

Referee Comments 1:

This manuscript presents a wake-up call to the hydrogeology community that is both disturbing in its findings and exceptionally thorough and helpful. Although many of the points made throughout the manuscript are also presented in hydrogeology textbooks, they have rarely been combined in one place nor have they been researched and updated so thoroughly. Many field practitioners have been lulled into complacency by the stated accuracies of sensors capable of measuring and recording water-level data at whatever time increment is desired, forgetting that errors in manual measurements, if they are still made at all, are in addition to automated sensor errors, some of which are rarely stated or considered. The thorough listing of sources of error and their potential relative magnitudes, particularly with regard to interpretations of horizontal and vertical hydraulic gradients, will be a useful resource.

At first, I wondered how much of an improvement this manuscript would be compared to the excellent review of this topic by Post and von Asmuth (2013). As I read further, I was impressed by the thorough coverage of sources of error, including a wide range of errors that most readers likely have not previously considered. The manuscript represents a substantial step forward in the somewhat mundane and yet very important process of collecting accurate water-level data from monitoring wells and piezometers.

The manuscript is also very well written and was a pleasure to read. Figures are clear and convey important points very convincingly. Citations to the literature are appropriate and the authors do a good job of presenting newer capabilities relative to those from decades ago.

We thank the reviewer for the time taken to assess our manuscript and are pleased to receive such positive feedback. Our detailed answers are below.

Specific comments

Page 3, L30: Errors associated with non-vertical boreholes seem to have largely been forgotten and I was glad to see mention of this problem here. Most of the time, this error is insignificantly small. However, this error is common and, if unknown, can lead to substantial misinterpretation as the authors point out, particularly for deeper boreholes.

We do not see the need to revise our manuscript in response to this comment.

Page 7 L5: In addition to a lag related to barometric efficiency, many wells also suffer from a lag in the water-level response to changes in formation pressure, either because the well screen is partially clogged or improperly sized, or because the well diameter is so large that water cannot flow fast enough through the surrounding porous media and well screen to allow rapid equilibration of the water level inside the well to surrounding pressure changes (e.g., Hvorslev, 1951). This point is not directly applicable to water-level measurement accuracy, but it could confound interpretation of barometric efficiency and could be added to this section for completeness.

We agree and will include the possibility of time lags between formation pressure and water level.

Page 9: This is a very nice description of the various “accuracy” indicators and how they differ. I particularly valued the ADC component and how resolution is dependent on sensor range. I had not seen that before.

Thank you. We do not see the need to revise our manuscript in response to this comment.

Page 10 L24: The authors may want to mention the commonly used surveying technique of closure, or leapfrogging from a known point to unknown points and then back to the known point. Ideally, the beginning and ending locations (or elevations if surveying on the vertical axis only) will be the same, and the difference will give the user a good indication of the total-survey accuracy.

We agree and will include a mention of leapfrogging into our revised manuscript.

Page 12 L31-33: What saves us with regard to measuring vertical gradients is that vertical gradients tend to be much larger than horizontal gradients (both because of anisotropy and because of greater formation heterogeneity in the vertical axis, and also because piezometers designed to measure vertical gradients commonly have very short screened intervals. In the case of grouted-in piezometers, many piezometers are open only on the bottom making the screened interval essentially zero. Although the authors’ points are valid, they might want to add that measurement errors can be minimized when determining vertical gradients with appropriate piezometer design and method of installation.

We agree and will include the influence of piezometre design on the vertical gradient detection into our revised manuscript.

Page 16, L19: Here, the authors first mention use of a “dip meter” to make manual water-level measurements. One common problem with these devices is that many designs require displacement of some water before the upper sensor makes contact with the water in the well, creating an audible beep. This displacement causes minimal error for larger-diameter wells, but when the well diameter is not much larger than the diameter of the sensing device, this can cause a substantial artificial rise in the water level in the well due solely to the dip-meter measurement. Authors might want to indicate this source of error here or perhaps later in the manuscript where they talk about dip meters not having improved over the years.

We note that this issue is discussed in some detail in Post and van Asmuth (2013). We will add a note of this into our revised manuscript.

Page 17 L1: The authors mention the need to re-survey well elevations in areas of unstable land surface. One source of instability that was not mentioned, but that should be included, is soil frost. Soil frost can result in vertical movement of well casing on the order of tens of cm per year. For monitoring wells installed where the depth to the water table is small, a common occurrence near lakes or streams or wetlands, well movement due to soil frost can be a large problem that requires

annual re-surveying of the altitude of the top of well casing. I have seen monitoring wells jacked completely out of the ground after several seasons. I was looking for a citation to a paper that discusses this substantial problem but the only place I found mention of this problem is in Rosenberry et al. (2008) where they write: “Shallow well casings can move vertically in response to pumping for water-sample collection, frost, and settling of well cuttings placed in the annular space between the well casing and undisturbed sediments. This is particularly common for wells installed in wetland sediments. Shallow wells constructed with plastic casing can break from ice expansion during subfreezing temperatures. Wells and surface-water staff gages located near a downwind shoreline also can be tilted, moved horizontally, or broken if surface ice is pushed onto the shoreline during fall freeze or spring thaw. Annual leveling surveys are necessary for surface-water staff gages, as well as many near-shore wells, in order to document changes in the elevation of the staff gage or the top of the well casing.”

Thank you. We have experienced another example of changing reference point on swelling clays in Australia. There, the whole monument was floating above the ground under dry conditions. We agree and will include these points into our revised manuscript.

Page 20 L15: Transducer and particularly barometer error due to exposure to large temperature variation is a problem that few practitioners are aware of. An easy solution is to hang the barometer inside a well casing below ground surface but above the highest expected water level. Figure 6 provides an excellent example of the effect of allowing the barometer to be exposed to large temperature variability. However, one sentence in the figure caption is not supported by the data. The writers state in the caption, “Note that the manual dips confirm that there was no diurnal variability in the water levels (blue dots).” The manual data do not show this, nor can they. Numerous manual measurements made on the same day would be required to show diurnal response to temperature or the lack thereof. I suggest this sentence be removed from the figure caption, or perhaps be altered to indicate that the manual measurements indicated that the corrected water-level data adequately reflected the changing water level in the well.

We agree and will remove the reference to confirmation from naula dips from our manuscript.

Page 26 L16: These data regarding sensor clock drift are very disturbing. I also notice clock drift and correct for it each time I download sensor data, but I’ve never encountered drift this bad. I hope this is atypically bad compared to other sensors. If so, you might want to state that this table represents a perhaps extreme example.

These data were acquired using standard industry loggers that are sold with the promise that clock accuracy is +/- 1 min/year. Because we have noted clock drift over the years of acquiring groundwater hydraulic heads, we decided to test our standard loggers with the results presented in our manuscript. We believe that this is a realistic example.

Page 27 L9: Another source of error that has not thus far been mentioned is the offset created by displacement when a transducer is lowered into a small-diameter well that is slow to respond. If a

manual measurement of depth to water is made prior to installation of the pressure transducer, output from the transducer will be related to that depth-to-water value. However, if the water level in the well rises due to displacement of water when the transducer is lowered into the well, there will be an artificial offset in the relation between transducer data and the manual water-level measurement. It is better to lower the transducer first, and then relate transducer output to a manual water-level measurement that is made at or close to the same time as a programmed sensor scan.

We agree and will discuss the possibility that water displacement can change a measurement in our revised manuscript.

Page 30 Fig. 12: It took me a while to figure out what is conveyed in this very useful figure. I think a slight change to the figure caption would help lead the reader to a more efficient understanding. I suggest you revise to write “Visual comparison of horizontal and vertical random errors based on precision values (from Table 1) (note that some errors are distance dependent) for the different steps (Figure 1) and method options (Table 1).” Also, a “for example” sentence might help, either in addition to or instead of the black boxes indicated in the panels. I found it somewhat surprising that your calculated errors of about 15 percent were the same for your examples for both horizontal and vertical gradients. If you could give an example of a measured delta h, a calculated percent error based on assumed conditions and values from Table 1, and then the resulting minimum HHG or VHG, that might more clearly convey the usage of the figure.

We agree and will revise the figure caption according to the suggestion.

Page 33 L2-3: The authors state that “improved standards for water level measurement would be an important step towards better hydro(geo)logical data quality and consistency.” Standards have existed for many years that remain robust and are still appropriate for modern use. One good example is from USGS (Freeman et al., 2004, p. 16). You might consider mentioning those standards as a goal that could extend more broadly throughout the hydrogeological community: “A water-level sensing and recording system should be capable of performing within a measurement error of + or – 0.01 ft. for most water-level measurement applications. For the case of large changes in water level (for example, during aquifer tests), this measurement error may not be achievable, and an accuracy of 0.1 percent of the expected range in water-level fluctuation is acceptable. Where the depth to water is greater than 100 ft, an accuracy of 0.01 percent of the estimated depth to water is generally acceptable.”

We agree and will add those goals to our manuscript alongside a reference to the source.

Page 34 L32: Just as you stated regarding use of “dippers,” it is somewhat surprising that we still are using primarily silicon strain-gage pressure transducers. Quartz oscillator pressure transducers have been available for many years but remain little used by hydrogeologists, largely because of cost. A mention here of the exceptional accuracy of these devices might generate increased interest and demand from hydrogeologists, which may lead to larger sales and eventual reduction in unit costs.

We agree and will add a mention of Quartz oscillator pressure transducers into our revised manuscript.

Page 35 L1-5: I agree that vented transducers are better where their use is appropriate, but mention of the concern over keeping the vent tube unclogged and the desiccant materials fresh should also be included here. Errors resulting from improper maintenance of vented transducers can be as large as errors associated with the use of a non-vented transducer and associated barometer.

We agree and will include a short discussion about the maintenance of vented transducers into our revised manuscript.

Page 35 L11: Use of a transducer with a smaller pressure range to improve accuracy is another important point that often is lost. Many studies make use of transducers that have a large operating range and that are installed near the bottom of a monitoring well, when a 34 kPa transducer could be deployed at a much shallower depth with substantially greater accuracy, and for no additional cost.

While we agree we caution in the case of variable density fluid inside boreholes (refer to Figure 7). Thanks for confirming our observations. We do not see a need to revise our manuscript in response to this comment.

Technical corrections

Page 1 L7: Change measurements to measurement.

Page 2, L 8, 22: Why do you write hydro(geo)logical with parentheses around geo? Hydrogeology is a commonly used word that is in virtually every dictionary. There is no reason for the parentheses when hydrogeology is used as an adjective.

Page 10 L12: You write, “reflecting of a target”. This should be changed to “reflecting off of a target” or perhaps “reflecting from a target.”

Page 15 L5: Change an to a to write “a gyroscope.”

Page 16 Fig. 4: You should add titles for the x and y axes to indicate the units used. I assume they are m and mm, but you should state that for clarity. Also, I do not understand what you are conveying with the second y axis on the right side of the chart where values are listed in the order 0, 25, 50, and 5.

Page 17 L10: Consider changing unimpeded to unattended.

Page 23 Fig. 8: The y axis in panel a of Figure 8 appears to be labeled incorrectly. The axis is titled “Depth to water” but that implies that the water level inside the well (the depth to water) changed on the order of 60 to 70 m with temperature. That clearly cannot be the case. I suspect this actually is the water temperature at various depths within the standing water column inside the well. Therefore, I

suggest the axis title be changed to something like “Depth below ground surface” or “Depth below water surface in well”.

Page 26 Fig. 9: This is another excellent example of a common problem that all too commonly is ignored or unknown.

Page 26 L2: You might want to mention that a field laptop used for this purpose should be set to not automatically update to societally driven artificial changes in the clock, such as daylight savings time.

Page 28 Fig. 10: I assume the units on the y axis are meters because those are the units for your previous figures. However, for clarity and consistency, this should also be indicated in this figure.

Page 33 L21-22: You might add that the concern about non-vertical boreholes is a minor concern for wells that are relatively shallow. You might even include a threshold depth to water of xx m, below which most situations would result in errors that are inconsequentially small.

Page 34 L3-6: This reminder that manual measurements are still required is a very important message to convey and I was happy to see it included and emphasized in the conclusions.

We agree with all these technical corrections and will revise our manuscript accordingly.

References cited:

Freeman, L.A., Carpenter, M.C., Rosenberry, D.O., Rousseau, J.P., Unger, R., and McLean, J.S., 2004, Use of submersible pressure transducers in water-resources investigations: U.S. Geological Survey Techniques of Water-Resources Investigations 8-A3, 50 p.

Hvorslev, M.J., 1951, Time lag and soil permeability in ground water observations: U.S. Army Corps of Engineers Waterways Experimental Station Bulletin No. 36, 50 p.

Post, V.E.A., and von Asmuth, J.R., 2013, Review: Hydraulic head measurements - new technologies, classic pitfalls: Hydrogeology Journal, v. 21, no. 4, ~ p. 737-750.

Rosenberry, D.O., LaBaugh, J.W., and Hunt, R.J., 2008, Use of monitoring wells, portable piezometers, and seepage meters to quantify flow between surface water and ground water, in Rosenberry, D.O., and LaBaugh, J.W., eds., Field techniques for estimating water fluxes between surface water and ground water Denver, U.S. Geological Survey Techniques and Methods 4-D2, p. 39-70.

We will add these literature suggestions as citations in appropriate places to our manuscript.

Referee replies to Discussion contributions

Additional contributions from Bian and Kennel et al. provide several helpful and insightful thoughts for the authors to consider. Regarding Kennel et al.'s comment no. 3 about vented versus non-vented

transducers, they make a good point about not needing barometric corrections when determining horizontal gradients using multiple non-vented transducers. I also do not bother with barometric corrections when I am using two transducers to provide data related to determination of vertical gradients. Their comment about substantial noise from the non-vented transducer in Figure 5 also raises an important point that I had missed in my review. My experiences have been similar to theirs; unless I am using a rather poor-quality transducer, I get much better response (smaller noise in the data) for barometrically corrected non-vented transducers than what is shown in Figure 5. Authors may want to provide specifications for the non-vented transducers that provided these data.

Our data is based on a cheap pressure transducer which shows more noise than the others. We will make sure to mention this so that the reader does not walk away with the impression that non-vented transducers are inferior.

In response to Kennel et al.’s Other comments and specific notes, their question no. 3 about grouted-in applications is also a concern of mine. That situation makes me very nervous. In such an installation, we have no chance to make manual measurements once the transducer is installed. We must simply trust that the transducer is operating according to specifications. One solution is to install grouted boreholes with transducers in triplicate for each measurement installation. The authors may want to raise this consideration in their concluding remarks.

We agree and will include a short discussion about the risks of grouted in piezos as well as the need for redundancy with regards to measurement instruments.

Kennel et al.’s comment/question no. 13 is also one that I had missed. Use of a wet-wet transducer is appropriate for many groundwater-surface-water installations where the need to measure vertical gradients exists and yet I see little evidence of their use in the literature. This comment is somewhat buried at the end of the concluding remarks. As Kennel et al. point out, it would be a good idea to mention the existence and special features of these transducers earlier in the manuscript.

We agree and will discuss the advantages of wet/wet pressure transducers in our revised manuscript.