Reply to comments of reviewer #4

(reviewer comments in black, replies in blue)

Dear Authors, I've read your manuscript with enthusiasm, as this topic is of my great interest. I do agree with the authors that there is no consensus in the design of monitoring networks, but I do not agree with the conclusions found in this paper. In general, I find that the scoping of the problem validate the obtained results, and therefore, beats its purpose. This is further detailed in the comments below. In addition, I believe that critical methods and literature are overlooked, especially in the use of metaheuristics for the design of sensor networks. Also, I believe that the document could be better structured, as sometimes results and methodology sections overlap.

We appreciate your thoughtful comments and the amount of time spent to provide such constructive suggestions and discussions. In the following section, reviewer comments are in black color, author responses are in red color, and changes planned to be made to the text in the manuscript are in *italics* and <u>underline</u> font.

We agree with you on many aspects of monitoring network design, though our different opinions on our theoretical arguments have led us to different conclusions. In our opinion, the results from (we now clarified that) a case study can only serve as illustration and not for normative arguments for the use of a particular objective function. In information-theoretical network design, performance metric usually is optimization by one of the objective functions. Therefore, we believe choice of the objective functions must be justified by theoretical argument otherwise, evaluation would become circular. We believe you agree with that, when you say the scoping of the problem validates the results.

We therefore emphasized in several places in the paper that it is not the case study results that lead us to our conclusions (they are merely for illustration), but it is the arguments for scoping the problem and interpretation of the information measures that lead to our conclusions.

We limited the scope of our paper to a substantial body of literature employing several information measures as objective functions. While we agree that several different objective functions may be justified to reflect the users and their decision problems in a specific case study, we believe these should then be explicitly derived from those decision problems. In the case of government-funded general purpose monitoring networks, these explicit formulations are often not possible. It then makes sense to maximize the collected information by the network, without judgement on what that information is used for.

This maximizing the total collected information is apparently also the underlying objective of the several papers dealing with information-based monitoring network design that we discuss here. We assume this, because in those papers, there is typically no mention of a decision problem to motivate the objective functions.

The main point of our paper is that, while the idea of minimizing redundant information for monitoring networks intuitively makes sense to promote efficiency, we believe the only reason to justify this is in fact the underlying desire to maximize the non-redundant information the network collects. Redundancy in itself does not hurt, it only hurts through the information loss it causes, by decreasing the amount of non-redundant information collected.

This amount of non-redundant information is precisely what is measured by the joint entropy. In other words, the only reason to include some objective to minimize redundancy, is already covered within the single objective of maximizing joint entropy. .

While we do not claim that joint-entropy is the only objective needed in all situations, we do claim it is the only measure needed in absence of motivations that go beyond the idea of collecting maximum information.

We found many previous studies did not contain these explicit other motivations, and often language that suggests that the objective to minimize redundancy is motivated from preventing inefficient information collection.

Some examples are given in the replies to specific comments.

On the greedy search strategies and metaheuristics: We found that most of the previous studies (listed in Table 1) used greedy search to find optimal ranking. We compared greedy and exhaustive approaches to raise awareness in the scientific community that greedy optimization might fall into local optimum; and to show why exhaustive optimization is not feasible in higher dimensions.

We agree with you about merits in using metaheuristics as a search tool, and that it should be mentioned as an alternative. In this paper, we intend to keep focus on our main message (importance of choice of objective function). Indeed, comparing greedy, exhaustive and metaheuristic would be an interesting research question, and we now promote this idea in further work section, as we agree it was an oversight not to discuss these. Also, we improved the paper structure in several other places by following your suggestions and other reviewers'.

We will add:

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. *Since the majority of previous research used greedy search tool to find optimal network ranking, we compared greedy and exhaustive search approaches to raise awareness* *in the scientific community that greedy optimization might fall into a local optimum, though its application can be justified considering computation cost of exhaustive approach. Banik et al. (2017) compared computation cost for greedy and metaheuristic optimization (Non-dominated Sorting Genetic Algorithm II). They reported greedy approach resulted in drastic reduction of the computational time for the same set of objective functions (metaheuristic computation cost was higher 58 times in one trial and 476 times in another). We recommend further investigation of these three search tools in terms of both optimality and computational cost.*

Objective functions for information-theoretical monitoring network design: what is optimal?

General comments

What is the definition of optimality?

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.

In the introduction you point out that monitoring objectives are posed as part of a wider decision problem (l18-19). Later, you state that the objective in the design of a sensor network is to maximise its joint entropy (l26-27) as sensor networks may support many decisions. At this point, this becomes a normative approach to the design problem, where you define what optimallity is, and that other objectives should be secondary (l347-348). As a consequence, it is clear that the problem is not a multi-objective optimisation problem anymore, and any trade-off with other objectives (such as minimising redundancy) will directly reflect in a performance loss. This is actually seen in Table 2, where you clearly point it out.

In this paper we are not trying to argue that maximizing joint entropy is the final answer to optimization of monitoring networks. What we do argue is that choosing an objective function is a normative activity, and should therefore be justified by arguments independent of numerical results.

We bring forward such theoretical arguments to argue that if maximizing the information collected by the network is the objective, then joint entropy is the mathematical expression to maximize. Minimizing redundancy only serves as one of the means to achieve that goal. When the goal itself is the objective function it is not necessary to add the means as a secondary objective. This will only lead to a loss in performance on the main goal.

Our scope in this paper is the information theoretical approaches to monitoring network design, which do not bring in external motivations to consider other objective functions, and as such we can only assume maximizing some idea of information is the objective. No other motivations for minimum redundancy were found in the previous literature.

Defining the design problem in these terms makes it sufficiently narrow to justify the use of single-objective optimisation, but the point is that not every problem is.
We agree that not every design problem is single objective. In absence of motivations based on decision problems supported by the network optimization of collected information could be a reasonable objective. Our main goal of this paper is to clear up the misunderstanding that minimum redundancy needs to be explicitly added as a separate objective to achieve an efficient network. We argue that the rationale for including minimum redundancy is missing.

This also connects with the three main arguments presented in the motivation (s 1.2). l44. "Firstly, we argue that objective functions for optimizing monitoring networks can, in principle, not be justified by case studies" - I do not agree with this postulate. The objective of monitoring is to provide information about the state of a system, to support a given action. Measuring for the sake of measuring do not serve any purpose.

l48. "However, from case studies, we cannot draw any normative conclusions as to what objective function should be preferred." - Preferences are not normative, but relative to the decision problem, objectives and context. These are particular to each case study.

We agree that the objective functions should be informed by the decision problem, objectives and context. We realize now that, to accurately convey what we mean, the sentence should be: “However, from **the results of** case studies, we cannot draw any normative conclusions as to what objective function should be preferred”. From your comments it looks like you agree this would be circular.

We do agree on the benefit of case studies to formulate case-specific objectives that are motivated by the stakeholders in the monitoring network. Eliciting these objectives would be a valuable use of case studies. However, that is not what the case studies are used for in the papers we discuss here. We also make our point about case studies to make it explicit to our readers that the case study in our paper should not be interpreted as a (circular) argument for use of joint-entropy, but merely as an illustration of the different objectives at work. The arguments for using joint entropy are given before and after the case study, but do not rely on its results.

Measuring for the sake of measuring indeed makes no sense, but if we are measuring for an unknown “set of sakes”, then maximizing information content gives maximum potential for use. Though it is possible to get information from the network without utility for decision, it is impossible to have utility from the network for decisions without it providing information.

l50. "Secondly, we argue that the joint entropy of all signals together is in principle sufficient to characterize information content and can therefore serve as single optimization objective" - I do agree with this postulate, as long as the objective is to maximise joint entropy. This goes back to the first comment.

Yes, our argument is that maximizing joint entropy is the only objective needed to maximize joint information content, if no other requirements are given about a target for prediction or a decision problem. Would you also agree with the statement if we said: “joint entropy characterizes joint information content without the need to separately account for redundancy.”? This is what we argue and disagree with much of the literature.

l53. "Thirdly, multi-objective approaches that use some quantification of dependency or redundancy as a secondary objective, next to joint entropy, could only be justified if redundancy is interpreted as beneficial for creating a robust network" - I do partially agree with this postulate, but then again is linked to the definition of optimality. Given the decision problem, a decision-maker may opt to trade some improvement in joint entropy for redundancy (as an example), and that is out of the scope of what this paper presents.

We agree, that the pareto front defined by a max JE, max Redundancy multi-objective problem could be interesting to explore. Our main point here was, however, that there is no justification for adding **Minimum Redundancy** as an objective. To emphasize this, we actually suggest that the opposite, **maximizing redundancy**, makes sense in certain contexts. Perhaps this was not the most clear way to formulate this and we will try to make this clearer.

In general, once you assume that joint entropy corresponds to the definition of optimality, the problem is self-validated. Leading to the conclusions that you are presenting such as: "Information theory provides a valid framework for monitoring network design" (l344), and that single-objective optimisation is sufficient to approach the sensor network design problem.

We do not assume that JE corresponds to the definition of optimality, but we argue that, within the scope of the 20+ papers that use minimization of dependency related information measures, JE is sufficient, since no motivation for reducing redundant information was given, except implicitly as a means to capture more information (see quotes below)..
We are aware of the self-validation problem, and this is why we emphasized at several points that our case study should not be interpreted as validation of our points, but the validation should be sought in our demonstration of the meaning of information measures and how they relate to each other.

We will make it clearer at the beginning of the conclusions that these come from interpretation of the measures, not from the case study results..

The justifications for including minimum redundancy-related measures that we found in previous literature, apart from the quote in our paper from Mishra and Coulibaly [2009], are in the following quotes:

- Alfonso et al 2010b: *"The main contribution of this paper is that joint entropy and total correlation are independent objectives that must be optimized."*. --- Note that we show they are not independent. They argue redundancy should be minimized to find an independent set of stations. They stated this argument comes from [Mishra and Coulibaly,2009], and they proposed to use total correlation instead of transinformation to achieve that goal.

- Li et al (2012): "*highest information content and **avoid dependent stations as much as possible**, guaranteeing while the stations within and outside of the optimal set has high common information*". They also argued "*The information-redundancy tradeoff weights **provide the user a flexible handle to balance the two conflicting objectives**: maximum information and minimum redundancy.*"

- Keum and Coulibaly (2017) rephrase goal of independent network: "*Therefore, the amount of duplicated or sharable information in a network explains the redundancy or **ineffectiveness of the network**.*". They also comment on MIMR: "*On the other hand, the MIMR reformulate the multiobjective problem to a single objective optimization by merging the different criteria into one objective using weighting factors [e.g., Li et al., 2012; Fahle et al., 2015]. However, the weight for each objective should be assumed in advance. In this study**, the former approach is applied not to make any prior assumptions** but to compare various optimal networks in decision processes*.".

- Banik et al. (2017) stated TC as single objective should not be used: "*Minimizing this objective means reducing the correlated information. The objective of the problem being to maximize the information furnished by the sensors, the **TC function is considered always in combination with JH**. In fact, TC as a single objective furnishes solutions with less-correlated sensors, for example, terminal nodes, **with a poor content of information***."

Note that all these argue that stations need to be independent, but do not give a reason why, except for effectivenes, which we argue is covered by looking at the total non-redundant information the the network delivers.

Below are some of the 20+ papers that we referred to earlier in this reply:

(1) Wang, W.; Wang, D.; Singh, V. P.; Wang, Y.; Wu, J.; Zhang, J.; Liu, J.; Zou, Y.; He, R. Information Theory-Based Multi-Objective Design of Rainfall Network for Streamflow Simulation. *Advances in Water Resources* **2020**, *135*, 103476. https://doi.org/10.1016/j.advwatres.2019.103476.

(2) Li, H.; Wang, D.; Singh, V. P.; Wang, Y.; Wu, J.; Wu, J.; He, R.; Zou, Y.; Liu, J.; Zhang, J. Developing a Dual Entropy-Transinformation Criterion for Hydrometric Network Optimization Based on Information Theory and Copulas. *Environmental Research* **2020**, *180*, 108813. https://doi.org/10.1016/j.envres.2019.108813.

(3) Werstuck, C.; Coulibaly, P. Assessing Spatial Scale Effects on Hydrometric Network Design Using Entropy and Multi-Objective Methods. *JAWRA Journal of the American Water Resources Association* **2018**, *54* (1), 275–286. https://doi.org/10.1111/1752-1688.12611.

(4) Wang, W.; Wang, D.; Singh, V. P.; Wang, Y.; Wu, J.; Wang, L.; Zou, X.; Liu, J.; Zou, Y.; He, R. Optimization of Rainfall Networks Using Information Entropy and Temporal Variability Analysis. *Journal of Hydrology* **2018**, *559*, 136–155. https://doi.org/10.1016/j.jhydrol.2018.02.010.

(5) Banik, B. K.; Alfonso, L.; Di Cristo, C.; Mynett, A. Evaluation of Different Formulations to Optimally Locate Sensors in Sewer Systems. *Journal of Water Resources Planning and Management* **2017**, *143* (7), 04017026. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000778.

(6) Alfonso, L.; Lobbrecht, A.; Price, R. Optimization of Water Level Monitoring Network in Polder Systems Using Information Theory. *Water Resources Research* **2010**, *46* (12). https://doi.org/10.1029/2009WR008953.

(7) Li, C.; Singh, V. P.; Mishra, A. K. Entropy Theory-Based Criterion for Hydrometric Network Evaluation and Design: Maximum Information Minimum Redundancy. *Water Resources Research* **2012**, *48* (5). https://doi.org/10.1029/2011WR011251.

(8) Fahle, M.; Hohenbrink, T. L.; Dietrich, O.; Lischeid, G. Temporal Variability of the Optimal Monitoring Setup Assessed Using Information Theory. *Water Resources Research* **2015**, *51* (9), 7723–7743. https://doi.org/10.1002/2015WR017137.

(9) Samuel, J.; Coulibaly, P.; Kollat, J. CRDEMO: Combined Regionalization and Dual Entropy-Multiobjective Optimization for Hydrometric Network Design. *Water Resources Research* **2013**, *49* (12), 8070–8089. https://doi.org/10.1002/2013WR014058.

(10) Stosic, T.; Stosic, B.; Singh, V. P. Optimizing Streamflow Monitoring Networks Using Joint Permutation Entropy. *Journal of Hydrology* **2017**, *552*, 306–312. https://doi.org/10.1016/j.jhydrol.2017.07.003.

(11) Keum, J.; Coulibaly, P. Information Theory-Based Decision Support System for Integrated Design of Multivariable Hydrometric Networks. *Water Resources Research* **2017**, *53* (7), 6239–6259. https://doi.org/10.1002/2016WR019981.

(12) Keum, J.; Coulibaly, P.; Razavi, T.; Tapsoba, D.; Gobena, A.; Weber, F.; Pietroniro, A. Application of SNODAS and Hydrologic Models to Enhance Entropy-Based Snow Monitoring Network Design. *Journal of Hydrology* **2018**, *561*, 688–701. https://doi.org/10.1016/j.jhydrol.2018.04.037.

(13) Huang, Y.; Zhao, H.; Jiang, Y.; Lu, X. A Method for the Optimized Design of a Rain Gauge Network Combined with Satellite Remote Sensing Data. *Remote Sensing* **2020**, *12* (1), 194. https://doi.org/10.3390/rs12010194.

(14) Banik, B. K.; Alfonso, L.; Torres, A. S.; Mynett, A.; Di Cristo, C.; Leopardi, A. Optimal Placement of Water Quality Monitoring Stations in Sewer Systems: An Information Theory Approach. *Procedia Engineering* **2015**, *119*, 1308–1317. https://doi.org/10.1016/j.proeng.2015.08.956.

(15) Banik, B. K.; Alfonso, L.; Di Cristo, C.; Leopardi, A. Greedy Algorithms for Sensor Location in Sewer Systems. *Water* **2017**, *9* (11), 856. https://doi.org/10.3390/w9110856.

(16) Keum Jongho; Coulibaly Paulin. Sensitivity of Entropy Method to Time Series Length in Hydrometric Network Design. *Journal of Hydrologic Engineering* **2017**, *22* (7), 04017009. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001508.

(17) Leach, J. M.; Coulibaly, P.; Guo, Y. Entropy Based Groundwater Monitoring Network Design Considering Spatial Distribution of Annual Recharge. *Advances in Water Resources* **2016**, *96*, 108–119. https://doi.org/10.1016/j.advwatres.2016.07.006.

(18) Leach, J. M.; Kornelsen, K. C.; Samuel, J.; Coulibaly, P. Hydrometric Network Design Using Streamflow Signatures and Indicators of Hydrologic Alteration. *Journal of Hydrology* **2015**, *529*, 1350–1359. https://doi.org/10.1016/j.jhydrol.2015.08.048.

(19) Maymandi, N.; Kerachian, R.; Nikoo, M. R. Optimal Spatio-Temporal Design of Water Quality Monitoring Networks for Reservoirs: Application of the Concept of Value of Information. *Journal of Hydrology* **2018**, *558*, 328–340. https://doi.org/10.1016/j.jhydrol.2018.01.011.

(20) Pádua, L. H. R. de; Nascimento, N. de O.; Silva, F. E. O. e; Alfonso, L.; Pádua, L. H. R. de; Nascimento, N. de O.; Silva, F. E. O. e; Alfonso, L. Analysis of the Fluviometric Network of Rio Das Velhas Using Entropy. *RBRH* **2019**, *24*. https://doi.org/10.1590/2318-0331.241920180188.

(21) Vivekanandan, N. Evaluation of Stream Flow Network Using Entropy Measures of Normal and Lognormal Distributions. *Bonfring International Journal of Industrial Engineering and Management Science* **2012**, *2* (issue 3), 33–37. https://doi.org/10.9756/BIJIEMS.10040.

(22) Vivekanandan, N.; Jagtap, R. S. Evaluation and Selection of Rain Gauge Network Using Entropy. *J. Inst. Eng. India Ser. A* **2012**, *93* (4), 223–232. https://doi.org/10.1007/s40030-013-0032-0.

(23) Wang, W.; Wang, D.; Singh, V. P.; Wang, Y. Spatial-Temporal Evaluation of Rain-Fauge Network Based on Entropy Theory. In *EPiC Series in Engineering*; EasyChair, 2018; Vol. 3, pp 2293–2300. https://doi.org/10.29007/1kc9.

(24) Xu, P.; Wang, D.; Singh, V. P.; Wang, Y.; Wu, J.; Wang, L.; Zou, X.; Chen, Y.; Chen, X.; Liu, J.; Zou, Y.; He, R. A Two-Phase Copula Entropy-Based Multiobjective Optimization Approach to Hydrometeorological Gauge Network Design. *Journal of Hydrology* **2017**, *555*, 228–241. https://doi.org/10.1016/j.jhydrol.2017.09.046.

(25) Werstuck, C.; Coulibaly, P. Hydrometric Network Design Using Dual Entropy Multi-Objective Optimization in the Ottawa River Basin. *Hydrology Research* **2017**, *48* (6), 1639–1651. https://doi.org/10.2166/nh.2016.344.

What are the alternatives for solving the optimisation problem?

From the methodological point of view, I see that you propose the greedy drop algorithm, as an alternative to greedy selection, especially attractive in the cases where exhaustive search is not feasible (l216-223). However, you omitted mentioning metaheuristics to approach this issue. These are often used in the design of monitoring networks when the combinatorial problems are too large. This being the argument for greedy approaches.
We fully agree with the reviewer on this one. We now mention metaheuristic approaches, and put in a few words about the complexity of the search space. Metaheuristic approaches indeed fall somewhere in between greedy approaches and exhaustive search. It would be interesting to explore the tradeoff between computational complexity and optimality. We will mention this where greedy approaches are discussed and in the future work section.

What are the cases of sensor network design that are considered?

When designing a monitoring network, you may have one of 3 possible scenarios: augmentation, reduction or relocation. Augmentation makes for the case when additional sensors are to be placed in the network. Reduction, accounts for the case when sensors would like to be removed from the network. Relocation, deals with the issue of changing the placement of the sensors. The case you illustrate in this paper corresponds to reduction, as the objective is to select a sub-set of sensors from a larger pool. However, you use justifications from the augmentation case (l221-222) to justify the use of a greedy algorithm. I think it is necessary to better define the optimisation problem in this respect.

Yes, we agree on those 3 scenarios. Note that, in this paper we show that to maintain optimality, reduction or augmentation may both need to go hand in hand with relocation. This is why no greedy approach to step by step expansion or reduction of sensors can be guaranteed to be optimal. We now added the case of reduction as well to line 221-222, since it also applies to that scenario.

About the definition of the optimization problem: If we keep the focus on information theoretical objective functions, which is the scope of this paper, then we will need data for the time series at sensor locations to calculate the information measures, whether they are candidates for removal or potential new sites.

The choice of interest in either augmentation, reduction or relocation case mainly influences the way we get the data (from measured time series vs. modeled ones), and the constraints on which locations we consider as candidates. Even in case of network expansion/augmentation,, stations will need to be selected from a pool modeled locations. In rain gauge network design, for example, gridded radar or satellite data can be assumed as hypothetical sensors to solve augmentation or relocation of existing network. For example, Yeh et al (2017) used simple entropy-based objective function to solve network augmentation. In terms of requirements for the objective function, there is no difference between these problems.

We do not argue that this is the only way to formulate the optimization problem, but given the amount of discussion the choice of objective function in the information-theoretical setting already raises, we keep our focus on that aspect.

We will add the thoughts above in the paper to clarify the relation between the 3 scenarios, constraints, and data sources.

Yeh, H.-C.; Chen, Y.-C.; Chang, C.-H.; Ho, C.-H.; Wei, C. Rainfall Network Optimization Using Radar and Entropy. *Entropy* **2017**, *19*, 553.DOI: **https://doi.org/10.3390/e19100553**

Specific comments

The section on presenting the basics of information theory can be better summarised. There is plenty of "well-known" material on it.

Since not all of the HESS audience is necessarily aware of information theory, and the interpretation of information measures is at the core of the debate here, we believe the inclusion of some well-known material is justified. For example, reviewer #1 states: “I enjoyed reading the manuscript, especially the introduction to the information theory terms.”.

Though readers like you, who have a deep background and opinion on the information measures is definitely are a key part of the intended audience we hope to convince, we also hope to serve a wider audience who would welcome some introductory material.

We deleted a few points that are not directly relevant to the argument here, such as the different units, and the thermodynamic origin of the entropy concept.

l172-173 requires a reference.
Thanks for catching this, it is fixed by adding the following references.

We will add:

Maximizing network information content, through either the sum of marginal entropy or joint entropy, is the common theme among existing methods *(Alfonso et al., 2010b; Li et al., 2012; Samuel et al., 2013; Keum and Coulibaly, 2017; Wang et al., 2018; Huang et al., 2020).*

l187 requires explanation about what is objective GR3

Thanks for catching this, we did explain GR1 to GR6 in the appendix but forgot to direct our readers there. We also noticed a critical word "not" that we forgot to insert. it is now modified:

Added in text:

This is equivalent to the GR3 objective proposed by Banik et al. (2017), as part of six other objectives *(see appendix A for more detail)* proposed in the same paper, which did not provide arguments or preference for its use.

l200-201 These are not MOO methods. These are objective functions.

We agree. We replaced "three multi-objective optimization methods" with "three other (sets of) objective functions from previously proposed methods".

l202- 203 requires references
Thanks, it is properly referenced now.

l218-220 It is not true that the only way of selecting stations is using greedy algorithms.

We agree with you that the greedy algorithm is not the only way. But, a greedy approach is used by the majority of studies in the literature. We modified the text to clarify this.

We will add:

In the majority of existing literature, another constraint has often implicitly been imposed: to treat the selection of stations as a greedy optimization, meaning that one station is added to the set of selected stations each time while trying to optimize the objective function, without reconsidering the already selected stations in the set.

l221 is not combinatorial "explosion". Instead we can argue that the problem is exponentially complex (Oˆn)

Will be modified as recommended.

We agree that the the search space for evaluating all sub-networks (the power set of the set of sensor locations) grows exponentially with 2^n. The greedy approach has a search space of O(n^2).

We will add::

A practical reason for this is numerical efficiency; an exhaustive search of all subsets of k stations out of n possible stations will need to consider a large number of combinations, since the search space grows exponentially with the size n of the full set of sensors (2^n combinations of sensors need to be considered).

l222 I think here you are mixing two design problems. One of the problems is of design (where to measure at several locations), and other of augmentation (Where to put additional stations). Of course these are clearly different processes.

You are correct that technically these are different problems where perhaps in one case, you would install a new sensor, and in the other, you would just switch it on. In this research, we are just focusing on the benefits side (the information obtained), and how to quantify that. While the pool of candidate locations and way to model what could be measured or is measured may be different, the way to formulate the objective functions would be the same. So the problems are indeed different, but not that much in the aspects that we consider in this paper. See also the answer for "What are the cases of sensor network design that are considered?" We will clarify this in the paper by reformulating line 222 and some lines around it.

l232 I think it should be necessary to include methods using metaheuristics for comparison.

As mentioned in our answer to comment 1, we will discuss the metaheuristic approach in the further work section.

We agree with the reviewer that for a full investigation of computationally efficient methods, we would need to include the metaheuristic approaches. We mention this in the future work section. For our current paper, we just want to highlight one other very computationally cheap greedy approach - O(n^2) - , which we prove non-optimal by counter example. We then use that to briefly discuss why greedy approaches cannot be optimal, but also what the implication for expanding or reducing a network is.

It would indeed be interesting to look at approaches of intermediate computational complexity and see how much optimality can be gained compared to either of the the greedy approaches.

l233 It is not clear what logistical reasons are. Should not these be included in the optimisation constraints?

We agree this was vague. We now clarified by adding:

... logistical reasons that may require a gradual strategy of expanding or reducing 1 station at a time. The exhaustive optimization yields a series of networks where an increase in size also involves relocating stations, which may not always be practically feasible or desired.

l235-236 This seems speculative at this point, and better be moved to other place in the document (perhaps introduction?)

Agreed. We will move it. . This is a hypothesis that we falsify in the paper, and is better placed in the introduction. .

l233 This "golden standard" expression seem somewhat loose talk. Can just point out is the only way to prove optimality?

Agreed. "golden standard" expression is replaced by "optimality benchmark"

Figure 4 can be improved. labels are hard to read, and would be more informative just to keep the ID's?. Also if its for monthly data, have you considered using flow duration curves instead? alternatively, please consider using a log y-scale, as discharge distributions are positively skewed. In addition, y-scale label is missing.

Thanks for the idea, we modified fig 4.

We changed y-scale to the log scale. The box plot and statistics show much of the information that can be presented by FDC. We tried the FDC, but it becomes quite busy. Also, the information measures are more directly linked to the pdf than to the cdf, so we find the stick to the current figure with the logarithmic y axis you suggested.

in Table 2 are presented the results of different optimisation criteria (defined as MOO), but there is no indication of the selecting strategies, thus the information of the whole Pareto set is unavailable. In top, that is the whole point of MOO, that there exists trade-offs between objectives that cannot be assessed by the modeller.

We agree with the reviewer about the role of trade-offs weights in MOO. But, we chose to accept Li et al. (2012)'s conclusion on insignificant effect of information-redundancy trade-off weights in this particular case study. They conducted sensitivity analysis and reported that MIMR results are stable with respect to trade-offs weights in MOO. We compared maxJE with two other highly cited single-objective methods in our field because they either directly minimize redundancy (minT) or indirectly minimize redundancy by imposing constraint in search space (WMP). We didn't discuss trade-offs weights and presented Li et al. (2012)'s conclusion since we argue that this trade-off is irrelevant in the first place. As you correctly pointed out, the subjectively issue can be raised on the modeler's decision of trade-off weights. Therefore, we decided to present the Li et al. (2012) results in our paper for the sake of having unbiased comparison, assuming they represent the use of the method as intended.

S4.2 includes parts that should have been presented in the methodology and not in the results section.

We will see if we can re-organize some material to improve the flow. This is somewhat challenging because the case study serves as illustration, not as an experiment is needed to support conclusions. To explain our rationale for including this material here: This section is meant to further explore our key point that minimizing dependence is not needed as an objective. This is a result from the reasoning in the paper, which is continued in the results section so it can benefit from numerical results for illustration. We bring in the pareto space plot to illustrate that the maximizing redundancy objective changes the pareto front considered. We see that as an illustration of the discussion, rather than a predefined experiment to test a hypothesis, hence it's placement here. We will look at this again carefully to see if the paper improves by moving material.

l276. H(X,X) = H(X), therefore two completely "dependent" are exactly as informative as one.

Your statement on two completely "dependent" is correct. But, our goal in L276 is to point out that minT would prefer a completely "independent" sensor with low entropy in its selection process over the one with higher entropy that is more dependent. The sentence in line 276 is meant to criticize preference on minimum dependence by negatively stating the obvious.

Visualisation using Venn diagrams are an excellent way of presenting concepts, but are really hard to follow to describe precise quantities. Will it be possible to re-think Figure 5 in a simpler manner?
We agree on the loss of precise quantities, but we think the figure as it is provides a good visual overview of how the info measures vary, by growing and shrinking colors. Many other ways of plotting lose the interconnections between the measures plotted. To have the best of both worlds, we will add all data behind that figure in tables in the supplementary material, so the precise quantities are available for further study.

l287-288 Those are precisely the trade-offs that decision-makers do in MOO, and the reason of its relevance. If you claim that maximum joint entropy is somewhat equivalent to minimum total correlation, then these two objective functions are not conflicting, and therefore, by maximising one, you are maximising the other. Precisely as shown in Figure 7.
In these lines we indicate that we do see the relevance of MOO in the case of **maximizing** total correlation next to maximizing joint entropy. However, the papers cited in this work **minimize** total correlation as a secondary objective. As seen from figure 7, this will reduce the joint entropy and, as argued in the text, it is doing so without independent justification for that secondary objective.
Low total correlation is one of the factors in max joint entropy, but they are not equivalent. There is indeed some correlation between the objectives, but there is still some trade off (as indicated by the dashed pareto front in figure 7, which previous approaches explored). In this paper, we argue that this trade-off is not relevant, and that an exploration of another pareto front in this space, could be justified from the point of view of maximizing robustness against failure.

l297-298 The whole point in obtaining the Pareto front between maximum joint entropy and minimum total correlation explores the trade-offs between a network that is able to capture most information, vs a network that has little "information overlap", which is not the same as the individual entropies are different. Therefore, these are different metrics, and is the reasoning behind finding the Pareto set in the dashed line of Fig 7.

We agree that this is a good summary of the justification given in most previous literature, and we do agree there is a trade-off, **but what is the motivation behind having little information overlap if it's not to capture more information** (which is already addressed by joint entropy)? We did not find such motivation or justification in the previous literature and don't believe there is one.

Hence this is the core point of our paper, which results from reasoning about the information measures and not from case study results. The case study results just illustrate the effects.

S4.3 A lot of the text in this section can be part of the methodology.

We agree that some material can be moved to methodology. We will experiment with this to see what best serves the clarity of our messages, keeping in mind that we think the points here are secondary in importance to our points about the objective functions.

Table 4 is quite hard to read. Also, this table is precisely showing you that optimallity is not found using greedy algorithms.

Indeed, the point of including table 4 and 5, for which we generated a dataset, is to show that greedy algorithms will not necessarily find the optimum. We now included this in the caption to make that clearer, and also reorganized table 4 slightly make it easier to read. We split rows of exhaustive optimization into added and removed stations), use light shading of cells, and improved layout.

Table 5 Be consistent in the amount of significant digits through your document

Thanks for catching this. We will make sure to be consistent at least within each table. We felt the need to include extra digits in this table, to show the small differences that occur, and hence provide a falsification of the hypothesis that greedy algorithms always lead to optimal solutions. Even a small difference would prove that. Alternatively we could reduce to 2 decimals for consistency, but color the suboptimal configurations, but then the sub-optimality could not be verified by the readers.

l344 No information system is justified if there is not objective to tackle. What problem am I addressing if I do not know what the problem is.

We reformulated to make it clearer what we mean: When the decision problem that is supported by the monitoring network is not simple and explicitly defined, and not fully known, then we may choose to maximize the information content of the network, as that provides potential value for the widest possible set of uses. You cannot derive value from a network that provides no information. On the other hand, giving information does not guarantee value for every decision.

An example is a hydrometric network of a national institute, which is maintained by government money to support a variety of a priori unknown decision problems.

We agree that attempt could be made to quantify an estimated value regardless, but our focus here to discuss purely information based methods, for which there is a wide existing literature.

l361 Large optimisation problems are tackled using metaheuristics, and has been a widely used approach. This has not been mentioned here at all.

Thank you for noticing this, we now mention metaheuristics as a potential other approach to address large problems. This is certainly worth further investigation in future work..

l364 the differences between the greedy and exhaustive search approaches have not been presented quantitatively. In this problem they may seem "little" (not being explicit about what little or big is in the context of the problem), but this cannot be ensured for larger problems that the ones presented here.
The differences between these approaches have been presented quantitatively in table 5. We agree this cannot be ensured for larger problems, hence our call for further research on this.
We modified line 361 to clarify that we only speak about our simple case study and not in general when we say differences were small.

l367-368 Language has to be precise (how to numerically calculate this objective function, or other objective functions used in other approaches)

Agreed, we will modify as follows:

Another important question that needs to be addressed in future research *is to investigate how the assumption made (i.e., data quantization which influences probability distribution) in the numerical calculation of objective functions would affect network ranking.*

l370 Any information metric is hard to calculate with limited data.
Yes we agree, but the hardness (or data need) exponentially grows with the number of dimensions in the PDF. Entropy and pairwise mutual information are not a big problem, but total correlation and joint entropy become problematic for a large number of sensors.

l378-379 I completely agree with this line "before thinking about how to optimize, we should be clear on what to optimize".
We are glad the reviewer agrees, this is also a reason why we placed the greedy vs non greedy discussion later in the paper as it is secondary to the main points about which objective function to use. We will see if we can move some material to the methods without breaking the flow of the main argument.

l381 I visited the GitHub repository, but I was unable to find the code to reproduce these results. Only a reply to a WRR paper of 2018, and a fork of pysheds.
We apologize. They will be posted shortly.