Veflausn með daglegum rennslisspám sem byggist á hliðstæðri greiningu veðurgagna

Morgane Priet-Mahéo, Andréa-Giorgio R. Massad, Tinna Þórarinsdóttir og Matthew J. Roberts

Greinargerð

Veðurstofa Íslands
Bústaðavegur 7–9
108 Reykjavík

+354 522 60 00
vedur@vedur.is

MPM/ofl/2020-01

Lykilsíða

Greinargerð nr.	Dags.	Dreifing: Opin ☒ Lokuð ☐
MPM/ofl./2020-01	Mars 2020	**Skilmálar:**

Heiti greinargerðar: Veflausn með daglegum rennslisspám sem byggist á hliðstæðri greiningu veðurgagna / Web solution for daily discharge forecasts built on analogue-based analysis of weather data	**Upplag:** Rafræn útgáfa **Fjöldi síðna:** 28
	Framkvæmdastjóri sviðs: Jórunn Harðardóttir
Höfundar: Morgane Priet-Mahéo, Andréa-Giorgio R. Massad, Tinna Þórarinsdóttir og Matthew J. Roberts	**Verkefnisstjóri:** TÞ/MJR
	Verknúmer: 4611-0-0002
Gerð greinargerðar/verkstig:	**Málsnúmer:** 2019-0118

Unnið fyrir: Vegagerðina
Samvinnuaðilar:

Útdráttur:

Veðurstofa Íslands hefur á undanförnum árum unnið að verkefnum er snúa að flóðagreiningum vatnsfalla á Íslandi í nánu samstarfi við Vegagerðina. Upplýsingar um tíðni og stærð flóða eru nauðsynlegar hönnunarforsendur fyrir vegaframkvæmdir og úrbætur sem og fyrir mat á áhættuviðmiðum og við svæðisskipulag. Þetta verkefni er unnið í framhaldi af verkefninu Daglegar rennslisspár með notkun hliðstæðrar greiningar Harmonie veðurgagna frá 2019. Í fyrra verkefni var sýnt fram á að unnt er að setja fram áreiðanlega spá um rennsli og flóð vatnsfalla með því að nota aðferðafræði sem byggir á hliðstæðri greiningu (e. analogue sorting) gagna. Aðferðin var prófuð á 13 vatnasviðum með góðum árangri þar sem spáð var fyrir um rennsli 1–5 daga fram í tímann. Í þessu verkefni er spáin útvíkkuð fyrir 20 vatnasvið og nær nú yfir 33 vatnasvið alls, þar á meðal eitt sem er án rennslismælinga. Enn fremur var aðferðafræðin endurbætt til að unnt sé að spá betur fyrir um tímasetningu rennslistoppa. Niðurstöður eru settar fram á vefsíðu sem er uppfærð daglega með nýrri rennslisspá fyrir hvert vatnasvið.

Lykilorð: Hydrological forecast, flood forecast, HARMONIE, flood warning, analogue sorting	**Undirskrift framkvæmdastjóra sviðs:**
	Undirskrift verkefnisstjóra:
	Yfirfarið af: SG

Contents

Höfundar skýrslunnar bera ábyrgð á innihaldi hennar. Niðurstöður hennar ber ekki að túlka sem yfirlýsta stefnu Vegagerðarinnar eða álit þeirra stofnana eða fyrirtækja sem höfundar starfa hjá.

1 Introduction

Weather-related flooding in Iceland causes significant material destruction and societal disruption. Such floods pose a threat to inhabitants and tourists, resulting occasionally in traffic disruption due to the closure of roads and bridges. With the rapid development of tourism in the country, floods at various times of year can be expected to cause further problems. Flood forecasting and the communication of flood risk is therefore a top priority.

In recent years, the Icelandic Meteorological Office (IMO) has focussed increasingly on flood-related research for practical purposes. Tools have been developed to support the work of forecasters in predicting floods, including the setup of flood forecasts for three catchments based on the physical hydrological model WaSiM (Schulla, 2017), supplemented by forecasts from the numerical weather prediction model HARMONIE (Bengtsson *et al.*, 2017; Nawri, 2014; Nawri *et al.*, 2017). This system produces a daily two- to three-day forecast for these three catchments and it has proven to be useful. However, the system is computationally heavy, requiring a large amount of data preparation and intensive processing. In response, a nowcasting system was developed, combining the latest measurements at selected hydrological stations throughout the country with runoff forecasts from HARMONIE. The system also offers historical, physical and statistical information about the catchments. The system is setup on a webpage with a simple color-coded map, and it provides a first-order assessment of catchment status.

In 2018, a new forecast system was developed, based on findings from previous research at IMO (Crochet, 2013). This system is based on the Mahalanobis distance to discriminate past events, which has been used previously for climate forecasts (Yates, 2003) and hydrological analysis (Akbari *et al.*, 2011; Karlsson & Yakowitz, 1987). It produces a daily five-day hydrological prediction, computed from the analogue sorting of historical discharge series and weather data. This system was setup for 13 catchments around the country (Priet-Mahéo *et al.*, 2019) via a website. The project was funded by The Icelandic Road and Coastal Administration (IRCA) and IMO.

The research presented here is a continuation of the 2018 project with the goals of improving the operational forecast and adding 20 additional catchment forecasts to the website, including one that is an ungauged catchment. The project continues along the same funding path, with support from IRCA and IMO. The original plan was to test the methodology on an ungauged catchment and make on-site discharge measurements to validate the results. Because the grant was smaller than applied for originally, it was decided to exclude the discharge measurements and test the method on a station that is presently ungauged but was monitored in the past.

This report is written as a sequel to the report published in 2019, *Daglegar rennslisspár með notkun hliðstæðrar greiningar HARMONIE veðurgagna* (Priet-Mahéo *et al.*, 2019); for further background details, see the previous report. In Section 2 of this report, a classification schema is presented for the entire group of catchments, allowing defining catchment characteristics to be delineated. The strengths and weaknesses of the existing forecast system are assessed in Section 3, both in terms of timeliness and accuracy since becoming operational at IMO. Based on these findings, a methodology for the correction of forecast results has been developed and it is described in Section 4, along with the forecast methodology for the ungauged catchment. Section 5 presents an enhanced forecast for the 13 catchments from the previous project and a hydrological forecast for 19 new gauged catchments, plus one ungauged site; it also summarises the updated webpage. The findings of the project are presented in Section 6.

2 Clustering of catchments

In addition to the 13 catchments that are currently within the hydrological forecasting system, 20 new catchments have been selected and added, in order to cover more of Iceland's main roads. Figure 1 presents a map of the country where the initial catchments are shown in light green and the newly added catchments in red.

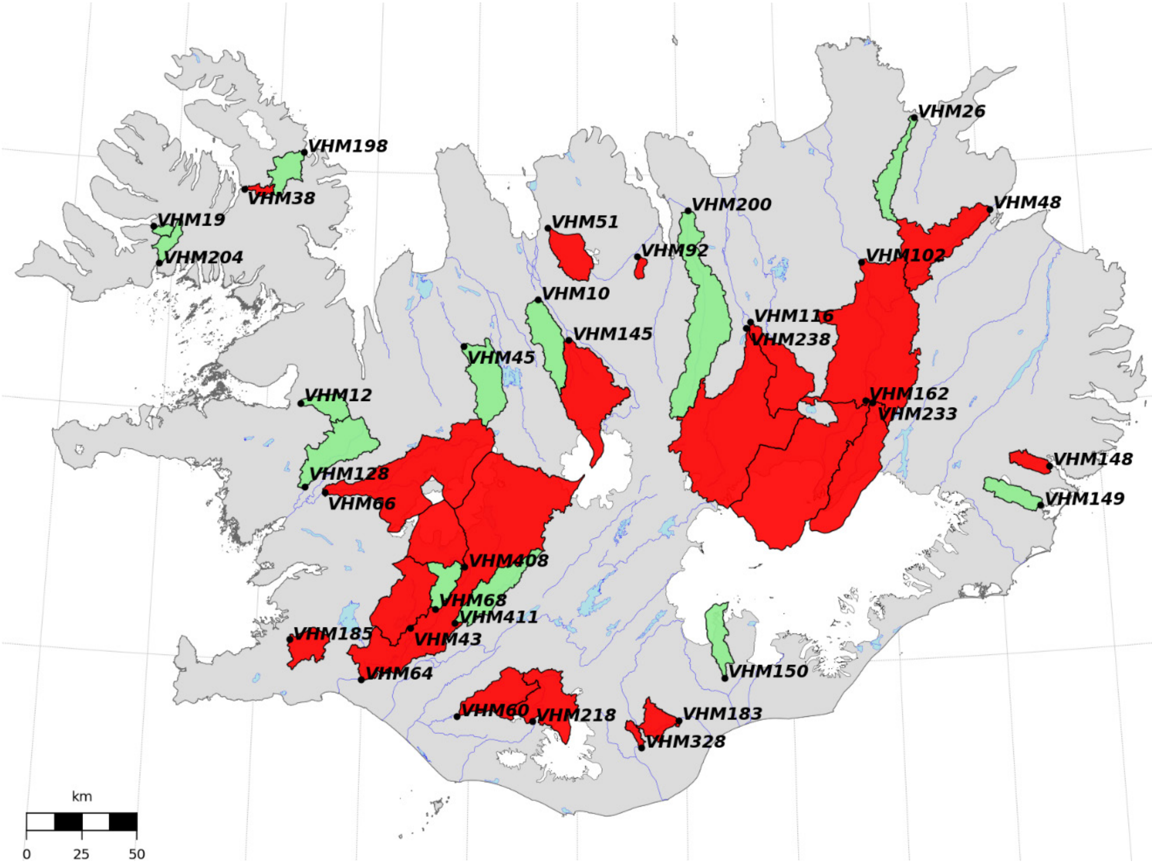

Figure 1. Map of Iceland showing the location of river catchments from the 2018 study in green and the 20 additional catchments as part of the 2019 project. Gauging stations are marked by a point and the labelling is according to IMO's station list. For river names, see Table 1.

Table 1 summarises the size and nature of the river catchments used in the 2018 project, as well as the present study. Catchment characteristics include area, aspect ratio, longest flow-path, mean elevation and other geological parameters.

Table 1. Physical characteristics of the selected catchments, both from the previous and present studies. For station locations, see Figure 1.

vhm	River	Area *km²*	Aspect ratio	Longest flow-path *m*	Average height *m a.s.l.*	Glacial Cover %	Old Bedrock %	Young Bedrock %	Total Bedrock %
10	Svartá	396.1	2.96	55,167	527	0	99.2	0.5	99.7
12	Haukadalsá	164.7	1.74	31,960	408	0	96.8	0	96.8
19	Dynjandisá	38.4	1.7	15,040	510	0	100	0	100
26	Sandá, Þistilfirði	266.3	3.36	64,555	387	0	61.3	38.7	100
38	Þverá á Langadalsströnd	42.8	2.39	20,971	428	0	100	0	100
43	Brúará	640.7	1.73	50,958	307	0	3.1	96.9	99.9
45	Vatnsdalsá	458.3	2.37	58,072	547	0	67.1	32.9	100
48	Selá	701.4	1.74	74,306	543	0	48.9	51.1	100
51	Hjaltadalsá	299.6	1.87	34,990	723	2.96	97	0	97
60	Eystri-Rangá	419.9	1.92	60,336	572	2.02	0	97.4	97.4
64	Ölfusá	5661.9	2.41	169,493	304	11.84	22	62.9	84.9
66	Hvítá í Borgarfirði	1574.4	2.24	123,017	650	20.3	22.7	53.7	76.4
68	Tungufljót	201.1	1.33	35,345	245	0	6.2	93	99.2
92	Bægisá	37.4	1.93	13,904	900	0	77.8	0	77.8
102	Jökulsá á Fjöllum	5097.1	2.6	189,195	538	28.64	0	71.3	71.3
116	Svartá í Bárðardal	527.1	1.87	62,858	645	0	1.2	98.8	100
128	Norðurá	513	2.01	58,289	338	0	93.7	1.7	95.4
145	Vestari-Jökulsá	843.8	1.86	70,333	752	10.9	44.1	40.6	84.6
148	Fossá	115.1	2.47	28,963	577	0	99.8	0	99.8
149	Geithellnaá	189.4	3.27	37,033	609	4.83	91	0	91
150	Djúpá	225.9	3.03	45,563	767	40.23	47	12.8	59.8
162	Jökulsá á Fjöllum	2023.1	2.01	110,507	1195	56.92	0	43.1	43.1
183	Skaftá	1627.2	2.36	133,710	249	26.31	10.9	62.1	73
185	Hólmsá, Reykjavík	216.8	1.42	31,189	710	0	0	100	100
198	Hvalá, Ófeigsfirði	192.9	1.31	31,543	399	0	100	0	100
200	Fnjóská	1102.2	3.51	131,238	723	0	97.1	0.4	97.6
204	Vatnsdalsá	102.3	2.47	28,106	466	0	100	0	100
218	Markarfljót	516.9	1.14	53,731	737	12.22	0	71.7	71.7
233	Kreppa	818.1	3.42	81,106	1130	37.65	0	62.3	62.3
238	Skjálfandafljót	2163	1.54	118,032	822	4.51	26.5	68.7	95.2
328	Eldvatn	1496	2.99	146,594	155	28.62	5.8	64.7	70.6
408	Sandá	581.3	1.13	58,363	756	49.27	0	50.7	50.7
411	Stóra-Laxá	387.1	3.71	73,405	559	0	97.7	2.3	100

2.1 Classification of the catchments

In order to collect rivers that share the most similarities and find the predictor sets that give the best results, a hierarchical cluster analysis has been performed using discharge data (seasonality, duration curves, and mass curves) as well as the watershed characteristics previously listed in Table 1. This clustering helps us identify the physical similarities of catchments, which can in turn simplify the identification of predictors important to the analogue sorting. It is expected that catchments within the same cluster behave in a similar way so that the same sets of predictors can be applied within each cluster. More details regarding the cluster analysis can be found in the previous report (Priet-Mahéo *et al.*, 2019). Figure 2 shows the results from the cluster analysis of all the catchments in the form of a dendrogram. A cophenetic correlation coefficient is defined as a measure of how well the analysis preserves the distances between the data. In this case, results show a value of 0.81 which indicates a good clustering (the closer to 1, the better the clustering).

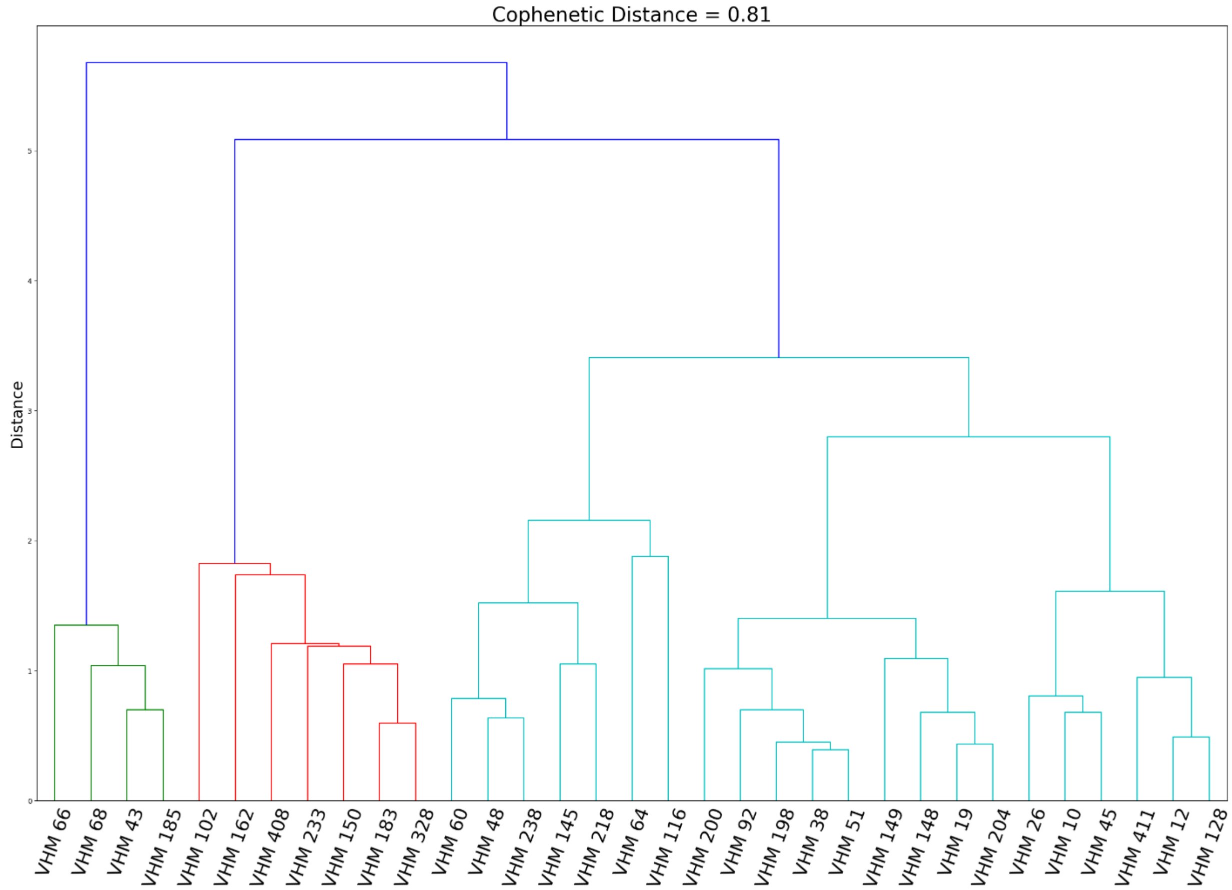

Figure 2. Results from the cluster analysis on a dendrogram. All the catchments group in five clusters with the cophenetic correlation distance equal to 0.81.

The five following clusters appear on the dendrogram in Figure 2:

- *Cluster A:* vhm 66, vhm 68, vhm 43, vhm 185
- *Cluster B:* vhm 102, vhm 162, vhm 408, vhm 233, vhm 150, vhm 183, vhm 328
- *Cluster C:* vhm 60, vhm 48, vhm 238, vhm 145, vhm 218, vhm 64, vhm 116
- *Cluster D:* vhm 200, vhm 92, vhm 198, vhm 38, vhm 51, vhm 149, vhm 148, vhm 19, vhm 204
- *Cluster E:* vhm 26, vhm 10, vhm 45, vhm 411, vhm 12, vhm 128

These results show that the gauging stations have been grouped according to river types and weather conditions. For further analysis, the clusters have been plotted on a soil map of Iceland (Figure 3).

Cluster A unites spring-fed rivers, some of them originating from glaciers. Those stations are all located in the southwestern part of the country and the rivers have a baseflow counting for a large part of the total discharge. Note that some rivers are regulated due to hydropower generation, although all gauges are located where discharge is unregulated.

In C*luster B*, most rivers are typical glacial rivers (some are affected by jökulhlaup) and all watersheds are partially covered by glaciers.

Cluster C includes rivers that are mostly a mix of spring water and glacial origin and they mostly have a direct runoff component. Many of the rivers are located at the border of volcanic soil, as seen in Figure 3.

In C*luster D* comprises direct runoff rivers influenced by snowmelt. There catchments have a long aspect ratio and they are in mountainous areas where the rivers reach high elevations and melting occurs later in the year.

Cluster E reflects catchments with mostly direct runoff rivers underlain by Histosol soils that encourage the formation of groundwater, wetlands, and lakes (Figure 3). Excluding stations vhm 26 and vhm 411, all sites in this cluster are in the northwest part of the country.

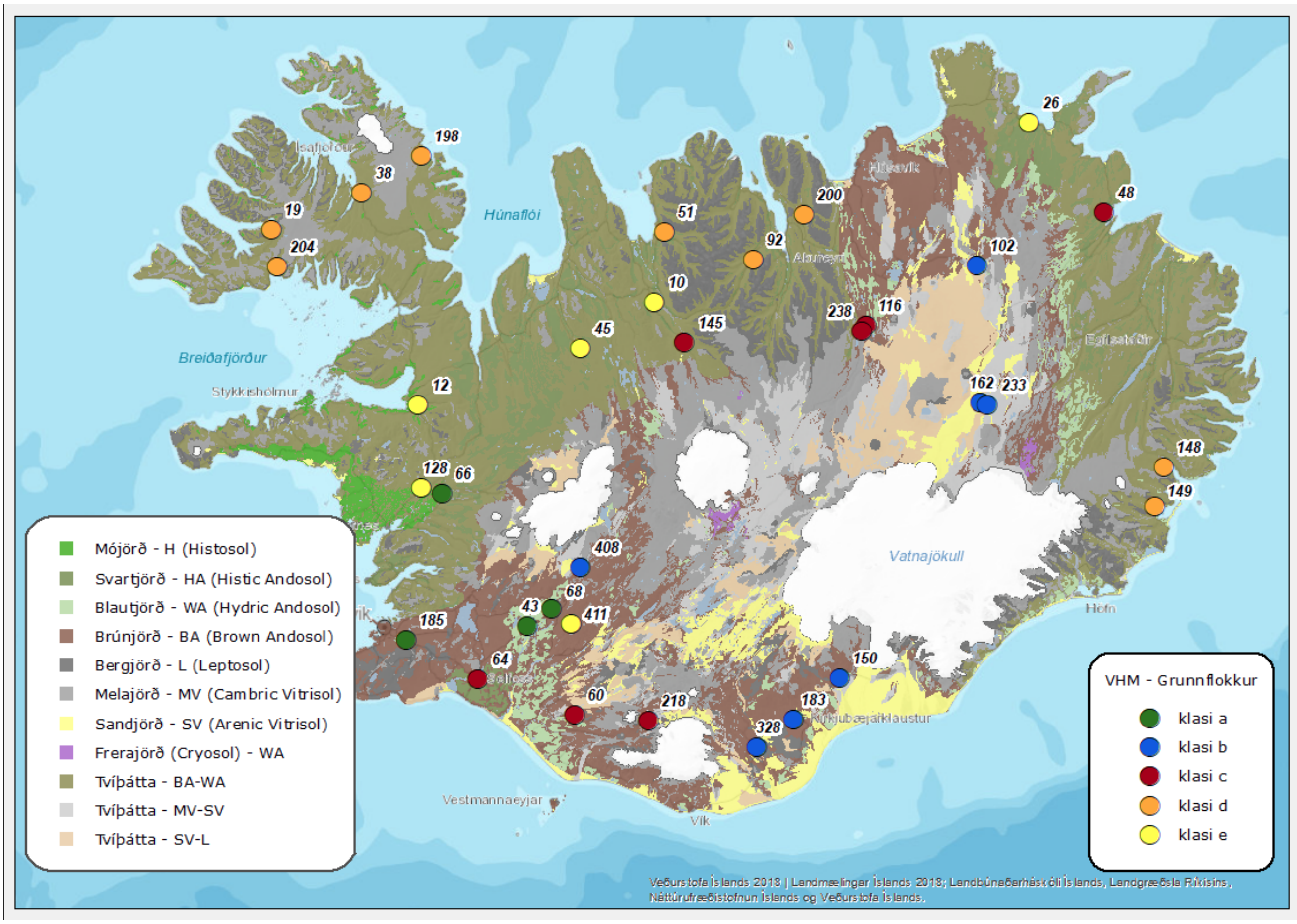

Figure 3. Soil map of Iceland with gauging stations shown according to their hydrological grouping from the cluster analysis. For station names, see Table 1.

3 Assessment of the operational forecasts

An operational forecast based on analogue sorting has been running at IMO since mid-2018. The main setbacks in forecast has been time-lags in estimated peak discharge. Other problems have been observed, such as the disruption of water-level measurements due to river ice. In this project, a focus was set on analysing and fixing the time-lag in peak estimation.

3.1 Lag in peak estimation

When analysing the results of the hydrological forecast, there is sometimes a time-lag between forecast and measured discharge. Lags between measurements and forecasts can be computed in a variety of ways, including: (i) cross-correlation; (ii) statistical coefficients such as the Nash-Sutcliffe Efficiency coefficient (NSE); (iii) the modified Nash-Sutcliffe Efficiency coefficient (mNSE); (iv) the Root Mean Square Error (RMSE); and (v) the Mean Error (ME). The mNSE uses mean absolute error instead of mean square error in order to minimise the effect of extreme values, while the RMSE and the ME are computed to address the performance of the model regarding the accuracy of discharge prediction. Table 2 summarises the main information concerning the 13 stations that have been run operationally since mid-2018. A lag of zero days means that the event was predicted accurately, a lag of minus one day means a day late and a positive lag of one day means a day too early. While the NSE is reasonable (above 0.7 in 62% of the cases for D1 and 23% for D2), a one-day lag in the 24-hour prediction was observed for most catchments. Stations vhm 150, 128 and 411 obtained the best forecasts in terms of timeliness. Stations vhm 68, 198 and 19, on the other hand, had the worst forecasts both in terms of timeliness and accuracy. However, it is important to consider that the lag value shown in Table 2, is the most significant one, it is however not representative of all days. While a main lag of minus one suggests a delay by one day of the 24-hour forecast – and therefore a miss – it does not necessarily mean that all events were missed. Figure 4 illustrates this phenomenon for the winter season 2018–2019, where the predominant lag is minus one for the entire operational season, suggesting that the predictions by the best set were mostly missed by a day; however, many of the winter peaks are predicted reasonably and they fell in the prediction range.

Table 2. Performances of the 24-hour (D1) and 48-hour (D2) streamflow forecasts for the 13 stations running operationally. The results are presented for the best sets (highest NSE for lag closest to zero). For details about the statistical parameters, see Section 3.1.

		D1: 24 hours						**D2: 48 hours**					
Class	**VHM**	**set**	**lag**	**NSE**	**mNSE**	**RMSE**	**ME**	**set**	**lag**	**NSE**	**mNSE**	**RMSE**	**ME**
A	vhm068	set15SR1	**-1**	0.5	0.58	3.8	0.007	set16S2	**0**	-0.62	0.3	5	0.71
B	vhm150	set16SR1	**0**	0.86	0.76	8.3	-1	set20SR2	**0**	0.71	0.67	12	-0.91
D	vhm019	set7SR1	**-1**	0.45	0.49	1.9	0.0024	set19SR2	**-2**	-0.17	0.17	2.7	-0.32
D	vhm149	set15SR1	**0**	0.81	0.71	9.4	0.28	set15SR2	**-1**	0.42	0.48	15	0.11
D	vhm198	set15SR1	**-1**	0.36	0.6	23	2.1	set15SR2	**-2**	-0.18	0.36	27	2.4
D	vhm200	set19SR1	**-1**	0.94	0.78	4.4	-0.2	set19SR2	**-2**	0.85	0.66	6.8	-0.14
D	vhm204	set19SR1	**-1**	0.7	0.72	3	0.066	set19SR2	**-2**	0.25	0.45	4.2	-0.02
E	vhm010	set19SR1	**-1**	0.51	0.58	2.6	0.046	set19SR2	**-2**	0.0035	0.36	3.6	0.076
E	vhm012	set20SR1	**-1**	0.88	0.79	2.3	-0.031	set15SR2	**0**	0.46	0.58	5.1	-0.26
E	vhm026	set20SR1	**-1**	0.88	0.68	2.7	-0.15	set20SR2	**-1**	0.76	0.56	3.7	-0.18
E	vhm045	set19SR1	**-1**	0.88	0.77	4.6	0.38	set3SR2	**-2**	0.57	0.6	8	0.43
E	vhm128	set15SR1	**0**	0.74	0.69	12	1.3	set15SR2	**0**	0.39	0.51	17	1.5
E	vhm411	set15S1	**0**	0.59	0.45	6.5	0.54	set15S2	**0**	0.24	0.37	9	0.51

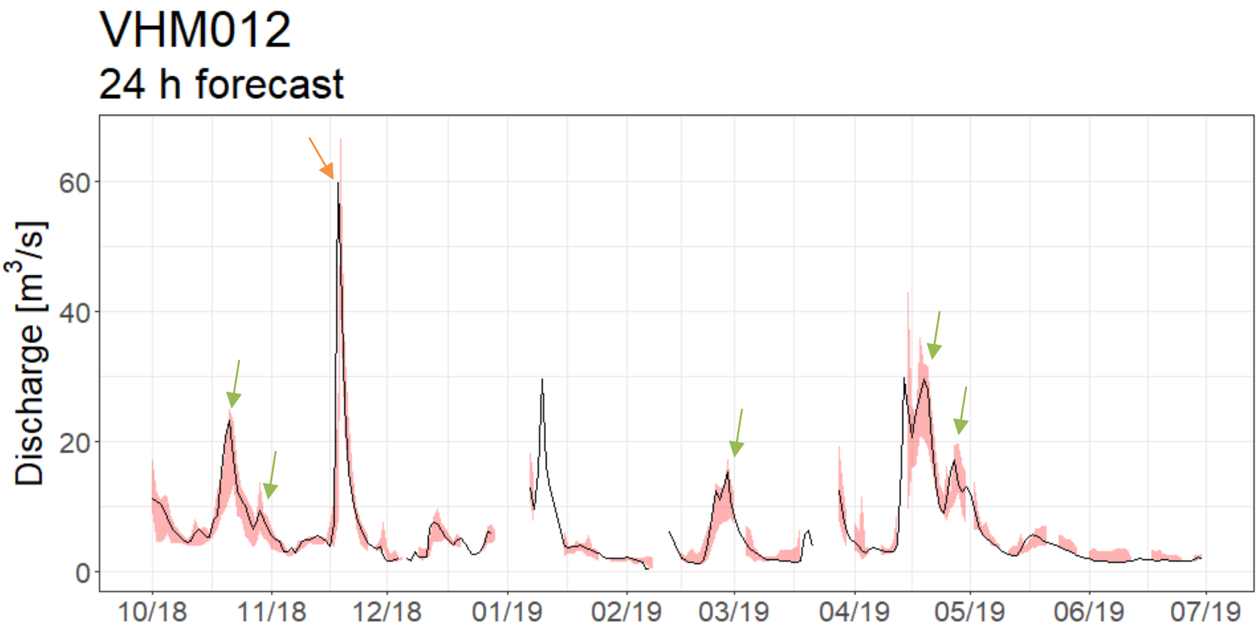

Figure 4. Twenty-four hour forecast for station vhm 12 during the winter season 2018–2019. The black line represents the daily averaged measurements, whereas the red band the prediction range (min-max) for the station. The green arrow shows peaks that were predicted at the correct time; the orange arrow a peak that was missed.

3.2 Technical problems and ice perturbations

Technical problems in relation to data collection and processing have affected the operational forecast. Disruptions have occurred when input data were not delivered, such as discharge measurements or meteorological forecast data from HARMONIE, resulting in large temporal gaps, ranging from days to weeks (Figure 5).

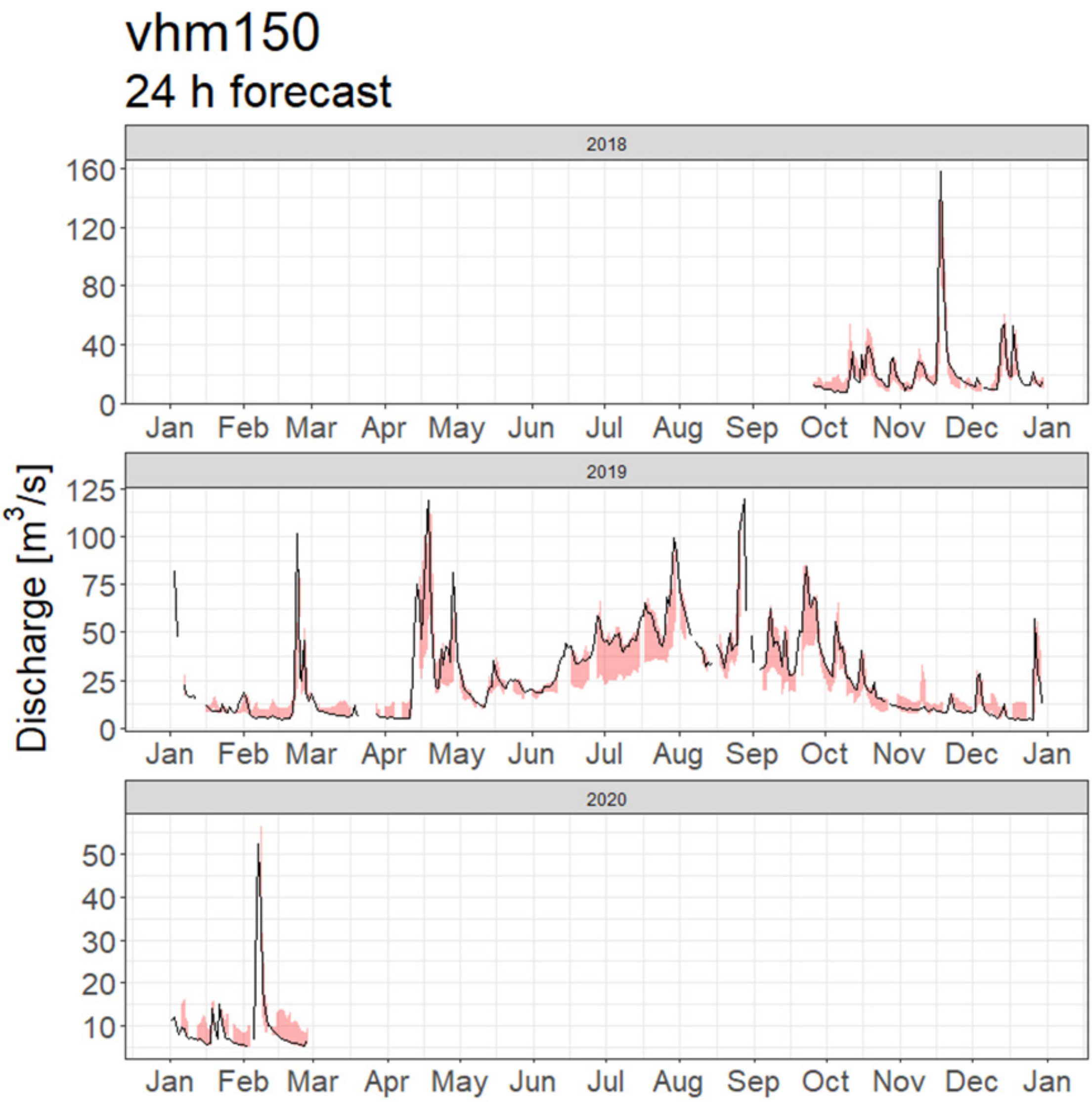

Figure 5. Twenty-four hour forecast for station vhm 150 during the operational period from September 2018 to March 2020. The black line represents the daily averaged measurements, the red band the minimum and maximum prediction range for the station.

A paucity of input data was not the only impact on the forecasting system. As the discharge results are derived from near-real-time measurements of water stage at a fixed cross-section, they can be influenced by ice perturbations during the wintertime. This has the effect of causing spikes in peak discharge, which need to be identified and corrected manually. The issues described in section 3.2 have not been attended to in this project, so they should be kept in mind for future improvements of the forecasting system.

4 Methodology

Test runs for analogue sorting were performed over a two-year period (01 September 2015 to 31 August 2017) for all stations, except the ungauged site. Based on this period, delays in the timeliness of the forecast was the main problem with the model, as described in Section 3. This suggests that the model does not always manage to discriminate and select the analogue-based events adequately. Consequently, a correction of the existing methodology is necessary. The following sub-sections outline the methodology for correcting the lag and the techniques used for discharge forecasts from the ungauged catchment.

4.1 Lag in peak estimation

The past events are sorted using the Mahalanobis distance, as described in the previous report (Priet-Mahéo *et al.*, 2019). This is the distance between two points in multivariate space. However, a major weakness of this distance is the lack of hierarchization between the variables in the dataset, as all variables have the same weight in the computation. A solution to this problem is to pre-sort the dataset and reduce the number of events that would qualify in the final sorting, thereby increasing the weight of a variable used to select past events. Therefore, the weight of some variables in the sorting process is increased. A series of tests were performed by pre-sorting the dataset once or twice consecutively based on one or several variables in the dataset.

4.2 Ungauged station

Station vhm 218 on Markarfljót, southern Iceland, was monitored from June 1982 to June 2001, and it was used as the test-site for an ungauged catchment in this project. Lacking modern-day discharge measurements, it was decided that tests with additional variables was needed. Variables such as minimum and maximum air temperature were compared between the ungauged catchment and other similar catchments within the same class. The goal was to find the station that had the most similar set of predictors. Stations vhm 64 and vhm 238 were selected as reference stations, and comparisons between vhm 238 and vhm 218 can be seen in Figure 6. It was not possible extrapolate discharge values from the existing measurements at vhm 218, so it was decided to attempt rescaling using other variables of the predictor set.

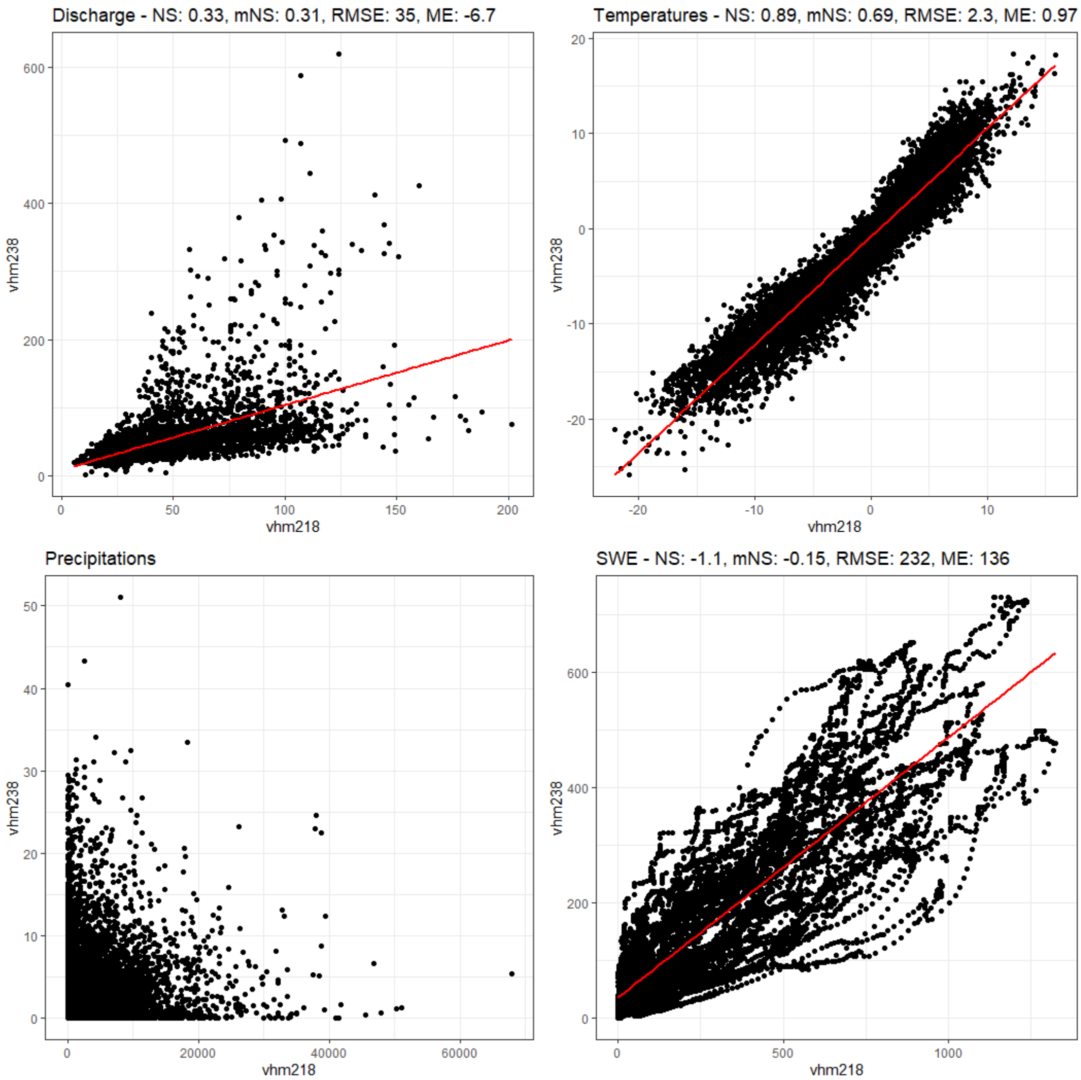

Figure 6. Comparison of predictors for stations vhm 218 and vhm 238.

5 Results

In this section, results for the correction of the lag of forecasted discharge peaks are discussed, as well as the discharge forecast for the ungauged catchment. A short summary of the expanded forecasting system of 33 catchments is also presented, together with a simple description of the operational website for daily discharge forecasts.

5.1 Rectification of the time-lag

The hierarchization (pre-sorting) of variables improved significantly the results in terms of timeliness (reduction of lag for peak events) and accuracy. Figure 7 shows the analogue sorting results for station vhm 48 with hierarchization and without hierarchization. While the improvement in the 24-hour forecast is minimal as the performance of the station was adequate already, hierarchization still improves the timeliness (*e.g.* peak discharge in 2016). For the 48-hour forecast, the hierarchization increases significantly the accuracy of the results, going from an NSE value of 0.65 to 0.92.

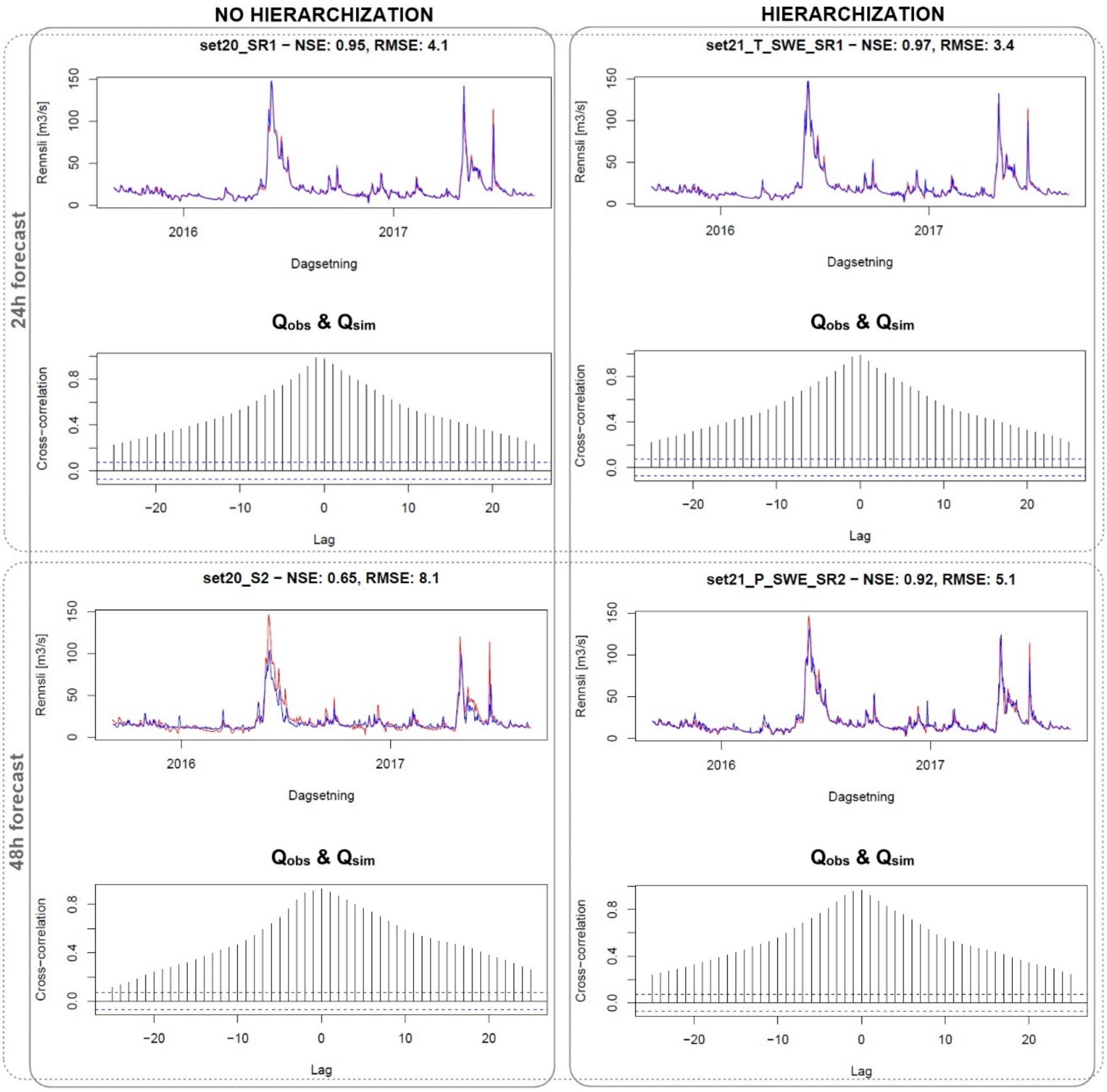

Figure 7. Comparison and cross-correlation of discharge measurements (red) and predictions (blue) (24-hour, up and 48-hour, down) for the analogue sorting with (left) and without (right) hierarchization of the predictors for station vhm 48. The lag is expressed in days; a negative lag suggests a delay in the predictions.

5.2 Forecast performance at the ungauged station

The analysis for the ungauged catchment shows that the added variables in the dataset do not necessarily improve the predictions; however, the tests underline the importance of the snow-water-equivalent (SWE) variable. The predictor sets comprised air temperature (T), precipitation (P) and SWE. Figure 8 presents the best results for the test period, based on SWE and temperature for the first three days of the forecast; note that the graphs cover three intervals: 24, 48 and 72 hours. Despite a relatively low coefficient (NSE ≈ 0.57), the first three days of predictions are mostly timely (lag = 0). The results can be assumed acceptable for an initial estimate of discharge over longer periods, but further research is needed before the approach can be used for short-term flood forecasting. Simple rescaling of the results was attempted using P, SWE and the discharge of the neighbouring catchment, but the findings were unsatisfactory.

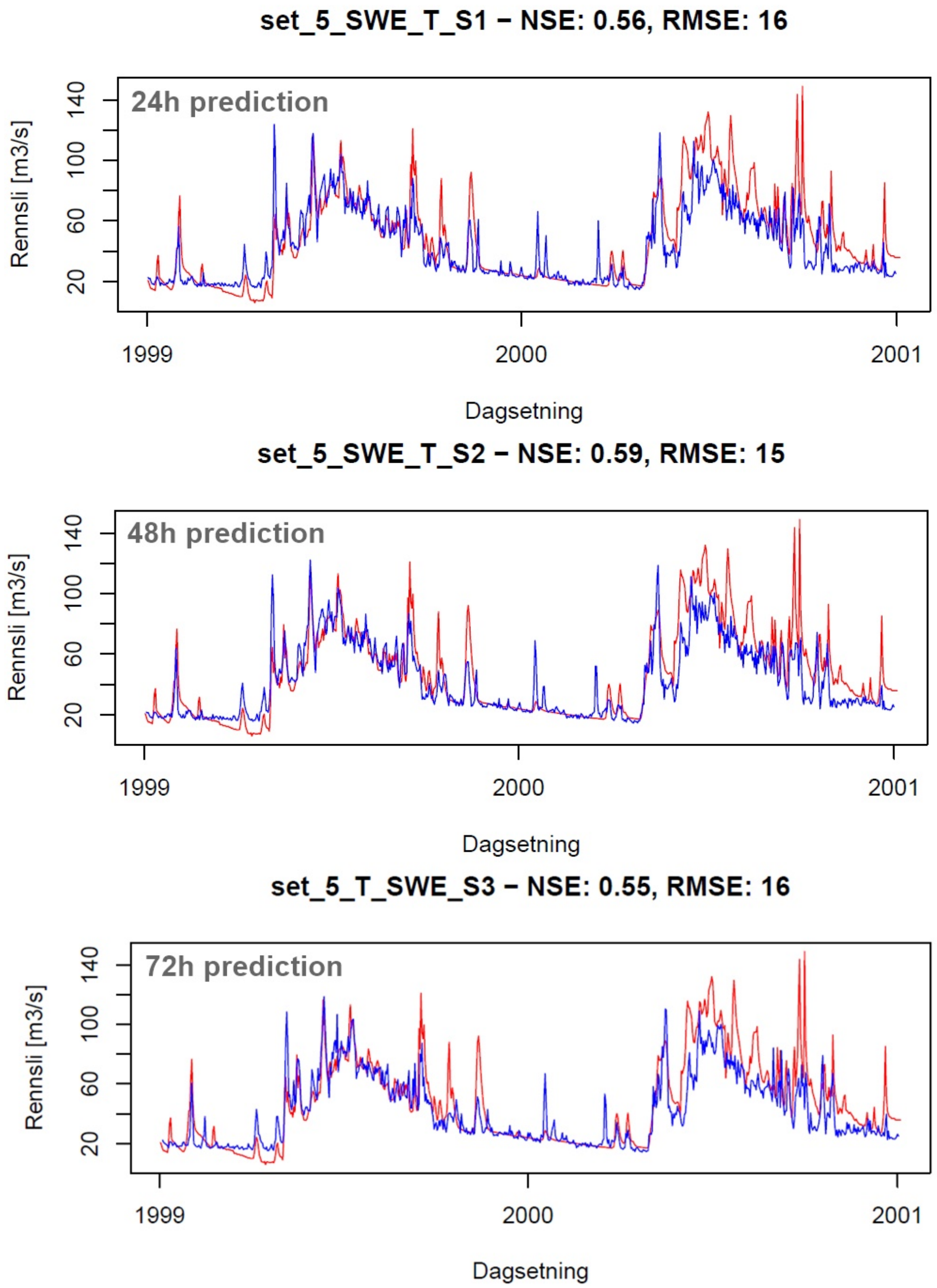

Figure 8. Best test results for ungauged station vhm 218 for the three first days of the forecast period; the measured discharge is shown in red and the forecast in blue.

5.3 Expanded forecasting system

The new sorting method was applied to all selected catchments to evaluate its validity and effectiveness. The hierarchization of variables has a positive effect on most of the stations, resulting in a reduced time-lag for most sites and a slight improvement in NSE values (Table 3). As in the previous report (Priet-Mahéo *et al.*, 2019), a trend was observed in the variables identified for successful predictions. This trend is also observed in the results for the new classification and appears in the variables used for the hierarchization as well as in the final set.

Table 3 summarises the best results for all the stations for 24-hour (D1) and 48-hour (D2) discharge forecasts. The name of the set combined the number of the final set used for the sorting followed by the acronym of the two variables used for the hierarchization. The table shows that sets 15, 19, 20 and 21 give the best results, and they can be divided into three categories:

1. Set 15 is the simplest and includes the 24-hour forecast for temperature and precipitation.
2. Set 19 focuses on temperature and SWE (present, past and future).
3. Sets 20 and 21 include the discharge for the previous days, as well as the forecasts for temperature, precipitation and SWE.

Set 20 yields the best results for most of the catchments (20 and 17 out of 33 catchments for the 24-hour and 48-hour forecasts, respectively).

Table 3. Summary of the best runs over the calibration period for the 33 gauging stations.

		D1: 24 hours						D2: 48 hours					
Class	VHM	Set	lag	NSE	mNSE	RMSE	ME	set	lag	NSE	mNSE	RMSE	ME
A	vhm043	set21_T_P_SR1	0	0.97	0.87	2.3	0.023	set21_SWE_P_SR2	0	0.95	0.82	3.2	0.15
A	vhm066	set15_P_T_SR1	0	0.75	0.63	8	-0.73	set20_P_T_SR2	0	0.72	0.6	8.5	-0.88
A	vhm068	set20_P_T_SR1	0	0.78	0.72	2.8	-0.067	set20_Q_T_SR2	0	0.72	0.67	3	0.047
A	vhm185	set20_SWE_P_SR1	0	0.97	0.89	0.57	0.045	set21_SWE_P_SR2	0	0.92	0.84	0.84	0.053
B	vhm102	set21_T_Q_SR1	-1	0.97	0.90	16	-1	set21_SWE_T_SR2	-1	0.94	0.85	22	-1.8
B	vhm150	set20_SWE_Q_SR1	0	0.88	0.78	10	-0.23	set20_P_T_SR2	0	0.88	0.75	10	-1
B	vhm162	set21_Q_T_SR1	-1	0.98	0.90	6.3	-0.64	set15_P_T_S2	-1	0.91	0.76	11	-2
B	vhm183	set20_T_P_SR1	0	0.89	0.76	10	-0.055	set21_T_SWE_SR2	0	0.81	0.67	14	-0.72
B	vhm233	set21_SWE_T_S1	0	0.93	0.82	11	-0.021	set19_T_Q_SR2	0	0.96	0.87	8.5	-1.1
B	vhm328	set21_T_P_SR1	-1	0.48	0.7	57	-0.64	set21_SWE_Q_S2	-1	-2.5	0.4	66	9.5
B	vhm408	set19_T_Q_SR1	0	0.98	0.91	0.81	-0.081	set15_T_P_SR2	0	0.93	0.83	1.7	-0.36
C	vhm048	set21_T_SWE_SR1	0	0.97	0.83	3.4	-0.35	set21_P_SWE_SR2	0	0.92	0.74	5.1	-0.64
C	vhm060	set20_T_P_SR1	0	0.83	0.7	2.1	-0.094	set20_Q_T_SR2	0	0.73	0.62	2.6	0.05
C	vhm064	set20_SWE_P_SR1	0	0.96	0.84	20	0.17	set20_P_SWE_SR2	0	0.91	0.75	29	1.8
C	vhm116	set21_T_SWE_SR1	-1	0.94	0.83	0.43	0.023	set21_T_SWE_SR2	-2	0.90	0.76	0.55	0.057
C	vhm145	set20_T_Q_SR1	-1	0.95	0.85	3	-0.018	set19_P_T_SR2	-1	0.89	0.77	4.7	-0.27
C	*vhm218*	*set_5_SWE_T_S1*	*0*	*0.56*	*0.49*	*16*	*4.6*	*set_5_SWE_T_S2*	*0*	*0.59*	*0.51*	*15*	*4.6*
C	vhm238	set20_P_SWE_SR1	0	0.95	0.80	8	-0.26	set19_T_P_S2	0	0.81	0.58	13	1.4
D	vhm019	set19_Q_SWE_SR1	-1	0.77	0.72	1.1	-0.017	set19_T_Q_SR2	-2	0.53	0.56	1.4	0.025
D	vhm038	set20_SWE_Q_SR1	-1	0.94	0.84	0.68	-0.04	set20_Q_SWE_SR2	-2	0.88	0.77	0.95	-0.07
D	vhm051	set20_Q_P_SR1	0	0.94	0.81	2.1	-0.11	set20_Q_P_SR2	0	0.89	0.74	2.7	-0.15
D	vhm092	set20_T_SWE_S1	0	0.88	0.73	0.6	0.049	set20_T_SWE_S2	0	0.81	0.69	0.74	0.045
D	vhm148	set15_T_Q_SR1	0	0.81	0.71	6.4	0.33	set20_T_Q_SR2	0	0.77	0.65	8	-0.26
D	vhm149	set20_P_T_SR1	0	0.73	0.7	12	0.1	set20_P_T_SR2	0	0.70	0.63	14	-0.35
D	vhm198	set20_P_SWE_SR1	0	0.94	0.84	4.3	-0.04	set20_T_P_SR2	0	0.89	0.77	6.2	-0.35
D	vhm200	set19_Q_SWE_S1	0	0.94	0.82	8.5	1.9	set19_SWE_Q_S2	0	0.90	0.78	11	2.1
D	vhm204	set20_SWE_P_SR1	0	0.95	0.84	1.7	-0.11	set20_Q_SWE_SR2	0	0.85	0.74	2.8	-0.091
E	vhm010	set20_Q_SWE_SR1	0	0.91	0.76	1.7	-0.0099	set20_Q_SWE_SR2	0	0.84	0.68	2.3	-0.067
E	vhm012	set20_Q_P_SR1	0	0.93	0.82	1.8	0.012	set20_SWE_Q_SR2	0	0.80	0.68	2.8	0.21
E	vhm026	set20_Q_P_SR1	0	0.98	0.83	1.7	-0.039	set20_SWE_T_SR2	0	0.95	0.76	2.5	-0.19
E	vhm045	set20_SWE_T_SR1	0	0.89	0.74	2.3	-0.21	set15_Q_SWE_SR2	0	0.84	0.61	2.7	-0.36
E	vhm128	set20_P_Q_SR1	0	0.84	0.73	8.8	0.68	set20_Q_T_SR2	0	0.77	0.63	12	0.62
E	vhm411	set20_Q_SWE_SR1	0	0.94	0.83	3.4	-0.063	set20_Q_SWE_SR2	0	0.89	0.76	4.5	-0.067

All catchment clusters benefit from the inclusion of discharge in the set. Forecasts for catchments of class A and B rely on temperature. These two classes contain catchments with glacial inputs and class A, being spring-fed rivers, can also depend on ice cover affecting the infiltration of runoff. Classes C and D have storages in the form of snow and or ice, and they perform better with the inclusion of SWE. In addition, classes D and E (direct run-off rivers) benefit from the inclusion of meteorological data; this applies to Class A (spring-fed rivers), as well.

Figure 9 summarises the results for all the gauging stations for the first day of forecast (D1), with a circle size representing the NSE obtained and the circle colour representing the main time-lag. The figure shows that most forecasts were temporally accurate, except for seven stations (marked in blue), which are all located in the northern half of the country, except one. The seven sites are vhm 19, vhm 38, vhm 102, vhm 116, vhm 145, vhm 162 and vhm 328. Several reasons could account for the misfits; for instance, stations vhm 38 and vhm 19 are affected by ice during the winter, and even though measurements used for these tests were screened and corrected, this introduces unreliability in the model. Furthermore, the rating curve for station vhm 19 is missing measurements during high discharge, which increases the uncertainty in stage-discharge comparisons. For sites influenced by glacial runoff, some discharge peaks could relate to the release of stored meltwater or changes in subglacial geothermal activity. Such influences cannot be included in the model, which relies on meteorological predictors. The stations represented with smaller circles in Figure 9 have lower NSE values. Some of these low values can be explained, for example, by the stations with the lowest score (vhm 328 and vhm 218), as they have shorter timeseries. Station vhm 328 is affected by jökulhlaups and vhm 218 is tested as ungauged, hence without a discharge reference. Further research is needed to understand some of the lower NSE results. A possible explanation could be the quality of the timeseries, as is the case for stations vhm 19, 148 and 149, where ice perturbations occur.

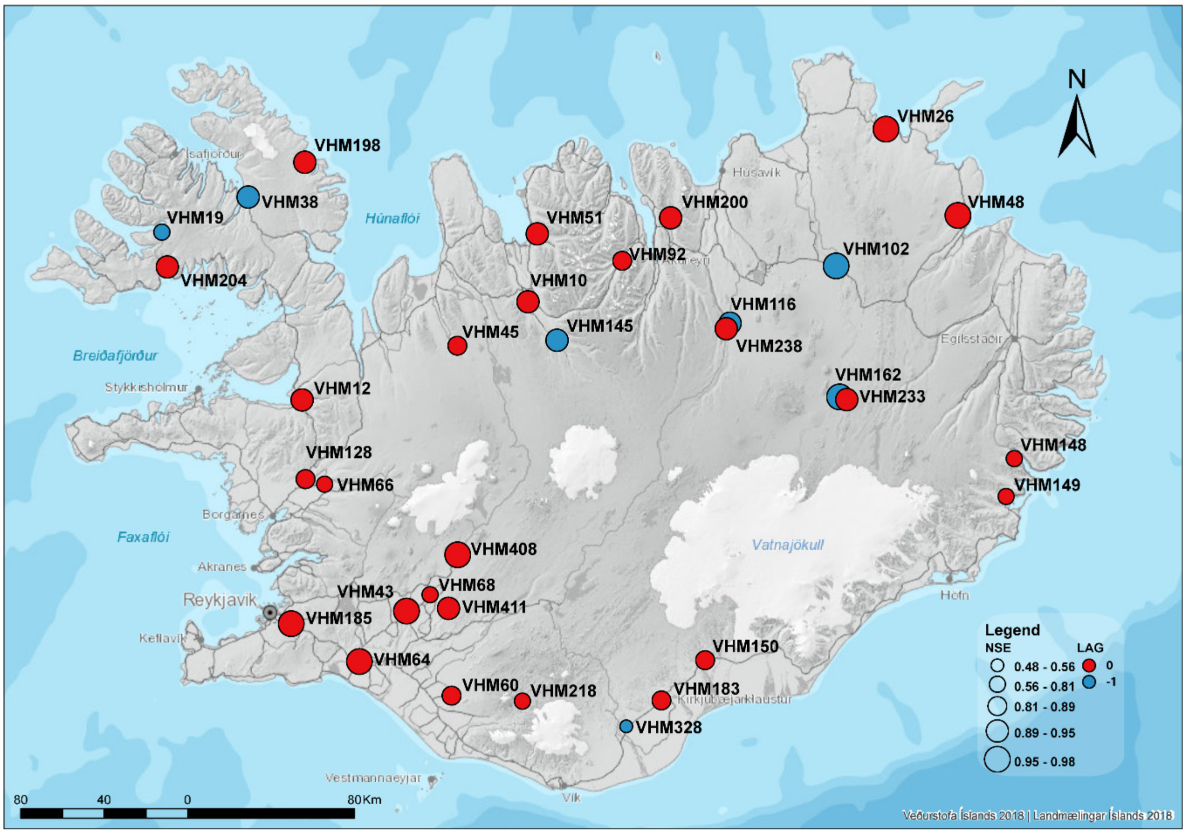

Figure 9. Geographical distribution of the quality of the 24-hour analogue sorting results. The circle size represents the Nash-Sutcliffe Efficiency coefficient while the colour indicates the lag; red for timely predictions and blue for missed predictions (one day late).

Most of the catchments benefit from a rescaling of the results for the first days of forecast (Priet-Mahéo *et al.*, 2019). The NSE coefficient for the first day of the forecast is generally higher than for the following days (Figure 10), but the coefficients remain high for most stations. Only four stations have a coefficient under 0.75 for the first day of the forecast; this number increases by three stations for the second day.

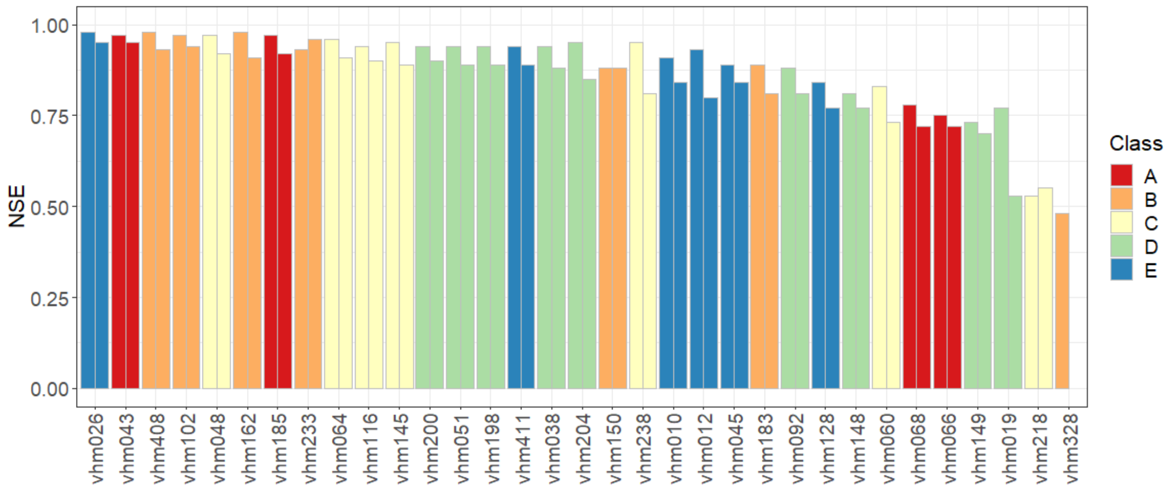

Figure 10. Nash Sutcliffe Efficiency coefficient (NSE) of the performance of the hierarchized analogue sorting method over the entire testing period for the 33 stations.

Figure 11 summarises the results visually for the five-day forecast. Each class returns correct predictions for the first three days, although a deterioration in accuracy (NSE) and timeliness (lag) occurs for most stations from day four onwards. For five stations (vhm 68, 148, 149, 183 and 328), the predictions for day five are poor and unreliable as the negative NSE indicates. Class E, representing direct run-off rivers, presents the best results over the five-day forecasts, both in terms of accuracy and timeliness, and all stations within that class, except for station vhm 10, have timely forecasts up to four days ahead with NSE above 0.5.

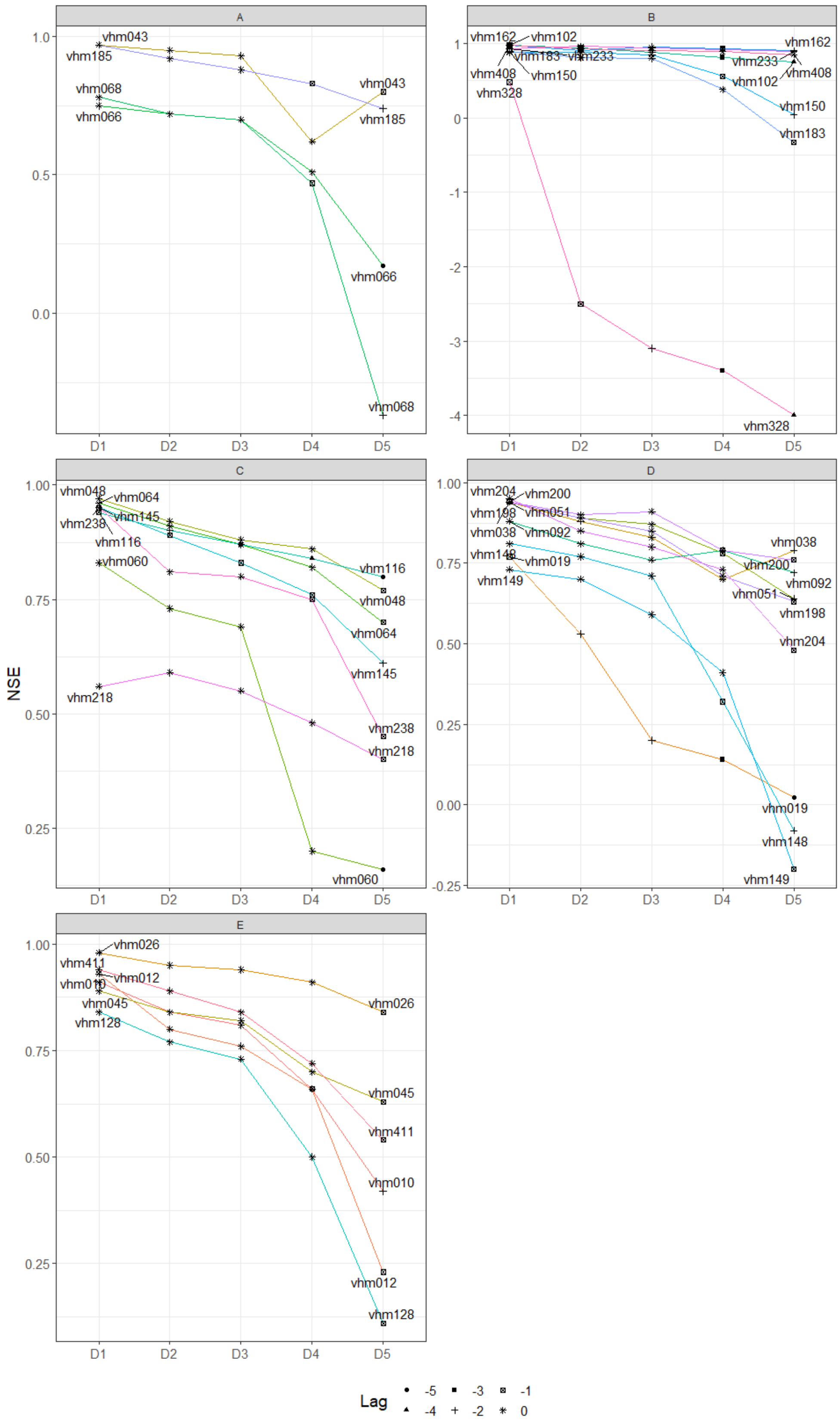

Figure 11. Nash-Sutcliffe Efficiency coefficient for the best-simulated discharge over the test period for the five days of forecast (D1 = the 24-hour forecast, D5 = the 120-hour forecast). The main lag associated with these results is represented by the scatter shape, and the results are organised by class.

5.4 Improved operational website

All figures presenting the results from the analogue forecast with the added catchments are hosted on the following website:

http://customer.vedur.is/vegag/analogue_forecast/analogue_sorting.html

As for the previous version of the webpage, the link opens a map of Iceland that shows the location of the catchments used in this project. The colour of each catchment is determined by the latest discharge measurements that have been recorded at the corresponding gauging station. If the latest discharge measurement is below the value of the 2-year return-period of the river, the watershed is coloured green. If the latest discharge measurement reaches 90% of the value of the 2-year return-period, the catchment is depicted in yellow. Similarly, if the observations reach 90% of the 5-year, 10-year and 25-year return-periods, the colours orange, red and brown are displayed, respectively. If a station has not sent data for more than a day, a symbol appears on the corresponding catchment as a visual warning.

By hovering your computer mouse over a catchment, results from the analogue forecast appear to the right of the flood-warning map. The lower subplot shows the measured daily-averaged discharge over the last 30 days, while results from the analogue sorting for the next five days are represented by a red line, which shows for each day the results from the most efficient predictor-set based on NSE values. A green shading area illustrates the minimum and maximum forecast interval for each day of the forecast. Additionally, results from past forecasts are shaded in light grey for the past day of the forecast interval. To help with the interpretation of discharge values, the horizontal dashed lines denote the 2-year return-period for daily averaged and instantaneous discharge values. If the daily threshold is reached, either by the observed discharge or by the forecast, the next threshold will be displayed and so on until the last threshold (25-year return-period) is reached. The upper subplot shows the simulated daily averaged temperature and daily summed rainfall for the last 30 days and for the next three days, as predicted by HARMONIE.

It is also possible to click on a catchment and open a new webpage where results are shown for the last three months and the past ten days. Three boxplots also feature on the page; they show temperature, precipitation and discharge values for the day in question relative to the same day of the year over the analysis period, thus placing today's values in statistical context. For further details about the configuration of the webpage and the plots, see the previous report (Priet-Mahéo *et al.*, 2019).

6 Conclusions

The analogue sorting method is a fast and powerful method to obtain discharge forecasts. The method has been in use for under two years at IMO in the form of a streamflow forecast for thirteen gauging stations. Overall, the results have been satisfactory; however, the presence of time-lag (a delay of one day) shows that the method was not always able to discriminate the relevant past events. In addition, interruptions in the computation of the forecast resulted from the data-flow disruptions. In this project, the operational streamflow forecast has been expanded from 13 to 33 gauged catchments. Additionally, a streamflow forecast has been setup for an ungauged catchment and improvements made on the forecasting system by correcting time-lags in peak flows.

The Mahalanobis distance does not introduce weight between its predictors, leading to some errors in the selection of analogue past events. In order to introduce some differential weight for these predictors, predictors set have been pre-selected ahead of final sorting, reducing the number of events to choose from. This simple approach introduces an hierarchization of the predictors. For most stations, the introduction of hierarchization of the predictors improves the forecasts significantly, both in terms of timeliness and accuracy. The hierarchization facilitates the discrimination of past events and each class of catchments shows some trend in preferential predictors that are consistent with their physical characteristics. A similar approach was used successfully for the ungauged station.

The usefulness of the streamflow forecast seems to be affected by factors such as the quality of the predictors (vhm 19 and vhm 38), and the existence of additional sources of discharge (*e.g.* vhm 116 and vhm 328). These disturbances would require additional attention before they could be included in the model. Ice perturbations could be accounted for through correction of the input data. Including water temperature measurements (when available) or air temperature could help to define temperature thresholds for possible ice growth. Volcanic and geothermal influences on discharge could also be investigated through the analysis of conductivity measurements, for example.

This research extended the analogue sorting forecast successfully to 33 catchments of diverse nature in various locations, underlining the usefulness of the method. Although the best results were associated with direct runoff catchments, all catchment types gave satisfactory forecasts up to three days ahead.

Further improvements of the forecast system could include: (i) the setup of back-up routines in case of data-flow interruptions; (ii) the deployment of the code on a development platform such as GitHub; (iii) the investigation of ice perturbations; (iv) correction or warnings and the introduction of predictors to account for volcanic and geothermal activity in some of the catchments; and (v) the extension of the system to all catchments monitored by IMO, including key ungauged catchments prone to flooding.

7 References

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

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., . . . Køltzowb, M. (2017). The HARMONIE--AROME model configuration in the ALADIN--HIRLAM NWP system. *Monthly Weather Review, 145*(5), 1919-1935. doi:https://doi.org/10.1175/MWR-D-16-0417.1

Crochet, P. (2013). *Probabilistic daily streamflow forecast using an analogue sorting method.* Reykjavík: Veðurstofa Íslands.

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

Nawri, N. (2014). *Evaluation of HARMONIE reanalyses of surface air temperature and wind speed over Iceland.* Reykjavik: Icelandic Meteorological Office.

Nawri, N., Pálmason, B., Petersen, G., Björnsson, H., & Þorsteinsson, S. (2017). *The ICRA atmospheric reanalysis project for Iceland.* Reykjavik: Icelandic Meteorological Office.

Priet-Mahéo, M., Massad, A.-G. R., Pétursdóttir, S., Þórarinsdóttir, T., & Egilson, D. (2019). *Daglegar rennslisspár með notkun hliðstæðrar greiningar Harmonieveðurgagna.* Reykjavík: Veðurstofa Íslands.

Schulla, J. (2017). *Model Description WaSiM.* Technical report.

Yates, D. (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research, 39*(7).