Received: 6 September 2021 | Accepted: 16 December 2021
DOI: 10.1002/hyp.14457

RESEARCH ARTICLE

WILEY

Magnitude, consequences and correction of temperature-derived errors for absolute pressure transducers under common monitoring scenarios

Joseph Shannon[1] | Fengjing Liu[1] | Matthew Van Grinsven[1,2] | Randall Kolka[3] | Thomas Pypker[4]

[1]College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan, USA

[2]Department of Earth, Environmental and Geographical Sciences, Northern Michigan University, Marquette, Michigan, USA

[3]Northern Research Station, USDA Forest Service, Grand Rapids, Minnesota, USA

[4]Department of Natural Resource Sciences, Thompson Rivers University, Kamloops, British Columbia, Canada

Correspondence
Joseph Shannon, College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan 4993, USA.
Email: jpshanno@mtu.edu

Funding information
U.S. Forest Service, Grant/Award Number: 17-JV-11 242307 -133; Graduate School, Michigan Technological University

Abstract

Continuous water level monitoring using absolute pressure transducers with onboard data logging is common practice in hydrologic studies. While there has been some discussion and study of temperature-derived error (TDE), there has not been a systematic evaluation of the problem. We sought to answer three questions: (1) are current best practices enough to avoid these errors, (2) can laboratory correction be used to correct field data from varying conditions, and (3) what is the scale of the additional uncertainty of the correction procedure? We evaluated the magnitude of such errors under laboratory conditions that mimicked common monitoring scenarios. Using field data, we also demonstrated the impact of TDEs on calculated daily mean water level and diurnal signal decomposition to estimate evapotranspiration (ET). To address instrument and model uncertainty, we fit 1000 possible correction models using a double-bootstrap approach. Correction models fit expected error as a function of water and air temperature and rate of change of air temperature. TDEs were a significant source of error, resulting in recorded data outside of manufacturer-stated instrument uncertainty, with 45% of bootstrap models showing significant but small TDEs under best-practice deployment. Correction equations did introduce additional error, often on a much smaller scale than instrument uncertainty. When tested against a validation data set, correction equations effectively reduced total measurement uncertainty below instrument uncertainty by up to 65%. The effects of TDEs on case-study field data resulted in 56% of daily mean values outside of instrument error bounds (errors: −1.5 to 4.2 cm). Our results suggest that a single laboratory correction equation can be used across monitoring scenarios, though we suggest matching deployment conditions as closely as practical. Identification and correction of TDEs is essential to avoid erroneous conclusions, downstream analyses and water resources management.

KEYWORDS
groundwater monitoring, measurement uncertainty, pressure transducer, stage, streamflow measurement, water level correction

1 | INTRODUCTION

Hydrologic monitoring for research and management, including monitoring wells, piezometers and stream gauges, often utilizes submerged pressure transducers to measure water levels. These instruments are easy to deploy, require minimum access and provide long-term continuous records in an easily analysed format via on-board data logging. They have been developed in two primary configurations: absolute (unvented) and gauge (vented). Absolute transducers measure pressure relative to a fixed pressure chamber on one side of a flexible membrane. Gauge transducers measure pressure across a similar membrane with one chamber connected to the atmosphere via a vent tube. The vent tube in a gauge transducer directly compensates for atmospheric pressure atop the water column. Absolute transducers require a secondary reading of atmospheric pressure for a post-monitoring correction. While the approach for both pressure transducers is simply stated, previous work has concluded that "measurement of the water level…is not simple" (Liu & Higgins, 2015, p. 72). Earlier guidance documents for monitoring water levels with pressure transducers warned that "these systems commonly are not adequately supported by quality-assurance procedures" (Freeman et al., 2004, p. 1). Freeman et al. (2004) go on to recognize the effects of temperature on pressure transducer function from theoretical and practical standpoints. Even in a closed tank, the relatively straightforward task of measuring water level may be confounded by propagating instrument uncertainty through each step required to arrive at final water levels (Tamari & Aguilar-Chávez, 2010).

Apart from manufacturer-stated instrument error, additional uncertainty caused by air and water temperatures, and the rates of temperature change have been theorized and observed in water level measurements. Freeman et al. (2004) highlighted water temperature gradients and rapid temperature fluctuations as potential sources of error, while Cain III et al. (2004) described air temperature-derived errors (TDEs) within gauge transducers. Within absolute transducers, errors were shown to arise from the temperature of the instrument (Moore et al., 2016), as well as from differing thermal regimes between absolute barometric and water pressure transducers (Cuevas et al., 2010; McLaughlin & Cohen, 2011). Liu and Higgins (2015) and Moore et al. (2016) went on to suggest that all transducers be tested for inaccuracies over the range of temperature measurements and that test individual correction equations be developed to apply to collected data. That conclusion is supported by others who have found that air temperature-induced errors can be as high as 19% in streamflow monitoring (Cuevas et al., 2010) or 1.5 cm day^{-1} in inundated wetland stages (McLaughlin & Cohen, 2011).

To pre-empt these errors for absolute transducers, it is recommended as 'best practice' deployment of barometric and water level pressure transducers to place both in similar thermal regimes (i.e. within a dry well at the same depth) (McLaughlin & Cohen, 2011). However, this type of deployment is not always possible due to cost or setting. In other cases, previous deployments may have not recognized the potential magnitude of error, but still represent important data sources. We would expect that barometric pressure transducer deployments and monitoring objectives could influence the magnitude of error, which would be controlled by the thermal range of both the water and air pressure monitoring. Monitoring objectives include stable temperature regimes in groundwater or more variable-temperature regimes in surface water, and instrument deployment can lead to air temperature regimes that are (i) variable and different from the water temperature regime, (ii) stable and also different from the water temperature regime, or (iii) stable and similar to the water temperature regimes. To date, studies on these types of errors have been focused on individual deployment scenarios without an attempt to compare the magnitude of errors under various common monitoring scenarios. In this study, we defined four such scenarios which are a cross of two factors: thermal setting and deployment practice. We grouped thermal setting into stable (groundwater) and variable (surface water) conditions. Deployment practice was considered as 'best practice' or 'temperature-difference-biased' (transducers deployed without regard to similar thermal settings). The four scenarios were then best-practice groundwater (GW_{best}), temperature-biased groundwater (GW_{bias}), best-practice surface water (SW_{best}), and temperature-biased surface water (SW_{bias}). Specifically, and to compare with best-practice deployment scenarios where temperature regimes of the air and water transducers are synchronized, the equipment deployment for the GW_{bias} and SW_{bias} scenarios were set up where the thermal regime of the barometric transducer was dissimilar than the thermal regime of the transducer deployed in the groundwater and surface water scenarios, respectively. For the SW_{bias} scenario the absolute pressure transducers deployed in air and water were both exposed to variable temperature regimes, where the variable air and water regimes were not synchronous. For the GW_{bias} scenario, the absolute pressure transducers deployed in air and water were both exposed to stable temperature regimes with slight asynchronous variation. Without a systematic comparison between these scenarios, it is difficult to determine the primary source of error, the potential magnitude of error, and the appropriate mitigation or correction. A systematic evaluation can also show if 'best practice' is enough to eliminate temperature-induced errors.

Based on the above theoretical and experimental results, the potential sources of error for absolute transducers are water temperature, air temperature, and the rate of change for both water and air temperatures. Liu and Higgins (2015) examined the influence of water temperature variations on vented pressure transducer observations for sensors deployed in surface water or shallow groundwater locations and demonstrated that rapid changes in water temperature introduce additional sensor error beyond the reported instrument resolution. Apparent error introduced by temperature differential is likely the response of individual transducers to out-of-phase variation in air and water temperatures. While a vent tube that directly links air and water temperature can be subject to temperature differentials along the length of the tube (Liu & Higgins, 2015), there is no analogous physical process for absolute transducers and for this reason correction of temperature-induced pressure measurement errors are not directly comparable between vented and absolute pressure sensor systems.

Asynchronous variation in air and water temperatures can result in a stronger signal that masks the effect of a weaker signal, obscuring the amplification of manufacturer-stated errors. The sequential correction shown in Figure 1 demonstrates water level measurement error based on air temperature (Figure 1a), water temperature (Figure 1b), and water temperature after removing the air temperature error (Figure 1c). There are two possible approaches to the final correction. One is to develop and apply univariate corrections to a reference pressure for each transducer. The second is to develop a single multivariate correction to a fixed water depth for each pair of air and water transducers. The single multivariate correction presents a simpler alternative because it can be performed based only on a known water level without requiring a pressure chamber or other suitable reference pressure.

Implied in the consideration of temperature bias above is that there is additional, quantifiable error that should be considered in the estimate of total measurement uncertainty. Without accounting for total uncertainty, an analysis built on water levels may lead to skewed results as the initial inaccuracies compound. For example, many studies have relied on the diurnal signal in well water levels and streamflow to estimate evapotranspiration and groundwater flow, but temperature-derived errors could also drive diurnal fluctuation (Kirchner, 2009; Loheide II et al., 2005; Mazur et al., 2014; McLaughlin & Cohen, 2014; Watras et al., 2017; White, 1932). Some researchers may recognize the potential for errors and propose using a rolling average or a daily mean to smooth out and cancel the errors. However, if there is a consistent bias in any source of error (e.g., Figure 1a), the result will be a systematic error that cannot be removed using a smoothing approach developed for random errors. Therefore, quantifying and correcting temperature-induced errors is essential to ensure that the observed data trends are real and not entirely or partially erroneous. While there are some attempts to provide that validation through high-frequency manual readings (Cuevas et al., 2010; Gribovszki et al., 2013), it is often impractical or impossible to capture the range of potential temperature and water level values, leaving laboratory-derived correction equations as the best alternative, as recommended by Moore et al. (2016) and Liu and Higgins (2015).

To our knowledge, no previous study has quantified the separate impacts of air and water TDEs to assess the influence of out-of-phase air and water temperatures for pressures recorded by absolute pressure transducers. Further, no previous study has used the amplification of water level measurement errors for the purpose of generating a broadly applicable compensation procedure for common water level monitoring scenarios. The results of this study are intended to support most common surface and groundwater monitoring applications that use absolute pressure transducers by providing a broadly applicable correction method to improve performance across wide range of environmental conditions caused by temperature-induced errors when water and air pressure measurements are required to monitor water levels.

In this study, we present a lab experiment to isolate and correct the separate impacts of transducer responses to water and air temperatures using absolute transducers. Our experiment is designed as a systematic comparison of errors under the four scenarios described above (GW_{best}, GW_{bias}, SW_{best}, SW_{bias}). We compare the corrected and uncorrected water levels on a test set of laboratory data as well as a case study using field data from an ongoing hydrologic project. The comparison between corrected and uncorrected laboratory and field data highlights the impacts on derived analyses and future conclusions. Finally, we calculate the propagated uncertainty that results from combining the separate impacts of transducer instrument responses to water and air temperatures with the uncertainty derived from the modelled correction procedure for common absolute transducer deployment scenarios and consider whether temperature correction can be used to increase measurement precision beyond the manufacturer-stated error.

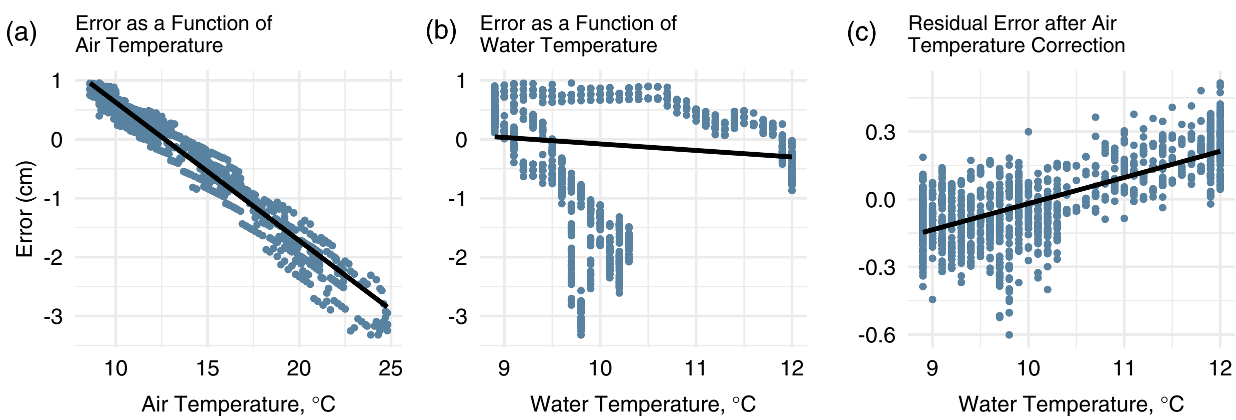

FIGURE 1 Error is driven by a combination of individual transducer errors, which can result in masking by the stronger trend. Panel (a) shows water level error as a function of air temperature, illustrating a strong linear trend. Panel (b) shows a dichotomous relationship between water level error and water temperature relative to the expected linear relationship. Panel (c) is the relationship between error and water temperature after the influence of air temperature is removed via OLS linear regression. Data taken from a barometric pressure transducer 1066019 and water pressure transducer 1062452 during the $SW_{bias\text{-}2}$ period

2 | DATA AND METHODS

2.1 | Data collection and preparation

We conducted a 5-day experiment recording constant water levels using absolute pressure transducers in four deployments, which represent a fully crossed design with stable and variable temperatures and similar and dissimilar air and water thermal regimes (Table 1). A replication of SW_{bias} condition ($SW_{bias\text{-}1}$ and $SW_{bias\text{-}2}$) was collected as a training set for developing new measurement uncertainty bounds (see details below). Eighteen pressure transducers of varying ages and models, all manufactured by Solinst, Ltd. (Georgetown, ON, Canada), were used to record water levels (Table 2). Two transducers, both Solinst Levellogger Juniors, were used to monitor barometric pressure. All instruments contained an absolute pressure transducer, temperature sensor and on-board data logging. These transducers also provide internal temperature compensation of absolute pressure measurements.

Water level transducers were suspended from a brace across a large container with the transducers' zero points aligned (Figure 2). The transducers were covered with a fixed depth of water and a layer of canola oil was added to prevent evaporation over the length of the experiment. Water level was measured to the millimetre at the start and end of the experimental period to assure no evaporation had occurred. To create the best-practice condition of thermally similar air and water environments, we placed a weighted, open-topped cylinder in the centre of the container. During the two periods with synchronous (similar) thermal regimes, the two barometric transducers were suspended in the cylinder below the level of the water surface. The entire apparatus was covered loosely with aluminium foil to prevent direct wind effects.

Best-practices and temperature-differential bias experiments were each performed twice under conditions designed to mimic common measurement targets. For stable thermal regimes, the experimental apparatus was placed in a non-climate-controlled underground room to minimize temperature variations. Varying thermal regime periods were conducted outdoors, out of direct solar radiation. For both SW_{bias} periods, the barometric transducers were placed under a black radiation shield to magnify temperature differentials and purposely induce errors from rapid temperature fluctuations such as those observed by Liu and Higgins (2015). In contrast, for the SW_{best} periods the barometric transducers were placed in the same variable-temperature location and were suspended within an inner open-top cylinder. For the GW_{bias} periods the barometric transducers were placed in the same stable-temperature location but not suspended within the inner cylinder, whereas for the GW_{best} periods the barometric transducers were placed in the same stable-temperature location

TABLE 1 Description of experimental conditions

Scenario	Monitoring analog	Deployment description	Temperature	Air/water temperature regime
GW_{best}	Groundwater	Best practice	Stable	Similar
GW_{bias}	Groundwater	Temperature-differential biased	Stable	Dissimilar
SW_{best}	Surface water	Best practice	Variable	Similar
$SW_{bias\text{-}1}$, $SW_{bias\text{-}2}$[a]	Surface water	Temperature-differential biased	Variable	Dissimilar

Note: Best practice describes air and water pressure transducers deployed under similar thermal regimes. All experimental periods were conducted over five consecutive days. Temperature-differential biased indicates that barometric and water pressure transducers were not deployed under similar thermal regimes.
[a]SW_{bias} conditions were repeated, and $SW_{bias\text{-}2}$ data were used to generate an independent correction model tested using data from all other periods.

TABLE 2 Summary of pressure transducer model type, measurement resolution and associated errors expressed as *SD*, derived from manufacturer-stated 99% instrument accuracy and manufacturer-stated 0.02%, 0.001% and 0.02% resolution at full scale for the Levelogger junior (LT_Jr) Levelogger Edge (LT_EDGE), and Levelogger junior Edge (LT_EDGE_JR) models, respectively

Model (range)	Number of loggers	Pressure resolution (cm)	Pressure error (cm)	Temperature resolution (°C)	Temperature error (°C)
LT_Jr (5 m)	8	1.4×10^{-1}	2.15×10^{-1}	1.0×10^{-1}	4.30×10^{-2}
LT_EDGE (2 m)	2	1.2×10^{-4}	4.30×10^{-2}	3.0×10^{-2}	2.15×10^{-2}
LT_EDGE_JR (5 m)	10	1.4×10^{-1}	2.15×10^{-1}	1.0×10^{-1}	4.30×10^{-2}
LT_Jr		**LT_EDGE**		**LT_EDGE_JR**	
1033239	1062534	2013939	2025928	2030899	2064739
1062452	1 065861			2059683	2064745
1062520	1066016			2064734	2069158
1062528	1066019			2064737	2100561
				2064738	2104452

Note: The lower portion of the table lists each transducer serial number under its instrument model.

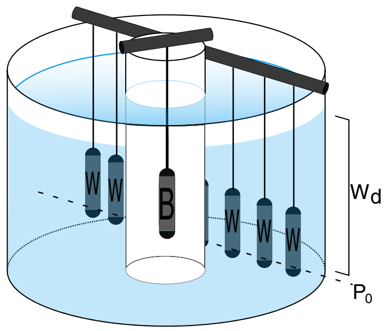

FIGURE 2 Diagram showing arrangement of pressure transducers within experimental apparatus (reduced number shown for clarity), where (b) represents a barometric transducer and W represents a water transducer. P_0 marks the water pressure transducer zero point and W_d denotes the water depth over the zero point of the transducers. During the study, two barometric transducers were deployed in the central tube in best-practice study scenarios. During all study periods, 18 water transducers were suspended from two cross braces

and were suspended within an inner open-top cylinder. Correction models (see below) were developed individually for each experimental period. To test correction model performance against observation data collected under different conditions, the final model coefficients developed from data collected during the $SW_{bias\text{-}2}$ experimental period were used to predict corrections for all other experimental periods. This approach allows for evaluating the common scenario of a lab-derived correction procedure being applied to monitoring situations that differ from lab conditions.

Transducers were deployed and data retrieved using Solinst Levellogger v.4.4.0 software. Transducers were set to record at 1-min intervals with time synchronized to computer time at launch. To avoid disrupting the experimental apparatus, the five experimental periods were run serially without retrieving data from the transducers. Experimental conditions were manipulated by relocating the barometric transducers for the bias and best practice scenarios, or the entire experimental apparatus for the groundwater and surface water scenarios.

2.2 | Water level correction

All of the sensors used in this study are designed to measure pressure at a relatively constant temperature and utilize a constant water density of 1000 kg m^{-3} to convert pressure to depth (Personal Communication, Solinst). To account for the larger range of water temperatures observed in this study, the recorded pressure was adjusted to account for the variable density of water (Equation 1).

$$P_c = \frac{P * \rho_o}{\rho_T} \tag{1}$$

where ρ_T is the density of water (kg m^{-3}) as a function of water temperature taken from Tanaka et al. (2001, eq. 1), ρ_o is the assumed density of water (1000 kg m^{-3}), and P and P_c are the observed and density-corrected pressure head readings, respectively in units of depth (cm for this study). Hereinafter, all references to transducer pressure head refer to the density-compensated pressure head.

Actual depth of water (W_d) was measured at a resolution of a millimetre separately for each transducer from the water surface to the transducer zero point (Figure 2). For each water and air transducer pair ($n = 36$), the measured absolute water pressure head (P_w) minus the measured absolute air pressure head (P_a) was used to determine the raw recorded water level ($WL_{raw} = P_w - P_a$). Water level error (ϵ_L) was then calculated as WL_{raw} minus W_d.

A multivariate linear model was fit between the response variable water level error (ϵ_L) and the explanatory variables air temperature (T_a, °C), water temperature (T_w, °C), and the temporal air temperature gradient (ΔT_a, 0.01°C min^{-1}) for each pair of water and air transducers using ordinary least squares (OLS). Due to the small change in air temperature at the minute timestep, the unit of ΔT_a is set as 0.01°C min^{-1} to place it on the same scale as ϵ_L, T_a and T_w. The 36 models have the form (Equation 2)

$$\epsilon_L = \beta_0 + \beta_a T_a + \beta_w T_w + \beta_{\Delta T_a} \Delta T_a \tag{2}$$

where β_0 is the intercept term and β_a, β_w and $\beta_{\Delta Ta}$ represent regression coefficients for respective variables. Figure 1c illustrates how the scale and direction of one source of error (air temperature, Figure 1a) masks the form of another source of error (water temperature, Figure 1b), supporting the use of a single multivariate correction such as Equation 2.

To account for the uncertainty of instrument error in the temperature correction models, Equation 2 was adapted to Equation 3 below and was fit repeatedly using a two-stage parametric bootstrapping approach (Buonaccorsi, 2010). The two-stage bootstrap first draws a sample of the observed data with replacement where a single data point can be included more than one time to create a data set of the same size as the observations. For each sampled coordinate ($\epsilon_L, T_a, T_w, \Delta T_a$), additional random error is added to each observation (Equation 3), drawn from a normal distribution centred on zero with *SD* equal to the known instrument error for each variable (Table 2). This second stage allows the observations to vary within the manufacturer-stated uncertainty. Each perturbation will affect the fit of the correction model. The dual bootstrap procedure was repeated 1000 times for each transducer pair. This procedure generates 1000 sets of potential model coefficients, allowing for 1000 sets of correction predictions that can be used to evaluate the variability of correction models when instrument error is incorporated. The final output is visualized in Figure 3, where we show coefficient estimates of an OLS model without accounting for instrument error and the distribution of 1000 bootstrap-generated coefficients, both from a single transducer pair.

$$\epsilon'_L = \epsilon_L + N(0, \sigma_{L'}); \quad \sigma_{L'} = \sqrt{2\sigma_L^2}$$

$$T'_a = T_a + N(0, \sigma_T)$$

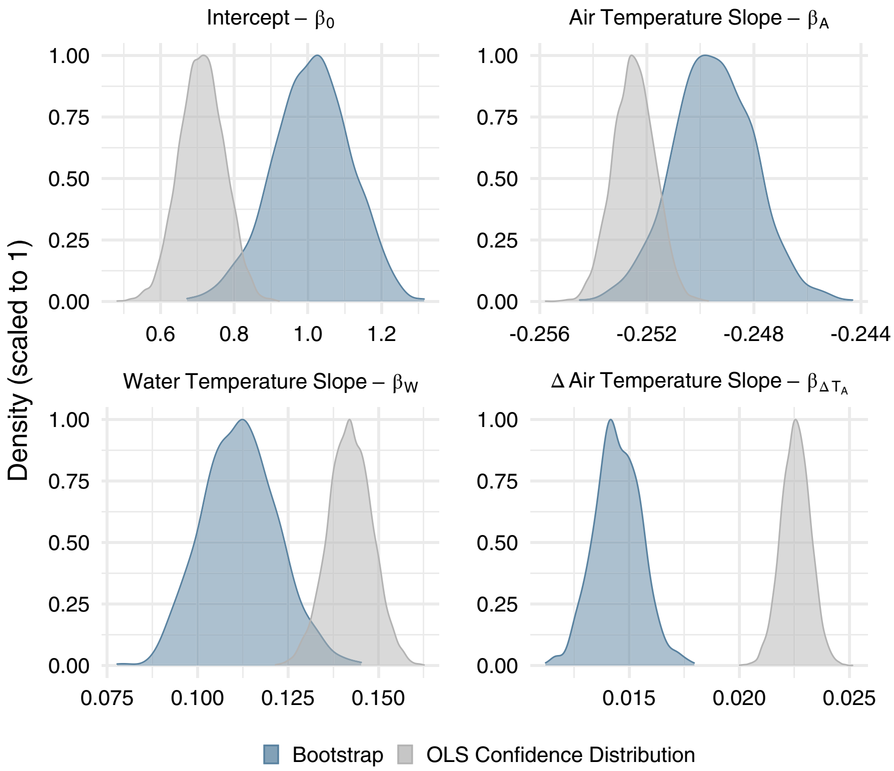

FIGURE 3 A comparison of correction model coefficients for a single transducer pair (chosen at random from the $SW_{bias\text{-}2}$ period, barometric SN: 1066019, water SN: 2069158) as derived from ordinary least squares (OLS) regression and dual bootstrap analysis. The bootstrap density curve represents the distribution of all 1000 bootstrap models. The OLS density curve represents the confidence distribution of the fitted coefficients without accounting for instrument errors [~N(μ, σ) where μ = best estimate of coefficient and σ = *SE* of coefficient estimate]

$$T'_w = T_w + N(0, \sigma_T)$$

$$\Delta T'_a = \Delta T_a + N(0, \sigma_{T'}); \sigma_{T'} = \sqrt{2\sigma_T^2}$$

$$\epsilon'_L = \beta_0 + \beta_a T'_a + \beta_w T'_w + \beta_{\Delta T_a} \Delta T'_a \tag{3}$$

where T_a', T_w', $\Delta T_a'$ denote bootstrap samples and instrument error. The bootstrap procedure was carried out independently within each of the five experiments. A backward elimination stepwise approach was used for each model fit using the bootstrap procedure. This approach removes any slope parameters not significant at $\alpha = 0.05$ and iteratively refits until only intercept and any significant predictors remain. To determine the full range of error (sum of correction error and instrument error), final predicted error values ($\hat{\epsilon}_L$) were drawn from a normal distribution centered on the expected value of ϵ'_L (Equation 3) with a *SD* equivalent to the residual *SE* of the respective model fit within the bootstrap iteration. This procedure resulted in 1000 estimates of ϵ'_L (Equation 3), representing instrument error, with each perturbed by the addition of random noise where the scale was determined by the *SE* of the respective bootstrap model $(\sigma_{\epsilon'_L})$. The resulting error represents the fully propagated error that is produced from combining the instrument measurement uncertainty with the uncertainty of the correction model estimates. The entire error estimate can be presented as $\hat{\epsilon}_L$ (Equation 4).

$$\hat{\epsilon}_{L,i} \sim N(\epsilon'_L, \sigma_{\epsilon'_L}) \tag{4}$$

where $\hat{\epsilon}_{L,i}$ is the predicted error (or inversely the predicted correction required) for the *i*th record. For each sample record, we used the median of the 1000 $\hat{\epsilon}_{L_i}$ values as the final predicted error within an experiment and took the 0.025 and 0.975 quantiles of the same 1000 $\hat{\epsilon}_{L_i}$ values as the upper and lower prediction intervals (Figure 4). For analysis and plotting, constant transducer offsets were removed from both WL_{raw} and WL_{corr} by subtracting the mean of ϵ_L and $\hat{\epsilon}_L$ from WL_{raw} and WL_{corr}, respectively, centering the estimates on the true water depth.

2.3 | Uncertainty analysis

The impact of TDEs on water level measurements was evaluated by comparing both the instrument error and the propagated error with the observed error of corrected data. Within each experimental period, the magnitudes of instrument error and propagated error from the bootstrap prediction bounds were compared using a *t* test. Differences between experimental periods were compared using a one-way ANOVA and with Tukey Honest Significant Differences (Tukey) method.

We believe correction to be a necessary step to reduce bias in water level measurements, but additional uncertainty introduced from correction equations could lead to higher variance around the final measurement value. To evaluate the bias-variance trade-off, WL_{corr} was calculated for the $SW_{bias\text{-}1}$ using the correction equations from $SW_{bias\text{-}2}$. New accuracy bounds for the period were computed as $\mu_{\hat{\epsilon}_L} \pm 1.96 * \sigma_{\hat{\epsilon}_L}$, where $\mu_{\hat{\epsilon}_L}$ is the mean of the predicted errors and $\sigma_{\hat{\epsilon}_L}$

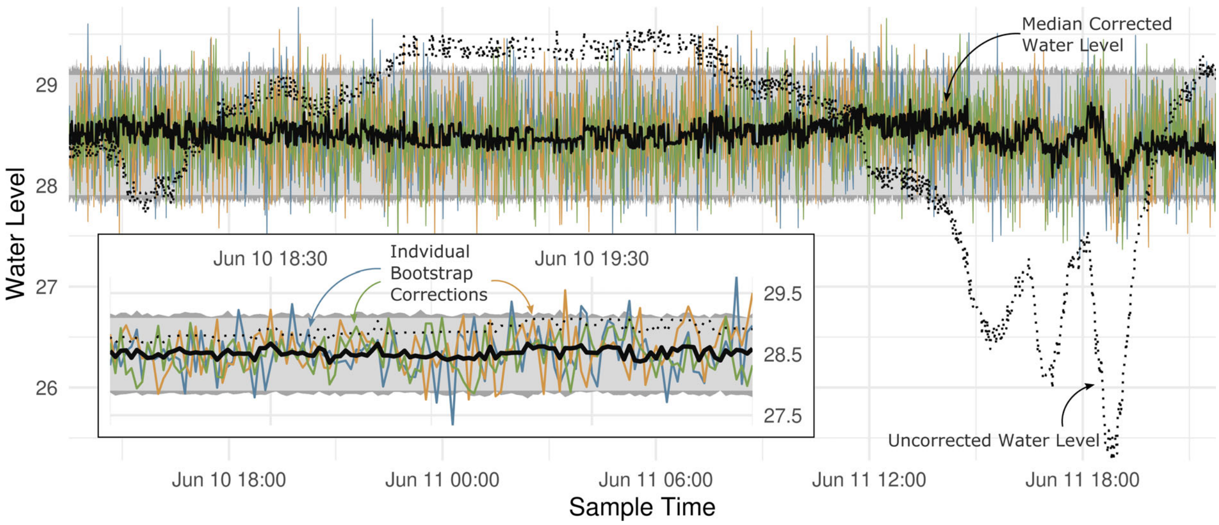

FIGURE 4 A visualization of the development of a single water-barometric transducer correction model to correct water level taken from the $SW_{bias\text{-}2}$ period. The known water depth for this example is ~28.5 cm. The dotted line represents the raw water level from a pressure transducer pair, which can be seen to vary greatly from a constant water level. Each of the three coloured lines represent a single instance of bootstrap model (a random selection of 3 out of 1000 correction models fit) and the solid black line is the median response from all 1000 models. Grey bands show the 95% confidence bands derived from the instrument error as stated by the manufacturer (light grey) and propagated (manufacturer-stated instrument error plus additional error from model uncertainty) error (dark grey). The expanded section shows the same series for a 2-h period for clarity

is the *SD* of predicted errors as in Equation (4). The new accuracy bounds were then tested against manufacturer-stated instrument error bounds using a paired *t* test. As $\hat{\epsilon}_L$ varies for each record in a time series based on the scale of correction, there could be a very high *n* for comparisons of model-derived and manufacturer-stated error bounds. To avoid artificially inflating significance through high *n*, all *t* tests were performed between the fixed instrument error band and the mean of the propagated error band to hold $n = 36$ within each experiment.

2.4 | Correction evaluation

To evaluate correction across conditions, the final models generated for the $SW_{bias\text{-}2}$ period experiment were used to predict $\hat{\epsilon}_L$ for all other experimental periods. For each water-air transducer pair, the $SW_{bias\text{-}2}$–derived correction model was used to predict the error ($\hat{\epsilon}_L$), which was in turn used to correct the recorded water levels such that $WL_{corr} = WL_{raw} - \hat{\epsilon}_L$. Correction effectiveness was evaluated based on two criteria: number of recorded points that fell outside of the range (out of range, OOR) of combined instrument errors ($n_{\epsilon,i}$) and the root mean squared error (RMSE) from the known water depth. Effect of correction was tested using the Tukey method as the pairwise difference in estimated marginal means of the RMSE for corrected and uncorrected data within a given experimental period (Lenth, 2021). Additionally, estimated marginal means of the corrected data RMSE were compared across experimental periods.

2.5 | Case study

We applied the proposed correction protocol to observed wetland water level data to highlight propagation of uncorrected errors into subsequent analyses. The data for this case study comes from a forested headwater depressional wetland (0.81 ha) in the western Upper Peninsula of Michigan, USA. The data were collected during the ice-free season of 2018. Transducer 1 066 016 was deployed 1.75 m below the soil surface in a 5.08-cm internal diameter (I.D.) monitoring well and transducer 1 065 861 was deployed in a sheltered open-air area with both transducers recording synchronous 15-min data (Grinsven et al., 2017). These wetlands are primarily groundwater fed with stable water temperatures that vary at seasonal rather than daily or sub-daily scales, making them most similar to the temperature-differential biased groundwater monitoring (GW_{bias}) experimental period. TDE correction was performed on these data using the final equations developed from the $SW_{bias\text{-}2}$ period, which were evaluated against all other experimental periods for independent validation (below). We will illustrate the impact of temperature bias on two common analyses: mean daily water level and evapotranspiration (*ET*) derived from diurnal water level variation. For the former analysis, we calculated mean uncorrected and corrected water levels on a daily timestep. For the latter analysis, we followed methods built on White (1932).

White (1932) proposed an approach for decomposing daily groundwater level records into evapotranspiration (*ET*) and groundwater recharge. The approach involves estimating net flow rate (G_{in}) during a period with negligible *ET*, typically the pre-dawn hours. This rate and the change in water level between recharge periods on subsequent days can be used to determine *ET*. Converting water level fluctuation to surface *ET* also requires accounting for the ecosystem specific yield (E_{Sy}). E_{Sy} links the scale of surface hydrologic drivers and water table fluctuations. It is a function of the substrate surrounding the well, surface and subsurface hydrologic connections, and for this case study the shape of the wetland basin. Loheide II (2008) and Loheide II et al. (2005) expanded this work to improve estimates of specific yield and account for variation in groundwater flow with water level change. The White and Loheide research focused on groundwater-phreatophyte interactions, but the method has been extended to use in wetland settings (Diamond et al., 2018; McLaughlin & Cohen, 2014; Telander et al., 2015). In these settings, Loheide's readily available specific yield was expanded to ecosystem specific yield to acknowledge the influence of factors beyond the substrate surrounding the well (McLaughlin & Cohen, 2014). We applied the adjusted White methodology to the 15-min wetland well data to calculate daily ET for the raw and corrected water level records (ET_{raw} and ET_{corr}, respectively).

Our estimates of E_{Sy} were derived as the ratio between the change in daily water level and the daily change in water availability (rainfall plus snowmelt less PET) during the drawdown period (start of water level records to minimum observed water level). This is an inverse analog of the rain-to-rise ratio used in other studies to estimate E_{Sy} (McLaughlin & Cohen, 2014). We used this analogous approach because our site is influenced by two factors known to obfuscate a clear rain-to-rise E_{Sy} relationship: surface water exchange (Watras et al., 2017) and additional hydrologic cycles at longer wavelengths (e.g. snowmelt and seasonal inundation patterns) (Zhu et al., 2011). E_{Sy} estimates were then fit to an asymptotic function with water level as the independent variable using iterated reweighted least squares non-linear regression (Maechler et al., 2020). Development of these models is intended to allow for continuous estimation of E_{Sy}. Curves for E_{Sy} were derived separately using the raw and corrected data using data from the same study from a different year to avoid comparing derived *ET* with the same *PET* value used to compute E_{Sy} (Figure S1). The respective E_{Sy} functions were then used to separate the raw and corrected diurnal fluctuations into G_{in} and *ET*. For validation of the correction procedure, we compared the ET_{raw} and ET_{corr} values to potential evapotranspiration (*PET*) calculated using the modified Hargreaves-Samani equation (Hargreaves & Allen, 2003) with data from nearby meteorological stations. For analysis, only dry days were used to avoid the complicating factors of interception, suppressed *ET* and variable water level increase. We defined dry days as days with daily rainfall or snowmelt of less than 1 mm in the preceding 2-day period. We assess the correlation and fit of quantile regression model intercept and slope between both corrected and raw *ET* and *PET*. We also perform a pairwise t-test of both *ET* values and *PET* to test for significant differences between derived *ET* and *PET*. Prior to the *t* test the *ET* values have the intercept of the respective quantile regression model subtracted. This removes any offset error so that the results are a comparison of the relationship between derived *ET* values *PET*.

2.6 | Software

All data processing and analysis were performed in R (R Core Team, 2019) using the packages *data.table* (Dowle & Srinivasan, 2019) for data manipulation, *ggplot2* (Wickham, 2016) and *patchwork* (Pedersen, 2019) for visualization, and *emmeans* (Lenth, 2021) for analysis. The *drake* (Landau, 2018), *rmarkdown* (Allaire et al., 2020; Xie et al., 2018), *rticles* (Allaire et al., 2019) and *rbbt* (Dunnington & Wiernik, 2020) packages were used to generate a reproducible project and publication.

3 | RESULTS

3.1 | Uncertainty analysis

If the instrument error was assumed as the only source of error, the 95% confidence band (CI) is calculated by combining the instrument errors stated in Table 2, where $CI = 2 * Z_{0.95} * \sqrt{(\sigma_1^2 + \sigma_2^2)}$. The resulting error was 0.859 cm for the Levelogger Edge (LT_EDGE) water pressure transducers and 1.19 cm for the Levelogger Junior (LT_Jr) and Levelogger Junior Edge (LT_EDGE_JR) water pressure transducers. When measurement errors and temperature effects were propagated to final water level estimates by including the correction model uncertainty derived from the double bootstrap approach, every experimental period showed a significant increase in uncertainty bounds. The magnitude of the increase varied among experimental treatments from 0.042 cm for GW_{best} to 0.48 cm for $SW_{bias\text{-}2}$ (Table 3). Significant differences were observed in propagated errors between experimental periods. Experimental period $SW_{bias\text{-}2}$ had significantly larger uncertainty than $SW_{bias\text{-}1}$, and the uncertainty of both $SW_{bias\text{-}1}$ and $SW_{bias\text{-}2}$ were significantly larger than GW_{best}, GW_{bias} and SW_{best}, respectively. This variability corresponds to higher variability in air temperature and change in air temperature during this experimental period (Figure S2).

Due to large discrepancies in correction model accuracy between the barometric transducers, we compared the propagated errors and the instrument errors separately for each set of 18 models (Table S2). For transducer 1066019, the propagated errors were found to be significantly less than the instrument error (mean decrease in 95% confidence bounds = 0.67 cm). Transducer 1065861 was found to have propagated errors that were significantly higher than instrument error (mean increase in 95% confidence bounds = 0.21 cm).

3.2 | Correction effectiveness

The model form described in Equation (2) showed a good fit to the $SW_{bias\text{-}2}$ data with bootstrapped estimates of adjusted R^2 ranging from 0.89 to 0.98. The bootstrap predictions of final WL_{corr} had RMSE values ranging from 0.32 to 0.51 cm, which includes both instrument error and correction model uncertainty (Table S1). For reference, observed (WL_{raw}) RMSE for that period ranged from 1.53 to 2.60 cm relative to known water depth. When the $SW_{bias\text{-}2}$ model was used to predict other experimental periods, the RMSEs of raw and corrected water level errors relative to true water level showed differences of −0.61 to 0.04, where a negative value indicates a smaller RMSE following correction (Table 4). The reduction in RMSE was greatest for the $SW_{bias\text{-}1}$ period and along with the GW_{bias} period was significant at the $\alpha = 0.05$ level. The SW_{best} period showed a significant increase in RMSE following correction. We also tested model performance after excluding two transducers (water transducers 2025928 and 2030899) from the analysis based on their status as outliers (Figure 5). With the two outlier transducers removed, the increase in RMSE in the SW_{best} period became a decrease of 0.02 and was no longer significant. When the RMSE of WL_{corr} was compared with instrument error bounds using a t test with $\alpha = 0.05$ and $H_A : \hat{\epsilon}_L - \epsilon_{instrument} < 0$, we found RMSE of WL_{corr} was significantly less than instrument error for all experimental periods.

Prior to correction, the $SW_{bias\text{-}1}$ period showed the largest percentage (56.92%) of measured points outside of the 95% confidence band defined by the combined stated instrument errors of the barometric and water pressure transducers, with all other periods except $SW_{bias\text{-}1}$ showing less than 1% of points outside of the error range (Table 4). Following the correction, only 6.17% of $SW_{bias\text{-}1}$ points fell outside of that same confidence band. A small increase of points outside of the band was observed for SW_{best} (2.48%) and GW_{best} (<0.005%) following correction. When the interval is expanded to include both the instrument uncertainty according to instrument's reported accuracy and the errors associated with temperature correction, only $SW_{bias\text{-}1}$ showed any raw data points (35.40%) outside of the expanded error band. Following correction this fell to 0.36% of points for $SW_{bias\text{-}1}$ but increased from 0.00 to 0.32% for SW_{best}. All corrected points that fell outside of the instrument and propagated error bands were associated with the two outlier transducers mentioned above.

3.3 | Case study

For much of the case study period, the overall trend in water levels tracks well between the raw and corrected levels (Figure 6a). Differences become apparent when the daily mean water level (Figure 6b) and the derived ET estimates (Figure 6c) are compared. When mean WL_{raw} and WL_{corr} are compared, 92 days (55.8%) were found to have

TABLE 3 Mean propagated and instrument errors for all 36 transducer-pair models by experimental period where the propagated error refers to the combined error from instrument uncertainty and the uncertainty of the correction model estimates

Experimental period	Mean propagated error (cm)	Mean instrument error (cm)	Difference ± *SE* (cm)	Significance of propagated error between groups
GW_{best}	1.20	1.15	0.042 ± 0.002*	a
GW_{bias}	1.20	1.15	0.047 ± 0.001*	a
SW_{best}	1.23	1.15	0.071 ± 0.008*	a
$SW_{bias\text{-}1}$	1.39	1.15	0.23 ± 0.02*	b
$SW_{bias\text{-}2}$	1.63	1.15	0.48 ± 0.03*	c

Note: Mean instrument error refers to the mean of the combined instrument error for the 36 tranducer pairs. Differences and significance were calculated pairwise between instrument and propagated error within transducer pair nested within experimental period. Difference and *SE* show the change in confidence band size when all errors are incorporated and tested for significance using a paired *t* test within each experimental period (denoted by '*', $\alpha = 0.05$). Difference in propagated errors between experimental periods was tested using a one-way ANOVA. Different alphabets indicate significance grouping where periods with the same letter are not significantly different.
*Indicates the propagated and instrument errors were significantly different at $\alpha = 0.05$.

TABLE 4 Root-mean-squared error of predicted water level and percentage of observations of raw and corrected water levels falling outside of instrument accuracy range and propagated error range

		Percentage of observed or corrected measurements outside 95% error range			
		Instrument error		Propagated error	
Experiment	RMSE of WL_{corr}[a] (cm)	WL_{raw} (cm)	WL_{corr} (cm)	WL_{raw} (cm)	WL_{corr} (cm)
GW_{best}	0.08 (−0.00)	0.00	0.00	0.00	0.00
GW_{bias}	0.09 (−0.06)	0.36	0.00	0.00	0.00
SW_{best}	0.18 (0.04)[b]	0.13	2.48	0.00	0.32[b]
$SW_{bias\text{-}1}$	0.24 (−0.61)	56.92	6.17	35.40	0.36

Note: RMSE and propagated error were determined using the final model developed using the $SW_{bias\text{-}2}$ data set to test a single correction equation across monitoring scenarios. Mean values of the 36 transducer pairs reported. Values in parentheses show the difference in RMSE (Δ RMSE) between observed (WL_{raw}) water level error and corrected water level error (WL_{corr}) using the $SW_{bias\text{-}2}$ models, where a negative value indicates a reduction in RMSE.
[a]Indicates that the change in RMSE was significantly different at $\alpha = 0.05$.
[b]Transducers 2025928 and 2030899 showed non-standard behaviour (discussed in text). When they are removed from the analysis Δ RMSE = 0.02 (not significant at $\alpha = 0.05$) and no corrected values fell outside of the propagated error range.

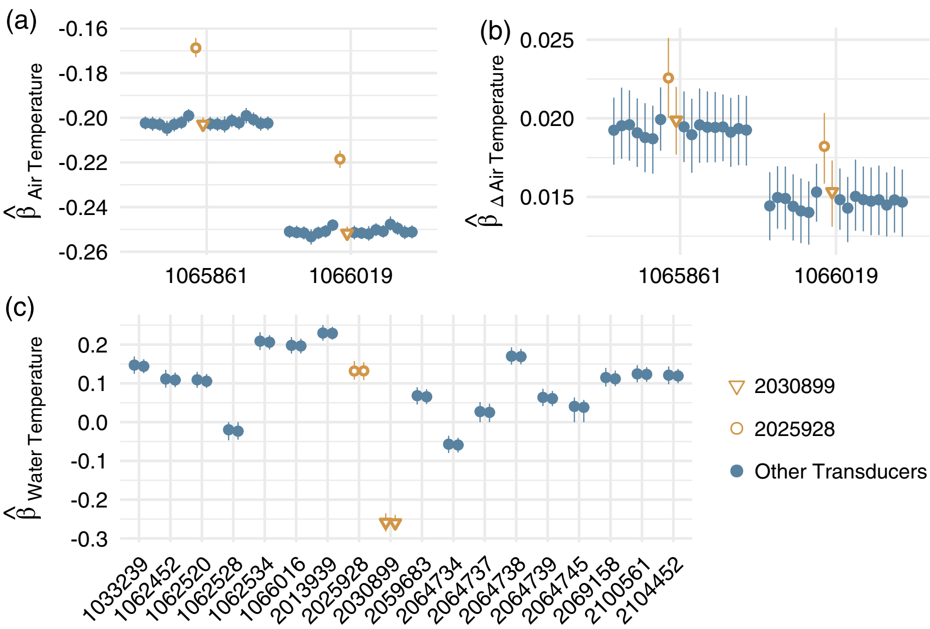

FIGURE 5 Mean and 95% confidence interval for correction model coefficients for the 1000 bootstrap models for the $SW_{bias\text{-}2}$ period. Panels (a) and (c) show coefficients grouping by barometric transducer and panel (b) shows models grouping in pairs for each water-barometric transducer pair. Water transducers 2025928 (hollow circle) and 2030899 (hollow inverted triangle) are highlighted to illustrate disagreement with other models for one or more coefficients

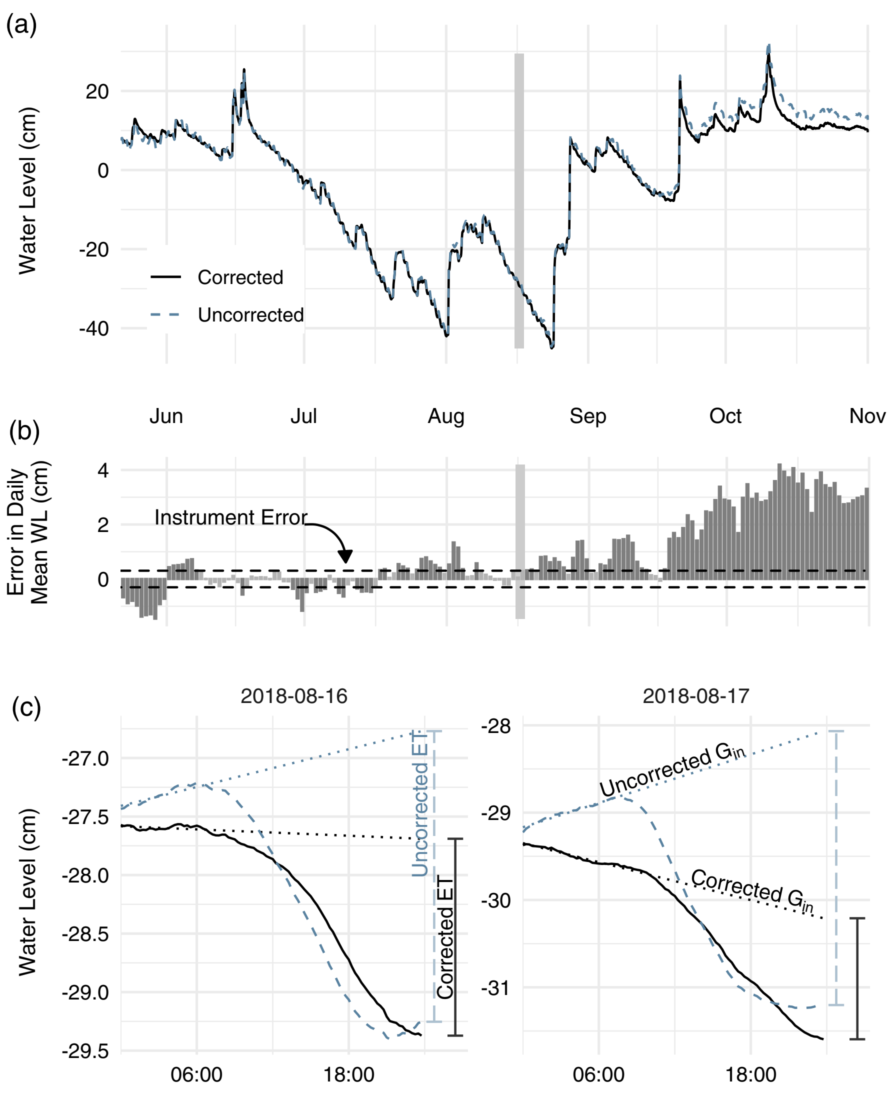

FIGURE 6 A comparison of corrected (using median SW_{bias} models to match data collection conditions) and uncorrected wetland water levels. Panel (a) shows the entire growing season at 15-minute interval. Panel B shows the error in daily mean water level. Panel B dashed lines represent manufacturer-stated instrument error and darker columns indicate errors greater than combined instrument error. The vertical shaded bar in panels (a) and (b) highlight the 2 days shown in panel (c). Panel (c) shows details of two diurnal water level cycles and the decomposed components of the White (1932) method: Inflow (G_{in}) and evapotranspiration (*ET*). In panel (c) the solid line represents corrected water level, the dashed blue line represents uncorrected water levels and the dotted lines illustrate the basic White methodology of estimating G_{in}

temperature corrections that exceeded combined instrument error (Figure 6b). Daily mean errors were greatest in the fall, and smallest during the summer months, ranging from a 1.5 cm underestimate to a 4.2 cm overestimate of mean daily water level.

Errors at the sub-daily range were also found when comparing WL_{raw} and WL_{corr}. The 2 days shown in Figure 6c have daily mean water level errors within the expected instrument error, but the estimates of G_{in} and ET derived from these diurnal fluctuations are qualitatively different upon visual inspection and the groundwater is seen to shift from an inflow (positive slope) to an outflow (negative slope). The uncorrected data show higher inflow and corresponding higher ET even when the days show similar raw and corrected water level change. During the drawdown period used to calculate E_{Sy}, the temperature-induced error was generally small (Figure 6b) resulting in similar E_{Sy} functions (Figure S1). While there is no validation data to compare estimated inflow to actual inflow, we can compare estimated ET to PET. ET_{raw} showed weaker correlation with PET (Spearman's $r = 0.27$), while ET_{corr} showed moderate correlation with PET (Spearman's $r = 0.44$). Fitting a quantile median regression between estimated ET and PET resulted in a unit-slope near-zero intercept fit for ET_{corr}. ET_{raw} showed a larger intercept and a slope with a large confidence interval which overlapped zero (Figure 7). After removing the respective intercept term from the corrected and raw ET values the mean pairwise difference between ET_{raw} and PET was -0.13 ± 0.096 and 0.02 ± 0.056 between ET_{corr} and PET. The difference between ET_{raw} and PET was significantly different than 0 ($p_{bonferonni} = 0.02$), but was not significantly different from zero for ET_{corr} and PET ($p_{bonferonni} = 0.92$).

4 | DISCUSSION

4.1 | Temperature-induced bias and correction equations

Our results show that it is possible to correct water-depth observations derived from a pair of water and barometric absolute pressure transducers without the need to reference to an external data source. Any transducer correction approach must be a trade-off between convenience, absolute accuracy and uncertainty. The described approach to correction achieves accuracy that reduces the SW_{bias-1} error rate from 56.92% to 6.17% of observed points that fall outside of the bounds of combined instrument error (an 89% reduction) and reduces up to 99% of erroneous uncorrected data to within the fully propagated error bounds (Table 4). An alternative correction would use a two-stage correction procedure where barometric pressure transducer readings need to first adjusted to a reference set of external pressure readings and then subsequently used to correct the water pressure transducer readings. Whereas the correction method presented here uses a single stage correction procedure where the water level serves as the reference point making this method comparatively advantageous because it can be performed at monitoring locations with little to no extra equipment. Moreover, by using a single-stage correction we also remove the dependence of using additional pressure readings from an external reference data source, effectively avoiding a tertiary source of error and thereby minimizing the total uncertainty of the corrected water level measurements.

Importantly, the presented approach captures the individual transducer corrections. One possibility of the dual transducer model is that the model would be fit to some artefact of the interaction between the transducers, rather than each transducer being represented in the air and water temperature coefficients. Figure 5 shows that the air temperature coefficients were grouped by barometric transducer (two distinct regions of overlapping ranges), and the water temperature coefficients were grouped by water transducer (18 distinct regions of overlap of the two models representing each water transducer). Clear transducer differentiation also supports the assertion above that water level errors using absolute pressure transducers are introduced through independent water and air temperature errors and not temperature differences. Correcting for water level error using temperature difference alone would disregard the individual differential transducer responses and the asynchronous temporal variation between air and water temperatures. Based on the scale of the coefficients for air temperature and water temperature, these errors would likely be small in this study. The error would be directly proportional to the sensitivity of each transducer to temperature and would likely vary between manufacturers and instruments.

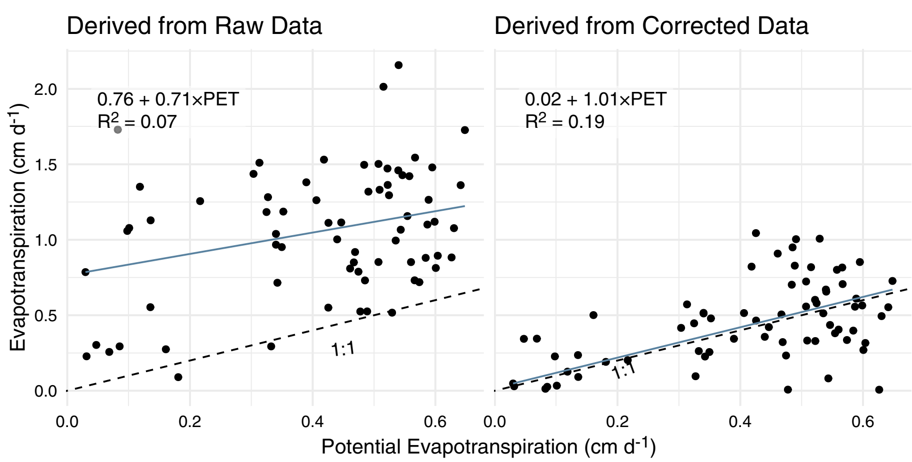

FIGURE 7 Quantile median ($\tau = 0.5$) regression fit between potential evapotranspiration and White (1932) method estimates of evapotranspiration. The solid line represents the quantile regression fit and the dashed line shows a 1:1 relationship to represent perfect estimation of ET from *PET*

Both Cain III et al. (2004) and Liu and Higgins (2015) recommend lab-based correction equations be developed for all deployed transducers. Building upon that, we have demonstrated that it is possible to develop a single lab-based correction with a structure suitable for varied deployment environments. Our correction model developed on $SW_{bias\text{-}2}$ performed well correcting other experimental periods. We were also able to reduce the total uncertainty for an independent data set below instrument error using our $SW_{bias\text{-}2}$ correction equations. However, we recommend developing a correction equation under conditions similar to field conditions whenever possible to ensure accurate correction. The degraded predictive performance of $SW_{bias\text{-}2}$ models under SW_{best} conditions relative to all other periods highlights this point (Table 4). Under SW_{best} conditions we observed a small but significant increase in RMSE for corrected water levels. We believe that the inaccuracies have two potential causes at their root. The first is that the SW_{best} conditions had higher temperatures than any of the other experimental periods. Models developed on $SW_{bias\text{-}2}$ data were therefore extrapolating when predicting under SW_{best} conditions. The second possibility is the internal temperature compensation of the pressure transducers caused inaccuracies. Freeman et al. (2004) anticipated that isolating and correcting temperature-driven errors would be more difficult in transducers with internal compensation. The data used to train our correction models were purposefully collected under the extreme conditions ($SW_{bias\text{-}2}$), with large, rapid temperature fluctuations in both the air and water temperature readings. Under these conditions, the internal temperature compensation algorithm is likely overwhelmed. Under conditions with slower temperature fluctuations (SW_{best}), the correction models may overcorrect the data because the internal temperature compensation has already dampened the temperature-induced errors. Identifying extrapolation and internal-compensation driven errors illustrates the importance of using an independent data set to validate correction models.

Model performance in this study was highly dependent on individual transducers. Based on the results of the uncertainty analysis, barometric transducer 1065861 appears to be more sensitive to air temperature and high rates of temperature change than transducer 1066019 as evidenced by the difference in results of propagated error bounds (Table S2). This led to larger errors at the beginning and end of the $SW_{bias\text{-}1}$ period when rates of temperature change were greatest (Figure S2). Transducer 2030899 shows larger errors and along with 2025928 accounts for all the increase in OOR measurements in the SW_{best} period. Further investigation into this logger suggests that during the SW_{best} period the direction of the relationship between water temperature and error reversed relative to all other experimental periods (Figure S3). Transducer 2064734 showed the same effect but to a lesser extent. Currently, it is impossible to tell if this is due to the high temperatures during this period, differential internal compensation among transducers, or some other error such as a trapped air bubble against the transducer membrane. The experimental and modelling approach laid out above can, and should, be used to adjust the correction equations for individual transducers such as 2030899 and 2064734. That scale of tuning is outside of the scope of this study, which is instead focused on commonalities in correction that can be applied across instruments and manufacturers and the role of TDEs in total uncertainty. The above discussion highlights an advantage of simultaneous correction of multiple instruments. By simultaneously including multiple pairs of barometric and water pressure transducers, we can identify and isolate additional errors specific to the barometric or water transducer.

4.2 | The role of uncertainty in water level measurements

We found that under certain conditions water level errors exceeded instrument accuracy bounds (Figures 4 and 6, Table 4). As expected from previous work (McLaughlin & Cohen, 2011), experimental periods showed the largest errors when air and water transducers were under separate thermal regimes. Our results indicate that when deploying transducers to variable temperature environments, similar air and water thermal regimes (best practices) are enough to prevent significant increases in uncertainty relative to stable temperature environments. Although total error is likely to be greater than stated instrument error, monitoring of small streams, surface waters and other water sources with diurnal temperature signals is not inherently more uncertain than groundwater or other stable-temperature environments when best practices are observed. The magnitude of error introduced by air temperature variation was greater and more uniformly influential when compared with the impact of water temperature variation in our results (Figure 5a,c). This pattern suggests that an untested deployment scenario with stable water temperatures and highly variable air temperatures would show errors closer in magnitude to the SW_{bias} condtions than GW_{bias} as described in this research.

The observed errors, both within and outside of instrument accuracy bounds, were correlated with environmental conditions. This correlation is especially vexing because the points that are most likely to lie outside of the error bounds occur simultaneously with periods of hydrologic interest. As the errors are associated with higher temperatures and high rates of temperature change, the errors will often co-occur with changes in flow following snowmelt or rain events or periods of peak evapotranspiration. To confidently attribute diurnal or seasonal signals to true variation rather than artificial errors requires additional validation. One source of validation is using additional instrumentation that is not susceptible to the same error source, such as frequent manual measurements of stage/level (e.g., Cuevas et al., 2010; Gribovszki et al., 2013). Where this is not practical or possible, we have shown that correction equations can quantify the additional uncertainty of environmentally-induced errors and rule them out as the cause of the signal of interest.

Even when transducers are deployed under 'best practices', temperature-induced errors will still occur. The magnitude of the error may be within the range of instrument error and thus may not be apparent. Our results show that even if not apparent, this error should be considered because it represents systematic bias rather than a random error. Of the 1000 models fit per experimental period, 55.1% of GW_{best}, 2.2% of GW_{bias} and 28.8% of SW_{best} final models were intercept-only. This trend shows that under increasingly ideal conditions more models showed no significant relationship with air temperature, water temperature or rate of change of air temperature. However, even under the most stable monitoring scenarios (GW_{best}) almost half of all fit bootstrap models showed significant correlation between error and environmental variables. The error associated with these intercept-only models still provides important information regarding true water level accuracy and they are included in the propagated error estimates. The intercept-only models are likely not devoid of temperature derived errors, and SW_{best} and GW_{bias} show slightly, but significantly higher σ_{mod} than other models [$\sigma_{intercept\ models} - \sigma_{slope\ models}$: GW_{best}: -0.005 ($p = 0.42$); GW_{bias}: -0.012 ($p = 0.02$); SW_{best}: -0.010 ($p = 0.010$)]. This small increase in σ_{mod} is likely the result of unmodeled temperature-driven errors. As temperature artefacts were present under all conditions, it is not unexpected to find that true uncertainty was greater than instrument error under all conditions. Taken together it is likely that even under ideal conditions thermal artefacts will affect measurement accuracy in a systematic way and correction is recommended.

Increases in measurement error can have important implications for research quantifying small changes or weak diurnal (or other cyclic) signals. Observations that fall outside of, but close to the stated instrument accuracy range, may not be outside of the range of full uncertainty. Using a validation data set collected under the same conditions, we demonstrated above that we could quantify the error of newly corrected observations and increase final measurement precision. In this study, our best performing correction models showed a 65% reduction in the uncertainty range, from 1.19 to 0.42 cm (Table S2). A reduction of this magnitude is not insignificant in the scale of diurnal signals often studied.

4.3 | Case study

For both analyses performed in the case study we found significant bias introduced due to temperature-induced errors in WL_{raw}. The purpose of water level records in monitoring and research is to enable further analysis. Unaccounted errors in instrument records have measurable impacts, which could lead to erroneous conclusions and potentially to real-world management decisions that are at best misguided and at worst harmful. The presence of detectable error is dependent on the scale of the analysis performed. In Figure 6, the two highlighted days in panel C show approximately a 1.7-fold difference in ET estimates, representing estimation errors of 170% and 166% of ET_{corr}. In Panel b, we see that the daily mean error on these days was within the range of expected instrument error. Quantifiable errors may be more likely to be detected at finer scales (sub-daily) than at coarser scales (daily) as aggregation smooths out evenly distributed errors. But during periods where the errors are systematically positive or negative no level of smoothing will bring temperature-induced errors within the range of instrument error. The errors observed in the fall are one such period for this case study (Figure 6a).

We observed the largest errors in the fall after wetland water levels had rebounded to the surface. This is due to a combination of seasonal air and water temperature trends. This period had the lowest air temperatures, and the highest water temperatures (not shown). Figure 1a,c show that these conditions led to the largest errors in our laboratory experiments. As the air temperatures decline in the fall and wetland water temperatures increase due to groundwater inputs, the temperature bias increases. Errors in the spring would likely be large, but somewhat smaller than fall errors with low air temperatures and low water temperatures due to snowmelt. Our observed daily mean error was of similar magnitude to the 1.5 cm increase in diurnal variation observed by McLaughlin and Cohen (2011). Our estimates of ET error were much larger than the 19% increase in streamflow diurnal fluctuation presented in Cuevas et al. (2010), though ET is often a smaller component of the water budget in such a setting so that larger percentage errors can be expected.

There were two potential sources of error when calculating ET using the modified White method: the magnitude and shape of the diurnal fluctuations and the determination of E_{Sy}. We have concluded that the poor correlation between ET_{raw} and PET is the result of the diurnal water level fluctuations. We did not see evidence that the E_{Sy} calculations played a role as the final E_{Sy} functions were similar for the raw and corrected data sets (Figure S1). It is important to note that while ET_{corr} showed a moderate correlation with PET and a significant slope near 1, the R^2 of the OLS model was low. This is in line with other White-method separation studies where low R^2 values are observed. Soylu et al. (2012) found R^2 values of less than 0.2 for the White method and 0.33–0.40 for their proposed sine-wave based method. McLaughlin and Cohen (2014) did not report R^2, but their ET/PET index showed high variability, suggesting low R^2. Even with direct eddy flux measurement of ET, Lafleur et al. (2005) found an R^2 as low as 0.56 between ET and PET in a Canadian shrub wetland. Based on other studies, our wetland could be expected to have poor performance for the White Method. Nearby surface water dynamics and connectivity to other surface water sources have been shown to have high levels of interference with determining E_{Sy} or White-method ET (Watras et al., 2017; Zhu et al., 2011).

4.4 | Uncertainty and transducer type

This study focused on absolute pressure transducers, but some of the findings presented here could inform similar work on gauge pressure transducers. Cain III et al. (2004) found that temperature changes within the vent-tube of gauge pressure transducers introduce water level artefacts, which was explored further by Liu and Higgins (2015) in multi-day laboratory and long-term field experiments. Their findings showed that errors can be introduced due to water temperature or rapid fluctuations in air temperature that would be more likely to occur under shallow groundwater monitoring scenarios. Their laboratory experiment found 1.4 mm $^{\circ}C^{-1}$ error as water temperature changed (with constant air temperature). These corrections corresponded to errors of up to 0.7 cm in their field experiments, with at least some of that error due to the rate of temperature change affecting the connection to the atmosphere via the vent tube. They also highlight the impact of rate of temperature change, with more rapid gradients inducing larger water level errors, which agrees with the results from our study. Agreement between the two logger types is expected as both contain internal electronics and membranes that must equilibrate to the new temperatures (Freeman et al., 2004). Under certain common monitoring scenarios, such as deep groundwater monitoring with a gauge pressure transducer, TDE would be unlikely to occur due to the stable temperature regime of both the water and the vent-tube, and therefore would not require additional temperature-induced correction procedures. However, the methodology for developing correction equations presented here could be easily adapted to gauge pressure transducers to similarly reduce temperature-induced errors from observation data collected under scenarios where rapid fluctuations in groundwater or air temperature are more likely to occur.

5 | CONCLUSIONS

We presented a laboratory experiment demonstrating that thermal bias can increase measurement uncertainty beyond instrument error. In the case of absolute pressure transducers, the added uncertainty is not the result of temperature difference between transducers, but is a combination of individual transducer biases. These biases can impact uncertainty across commonly used deployment scenarios for hydrologic monitoring. Even under best deployment practices, individual logger response to similar thermal regimes will result in increased uncertainty, and can result in measurable water level errors. Transducer and transducer-pair specific correction equations can be used to reduce errors and fully quantify measurement uncertainty. In addition, laboratory-derived correction can be used to reduce final uncertainty below stated instrument error, providing a mechanism for higher resolution analyses. We provide the following recommendations for addressing and correcting TDEs.

- For each transducer, check for *all* potential error-inducing responses to temperature and rates of change in temperature.
- Even under best practice deployments thermal error should be considered. This is especially true if monitored water sources vary in temperature, for example surface water and shallow groundwater.
- Correction equations should be developed under conditions that capture the full range of temperatures likely to be monitored.
- To maximize correction benefits, collect separate validation data sets not used to derive equations. These data sets can be used to create new and narrower uncertainty estimates.
- When possible, correct multiple transducers simultaneously to better isolate problematic instruments via cross-reference. If that is not possible, compare with local weather station(s) for validation of your observations.

This study focused on trend correction rather than offset. Often systematic offsets are corrected through manual calibration measurements during deployment. If deployed conditions will not allow for manual calibration measurements, the above approaches are applicable, but offsets must be carefully calculated.

ACKNOWLEDGEMENTS

The authors would like to thank the Great Lake Restoration Initiative, the MTU Ecosystem Science Center and the MTU Graduate School Finishing Fellowship program for supporting this research.

DATA AVAILABILITY STATEMENT

Raw data and analysis code used in this analysis is available at https://www.github.com/jpshanno/levellogger_correction.

ORCID

Joseph Shannon https://orcid.org/0000-0002-0052-9962
Fengjing Liu https://orcid.org/0000-0001-8743-321X

REFERENCES

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *Rmarkdown: Dynamic Documents for R*. Manual.

Allaire, J. J., Yihui X., Wickham H., Vaidyanathan R., Boettiger, C., Broman, K., Mueller, K., Quast, B., Pruim, R., Marwick, B., Wickham, C., Keyes, O., Yu, M., Emaasit, D., Onkelinx, T., Gasparini, A., Desautels, M., Leutnant, D., Ogreden, O, Journal of Statistical Software, MDPI & Taylor and Francis. (2019). Rticles: Article Formats for R Markdown. Manual.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications. Interdisciplinary statistics series*. Chapman and Hall/CRC.

Cain, S. F., III, Davis, G. A., Loheide, S. P., II, & Butler, J. J., Jr. (2004). Noise in pressure transducer readings produced by variations in solar radiation. *Ground Water*, *42*(6), 939–944. https://doi.org/10.1111/j.1745-6584.2004.t01-12-.x

Mazur, C., Martha, L., Wiley, M. J., & Wilcox, D. A. (2014). Estimating evapotranspiration and groundwater flow from water-table fluctuations for a general wetland scenario. *Ecohydrology*, *7*(2), 378–390. https://doi.org/10.1002/eco.1356

Cuevas, J., Calvo, M., Little, C., Pino, M., & Dassori, P. (2010). Are diurnal fluctuations in Streamflow real? *Journal of Hydrology and Hydromechanics*, *58*(3), 149–162. https://doi.org/10.2478/v10098-010-0014-0

Diamond, J. S., McLaughlin, D. L., Slesak, R. A., D'Amato, A. W., & Palik, B. J. (2018). Forested versus herbaceous wetlands: Can management mitigate ecohydrologic regime shifts from invasive emerald ash borer? *Journal of Environmental Management*, *222*, 436–446. https://doi.org/10.1016/j.jenvman.2018.05.082

Dowle, M, and Srinivasan A. 2019. *Data.Table: Extension of 'data.Frame'*. Manual.

Dunnington, D, and Wiernik B.M. 2020. *Rbbt: R Interface to the Better BiBTex Zotero Connector*. Manual.

Freeman, L. A., Carpenter, M. C., Rosenberry, D. O., Rousseau, J. P., Unger, R., & McLean, J. S. (2004). Use of submersible pressure transducers in water-resources investigations. In *In book 8: Instrumentation, A3*. US Geological Survey.

Gribovszki, Z., Kalicz, P., & Szilágyi, J. (2013). Does the accuracy of fine-scale water level measurements by vented pressure transducers permit for diurnal evapotranspiration estimation? *Journal of Hydrology*, *488*(April), 166–169. https://doi.org/10.1016/j.jhydrol.2013.03.001

Hargreaves, G. H., & Allen, R. G. (2003). History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering*, *129*(1), 53–63. https://doi.org/10.1061/(ASCE)0733-9437

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*(2), 1–34. https://doi.org/10.1029/2008WR006912

Lafleur, P. M., Hember, R. A., Admiral, S. W., & Roulet, N. T. (2005). Annual and seasonal variability in evapotranspiration and water table at a shrub-covered bog in southern Ontario, Canada. *Hydrological Processes*, *19*(18), 3533–3550. https://doi.org/10.1002/hyp.5842

Landau, W. M. (2018). The drake R package: A pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *3*(21), 550.

Lenth, Russell V. 2021 "Emmeans: Estimated Marginal Means, Aka Least-Squares Means. Manual." https://github.com/rvlenth/emmeans.

Liu, Z., & Higgins, C. W. (2015). Does temperature affect the accuracy of vented pressure transducer in fine-scale water level measurement? *Geoscientific Instrumentation, Methods and Data Systems*, *4*(1), 65–73. https://doi.org/10.5194/gi-4-65-2015

Loheide, S. P., II. (2008). A method for estimating subdaily evapotranspiration of shallow groundwater using diurnal water table fluctuations. *Ecohydrology*, *1*(1), 59–66. https://doi.org/10.1002/eco.7

Loheide, S. P., II, Butler, J. J., & Gorelick, S. M. (2005). Estimation of groundwater consumption by Phreatophytes using diurnal water table fluctuations: A saturated-unsaturated flow assessment. *Water Resources Research*, *41*(7), 1–14. https://doi.org/10.1029/2005WR003942

Maechler, M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao E.L.T, and di Palma M.A.. 2020. Robustbase: Basic robust statistics. Manual.

McLaughlin, D. L., & Cohen, M. J. (2011). Thermal artifacts in measurements of fine-scale water level variation. *Water Resources Research*, *47*(9), 1–3. https://doi.org/10.1029/2010WR010288

McLaughlin, D. L., & Cohen, M. J. (2014). Ecosystem specific yield for estimating evapotranspiration and groundwater exchange from diel surface water variation. *Hydrological Processes*, *28*(3), 1495–1506. https://doi.org/10.1002/hyp.9672

Moore, M. F., Vasconcelos, J. G., Zech, W. C., & Soares, E. P. (2016). A procedure for resolving thermal artifacts in pressure transducers. *Flow Measurement and Instrumentation*, *52*, 219–226. https://doi.org/10.1016/j.flowmeasinst.2016.10.010

Pedersen, T.L.. 2019. Patchwork: The composer of plots. Manual.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Soylu, M. E., Lenters, J. D., Istanbulluoglu, E., & Loheide, S. P. (2012). On evapotranspiration and shallow groundwater fluctuations: A Fourier-based improvement to the White method. *Water Resources Research*, *48*(6), 1–17. https://doi.org/10.1029/2011WR010964

Tamari, S., & Aguilar-Chávez, A. (2010). Probando Transductores de Presión Sumergibles Para Monitorear El Nivel Del Agua En Tanques. *Tecnología Y Ciencias Del Agua*, *1*(3), 71–88.

Tanaka, M., Girard, G., Davis, R., Peuto, A., & Bignell, N. (2001). Recommended table for the density of water between 0 C and 40 C based on recent experimental reports. *Metrologia*, *38*(4), 301–309. https://doi.org/10.1088/0026-1394/38/4/3

Telander, A. C., Slesak, R. A., D'Amato, A. W., Palik, B. J., Brooks, K. N., & Lenhart, C. F. (2015). Sap flow of black ash in wetland forests of northern Minnesota, USA: Hydrologic implications of tree mortality due to emerald ash borer. *Agricultural and Forest Meteorology*, *206*, 4–11. https://doi.org/10.1016/j.agrformet.2015.02.019

Grinsven, V., Matthew, J., Shannon, J. P., Davis, J. C., Bolton, N. W., Wagenbrenner, J. W., Kolka, R. K., & Pypker, T. G. (2017). Source water contributions and hydrologic responses to simulated emerald ash borer infestations in depressional black ash wetlands. *Ecohydrology*, *10*(7), e1862. https://doi.org/10.1002/eco.1862

Watras, C. J., Morrison, K. A., Rubsam, J. L., & Buffam, I. (2017). Estimates of evapotranspiration from contrasting Wisconsin Peatlands based on Diel water table oscillations. *Ecohydrology*, *10*(4), e1834. https://doi.org/10.1002/eco.1834

White, W. N. (1932). *A method of estimating ground-water supplies based on discharge by plants and evaporation from soil: Results of investigations in* (Vol. 659). US Government Printing Office.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC.

Zhu, J., Young, M., Healey, J., Jasoni, R., & Osterberg, J. (2011). Interference of river level changes on riparian zone evapotranspiration estimates from diurnal groundwater level fluctuations. *Journal of Hydrology*, *403*(3–4), 381–389. https://doi.org/10.1016/j.jhydrol.2011.04.016

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Shannon, J., Liu, F., Van Grinsven, M., Kolka, R., & Pypker, T. (2022). Magnitude, consequences and correction of temperature-derived errors for absolute pressure transducers under common monitoring scenarios. *Hydrological Processes*, *36*(1), e14457. https://doi.org/10.1002/hyp.14457