Reply to comments of reviewer #3

(reviewer comments in black, replies in blue)

This paper deals with the study of monitoring and design hydrometric networks problems. The authors used information theoretical methods to discuss the objective functions support the choice of a single-objective function to maximize the informative sensor network. The topic of the paper is very interesting and the problem of finding an optimal monitoring network is certainly a stimulating challenge. The authors performed an interesting analysis comparing exhaustive optimization and backward greedy approach using many data but probably I miss the point: why this approach lead to the optimum? And what is the optimal design?

We are glad that reviewer #3 has found the topic of our paper very interesting. We really appreciate your thoughtful comments that have guided us to improve the manuscript. In the following section, reviewer comments are in black color, author responses are in blue color, and changes made to the text in the manuscript are in *italics* and <u>underline </u>font.

In our paper, true optimum refers to optimal network ranking, which is influenced by the choice of objective function. Our goal is to advocate for choosing objective function based on theoretical argument. In this context, we presented three main arguments, which subsequently resulted in proposing maxJE method. We argued notions of minimizing dependence or redundant information by a secondary objective function between monitored signals are not desirable and have no intrinsic justification.

The greedy approach does not give the true optimum, but is computationally more tractable for larger problems. Based on suggestions of reviewer 4, we also will mention some other approached of intermediate complexity.

General comments:

1.The paper is mostly well written but there are sometimes redundant informations and figures, tables and formula are presented in an order that confuse the reader. In particular it is convenient that the explanation of the equation and the symbols involved are immediately after the equation itself. Some figures (for example Fig. 4, 5, 6) need clearer captions and a more detailed description.

Thanks for the evaluation of our paper as mostly well written. We have made significant changes to the manuscript in response to your suggestions. The changes are summarised as follows:

- We now provide explanation of the equation and the symbols directly after the first appearance.

- Regarding the confusion in table orders, it was 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.

- We modified fig 4 caption and provided more information on how the presented information is calculated.

Planned additions to figure 4 caption:

Brazos River streamflow (m3/s) statistics and resulting entropy values (bits). *The stations' IDs are organized from upstream to downstream gauges in the watershed. Entropy values are calculated by floor function and parameter a=150 m3/s.*

- We expanded our discussion on fig 5 and 6 (after adding new figure they become fig 6 and 7) by adding the following changes:

Planned edits:

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 yellow area in each iteration (i.e. remaining uncertainty in the network) in Figure 6 corresponds to growing of joint entropy values in Table 2 for each method. maxJE and minT have the fastest and the slowest rate of reduction of remaining uncertainty, respectively. Method's preference for reaching minimum redundancy or growing joint information (red area in Figure 6) governs the reduction rate of remaining uncertainty in each method.* 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, maxJE method tries to grow joint entropy, which in turn, translates to finding a station that has maximum conditional entropy (green links in Figure 7). Other methods opt to combine two approaches by either imposing a constraint (WMP) or having a trade-off between them (MIMR).*

2. The authors argue that a single-objective optimization of the joint entropy of all selected sensors will lead to a maximally informative sensor network and that the objective function indirectly minimizes redundant information: in my opinion it is not very clear why this happens. And it seems in contrast with the sentence at line 55 "Minimization of redundancy would mean that each sensor becomes more essential, and therefore the network as a whole more vulnerable to failures in delivering information".
We agree that this is not the most accurate formulation and we will clarify.

We do not intend to claim that minimizing redundancy is equivalent to maximizing joint entropy. Rather, the effects of redundancy on the amount of information captured, which we could see as some form of inefficiency, is already captured in the joint entropy measure, which quantifies total information net of any redundancies.
Because also the marginal entropy of the stations plays a role in joint entropy, there is still a trade-off, but this is not a trade-off that represents pareto-optimality in any sensible way: the only reason to consider minimizing total correlation is already captured fully in the competing objective, joint entropy.
On the other hand, putting an extra focus on minimizing redundancy by adding it as an extra objective, will focus on independence (and thus on how essential the sensors are, increasing vulnerability to failure), without caring for whether it makes a positive contribution to the total non-redundant information collected (which in our opinion is the only motivation for decreasing dependence).

We will modify our text to make it clear that it is not equivalent, but the reason to include minimum redundancy is already covered by maximizing joint entropy.

3. The greedy algorithm proposed it is not very clear for me. It is not clear why the optimum found by the algorithm is the global one instead of the local one. Also it is not clear why "remove" a station should be better than a network with a large number of sensors. Probably this is link to other costs (like installation or maintenance costs) but I missed them if specified in the paper.

The main goal of this paper is to compare and argue about choice of objective function in the information-theoretical monitoring network. We found that most of the previous approaches (listed in Table 1) used greedy optimization as a tool for their respective objective functions. Therefore, we decided to study greedy and exhaustive optimization to raise awareness in the scientific community that greedy optimization might fall into local optimum. The results show that although the greedy algorithm found a global optimum in our case study, it only found a local optimum in the added case with a artificial data set. As you correctly state, it is not generally true that greedy algorithms will find the global optimum.

The "removal" of a station is not a literal removal, it is meant to indicate ranking stations starting from the full set as opposed to starting from the empty set. We indicated that both greedy add and drop might not produce true optimum, and only exhaustive optimization will give the true optimum since it does not impose a constraint on search space.

4. All the data used in the paper should be used to compare the optimum found by the algorithm with the existent network but they not ensure the optimality.

We used two data sets (one original and one artificial) in our paper with three algorithmic variants (exhaustive, greedy add, and greedy drop). For original data, all three algorithms have produced the same result. Therefore, we only show one optimum ranking in Table 2. However, these algorithms have produced different results (Table 4) for the artificial data set. We stated in the abstract that only exhaustive optimization will give the true optimum.