Dear Reviewer,

Thank you for taking the time to review our manuscript, and the useful suggestions and comments, which we will address in the revised manuscript. Please find our answers to your comments below.

Original comment: *The work seems to bring causality research in AI (more specifically Pearl, 2009) into hydrologic analysis. While I am no expert on causality analysis, it occurs to me there is some novelty in the authors' valiant effort in venturing into this realm alone and presenting a stab for hydrology, but there are also some concerns regarding clearly defining the real merit of the method. If the authors call for more research in this direction, the limitations and potential should be carefully discussed. The grand goal of the paper was to "learn causality", but the reality is that this is still very difficult from purely data-driven basis. I personally appreciate such explorations and think this concept is new to hydrology. I think the paper can be considered for publication after some substantial revisions. However, the authors will have to carefully qualify the applicability and limitations of the technique.*

Answer: Thank you for your assessment of our work. We will expand the discussion of merit, applicability and limitations of the methodology in the revised manuscript (see our answers below, in particular the next one).

Original comment: *The most important issue — as far as I can read, the key appears to be defining a sufficient set, which requires lots of subjective decisions and prior assumptions. The authors included previous-day precipitation and previous-day soil moisture because they think these variables will influence today's soil moisture. Also included are precipitation, daily temperature, humidity, wind. By the time you are done providing the sufficient set, you already need to inject lots of knowledge. We might wonder why we still need to run this causality test in the first place. I do see the point – some of the decisions can be based on prior knowledge while the main causality gradient of interest (is soil moisture leading to more rainfall) may be unclear from our prior knowledge. This raises two issues: (i) there is only a niche of questions where this approach is meaningful: where we know enough to identify a causal graph and a sufficient set, but do not know the answer to the main question. This niche does exist; (ii) it will be much harder to apply where the causality or even the important factors are unknown, so the sales language of "learning a causality link" does not fit reality and should be carefully qualified.*

Answer: It is correct that the proposed methodology assumes that a sufficient set can be identified. Identifying such a sufficient set requires knowledge on the existence of causal links between variables, but not on the strength or sign of these links, which can then be determined with the proposed methodology. In our opinion, a sufficient set can be identified for many geoscientific questions, thus, the proposed methodology may provide many new insights into the Earth system.
In the revised manuscript, we will provide guidelines for choosing sufficient sets for different geoscientific questions and improve the description of the choice of input variables in the example of soil moisture-precipitation coupling in order to make this process more comprehensible. The idea is to start with a set of causal parents of the considered variable, i.e. any set of other variables suitable to determine the considered variable, which always forms a sufficient set. For example, for current soil moisture, causal parents could be previous soil mositure, precipitation, evaporation and runoff. There are *two* sources of errors in the proposed methodology that are the approximation of a sufficient set *and* the approximation of the function in Equation 13 of the submitted manuscript, which maps the input variables to the conditional expectation of the target variable given the input variables. Therefore the identified set might still have to be modified, but we believe that it provides a good starting point and makes the choice of additional input variables more comprehensible. There are still subjective decisions, but these decisions and resulting assumptions are clearly communicated by the causal graph. We will expand the discussion of these aspects in the revised manuscript.
When a sufficient set cannot be identified from prior knowledge because "causality or even the important factors are unknown", methods from causal discovery (Guo 2020) might be used to identify a

causal graph (and then a sufficient set). While this is not topic of the manuscript, we will add this in the discussion of the revised manuscript.

Original comment: *As an initial demonstration the study also lacked a control experiment. In other words, if you replace today's soil moisture with a potential highly-correlated confounder, will the analysis show it is non-causal? This has not been demonstrated.*

Answer: We agree that control experiments are important for any novel methodology. However, constructing control experiments for the proposed methodology is difficult due to the following reasons. From a theoretical point of view, the proposed methodology will always identify the causal impact of e.g. soil moisture on precipitation including potential errors. These errors result from an incomplete or incorrect sufficient set, and errors in the approximation of the function in Equation 13 of the submitted manuscript, which maps the input variables to the conditional expectation of the target variable given the input variables. Both errors are expected to vary when replacing soil moisture by a different variable or considering other relations than soil moisture-precipitation coupling, because the additional input variables may no longer form a sufficient set, and because the function in Equation 13 will be different. Therefore, defining a control experiment, which confirms that the methodology works for the considered example is not possible. Instead, we performed additional analyses to assess the correctness of obtained results (Section 4), which indicate that the results do indeed not only reflect correlations, but causal relations between soil moisture and precipitation.
We performed a very simple control experiment (not mentioned in the manuscript), where we replaced the target variable precipitation by random noise. As expected from the missing correlations between soil moisture and random noise, the methodology identified no causal impact of soil moisture on the target variable in this case. We will briefly mention this finding in the revised manuscript.

Original comment: *there should be a simple logical explanations for Theorem 1. I mean, the mathematical form can be accurate but does not help many people to understand the logic. You should translate this into simple, ordinary language. I don't believe the underlying logic is that remote.*

Answer: To understand the rationale of Theorem 1, it is necessary to understand when confounding happens, i.e. when the expected value of variable Y given X and some other variables $\{\mathrm{C}_i\}_{i=1}^{k}$ does not reflect the causal impact of X on Y. There are two cases where this happens, illustrated by their simplest examples in Figure 1 (of this answer). In the first example, there is no causal impact of X on Y, but there is a variable N affecting X and Y and leading to a spurious relation between X and Y. If we exclude N when studying the causal impact of X on Y, a causal impact will (erroneously) be identified because X can be used to make inference about the value of Y (by first making inference about the value of N). This can be prevented by including N as an additional input variable C_1. If N does not directly affect Y, but affects another variable M, which then affects Y, one can prevent confounding by including N or M as additional input variable, and so forth. These examples are covered by the second condition in the definition of a sufficient set (see lines 139 to 142 of the submitted manuscript).
In the second example, there is a causal impact of X on Y via variable D. The same reasoning applies as above: if D is excluded when studying the causal impact of X on Y, the causal impact will correctly be identified because X is used to make inference about the value of Y (by first making inference about the value of D). On the other hand, when D is included, no causal impact of X on Y will be identified (which is wrong). This example is covered by the first condition in the definition of a sufficient set. We will clarify this in the revised manuscript and add references to (Pearl 2009a) and (Pearl 2009b), where this is discussed in detail.

Original comment: *the Methods and Results are intermingled in an unhelpful way. Try to have more clear sections with dedicated functions.*

Answer: We assume that the Reviewer is refering to the additional analyses described in Section 4.

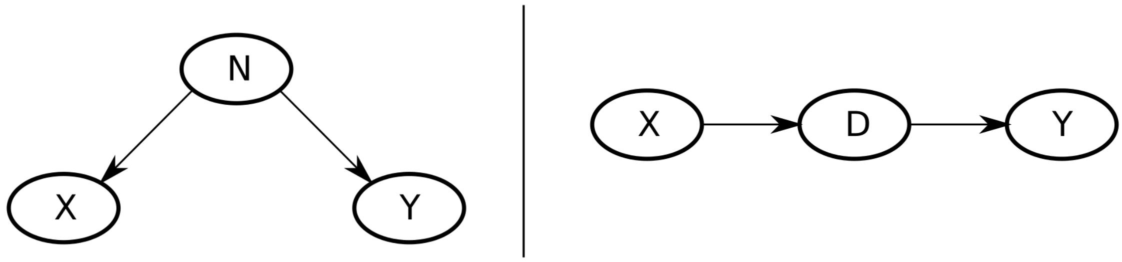

Figure 1: Examples where confounding happens. Left: if we exclude N when studying the causal impact of X on Y, a causal impact of X on Y will be identified although there is only a spurious relation between X and Y (via the confounder N). Right: if we include D when studying the causal impact of X on Y, no causal impact of X on Y will be identified although there is one. See text for more details.

Currently, the manuscript is structured as follows: Section 2.1 gives the required causal background, Section 2.2 describes the general methodology, and Section 3 describes the application of the general methodology to the example of soil moisture-precipitation coupling. Within Sections 2.2 and 3, we proceed along the steps of the methodology, i.e. we first detail the training procedure for a causal model and then the sensitivity analysis of the trained model. Section 4 describes "Further analyses to assess the correctness of obtained results", both in general and for the considered example. Finally, Section 5 compares the results obtained for the considered example with results obtained from a linear correlation analysis.
We decided for this structure, because we believe that first describing the general methodology *and* the further analyses for assessing the correctness of obtained results in general in Section 2, and only then introducing the application of the general methodology to the illustrative example will make Section 2 even harder to digest for people new to causality. Moreover, we believe that the additional analyses described in Section 4 are easier to understand when directly describing the respective analysis in the case of the considered example. The results of the additional analyses are only briefly mentioned in the subsection describing the analysis because the focus of this manuscript is the methodology and not the results. Nevertheless, we may change the structure of the revised manuscript if this is desired.

Original comment: *By the time I reach section 4 I am totally tired and cannot understand the rather complicated logic. Can you make this simpler?*

Answer: We will revise the manuscript and make a serious effort to improve the readability to help the reader.

Original comment: *How does the UNet represent the causal links in Figure 2? To my understanding all the inputs were treated in the same way.*

Answer: We assume that the Reviewer is referring to Figure 5. It is correct that all inputs are treated in the same way. The causal graph in Figure 5 is used to find a sufficient set of input variables in addition to soil moisture (and to communicate the assumptions underlying the methodology in the illustrative application to soil moisture-precipitation coupling), such that we can apply Theorem 1. We will clarify this in the revised manuscript.

Original comment: *define "blocking a path"*

Answer: "Blocking a path" is defined in lines 146 to 148 of the submitted manuscript. We described the intuitive meaning in our answer related to the comment on Theorem 1 above and will add it in the revised manuscript.

Original comment: *line 204 "further input variables" like what?*

Answer: "Further input variables" forming a sufficient set. Our particular choice is described in Section 3.3, "Choice of input variables". We will add a reference to Section 3.3 in line 204 of the revised manuscript to prevent confusion.

Original comment: *Page 6 needs lot of plain-language explanations.*

Answer: Page 6 is the page with Theorem 1 and the definition of a sufficient set. In the revised manuscript, we will add plain language explanations as detailed in our answer concerning the comment on Theorem 1 above.

Original comment: *don't understanding "By including antecedent precipitation as input variable, or, in other words, conditioning on antecedent precipitation, we can exclude this correlation from our analysis."*

Answer: Please see our answer concerning your comment on Theorem 1 above, and replace X by soil moisture, Y by subsequent precipitation and N by antecedent precipitation. In this example, antecedent precipitation has a confounding effect when analyzing soil moisture-precipitation coupling, which can be avoided by including antecedent precipitation as an additional input variable.

Sincerely,

Tobias Tesch

References

Guo R, Cheng L, Li J, Hahn P R, and Liu H (2020) A Survey of Learning Causality with Data: Problems and Methods. ACM Comput. Surv. 53, 4, Article 75 (July 2021), 37 pages. https://doi.org/10.1145/3397269

Pearl J (2009a) Causal inference in statistics: An overview, Statistics Surveys, 3, https://doi.org/10.1214/09-ss057

Pearl J (2009b), Excerpts from the 2nd edition of Causality (Cambridge University Press), Chapter 11.3.1, http://bayes.cs.ucla.edu/BOOK-09/ch11-3-1-final.pdf