

First off, we thank the Editor, reviewers and readers for their efforts in handling, reading, assessing, reviewing and commenting on our discussion article published in HESSD. We have received many positive and constructive comments during the public discussion and review. We now take the opportunity to reply to all these comments in more detail. We note that none of the comments have provided criticism that could alter the overall message. We hope that our replies will lead to the decision that we can revise the manuscript and, finally, to a consideration for publication in HESS. To make the assessment easier, we have colour coded our replies in the categories agreed (**green**), partially agreed (**orange**) and disagreed (**red**). Our explanations of changes to the manuscript are highlighted in **blue** colour.

Additional revisions that go beyond the reviewers' recommendations:

- We added to the acknowledgements: “MAS was supported by funding from the Australian Research Council, grant DE150100302. We thank the editor Brian Berkowitz for handling, the reviewers Don Rosenberry and anonymous for assessing and some attentive readers for suggesting improvements for this work.”
- We added a citation to Bredehoeft (1967) and McMillan et al. (2019) in appropriate places (page 21 lines 4, 5 and page 34 line 25)

Bredehoeft, J. D.: Response of well-aquifer systems to Earth tides, *J. Geophys. Res.*, 72(12), 3075–3087, doi:10.1029/JZ072i012p03075, 1967.

McMillan, T. C., Rau, G. C., Timms, W. A. and Andersen, M. S.: Utilizing the Impact of Earth and Atmospheric Tides on Groundwater Systems: A Review Reveals the Future Potential, *Rev. Geophys.*, 2018RG000630, doi:10.1029/2018RG000630, 2019.

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.

A new paragraph was inserted on page 30 in line 5 to describe time lag effects. Since it applies to multiple processes in addition to barometric effects, we chose to include the paragraph under section 6.6 Miscellaneous errors.

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.

The following sentence was inserted on page 10 line 15 “*An indication of the measurement error can be obtained by returning to the starting location of the survey and determine the difference between the recorded positions at the start and end.*”

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.

We have included the following sentence on page 14 line 15 “*When gradients are higher the error can be minimised by using as short a screen as possible, taking care that any pressure difference between the GMI and the formation is rapidly equilibrated by water movement (see also Section 6.6).*”

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 von Asmuth (2013). We will add a note of this into our revised manuscript.

We replaced “*(with saltwater, for example)*” by a citation to Post and von Asmuth (2013) on page 17 line 15.

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.

We expanded this sentence into a new paragraph starting on page 17 line 20. This also accommodates the short comment by Fang Bian (see below).

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 manual dips from our manuscript.

The sentence about manual dips was removed from page 21 line 20 and the caption of Figure 6 and replaced by “Manual dips are indicated by blue dots”.

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.

We inserted “*Such deviations are unfortunately not unusual for commercial PTs (Post and von Asmuth, 2013).*” On page 27 line 15.

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.

Inserted “*Observation wells may also take appreciable time to readjust after the water level inside was raised by inserting measurement instruments.*” On page 30 line 5.

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.

We have revised the figure caption to “*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).*”. We note that the requested example is already stated in the text (page 33 line 20). To clarify this, we added “*Please note the example error calculation in the main text.*” to the caption.

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.

We have revised this paragraph as follows: *“However, the quantification and reporting of measurement error does not seem to be commonplace yet. Moreover, existing standards like Spane and Mercer (1985) or Freeman et al. (2004) contain useful guidelines for a maximum error as follows: (1) +/-3 mm (0.01 ft) for general applications, (2) 0.1 % of expected water level changes, (3) 0.01 % for cases where the depth to water exceeds ~30 m (100 ft). Such standards must see wider uptake, and the development of more sophisticated or site-specific standards, suited for a particular study area or research objective, would be important steps towards better hydrogeological data quality and consistency.”* On page 34 line 20.

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.

We inserted *“Quartz oscillator PTs are much more accurate than the commonly-used strain gauge type PT. However, they have hardly been used in groundwater studies to date, probably because of their higher cost.”* On page 36 line 25.

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.

We inserted *“(…) , as long as there is no problem with keeping the venting tube dry.”* On page 36 line 20.

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.

We deleted the extra ‘s’.

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.

We removed parentheses throughout the manuscript.

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.”

We changed “of” to “off”:

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

Changed “an” to “a”.

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.

Figure 4 already has titles including units. We do not understand why the reviewer did not see this. Further, the second y axis is already labelled with “Precision” and it conveys the standard deviation of the statistics shown in the first y axis. We did not see a need to add more information to this figure.

Page 17 L10: Consider changing unimpeded to unattended.

We changed „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”.

We changed the y axis label to “*Depth below water surface*” and omitted “in well” because it did not fit properly.

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

Thank you! We made no further changes in response to this.

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.

We agree and inserted “ (...) *that is always set to the same time zone and does not update to daylight savings time.*” Page 27 line 4.

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.

As indicated, the units for pressure head are metres and the head gradients are dimensionless. We assumed that stating [-] would be understandable, but have changed this to [m/m] to avoid further confusion.

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.

Note that our error comparison and propagation example in Figure 12 shows that non-verticality is potentially the largest error even for a borehole that is only 10 m deep and has been logged appropriately (Figure 12a see “Point of head”). Because we consider 10 m deep wells to be classified as shallow, we wish to refrain from adding a threshold to our manuscript.

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.

Thank you, we are happy you share this view, too many people put too much confidence in modern electronics. We have not made any changes in response to this comment.

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.

We have added these citations in accordance with our revisions made in response to the reviewers' comments.

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.

We inserted “*The low precision achieved by the non-vented PT is specific for the particular instrument used and not representative for all PTs of this type.*” in the caption of Figure 5.

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.

Inserted “One strategy to verify PT performance in that case is to install three instruments at the same depth.” On page 18 line 34.

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

We have added a short explanation of the usefulness of wet/wet pressure transducers starting on page 19 line 4: “Wet/wet pressure transducers measure the pressure difference between two points that are both exposed to water (Cuthbert et al., 2011). Such devices are ideal for obtaining small head gradients, such as is required for measuring groundwater-surface water interactions, because they eliminate the uncertainties arising from barometric correction or the spatial positioning of two individual measurement points.”

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.

In response to the point about the interplay between systematic errors, we have rephrase the sentence on page 3 line 7 to “*Unrecognised and unaccounted for systematic errors can accumulate or cancel, leading to unquantifiable inaccuracies.*”

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 10^{-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.

Note that head measurements are made and interpreted regardless of system heterogeneity, transience or anthropogenic impact. Errors in head measurements will be present and affect groundwater flow estimates under all conditions. Therefore, minimising the error by adhering to our recommendations would be of interest in all situations. In contrast, one could argue that minimising head measurement errors could enable improved analysis of heterogeneity. To further clarify the intent of our manuscript, we:

- Changed the criticised sentence in the abstract to “*There is sufficient information in the literature to suggest that head measurement errors can impede the reliable detection of flow directions and significantly increase the uncertainty of groundwater flow rate calculations.*”
- We inserted “*We acknowledge that quantifying groundwater flow requires knowledge of the distribution of hydraulic conductivity in addition to hydraulic gradients. While this can be highly heterogeneous and could further complicate investigations, we focus on minimising hydraulic head and gradient measurement errors because doing so increases the accuracy of flow estimates or hydraulic property inversions.*” Starting on page 3 line 21.
- We further inserted “*In practice, heterogeneity of the hydraulic conductivity will further add to the uncertainty of groundwater flow estimates.*” Starting on page 33 line 26.

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.

We note that our revisions from the previous comment also covers the current recommendation.

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.

We have not made any revisions in response to this comment.

After equation 6, the authors state that $\Delta(H)$ is continuous. In the presence of any type of heterogeneity, this is not necessarily the case (think for example, about the instantaneous change in $\Delta(H)$ 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 (with the exception perhaps of 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.

We believe that our revisions to an earlier comment made by this reviewer sufficiently cover revisions required for this comment. See the new statements in the introduction (page 3 line 21) and conclusion (page 33 line 26) which further clarify this and delineate the focus of our 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.

We note also that we specifically mention “(..), it is rare for field studies to determine ∇h in three dimensions.” On page 6 in line 15.

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.

We have not made any revisions in response to this comment.

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.

After careful consideration we have refrained from making substantial changes to the manuscript because adding a “range of likely precisions” as requested by the reviewer proved impossible due to the many factors that play a role. We like to point out that we already provided an estimate of the best achievable performance (deviations between sensors on the order of mm over a 15 month measurement period) in the last paragraph on page 17 of the original manuscript. In response to this comment we inserted “(e.g., within a few centimetres)” on page 34 line 15 to indicate the best achievable accuracy based on our own experience.

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.

We have changed the extents of the y axes in Figure 10 to 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.

We rephrased the sentences starting on page 29 line 8 to *“To avoid such errors, a suitable measurement interval must be chosen upon initial logger deployment, which depends on the hydrogeological conditions at the measurement location. Only when it becomes clear that there is no temporal variability at this timescale can the sampling interval be increased to avoid unnecessary data handling and storage requirements.”*

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.

Short Comments 1:

Hello: First of all, thanks for the precious recommendations for minimising the systematic errors. It's quite practical but not many people ever considered. I've three aspects to ask:

1) Change on the surface elevation: It's generally not considered for the most case, but there's a special occasion, that we recorded a surface subsidence from a 8 m GMI due to the severe drought in Thirlmere Lakes, NSW, Australia. It's been noticed because it's so visible, that without any changes on the GMI, the water level suspiciously increased for 3 days around 6 pm for 3 days. Then no more changes were monitored. Every change lasts around 1 hour with more than 100 mm increase. I think this phenomenon might associate with temperature effects or simply aquifer changes. Not sure if there's good method to mitigate?

We agree and will include the possibility of a change in reference point into our revised manuscript (together with the issue of freezing conditions raised by reviewer 1).

We expanded the discussion of the contributing factors to land surface movement. It starts on page 17 line 20 with “*The causes for such movements (...)*”

2) Change on the logger position from the wire connected: It’s been mentioned, but not fully discussed in the main content. It mainly occurs with a vented logger and the venting cable, which has larger diameters than regular wires. Especially when the piezometer is narrow, the changes from that might be a big problem for this bore during a relatively large water recharge/pumping. That’s why I personally don’t recommend vented logger.

We agree and will briefly discuss the possibility that a transducer can change its vertical position in our revised manuscript.

Inserted “*The cables of vented PTs may be large relative to the well diameter, and sometimes there is little room for the desiccation unit at the top, which may mean that the logger is not always returned to exactly the same position after the GMI was accessed for maintenance or other measurements (e.g. water sampling).*” Starts on page 26 line 12.

3) As some errors might offset other errors, and we already have a brief idea for some general ranges of errors, is that possible to have an universal accepted errors for the whole records?

We agree and point out that this is an extremely complicated issue. Once multiple errors are superimposed it is likely impossible to disentangle the individual effects. Therefore, it is extremely important to understand which systematic errors can occur, so that they can be identified and eliminated as much as possible. This avoids the superposition problem, at least partially. We will mention this in our revised manuscript.

We inserted “*It is difficult to establish if systematic errors are accumulating or cancelling, and hence they must be avoided, or identified and corrected.*” On page 34 line 7.

Regards, Fang

Short Comments 2:

Comments by: Jonathan Kennel, Jessica Meyer, Christopher Neville, Beth Parker

This paper provides a good review of many of the factors that can influence hydraulic head measurements and it is nice to see increased focus on such a fundamental measurement. There are three areas in the paper where we feel further clarification is warranted: time-lag of the monitoring well; calculation of hydraulic gradient; and the comparison of vented and non-vented transducers. Our comments address relatively subtle points and not intended to detract from the overall emphasis of the paper.

Many thanks for your positive assessment, and for taking the time to provide the valuable comments below.

1 Time-lag

It should also be emphasized that while the equivalent freshwater water level (corrections for density) or equivalent piezometric level (for shut in measurements) are often used to infer flow directions and calculate hydraulic gradients, the water level elevation in an open well is not the fundamental parameter of interest. The hydraulic head in the formation is what drives flow. A key distinction is that often a monitoring well or piezometer requires groundwater flow between the formation and monitoring well to record a change in water level surface. This flow is not instantaneous and is commonly referred to as time-lag (Hvorslev, 1951). This time-lag is different than the timelag associated with the propagation of barometric pressure through the vadose zone mentioned in the current paper. The length of the time-lag is dependent on well-bore storage and formation properties; the water level measured from an open monitoring device (or even a grouted transducer to a much smaller extent) will incorporate the effects of antecedent changes in formation pressure. Monitoring wells do not all have the same time-lag associated with them, which also adds temporal uncertainty when comparing measurements. As monitoring frequency increases the uncertainty associated with variable time-lags will become more important and apparent.

We agree and will discuss the possibility of time lags between formation pressure and water levels into our revised manuscript. This point was also raised by reviewer 1.

We inserted a new paragraph to address this issue. It starts on page 30 line 1 with “*Open GMI may suffer from a lag in the water-level response (...)*”

2 Hydraulic gradients

The locations of the monitoring wells play a very important role in the calculation of the hydraulic gradient, not only the knowledge of their true position (x and y, z, and time coordinates) but how the wells are oriented in relation to each other. For example, it is worth mentioning that for calculating a planar hydraulic gradient from three wells in the same hydrogeologic unit the optimal arrangement is in the form of an equilateral triangle. As the locations deviate further from an equilateral triangle the uncertainty of water level measurements plays an increasingly important role in gradient calculations. Gradients vary in space and time, so with increasing monitoring distances the spatial confidence of the calculated gradient actually declines (i.e., uncertainty around the representativeness of the gradient). We are limited by our devices but we strive to get gradients across appropriate scales. It should be mentioned in the discussion on vertical gradients that avoiding blending of distinct hydrogeologic/hydrostratigraphic units is critical to accurately calculating meaningful vertical gradients.

We agree and will include a brief discussion of the influence of screen position of the calculation of gradients into our revised manuscript.

We inserted “(...), which are best arranged in the form of an equilateral triangle, (...)” on page 6 line 27 and “For accurate vertical gradients it is important to use short screens that are within a single hydrogeological unit.” On page 7 line 2.

3 Vented and non-vented transducers

One of the main arguments against non-vented transducers is that you need to convert these values to an open hole water level measurement. While this may be the case when you want to compare the value to a manual measurement in a conventional piezometer, in many cases this is not necessary. For example, if you are trying to calculate a horizontal gradient between three wells in the same aquifer, and all of the transducers deployed are non-vented, there is no need to first remove the barometric component from the results if the elevations of the sensor of each transducer are known. Converting to an equivalent water level may just add uncertainty that is not necessary. Another example is for calculating barometric/loading efficiency. The main issue is related to time shift between equipment, which affects both methods similarly. Given the same transducer specifications, a similar uncertainty will be associated with the result. Figure 5 is particularly damning for non-vented transducers and we think that it needs the raw data and transducer specifications to be provided as well, or the data should not be included at all. This figure runs counter to our experience with hundreds of transducers, both vented and non-vented. It may be that you are comparing differences in full-scale, transducer type, transducer location, or perhaps the barometric compensation procedure could be improved to account for temporal offsets. Likely it is a full-scale or transducer type issue given the smoothness of the barometric pressure data. Care needs to be taken with a figure like this to be as transparent as possible in what is being compared to not overemphasize a preference or mis-interpretation of the cause/effect. Both vented and non-vented transducers exposed to large temperature changes will have increased uncertainty about their measurements; non-vented should not be singled out in section 6.4.1 line 21. This is more a question of deployment location (protected vs. unprotected) and less of a vented, non-vented issue. If possible, both transducer types should be deployed in protected environments that minimize the effect of the external environment while still capturing the measurements of interest (i.e., adherence to data quality objectives). While we agree that the smaller full scale (and thus typically better accuracy and resolution) of vented transducers is a key preference and leads to some simplified calculations, the vent tube and increased cabling, particularly for even moderately deep applications, is a major downside that should be considered when selecting the optimum transducer type. We would suggest not having such a conclusive recommendation of one transducer type over the other.

We agree that we should be careful to make recommendations about transducer types, and indeed there are other factors at play that determine a transducer’s performance. What we meant to illustrate in Figure 5b was that not just any logger should be used to measure water level changes. This may seem like an obvious point, but we have noted all too often that the choice for an instrument is made based on practical and logistical considerations (e.g., availability, affordability, etc.) rather than scientific objectives. We will strengthen this message, and will make sure that we do not provide a recommendation of vented over non-vented based on false comparisons.

In addition to the changes in response to a number of suggestions by reviewer #1 that were similar to this one, we included the following sentence on page 21 line 8 *“It should be noted that when PTs are used to calculate gradients, the readings from non-vented PTs may be used directly without compensating for atmospheric pressure changes as long as the PTs all experience the same atmospheric pressure change.”*

4 Other comments and specific notes

1) We feel that there is little reason to record with such an infrequent or low monitoring frequency as 1 hour given current technology. The optimal monitoring frequency is dependent on the device hardware and the tools available for the analysis of the data. With improved tools, the hope is that monitoring at higher frequency becomes more common so that more complete water level histories are obtained. While you say that it is the maximum monitoring frequency, should this paper suggest a higher frequency of monitoring to push the profession forward?

We believe that monitoring should start at the highest feasible frequency until the dynamics of a system are revealed. After that there is the potential to reduce sampling intervals. There is a trade-off between generating too much data in deep (static) and missing the water level history in a shallow (dynamic) groundwater system. We will clarify this in our revised manuscript.

We rephrased the sentence on page 29 line 9 to *“To avoid such errors, a suitable measurement interval must be chosen upon initial logger deployment, which depends on the hydrogeological conditions at the measurement location. Only when it becomes clear that there is no temporal variability at this timescale can the sampling interval be increased to avoid unnecessary data storage requirements.”*

2) The HEADCO manual by Spane 1985 also provides a thorough review of many of the common issues related to hydraulic head measurements and should be cited as an excellent reference for the readers of this manuscript.

We agree and will cite this reference in appropriate places in our revised manuscript.

We inserted a reference to Spane (1985) on page 34 in line 18.

3) For the grouted-in application, would you still recommend vented transducers?

That question cannot be answered simply with yes or no because it depends on the case. We do not see a need to revise our manuscript in response to this question.

4) Page 2: “This is by no means trivial, and certainly much more complex than collecting manual measurements”. Understanding what a manual measurement represents is also quite complex, in part because we are often missing the appropriate additional information necessary for their interpretation. Manual measurements, tend to have increased temporal uncertainty and also lack an associated barometric pressure value taken at the same time.

The purpose of manual measurements is to determine the water level in the monitoring bore and to transform automated time series into pressure heads. This can subsequently be used to conduct barometric corrections using an appropriate pressure record. We believe that this is sufficiently covered in our manuscript.

5) Page 4: “The hydraulic head is defined as (e.g., Freeze and Cherry, 1979)”. Consider citing the original work here (Hubbert, 1940) rather than an introductory level textbook.

We agree and will cite the provided reference in our revised manuscript. Note that we already cited this textbook for the reader’s convenience (especially since it has become freely available online).

We added the citation Hubbert (1940) on page 4 line 4.

6) Page 6-7: “Air pressure changes act differently on the water column in open GMI than on the groundwater, because in the open GMI the air pressure change is transmitted instantaneously to the water, whereas the groundwater pressure response is more complex and can be delayed.” We disagree with this statement. The air pressure changes result in formation pressure changes that are reflected in the open hole water level measurement, just perhaps at a later time period. We would argue that the water level response in an open hole would be more complex than the actual formation head. This is because open hole response contains both the formation response in addition to the responses resulting from the direct atmospheric connection and well-bore storage.

Our statement is in principal correct when you consider a semi-confined systems which is poroelastic. We disagree with your statement “The air pressure changes result in formation pressure changes that are reflected in the open hole water level measurement, just perhaps at a later time period”. This is because the formation pressure changes will depend on the barometric efficiency. We do agree with your last statement and will clarify this in our revised manuscript.

We changed our formulation to “Air pressure changes are transmitted instantaneously to the water column in open GMI. In contrast, the formation response is more complex because air pressure changes must propagation through the subsurface to the point of measurement which can result in a delay.” On page 7 line 5.

7) Page 12: “Vertical head gradients in an aquifer tend to be small under natural (i.e., not pumped) conditions, often less than 10^{-3} ”. Followed by “this can be taken as an indication of the maximum head error for a typical piezometer caused by uncertainty about the elevation of the point of measurement”. Given that this reasoning is used to quantify one of the forms of uncertainty based on standard practice, some basis for the gradient of 10^{-3} should be provided. Also, it seems very limiting to constrain this discussion to ‘aquifer’ units in non-pumped systems. Larger vertical gradients should be expected across units with lower bulk vertical hydraulic conductivity and in recharge areas of a flow system or where units are being pumped (which is often the case). There are examples in the literature that show vertical gradients larger than 10^{-3} (see references provided in Meyer et al.

2014). Also, blending of distinct hydrostratigraphic/hydrogeologic units in a single well screen seems like an important but neglected aspect of this discussion.

The basis for this value is the upper end of the range of recharge rates (1 mm/d) for a $K = 1$ m/d. We will clarify this in our revised manuscript.

We inserted “(...) (this value would be typical for an aquifer with a rainfall recharge rate of 1 mm d^{-1} and a vertical hydraulic conductivity of 1 m d^{-1})” page 14 line 8.

8) Page 14: “In layered aquifer systems, the water level in wells with long screens was found to depend on the transmissivities of the layers intersected by the wells (Sokol, 1963)” This is the key point! This can have a dramatic influence on the head recorded even for short screens if cross-connecting the system. The measured head value becomes biased toward the highest transmissivity intersected and much of the earlier text on the monitoring point tends to confuse this issue. For a monitoring well the head is representative of the open interval and assigning the location to a point is inappropriate.

We agree and have tried to make this point clear. We will further clarify this in our revised manuscript.

We inserted “For accurate vertical gradients it is important to use short screens that are within a single hydrogeological unit.” Page 7 line 2.

9) Page 21 Figure 6 caption: “Note that the manual dips confirm that there was no diurnal variability in the water levels (blue dots).” With the sparsity of measurements we don’t think this statement is justified.

We agree that this is not clear, and will remove this statement. This was also raised by reviewer 1.

The sentence about manual dips was removed from page 21 line 23 and the caption of Figure 6.

10) Page 26 Line 3: “Clock stability is an important consideration when using multiple instruments. Examples include the barometric correction of absolute pressure measurements from a non-vented transducer, or the calculation of hydraulic gradients using two different time series.” Barometric correction requires the calculation of the barometric efficiency. You need barometric pressure measurements to do this calculation and therefore both non-vented and vented transducers will be affected.

We agree and will clarify this point in our revised manuscript.

...

11) Page 33 Line 7: Probably should site some earlier works here - for example Jacob 1940, Rojstaczer et al. 1988.

We agree and will cite these works in our revised manuscript.

We inserted a citation to Jacob (1940) page 7 line 28.

12) Page 35: “Because vented PTs measure a relative pressure instead of an absolute pressure, they have a smaller range and do not require a separate instrument to simultaneously record the atmospheric pressure.” We should encourage barometric pressure to be monitored at every site. In addition, recording barometric pressure at a higher frequency can provide certain advantages related to barometric response function calculation as a more complete barometric history is obtained.

We agree that barometric data is very important and will clarify this statement in our revised manuscript.

We inserted “Nevertheless, barometric pressure must still be acquired in order to perform a barometric correction.” Page 36 line 23.

13) Page 35: “For reliably resolving head gradients and flow direction at small vertical distances, for example when assessing surface water-groundwater interactions, we recommend the use of wet/wet differential pressure sensors (e.g., Cuthbert et al., 2011).” We don’t think “wet/wet differential pressure sensors” were discussed in the body of the manuscript. If not, this comment is out of place in the conclusions section. Consider adding a brief discussion to the main body of the paper or removing from the conclusions.

We agree and will include some more discussion of wet/wet pressure transducers into our revised manuscript. This was also raised by reviewer 1.

We have inserted “Wet/wet pressure transducers measure the pressure difference between two points that are both exposed to water {citep{Cuthbert2011}}. Such devices are ideal for obtaining small head gradients, such as is required for measuring groundwater-surface water interactions, because they eliminate the uncertainties arising from barometric correction or the spatial positioning of two individual measurement points.” Page 19 line 5.

Thank you for this important paper,

Regards, Jonathan Kennel, Jessica Meyer, Christopher Neville, Beth Parker

Affiliations: University of Guelph, University of Iowa, S.S. Papadopoulos and Associates, University of Guelph References

References

Hvorslev, M.J., 1951. Time lag and soil permeability in ground-water observations.

Jacob, C.E., 1940. On the flow of water in an elastic artesian aquifer. *Eos, Transactions American Geophysical Union*, 21(2), pp.574-586.

Meyer, J.R., Parker, B.L., and Cherry, J.A. 2014. Characteristics of high resolution hydraulic head profiles and vertical gradients in fractured sedimentary rocks. *Journal of Hydrology*, 517, 493-507.

Rojstaczer, S., 1988. Determination of fluid flow properties from the response of water levels in wells to atmospheric loading. *Water Resources Research*, 24(11), pp.1927- 1938.

Spane Jr, F.A. and Mercer, R.B., 1985. HEADCO: A program for converting observed water levels and pressure measurements to formation pressure and standard hydraulic head (No. RHO-BW-ST-71-P). Rockwell International Corp.

We will include the suggested references at appropriate places in our revised manuscript.

Error in hydraulic head and gradient time-series measurements: a quantitative appraisal

Gabriel C. Rau^{1,2}, Vincent E. A. Post³, Margaret A. Shanafield⁴, Torsten Krekeler³, Eddie W. Banks⁴, and Philipp Blum¹

¹Karlsruhe Institute of Technology (KIT), Institute of Applied Geosciences (AGW), Karlsruhe, Germany

²Connected Waters Initiative Research Centre (CWI), The University of New South Wales (UNSW), Sydney, Australia

³Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany

⁴National Centre for Groundwater Research and Training (NCGRT) and College of Science and Engineering, Flinders University, Adelaide, Australia

Correspondence: Gabriel C. Rau (gabriel.rau@kit.edu)

Abstract. Hydraulic head and gradient measurements underpin practically all investigations in ~~hydro(geo)logy~~hydrogeology.

There is sufficient information in the literature to suggest that head measurement errors ~~may be so large that flow directions can not be inferred reliably, and that their magnitude can have as great an effect on~~ can impede the reliable detection of flow directions and significantly increase the uncertainty of ~~flow rates as the hydraulic conductivity~~groundwater flow rate

5 calculations. Yet, educational text books contain limited content regarding measurement techniques and studies rarely report on measurement errors. The objective of our study is to review currently-accepted standard operating procedures in hydrological research and to determine the smallest head gradients that can be resolved. To this aim, we first systematically investigate the systematic and random ~~measurements~~ measurement errors involved in collecting time series information on hydraulic head at a given location: (1) geospatial position, (2) point of head, (3) depth to water, and (4) water level time series. Then, by

10 propagating the random errors, we find that with current standard practice, horizontal head gradients $< 10^{-4}$ are resolvable at distances $\gtrsim 170$ m. Further, it takes extraordinary effort to measure hydraulic head gradients $< 10^{-3}$ over distances < 10 m. In reality, accuracy will be worse than our theoretical estimates because of the many possible systematic errors. Regional flow on a scale of kilometres or more can be inferred with current best-practice methods, but processes such as vertical flow within an aquifer cannot be determined until more accurate and precise measurement methods are developed. Finally, we offer a concise

15 set of recommendations for water level, hydraulic head and gradient time series measurements. We anticipate that our work contributes to progressing the quality of head time series data in the ~~hydro(geo)logical~~ hydrogeological sciences, and provides a starting point for the development of universal measurement protocols for water level data collection.

Copyright statement.

1 Introduction

Water level and hydraulic head time series are critical for understanding water flow-related processes and properties in both surface and subsurface aquatic environments. At the surface, water levels are important to understand relationships between water level and flow, and for estimating surface water-groundwater interactions (e.g., Kalbus et al., 2006; McCallum et al., 5 2014). In the subsurface, measurements of hydraulic head are used to determine groundwater flow, estimate aquifer properties, and investigate aquifer processes such as the response to pumping or groundwater recharge (e.g., Freeze and Cherry, 1979; Domenico and Schwartz, 1997). While it has been confirmed in several studies that the accuracy of water level measurements is a limiting factor for drawing conclusions about ~~hydro(geo)logieal~~ hydrogeological processes (e.g., Saines, 1981; Silliman and Mantz, 2000; Devlin and McElwee, 2007), measurement errors are not always properly recognised (Post and von Asmuth, 10 2013).

Pressure transducers (PTs) have been used since the 1960s to measure water level (Liu and Higgins, 2015), and the development and availability of a wide variety of commercial instruments has made collection of high temporal resolution water level time series common practice. This has been a major advancement of our capability to study ~~hydro(geo)logieal~~ hydrogeological processes, but the proper use of automated sensors means that researchers need to have a good understanding of instrument 15 technology and operating procedures. This is by no means trivial, and certainly much more complex than collecting manual measurements. Knowledge is already required during the procurement phase, as there are many brands and logger types available, and the specific research objectives of a project determine which sensors are suitable and which are not (Dunncliff and Green, 1993). The same is true for modern positioning and levelling instruments, needed to establish the horizontal and vertical position of the monitoring well (e.g., Brinker, 1995; Hegarty, 2017). The storage and quality assurance of the large volume 20 of time-series data is not straightforward either, and can require programming skills to process data in an efficient manner. All things considered, modern hydrologists and hydrogeologists require a broad skill set, which is typically too extensive to be comprehensively covered in standard textbooks and water-related educational programs.

Yet, water level measurement lies at the heart of all ~~hydro(geo)logieal~~ hydrogeological investigation and knowledge of the measurement error associated with modern instruments is fundamental to the collection of reliable time series data. Studies 25 on the topic published in the literature focus mainly on the instruments themselves. One of the first estimates of PT drift was published by Rosenberry (1990), who showed how these errors would have led to incorrect interpretation of water levels at several sites. More recently, Sorensen and Butcher (2011) examined the accuracy and drift of different brands of PTs and found that the manufacturers' specifications were not met during field deployment. The effect of temperature on sensor performance has also received some attention (Cuevas et al., 2010; McLaughlin and Cohen, 2011; Liu and Higgins, 2015). These studies 30 concluded that strong temperature fluctuations such as those that occur under field conditions affect PTs of all types.

More comprehensive treatments of the subject tend to be published as reports by national research organisations. Prime examples include Freeman et al. (2004) and Cunningham and Schalk (2016) who not only discussed sensor technology, but also provided technical procedures for collecting water levels and some of the errors involved. Moreover, some relevant works were published in a non-English language (e.g. Bouma et al., 2012; Ritzema et al., 2012; Morgenschweis, 2018) or as conference

proceedings (e.g. Atwood and Lamb, 1987; Simeoni, 2012; Mäkinen and Orvomaa, 2015) so that, despite their usefulness, their findings did not permeate the indexed international literature.

When collecting hydraulic head time series in the field, many different factors apart from instrument drift influence the stability of the measurement setup (Post and von Asmuth, 2013). These include cable stretch, well clogging, sensor fouling, variable-density effects and even changes in the vertical position of the observation well. This requires regular field site maintenance, re-calibration, and record-keeping. However, without knowledge of the magnitude of the water level error caused by such effects, there is no general guidance to develop adequate systematic field procedures. Unrecognised and unaccounted for systematic errors can accumulate (or cancel), leading to unquantifiable inaccuracies, while the random errors increase the uncertainty. Sweet et al. (1990) contended that the propagation of measurement errors can result in $\pm 100\%$ uncertainty in calculated flow velocities, and that the uncertainty of the head gradient may be of a similar magnitude as that of the hydraulic conductivity.

While the large uncertainty of head gradients due to water level measurement error has also been confirmed by others (Silliman and Mantz, 2000; Devlin and McElwee, 2007), there is currently no single resource that ties together the lessons learned during decades of experience ~~together~~. The objective of the present paper is to address this gap by quantifying the smallest possible head gradients that can be resolved using currently-accepted standard operating procedures in hydrological research. Using data collected in a wide range of field settings we provide a comprehensive and quantitative analysis of the systematic and random errors that must be considered when collecting water level time series using automated instruments. The emphasis is on transient effects and errors that can change with time. We further add to the existing literature by highlighting sources of error that are generally overlooked. Furthermore, we propagate the random errors to quantify the best-possible composite uncertainty of horizontal and vertical head gradients considering error magnitudes from good field practice and a wider spatial extent than Silliman and Mantz (2000) and Devlin and McElwee (2007). We acknowledge that quantifying groundwater flow requires knowledge of the distribution of hydraulic conductivity in addition to hydraulic gradients. While this can be highly heterogeneous and could further complicate investigations, we focus on minimising hydraulic head and gradient measurement errors because doing so increases the accuracy of flow estimates or hydraulic property inversions.

We anticipate that our analysis is helpful to field practitioners at all levels, and can be used as an educational resource. By providing a concise list of best practice recommendations at the end of the manuscript, we intend to provide a starting point for the development of comprehensive and universal international standard procedures, which are currently lacking.

2 Review of measurements and error terminology

2.1 From measurements to heads

In this work we use the term *groundwater monitoring infrastructure* (GMI) as an umbrella term for open and cased boreholes, wells, standpipe or grouted-in piezometers (Section 4). The most typical GMI in hydrogeology consists of boreholes equipped with a standpipe piezometer, where the standing water level is in contact with the atmosphere (open GMI, see location 1 in Figure 1) and therefore readily accessible for measurement. Fully grouted-in piezometers contain a single or multi-array string

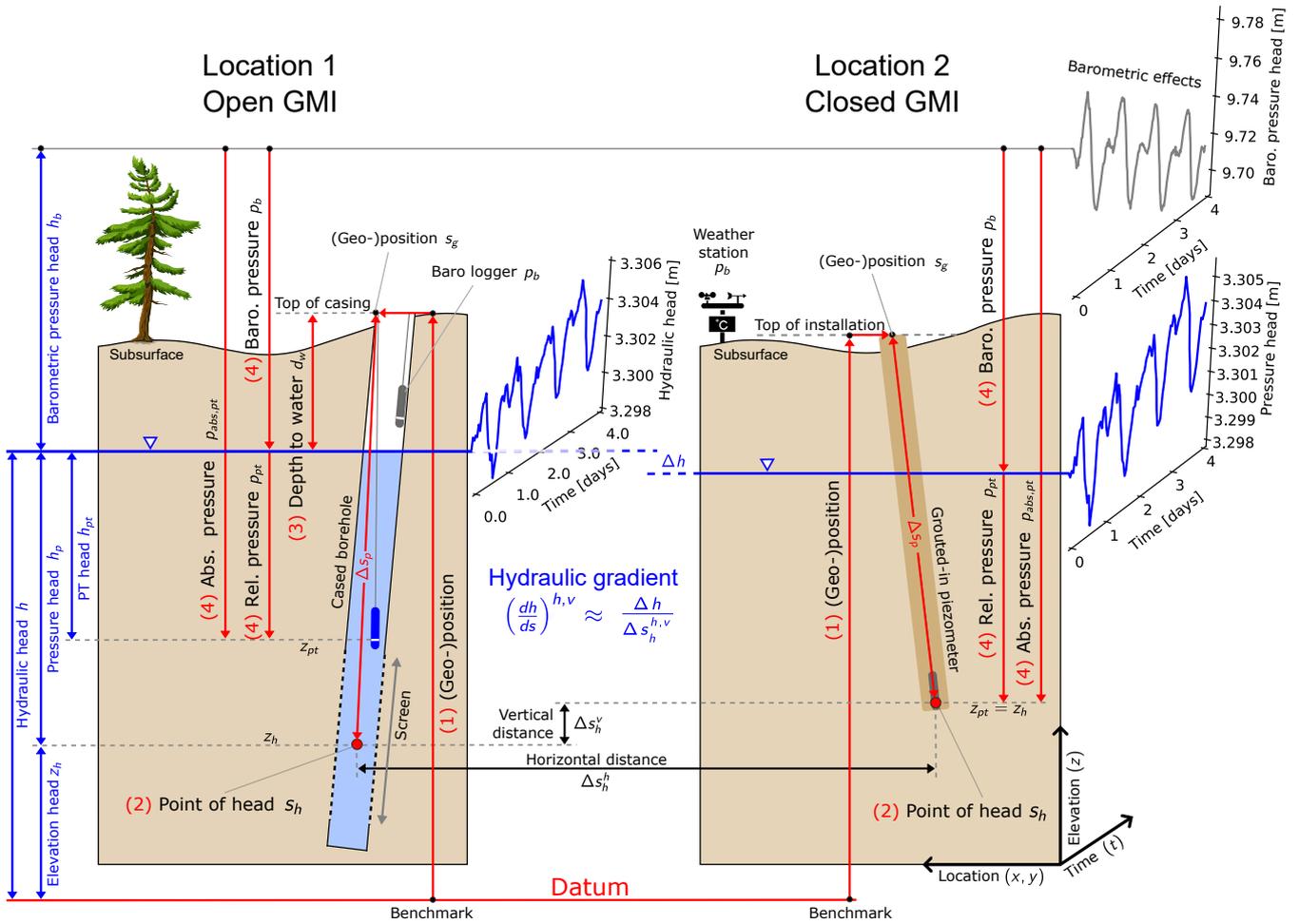


Figure 1. Overview of the four individual measurements (enumerated as steps and marked in red) required to calculate time series of hydraulic head (one location) and gradient (two locations) using two different types of groundwater monitoring infrastructure (GMI). Location 1 shows a cased borehole that is open to the atmosphere (open GMI), whereas Location 2 illustrates a fully grouted-in piezometer (closed GMI). The boreholes are drawn at an angle to highlight the importance of errors caused by inclination during construction of the borehole.

of PTs and are closed to the atmosphere (closed GMI, see location 2 in Figure 1) and are often used in mining and geotechnical engineering (e.g., McKenna, 1995; Mikkelsen and Green, 2003).

GMI allows access to measuring *depth to water* or *groundwater pressure* from which the *hydraulic head* can be calculated (terminology is illustrated in Figure 1). The hydraulic head is defined as (e.g., Freeze and Cherry, 1979) (e.g., Hubbert, 1940; Freeze and Cherry, 1979)

$$\begin{aligned}
 h(x, y, z, t) &= z_h(x, y) + \frac{p(x, y, z, t) - p_b(x, y, z, t)}{\rho(x, y, z, t)g} \\
 &= z_h(x, y) + h_p(x, y, z, t)
 \end{aligned}
 \tag{1}$$

where (x, y, z) are the Cartesian coordinates [m] of the measurement point, t is time [s], z_h is referred to as elevation head [m], p is the total groundwater pressure [Pa], and p_b is the barometric pressure [Pa], ρ is the groundwater density [kg m^{-3}] across the water column above z_h , g is the gravitational constant [$\approx 9.81 \text{ m s}^{-2}$]. The term h_p [m] is the *pressure head*.

The dependence of the variables in Equation 1 on (x, y, z, t) has been deliberately emphasised to stress the point that their magnitude varies in space and time. Determining hydraulic head time series requires four measurements (hereafter also referred to as steps), which are conceptualised in Figure 1 and can be summarised as follows:

- (1) Geo- or relative positioning of the GMI, i.e. determining its location at the Earth's surface $s_g = (x_g, y_g, z_g)$ (Section 3);
- (2) Establishing the point of (or location representative for) head measurement $s_h = (x_h, y_h, z_h) = s_g + \Delta s_p$ with $\Delta s_p = (\Delta x_p, \Delta y_p, \Delta z_p)$ being the vector that represents the location offset s_h with respect to s_g (Section 4);
- 10 (3) Measurement of the water depth below the top of casing $d_w(t_j)$ at a discrete times t_j (open GMI only, Section 5);
- (4) Automated pressure measurements at PT location $s_{pt} = (x_{pt}, y_{pt}, z_{pt})$ of $p_{pt}(s_{pt}, t_i)$ at discrete times t_i (Section 6).

There are two methods to obtain $h(x_h, y_h, z_h, t) = h(s_h, t)$ based on field measurements:

Method 1 (only for open GMI): When only d_w has been measured in the field (for example, by taking regular manual water level measurements) the hydraulic head simply follows from

$$15 \quad h(s_h, t_j) = z_g - d_w(t_j), \quad (2)$$

where t_j is the distinct time at which the water level measurement was made.

Hydraulic head time series are nowadays commonly determined from the pressure readings of a transducer located at elevation z_{pt} [m]. The head is then calculated using

$$h(s_h, t_i) = z_g - d_w(t_j) + [h_{pt}(s_{pt}, t_i) - h_{pt}(s_{pt}, t_j)], \quad (3)$$

20 where h_{pt} is the transducer pressure head, i.e., the pressure recorded by the PT expressed as a water column height (e.g., Hölting and Coldewey, 2013)

$$h_{pt}(s_{pt}, t_i) = \frac{p_{pt,abs}(s_{pt}, t_i) - p_b(s_b, t_i)}{\bar{\rho}_w(t_i)g} = \frac{p_{pt}(s_{pt}, t_i)}{\bar{\rho}_w(t_i)g}, \quad (4)$$

25 where $p_{pt,abs}$ and p_{pt} are, respectively, the absolute and relative transducer recorded pressures and $\bar{\rho}_w$ is the average density [kg m^{-3}] across the water column above the transducer's elevation z_{pt} . Application of Equation 4 is referred to as *barometric compensation*. The location of the barometric pressure measurement $s_b = (x_b, y_b, z_b)$ must be chosen so that it is representative for the barometric pressure experienced by the PT (Post and von Asmuth, 2013).

Method 2 (for open and closed GMI): For a PT installed at location $s_{pt} = s_h$, the hydraulic head follows from

$$h(s_h, t_i) = z_{pt} + \frac{p_{pt}(s_{pt}, t_i)}{\bar{\rho}_w(t_i)g} = z_{pt} + h_{pt}(s_{pt}, t_i). \quad (5)$$

This is the only way by which heads can be measured in closed GMI for which d_w can not be determined.

For open GMI $\bar{\rho}_w$ can be measured. For closed GMI, however, $\bar{\rho}_w$ is the average density of the groundwater above z_{pt} , which has to be estimated in the absence of direct measurements. Because the PT is at elevation $z_{pt} = z_h$, $p_{pt} = p - p_b$, and Equation 5 is identical to Equation 1 when $\bar{\rho}_w = \rho$. These considerations have important implications when density effects influence the pressure-head relationship of a GMI (Section 6.4.2).

2.2 Hydraulic head gradient

Hydraulic head is a scalar quantity and the gradient of the head field in combination with hydraulic conductivity enables quantification of groundwater flow rates using *Darcy's Law*. In three dimensions the hydraulic head gradient (or simply head gradient) is a vector defined as (e.g., Domenico and Schwartz, 1997)

$$\nabla h = \mathbf{i} \frac{\partial h}{\partial x} + \mathbf{j} \frac{\partial h}{\partial y} + \mathbf{k} \frac{\partial h}{\partial z}, \quad (6)$$

where the bold \mathbf{i} , \mathbf{j} and \mathbf{k} symbols denote the unit vectors in the x , y , and z direction, respectively. Since h and ∇h are continuous field variables, and, in practice, h can only be measured at discrete points s_h , head measurements can only be used to approximate ∇h . Moreover, it is rare for field studies to determine ∇h in three dimensions. Therefore, for the purpose of error propagation (Section 7) we consider the horizontal (in the $x - y$ plane) and vertical components (indicated by either a superscript h or v , respectively) separately by

$$\left(\frac{dh}{ds} \right)^{h,v} \approx \frac{\Delta h}{\Delta s_h^{h,v}}, \quad (7)$$

where the term on the left-hand side represents the rate of head change per unit of distance s , which is approximated by the ratio of Δh , the head difference between two points of measurement, over

$$\Delta s_h^h = \sqrt{\Delta x_h^2 + \Delta y_h^2}, \quad (8)$$

or

$$\Delta s_h^v = \Delta z_h, \quad (9)$$

where Δx_h , Δy_h and Δz_h are the distances between two points of head measurement in the x , y and z direction, respectively.

It must be emphasised that considering the horizontal head difference between two points is only meaningful when they are located along the direction of maximum rate of head change, i.e. perpendicular to the contour planes of equal head (assuming isotropic and constant-density conditions). Hydraulic head measurements from at least three different locations are, which are best arranged in the form of an equilateral triangle, are required to determine the head gradient in two dimensions (e.g., Freeze

and Cherry, 1979), or four locations in three dimensions (Silliman and Mantz, 2000; Devlin and McElwee, 2007). Even more locations are required for head contour maps (e.g., Ohmer et al., 2017). For accurate vertical gradients it is important to use short screens that are within a single hydrogeological unit.

2.3 Barometric effects

5 The following discussion is only applicable to open GMI (i.e., open to the atmosphere; location 1 in Figure 1). Air pressure changes ~~act differently on~~ are transmitted instantaneously to the water column in open GMI ~~than on the groundwater, because in the open GMI the air pressure change is transmitted instantaneously to the water, whereas the groundwater pressure.~~ In contrast, the formation response is more complex ~~and can be delayed~~ because air pressure changes must propagate through the subsurface to the point of measurement which can result in a delay. Barometric pressure can change as part of the local weather (e.g., the passing of high and low pressure systems) by as much as 1.5 m water level equivalent for the extreme weather events. If a barometric pressure change propagates through the unsaturated zone of an unconfined system without delay, the water level in an open GMI is a direct representation of the groundwater pressure. However, since the unsaturated zone can resist air movement, for example under low (air) permeability or variably saturated conditions (e.g., Weeks, 1979), there can be a time lag between barometric pressure changes and the associated GMI water level response (e.g., Rasmussen and Crawford, 1997). This can be quantified using the barometric response function, which can change over time (Rasmussen and Crawford, 1997; Spane, 2002; Butler et al., 2011).

In addition to this, the response to air pressure changes of an open GMI's water level is fundamentally different than the response of the hydraulic head due to the elastic storage behaviour of the subsurface. This can be understood by considering that an increase in barometric pressure raises the total stress acting on both the GMI's water column and the subsurface. The additional stress is borne exclusively by the water column inside the GMI, whereas it is shared between the water and the formation in the surrounding subsurface (e.g., Freeze and Cherry, 1979; Domenico and Schwartz, 1997). As a result, the pressure increase inside the GMI is larger than the groundwater pressure increase, which induces water flow from the GMI into the formation, thus leading to a lowering of the measured water level. The result is an inverse relationship between changes in water level inside open GMI and the changing barometric pressure (e.g., Meinzer, 1939; Gonthier, 2003). This relationship can be exploited to detect aquifer confinement (Acworth et al., 2017), but also necessitates the correction of water levels measured in open GMI to faithfully infer the hydraulic head in the formation.

The barometric efficiency (BE) expresses the ratio between the water level change in a GMI Δh_{pt} and the barometric pressure change Δp_b causing it (Jacob, 1940; Clark, 1967; van der Kamp and Gale, 1983)

$$BE = -\frac{\Delta h_{pt}}{\Delta p_b} \bar{\rho}_w g = \frac{\Delta h}{\Delta p_b} \bar{\rho}_w g = \frac{n\beta}{n\beta + \alpha} \quad (10)$$

30 where n is the total porosity of the formation [-], β is the compressibility of water ($\approx 4.59 \cdot 10^{-10} Pa^{-1}$) and α is the (undrained) compressibility of the formation [Pa^{-1}]. The minus sign is due to the discussed inverse relationship between h_{pt} and p_a .

The BE quantifies the partitioning of the total stress change between the formation and the groundwater (Domenico and Schwartz, 1997; Acworth et al., 2016a). If the subsurface is assumed to be incompressible ($\alpha = 0$ so $BE = 1$, an often-

made assumption), the inverse relationship between water level measured in the GMI and hydraulic head in the subsurface is most pronounced. However, the majority of geological materials are more compressible than water ($\beta > \alpha$), so realistically $0 < BE < 1$ (Rau et al., 2018). Methods to reduce barometric effects on hydraulic head measurements were suggested in the literature and are referred to as *barometric correction* (not to be confused with barometric compensation, Equation 4) (e.g., 5 Hubbell et al., 2004; Toll and Rasmussen, 2007; Noorduyn et al., 2015). This discussion highlights that the *BE* of a formation is an important property, and ignoring it can have significant implications, when hydraulic heads or gradients are derived from water level measurements with the aim to interpret groundwater processes (Spane, 2002).

Avoiding barometric effects require GMI with a specific design. Hubbell et al. (2004) suggested a sealed well and showed that their design reduced barometric pressure effects by an order of magnitude, especially for sites with deep vadose zones. 10 Furthermore, a laboratory study by Noorduyn et al. (2015) demonstrated that measured total pressure recorded in sealed and unsealed wells are equal assuming barometric pressure is also measured, water levels can be accurately measured in either sealed or unsealed standpipes. This is convenient for fluvial environments, where long standpipes are subject to the forces of river flows, which can be quite violent in ephemeral streams especially (e.g., Shanafield and Cook, 2014).

2.4 Clarification of error terminology

15 Despite their importance, the terms related to measurement error are often mixed up or used ambiguously. Thus, before proceeding, it is crucial to clarify their meanings within the context of head measurement.

Accuracy is a measure of how closely the mean of the measured head corresponds to the real head. The deviation between the true value and the mean of its measurements is the *systematic* (or *absolute*) error (Figure 2).

20 **Precision** is the *spread* of the measured heads around their mean value. When the measurements are normally distributed, it can be expressed by the standard deviation of a Gaussian distribution. It is also referred to as *random* error (Figure 2).

Resolution is the smallest numerical separation at which the change of real value can be distinguished.

Range is the difference between the minimum and maximum value an instrument can measure.

Electronic measurements use analogue-to-digital converters (ADC), which convert continuous analogue signals into discrete (digital) values. ADCs generally have limited steps (resolution bins, Figure 3), leading to an inverse relationship between the measurement range and resolution. Consequently, the larger the range of measurement, the coarser the resolution. For example, 25 a 12-bit ADC has $2^{12} = 4,096$ resolution bins, which equates to a theoretical resolution of 2.4 mm, when the range is 10 m, or a resolution of 12.2 mm, when the range is 50 m. As Figure 3 demonstrates, the difference between continuous and instrument-reported (quantised) head, and thus the measurement error, decreases with increasing resolution.

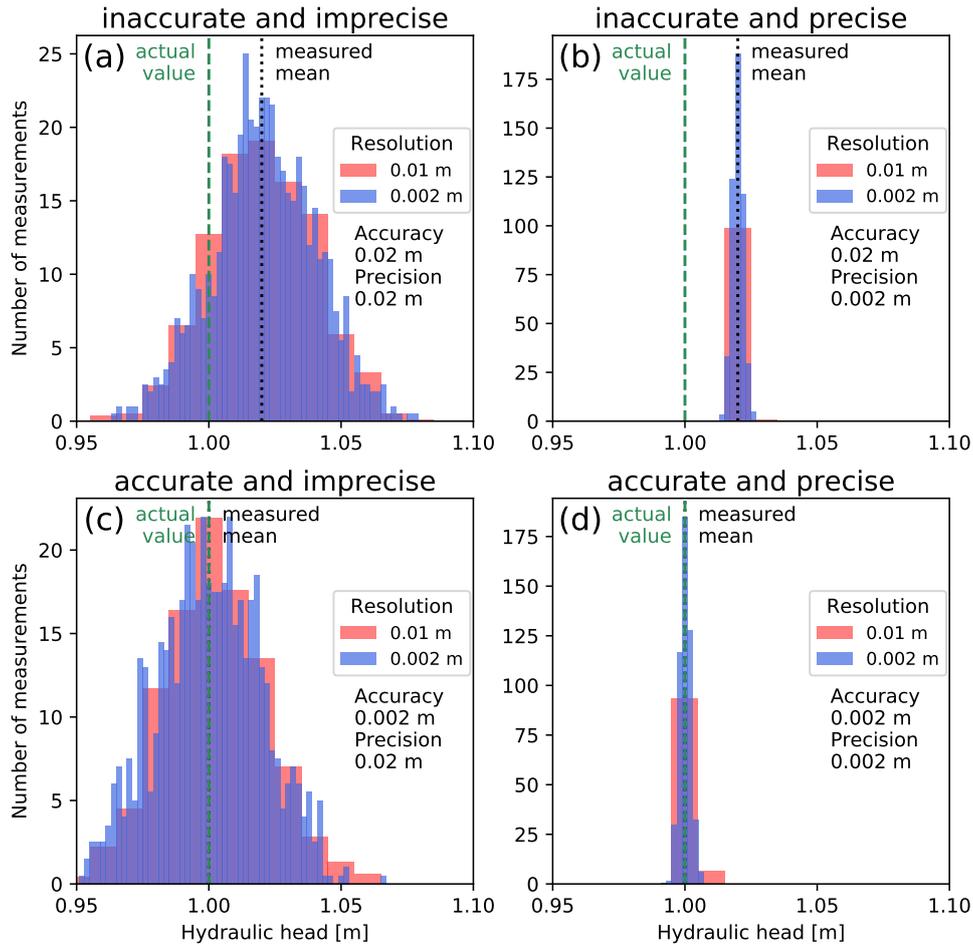


Figure 2. Possible combinations of accuracy, precision and resolution illustrated in a matrix, when 1,000 measurements of the same head ($h = 1$ m) are made. Measurements are (a) inaccurate and imprecise, (b) inaccurate and precise, (c) accurate and imprecise, (d) accurate and precise. Examples are illustrated with two different values of accuracy, precision and resolutions (equal to bin width in histogram).

3 Geo-spatial positioning of groundwater monitoring infrastructures

There are two ways to determine a GMI's position (s_g , Figure 1). The first is surveying, which is the determining of the three-dimensional distance between points of interest. The second is to use navigation satellites. This section briefly summarises both. More details on surveying can be found in Brinker (1995), and on satellite system technology and applications in Hegarty (2017), Bock and Melgar (2016) and Misra and Enge (2010).

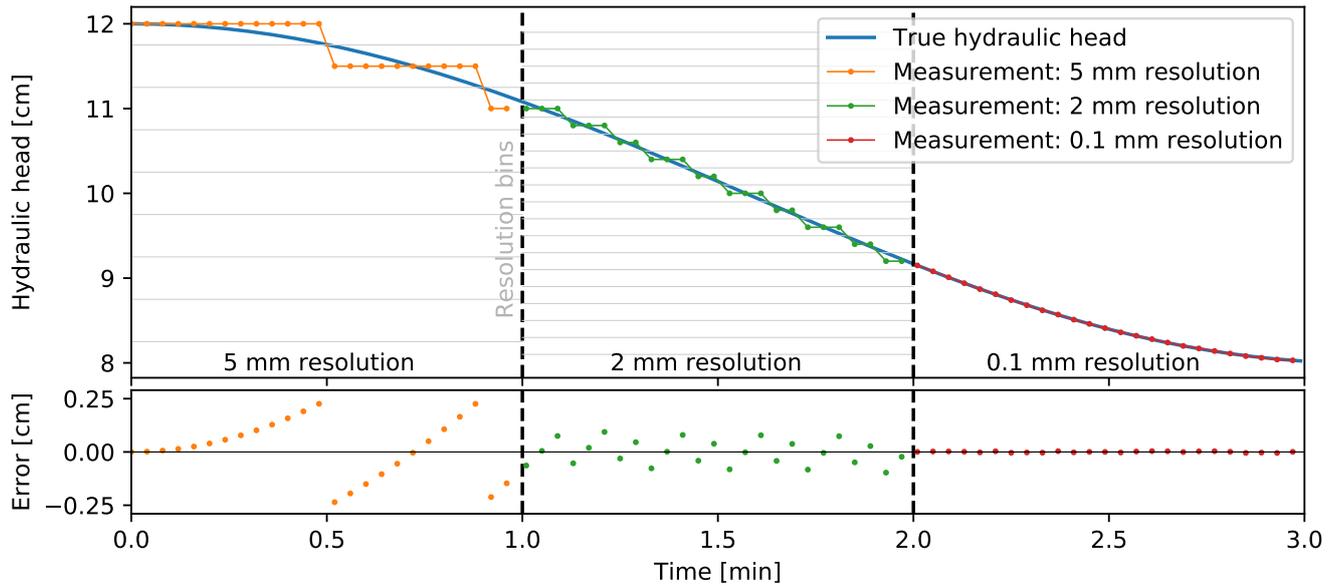


Figure 3. Illustration of the influence that the instrument resolution has on the measurement error: A continuous, time variable head is measured at discrete time intervals by instruments with analogue to digital conversion resolution of 5 mm, 2 mm and 0.1 mm.

3.1 Relative positioning using traditional surveying

Determining the horizontal and vertical distances to a reference point (known as trigonometric levelling) can be done using a total station theodolite. They are equipped with a precision telescope that can rotate in the horizontal and vertical direction, allowing visual adjustment of the telescope to points of interest. Precise optical sensors can pinpoint a bar-code on the staff and digitise the angle and azimuth readings from which the horizontal and vertical distances are calculated using a built-in computer. They further include an electronic distance measurement (EDM) device, based on the travel time of laser pulses reflecting off a target, and have satellite receivers to determine geo-coordinates (Section 3).

Levelling, the technique of measuring vertical distances (heights) relative to a known survey benchmark, can be conducted using optical or light-based instruments operating from a tripod. The latest generations of optical levelling instruments use a rotating precision telescope to magnify the scale printed onto a levelling rod (staff) that is held vertically on top of a point of interest. The telescope is used to read the vertical distance above the point of interest of a laser beam rotating in a horizontal plane. The levelling rod is equipped with a receiver that can be moved vertically until it detects the beam.

The maximum measurement distance of digital levels or total stations is limited to hundreds of meters, depending on the telescope, the range of the laser beam and the visibility of the target (El-Ashmawy, 2014). Longer distances are surveyed by leap-frogging survey devices along multiple points (traversing) (Brinker, 1995). Measurement error is a function of distance and accuracy and precision of leap-frog surveys tend to be poorer than surveys where the instrument does not require moving.

[An indication of the measurement error can be obtained by returning to the starting location of the survey and determining](#)

the difference between the recorded positions at the start and end. When GMI locations are to be referenced with respect to a national datum, the accuracy is further dependent on the quality of the known benchmarks that provide the link between the local survey to the national datum (Figure 1]).

It is difficult to determine the accuracy of high-precision surveying, because this must be compared to a more accurate benchmark method. The measurement error for state-of-the-art survey devices depends on many factors, including instrument setup, calibration, sun position, temperature elevation gradient, battery level, and most importantly, operator's expertise (Beshr and Abo Elnaga, 2011; Bitelli et al., 2018). The literature contains very few peer-reviewed investigations that test manufacturers' specifications. However, one assessment has illustrated that digital levelling can reach an accuracy of 2 mm/km with precision of 1 mm + 1 mm/km, respectively (Bitelli et al., 2018). Leap-frogging using 150 m distance steps found an elevation precision of 1.9 mm / \sqrt{km} (Ceylan and Baykal, 2006).

Estimating the positioning errors of total stations is even more complicated due to the combination of EDM and angle sensors (Walker and Awange, 2018). Braun et al. (2015) thoroughly investigated the accuracy and precision of industry standard EDM devices over a well-calibrated distance of 40 m. They found that the accuracy varied from 0-4 mm, with some devices showing dependence on the measurement distance. We use the precision of 0.5 mm stated by Braun et al. (2015) for our error analysis (Table 1).

3.2 Navigation satellite positioning

Global Navigation Satellite Systems (GNSS) currently available include the widely-used Global Positioning System (GPS; USA) and Globalnaya Navigazionnaya Sputnikovaya Sistema or Global Navigation Satellite System (GLONASS; Russia), as well as Galileo (European Union) and BeiDou (China), which are currently being deployed. Additionally there are local systems such as Indian Regional Navigation Satellite System (IRNSS; India) and the Quasi-Zenith Satellite System (QZSS; Japan). Each system type consists of a network of satellites that orbit the Earth at between 18,000 - 25,000 km altitude.

The network satellites transmit their location and absolute, synchronised time, encoded in radio signals with at least two different frequencies. A GNSS receiver can decode these signals and calculate the distance to multiple satellites using the signal arrival times. In the case of global systems, the intersect of distances from at least four individual satellites enables a GNSS device to calculate location in geo-coordinates via trilateration. Single-Point Positioning (SPP) requires only one GNSS receiver (Hegarty, 2017). The horizontal positioning accuracy is at best within 5-8.5 m (Zandbergen and Barbeau, 2011) and the vertical accuracy is poorer still. This is because the visible satellites are more closely aligned in a horizontal plane and the Earth shields the signals from remaining satellites, which would provide more vertical information. Recent developments have focused on eliminating the need for multiple GNSS receivers and speeding up the time required to achieve accurate positioning (Kouba et al., 2017).

Measuring locations relies on a reference system (georeferencing) that is Earth-centered Earth-fixed. A catalogue of 3D positions is given by the International Terrestrial Reference Frame (ITRF). The latter falls to within ± 1 m of the World Geodetic System 1984 (WGS84) and is therefore used as the common reference frame for geo-positioning (Bock and Melgar,

2016). The *International Hydrographic Organisation* mandates the use of WGS84 as the horizontal reference for hydrographic mapping (Rizos, 2017).

Geo-positioning is based on the geographical coordinate system, which delivers the spherical coordinates latitude, longitude and height (geoidal geometry as global reference point). Measuring lengths and areas in spherical coordinates is not straight-
5 forward. For the purpose of ~~hydro(geo)logical~~ hydrogeological investigations, these coordinate points are transformed into a projected coordinate system, a 2D representation of the Earth's surface. Although, there is some uncertainty as to the origin of this projection (Buchroithner and Pfahlbusch, 2017), the most commonly used projection is the Universal Transverse Mercator (UTM) system, which divides the Earth into 60 zones and 20 latitude bands. Each zone is then assumed to be planar and coordinates are expressed in meters as Northing, Easting and Elevation (projected from the geoid to a flat surface with local zone
10 as the reference point). Note that height (vertical distance above the ground surface) and elevation (vertical distance above sea level) should not be confused.

Differential global navigation satellite system (DGNS) positioning can provide much better accuracy and precision than GNSS. This approach requires at least two GNSS receivers, one of which is stationary and located at a known point (base station). The base station uses single-point positioning in conjunction with its known location to calculate an error correction.
15 The second, mobile GNSS receiver (rover) uses the GNSS signals in conjunction with the error correction to calculate its distance from the base station. The error correction is determined from signal phase observations at both stations (Remondi, 1985). This can be achieved offline by post-processing the stored satellite signals in both receivers, or in real-time through a radio link between the rover and the base.

Recent developments in many countries have resulted in continuous operating reference stations (CORS) at strategic lo-
20 cations, whose error corrections can be accessed via mobile data networks, as long as there is network coverage. The most sophisticated GNSS devices can nowadays provide positioning with millimetre horizontal and sub-centimetre vertical accuracy (Li et al., 2015; Siejka, 2018). However, these innovations have yet to make it into commercial receivers. More typically, best-achievable horizontal accuracy and precision are 15 mm and 10 mm, respectively, whereas vertical accuracy and precision are 30 mm and 40 mm, respectively (Garrido et al., 2011). These numbers have been adopted for the purpose of our error anal-
25 ysis (Table 1). Interestingly, Kim Sun and Gibbings (2005) found that accuracy and precision did not show any dependence on the distance to the base station within their test area of about 11 km. It should be noted that these accuracies are achievable only when there is a sufficient number of visible satellites (for both receivers). When points of interest are near or under vegetation, the geo-positioning accuracy is significantly degraded (Bakuła et al., 2009).

When traditional surveying is undertaken there appears to be a horizontal or vertical distance dependent error whereas for
30 DGNS this is not the case (Table 1). Using the random error estimates, a horizontal cut-off distance at which the precision from state-of-the-art DGNS is better than that of a total station theodolite is ≈ 700 m. For vertical distances, DGNS become more precise than digital levelling when two locations are further apart than ≈ 15 km in the horizontal direction. In this case it is meaningless to derive vertical head gradients. Consequently, the surveying approach should be chosen according to the distance between the locations.

Table 1. Summary of precisions for the four different steps and methods required to calculate hydraulic heads and gradients. For a graphical explanation see Figure 1. Values are best possible estimates or collated from the literature.

Measurement	Step	Option	Method	Hor. Precision	Vert. Precision	Comment/Reference
Geo-position of GMI (Section 3)	1	A	DGNSS positioning	15 mm	40 mm	Using state-of-the-art DGNSS systems (Garrido et al., 2011)
		B	Digital total station	1 mm + 1 mm/50 m	1.9 mm/ \sqrt{km}	Horizontal (Braun et al., 2015), vertical (Ceylan and Baykal, 2006)
		C	Digital levelling	-	1 mm + 1 mm/km	Vertical only, error dependent on horizontal distance (Bitelli et al., 2018)
Point of head (Section 4)	2	A	Driller's record	-	≈ 500 mm	This is a crude estimate
		B	Assuming verticality	872 mm	38 mm	Based on 10 m borehole length and a 5° inclination (Section 4.2)
		C	Down-hole camera	-	20 mm	Precision of winch and visual detection of screen dimensions
Depth to water (Section 5)	3	D	Verticality log	87 mm	-	Based on 10 m depth and 0.5° inclination precision
		A	Dip meter	-	8.4 mm	Based on 256 manual measurements (Knotters et al., 2013)
		B	Chalk on steel tape	-	2 mm	Assuming twice the resolution of standard metric steel tape (Cunningham and Schalk, 2016)
		A	Vented	-	1.5 mm	Based on 15 months of data from 3 PTs (Section 6.2)
Automated water level (Section 6)	4	B	Non-vented/baro	-	3 mm	Precision is doubled due to the use of 2 devices
		C	Vibrating wire	-	7 mm	Standard deviation of best sensor during stability test (Zarriello, 1995)
		D	Laser-based	-	0.5 mm	As tested by Benjamin and Kaplan (2017)

4 Point of head measurement

4.1 Representative point of measurement

For a grouted-in piezometer (Location 2 in Figure 1), the measured pressure reflects the groundwater pressure at the vertical position of the sensor (Simeoni, 2012), and therefore this represents a true point measurement. By contrast, the water column in a GMI that is open to the atmosphere (Location 1 in Figure 1) equilibrates to the vertical groundwater pressure distribution along the subsurface screen. The mid-point of the screen is often selected as the representative point for the measurement. However, the appropriateness of this assumption has to be considered on a case by case basis. Vertical head gradients in an aquifer tend to be small under natural (i.e., not pumped) conditions, often less than 10^{-3} (this value would be typical for an aquifer with a rainfall recharge rate of 1 mm d^{-1} and a vertical hydraulic conductivity of 1 m d^{-1}). Having a lower resistance to flow than the surrounding aquifer, a piezometer provides a flow conduit (Freeze and Cherry, 1979; Elçi et al., 2003). These associated flow head losses are very small, thus the head within a piezometer is constant. Outside of the piezometer, the total head change in the aquifer along its screen depends on the screen length. For example, for a 2 m screen the head varies by no more than 2 mm for the quoted vertical head gradient in the aquifer. This can be taken as an indication of the maximum head error for a typical piezometer in an aquifer caused by uncertainty about the elevation of the point of measurement. When gradients are higher the error can be minimised by using as short a screen as possible, taking care that any pressure difference between the GMI and the formation is rapidly equilibrated by water movement (see also Section 6.6).

However, larger errors can be expected when vertical gradients are higher than the example value used, which may be the case near groundwater discharge zones, under pumped conditions, and in formations of low permeability. Rowe and Nadarajah (1994) found that for aquitard hydraulic conductivity tests, where the propagation of an induced head drop in a piezometer is recorded as a function of time, the representative point of measurement was biased towards the bottom of the screen, and that this significantly influenced the outcomes of the parameters to be determined. Moreover, as the gradients changed in time, so did the representative point of measurement. In layered aquifer systems, the water level in wells with long screens was found to depend on the transmissivities of the layers intersected by the wells (Sokol, 1963). These findings highlight the need for using short screens. However, the finite screen length of standpipe piezometers means that some uncertainty remains about the representative vertical position of the head measurement.

4.2 Borehole verticality and screen location

Despite best efforts in many countries on the mandatory requirement to report accurate information on the drilling and completion of GMIs, construction details are often reported at a precision of decimetres (not centimetres) and prone to significant systematic error. Due to the variety of different field and environmental conditions as well as the different qualifications and experience of drillers, we assume that the vertical screen locations can be estimated from driller's logs with a precision no better than about 0.5 m (Table 1).

The deviation from vertical of a GMI further results in uncertainty about s_h (NUDLC, 2011). The importance of borehole deviation surveys is critical in other industries such as oil and gas, where errors in the observed inclination angle and other

parameters of the monitored fracture-system geometry impact on the monitoring and interpretation of hydraulic fracturing (Bulant et al., 2007). Yet, in hydrogeology borehole verticality is typically ignored in particular when calculating heads gradients. Poorly aligned boreholes impact significantly on the integrity of casing, and hence increase the risk of flow short-circuiting and water column density-stratification (Section 6.4.2).

5 A thorough investigation of borehole deviation was conducted by Twining (2016), who applied a correction factor to water level data when a borehole deviation survey indicated a change of more than 0.06 m between the measured borehole length and the true vertical depth. From the 177 boreholes surveyed, correction factors to the historical water levels of these wells ranged from 0.06 to 1.8 m and inclination angles from 1.6 to 16 degrees. A comprehensive examination of borehole deviation was conducted in more than 100 boreholes drilled (up to 1,000 m deep) at the Swedish nuclear repository site by the Swedish
10 Nuclear Fuel and Waste Management Company (SKB) (Nilsson and Nissen, 2007). Their investigation provided an uncertainty of deviation measurement of the inclination of the borehole (up to 3 degrees) as well as an elevation uncertainty at the bottom of the borehole (up to 15 m for the boreholes measured).

Guidelines for drilling and water bore construction for plumbness and straightness is generally a '*do the best you can*' approach within practical limits using appropriate equipment and drilling operation (e.g. drilling centralisers, correct collar
15 and feed pressure) for the geological conditions (NUDLC, 2011; Treskatis, 2006; BDA, 2017). Drillers consider angles of less than 5° as acceptable (Bulant et al., 2007). Hence, the horizontal positioning error of a 10 m long borehole would become: $\sin(5^\circ) \cdot 10 \text{ m} \approx 872 \text{ mm}$, whereas the vertical error is: $[1 - \cos(5^\circ)] \cdot 10 \text{ m} \approx 38 \text{ mm}$. In the absence of more literature reporting on borehole deviations, we use this reported figure as the random error when determining the point of head (Table 1).

Since this uncertainty is much greater than the achievable accuracy of the GMI's geo-position, s_g , the verticality (plumbness)
20 of a borehole should be measured using downhole geophysical tools such as verticality probes or inclinometers. This includes an a gyroscope or an accelerometer to measure the vertical angle combined with a magnetometer to provide the probe's rotational position around the vertical axis. Both measurements can be logged continuously while lowering the probe. For example, assuming an industry standard precision of 0.5° (e.g., Verticality Sonde by GeoVista, UK) an otherwise straight 10 m deep borehole, the resulting precision in identifying the horizontal screen offset would be $\approx 87 \text{ mm}$ which is an improvement over
25 the crude guess based on a 5° angle according to best drilling practice. The borehole deviation survey should be combined with a down-hole camera to determine the position of the screen relative to the top of the GMI. We estimate the depth measurement precision of a typical system to be approximately 20 mm (Table 1).

5 Depth-to-water measurements

There are a number of different ways to measure depth to water (d_w , Equation 3). It is commonly done by hand and involves
30 the use of a measurement tape (Nielsen and Nielsen, 2006). Most groundwater projects today use electric water level meters colloquially called *dip meters*, which provide an audible or visible signal when a sensor touches the water surface. When the acoustic signal is not electronic but an audible noise is generated mechanically, for example by lifting and dropping a hollow

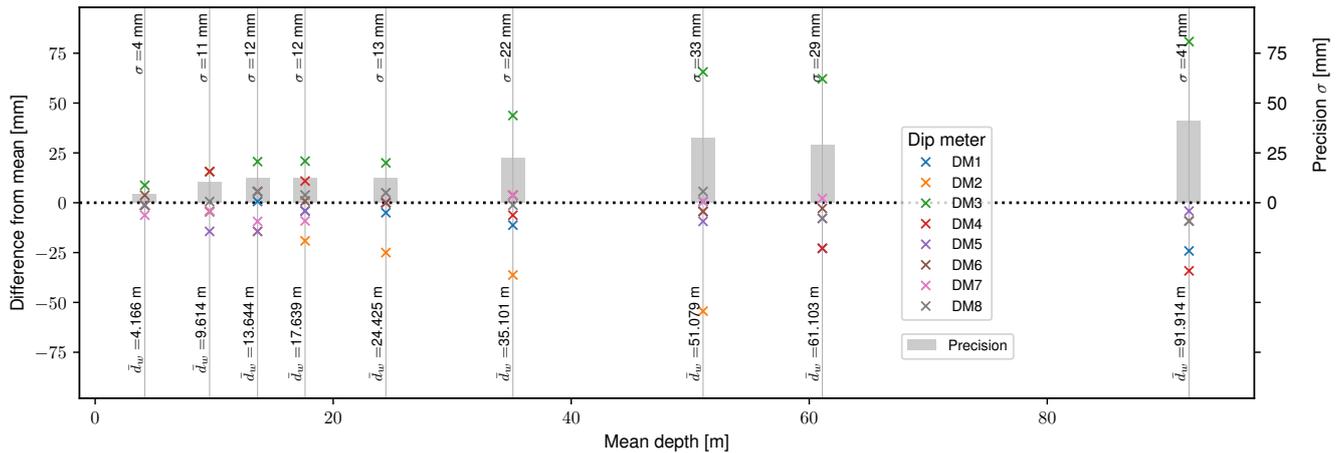


Figure 4. Manual measurement of 9 different groundwater depths using 8 different dip meters: Difference from mean and standard deviation (precision) over mean depth.

brass cylinder just touching the water surface, the instrument is called a *plopper*. Another inexpensive method uses a steel tape that is covered with chalk (Cunningham and Schalk, 2016).

Depth-to-water measurements should be performed frequently (at a minimum every 3 months) for checking the performance, and adjustment, of automatic sensors. Good-quality measuring tapes are marked every 1 mm (metric) or every 0.01 ft (imperial). The chalked-tape method can potentially deliver a precision that corresponds to the resolution of the graduated steel tape (Nielsen and Nielsen, 2006) (Table 1 and Figure 2), whereas dip meters and ploppers are generally read to the nearest centimetre. This may involve human measurement errors such as the switching of digits (e.g., noting 57 instead of 75) or reading to the wrong decimetre/metre marking on the tape (Knotters et al., 2013).

There is only minor information in the literature about the errors associated with manual head measurements. The lack of assessment is surprising given that manual measurement is the most important link which ties automated pressure time-series to a benchmark (Equation 3). Some controlled experiments have been conducted though, most recently in the Netherlands by Knotters et al. (2013). Sixteen operators, with varying degrees of experience, each took a reading in a total of 16 standpipes. Half of the readings were done with an electronic dip meter, the other half with a plopper. After discarding the obvious mistakes from the data set, the errors were fitted to a normal distribution with a mean and standard deviation of 5 mm and ± 8.4 mm, respectively (Table 1) for the electronic dip meter, versus 0.3 mm and ± 9.5 mm for the plopper. The measurements by Knotters et al. (2013) were representative for very shallow water tables. A poorer precision (0.05 ft = 15 mm) was reported by Atwood and Lamb (1987) (cited in Sweet et al. (1990)) for water levels more than 120 m below the surface measured by two observers using the same instrument within a short time period. Sweet et al. (1990) conducted a rather comprehensive experiment themselves but reported the errors as percentages which makes the figures difficult to compare to the other studies.

Knotters et al. (2013) noted that the graduations on some of the tapes that were used showed noticeable differences and that this caused systematic measurement error. Plazak (1994) compared three water level probes to a reference probe and found differences that increased with depth, reaching a maximum value of 0.1 ft = 0.03 m at a depth of 61 ft = 19 m. Comparable findings based on our own experiments are shown in Figure 4, which summarises variations in manual measurements using several commercial electric dip meters of various lengths to measure water levels at 9 depths between 5 and 90 m. Electric water level loggers by the same manufacturer differed by several centimetres, with differences of up to 0.12 m observed overall (one person taking the measurements sequentially for each borehole, using the same location at the top of casing for each bore). The differences increased with depth to the water table for several instruments, confirming the observations made by Plazak (1994) 25 years ago. Discrepancies of this magnitude preclude the use of data for accurately identifying small head gradients and call for replacement of the measuring instrument.

Wear (e.g. kinks and tears) on electric water level tapes causes additional discrepancies in measurements over time. Cunningham and Schalk (2016) detail procedures for calibrating electric water level devices before each use; this involves measuring the electric tape against a steel measuring tape kept in the office only for this purpose. For consistent time series, it is extremely important that a datum on the casing is used to always measure water level from the same point (Nielsen and Nielsen, 2006; Cunningham and Schalk, 2016). This seems obvious but there can be confusion when different operators are involved, and regular maintenance is required to make sure the mark stays visible. Repeating the measurement a number of times can avoid tape reading errors and ensures proper functioning of the electronic dip meter, which may give inconsistent readings sometimes (with saltwater, for example). Post and von Asmuth (2013).

In areas prone to vertical land surface movement (~~i.e., those that are tectonically active or experience land subsidence~~) a regular check of the elevation of the casing (z_g) is necessary. The causes for such movements can be manifold, and include tectonic processes, slope instability, freezing and thawing cycles (Rosenberry et al., 2008), or clay swelling and shrinking with changing moisture conditions. In peat areas, subsidence by compaction, oxidation or drying is a well-known cause for movement of the well casing (Drexler et al., 1999). Moreover, damage can occur to standpipes, either by vandalism or natural processes, such as ice expansion when the water inside the GMI freezes (Rosenberry et al., 2008).

6 Automated water level time-series measurement

6.1 Automated measurements

Automated measurement of water levels or pressures in GMI requires electronic devices capable of time keeping and sensing. Many commercial instruments combine a stand-alone clock, a sensor, an ADC unit, memory and a power supply in a single housing. There are also instruments that house only the sensor and are connected to a data logger that converts the sensor signal and stores the quantised readings. The focus of this section is on systematic errors that occur during automated collection of water level data in the field.

Automated instruments have the capacity to record for a long period of time ~~unimpeded~~unattended. Data loss as a result of logger chip or battery failure can be prevented by the automated transmission of instrument readings to a receiving data

management system via radio, infrared signals, a GSM network or satellites (Morgenschweis, 2018; Bailey, 2003). This is referred to as telemetry. The expansion of cellular network provider coverage and the reduced cost of data-only plans in the recent past have made telemetry systems a more accessible and viable option for remote ~~hydro(geo)logical~~ hydrogeological monitoring. Nowadays, transmitted data is stored on network servers which can be accessed in real time via computer or smart phones. Telemetry systems allow remote modification of sensor settings and identification of sensor problems, early identification of logger failure as a safeguard against data loss, re-synchronisation of time keeping to avoid time based errors (Section 6.6). However, the deployment of a telemetered system does not avoid certain errors such as sensor drift (Section 6.4.3). Consequently, to ensure the accuracy of automated water level measurements, frequent site inspections and manual measurements are still required.

10 6.2 Types of devices

At present, water level time series are typically determined using pressure measured using submersible or grouted-in PTs (Figure 1) and rely on Equations 3 and 4 to determine the water level. A comprehensive overview of PTs can be found in Freeman et al. (2004) or Hölting and Coldewey (2013).

The most popular type of PT consists of a piezo-resistive crystal made from silicone or ceramics, which acts as a strain gauge as it deforms under pressure. The deformation causes the electrical resistance of a Wheatstone bridge to change, which is gauged by recording the changing voltage due to a constant current. *Vented PTs* are connected to a venting tube that connects the air chamber of the submerged PT to the atmosphere. The measured pressure is the relative pressure (p_{pt} in Equation 4) and there is no need for barometric compensation. We estimate a best possible precision for vented PTs to be 1.5 mm (Table 1). This value is based on water level time series recorded with three vented PTs (with a range of either 10 or 20 m) inside a standpipe piezometer in Hannover (Germany) over a period of more than 15 months. Using one logger as a reference and calculating the difference with the remaining two loggers resulted in standard deviations of 1.4 and 1.7 mm ($n = 11,279$), thereby demonstrating excellent performance. However, readings can be influenced when the venting tube does not remain dry. A desiccant capsule is therefore attached to the tube and this can cause some practical difficulties when measuring in river beds or other areas subject to flooding, as well as when freezing occurs (Liu and Higgins, 2015).

Non-vented PTs measure absolute pressure $p_{pt,abs}$ and are converted to relative pressure p_{pt} by subtracting the barometric pressure p_b (Equation 4, barometric compensation), typically measured with a barometric PT nearby the GMI. The subtraction of barometric pressure from absolute pressure results in a loss of water level measurement precision, because the two instrument measurement errors accumulate (Section 7). Because of this, we estimate the best possible water level measurement precision as 3 mm (double that of vented PTs; Table 1).

A second type of PT is the so-called ~~vibrating wire piezometer~~ vibrating wire piezometer (VWP) that uses electromagnetic coils to excite a wire exposed to differing strain resulting from pressure changes. The square of the resonant frequency is linearly proportional to the pressure (Zarriello, 1995). VWPs are designed for long-term stability and are therefore used for closed GMI when the instrument is fully grouted-in (Figure 1b). However, re-calibration becomes impossible once installed (Contreras et al., 2007). One strategy to verify PT performance in that case (if the budget permits it) is to install three instruments at the

[same depth](#). VWPs of the non-vented and vented type exist, and some models have a pressure range of 10 MPa (equivalent to about 1 km of water) or sometimes higher. We estimate the best possible water level measurement precision of VWPs to 7 mm (Table 1).

[Wet/wet pressure transducers measure the pressure difference between two points that are both exposed to water \(Cuthbert et al., 2011\)](#)
5 . [Such devices are ideal for obtaining small head gradients, such as is required for measuring groundwater-surface water interactions, because they eliminate the uncertainties arising from barometric correction or the spatial positioning of two individual measurement points.](#)

There are also electronic water level measurement devices that emit a laser pulse and determine the depth of water from the time it takes the pulse to reach the water table and return to the sensor (known as LiDAR: light detection and ranging).

10 When connected to a time-keeping data logger, this technology is suitable for time-series collection (Benjamin and Kaplan, 2017). A recent development and test of a LiDAR-based system has demonstrated an outstanding precision of 0.5 mm (Table 1; Benjamin and Kaplan, 2017), but condensation and, in groundwater studies, borehole non-verticality, can interfere with the light reaching the water surface.

Another type of water level sensor is based on electronic capacitance measurement. It consists of two electrically isolated
15 plates or wires that are aligned in parallel at close proximity. Submergence of the wires in water creates a contrast in electrical capacitance compared to air, with values that are proportional to the submerged length. Their range (typically 1 to 2 m of water level) is smaller than piezo-resistive PTs (which can be used in water depths of 100 m or even more). An important advantage of capacitance probes is that they are rugged and can withstand overload, drying and freezing. In contrast to the measurement-tape and LiDAR technique, which measure d_w from the top downward, the capacitance probe sits in the water
20 column and records water levels on a data logger (similar to PTs).

6.3 Instrument range and resolution

In their instrument specifications, manufacturers typically provide the accuracy of a PT as a percentage of the full-scale (FS) range. Unfortunately, this number is not defined unambiguously. Typically, it may comprise a combination of a sensor's non-linearity (the relationship between p and voltage V not being a straight line), hysteresis (differing p - V relations during p
25 increases or decreases) and repeatability (the closeness of measured p values for the same V), and thermal artefacts (influence of temperature on the p - V relation). These are non-adjustable errors and are therefore not related to accuracy in the sense that they can be corrected by applying a simple offset to calibrate the instrument to a known value. Moreover, since there are different ways to quantify the instrument's deviation from the ideal p - V relation, the number specified as the instrument's accuracy can have a different meaning depending on a manufacturer's definition. This can even mean that an instrument with
30 0.5% FS range accuracy is as accurate as an instrument with 0.1% FS range accuracy (STS Sensors, 2017).

As a practical example of precision and resolution (see Section 2.4), Figures 5a and 5b show several days of automated depth to water level measurements made using different logger types. The difference between the graphs is the vertical scale, with the water levels in Figure 5a fluctuating at the mm scale and those in Figure 5b showing an overall decline of a few centimetres. The curves recorded by different water level measurement devices illustrate the discrete time and magnitude nature of automated

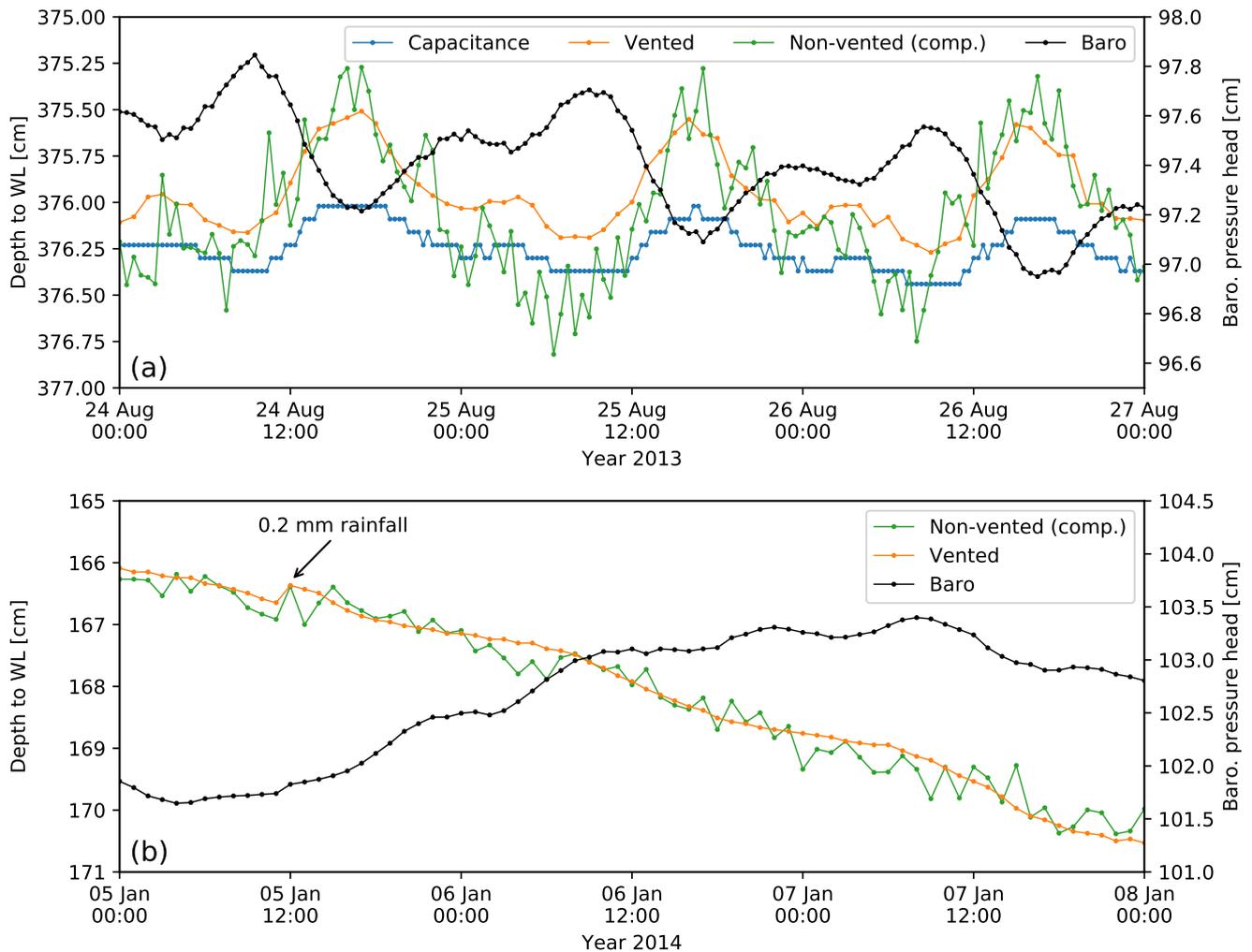


Figure 5. (a) Depth to water measured by three different instrument types, illustrating the influence of precision and resolution on head time series: Capacitance PT (blue line, high precision and low resolution), vented PT (orange line, high precision and high resolution), non-vented PT (corrected for barometric pressure) (green line, low precision and high resolution). The examples are from (a) Ti Tree Basin in the Northern Territory (Australia) and (b) a farm dam in South Australia (note the effect of a 0.2 mm rainfall event on 5 January, which is visible in the orange line but not in the green line). [The low precision achieved by the non-vented PT is specific for the particular instrument used and not representative for all PTs of this type.](#)

measurement. Here, the blue line illustrates high precision and low resolution values, the orange line shows high precision and high resolution measurements, whereas the green line represents low resolution and low precision data.

Current standard practice allows the quantification of subsurface processes and properties from time changes in heads due to either natural or induced causes. Common examples such as aquifer tests rely on large head changes that are sufficiently

resolved using off-the-shelf instruments. However, more recent research advances have demonstrated that subtle signals in hydraulic heads can also be used to passively quantify hydrogeological processes and properties. For example, Figure 5 demonstrates sub-centimetre diel (i.e., daily) fluctuations that originate from phreatophyte evapotranspiration (e.g. Gribovszki et al., 2013) or Earth and atmospheric tides (e.g. Acworth et al., 2015)(e.g. Acworth et al., 2015; McMillan et al., 2019)

5 . Such subtle signals can only be detected with appropriately high sensor resolution. Currently, it is advisable to deploy vented transducers to minimise errors resulting from clock differences and ~~precisions~~-imprecisions due to barometric compensation (refer to Section 2.1). For such intentions the measurement range must be minimised in favour of maximum resolution, which reduces the measurement error (Figure 3). It should be noted that when PTs are used to calculate gradients, the readings from non-vented PTs may be used directly without compensating for atmospheric pressure changes as long as the PTs all experience the same atmospheric pressure change.

10

6.4 Issues related to pressure transducers

6.4.1 Temperature effects

The response of piezo-resistive sensors to pressure changes is a function of temperature, hence most instruments record temperature alongside pressure and use this to compensate the readings. Nevertheless, the operation of PTs in transient temperature environments has been found to affect water levels that are calculated from pressure readings. For example, Cain et al. (2004) showed that when PTs are exposed to direct sunlight, thermal effects add noise to water level measurements. Sorensen and Butcher (2011) noted that temperature compensation often significantly compromises the accuracy of pressure readings.

15

For non-vented PTs, it is especially important to consider placement of the barometric PT to prevent adding noise from thermal effects into water levels during barometric compensation (Cuevas et al., 2010; McLaughlin and Cohen, 2011), which is demonstrated in Figure 6. The graph of water level versus time shows a clear diurnal variation in the time series recorded by a non-vented/barometric PT pair. The data were converted to water levels by subtracting the recorded barometric pressures from the total pressure recorded by the non-vented PT, which was suspended in a surface water pond. ~~The manual measurements taken between 3 and 6 November do not confirm the variations inferred by the logger data.~~ Inspection of the diurnal temperature variations of the barometric PT shows daily temperature variations of 10°C or more prior to 16 November, on which date the PT was placed in a more constant temperature environment. As a result, the periodicity that characterised the water level data before that date disappears from the water level time-series. Gribovszki et al. (2013) argued that such thermal effects should not affect vented PTs, which therefore should be suitable for fine-scale (e.g. sub-daily) measurements as required for evapotranspiration calculations. However, Liu and Higgins (2015) found that rapid changes in temperature on a sub-daily time scale can cause the air in the line to expand or contract, and that the relationship between temperature fluctuation and logger error varies between loggers.

20

25

30

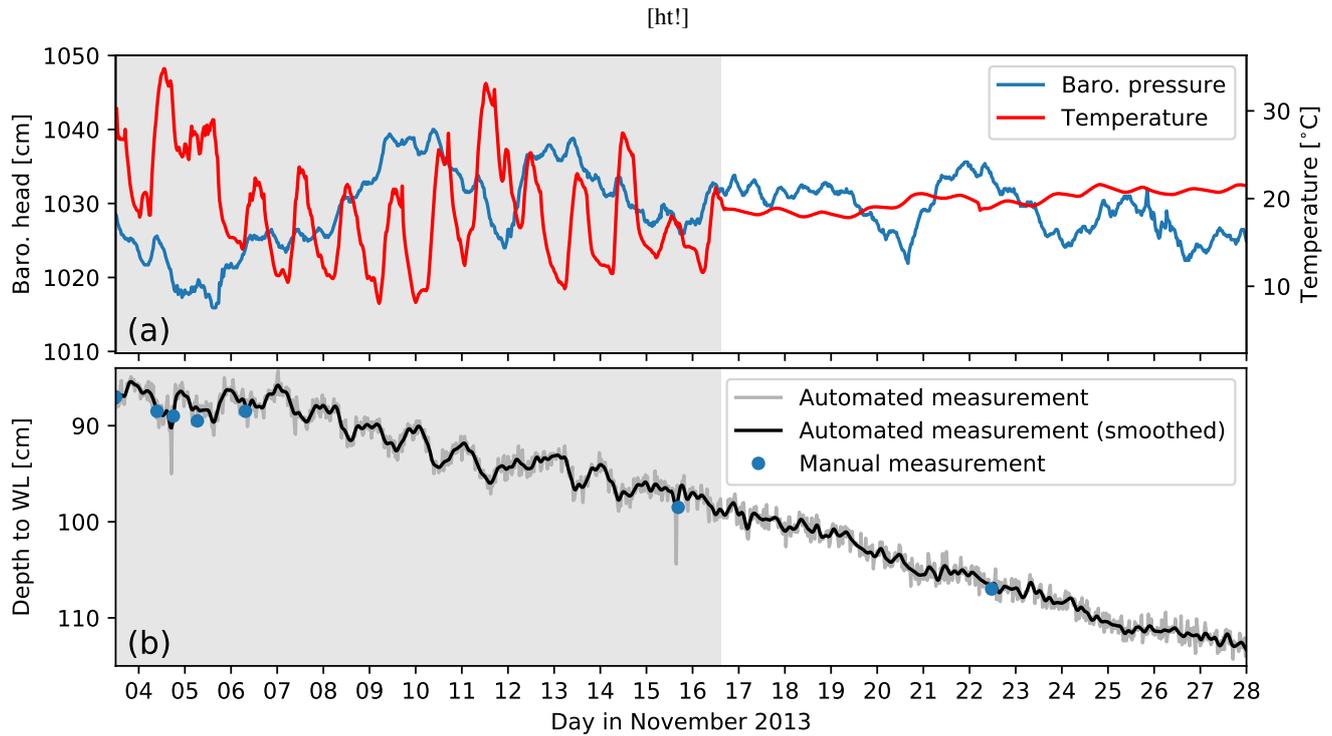


Figure 6. (a) Barometric pressure and temperature and (b) water level versus time. The data in (a) were collected using a logger that was exposed to significant temperature fluctuations prior to 16 November 2013. The water head time series in (b) was derived by subtracting the measured pressures shown in (a) from the total pressures measured using a non-vented PT. The artefacts caused by the temperature variations prior to 16 November 2013 are clearly reflected by diurnal oscillations of the water levels (grey shaded area). ~~Note that the manual~~ Manual dips confirm that there was no diurnal variability in the water levels (are indicated by blue dots).

6.4.2 Water column density changes

For internally consistent hydraulic head time series, it is imperative that the average density across the water column $\bar{\rho}_w$ in Equation 4 stays constant in time. Strictly speaking, this is never the case and the impact of the changes of $\bar{\rho}_w$ represent a systematic measurement error that has to be assessed and, when not negligible, corrected for. The effects are largest in groundwater systems with changes in salinity, such as coastal aquifers (Post et al., 2018). When the density varies within the water column and with time, $\bar{\rho}_w$ is given by (e.g., Post and von Asmuth, 2013)

$$\bar{\rho}_w(t) = \frac{1}{h_{pt}} \int_0^{h_{pt}} \rho(z, t) dz, \quad (11)$$

where $\rho(z, t)$ is the density as a function of the vertical dimension and time, h_{pt} is the vertical distance between the top of the water column and the PT (Equation 4).

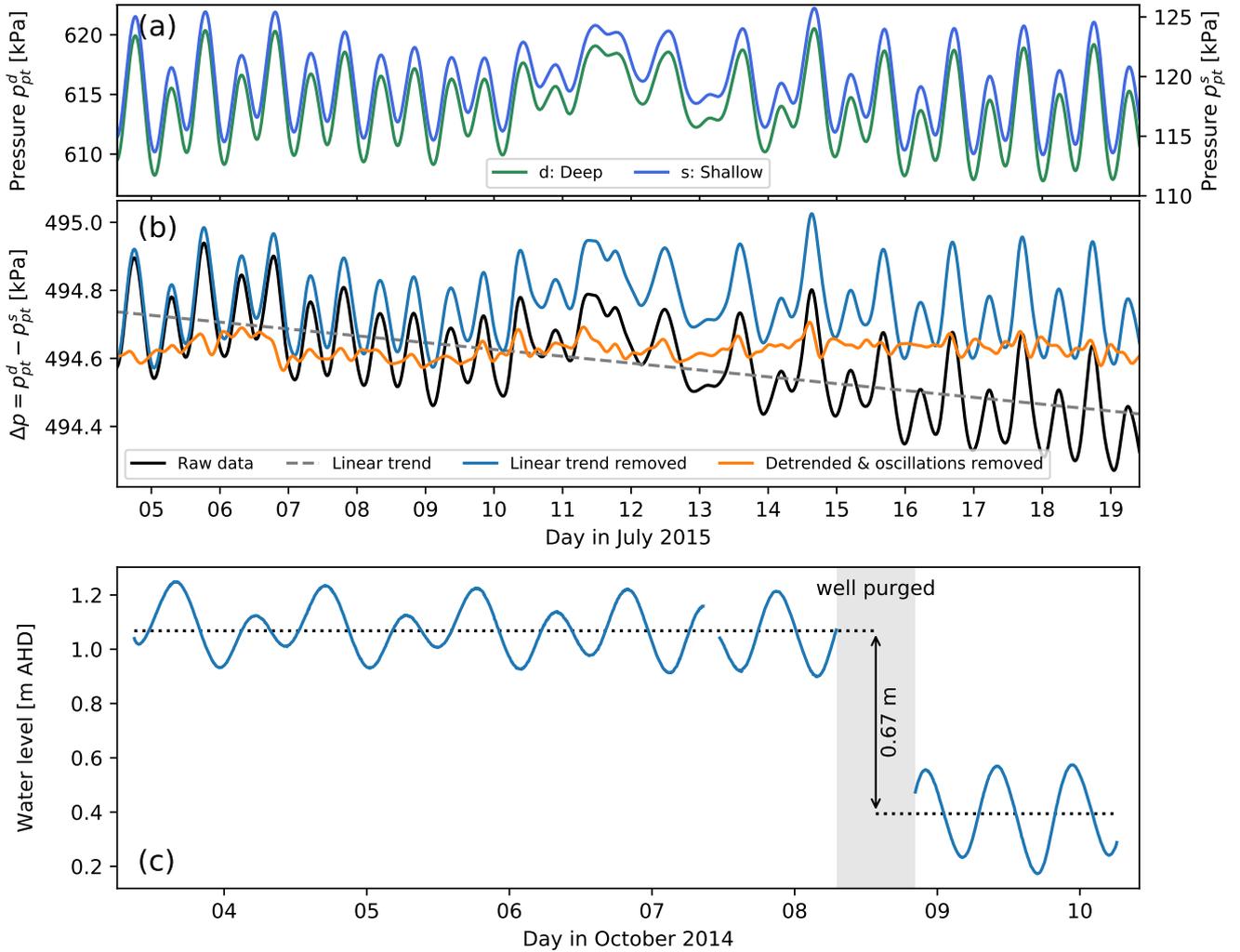


Figure 7. (a) Time series of water level pressure recorded by two transducers located shallow (superscript S) and deep (superscript d) in the same well. (b) Difference between the deep and shallow pressure measurements. (c) Water level in an observation well showing offset in mean water level after the well was purged for hydrochemical sampling (AHD is *Australian Height Datum*). Note that due to the high temporal resolution, the discrete measurement points are not shown and the time-series are drawn as lines instead.

Application of Equation 11 requires knowledge of the density distribution across the length of the water column at multiple times, which is seldom collected. Pressure to head conversion errors due to unknown knowledge of $\bar{\rho}_w(t)$ is therefore probably one of the most overlooked issues in head time series measurement. Some instruments provide a correction function based on the change of the electrical conductivity and temperature measured by sensors housed in the same instrument as the PT, but this is only meaningful if the density of the water column above the PT is constant. It is important to distinguish the effects

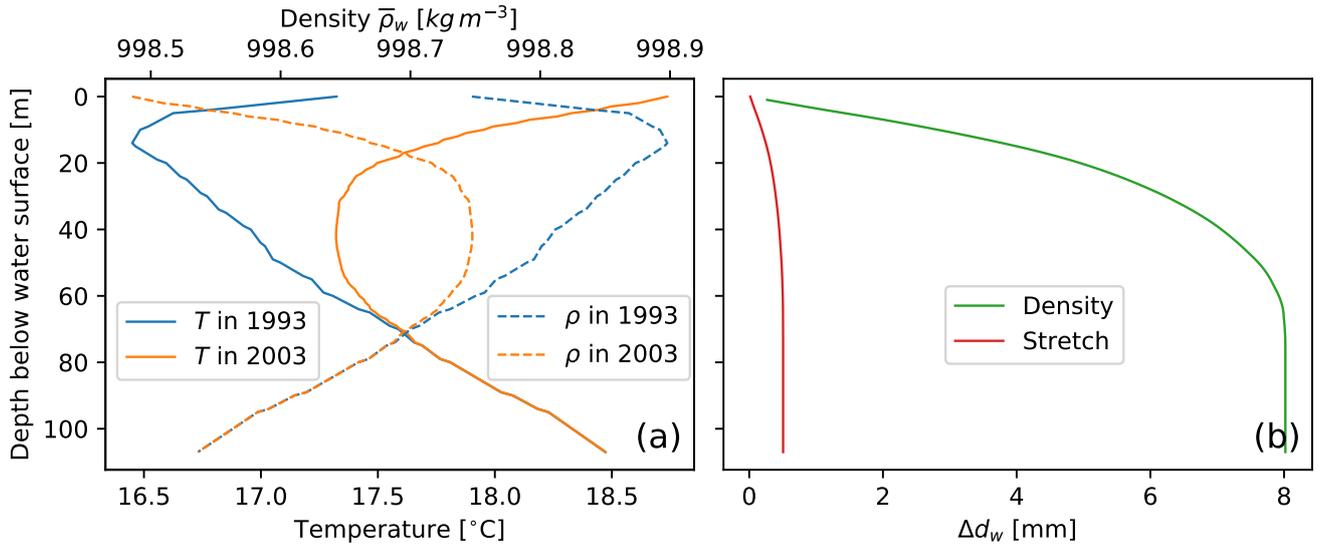


Figure 8. (a) Temperature and average density as a function of depth for an observation well in Japan that experienced significant warming due to urbanisation between 1993 and 2003. Temperature data from Yamano et al. (2009). (b) Theoretical head increase due to the temperature-related density decrease and head increase that would be perceived if the PT's vertical position is lowered due to thermal expansion of the steel suspension wire.

discussed here from the head corrections that must be applied when studying flow in variable-density groundwater systems (Luszczynski, 1961; Post et al., 2018).

A subtle effect of the change of $\bar{\rho}_w$ with time is shown in Figure 7. The upper graph shows the pressures recorded during an experiment whereby two PTs were hanging inside the same standpipe piezometer. One was located just beneath the air-water interface, and one was just above the bottom of the piezometer. The latter case corresponds to the way the pressures are recorded when heads are calculated using Method 2 (Section 2.1), whereas the former is representative for Method 1. Because of the well's vicinity to the sea, the recorded pressures vary with the tide. The difference between the PT readings is shown in Figure 7b. In a well with a constant $\bar{\rho}_w$, the pressure difference would be constant in time. Clearly, this is not the case here and two effects are notable: (i) a linear trend (grey dashed line in Figure 7b), causing the pressure difference to become smaller and (ii) oscillations that are superimposed on this linear trend.

The linear trend was due to leaking casing joints, which led to the ingress of fresh groundwater in the upper parts of the piezometer, as a result of which a salinity stratification developed. As more freshwater seeped in with time, $\bar{\rho}_w$ decreased by an amount $\Delta\bar{\rho}_w$ per unit of time Δt . In fact, the slope of the linear trend line is equal to $\Delta\bar{\rho}_w g h_{pt} / \Delta t = 200\text{ Pa d}^{-1}$, which for $h_{pt} \approx 50\text{ m}$ gives $\Delta\bar{\rho}_w / \Delta t = 0.04\text{ kg m}^{-3}\text{ d}^{-1}$, and this is roughly consistent with the estimate of $\Delta\bar{\rho}_w / \Delta t = 0.03\text{ kg m}^{-3}\text{ d}^{-1}$ derived from consecutive downhole probe measurements on 1 July and 5 August 2015. The superimposed tidal oscillations were caused by the change of the density stratification inside the piezometer standpipe with the tide: as the tide rose,

groundwater with an ambient, high salinity entered across the well screen and this caused more saltwater to stand above the deepest PT. The shallow PT, however, remained in the freshwater part of the stratified water column. Both PTs experienced the same increase in water column height above the sensor, but because this added height consisted of freshwater for the shallow PT, it sensed a smaller pressure change than the deeper PT. Correcting for these effects shows that the pressure difference becomes virtually constant although not all fluctuations disappear. The fluctuations around the mean difference decrease to become around 0.1 kPa (1 mm of water column height). The cause of the remaining fluctuations is not clear; they may be due to clock synchronisation issues.

While the previous example showed a subtle trend of relatively low magnitude, the time series in Figure 7c show the potentially large magnitude of an abrupt change of $\bar{\rho}_w$. In this case it was caused by the purging of the well for hydrochemical sampling. Prior to sampling, the water inside the well had a non-constant salinity, because it had not been properly developed at the time of construction. After sampling, the well was filled with water with the same salinity as the groundwater at the well screen. As a result, $\bar{\rho}_w$ increased from 1006.7 to 1015.1 kg m⁻³. Based on these density values, the length of the water column inside the well would have changed from 72 m to $(72 \cdot 1006.7)/1015.1 = 71.4$ m, i.e. a decrease of 0.6 m, which corresponds to the measured change of 0.67 m; the additional difference may be due to the removal of silt and other fouling material from the well screen by the pumping.

An example of the effect of temperature-related density changes on the head error is shown in Figure 8. In this example, the change in temperature was caused by urbanisation, resulting in a noticeable warming of the upper 75 m of the subsurface. This caused the density of the water column to decrease, and hence a longer column of water is required to balance the pressure at the screen. Figure 8b shows the magnitude of this effect as a function of depth, which in this example remains limited to less than one centimetre. Another effect that plays a role is the lengthening of the logger's metal suspension wire as it warms. Assuming a linear expansion coefficient for steel of $11 \cdot 10^{-6} K^{-1}$, the increase in wire length as a function of depth is showing in Figure 8c. The magnitude of this effect is less than 1 mm. Because this example was chosen to represent a case of relatively strong temperature increase for groundwater, it is expected that these values represent the upper bounds for thermal expansion effects, which thereby represent relatively small errors under typical groundwater conditions. Larger effects could occur though near aquifer thermal storage facilities, geothermal areas, or in very deep wells.

6.4.3 Measurement drift

Sensor drift is one of the most common errors in automated hydraulic head measurements. Here it is expressed as Δd_w , which is defined as the depth below TOC measured manually with a dip meter minus the depth measured by the PT. Sorensen and Butcher (2011) tested 14 different transducer brands commonly used in hydrogeological studies. For PTs with a range < 15 m H₂O, the drift was observed to be $-8 \leq \Delta d_w \leq 27$ mm after 99 days in the field, but the models with a greater range showed up to 5 times more drift. Data available to the present study from Syria, where 11 observation wells were equipped with vented PTs in January 2009 and were not inspected until June 2010, showed $-199 \leq \Delta d_w \leq 153$ mm, with only one of the PTs showing a negative Δd_w value.

In an extensive study of 473 piezometers all equipped with the same brand logger and inspected every three months for a total of two years, ? statistically evaluated Δd_w values based on a data set of 5,583 measurements. For 144 piezometers, a statistically-significant linear trend could be identified. The slope of the trend line was negative for 95 and positive for 49 of the piezometers. The drift was reported (as the median of the trend line slopes) to be -3.6 cm yr^{-1} and 4.4 cm yr^{-1} for the
5 negative and positive trends, respectively.

Apart from technical reasons that cause PT drift, fouling of instruments is a well-known problem that affects the quality of head time series. This can be due to the formation of mineral precipitates by hydrochemical processes (Sorensen and Butcher, 2011). Biological processes often build biofilms of microorganisms, or larger organisms such as snails attach themselves to a sensor. Biofouling filters consisting of copper coiled wire that can slow down these effects, but regular inspection and cleaning
10 are a requirement to prevent measurements from being compromised. Moreover, improper suspension cables may stretch, hence causing the logger to sit deeper below the water surface, or frequent removal of loggers for downloading may cause the wire length to change due to kinks. The cables of vented PTs may be large relative to the well diameter, and sometimes there is little room for the desiccation unit at the top, which may mean that the logger is not always returned to exactly the same position after the GMI was accessed for maintenance or other measurements (e.g. water sampling).

Drift introduces errors of unknown magnitude that remain unnoticed unless identified by frequent checks using an independent measurement (Rosenberry, 1990). The examples of field-observed drift in Sorensen and Butcher (2011) and ? show that drift is not generally linear. The rate of change can vary in time, sometimes suddenly, and even reverse direction. Frequent, independent d_w measurements by manual dipping provides the only means to correct for drift. Drift correction involves removal of the linear trend between manual measurements, which introduces uncertainty because of the linearity assumption.
15 Drift corrections must be carefully documented and at all times must the original data be stored alongside with the corrected data. Data downloaded at different times must be stored separately to ensure that a drift correction applicable to a particular block of data is not inadvertently applied to other blocks of data.

Another form of drift, which is related to the GMI and not the PT itself, occurs if the conditions in the GMI change such that the relationship between the recorded water level h_{pt} in the well and the groundwater pressure p in the aquifer (Equation
25 4) is not constant over time. When the change of $\bar{\rho}_w$ with time (Section 6.4.2) is responsible for this, it will cause $\Delta d_w \neq 0$. However, $\Delta d_w \neq 0$ cannot be used to detect measurement errors due to clogging of the well screen by suspended sediment particles, geochemical processes (e.g., iron oxidation) or biofilm growth. An example of the detrimental effect on time series because of the latter phenomenon is illustrated in Figure 9, which shows the water level in a piezometer as a function of time. The temporal dynamics remained very much subdued until the well screen was mechanically rehabilitated in June 1996
30 (Willemsen, 2006). As soon as the hydraulic connection between the piezometer and the aquifer was restored, the temporal variability of the head in the aquifer became apparent.

6.5 Clock drift

Automated water level and pressure recorders use an autonomous or external clock which relies on crystal oscillators (commonly made of quartz) for a counter measuring process that forms the basis for digital time keeping. The crystal oscillators are

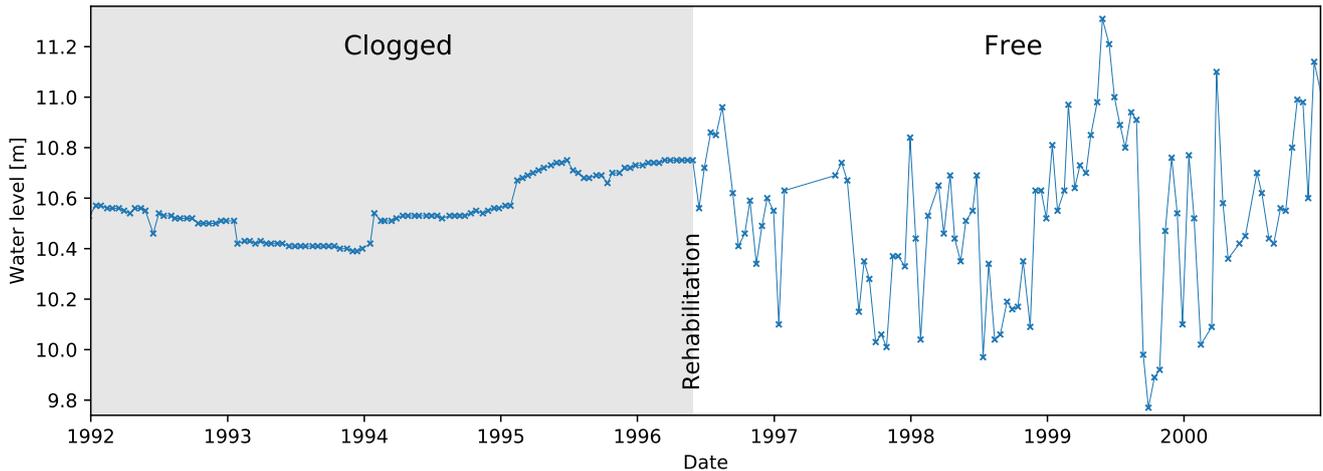


Figure 9. Water level versus time for a piezometer in the Netherlands that had a clogged well screen until it was rehabilitated in June 1996 (Willemsen, 2006). The temporal dynamics of the head in the aquifer were not registered by the piezometer until that time. Data were obtained from <http://www.dinoloket.nl> (downloaded on 16 January 2019).

highly accurate, yet small deviations in their oscillation frequency, which changes with time (a phenomenon known as ageing), can add up over long measurement periods. This results in a gradual drift of the internal clock in relation to the real time. Elimination of this form of systematic error can be achieved by synchronising the transducer clock with a more accurate clock such as in a field laptop [that is always set to the same time zone and does not update to daylight savings time](#). Clock stability is an important consideration when using multiple instruments. Examples include the barometric correction of absolute pressure measurements from a non-vented transducer, or the calculation of hydraulic gradients using two different time series.

The clock stability of 8 PTs was assessed during a long-term surface water-groundwater exchange monitoring program in the arid zone of Australia (Fowlers Creek at Fowlers Gap, New South Wales, Australia), where streams are dry for most parts of the year, but flow if there is enough rainfall (Acworth et al., 2016b). Monitoring stream flow under such conditions relies on long-term and accurate resolution hydraulic gradients. To monitor the spatial and temporal dynamics of stream flow we used streambed arrays similar to those reported in McCallum et al. (2014). Before deploying the PTs, the internal clock of the field laptop was synchronised to an online time server. This ensured that all loggers had the same time stamp. The transducers were setup to start logging on 21 October 2014 at 18:00 AEDT (*Australian Eastern Daylight Time*) with a sampling interval of 30 minutes.

Due to the remoteness of the field site, monitoring continued for over 2 years. After removal and disassembly of the streambed arrays, the internal clock of each PT was compared to that of a synchronised computer. The findings demonstrate that the majority of the PTs did not comply with the manufacturers specifications of $\pm 1 \text{ min y}^{-1}$, with most of the clocks running slower and the worst clock drift being $+7.5 \text{ min y}^{-1}$ (Table 2). [Such deviations are unfortunately not unusual for commercial PTs \(Post and von Asmuth, 2013\).](#)

Table 2. Example of an assessment of clock stability for 8 different standard PTs (AEDT is *Australian Eastern Daylight Time*). None of the PTs complied with the clock stability of $\pm 1 \text{ min yr}^{-1}$, as specified by the manufacturer.

Logger serial	PT end time AEDT	Actual time AEDT	Time difference [min]	Record duration [days]	Clock drift [min yr ⁻¹]
2004398	27/05/2016 14:24	27/05/2016 14:32	8	584	5.0
2004777	27/05/2016 15:10	27/05/2016 15:21	11	584	6.9
2020185	27/05/2016 15:15	27/05/2016 15:11	-4	584	-2.5
2020180	27/05/2016 15:39	27/05/2016 15:51	12	584	7.5
2005097	27/05/2016 15:39	27/05/2016 15:41	2	584	1.2
2005116	27/05/2016 14:38	27/05/2016 14:50	12	584	7.5
2005086	17/11/2016 10:13	17/11/2016 10:24	11	758	5.3
2020164	17/11/2016 10:30	17/11/2016 10:36	6	758	2.9

The influence of clock stability on measuring hydraulic heads and gradients is illustrated in Figure 10. In this example, a vertical array similar to that used in McCallum et al. (2014) was deployed in a streambed at Maules Creek (New South Wales). Resolving vertical head gradients over small distances is a significant challenge. Both PTs were calibrated against each other by placing the array inside a water bath overnight.

5 Figure 10a shows the pressure heads as well as the vertical head gradient during the experiment. Figure 10b illustrates the outcome if either one of the PTs was synchronised at a different time or as a result of clock drift. It is clear that the largest error arises during largest head changes with time, where the gradient disregarding time errors could be interpreted as either gaining or losing conditions with different magnitudes. Similar to this example, Post et al. (2018) showed how clock drift led to erroneous flow estimates in a coastal aquifer subject to [ocean](#) tides. Hydrological processes could be fundamentally
10 misinterpreted if time related monitoring errors are ignored, which is not always properly recognised.

6.6 Miscellaneous errors

The importance of setting the logger to the appropriate time resolution (sampling rate) is illustrated by Figure 11. Both lines show the same water level time series, but the red line shows how the curve would look if the measurement frequency had been set to twice daily, whereas the blue line shows the data as measured using an interval of 30 minutes. Obviously, the short-term
15 variations caused by the operations of a nearby production bore are not captured when an inappropriate measurement interval is chosen. Similar issues can arise in aquifers affected by [ocean](#) tides or river stage fluctuations.

When unresolved temporal head fluctuations occur between two consecutive automated measurement intervals t_i , a large discrepancy can arise between a manual measurement taken at time t_j and the nearest measurement at time t_i . When field

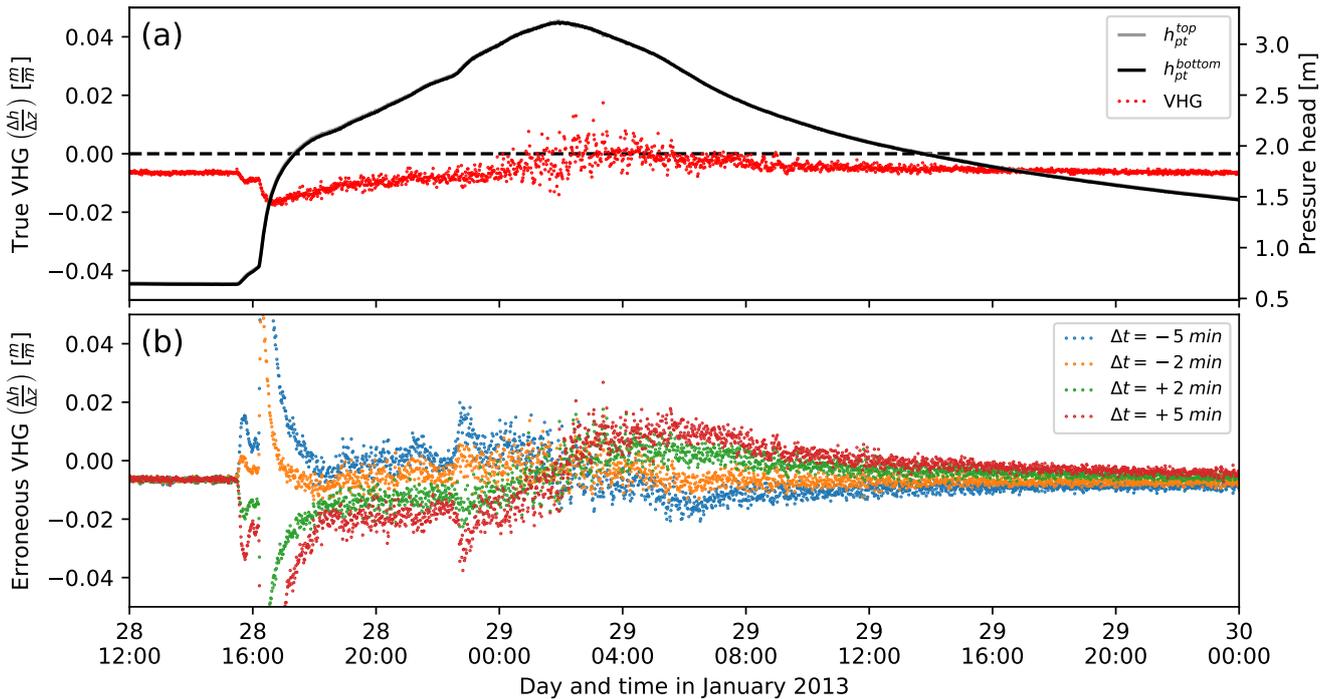


Figure 10. The influence of clock stability on calculating a vertical head gradient: (a) Pressure heads measured in a streambed using two PTs (note that the two lines are too close together to distinguish) with clocks that are in sync and the calculated vertical head gradient. (b) Erroneous vertical head gradients arising from time differences Δt due to out-of-sync instrument clocks caused by clock drift (the data was synthetically shifted).

personnel take manual measurements during their regular site visits, the timing of which is usually not determined by the logger recording settings but by logistical factors instead, considerable differences can arise from the fact that $t_j \neq t_i$. Using the data in Figure 11 as an example, a manual head measurement taken at time (t_j) of 09:11 on 5 September 2015 would be 0.83 m higher than the closest automated reading of the logger (set to a 12 hour sampling interval) at the time (t_i) of 12:00.

- 5 The difference is unrelated to any instrument error and is solely due to unresolved temporal variability. A manual dip taken at any time between the two sampling times would result in an error that falls within the grey box in Figure 11. While this hypothetical example represents an extreme case of this effect, misalignment of t_i and t_j is very common and it supports the contention by Sweet et al. (1990) that unrecognised hydrological processes are a form of measurement noise. To avoid such errors, a measurement interval of at most 1 hour suitable measurement interval must be chosen upon initial logger deployment,
- 10 which depends on the hydrogeological conditions at the measurement location. Only when it becomes clear that there is no temporal variability at this timescale can the sampling interval be increased to avoid unnecessary data handling and storage requirements.

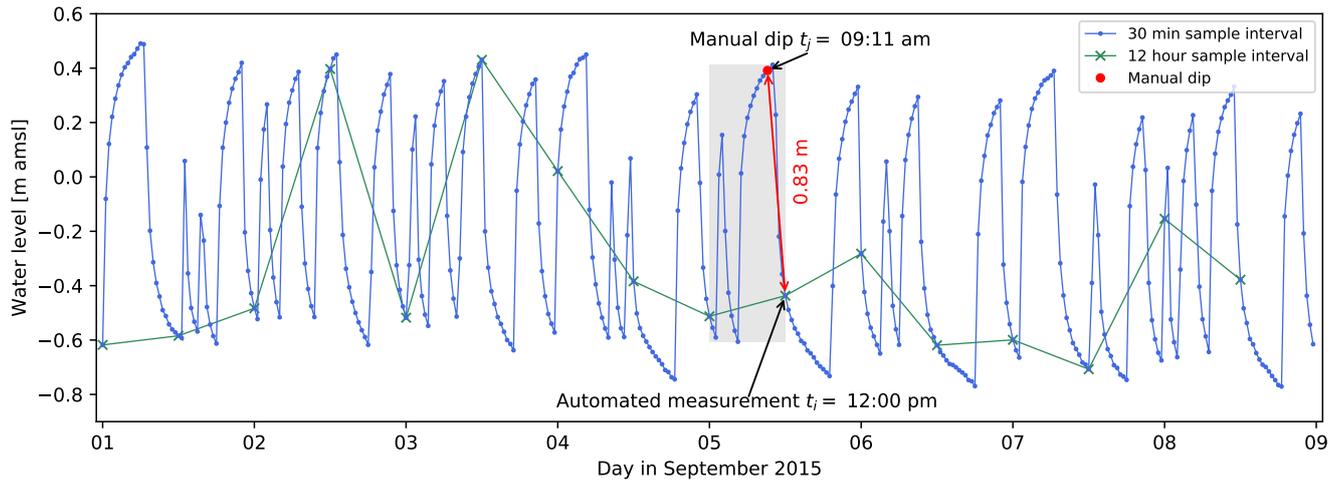


Figure 11. Water level in an observation well responding to irregular and frequent pumping from a nearby production bore automatically measured at two different sampling intervals (12 hours or 30 minutes). Note that a manual dip which is taken at a time that falls between samples of automated measurement (e.g., the time period indicated by the grey box) can result in a significant measurement error in dynamic hydro(geo)logical-hydrogeological systems.

Open GMI may suffer from a time delay in the water level response to changes in subsurface pore pressure, either because the well screen is partially clogged or improperly sized, or because the well volume is so large that water cannot flow fast enough through the surrounding porous media and well screen to allow equilibration of the water level inside the well with the groundwater pressure (e.g., Hvorslev, 1951). In low-permeability materials like compacted peat or fine-grained sediments, time lags can be on the order of hours or even longer, which precludes the registration of the response to rapid processes such as for example river flooding events (Hanschke and Baird, 2006), but also to water level changes induced by pumping, ocean tides or atmospheric pressure changes (e.g., Bredehoeft, 1967). Observation wells may also take appreciable time to readjust after the water level inside was raised by water displaced from inserting measurement instruments.

There is a variety of reasons why PTs do not always accurately record the water level in open GMI, many of which can be prevented by proper installation. When suspension cables are attached to well caps, the logger may not always be in exactly the same position after having been removed from the GMI. Some lightweight pressure transducers may experience buoyancy, especially in saltwater, and hence their vertical position is not constant in time. As a consequence of suspension cables being too short, PTs may end up being suspended above the water surface inside the GMI when the water level falls, and hence record the atmospheric pressure (Mäkinen and Orvomaa, 2015). Air bubbles that become entrapped in the PT after the water level rises again can cause inaccurate readings and must be removed.

When open GMI becomes artesian, which can occur when water levels rise higher than the standpipe's top, the PT no longer indicates the true water level. When the PT is too deep to withstand the pressure of the water column (the so-called burst pressure, usually about twice the measurement range), the sensor may damage and the logger will malfunction. Freezing

of the water column and lightning strikes can also cause damage to the PT (Freeman et al., 2004). Sometimes PTs show erratic readings for no apparent reason, which can be due to the overheating of electronics. Temperatures in the standpipe sticking up above the land surface can easily exceed 40°C, the upper temperature threshold for correct functioning of electronic parts, due to sun exposure. Shading or ventilation measures are therefore also an important part of GMI.

- 5 One issue not discussed in the literature is the considerable confusion that arises due to daylight saving time (DST) related clock adjustments. Perhaps this is because it is considered a trivial point, but an important one nonetheless that must be specifically addressed in the measurement protocol. Some devices automatically adjust to daylight savings time, whereas others do not. When they do, the instrument's clock setting depends on the time of year it was set up. Some manufacturers apply DST corrections only when the recorded data are exported to a file and this depends on the computer's DST settings.
- 10 The same data could therefore also end up with different time stamps if multiple computers are used.

7 Random error propagation

Figure 1 illustrates steps required to calculate a head gradient. In what follows it is assumed that all systematic errors have been eliminated from this process in a way that the only error remaining is the random error. For the sake of simplicity and in the absence of further information, we also assume that all random errors are normally distributed and not correlated (Table 1). Furthermore, horizontal errors stated in Table 1 are isotropic, i.e. do not vary in the x and y directions. Note that all error quantification in this subsection assumes that standard practice consists of (see Table 1 and Figure 12a):

15

1. Horizontal distance measurement using a total station (step 1 option B), vertical distance measurement using digital levelling (step 1 option C). Distance errors are limited to the respective errors from DGNS (step 1 option A). While GNSS surveying with a single receiver is useful for mapping, the precision of coordinates is not high enough to determine the distances between GMI for head gradient calculations.
- 20 2. Point of head measurement using downhole camera for vertical (step 2 option C) and verticality for horizontal error (step 2 option B) assuming a 10 m deep well;
3. Manual water level measurement using a dip meter (step 3 option A) assuming no depth dependency of the random error;
4. Automated pressure measurement using a vented transducer (step 4 option A).

25 Figure 12a graphically compares the random measurement error magnitudes for the different steps and methods summarised in Table 1. We stress that the adopted values reflect the absolute best case scenario from current standard field practice.

The random error associated with head differences arises from steps 1, 3 and 4 and can be expressed as

$$\delta\Delta h = 2\sqrt{(\delta z_g)^2 + (\delta d_w)^2 + (\delta h_{pt})^2}. \quad (12)$$

We use the random errors associated with standard practice in this equation to estimate the minimum achievable precision (combined random error) when calculating head differences as $\delta\Delta h = 0.017$ m. This error is somewhat higher than the findings

30

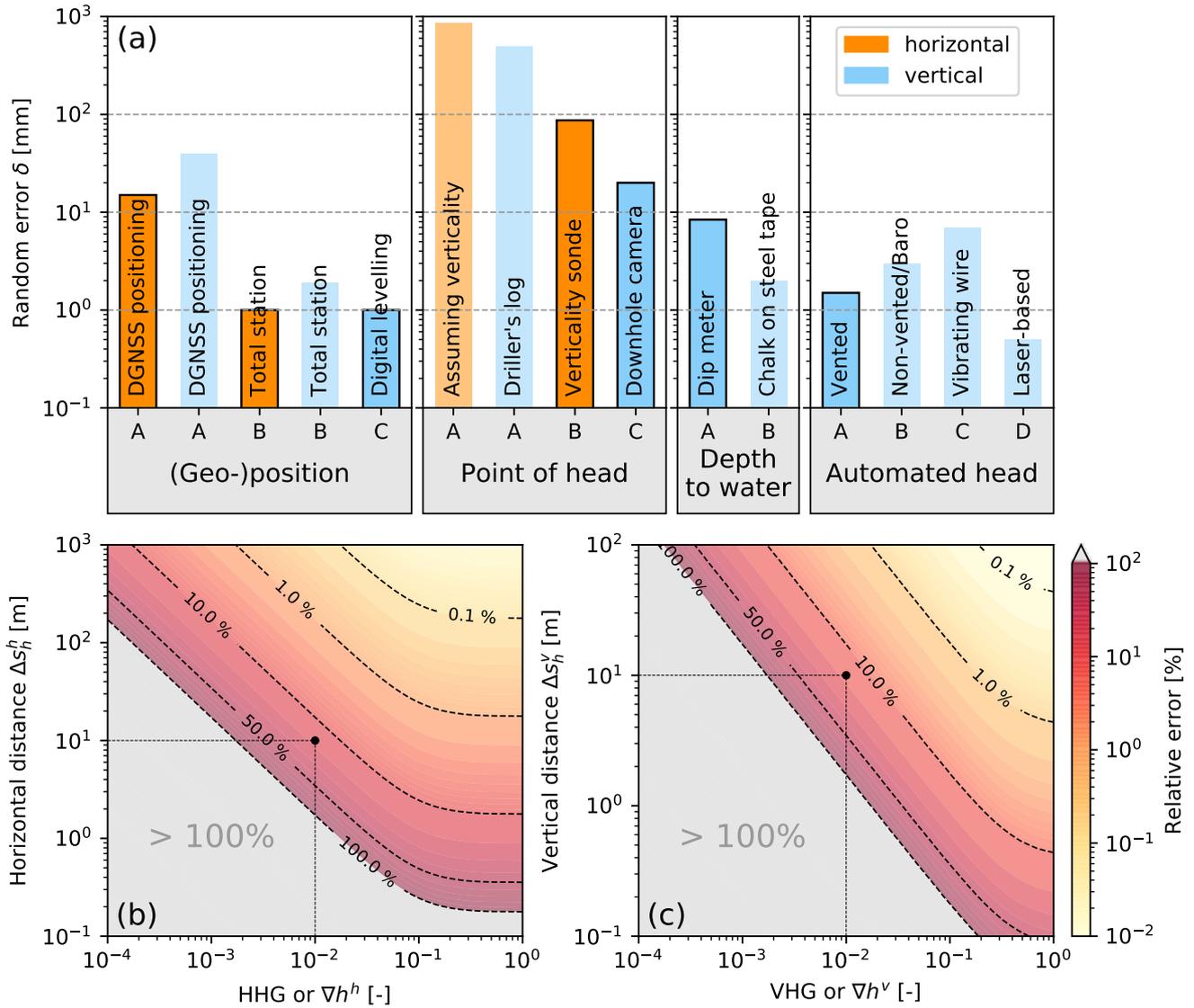


Figure 12. (a) Visual comparison of horizontal and vertical random errors [listed as based on](#) precision values in Table 1 (note that some errors are distance dependent) for the different steps (Figure 1) and method options. Minimum achievable relative random head gradient error in the horizontal (b) and vertical (c) direction calculated using standard practice measurements (highlighted by the black frame: step 1 options A,B,C; step 2 options B,C; step 3 option A; step 4 option A). Note that HHG and VHG values with errors exceeding 100% are blanked out in (b) and (c). [Please note the example error calculation in the main text.](#)

by Devlin and McElwee (2007), but lower than the field-based values ($\delta\Delta h = 0.022$ m) reported by Post et al. (2018). Measured head differences that are smaller than this value will not allow much confidence in detecting the direction of groundwater flow.

To improve this precision, approaches to reduce the achievable random errors when measuring steps 1, 3 and 4 must be found (Figure 1), likely resulting higher effort and cost than what is currently standard practice.

Using the measurements illustrated in Figure 1, the horizontal hydraulic head gradient (HHG) is calculated as

$$\nabla h^h = \frac{[z_g - d_w + h_{pt}]_2 - [z_g - d_w + h_{pt}]_1}{\Delta s_h^h}, \quad (13)$$

- 5 where the numeric subscripts depict the two locations. Analogously, the vertical hydraulic head gradient (VHG) can be determined as

$$\nabla h^v = \frac{[z_g - d_w + h_{pt}]_2 - [z_g - d_w + h_{pt}]_1}{[z_g + \Delta z_p]_2 - [z_g + \Delta z_p]_1}, \quad (14)$$

A propagation of random errors accounts for the errors involved in measuring the different variables explained in Figure 1. The relative error for the head gradients is as follows

$$10 \frac{\delta \nabla h^{h,v}}{|\nabla h^{h,v}|} = \sqrt{\left(\frac{\delta \Delta h}{\Delta h}\right)^2 + \left(\frac{\delta \Delta s_h^{h,v}}{\Delta s_h^{h,v}}\right)^2}, \quad (15)$$

where

$$\delta \Delta s_h^{h,v} = 2\sqrt{\left(\delta s_g^{h,v}\right)^2 + \left(\delta \Delta s_p^{h,v}\right)^2}. \quad (16)$$

- For the VHG case, the $\delta \Delta s_h^v = \delta \Delta z_h$, $\delta s_g^v = \delta z_g$ and $\delta \Delta s_p^v = \delta \Delta z_p$. The latter are the vertical positioning errors of the GMI and point of measurement resulting from a non-vertical borehole (steps 1 and 2). We use equation 15 to calculate the minimum achievable random relative error for HHGs and VHGs as a function of horizontal or vertical distance between two points of head measurement (Figures 12b and Figure 12c).
- 15

- Figure 12 clearly demonstrates the relationship between HHGs or VHGs and distance between the points of head. In general, the greater the distance between screens, the smaller the relative head gradient error. For example, the random error of determining a HHG or VHG of 10^{-2} at a 10 m horizontal or vertical point of head distance is $\approx 17\%$ (see examples in Figure 12).
- 20 Figure 12b further illustrates that measuring a HHG $< 10^{-4}$ with an error less than 100% requires a distance $\gtrsim 170$ m between points of head. VHGs of 10^{-4} are unresolvable within the considered maximum vertical distance of 100 m (Figure 12c). We stress that these errors are the best case scenario, as in reality there is a likeliness of additional systematic errors contained in the measurements. The errors calculated here are thus unlikely to be achieved in practice. Extraordinary effort must be put towards improving the precision of measurements when head gradients less than 10^{-2} are to be calculated for distances smaller
- 25 than 10 m (Figure 12b). Note that in order to additionally determine the direction of the gradient, a minimum area between GMI is required (Devlin and McElwee, 2007). [In practice, heterogeneity of the hydraulic conductivity will further add to the uncertainty of groundwater flow estimates.](#)

8 Concluding remarks

- Reliable water level measurements are at the core of every ~~hydro(geo)logical~~ [hydrogeological](#) investigation and the measurement error determines which processes or properties can be resolved. We have analysed unpublished and published data to
- 30

quantify the best possible accuracy and precision of hydraulic head measurements using commonly available, state of the art, commercial instruments. By propagating the random errors, we find that with current standard practice, horizontal head gradients $< 10^{-4}$ are only resolvable at distances $\gtrsim 170$ m, and that it takes extraordinary effort to measure hydraulic head gradients $< 10^{-3}$ over distances < 10 m. However, we consider these estimates very optimistic, as they assume that systematic errors are absent or that systematic error corrections do not introduce additional error.

The magnitude of systematic errors tends to be much larger than that of random errors and hence failure to recognise systematic errors can seriously compromise the outcomes of an investigation. It is difficult to establish if systematic errors are accumulating or cancelling, and hence they are must be avoided, or identified and corrected. In part, systematic errors are due to the measurement conditions in the field, which are not easy to control and negatively affect instrument performance. But, other factors play a role too, including improper instrument use, faulty or unsuitable GMI (e.g., long well screens), and the lack of measurement protocols that pay due consideration to all sources of error. Some measurement techniques have not seen performance improvement in decades, and there does not seem to be the same quest for measurement error reduction in ~~hydro(geo)logy~~ hydrogeology as there is in other fields of science, where the smallest of dimensions are measured with ever-better accuracy and precision and advances in measurement technology are pushing the frontiers of science.

We acknowledge that the measurement error with available technology could already be sufficiently small (e.g., within a few centimetres) for a lot of practical applications. ~~Still, improved standards for water level measurement would be an important step towards better hydro(geo)logical~~ However, the quantification and reporting of measurement error does not seem to be commonplace yet. Moreover, existing standards like Spane and Mercer (1985) or Freeman et al. (2004) contain useful guidelines for a maximum error as follows: (1) ± 3 mm (0.01 ft) for general applications, (2) 0.1 % of expected water level changes, (3) 0.01 % for cases where the depth to water exceeds ≈ 30 m (100 ft). Such standards must see wider uptake, and the development of more sophisticated or site-specific standards, suited for a particular study area or research objective, would be important steps towards better hydrogeological data quality and consistency. Moreover, technological advances are necessary to enable the measurement of vertical flow within an aquifer, the subtle temporal head fluctuations related to tidal cycles. Increased sensor performance and sensitivity would underpin new developments, such as the use of the ground-water response to Earth and atmospheric tides to characterise the degree of groundwater confinement (e.g. Butler et al., 2011; Acworth et al., 2017) and quantify compressible subsurface properties (e.g. Aeworth et al., 2016a; Rau et al., 2018) (e.g. Acworth et al., 2016a; Rau et al., 2018; McMillan et al., 2019). Such advances highlight the need to innovate beyond standard practice to support research in the hydrogeological sciences. We believe that researchers and industry should work together and find ways to increase instrument performance.

The following list of recommendations synthesises the findings from our study and focuses on aspects that could considerably improve the current practice of hydraulic head measurement. These are:

Elimination of systematic errors: Our estimation of the minimum achievable random error across all measurements presumes the absence of systematic errors (Table 1, Figure 12). Not all systematic errors (e.g., sensor drift) can be eliminated, but to minimise human error, measurements should be conducted exclusively by personnel that has received formal training. Moreover, a detailed measurement protocol must be designed and periodically evaluated, which outlines

the procedures for measurement, maintenance, note keeping (using standardised field data sheets) and data storage and handling.

Point of measurement: Our review of the literature demonstrates that GMI can significantly deviate from vertical (Section 4).

The resulting error in the point of head measurement is larger in the horizontal compared to the vertical direction (step 2 in Figure 1). While this potentially introduces one of the largest errors, it is generally ignored when calculating head gradients (Table 1 and Figure 12). If investigations necessitate the detection of small HHGs, we recommend measuring the borehole verticality using down-hole profiling tools in open GMI. The best possible precision in measuring the point of head is achieved by combining a verticality sonde with a downhole optical camera. Geophysical logging (Keys, 2017) or flow meter measurements can identify GMI construction errors and ageing issues, such as casing leaks. For open GMI, joint interpretation of barometric pressure and water level time series is required to determine hydraulic heads (*barometric correction*, Section 2.3). For closed GMI, the point of head measurement accuracy depends on the details contained in the original drilling report if inclinometer casing is not used in the installation (McKenna, 1995; Mikkelsen and Green, 2003).

Geo-spatial positioning: Our error propagation analysis (Figure 12) clearly demonstrates that precise measurement of the

horizontal and vertical distances between GMI is paramount to resolve the small hydraulic gradients inherent to ground-water investigations. This is particularly important when the GMI locations are in close proximity. Geo-spatial data from single-receiver GNSS should only be used for mapping but not to calculate distances. Traditional surveying techniques deliver more precise results for horizontal distances <700 m compared to DGNSS. Vertical distances should only be calculated using data from digital levelling and not DGNSS. If possible, leap-frogging the survey device should be avoided (Section 3).

Automated head measurements: The widespread use of automated PTs for hydraulic head and gradient measurement has

perhaps led to the impression that manual measurements have become less important. Our analysis demonstrates that regular, frequent manual water level measurement remains important as it is the only way to verify that PTs are accurately recording the correct water level (Section 6.4.3). It also improves the precision of the automated measurements by averaging out the error introduced from manual dipping (Table 1). Given the significance of manual measurement, it is surprising to note that commercially available dip meters show as much error today (Figure 4) as a quarter century ago (Plazak, 1994). Telemetry does not obviate field site visits, but only offers the convenience of not having to download devices and the advantage of being able to detect potential problems remotely, albeit at a higher cost of installation and maintenance (e.g., data service and connection problems).

Time related errors: Automated transducers rely on the stability of their internal or external time base once synchronised

with the clock of the device that is used to set up the logging protocol. We demonstrate that clocks can drift significantly (Section 6.5), which leads to silent measurement errors and false interpretations (Figure 10), especially for highly dynamic systems, where uninterrupted long-term monitoring is required. In reality, the clock stability error must be

doubled when non-vented PTs are used to assess barometric effects. We recommend that the clock is re-synchronised as frequently as possible or, where this is impossible, careful documentation of the device's internal clock status when monitoring is finished. Such practice is not always supported by off-the-shelf devices and the limitations of the software and device have to be trialled before deployment. Good time keeping practice also includes the use of one and the same field laptop, which is regularly synchronised with a time server. Moreover, we recommend the use of an absolute time base, for example *Universal Time Coordinated* (UTC), to avoid systematic errors arising from daylight savings time confusion.

Density and temperature effects: We demonstrate that automated pressure measurement and accurate conversion into water levels necessitates knowledge of the average density of water inside the borehole ($\bar{\rho}_w$, Equation 11). Since water density depends on the amount of dissolved substances as well as temperature, there is a need for measuring water temperature and electric conductivity across the length of the water column to establish their potential influence. If the water density inside the open GMI is not constant, the best solution is to position the PT such that it measures p_{pt} at elevation z_h (Method 2 in Section 2.1), although this may come at the cost of greater measurement error due to a larger PT range. Further, PT readings are often affected by temperature despite internal compensation (Section 6.4.1). While subsurface temperatures beyond 2-3 m depth are generally roughly stable, avoiding temperature effects can be a significant problem when measuring water levels or barometric pressure at or near the surface.

Type of pressure transducer: The type of PT is an important consideration that should be made according to the purpose of the investigation. For general groundwater monitoring away from topographic depressions and waterways (no risk of borehole over-topping) we recommend vented PTs, as long as there is no problem with keeping the venting tube dry. Because vented PTs measure a relative pressure instead of an absolute pressure, they have a smaller range and do not require a separate instrument to simultaneously record the atmospheric pressure. As such, they have better accuracy, precision, and resolution. Also, there is less risk of human error and their use avoids the problems that are introduced with two PTs (i.e., sensor and clock drift). Nevertheless, barometric pressure must still be acquired in order to perform a barometric correction. For reliably resolving head gradients and flow direction at small vertical distances, for example when assessing surface water-groundwater interactions, we recommend the use of wet/wet differential pressure sensors (e.g., Cuthbert et al., 2011). Quartz oscillator PTs are much more accurate than the commonly-used strain gauge type PT. However, they have hardly been used in groundwater studies to date, probably because of their higher cost.

Technical specifications: In the technical specification of PTs, much of the focus is on accuracy as a percentage of the full-scale range. We noted that this value is not consistently defined between manufacturers and may contain adjustable (i.e., errors that can be corrected using manual depth-to-water measurements) as well as non-adjustable errors (hysteresis, repeatability, non-linearity). Before purchasing, it is wise to approach manufacturers and enquire about the various technical details. Consideration must be paid to minimising the measurement range in favour of maximum possible resolution (Section 2.4). Practice has shown that PTs have high failure rates, hence reliability is also an important selection criterion.

As a final remark, documentation of measurement procedure is critical for data validation. Without any assessment of the measurement uncertainty it is impossible to assign a quality label to the data, which severely limits their worthiness for consideration in public databases. In addition to data collection protocols, quality control procedures must be in place to ensure the reliability of the distributed water level data. The development of such procedures should be considered in future work.

- 5 *Author contributions.* The idea for this manuscript was conceived by GCR and VEAP, and they were the lead authors. Everyone contributed with ideas, data and writing. GCR coordinated the effort and produced the figures and tables. MAS was supported by funding from the Australian Research Council, grant DE150100302. We thank the editor Brian Berkowitz for handling, the reviewers Don Rosenberry and anonymous for assessing and some attentive readers for suggesting improvements for this work.

Competing interests. No competing interests.

- 10 *Acknowledgements.* Some of the data used in this analysis were collected with equipment provided by the Australian Federal Government financed National Collaborative Research Infrastructure Strategy (NCRIS). Some of the groundwater data are available through the NCRIS Groundwater Database: <http://groundwater.anu.edu.au>. [This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 835852.](#) GCR acknowledges support from the Australian NSW State Government's Research Acceleration and Attraction Program in ~~2018.~~2019. EWB acknowledges funding support from the
15 Australian Research Council (ARC) Linkage Project LP150100588. The authors thank Mr Nick White for collecting the data shown in Figure 4 and Mr Michael Teubner for the manual measurements shown in Figure 6.

References

- Acworth, R. I., Rau, G. C., McCallum, A. M., Andersen, M. S., and Cuthbert, M. O.: Understanding connected surface-water/groundwater systems using Fourier analysis of daily and sub-daily head fluctuations, *Hydrogeology Journal*, 23, 143–159, <https://doi.org/10.1007/s10040-014-1182-5>, 2015.
- 5 Acworth, R. I., Halloran, L. J. S., Rau, G. C., Cuthbert, M. O., and Bernardi, T. L.: An objective frequency domain method for quantifying confined aquifer compressible storage using Earth and atmospheric tides, *Geophysical Research Letters*, 43, 611–671, <https://doi.org/10.1002/2016GL071328>, 2016a.
- Acworth, R. I., Rau, G. C., Cuthbert, M. O., Jensen, E., and Leggett, K.: Long-term spatio-temporal precipitation variability in arid-zone Australia and implications for groundwater recharge, *Hydrogeology Journal*, 24, 905–921, <https://doi.org/10.1007/s10040-015-1358-7>,
10 2016b.
- Acworth, R. I., Rau, G. C., Halloran, L. J. S., and Timms, W. A.: Vertical groundwater storage properties and changes in confinement determined using hydraulic head response to atmospheric tides, *Water Resources Research*, 53, 2983–2997, <https://doi.org/10.1002/2016WR020311>, 2017.
- Atwood, D. and Lamb, B.: Resolution problems with obtaining accurate ground water elevation measurement in a hydrogeologic site investigation, in: *Proceedings, First National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Westerville, OH, pp. 185–193, 1987.
- 15 Bailey, D.: Preface, in: *Practical Radio Engineering and Telemetry for Industry*, edited by Bailey, D., p. xi, Elsevier, Oxford, <https://doi.org/10.1016/B978-075065803-4/50013-5>, 2003.
- Bakuła, M., Oszczak, S., and Pelc-Mieczkowska, R.: Performance of RTK Positioning in Forest Conditions: Case Study, *Journal of Surveying Engineering*, 135, 125–130, [https://doi.org/10.1061/\(ASCE\)0733-9453\(2009\)135:3\(125\)](https://doi.org/10.1061/(ASCE)0733-9453(2009)135:3(125)), 2009.
- 20 BDA: Guidance For The Operation Of Cable Percussion Rigs And Equipment (rev 1.4), Tech. rep., British Drilling Association (BDA), 2017.
- Benjamin, J. and Kaplan, D.: Development of a laser-based water level sensor for fine-scale ecohydrological measurements, in: *2017 IEEE Conference on Technologies for Sustainability (SusTech)*, pp. 1–8, IEEE, <https://doi.org/10.1109/SusTech.2017.8333522>, 2017.
- 25 Beshr, A. A. and Abo Elnaga, I. M.: Investigating the accuracy of digital levels and reflectorless total stations for purposes of geodetic engineering, *Alexandria Engineering Journal*, 50, 399–405, <https://doi.org/10.1016/j.aej.2011.12.004>, 2011.
- Bitelli, G., Roncari, G., Tini, M. A., and Vittuari, L.: High-precision topographical methodology for determining height differences when crossing impassable areas, *Measurement: Journal of the International Measurement Confederation*, 118, 147–155, <https://doi.org/10.1016/j.measurement.2018.01.013>, 2018.
- 30 Bock, Y. and Melgar, D.: Physical applications of GPS geodesy: A review, *Reports on Progress in Physics*, 79, 106 801, <https://doi.org/10.1088/0034-4885/79/10/106801>, 2016.
- Bouma, J., Maasbommel, M., and Schuurman, I.: *Handboek meten van grondwaterstanden in peilbuizen* (In Dutch. Translated title: Handbook on measurement of groundwater levels in piezometers), Rapport / STOWA;2012-50, STOWA, Amersfoort, 2012.
- Braun, J., Štroner, M., Urban, R., and Dvořáček, F.: Suppression of Systematic Errors of Electronic Distance Meters for Measurement of Short Distances, *Sensors*, 15, 19 264–19 301, <https://doi.org/10.3390/s150819264>, 2015.
- 35 Bredehoeft, J. D.: Response of well-aquifer systems to Earth tides, *Journal of Geophysical Research*, 72, 3075–3087, <https://doi.org/10.1029/JZ072i012p03075>, 1967.

- Brinker, R. C.: *The Surveying Handbook*, Springer US, Boston, MA, <https://doi.org/10.1097/BCR.0b013e3182a2ad17>, 1995.
- Buchroithner, M. F. and Pfahlbusch, R.: Geodetic grids in authoritative maps—new findings about the origin of the UTM Grid, <https://doi.org/10.1080/15230406.2015.1128851>, 2017.
- Bulant, P., Eisner, L., Pšenčík, I., and Calvez, J. L.: Importance of borehole deviation surveys for monitoring of hydraulic fracturing treatments, *Geophysical Prospecting*, 55, 891–899, <https://doi.org/10.1111/j.1365-2478.2007.00654.x>, 2007.
- 5 Butler, J. J., Jin, W., Mohammed, G. A., and Reboulet, E. C.: New insights from well responses to fluctuations in barometric pressure, *Ground Water*, 49, 525–533, <https://doi.org/10.1111/j.1745-6584.2010.00768.x>, 2011.
- Cain, S. F., Davis, G. A., Loheide, S. P., and Butler, J. J.: Noise in Pressure Transducer Readings Produced by Variations in Solar Radiation, *Ground Water*, 42, 939–944, <https://doi.org/10.1111/j.1745-6584.2004.t01-12-.x>, 2004.
- 10 Ceylan, A. and Baykal, O.: Precise Height Determination Using Leap-Frog Trigonometric Leveling, *Journal of Surveying Engineering*, 132, 118–123, [https://doi.org/10.1061/\(ASCE\)0733-9453\(2006\)132:3\(118\)](https://doi.org/10.1061/(ASCE)0733-9453(2006)132:3(118)), 2006.
- Clark, W. E.: Computing the barometric efficiency of a well, *Journal of the Hydraulics Division*, 93, 93–98, 1967.
- Contreras, I. A., Grosser, A. T., and Ver Strate, R. H.: The Use of the Fully-Grouted Method for Piezometer Installation, in: 7th FMGM 2007, pp. 1–20, American Society of Civil Engineers, Reston, VA, [https://doi.org/10.1061/40940\(307\)67](https://doi.org/10.1061/40940(307)67), 2007.
- 15 Cuevas, J., Calvo, M., Little, C., Pino, M., and Dassori, P.: Are diurnal fluctuations in streamflow real?, *Journal of Hydrology and Hydromechanics*, 58, 149–162, <https://doi.org/10.2478/v10098-010-0014-0>, 2010.
- Cunningham, W. L. and Schalk, C. W.: *Groundwater Technical Procedures of the U . S . Geological Survey Techniques and Methods 1 – A1*, Tech. rep., U.S. Geological Survey, 2016.
- Cuthbert, M., Greswell, R., and Mackay, R.: A Wet/Wet Differential Pressure Sensor for Measuring Vertical Hydraulic Gradient, *Ground*
- 20 *Water*, 49, 781–782, <https://doi.org/10.1111/j.1745-6584.2011.00789.x>, 2011.
- Devlin, J. F. and McElwee, C. D.: Effects of Measurement Error on Horizontal Hydraulic Gradient Estimates, *Ground Water*, 45, 62–73, <https://doi.org/10.1111/j.1745-6584.2006.00249.x>, 2007.
- Domenico, P. A. and Schwartz, F. W.: *Physical and Chemical Hydrogeology*, John Wiley & Sons, Inc., 2nd edn., 1997.
- Drexler, J. Z., Bedford, B. L., Scognamiglio, R., and Siegel, D. I.: Fine-scale characteristics of groundwater flow in a peatland, *Hydrological*
- 25 *Processes*, 13, 1341–1359, [https://doi.org/10.1002/\(SICI\)1099-1085\(19990630\)13:9<1341::AID-HYP810>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-1085(19990630)13:9<1341::AID-HYP810>3.0.CO;2-5), 1999.
- Dunnicliff, J. and Green, G. E.: *Geotechnical Instrumentation for Monitoring Field Performance*, A Wiley-Interscience publication, Wiley, 1993.
- El-Ashmawy, K. L. A.: Accuracy, time cost and terrain independence comparisons of levelling techniques, *Geodesy and Cartography*, 40, 133–141, <https://doi.org/10.3846/20296991.2014.962727>, 2014.
- 30 Elçi, A., Flach, G. P., and Molz, F. J.: Detrimental effects of natural vertical head gradients on chemical and water level measurements in observation wells: identification and control, *Journal of Hydrology*, 281, 70–81, [https://doi.org/https://doi.org/10.1016/S0022-1694\(03\)00201-4](https://doi.org/https://doi.org/10.1016/S0022-1694(03)00201-4), 2003.
- Freeman, L. A., Carpenter, M. C., Rosenberry, D. O., Rousseau, J. P., Unger, R., and McLean, J. S.: *Use of Submersible Pressure Transducers in Water-Resources Investigations*, Tech. rep., U.S. Geological Survey, Reston, Virginia, USA, 2004.
- 35 Freeze, R. A. and Cherry, J. A.: *Groundwater*, Prentice Hall, Inc., Upper Saddle River, NJ 07458, 1979.
- Garrido, M. S., Giménez, E., de Lacy, M. C., and Gil, A. J.: Surveying at the limits of local RTK networks: Test results from the perspective of high accuracy users, *International Journal of Applied Earth Observation and Geoinformation*, 13, 256–264, <https://doi.org/10.1016/j.jag.2010.12.001>, 2011.

- Gonthier, G.: A Graphical Method for Estimation of Barometric Efficiency from Continuous Data - Concepts and Application to a Site in the Piedmont, Air Force Plant 6, Marietta, Georgia, Tech. rep., US Geological Survey, <https://doi.org/10.3133/sir20075111>, 2003.
- Gribovszki, Z., Kalicz, P., and Szilágyi, J.: Does the accuracy of fine-scale water level measurements by vented pressure transducers permit for diurnal evapotranspiration estimation?, *Journal of Hydrology*, 488, 166–169, <https://doi.org/10.1016/J.JHYDROL.2013.03.001>, 2013.
- 5 Hanschke, T. and Baird, A. J.: Time-lag errors associated with the use of simple standpipe piezometers in wetland soils, *Wetlands*, 21, 412–421, [https://doi.org/10.1672/0277-5212\(2001\)021\[0412:tleawt\]2.0.co;2](https://doi.org/10.1672/0277-5212(2001)021[0412:tleawt]2.0.co;2), 2006.
- Hegarty, C. J.: The Global Positioning System (GPS), in: *Springer Handbook of Global Navigation Satellite Systems*, pp. 197–218, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-42928-1_7, 2017.
- Höltling, B. and Coldewey, W. G.: *Hydrogeologie*, Spektrum Akademischer Verlag, Heidelberg, <https://doi.org/10.1007/978-3-8274-2354-2>,
10 2013.
- Hubbell, J. M., Sisson, J. B., Nicholl, M. J., and Taylor, R. G.: Well Design to Reduce Barometric Pressure Effects on Water Level Data in Unconfined Aquifers, *Vadose Zone Journal*, 3, 183–189, <https://doi.org/10.2113/3.1.183>, 2004.
- Hubbert, M. K.: The theory of ground???water motion, *Eos, Transactions American Geophysical Union*, 21, 648–648, <https://doi.org/10.1029/TR021i002p00648-1>, 1940.
- 15 Hvorslev, M.: *Time Lag and Soil Permeability in Ground-water Observations*, Tech. rep., U. S. ARMY CORPS OF ENGINEERS, Vicksburg, Mississippi, USA, 1951.
- Jacob, C. E.: On the flow of water in an elastic artesian aquifer, *Eos, Transactions American Geophysical Union*, 21, 574–586, 1940.
- Kalbus, E., Reinstorf, F., and Schirmer, M.: Measuring methods for groundwater – surface water interactions: a review, *Hydrology and Earth System Sciences*, 10, 873–887, <https://doi.org/10.5194/hess-10-873-2006>, 2006.
- 20 Keys, W. S.: *A Practical Guide to Borehole Geophysics in Environmental Investigations*, Routledge, <https://doi.org/10.1201/9781315136417>, 2017.
- Kim Sun, G. O. and Gibbins, P.: How well does the virtual reference station (VRS) system of gps base stations perform in comparison to conventional RTK?, *Journal of Spatial Science*, 50, 59–73, <https://doi.org/10.1080/14498596.2005.9635038>, 2005.
- Knotters, M., Meij, T. v. d., and Pleijter, M.: *Nauwkeurigheid van handmatig gemeten grondwaterstanden en stijghoogtes : verslag van een veldexperiment*, Tech. rep., 2013.
- 25 Kouba, J., Lahaye, F., and Tétreault, P.: Precise Point Positioning, in: *Springer Handbook of Global Navigation Satellite Systems*, chap. E(25), pp. 724–751, Springer International Publishing, https://doi.org/10.1007/978-3-319-42928-1_7, 2017.
- Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., and Schuh, H.: Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo, *Journal of Geodesy*, 89, 607–635, <https://doi.org/10.1007/s00190-015-0802-8>, 2015.
- 30 Liu, Z. and Higgins, C. W.: Does temperature affect the accuracy of vented pressure transducer in fine-scale water level measurement?, *Geoscientific Instrumentation, Methods and Data Systems*, 4, 65–73, <https://doi.org/10.5194/gi-4-65-2015>, 2015.
- Luszczynski, N. J.: Head and flow of ground water of variable density, *Journal of Geophysical Research*, 66, 4247–4256, <https://doi.org/10.1029/JZ066i012p04247>, 1961.
- Mäkinen, R. and Orvomaa, M.: Experiences and recommendations on automated groundwater monitoring, in: *20th International Northern Research Basins Symposium and Workshop – Kuusamo, Finland August 16–21, 2015*, edited by Korhonen, J. and Kuusisto, E., pp. 71–75, Finnish Environment Institute (SYKE), Kuusamo, Finland, 2015.

- McCallum, A. M., Andersen, M. S., Rau, G. C., Larsen, J. R., and Acworth, R. I.: River-aquifer interactions in a semiarid environment investigated using point and reach measurements, *Water Resources Research*, 50, 2815–2829, <https://doi.org/10.1002/2012WR012922>, 2014.
- McKenna, G. T.: Grouted-in installation of piezometers in boreholes, *Canadian Geotechnical Journal*, 32, 355–363, <https://doi.org/10.1139/t95-035>, 1995.
- McLaughlin, D. L. and Cohen, M. J.: Thermal artifacts in measurements of fine-scale water level variation, *Water Resources Research*, 47, <https://doi.org/10.1029/2010WR010288>, 2011.
- McMillan, T. C., Rau, G. C., Timms, W. A., and Andersen, M. S.: Utilizing the Impact of Earth and Atmospheric Tides on Groundwater Systems: A Review Reveals the Future Potential, *Reviews of Geophysics*, p. 2018RG000630, <https://doi.org/10.1029/2018RG000630>, 2019.
- Meinzer, O. E.: Ground water in the United States, a summary of ground-water conditions and resources, utilization of water from wells and springs, methods of scientific investigation, and literature relating to the subject, Tech. rep., U.S. G.P.O., <https://doi.org/10.3133/wsp836D>, 1939.
- Mikkelsen, P. and Green, G.: Piezometers in Fully-Grouted Boreholes, *International Symposium on Geomechanics*, Oslo, Norway, 2003.
- Misra, P. and Enge, P.: *Global Positioning System: Signals, Measurements, and Performance (Revised Second Edition)*, Ganga-Jamuna Press, 2010.
- Morgenschweis, G.: Messung des Wasserstands (in German. Translated title: Measurement of water level), in: *Hydrometrie (in German. Translated title: Hydrometry)*, edited by Morgenschweis, G., pp. 25–114, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-55314-5_3, 2018.
- Nielsen, D. M. and Nielsen, G.: *The Essential Handbook of Ground-Water Sampling*, Cambridge University Press, Cambridge, 2006.
- Nilsson, G. and Nissen, J.: Revision of borehole deviation measurements in Forsmark. Forsmark site investigation., Tech. rep., Swedish Nuclear Fuel and Waste Management Company, Stockholm, 2007.
- Noorduyn, S. L., Cook, P. G., Wood, C., and White, N.: Using Sealed Wells to Measure Water Levels Beneath Streams and Floodplains, *Groundwater*, 53, 872–876, <https://doi.org/10.1111/gwat.12287>, 2015.
- NUDLC: Minimum Construction Requirements for Water Bores in Australia, Tech. rep., National Uniform Drillers Licensing Committee, Australian Government National Water Commission, 2011.
- Ohmer, M., Liesch, T., Goepfert, N., and Goldscheider, N.: On the optimal selection of interpolation methods for groundwater contouring: An example of propagation of uncertainty regarding inter-aquifer exchange, *Advances in Water Resources*, 109, 121–132, <https://doi.org/10.1016/j.advwatres.2017.08.016>, 2017.
- Plazak, D.: Differences Between Water- Level Probes, *Groundwater Monitoring & Remediation*, 14, 84, <https://doi.org/10.1111/j.1745-6592.1994.tb00091.x>, 1994.
- Post, V. E. A. and von Asmuth, J. R.: Review: Hydraulic head measurements—new technologies, classic pitfalls, *Hydrogeology Journal*, 21, 737–750, <https://doi.org/10.1007/s10040-013-0969-0>, 2013.
- Post, V. E. A., Banks, E., and Brunke, M.: Groundwater flow in the transition zone between freshwater and saltwater: a field-based study and analysis of measurement errors, *Hydrogeology Journal*, 26, 1821–1838, <https://doi.org/10.1007/s10040-018-1725-2>, 2018.
- Rasmussen, T. C. and Crawford, L. A.: Identifying and Removing Barometric Pressure Effects in Confined and Unconfined Aquifers, *Ground Water*, 35, 502–511, <https://doi.org/10.1111/j.1745-6584.1997.tb00111.x>, 1997.

- Rau, G. C., Acworth, R. I., Halloran, L. J. S., Timms, W. A., and Cuthbert, M. O.: Quantifying Compressible Groundwater Storage by Combining Cross-Hole Seismic Surveys and Head Response to Atmospheric Tides, *Journal of Geophysical Research: Earth Surface*, 123, 1910–1930, <https://doi.org/10.1029/2018JF004660>, 2018.
- Remondi, B. W.: Performing Centimeter-Level Surveys in Seconds with GPS Carrier Phase: Initial Results, *Navigation*, 32, 386–400, <https://doi.org/10.1002/j.2161-4296.1985.tb00918.x>, 1985.
- Ritzema, H. P., Heuvelink, G. B. M., Heinen, M., Bogaart, P. W., Bolt, F. J. E. v. d., Hack-ten Broeke, M. J. D., Hoogland, T., Knotters, M., Massop, H. T. L., and Vroon, H. R. J.: Meten en interpreteren van grondwaterstanden : analyse van methodieken en nauwkeurigheid, *Tech. rep.*, 309, ,, 2012.
- Rizos, C.: Surveying, in: *Springer Handbook of Global Navigation Satellite Systems*, chap. F(35), pp. 1011–1038, Springer International Publishing, https://doi.org/10.1007/978-3-319-42928-1_7, 2017.
- Rosenberry, D. O.: Effect of Sensor Error on Interpretation of Long-Term Water-Level Data, *Ground Water*, 28, 927–936, <https://doi.org/10.1111/j.1745-6584.1990.tb01729.x>, 1990.
- Rosenberry, D. O., LaBaugh, J. W., and Hunt, R. J.: Use of Monitoring Wells, Portable Piezometers, and Seepage Meters to Quantify Flow Between Surface Water and Ground Water, in: *Field Techniques for Estimating Water Fluxes Between Surface Water and Ground Water*, Chapter 2 of *Field Techniques for Estimating Water Fluxes Between Surface Water and Ground Water*, pp. 39–70, 2008.
- Rowe, R. K. and Nadarajah, P.: Evaluation of the hydraulic conductivity of aquitards, *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 31, 139, [https://doi.org/10.1016/0148-9062\(94\)90614-9](https://doi.org/10.1016/0148-9062(94)90614-9), 1994.
- Saines, M.: ERRORS IN INTERPRETATION OF GROUND-WATER LEVEL DATA, *Groundwater Monitoring & Remediation*, 1, 56–61, <https://doi.org/10.1111/j.1745-6592.1981.tb00798.x>, 1981.
- Shanafield, M. and Cook, P. G.: Transmission losses, infiltration and groundwater recharge through ephemeral and intermittent streambeds: A review of applied methods, *Journal of Hydrology*, 511, 518–529, <https://doi.org/10.1016/j.jhydrol.2014.01.068>, 2014.
- Siejka, Z.: Validation of the Accuracy and Convergence Time of Real Time Kinematic Results Using a Single Galileo Navigation System, *Sensors*, 18, 2412, <https://doi.org/10.3390/s18082412>, 2018.
- Silliman, S. E. and Mantz, G.: The Effect of Measurement Error on Estimating the Hydraulic Gradient in Three Dimensions, *Ground Water*, 38, 114–120, <https://doi.org/10.1111/j.1745-6584.2000.tb00208.x>, 2000.
- Simeoni, L.: Laboratory tests for measuring the time-lag of fully grouted piezometers, *Journal of Hydrology*, 438–439, 215–222, <https://doi.org/10.1016/j.jhydrol.2012.03.025>, 2012.
- Sokol, D.: Position and fluctuations of water level in wells perforated in more than one aquifer, *Journal of Geophysical Research*, 68, 1079–1080, <https://doi.org/10.1029/JZ068i004p01079>, 1963.
- Sorensen, J. P. and Butcher, A. S.: Water Level Monitoring Pressure Transducers-A Need for Industry-Wide Standards, *Ground Water Monitoring & Remediation*, 31, 56–62, <https://doi.org/10.1111/j.1745-6592.2011.01346.x>, 2011.
- Spane, F. A.: Considering barometric pressure in groundwater flow investigations, *Water Resources Research*, 38, 14–1, <https://doi.org/10.1029/2001WR000701>, 2002.
- Spane, F. J. and Mercer, R.: HEADCO: a program for converting observed water levels and pressure measurements to formation pressure and standard hydraulic head, *Tech. rep.*, Rockwell International Corporation, Richland, WA, USA, 1985.
- STS Sensors: Genauigkeitsangaben bei Drucksensoren richtig deuten (in German. Translated title: Correct interpretation of accuracy specifications of pressure transducers): <https://campaign.stssensors.com/blog/de/druckmesstechnik-genauigkeitsangaben-drucksensor>, 2017.

- Sweet, H., Rosenthal, G., and Atwood, D.: Water level monitoring - Achievable accuracy and precision, in: *Ground Water and Vadose Zone Monitoring*, edited by Nielsen, D. and Johnson, A., ASTM STP 1053, pp. 178–192, ASTM, 1990.
- Toll, N. J. and Rasmussen, T. C.: Removal of Barometric Pressure Effects and Earth Tides from Observed Water Levels, *Ground Water*, 45, 101–105, <https://doi.org/10.1111/j.1745-6584.2006.00254.x>, 2007.
- 5 Treskatis, C.: *Bewirtschaftung von Grundwasserressourcen: Planung, Bau und Betrieb von Grundwasserfassungen*, 2006.
- Twining, B. V. B.: Borehole deviation and correction factor data for selected wells in the eastern Snake River Plain aquifer at and near the Idaho National Laboratory, Idaho, Tech. rep., U.S. Geological Survey, <https://doi.org/10.3133/sir20165163>, 2016.
- van der Kamp, G. and Gale, J. E.: Theory of earth tide and barometric effects in porous formations with compressible grains, *Water Resources Research*, 19, 538–544, <https://doi.org/10.1029/WR019i002p00538>, 1983.
- 10 Walker, J. and Awange, J. L.: Total Station Differential Levelling, in: *Surveying for Civil and Mine Engineers*, pp. 93–101, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-53129-8_6, 2018.
- Weeks, E. P.: Barometric fluctuations in wells tapping deep unconfined aquifers, *Water Resources Research*, 15, 1167–1176, <https://doi.org/10.1029/WR015i005p01167>, 1979.
- Willemsen, J.: Basismetnet grondwater (in Dutch, translated title Primary monitoring network groundwater), Tech. rep., Waterschap Hol-
- 15 landse Delta, 2006.
- Yamano, M., Goto, S., Miyakoshi, A., Hamamoto, H., Lubis, R. F., Monyrath, V., and Taniguchi, M.: Reconstruction of the thermal environment evolution in urban areas from underground temperature distribution, *Science of the Total Environment*, 407, 3120–3128, <https://doi.org/10.1016/j.scitotenv.2008.11.019>, 2009.
- Zandbergen, P. A. and Barbeau, S. J.: Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones, *Journal of*
- 20 *Navigation*, 64, 381–399, <https://doi.org/10.1017/S0373463311000051>, 2011.
- Zarriello, P. J.: Accuracy, Precision, and Stability of a Vibrating-Wire Transducer Measurement System to Measure Hydraulic Head, *Groundwater Monitoring & Remediation*, 15, 157–168, <https://doi.org/10.1111/j.1745-6592.1995.tb00527.x>, 1995.