Combine answers to all 4 reviewers

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 <u>underline</u> 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 even 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 expanded the first paragraph of the further work section to discuss these issues.

Added in L240:

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.

Added in L269:

To assess and illustrate the workings of the different objectives in retrieving information from the water system, we compared three existing methods with a direct maximization of the joint entropy of selected sensors, H (S,Fc), indicated with maxJE in the results, such as Tables 2 and 3. The joint entropy results in table 2 indicate that maxJE is able to find a combination of 8 stations that contains joint information of all 12 stations ranked by other existing methods*. Figure 5 displays spatial distribution of the top 8 stations chosen by different methods. Before any interpretation of the placement, we must note that the choices made in quantization and the availability of data play an important role in the optimal networks identified. Whether the saturation that occurs with 8 stations has meaning for the real world case study depends on whether the joint probability distribution can be reliably estimated. This is highly debatable and merits a separate detailed discussion which is out of the scope of this paper. We present this case study solely to illustrate behaviour of the various objectives.*

The most notable difference between MaxJE and the other methods is the selection of all 3 of the stations located most downstream. While other methods would not select these together due to high redundancy between them, maxJE still selects all stations, because despite the redundancy, there is still found to be enough new information in the second most downstream station. This can be in part attributed to the quantization choice of equally sized bins throughout the network, leading to higher information contents downstream. While this quantization choice is debatable, it is important, in our opinion, to not compensate artifacts from quantization by modifying the objective function, even if the resulting network may seem more reasonable, but rather to address those artifacts in the quantization choices themselves. To repeat the key point: An objective function should not be chosen based on whether it yields a "reasonable network" but rather based on whether the principles that define it are reasonable.

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. When considering all sub-network sizes, the number of combinations to consider is 2^n, so an exponential growth. We could make an optimistic estimate, only considering the scaling from station combinations to evaluate, 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 1 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 in L360:

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:
Added in L387:

The aim of this paper was to contribute to better understanding the problem of optimal monitoring network layout using information-theoretical methods. Since using resulting networks and performance metrics from case studies to demonstrate that one objective should be preferred over the other would be circular, the results from our case study served as an illustration of the effects, but not as arguments supporting the conclusions we draw about objective functions. We investigated the rationale for using various multiple-objective and single-objective approaches, and discussed the advantages and limitations of using *exhaustive vs. greedy search. The main conclusions for the study can be summarized as follows:*

- *The purpose of the monitoring network governs which objective functions should be considered. When no explicit information about users and their decision problems can be identified, maximizing the total information collected by the 395 network becomes a reasonable objective. Joint entropy is the only objective needed to maximize retrieved information, assuming that this joint entropy can be properly quantified.*
- *We argued that the widespread notion of minimizing redundancy, or dependence between monitored signals, as a secondary objective is not desirable and has no intrinsic justification. The negative effect of redundancy on total collected information is already accounted for in joint entropy, which measures total information net of any redundancies.*
- *When the negative effect on total information is already accounted for, redundant information is arguably beneficial, as it increases robustness of the network information delivery when individual sensors may fail. Maximizing redundancy as an objective secondary to maximizing joint entropy could therefore be argued for, and trade-off between these objectives could be explored depending on the specific case.*
- *The comparison of exhaustive and greedy search approaches shows that no greedy approach can exist that is guaranteed to give the true optimum subset of sensors for each network size. However, the exponential computational complexity, which doubles the number of sensor combinations to evaluate with every sensor added, makes exhaustive search prohibitive when the number of possible locations become larger than about 25. The complexity of the greedy approach is quadratic in the number of locations, and therefore feasible for large search spaces.*
- *The constraints to the search space imposed by the greedy approach could also be interpreted as a logistical constraint. In a network expansion scenario, it disallows replacement of stations already selected in the previous iteration.*
- *We introduced the "greedy drop" approach that starts from the full set and deselects stations one by one. We have demonstrated that the two types of greedy approaches do not always lead to the same result, and neither approach guarantees the unconstrained true optimal solution. Synergistic interactions between variables may play a role, although this is not the only possible explanation. In our case study, the suboptimality of greedy algorithms was not visible in original data, but we demonstrated its existence with artificially generated data. In our specific case studies, differences between exhaustive and greedy approaches were small; especially when using a combination of the greedy add and greedy drop strategy. It remains to be demonstrated in further research how serious this loss of optimality is in a range of practical situations, and how results compare to intermediate computational complexity approaches such as metaheuristic algorithms.*

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.

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 <u>underline</u> 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.

Added in L23:
<u>In this paper, we do not answer the question "what is optimal?" with an optimal network. Rather, we reflect on the question of how to define optimality in a way that is logically consistent and useful within the monitoring network optimization context, thereby questioning the widespread use of minimum dependence between stations as part of the objectives.</u>

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 made several edits to section 3 and 4.1 to clarify these contributing factors to our readers.

Some further related thoughts, that we integrated at several points 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.

Changed start of section 3::

In previous studies, the focus of the research has been on finding an optimal network for the subject
case study without sufficiently addressing the theoretical justification of applying a new methodology.

For this reason and the primary goal of this paper, which is highlighting the unnecessary use of multiple objective functions in monitoring network design, we decided to apply our methodology in Brazos River streamflow network (Figure 3) since this network was subject of study for the MIMR method. This network is under-gauged, according to the World Meteorological Organization density requirement. However, it provides an opportunity to eliminate subjectivity issues on network design decisions such as network density, data quantization, and temporal variability of data.

We also added to section 4.1:
Whether the saturation that occurs with 8 stations has meaning for the real world case study depends on whether the joint probability distribution can be reliably estimated. This is highly debatable and merits a separate detailed discussion which is out of the scope of this paper. We present this case study solely to illustrate behaviour of the various objectives.

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 modified 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, λ1, 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.

Added in L240:

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.*

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.

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 underline 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.

We added in L372:

Based on our limited case study, the questions remain open: 1) whether faster algorithms can be formulated that yield guaranteed optimal solutions, and 2) in which cases the greedy *algorithm provides a close approximation. It is also possible to formulate modified greedy methods with the ability of replacing a limited number of monitors instead of just adding monitors. This leads to a significantly reduced computational burden compared to exhaustive optimization, while reaching the optimum more often than when adding monitors one at a time. In Table 5, it can be seen that allowing a maximum number of two relocated monitors would already reach the optimal configurations for this specific case. Another limitation of this comparison is that we did not consider metaheuristic search approaches (Deb et al., 2002; Kollat et al., 2008), which fall in between greedy and exhaustive approaches in terms of computational complexity, could serve to further explore the optimality versus computational complexity trade-off. It would be interesting to further investigate what properties in the data drive the suboptimality of greedy algorithms. Synergistic interactions (Goodwell and Kumar, 2017) are a possible explanation, although our generated data example shows that even when moving from 1 to 2 selected stations, a replacement occurs. Since there are only pairs of variables involved, synergy is not needed in the explanation of this behaviour. Rather, the pair with maximum joint entropy does not always include the station with maximum entropy, which could perhaps be too highly correlated with other high entropy variables.*

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:

Added in L287

Though already necessarily true from the formulation of the objective functions, we use the case study to illustrate how other methods with a separate minimum redundancy objective lead to the selection of stations with lower new information content (green area in Figure 6). 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 (by definition) has the fastest, and minT the slowest rate of reduction of information loss. Methods' 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).*

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.

Added in L56:

Secondly, we argue that, in purely information-based approaches, the joint entropy of all signals together is in principle sufficient to characterize information content and can therefore serve as single optimization objective. *Notions of minimizing dependence between monitored signals through incorporation of other information metrics in the objective function lack justification and are therefore not desirable. Thirdly, because the undesirable information inefficiencies associated with high dependency or redundancy are already accounted for in maximizing joint entropy, we could actually argue for maximizing redundancy as a secondary objective, because of its associated benefits for creating a network robust against failures of individual sensors. 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. Adding a trade-off with maximum redundancy is outside the scope of this paper, but serves to further illustrate the argument against use of minimum redundancy.*

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.

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, or is presented as self-evident.

Some examples of this language 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 also clarify why no greedy approach could exist that is guaranteed to find all optimum sensor subsets. We now added that more explicitly in the conclusions.

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 added in section 4.3

Another limitation of this comparison is that we did not consider metaheuristic search approaches (Deb, 2002; Kollat, 2008), which fall in between greedy and exhaustive approaches in terms of computational complexity, could serve to further explore the optimality versus computational complexity trade-off.

We added in the conclusions:

It remains to be demonstrated in further research how serious this loss of optimality is in a range of practical situations, and how results compare to intermediate computational complexity approaches such as metaheuristic algorithms.

We added in further work:

In this paper, we focused on the theoretical arguments for justifying the use of various objective functions, 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 tools to find optimal network configurations, we compared greedy and exhaustive search approaches to raise awareness in the scientific community that greedy optimization might fall into 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 that the 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 (for the maxJE objective) and computation 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.

Added in L23:
In this paper, we do not answer the question ``what is optimal?" with an optimal network. Rather, we reflect on the question of how to define optimality in a way that is logically consistent and useful within the monitoring network optimization context, thereby questioning the widespread use of minimum dependence between stations as part of the objectives.

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 for an optimization that is intended to help decision makers, rather than describing their existing choices, necessarily has some normative elements to it, 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 one of the means as a secondary objective. This will only lead to a loss in performance on the main goal (in a way by overly accounting for the effect of dependency).
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.
So you are correct that at least part of our approach is normative, in the sense that we argue against use of minimum dependency as an objective. We think that was also clear from our wording, but made some small textual modification to make it clearer.

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.

We edited the first argument as follows:
Firstly, we argue that objective functions for optimizing monitoring networks can, in principle, not be justified by analysing the resulting networks from application case studies. Evaluating performance of a chosen monitoring network would require a performance indicator which in itself is an objective function. Case studies could be helpful in assessing whether one objective function (the optimization objective) could be used as an approximation of another, underlying, objective function (the performance indicator). However, from results of case studies we should not attempt to draw any conclusions as to what objective function should be preferred. In other words: the objective function is intended to assess the quality of the monitoring network, as opposed to a practice where the resulting monitoring networks are used to assess the quality of the objective function.

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. We now worded argument 2 as follows:
Secondly, we argue that, in purely information-based approaches, the joint entropy of all signals together is in principle sufficient to characterize information content and can therefore serve as single optimization objective. Notions of minimizing dependence between monitored signals through incorporation of other information metrics in the objective function lack justification and are therefore not desirable.

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 now rewrote the third point as following:

Thirdly, because the undesirable information inefficiencies associated with high dependency or redundancy are already accounted for in maximizing joint entropy, we could actually argue for maximizing redundancy as a secondary objective, because of its associated benefits for creating a network robust against failures of individual sensors. 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. Adding a trade-off with maximum redundancy is outside the scope of this paper, but serves to further illustrate the argument against use of minimum redundancy.

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 agree that the sentence you quoted is not directly a conclusion from the main arguments presented is this paper, and hence we removed it. To cl
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..

Added in L387:
The aim of this paper was to contribute to better understanding the problem of optimal monitoring network layout using information-theoretical methods. Since using resulting networks and performance metrics from case studies to demonstrate that one objective should be preferred over the other would be circular, the results from our case study served as an illustration of the effects, but not as arguments supporting the conclusions we draw about objective functions. We investigated the rationale for using various multiple-objective and single-objective approaches, and discussed the advantages and limitations of using exhaustive vs. greedy search.

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 effectiveness, which we argue is covered by looking at the total non-redundant information 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 mention this where greedy approaches are discussed and in the future work section. See the quotes of added text earlier in this reply.

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 exist that 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 made several smaller edits throughout the text to make the points above clearer.

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.

added in L171:

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 L187:

This is equivalent to the GR3 objective proposed by Banik et al. (2017), as part of six other objectives (see appendix B for more detail) proposed in the same paper, which did not provide 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 changed as follows:

In the majority of existing literature listed in Table 1, one constraint has often implicitly been imposed: to treat the selection of stations as 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 added::

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?”

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 changing as follows:

In this comparison, we will investigate whether the exhaustive optimization yields a series of networks where an increase in network size may also involve relocating stations. This may not always be practically feasible or desired in actual placement strategies, where networks are slowly expanded one station at a time. Occurrence of relocation in the sequence of growing subsets would also show that no greedy algorithm could exist that guarantees optimality.

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

Agreed. This is a hypothesis that we falsify in the paper, and we removed it here .

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.
We added:
For the multi-objective approaches used in the case study, we used the same weights as the original authors to identify a single solution.

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 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 looked at this carefully and moved some 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 and sharper than our original formulation. We modified it as suggested.

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 pointed to the code repository in which all data behind that figure and tables can be found, 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. We did subtle rewording throughout the paper to clarify this point.

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 moved some material around to best serve 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 no greedy algorithm can exist that is guaranteed to find the optimum. We now included this in the caption to make that clearer, and also reorganized table 4 (now 5) slightly to make it easier to read. We split rows of exhaustive optimization into added and removed stations), and improved layout, to make the table easier to read..

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.

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 attempts could be made to quantify an estimated value regardless, but our focus here is to discuss purely information based methods, for which there is a wide existing literature. We did significant rewording of the whole conclusions section.

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.

In <u>our specific case studies,</u> differences between exhaustive and greedy approaches were small; especially when using a combination of the greedy add and greedy drop strategy. It remains to be demonstrated in further research how serious this loss of optimality is in a range of practical situations, and how results compare to intermediate computational complexity approaches such as metaheuristic algorithms.

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 choices and assumptions 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. We now clarified as follows:

These probability distributions are hard to reliably estimate from limited data, especially in higher dimensions, *since data requirements grow exponentially.*

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 reorganised some material, 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.

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

Hossein Foroozand[1] and Steven V. Weijs[1]

[1]Department of Civil Engineering, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence: Steven V. Weijs (steven.weijs@civil.ubc.ca)

Abstract. This paper concerns the problem of optimal monitoring network ~~lay- out~~ layout using information-theoretical methods. Numerous different objectives based on information measures have been proposed in recent literature, often focusing simultaneously on maximum information and minimum dependence between the chosen locations for data collection. We discuss these objective functions and conclude that a single objective optimization of joint entropy suffices to maximize the collection of information for a given number of sensors. ~~Minimum dependence is~~ We argue that the widespread notion of minimizing redundancy, or dependence between monitored signals, as a secondary objective ~~that automatically follows from the first, but~~ is not desirable and has no intrinsic justification. The negative effect of redundancy on total collected information is already accounted for in joint entropy, which measures total information net of any redundancies. In fact, for two networks of equal joint entropy, the one with a higher amount of redundant information should be preferred for reasons of robustness against failure. In attaining the maximum joint entropy objective, we investigate exhaustive optimization, a more computationally tractable greedy approach that adds one station at a time, and we introduce the ~~"greedy drop"~~ "greedy drop" approach, where the full set of sensors is reduced one at a time. We show that ~~only exhaustive optimization will give the true~~ no greedy approach can exist that is guaranteed to reach the global optimum. The arguments are illustrated by a comparative case study.

1 Introduction

Over the last decade, a large number of papers on information theory based design of monitoring networks have been published. These studies apply information-theoretical measures on multiple time series from a set of sensors, to identify optimal subsets. Jointly, these papers (Alfonso et al., 2010a, b; Li et al., 2012; Ridolfi et al., 2011; Samuel et al., 2013; Stosic et al., 2017; Keum and Couli have proposed a wide variety of different optimization objectives. Some have suggested that either a multi-objective approach or a single objective derived from multiple objectives is necessary to find an optimal monitoring network. These methods were often compared to other existing methods in case studies, used to demonstrate, based on the resulting networks, that one objective should be preferred over the other.

In this paper, we do not answer the question "what is optimal?" with an optimal network. Rather, we reflect on the question of how to define optimality in a way that is logically consistent and useful within the monitoring network optimization context, thereby questioning the widespread use of minimum dependence between stations as part of the objectives.

The objective of a hydrological monitoring network depends on its purpose, which can usually be framed as supporting decisions. The decisions can be relating to management of water systems, as for example considered by Alfonso et al. (2010a) or flood warning and evacuation decisions in uncontrolled systems. ~~However, also~~ Also purely scientific research can be formulated as involving decisions to accept or reject certain hypotheses, focus research on certain aspects, or collect more data (Raso et al., 2018). In fact, choosing monitoring locations is also a decision, whose objective can be formulated as choosing monitoring locations to optimally support subsequent decisions.

The decision problem of choosing an optimal monitoring network layout needs an explicit objective function to be optimized. While this objective could be stated in terms of a utility function (Neumann and Morgenstern, 1953), this requires knowledge of the decision problem(s) at hand and preferences of the decision maker, which are often not explicitly available, for example in case of a government operated monitoring network with a multitude of users. As a special case of utility, it is possible to state the objective of a monitoring network in terms of information (Bernardo, 1979). This can be done using the framework of information theory, originally outlined by Shannon (1948), who introduced information entropy H(X) as a measure of uncertainty or missing information in the probability distribution of random variable X, as well as many related measures.

Although ultimately the objective will be a more general utility, the focus of this paper is on information-theoretical methods for monitoring network design, which typically do not optimize for a specific decision problem supported by the network. Because information and utility (value of information) are linked through a complex relationship, this does not necessarily optimize decisions for all decision makers. Since we do not consider a specific decision problem, the focus in the present paper is on methods for maximization of information retrieved from a sensor network.

1.1 ~~Background~~

~~Over the last decade, a large number of papers on information theory based design of monitoring networks have been published. These studies apply information theoretical measures on multiple of time series from a set of sensors, to identify optimal subsets. Jointly, these papers (Alfonso et al., 2010a, b; Li et al., 2012; Ridolfi et al., 2011; Samuel et al., 2013; Stosic et al., 2017; Keum ar have proposed a wide variety of different optimization objectives. Some have suggested that either a multi-objective approach or an single objective derived from multiple objectives is necessary to find an optimal monitoring network. These methods then were often compared to other existing methods in case studies, used to demonstrate that one objective should be preferred over the other.~~

In this paper, the rationale behind posing ~~these information theoretical~~ various information-theoretical objectives is discussed in detail. While measures from information theory ~~provides~~ provide a strong foundation for mathematically and conceptually rigorous quantification of information content, it is important to pay attention to the exact meaning of the measures used. This paper is intended to shed some light on these meanings in the context of monitoring network optimization and provides new discussion motivated in part by recently published literature.

1.1 ~~motivation~~

We present three main arguments in this paper. Firstly, we argue that objective functions for optimizing monitoring networks can, in principle, not be justified by analysing the resulting networks from application case studies. Evaluating performance of a chosen monitoring network would require a performance indicator which in itself is an objective function. Case studies could be helpful in assessing whether one objective function (the optimization objective) could be used as an approximation of another, underlying, objective function (the performance indicator). However, from results of case studies we ~~cannot draw any normative~~ should not attempt to draw any conclusions as to what objective function should be preferred. In other words: the objective function is intended to assess the quality of the monitoring network, as opposed to a practice where the resulting monitoring networks are used to assess the quality of the objective function.

Secondly, we argue that, in purely information-based approaches, the joint entropy of all signals together is in principle sufficient to characterize information content and can therefore serve as single optimization objective. Notions of minimizing dependence between monitored signals through incorporation of other information metrics in the objective function ~~are~~ lack justification and are therefore not desirable.

~~Thirdly, multi-objective approaches that use some quantification of~~ Thirdly, because the undesirable information inefficiencies associated with high dependency or redundancy ~~as a secondary objective, next to~~ are already accounted for in maximizing joint entropy, ~~could only be justified if redundancy is interpreted as beneficial~~ we could actually argue for maximizing redundancy as a secondary objective, because of its associated benefits for creating a ~~robust network , and therefore an objective to be maximized~~network robust against failures of individual sensors. 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. Adding a trade-off with maximum redundancy is outside the scope of this paper, but serves to further illustrate the argument against use of minimum redundancy.

1.0.1 ~~scope~~

~~In monitoring network design,~~ The manuscript is organized as follows. In the following methodology section, we introduce the methods used to investigate and illustrate the role of objective functions. In section 3, we discuss the case study on the streamflow monitoring network of Brazos River. Section 4 introduces the results for the various methods, and then discusses the need for multiple objectives, the interpretation of trade-offs, and the feasibility of greedy algorithms reaching the optimum. The article concludes with summarizing the key messages and raising important questions about the calculation of the metrics, to be addressed in future research.

2 Methodology

2.1 Choice of scope and role of the case study

In monitoring network design, also other objectives, not relating to information measures, have been used. Examples are cost, geographical spread, and squared error based metrics. ~~Also some~~ Some approaches use models describing spatial variability with certain assumptions, e.g. kriging (Bayat et al., 2019). In the case of network expansion to new locations, models are always needed to describe what could be measured in those locations. ~~This~~ These could vary from simple linear models to full hydrodynamic transport models, such as for example done in (Aydin et al., 2019).

In this paper, our main focus is ~~discussing~~ to discuss the formulation of information-theoretical objective functions and previous literature on that topic. Therefore, we restrict our scope to those information-theory based objective functions, based on spatially distributed observed data on one single variable. Keeping this limited scope allows us to discuss the interpretation of these objective functions, which formalize what we actually want from a network. Furthermore, we investigate whether the desired optimum in the objective function can be found by greedy approaches, or whether exhaustive search is needed to prevent a loss of optimality.

Only after it is agreed on what is wanted from a network and this is captured as an optimization problem, other issues such as the solution or approximation of the solution to the problem become relevant. The numerical approach to this solution and calculation of information measures involved, warrants another, independent discussion, which is outside our current scope and will be presented in a future paper.

Our discussion is numerically demonstrated by using data from the case study for Brazos River in Texas, as presented in Li et al. (2012), to allow for comparisons. However, as we will argue, the case study can only serve as illustration, and not for normative arguments for use of a particular objective function. Such an argument would be circular, as the performance metric will be one of the objective functions.

2.1.1 ~~Manuscript organization~~

~~The manuscript is organized as follows. In the following methods section, we introduce the methods used to investigate and illustrate the role of objective functions. In section 3, we discuss the case study on the streamflow monitoring network of Brazos River. Section 4 introduces the results for the various methods, and then discusses the need for multiple objectives, the interpretation of trade-offs, and the feasibility of greedy algorithms reaching the optimum. The article concludes with summarizing the key messages and raising important questions about the calculation of the metrics, to be addressed in future research.~~

3 ~~Methodology~~

In this paper, since we are discussing the appropriate choice of objective function, there is no experimental setup that could be used to provide evidence for one objective version versus the other. ~~This is because to define a "best" or "optimal" golden standard network to aspire to, we need an objective function.~~ Rather, we must make use of normative theoretical reasoning, and shining light on the interpretation of the objectives used and their possible justifications. The practical case studies in this paper therefore serve as illustration, but not as evidence for all the conclusions advocated in this paper, some of which are arrived at through interpretation and argumentation in the discussion section. This methodology section introduces the elementary information measures used in this paper and previous literature we compare with. After, we discuss visualization of these information measures, and finally, we discuss the proposed and previously used objective functions for monitoring network design, which are composed of these elementary information measures.

2.1 Information theory terms

~~The concept of entropy was introduced in thermodynamics as a measure of thermal disorder or randomness of a system.~~ Shannon (1948) developed information theory (IT) based on entropy, the concept that explains a system's uncertainty reduction as a function of added information. To understand how, consider set of N events for which possible outcomes are categorized into m classes, uncertainty is a measure of our knowledge about which outcome will occur. Once an event is observed, ~~the class the outcome~~ and which of the m classes it belongs to is identified, ~~and~~ our uncertainty about the outcome decreases to 0. Therefore, information can be characterized as ~~decrease~~ the decrease of an observer's uncertainty about the outcome ~~.~~ (Krstanovic and Singh, 1992; Mogheir et al., 2006; Samuel et al., 2013; Foroozand and Weijs, 2017; Foroozand et al., 2018). For monitoring networks, the information each sensor provides through its observations (outcomes) is therefore linked to the uncertainty of those outcomes before measurement. These are quantified through the probability distributions of the data.

In monitoring network design, IT has been applied in the literature to evaluate data collection networks that serve a variety of purposes, including rainfall measurement, water quality monitoring, and streamflow monitoring. These evaluations are then used to optimize placement of sensors. ~~The purpose of the network often governs which of information theory's expressions are considered.~~ In the monitoring network optimization literature, three expressions from IT are often used in monitoring network design: (1) entropy (H), to estimate the expected information content of observations of random variables; (2) Mutual information, often called transinformation (T), to measure redundant information or dependency between two variables; (3) total correlation (C), a multivariate analogue to mutual information, to measure the total nonlinear dependency among multiple random variables. Objective functions are often composed from these basic expressions. Details of each expression are presented below.

The Shannon entropy H (X) is a nonparametric measure, directly on the discrete probabilities, with no prior assumptions on data distribution. It is also referred to as discrete marginal entropy, to distinguish it both from continuous entropy and from conditional entropy. Discrete marginal entropy (Eq.1), defined as the average information content of observations of a random variable X, is given by:

$$\mathrm{H(X)} = -\sum_{x \in X} p(x) \log_2 p(x) \qquad (1)$$

where $p(x)$ $(0 \leqslant p(x) \leqslant 1)$ is the probability of occurrence of outcome x of random variable X. Equation 1 gives the entropy in the units of "bits" since it uses a logarithm of base 2. The choice of ~~logarithm's base for entropy calculation~~ base is determined by the desired unit~~— other information units are "nats" and "Hartley" for the natural and base 10 logarithms, respectively. For~~. In the monitoring network design ~~, logarithm of base 2 is common in the literature~~ literature, the bit is a common unit, since it can be interpreted as the needed number of answers to a series of binary questions.

~~Joint entropy~~ Joint entropy (Eq.2) measures the number of questions needed to determine the outcome of a multivariate system.

$$\mathrm{H}(X_1,X_2) = -\sum_{x_1\in X_1}\sum_{x_2\in X_2} p(x_1,x_2)\log_2 p(x_1,x_2) \tag{2}$$

where $p(x_1)$ and $p(x_2)$ constitute the marginal probability distribution of random variable X_1 and X_2, respectively; and $p(x_1,x_2)$ form their joint probability distribution. For a bivariate case (X_1,X_2), if two random variables are independent, then their joint entropy, ~~$\mathrm{H}(X_1|X_2)$, (Eq.2)~~ $\mathrm{H}(X_1,X_2)$, is equal to the sum of marginal entropies $\mathrm{H}(X_1)+\mathrm{H}(X_2)$. Conditional entropy (Eq.3), ~~which~~ $\mathrm{H}(X_1|X_2)$, explains the amount of information ~~one variable delivers that other variable~~ X_1 delivers that X_2 can not explain~~,~~.

$$\mathrm{H}(X_1|X_2) = -\sum_{x_1\in X_1}\sum_{x_2\in X_2} p(x_1,x_2)\log_2 \frac{p(x_1,x_2)}{p(x_2)} \tag{3}$$

$$0 \leqslant \mathrm{H}(X_1|X_2) \leqslant \mathrm{H}(X_1) \tag{4}$$

$\mathrm{H}(X_1|X_2)$ can have a range (Eq.4) between zero when both variables are completely dependent and marginal entropy $\mathrm{H}(X_1)$ when they are independent. Mutual information, in this field often referred to as transinformation (Eq.5), $\mathrm{T}(X_1,X_2)$, explains the level of dependency and shared information between two variables by considering their joint distribution. ~~The metrics are defined as follows,~~

$$\mathrm{H}(X_1,X_2) = -\sum_{x_1\in X_1}\sum_{x_2\in X_2} p(x_1,x_2)\log_2 p(x_1,x_2)$$

$$\mathrm{H}(X_1|X_2) = -\sum_{x_1\in X_1}\sum_{x_2\in X_2} p(x_1,x_2)\log_2 \frac{p(x_1,x_2)}{p(x_2)}$$

$$0 \leqslant \mathrm{H}(X_1|X_2) \leqslant \mathrm{H}(X_1)$$

$$\mathrm{T}(X_1;X_2) = -\sum_{x_1\in X_1}\sum_{x_2\in X_2} p(x_1,x_2)\log_2 \frac{p(x_1,x_2)}{p(x_1)*p(x_2)} \tag{5}$$

~~, where $p(x_1)$ and $p(x_2)$ constitute the marginal probability distribution of random variable X_1 and X_2, respectively; and $p(x_1,x_2)$ form their joint probability distribution.~~ The assessment of the dependencies beyond three variables can be estimated by the concept of Total Correlation (Eq.6) (proposed by McGill (1954) and named by Watanabe (1960)).

$$\mathrm{C}(X_1,X_2,\ldots,X_n) = \left[\sum_{i=1}^{n} H(X_i)\right] - \mathrm{H}(X_1,X_2,\ldots,X_n) \tag{6}$$

Total Correlation (C) gives the amount of information shared between all variables by taking into account their nonlinear dependencies. C can only be non-negative since sum of all marginal entropies cannot be smaller than their multivariate joint entropy~~(Eq.??)~~, though in the special case of independent variables, C would become zero.

$$\mathrm{C}(X_1, X_2, \ldots, X_n) = \left[\sum_{i=1}^{n} H(X_i)\right] - \mathrm{H}(X_1, X_2, \ldots, X_n)$$

$$\mathrm{H}(X_1, X_2, \ldots, X_n) = -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} \cdots \sum_{x_n \in X_n} p(x_1, x_2, \ldots, x_n) \log_2 p(x_1, x_2, \ldots, x_n)$$

2.2 Understanding and visualizing the measures

In this paper, we argue that, due to the additive properties of information measures, the proposed objectives functions in the literature are unnecessarily complicated, and a single-objective optimization of the joint entropy of all selected sensors will lead to a maximally informative sensor network. The additive relations between some of the information measures discussed in this paper are illustrated in Figure 1. In this figure and later is this paper, we use a shorthand notation: we use the sets of stations directly in the information measures, as a compact notation for the multivariate random variable measured by that set of stations. Various types of information interactions for three variables are conceptually understandable using Venn diagram (Figure 2.a). Although a Venn diagram can be used to illustrate information of more than three variables when they are grouped in three sets (Figure 1), it can't be used to illustrate pairwise information interactions beyond three variables. A chord diagram, on the other hand, can be useful to better understand pairwise information interaction beyond three variables. Figure 2 provides simple template to interpret and compare Venn and chord diagrams.

There are two important caveats with these visualizations. In the general Venn diagram of 3-variate interactions, the "interaction information", represented by the area where 3 circles overlap, can become negative. Hence, the Venn Diagram ceases to be an adequate visualization. For similar reasons, in the chord diagram, the sector size of outer arc lengths should not be interpreted as a total information transferred (Bennett et al., 2019). Information that can contribute to ~~this length~~ interactions is a combination of unique, redundant and synergistic components (Goodwell and Kumar, 2017; Weijs et al., 2018). Their information entanglement is an active area of research in 3 or more dimensions. In this paper, the total size of the outer arc lengths is set to represent the sum of pairwise information interactions (used in Alfonso et al. (2010a)) and conditional entropy of each variable. This size may be larger than the total entropy of the variable and does not have any natural or fundamental interpretation.

In this paper, we use Venn diagrams to illustrate information relations between 3 groups of variables. Group one is the set of all sensors that are currently selected as being part of the monitoring network, which we denote as S. Group two is the set of all sensors that are currently not selected, denoted as F, and group 3 is the single candidate sensor that is currently considered for addition to the network, F_c; see appendix A for an overview of notation. Since group 3 is a subset of group 2, one Venn circle is contained in the other, and there are only 5 distinct areas vs 7 in a general 3-set Venn diagram. In this particular case, there is no issue arising from negative interaction information.

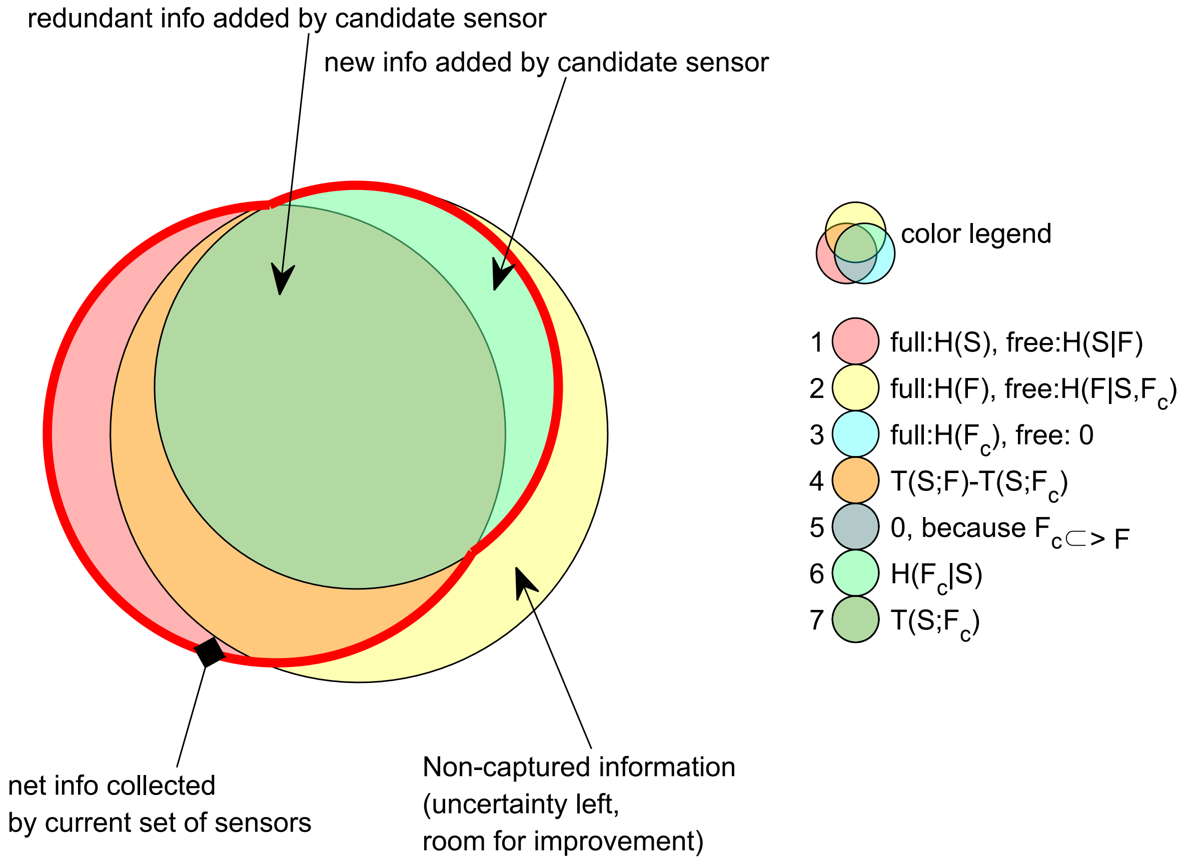

Figure 1. Venn diagram illustrating the relations between the relevant information measures. In the legend, the joint and marginal information-theoretical quantities (joint) entropy H(X), conditional entropy H(X|Y), and transinformation T (X;Y) for the variable from sets of already selected sensors S, not yet selected sensors F and the current candidate sensor F_C are represented by the surfaces in the Venn diagram. For the 3 basic circle colors (first three circles in the legend), "free" gives the quantity represented by the non-covered part and "full" gives the quantity represented by the entire circle surface. The joint entropy that is proposed to be maximized in this paper is the area enclosed in the thick red line.

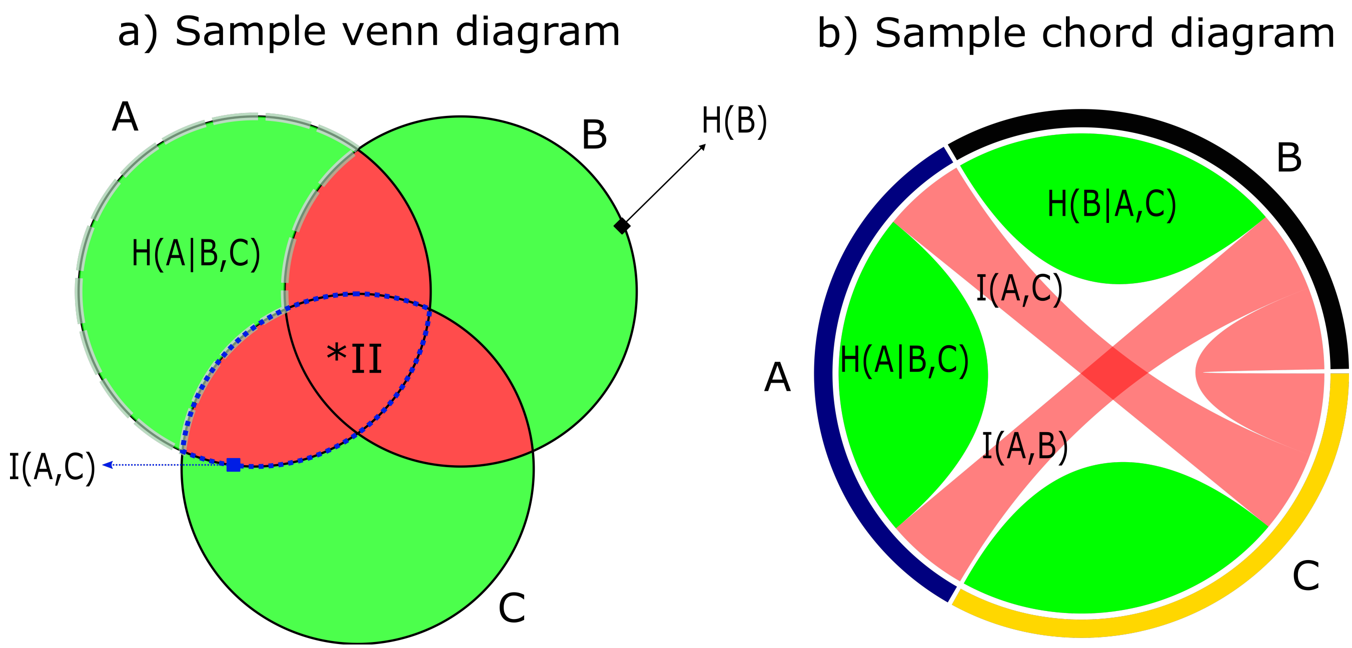

Figure 2. Template illustrations of information interactions with (a) Venn diagram, (b) chord diagram. The green and red areas in both diagrams show a graphical representation of conditional entropy and mutual information respectively. The solid line circles in Venn diagram depict single-variable entropy. *II is information interaction between three variables. The sector size in the outer circle in chord diagram is composed of arcs whose relative lengths correspond to the sum of pairwise information interactions and conditional entropy of each variable, and are best not interpreted.

2.3 Multi-objective optimization

Information theory-based multi-objective optimization methods for monitoring networks have gained significant attention in the literature. 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., 20 . However, there is no consensus on how to minimize redundant information. Table 1 ~~given~~ gives an overview of the large number of objectives and combinations of objectives used in the last decade. On the one hand, water monitoring in polders (WMP) method (Alfonso et al., 2010a) and joint permutation entropy (JPE) method (Stosic et al., 2017) used normalized transinformation to minimize redundant information. While, on the other hand, multi-objective optimization problem (MOOP) method (Alfonso et al., 2010b), Combined regionalization and dual entropy-multi-objective optimization (CRDEMO) method (Samuel et al., 2013), multivariable hydrometric networks (MHN) method (Keum and Coulibaly, 2017) and greedy rank based optimization (GR 5 and 6) method (Banik et al., 2017) adopted total correlation to achieve minimum redundancy. Interestingly, both C and T were used as competing objectives in maximum information minimum redundancy (MIMR) method proposed by Li et al. (2012). They argued that transinformation between selected stations in the optimal set and non-selected stations should be maximized to account for the information transfer ability of a network. Meanwhile, recently proposed methods in the literature attempted to improve monitoring network design by introducing yet other additional objectives (Huang et al., 2020; Wang et al., 2018; Banik et al., 2017; Keum and Coulibaly, 2017). These additional objective are further discussed in ~~the appendix~~Appendix B.

2.4 Single-objective optimization

In this paper, we argue for the Maximum Joint Entropy (maxJE) objective for maximizing the total information collected by a monitoring network. This is equivalent to the GR3 objective proposed by Banik et al. (2017), as part of six other objectives (see appendix B for more detail) proposed in the same paper, which did ~~provide arguments or~~ not provide preference for its use. In ~~this paper, we provide theoretical argument for using single-objective optimization of joint entropy (maxJE) instead of multi-objective optimization. The philosophy of the maxJE objective function is rooted in using multivariate joint entropy defined by Shannon's information theory (1948) for network evaluation. In the~~ the discussion, we argue that a single-objective optimization of the joint entropy of all selected sensors will lead to a maximally informative sensor network. Also, it should be noted that the maxJE objective function ~~indirectly minimizes~~ already penalizes redundant information through its network selection process, which aims ~~for finding~~ to find a new station that produces maximum joint entropy when it is combined with already selected stations in each iteration. ~~This~~ When applied in a greedy search, adding one new station at a time, this approach ranks stations based on growing joint information ~~which is achievable when a~~ as quickly as possible. This is mathematically equivalent to add to the selection, in each iteration, the new station F_C ~~can provide~~ that provides maximum conditional entropy $\cancel{H(S|F_c)}$ $H(F_c|S)$ on top of an already selected set (S) of stations (see Figure 1 for visual illustration). ~~For ease of addition of a single new station to an already existing set, it is therefore mathematically equivalent to maximizing conditional entropy of the new station by selecting it from the pool of non-selected stations F.~~

2.5 Objective functions used in comparison for this study

For the purpose of illustrating the main arguments of this study, we compare maxJE objective function (Eq.7) with three ~~multi-objective optimization~~ other (sets of) objective functions from previously proposed methods: MIMR (Eq.8), WMP (Eq.9) and minT (Eq.10). These methods were chosen since they are highly cited methods in this field, and more importantly, recent new approaches in the literature have mostly been built on one of these methods with additional objectives ~~(see Table 1 for more details of examples)~~Alfonso et al. (2010a, b); Ridolfi et al. (2011); Li et al. (2012); Samuel et al. (2013); Stosic et al. (2017); Keum and C

.

$$\text{Objective function (maxJE)} := \mathit{maximize}\, \mathrm{H}\left(\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle, X_{F_C}\right) \tag{7}$$

$$\text{Objective function (MIMR)} := \begin{cases} \mathit{maximize}\, \mathrm{H}\left(\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle, X_{F_C}\right) \\ \mathit{maximize}\, \sum_{i=1}^{m} \mathrm{T}\left(\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle, X_{F_i}\right) \\ \mathit{minimize}\, \mathrm{C}\left(\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle, X_{F_C}\right) \end{cases} \tag{8}$$

$$\text{Objective function (WMP)}: = \begin{cases} maximize\, \mathrm{H}(F_C) \\ subject\ to\ \sum_{i\in S} \frac{\mathrm{T}(S_i;F_C)}{\mathrm{H}(S_i)} < SBM \end{cases} \tag{9}$$

$$\text{Objective function (minT)}: = \begin{cases} maximize\ first\, \mathrm{H}(F_C) \\ minimize\, \mathrm{T}\left(\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle, X_{F_C}\right) \end{cases} \tag{10}$$

Where $\langle X_{S_1}, X_{S_2}, \ldots, X_{S_k}\rangle$ ~~denotes~~ refers to selected stations in the previous iterations~~, and~~ . X_{F_C} ~~denotes~~ and $\mathrm{H}(F_C)$ denote the variable at the current candidate station and its marginal entropy, respectively. SMB stands for constraint where only stations are considered that are below the median score of all potential stations on that objective. m is equal to the number of non-selected station in each iteration ($m + k = n$ total number of stations). For the multi-objective approaches used in the case study, we used the same weights as the original authors to identify a single solution. It can be seen that a large number of different combinations of information-theoretical metrics are used as objectives.

2.6 Exhaustive search vs greedy add and drop

Apart from the objective function, the optimization of monitoring networks is also characterized by constraints. These constraints can either be implemented for numerical reasons or to reflect ~~the~~ practical aspects of the real world problem. In ~~existing literature~~the majority of existing literature listed in Table 1, one constraint ~~that~~ has often implicitly been imposed~~is that~~ : to treat the selection of stations ~~is greedy~~ as 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. A practical reason for this is numerical efficiency; an exhaustive search of all subsets of ~~k~~ k stations out of ~~n~~ n possible stations will need to consider a large number of combinations~~due to combinatorial explosion. Also in practice, when gradually expanding a network, it may be undesirable to relocate existing stations each time a new station is added~~, 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).

In this paper, for the maximization of joint entropy that we advocate, we will consider and compare 3 ~~constraints that reflect strategies of placement~~cases for constraints with a large influence on computational cost, with the purpose of investigating whether these influence the results. We also interpret the constraints as reflections of placement strategies. Firstly, the ~~"greedy add"~~ “greedy add” strategy is the commonly applied constraint that each time the network expands, the most favorable additional station is chosen, while leaving the already chosen network intact. The optimal network for k stations is found by expanding ~~from 1 station~~ one station at a time. ~~Secondly, "greedy drop"~~This approach can for example be useful in Alpine terrain, where relocating a sensor requires significant effort (Simoni et al., 2011). Secondly, “greedy drop” is the reverse strategy, not previously discussed in literature, where the starting point is the full network with all n stations, and the optimal network for k stations is found by reducing the full network one step at a time, each step dropping the least informative station. Since all of the discussed monitoring design strategies use recorded data and hence discuss networks whose stations are already established, network reduction is perhaps ~~a more realistic scenario~~the more realistic application scenario for information-based design methods. Thirdly, ~~"exhaustive search"~~ "exhaustive search" is the strategy where the optimal network of k stations is found by considering all subsets of k stations out of n. This unconstrained search is far more computationally expensive, and may not be feasible in larger networks for computational reasons~~, or not possible in actual placement strategy for logistical reasons~~. It can therefore be seen as a ~~golden standard~~optimality benchmark. Because all options are considered, this is guaranteed to find the optimal combination, given the objective function. ~~Previous research, such a Fahle et al. (2015); Wang et al. (2018) already discussed sub-optimality of greedy-add. Whether greedy-drop or a combination of the two greedy strategies yields the the fully optimal solutions will be investigated in this paper~~

In this comparison, we will investigate whether the exhaustive optimization yields a series of networks where an increase in network size may also involve relocating stations. This may not always be practically feasible or desired in actual placement strategies, where networks are slowly expanded one station at a time. Occurrence of relocation in the sequence of growing subsets would also show that no greedy algorithm could exist that guarantees optimality.

3 Study area and data description

In previous studies, the focus of the research has been on finding an optimal network for the subject case study ~~without sufficiently addressing~~ with only little discussion on the theoretical justification of applying a new methodology. For this reason~~and~~, the primary goal of this paper ~~, which is highlighting the unnecessary use of multiple~~ is critically discussing the rationale for use of several objective functions in monitoring network design. To illustrate differences between the methods, we decided to apply our methodology ~~in~~ to the Brazos River streamflow network (Figure 3) since this network was subject of study for the MIMR method. ~~To isolate the effect of temporal variability of data~~ This network is under-gauged, according to the World Meteorological Organization density requirement. However, using the exact same case study eliminates the effect of other factors besides the objective function on the comparison. Such factors could be initial network density, temporal, and spatial variability. To isolate our comparison from those effects, as well as from methodological choice such as resolution, time period considered, and quantization method~~on methodology comparison~~, we used the same data period and floor function quantization (Eq.11) proposed by Li et al. (2012). ~~In that paper~~

$$x_q = a\left\lfloor \frac{2x+a}{2a} \right\rfloor \tag{11}$$

Here, a is the histogram bin-width for all intervals except the first one, for which the bin-width is equal to $\frac{a}{2}$. x is station's streamflow value, and x_q is its corresponding quantized value; and $\lfloor\,\rfloor$ is the conventional mathematical floor function.

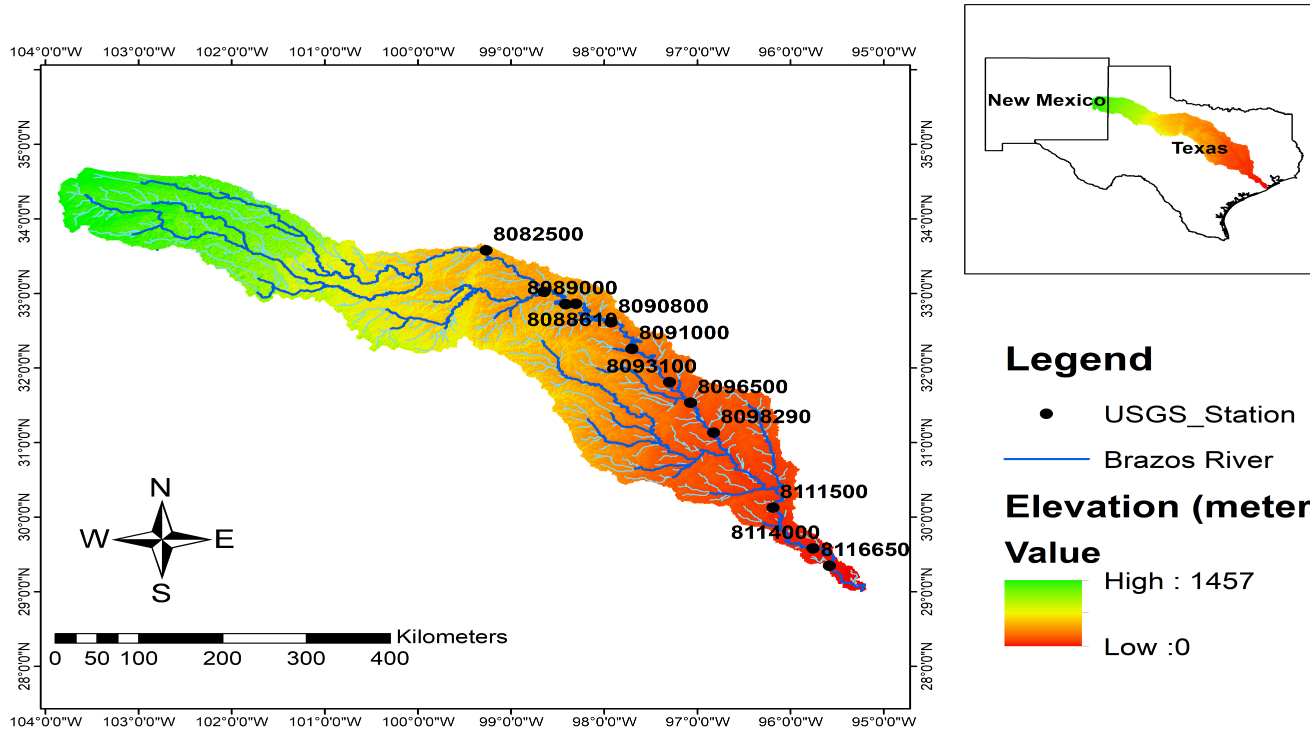

Figure 3. Brazos streamflow network and USGS stream gauges locations.

In Li et al. (2012), 12 USGS stream gauges on the Brazos River were selected for the period of 1990-2009 with monthly temporal resolution; some statistics of the data are presented in Figure 4. For the discretization of the time series, they used a binning approach ~~were~~ where they empirically optimized parameter a to satisfy three goals: (1) to guarantee all 12 stations have distinguishable marginal entropy, (2) to keep spatial and temporal variability of stations' time series, bin-width should be fine enough to capture the distribution of the values in the time series while being coarse enough so that enough data points are available per bin to have a representative histogram, and (3) to prevent rank fluctuation to due to the bin-width assumption, sensitivity analysis must be conducted. They carried out the sensitivity analysis and proposed $a = 150\ m^3/s$ for this case study, the resulting marginal entropy for each station is illustrated in Figure 4.

$$x_q = a \left\lfloor \frac{2x+a}{2a} \right\rfloor$$

~~Where a is histogram bin-width for all intervals except the first one which its bin-width is equal to $\frac{a}{2}$. x is station's streamflow value, and x_q is its corresponding quantized value; and $\lfloor\ \rfloor$ is the conventional mathematical floor function. Brazos streamflow network and USGS stream gauges locations.~~

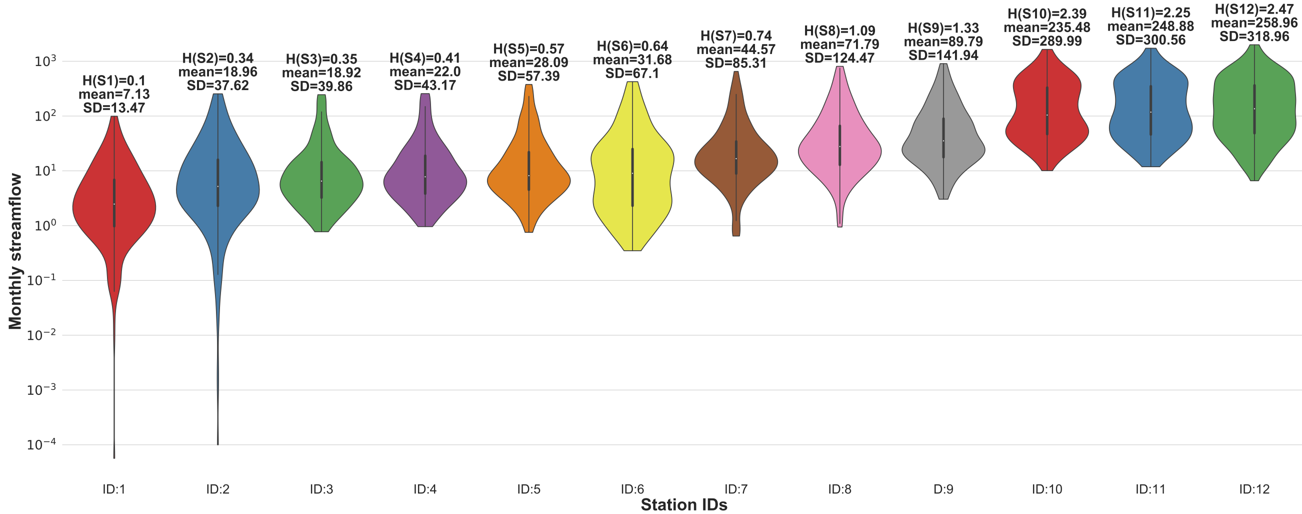

Figure 4. Brazos River streamflow (m^3/s) statistics and resulting entropy values (bits).The stations' IDs are organized from upstream to downstream gauges in the watershed. Entropy values are calculated by floor function and parameter $a = 150\ m^3/s$.

4 Results and Discussion

4.1 Comparison of the objectives for Brazos River case

As indicated in the introduction, we should not attempt to gauge the merits of the objective functions by the intuitive optimality of the resulting network. Rather, the merits of the networks should be gauged by the objective functions. Still, the case study can provide insight in some behaviours resulting from the objective functions.

To assess and illustrate the workings of the different objectives in retrieving information from the water system, we compared three existing methods with a direct maximization of the joint entropy of selected sensors, $\mathrm{H}(S, \mathrm{F}_c)$, indicated with maxJE in the results, such as ~~tables~~ Tables 2 and 3. The joint entropy results in ~~table~~ Table 2 indicate that maxJE is able to find a combination of 8 stations that contains joint information of all 12 stations ranked by other existing methods. ~~We demonstrate that~~ Figure 5 displays spatial distribution of the top 8 stations chosen by different methods. Before any interpretation of the placement, we must note that the choices made in quantization and the availability of data play an important role in the optimal networks identified. Whether the saturation that occurs with 8 stations has meaning for the real world case study depends on whether the joint probability distribution can be reliably estimated. This is highly debatable and merits a separate detailed discussion which is out of the scope of this paper. We present this case study solely to illustrate behaviour of the various objectives.

The most notable difference between MaxJE and the other methods is the selection of all 3 of the stations located most downstream. While other methods would not select these together due to high redundancy between them, maxJE still selects all stations, because despite the redundancy, there is still found to be enough new information in the second-most and third-most downstream station. This can be in part attributed to the quantization choice of equally sized bins throughout the network, leading to higher information contents downstream. While this quantization choice is debatable, it is important, in our opinion, to not compensate artifacts from quantization by modifying the objective function, even if the resulting network may seem more reasonable, but rather to address those artifacts in the quantization choices themselves. To repeat the key point: An objective function should not be chosen based on whether it yields a "reasonable network" but rather based on whether the principles that define it are reasonable.

Though already necessarily true from the formulation of the objective functions, we use the case study to illustrate how 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, given by the yellow~~ 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 (by definition) has the fastest, and minT the slowest rate of reduction of information loss. Methods' 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 $\cancel{\langle X_1, X_2, \ldots, X_{i-1} \rangle}$ $X_1, X_2, \ldots, X_{i-1}$ 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).

Table 2. Resulting maximum joint entropy (bits) for different number of gauges found with different methods for Brazos River case study (JE used exhaustive optimization)

Method	Multivariate dimensions											
	1	2	3	4	5	6	7	8	9	10	11	12
MIMR	2.47	2.84	2.87	3.21	3.23	3.23	3.23	3.32	3.33	3.52	3.93	4.1
WMP1/2	2.47	3.07	3.21	3.36	3.38	3.38	3.38	3.38	3.38	3.52	3.82	4.1
minT	2.47	2.53	2.69	2.72	2.76	2.89	3.06	3.08	3.33	3.52	3.93	4.1
maxJE	2.47	3.07	3.5	3.7	3.88	4.02	4.09	4.1	4.1	4.1	4.1	4.1

Table 3. Optimal gauge orders found with different methods for Brazos River case study.

Method	Station ranking in multivariate dimensions											
	1	2	3	4	5	6	7	8	9	10	11	12
MIMR	12	6	1	8	2	3	4	7	5	9	10	11
WMP1/2	12	9	7	6	5	4	3	2	1	8	11	10
minT	12	1	2	3	4	5	7	6	8	9	10	11
maxJE	12	9	10	11	8	5	7	2*	1*	3*	4*	6*

Note that for the last 5 stations, indicated with *, multiple optimal orders are possible.

~~Approximately proportional Venn diagrams showing the evolution of information measures when progressively (going downwards on the rows) selecting stations (selected station for each step indicated by the numbers) using four different methods (in the different columns). The interpretation of the color-coded areas representing the information measures is the same as in figure 1. All methods select station 12 as the initial station (entropy given by pink circle on row 1). As can be seen from the diagram on the bottom right, the method maximizing joint entropy leaves almost no information unmeasured (yellow part) with just 6 stations, while the other methods still miss capturing this information.~~

~~Evolution of pairwise information interaction between already selected stations in the previous iterations and new proposed station. Green and red links represent proportional conditional entropy and mutual information,respectively. Links with black border emphasizes on the information interaction of new proposed station in each iteration.~~

4.2 Is minimization of dependence needed?

The existing approaches considered above have in common that they all involve some form of dependence criterion to be minimized. ~~Mishra and Coulibaly (2009) stated that "The fundamental basis in designing monitoring networks based the entropy approach is that, the stations should have as little transinformation as possible, meaning that the stations must be independent of each other".~~ For example, the total correlation gives a measure of total redundant information within the selected set. This is information that is duplicated and therefore does not contribute to the total information content of the sensors, which is given by the joint entropy. Focusing fully on minimizing dependence, such as done in the minT objective optimization, makes the optimization insensitive to the amount of non-duplicated information added. This results in many low entropy sensors being selected. It is important to note that the joint entropy already accounts for duplicated information and only quantifies the non-redundant information. This is exactly the reason why it is smaller then the sum of individual entropies. In terms of joint entropy, two completely dependent monitors are ~~not~~ considered to be ~~more informative than~~ exactly as informative as one of them. This means that the negative effect that dependency has on total information content is already accounted for by maximizing joint entropy only.

~~The~~

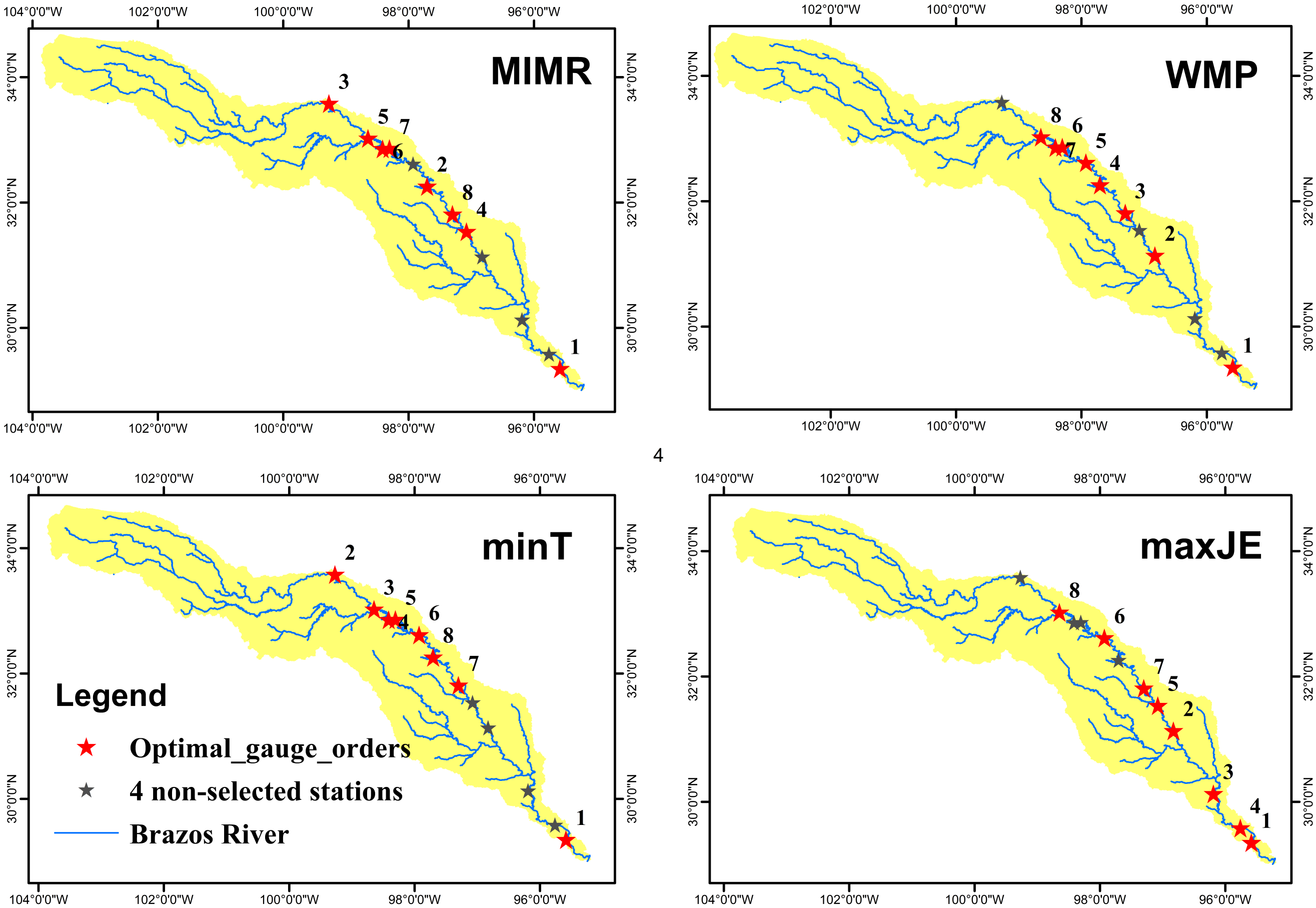

Figure 5. Spatial distribution of the top 8 streamflow gauges ranked different objectives.

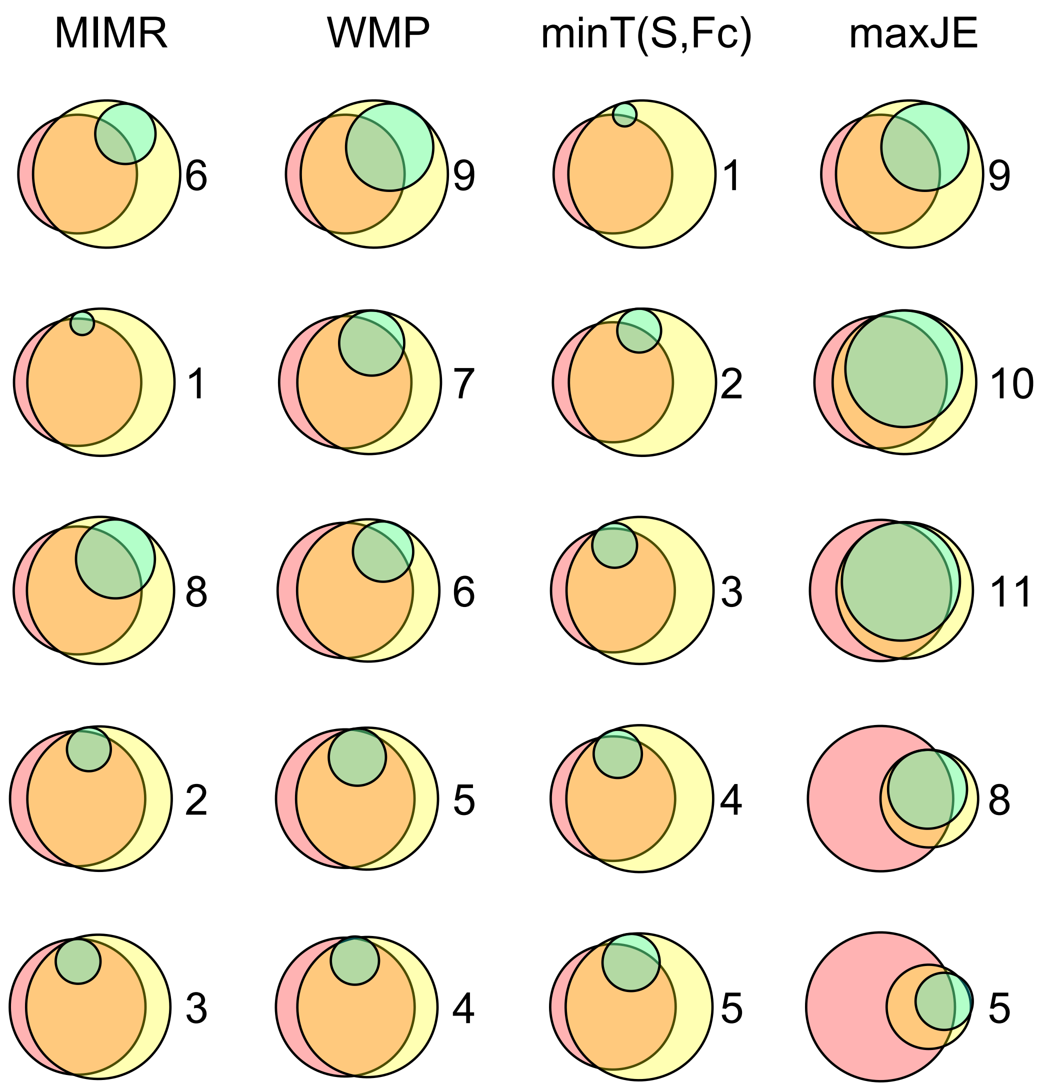

Figure 6. Approximately proportional Venn diagrams showing the evolution of information measures when progressively (going downwards on the rows) selecting stations (selected station for each step indicated by the numbers) using four different methods (in the different columns). The interpretation of the color-coded areas representing the information measures is the same as in figure 1. All methods select station 12 as the initial station (entropy given by pink circle on row 1). As can be seen from the diagram on the bottom right, the method maximizing joint entropy leaves almost no information unmeasured (yellow part) with just 6 stations, while the other methods still miss capturing this information. Exact numbers behind the Venn diagram can be found with the code available with this paper.

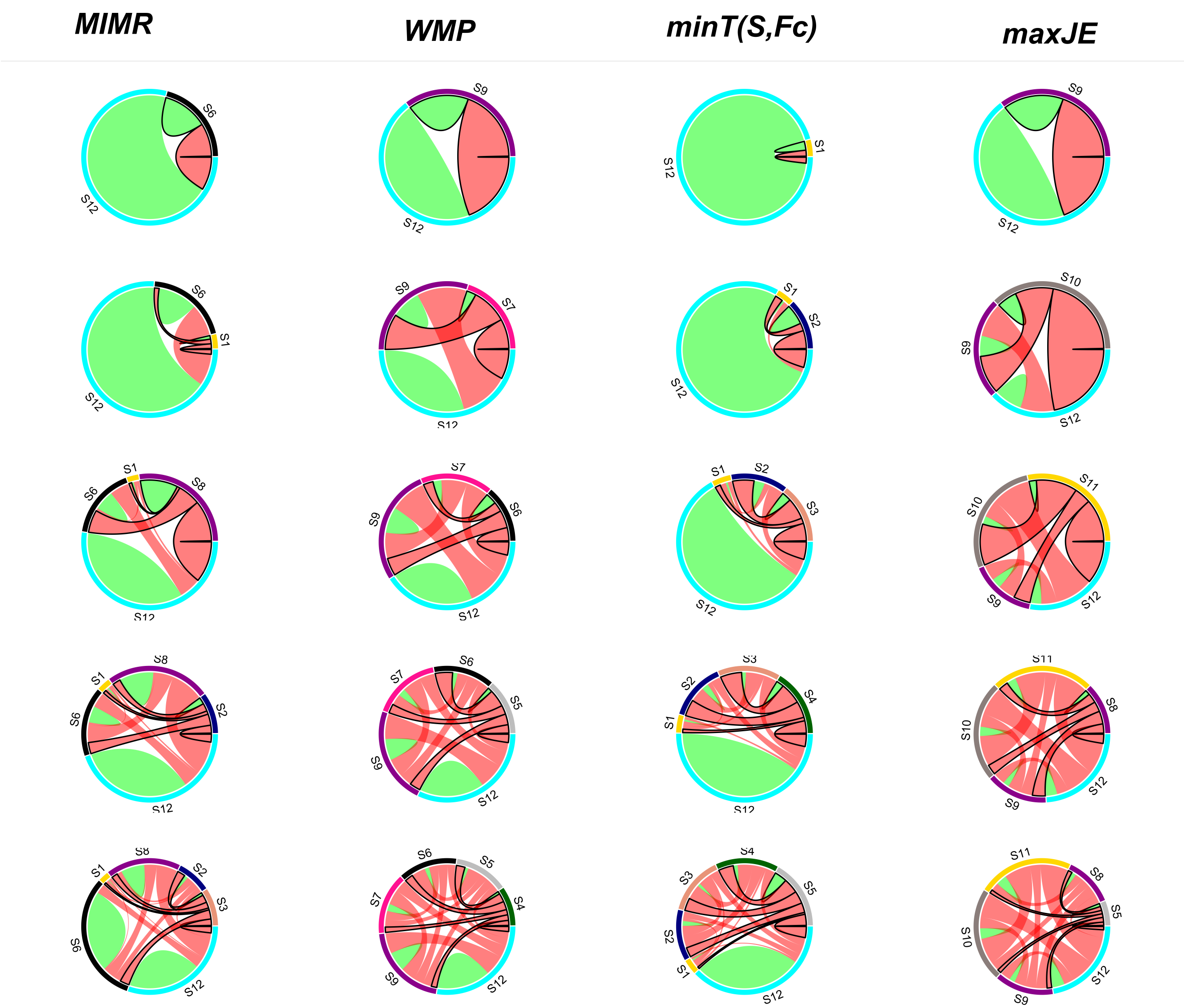

Figure 7. Evolution of pairwise information interaction between already selected stations in the previous iterations and new proposed station. Green and red links represent proportional conditional entropy and mutual information,respectively. Links with black border emphasizes on the information interaction of new proposed station in each iteration.

Mishra and Coulibaly (2009) stated that "The fundamental basis in designing monitoring networks based the entropy approach is that, the stations should have as little transinformation as possible, meaning that the stations must be independent of each other". However, no underlying argument for this fundamental basis is given in the paper. The question is then whether there is another reason, apart from information maximization, why the total correlation should be minimized. In three of the early papers (Alfonso et al., 2010a, b; Li et al., 2012) introducing the approaches that employed or evaluated total correlation, no such reason was given other than the one by Mishra and Coulibaly (2009). Also in later citing research, no such arguments have been found. Traditional reasons for minimizing redundancy are reducing the burden of data storage and transmission, but these are not very relevant in monitoring network design, since those costs are often negligible compared to the costs of the sensor installation and maintenance (see (Barrenetxea et al., 2008; Nadeau et al., 2009; Simoni et al., 2011)). Moreover, information theory tells us that, if needed, redundant information can be removed before transmission and storage by employing data compression. The counter-side of minimal redundancy is less reliability, a far more relevant criterion for monitoring network design. Given that sensors often fail or give erroneous values, one could argue that redundancy (total correlation) should actually be maximized, given a maximum value of joint entropy. We might even want to gain more robustness at the cost of losing some information. One could for example imagine placing a new sensor directly next to another to gain confidence in the values and increase reliability, instead of using it to collect more informative data in other locations.

The Pareto front that would be interesting to explore in this context is the trade-off between ~~maximum~~ *maximum* total correlation (robustness) vs. joint entropy (expected information gained from the network), indicated by the red line in Figure 8. Different points on this Pareto front reflect different levels of trust in the sensors' reliability. Less trust requires more robustness and leads to a network design yielding more redundant information. Previous approaches, such as the MOOP approach proposed by Alfonso et al. (2010b), explore the Pareto front given by the black dashed line, where ~~minimum~~ *minimum* total correlation is conflicting with maximizing joint entropy. As argued in this section, this trade-off is not a fundamental trade-off in information-theoretical terms, but results from the fact that usually there is some redundant information as a by-product of new information, so highly informative stations also carry more redundant information. This redundant information does not reduce the utility of the new information, so does not need to be included as a minimization objective in the optimization.

Summarizing, the maximization of joint entropy while minimizing redundancy is akin to maximizing effectiveness while maximizing a form of efficiency = bits of unique info / bits collected. However, bits collected do not have any significant associated cost. If installing and maintaining a monitoring location has a fixed cost, then efficiency should be expressed as unique information gathered per sensor installed, which could be found by maximizing joint entropy (effectiveness) for a given number of stations, as we suggest in this paper.

~~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 is considered, the pareto front of interest is in the top-right corner.~~

4.3 Greedy algorithms vs. exhaustive optimization of maximum joint entropy

~~For the objective function of maximum joint entropy, we~~ 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 ~~for different numbers of sensors. The greedy-add optimization works by adding one location at a time, maximizing joint entropy, but constrained by having the already selected locations in the new set. The greedy-drop strategy has the same constraint, but~~

~~works backward, starting from the maximum number of sensors and each time dropping the one sensor that retains most joint entropy in~~ in the ~~reduced set~~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.

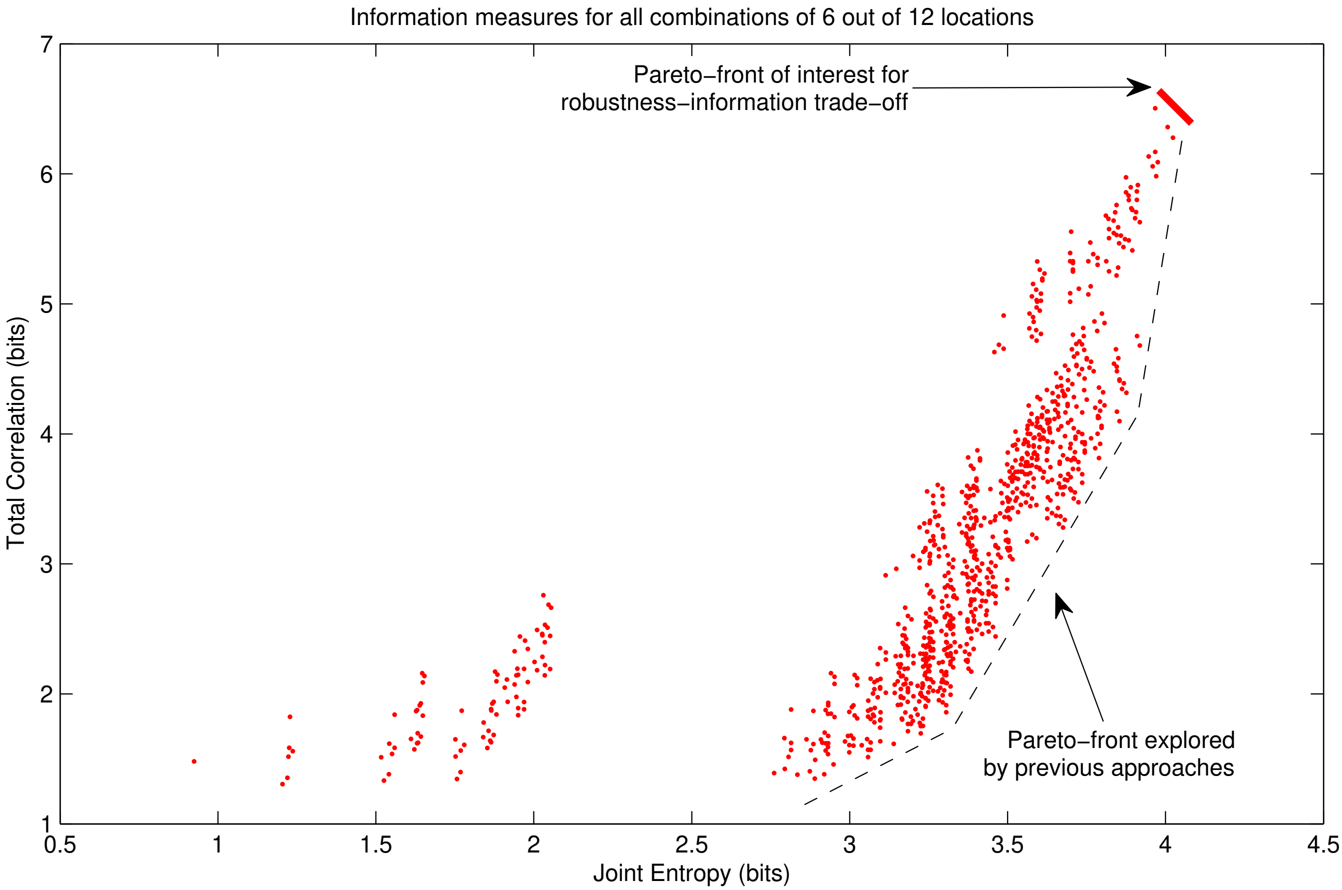

Figure 8. The resulting total correlation and joint entropy for all 924 possible combinations of 6 out of 12 sensor locations. In some past approaches, a Pareto front in the lower right corner is given importance. In this paper, we argue that this trade off is irrelevant, and information can be maximized 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.

The exhaustive optimization tests all possible new combinations when increasing network size by one station, not restricted to those combinations containing the set that was already selected in a smaller network. Since the joint entropy of a set of locations does not depend on the order in which they are added, the number of possible combinations is $\binom{n}{k}$ (i.e. n choose k), where n is the number of potential stations in the pool and k is the number of selected stations. 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. When considering all sub-network sizes, the number of combinations to consider is 2^n, so an exponential growth. We could make an optimistic estimate, only considering the scaling from ~~combinatorial explosion of station sets~~station combinations to evaluate, 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.

Greedy approaches might be candidates for ~~such algorithms. Most of the previous approaches listed in Table 1 can be categorized as greedy optimizations.~~ efficient algorithms. For the proposed joint entropy objective, we tested the optimality of greedy approaches against the benchmark of exhaustive optimization of all possible station combinations. For the Brazos River case study, both the ~~”add” and ”drop”~~ “add” and “drop” greedy selection strategies resulted in the global optimum sets, i.e. the same gauge order and resulting joint entropy as was found by the exhaustive optimization. These results can be read from last row of ~~tables~~ Tables 2 and 3. Therefore, for this case, the greedy approaches did not result in any loss of optimality. For the last few sensors, multiple different optimal sets could be identified, which are detailed in Table 4.4. Results in Table 4.4 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.

In a further test, using artificially generated data, we experimentally falsified the hypothesis that ~~this result is general~~greedy approaches can guarantee optimality. For this test~~.~~, we generated a correlated random ~~gaussian~~ Gaussian dataset for 12 monitors, based on the covariance matrix of the data from the case study. We increased the number of generated observations to 860 ~~timesteps~~time steps, to get a more reliable multidimensional probability distribution. ~~Tables 5 show~~ Table 5 shows the resulting orders for twelve monitors for the three different approaches. Note how for the exhaustive optimization in this example, in some instances~~2~~, one or two previously selected gauges are dropped in favor of selecting ~~3~~ new stations. The resulting joint entropies for the selected sets are shown in Table 6. This means no greedy approach can exist that finds results equivalent to the exhaustive approach.

~~Further research should point out~~ Based on our limited case study, the questions remain open: 1) whether faster algorithms can be formulated that yield guaranteed optimal solutions, and 2) in which cases the greedy algorithm provides a close approximation. It is also possible to formulate modified greedy methods with the ability of replacing a limited number of monitors instead of just adding monitors. This leads to a significantly reduced computational burden compared to exhaustive optimization, while reaching the optimum more often than when adding monitors one at a time. In Table 5, it can be seen that allowing a maximum number of two relocated monitors would already reach the optimal configurations for this ~~case~~specific case. Another limitation of this comparison is that we did not consider metaheuristic search approaches (Deb et al., 2002; Kollat et al., 2008), which fall in between greedy and exhaustive approaches in terms of computational complexity, could serve to further explore the optimality versus computational complexity trade-off. It would be interesting to further investigate what properties in the data drive the sub-optimality of greedy algorithms. Synergistic interactions (Goodwell and Kumar, 2017) are a possible explanation, although our ~~generate~~ generated data example shows that even when moving from 1 to 2 selected stations, a replacement occurs. Since there are only pairs of variables involved, synergy is not needed in the explanation of this behaviour. Rather, the pair with maximum joint entropy does not always include the station with maximum entropy, which ~~is perhaps to~~ could perhaps be too highly correlated with other high entropy variables.

Table 4. ~~Resulting monitor orders~~ All optimal combinations of sensors for ~~random uniform dataset~~the joint entropy objective. For number of sensors above 7, ~~using 12 monitors with 240 data points~~ multiple optimal combinations can be found due to saturation of joint entropy. Black squares are selected sensors.

~~1 2 3 4 5 6 7 8 9 10 11 12Exhaustive31&12 -3*65&7&11 -6* & -12* 2&6 -5*9&12 -11*31058411Greedy Add311176921058124Greedy Drop112672931058411~~

~~Resulting joint entropy for random uniform dataset, using 12 monitors with 240 data points~~

~~1 2 3 4 5 6 7 8 9 10 11 12Exhaustive 1.5383.00254.22515.05815.73226.17176.5156.69476.83186.9337.02377.0834~~

~~Greedy add 1.5382.8984.16355.04315.68116.1116.48586.64626.78926.8996.99567.0834~~

~~Greedy drop 1.52983.00254.22515.04075.72396.17176.5156.69476.83186.9337.02377.0834 Information theory provides a valid framework for monitoring network design, especially when no single users with explicit decision problems can be identified. Within this framework, maximizing the joint entropy is the only objective needed to maximize retrieved information, assuming that this joint entropy can be properly quantified. A large part of the literature on monitoring network design has put much focus on minimizing various pairwise or joint redundancy measures, while this should actually be a secondary objective, that is already considered in the first: maximizing the obtained information. Since this total information is directly given by joint entropy of all selected locations, this measure can be directly optimized.~~

~~The optimal solution for maximizing joint entropy can be found by exhaustively testing all possible combinations of monitors that are feasible given the current network. The number of possible combinations of monitors, however, becomes prohibitively large for more than some 25 possible locations, especially when around half of the stations are selected. Practical constraints on sensor placement may reduce the computational burden somewhat by limiting the combinatorial explosion. One of these constraints could be that the sensors should be placed one by one, each time optimizing the joint entropy. This so-called greedy optimization approach adds the constraint that the chosen set for~~ $n + 1$ ~~stations contains the chosen set for n stations. This approach can for example be useful in Alpine terrain, where relocating a sensor requires significant effort (Simoni et al., 2011). In this work we introduced the "greedy drop" approach that start from the full set and deselects stations one by one. We have demonstrated that the two types of greedy approaches do not always lead to the unconstrained true optimal solution. Synergistic interactions between variable may play a role, although this is not the only possible explanation. In our case study the suboptimality of greedy algorithms was not visible, but we demonstrated its existence with artificially generated data. Differences between exhaustive and greedy approaches were small, especially when using a combination of the greedy add and greedy drop strategy, It remains to be demonstrated in further research how serious this loss of optimality is in practical situations.~~

4.4 ~~further work~~

~~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.~~

~~Numerically, this presents a problem for the calculation of multivariate information measures. Estimating multivariate discrete joint distributions exclusively from data requires quantities of data that exponentially grow with the number of variables, i.e.~~

Table 5. Resulting monitor orders for random uniform dataset, using 12 monitors with 860 data points

Method		Station selection for various network sizes											
		1	2	3	4	5	6	7	8	9	10	11	12
Exhaustive	Added	3	1;12	6	5;7;11	2;6	9;12	3	10	5	8	4	11
	Removed*		3		6; 12	5	11						
Greedy Add		3	11	1	7	6	9	2	10	5	8	12	4
Greedy Drop		1	12	6	7	2	9	3	10	5	8	4	11

* means a previously selected station is removed from optimal set when expanding the network.

Table 6. Resulting joint entropy for random uniform dataset, using 12 monitors with 860 data points

Method	Multivariate dimensions											
	1	2	3	4	5	6	7	8	9	10	11	12
Exhaustive	1.538	3.003	4.225	5.058	5.732	6.172	6.515	6.695	6.832	6.933	7.024	7.083
Greedy add	1.538	2.898	4.164	5.043	5.681	6.111	6.486	6.646	6.789	6.899	6.996	7.083
Greedy drop	1.530	3.003	4.225	5.041	5.724	6.172	6.515	6.695	6.832	6.933	7.024	7.083

5 Conclusions

The aim of this paper was to contribute to better understanding the problem of optimal monitoring network layout using information-theoretical methods. Since using resulting networks and performance metrics from case studies to demonstrate that one objective should be preferred over the other would be circular, the results from our case study served as an illustration of the effects, but not as arguments supporting the conclusions we draw about objective functions. We investigated the rationale for using various multiple-objective and single-objective approaches, and discussed the advantages and limitations of using exhaustive vs. greedy search. The main conclusions for the study can be summarized as follows:

- The purpose of the monitoring network governs which objective functions should be considered. When no explicit information about users and their decision problems can be identified, maximizing the total information collected by the network becomes a reasonable objective. Joint entropy is the only objective needed to maximize retrieved information, assuming that this joint entropy can be properly quantified.

- We argued that the widespread notion of minimizing redundancy, or dependence between monitored signals, as a secondary objective is not desirable and has no intrinsic justification. The negative effect of redundancy on total collected information is already accounted for in joint entropy, which measures total information net of any redundancies.

- When the negative effect on total information is already accounted for, redundant information is arguably beneficial, as it increases robustness of the network information delivery when individual sensors may fail. Maximizing redundancy as an objective secondary to maximizing joint entropy could therefore be argued for, and trade-off between these objectives could be explored depending on the specific case.

- The comparison of exhaustive and greedy search approaches shows that no greedy approach can exist that is guaranteed to give the true optimum subset of sensors for each network size. However, the exponential computational complexity, which doubles the number of sensor combinations to evaluate with every sensor added, makes exhaustive search prohibitive when the number of possible locations become larger than about 25. The complexity of the greedy approach is quadratic in the number of locations, and therefore feasible for large search spaces.

- The constraints to the search space imposed by the greedy approach could also be interpreted as a logistical constraint. In a network expansion scenario, it disallows replacement of stations already selected in the previous iteration.

- We introduced the "greedy drop" approach that starts from the full set and deselects stations one by one. We have demonstrated that the two types of greedy approaches do not always lead to the same result, and neither approach guarantees the unconstrained true optimal solution. Synergistic interactions between variables may play a role, although this is not the only possible explanation. In our case study, the suboptimality of greedy algorithms was not visible in original data, but we demonstrated its existence with artificially generated data. In our specific case studies, differences between exhaustive and greedy approaches were small; especially when using a combination of the greedy add and greedy drop strategy. It remains to be demonstrated in further research how serious this loss of optimality is in a range of practical situations, and how results compare to intermediate computational complexity approaches such as metaheuristic algorithms.

5.1 Further work

In this paper, we focused on the theoretical arguments for justifying the use of various objective functions, 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 tools to find optimal network configurations, we compared greedy and exhaustive search approaches to raise awareness in the scientific community that greedy optimization might fall into 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 that the 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 (for the maxJE objective) and computation cost. Another important question that needs to be addressed in future research is to investigate how the choices and assumptions made (i.e., data quantization which influences probability distribution) in the numerical calculation of objective functions would affect network ranking. What many of these objective functions have in common, is that they rely on multivariate 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, especially in higher dimensions, since data requirements grow exponentially. Also, these probability distributions and the resulting information measures are influenced by multiple factors, including choices about the data's temporal scale and quantization. To have an unbiased comparison framework of objective functions, we kept data and quantization choices from a case study previously described in the literature. It is worth acknowledging that these assumptions, as well as data availability, can greatly influence optimal network ranking, and require more attention in future research.

Numerically, the limited data size in the case study presents a problem for the calculation of multivariate information measures. Estimating multivariate discrete joint distributions exclusively from data requires quantities of data that exponentially grow with the number of variables, i.e. potential locations. When these data-requirements are not met and joint distributions are still estimated directly based on frequencies, independent data will be falsely qualified as dependent and joint information content severely underestimated. This can also lead to apparent earlier saturation of joint entropy, at a relatively low number of stations. For the case study presented here, we do not recommend interpreting this saturation as reaching the number of needed stations, since it could be a numerical artifact. This problem applies to all methods discussed in this paper. Before numerics can be discussed, clarity is needed on the interpretation and choice of the objective function. In other words, before thinking about how to optimize, we should be clear on what to optimize. We hope that this paper helped illuminate this.

Code and data availability. Data and code availability

The code and data that were used to generate the results in this manuscript are available from https://github.com/hydroinfotheory and the USGS https://waterdata.usgs.gov/nwis.

Author contributions. SW conceptualized study, HF and SW jointly performed analysis and wrote manuscript. SW supervised HF

Competing interests. The authors declare no competing interests

Acknowledgements. This research was supported by funding from Hossein Foroozand's NSERC CGS - Doctoral award and Steven V. Weijs's NSERC discovery grant.

Appendix A: Notation and definitions

S	Set of indices of selected monitoring locations
F	Set of indices of potential monitoring locations not yet selected
F_C	The index of the monitoring station currently under consideration for addition
X_S, X_F, X_{F_C}	The (sets of) time series (variables) measured at the monitor(s) in the respective sets
$p(x_1)$	The marginal probability distribution of random variable X_1
$p(x_1, x_2)$	The joint probability distribution of variable X_1 and X_2
$\mathrm{H}(F_C)$	A shorthand for $\mathrm{H}(X_{F_C})$. In information measures, the set is used as shorthand for variables in that set
$\mathrm{H}(X_{F_C})$	The entropy of the marginal distribution of time series in F_C
$\mathrm{H}(X_F \mid X_S)$	The conditional joint entropy of variables in F, given knowledge of variables in S
$\mathrm{T}(X_F; X_S)$	Mutual information or transinformation between set of variables in F and set of variables in S
$\mathrm{C}(X_1, X_2, \ldots, X_n)$	Total Correlation, the amount of information shared between all variables
SMB	Stands for constraint where only stations are considered that are below the median score of all potential stations
a	Histogram bin-width
x	Station's streamflow value
x_q	Quantized value after discretization
AE	Apportionment entropy
RDI	Ranking disorder index
$\mathrm{SI}_z(X)$	Local spatiotemporal information of the grid X in local window z in the time series
$p(\sigma_z)$	Probability distribution of the standard deviation σ_z in time series
$\mathrm{A}_{network}$	Network accuracy
Var	Kriging variance
D	Detection time
$D_{sp}(\gamma)$	The average of the shortest time among the detection times for monitoring station
R	Reliability
δ_s	Binary choice of 1 or 0 for whether the contamination is detected or not

Appendix B: Additional objectives used in recent literature

Recent literature has expanded the ~~information theoretical~~ information-theoretical objectives with additional objectives. For instance: (1) Wang et al. (2018) proposed dynamic network evaluation framework (DNEF) method that follows MIMR method for network configuration in different time windows and optimal network ranking is determined by maximum Ranking disorder index (RDI) (Eq.B2), which is normalized version of apportionment entropy (AE). RDI was proposed by Fahle et al. (2015) and named by Wang et al. (2018) to analyze the uncertainty of the rank assigned to a monitoring station under different time windows; (2) Huang et al. (2020) proposed information content, spatiotemporality, and accuracy (ISA) method, which extends MIMR method by adding two objectives: maximizing spatiotemporality information (SI), and maximizing accuracy (A). The SI (Eq.B4) objective is introduced to incorporate spatiotemporality of satellite data into network design, and A (Eq.B5) objective is proposed to Maximize the interpolation accuracy of the network by minimizing the regional kriging variance; (3) Banik et al. (2017) proposed six combinations (GR 1-6) of four objectives: detection time (D) (Eq.B6), reliability (R) (Eq.B7), H (Eq.~~??~~2) and C (Eq.6) for locating sensors in sewer systems; and (4) Keum and Coulibaly (2017) proposed to maximize conditional entropy as a third objective in dual entropy-multi-objective optimization to integrate multiple networks (in their case: raingauge and streamflow networks). Although maximizing conditional entropy can indirectly be achieved in other used-objective (joint entropy), this new objective gives more preference to maximizing unique information that one network can provide when another network can't deliver. These multi-objective optimization problems are solved by either finding an optimal solution in a Pareto front (Alfonso et al., 2010b; Samuel et al., 2013; Keum and Coulibaly, 2017) or by merging multiple objectives with weight factors into a single objective function (Li et al., 2012; Banik et al., 2017; Stosic et al., 2017).

$$\mathrm{AE} = -\sum_{i=1}^{n} \frac{r_i}{M} \log_2 \frac{r_i}{M} \tag{B1}$$

$$\mathrm{RDI} = \mathrm{nAE} = \tfrac{\mathrm{AE}}{\log_2 n} \tag{B2}$$

Where n is the number of possible ranks that a station can have (i.e., n is equal to the total number of stations). $\frac{r_i}{M}$ ratio is an occurrence probability of the outcome, where M is the number of ranks under different time windows, and r_i is the number of a certain i$^{\text{th}}$ rank. Therefore, AE takes on its maximum value when the ranking probability of a station has equally probable outcome while minimum AE happens when the station's rank is constant. RDI ranges from 0 to 1, and higher RDI values indicate ranking sensitivity of a station to temporal variability of the data.

$$\mathrm{SI}_z(X) = -\sum_{i=1}^{l} p(\sigma_z) \log_2 p(\sigma_z) \tag{B3}$$

$$\mathrm{SI}_{network}(X, \gamma_{F_i}) = \frac{1}{n+1}\left[\sum_{j=1}^{n} SI_z(X_{S_j}) + SI_z(\gamma_{F_i})\right] \tag{B4}$$

$$\mathrm{A}_{network}(X, \gamma_{F_i}) = -\frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{k} Var_{ij} \tag{B5}$$

Where $\mathrm{SI}_z(X)$ is the local spatiotemporal information of the grid X in local window z in the time series; and $p(\sigma_z)$ is probability distribution of the standard deviation σ_z in time series l. $\mathrm{SI}_{network}(X, \gamma_{F_i})$ is spatiotemporal of the network, which is calculated by the average of spatiotemporal information of already selected sites $SI_z(X_{S_j})$ and a potential site $SI_z(\gamma_{F_i})$.

$\mathrm{A}_{network}(X, \gamma_{F_i})$ is network accuracy, and Var is kriging variance over time series l and number of grids k in the study area.

$$D(\gamma) = \frac{1}{S}\sum_{s=1}^{S} D_{sp}(\gamma) \tag{B6}$$

$$R(\gamma) = \frac{1}{S}\sum_{s=1}^{S} \delta_s \tag{B7}$$

Where S is the total number of scenarios considered,and $D_{sp}(\gamma)$ is the average of the shortest time among the detection times for monitoring stations, and δ_s is binary choice of 1 or 0 for whether the contamination is detected or not.

Appendix C: ~~Notation and definitions~~

~~SSet of indices of selected monitoring locations FSet of indices of potential monitoring locations not yet selected F_CThe index of the monitoring station currently under consideration for additionX_S, X_F, X_{F_C}The (sets of) time series (variables) measured at the monitor(s) in the respective sets$p(x_1)$The marginal probability distribution of random variable $X_1$$p(x_1, x_2)$The joint probability distribution of variable X_1 and $X_2$$\mathrm{H}(X_{F_C})$The entropy of the marginal distribution of time series $X_{F_C}$$\mathrm{H}(X_F)$ The joint entropy of the marginal multivariate distribution of variables in $F$$\mathrm{H}(X_F|X_S)$The conditional joint entropy of variables in F, given knowledge of variables in $S$$\mathrm{T}(X_F; X_S)$Mutual information or transinformation between set of variables in F and set of variables in S $\mathrm{C}(X_1, X_2, \ldots, X_n)$Total Correlation, the amount of information shared between all variablesSMB Stands for constraint where only stations are considered that are below the median score of all potential stations aHistogram bin-width$x$Station's streamflow valuex_q Quantized value after discretizationAEApportionment entropyRDIRanking disorder index$\mathrm{SI}_z(X)$ Local spatiotemporal information of the grid X in local window z in the time series$p(\sigma_z)$ Probability distribution of the standard deviation σ_z in time series$\mathrm{A}_{network}$ Network accuracyVar Kriging variance$D$ Detection time$D_{sp}(\gamma)$ The average of the shortest time among the detection times for monitoring station$R$ Reliabilityδ_s Binary choice of 1 or 0 for whether the contamination is detected or not~~

~~This research was supported by funding from Hossein Foroozand's NSERC CGS - Doctoral award and Steven V. Weijs's NSERC discovery grant.~~

References

Alfonso, L., Lobbrecht, A., and Price, R.: Information theory–based approach for location of monitoring water level gauges in polders, Water Resources Research, 46, https://doi.org/10.1029/2009WR008101, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008101, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008101, 2010a.

Alfonso, L., Lobbrecht, A., and Price, R.: Optimization of water level monitoring network in polder systems using information theory, Water Resources Research, 46, https://doi.org/10.1029/2009WR008953, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008953, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008953, 2010b.

Aydin, B. E., Hagedooren, H., Rutten, M. M., Delsman, J., Oude Essink, G. H. P., van de Giesen, N., and Abraham, E.: A Greedy Algorithm for Optimal Sensor Placement to Estimate Salinity in Polder Networks, Water, 11, 1101, https://doi.org/10.3390/w11051101, https://www.mdpi.com/2073-4441/11/5/1101, number: 5 Publisher: Multidisciplinary Digital Publishing Institute, 2019.

Banik, B. K., Alfonso, L., Di Cristo, C., and Leopardi, A.: Greedy Algorithms for Sensor Location in Sewer Systems, Water, 9, 856, https://doi.org/10.3390/w9110856, https://www.mdpi.com/2073-4441/9/11/856, number: 11 Publisher: Multidisciplinary Digital Publishing Institute, 2017.

Barrenetxea, G., Ingelrest, F., Schaefer, G., and Vetterli, M.: The hitchhiker's guide to successful wireless sensor network deployments, in: Proceedings of the 6th ACM conference on Embedded network sensor systems, SenSys '08, pp. 43–56, Association for Computing Machinery, Raleigh, NC, USA, https://doi.org/10.1145/1460412.1460418, https://doi.org/10.1145/1460412.1460418, 2008.

Bayat, B., Hosseini, K., Nasseri, M., and Karami, H.: Challenge of rainfall network design considering spatial versus spatiotemporal variations, Journal of Hydrology, 574, 990–1002, https://doi.org/10.1016/j.jhydrol.2019.04.091, http://www.sciencedirect.com/science/article/pii/S0022169419304317, 2019.

Bennett, A., Nijssen, B., Ou, G., Clark, M., and Nearing, G.: Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models, Water Resources Research, 55, 4613–4629, https://doi.org/10.1029/2018WR024555, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024555, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR024555, 2019.

Bernardo, J. M.: Expected Information as Expected Utility, The Annals of Statistics, 7, 686–690, https://doi.org/10.1214/aos/1176344689, https://projecteuclid.org/euclid.aos/1176344689, publisher: Institute of Mathematical Statistics, 1979.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 182–197, https://doi.org/10.1109/4235.996017, conference Name: IEEE Transactions on Evolutionary Computation, 2002.

Fahle, M., Hohenbrink, T. L., Dietrich, O., and Lischeid, G.: Temporal variability of the optimal monitoring setup assessed using information theory, Water Resources Research, 51, 7723–7743, https://doi.org/10.1002/2015WR017137, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017137, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR017137, 2015.

Foroozand, H. and Weijs, S. V.: Entropy Ensemble Filter: A Modified Bootstrap Aggregating (Bagging) Procedure to Improve Efficiency in Ensemble Model Simulation, Entropy, 19, 520, https://doi.org/10.3390/e19100520, https://www.mdpi.com/1099-4300/19/10/520, number: 10 Publisher: Multidisciplinary Digital Publishing Institute, 2017.

Foroozand, H., Radić, V., and Weijs, S. V.: Application of Entropy Ensemble Filter in Neural Network Forecasts of Tropical Pacific Sea Surface Temperatures, Entropy, 20, 207, https://doi.org/10.3390/e20030207, https://www.mdpi.com/1099-4300/20/3/207, number: 3 Publisher: Multidisciplinary Digital Publishing Institute, 2018.

Goodwell, A. E. and Kumar, P.: Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables, Water Resources Research, 53, 5920–5942, https://doi.org/10.1002/2016WR020216, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR020216, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR020216, 2017.

Huang, Y., Zhao, H., Jiang, Y., and Lu, X.: A Method for the Optimized Design of a Rain Gauge Network Combined with Satellite Remote Sensing Data, Remote Sensing, 12, 194, https://doi.org/10.3390/rs12010194, https://www.mdpi.com/2072-4292/12/1/194, number: 1 Publisher: Multidisciplinary Digital Publishing Institute, 2020.

Keum, J. and Coulibaly, P.: Information theory-based decision support system for integrated design of multivariable hydrometric networks, Water Resources Research, 53, 6239–6259, https://doi.org/10.1002/2016WR019981, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019981, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019981, 2017.

Khorshidi, M. S., Nikoo, M. R., Taravatrooy, N., Sadegh, M., Al-Wardy, M., and Al-Rawas, G. A.: Pressure sensor placement in water distribution networks for leak detection using a hybrid information-entropy approach, Information Sciences, 516, 56–71, https://doi.org/10.1016/j.ins.2019.12.043, http://www.sciencedirect.com/science/article/pii/S0020025519311545, 2020.

Kollat, J. B., Reed, P. M., and Kasprzyk, J. R.: A new epsilon-dominance hierarchical Bayesian optimization algorithm for large multiobjective monitoring network design problems, Advances in Water Resources, 31, 828–845, https://doi.org/10.1016/j.advwatres.2008.01.017, http://www.sciencedirect.com/science/article/pii/S0309170808000298, 2008.

Krstanovic, P. F. and Singh, V. P.: Evaluation of rainfall networks using entropy: I. Theoretical development, Water Resources Management, 6, 279–293, https://doi.org/10.1007/BF00872281, https://doi.org/10.1007/BF00872281, 1992.

Li, C., Singh, V. P., and Mishra, A. K.: Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy, Water Resources Research, 48, https://doi.org/10.1029/2011WR011251, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011251, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR011251, 2012.

McGill, W. J.: Multivariate information transmission, Psychometrika, 19, 97–116, https://doi.org/10.1007/BF02289159, https://doi.org/10.1007/BF02289159, 1954.

Mishra, A. K. and Coulibaly, P.: Developments in hydrometric network design: A review, Reviews of Geophysics, 47, https://doi.org/10.1029/2007RG000243, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007RG000243, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007RG000243, 2009.

Mogheir, Y., Singh, V. P., and de Lima, J. L. M. P.: Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine, Hydrogeology Journal, 14, 700–712, https://doi.org/10.1007/s10040-005-0464-3, https://doi.org/10.1007/s10040-005-0464-3, 2006.

Nadeau, D. F., Brutsaert, W., Parlange, M. B., Bou-Zeid, E., Barrenetxea, G., Couach, O., Boldi, M.-O., Selker, J. S., and Vetterli, M.: Estimation of urban sensible heat flux using a dense wireless network of observations, Environmental Fluid Mechanics, 9, 635–653, https://doi.org/10.1007/s10652-009-9150-7, https://doi.org/10.1007/s10652-009-9150-7, 2009.

Neumann, J. v. and Morgenstern, O.: Theory of Games and Economic Behavior, Princeton University Press, 1953.

Raso, L., Weijs, S. V., and Werner, M.: Balancing Costs and Benefits in Selecting New Information: Efficient Monitoring Using Deterministic Hydro-economic Models, Water Resources Management, 32, 339–357, https://doi.org/10.1007/s11269-017-1813-4, https://doi.org/10.1007/s11269-017-1813-4, 2018.

Ridolfi, E., Montesarchio, V., Russo, F., and Napolitano, F.: An entropy approach for evaluating the maximum information content achievable by an urban rainfall network, Natural Hazards and Earth System Sciences, 11, 2075–2083, https://doi.org/https://doi.org/10.5194/nhess-11-2075-2011, https://www.nat-hazards-earth-syst-sci.net/11/2075/2011/, publisher: Copernicus GmbH, 2011.

Samuel, J., Coulibaly, P., and Kollat, J.: CRDEMO: Combined regionalization and dual entropy-multiobjective optimization for hydrometric network design, Water Resources Research, 49, 8070–8089, https://doi.org/10.1002/2013WR014058, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013WR014058, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013WR014058, 2013.

Shannon, C. E.: A mathematical theory of communication, The Bell System Technical Journal, 27, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x, conference Name: The Bell System Technical Journal, 1948.

Simoni, S., Padoan, S., Nadeau, D. F., Diebold, M., Porporato, A., Barrenetxea, G., Ingelrest, F., Vetterli, M., and Parlange, M. B.: Hydrologic response of an alpine watershed: Application of a meteorological wireless sensor network to understand streamflow generation, Water Resources Research, 47, https://doi.org/10.1029/2011WR010730, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR010730, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR010730, 2011.

Stosic, T., Stosic, B., and Singh, V. P.: Optimizing streamflow monitoring networks using joint permutation entropy, Journal of Hydrology, 552, 306–312, https://doi.org/10.1016/j.jhydrol.2017.07.003, http://www.sciencedirect.com/science/article/pii/S0022169417304547, 2017.

Wang, W., Wang, D., Singh, V. P., Wang, Y., Wu, J., Wang, L., Zou, X., Liu, J., Zou, Y., and He, R.: Optimization of rainfall networks using information entropy and temporal variability analysis, Journal of Hydrology, 559, 136–155, https://doi.org/10.1016/j.jhydrol.2018.02.010, http://www.sciencedirect.com/science/article/pii/S0022169418300866, 2018.

Watanabe, S.: Information Theoretical Analysis of Multivariate Correlation, IBM Journal of Research and Development, 4, 66–82, https://doi.org/10.1147/rd.41.0066, conference Name: IBM Journal of Research and Development, 1960.

Weijs, S. V., Foroozand, H., and Kumar, A.: Dependency and Redundancy: How Information Theory Untangles Three Variable Interactions in Environmental Data, Water Resources Research, 54, 7143–7148, https://doi.org/10.1029/2018WR022649, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022649, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR022649, 2018.

Table 1. Various information-theoretical objectives used by methods proposed in recent literature.

Method	Reference	Objective Function								
		$\mathrm{H}(F_C)$	SBM	$\mathrm{H}(S,F_C)$	C(S)	$\mathrm{T}(F_C;S)$	$\mathrm{C}(F_C;S)$	$\mathrm{PE}(F_C;S)$	$\mathrm{H}(Q\|P)$	Others
WMP1	Alfonso et al. (2010a)	max	W_1							
WMP2	Alfonso et al. (2010a)	max	W_2							
WMP3	Alfonso et al. (2010a)	max	W_3							
MOOP	Alfonso et al. (2010b)			max			min			
minT	Ridolfi et al. (2011)	max 1^{st}				min				
MIMR	Li et al. (2012)			λ_1	$-(1-\lambda_1)$	λ_1				
CRDEMO	Samuel et al. (2013)			max			min			
JPE	Stosic et al. (2017)					min		max		
MHN	Keum and Coulibaly (2017)			max			min		max	
GR1	Banik et al. (2017)									min D
GR2	Banik et al. (2017)									max R
GR3	Banik et al. (2017)			max						
GR4	Banik et al. (2017)									min D & max R
GR5	Banik et al. (2017)			max			min			
GR6	Banik et al. (2017)			max			min			min D & max R
DNEF	Wang et al. (2018)			λ_1	$-(1-\lambda_1)$	λ_1				max RDI
ISA	Huang et al. (2020)			λ_1	$-(1-\lambda_1)$	λ_1				max SI & max A
maxJE	this paper			max						

WMP objectives: $W_1 = \sum_{i\in S}\mathrm{T}(S_i;F_C)$, $W_2 = \sum_{i\in S}\frac{\mathrm{T}(S_i;F_C)}{\mathrm{H}(S_i)}$, $W_3 = \sum_{i\in S}\frac{\mathrm{T}(S_i;F_C)}{\mathrm{H}(F_C)}$

The table shows whether an objective is maximized (max) or minimized (min) or forms part of a weighted objective function that is maximized with weights λ. SBM stands for constraint where only stations are considered that are below the median score of all potential stations on that objective. D is detection time, and R is reliability. RDI stands for ranking disorder index. SI is spatiotemporality information, and A is accuracy. WMP=Water Monitoring in Polders; MOOP = Multi Objective Optimization Problem; minT = Minimum Transinformation; MIMR = Maximum Information Minimum Redundancy; CRDEMO = Combined Regionalization and Dual Entropy-Multi-objective Optimization; JPE = Joint Permutation Entropy; MHN = Multivariable Hydrometric Networks; GR = Greedy Rank; DNEF = Dynamic Network Evaluation Framework; ISA = information content, spatiotemporality and accuracy ; and maxJE = Maximum Joint Entropy.