## Reply to comments of reviewer #2

(reviewer comments in black, replies in blue)

We are glad that reviewer #2 was interested in reading our paper. We thank reviewer #2 for giving an example that our argument may become agreeable while raising theoretical questions about some hypothetical situations that are not discussed in our paper. Although some of the pointed questions are out of the scope of this paper, we think addressing these concerns would clarify our underlying assumptions and help our readers. In the following section, reviewer comments are in black color, author responses are in blue color, and changes planned for the text in the manuscript are in *italics* and underline font.

This paper discussed the objective functions in information theory-based hydrometric network design problems and suggested a backward greedy approach instead of other optimization methods. While I was interested in reading this paper because taking more correct, reasonable, and meaningful objective functions and proper optimization techniques is very important in the network design using information theory, unfortunately, I couldn't get the answer of the question in the title, "what is optimal?", throughout this paper.

The rhetorical question "what is optimal?" in the title can be understood in two ways. Firstly it may refer to the question "what is the optimum network configuration?" (given that we have defined what we want from the network, i.e. the objective function). Secondly it might ask the question "what do we want from the network?"; "what do we consider optimal?"
Putting this question in the title was our way of drawing the reader's attention to the core difference between these two. Also it serves to hint at the fact that choice of objective function controls optimal answer—changing the objective function would change the subsequent optimal answer, so the objective function must be justified by reasoning independent of the numerical results. We will add a brief explanation in the beginning of the introduction.

1. The authors argue that a higher amount of redundant information should be preferred because it reinforces the robustness of the network. However, this is not a general statement but can be applied only for a specific condition.

In the abstract section, we stated that " for two networks of equal joint entropy, one with a higher amount of redundant information should be preferred for reasons of robustness against failure." We did not intend to make a general statement. We made this conditional statement about two networks of equal joint entropy to point out the fact that minimizing dependence as a secondary objective would give preference to one with lower shared information. We think this is undesirable in any kind of situation. Our main point of the paper is that there is, in general, no reason for minimizing redundancy that is not already reflected in the max joint entropy objective. We emphasize this point by saying that we may even prefer maximizing instead of minimizing redundancy.

World Meteorological Organization has recommended the minimum network densities, which do not represent the optimal number of stations, rather they are suggested to avoid any serious deficiency in water resources management. That is, just meeting the guideline couldn't be sufficient while there is no clear definition of sufficient density. However, many regions in the world, such as developing countries, do not meet even the WMO guideline. Even in the study area in this paper, Brazos River Watershed, the drainage area is about 116,000 sq.km. and there are 12 USGS stations considered as the existing stations. The current network density becomes 9,667 sq.km per station while the minimum network density of the WMO guideline for the Interior Plains is 1,875 sq.km., which is more than five times sparser network. Considering the WMO's minimum network is the baseline not to lose critical information, it should be noted that the network is seriously under-gauged, such that we need to consider network expansion and network efficiency rather than network robustness. In this case, minimizing total correlation is more meaningful to optimize network efficiency. Besides, installing and maintaining monitoring stations often highly depend on financial budgets which cannot be satisfactory in practice. On the other hand, if the network density largely exceeds the minimum network density, and the water resources managers consider shrinking network by closing stations, the authors' argument may become agreeable.

Thanks for your comment. In general, network density is an important issue, and we acknowledge that this network is under-gauged. However, decision on network density is mainly derived by budget and network's purpose, and formulating objective function for any given budget is discussed in this paper. In fact, our main focus is to discuss the formulation of information-theoretical objective functions and previous literature on that topic. In monitoring network design problems, there are multiple important contributing factors, including assumptions on network density, objective function, data quantization, temporal variability of data, etc. We restricted our scope to avoid multifarious analysis and isolate the effect of all contributing factors except assumption on the objective function. Therefore, we expand the first paragraph of the study area section to clarify these contributing factors to our readers.

Some further related thoughts, that we hope to integrate clearly in the text:

- We agree that network robustness it's the secondary objective that only becomes relevant when the first objective is met, and is less important in under gauged situations.
- We do not propose a multi objective approach here, but added this only to emphasize that minimizing redundancy is not a helpful objective.
- The first objective of maximizing information content per sensor, or network efficiency, is optimized by joint entropy. Adding other objectives in a multi-objective optimization will detract from that primary objective.
- The question of how to properly define network efficiency is independent from whether the current network is under gauged or over gauged or whether we are considering expanding or reducing the network.

2. The authors also argue that maximizing joint entropy already connotes minimizing total correlation; however, this is not an absolute case, even in the case study in this paper. In Figure 7, the red bar on the top right represents Pareto-front given by maximizing both joint entropy and total correlation. If maximizing joint entropy is equivalent to minimizing total correlation, there should be only one optimal solution rather than Pareto-fronts, and its total correlation should be minimum. It seems like there are three (authors') optimal solutions in Figure 7, and it represents the solution which has the maximum joint entropy does not have maximum total correlation, this conflicts with the authors' argument.

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.
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.

As mentioned in caption in figure 7, past approaches, in the literature, gave importance to the maximum joint entropy and minimum total correlation. We make a conditional statement that if a trade-off between objectives is to be considered, the red Pareto-front should be preferred instead of the black line. Also, before making this conditional statement, we made a general statement that tradeoff is irrelevant in our paper and information can be maximized with the horizontal direction only. Therefore, the conditional statement is not in conflict with our main argument since our method does not consider a tradeoff and will only produce one optimal solution. We modified the figure's caption to clarify this issue.

Added in figure's caption:

The resulting total correlation and joint entropy for all 924 possible combinations of 6 out of 12 sensor locations. In some past approaches, a pareto front in the lower right corner is given importance. In the paper, we argue that this trade-off is irrelevant, and information can be maximized with the horizontal direction only. If a trade-off with reliability *needs to be* considered, the Pareto front of interest is in the top-right corner *instead of the lower right corner that is previously recommended in the literature.*

3. The original objective function of the MIMR method by Li et al. (2012) is not the one in Function (9) in the paper. To convert the multiobjective problem into a singleobjective problem, Li et al. applied the information-redundancy tradeoff weights and maximized the single objective function. In this case, proper weights should be predefined because the optimal solutions can differ due to the weights. If the weight, $\lambda 1$, is equal to one, the problem will become the same with the maxJE what the authors are proposing. To apply the objective functions in Function (9), multiobjective optimization technique should be employed and it will of course yield multiple optimal solutions on Pareto-fronts. In this case, which optimal solution was selected and discussed, such as in Tables 2 and 3, and why the optimal solution was selected should be addressed.

Our responses to this comment are three-fold:

1) Function (9) in our paper is the same as Equation 14 in Li et al. (2012). We directed our reader to Table1 for more details where tradeoff weights are explained.

2) Li et al. (2012) stated that "*the sensitivity analysis was only carried out for information weight falling between 0.5 and 1.0. Results are summarized in Table 4, which mainly signify the stability of MIMR with respect to information weight.".* Therefore, we chose to accept Li et al. (2012)'s conclusion on insignificant effect of information-redundancy tradeoff weights in this particular case study. MIMR's information presented in our Tables 2 and 3 can be verified by the information presented in Li et al. (2012)'s Table 3.

3) We disagree with the reviewer's statement: "*If the weight, λ1, is equal to one, the problem will become the same with the maxJE what the authors are proposing.*". Li et al. (2012) proposed λ1 as trade-ff weight for both maximum joint entropy (H) and maximum transinformation (T). Therefore, if λ1 is equal to one, MIMR method would have two terms (H and T) in its objective function while maxJE method has only one term (H). In any circumstance, MIMR will not become the same as the maxJE method.

4. Calculating streamflow information for network design from the monthly time series is quite skeptical. Is a hydrometric network which was numerically designed by monthly time series also good for short-term analysis, such as flood forecasting?

Hydrometric network design with the purpose of flood forecasting not only requires data set with high temporal resolution, it may also need a combination of rain gauges and streamflow gauges. As we outlined in section 3, our goal was to isolate the effect of temporal variability of data and quantization method from the methodology comparison; we used the same data period with the same temporal scale and quantization method proposed by Li et al. (2012). Although we intended to only focus 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 address these issues.

5. The authors finally suggested the greedy optimization algorithm. However, the greedy algorithm is not guaranteed to find the global optimum solution and is easy to fall into a local optimum, even though global optimum can be found in the case study of this paper. Also, in the reviewer's opinion, taking 20 years of monthly time series at 12 stations could be not enough to make a general conclusion. Why do we need an optimization technique if we can calculate the objective values of all populations?

We agree with the reviewer's comment on greedy algorithms. In section 4.3, we reiterated that only exhaustive optimization will give the true optimum, which is global optimum solution. We stated (Line 319) most of the previous approaches listed in Table 1 can be categorized as greedy optimizations. The comparison between greedy and exhaustive optimization has two benefits: (1) it raises awareness in the scientific community that greedy optimization might fall into local optimum; (2) it shows why exhaustive optimization is not feasible in higher dimensions.

Our response to reviewer's opinions is as follows:

Regarding the first opinion "taking 20 years of monthly time series at 12 stations could be not enough to make a general conclusion.", we explained our reasons for choosing this data set in section 3 as well as our answer to comment #1. It is important to emphasize that unbiased comparison between objective functions is the goal of our paper. Therefore, we decided to use a data set that was subject of study for the MIMR method (a highly cited paper in our field). We agree that data scarcity is an issue in drawing conclusions on the case study, be the objective of the case study is just for illustration of the discussion on objective functions.

Regarding the second opinion: we agree we can just use exhaustive optimization here, where all objective function values of the search space are calculated. However, due to the exponential growth in search space, this is practically impossible in other cases. That is why we discuss other, more tractable approaches.