Referee Comments 2:

The authors of "Error in hydraulic head and gradient time-series measurements: a quantitative appraisal" provide interesting discussion of a fundamental concern in evaluation of field hydrogeologic data. As such, the paper has potential to make a significant contribution to the hydrologic literature. In presenting the following comments, I should note that I am principally an academic with substantial field experience. I believe that I may have approached review of this manuscript from a different viewpoint than did the other referee and the other reviews already received. I hope that this difference in viewpoint is useful to the authors.

We thank the reviewer for her/his valuable assessment.

In this vein, one overall comment that I would introduce beyond those comments already provided by Kennel and the other reviewers is that it is somewhat unclear whether this paper is intended to be a basic discussion for those working in the field (in which case some of the additional sources of error suggested by the other reviewers might be considered) or a more theoretical analysis to help inform further study for improving estimation of hydraulic gradients and fluid fluxes in complex groundwater systems (for example, use the sources of error identified by these authors, but place them into a random numerical analysis in an effort to provide more insight into the most important errors within multiple field scenarios such as local, three-dimensional flow versus regional flow). Specifically, in reviewing this manuscript, I found that the discussion of the details of the field technologies (tapes, transducers, dip instruments) was quite fundamental (e.g., discussing the increment of measurement on a depth-to-water tape) without discussion of possible improvement, while the discussion of the magnitude of errors (and lack of discussion of interaction among errors) involved a number of assumptions. The paper has potential to be a valuable contribution, but I believe that it would benefit substantially through a bit more clarity on the intended audience (e..g, field technician versus more theory-based hydrologist) and a bit greater effort to more thoroughly understand the interplay in the identified errors. I also believe that this suggestion is reflected in some of the comments of the other reviewers (e.g., interplay of errors as suggested by Fang, the suggestions for additional types of errors by Rosenberry, the comments by Kennel et al that the example magnitude of errors should be based on a broader range of field experience and placed within context of reasonable error expectations).

Our target audience is the groundwater community at large. We believe that our review will benefit those that collect and use field data as it raises awareness about potential sources of errors, some of which are often overlooked, as noted by the first reviewer. But at the same time we think that this paper is equally relevant for academics and theoretical groundwater modellers, even if they never collect any field data themselves. Hydraulic head data is frequently taken from public databases or other third-party sources to be used in model calibration and without in-depth understanding of the causes and magnitudes of measurement errors, the limitations posed by the data accuracy may not be fully appreciated by the user(s).

We note that the scientific objectives determine the required accuracy. Hence, scientists must be as aware of the operating procedures as the field practitioners who make the measurements. Otherwise,

they cannot design their research. Essentially, the paper is written for both cases but mainly for academics. The aspects that we highlight are often unknown or ignored which leads to bad data (and thus conclusions). We hope that this aim fulfils the reviewer's expectations.

More specific comments:

For many of the conclusions put forward by the authors, it might be beneficial to both suggest the implications for measurement precision in the field and avoid comparisons / generalizations that cover only a partial range of field experiences. For example, the abstract suggests that uncertainty in the hydraulic gradient "magnitude can have as great an effect on the uncertainty of flow rates as the hydraulic conductivity". I would note that these two aspects of groundwater flow analysis are fundamentally different in terms of impact on flow rates, flow direction, and response to hydraulic stimuli. I might avoid this simplification of error comparison as it will require substantially great discussion in terms of impact on the type of final analysis desired. Further, the authors suggest that 170 meters measurement point separation is required to achieve an estimate of 10ˆ-4 in the field. Note that this implies (in perfectly one-dimensional flow) an error in differential head measurements of approximately 1.7 cm. The authors might be direct about this allowable error and briefly discuss whether this is a reasonable field result.

We respectfully disagree. Our analysis clearly shows that the accuracy of the head change is dependent on the distance between the measurement points. While it is impossible to measure 1.7 cm over 10 km, it might work better over a distance of 10 m. Therefore, casting the discussion in terms of gradients is much more useful than head differences.

As noted by the other reviewers, situations in which gradients in vertical flow are of interest will often involve impact of geologic heterogeneity, natural transients (e.g., due to precipitation), and anthropogenic impacts (e.g, pumping from wells or differential densities near contamination sources). Clarification of concerns, and where those concerns are important, need to be clarified in terms of vertical gradients / vertical flow. Once again, the authors are encouraged to avoid making generalizations. This is particularly of concern in that suggestions from this manuscript might be adopted by field technical staff without careful review of the field conditions assessed by the authors and applied under conditions that are not appropriate.

We agree and will include discussion of realistic gradients into our manuscript. Please note that our generalisations are required in order for the complexity of this topic to be simplified to the reader. We will aim to include appropriate caveats wherever necessary.

I would agree with comments in one of the other reviews regarding the authors' suggestion that complexity in automated water-level measurement is significantly more complex than manual measurements. Specifically, this is perhaps inappropriate and likely ignores the complexities involved in making repeated manual measurements.

In our manuscript we say that the technology involved in making automated measurements is more complex. This includes the data processing, for which automated measurements and QA are much

more complicated compared to manual measurements. We hope that our manuscript reflects this viewpoint accurately.

After equation 6, the authors state that del(H) is continuous. In the presence of any type of heterogeneity, this is not necessarily the case (think for example, about the instantaneous change in del(H) in the vertical direction as we move from a high K to a low K material). This statement should be corrected. More importantly, as noted in one of the other reviews, heterogeneity makes the discussion of errors in the hydraulic gradient far more complicated than even presented in this manuscript. Well geometry, well screen length and location relative to changing hydrologic units in the subsurface, screen clogging, regional variation in pumping (other wells not part of a given study), the distance of well separation relative to the scale of heterogeneity, are all errors that make this analysis far more complex than presented here.

While there may be rapid head changes in space, we believe that true head discontinuities are rare (e.g. seepage faces). We note that the focus of our manuscript is explicitly on hydraulic heads and not on geological heterogeneities. We will include this as a caveat into our revised manuscript.

There is substantial concern that the authors have artificially separated "horizontal" from "vertical" gradients (e.g., equations 6-9). Certainly at the lengths scales for which errors in water levels have substantial, negative impact on our field analysis, there is no reason to make an assumption in advance that the flow field can be separated into horizontal and vertical flow and that such analysis does not vary rapidly in space. Why not base the discussion in the paper on analysis of the error in the direction and magnitude of the three-dimensional hydraulic gradient?

We do not understand the difference between what we have done and what the reviewer requests us to do. We have merely broken down a 3D flow field into its cartesian components, horizontal and vertical. Many measurement techniques require separate horizontal and vertical errors. This is standard practise in groundwater investigations and modelling. We do not see the need to revise our manuscript in response to this comment.

On page 15, the authors make some strong, sweeping conclusions about nonverticality of wells. I agree with one of the other reviewers that the authors could assist the reader by providing a bit more insight here. For example, for what minimum depth of well and in what geologic conditions will this error be most likely to impact field analysis? Further, the suggestion to use geophysical measurements to measure vertical deviation in all wells in all projects is likely beyond the financial capacity of many field efforts.

We were surprised to see very little literature regarding this topic. We use published statistics to show that non-verticality is a serious issue that has been neglected by the hydrogeological community. In our error example, we use a realistic deviation and a small well depth to show just how large the error from non-verticality can be. We therefore disagree that our statements are sweeping and refer to RC1 who specifically appreciates us raising this issue. We agree that there are serious financial

implications, and certainly it may not always be feasible to fulfill this recommendation, but in that case the uncertainty should be acknowledged and an assessment of the potential error included.

Starting on page 17, the authors make several assumptions regarding the error in the measurements of pressure transducers. As noted in at least one of the other reviews, the precision and time drift in a pressure transducer is strongly dependent on a number of factors including the type of transducer, the maximum range, and quality of construction. A bit more discussion of the range of likely precisions to be observed in the field and, as suggested by Rosenberry, careful field design can provide an opportunity to optimize field instrument design to minimize instrument errors.

We will try to include more discussion about factors that degrade the precision of pressure transducers into our revised manuscript. We had explored the idea of providing a comprehensive list of the most common manufactures and pressure transducer types with reported specs, however, this was beyond the scope of this paper. It would however, be a valuable follow up to this study.

I would prefer if figures 10a and 10b were presented on the same vertical scale (with some data in figure 10b shown as off range on the graph) so that the reader can actually compare the majority of the data presented.

We agree and will make the scales in Figures 10a and 10b equal.

Page 28 - I agree with the other review comment that the sampling interval of 1 hour seems arbitrary and too long. Perhaps reconsider this suggestion.

An optimal sampling interval is a controversial issue. Here, a balance has to be found between generating too much data for groundwater responses that are slow and missing details for dynamic systems. We will attempt to discuss this in a bit more detail in our revised manuscript.

Again, I believe that there is a potentially valuable paper presented here. I would, however, encourage the authors to consider the comments of the other reviewers as well as the comments presented here as an opportunity to substantially increase the applicability and value of the discussion presented.

We thank the reviewer and will do our best to address all the comments received during the review process.