Responses to Reviewer #1 (hess-2021-353)

Our responses are in blue and proposed revisions are underlined.

Summary:

1. Guo et al. investigate the variability in C-Q relationships in relation to the catchment hydrological conditions (more specifically the BFI) for several water quality parameters across several climate regions in Australia. The authors make use of an impressive data set from an arid region and apply a Bayesian Hierarchical Approach including the BFI, which allows understanding spatial patterns of export dynamics. This study can thus provide an important contribution for understanding solute transport beyond temperate regions. However, the manuscript still needs more clarity on the research questions and key messages and methodological improvements. I would suggest the manuscript for major revisions as substantial improvements are still necessary.

Thank you for acknowledging the value of the study, and your comprehensive suggestions for further improvement. We provide a point-to-point response to your comment below.

General comments:

2. One of my concerns is that given the fact that previous studies in the same region, using the same dataset and similar methods (as I read from the text), are not accessible or provided (under review or in preparation), it is not possible to judge the additional value of this study. The preceding studies (Lintern and Liu) are referenced both when defining the research goals and in the method sections. I definitely see the value of investigating C-Q relationships in various climate zones, but this was also done by these referenced studies. It is hard to judge the additional value without knowing what was shown already.

I believe this comment refers to L65-69 in the Introduction, which introduced our preceding studies, Lintern et al. (in review) and Liu et al. (in preparation), which both focused on how C-Q relationships vary across the Australian continent. Considering the paper timelines, we propose to keep only Lintern et al. (in review) here and add a summary of this study in the revised manuscript as:

- *"Lintern et al. (in review) focused on differences in water quality and the C-Q relationships across different climate zones. One remaining question that Lintern et al. (in review) highlighted is the high variation of the C-Q relationships within each climate zone, which required further explorations."*

We also propose to add the following discussion on new insights added by this current study to the end of the Introduction, as:

- *"We use a subset of the grand dataset that Lintern et al. (in review) used, which enabled us to focus on the representative observations for a wider range of flow conditions at individual catchments. As such, this study by analysing the impacts of baseflow contribution, adds new insights to explain the variations in C-Q relationships within individual climate zones."*

We also provide a copy of Lintern et al (in review) as a supplementary document for the reviewer's consideration. Note that Lintern et al (in review) is likely to be published before the publication of the current study, in which case we will include the full reference in the revised manuscript.

3. The motivation of investigating BFI impact on C-Q relationships was not convincing for me. It needs to be clear 1.) why we need to know that and 2.) what exactly we do not know yet. The first question is not satisfyingly presented: Why do you want to focus on BFI, why is it useful to

investigate this relationship? For the second: From my knowledge and in contrast of what you state (see also my comments below), the influence of BFI on the spatial variability of C-Q relationships has been discussed in several previous studies. However, I agree that studies have been biased towards temperate climates. I think the latter should be the main motivation, while generally the literature review on the control of BFI needs to be extended. It is not right, that is has not been investigated. There are studies using BFI as a descriptor for explaining the variability in export behaviour of different solutes, including several studies that you have cited in the introduction but considering other statements. For example, Minaudo et al. 2019 stated, "we found for NO3− that high BFI values, low W2, and low erosion differentiated C-Q dilution patterns from non-significant and mobilization types". But also Ebeling et al. 2021, Moatar et al. 2017, Musolff et al. 2015 have used the BFI to explain variability in C-Q relationships among catchments for several solutes. Moatar et al. 2020 also investigated the impact of discharge flashiness on C-Q slopes and subsequently load flashiness. As BFI and Q flashiness are closely linked, this needs to be mentioned in the introduction. These also need to be discussed in relation to your study in the discussion. Also see further comments below

This study is important because understanding the impact of baseflow contribution on CQ relationships can help explain both the spatial and temporal variation of CQ relationships, and to infer key transport pathways.

We agree with the reviewer that previous studies have investigated how variation in baseflow contribution affects CQ relationships across multiple catchments. The link between CQ patterns and interannual baseflow index (or similar metrics) were well studied (e.g., Moatar et al. 2017, 2020, Musolff et al. 2015, Ebeling et al. 2021), while Minaudo et al. (2019) explicitly explored the link between instantaneous BFI and CQ relationships. However, the existing studies on the impact of baseflow contribution are largely focused on the temperate climate. Therefore, we haven't yet well understood the effects of baseflow contribution across a wider range of climate zones, nor been able to quantify this impact.

The key motivation of this study is to understand and quantify how catchment-level metrics of baseflow contribution affect CQ relationships across catchments in a wide range of climate zones. Through synthesizing continental water quality data, we were able to gain large-scale understanding of how CQ changes in catchments with different hydrological regimes.

To clarify these, we propose to revise the Introduction to:

1) highlight the importance of the key issue that this study explores (impact of catchment baseflow contribution on the CQ relationship).
2) revise the literature review and add discussions on the state-of-art research on the impacts of baseflow contribution, including existing studies on the variation of CQ relationships across catchment as well as the role of flow flashiness.
3) improve clarification of the knowledge gap on the lack of understanding of how CQ relationships vary with baseflow contribution across a wide range of climate zones.

4.
  a) Some of the methods seem inappropriate, especially as there is too few data for some climate-solute combinations to fit robust models/regressions and interpret them (further comments below).
  b) Besides, I do not see the value of investigating the BFI impact within each climate zone individually, i.e. separating the climate zones and fitting different models, instead of investigating the BFI impact across the whole climate variability. I think it would be more

valuable to know what effect the BFI has across the whole climatic variability, i.e. the continuum of variations. The climate zones, could be represented by their characteristics such as precipitation amount, seasonality, aridity, temperature etc.. Even within the climate zones those variables vary and could potentially explain the deviations not explained by BFI.

a)

In this study, we used a single Bayesian Hierarchical Model (BHM) to fit all data points for each constituent simultaneously, across all climate zones. Although we used different parameters for individual climate zones, this model structure has a great advantage of 'borrowing power', meaning that information is shared across climate zones, and parameter fitting for one climate zone can utilize data from all other climate zones. As illustrated in Figures 5 and 6, this approach has been effective in estimating the impact of BFI_m (catchment median BFI) on CQ slopes for climate zones with limited data, while acknowledging the greater uncertainty. This model enables us to estimate the BFI impacts for climate zones with limited data by making use of data from other climate zones; in this way we made the best use of all data available. BHM is thus ideally suited to data-limited situations in dealing with spatial data, which was also illustrated in our previous works (Guo et al., 2018; Guo et al., 2020; Liu et al., 2021).

We plan to add further discussion to the Method section on why BHM is suited to the dataset analysed in this study. Specifically, in the introduction paragraph of Section 2.2 (which introduces the model), we will add discussions on the key advantages of BHM in effectively handling data-limited situations and spatio-temporal data with uneven coverage. We will also add references to our previous works which have illustrated these advantages.

References:

Guo, D., Lintern, A., Webb, J. A., Ryu, D., Bende-Michl, U., Liu, S. & Western, A. W. (2020). A data-based predictive model for spatiotemporal variability in stream water quality. Hydrology and Earth System Sciences, 24(2), pp. 827-847.doi:10.5194/hess-24-827-2020

Guo, D., Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., . . . Western, A. W. (2019). Key Factors Affecting Temporal Variability in Stream Water Quality. Water Resources Research, 55(1), 112-129. doi:10.1029/2018wr023370

Liu, S., Ryu, D., Webb, J. A., Lintern, A., Guo, D., Waters, D., & Western, A. W. (2021). A Bayesian approach to understanding the key factors influencing temporal variability in stream water quality – a case study in the Great Barrier Reef catchments. Hydrol. Earth Syst. Sci., 25(5), 2663-2683. doi:10.5194/hess-25-2663-2021

b)

The reason to investigate BFI effects for individual climate zones separately is to see whether we could identify any statistically significant differences of the effects for different climate zones, and the results suggested that this is the case (from the CQ slopes in Figure 5 and the simulated CQ~BFI relationships in Figure 6). If there is no significant difference of the BFI effects between climate zones, the model is also capable to indicate this – as would be shown with similar, undistinguishable modelled effects of BFI for individual climate zones. Our current model structure enables us to test this hypothesis, since we are essentially developing different models for each climate zone and comparing them within one larger modelling framework. In contrast, a model with only a single parameter for the BFI effects across all climate zones cannot be used to test this hypothesis.

We further explored the performance of a model which only allows a single parameter for the BFI effects across all climate zones. As shown in Table R1, a model that uses a lumped effect of BFI_m across all climate zones can barely explain any variation in the CQ slopes. However, we see this as a useful result to include and to illustrate any value added by exploring climate-specific effects of BFI. Therefore, we propose to include in Section 3.3.1 a new comparison of performance between the lumped and climate-specific models. This new comparison will replace the existing results in Section 3.3.1 which have caused confusion (as highlighted in your Comments #24 and #26, see our further responses to those comments).

Table R1. The performance in explaining variation in the CQ slopes (summarized by R2), for a model that only allows a single effect of BFI_m across all climate zones.

WQ variable	Baseline (lumped)
EC	0
TP	0
FRP	0.03
TSS	0
NOx	0.03
TN	0.02

We agree with you that variations in the CQ slopes can also be explored together with other climate drivers (e.g., temperature, rainfall, aridity), but a key challenge we see in other water quality modelling studies with these potential drivers is the high-cross correlation between these variables, thus we decided to use climate zones as an integrator of many hydroclimatic variables. The focus of this study is BFI and we hypothesise that BFI has a significant effect on CQ slopes – which is supported by the study results. To address your comment, we will add further justification to the use of climate zones in Section 2.2 (Methods) for exploring the BFI effects. We will also add discussions on the need for further studies on the impact of individual climate drivers on CQ slopes in the Conclusion.

5. The interpretation of BFI_m in terms of variability of flow paths is not convincing to me as you could easily and more directly use the range of BFI to determine the relationships between C-Q slope variability with BFI ranges. I think it would be good to look at this instead of speculating, as you have the data at hand and Figure 3b is not convincing enough for this interpretation, in my opinion. Instead of the range BFI_h-BFI_l you could also consider other metrics of variability.

To address this comment, we performed further analyses to explore the impact of catchment BFI range on the CQ slopes. Specifically, we explored a new model with the same structure as our original model based on BFI_m, which is now based on BFI_range (i.e. the difference between BFI_h and BFI_l). The modelled effects of BFI_range on CQ slopes (see Figure R1 below) are highly consistent with the modelled effects of BFI_m on CQ slopes (see Figure R2 below, with minor updates from the original Figure 6 after correcting some previous errors in NOx data). These new results provide more concrete evidence on our previous 'speculation' on how the BFI range in a catchment can impact its CQ slope. We will add the new modelled effects to the Results section and update our discussion accordingly.

We also further investigated on the non-linear relationship between BFI_m and the BFI_range beyond the current results in Figure 3b) in the manuscript. We found that the non-linearity is a result of BFI_l and BFI_h both increasing with BFI_m, but the increase in BFI_h slows down at high BFI_m (see Figure R3 below). This result further highlights the need to explicitly consider the impact of the BFI_range. Thus, in addition to the proposed addition of the BFI_range model, we also plan to replace the current
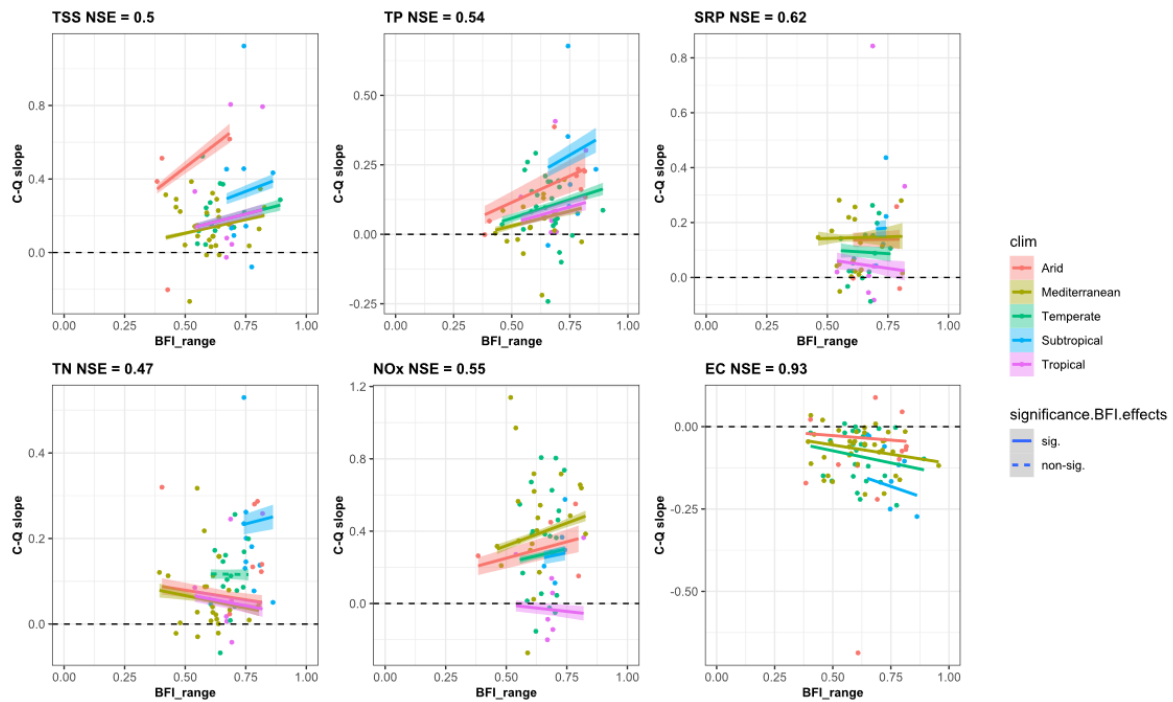
Figure R1. Catchment C-Q slope vs. catchment BFI range (BFI_range, as the difference between BFI_h and BFI_l), coloured by climate zones. The lines represent the modelled C-Q slope~BFI_range regression lines for individual climate zones,
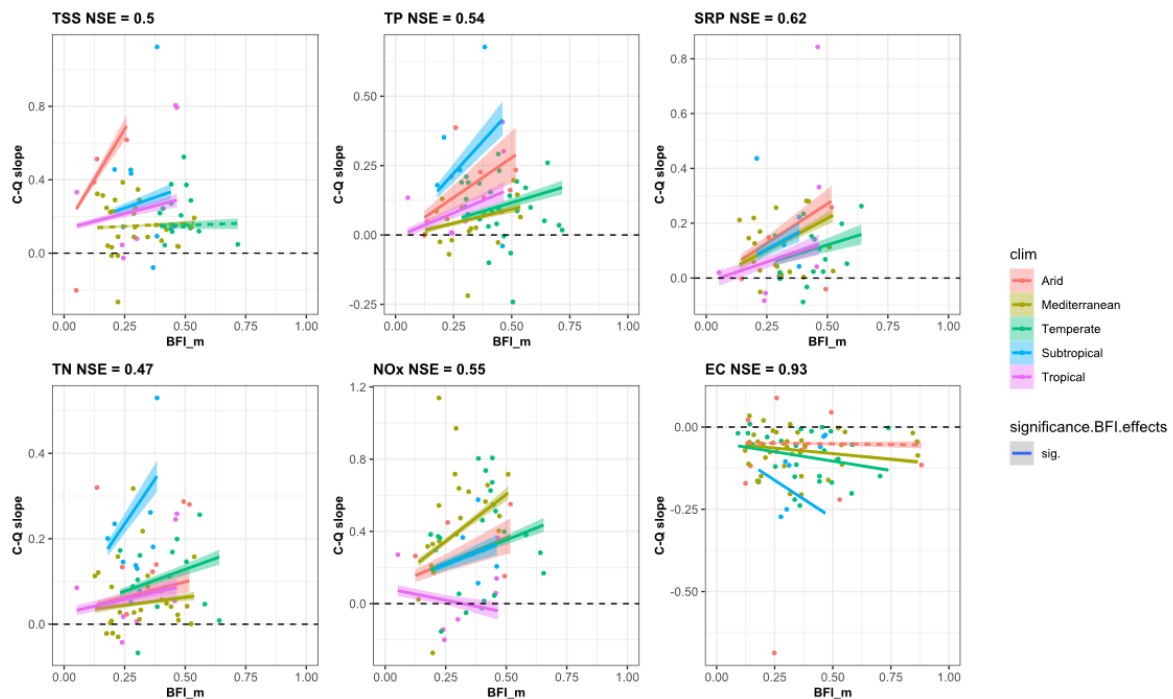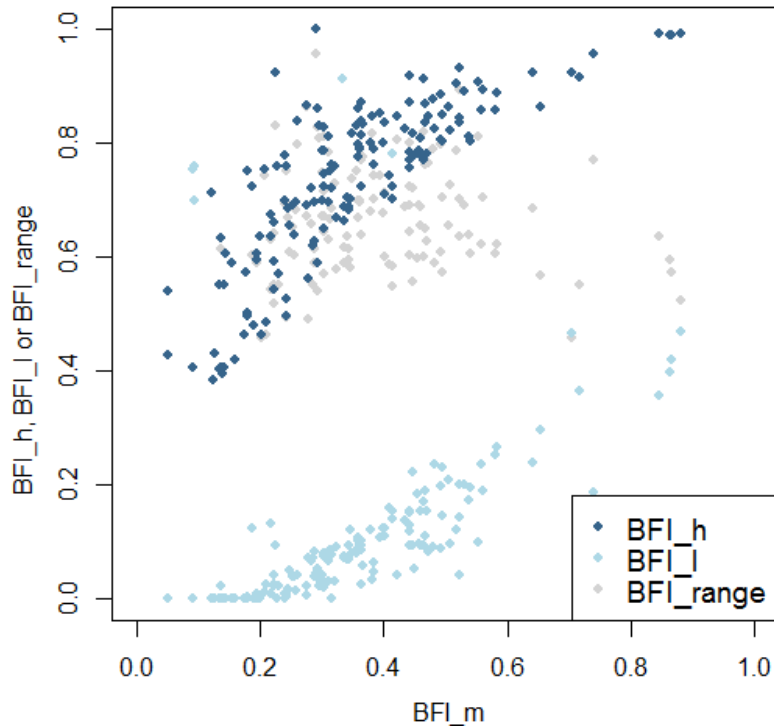

Figure R2. Catchment C-Q slope vs. catchment median BFI (BFI_m), coloured by climate zones. The lines represent the modelled C-Q slope~BFI_m regression lines for individual climate zones,

Figure R3. The low (10th percentile) and high (90th percentile) of instantaneous BFI (BFI_l and BFI_h), and the range of instantaneous BFI (BFI_h - BFI_l) versus BFI_m. The plot includes all 157 catchments across six water quality variables studied.

6. Linked to the methodology, I have a concern about the conceptualisation in Figure 7 and main conclusions. I think some methodological approaches and results/evidence are not robust and clear enough to generalise the results in the given way.

This figure will be heavily revised to incorporate our new results with the model that focuses on the impact of BFI_range on catchment CQ slopes (as detailed in our response to your Comment #5). To further improve the clarity, we will remove the panels on specific constituents to simplify the plot, and also revise the organization of the figure to improve the connection of individual panels.

7. The discussion misses comparison to relevant previous studies. Previous studies investigating hydrological controls (such as the BFI, flashiness etc.) spatial variability of C-Q relationships. The discussions needs extension

To address this comment, we propose to add some in-depth discussion of the studies summarized in the Introduction, which looked at various hydrological controls of CQ relationships and how they vary across catchments. We will highlight the key findings of these studies, compare our results to the existing studies, and discuss the added value of our work.

Specific comments:

8. L17: Does the baseflow contribution in a catchment impact the concentration itself or the C-Q relationship? For me, it seems like spatial and temporal dimensions are mixed up here.

Here we aim to provide a general statement of the importance of understanding C-Q relationships. For generality and clarity, we will remove the 'spatial and temporal variation' and revise the sentence as:

- *'Understanding the concentration-discharge (C-Q) relationships informs solute and particulate export processes'*

9. L18: This is not true "these patterns have not yet been investigated across large spatial scales", e.g. Minaudo et al. 2019 (see also related comments)

We propose to revise this sentence as:

- *"Previous studies have shown that the extent to which baseflow contributes to streamflow can affect C-Q relationships in some catchments. However, existing understanding on the effects of baseflow contribution in shaping the CQ pattern is largely derived from studies in temperate areas, while we still lack quantitative understanding of these effects across a wide range of climates."*

10. L48: "variable, which" reference is unclear, you mean here the studies? Please revise

We propose to start a new sentence from "which", as:

- *"These studies highlighted a number of critical drivers for these spatial variations, such as land use, land management, lithology and topography (e.g., Ebeling et al., 2021; Minaudo et al., 2019)."*

11. L 57: This sentence needs revision. The concentration variability within one catchment regarding the contribution of baseflow or quickflow to the current discharge is represented by its C-Q relationship, i.e. the variability "within a particular catchment". However, the cited studies also investigate the differences in C-Q relationships among catchments. Therefore, the provided references do not fit to this sentence, in my opinion. This also leads to the next sentence being incorrect. It defines the research gap as the differences among the catchments regarding the hydrological average behaviour not being investigated and understood. E.g. Minaudo et al. 2019 (others, see main comment) considered BFI in the analysis of C-Q relationship variability among catchments.

We agree with the reviewer that previous studies have investigated how variation in baseflow contribution affects CQ relationships across multiple catchments. As in our response to your Comment #3, we acknowledge the key gap that this current study addresses is the lack of understanding of the effects of baseflow contribution on CQ slopes across a wider range of climate zones, as well as the lack of quantification of the impact of baseflow. The key motivation of this study is thus to understand and quantify how catchment-level metrics of baseflow contribution affects CQ relationships across catchments in a wide range of climate zones.

We propose to revise the Introduction (including this sentence) to:

1) highlight the importance of the key issue that this study explores (impact of catchment baseflow contribution on the CQ relationship).
2) revise the literature review and add discussions on the state-of-art research on the impacts of baseflow contribution, including existing studies on the variation of CQ relationships across catchment as well as the role of flow flashiness.
3) improve clarification of the knowledge gap on the lack of understanding of how CQ relationships vary with baseflow contribution across a wide range of climate zones.

12. L91: "nitrate-nitrite" I am not sure what you mean with that. Is it the sum of both nitrate and nitrite concentrations? It should be defined once in the manuscript

Thank you for highlighting this, we will add this in the revised manuscript as:

- *"... total suspended solids (TSS), total phosphorus (TP), soluble reactive phosphorus (SRP), total nitrogen (TN), the sum of nitrate and nitrite (NOx) and electrical conductivity (EC)"*

13. L105 "unaffected" this is a strong word, I suggest to say "more robust"

We will revise this phrase as:

- *"...this is to ensure that the C-Q relationships are more robust with the presence of outliers..."*

14. L106: plural "span"

Thank you. We will revise this phrase as:

- *"Having water quality time-series that span at least 3 years…"*

15. L111: "met the above criteria across all the six water quality variables" for me this sounds as if the stations needed to meet the criteria for all the six variables, which was not the case from what I read in the next sentence and the following. I suggest to revise this formulation

We will revise this sentence as:

- *"We performed the above catchment selection for each water quality variable and found a total of 157 sites (catchments)."*

16. L144: Does that mean you fit equation 1 only to baseflow discharge? How can this work, if C is a mix of baseflow and quickflow concentrations? This sentence in unclear, please revise.

No, Eqn 1 was fitted to all available pairs of C-Q data regardless of whether the flow is baseflow dominated or quickflow dominated. To clarify this, we propose to revise the sentence in L145 as:

- *"Our model is based on such a single C-Q relationship at individual catchments. However, our model enables the slope term (β$_s$) for different catchments to change with the overall catchment-level baseflow contribution. Such model conceptualization is based on previous literature on the effects of baseflow contribution on C-Q slopes within individual catchments (Gorski & Zimmer, 2021; Minaudo et al., 2019), while aiming to further explore the impact of BFI across multiple catchments."*

17. L146: "the C-Q slopes of all catchments are following a normal distribution with a 'grand mean'" This works if the represented catchments cover the range of variability well. This would not be true if catchment types are overrepresented, would it?

The use of this 'grand mean' for each water quality variable is based on a finding from our preceding study, that the C-Q slopes for each water quality variable do not vary substantially across climate zones (Lintern et al. in review). We clarify that we used the mean in our model to represent a baseline level of CQ slopes for each water quality variable; whereas the CQ slope of each catchment is the sum of the grand mean and the BFI effect (rather than sampled from a normal distribution), following Eqn 2. To clarify this, we propose to revise this sentence to a more accurate description of the model structure as:

- *"We assume that for each water quality variable, the C-Q slopes of all catchments have a 'grand mean', β0. Then the variation of C-Q slopes between catchments, away from β0, are explained by changes in catchment BFI."*

As a further comment to the representativeness of catchments that we analysed, we acknowledge the lack of data for some regions due to limited monitoring (i.e., regions of low management prioritization/interest for monitoring, and regions inaccessible for monitoring). Further, climate zones and catchments are unequally distributed across Australia; catchments are more clustered towards the coastline due to higher rainfall and topography, while the large arid region in the middle of the

continent does not have many flowing rivers. These mean that we expect the catchment representation to be biased towards regions that have more catchments and are monitored more frequently (e.g., in the coastal regions of Australia).

18. L146f: "Then the variation of C-Q slopes between catchments, away from beta_0, are explained by changes in catchment BFI. " This would only be true, if the BFI is the "only" controlling variable

We intend to use BFI as the only controlling variable because this study focuses on the impacts of BFI on the variation of CQ slopes across catchments. To clarify, we will add a clarification of this assumption after this sentence:

- *"This model conceptualization also assumes that the BFI is the only controlling variable and thus will inform how well BFI explains the spatial variation of C-Q slopes."*

19. Fig2: I suggest to add axis labels and ticks for panel a and b

We will revise the figure as suggested.

20. L185: "together with flow" I do not understand this. If c=f(Q) in the C-Q relationship this is already included.

We will delete this phrase. This entire paragraph 'Model performance' will be replaced with the new model performance comparison to be presented in Section 3.3.1 (comparison of performance between the lumped and climate-specific models, as detailed in our response to your Comment #4a).

21. L214 "surface flow" is imprecise. There is not just baseflow and surface flow

We will replace this phrase as 'quickflow' for consistency with the rest of the paper.

22. L217: "In contrast, a catchment with a high BFI_m generally has a large range in instantaneous BFIs" I see several high BFI_m with not very high ranges in BFI. Fig3b rather looks like a bell shape with highest ranges for medium values, not like a linear relationship.

We agree with your observation. This nonlinearity highlights the need to explicitly consider the impact of the range of BFI. This is further justified with our additional investigated on the non-linear relationship between BFI_m and the BFI_range as below in Figure R3. We found that the non-linearity is a result of BFI_l and BFI_h both increasing with BFI_m, but the increase in BFI_h slows down at high BFI_m. As detailed in our response to your Comment #5, we plan to update the current Figure 3b) as Figure R3, as well as adding an alternative model structure with the range of BFI as the key predictor, so that we can explicitly consider the impact of the range of BFI on CQ slopes.
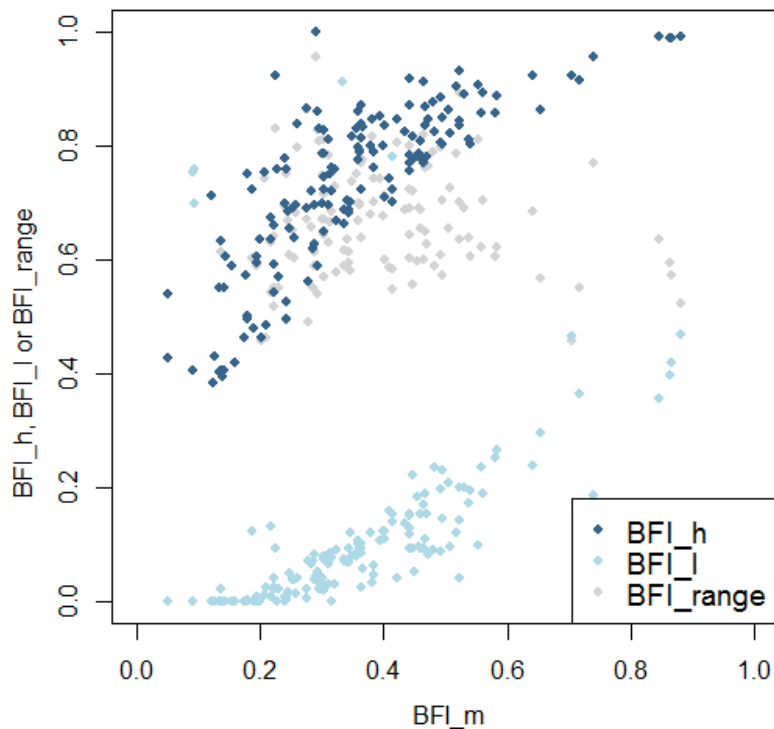
23. Fig 3: Boxplots can create confounding impressions if the sample size is very different, which I see from Figure 1. I would suggest adapting the boxplot widths according to the sample size and/or writing the sample size number to the plot for each climate zone (probably the numbers in the x axis labels?). For the right panel, I would suggest to colour the dots according to their climate zone as in Figure 1, possibly also for the left panel with different hue or saturation values to distinguish the different BFI quantiles.

Thank you for the suggestions. We already have the sample size for each climate zone labelled on the x-axis, we will add a sentence to the figure captions to clarify this.

We will also use different colours for individual climate zones in both the left and right plots as suggested.

24. L255: "BFI-based model has only marginally lower performance … BFI-based model, while having the capacity to predict C-Q slope across space, can predict water quality almost as well as using the observed C-Q slope." I do not understand this comparison of individual C-Q slope with a model explaining the variability in C-Q slopes. It sounds like you were expecting worse performance, while actually a more complex model should improve performance to be a valid approach.

As detailed in our response to your Comment #4b), we propose to remove the current results in Section 3.3.1 and instead present new results that compare the performance of models with i) climate-specific impact of BFI; and ii) a single impact of BFI across all climate zones. This update will hopefully clarify our approach and avoid any confusion on our communication of model performance.

25. L269: "confidence intervals"? I do not know credible intervals

"Credible interval" is the term used to describe uncertainty in the posterior samples (including simulations and parameter values) from Bayesian modelling (Gelman et al., 2013). We propose to add a sentence to clarify this as:

- *"The significance of the modelled effect of catchment median BFI is assessed by the 95th credible intervals (the range between the 2.5th and 97.5th percentiles of the posterior parameter values, which measured the uncertainty in Bayesian posterior distribution, as Gelman et al., 2013). The effect of catchment median BFI on C-Q slope is almost always significant, with the 95th credible intervals not crossing over 0 for most combinations of water quality variables and climate zones."*

Reference: Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis, Third Edition: Taylor & Francis.

26. Table 1: Why do you think the NSE of your BFI-base model are lower than the baseline model? Does it actually make sense to fit a more complex model in this case?

As detailed in our response to your Comment #4b), the current results in Section 3.3.1 will be removed and replaced with new results that compare the performance of models with i) climate-specific impact of BFI; and ii) a single impact of BFI across all climate zones. We believe this update will clarify our approach and avoid any confusion on our communication of model performance.

27. Section 3.2 I do not like this whole section, I think the statements derived from selected examples are not representative for the catchments distribution of TSS export patterns within the corresponding climate zones (Fig6)

We intend to present these cases only as examples to show the potential impacts of BFI. We expect that showing the actual data of C, Q and BFI can better help the readers to understand the key model components and relationships we aim to represent. We plan to add a final sentence to this section to highlight the limitation of the examples and need to explore a wider range of catchments to generalize the pattern:

- *"However, the impacts of BFI seen in these examples cannot meaningfully represent large-scale patterns over a wide range of catchment conditions. Section 3.3 presents our modelled results in which these large-scale patterns are synthesized."*

28. Figure 5: What is the modelled effect? Is it dBFI_climate from equation 3? What is the NSE above each subplot describing?

Yes, the modelled effect is $\delta BFI_{climate}$ in Eqn 3. This is already described in L269 before the Figure as: "Figure 5 presents the median and the 95th credible intervals of these modelled impacts for each water quality variable, derived from the Bayesian posterior estimates of $\delta BFI_{climate}$ (Eqn. 3)."

The NSE above each panel summarizes the model performance to predict the concentration of individual water quality variables. We will revise the figure captions to clarify these points as below. Note that the term $\delta BFI_{climate}$ will be revised as $\delta_{\mathrm{BFI\_climate}}$ throughout the manuscript to avoid the confusion highlighted in Comment #10 of Reviewer 2.

- *"Figure 5. Modelled effects of BFI_m on C-Q slope for each climate zone, i.e., $\delta_{\mathrm{BFI\_climate}}$. The bars show the 95% credible interval (2.5th to 97.5th percentile) of modelled effect for each climate zone for each water quality variable, and the dots indicate the corresponding median levels. The colours indicate whether an effect is significantly positive, significantly negative, or non-significant; a positive effect means that C-Q slope increases with higher catchment BFI,*

and vice versa. Black dashed lines show zero-effect i.e. no effect at all. The NSE above each panel summarizes the model performance to predict the concentration of individual water quality variables."

29. L278-285: I do not agree with this approach and subsequent observation. Fitting linear relationships is not appropriate for the given observations and "consistent diverging" behaviour goes beyond what can be interpreted here. Especially for the last two sentences: When looking at the overall point clouds, there is not clearly increasing variance (diverging behaviour) with higher BFI_m. In my opinion, weak relationships are overinterpreted here. These are also transferred to conclusions.

We'd like to clarify that L278-285 do not intend to make inference on the BFI effects here; instead, we clarify that the diverging patterns are a result of the model structure. To clarify, we propose to add a further sentence after L285 as:

- *"Since these diverging patterns are a result of the model structure, we do not further interpret any physical representations on the impacts of BFI."*

We have not included any summary of these results in the Conclusion section.

We understand the concern about the limitation of the linear model used, but we consider this model structure to be appropriate for our study purpose, which is to synthesize large-scale patterns of BFI impacts across different climate zones. The model structure is suitable for our purpose, with demonstrated ability to identify significance effects of catchment baseflow contribution on CQ slope (i.e., these modelled effects are mostly constrained to a very limited range as seen in Figure 5 in the manuscript). We do acknowledge the limitation of this structure and the potential issue of outliers, and the need of further studies to explore model improvement possibly via more complex model structures - we proposed to add further discussions on these points in the revised Discussion section.

30. L283: 'grand mean' I know you have used this term before, but I think it is not well chosen, as it does not tell that it is about the "solute-specific base C-Q slope". Consider changing the term

Thank you for the suggestion, we will revise this term throughout the manuscript.

31. Figure 6: In my opinion, it is not justifiable to fit linear regressions to all combinations of solutes and climate zones, because several combinations have 1) too little sample sizes, 2) clearly non-linear relationships, and 3) in some cases plus influenced by outliers, e.g. the Tropical NOx fit. The legend titles should be improved.

Firstly, small sample size for some climate zones is not a real limitation of the Bayesian Hierarchical Model due to its ability to use information across data groups, as explained with more details in our response to your Comment #4a). We also propose to add more details on the advantage of the Bayesian Hierarchical Model in working with small sample size in the Method section. Specifically, in the introduction paragraph of Section 2.2 (which introduces the model), we will add a brief justification for applying BHM. We will discuss the key advantages of BHM in effectively handling data-limited situations and spatio-temporal data with uneven coverage, which are illustrated in our previous studies (Guo et al., 2018; Guo et al., 2020; Liu et al., 2021).

Secondly, we consider a linear model suitable for the purpose of this study, which was to synthesize large scale patterns with a simple model structure applied to the whole country. The model structure is suitable for our purpose, with demonstrated ability to identify significance effects of catchment baseflow contribution on CQ slope (i.e., these modelled effects are mostly constrained to a very limited range as seen in Figure 5). We acknowledge the limitation of this structure and the potential

issue of outlier, and the need of further studies to explore model improvement possibly via more complex model structures. We will add discussions on this limitation and future studies to the revised Discussion section.

We will also add the legend title in the revised figure.

32. How does the modelled effect from Figure 5 relate to the slope of the linear regressions in Figure 6? This seems somewhat redundant to me.

The modelled effects in Figure 5 are the same as the slopes of the linear regression in Figure 6. Although the two figures overlap on the impacts of BFI, they focus on different aspects. Figure 5 focuses on the direction and significance of the BFI impacts themselves, while Figure 6 focuses on illustrating these effects with CQ slopes and BFI. Further, Figure 6 is a more intuitive way to show the impacts on BFI. Therefore, we believe it's necessary to keep both figures.

33. Figure 7: Firstly: The generalisation shown is questionable (see main and other comments). E.g. Figure 3b shows that highest BFI ranges are for medium BFI_m, suggesting that high BFI might also have more stable flow conditions with generally higher groundwater contributions. Secondly: This Figure takes a lot of time to understand and could benefit from some reworking including/according to other adaptations. The Figure text is unclear without reading the main text, as well as the meaning of a1, a2, b1, b2 only from the second reading. I suggest selecting other identifiers. The spatial organisation, e.g. link between the upper and the bottom panel and left, middle and right column, is not visually clear.

This figure will be heavily revised to incorporate our findings with the additional BFI_range model (as detailed in response to your Comment #5). For further clarity, we will remove the panels on specific constituents to simplify the plot, and also revise the organization of the figure to improve the connection of individual panels.

34. L332-334: This statement could benefit from checking also mean concentrations: are activated sources low or high in more quickflow dominated catchments?

We have checked this and included the results in Figure S11 in the original manuscript. In general, we see high median concentration at the more quickflow dominated catchments (note: since we work in the log-10 space for both concentration and flow in deriving the CQ slopes, the median concentration is a more useful indicator than the mean). To address this comment, we propose to revise this description as below to include the pattern on median concentration:

- *"catchments with lower BFI_m generally have narrower ranges of instantaneous BFI, which thus tend to always have low baseflow contributions (i.e. surface flow dominated) regardless of dry or wet conditions (Catchment a, Figure 7). This can lead to a limited range of water sources, with small variation of flow pathways to transport water chemistry to rivers, resulting in a relatively stable export pattern across low and high flows at these catchments. Note that these catchments with lower BFI_m tend to have higher median concentration (Figure S11), suggesting relatively high activated sources from these limited range of flow pathways'*

35. L398-391: I cannot follow this point unfortunately due to missing information. Moreover, the characteristics land use, geology and climate are all integrated in the base flow index, which is a resulting hydrologic characteristic. This point definitely also needs a discussion part, including further literature and potential controls.

We propose to delete this sentence as the referenced study (Liu et al., in preparation) is still ongoing. Instead, we will highlight the need of further investigating other potential drivers for spatial variation of the CQ relationships, such as land use, land cover, geology and climate.

36. L392: Why is the Bayesian hierarchical model more effective than multiple linear regressions with BFI or other multivariate models? For me, Figure 5 (outcome of the Bayesian model) and Figure 6 (not a direct model output) were somewhat redundant. This is not covered in enough in the discussion section.

Bayesian Hierarchical Model (BHM) and multivariate regressions are not exclusive to each other i.e. BHM is also capable of including multiple predictors. We are not intending to compare BHM with any other model with this statement, instead we highlight the efficiency of BHM in making inferences and synthesizing large-scale patterns with limited and heterogeneous data availability across space. We will revise this paragraph to better clarify these points.

Figures 5 and 6 are intended to show different aspects of the results although they have overlapped contents, we explained this in detail in our response to your Comment #32.

37. SUPPLEMENTS: Figure S3: I do not understand why the BFI_m should be shown per solute, if the BFI depends on discharge and not on concentration.

BFI depends on discharge only, but the catchments analysed for individual water quality variables differ because the catchment selection criteria (Section 2.1.1) considered water quality data availability. This figure shows each water quality variable in as a subplot to incorporate the difference in catchments analysed.

38. REFERENCES: Ehrhardt et al. is already finally published, please change reference to: Ehrhardt, S., Kumar, R., Fleckenstein, J. H., Attinger, S., & Musolff, A. (2019). Trajectories of nitrate input and output in three nested catchments along a land use gradient. Earth Syst. Sci., 23(9), 3503-3524. https://www.hydrol-earth-syst-sci.net/23/3503/2019/

Thank you, we will update this reference in the revised manuscript.