Hydrol. Earth Syst. Sci. Discuss., 12, C4545–C4568, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C4545/2015/

Hydrology and Earth System Sciences
Discussions

HESSD

12, C4545–C4568, 2015

Interactive Comment

Interactive comment on "Accounting for three sources of uncertainty in ensemble hydrological forecasting" *by* A. Thiboult et al.

A. Thiboult et al.

antoine.thiboult.1@ulaval.ca

Received and published: 31 October 2015

The authors would like to thank the two anonymous referees for their careful reading and the interesting comments they provided that will contribute to the quality of the paper.

Shortly after initial submission, while working on the same database, we realize that the streamflow measurements of two of the 20 catchments where of dubious quality. Several clues led us conclude that they should not be longer included in the catchment pool. We sincerely apologize for any inconvenience this may have caused.

The two catchment that have been withdrawn from the initial submission are the ones that often behaved like outliers and were the most unreliable. Thus, the new results

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

are more homogeneous. The two problematic catchments have been substituted by two new. Also, Figures 3, 4, 5, 6, 8, 9, 10, 11, 12, 13 were updated and are provided along this answer. Even if this does not affect any of the conclusion, the author suggest several modifications of the text:

- page 7193, line 17-19: " Exceptions can be occasionally observed for catchment 3, 17, and 20 where only one or two models outperform the ensemble." should be replaced by " Exceptions can be occasionally observed for catchment 3 and 17 where only one or two models outperform the ensemble."
- page 7194, line 5-6: we suggest to modify "For the first lead time, most of the catchments are close to reliability while there is a clear outlier for which accuracy skills do not match its corresponding spread. In fact, this low performing catchment exhibits a constant hydrological wet bias – partially explained by a meteorological forecast wet bias that over-forecasts precipitations by 15% – that is not captured by any of the models even if the global tendency is respected." to "For the first lead time, most of the catchments are close to reliability while there are two outliers for which accuracy skills do not match their corresponding spread. In fact, these catchments exhibits a constant hydrological bias – partially explained by an inaccurate meteorological forcing – that is not captured by any of the models even if the global tendency is respected"
- page 7194, line 20: "(outlier)" should be deleted.
- page 7194, line 20-23: "Data assimilation is particularly effective on catchments that present a systematic bias. For example, catchment number 11 that was problematic from the first lead time lies among the other catchments in terms of performance." should be deleted, even if we think that DA is particularly effective on catchment that have a systematic bias, but this assertion is no longer explicitly supported by the new Figure.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

- page 7196, line 12-13: values should be replaced from “While they are almost identical with a value of 0.55 and 0.57 mm day-1 respectively for the day 3, G spread drops to 0.44 mm day-1 for day 9 while the use of the MEPS maintains the spread to 0.55 mm day-1” to “While they are almost identical with a value of 0.58 mm day-1 and 0.59 mm day-1 respectively for the day 3, G spread drops to 0.45 mm day-1 for day 9 while the use of the MEPS maintains the spread to 0.59 mm day-1.”
- page 7197, line 18-19: “and only model 1 and 5 perform” should be replace to “and only models 1, 5, and 17 perform”
- page 7198, line 3: “catchment 19”, should be replaced to “catchment 20”.
- page 7199, line 1: “reducing the overdispersion with a sensible decrease in the ensemble spread from 0.65 to 0.54 mm day-1” should be replaced with “reducing the overdispersion with a sensible decrease in the ensemble spread from 0.72 to 0.57 mm day-1”
- page 7199, line 8-11: “The two outlier catchments that exhibit poorer reliability present an underdispersed forecast that is a bit more pronounced for the H system than the H system (see Fig. 9). This indicates that uncertainties used to define the EnKF perturbations are under-estimated.” should be suppressed. We also suggest to replace by “As a matter of fact” by “Finally”.

This manuscript investigated impacts of different uncertainty sources on streamflow forecasting comparing various combinations including multi-model ensemble, data assimilation, and meteorological ensembles. It fits well the scope of Hydrology and Earth System Sciences and the topic is of interest to a broad ranges of the scientific and engineering community. Their research questions and methodologies are of importance to better improve understanding on prediction uncertainty. However, for some study materials, description

and information are not enough to convince general readers of their results. Especially, I have concerns on excessively simplified application of hydrologic models in terms of spatial and temporal scales and interpretation of contribution of different uncertainty sources. Therefore, revisions should be required to clarify several issues shown below before possible publication:

Major comments:

1. Multimodel ensemble:
Abstract: One of main findings of this manuscript is that the multimodel approach to take into account structural uncertainty supports the streamflow forecasts to maintain the required dispersion throughout the entire forecast horizon. However, such a statement might mislead a conclusion as if structural uncertainty is a dominating factor rather than forcing uncertainty, which could not convince readers with given results of this study. The fact that contribution of the meteorological ensemble forcing was negligible compared to deterministic one (Fig. 8) could strengthen such misinterpretation. In this study, as I understood correctly, input uncertainty (e.g. forecast forcing) seemed to be compensated by structural uncertainty (e.g. multimodel) to enhance performance metrics. In addition, when we recall one of aims of this study is to decipher the traditional hydrometeorological sources of uncertainty (Page 7183), it is a bit doubtful if their aim was achieved and demonstrated successfully. Please clarify your findings and opinions on hydrologic prediction uncertainty which can be concluded from your study results.

The authors do not claim that the meteorological uncertainty should be neglected, far from it, and we agree that this source of uncertainty is among dominant ones. Nevertheless, the authors also believe and show that the uncertainty arising from the structure of hydrological model have to be taken into account. If not, system outputs will clearly underestimate total predictive uncertainty.

In this article, rather than trying to compensate for unaddressed sources of uncertainty (like overestimating structural uncertainty to balance a lack located in inputs), the authors try to decipher the different sources of uncertainty explicitly and coherently with dedicated tools that are meant to prevent overlapping in their respective action.

The contribution of meteorological ensemble forecasting may appear smaller than it is in reality for two reasons:

- The superiority of MEPS over deterministic NWP systems has already been demonstrated and MEPS are recognized particularly useful to estimate uncertainties in rainfall prediction. Here, results are assessed on the whole time period and this may contribute to dilute the importance of rainy days. We attached to this answer a complementary plot to Figure 8. The same comparison between system G and H is carried out but with 5 sub-periods of assessment (spring, summer, fall, winter, and days for which streamflow is higher than the yearly median streamflow). One can notice that the benefits of ensemble forcing are clearer during time of the year where streamflows are the higher (spring, $Q > Q_{50}$).

- The meteorological ensemble was not post-processed and this may contribute to underestimate its value. It is likely that HEPS performance could have been enhanced with improved MEPS that benefited from a suitable processing. Yet, we also would like to emphasize that the contribution of meteorological ensemble forcing is not only about a subtitle gain in the MCRPS metric but it provides a substantial improvement in reliability for longer lead times (page 7196, line 6-9). Nevertheless, we agree that removing the bias and correcting the dispersion of the meteorological ensemble could improve the reliability further. Thus MEPS forcing is a piece that cannot be overlooked for medium range forecasting.

2. Specification of individual model:

- **Line 7-8 at page 7186: It is not clear whether or not spatial discretization is considered to construct catchment applications by 20 conceptual lumped models. There are lots of ways and examples to apply lumped hydrologic models considering spatial heterogeneity. Please clarify this sentence and relevant comments below:**
- **If spatial discretization 'IS' considered related to 1st comment, please clarify what spatial resolution was used. Additionally, how spatial heterogeneity was resolved in parameterization using lumped models?**
- **If spatial discretization is 'NOT' considered, please clarify how large catchments (>10,000 square kilometers) were conceptualized and parameterized.**
- **Regardless of spatial discretization, please clarify which flow routing methods were used in each model.**

To clarify line 7-8: The 20 models were derived from pre-existing models found in the literature. The models, in their original form, are either lumped (GR4J, GARDENIA, HBV, MOHYSE,...) or use a spatial discretization of the watershed (CEQUEAU, TOPMODEL, SACRAMENTO,...). For the models that are initially semi-distributed, they have been converted into lumped models. This has been done in order to facilitate their integration in a common framework (which is the multimodel framework that is used in this study) and for computational requirement. This is also why we emphasize line 4-5 that they are not the original models but only derived from the original models.

The 20 conceptual lumped models are applied in a traditional way. No spatial discretization has been done and hydrological processes are computed at the catchment scale. Consequently, the parameterization is uniform over the entire catchment, even for the largest one that spans over 15000 km^2.

We recognize that model spatial discretization can be useful and/or necessary for many purposes in hydrological sciences. However, conceptual lumped models are still very competitive, especially in the case where they are combined. We recently published a paper where we compare the multimodel hydrological forecast performance with a semi-distributed and physically based model. Results shown clearly that in our case, the multimodel was superior (Assessment of a multimodel ensemble against an operational hydrological forecasting system. A. Thiboult , F. Anctil Canadian Water Resources Journal / Revue canadienne des ressources hydriques Vol. 40, Iss. 3, 2015).

We dedicated a particular attention to the structural diversity of the lumped conceptual models, including the diversity of representation of flow routing as it is a driving process in hydrology. This ensures, or at least maximizes the chance to encompass the most effective way to achieve routing for a catchment by providing an ensemble of likely descriptions of the process. In depth description of the routing of each model may be too long for a discussion but these information can be found in G. Seiller's Ph.D. thesis annex (Évaluation de la sensibilité des projections hydrologiques au choix des outils hydro-météorologiques globaux conceptuels, http://theses.ulaval.ca/archimede/) from page 288 to 312 (in English).

3. Meteorological ensemble

- **In page 7185, rainfall forecasts seemed to be aggregated in space (one point per catchment) and time (daily). Please clarify possible impacts of excessive aggregation of rainfall forecasts on study results.**
- **For evaluating contribution of different uncertain sources on forecasts, it**

is essential to check bias of input forcing forecasts for varying lead times, while the manuscript only showed MCRPS. Please clarify the detailed analysis on it.

- **In Line 8-9 Page 7186, please clarify the sentence such as “modifications include their spatial discretization if they were initially distributed and their evapotranspiration formulation”.**

It is expected that excessive spatial aggregation deteriorate the results with the possibility to miss local but driving events. However, in the article, several meteorological grid points are systematically situated within the catchment boundaries. Consequently, it is unlikely that the contribution of local event to the streamflow out the outlet is neglected, in the case that particular meteorological event was predicted by the MEPS. Concerning temporal aggregation, it is possible that there is some loss of information, especially for small catchments, but the hydrological models used in this study are designed to work with such time step. We are currently working on the conversion from a daily to a hourly time step multimodel framework.

MEPS performance are mentioned (indeed in terms of CRPS) but their performance respect to meteorological observations are not detailed in the paper. Other scores have been evaluated (NSE, RMSE, MAE, Normalized Root-mean-square error Ratio) and are in agreement with the CRPS values. The choice of the CRPS metric relies on the fact that it is a popular score that is strictly proper. Moreover, it has the advantage that it can be reduced to the MAE and thus allows a straightforward comparison of deterministic (system A) and probabilistic forecasts (system B to H’). Also, to prevent too inaccurate forcing, a selection in the catchments has been carried out. As explained line 15-18 of page 7185, the meteorological forcing quality was too poor for 18 of the 38 catchments. Thus they have been withdrawn from the catchment pool. It is likely that this could have been enhanced with meteorological post-processing but this was deemed out of scope of this study.

Line 8-9. This belongs to the modification of hydrological model part. For the same reason as previously mentioned, models are not the original models but were derived from these models. Only the hydrological component of the model is kept. For the models that originally included a module to compute PET or snowmelt, this module has been omitted and replaced by a common one, CemaNeige for snow, and the formulation proposed by Oudin for the PET.

4. Information and analysis on catchments

- **In Section 2.1, information on catchments is limited. A new table showing information of each catchment such as catchment size, river length, low and high flow, typical time of concentration of flood, and etc, is required.**
- **Please clarify whether there are critical human intervention facilities such as dam reservoir, water gate, or weir in catchments. If there are, please clarify how such intervention was considered or affected in model configuration or results.**
- **Analysis on catchments in Results (e.g. Fig. 3, 5, 8, 10, 11, and 12) should be revised with additional analysis or figures regarding catchment characteristics such as catchment size or human intervention (e.g. Rakovec et al. 2015). References: Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schafer, D., Schron, M., Samaniego, L. (2015): Multiscale and multivariate evaluation of water fluxes and states over European river basins, J. Hydrometeorol., in press,doi:10.1175/JHM-D-15-0054.1.**

The required table that describes some catchment characteristics is provided. A criterion in the selection of the catchments was the absence or the very deem influence of human intervention on streamflows. Considerable efforts have been

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

paid to link estimated time of concentration, size of catchments and other variables without any clear results. We were not able to relate any catchment feature with particular results. This is a reason why catchments are presented anonymously in the paper.

5. DA

- **Please clarify how observation uncertainty was considered in EnKF. In conventional EnKF, noise for observation is commonly added to each ensemble which may lead to increase additional uncertainty. Otherwise, square root formulation can be used to avoid instability coming from observation noise.**
- **In Section 3.5, please clarify how EnKF perturbation was optimized in details in the case of H. Please remind that authors already mentioned that "the optimal setting may use unrealistically high perturbations that compensate partially for the structural error".**
- **In Conclusion, authors described quick decrease of reliability is found in EnKF. However, it might be accelerated by coarse spatial and temporal resolution of models and input. Please clarify this issue.**

The formulation that was used is indeed the conventional EnKF. The additional noise to each member of the ensemble is not meant to "increase" uncertainty but rather to take it into account, in particular the uncertainty related to the catchment state. For the system H', streamflow observations are perturbed with random sampling from a normal law with a standard deviation equal to 10% of the observed streamflow and 0 mean. Added perturbations for precipitation are sampled from a gamma law with a standard deviation of 25% of the initial precipitation and finally, temperatures with a normal law with a 2°C standard deviation, still with 0 mean (page 7198). For the system H, details on the perturbations added to streamflow, temperature and precipitation, but also about coupling between individual model and the EnKF can be found in a recently published article that was still under review at the time of initial submission of this article (On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments, Journal of Hydrology, Volume 529, Part 3, October 2015, Pages 1147-1160 A. Thiboult, F. Anctil).

We do not think that the decrease of reliability is due to spatial or temporal resolution but rather to the resilience of the models. Same results were found with a semi-distributed models with a 3h time step (Abaza, M.; Anctil, F.; Fortin, V. & Turcotte, R. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. Journal of Hydrology, 2014, 519, 2692-2706). In the aforementioned article, the EnKF spread decreases quickly leading to unreliable forecast from day 2 but has been compensated by direct perturbations of states variable and meteorological ensemble forcing.

6. Figures and analysis

- **Model diagnostic metrics were drawn by aggregating results of all simulation periods. Additional analysis and description on conditional statistics of different flow regimes and seasons are highly recommended.**
- **Similar figures on reliability and catchment comparison are suggested to be removed or merged together.**

We suggest to add the figure that assesses the MCRPS according to the time of the year to the supplementary material of the article.

We would prefer to keep them separate. Even if the same metrics are presented, they represent the evolution and complexification of the systems and follow the article

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

progression.

Minor comments:

1. Please use a consistent term between catchment and watershed throughout the manuscript.

This will be corrected.

2. In Fig. 7, the legend of a dotted line is not shown.

We are not sure to understand the remark. The dotted line corresponds to the RMSE of the ensemble as stated in the legend.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 7179, 2015.

River name	Area (km2)	River length (km)	Average slope (%)	Mean ann. Q (m3/s)	Coeff. of variation	Mean ann. P (mm)	Mean ann. snow (cm)
Trois Pistoles	923	52	0.52	18	1.81	1109	382
Du Loup	512	45	0.78	10	1.47	1050	378
Gatineau	6796	190	0.12	127	1.08	1023	332
Dumoine	3743	145	0.13	50	0.81	968	297
Kinojévis	2572	83	0.12	39	1.12	921	324
Matawin	1383	68	0.29	24	1.11	1025	328
Croche	1551	102	0.33	29	1.24	996	360
Vermillon	2650	145	0.20	39	1.10	957	312
Batiscan	4483	167	0.45	96	1.03	1162	381
Saint-Anne	1539	84	0.81	51	1.20	1412	502
Bras du Nord	643	77	0.82	19	1.21	1385	499
Du loup	767	57	0.78	12	1.27	1020	332
Aux Ecorces	1107	54	1.04	28	1.09	1236	450
Métabetchouane	2202	155	0.43	48	1.19	1168	420
Péribonka	1010	101	0.50	19	1.16	1000	376
Ashuapmushuan	15342	342	0.16	300	0.92	984	379
Ashuapmushuan	11200	232	0.12	227	0.88	1001	394
Au Saumon	586	69	0.65	8	1.36	877	334
Mistassini	9534	278	0.20	200	1.08	1004	409
Valin	761	59	1.06	24	1.13	1123	453

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

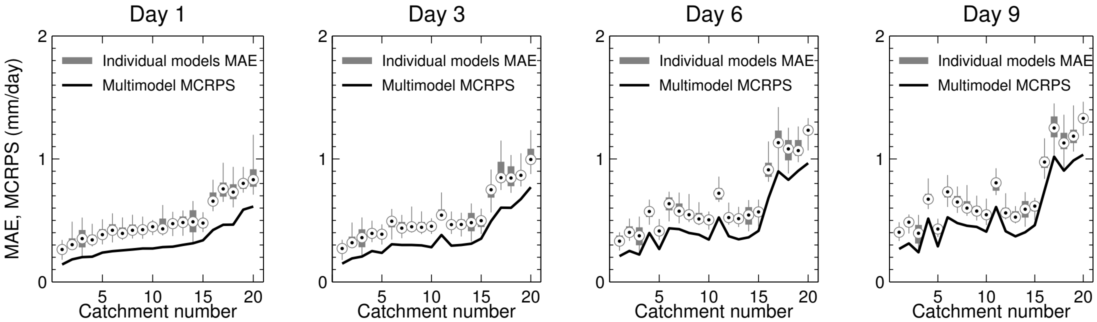

Fig. 1. Comparison of individual models MAE and multimodel MCRPS sorted by increasing multimodel MCRPS for the first day (version A vs. E).

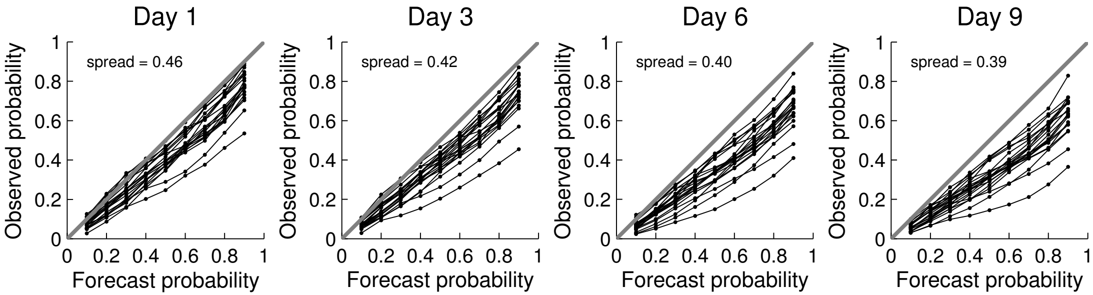

Fig. 2. Reliability of the multimodel ensemble (system E) for all individual catchments. The spread represents the square root of mean ensemble variance averaged over all catchments.

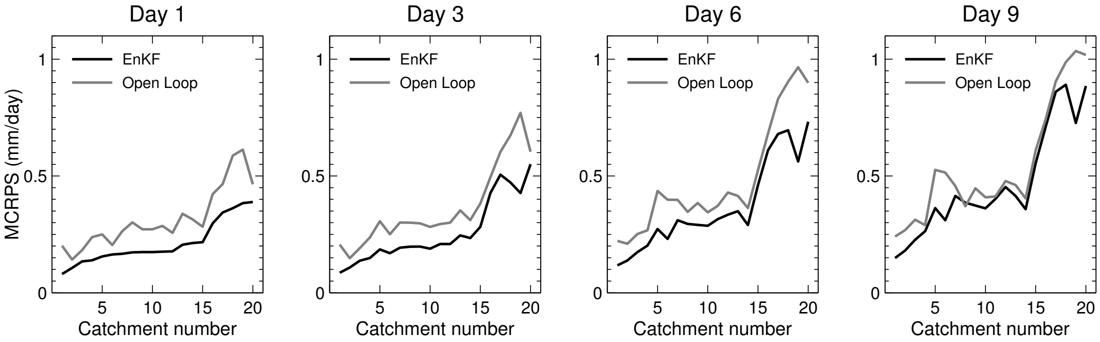

Fig. 3. Comparison of open loop and EnKF multimodel MCRPS sorted by increasing EnKF MCRPS (system E vs. G).

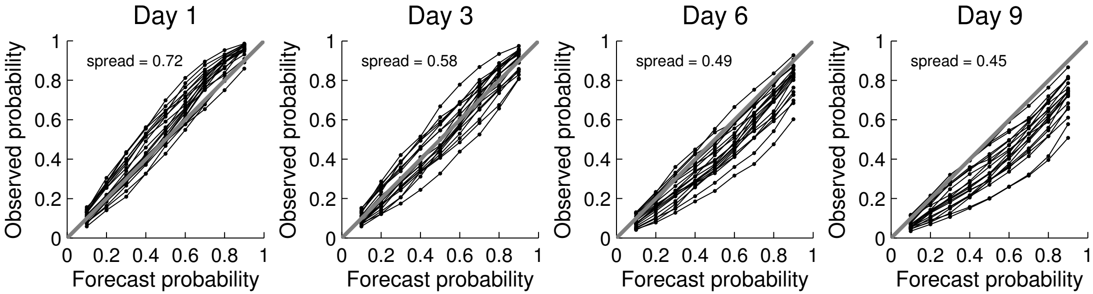

Fig. 4. Reliability of the EnKF multimodel ensemble (system G) for all individual catchments. The spread represents the square root of mean ensemble variance averaged over all catchments.

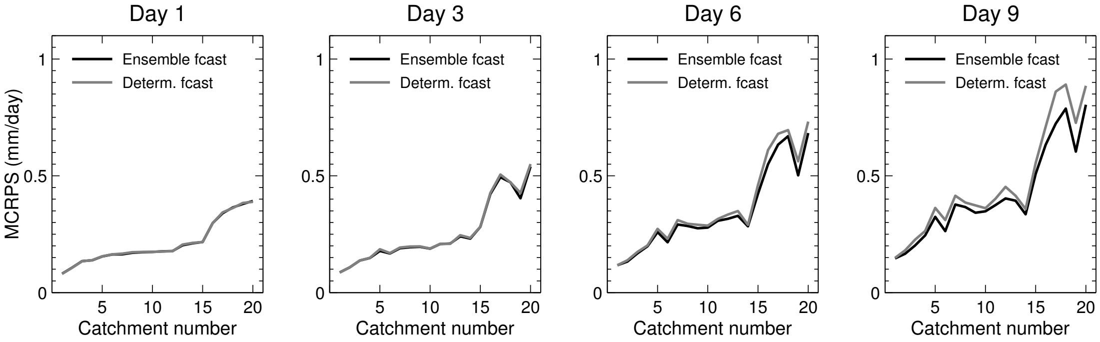

Fig. 5. Comparison of EnKF multimodel MCRPS with deterministic and ensemble meteorological forcing (system G vs. H).

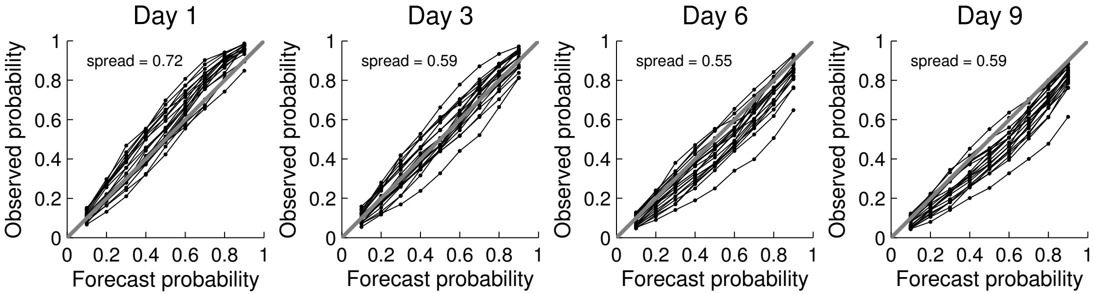

Fig. 6. Reliability of the EnKF multimodel ensemble with MEPS forcing (system H).

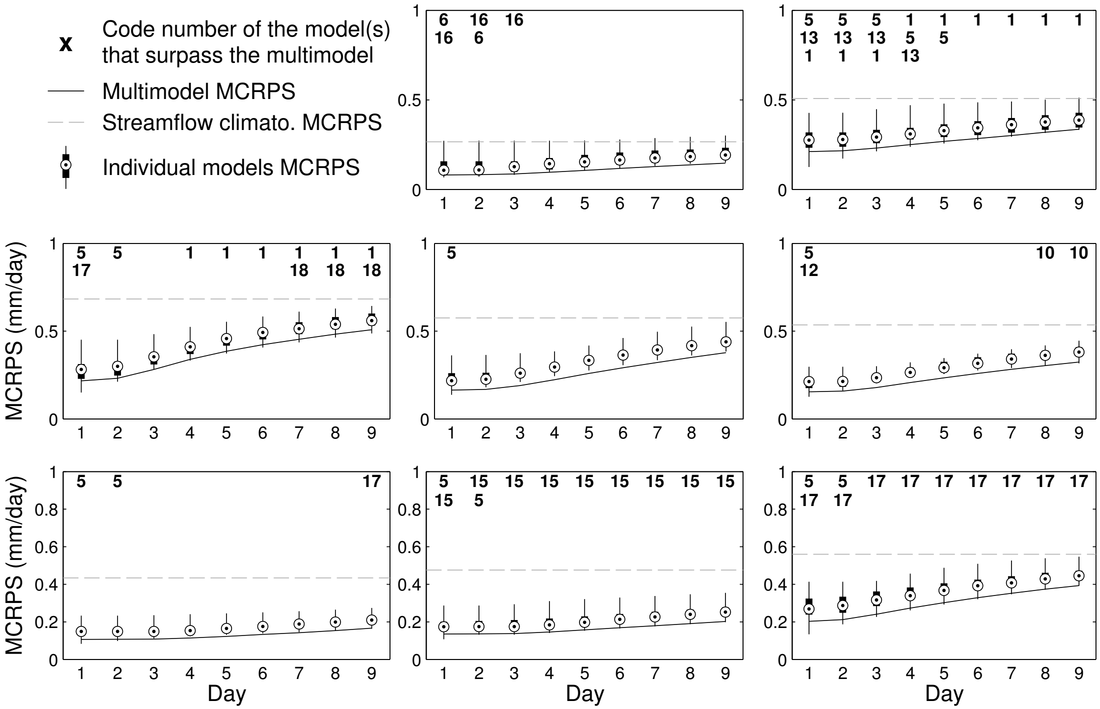

Fig. 7. Comparative examples of the MCRPS on 8 watersheds of the EnKF individual models and the EnKF multimodel, both using MEPS forcing (system D vs. H).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

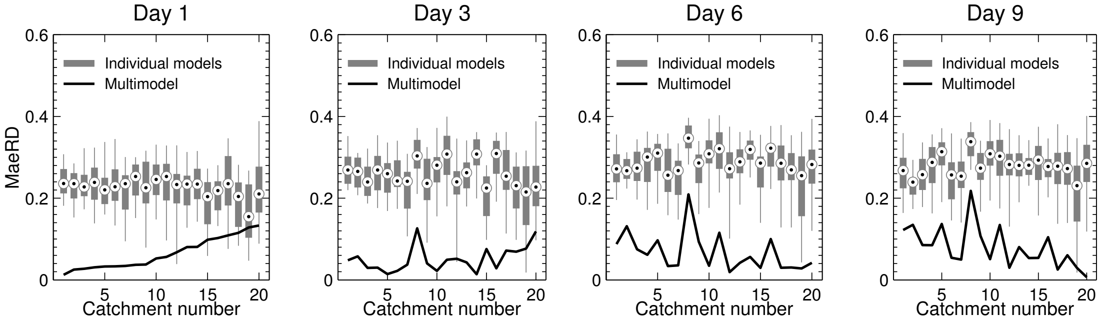

Fig. 8. Comparison of the deviation from perfect reliability of EnKF individual models and the EnKF multimodel, both using MEPS forcing sorted by increasing EnKF multimodel MaeRD for the first day (sys

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

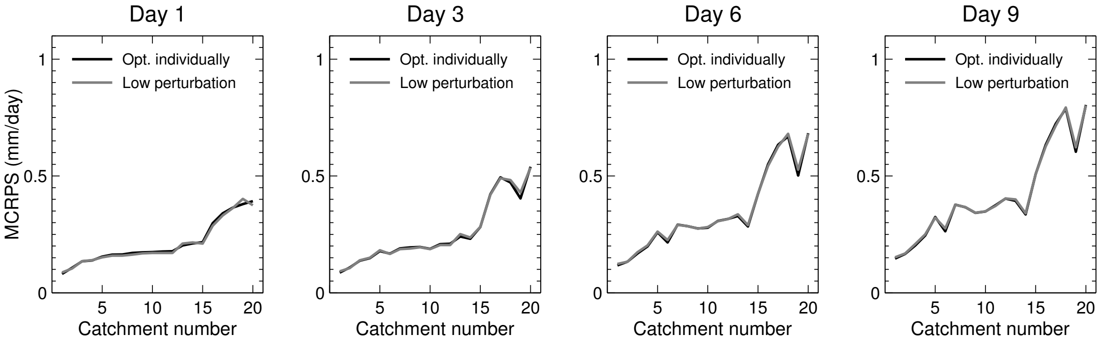

Fig. 9. Comparison of EnKF multimodel MEPS systems using either individually optimized EnKF perturbations or lower input-output perturbations (system H vs. H).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

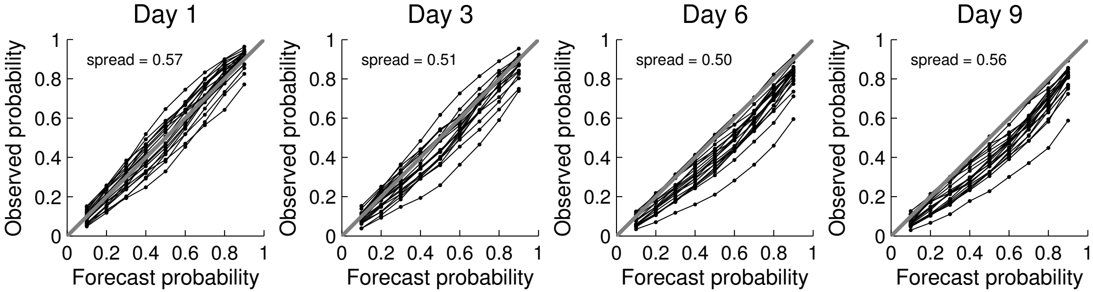

Fig. 10. Reliability of the EnKF multimodel ensemble with MEPS forcing and lower input-output perturbations (system H).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

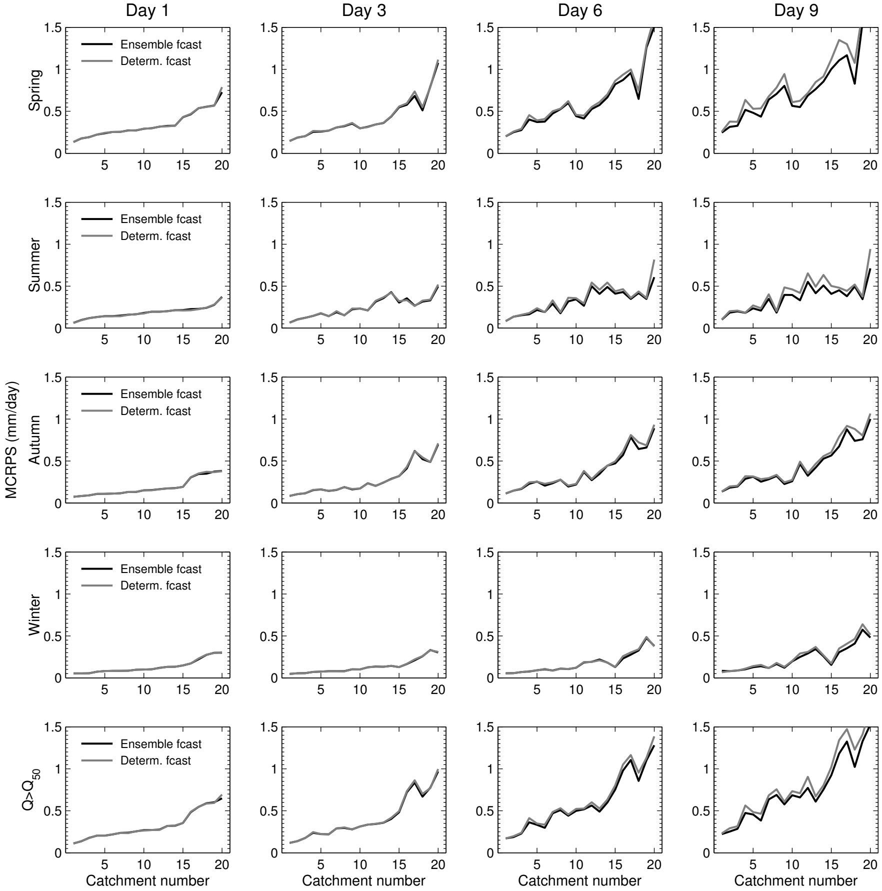

Fig. 11. Comparison of EnKF multimodel MCRPS per season with deterministic and ensemble meteorological forcing (system G vs. H).

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper