page 10, line 239 - “name” → “named”

Corrected, thanks.

Page 11, line 263 - “cross-compatibility and spatial coherence of the optimized parameters”

Information added.

Page 11, line 276 - “significant preference in the AM” is not clear.

The sentence has been removed.

Page 14, lines 360-365 - Does this mean that the two levels of analogy bring complementary information (not independent)? This is a good finding.

Thanks. A note has been added.

Page 15, line 382 - “spatial shift”?

No, a vertical shift (clarification added)

Page 18, line 481 - “does not”

Corrected, thanks.

Page 20, line 522 - “what the sequential calibration” is a strange expression to me.

Corrected, thanks.

Page 21, line 544 - “dependence in the selected parameters”

Corrected, thanks.

Page 21, line 555 - “significantly more improved” → “improved further” or other

Corrected, thanks.

Figure 2 and 7 - Would it be better to put the actual pressure levels (Z500, Z1000, etc) instead of the four levels (Z1, ..., Z4)?

Corrected, thanks.

Table 1-3 - It is not clear if the provided hour (12h, 24h) is for the day before the target day that we want to forecast.

It has been clarified. It is the hour within the target day. In a forecasting application, the AM is applied to the NWP model outputs, thus letting the temporal extrapolation to the NWP model. It is then more an “adaptation technique”.



Table 6 - In the caption I would make clear whether the improvement is w.r.t. climatology or the reference method.

Clarification added.