Author Response – EGU Sphere – 2022-617

Developing a Bayesian network model for understanding river catchment resilience under future change scenarios

Kerr J. Adams, Christopher (Kit) A. J. Macleod, Marc J. Metzger, Nicola Melville, Rachel C. Helliwell, Jim Pritchard, Miriam Glendell

Referee Comments #1

Dear Dr Laura Uusitalo,

Many thanks for taking the time to review our manuscript and provide constructive comments to help improve its content. Included in Table 1 is our response and details on how we have addressed each of your comments and questions. We have updated the manuscript and supporting material as appropriate and provided updated versions of both. All updates to the manuscript and supporting material documents will be highlighted in yellow, while Table 1 includes the line number for where updated material starts and ends.

We hope our latest edits improve the manuscript and address your questions.

With best wishes,

Kerr Adams, on behalf of authors.

Table 1: Responses to referee #1 comments and questions

Referee Comment	Author Response
1) I think a picture of the model should be presented. I understand it can be complex, but I also understand it was presented for the stakeholders in the workshops, so it should be possible to present it also in the paper, or at the minimum in the supplement. It would make it easier for the reader to understand the model.	We agreed that a representation of the model would be beneficial for the reader and have included a simplified visualisation of the model in the supplementary material, now S3, Figure S2, which we refer to in lines 303-304 in the manuscript. Figure S2 includes an example of the future Business as Usual scenario being performed by the model to represent the different continuous and discrete variables and how we used both to compare current future scenarios. Our model contains 417 nodes, 623 arcs and 23 sub-models. Despite not being a spatial model, there are some geographical considerations included to represent the sub-catchment scale and individual wastewater assets, which results in repetition of nodes, arcs and sub-models. These geographical considerations do make it complex to represent the full model visually, which is why we decided to include a simplified version in the supplementary material. We've highlighted where there is repetition in the supporting text box. We add information regarding the complexity of the model in lines 200-204 of the manuscript. The GeNIe software was effective for building the conceptual model during focus groups with

	each sub-system group. When presenting the full model, it was difficult for stakeholders to follow and comment on important variables and cause and effect relationships, as is evident in Figure S2. We therefore used simplified versions, such as in Figure 4 of the manuscript to visually represent the model, giving stakeholders the opportunity to input their opinion on the model structure during workshops using the collaborative software Miro, then used the feedback to update the model in GeNIe. We added this context to the discussion in manuscript lines 505-511.
2) It seems from the supplement that the model was parameterized using deterministic equations. Usually Bayesian Networks are use specifically to model also the uncertainty that is related to the model parameters. Please discuss this and explain your modelling choice.	Where we do have data available to represent uncertainty, we fitted a truncated normal prior distributions - donated as β in equations represented in the supplementary material TableS3 - to the available data by calculating the mean and standard deviation from the small number of available values. Truncated normal distributions were fitted to avoid negative values, where appropriate. Secondly, where longer data records were available, we used a built in GeNIe function to fit a custom prior distribution (histogram) to time-series data , such as surface water flows. Where available data was limited to a single deterministic value and statistical moments could not be calculated, we applied scenario modelling using the diverse coupled future pathways as a best available method for representing uncertainty. We have included details of our use of fitted truncated normal prior distributions and scenario modelling in lines 251-257 of the manuscript.
3) The use of simulations to evaluate the results is a bit unclear. We don't usually use simulations as such to evaluate the outputs of a BN, but we aim to compute the total probability distribution over the modelled domain, given the conditional probability distributions and the model structure. This way, we can then reason "backwards" (what is the most probable cause given the consequences), compute the probabilities of outcomes given a number of causes or observations, etc. In the case of discrete models, this can be done analytically, and in the case of continuous models, the distributions are often approximated using simulations, but BNs are not usually simulated as such. When continuous BNs are run/solved, often using Monte Carlo Markov chain computation, the early part of the Markov	Many thanks for highlighting our confusing use of the term ‘simulations’, we have updated the manuscript to replace *simulation* with *scenario* and *samples* where appropriate throughout as our results are describing outputs comparisons for both the current and future (coupled RCP and SSP) scenarios. The modelling technique we use is a **hybrid forward sampling algorithm**, which is the best available algorithm for hybrid models using the GeNIe software. We have added details of the forward sampling algorithm in the methods section lines 259-265 of the manuscript explaining the following:

chain is usually thrown out to make sure that the chain has converged to the true distribution (burn-in). This wasn't mentioned in this paper, and I was left uncertain about the modelling technique. Please explain it more clearly. Also, BNs are supposed to give the best available assessment of the *probabilities* of the events (given the scenarios etc.), so it should not be necessary to refer to "x out of y simulations" whan discussing the results.	The hybrid forward sampling algorithm generates samples from the probability distributions of parentless nodes, which it then uses to generate samples in child nodes of the parent nodes that have been sampled, generating conditional probability distributions. The algorithm is hybrid, because the algorithm can generate samples from both discrete and continuous distributions. As the algorithm generates 10,000 samples, stakeholders enquired what was meant by, for example, a 51% probability of a variable being resilient. Stakeholders were more receptive to with phrases such as, 51% of the 10,000 samples, which we have retained when explaining results.
4) Maybe go further back to the roots (such as Perl 1986) when explaining what BNs are in the introduction.	Reference now made to the work of Pearl (1986) in Line 42 of the updated manuscript describing BNs as directed acyclic graphs and conditional probability quantification.