Reply to comments of reviewer #1

(reviewer comments in black, replies in blue)

This paper deals with the optimisation of monitoring networks using information-theoretical methods with a focus on the analysis of useful objective functions. The authors argue that a single-objective optimization of the joint entropy of all selected sensors will lead to a maximally informative sensor network. They compare exhaustive optimization, a greedy approach and a new ”greedy drop” approach using available monthly runoff data.

I enjoyed reading the manuscript, especially the introduction to the information theory terms. This study adds some interesting new views on the optimisation of monitoring networks and fits well to the scope of HESS. However, although the paper is mostly well written, it is sometimes poorly structured. In addition, I have some general and specific remarks that needs to be considered before publication. I therefore recommend a minor revision of the paper.

We thank Dr. Heye Bogena for his positive review and constructive suggestions, which allowed us to improve the quality of the manuscript. In the following section, reviewer comments are in black color, author responses are in blue color, and changes proposed for the text in the manuscript are in *italics* and underline font.

General comments: For this study, monthly runoff data were used from gauging stations with very different catchment areas. It can be assumed that the gauging stations with smaller catchment areas will show a higher temporal variability of discharge. Therefore, their sub-monthly data should have a higher information value than the data of the stations with larger catchments. Thus, the authors should also discuss the dependence of their results on the temporal scale of the discharge data used.

We agree with the reviewer's comment on the connection of temporal scale and having a higher information value. Increasing resolution on a temporal scale means having more data points, which in turn, may translate to having higher information value and better estimation of network dependencies. However, in our case, the low entropy value of gauging stations with smaller catchment areas is the outcome of having coarse discretization rather than the effect of temporal scale. In fact, if we use finer discretization, we will get higher entropy values for gauging stations with smaller catchment areas and with a monthly temporal scale. As we outlined in section 3, our goal in this paper was to control for the effect of temporal variability of data and quantization method in our methodology comparison. To do this, we used the same data period, with the same temporal scale and quantization method proposed by Li et al. (2012). Although our paper focuses on the role of the objective function in the network design, we think it is helpful for the readers to highlight the effect of decisions we made to isolate the impact of temporal variability of data and quantization method. Therefore, we will expand the first paragraph of the further work section to discuss these issues.

Added in Lxxx:

In this paper, we focused on the theoretical arguments for choosing the right objective functions to optimize, and compared a maximization of joint entropy to other methods, while using the same data set and quantization scheme. Another important question that needs to be addressed in future research is how to numerically calculate this objective function, or other objective functions used in other approaches. What many of these objective functions have in common, is that they rely on multi-variate probability distributions. For example, in our case study, the joint entropy is calculated from a 12-dimensional probability distribution. These probability distributions are hard to reliably estimate from limited data. *Also, these probability distributions are influenced by multiple factors, including data's temporal scale and quantization assumptions. To have an unbiased comparison framework of objective functions, we chose to isolate the effect of temporal variability of data and quantization method on methodology comparison. It is worthy to acknowledge that these assumptions as well as data availability can greatly influence optimal network ranking, and require more attention in future research.*

The order of presentation of the tables and figures is sometimes confused.

Thanks for pointing this out. The confusion is caused by the placing of table1 at the end of the manuscript. This movement is forced by the Copernicus Latex standard for the discussion manuscript format (partly because it takes a whole page). We'll make sure all tables are in a proper position in the final publication format.

Specific comments:
L1: "layout"

Thanks for catching this, we will correct it.

L11-83: Subtitles within the introductory chapter are rather uncommon.
Thank you for your comments. We now removed all subtitles and modified the introduction section, and moved down the scope subsection under the methodology section.

L123-145: Please always give the equations directly after their first citation.
Will be fixed as recommended.

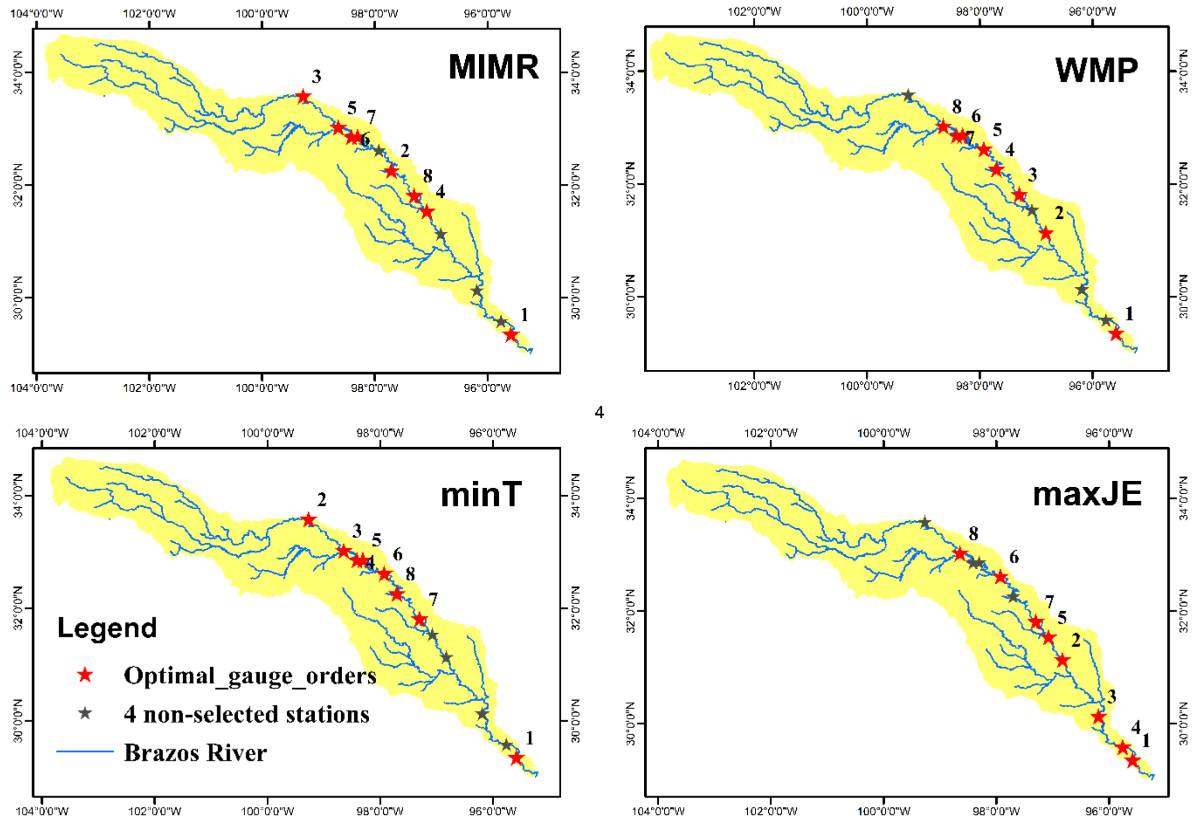
L166: You are first referring to Appendix B instead of A. Should be reversed.
Will be fixed as recommended.

L184: Which part of the Appendix?
Thanks, it is properly referenced now.

L243: Eq. 12 and the corresponding variable description should directly follow this sentence.
Will be fixed as recommended.
L259: "Tables"
Thanks for catching this, it is fixed.

L259-260: Please indicate the location of the eight stations containing the joint information of all 12 stations in Figure 3 and discuss whether the result is meaningful, e.g. in terms of the distance between stations, their respective catchment areas, etc..

Thanks for your comment. We add a new figure that contains the information of the top eight stations of all four optimization methods. Although information theory-based methods do not consider the distance between stations as a factor in network design, there are noticeable differences in terms of the distance between selected stations by different methods. In addition to the new figure, we will expand our discussion on this.



L260-265: Results shown in Figs. 5 and 6 need to be explained in more detail.

Agreed, we expanded our discussion on these two figures (after adding new figure they become fig 6 and 7) by adding the following changes:

We propose to modify as follows:

We demonstrate that other methods with a separate minimum redundancy objective lead to the selection of stations with lower new information content (green area in Figure 6). This leads to slower reduction of the remaining uncertainty that could be resolved with the full network. _Reduction of the yellow area in each iteration (i.e. the information loss compared to the full network) in Figure 6 corresponds to the growth of joint entropy values in Table 2 for each method. maxJE has the fastest, and minT the slowest rate of reduction of information loss. A method's preference for reaching minimum redundancy or growing joint information (red area in Figure 6) governs the reduction rate of information loss._ Also, Figure 7 provides auxiliary information about the evolution of pairwise information interaction between already selected stations $<X_1,X_2,...,X_{i-1}>$ in the previous iterations and new proposed station X_i . Figure 7 illustrates the contrast between the choice of the proposed stations in the first six iterations by different methods. *For instance, minT method aims to find a station that has minimum mutual information (red links in Figure 7) with already selected stations. In contrast, the maxJE method tries to grow joint entropy, which translates to finding a station that has maximum conditional entropy (green segments in Figure 7). Other methods opt to combine two approaches by either imposing a constraint (WMP) or having a trade-off between them (MIMR).*

L305-309: Repetition

Thanks for pointing this out. We now modified the paragraph to solve the issue:

Different search strategies have been adopted in the literature for monitoring network design. The most commonly used greedy algorithms impose a constraint on exhaustive search space to reduce computational effort. We investigated three different search strategies to obtain the optimal network in the context of using maxJE as an objective function. We discuss the advantages and limitations of each search strategy in terms of optimality of the solution and computational effort.

L315-319: This estimate is of limited value because it depends largely on the programming code (i.e. Matlab is very slow compared to e.g. FORTRAN).

Agreed, Matlab may not be the most time-efficient programming code. But, our main message was to highlight the exponential explosion in the number of combinations as the number of stations increases linearly. So we can see the two numbers relative to each other. We will clarify in the text the the relative differences and scaling are more important than the absolute numbers. These are language and implementation independent. :.

The computational burden is therefore greatest when about half of the stations are selected. For a number of potential sensors under 20, this is still quite tractable (4 minutes on normal PC, implemented -by a hydrologist- in MATLAB, with room for improvement by optimizing code, language, and programmer), but for larger numbers, the computation time increases very rapidly. We could make an optimistic estimate, only considering the scaling from combinatorial explosion of station sets, but not considering the dimensionality of the information measures. For 40 stations, this estimate would yield a calculation time of more than 5 years, unless a more efficient algorithm can be found. Regardless of potential improvements in implementation, the exponential scaling will cause problems for larger systems.

L323: “Tables”
Thanks for catching this, it is fixed.

L325: Table 6 deserves more explanation. The numbering of the tables is confusing. This should be rather Table 4.
Thanks for catching this, the numbering issue is fixed (it becomes Table 4). Also, we expand our discussion:

We will add:

Results in Table 4 show multiple network layouts with equal network size and joint information exist. For this case, network robustness could be an argument to prefer the network with maximum redundancy. Also, it should be acknowledged that the assumptions in data quantization would influence reaching equal joint information, and further research is warranted to investigate the network's susceptibility to quantization assumptions.

L328-329: If find Table 4 difficult to understand. In addition, the captions of Table 4 and 5 indicate 240 data points, which should be rather 860*12 data points, if I understand correctly. I suggest combining Table 4 and 5.

Thanks for catching this (it was a mistake in our first submission before the start of the online discussion). it is already corrected in the downloadable pdf file. We artificially generated 860 data points per station based on statistics from the original data (240 data points).

L339: "generated"
Will be fixed as recommended.

L344-356: This section is more like introduction and discussion and thus not appropriate for a concluding chapter. L357-363: You must present the most important results of your study more clearly, e.g. by using also bullet points.

Thanks for pointing this out. We will thoroughly revise both structure and language of this section (L344-363) to be more direct and clear. The following is the modified version:

L365-379: After removing any redundancies this section should be moved to the discussion section.
We modified Further work section by addressing reivewer's first comment on the need to discuss the dependence of our results on the temporal scale of the data. In an effort to bring the important points about objective function across, we left other issues undiscussed. We think having a separate section to give direction to future research would be helpful for our readers.