Regina Jason (Orcid ID: 0000-0002-4091-8647)

Automated Correction of Systematic Errors in High Frequency Depth Measurements Above V-Notch Weirs using Time Series Decomposition

Jason A. Regina[1,✉] and Fred L. Ogden[1,2]

Hydrologic studies require high quality streamflow measurements. V-notch weirs provide accurate discharge measurements and when properly equipped produce continuous records of streamflow. However, sensor drift and accumulations of floating woody debris cause systematic errors in depth measurements behind weirs, particularly during periods of base flow between storms. Manual processing of high-frequency streamflow data is often subjective and difficult to reproduce. We developed a method to automatically correct erroneous depth time series to accelerate data processing and promote objectivity and reproducibility. The method inspects depth measurements, isolates and preserves periods of direct runoff during storms, and corrects erroneous inter-event baseflow. We applied experience obtained during manual data processing to implement an automatic correction method based on time series decomposition to produce a consistent time series of stage data that eliminates obvious errors due to sensor drift and clogging by woody debris. The automatic method offers an objective and reproducible procedure capable of efficiently processing one year of stage data in seconds. This method promotes reproducibility and may appeal to water resources experts working with large amounts of data especially at sites where distance, cost, or inclement weather prevents regular instrument maintenance.

[1] Dept. of Civil & Architectural Engineering, University of Wyoming
[2] NOAA-NWS Office of Water Prediction

[✉] Correspondence: Jason A. Regina <jasonaregina@gmail.com>

1. Introduction

Hydrologic science suffers from many problems related to hydrometry. Many experimental catchment-scale studies struggle with the primary requirement of water balance closure (Beven et al. 2020). Hydrometry errors negatively impact the utility of datasets. The inability to close the water balance in catchment studies diminishes the utility of datasets hypotheses testing at the catchment scale (Beven and Lane, 2019). Many physiographic factors impede water balance closure (Safeeq et al., 2021), and limits on observability of those factors (e.g. deep groundwater exchanges) requires that we minimize errors associated with measurements of observable quantities, including streamflow (Westerberg et al., 2020).

Water professionals frequently use streamflow measurements for flood and drought prediction (McEnery et al., 2005), allocation of water rights (Wurbs & Bria Wallsn, 1989), hydraulic design (Close et al., 1970), energy production (Pereira et al., 1984), and enforcement of water quality policies (Bosch et al., 2002). Experimental watershed managers and hydrological modelers

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hyp.14405

require accurate streamflow measurements to evaluate hydrological models and inform policy decisions. Acquiring and processing streamflow measurements are necessary steps to address a multitude of water resources issues. Raw streamflow measurements require preprocessing to correct known errors (Famili et al., 1997).

Streamflow measurements are fundamental to hydrologic science and engineering, and to water management. Dobriyal et al. (2017) categorized methods of streamflow measurement as direct measurements, velocity-area methods, constricted flow methods, or non-contact methods. Weirs such as shown in Fig. 1 represent a constricted flow method of streamflow measurement. Weirs are hydraulic structures that obstruct streamflow, increase water depth, and force flow through a precisely designed control section (Bos, 1976). Calculation of discharge over the weir uses upstream measurements of water depth and known head-discharge relationships for a specific weir (United States Bureau of Reclamation, 2001). Well constructed and maintained concrete short-crested weirs can provide discharge measurements with less than 5% uncertainty in discharge coefficient when the stage over the weir is greater than about 1.5 times the crest thickness measured in the flow direction (Ogden et al. 2017).

Rantz (1982) identified several sources of uncertainty affecting weir discharge measurements. These include non-standard weir designs, sensors and data-loggers, upstream velocity distribution, head-discharge relationships, field-observed head measurements, construction, and maintenance. Several prior studies focused on quantifying or reducing uncertainty associated with installed sensors or head-discharge relationships (Haddad et al., 2010; Hamilton & Moore, 2012; Oberg & Mueller, 2007; Ogden et al., 2017; Petersen-Øverleir et al., 2009; Pfister et al., 2013). Weir maintainers employ preventative measures to mitigate errors associated with debris and sensor drift. Hydrological professionals can correct erroneous data after collection in cases where remote installations, inclement weather, or available resources inhibit implementation of preventative measures.

A literature review turned up no example solutions to the problem of automated correction of water stage measurements above weirs. Methods to detect and correct systematic errors vary by measurement method and the characteristic or event under observation. Several studies introduced methods using techniques from digital signal processing to identify components of systematic error. Yu et al. (2014) applied a weighted moving average model with confidence bounds to detect individual outliers in depth and discharge data. Chen et al. (2010) used non-parametric smoothing filters to estimate underlying components of corrupt load curves in power systems. Dorninger et al. (2008) applied Fourier analysis to detect and correct cyclical components of systematic error in distance measurements from terrestrial laser scanners. Makarenkov et al. (2007) used regression analysis and normalization to detect and correct systematic errors in chemical compound measurements. Lu et al. (2003) tested an additive model to decompose Normalized Difference Vegetation Index (NDVI) into perennial and annual components to detect contaminated NDVI data.

The objective of this study was to identify and test a method to remove systematic biases in the underlying depth measurements from two sources (1) sensor drift, and, (2) intermittent partial clogging of weirs by floating debris, which predominantly occurs during periods of low flow. We developed a method that uses an additive model of depth to isolate and correct systematic depth errors during times of low-flow or “baseflow.” Criteria for the method included minimal

hydrological assumptions, repeatability, and retention of important hydrograph characteristics. We desired a method based on the same assumptions used to manually correct stage data and eliminate periods of physically-unrealistic changes in discharge. Ideally, such an algorithm would produce small, approximately normally-distributed corrections, without significant changes to flow-duration curves or commonly used flashiness indices.

2. Data Source and Error Characteristics

Motivation for this study comes from establishment of instrumented research catchments as part of the Smithsonian Tropical Research Institute, "Agua Salud" ecosystem services study in the steep, humid-seasonal tropical lowlands of Panama (Stallard et al., 2010). A total of 13 instrumented weirs established streamflow measurement points from catchments that differed in landuse and scale. Initial plans called for manual editing of stage measurements to eliminate occasional problems associated with partial blockages. In reality, blockages occur frequently. Sensor drift proved an unanticipated and common error source. Manual processing of the stage record to eliminate these frequent errors proved to be too time intensive given the rate of data generation, and essentially impossible to perform in a systematic and repeatable fashion.

The manual approach was the standard method we employed between approximately 2010 to 2015. Our experience with manual processing indicated that it required many months and maybe as much as one year of experience-based learning to produce physically consistent depth time series. Evaluating and implementing necessary corrections was subjective, dependent on user fatigue, and varied by user. The data editor interface offered limited tools for manipulating large datasets. Users employed the offset and rotation tools for most corrections. We found the point-and-click interface and simple transformations failed to scale with increasing data volume.

2.1 Data Source

The "Agua Salud" Project (ASP) studied hydrologic ecosystem services provisioning in catchments with different land covers (Adamowicz et al. 2019). The study area included 13 experimental catchments in Central Panama with varying land-covers (Stallard et al., 2010; Regina et al., 2020). Panama lies in the seasonal tropics, with an early May to mid/late-December "wet" season during which about 90% of annual rain falls (Ogden et al., 2013). The rest of the year from mid/late-December through early May represents the "dry" season, when the occurrence of rainfall almost stops and streamflow in first order catchments decays as groundwater storage depletes.

The ASP defined each catchment outlet by a V-notch weir equipped with a non-vented pressure transducer (In-Situ LevelTROLL, model 100, 300, or 500) located behind the weir and below the water surface. Pressure transducers recorded pressure every five minutes. A field technician retrieved pressure transducer data and recorded the water depth using a staff gage on a semi-monthly basis. We used three In-Situ BaroTROLL instruments installed no more than 15 km away to remove the effects of barometric pressure from the water pressure record. We used field measurements of water depth as a reference to convert baro-corrected absolute pressure measurements to depth measurements. Fig. 2 shows a record of erroneous depth measurements

from April 20, 2016 to May 25, 2017 used to test the method.

The data shown in Fig. 2 represent observations at the outlet of a 144.0 ha old-regrowth forested catchment in the ASP (Regina et al., 2020, 2021). We defined the outlet at a two-stage V-notch weir. The two-stage weir consisted of a high stage concrete short-crested 143° nominal (142.66° ± 0.41° measured) V-notch weir of the USDA-ARS design (Brakensiek et al., 1979) and a low stage box equipped with a 90° sharp-crested V-notch weir with a capacity of 0.03 m^3s^{-1}. We only used high stage data to generate the depth time series shown in Fig. 2 and test the correction method.

2.2 Known Error Sources

Initial reliance on manual data correction provided evidence of error sources. Systematic errors or biases appear in the depth record as behavior inconsistent with unmanaged flows from a small headwater catchment in the humid tropics. Fig. 2 includes examples of temporary errors evident by sudden decreases in depth after a storm event, sudden decreases in depth following a field measurement, sudden increases in depth not associated with a storm event, and gradual increases in depth ending at a field observation. Debris in the weir and sensor drift caused temporary erroneous trends in the depth hydrograph. Storm events, sensor redeployment, or debris releases reset or altered specific biases in the overall hydrograph trend. We expected depth hydrographs from the test catchment to exhibit "flashy" storm events with short rising limbs followed by longer falling limbs. We assumed perennial flows that receded gradually during dry periods.

Other known sources of error included staff gage readings and timing observations. Under ideal conditions, a field technician took a depth measurement during baseflow at an unclogged weir, then retrieved and redeployed the pressure transducer as quickly as possible. In practice, a technician's activities at a particular weir may have taken an hour or more and may have coincided with event flows. Technicians typically retrieved data between storms during relatively calm periods and avoided data retrieval if flows changed rapidly. We assumed a constant depth value during data retrieval.

Turbidity, surface turbulence, viewing angles, and measurement duration effect staff gage measurements (Sauer & Turnipseed, 2010). A field technician reported depths to within ±0.1 cm, or more typically ±0.5 cm (single measurement), using a staff gage graduated in centimeters under ideal conditions. Field technicians typically removed clogs and waited to record depth until the flow returned to a near steady-state. Perceptions of steady-state varied in the field. We omitted field depth measurements that seemed inconsistent with the site history, adjacent measurements, and season. Manual measurements of stage during high flows remain impractical because of rapid changes in discharge over the measurement period. In addition, using standard measurement approaches during high flows exposes personnel and equipment to risks associated with flood waves and debris. For this reason manual observations of stage occurred exclusively during recession periods.

The pressure transducers used in this study included on-board data loggers that obtained time settings from a computer when connected. Computers with incorrectly set clocks introduced timing errors of ±5.0 hours if time zones were set incorrectly or ±12.0 hours if a machine switched from a 12-hour based time to a 24-hour based time. Detecting these errors required comparison to data from other sources such as rainfall data from nearby meteorological stations

operated by the Panama Canal Authority.

3. Methods

3.1 Initial Attempts at Data Correction

We attempted to correct these data using three methods prior to employing the current method based on time series decomposition (TSD). Alternative methods included manual correction, deep learning neural networks and forward-backward moving averages. We desired a correction method suitable for use on a large number of measurements that added minimal uncertainty and avoided inconsistent flow behavior.

3.1.1 Manual Correction

Several published studies to date describe results of analyses and model comparisons using depth data produced by manual correction using a time series data editor (Adamowicz et al., 2019; Cheng, 2018; Crouch, 2012; Litt, 2016; Regina, 2017). These studies used closed-source proprietary software developed for Microsoft Windows called the Time Series Data Editor program produced by Aquaveo LLC as part of the Watershed Modeling System (Aquaveo, 2014). The data editor offered many utilities to manipulate high resolution measurements using a graphical interface.

The data editor provided sufficient data correction tools for manual correction of relatively small datasets. The tools available included barometric pressure corrections, shifting, linear trend adjustment, time-shifting, and smoothing. The interface allowed a user to zoom, select, alter, and delete individual measurements, as well as groups of measurements. Manual correction using the data editor required one to three hours per year of 5-minute depth data depending upon the frequency of perceived errors in the raw data.

3.1.2 Recursive Neural Networks

A previous attempt at automatic correction of erroneous depth data used recursive neural networks (RNN) (Frazier et al., 2017). The RNN model required uncorrected data, rainfall data, and manually corrected data as inputs. Use of RNN resulted in low root mean squared errors (RMSE) with manually processed data and high absolute errors around storm events. Errors increased significantly with a reduction in available manually processed training data.

We attributed the low RMSE achieved using RNN to the low frequency of large corrections, high frequency of small corrections, and limitations of using RMSE to optimize model performance. Many measurements required small corrections, which produced small model errors that became smaller when squared. The range of absolute errors better demonstrated the limitations of using RNN to correct erroneous depth data. The RNN model required substantial amounts of manually processed data as input. We found the need for manually processed input ran contrary to the motivations of using an automated correction method.

3.1.3 Event Detection using Forward-Backward Moving Averages

Correction of systematic errors in baseflow requires a method to detect and preserve direct runoff

events in a flow record. The forward-backward moving averages method (FBMA) (Maestre et al., 2019) is comparable to the Moving Average Convergence/Divergence method (MACD) used as a stock index to detect changing market trends (Appel, 2005). The MACD uses two forward MA of different window lengths. The FBMA uses a forward and a backward MA that may or may not have different window lengths. We used non-zero differences between two MA to indicate a significant change in depth associated with a storm event called a *delta* (or *histogram* under MACD).

Using FBMA for event detection required more coefficients than TSD. FBMA used a threshold *delta* and the slope of *delta* to detect the beginning and end of individual events. The forward and backward MA window lengths, *delta* and *delta*-slope thresholds required calibration for each time series. The convergence of pre-peak moving averages occurred much slower than pre-peak divergence and resulted in a failure of the method to adequately detect event end-times. The FBMA method resulted in more frequent misidentification of errors as storm events compared to decomposition.

3.2 Automatic Baseflow Correction Using Time Series Decomposition

We assumed streamflow at the catchment outlet consisted of four components: baseflow, daily periodic components, direct runoff, and noise. This multicomponent model of depth is similar to several models of streamflow partitioning reviewed by Nejadhashemi et al. (2003). Baseflow correction isolates and replaces the baseflow component of this model. We used an additive decomposition model (Hyndman & Athanasopoulos, 2018) of flow depth behind a weir (y_t), such that at time t:

$$y_t = B_t + E_t + R_t + N_t \quad (1)$$

Where: B_t = long term cyclical and seasonal trend including groundwater contributions to streamflow, also called baseflow, E_t = daily periodic components such as evapotranspiration and residual barometric pressure fluctuations, R_t = response of the system due to precipitation including overland flow and subsurface stormflow, also called direct runoff, and N_t = residual components and random noise.

We employed the digital filters described below to approximate and isolate each component of Eqn. 1. We assigned a high priority to retaining as much of the original direct runoff component (R_t) as possible and focused on correction of B_t. The complete method consisted of three high-level steps: event detection, baseflow correction, and hydrograph reconstruction.

3.3 Event Detection

Event detection approximates components of Eqn. 1 to isolate direct runoff (R_t). We modeled erroneous baseflow and daily periodic components using local minimums and an exponentially weighted moving average filter (EWMA). We used an equally weighted moving average filter (MA) and applied a threshold to reduce the noise component. We computed the sum of the direct runoff and noise by manipulating Eqn. 1:

$$R_t + N_t = y_t - B_t - E_t \quad (2)$$

3.3.1 Baseflow component, B_t

Baseflow separation techniques frequently apply digital recursive filters to detect periods of direct runoff. Determining filter coefficients commonly involves manual or automated calibration or uses assumptions or observations of physical aquifer characteristics such as porosity or linearity of baseflow recession (Eckhardt, 2005). The greater relative impact of weir clogs and sensor drift on lower flows precluded extraction of baseflow coefficients from the uncorrected data. We modeled the erroneous baseflow component using a local minimum given by:

$$B_t = max\left(min(Y_{-24}), min(Y_{+24})\right) \quad (3)$$

Where: Y_{-24} = set of recorded depths over the previous 24 hours and Y_{+24} = set of recorded depths over the following 24 hours.

3.3.2 Periodic component, E_t

Daily periodic fluctuations in groundwater due to barometric pressure (Weeks, 1979) and evapotranspiration (Boronina et al., 2005) influence streamflows. This method approximated E_t using an EWMA on a detrended record of water depth. We employed an EWMA given by (Roberts, 1959):

$$E_t = \begin{cases} E_1 = D_1 \\ E_t = \lambda D_t + (1-\lambda) E_{t-1} \end{cases} \quad (4)$$

Where: $D_t = y_t - B_t$, detrended depth record $0 < \lambda \leq 1$, smoothing coefficient

The smoothing coefficient (λ) is frequently specified using a time-based span coefficient. The resulting EWMA is referred to as a N-day or N-*time* EWMA. The span coefficient should be large enough to smooth out sub-daily noise, but small enough to approximate daily periodicity. The span term s is related to the smoothing coefficient λ by:

$$\lambda = \frac{2}{s+1}, s \geq 1 \quad (5)$$

3.3.3 Noise, N_t

Using Eqn. 2, Eqn. 3, and Eqn. 4 resulted in a record of direct runoff (D_t) that included multiple small fluctuations associated with measurement noise. We used MA and a minimum threshold value to filter out the noise component (Natrella & others, 2013). Fig. 3 provides a graphical example of the entire decomposition process. We scaled the size of the MA window and threshold value to the highest depth measurements with consideration for instrument noise. Given a window of size n, a first approximation of direct runoff R_t' has $n - 1$ initial values computed by:

$$R_{t+n-1}' = \frac{D_t + D_{t-1} + \cdots + D_{t-n+1}}{n} \tag{6}$$

The resulting series R_t' consequently loses the first $n - 1$ measurements. The following values of R_t' are computed recursively by:

$$R_t' = R_{t-1}' + \frac{1}{n}(D_t - D_{t-n}) \tag{7}$$

The type of simple moving average demonstrated by Eqn. 6 and Eqn. 7 tended to impart a phase distortion on the input data. Using a moving average filter on flow data advanced event start and end times forward in time. We found the resulting lag produced more accurate event start and end times on "flashy" flow records that exhibited shorter duration rises and longer recessions.

We generated the final approximation of R_t by applying a threshold:

$$R_t = \begin{cases} 0, & R_t' < \beta\left(max(R_t')\right) \\ R_t', & R_t' \geq \beta\left(max(R_t')\right) \end{cases} \tag{8}$$

Where: $0 > \beta \leq 1$, is the allowable noise threshold.

Eqn. 8 used the maximum of R_t', but lower quantile measurements could be employed to reduce the influence of outliers on the noise threshold.

3.3.4 Event refinement

With reference to Fig. 3 the non-zero values of "Direct runoff" (R_t) corresponded to event measurements in the original depth record y_t. The preceding steps focused on differentiating fluctuations in depth due to precipitation from those attributable to debris or sensor drift. This method worked independent of precipitation measurements. We enforced more strict limits on event detection by eliminating (setting $R_t = 0$) consecutive event points whose total duration was less than a minimum storm length and merging sufficiently close event points into single

compound events. We refined event start times by adjusting each event start to a local minimum. Fig. 4 demonstrates the results of event detection.

3.4 Baseflow Correction

Baseflow correction assumed an erroneous baseflow component B_t in Eqn. 1. The corrected B_t chosen minimally reflected basic assumptions about volumetric discharge at the catchment scale. We generated B_t using four assumptions:

1 Independent field observations of depth reflected measurements obtained during inter-event ($R_t = 0$) flows
2 Baseflow increased during storm events
3 Baseflow receded during dry periods and never reached zero flow
4 Events separated by a short amount of time exhibit similar characteristics

Assumption (1) depends on the frequency and timing of data collection. Assumption (2) applies to streams fed by shallow aquifers in humid regions (Scanlon et al., 2002). Assumption (3) applies to unmanaged perennial streams following normal rainfall conditions. Assumption (4) applied if catchment characteristics did not change quickly between events. We split baseflow correction into *event* baseflow correction and *inter-event* baseflow correction under these assumptions.

We adopted models of baseflow increase or recession consistent with the above assumptions and corrected outlying or non-physical event coefficients. The method replaced event model coefficients in the 2.5th percentile and coefficients outside the 97.5th percentile with interpolated coefficients from adjacent events. We associated negative, extremely high or low coefficients with possible sensor drift or debris. The method also smoothed event coefficients using a 5-event window.

We applied field-observed depth measurements in either a forward-looking or backward-looking fashion during preprocessing. Use of a forward-looking depth conversion frequently resulted in a need to raise depths after the occurrence of systematic errors. Use of a backward-looking depth conversion frequently required a decrease in depths before the occurrence of systematic errors.

3.4.1 Post-event Baseflow Correction

Cross-referencing the non-zero values of direct runoff (R_t) with total depth (y_t) resulted in a series of storm-event depth measurements. Collating consecutive event measurements produced a record of discrete storm events. We assumed the final depth at cessation of direct runoff was greater than the initial event depth. We used a first-order approximation of baseflow increase during storm events to model greater increases in baseflow due to longer events:

$$B_f = B_o + md \tag{9}$$

Where: B_f = flow depth measurement at event end (L), B_o = flow depth measurement at event start (L), m = event baseflow increase coefficient (LT^{-1}), and d = event duration (T)

We used Eqn. 9 to derive baseflow increase coefficients (m) for all events. The method replaced negative, extremely low, and extremely high coefficients with interpolated values from adjacent events. We used the corrected coefficients and Eqn. 9 to compute a final baseflow value (B_f) for each event requiring correction. The method offset all post-peak flow depths to match the new B_f.

3.4.2 Inter-event Baseflow Correction

Comparing the zero values of direct runoff (R_t) with total depth (y_t) produced a series of *inter-event* depth measurements. We collated consecutive inter-event measurements to produce a record of individual recessions. We assumed flow decreased exponentially during recessions:

$$B_f = B_o e^{-bd} \tag{10}$$

Where:

B_f = flow depth measurement at end of recession (L), B_o = flow depth measurement at start of recession (L), b = baseflow recession coefficient (T^{-1}), and d = recession duration (T).

We used Eqn. 10 to derive baseflow decrease coefficients (b) for each recession. We inferred values of B_f and B_o from the corrected storm event depths that flanked each recession. The method used the B_f of the preceding event as the B_o of the recession and the B_o of the following event as the B_f. Fig. 5 shows an example baseflow recession correction.

3.5 Hydrograph Reconstruction

Hydrograph reconstruction combines the modeled baseflow with the original components of total depth to produce a new corrected hydrograph with the effects of debris and sensor drift removed. Hydrograph reconstruction consisted of four steps:

1. Approximate the seasonal component of B_t
2. Superimpose modeled event baseflow on seasonal baseflow
3. Connect event baseflow rises with baseflow recessions
4. Add remaining components of y_t to new B_t

We approximated the seasonal component of B_t by using independent field observations of depth. The method linearly interpolated between field measurements, then smoothed the interpolations using a 28-day EWMA.

We assumed inter-event baseflow recessions trended toward the seasonal component of baseflow. The method incorporated baseflow rises by using Eqn. 9 and assumed B_o is equal to the seasonal component at the event start. The method used Eqn. 10 to model inter-event depths between events. The method constructed a final corrected record of flow depths by subtracting Eqn. 3 from y_t and adding the new modeled baseflow. Fig. 6 shows an example of raw, manually corrected, and TSD corrected time series.

3.6 Discharge and Runoff

Processed stage data are often used to compute volumetric discharge and areal average catchment runoff. Characterizing the potential effects of the automatic method on hydrological research conclusions required computation of stage-derived values. We converted depth measurements to volumetric flow rates using a weir rating curve (Clemmens et al., 2001) accounting for whether or not the weir was full of sediment on the upstream side (Ogden et al., 2017). We limited measurement precision for discharge to 0.0001 m^3s^{-1} (0.1 Ls^{-1}). Changes in depth that resulted in changes in discharge less than measurement precision resulted in no perceptible change in discharge from uncorrected to corrected data.

We produced total runoff flow duration curves (Bedient et al., 2008) assuming a catchment area of 144 ha. We expected to see patterns in total runoff consistent with streams in Central America and the Caribbean region where flows follow seasonal rainfall (Dettinger & Diaz, 2000). We also derived flashiness indices (Richards, 1990) from the flow duration curves. Finally, event-based analyses require some form of baseflow separation. We computed baseflow separation coefficients (Collischonn & Fan, 2013; Eckhardt, 2005, 2008) and computed annual storm runoff efficiencies (Ogden et al., 2013) for automatic and manually corrected depth data to compare the effects of the two methods on storm event analyses.

4. Results

4.1 Water Level Adjustment

The automatic correction method constructed the final time series shown in Fig. 6 using corrected coefficients and estimates of seasonal baseflow. This particular depth record originally contained many negative or erroneously low values. Both manual and automatic methods increased average ($\bar{x}$) and median (*Md*) depth measurements. The automated method increased average depth by 11.3 mm (+8.1%) and median depth by 4.6 mm (+3.6%). Manual correction increased average depth by 9.5 mm (+6.8%) and median depth by 4.2 mm (+3.3%).

Sub-hourly noise resulted in numerous false positives during event identification. Event detection using an hourly EWMA significantly reduced false positives. Automated correction identified numerous erroneous events and baseflow recessions with negative recession coefficients. The method enforced a minimum baseflow rise on identified events and computed new recession coefficients using adjacent storm events and seasonal baseflow. Enforcing limits on acceptable coefficients resulted in a more narrow range of baseflow coefficients. Events or recessions that had negative coefficients prior to correction all had the same minimum coefficient after correction. The method also smoothed baseflow coefficients across events resulting in more gradual changes in coefficients over time. Fig. 7 shows distributions of modeled baseflow coefficients derived from Eqn. 9 and Eqn. 10.

Fig. 8 shows distributions of all corrections for both manual and automatically corrected depth data. The manual method resulted in more conservative corrections with a bias for particular adjustments. Manual correction resulted in many corrections of 5 mm or factors of 5 mm. The automated method performed more corrections across a broader range of values than manual correction. Furthermore, the distribution in differences resulting from the TSD method shown in

Fig. 8a are less skewed and more normally distributed than those differences from manual correction shown in Fig. 8b.

Both methods resulted in values of r^2, corrected regressed uncorrected, greater than 90%. The TSD approach resulted in data with a lower r^2 (0.924) compared to the manual method (0.965).

Manual correction resulted in perfect compliance with field observations of depth. Automatic correction treated field observations as fitted points for the seasonal component of baseflow. Automatically corrected measurements regressed field observations resulted in a r^2 of 0.966.

4.2 Discharge and Runoff Estimation

Fig. 9 shows differences in uncorrected discharge for different ranges of corrected volumetric flows. The automated method imposed similar corrections on flows less than 10 Ls^{-1} compared to manual correction. Automatic correction produced more diverse and sometimes greater adjustments on final discharges greater than 10 Ls^{-1} compared to the manual method. Approximately 85% of all automatic and manual corrections occurred on discharges less than 100 Ls^{-1}. Ephemeral clogs due to woody debris became less likely with increasing discharges.

Fig. 10 shows resulting total runoff flow duration curves assuming a catchment area of 144 ha. Both manual and automatic correction resulted in total runoff flow duration curves that were consistent with streams in Central America and the Caribbean region. Manual correction resulted in total runoff of 1409.9 mm. Automatic correction correction resulted in total runoff of 1400.2 mm or a decrease of -0.69% compared to manual correction.

Tbl. 1 shows flashiness indices for automatically corrected, manually corrected, and historical data from the period 18 Jan 2009 to 16 Mar 2016 (Regina, 2017). A technician manually corrected the historical data using a time series data editor. Flashiness indices for manually and automatically corrected data followed similar patterns of increase or decrease when compared to historically averaged indices with the exception of $F_{10/90}$. $F_{10/90}$ which is a ratio of extreme flows near the ends of a flow duration curve. The increases in low flows and decreases in high flows imparted by automatic correction resulted in a lower $F_{10/90}$ compared to manually corrected data.

The results of recession analysis and baseflow separation are listed in Tbl. 2. Both automatic and manually corrected data resulted in baseflow coefficients similar to those derived from historical data. Manual correction resulted in an annual storm runoff efficiency of 10.9%. Automatic correction resulted in an annual storm runoff efficiency of 10.6%. Both automatic and manual correction produced annual storm runoff efficiencies within the range of historical runoff efficiencies.

5. Discussion

Corrections occurred most frequently during the December-May Panamanian dry season. Systematic errors during the dry season exhibited larger relative impacts on depth measurements and resulting discharge measurements due to lower total flows (Fig. 9). These errors are particularly problematic for studies of land use and land cover effects on dry season base flow

aimed at studies of the so-called "forest sponge effect" (Bruijnzeel, 2004; Ogden et al., 2013; Liu et al., 2015; Peña-arancibia et al., 2019). Data correction using manual stage measurements are essential in this case because small discharge measurement errors may represent a significant fraction of small flows and may confound such studies. The results shown in Fig. 10 indicate that above a probability of exceedance of approximately 0.8, data correction was essential. Fig. 10 also shows that manual and TSD data correction yielded similar results, and that both are far superior to using raw data without adjustment against field observations. However, manual correction can introduce unquantifiable or unknown biases into the flow record.

Automatic correction generated smoother and more continuous recessions compared to a single revision of manually corrected data. We optimized this TSD implementation on the falling limb and during periods of low flows associated with hydrograph recession on data from V-notch weirs, which are self cleaning during the rising limb of the hydrograph. The models of flow separation used by this TSD method may not work in natural channels or other types of flow measurement structures that do not tend to shed debris on the rising limb of the hydrograph.

In seasonal climates, base flows tend towards meta-stability depending on seasonal variation in time between storms and mean catchment average precipitation per storm (Cheng et al., 2020). This gives rise to a seasonal-dependence of the differences between TSD and MAN that appear in Fig. 10 between exceedance probabilities of 0.2 to 0.4 and again from 0.6 to 0.8.

Manual correction of automatically-collected stage data is a time consuming process. Large field projects such as the Agua Salud project can collect data at a rate that overwhelms project staff tasked with manual data correction. A technician with a graduate degree in hydrologic science and several years experience performed manual correction of one year of data using a time series data editor in approximately two hours. These manually corrected data represented a single revision. Production of final quality depth data frequently require two to four additional revisions. The automated TSD method completed correction of ~105,000 measurements using a 4-core/8-thread consumer processor (AMD 2500U) in approximately 1 minute, averaged across three executions.

Our implementation of TSD included non-volatile memory express storage (NVME), a RAM-based temporary file system, and multi-threaded optimizations from the NumPy (Van Der Walt et al., 2011) and Pandas (McKinney & others, 2010) Python libraries. The methods described in this paper allowed correction of over 17.5 million weir stage measurements from 13 experimental catchments. Both original and processed data series are available online as discussed in Regina et al. (2021).

6. Conclusions

Streamflow data correction is essential, especially on low flows where small discharge measurement errors may represent a significant fraction of flows. These measurement errors have implications on low flow focused studies, including the so called "forest sponge effect" hypothesis. We developed an automated correction method using time series decomposition (TSD) to correct systematic measurement errors of water depth behind a weir due to debris and sensor drift. The TSD method applied objective, predictable and repeatable transformations across all measurements (Fig. 6). Manual correction resulted in subjective corrections that are

difficult to duplicate, essentially non-repeatable. Automatic correction employed the same basic assumptions about flow behavior as those used by a trained technician during manual correction.

We found automatic correction using TSD provided a more robust and efficient method to process streamflow measurements in a repeatable way when compared to manual processing and did not significantly alter annual mass balances and hydrological coefficients. Compared to manual processing, the differences between unprocessed and processed data are less skewed and approximately normally distributed, a distinct improvement over the manual processing which tended to increase stage values in fixed increments (Figs. 8a and 8b). Resulting changes to flow duration curves were small (Fig. 10), as were changes to commonly used flashiness indices (Tbl. 2).

Automatic correction of depth measurements using TSD may be applicable to other records that can be decomposed into predictable components. Appropriate replacement components should be informed by site specific considerations and make minimal assumptions about underlying characteristics. Alternative versions of the method might use different assumptions about baseflow behavior, include more or higher order components, or use a different decomposition model.

Acknowledgements

This study was made possible by the U.S. National Science Foundation through Grant EAR-1360384. The first author was supported during part of this study by the University Corporation for Atmospheric Research through a cooperative agreement with the National Oceanic and Atmospheric Administration.

Data and code availability: The source code to duplicate these results is available at https://gitlab.com/jarq6c/unclog. The data required to duplicate this study are available at https://www.hydroshare.org/resource/693ed96d8c85447b9a62c29c5751a4ff/

References

Adamowicz, W., Calderon-Etter, L., Entem, A., Fenichel, E. P., Hall, J. S., Lloyd-Smith, P., Ogden, F. L., Regina, J. A., Rad, M. R., & Stallard, R. F. (2019). Assessing ecological infrastructure investments. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5254–5261. https://doi.org/10.1073/pnas.1802883116

Appel, G. (2005). *Technical analysis: power tools for active investors*. FT Press.

Aquaveo. (2014). *Watershed Modeling System v9.0.11 (64-bit)*. https://www.aquaveo.com/

Bedient, P. B., Huber, W. C., Vieux, B. E., & others. (2008). *Hydrology and floodplain analysis, 5th edition*. Prentice Hall Upper Saddle River, NJ.

Beven K., Asadullah A., Bates P., et al., (2020). Developing observational methods to drive future hydrological science: Can we make a start as a community? *Hydrological Processes*.

34:868–873. https://doi.org/10.1002/hyp.13622

Beven, K. J., and Lane, S. (2019). Invalidation of models and fitness-for-purpose: A rejectionist approach, Chapter 6. In C. Beisbart, & N. J. Saam (Eds.), Computer simulation validation—Fundamental concepts, methodological frameworks, and philosophical perspectives (pp. 145–171). Cham:Springer.

Boronina, A., Golubev, S., & Balderer, W. (2005). Estimation of actual evapotranspiration from an alluvial aquifer of the Kouris catchment (Cyprus) using continuous streamflow records. *Hydrological Processes: An International Journal*, *19*(20), 4055–4068.

Bos, M. G. (1976). *Discharge measurement structures* (Nos. 4; Vol. 20). Ilri.

Bosch, D., Lowrance, R., Vellidis, G., Sheridan, J., & Williams, R. (2002). Dissolved oxygen and stream flow rates: Implications for TMDL's. *Total Maximum Daily Load (TMDL): Environmental Regulations, Proceedings of 2002 Conference*, 92.

Brakensiek, D. L., Osborn, H. B., and Rawls, W. J. (1979). "Field manual for research in agricultural hydrology." Agriculture handbook 224, USDA, Washington, DC.

Bruijnzeel, L.A. (2004). Hydrological functions of tropical forests: not seeing the soil for the trees? *Agric. Ecosyst. Environ*. 104, 185-228.

Chen, J., Li, W., Lau, A., Cao, J., & Wang, K. (2010). Automated load curve data cleansing in power systems. *IEEE Transactions on Smart Grid*, *1*(2), 213–221. https://doi.org/10.1109/TSG.2010.2053052

Cheng, Y. (2018). *Effects of Preferential Flow Paths and Land Use on Hydrological Behaviors in Tropical Catchments* [PhD thesis]. University of Wyoming; University of Wyoming.

Cheng, Y., Ogden, F.L. and Zhu, J. (2020). Characterization of sudden and sustained base flow jump hydrologic behaviour in the humid seasonal tropics of the Panama Canal Watershed. *Hydrological Processes*, 34(3), pp.569-582.

Clemmens, A. J., Wahl, T. L., Bos, M. G., & Replogle, J. A. (2001). *Water measurement with flumes and weirs No. 58*. International Institute for Land Reclamation and Improvement/ILRI.

Close, E. R., Beard, L., & Dawdy, D. R. (1970). Objective determination of safety factor in reservoir design. *Journal of the Hydraulics Division*, *96*(5), 1167–1177.

Collischonn, W., & Fan, F. M. (2013). Defining parameters for Eckhardt's digital baseflow filter. *Hydrological Processes*, *27*(18), 2614–2622.

Crouch, T. D. (2012). *Quantifying Hydrological Ecosystem Services of Various Land Covers and Uses Within the Panama Canal Watershed* [PhD thesis]. University of Wyoming; University of Wyoming.

Dettinger, M. D., & Diaz, H. F. (2000). Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, *1*(4), 289–310.

Dobriyal, P., Badola, R., Tuboi, C., & Hussain, S. A. (2017). A review of methods for monitoring streamflow for sustainable water resource management. *Applied Water Science*, *7*(6), 2617–2628.

Dorninger, P., Nothegger, C., Pfeifer, N., & Molnár, G. (2008). On-the-job detection and correction of systematic cyclic distance measurement errors of terrestrial laser scanners. *Journal of Applied Geodesy*, *2*(4), 191–204.

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrological Processes*, *19*(2), 507–515. https://doi.org/10.1002/hyp.5675

Eckhardt, K. (2008). A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. *Journal of Hydrology*, *352*(1-2), 168–173. https://doi.org/10.1016/j.jhydrol.2008.01.005

Famili, A., Shen, W.-M., Weber, R., & Simoudis, E. (1997). Data Preprocessing and Intelligent Data Analysis. *Intelligent Data Analysis*, *1*(1), 3–23. https://doi.org/10.3233/IDA-1997-1102

Frazier, N., Ogden, F. L., Regina, J. A., & Cheng, Y. (2017). Validating the Use of Deep Learning Neural Networks for Correction of Large Hydrometric Datasets. *AGU Fall Meeting Abstracts*.

Haddad, K., Rahman, A., Weinmann, P. E., Kuczera, G., & Ball, J. (2010). Streamflow data preparation for regional flood frequency analysis: Lessons from southeast Australia. *Australasian Journal of Water Resources*, *14*(1), 17–32.

Hamilton, A. S., & Moore, R. D. (2012). Quantifying uncertainty in streamflow records. *Canadian Water Resources Journal*, *37*(1), 3–21. https://doi.org/10.4296/cwrj3701865

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice, 2nd edition*. OTexts.com/fpp2; OTexts: Melbourne, Australia.

Litt, G. F. (2016). *Hydrometric, hydrochemical, and hydrogeophysical runoff characterization across multiple land covers in the Agua Salud Project, Panama* [PhD thesis]. University of Wyoming.

Liu, W.F., Wei, X.H., Fan, H.B., Guo, X.M., Liu, Y.Q., Zhang, M.F., Li, Q. (2015). Response of flow regimes to deforestation and reforestation in a rain-dominated large watershed of subtropical China. *Hydrol. Proc.* 29, 5003-5015

Lu, H., Raupach, M. R., McVicar, T. R., & Barrett, D. J. (2003). Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment*, *86*(1), 1–18.

Maestre, A., Regina, J. A., Ogden, F. L., Frazier, N. J., Wolford, R., Kim, J., & Flowers, T. (2019). Runoff Identification and Delimitation in Discharge Series Using Forward and Backward Moving Averages. *AGU Fall Meeting Abstracts*, *2019*, H43I–2143.

Makarenkov, V., Zentilli, P., Kevorkov, D., Gagarin, A., Malo, N., & Nadon, R. (2007). An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics*, *23*(13), 1648–1657.

McEnery, J., Ingram, J., Duan, Q., Adams, T., & Anderson, L. (2005). NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. *Bulletin of the American Meteorological Society*, *86*(3), 375–386.

McKinney, W., & others. (2010). *Data structures for statistical computing in python* (Vol. 445, pp. 51–56). Austin, TX.

Natrella, M., & others. (2013). *NIST/SEMATECH e-handbook of statistical methods*. National Institute of Standards and Technology; http://www.itl.nist.gov/div898/handbook/.

Nejadhashemi, A. P., Shirmohammadi, A., & Montas, H. J. (2003). Evaluation of streamflow partitioning methods. *2003 Asae Annual Meeting*. https://doi.org/10.13031/2013.14955

Oberg, K., & Mueller, D. S. (2007). Validation of streamflow measurements made with acoustic Doppler current profilers. *Journal of Hydraulic Engineering*, *133*(12), 1421–1432.

Ogden, F. L., Creel, J. N., Kempema, E. W., & Crouch, T. D. (2017). Sedimentation Effects on Triangular Short-Crested Flow-Measurement Weirs. *Journal of Hydrologic Engineering*, *22*(8), 4017020.

Ogden, F. L., Crouch, T. D., Stallard, R. F., & Hall, J. S. (2013). Effect of land cover and use on dry season river runoff, runoff efficiency, and peak storm runoff in the seasonal tropics of Central Panama. *Water Resources Research*, *49*(12), 8443–8462.

Peña-arancibia, J. L., Bruijnzeel, L. A., Mulligan, M., & Van Dijk, A. I. J. M. (2019). Forests as 'sponges' and'pumps': Assessing the impact of deforestation on dry-season flows across the tropics. *Journal of Hydrology*, 574, 946-963. https://doi.org/10.1016/j.jhydrol.2019.04.0

Pereira, M. V. F., Oliveira, G. C., Costa, C. C. G., & Kelman, J. (1984). Stochastic streamflow models for hydroelectric systems. *Water Resources Research*, *20*(3), 379–390.

Petersen-Øverleir, A., Soot, A., & Reitan, T. (2009). Bayesian rating curve inference as a streamflow data quality assessment tool. *Water Resources Management*, *23*(9), 1835–1842.

Pfister, M., Schleiss, A. J., & Tullis, B. (2013). Effect of driftwood on hydraulic head of Piano Key weirs. *Labyrinth and Piano Key Weirs II*, 255–264.

Rantz, S. E. (1982). *Measurement and computation of streamflow* (Vol. 2175). US Department of the Interior, Geological Survey.

Regina, J. A. (2017). *Effect of Non-native Land Covers on Runoff Efficiency and Flows in the Humid Tropics in the Panama Canal Watershed* [PhD thesis]. University of Wyoming; University of Wyoming.

Regina, J. A., Ogden, F. L., Hall, J. S., & Stallard, R. F. (2020). *Agua Salud Hydrometric Data*. HydroShare. https://doi.org/10.4211/hs.0daa4ade63ec4ff3a3aec70ddad0a93b

Regina, J.A., Ogden, F.L., Hall, J.S., and Stallard, R.F. (2021) . Agua Salud Project Experimental Catchments Hydrometric Data, Panama, Hydrol. Proc., DOI: 10.1002/hyp.14359

Richards, R. P. (1990). Measures of flow variability and a new flow-based classification of Great Lakes tributaries. *Journal of Great Lakes Research*, *16*(1), 53–70.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, *1*(3), 239–250.

Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., & Wagenbrenner, J. W. (2021). How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. *Hydrological Processes*, 35: e14199. https://doi. org/10.1002/hyp.14199

Sauer, V. B., & Turnipseed, D. P. (2010). *Stage measurement at gaging stations* (Nos. A7; Vol.

3). U.S. Geological Survey; US Department of the Interior, US Geological Survey. http://pubs.usgs.gov/tm/tm3-a7/

Scanlon, B. R., Healy, R. W., & Cook, P. G. (2002). Choosing appropriate techniques for quantifying groundwater recharge. *Hydrogeology Journal*, *10*(1), 18–39.

Stallard, R. F., Ogden, F. L., Elsenbeer, H., & Hall, J. (2010). Panama canal watershed experiment: Agua Salud project. *Water Resources IMPACT*, *12*(4), 18–20.

United States Bureau of Reclamation. (2001). Water Measurement Manual, 3rd Edition Revised Reprinted. *Technical Report, Water Resources Research Laboratory, US Department of the Interior*, 317. https://www.usbr.gov/pmts/hydraulics_lab/pubs/wmm/index.htm

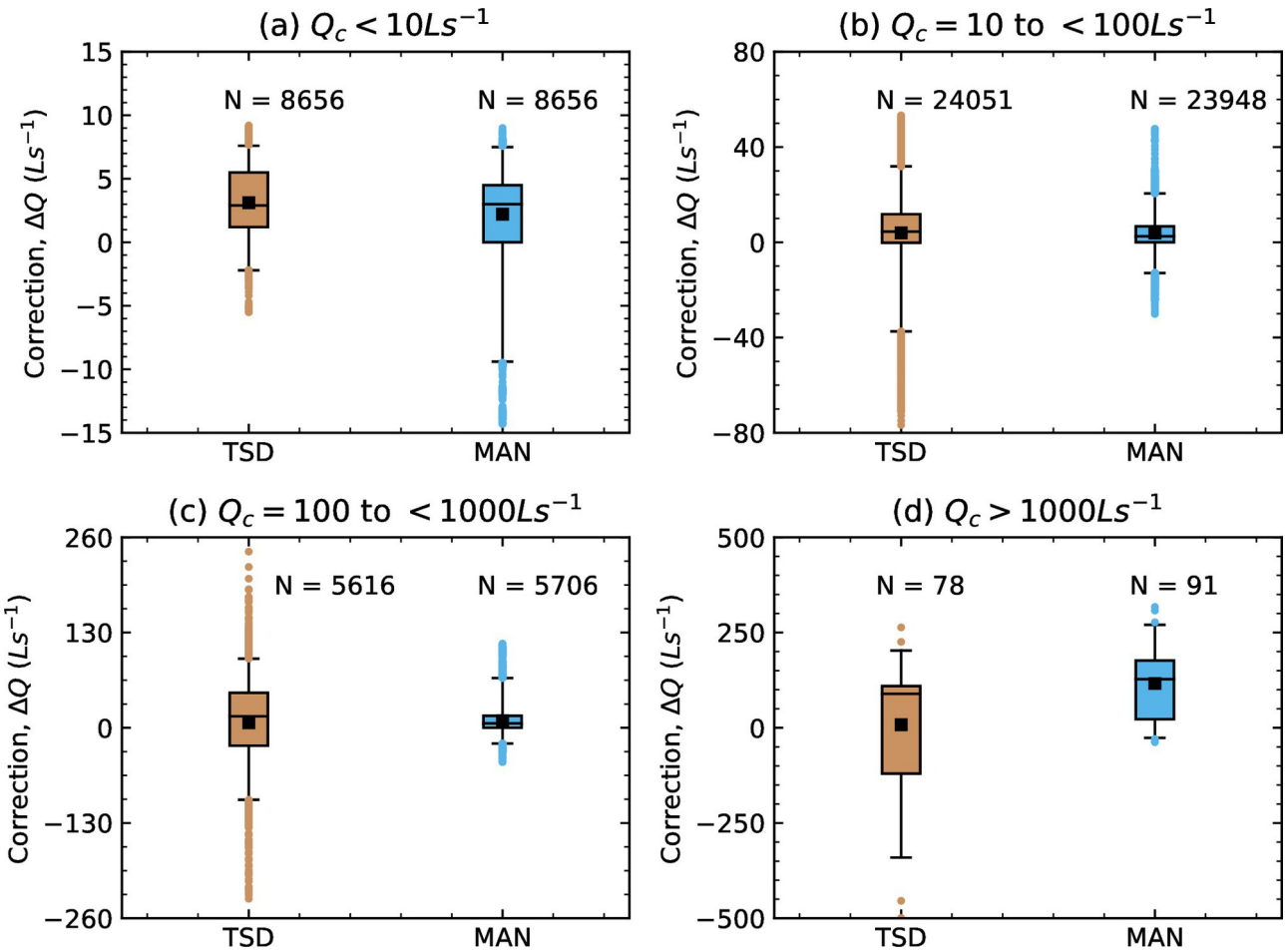
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). *The NumPy array: a structure for efficient numerical computation* (Nos. 2; Vol. 13, p. 22). IEEE Computer Society.

Weeks, E. P. (1979). Barometric fluctuations in wells tapping deep unconfined aquifers. *Water Resources Research*, *15*(5), 1167–1176.

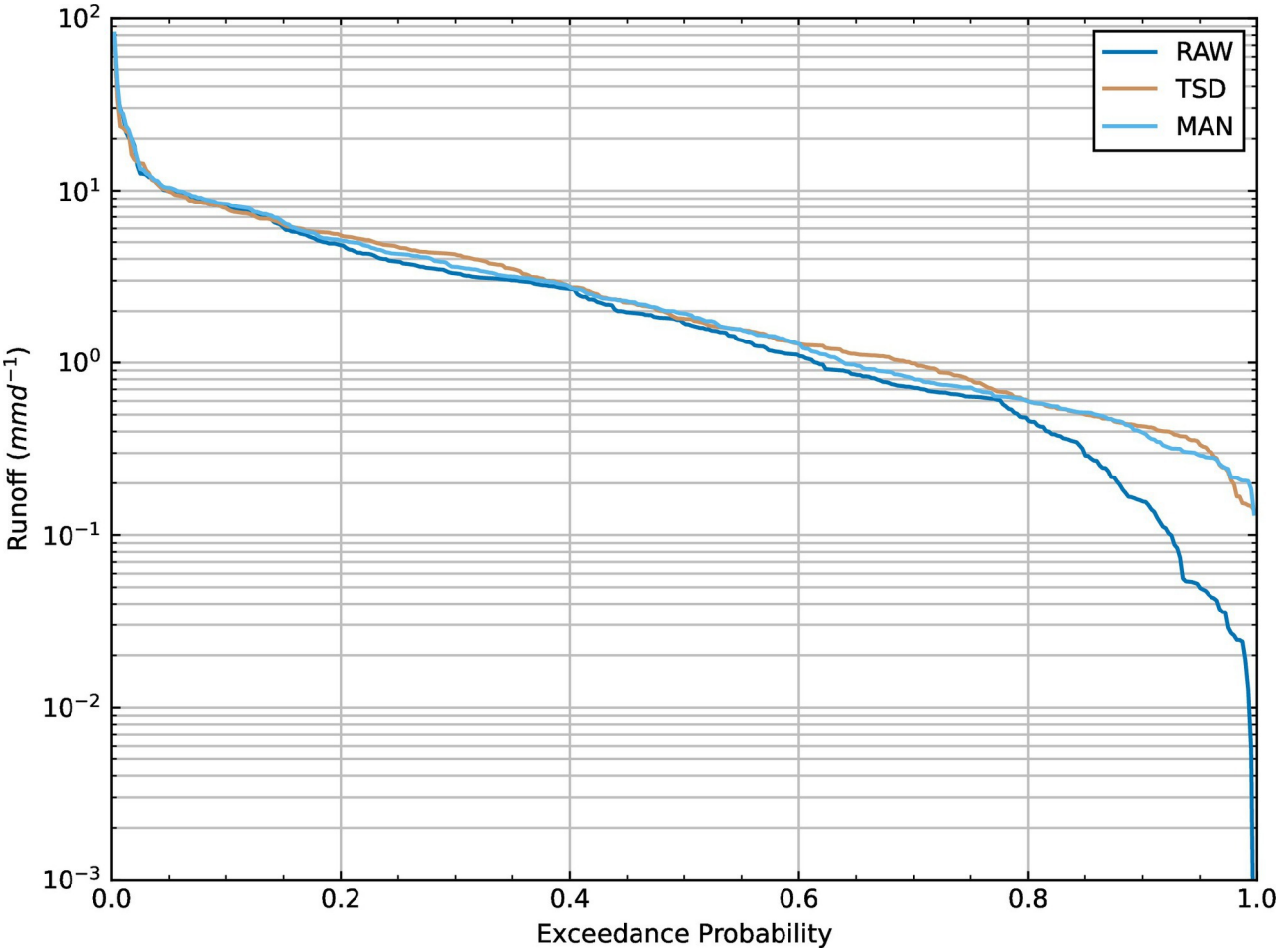
Westerberg, I.K., Sikorska-Senoner, A.E., Viviroli, D., Vis, M. and Seibert, J., 2020. Hydrological model calibration with uncertain discharge data. *Hydrological Sciences Journal*, pp.1-16. Doi:10.1080/02626667.2020.1735638.

Wurbs, R. A., & Bria Wallsn, W. (1989). Water rights modeling and analysis. *Journal of Water Resources Planning and Management*, *115*(4), 416–430.

Yu, Y., Zhu, Y., Li, S., & Wan, D. (2014). Time series outlier detection based on sliding window prediction. *Mathematical Problems in Engineering*, *2014*.

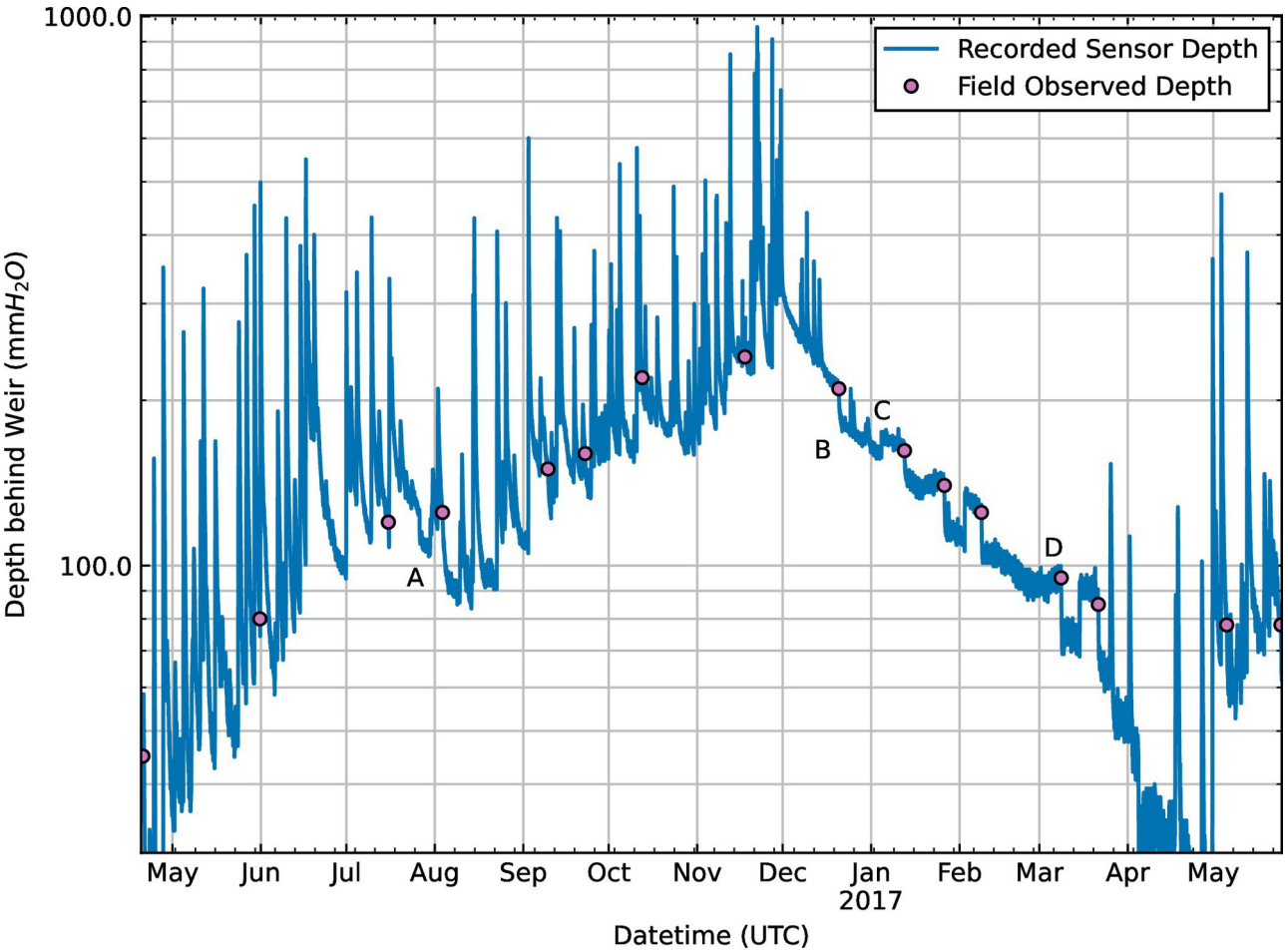


hyp_14405_delta_q.eps

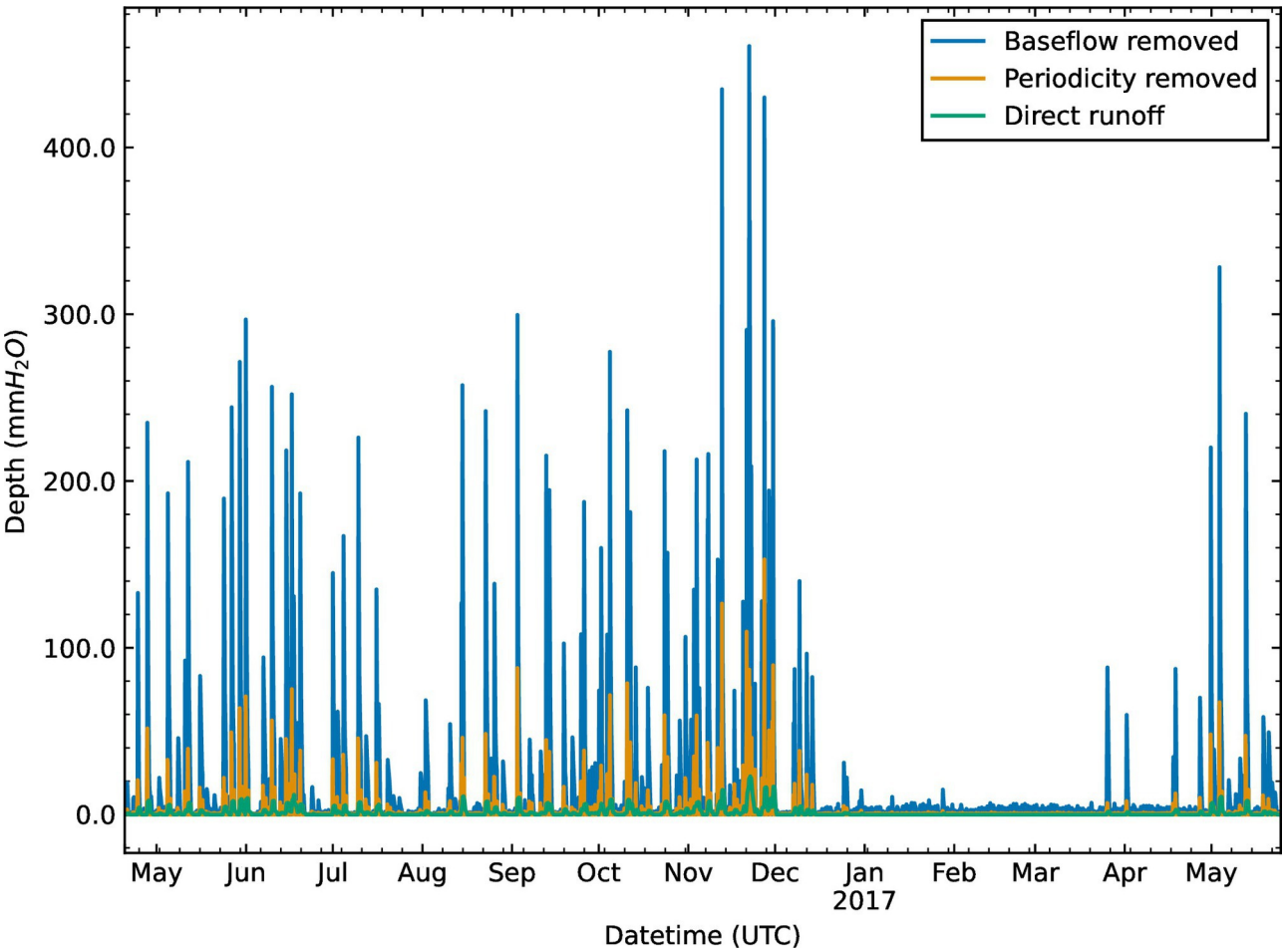


hyp_14405_fdc.eps

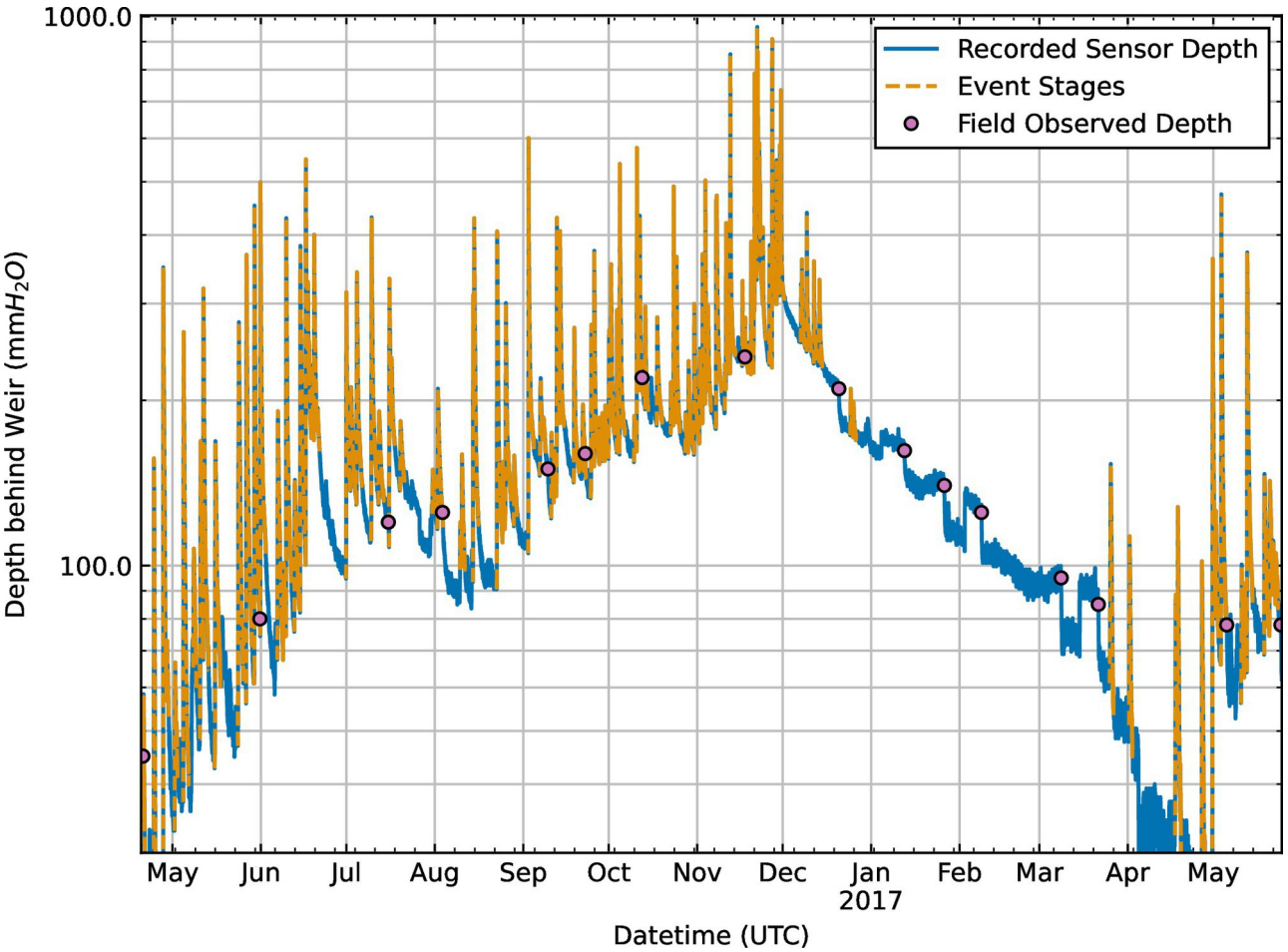
HYP_14405_Figure_1.tif



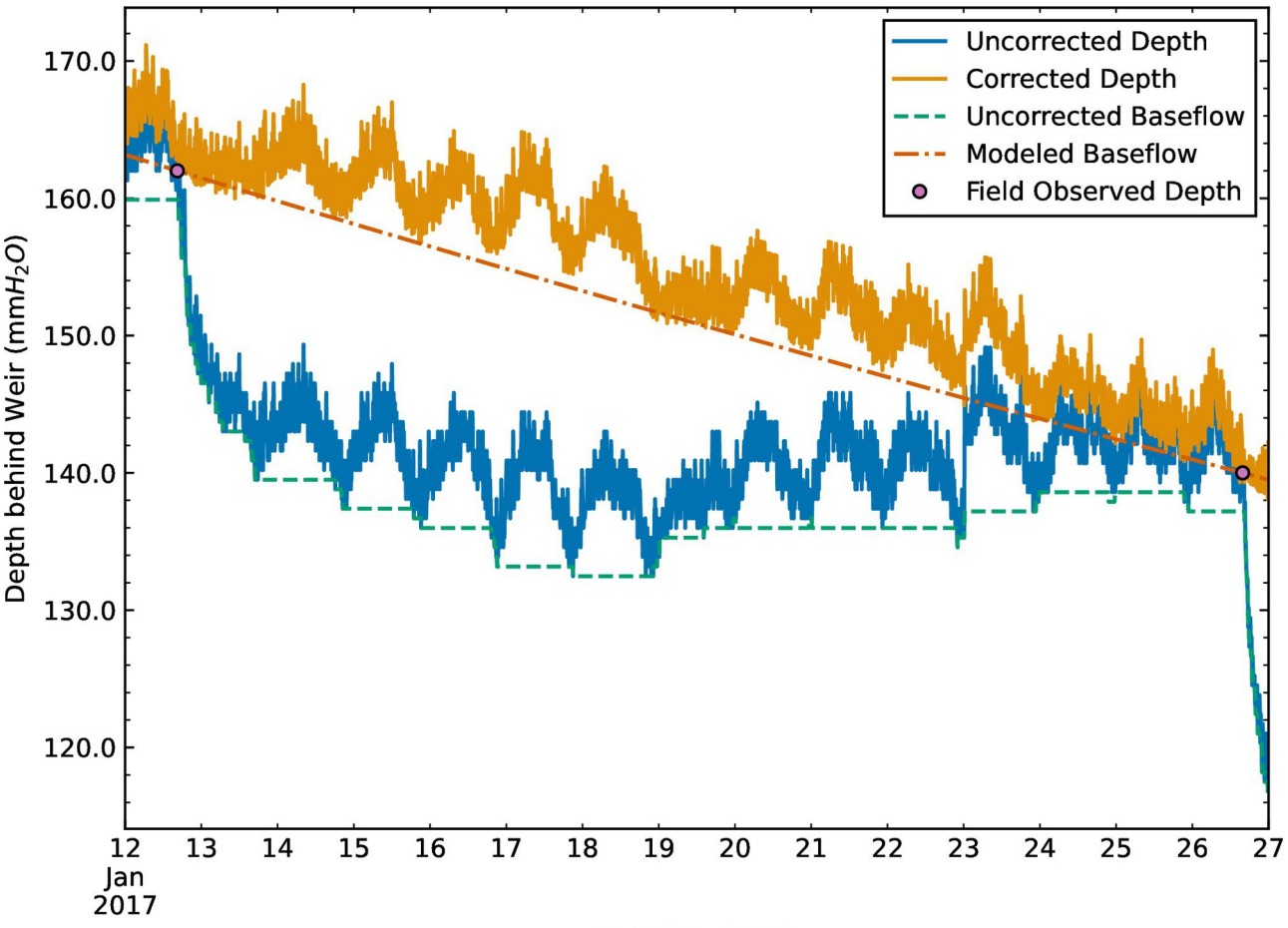
hyp_14405_figure_2.eps



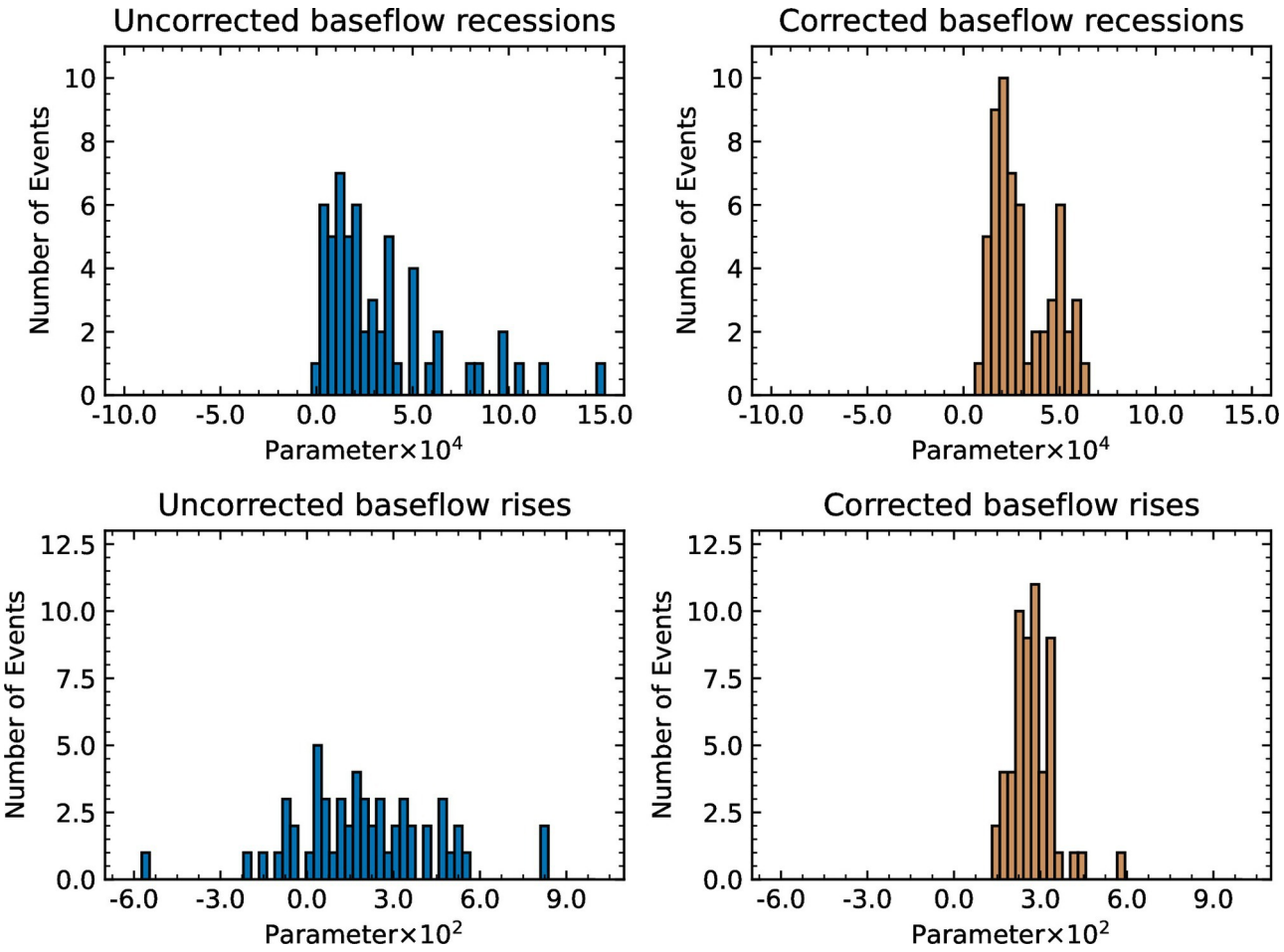
hyp_14405_figure_7.eps



hyp_14405_figure_8.eps



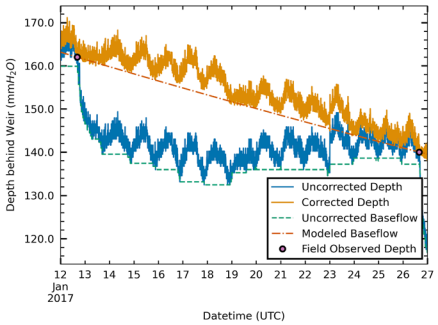
hyp_14405_figure_10.eps



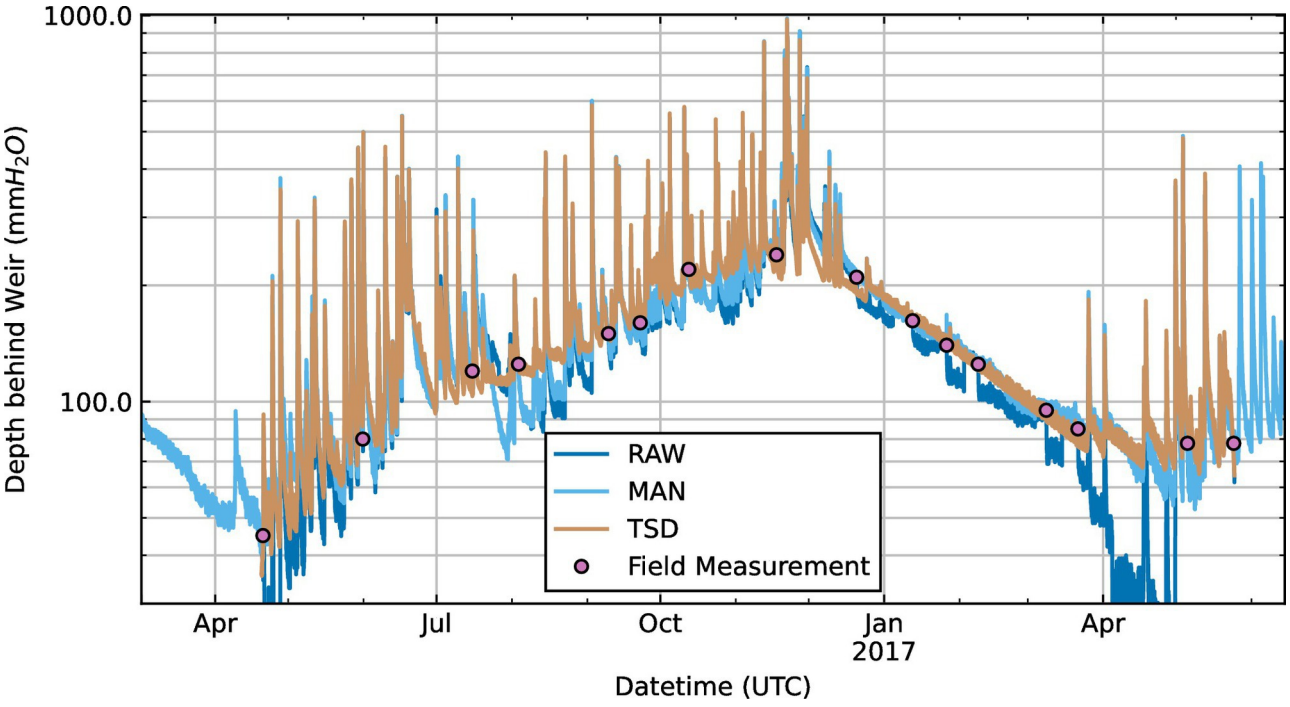
hyp_14405_figure_15.eps

Automated Correction of Systematic Errors in High Frequency Depth Data from V-Notch Weirs using Time Series Decomposition

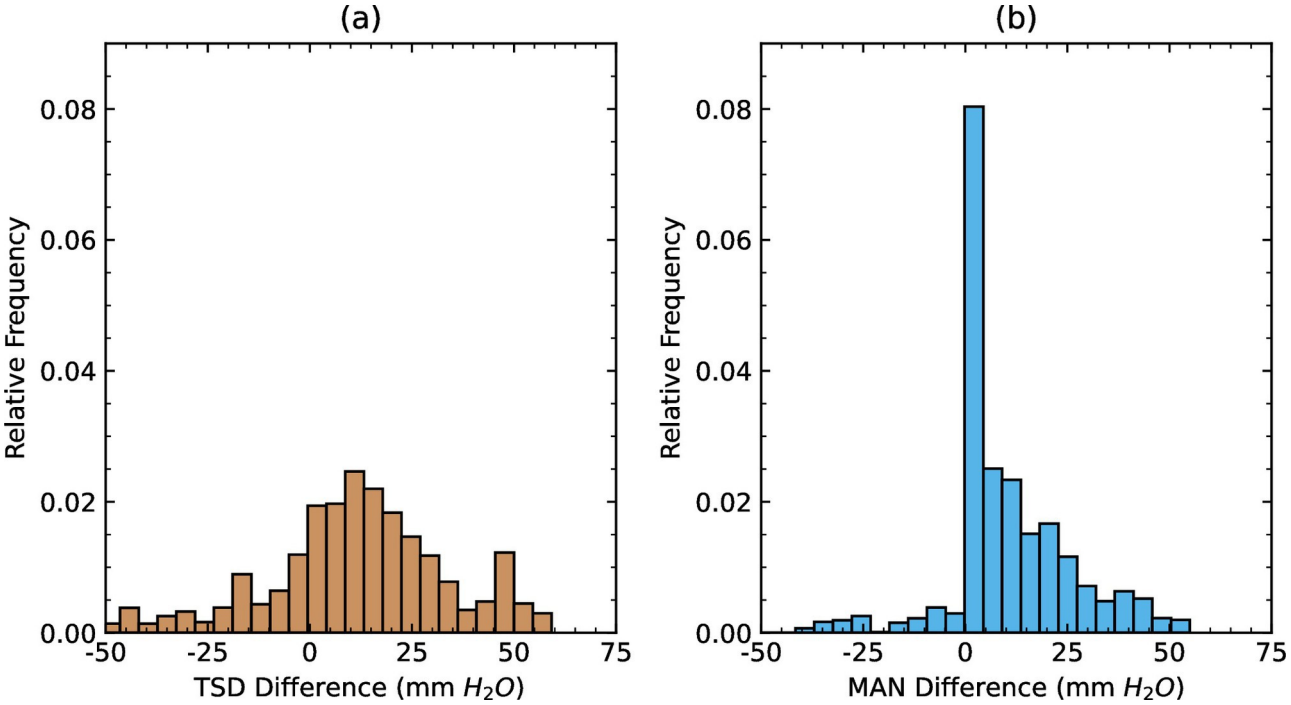
Jason A. Regina* & Fred L. Ogden



Demonstrate a method to automatically correct systematic errors due to woody debris and sensor drift in depth time series, while isolating and preserving periods of direct runoff during storm events.



hyp_14405_manned_corrected_data.eps



hyp_14405_residuals_hist.eps

Table 1: Runoff flashiness indices derived from automatically corrected (TSD), manually corrected (MAN), and historical (HIS) stage data.

Index	**TSD**	**MAN**	**HIS**
$F_{10/90}$	17.96	20.88	20.1
$F_{20/80}$	8.93	8.47	6.9
$F_{25/75}$	5.84	5.94	5.2
$F_{.5}$	2.14	1.83	1.5
$F_{.6}$	2.67	2.32	1.8
$F_{.8}$	4.04	4.11	3.5
CVLF5	0.90	0.96	1.56

Table 2: Baseflow recession coefficients and annual storm runoff efficiencies derived from runoff data corrected using time series decomposition (TSD), manual correction (MAN), and manually corrected historical data (HIS).

Statistic	**TSD**	**MAN**	**HIS**
Number of recessions (-)	127	141	757
Recession constant (-)	0.98	0.98	0.97
Maximum baseflow index (-)	0.81	0.81	0.84
Annual storm runoff efficiency (%)	10.6	10.9	8.45 to 11.1