Hydrol. Earth Syst. Sci. Discuss., 8, C2764–C2782, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C2764/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on " Simplifying a hydrological ensemble prediction system with a backward greedy selection of members – Part 2: Generalization in time and space" *by* D. Brochero et al.

D. Brochero et al.

darwin.brochero.1@ulaval.ca

Received and published: 14 July 2011

It was shown in the companion paper (Brochero et al., 2011a) how the interchangeability of members of the Hydrological Ensemble Prediction System (HEPS) at hand can be exploited through the participation of hydrological models in the subset of selected members, possibly as a result of the interchangeability of the Meteorological

EPS (MEPS) that serves it as input. In a similar fashion, this paper proposes the evaluation of random selections with and without the guidance of the response found with Backward Greedy Selection and Cross Validation, hereafter BGS-CV, evaluating performance and interchangeability of HEPS members (see Fig. 4 and 5 of this report).

Additionally, to avoid confusion that can bring the explicit analysis of members of the MEPS in the selected subset, on account of its inherent interchangeability, Fig. 4 and paragraphs that considered their interpretation will be removed.

1 General Comments

1.1 The presentation of the scores used in this article (section 3) is very similar to the section 2 of the part 1 of this study (companion paper). This is not necessary; the authors could reduce this section to a few lines, referring to Part 1. As detailed in the specific comments, some other elements (one figure, one table and the section 2) are not necessary because already given in Part 1 or would be more useful in Part 1. It will reduce the size of this article. Maybe it would thus be worth adding some complementary study?

We tried to give independence to each paper to facilitate its evaluation and reading. However, taking into account this comment, the new version of this paper considers the following modifications:

- Elimination of Table 1.
- Sections 2, 3 and 4 will be reduced to a new section called "Experimental set-up".
- Modification of Fig. 2.
- Elimination of Fig. 4.

- Insert a new figure showing the performance of initial HEPS (800-member) on different FTH (see Figure 3 of this report).
- Insert a new figure showing the generalization of the BGS-CV selection on different FTH (see Figure 4 of this report).

2 Specific Comments

2.1 Abstract, line 7: please clarify already in the abstract what is a 94% simplification (i.e. that it concerns the number of members)

This sentence will be reformulated as follows:

This article evaluates a simplification equivalent to 94% of members of the initial 800-member HEPS (i.e. 750 members are removed). The initial HEPS is obtained forcing 16 lumped rainfall-runoff models with the 50 perturbed members from the European Centre for Medium-range Weather Forecasts (ECMWF) EPS.

2.2 Section 1, last paragraph: the goal of this article does not appear clearly to me

The penultimate and the last paragraph will be rewritten as follows:

Vrugt et al. (2006) posed the BMA inverse problem in a multi-objective framework, examining the Pareto set of solutions between the Continuous Ranked Probability Score (CRPS), the Mean Absolute Error (MAE), and the Ignorance Score with the AMALGAM method (Vrugt and Robinson, 2007). In this line, the companion paper shows that a combined criterion which groups various characteristics of the probabilistic forecast is adequate to guide the selection of members with BGS-CV method. At this point it is important to note that the BGS-CV method offers the possibility of combining results from

different studies, which is highlighted as one of the aspects related to the improvement of HEPS (Cloke and Pappenberger, 2009).

In this paper we evaluate the generalization of the selection of members with BGS-CV with the methodology presented in Sect. 2. Thus, we test the members' selection obtained in sixteen catchments for the 9-day lead time, for the other 8 lead times. Additionally we evaluate the ability to extrapolate the selections to neighbouring catchment. Finally we present the integration of results from different catchments within a regional framework. Results and discussion are gathered in Sect. 3, while conclusions and a guideline for future work are drawn in Sect. 4.

2.3 The Velazquez et al. (2010) reference used in this article is not enough. Consider replacing this EGU abstract reference with the Velazquez et al. (2011) paper published C1478 Discussion Paper in Advances in Geosciences, which is much more complete.

Done.

2.4 Page 2788 end of 1st paragraph: the authors state that "it is important to note that some models. . .". Why is it important? Could you explain? This statement is not used in the rest of the article for any further interpretation of the results, whereas it should be if you keep this sentence.

The following paragraph " It is important to note that some models such as the HM02 were specifically devised for the catchment scale, whereas others such the HM06 or HM08, inspired from distributed models, have suffered a series of substantial changes bringing them to a lumped state." will be rewritten as follows :

It is important to note that this study focuses on evaluating the probabilistic hydrological forecasting from a cooperative point of view seeking diversity in the final selection, i.e. that each member acts as a complement to the other. This clarification is relevant in order to avoid misinterpretation of competitiveness in the different conceptualizations of the sixteen hydrological models used. It should be clear that the comparison would not be fair because some models such as the HM02 were specifically devised for the catchment scale, whereas others such the HM06 or HM08, inspired from distributed models, have suffered a series of substantial changes bringing them to a lumped state.

2.5 Section 2: please add this reference for SAFRAN (Quintana-Seguí et al., 2008). If the 50 year reanalysis has been used, please add: Vidal et al., 2010.

Done.

2.6 Page 2798 line 10: the 50% is the minimum gain and 87% is the maximum gain The RD_{mse} shows minimum gains of 50% (catchment B21) and 87% (catchment K17)

You are right. This sentence will be rewritten as follows: RD_{mse} shows a minimum gain of 50% (catchment B21) and a maximum gain of 87% (catchment K17), reflecting the emphasis given to this property in the formulation of the combined criterion used for optimizing where this component was given a weight twice that of the others. With respect to the IGNS, index gains between -5% and 27% (excluding the catchment B21) reflect an acceptable behaviour.

2.7 Table 1 is already given in the companion paper; I don't think it is necessary to give it in Part 2.

You are right.

2.8 Table 4: could you explain the huge difference we observe for FTH 4 for four basins of cluster 2? The score is much better than for the other FTHs, which is surprising.

After verifying the peculiar behaviour of the catchment J85 for the 4-day lead time (Fig. 3) and of the catchments that you mentioned in Table 4 in Brochero et al. (2011b), we have identified a problem with the definition of the threshold z_1 used to manipulate the IGNS when we extrapolate results to the first lead times.

For example in the case of the 4-day lead time for catchment, the reference IGNS was equal to -2.0112 bits and the 50-member regional selection was equal to -1.9405 bits, so the weight of this component in the normalized sum would be equal to -5.31 ($(-2+1.9405)/(-2+2.0112)$). A value that lacks interpretability because the normalization of scores is a measure that represents the degree of improvement achieved by the members' selection with respect to the initial HEPS of 800 members. The measurement scale of each component of the normalized sum would then fall between zero and infinity. A value of zero would indicate that the selection reaches the minimum value of the score in evaluation, while a unit value represents a performance of the members' selection equal to the reference set, and finally a value greater than one indicates that the selection deteriorates the score in evaluation.

In conclusion, the results presented in the original paper, which focus on the extrapolation of the BGS-CV methodology, will be corrected using a threshold z_1 equals to -4 to avoid the problem mentioned above. Tables 2 and 3, presented in this report, provide the corrected results. Note that differences in the delta ratio with regard to the Tables 3 and 4 presented in Brochero et al. (2011b) are due to the randomness that is assigned to the observation when it is equal to several members of the forecast.

More importantly, note that this modification does not alter the members' selection at the 9-day lead time with the threshold $z_1 = -2$, which was the scenario in which the selection of members was run, because the IGNS never reached values lower than -2.

2.9 Fig. 1: please draw the area of the basins. Are some of these basins included in other ones used in this study?

Done. see Figures below.

2.10 Please consider putting Figure 2 in the Part 1 companion paper instead of Part 2.

As you proposed, it is appropriate to show this figure in the companion paper under a slight modification in the choice of the score. Thus Fig. 2 of this report will be presented in the new version. It basically shows the methodology for evaluating the generalization of the members' selection.

2.11 Fig. 5: it is difficult to see anything on the time series plots of CRPS and IGNS, please consider improving them. The first two lines of the legend do not correspond with the description of this figure that is given page 2799 lines 24-25.

Done. see Figures below.

2.12 Fig. 6: The legend should be: "Hydrological models participation. (a) (b) (c) (d) (e) Distribution in the five regions (clusters) presented at the Fig. 4. (f) Model performance evaluated as the mean rank."

Done. see Figures below.

New figures (designed in colour for easy viewing) are given at the end of this document.

The complete captions for the figures below are:

Fig. 1. Selected catchments. Each catchment is identified with the first three digits of each code used in Table 1.

Fig. 2. Generalization of members selection with BGS-CV.

Fig. 3. Interquartile range of $\mathrm{RD}_{\mathrm{mse}}$ and δ ratio assessed in the 28 catchments studied by Velázquez et al. (2011) under two HEPS schemes: 16-member ensemble (16 hydrological models are driven by the deterministic forecast from ECMWF) and the 800-member ensemble (16 hydrological models are driven by the 50-perturbed member forecast from ECMWF).

Fig. 4. Evolution of the normalized sum (NS) to evaluate the response sensibility with regard to the interquartile range (iqr) of 200 random experiments in different forecast time horizons (FTH) with and without any guidance of the Backward Greedy members' Selection. Each catchment is identified with the first three digits of each code used in Table 1.

Fig. 5. Evolution of the normalized sum (NS) to evaluate the response sensibility of the extrapolation of results in others catchments with regard to the interquartile range (iqr) of 200 random experiments in different forecast time horizons (FTH) with and without any guidance of the Backward Greedy members' Selection. Each catchment is identified with the first three digits of each code used in Table 1.

Fig. 6. Comparison between the initial ensemble (800 members) and the ensemble selected (50 members) for a lead time of 9 days. **(a)** Figure above: observed flow; figure below: CRPS (x-axis formatted as: day/month). Note the correspondence between higher observed flows and higher CRPS. **(b)** Figure above: observed flow; figure below: IGNS. Note that there is no full correspondence between the higher IGNS and higher observed flow (x-axis formatted as: day/month). **(c)** Reliability diagram error (MSE based on vertical distances between the points). **(d)** Rank histogram for the 30 selected members. The horizontal dashed line indicates the frequency $(N/d+1)$ attained by a uniform distribution. **(e)** Occurrences of the employed models in the final

solution of 50 members.

Fig. 7. Hydrological Models participation. Distribution in the five regions (clusters) are presented in **(a)**, **(b)**, **(c)**, **(d)**, and **(e)**. Model performance evaluated as the mean rank index is shown in **(f)**.

References

Brochero, D., Anctil, F., and Gagné, C.: Simplifying hydrological ensemble prediction system with a backward greedy selection of members, Part I: Optimization criteria, Hydrol. Earth Syst. Sci. Discuss., 8, 2739–2782, 2011a.

Brochero, D., Anctil, F., and Gagné, C.: Simplifying hydrological ensemble prediction system with a backward greedy selection of members, Part 2: Generalization in time and space, Hydrol. Earth Syst. Sci. Discuss., 8, 2783-2820, 2011b.

Cloke, H. and Pappenberger, F.: Ensemble flood forecasting: a review, J. Hydrol., 375, 613–626, 2009.

Velázquez, J. A., Anctil, F., Ramos, M. H. and Perrin C.: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, Advances in Geosciences, 29, 33–42, 2011.

Vrugt, J. A., Clark, M. P., Diks, C. G. H., Duan, Q., and Robinson, B. A.: Multi-objective calibration of forecast ensembles using Bayesian model averaging, Geophys. Res. Lett., 33, L19817, 2006.

Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, P. Natl. Acad. Sci. USA, 104(3), 708–711, 2007.

Table 1. Main characteristics of the studied basins (mean annual values) based on a 36 year length of the series (1970–2006). P: precipitation, ET: potential evapotranspiration, Q: flow. For the distinction of the basins used in training and testing, the latter are highlighted in bold.

Catchment codes	Area (km^2)	P (mm)	ET (mm)	Q (mm)	Catchment codes	Area (km^2)	P (mm)	ET (mm)	Q (mm)
A6921010	2780	3.04	1.79	1.18	M0680610	7380	2.04	1.93	0.56
A7930610	9387	2.78	1.80	1.21	O3401010	2170	3.19	1.80	1.90
A9221010	1760	2.49	1.83	0.91	Q2593310	2500	2.52	2.24	0.75
B2130010	2290	2.57	1.80	0.87	U2542010	4970	3.63	1.75	1.88
B3150020	3904	2.58	1.80	1.09	**A7010610**	6830	2.99	1.78	1.46
H2482010	2982	2.31	1.89	0.84	**H6221010**	2940	2.50	1.83	0.92
H3621010	3900	1.98	1.95	0.45	**M3600910**	3910	2.31	1.88	0.80
H5321010	8818	2.41	1.85	0.93	**K1341810**	2277	2.65	1.89	1.02
J8502310	2465	2.36	1.89	0.81	**M1531610**	7920	1.85	1.95	0.36
K1773010	1465	2.65	1.94	1.07	**P7001510**	1863	2.88	2.08	1.19
K7312610	1712	2.13	2.01	0.68	**P7261510**	3752	2.65	2.14	0.87
M0421510	1890	2.04	1.89	0.62	**U2722010**	7290	3.63	1.79	2.07

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 2783, 2011.

Table 2. Selection of 50 members based on combined criterion and the combination of k-fold cross-validation results on the forecast time horizon 9. Beside each score is presented the gain index evaluated by Eq. 10. NS represents the normalized sum (Eq. 6 with unit weights). NHM indicates the number of hydrological models participating in the solution.

Catchment codes	Scores: CRPS	$RD_{mse}[e-3]$	δ	IGNS	MDCV function	NS
A69	0.284 (+0%)	1.3 (+81%)	1.5 (+18%)	0.67 (+14%)	0.39 (+5%)	4.0
800 memb.	0.284	7.0	1.8	0.78	0.37	5.0
A79	0.254 (+3%)	1.5 (+69%)	3.6 (−11%)	0.34 (+23%)	0.41 (−1%)	4.4
800 memb.	0.263	5.1	3.3	0.44	0.41	5.0
A92	0.183 (+4%)	0.3 (+86%)	2.3 (−28%)	-0.42 (+27%)	0.57 (+0%)	4.4
800 memb.	0.192	2.4	1.8	-0.33	0.57	5.0
B21	0.232 (−1%)	1.2 (+49%)	2.6 (−16%)	-0.18 (−38%)	0.63 (+9%)	4.6
800 memb.	0.230	2.4	2.2	-0.29	0.57	5.0
B31	0.134 (+1%)	1.3 (+72%)	2.0 (+27%)	-0.84 (− 5%)	0.24 (+7%)	4.0
800 memb.	0.135	4.5	2.7	-0.88	0.22	5.0
H36	0.157 (+2%)	0.7 (+80%)	2.0 (−37%)	-1.02 (+ 2%)	0.36 (−1%)	4.5
800 memb.	0.161	3.5	1.5	-0.99	0.37	5.0
H53	0.165 (+3%)	1.9 (+74%)	4.3 (−39%)	-0.76 (+ 8%)	0.36 (+8%)	4.6
800 memb.	0.171	7.4	3.1	-0.71	0.33	5.0
H24	0.180 (+2%)	2.2 (+68%)	3.8 (−32%)	-0.82 (+ 9%)	0.37 (+6%)	4.6
800 memb.	0.185	7.1	2.9	-0.76	0.35	5.0
K17	0.205 (+4%)	0.5 (+87%)	1.8 (− 9%)	-0.73 (+12%)	0.38 (−2%)	4.2
800 memb.	0.213	3.6	1.7	-0.65	0.39	5.0
U25	0.290 (+0%)	0.9 (+74%)	2.6 (− 1%)	-0.40 (+13%)	0.38 (+7%)	4.2
800 memb.	0.289	3.4	2.5	-0.36	0.35	5.0
J85	0.159 (+2%)	0.4 (+80%)	1.7 (− 5%)	-1.00 (+ 2%)	0.40 (+8%)	4.2
800 memb.	0.163	2.2	1.7	-0.98	0.37	5.0
K73	0.160 (+3%)	0.9 (+70%)	2.1 (− 5%)	-0.93 (+ 0%)	0.38 (+9%)	4.3
800 memb.	0.165	3.1	2.0	-0.93	0.35	5.0
M04	0.158 (+1%)	0.6 (+68%)	1.6 (−2 %)	-0.98 (− 1%)	0.37 (+2%)	4.3
800 memb.	0.160	1.7	1.6	-0.99	0.37	5.0
M06	0.153 (+4%)	0.3 (+79%)	1.6 (−4 %)	-1.09 (+ 6%)	0.39 (+1%)	4.2
800 memb.	0.159	1.4	1.5	-1.03	0.38	5.0
O34	0.166 (+2%)	1.0 (+71%)	1.6 (+1 %)	-0.91 (+ 5%)	0.37 (+3%)	4.2
800 memb.	0.169	3.5	1.6	-0.86	0.36	5.0
Q25	0.159 (+3%)	0.6 (+73%)	1.1 (+22%)	-0.94 (− 5%)	0.39 (+4%)	4.0
800 memb.	0.163	2.1	1.4	-0.98	0.37	5.0

Table 3. Test based on the normalized sum with unit weights in new catchments and different forecast time horizons (FTH) of regional integration given by the analysis of clusters by location. Values lower than 5 determined that the scores of selection are better than the reference set. See clusters' distribution at the Fig. 2. In each cluster, the catchments highlighted in bold represent the series that are not used by the members' selection training.

FTH	Cluster 1								Cluster 2						
	H24	K17	U25	**K13**	**K52**	**U06**	**U24**	**U27**	J85	K73	M04	M06	**H93**	**M15**	**M36**
1	5.08	5.25	5.06	5.19	5.36	5.20	5.15	5.12	4.96	5.19	5.09	5.07	5.06	5.09	4.96
2	5.17	5.18	5.12	5.07	5.24	5.02	5.36	5.04	5.03	4.97	4.97	4.89	4.85	4.90	5.00
3	4.89	4.85	4.87	4.71	5.01	4.60	4.86	4.78	4.66	4.63	4.67	4.73	4.71	4.70	4.67
4	4.50	4.56	4.69	4.26	4.76	4.53	4.68	4.59	4.67	4.57	4.72	4.71	4.70	4.71	4.60
5	4.82	4.56	4.56	4.31	4.85	4.54	4.76	4.68	4.70	4.33	4.51	4.54	4.40	4.43	4.29
6	4.99	4.74	4.86	4.59	4.87	4.59	4.76	4.79	4.41	4.47	4.53	4.29	4.49	4.53	4.34
7	4.50	4.52	4.42	4.58	4.74	4.50	4.52	4.50	5.01	5.04	5.00	4.81	4.77	4.80	4.80
8	4.38	4.25	4.27	4.16	4.71	4.22	4.33	4.33	4.43	4.61	4.78	4.62	4.47	4.84	4.41
9	4.50	3.97	4.09	4.04	4.36	4.07	4.32	4.17	4.09	4.32	4.59	4.39	4.31	4.39	4.22

FTH	Cluster 3				Cluster 4				Cluster 5				
	O34	Q25	**P70**	**P72**	B31	H36	H53	**H62**	A69	A79	A92	B21	**A70**
1	4.88	4.68	4.74	4.78	5.69	5.21	4.92	5.09	4.20	4.78	4.42	4.98	4.94
2	4.83	4.61	4.73	4.81	5.85	5.11	4.64	5.15	4.40	4.98	4.78	4.52	5.22
3	4.16	4.36	5.98	4.74	5.83	4.69	7.24	4.65	5.03	5.42	5.02	4.96	5.45
4	4.77	3.43	4.47	4.28	5.97	4.49	5.23	7.01	5.19	5.57	5.58	5.11	6.22
5	4.80	4.53	4.69	4.68	5.71	5.29	5.24	5.60	5.10	5.80	4.74	5.50	5.60
6	4.68	4.47	4.59	4.55	5.78	4.96	5.41	5.45	4.78	5.62	5.32	5.31	5.45
7	4.62	4.74	4.45	4.32	5.24	4.60	4.81	5.16	5.12	5.11	4.35	5.53	5.57
8	4.70	4.34	4.39	4.28	4.58	4.57	4.91	5.46	4.97	5.22	4.25	5.50	5.08
9	4.36	4.15	4.28	4.12	4.26	4.08	4.50	4.74	4.87	4.66	4.45	4.92	5.38

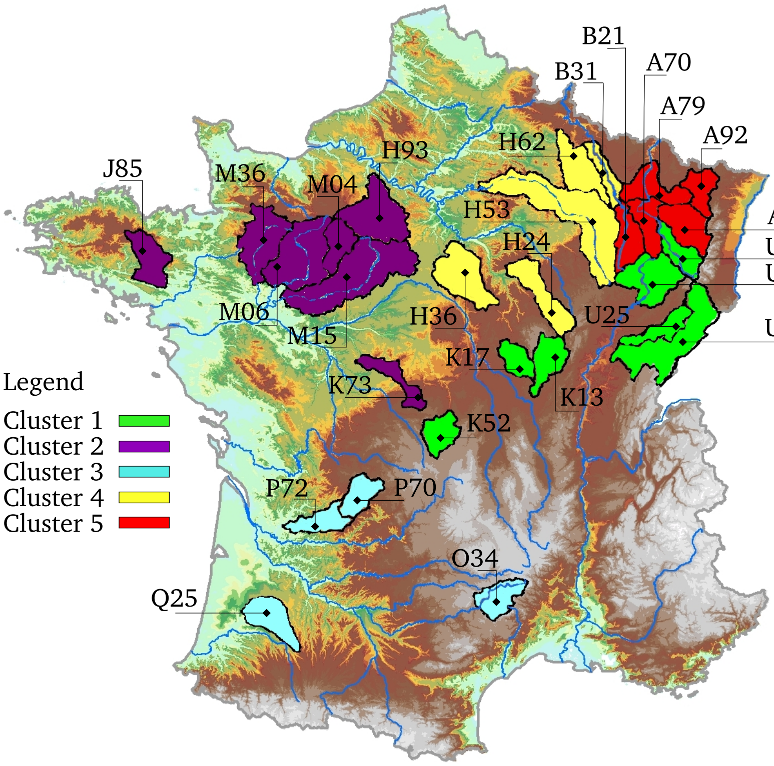

Fig. 1. Selected catchments.

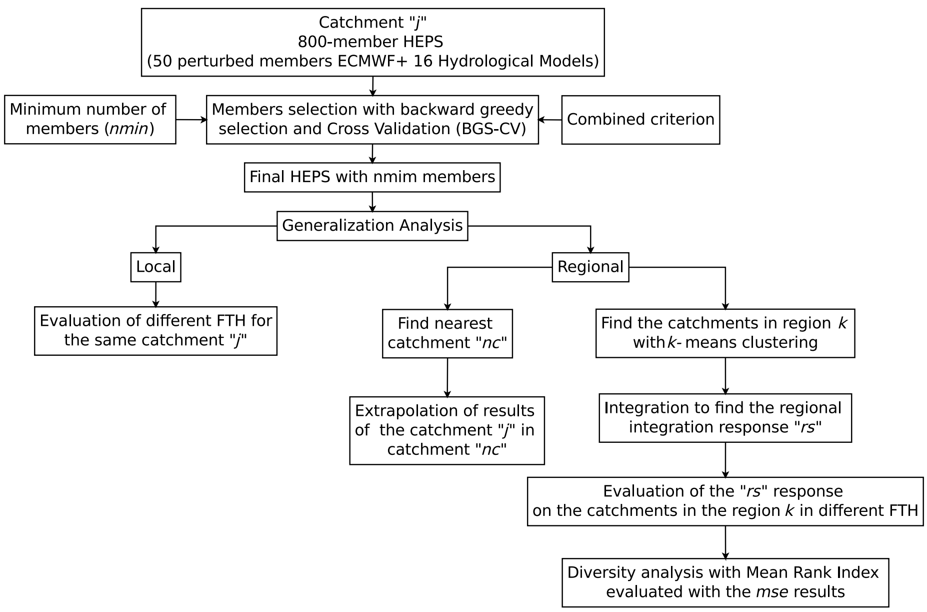

Fig. 2. Generalization of members' selection with BGS-CV.

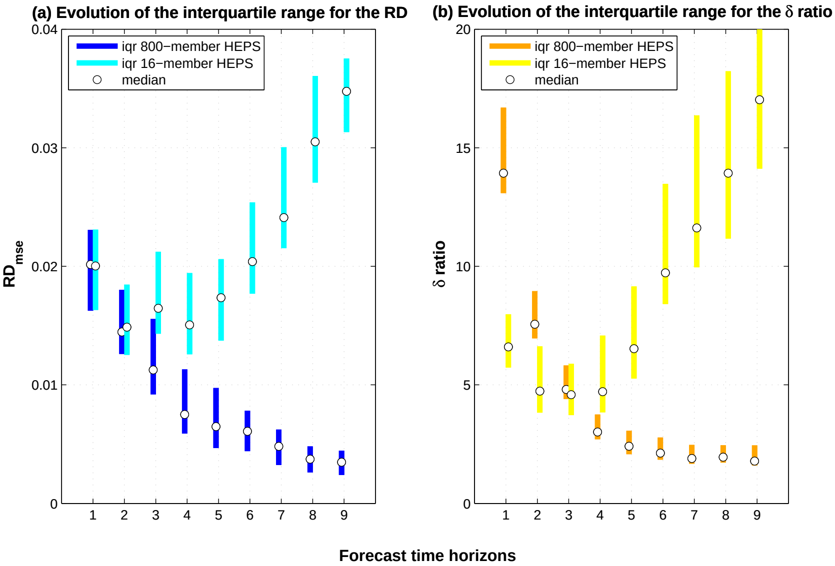

Fig. 3. Interquartile range of RD and delta ratio assessed in the 28 catchments studied by Velazquez et al. (2011).

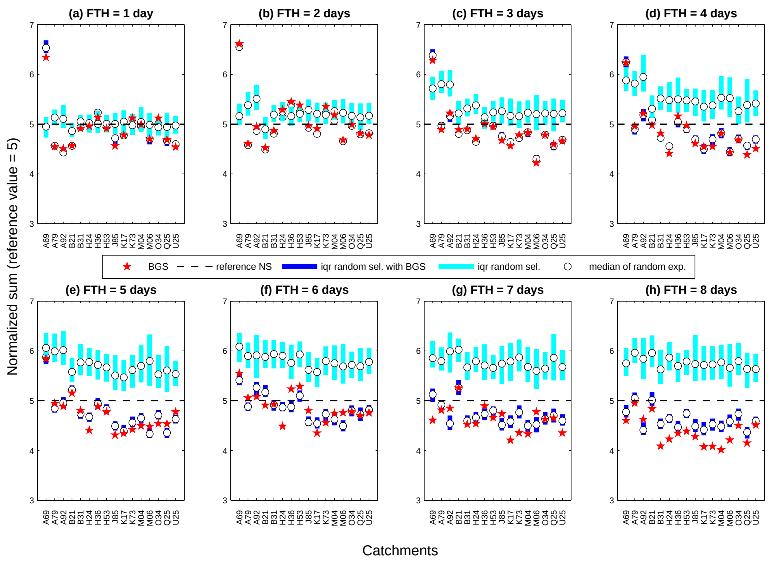

Fig. 4. Evolution of the normalized sum (NS) to evaluate the response sensibility with regard to the interquartile range (iqr) of 200 random experiments in different forecast time horizons (FTH).

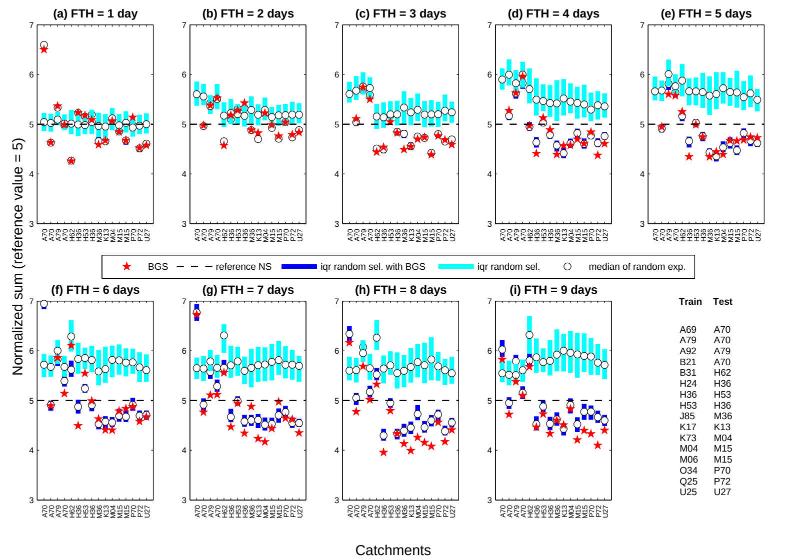

Fig. 5. Evolution of the normalized sum (NS) to evaluate the response sensibility of the extrapolation of results in others catchments.

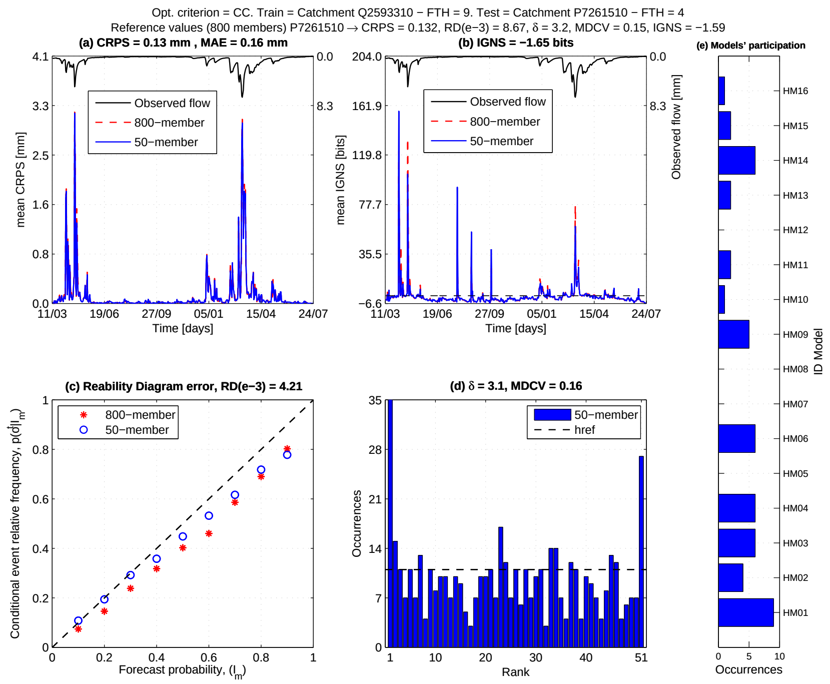

Fig. 6. Comparison between the initial ensemble (800 members) and the ensemble selected (50 members) for a lead time of 9 days.

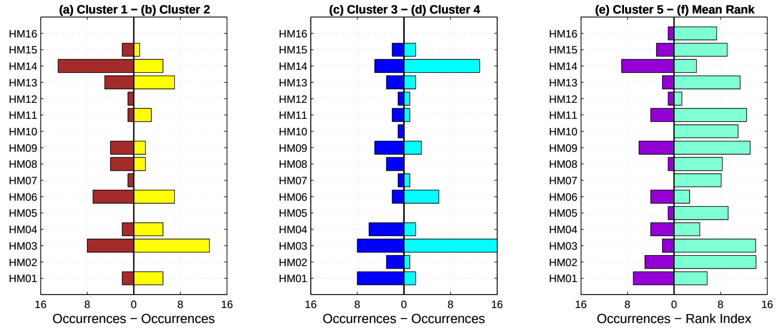

Fig. 7. Hydrological Models participation.