Response to all reviewers
"Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management"

Initial responses are in blue; new responses (after revision) are in red.

Response to Editor

Editor Decision: Publish subject to minor revisions (further review by editor) (02 Mar 2021) by Giuliano Di Baldassarre

Comments to the Author:

Dear Authors,

three Referees reviewed the article and provided minor comments that should be addressed before publication.

Kind regards,

Giuliano

We greatly thank the Editor. We have carefully addressed the Reviewers' suggestions as follows:

Response to Reviewer 1

We gratefully thank Reviewer 1 for these supportive and constructive comments, which will improve the quality of the Review. Below we provide the Reviewer's comments verbatim in black font, and our responses immediately below each comment in blue font.

Whereas most reviews focus on a single aspect of hydrologic non-stationarity (e.g. the type of trend tests applicable) this manuscript I believe is the first manuscript to review the entire non-stationary management process from identifying metrics, data considerations, through to analysis and management.

I have provided a lot of comments, but they are mainly editorial in nature. I understand it is not possible to capture all published literature in a review. Hence where I have suggested references or additional methods to be described the authors should not feel these recommendations are prescriptive.

Maybe one general comment: The manuscript described changes to wind in detail at the start, but it didn't seem that wind was mentioned in Section 5. This may be something the authors wish to consider in a revised manuscript.

Otherwise, thank you for the allowing me to review this manuscript which I am confident will be a very welcome addition to the literature. Please see below for my specific comments.

We thank the Reviewer for their positive feedback. We will make sure that wind is mentioned in Section 5, so that all extremes are described throughout the manuscript.

Page 1, Abstract: As the title mentions “management” I think the abstract should mention management also.

We will mention management in the abstract.

Done.

Page 1, Line 18: I think you should define natural nonstationarity. My guess is you mean “large scale climate variability” like ENSO and Milankovitch cycles? I am not sure.

This sentence and section will be clarified, to make it clear that it is the model used to describe the extremes that is stationary/nonstationary. Instead of ’natural nonstationarity’ we will discuss ’natural drivers of nonstationarity’. We will explain the role of large scale climate modes (which are both natural and affected by anthropogenic climate change to varying extents).

This sentence has been modified to focus on artificial patterns and trends induced by human activities.

Page 2, Figure 1: I wonder if this figure would work better as more generic: “magnitude” on the y-axis and “time” on the x-axis rather than specific numbers.

We agree, and will revise Figure 1 as suggested.

Done.

Page 2, Line 4: “death of stationarity in water management” – I would add water management as I think that is what Milly et al was specifically referring to and it ties to your title.

Thank you - we will clarify the sentence.

Done.

Page 3, Line 3: I would remove “somewhat” and just state “are predictable”.

We will remove the term.

Done.

Page 3, Line 4-11: My feeling is this paragraph could be deleted. It is repeating what was said before.

We will delete/merge the two paragraphs to ensure there is no repetition.

The two paragraphs have been merged.

Page 3, Line 13: Where you state your manuscript as an “introductory overview” maybe you could say something stronger like the manuscript is the first to review the entire non-stationary management process from identifying metrics to analysis through to management. I am not sure my wording is the best, but I hope the authors understand the sentiment.

Yes, thank you for the suggestion! We will strengthen the sentence.

Done.

Page 3, Figure 2: The words “issues” under attribution sounded a bit vague in the context of the other headings and subheadings.

Yes, good point. We will update the figure.

Done.

Section 2. I somewhat disagree with the choice of subheadings. I don’t think you answer the questions but rather summarise the metrics used to investigate the symptoms (as you state at the start). Hence, I feel the headings would be better as “event magnitude”, “frequency”, and “timing”. But I respect the authors may disagree.

We may adjust the headings.

Done.

Page 4, Line 22: I think a sentence defining the PMP is needed before the sentence which states “For a complete state-of-the-art review. . .”

We will define the PMP at this point.

Done: we have defined PMP as the greatest depth of precipitation that is possible in a given place and time and for a given storm duration.

Page 4, Line 25: This doesn’t quite fit in this section. It would fit however if you stated what metric they used and then it would be an example of the application of the metrics above.

The Reviewer is referring to the sentence "A recent global analysis of observed rainfall data indicated that...". This is a fair point and we will adjust the sentence as suggested.

We restructured the paragraph so this is now given as an example of changes in Rx1day and Rx5day precipitation, and the PMP is also illustrated.

Page 6, Line 10: I didn’t understand what “both sides” meant.

We will rephrase the sentence.

We adjusted the sentence so it just describes the inputs to the Palmer Drought Severity Index.

Page 7, Line 22: “percentage of time” (remove “the”).

Will do.

Done.

Page 7, Section 2.2. It could be worth noting the identification of independent events for pot analysis can be performed by fitting Poisson models https://doi.org/10.1016/0022-1694(82)90136-6

This will be noted, and a reference to Restrepo-Posada and Eagleson (1982) will be added, as suggested by the Reviewer.

Done.

Page 8, Line 28: I don't agree with the AEP definition as an average waiting time as I think a better definition is the probability of occurrence above a threshold in a given year.

We will update the definition and clarify this point.

We agree; this has been updated.

Page 9, Figure 4: I am not sure all the panels of the Figure were referred to in Section 2?

We will make sure that all panels are referenced in the text.

We have updated the text so that each of the panels in Figure 4 is referred to at an appropriate location in Section 2. We also modified Figure 4d to include a global figure of flood trends, rather than just one country.

Page 10, Line 5: A possible useful reference: https://doi.org/10.1038/s41598-019-52277-4

A reference to Myhre et al. (2019) will be added.

Done.

Page 10, Section 2.3: Some recent references on changed rainfall timing are:
https://doi.org/10.1029/2018GL079567; https://doi.org/10.5194/nhess-18-2047-2018.
Recent references on changed flood timing are:
https://doi.org/10.1029/2019WR026300; https://doi.org/10.1029/2020WR027233.

These are helpful suggestions. References to Marelle et al. (2018), Brönnimann et al. (2018), Wasko et al. (2020a) and Wasko et al. (2020b) will be added, thank you.

All four references have been added at appropriate locations in the text – thank you.

Page 11, Line 11: I feel the word 'essential' in the title is superlative.

The Reviewer refers to section 3, "Essential data considerations before detecting and attributing nonstationarity". The word "essential" will be removed.

Done.

Page 11, Line 15: Could insert "particularly" in front of prevalent (as data issues tend to be worse for extremes).

We will make the change.

Done.

Page 11, Line 21: “channel” cross-sections? This might help make it clearer that this refers to streamflow specifically (the factors listed before this apply to all weather stations).

Good point, we will alter this.

Done.

Page 11, Line 27: I wasn’t quite sure what “system drift” means.

We will clarify the sentence.

We removed "system drift" because it is not actually mentioned in these references. It is defined in another section of the manuscript about climate models.

Page 12, Figure 5: I don’t know what “recon”, “eval”, “cal” stand for. I feel this figure could be better explained (at least in the caption).

The figure will be clarified.

The figure has been modified and is now fully labelled.

Page 11, Line 29: Figure 5 presents an “approach” for dealing with homogeneity not “tests”.

Will alter this.

Done.

Page 12, Line 11: I agree with the statement, but I don’t actually know what climatologists are doing better than hydrologists – could this be stated explicitly?

We will clarify this statement.

The sentence has been modified, to avoid giving an opinion, and instead simply refers to the need for creating integrated datasets of observed variables

Page 13, Line 13: “Climate is non-stationary by definition”: or does it depend on the record length as you have stated? That is, if you pick the correct length of record it will be stationary? This statement probably just needs explaining to fit in the context here.

We will rewrite this sentence, which is currently flawed.

This sentence has been deleted because it was misleading.

Page 14, Line 20: I am not sure Theil Sen is a “test”. Maybe a test statistic? But even then, this is not strictly correct. Maybe remove the following text: “An additional test called the”

Correct - this was an error and will be altered.

Done.

Page 14, Line 21: “slope” -> “magnitude”?

Yes - we will alter this.

Done.

Page 14, Line 27: I think the statement “in cases where the assumptions of OLS are not met QR may be used” is correct but I want to note that when the assumptions of OLS aren’t met (e.g. linearity) many authors argue non-parametric methods (such as MK) should be used. It would be helpful to state what the assumptions are and point out that the Theil-Sen slope estimator is an alternative also?

We will state the assumptions and introduce the alternatives.

Done (four assumptions, and mention of Theil-Sen and Quantile regression as alternatives).

Page 15, Line 8: Or climate covariate? http://dx.doi.org/10.1016/j.jhydrol.2015.04.041 (the advantage being you can then covariate with GCM projections).

We will modify the text and add a reference to Du et al. (2015) in the section where we discuss statistical-dynamical projections.

We have indicated mentioned that climate covariates can be used in this way and referred the reader to the section on dynamical-statistical models.

Page 15, Line 29: After 2018 can add a reference to Figure panel 6f and maybe start a new paragraph.

Will do, thank you.

Done.

Page 17, Section 4.2: Max-stable models (after Coles, 2001) are another method of pooling data directly (e.g. https://doi.org/10.1016/j.jhydrol.2011.06.014)

Thank you! The references, incl. Coles (2001) and Westra and Sisson (2011) will be added.

We have included max-stable models and the references within our section on pooling approaches.

Page 18, Section 4.4: I think this section mentions circular statistics used but not the methods used to calculate the trend (e.g. circular regression: https://doi.org/10.1029/2019WR026300; circular-linear associations (Villarini, 2016); Theil-Sen (Bloschl et al 2017); Linear regression before https://doi.org/10.1126/science.1152538
or after standardization https://doi.org/10.1029/2020WR027233. I think alongside Gu et al., 2017 the following apply https://doi.org/10.1029/2018GL079567; https://doi.org/10.5194/nhess-18-2047-2018.

Thank you for pointing this out. References to Wasko et al. (2020a), Barnett et al. (2008), and Wasko et al. (2020b) will be added (as well as the others). Alongside Gu et al. (2017) we will also add Marelle et al. (2018) and Brönnimann et al. (2018).
This was very helpful, thank you. We have included information on the methods and further expanded this section, as indicated.

Page 19, Line 10: Not sure the other sections had this transitional statement so not sure it is needed here.
Thank you. We will consider deleting the sentence.
The sentence was removed.

Page 19, Line 25: "The effect of these nonstationary drivers, and their interaction, is complex. . ." Could insert "their interaction" to tie this to Section 5.5.
Will do.
Done.

Page 20, Line 5: I recognise most studies approximate this to 7% but 6-7% is probably more in line with background land temperatures.
We will update this.
Done.

Page 20, Line 12: Can it just be made clearer that the "longer duration" events statement is specific to the UK example. The way it is phrased it sounds like a global change which may not be true.
We will update this.
We removed this example, because the lengthening of events might actually be insignificant.

Page 22, Line 19: I appreciate the following reference isn't for extremes but it does provide a mechanism for more greening under climate change: https://doi.org/10.1038/nclimate2831
Thank you. A reference to Ukkola et al. (2016) will be added.
Thank you, this has been included.

Page 24, Line 9: The following reference could be relevant to the soil moisture-rainfall feedbacks: https://doi.org/10.1029/2018JD029762
Thank you. A reference to Holgate et al. (2019) will be added.
We have included this example: In Australia, both positive and negative correlations have been found between daily soil moisture and next-day rainfall, depending on spatial scale, location, and season (Holgate et al., 2019).

Page 24, Line 29: "Climate modelling framework in Section 4.2". I might have missed something as I am not aware of a climate modelling framework being introduced in Section 4.2. Maybe some rephrasing would help the reader follow here.

This will be rephrased/clarified.

This sentence was re-written for clarity (removing "climate modelling framework").

Page 26, Section 6.3: I think I expected a reference to the following which I think was the first study to attribute extreme rainfalls to human induced climate change: https://doi.org/10.1038/nature09763.

Thank you. A reference to Min et al. (2011) will be added.

We have used this example to introduce optimal fingerprinting, in a new paragraph.

Page 29, Line 4: Again, I am not sure the transitionary sentence is required.

We will consider whether to delete this sentence.

The sentence has been removed.

Page 29, Section 7: In the context of whether or not a system is vulnerable and requires management this following might be of interest: https://doi.org/10.1007/s10584-019-02497-4

Thank you. A reference to Nathan et al. (2019) will be added.

This reference has been added (explaining how the degree of "hydrologic stress" can be estimated).

Page 30, Lines 7-19: I did fail to see the relevance of much of this section given that non-stationarity deals with “climate” and hence longer time scales than those in numerical weather prediction. Maybe this section of text could be shortened as I feel it is more a pointer to Section 7.3.

Thank you. We will shorten the text accordingly.

We agree, and have removed the less relevant sections.

Page 30, Line 30: I would have thought the main limitation is the fact that in hydrology we are interested in projections at the catchment scale, but climate models work on resolutions much greater than this necessitating complex downscaling. I appreciate you pointed to this in the previous section, but I can’t help but feel this is a natural place for mentioning this point and describing it in a few sentences. I think a discussion on downscaling would not go astray here. See also: http://dx.doi.org/10.1016/j.jhydrol.2014.11.003

We will discuss the issue of downscaling here, and will also refer to Madsen et al. (2014).

A discussion on downscaling has been included, with the reference.

Page 32, Line 29: I am not sure that the presence of non-stationarity itself is controversial but whether it should be considered?

Yes. We will clarify this sentence.

Sentence has been updated accordingly.

Page 33, Line 3: Would it make sense to write out the definition of functional non-stationarity again here?

We might remove the term and instead highlight (and define) the different approaches.

We have removed this discussion of functional non-stationarity, as Francesco Serinaldi made some valuable suggestions in his Short comment.

Page 33, Line 31: Maybe "The" final step?

This will be altered.

Done.

All references are merged at the end of this document.

Response to Reviewer 2

We would like to thank Reviewer 2 for these positive and helpful comments which will improve the quality of our review. Below we provide the Reviewer's comments verbatim in black font, and our responses immediately below each comment in blue font.

A review paper serves a very specific purpose. A reader seeks to read a well-structured paper providing a well-developed synthesis of the literature in a clear and concise manner. This paper in my opinion succeeds in its role and offers an excellent addition to the literature. Since there aren't any major issues with the manuscript, I can only offer some minor points that might be useful to the authors. These are mainly some thoughts emerged as I was reading the manuscript line-by-line.

We are grateful to the Reviewer for this very positive feedback and for the suggestions made below.

1.11 This definition is too simplistic and misses the join properties. Yes, indeed this is a property of a stationary process, but we can easily create a process that has the same distribution over time but a changing autocorrelation. So, I you wish to keep it as simple as possible and avoid the formal definition just add ". . .statistical properties of the distribution and correlation do not. . ."

We will modify the sentence accordingly.

The sentence has been modified.

12.14 I am not sure if I understand this. Definition of nonstationarity is mathematical and precise; in simple terms any process that does not fulfill the formal mathematical conditions of stationarity is a nonstationary process. Thus this should not be linked with the data. Identifying nonstationarity or stationarity from data is another issue and doesn't differ than any other data-driven inference. Please clarify.

We agree and will remove the sentence.

This part has been completely re-written.

12.16 Whether a record is short or long, or sufficiently large for trend detection does not depend only on the absolute record length. So, I am a bit skeptical about such statements. Heavy tailed distributions for example introduce larger uncertainties.

This is a good point. The sentence will be removed.

Again, we have re-written this part to make it clear that the record length required for assessments of nonstationarity depends on the the type of process under consideration.

13.7 Not necessarily, the variance does not affect the significance of a trend and it is incorporated in the test. Or a process might change only in its very high values. Anyway, if you have a reference about this statement please add it as I am not sure if it is absolutely correct.

The original sentence was "Highly variable time series ... require a longer period of time for a significant signal to emerge....". Instead, we will simply state that for a highly variable time series it takes a larger percentage change in the mean of the data to identify a statistically significant change compared with a less variable time series (e.g. Chiew and McMahon, 1993).

The sentence has been updated accordingly.

13.19 This depends on what you define as trend here and is a bit confusing. If trends refer to a local systematic increase or decrease this is a property of the dynamics of a process or of the external factors causing this change. So it “feels” a bit confusing saying that trend depends on period of record. To clarify, a trend in the record is a trend anyway, but we assess if this is significant (for whatever reason) based on the properties of the process (inferred from the record).

Agreed. This comment will be removed ("trends depend intrinsically on the period of record and indeed on the quality of the dataset").

The sentence was removed to avoid confusion.

14.21 Just recheck this. I think the Theil-Sen is not a test, it’s just an estimator based on the median slope of all pairwise point of the record. Sometimes provides better results that the regression slope sometimes not. Still you can use also an intercept estimate based on Theil-Sen and show the fitted trend line.

Agreed; this is an error in the text. We will re-write and clarify this point.

This has been updated accordingly.

19.22 Large scale variability cannot be a driver of nonstationarity, unless you define nonstationarity in a “local” or short-term way. Long term variability causes local trends, but the process can still be stationary. This is very easy to show with MC simulations. For example, multidecadal oscillation can could cause multi decadal trends but this does not imply nonstationarity based on the formal definition.

We agree and will re-write this sentence.

The paragraph has been updated accordingly.

20.7 This is an assumption that circulates but it need more investigation as there are contrasting results regarding the light precipitation https://doi.org/10.1029/2019JD030855

The Reviewer is referring to the statement that "Extreme precipitation is expected to become more intense, and weaker rainfall less intense". We will clarify that this assumption needs further investigation, with reference to Markonis et al. (2019).

The sentence has been updated as suggested.

20.10 Also you can add information of the literature on convective non convective events and changes e.g. https://doi.org/10.1175/JCLI-D-17-0075.1, https://doi.org/10.1038/ngeo1731

Thank you. We will discuss these, including references to Park and Min (2017) and Berg et al. (2013).

These references have been added with the discussion of Clausius-Clapeyron scaling rates.

27.12 KS has a lot of theoretical issues and the AD should be favor. If I'm correct there's a tendency to slowly stop using the KS test.

Yes; we will mention this. The AD test is more powerful when comparing two distributions than the KS test (Engmann and Cousineau, 2011).

We have added a reference highlighting this point.

If the authors wish they can add more information on downscaling of climate model since it relates with nonstationarity, see e.g., https://doi.org/10.1029/2009RG000314

We agree and will include more information on downscaling, including a reference to Maraun et al. (2010).

A discussion of downscaling has been included, with the reference to Maraun et al.

Summarizing, I am happy I do not have much to report. This is well-written review paper on the topic. It was joy to read and it offers an excellent addition to the literature.

We are very grateful to the Reviewer for these kind comments and for their supportive review!

All references are listed at the end of this document.

Response to Reviewer 3

We thank Reviewer 3 for constructive comments on our manuscript. These comments are helpful and will improve the quality of our review article. Below we provide the Reviewer's comments verbatim in black font, and our responses immediately below each comment in blue font.

The authors present a review of methods and metrics for the detection, attribution and management of extreme events. Given the large body of literature on these topics, the authors have identified an area ripe for a solid review that will benefit the community.

The manuscript is a very large piece of work that contains a lot of useful and interesting material. In order for the reader to extract the most from this material, particularly new readers to this field, I would like the authors to enhance the educative components of this review. To this end I recommend improving the graphics (or adding tables) to distill the complexity for the reader into clear, concise overviews of the topics and discussions.

We are grateful to the Reviewer for their positive comments on our manuscript. We will strengthen the educational components of the review by enhancing figures or adding tables, as appropriate (details below).

Figures 1, 2, 5b, 6, 7, 8 are good, but Figures 3, 4, 5a should be revised into conceptual diagrams to help the reader. Figure 7 is a good conceptual diagram, but it is not tightly linked to the text, so could be replaced. For example, Figure 2 gives an overview of the workflow in this type of analysis, which is excellent, and the structure of the manuscript follows this workflow. Within each main section, or large sub-section, it would be good to have a conceptual diagram that helps the reader understand the relevant concepts within that section. The text could then be more tightly linked to these figures and then the reader would see the concepts and understand more the logic flow within each section.

We agree about ensuring the figures are consistently referred to throughout each section, to help guide the reader.

- Figure 3 (Metrics): we will ensure each of the panels is referred to within the text (if it is not already)
- Figure 4 (Examples of trends): we will ensure each of the panels is referred to within the text and we will also include a table which summarises the different indices, to help guide the reader.
- Figure 5a (Homogeneity): we will improve this figure and make sure it is clear.
- Figure 7 (Drivers of extremes): we will make sure this diagram is referred to and explained within the text; we think it serves a purpose and do not wish to replace it.
- Figure 3 (Metrics): We simplified the figure and made sure it was referred to at appropriate locations in the text.
- Figure 4 (Examples of trends): We have referred to each of the panels within the text. We have also added a new table summarising the different metrics.
- Figure 5a (Homogeneity): This figure has been improved as mentioned.
- Figure 7 (Drivers of extremes): we made sure this diagram was referred to at appropriate locations in the section on Drivers.

Another aspect that I would encourage further work on is linking more to existing reviews of detection and or attribution of extreme events to climate change. I was expecting reference to works like Hulme (2014), Zhai et al. (2018), Ummenhofer Meehl (2017) and Easterling et al. (2016).

Thank you. We will make sure we cite all existing works as appropriate, including references to Hulme (2014), Zhai et al.(2018), Ummenhofer and Meehl (2017), and Easterling et al. (2016).

We have referred to these reviews at appropriate locations in the text (such as in the new section titled "event attribution").

Overall, I am very happy with the content of the review – it is a major piece of work that draws together a lot of diverse material. I think this review could be made even more useful by improving the Figures to help the reader get the most out of this material.

We are very grateful for these positive comments and will implement the suggestions provided by the Reviewer.

Specific comments Page 2, line 12-13: Is deciding whether a time series should be treated as stationary or not for the purposes of managing extremes “one of the greatest challenges” facing scientists and practitioners today? I agree it is a challenge, but as you point out in this paragraph, the impact of picking the wrong model is a difference in uncertainty. Given the numerous uncertainties in this topic, I am not convinced this is one of the greatest challenges we face.

We will rephrase this sentence and instead highlight that it is a difficult topic.

This sentence has been modified; it is now described as "a key challenge".

Page 2, line 14: apparent trends in, and or correlations between, stationary longmemory / auto-correlated / smoothed time series has a long history, which could be highlighted by adding references to Slutzky (1937), Wunsch (1999), Yule (1926) and the Slutzky-Yule effect.

Thank you. We will discuss these points, with references to Slutzky (1937), Wunsch (1999), Yule (1926), and the Slutzky-Yule effect.

A discussion of these points, and the citations, have been added.

Page 3, line 9: another example of interdecadal to multidecadal hazard-rich and hazard-poor fluctuations is the flood- and drought-dominated regimes of rivers in eastern NSW, Australia (Warner, 1987).

Thank you, we will add reference to Warner (1987).

Added.

Page 3, line 12: the sentence “However, there is yet no comprehensive, introductory overview of these methods across hydroclimatic extremes, or overarching discussion of key challenges that can arise.” is a challenge to read. How about “However, a comprehensive, introductory overview of these methods across hydroclimatic extremes, including an overarching discussion of the key challenges that can arise, has not been published to date.”

Thank you, we will adjust the sentence structure as suggested.

Done.

Page 3, last paragraph: in the paper roadmap presented in this paragraph, no mention is made of the “discussion of key challenges” highlighted in the justification for the paper. Is the discussion scattered throughout the sections or presented in a particular location. Please signpost that discussion in the paper roadmap.

Yes, the challenges are described throughout the manuscript. We will rephrase the roadmap so this is clear.

A sentence has been added to indicate that challenges are presented for each step throughout the paper.

Section 2.1: There are a lot of different metrics for each variable extreme presented here. A table with one row per metric that summaries the variables, metrics, inputs, etc would be useful here to distill the complexity for the reader. An alternative to a table would be to revise Figure 3 (see next comment).

We will revise Figure 3 (next comment) and the accompanying text. Additionally, we may include a table that summarises the variables and metrics.

Figure 3 has been revised, and a table included.

Figure 3: many of the terms marked in 3a are not actually used in the text – loss curve, excess rainfall, centroid of rainfall excess. If the Table suggestion above is too difficult, then an improved figure to tightly link the ideas discussed in Section 2.1 with the image shown would help the reader. The idea of example conceptual time series for each variable with features of interest highlighted is good, but the features highlighted need to align with the text better. An improved version of this diagram could enhance the readers understanding of the material in this section.

We will improve Figure 3 as well as the text to make sure that all features highlighted on the figure are systematically described within the text.

We simplified Figure 3 to retain only the important terms, and then made sure these were mentioned in the text. We also included a new table describing different metrics.

Section 2.2 Figure 4: Again there is a lot to digest in this section and I don't think Figure 4 is helping the reader to understand the concepts. Rather than showing example results, I think a conceptual diagram of some of the concepts discussed would be more useful here. Then the text could be tightly linked to the diagram to help the reader follow the narrative and understand the differences in metrics.

We believe the current Figure 4 (examples of trends) is helpful for a reader who is new to the topic. It is difficult to envisage a what a conceptual diagram here might look like, but if we have an idea we will try to implement it. Additionally, we will ensure the current figure is better linked to the text.

We retained Figure 4 because we believe it is helpful for the reader to see applications of some of the different trend tests to different metrics. We have made sure that each of the panels is described in the text. We did not include a conceptual diagram, however we have included a table with examples of the different metrics.

Page 10, line 7: change "more strongly that the individual" to "more strongly than the individual".

We will alter this.

Done.

Page 12, line 11-12: you could mention the CAMELs initiative/papers here, which seek to publish large integrated hydrologic datasets for regions of the world.

We will mention the CAMELs papers in the revised manuscript, including CAMELS USA, GB, Australia, Brazil, Chile (Addor et al., 2017; Fowler et al., 2021; Coxon et al., 2020; Chagas et al., 2020; Alvarez-Garreton et al., 2018).

Description of CAMELS for different regions has been added.

Page 12, line 18: the paper by Thyer et al (2006) on how long a record needs to be for stochastic model identification (random, AR1 or HMM) could also be mentioned here as it suggests we need very long records to adequately identify our stationary models (let alone our non-stationary ones).

We will include a discussion of this with reference to Thyer et al (2006).

We have included this citation and explained this point in the text.

Page 13, line 2: change “with and perhaps a global "change point" in climate” to “with perhaps a global "change point" in climate”.

We will alter this.

The sentence was actually removed in the end for clarity.

Page 13, line 6: Variability of the time series is taken into account in the calculation of significance. It is better to say that for a high variability time series it takes a longer record to statistically identify a change of a given magnitude than the same change in a lower variability time series. See Chiew McMahon (1993) for a discussion of this issue.

Thank you. We will alter this sentence, with reference to Chiew & McMahon (1993).

We agree and have modified the sentence, with this citation.

Page 13, line 22-27: I totally agree about the importance of using long records for trend detection and the danger of short record lengths. Another area being explored to extend data back in time for insights into current conditions is palaeo-hydroclimatic reconstructions. For example, Freund et al (2017) reconstructed warm and cool season rainfall in Australia to then investigate recent observed trend magnitude in the context of palaeoclimatic variability. Hydroclimatic reconstructions of the last 500 years have great potential to place recent observations into a long-term context that is not achievable from short observation based record lengths alone.

Thank you, yes. We will include discussion of palaeo-hydroclimatic reconstructions and will include a reference to Freund et al. (2017).

We have included this helpful example in the revised text.

Page 14, line 15-23: I think people use the Mann-Kendall test because it does not assume a linear trend. The stated reason here is skewness – however, skewness can be resolved by transforming the data (Box-Cox). Linear regression only identifies linear trends, whereas Mann-Kendall identifies any monotonic trend, which is a much more useful/general feature of the Mann-Kendall test. Also the Thiel-Sen slope is not a test, it is just a way to estimate the magnitude of the trend. Also a reference to Hamed (2009a, b) for applying the Mann-Kendall test to auto-correlated data would be good to add.

Yes - we will clarify these points, including references to Hamed (2009a,b).

These aspects have been clarified, with the added citations.

Page 15, line 9: it is worth mentioning that the AIC and BIC assess the trade-off between goodness of fit and model complexity. More complex models are penalised, so they have to improve the goodness of fit enough to overcome the complexity penalty. At the moment the sentence is all about better fit, when it should be about better fit even when the increase in model complexity is taken into account.

Yes, indeed. We will make sure the trade-off is described in the text.

We have improved the explanation as suggested; thank you. This was an oversight.

Figure 6: While I agree the GA nonstationary model has the lowest BIC value and has a nice flat worm, it is interesting to note that the BIC values of the other three models are fairly similar to the GA nonstationary value. If you were to plot 6b for the other three models, would you also achieve an acceptable fit? Are any of the four models not an acceptable fit? It may well be that all four models are "acceptable", but the GA nonstationary is slightly more acceptable than the rest. If you present this Figure as an example of what can be done with GAMLSS models, then it would be good to more fully explore the results to justify the argument that the GA nonstationary model is the best model and the others are not.

We agree that when the difference in AIC/BIC is small, then other models may be equally acceptable (e.g. Wagenmakers and Farrell, 2004). We will include a discussion of this point in the text.

We have better explained in the manuscript that there are cases where multiple models may be equally acceptable.

Page 20, line 13+: in this discussion of changes in flooding in a warmer world, it would be good to highlight the finding from Wasko & Nathan (2019, Figure 7) that lower annual recurrence interval floods are more likely to be reduced due to drier antecedent soil moisture conditions, whereas higher annual recurrence interval floods are more likely to increase due to increases in extreme rainfall.

We will include this finding with reference to Wasko and Nathan (2019).

This reference has been included and discussed.

Page 23, line 23: remove the repetition of "to continue".

We will make the change.

Done.

Section 7.2: I was expecting to see mention of climateprediction.net and or weather@home (https://www.climateprediction.net/models/weatherathome/), which are ensemble runs of global or regional climate models that can be used to quantify internal ensemble variance and compare it to observed events.

A discussion of climateprediction.net and weather@home will be added, with reference to the key papers.

We have now described these ensemble experiments in section 6.3 on attribution, and included references.

References

Chiew FHS, McMahon TA 1993. Detection of trend or change in annual flow of Australian rivers. International Journal of Climatology, 13(6): 643-653.

Easterling DR, Kunkel KE, Wehner MF, Sun L 2016. Detection and attribution of climate extremes in the observed record. Weather and Climate Extremes, 11: 17-27.

Freund M, Henley BJ, Karoly DJ, Allen KJ, Baker PJ 2017. Multi-century cool-and warmseason rainfall reconstructions for Australia's major climatic regions. Climate of the Past, 13: 1751-1770.

Hamed KH 2009a. Exact distribution of the Mann–Kendall trend test statistic for persistent data. Journal of Hydrology, 365(1-2): 86-94.

Hamed KH 2009b. Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. Journal of Hydrology, 368(1-4): 143-155.

Hulme M 2014. Attributing weather extremes to 'climate change' A review. Progress in Physical Geography, 38(4): 499-511.

Slutzky E (1937). The summation of random causes as the source of cyclic processes. Econometrica, 5(2): 105–146.

Thyer M, Frost AJ, Kuczera G 2006. Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed record long enough?. Journal of Hydrology, 330(1-2): 313-328.

Ummenhofer CC, Meehl GA 2017. Extreme weather and climate events with ecological relevance: a review. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1723): 20160135.

Wasko C, Nathan R 2019. Influence of changes in rainfall and soil moisture on trends in flooding. Journal of Hydrology, 575: 432-441.

Warner RF 1987. The impacts of alternating flood-and drought-dominated regimes on channel morphology at Penrith, New South Wales, Australia. IAHS-AISH publication, (168): 327-338.

Wunsch C (1999). The interpretation of short climate records, with comments on the North Atlantic and southern oscillations. Bulletin of the American Meteorological Society, 80(2): 245–255.

Yule GU (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. Journal of the Royal Statistical Society, 89(1): 1–63.

Zhai P, Zhou B, Chen Y 2018. A review of climate change attribution studies. Journal of Meteorological Research, 32(5):

671-692.

Additional modifications

In response to other feedback from the community we have also improved and clarified our discussion of the concept of nonstationarity, and removed/improved various other parts of the text (all additions are highlighted in red font).

Combined References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., et al.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies–Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, 2018.

Barnett, T., Pierce, D., Hidalgo, H., Bonfils, C., Santer, B., Das, T., and Bala, G.: Wood, AW, Nozawa, T., Mirin, AA, Cayan, DR, and Dettinger, MD: Human induced changes in the hydrology of the western United States, Science, 319, 1080–1083, 15.

Berg, P., Moseley, C., and Haerter, J. O.: Strong increase in convective precipitation in response to higher temperatures, Nature Geoscience, 6, 181–185, 2013

Brönnimann, S., Rajczak, J., Fischer, E. M., Raible, C., Rohrer, M., and Schär, C.: Changing seasonality of moderate and extreme precipitation events in the Alps, Natural Hazards and Earth System Sciences, 18, 2047–2056, 2018.

Chagas, V. B., Chaffe, P. L., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth System Science Data, 12, 2075–2096, 2020.

Chiew, F. and McMahon, T.: Detection of trend or change in annual flow of Australian rivers, International Journal of Climatology, 13,643–653, 1993

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics, Springer-Verlag, London, //www.springer.com/gp/book/9781852334598, 2001.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J., Lane, R., Lewis, M., Robinson, E. L., et al.:CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth System Science Data,12, 2459–2483, 2020.

Du, T., Xiong, L., Xu, C.-Y., Gippel, C. J., Guo, S., and Liu, P.: Return period and risk analysis of nonstationary low-flow series under climate change, Journal of Hydrology, 527, 234–250, 2015.

Easterling, D. R., Kunkel, K. E., Wehner, M. F., and Sun, L.: Detection and attribution of climate extremes in the observed record, Weather and Climate Extremes, 11, 17–27, 2016

Engmann, S. and Cousineau, D.: Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test, Journal of applied quantitative methods, 6, 1–17, 2011.

Fowler, K. J., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth System Science Data Discussions, pp. 1–30, 2021.

Freund, M., Henley, B. J., Karoly, D. J., Allen, K. J., and Baker, P. J.: Multi-century cool-and warm-season rainfall reconstructions for Australia's major climatic regions, 2017.

Hamed, K.: Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data, Journal of hydrology, 368, 143–155, 2009a.

Hamed, K.: Exact distribution of the Mann–Kendall trend test statistic for persistent data, Journal of hydrology, 365, 86–94, 2009b.

Holgate, C., Van Dijk, A., Evans, J., and Pitman, A.: The Importance of the One-Dimensional Assumption in Soil Moisture-Rainfall Depth Correlation at Varying Spatial Scales, Journal of Geophysical Research: Atmospheres, 124, 2964–2975, 2019.

Hulme, M.: Attributing weather extremes to 'climate change' A review, Progress in Physical Geography, 38, 499–511, 2014.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T.: Review of trend analysis and climate change projections of extreme precipitation and floods in Europe, Journal of Hydrology, 519, 3634–3650, 2014.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user,Reviews of geophysics, 48, 2010.

Marelle, L., Myhre, G., Hodnebrog, Ø., Sillmann, J., and Samset, B. H.: The changing seasonality of extreme daily precipitation, Geophysical Research Letters, 45, 11–352, 2018.

Markonis, Y., Papalexiou, S., Martinkova, M., and Hanel, M.: Assessment of water cycle intensification over land using a multisource global gridded precipitation dataset, Journal of Geophysical Research: Atmospheres, 124, 11 175–11 187, 2019.

Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C.: Human contribution to more-intense precipitation extremes, Nature, 470, 378–381,2011

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., et al.: Frequency of extreme precipitation increases extensively with event rareness under global warming, Scientific reports, 9, 1–10, 2019.

Nathan, R., McMahon, T., Peel, M., and Horne, A.: Assessing the degree of hydrologic stress due to climate change, Climatic Change, 156, 2087–104, 2019.

Park, I.-H. and Min, S.-K.: Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature,Journal of Climate, 30, 9527–9537, 2017

Restrepo-Posada, P. J. and Eagleson, P. S.: Identification of independent rainstorms, Journal of Hydrology, 55, 303–319, 1982.

Slutzky, E.: The summation of random causes as the source of cyclic processes, Econometrica: Journal of the Econometric Society, pp.105–146, 1937.

Thyer, M., Frost, A. J., and Kuczera, G.: Parameter estimation and model identification for stochastic models of annual hydrological data: Is25the observed record long enough?, Journal of Hydrology, 330, 313–328, 2006.

Ukkola, A. M., Prentice, I. C., Keenan, T. F., Van Dijk, A. I., Viney, N. R., Myneni, R. B., and Bi, J.: Reduced streamflow in water-stressed climates consistent with CO 2 effects on vegetation, Nature Climate Change, 6, 75–78, 2016.

Ummenhofer, C. C. and Meehl, G. A.: Extreme weather and climate events with ecological relevance: a review, Philosophical Transactions of the Royal Society B: Biological Sciences, 372, 20160 135, 2017.

Wagenmakers, E.-J. and Farrell, S.: AIC model selection using Akaike weights, Psychonomic bulletin & review, 11, 192–196, 2004.

Warner, R. F.: The impacts of alternating flood-and drought-dominated regimes on channel morphology at Penrith, New South Wales, Australia, IAHS-AISH publication, pp. 327–338, 1987.

Wasko, C. and Nathan, R.: Influence of changes in rainfall and soil moisture on trends in flooding, Journal of Hydrology, 575, 432–441,2019.

Wasko, C., Nathan, R., and Peel, M. C.: Changes in antecedent soil moisture modulate flood seasonality in a changing climate, Water Resources Research, 56, 2020a.

Wasko, C., Nathan, R., and Peel, M. C.: Trends in global flood and streamflow timing based on local water year, Water Resources Research,56, e2020WR027 233, 2020b.

Westra, S. and Sisson, S. A.: Detection of non-stationarity in precipitation extremes using a max-stable process model, Journal of Hydrology, 406, 119–128, 2011

Wunsch, C.: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations, Bulletin of the American Meteorological Society, 80, 245–256, 1999

Yule, G. U.: Why do we sometimes get nonsense-correlations between Time-Series? A study in sampling and the nature of time-series, Journal of the royal statistical society, 89, 1–63, 1926

Zhai, P., Zhou, B., and Chen, Y.: A review of climate change attribution studies, Journal of Meteorological Research, 32, 671–692, 2018.