Hydrol. Earth Syst. Sci. Discuss., 11, C3703–C3709, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C3703/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Higher statistical moments and an outlier detection technique as two alternative methods that capture long-term changes in continuous environmental data" *by* I. Arismendi et al.

Anonymous Referee #2

Received and published: 10 September 2014

The work presented by Arismendi et al. represents a significant contribution to the discussion of environmental statistics, but I do not feel that it is ready for full publication yet. The authors present two techniques for assessing important shifts in the distributional properties of environmental variables. They first argue that higher-order moments, beyond the mean and variance, are better at capturing distributional shifts. They then present a technique for outlier detection, which is useful for the identification of potentially anomalous years. In my review, I am concerned with a number of statistical questions that I feel must be mentioned and addressed in the manuscript.

In general, the paper is well-written and the material is presented in a logical fashion, but I feel that a more rigorous justification and discussion of the results in necessary. Furthermore, I feel that there are a few statistical questions that must be considered. By more thoroughly discussing the points presented below, I feel that the authors will improve the impact and presentation of their findings. This is a valuable discussion, and with improvements, I feel it should be considered for publication.

Firstly, I would request that the authors revise the manuscript to reflect a more precise use of statistical language. I do not intend this to be a question of nit-picking, but I feel that more precise language will more clearly reflect the authors' intent and further substantiate their findings. To begin with, the authors state their intent to capture the 'variability' of the data. This term is not a concrete statistical term. As stated later in the manuscript, the authors appear to be more concerned with 'changes in [the] empirical distributions' (P:4731, L:7-8) and the 'shift in the shape of the . . . distribution' (P:4732, L:23). The term 'variability' does not describe the behavior of the entire distribution in a statistical sense; typically, this term is most closely related only to the second moment. In a similar vein, the authors note (P:4731, L:2-3) that metrics of central tendency do not capture this variability. I would argue that this is, of course, true, as metrics of central tendency are intentionally designed to capture just that, central tendency, not variability. By adjusting the language to reflect an interest in higher-order shifts in the distributional properties beyond the location, I believe the authors will make a better case for the limited utility of first and second-order moments.

The use of statistical terms and the discussion of statistical techniques could be improved elsewhere in the manuscript as well. By doing so, I think that the authors will make a more clear statement of their findings and avoid the pitfalls of loose language. For example, on page 4730, line 22, the authors refer to 'most traditional statistics' as relying on parametric assumptions of 'variability [which I read to mean distributional, or parametric, assumptions]. I think, as written, this statement is too broad and unsubstantiated. The term 'statistic' typically refers to a particular number or metric derived from data. Based on the next sentence, it seems that the authors mean to refer to statistical methods, not particular statistics. The distinction between statistics and statistical methods is important, the latter being potentially subject to parametric assumptions while the former are typically not subject to such assumptions (e.g. the mean can be calculated for any dataset, but interpretations and testing of the mean relies of parametric assumptions). Even if referring only to statistical methods, I would argue that the statement is still too broad: there exist whole field of non-parametric statistics. The authors appear to recognize this between pages 4730 and 4731, but dismiss this field and the use of non-parametric transformations as removing 'variability'. Certainly the authors would agree that ranks and transformations still possess distributional properties that can be assessed and, in some cases, translated to statements about the original data (e.g. if the logarithms of a variable are normal, the original variable is lognormal). It might be useful for the authors to point out particular methods or classes of methods that they are trying to improve, citing examples of usage in the literature.

In addition to precise terminology, I would like the authors to reconsider their application of the statistical tests presented on pages 4735 and 4736. Though I was not able to check that exact citation in Cramer (1998), from my reading of sampling properties in other statistical texts, the estimators of the standard error of skewness and kurtosis rely on parametric assumptions of normality. That is to say that the calculations of 'SES' and 'SEK' are only valid when a sample of size n is drawn from a normal population. I would ask that the authors provide further support for the use of these tests by showing normality or presenting evidence that this parametric assumption is not necessary. As it is, this seems to contradict the authors previous decrying of parametric methods (P:4730, L:22). Furthermore, the discussion of the null hypotheses and type-I errors could be improved. The null hypotheses should be that the skewness is zero and that the kurtosis is zero, respectively. With regard to type-I errors, the authors present a two-tailed alternative at a significance level of 0.05 and then draw one-tailed conclusions. In reality they are then conducting two 0.025-level tests, meaning that their conclusions are stronger than they state. Similarly, the authors conducted these tests on concurrent

decades. Typically tests are applied on independent datasets, but I wonder if temporal correlation may affect the power of these tests. It may be worthwhile for the authors to consider the true type-I error in the presence of temporal correlation.

With the above improvements to terminology and methodology, I would like the authors to consider the broader issues of temporal trends in statistical moments. The skewness and kurtosis, in a formal sense, are moment ratios, not raw moments. As such, the skewness and kurtosis are functions of the lower-order moments. Similarly, the variance, a centralized moment, is a function of the first moment. This functional dependence means that changes or temporal trends in higher-order moment ratios or centralized moments may be the result of changes in only the lower-order moments or higher-order moments. For example, a trend in the skewness could be attributed solely to a trend in the mean, with the second and third raw moments remaining unchanged, or attributed to a change in the raw third moment. Because skewness and kurtosis, as opposed to the third and fourth raw moments, might be more environmentally relevant, this concern may not be huge problem. Still, I think it is important to note that trends in these moment products are inter-related, making it difficult to attribute changes to any single driver; a comment to this effect might prove valuable.

The authors present a technique for identifying anomalous years in a time series. While this is indeed a useful technique, I am concerned that it only identifies years in the tails of the distribution rather than true outliers. Merely being in the tail of a distribution does not make a point an outlier. I feel that the term outlier may inadvertently connote an erroneous or otherwise concerning value. As executed, the method presented identifies points beyond the 95% region as outliers. We would, by definition of the 95% region, expect to find 5% of the observations outside of this region due to pure chance. Looking at the figures and tables in the manuscript and supplement, I do not see strong evidence that significantly more or less than 5% of the years at a given site, on average, were outside this region. For this reason, I think it is important to note on that the years identified are the most extreme, not necessarily 'outliers'. Additional tests would be needed in order to identify outliers as such. In large part, this is an issue of terminology, but I consider it an important distinction. With this distinction, I think that the authors have not shown that more or fewer points are 'outliers' in the regulated and unregulated sites. More interesting, to my mind, is the irregularity of the region at unregulated sites, when compared to the regulated sites. The authors make this last observation (P:4739, L:27-29), but I would love to read more discussion.

I think that this manuscript could benefit from a more thorough introduction and a more concrete discussion. While useful and well-written, I felt that the introduction left me with lingering questions supporting the justification and applicability of this work. I found myself struggling with some key questions: By providing clear answers, I think the manuscript would motivate the work more strongly and imperatively. What is the exact goal of this work? To identify change in the distribution not captured by trends in the first and second moment. What is this work important? Environmental changes may affect the distribution beyond the first and second moment. Ecosystems and organisms are sensitive to such distributional changes. How is this work different from previous works? The authors made this point clearly, but did not demonstrate a marked improvement: This work looked at higher-order moments. By providing a clear, well-referenced discussion of these points, I believe the manuscript will provide a much stronger case.

In the Results and Discussion section, I feel that some of the conclusions are only loosely substantiated. The authors may be served by expanding the Results section and relabeling it a Results and Discussion and relabeling the Discussion as a Summary and Conclusions. As it stands, the Discussion makes a series of claims that I am concerned with. In the first paragraph, and throughout the manuscript, the authors argue that they have showed higher moments to be an alternative to lower-moment analysis. I think this reads too much like lower-moment analysis should be rejected in favor of higher-moment analysis. Because no comparative analysis was presented, showing that higher-moment analysis identified trends were lower-moment analysis did

not, I do not feel that this claim is substantiated. Unless this comparative analysis is presented, I think it important to claim instead, as the authors imply on page 4741, line 18, that higher-moment analysis provides only complementary information. In the second paragraph, the authors state that the outlier detection technique presents a 'more complete and realistic view' of the data. Against what is this comparative statement made? The authors then argue the distributional analysis is more appropriate than single-metric analysis. I agree with the latter, but do not see how it is directly related to outlier detection. Finally, on page 4741, lines 21 through 24, the authors claim to have shown that water regulation masks climate influence. I believe this, but I do not think this is substantially supported by the results. The authors showed differences in temporal trends across regulated and unregulated sites, but I do not think a definitive conclusion is justified. It might be better to include this thought only as a discussion point or conjecture.

I would like to thank the authors for their work. I believe that this is a very interesting study. The call for distributional analysis of environmental variables will have a significant effect on how we, as a field, consider stationarity. I commend the authors on this excellent first effort. I hope that my thoughts are helpful in improving the project. The authors' writing was well-executed, presenting a concise treatment of their work. In addressing my comments, I hope that the authors will retain this style. I look forward to the next iteration of this manuscript.

Technical comments: (I have tried not to tread the same ground as the previous reviewers.)

P:4731, L:9: The authors introduce the concept of 'regimes'. I am not sure what is meant by these, as it seems to be equivalent to distributions of the environmental variable. I do not think it is necessary to introduce the idea of regimes. This idea does not seem relevant later in the manuscript.

P:4731, L:9-11: This sentence starts with the word 'Typically' and concludes that these ‘typical’ methods are not used in practice. This strikes me as inconsistent: if they are not used, are they typical?

P:4731, L:21: Change ‘in stream temperature’ to ‘in the distribution of stream temperature’

P:4732, L:14-17: The sentence starting with ‘For example…’ is a bit awkward. The statement is not clear. I think that the readability could be improved by revising or adding something like ‘by’ in ‘… captured [by] only using…’

P:4734, L:1-8: Why is this introduced? The authors state that this transformation is used to compare values across sites, but I do not think that this comparison is ever presented.

P:4734, L:11: Change ‘changes in environmental variables’ to ‘changes in the distribution of environmental variables’.

P:4736, L:20-23: This sentence could use revision. I think that the comma splice midway through makes the sentence seem too general. It reads almost like a result or discussion point, though this is in the methods section.

P:4738, L:2: Change ‘Stream temperature empirical distributions’ to ‘Empirical distributions of stream temperature’

P:4740, L:26: Change ‘here take advantage’ to ‘here takes advantage’

P:4741, L:15: To whom is ‘their’ referring? The antecedent is vague, not agreeing in quantity with the opening clause.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 4729, 2014.