Streamflow forecast for Icelandic catchments using an analogue sorting prediction method

A progress report to Vegagerðin

Morgane Priet-Mahéo, Andréa-Giorgio Massad, Sif Pétursdóttir, Davíð Egilson, Matthew James Roberts

Veðurstofa Íslands – November 2018

Introduction

Floods in Iceland have led to significant material destruction and disruption in the past. With the rapid development of tourism in the country, we can expect floods to cause further threats, so informing people of imminent dangers remains a top priority.

To forecast potential floods, the Icelandic Met Office (IMO) has historically relied on the daily monitoring of water levels along with weather forecasts.

Analogue sorting methods are traditionally used for meteorological forecasts, but similar methods, often named “k-nearest-neighbor” methods, have been successfully tested for streamflow forecasts (Karlsson & Yakowitz, 1987; Akbari et al., 2011; Oyebode et al., 2014). In 2013, Crochet (2013) researched successfully the use of these methods at IMO.

The numerical weather prediction model HARMONIE has been in use at IMO since 2011, with nationwide forecasts available for the next 72 hours. Additionally, hindcasting of previous weather conditions has resulted in a 30-year record of gridded, 2.5 by 2.5 km forecast data for the whole of Iceland. Combining previous weather forecasts and hydrological data with the latest meteorological forecast and hydrological measurements enables new streamflow forecasting possibilities up to three days ahead.

In this progress report, we outline the methodoloy used to set-up the analogue forecast, as well as the first results obtained. We also explain the development of an operational flood forecasting system for catchments in Iceland.

Study area

Thirteen catchments have been selected all around the country (Figure 1). Their characteristics are diverse as seen on Table 1: from simple direct runoff catchments (such as VHM 128) to catchments fed mostly by springs (i.e. VHM 10), lakes (i.e. VHM 204) or glacial rivers (i.e. VHM 150). Their size vary from 38 km^2 (VHM 19) to 1102 km^2 (VHM 200) as well as their aspect ratio (between 1.31 and 3.71) and longest flowpaths (from 15 to 131 km).

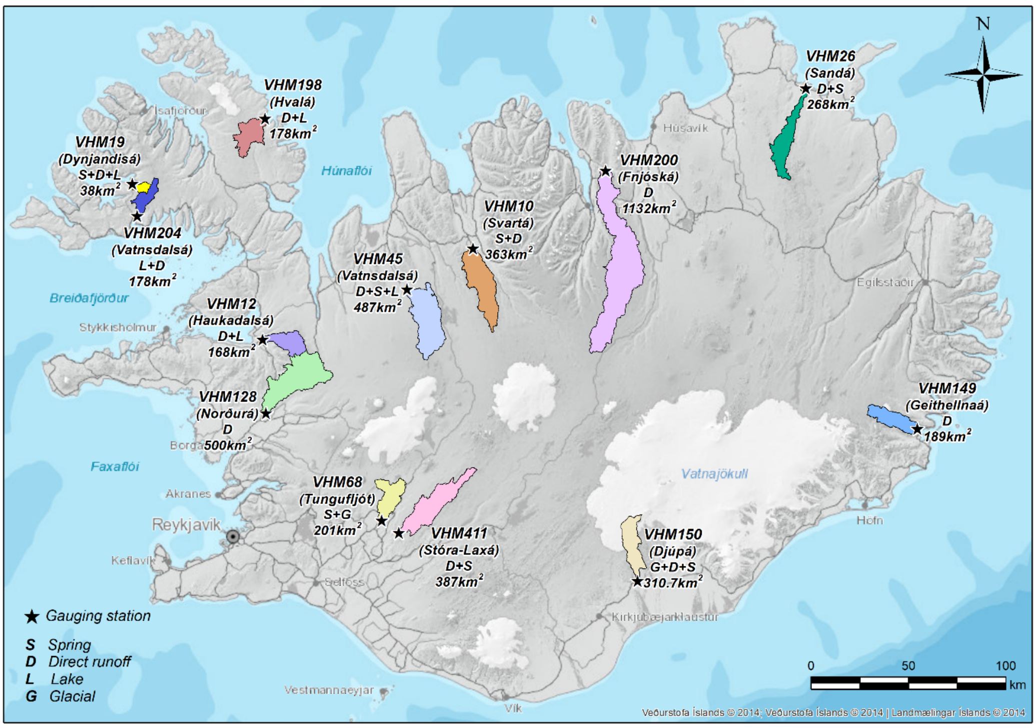

Figure 1 – Location of the catchments

Station number	River type	Area (km2)	Aspect ratio	Longest Flowpath (m)
VHM10	S+D	396	2.95	55166
VHM12	D+L	165	1.74	31960
VHM19	S+D+L	38	1.70	15040
VHM26	D+S	266	3.36	64555
VHM45	D+S+L	458	2.37	58072
VHM68	S+G	201	1.33	35345
VHM128	D	513	2.01	58288
VHM149	D	189	3.27	37033
VHM150	G+D+S	226	3.03	45563
VHM198	D+L	193	1.31	31543
VHM200	D	1102	3.51	131238
VHM204	L+D	102	2.47	28106
VHM411	D+S	387	3.71	73405

Table 1 – Catchments characteristics.

River types, S: spring-fed, D: direct-runoff, L: lake, G: glaciers

Methodology

Analogue sorting is a simple method in which the main assumption is that an event is the result of a combined of factors, and that comparing present conditions with similar past conditions will give us a good insight on the forecast of the event.

To forecast the discharge, past discharges measured at the selected stations and several meteorological variables (air temperature T, precipitation P and snow-water equivalent SWE, melting MT) computed by HARMONIE have been combined in numerous ways to create a range of 20 predictor-sets.

The Mahalanobis distance (Mahalanobis, 1936) is used to determine which past events ***x(u)*** are closest to the present event ***x(t)*** (Figure 2). The past events are then sorted according to their distance to ***x(t)*** and a number k of events is kept to derive the deterministic and ensemble forecasts (Figure 2).

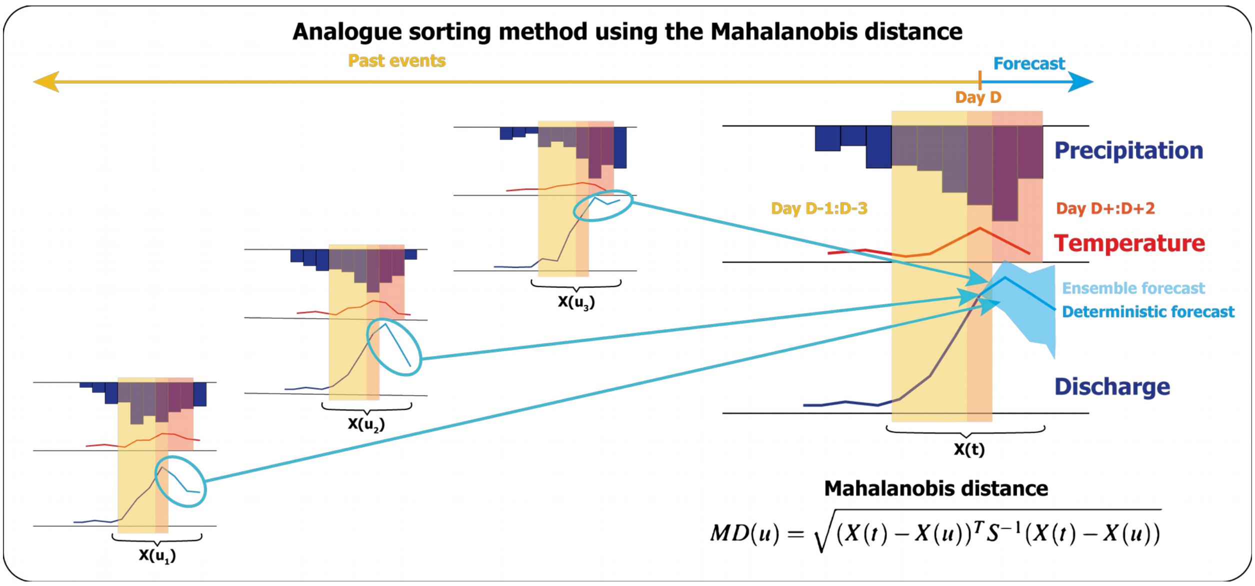

Figure 2: Schematic representation of the analogue sorting method based on the Mahalanobis distance

Statistical calibration

To test the results and determine which combinations of predictor-sets work the best, the five-day forecasts have been computed over a whole year for each predictor-sets and every stations.

Table 2 shows the best results for each station in term of predictor-sets by using the Nash-Sutcliffe Efficiency (NSE) coefficient. The analogue sorting method gives good results, especially for the first day of forecast with a NSE between 0.68 and 0.97. These values drop with each day of forecast and, for the fifth day, they range between 0.12 and 0.83.

A statistical analysis has been conducted to determine the best predictor-sets for each watershed, based on catchment characteristics such as area, aspect ratio, longest flowpath, river type, height difference and orientation (Table 1).

The results were most conclusive when the catchments were clustered according to river-type. If we focus on the D+1 results (Table 2), for snow-influenced direct-runoff catchments (VHM 204, 198, 200 and 19) the predictors-sets including the temperature, melt or SWE variables are the most efficient while for catchments with storage (groundwater, wetlands and lakes such as VHM 128, 411, 12, 54, 10 and 26), the predictors-sets including the previous days discharges improve the forecast. For catchments receiving glacial run-off (VHM 68, 150 and 149) better results are found when the weather forecasts are used in the predictors-sets. Regarding the D+2 – D+5 forecasts, the inclusion of weather forecasts in the predictor-sets gives better results.

	First day of forecast D+1		Fifth day of forecast D+5	
Station number	**Best predictors-set**	**Best NSE**	**Best predictors-set**	**Best NSE**
VHM10	2	0.75	17	0.34
VHM12	4	0.68	17	0.24
VHM19	11	0.76	20	0.35
VHM26	6, 7	0.88	17	0.59
VHM45	12	0.93	19	0.63
VHM68	15	0.76	6	0.34
VHM128	4	0.73	16, 17, 19, 20	0.17
VHM149	15	0.72	16	0.12
VHM150	15	0.73	9	0.43
VHM198	6, 7	0.88	18	0.64
VHM200	4, 12, 13	0.97	18	0.83
VHM204	5, 6, 10	0.87	18	0.41
VHM411	4	0.9	20	0.53

Table 2: Results of the analogue sorting for the first and last forecast days.

*Predictor-sets (**PS**): **PS1**: Q, P; **PS2:** Q, T; **PS3**: Q, SWE; **PS4**: Q, MT; **PS5**: Q, P, T; **PS6**: Q, P, T, SWE; **PS7**: Q, P, T, SWE, MT; **PS8**: Q, P, P3[1]; **PS9**: Q, P, P2[1]; **PS10** Q, P, T, MAN[2]; **PS11**: Q, P, T, ARS[3]; **PS12**: Q, Q1[4]; **PS13**: Q, Q1, Q2[4]; **PS14**: Q, P, Q1; **PS15**: Q, P, T, TS1[5], PS1[6]; **PS16**: Q, P, T, TS1, PS1, TS2[5], PS2[6]; **PS17**: Q, P, T, SWE, MAN, TS1, PS1, TS2, PS2; **PS18**: Q, P, T, MT, MAN, TS1, PS1, TS2, PS2; **PS19**: Q, T, P, SWE, TS1, PS1, TS2, PS2, Q1, Q2, MAN; **PS20**: all parameters.*

1: 2 and 3 past days of accumulated precipitations
2: months
3: seasonality (winter or summer)
4: 1 and 2 past days of discharge
5: 24 and 48 hours air temperature forecasts
6: 24 and 48 hours precipitation forecasts

Different number of events "k" were tested from 10 to 100 and it was estimated that a "k" between 40 and 50 would be the most efficient, a larger number of events would simply decrease the computing efficiency with no significant improvement of the results.

Rescaling has also been tested for all stations and all predictor-sets. As an event is not exactly the same as in the past, the forecasted discharge is being extrapolated by the same factor that make the present event differ from its closest past event:

$$\boxed{Q_{fd} = \frac{Q}{Q_p} * Q_{p+d}}$$

with Q *the current discharge,*

Q_p *the closest match in the past discharge event,*

Q_{fd} *the forecast discharge for day d and*

Q_{p+d} *the discharge d days after the closest past discharge.*

Rescaling improves the results for station with shorter timeseries or in cases of unusual events. On Figure 3, an example is given for the results obtained with predictor-set 17 at station VHM 150 and we see a clear improvement of the results with rescaling for the first two days of forecast (Sr1 and Sr2).

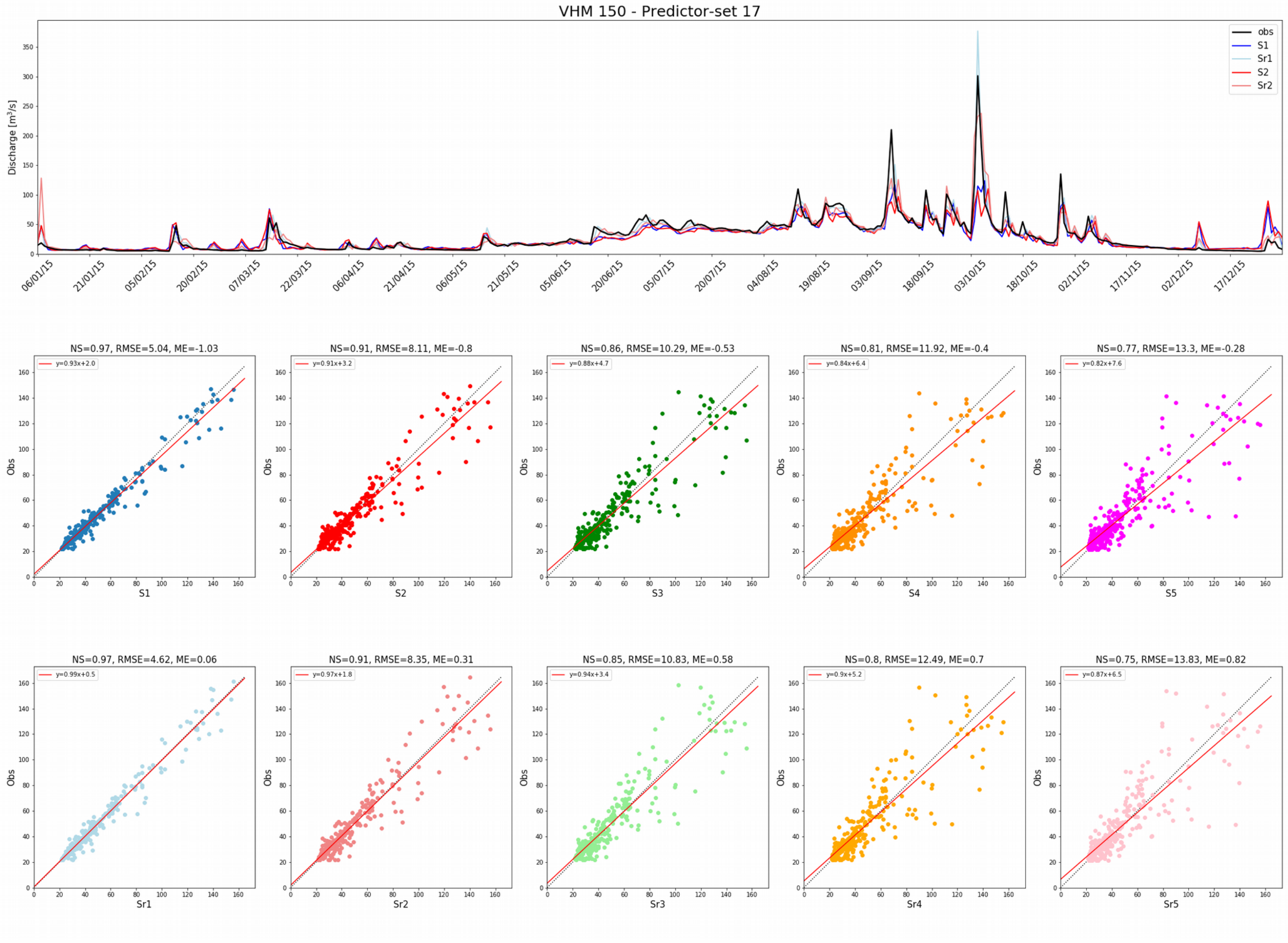

Figure 3 – Results for year 2015 obtained by analogue sorting using predictor-set 17 at station VHM 150.

Operational setup

Results of the analogue forecast will be displayed on a webpage similar to Figure 4. A map of Iceland shows all the catchments selected for the study (Figure 4, left). When the mouse hovers over a catchment, the results are presented to the right of the map for the highlighted watersheds. The lower sub-plot shows the measured discharge over the last 30 days (black lines) along with the analogue forecast for the next five days (each color corresponds to the results from a different predictors-set). The grey shading shows the minimum and maximum values of the past forecast, giving a visual indication on how reliable the results are. In addition, the yellow horizontal dashed line indicates the 2-year return-period and corresponding discharge threshold is also shown. If this first threshold is reached, the next one will be displayed (in this case the 5 years-return period) and so on. The upper graph shows the WaSiM simulated temperature (purple line) and precipitation (gray bars) over the last 30 days and for the next 3 days as predicted by the WaSiM model. The 0°C isotherm is represented by the dashed purple line. On both subplots, the vertical dashed line indicates what the values of yesterday's meteorological data on which the analogue forecast for the next 5 days (including the present day) are based.

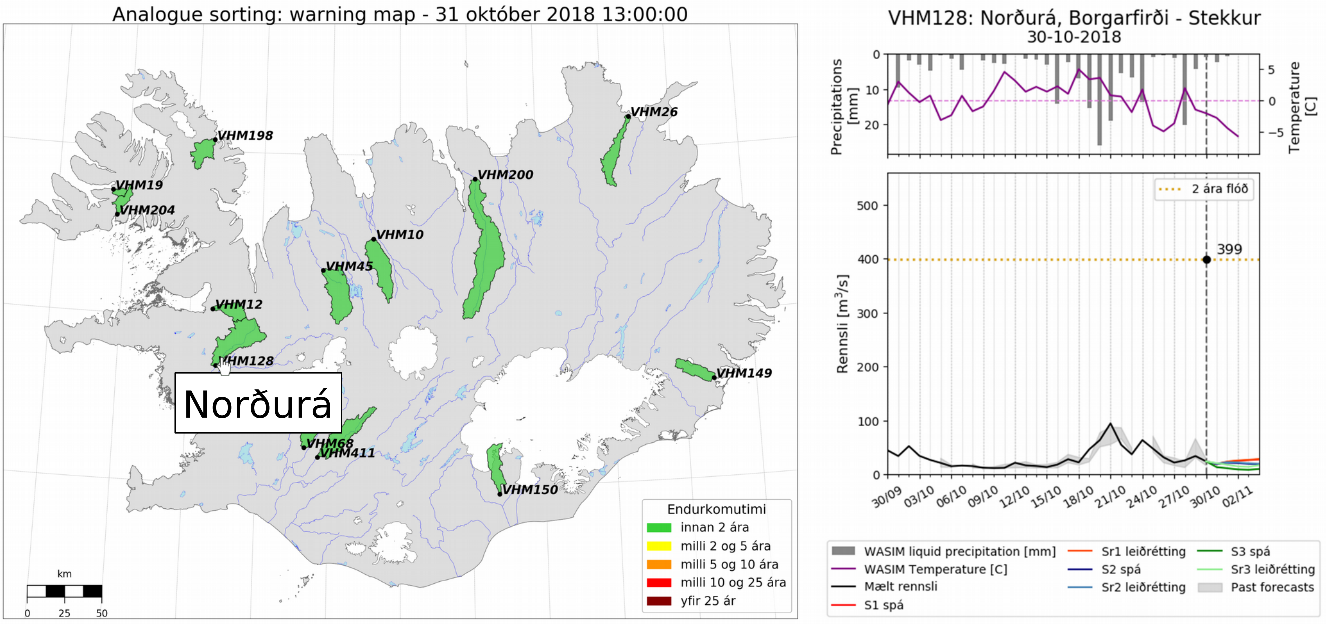

Figure 4: Screenshot of the operational forecast system, showing station VHM 128 as an example

In addition, the colors of the catchments (Figure 4, left) are updated hourly in relation to the latest measured discharge. A green catchment indicates a discharge value under the 2-year return period. As this first threshold is reached, the catchment will change color to become yellow and so on. This map serves as a warning map that could assist Vegagerðin and Veðurstofan with travel advisories to the public.

Conclusions and next steps

Streamflow forecasting based on analogue sorting methods requires little to no pre-processing and it is not especially computer-intensive. It also enables longer forecast periods, going from 2 to 5 day forecasts.

This method gives promising results for a variety of catchments. Presently, the model runs with the same predictor-sets for all catchments. The next step is to put the results from the statistical analysis in use so that each catchment is simulated with the most appropriate predictor-set. Further improvements of the forecasts will be investigated using post-processing methods and the webpage will be refined. The results of the project will be delivered to Vegagerðin in March 2019 as a report and a test-version of the forecasting web-site.

References

Akbari, M., Van Overloop, P.J. & Afshar, A. (2011). Clustered K-nearest neighbor algorithm for daily inflow forecasting. Water resources management, 25(5), 1341-1357.

Crochet, P. (2013). Probablistic daily streamflow forecast using an analogue sorting method. Icelandic Meteorological Office, Reykjavík.

Karlsson, M. & Yakowiz, S. (1987). Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. Water Resources Research, 23(7), 1300-1308.

Mahalanobis, P.C. (1936). On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India, 2, °1, 49–55

Oyebode, O.K., Otieno, F.A.O., Adeyemo, J. (2014). Review of three data-driven modeling techniques for hydrological modeling and forecasting.