Title
Resolving hydrologic water balances through a novel error analysis approach, with application to the Tahoe basin

Permalink
https://escholarship.org/uc/item/4570x1c8

Journal
Journal of Hydrology, 546(J. Hydrol. 237 2000)

ISSN
0022-1694

Authors
Trask, JC
Fogg, GE
Puente, CE

Publication Date
2017-03-01

DOI
10.1016/j.jhydrol.2016.12.029

Peer reviewed
Resolving hydrologic water balances through a novel error analysis approach, with application to the Tahoe basin

James C. Trask *, Graham E. Fogg, Carlos E. Puente

Hydrology Program (c/o Graham Fogg), Department of Land, Air, & Water Resources, University of California, Davis, CA 95616, USA

A R T I C L E   I N F O
Article history:
Received 1 August 2016
Received in revised form 14 December 2016
Accepted 16 December 2016
Available online 21 December 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Water balance
Tahoe
Error analysis
Uncertainty
Inter-annual variation
Watershed memory

A B S T R A C T
We introduce a new approach for improving estimates of water balance components, applicable to a multi-period water balance series for a lake, watershed, or other area of any size. It consists of making use of statistical relationships between a component series estimate and the residual errors of the water mass balance series. Through this approach, two novel specific techniques are developed. The first, ‘precipitation-decorrelation’, entails implementing a criterion of decorrelation of residual errors from precipitation estimates. The second, ‘residual-redistribution’, consists of redistributing each residual error over initial water balance component estimates, in accord with an error minimization criterion for each component series. Efficacy is tested using series of annual water balances for the Tahoe Basin. Upon implementation of precipitation-decorrelation, a tightly bounded statistical estimate of mean annual Lake Tahoe evaporation is obtained, which closely matches independent measurement-based estimates. Residual-redistribution yields revised estimates of annual series of Tahoe areal precipitation and watershed runoff, which are each shown to have substantially reduced random error variance. Highly precise revised estimates of inter-annual variations in Lake Tahoe precipitation have enabled resolution of the watershed multi-year ‘memory’ of precipitation, and more reliable separation of inter-annual changes in watershed storage from inter-annual variations in atmospheric loss.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In water balances of lakes and watersheds, large uncertainty is usually associated with the estimated magnitude of one or more components. For instance, groundwater flow and evaporation are often difficult to estimate reliably and with high accuracy, even with extensive and intensive field investigations (e.g., Winter, 1981; LaBaugh et al., 1997). More accurate water balance component estimates improve water mass accounting, and thus may also improve coupled energy, geochemical, and biochemical balances, and help to resolve effects of climate or land cover changes on such balances.

A condition of water mass balance closure (i.e., mass conservation) is often used to estimate the magnitude of a water balance component for which there is not a reliable measurement-based estimate (e.g., Schanen et al., 1998; Ruud et al., 2004). Otherwise, the sum of estimates of all water balance inputs generally does not exactly match that of the outputs (and storage changes); and the discrepancy defines the residual error of the water mass balance. The magnitude of this residual error can be used as a gage of the accuracy of input and output component estimates (e.g., Flерchinger and Cooley, 2000).

In this article, we address whether it is possible to make more effective use of residual errors in multi-period water balance series. The new techniques introduced are presented in a context of retrospective water balance models. Related methods have been developed for predictive rainfall-runoff models. During calibration of predictive rainfall-runoff models, including routing models that conserve the water mass balance, typically the model parameters and sometimes also the state (storage) variables are adjusted with the aim of minimizing the variance of the output (streamflow) prediction error (e.g., Moradkhani et al., 2005). Although input estimates such as precipitation have not typically been so modified, recent developments in particle filtering techniques have used prediction errors in streamflow output to inform and revise prior estimates of areal precipitation inputs (e.g., Weerts and El Serafy, 2006; Salamon and Feyen, 2009).

Although the spirit of the adjustments is similar, the techniques proposed in this article are simpler and more accessible, not requiring user familiarity with stochastic methods, and can be readily applied to precipitation and other water balance inputs (e.g., runoff into a lake for a lake water balance), as well as output...
components. Updates of evapotranspiration (ET) and routing parameters and state variables performed by predictive rainfall-runoff models can be used to revise prior estimates of ET and storage changes; however the updates are applied to parameters or state variables, not directly to the numerical component estimates as is done by the new techniques in this article. Predictive rainfall-runoff models typically use hourly or daily data such that very large data series are available to facilitate effective calibration of routing and state parameters and to improve their associated histograms. By contrast, the techniques in this article can be fruitfully applied using a small number (as few as \( \sim 10 \)) of time steps, and thus are also applicable to large time-step (e.g. annual) water balance series.

The main innovation in this article, for which we found no precedent in the literature for retrospective water balances, is the approach of identifying and using statistical relationships between numerical estimates of one or more component series and values of the consequent residual errors of the water balance series, with the aim of generating improved revised estimates of one or more component series. Specifically, two novel techniques, herein named (a) ‘precipitation-decorrelation’ and (b) ‘residual-redistribution’ are introduced. While (a) entails balancing the covariance of water balance inputs and outputs with precipitation, (b) denotes, for each time step, redistributing the residual error to the components. We first develop these related techniques on a coherent mathematical foundation, and then test them for efficacy in generating accurate component estimates, and in reducing random errors of component estimate series.

The outline of this paper is as follows: Section 2 presents general water balance formulations and associated errors. Section 3 introduces two novel water balance techniques. Section 4 gives pertinent information for application of the techniques to water balances for the Tahoe Basin. Section 5 illustrates application of precipitation-decorrelation to Tahoe water balances. Section 6 describes validation testing of residual-redistribution results for Tahoe precipitation and runoff series, and effects of the precipitation series estimate on hydrologic model performance. Section 7 comments on applicability of the new techniques, and Section 8 provides a summary and conclusions.

2. Formulation of water balances and associated errors

Water balance formulations and component error structure are specified here, facilitating later development and application of the new water balance techniques.

2.1. Initial water balances

Each annual water mass balance in a multi-annual sequence is formulated as

\[
\sum_{i} C_{i}(a) = \epsilon_{\text{res}}(a),
\]

where \( C_{i}(a) \) (units mass/year or volume/year) is an estimate of the true water balance component value \( C_{i}(a) \) for year \( a \), and the sum is over all components \( C_{i} \) in the water balance (e.g. precipitation, runoff, evaporation, etc.). The term \( \epsilon_{\text{res}}(a) \) is the residual error in the water balance estimate for year \( a \), which would be zero if \( C_{i}(a) = C_{i}(a) \) for every component in the water balance. When summing over \( i \) as in Eq. (1), water balance outputs are subtracted from inputs (sign convention: inputs +, outputs –).

The mean annual water balance results from summing Eq. (1) over \( a \) for a period of \( N \) years and then dividing by \( N \). The mean annual magnitude of component \( C_{i} \) for these \( N \) years is denoted \( \bar{C}_{i} \), such that the mean annual water balance for \( N \) years is

\[
\sum_{i} \bar{C}_{i} = \epsilon_{\text{res}},
\]

where \( \epsilon_{\text{res}} = 0 \) for a specified period of large \( N \); i.e., mean annual inputs, outputs, and storage changes are exactly balanced. This mass balance closure statement entails that one component \( C_{j} \) is determined using Eq. (2) in the form

\[
\sum_{i} \epsilon_{\text{res}}(a) + C_{j} = 0.
\]

2.2. Systematic and random errors

Each true water balance component \( C_{i}(a) \) is modeled in terms of its estimate \( \hat{C}_{i}(a) \) and associated errors, similar to formulations by Milly and Dunne (2002), as

\[
\hat{C}_{i}(a) = C_{i}(a) - \delta_{\text{sys}} C_{i} - \delta_{\text{cl}} \cdot [\hat{C}_{i}(a) - \bar{C}_{i}] - \epsilon_{\text{res}}(a),
\]

where \( \epsilon_{\text{res}}(a) \) denotes systematic error in \( C_{i} \), \( \delta_{\text{cl}} \) is a factor for quantifying systematic error in annual deviations of \( C_{i}(a) \) from \( C_{i} \), and all remaining error in \( C_{i}(a) \) is assigned to \( \epsilon_{\text{res}}(a) \), which is called the random error associated with \( C_{i}(a) \). Each \( \epsilon_{\text{res}}(a) \) term consists of error in \( C_{i}(a) \) that is not accounted for by the \( \delta_{\text{sys}} C_{i} \) term or \( \delta_{\text{cl}} \) factor. Each \( \epsilon_{\text{res}}(a) \) series is defined with zero annual mean (i.e., \( \sum_{a} \epsilon_{\text{res}}(a) = 0 \)) by re-assigning any nonzero mean in the value to \( \delta_{\text{sys}} C_{i} \). Supplement S6B presents a discussion of error model Eq. (3) and the suitability of this model for water balance component estimates.

Given that \( \epsilon_{\text{res}}(a) = 0 \), the mean annual version of Eq. (3) becomes

\[
\bar{C}_{i} = \bar{C}_{i} - \delta_{\text{sys}} C_{i}.
\]

which, coupled with the identity for true components \( \sum \epsilon_{\text{res}} = 0 \) and for the component estimates \( \sum \bar{C}_{i} = 0 \) for large \( N \) as for Eq. (2), implies \( \sum_{a} \delta_{\text{sys}} C_{i} = 0 \). As a consequence, summing Eq. (3) over all components and using Eq. (1) yields

\[
\epsilon_{\text{res}}(a) = \sum_{i} \epsilon_{\text{res}}(a) + \sum_{i} \delta_{\text{cl}} \cdot [\hat{C}_{i}(a) - \bar{C}_{i}],
\]

which hence is a fundamental expression linking annual water balance residual errors to random \( \epsilon_{\text{res}}(a) \) and systematic (\( \delta_{\text{cl}} \)) errors of the constituent component estimates.

3. Novel water balance techniques

Two new statistical techniques for improving water balance component estimates are introduced, each of which makes use of the annual residual errors of a multi-annual water balance series. Each technique can operate on multi-annual water balances that are formulated (or can be re-formulated) as Eqs. (1) and (2), with component errors that are approximated by model form Eq. (3) for major water balance components.

3.1. Technique of precipitation-decorrelation

First, bounds on the variance of the residual error are presented and conditions are specified for which the residual errors consist mainly of random errors. Next a criterion of ‘precipitation-decorrelation’ of the residual errors is proposed as a condition that is likely to be associated with a small residual variance. Finally, two methods for practical application of this criterion to water balances are outlined.

3.1.1. Bounds on the variance of the residual error series

In Supplement S1A, it is shown how Eq. (5) is used to determine the following bounds on the magnitude of \( \sigma_{\text{res}}^{2} = \text{var}\{\sum_{a} \epsilon_{\text{res}}(a)\} \)

(\( \text{where } \sigma_{\text{res}}^{2} \equiv \text{var}\{\epsilon_{\text{res}}(a)\} \)).
\[
\begin{align*}
\sigma_{res}^2 - \text{var}\left\{ \sum_i c_i(a) \right\} & \leq \sum_i + \delta_{ci} \cdot \text{cov}\left\{ c_i(a), \varepsilon_{res}(a) + \sum_j c_j(a) \right\} \\
-2 \sum_i + \delta_{ci} \cdot \text{cov}\left\{ c_i(a), \sum_j c_j(a) \right\} & \leq \sigma_{res}^2 - \text{var}\left\{ \sum_i c_i(a) \right\} \\
\leq +2 \sum_i + \delta_{ci} \cdot \text{cov}\left\{ \varepsilon_{res}(a), \sum_i c_i(a) \right\} 
\end{align*}
\]

(6)

(7)

where absolute values are required since it is generally not known whether any \( \delta_{ci} \) is positive or negative, and where the \( i' \) in the summand denotes all terms added (rather than outputs subtracted). Conditions on two terms in inequalities (6) and (7) are described below, which may each hold for some (but not all) water balance series:

**Condition (i):** \( \text{cov}\{ c_i(a), \sum_j c_j(a) \}/\sigma_{res} \ll 1 \) for each component with a large bound on \( |\varepsilon_{ci}^a| \). This condition cannot be checked numerically since \( \varepsilon_{ci}^a \) values are unknown. However, \( E(\text{cov}\{ c_i(a), \varepsilon_{ci}^a \}) = 0 \) can typically be inferred for many or all \( i, j \) pairs.

**Condition (ii):** \( \text{cov}\{ c_i(a), \varepsilon_{res}(a) \}/\sigma_{res} \ll 1 \) for each component with a large bound on \( |\varepsilon_{ci}^a| \). This condition can be checked numerically using water balance data for \( c_i(a) \) and \( \varepsilon_{res}(a) \) values. From the right side of Eq. (7), this condition entails \( \sigma_{res}^2 \ll \text{var}\{ \sum_i c_i(a) \} \).

When conditions (i) and (ii) both hold, then from inequalities (6) or (7) this is sufficient to ensure \( \sigma_{res}^2 \sim \text{var}\{ \sum_i c_i(a) \} \), from which it follows (Supplement S1B) that

\[
\varepsilon_{res}(a) \sim \sum_i c_i(a),
\]

(8)

i.e., each residual error is comprised mainly of the random errors of the component estimates. If additionally the \( c_i(a) \) series are mutually independent, such that \( \text{var}\{ \sum_j c_j(a) \} \sim \sum_i + \sigma_{ci}^2 \) (where \( \sigma_{ci}^2 \equiv \text{var}\{ c_i(a) \} \)), then from Eq. (8) it follows that

\[
\sigma_{res}^2 \sim \sum_i + \sigma_{ci}^2.
\]

(9)

When numerical estimates or bounds for \( \sigma_{ci}^2 \) are available for all water balance components, then the accuracy of Eq. (9) can be checked numerically. Otherwise, it is sufficient that conditions (i) and (ii) hold in order for approximation Eq. (8) to hold, and additionally that the \( c_i(a) \) series are mutually independent in order for Eq. (9) to hold.

### 3.1.2. The criterion of precipitation-decorrelation

Typically many component \( c_i(a) \) series (e.g. streamflow, groundwater flow, ET) are significantly correlated with areal precipitation \( P(a) \) series. Therefore when \( \text{cov}\{ P(a), \varepsilon_{res}(a) \} \) is small, it is likely that \( \text{cov}\{ c_i(a), \varepsilon_{res}(a) \} \) is also small for such precipitation-correlated components. This helps to ensure that condition (ii) (Section 3.1.1) is met, which constrains the size of \( \sigma_{res}^2 \) and thus motivates the following criterion:

\[
\text{cov}\{ P(a), \varepsilon_{res}(a) \} = 0
\]

(10)

which is called **Criterion I**, or the criterion of precipitation-decorrelation. Using Eqs. (9), (10) can be equivalently expressed as \( \text{cov}\{ P(a), \sum_{i,\text{inputs}} c_i(a) \} = \text{cov}\{ P(a), \sum_{i,\text{outputs}} c_i(a) \} \), i.e., the water balance output series varies with \( P(a) \) in the same way as does the input series. This approximates \( \text{cov}\{ P(a), \sum_i c_i(a) \} = 0 \); an identity since \( \sum_i c_i(a) = 0 \).

Substituting Eq. (1) for \( \varepsilon_{res}(a) \) into Eq. (10) and using statistical identities (see Supplement S1C), a weighted mean annual water balance results:

\[
\sum_i k_i \cdot C_i = 0,
\]

(11)

where \( k_i = \text{cov}\{ N_i(a), N_j(a) \}/\sigma_{N_p}^2 \), wherein \( N_i(a) = c_i(a)/C_i \) denotes a normalized annual component estimate, and \( \sigma_{N_p}^2 \) is the variance of the \( N_p(a) \) series. Each weighting factor \( k_i \) (dimensionless) is equivalent to the linear least-squares regression result for the slope in a plot of \( N_i(a) \) vs \( N_j(a) \), such that \( k_p = 1 \). A water balance for which condition Eq. (10) (or equivalently Eq. (11)) holds (within a specified level of statistical significance) is henceforth referred to as ‘precipitation-decorrelated’, i.e., the residual errors are not correlated with the precipitation estimates.

### 3.1.3. Implementing a criterion of precipitation-decorrelation

Fulfillment of criterion I by a water balance series helps to ensure that the variance of the water balance residual \( \varepsilon_{res}(a) \) series is no larger than the variance of the series of summed random component errors. We propose that the criterion of precipitation-decorrelation can be implemented to estimate component values, as follows:

**Application 1:** In the absence of reliable measurement-based or model estimates of the means of two water balance components (e.g. ET and groundwater), the weighted water balance Eq. (11) is used together with the conventional nonweighted water balance Eq. (2) to solve for each of these two components, as illustrated in Section 5.1.1.

**Application 2:** Alternatively, if the relationship of an annual component \( Cm(a) \) series to annual precipitation \( P(a) \) is not known reliably (e.g. for subsurface water storage changes), condition Eq. (10) is first checked. If condition Eq. (10) does not hold (within some desired level of statistical significance), then the relationship of the \( Cm(a) \) series to \( P(a) \) is revised so that it does, as outlined in Section 5.2 and detailed in Supplement S1D.

There is no requirement that the relationship of any of the component estimate series to the \( P(a) \) series need be linear in order for condition Eq. (10)/(11) to be used as in applications 1 or 2. If the relationship of the residuals to \( P(a) \) is nonlinear, then a nonlinear revision of the relationship of \( Cm(a) \) to \( P(a) \) can be used (application 2) in order to remove this nonlinear relationship of the residuals to \( P(a) \) (see Supplement S1D).

We emphasize that applications 1 and 2 yield reliably accurate estimates of water balance components under some, but not all conditions. Uncertainties associated with components estimated using applications 1 or 2 are presented in Supplement S2, wherein conditions for which uncertainties are small and large are detailed and quantified.

### 3.2. Technique of residual-redistribution

‘Residual-redistribution’ is a technique for revising a component series estimate such that its associated random error variance is reduced. This section presents the mathematical basis for residual-redistribution, and how to implement it.

#### 3.2.1. Model form for a revised component estimate and Hypothesis 1

The technique of residual-redistribution exploits the identity that each water mass balance residual error \( \varepsilon_{res}(a) \) equals the sum of the component estimate errors (Eq. (1)). This identity suggests that component estimate error might be reduced by adding
back a fraction \( F(1 - \delta) \) of \( \varepsilon_{\text{res}}(a) \) to each initial component estimate \( C_i(a) \), as follows:

\[
C_i(a) \equiv C_i(a) + (\pm)\left[F/(1 - \delta)\right] \cdot \varepsilon_{\text{res}}(a) \tag{12}
\]

where \( C_i(a) \) is a revised estimate of the actual component value \( C_i(a) \) and \( F \) is a dimensionless parameter, with \((\pm)\) for water balance inputs/outputs. Variances of the \( \varepsilon_{C_i}(a) \) and \( \varepsilon_{C_i\text{r}}(a) \) random error series associated with initial and revised component estimate series are denoted \( \sigma^2_{\varepsilon_{C_i}} \equiv \text{var}\{\varepsilon_{C_i}(a)\} \) and \( \sigma^2_{\varepsilon_{C_i\text{r}}} \equiv \text{var}\{\varepsilon_{C_i\text{r}}(a)\} \), respectively.

**Hypothesis 1.** Part 1: Given a water balance series for which each \( \varepsilon_{\text{res}}(a) \) is composed of a sum of mutually independent random errors \( \varepsilon_{C_i}(a) \) of the component estimates \( C_i(a) \), then for any \( C_i(a) \) series formed using model \( 12 \), there are values of \( F \) for which \( \sigma^2_{\varepsilon_{C_i\text{r}}} < \sigma^2_{\varepsilon_{C_i}} \). Part 2: There is a value of \( F \) (optimal value) for which \( \sigma^2_{\varepsilon_{C_i\text{r}}} \) is minimized.

### 3.2.2. Proof of Hypothesis 1: Minimization of random error variances

The error structure associated with the above \( C_i(a) \) estimate is modeled as for \( C_i(a) \) by Eq. (3). Using model \( 12 \) and results from Section 2 above, it is shown in Supplement S3A that: (i) \( C_i = C_i \), (ii) the systematic error terms \( \varepsilon_{\text{sys}} \) and \( \delta \) are the same for \( C_i(a) \) as for \( C_i(a) \), and therefore (iii) the random error constituent \( \varepsilon_{C_i\text{r}}(a) \) associated with each annual revised estimate \( C_i(a) \) is

\[
\varepsilon_{C_i\text{r}}(a) = \varepsilon_{C_i}(a) + (\pm)F \cdot \varepsilon_{\text{res}}(a) \tag{13}
\]

The variance \( \sigma^2_{\varepsilon_{C_i\text{r}}} \equiv \text{var}\{\varepsilon_{C_i\text{r}}(a)\} \) equals the variance of the right side of Eq. (13). Using Eqs. (8) and (9), Supplement S3B shows

\[
\sigma^2_{\varepsilon_{C_i\text{r}}} = (1 - 2F) \cdot \sigma^2_{\varepsilon_{C_i}} + F^2 \cdot \sigma^2_{\varepsilon_{\text{res}}} \tag{14}
\]

from which \( \sigma^2_{\varepsilon_{C_i\text{r}}} < \sigma^2_{\varepsilon_{C_i}} \) for \( 0 < F < 2(\sigma^2_{\varepsilon_{C_i}}/\sigma^2_{\varepsilon_{\text{res}}} ) \), thus proving Part 1 of Hypothesis 1.

In order to obtain \( \varepsilon_{C_i\text{r}}(a) \) magnitudes that are (on average) as small as possible, the variance \( \sigma^2_{\varepsilon_{C_i\text{r}}} \) of the \( \varepsilon_{C_i\text{r}}(a) \) series is minimized with respect to \( F \); i.e.,

\[
\partial(\sigma^2_{\varepsilon_{C_i\text{r}}})/\partial F = 0 \tag{15}
\]

which is an error variance minimization criterion, herein called criterion II. Substituting Eq. (14) for \( \sigma^2_{\varepsilon_{C_i\text{r}}} \) into Eq. (15) and solving for \( F \), criterion II is satisfied for

\[
F \approx F_C \equiv (\sigma^2_{\varepsilon_{C_i}}/\sigma^2_{\varepsilon_{\text{res}}}) \tag{16}
\]

From Eqs. (9) and (16), \( 0 \leq F_C \leq 1 \). Substituting \( F = F_C \) and \( \sigma^2_{\varepsilon_{\text{res}}} = \sigma^2_{\varepsilon_{C_i}}/F_C \) (Eq. (16)) into Eq. (14) yields the following approximation for \( \sigma^2_{\varepsilon_{C_i\text{r}}} \), minimized as per criterion II:

\[
\sigma^2_{\varepsilon_{C_i\text{r}}} \approx \sigma^2_{\varepsilon_{C_i}} \cdot (1 - F_C) \tag{17}
\]

and thus Part 2 of Hypothesis 1 is proved with optimal value \( F = F_C \).

The important result Eq. (17) is a statistical consequence of using an initial water balance estimate series for which residual errors have the characteristics defined by Eqs. (8) and (9), and of revising initial component estimate series as in model Eq. (12) with \( F = F_C \).

### 3.2.3. Implementing the residual-redistribution technique

Revising component estimates according to model Eq. (12) with \( F = F_C \) from Eq. (16) constitutes the technique of residual-redistribution. The technique is thus simple to implement, and requires only that each water balance residual error consists mainly of a sum of mutually independent random errors. Each revised component estimate series has a smaller associated random error variance (Eq. (17)), as validated in Section 6.

### 4. Tahoe basin water balances and measurement-based component estimates

#### 4.1. Tahoe basin location and hydrology

The Tahoe Basin (Fig. 1) is bounded by the Sierra Nevada and Carson mountain range crests of California and Nevada, USA, and consists of Lake Tahoe and the surrounding watershed. Lake Tahoe covers 38% of the 1314 km² basin area. Elevations range from 1.90 km at Lake Tahoe to 3.32 km at Free Peak. Mean annual precipitation ranges from under 45 cm/yr near the eastern shore of Lake Tahoe to over 200 cm/yr near the Sierra Nevada crest (Craig and Pavelka, 1970; Marjanovic, 1989; Thodal, 1997); falling mainly as winter and spring snow. Sixty-three streams drain into Lake Tahoe.

Establishment of baseline water balances with accurate characterization of inter-annual variability are needed to evaluate effects of land cover and climate change on Tahoe Basin water resources. Tahoe Basin mean water yield is about 23% of mean basin precipitation, thus a small shift in mean precipitation, or in evaporation from Lake Tahoe or the watershed would result in large changes.

![Fig. 1. Location (inset) and map of Tahoe Basin drainage, California and Nevada, USA. The numerous watershed stream networks (some shown) all terminate at Lake Tahoe. Lake Tahoe drains into the Lower Truckee River near gage site ‘SD’. Annual data series from gages on three streams (S1, S2, S3) and three precipitation gages (P1, P2, P3) were linearly adjusted to estimate inter-annual variations of runoff from the entire watershed into Lake Tahoe and of areal precipitation onto Lake Tahoe and the surrounding watershed.](image)
in mean annual stream outflow from the lake. Many estimates of the magnitudes of major components of Tahoe Basin and sub-basin water budgets have been performed over the past ~100 years, reviewed by Thodai (1997) and Trask (2007). Large uncertainties were associated with prior estimates of lake evaporation, watershed atmospheric losses, and groundwater flow into Lake Tahoe.

4.2. Tahoe annual water balance formulations

Each annual water mass balance for Lake Tahoe itself (Fig. 1) is formulated as

\[ PL(a) + R(a) + G(a) + GS - \{E(a) + D(a) + Y(a)\} = e_{res, L}(a) \]  
(18)

where water-year (WY) ‘a’ is defined from 1 Oct. of calendar year ‘a’ – 1 through 30 Sept. of calendar year ‘a’. For each WY a, the following are component estimates: \( PL(a) \) is lake precipitation, \( R(a) \) is incoming surface water runoff, \( G(a) \) and \( GS \) are time-varying and steady groundwater inflow, respectively, \( E(a) \) is lake evaporation, \( D(a) \) is engineered water diversion, \( Y(a) \) is yield of Lake Tahoe and the Tahoe Basin, and \( e_{res, L}(a) \) is the lake water balance residual error.

\( Y(a) \) is the sum of \( SD(a) \) and \( SL(a) \), where \( SD(a) \) is estimated surface water discharge from Lake Tahoe and \( SL(a) \) is estimated water storage (stage) change in Lake Tahoe over the water-year (+ increase, – decrease).

The Tahoe watershed is defined as the Tahoe Basin area exclusive of Lake Tahoe. Each Tahoe watershed annual water mass balance is formulated as

\[ PW(a) - R(a) + G(a) + GS - \{E(a) + AT(a) + \Delta SW(a)\} = e_{res,W}(a) \]  
(19)

where \( PW(a) \) is watershed precipitation, \( AT(a) \) is total atmospheric loss (including snow sublimation, ET, and land surface evaporation), \( \Delta SW(a) \) is watershed water storage change (+ increase, – decrease) over the period of water-year ‘a’, and \( e_{res,W}(a) \) is the watershed water balance residual error. \( \Delta SW(a) \) includes all water storage changes, including changes in water table depth, soil water content, and water levels of wetlands, ponds and lakes (excluding Lake Tahoe). The terms \( R(a) \), \( G(a) \) and \( GS \) in Eq. (19) are the same as in Eq. (18), since Lake Tahoe is the common drainage terminus for the surrounding Tahoe watershed area.

Each Tahoe Basin annual water balance is the sum of Eqs. (18) and (19),

\[ PB(a) - \{E(a) + AT(a)\} - \{D(a) + Y(a) + \Delta SW(a)\} = e_{res,B}(a) \]  
(20)

where \( PB(a) = PL(a) + PW(a) \) is total precipitation onto the entire Tahoe Basin, and \( e_{res,B}(a) = e_{res,L}(a) + e_{res,W}(a) \) is the basin water mass balance residual error.

Mean annual Tahoe water balances (mean annual Eqs. (18)–(20)) are for the period WY 1968–2000 as described below, with mean annual component values denoted \( \bar{C} \) and mean annual values of \( e_{res,L}(a), e_{res,W}(a), e_{res,B}(a) \) set to zero as for Eq. (2).

4.3. Measurement-based estimates of Tahoe water balance components

High quality meteorologic and hydrologic data were available from Oct. 1967 through Sept. 2000 for most components of the Tahoe areal water balances above, enabling accurate measurement-based estimates of most components for each of these 33 water-years. Reliable measurement-based estimates were not available for total groundwater flow from the entire Tahoe watershed area into Lake Tahoe, or for atmospheric losses from and water storage changes in the Tahoe watershed. However, estimates of most (and not all) components in the water balance are sufficient in order to fruitfully apply the new water balance techniques introduced in Section 3.

All measurement-based component estimates are from Trask (2007). Inter-annual variations of several components were determined first in normalized form: annual point precipitation data series are highly spatially correlated across the Tahoe Basin, as are annual streamflow data series for most gaged sub-basins. This facilitated very accurate estimates of inter-annual variations in normalized areal precipitation \( N_P(a) \) and in runoff \( N_R(a) \) from the entire watershed, using high-quality data from just three National Weather Service (NWS) precipitation gages and three United States Geological Survey (USGS) stream gage stations (locations Fig. 1). For groundwater inflow to the lake, a normalized \( N_G(a) \) series was modeled by assuming that \( N_G(a) \) responds to \( N_R(a) \) in the same way as does normalized annual stream baseflow in Tahoe sub-basins.

Estimates of mean annual Lake Tahoe evaporation \( E \) from class A pan, energy balance/Bowen Ratio, and Dalton mass-transfer methods all agree closely, and were combined to form a measurement-based estimate \( \bar{E}_m = 92.9 \pm 10.7 \) cm/yr (~90% confidence). Normalized inter-annual variations of the Dalton estimates \( N_E(a) \) are accurate, as shown by error variance and covariance analyses (see Supplement S4C).

Measurement-based \( \bar{C}(a) \) for annual areal precipitation, runoff and lake evaporation are each computed using a measurement-based mean annual estimate (listed in Table 1 for \( PL \) and \( R \) and the \( N_C(a) \) described above, as \( \bar{C}(a) = CI \cdot N_C(a) \).

5. Application of precipitation-decorrelation to Tahoe water balances

Results and analysis are presented for application of precipitation-decorrelation; first to Lake Tahoe in order to estimate mean annual Lake Tahoe evaporation and groundwater inflow, and next to the whole Tahoe watershed in order to estimate inter-annual variations in watershed moisture storage and annual atmospheric loss.

5.1. Results and evaluation for Lake Tahoe

This section for the lake includes: (1) computing mean annual evaporation and groundwater inflow, (2) a confidence interval on the result for mean evaporation, (3) corroboration of the result for mean evaporation, (4) combining independent estimates of mean evaporation, (5) comparison of several mean annual water balance component estimates and their uncertainties.

5.1.1. Mean annual evaporation and groundwater inflow

Application 1 (Section 3.1.3) of the precipitation-decorrelation technique is used to estimate magnitudes of two components of the mean annual Lake Tahoe water balance. \( \bar{C} \) is selected to be so estimated because a well-constrained measurement-based estimate of \( \bar{C} \) is not available. Measurement-based estimates are available for all other lake water balance components, including \( \bar{F} \) (Section 4.3). Here \( \bar{F} \) is also estimated using precipitation-decorrelation, both because of the small uncertainty in use of this technique for \( \bar{F} \) specifically, and because several independent measurement-based estimates of Lake Tahoe evaporation are available as checks on the accuracy of the result.

The mean WY 1968–2000 conventional and weighted water balance Eqs. (2) and (11) for Lake Tahoe are solved simultaneously for \( \bar{F} \) and \( \bar{C} \). Eliminating \( \bar{C} \) from the two equations and solving for \( \bar{F} = \bar{F}_{pd} \), where ‘pd’ denotes precipitation-decorrelation,

### 5.1.4. Combined estimate of mean annual evaporation

Because estimates of \( E \) from Sections 4.3 and 5.1.1 were obtained independently, they can be used together to form a combined estimate \( E_c \) that has smaller associated uncertainty than either estimate alone. Following Ross (1987), a maximum-likelihood estimate of \( E \) that makes use of the two independent estimates \( E_m \) and \( E_p \) is \( E_c = \sigma_E \cdot \left[ (\bar{E}_m/\sigma_m^2) + (\bar{E}_p/\sigma_p^2) \right] \) where \( \sigma_E^2 = 1/\left(1/\sigma_m^2 + 1/\sigma_p^2\right) \) is the variance of uncertainty in \( E_c \), and \( \sigma_m^2, \sigma_p^2 \) are the variances of uncertainty in \( E_m, E_p \). Inserting mean \( E \) estimates and associated uncertainties from Sections 4.3.5.1.1. and 5.1.2 into this expression for \( E_c \) yields \( E_c = 93.5 \pm 6.6 \) cm/yr (~90% confidence) for WY 1968–2000. With \( E = E_c \), then \( \bar{E} = \bar{E}_c = 1.4 \) cm/yr (over the area of Lake Tahoe) is computed as the residual in Eq. (2) for the Lake Tahoe water balance.

### 5.1.5. Comparison of Lake Tahoe mean annual water balance estimates

Fig. 2 compares results of the three methods, namely measurement-based, precipitation-decorrelation, and combined approaches and highlights the agreement and potential discrepancies.

### Table 1


<table>
<thead>
<tr>
<th>Component</th>
<th>Mean (cm/yr)</th>
<th>( \sigma_m ) (cm/yr)</th>
<th>( k )</th>
<th>( \text{corr} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_L )</td>
<td>53.3</td>
<td>22.0</td>
<td>1.000</td>
<td>-0.022</td>
</tr>
<tr>
<td>( R )</td>
<td>92.8</td>
<td>48.4</td>
<td>1.214</td>
<td>-0.099</td>
</tr>
<tr>
<td>( G(a) )</td>
<td>1.4</td>
<td>0.66</td>
<td>1.042</td>
<td>-0.026</td>
</tr>
<tr>
<td>( G_s )</td>
<td>0.75</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( D(a) )</td>
<td>43.5</td>
<td>3.92</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td>( E(a) )</td>
<td>50.2</td>
<td>69.3</td>
<td>3.223</td>
<td>-0.181</td>
</tr>
<tr>
<td>( \bar{E}_{\text{res}} )</td>
<td>6.67</td>
<td>1.4</td>
<td>0.66</td>
<td>1.000</td>
</tr>
</tbody>
</table>

\( a \) Inter-annual standard deviation.
\( \text{corr}(\bar{G}(a), \bar{X}(a)) \) For \( n = 33 \) data pairs \( \langle Y(a), X(a) \rangle \), hypothesis \( \rho = 0 \) accepted if \( |\text{corr}(Y(a), X(a))| < -29 \) (90% interval).
\( c \) cm/yr over Lake Tahoe surface area of 494 km\(^2\) (mean for WY 1968–2000).
\( d \) Combined mean \( E, G \) (Section 5.1.4).
(Sections 4.3.5.1.1, and 5.1.4) estimates of some mean annual water balance component (and component group) estimates, including the uncertainty associated with each estimate. Fig. 2 shows each of the three methods generates a similar estimate of the mean of each component or component group.

Results for uncertainty from left to right in Fig. 2 are as follows: for lake evaporation, $E_{pd}$ has an associated uncertainty that is slightly smaller than that of $E_{as}$, and $E_{as}$ has the smallest uncertainty. This highly accurate $E_{as}$ estimate is enabled by the availability of the $E_{pd}$ estimate (Section 5.1.4). Systematic uncertainty of sums or differences of components is determined as the square root of the sum of squares of uncertainties of contributing components, thus the same pattern is seen for total water balance outputs. Because uncertainty in $E_{as}$ is just $\pm 6.6$ cm/yr ($\sim$90% confidence) and uncertainties in remaining lake water balance output components ($D, T$) are even smaller, resultant uncertainty in the estimate of 148.2 cm/yr for WY 1968–2000 mean annual outputs $E + D + T$ from Lake Tahoe, and thus also of mean annual inputs $PL + GS + C + R$ to Lake Tahoe, is only $\pm 7.1$ cm/yr ($\sim$90% confidence). Uncertainty in watershed drainage into the lake $GS + C + R$ is computed by combining this uncertainty in water balance outputs with uncertainty in $PL$ ($\pm 10.6$ cm/yr), and is large for all methods but retains the same relative order of size. Uncertainty in $G$ is comparably large among all three methods; and is not reduced by using precipitation-decorrelation.

5.2. Results for Tahoe watershed

With $G = G_{C}$ (Section 5.1.4), all other components of the Tahoe watershed water balance Eq. (19) are available from measurements (Section 4.3) except annual and mean annual atmospheric loss and water storage changes. Mean annual $AT(a) + ASW(a)$ is determined as the residual in Eq. (2) for the watershed water balance; the result is $AT + ASW = 43.6$ cm/yr. Since measurement-based estimates of $AT(a) + ASW(a)$ were not available, a trial estimate of invariant $AT(a) + ASW(a) = 43.6$ cm/yr was tested for use in the annual watershed water balance Eq. (19) series. However, the resulting residuals are highly correlated with $PW(a)$ ($p < 0.001$), thus violating criterion I (Section 3.1.2).

Application 2 (Section 3.1.3) of the precipitation-decorrelation technique is used to estimate $AT(a) + ASW(a)$. Supplement S5 presents an approach for determining the approximate relationship of any poorly resolved water balance component $Ck(a)$ series to the precipitation $P(a)$ series. This approach entails transferring the relationship (nonlinear or linear) of the initial residuals with $P(a)$ to $Ck(a)$, such that criterion I holds for the resultant revised residuals. For the Tahoe watershed trial estimate $AT(a) + ASW(a) = 43.6$ cm/yr, the associated residual series relationship to $PW(a)$ is linear; furthermore these residuals are also correlated with precipitation from prior years and with $E(a)$. Thus the following multi-linear model is used to fit the residual $AT(a) + ASW(a)$ values:

$$AT(a) + ASW(a)_{res,a} = 43.9 + 92.2[N_{E}(a) - 1] + 17.3[N_{PW}(a) - 1] - 5.99[N_{PW}(a - 2) - 1]$$

where $AT(a) + ASW(a)_{res,a}$ values (for estimation of coefficients) were computed as the residuals in Eq. (19), and the fitted coefficient values (units cm/yr over the watershed area) are for WY 1968–2000, computed using least-squares regression. Bounds on random error variance of this series are estimated in Supplement S5d. Further discussion of this model form and systematic uncertainties is deferred to Section 6.5.2.

5.3. Initial water balances and residual errors

Initial Lake Tahoe and watershed annual water balance component estimates for each water-year from 1968 to 2000 are graphed in Fig. 3a and b, where $E(a)$, $G(a)$ values are computed by multiplying $N_{E}(a)$ and $N_{G}(a)$ (Section 4.3) by $E_{C}$ and $G_{C}$ (Section 5.1.4), respectively. $AT(a) + ASW(a)$ values are computed from Eq. (22), and remaining component estimates are measurement-based (Section 4.3).

Each annual residual value $e_{res,l}(a)$, $e_{res,wh}(a)$, or $e_{res,d}(a)$ is computed as the sum of the $C_{k}(a)$ in Eqs. (18), (19), or (20). These residual series have no significant correlation with precipitation or other major water balance component series, suggesting that they can be used for residual-redistribution (see Section 6.1.1). Fig. 3a and b illustrate that most residual values (solid black circles in graphs) are small, and yet as is illustrated in Section 6, component series that are revised using such small residuals have less associated random error ‘noise’, and as such are valuable for use in hydrologic models.

6. Application of residual-redistribution to Tahoe water balances

This section illustrates application of residual-redistribution to initial Tahoe water balances, validation of random error reduction
in resulting revised component estimates, and the value of such revised estimates for use in watershed hydrologic models.

6.1. Results for residual-redistribution

Tahoe water balance residual and component error variance relationships are first checked for suitability of application of residual-redistribution. Next, revised Tahoe component estimate series are described, and finally three distinct revised estimates of Tahoe watershed runoff series are presented.

6.1.1. Evaluation of suitability of the residuals for redistribution

The suitability of using the initial water balance residuals defined in Section 5.3 to generate residual-redistribution revised component estimates is evaluated using two approaches: (I) Assessment of fulfillment of conditions (i) and (ii) (and independence) from Section 3.1.1, and (II) Direct numerical evaluation of the accuracy of Eq. (9).

(I) Each residual series (for lake, watershed, and basin) has no statistically significant correlation (e.g. Table 1) with the precipitation series or with any other component Ci(a) series, fulfilling condition (ii) in Section 3.1.1. Furthermore, Kendall Tau tests and plots of residual values vs. Ci(a) values show that no apparent trend or nonlinear relationship exists between the residuals and any of the Ci(a). The methods used to estimate Ci(a) (Section 4.3) indicate that condition (i) from Section 3.1.1 is also met, and furthermore that component random errors are mutually independent.

(II) The left-side bar of the three bar-pairs in Fig. 4 displays, for the initial Tahoe lake, watershed, and basin water balances, the values (stacked) of the random error variances \( \sigma_{\text{res}}^2 \) associated with each component series, as well as the residual variances \( \sigma_{\text{est}}^2 \). The \( \sigma_{\text{res}}^2 \) values for precipitation, runoff, and yield (listed in Table 2) were reliably determined from the data and methodology for estimation of the initial component series values (see Supplement S4A). Estimates of \( \sigma_{\text{res}}^2 \) for \( E(a) \), \( AT(a) + \Delta SW(a) \), and \( G(a) \), each have large associated uncertainty (Supplements S4B, C, D). Fig. 4 shows that random errors in precipitation and runoff contribute most to the residual error variance \( \sigma_{\text{res}}^2 \), while \( \sigma_{\text{est}}^2 \) values are systematically small for each component estimate series, with the sum of the residual variance \( \sigma_{\text{res}}^2 \) being smaller than for \( \sigma_{\text{est}}^2 \), as shown below for watershed and basin precipitation; which have been tabulated for some major water balance components in Table 2.

\[
\text{Residual minus comp. errors} = \text{Evaporation} E(a) \text{ (lake)} + \text{AT(a)+delSW(a) (wshed)} + \text{Yield} Y(a) \text{ (lake & basin)} + \text{Diversions} D(a) \text{ (lake & basin)} + \text{Groundwater} G(a) \text{ (washed to lake)} + \text{Runoff} R(a) \text{ (washed to lake)} + \text{Precipitation} (lake, wshed, basin)
\]

The two sets of observations described in the two paragraphs above strongly indicate that Eqs. (8) and (9) are approximated for each of the initial lake, watershed, and basin water balance residual series, i.e., the residual error series are composed mainly of sums of mutually independent random component errors. Thus for each of these water balances, the residuals are suitable for redistribution over the initial component estimates, so that random errors of the component estimates can be reduced.

6.1.2. Results for revised annual component estimates

Residual-redistribution is applied to the initial annual component estimates from Section 5.3 (subsequently denoted \( \text{Ci(a)} \)) as follows: using the \( \sigma_{\text{res}}^2 \) and initial \( \sigma_{\text{est}}^2 \) values just described above, \( \sigma_{\text{e}}^2 \) values are computed from Eq. (16) for each component of the Tahoe lake, watershed, and basin water balances. Table 2 shows estimates of the annual mean and inter-annual standard deviation of Tahoe precipitation, runoff, and yield series, as well as associated \( \sigma_{\text{e}}^2 \) and \( \sigma_{\text{c}}^2 \).

Using these \( \sigma_{\text{c}}^2 \) values, a revised estimate \( \text{Ci(a)} \) of each of these components is formed using model Eq. (12), with results shown below for watershed and basin precipitation; which have the largest \( \sigma_{\text{e}}^2 \).

\[
\text{PW}_{\text{w}}(a) = \text{PW}_{\text{w}}(a) - 0.662\sigma_{\text{res,w}}(a)/(1 - \delta_{\text{PW}}) \quad (23)
\]

\[
\text{PB}_{\text{b}}(a) = \text{PB}_{\text{b}}(a) - 0.746\sigma_{\text{res,b}}(a)/(1 - \delta_{\text{PB}}) \quad (24)
\]

The \( \delta_{\text{Ci}} \) are systematic error factors (Section 2.2), which have expected values of zero and limits \( |\delta_{\text{Ci}}| < 0.2 \) (90% confidence) for all major Tahoe water balance components (Supplement S2D). Generally, values of \( \delta_{\text{Ci}} \) can be varied (within a range set by a confidence interval) to test the effect of \( \delta_{\text{Ci}} \) on hydrologic models that use \( \text{Ci(a)} \) series.

The random error variance \( \sigma_{\text{res,ci}}^2 \) associated with each revised component series is predicted using Eq. (17). Predicted \( \sigma_{\text{res,ci}}^2 \) values are tabulated for some major water balance components in Table 2 and illustrated in the right-side bar of each pair in Fig. 4. Fig. 4 illustrates, for each water balance, that the sum of component random error variances is predicted to be much smaller for revised than for initial water balances, reflecting the smaller error variances predicted for each revised component series.

6.1.3. Three revised estimates of the Tahoe watershed runoff series

Watershed runoff is a component of both the Lake Tahoe and Tahoe watershed water balances (Eqs. (18) and (19)), therefore...
runoff estimates can be revised in several ways, as shown in this section. From the lake and watershed water balances, two estimates of revised runoff series follow directly from model Eq. (12),

$$R_{L}(a) = R_o(a) - 0.418e_{est,L}(a)/(1 - \delta_k)$$

$$R_{W}(a) = R_o(a) + 0.161e_{est,W}(a)/(1 - \delta_k)$$

(25)

(26)

wherein the revised estimate $R_{L}(a)$ results from redistribution of the lake water balance residual $e_{est,L}(a)$, and revised estimate $R_{W}(a)$ results from redistribution of the watershed water balance residual $e_{est,W}(a)$. Since each revision is an estimate of the same actual runoff quantity, a single revised runoff series estimate should be definable, with a smaller associated random error variance than either $R_{L}(a)$ or $R_{W}(a)$. Analogous to Model Eq. (12), it is proposed that such a revised runoff estimate can have the following form:

$$R_{c}(a) = R_o(a) + [-F_{L} \cdot e_{est,L}(a) + F_{W} \cdot e_{est,W}(a)]/(1 - \delta_k)$$

(27)

Values of $F_{L}$ and $F_{W}$ that minimize the random error variance of this ‘combined’ $R_{c}(a)$ series estimate are determined to be 0.524 and 0.246 (Supplement S3C), using error variance minimization criteria analogous to Eq. (15). The $R_{c}(a)$ series is predicted to have a smaller associated $\sigma_{est}^2$ than either the $R_{L}(a)$ or $R_{W}(a)$ series (Table 2).

### 6.2. A Method for validation testing of residual-redistribution

The validity of residual-redistribution is defined by the efficacy of random error variance of component series estimates. The difference in random error variance between initial $C_i(a)$ and residual-redistribution revised $C_r(a)$ component series is denoted $\Delta \sigma_{est}^2 = \sigma_{est,i}^2 - \sigma_{est,r}^2$ (for brevity, the ‘$C_i$’ portion of the subscript is omitted from error terms). From Eq. (14) and substituting $\sigma_{est,i}^2 - F_{ci} \cdot \sigma_{est,i}^2$ from Eq. (16), ‘predicted’ $\Delta \sigma_{est}^2$ values are computed for any value of $F$ as

$$\Delta \sigma_{est}^2 \approx 2F \cdot F_{ci} - F_{ci}^2 \cdot \sigma_{est}^2$$

(28)

from which maximum $\Delta \sigma_{est}^2 \approx F_{ci}^2 \cdot \sigma_{est}^2$ at $F = F_{ci}$; and $\Delta \sigma_{est}^2 > 0$ for $0 < F < 2F_{ci}$.

An additional independent estimate of $\Delta \sigma_{est}^2$ requires the availability of another estimate $C_{int}(a)$, in addition to $C_i(a)$, of the same actual component series $C_i(a)$ for the same time period. The data for the $C_{int}(a)$ must be independent from that used for $C_i(a)$. Assuming that the $C_{int}(a)$ series has the same error structure Eq. (3) as does the $C_i(a)$ series, then each of the three component series $C_{int}(a)$, $C_i(a)$, $C_r(a)$ are linear combinations of each other, plus random error terms (see Supplement S5A). As shown in Supplement S5B, this leads to the following ‘observed’ approximation for $\Delta \sigma_{est}^2$, which is computed using data observations for $C_{int}(a)$, $C_r(a)$ and $C_i(a)$:

$$\Delta \sigma_{obs}^2 \approx \Delta MSE_{est} \cdot [(1 - \delta_k)^2/\beta_{est}]$$

(29)

where $\Delta MSE_{est} \equiv MSE_{est,i} - MSE_{est,r}$ wherein $MSE_{est,i} \equiv var(C_i(a) - \beta_{est} \cdot C_r(a))$, $MSE_{est,r} \equiv var(C_r(a) - \beta_{est} \cdot C_i(a))$, and $\beta_{est}$ is an estimate of $\beta \equiv (1 - \delta_b)/(1 - \delta_o)$ wherein systematic error factors $\delta_b$ and $\delta_o$ are defined as in Section 2.2 for initial and additional component series, respectively. As shown in Supplements S5C, D, E, highly accurate estimates $\beta_{est}$ of $\beta$ can be obtained by using either of the two expressions below:

$$\beta_{est} = [\sigma_{est,i}^2 + \sigma_{est,r}^2]/[\sigma_{est,i}^2 + \sigma_{est,r}^2/(1 - \delta_k)^2]$$

(30)

$$\beta_{est} = \beta_{reg}/[1 - \sigma_{est}^2/\beta_{est}^2 \cdot \sigma_{est}^2]$$

(31)

where $\sigma_{est,i}^2 \equiv var(C_i(a))$, $\sigma_{est,r}^2 \equiv var(C_r(a))$, $\sigma_{est}_{int} \equiv var(C_{int}(a))$, and $\beta_{reg}$ is the least-squares estimate of the slope in linear regression of $C_{int}(a)$ vs. $C_i(a)$. Eq. (31) must be solved quadratically for $\beta_{est}$, unless both $\sigma_{est,r}^2 < \sigma_{est,i}^2$ and $\sigma_{est,r}^2 < \sigma_{est}_{int}^2$, wherein a close approximation of Eq. (31) is $\beta_{est} \approx \beta_{reg}/[1 - \sigma_{est,r}^2/\sigma_{est}_{int}^2]$.

Under conditions where the value of $\beta_{est}$ is highly accurate (i.e. $\beta_{est} \approx \beta$), then $\Delta \sigma_{obs}^2$ may be a more accurate estimator of $\Delta \sigma_{est}^2$ than $\Delta \sigma_{pred}^2$. Accordingly, to gauge the efficacy of residual-redistribution and the accuracy of the $\Delta \sigma_{pred}^2$ value, the ‘observed’ $\Delta \sigma_{obs}^2$ value from Eq. (29) is to be compared with the prediction $\Delta \sigma_{pred}^2$ from Eq. (28). Whether or not the predicted value from Eq. (28) closely approximates the observed value from Eq. (29), some reduction in random error variance of the revised component estimates is confirmed if the value of $\Delta MSE_{est}$ (and thus $\Delta \sigma_{obs}^2$) is positive.

### 6.3. Testing random error reduction in revised estimates of Tahoe precipitation

In this section, the validation method just described above is applied to the initial and revised precipitation series estimates for both the Tahoe watershed and basin, to determine whether or not the random error variances of these revised precipitation series estimates are actually reduced by the predicted amounts listed in Table 2.

### Table 2


<table>
<thead>
<tr>
<th>$\rho_i(t)$</th>
<th>Water balance</th>
<th>Mean (cm/yr)</th>
<th>$\sigma_{est,i}(cm/yr)$</th>
<th>$\sigma_{est,r}(cm/yr)$</th>
<th>$\sigma_{est,obs}(cm/yr)$</th>
<th>$F_{ci}(-)$</th>
<th>$\sigma_{est,obs}(cm/yr)$</th>
<th>$1 - (\sigma_{est,i}/\sigma_{est,obs})^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{L}(a)$</td>
<td>Lake</td>
<td>54.5</td>
<td>22.5</td>
<td>4.00</td>
<td>6.66</td>
<td>0.360</td>
<td>3.20</td>
<td>36.0</td>
</tr>
<tr>
<td>$PW(a)$</td>
<td>Wished</td>
<td>82.9</td>
<td>29.8</td>
<td>5.29</td>
<td>6.50</td>
<td>0.662</td>
<td>3.08</td>
<td>66.2</td>
</tr>
<tr>
<td>$RB(a)$</td>
<td>Basin</td>
<td>72.2</td>
<td>26.9</td>
<td>4.58</td>
<td>5.30</td>
<td>0.746</td>
<td>2.34</td>
<td>74.6</td>
</tr>
<tr>
<td>$R(a)$</td>
<td>Lake</td>
<td>92.8</td>
<td>48.4</td>
<td>4.31</td>
<td>6.66</td>
<td>0.418</td>
<td>3.29</td>
<td>41.8</td>
</tr>
<tr>
<td>$R(a)$</td>
<td>Wished</td>
<td>56.1</td>
<td>29.4</td>
<td>2.61</td>
<td>6.50</td>
<td>0.161</td>
<td>2.39</td>
<td>16.1</td>
</tr>
<tr>
<td>$Y(a)$</td>
<td>Lake</td>
<td>50.2</td>
<td>69.3</td>
<td>1.82</td>
<td>6.66</td>
<td>0.075</td>
<td>1.75</td>
<td>75</td>
</tr>
<tr>
<td>$Y(a)$</td>
<td>Basin</td>
<td>18.9</td>
<td>26.1</td>
<td>0.69</td>
<td>5.30</td>
<td>0.017</td>
<td>0.68</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Note: Values in italics denote that the subscripts ‘o’ & ‘r’ denote initial & revised; wshed denotes watershed.

* Values of $F_{ci}$ and $\sigma_{est,i}$ predicted using Eqs. (16) and (17), respectively.

* Predicted reduction (as a proportion of $(\sigma_{est,i})^2$) in random error variance.

* Values for mean $(\text{and} \sigma)$ adjusted for formulation of $G(a)$ and $A(t) = \Delta SW(a)$ on precipitation.

* For ‘combined’ revised runoff $R_{c}(a)$ series (Section 6.1.3.1).
Additional 'Snotel' precipitation data series

Additional high-quality point precipitation data are available for WY 1979–2004 from ten 'Snotel' stations (Supplement S5F) widely distributed throughout the Tahoe watershed area, which are distinct from the NWS station data used for the initial precipitation estimates (Section 4.3). The data series from each Snotel station is linearly transformed to a common mean and variance (for either the watershed or basin), then for each year the average of the transformed data from all stations constitutes the ‘additional’ watershed or basin precipitation estimate $PW_{ad}(a)$ and $PB_{ad}(a)$ for that year, as described by Trask (2007). The resultant $PW_{ad}(a)$ and $PB_{ad}(a)$ series have the same annual mean and inter-annual variance values for WY 1979-2000 as the $PW(a)$ and $PB(a)$ series, respectively.

6.3.2. Regressions of Snotel precipitation against other precipitation series

Each $P_{ad}(a)$ series ($PW_{ad}(a)$ or $PB_{ad}(a)$) is regressed against several NWS-based precipitation series for WY 1979–2000: (a) data from a NWS gage at Tahoe City (b) a $Pw(a)$ series ($PW(a)$ or $PB(a)$ from Section 4.3) (c) a $Pw(a)$ series ($PW(a)$ from Eq. (23) or $PB(a)$ from Eq. (24)). For each year, each of these precipitation estimates is an estimate of the same $PW(a)$ or $PB(a)$ with error structure (3), thus scatter of data about the linear relationship between the $P_{ad}(a)$ and each of series (a) thru (c) is entirely due to random errors in $P_{ad}(a)$ and in (a), (b), or (c), which constitute the MSE of regression.

Regression of the $P_{ad}(a)$ series against series (a) (coefficient of determination $R^2 = 0.9402$) is much less ‘tight’ than regression against series (b) ($R^2 = 0.9813$), as expected since each $P_{ad}(a)$ series is formed using data from three gage stations and thus has smaller associated random errors. Further increase in $R^2$ is obtained upon regression against series (c) ($R^2 = 0.9919$ and 0.9874, respectively). As detailed in Supplement S5B, such observed increases of $R^2$ for series (c) strongly suggest that $\Delta_{ad} \sigma^2 > 0$.

Next, estimates of $\Delta_{pred} \sigma^2$ are computed and compared with predictions $\Delta_{pred} \sigma^2$.

6.3.3. Comparison of observed to predicted error variance values

Observed $\Delta_{obs} \sigma^2$ and predicted $\Delta_{pred} \sigma^2$ changes in error variance are compared between precipitation series (b) and (c) described just above. Supplement S5F shows $\beta_{est}$ computed using Eqs. (30) and (31) range from 0.9934 to 0.9973. Using $\beta_{est} = 0.995$, $\Delta MSE_{est}$ values are computed as defined under Eq. (29), from which $\Delta_{obs} \sigma^2$ values are estimated using Eq. (29). Values of $\Delta_{pred} \sigma^2$ are from Eq. (28).

Fig. 5a and b each shows results for $\Delta_{obs} \sigma^2$ and $\Delta_{pred} \sigma^2$ (each divided by $\sigma_{res}^2$ to form dimensionless values) as a continuous function of the fraction $F(1-\delta_F)$ of $\epsilon_{res}(a)$ added to $P_{ad}(a)$. Fig. 5a and b each illustrate that $\Delta_{obs} \sigma^2$ increases with $F(1-\delta_F)$ up to a maximum value near $F_{p}(1-\delta_F)$ then decreases upon further increase in $F_{p}(1-\delta_F)$, consistent with predictions $\Delta_{pred} \sigma^2$. If the $\epsilon_{res}(a)$ series had no correlation with the $\epsilon_{res}(a)$ series, then $\Delta_{obs} \sigma^2$ would be negative for all values of $F$; decreasing linearly with increasing $|F|$. It is the positive correlation of $\epsilon_{res}(a)$ with $\epsilon_{res}(a)$ (see Eq. (8)) that results in $\Delta_{obs} \sigma^2 > 0$ (i.e. the reduced random error variance of each $P_{ad}(a)$ series) for $0 < F(1-\delta_F) < 2F_{p}(1-\delta_F)$.

For $PW(a)$, Fig. 5a illustrates $\Delta_{pred} \sigma^2$ (for $F_{p}(1-\delta_F) = 0.662$) is slightly larger than $\Delta_{obs} \sigma^2$. However for $PB(a)$, Fig. 5b shows $\Delta_{pred} \sigma^2$ (for $F_{p}(1-\delta_F) = 0.746$) is much larger than $\Delta_{obs} \sigma^2$ due to mismatch between $\Delta_{obs} \sigma^2$ and $\Delta_{pred} \sigma^2$ (see below).

6.3.4. Discrepancies between observed and predicted error variance values

$\Delta_{obs} \sigma^2$ and $\Delta_{pred} \sigma^2$ values will match closely under the following conditions:

(i) Any nonlinearities between the $Ci_{ad}(a)$ series and the $Ci(a)$ series are insignificantly small, and $\beta_{est}$ is accurate (Supplements S5E, F describe how to evaluate this condition).

(ii) The estimate of $\sigma^{2}$ (and thus of $F_{CI}$ and $\Delta_{reg} \sigma^2$) is accurate, and $\{ci\} < -0.1$.

(iii) The term $2 \cdot (1-\delta \cdot F \cdot \text{cov}(\epsilon_{res}(a), \epsilon_{ad}(a)) / \sigma_{est}$ is $\ang\Delta_{obs} \sigma^2$.

Supplement SSH presents an assessment of the magnitude of departures of Tahoe precipitation series from conditions (i) thru (iii), and the effect of these departures on validation results. With reference to condition (ii), Fig. 5a and b illustrates that very close matches of $\Delta_{reg} \sigma^2$ to $\Delta_{obs} \sigma^2$ are attained for $F_{p}(1-\delta_F) = 0.590$ and $F_{p}(1-\delta_F) = 0.512$. With reference to condition (iii), comparably close matches (Fig. 5a and b) are also attained by replacing...
\( \Delta \sigma^2 \) with \( \Delta \sigma^2 = (1 - \delta_e \cdot F \cdot \text{cov}(\text{est}(a), \text{raw}(a)) / \sigma_{\text{tot}}^2 \), which is statistically expected to yield a more accurate match (than \( \Delta \sigma^2 \)) of \( \sigma_{\text{tot}}^2 \) (see Supplements S5G, H). For the \( \text{PW}_a(a) \) series in particular, it is likely that \( \Delta \sigma^2 \) substantially underestimates the actual change in random error variance.

In summary, random error variances of the revised precipitation series for the watershed and basin (Eqs. (23) and (24)) are each shown to be substantially smaller (~60% to ~70% and ~50% to ~75%, respectively) than those of the initial precipitation series. Indeed, the error variance reductions are close to the predictions shown to be substantially smaller (~60% to ~70% and ~50% to ~75%, respectively) than those of the initial precipitation series.

6.4. Testing random error reduction in revised estimates of Tahoe runoff

Testing of predicted random error variance reduction of revised runoff estimates is performed as described in Section 6.3 for precipitation. Additional USGS gage data are from seven streams (not used for the \( \text{PW}_a(a) \)) distributed about the Tahoe watershed (see Supplement S5I), which are combined to form a \( \text{PW}_a(a) \) runoff series for WY 1969–2000. The \( \text{R}_a(a) \) series is regressed against the initial runoff series \( \text{R}_a(a) \) (Section 4.3), and against each of the three revised runoff series defined by Eqs. (25)–(27).

Resulting observed \( AMSE_{\text{est}} \) values are positive, indicating that each revised runoff series estimate has a smaller random error variance than the initial runoff series. \( AMSE_{\text{est}} \) values range from about the same as to as much as 2-fold larger than \( AMSE_{\text{est}} \) values from Eq. (28). Discrepancies are due to violations of condition (i) (Section 6.3.4), since the \( \text{R}_a(a) \) data have significantly larger positive skew than the \( \text{R}_a(a) \) data, and violations of condition (iii). Supplement S5I presents additional details and discussion. In summary, the random error variances of the three revised runoff series estimates are about the same or smaller than the predicted values listed in Table 2.

6.5. Effects of random errors in precipitation series on regression results

Having confirmed above that the revised Tahoe precipitation (and runoff) series have reduced random errors, we now present two simple examples of the value of such estimates. Highly resolved hydrologic multi-regression models are presented for (1) Tahoe streamflow and basin yield data series, and (2) residual values of Tahoe watershed atmospheric loss and storage changes; wherein a precipitation \( \text{P}(a) \) series is one of the regressors. Results are compared among several different \( \text{P}(a) \) series; each \( \text{P}(a) \) series having a different associated random error variance size, ranging from large to small.

6.5.1. Tahoe streamflow and basin water yield regression models

Data for annual Tahoe yield \( Y(a) \) and volumetric flow \( \text{SF}(a) \) for Tahoe watershed streams (streams identified Fig. 6) are each regressed against initial evaporation \( E(a) \) and several \( \text{P}(a) \) estimates. The number of stations \( s \) used for each \( \text{P}(a) \) series below varies:

(a) Tahoe City NWS precipitation data series for WY 1915–2004 (\( s = 1 \)).
(b) Initial precipitation series \( \text{PW}_a(a) \) for WY 1946–2004 described in Section 4.3 (\( s = 3 \)).
(c) ‘Snotel Plus’ series, composed of annual mean values of linearly adjusted data (WY 1981–2004) from NWS and Snotel precipitation stations (Sections 4.3 and 6.3.1) (\( s = 13 \)).
(d) The watershed revised series \( \text{PW}_a(a) \) (Eq. (23)) for WY 1968–2000 (\( s = 3 \)).
(e) The basin revised series \( \text{PB}_a(a) \) (Eq. (24)) for WY 1968–2000 (\( s = 3 \)).

Precipitation data series among all the stations in (a), (b), and (c) are very highly inter-correlated, therefore the relative sizes of

![Fig. 6. Ratio of mean square error (MSE) of linear multi-regression to inter-annual variance of annual streamflow data, for each of ten streams (U.T. = Upper Truckee) in the Tahoe watershed, and for Tahoe Basin water yield \( Y(a) \). Annual (WY) streamflow data for each stream are multi-regressed against \( E(a) \) and various precipitation estimates (legend) for current and prior WY (Eq. (32)). Period of available data differs between streams; ranging in duration from 11 to 31 years between WY 1969 to 2004 (Trask, 2007).](image-url)
random error variances among series (a), (b), (c) are approximately proportional to 1/s (see appendices A1 and A2 in Trask (2007)). Like series (c), series (d) and (e) each have a random error variance much smaller than that of (b), as shown in Section 6.3.

A linear model for Tahoe Basin annual streamflow and yield data is

$$\{SF(a), Y(a)\} = \beta_1 \cdot E(a) + \beta_2 \cdot P(a) + \beta_3 \cdot P(a - 1) + \beta_4 \cdot P(a - 2) + \beta_5 + \varepsilon_Y(a)$$

(32)

where $\beta_1$ to $\beta_5$ are coefficients whose values are determined by least-squares regression, and $\varepsilon_Y(a)$ is the regression error term.

Results show that $Y(a)$ and many $SF(a)$ data series are closely fit by model (32), with $\beta_1 < 0$ and $\beta_2 > \beta_3 > \beta_4 \geq 0$, and that the $\varepsilon_Y(a)$ are small and approximately normally distributed. For $Y(a)$ and for $SF(a)$ of most streams, $R^2$ values are larger and confidence intervals for $\beta$ to $\beta_5$ are narrower when regressing against (c), (d), (e) as compared to (a), (b), as expected. Also inter-stream differences in dependence of $SF(a)$ on each regressor is significantly better resolved using (c), (d), (e) as compared with (a), (b), as detailed by Trask (2007). $Y(a)$ has a statistically significant positive correlation with $P(a - 2)$ which is obscured by random error ‘noise’ when using (a), (b). The positive correlations of $Y(a)$ with $P(a - 1)$ and $P(a - 2)$ data series (c), (d), (e) emanate from positive correlations of $SF(a)$ with $P(a - 1)$ and $P(a - 2)$ data series (b) thru (e), since the streams all flow into Lake Tahoe and thus contribute directly to Tahoe water yield $Y(a)$.

In regressions against component estimates with error structured as in Section 2.2, systematic errors $e_{C0,i}$ and $e_{C1,i}$ affect regression coefficient values for slope and intercept, but do not affect the scatter of data points about least squares regression lines. Scatter of data $\varepsilon_Y(a)$ about each regression line fit of model (32) is attributable to:

(I) A weighted sum of the random errors $e_E(a)$ in each $C_i(a)$ used in Eq. (32), i.e., $e_{C_{\text{sum}}}(a) = e_{C_{\text{sum}}}(a) - e_E(a) - e_F(a) - e_{C_{\text{sum}}}(a - 1) - \beta_4 \cdot e_{C_{\text{sum}}}(a - 2);$ and

(II) Model structural error $e_{S_{\text{mod}}}(a)$; i.e., error or incompleteness in the form of model (32).

So $\varepsilon_Y(a)$ is attributable to both random and model errors; i.e., $\varepsilon_Y(a) = e_{C_{\text{sum}}}(a) + e_{S_{\text{mod}}}(a)$. Fig. 6 shows $MSE/\sigma_{SF}^2$, results for fitting of model form (32) to streamflow data $SF(a)$ for ten Tahoe sub-basin streams and to $Y(a)$, where $MSE$ approximates the variance of the $e_{C_i}(a)$ values, and $\sigma_{SF}^2$ is the variance of the $SF(a)$ or $Y(a)$ data series. Each cluster of five bars in Fig. 6 shows $MSE/\sigma_{SF}^2$ for one $SF(a)$ data series, fit by each of the five precipitation series (a) thru (e). Smaller $MSE$ results when using (c), (d) or (e) as compared with (a) or (b), attributable to smaller random errors $e_E(a)$, $e_F(a - 1)$, $e_{C_{\text{sum}}}(a - 2)$ in (c), (d), and (e). Since $MSE/\sigma_{SF}^2$ values for (c), (d) and (e) are similar (last three bars in each cluster in Fig. 6), it follows that each of the 3-station revised (by residual-redistribution) precipitation series has random errors about as small as a 13-station unrevised series. This demonstrates that residual-redistribution can reduce the number of stations needed to achieve high-precision estimates of areal precipitation anomalies.

The contribution of $e_{C_E}(a)$ variances to $MSE$ is computed, assuming that the $e_{C_E}(a)$ are mutually independent, using random error variances of gaged yield and streamflow data and of initial evaporation estimates (Supplement S4) and of revised precipitation estimates $PW_i(a)$ and $PB_i(a)$ (Table 2). For each of six streams in the western or southern watershed (Fig. 6), the $e_{C_{\text{sum}}}(a)$ constitute ~35% to over 90% of $MSE$. Errors $\varepsilon_Y(a)$ in model form (32) account for the remainder (under 10% to ~65%) of $MSE$, such that $\text{var}(\varepsilon_Y(a))/\sigma_{SF}^2 < 0.02$. It follows that over 98% of inter-annual variations in $SF(a)$ of western and southern watershed streams (over 99.5% for General Ck., Upper Truckee R., and $Y(a)$) are correlated with inter-annual variations in precipitation and lake evaporation.

By contrast, for each of four eastern streams the large $MSE$ (Fig. 6) is attributable mainly (~70% to ~96%) to errors $\varepsilon_Y(a)$ in the model form of Eq. (32), with $\text{var}(\varepsilon_Y(a))/\sigma_{SF}^2$ ranging from 0.028 (Incline Ck.) to 0.18 (Logan Ck). The eastern watershed has more deeply weathered bedrock (due to lack of Pleistocene glacial scouring), less precipitation and a much smaller streamflow: precipitation mass ratio than the rest of the watershed. Flow data for Edgewood and Logan Ck. are marginally correlated with $P(a - 3)$ (Trask, 2007). Inter-annual variations in eastern streamflow are sensitive to factors in addition to those in Eq. (32), such as inter-annual variations in seasonal distribution and form (rain or snow) of precipitation.

Notably, $SF(a)$ data of southern and western watershed streams are negatively correlated with $E(a)$ with moderate to high statistical significance; however $SF(a)$ of the four eastern streams do not have any significant correlation ($p > 0.1$) with $E(a)$. This strongly suggests that atmospheric losses in the semi-arid eastern watershed are controlled primarily by water availability; whereas in the much more humid western and southern watershed, inter-annual changes in atmospheric demand (as approximated by $E(a)$) play a comparable role to that of variations in water availability (as approximated by $P(a)$) in determining inter-annual changes in atmospheric loss (see below).

6.5.2. Tahoe watershed annual storage change and atmospheric loss models

Numerical estimates of annual watershed storage changes and atmospheric losses are first computed as raw residuals in the watershed water balance Eq. (19), using either the initial $PW(a)$ or an adjusted revised $PB(a)$ precipitation series estimate. The $PB(a)$ series is a linear transformation of the $PW(a)$ series defined by Eq. (24) to the same mean and variance as the $PW(a)$ series. Least-squares multi-regression of the residual values of $\Delta SW(a) + AT(a)$ for WY 1970–2000 is performed using model form (22): the analogous result using normalized $PB(a)$ (denoted $NP_{PB}(a)$) is $|\Delta SW(a) + AT(a)|_{\text{max}} = 41.9 + 95.6 N_{NP_{PB}(a)} - 16.9 N_{NP_{PB}(a)} - 9.61 N_{NP_{PB}(a)} - 5.29 N_{NP_{PB}(a)} - 2$.

in cm/yr units. In contrast with use of $N_{NP_{PB}}(a)$ (and $NP_{PB}(a)$) for which $R^2 = 0.663$, use of $NP_{PB}(a)$ results in a ‘tighter’ regression, with $R^2 = 0.874$. Due to its very small random error variance, the revised precipitation series has enabled substantially improved resolution of the relationship of Tahoe watershed atmospheric losses plus storage changes to lake evaporation and precipitation from prior water-years. As illustrated in Fig. 7, regression coefficient values are similar for use of $NP_{PB}(a)$ (left-side bars) or $NP_{PB}(a)$ (right-side bars) as regressors, however the confidence intervals (denoted by span of error bars) for all regression coefficient values are significantly narrower for $NP_{PB}(a)$. The $p$ values (for hypothesis coefficient is zero) of coefficients for $NP_{PB}(a)$ and $NP_{PB}(a)$ change from marginal ($p = 0.13$) to significant ($p < 0.01$) for $NP_{PB}(a)$ replaced by $NP_{PB}(a)$.

This improved resolution enables more reliable partitioning of Eq. (33) into separate models for $\Delta SW(a)$ and $\Delta AT(a)$. Physically plausible proviso described in Trask (2007) are applied to yield

$$\Delta AT_{\text{mod}}(a) = 95.6 \cdot (N_{NP_{PB}(a)} - 1) + C_1 \cdot N_{NP_{PB}(a)} + C_2 \cdot N_{NP_{PB}(a)} - (C_1 + C_2)$$

(34)
The new techniques in this article should be readily applicable for a wide variety of planned, ongoing or archived lake and catchment water balance investigations, including those that employ physical, statistical, or hybrid models of component estimates. For archived water balance series, the techniques can be employed to update and reduce the uncertainty of archived numerical component estimates, using just the original numerical component estimates and estimates of error or uncertainty statistics for one or more of these component series.

7.1. Precipitation-decorrelation

An existing or preliminary lake or watershed multi-annual water balance series estimate is easily checked directly for the presence of correlation of annual water mass balance residual error values with annual values of precipitation estimates. If a statistically significant nonzero correlation exists, the technique of precipitation-decorrelation may be of value to implement (see Section 3.1.3) in order to obtain more accurate water balance component estimates. Additionally, a precipitation-decorrelation estimate can be combined with an independent measurement-based estimate, yielding a ‘combined’ estimate of a mean component that has smaller associated uncertainty than either estimate alone (see Section 5.1.4, Fig. 2).

7.2. Residual-redistribution

Residual-redistribution is broadly applicable to water balances. The technique requires only that water mass balance residual errors consist mainly of independent random errors. Those components with the largest associated random error variances have their random errors reduced the most upon application of residual-redistribution.

For components $C_k$ for which $\sigma_{C_k}^2/\sigma_{\text{res}}^2$ changes with the magnitude of $C_k(a)$, a refinement to Eq. (16) for determination of $F_k$ parameter values can help to ensure maximal reduction in random error variance, as detailed in Supplement S6B.2. The ratio $\sigma_{C_k}^2/\sigma_{\text{res}}^2$ may vary substantially with $C_k(a)$ for some components of water balances in which one or more $C_i(a)$ series ($i = k$ or $i \neq k$) has both (i) heteroscedastic $\sigma_{C_i}^2$, and (ii) very large variance and/or skew (e.g., annual streamflow series in some semi-arid to arid regions, which may include years of extremely small or large total volumetric flow).

Nonlinear watershed hydrologic models have previously been shown to exhibit better performance when using input data series with smaller associated random errors (e.g., Oudin et al., 2006). Use of residual-redistribution revised component series as input can improve the performance of not only simple linear regression models (as shown in Section 6.5 above), but therefore also of more complex nonlinear models. When calibrating parameter values in such hydrologic models, it is important that the ‘dependent’ (output) and independent (input) variable(s) do not both include revised component series, so that consequent spurious calibration statistics are avoided, and bias in resulting model parameter values will not be so introduced.

7.3. Use of both techniques to estimate inter-annual variations

In the absence of direct data to estimate inter-annual variations of a water balance component, statistical techniques can be used to generate rough estimates of these inter-annual variations. Inter-annual variations of Lake Tahoe evaporation for WY 1968–2000 are estimated as detailed in Supplement S6A by first applying...
precipitation-decorrelation and then residual-redistribution. The resultant statistically based $E_2(a)$ series estimate is moderately correlated ($R = 0.754$) to the Dalton mass-transfer $E_2(a)$ series estimate, with inter-annual standard deviation 3.8 cm/yr nearly matching that of the Dalton series (3.9 cm/yr). Inter-annual variations of $E_2(a)$ are more accurate than evaporation variations computed as simple residuals of the annual Lake Tahoe water mass balance series.

### 7.4. Non-stationary and stationary water balances

No assumptions of water balance stationarity were used to develop the two new water balance techniques, and each technique is broadly applicable to both stationary and non-stationary water balance series. If the residual errors trend over the water balance period, it is recommended that the trend should first be transferred to the appropriate component(s) series, followed by redistribution of the de-trended residuals (see Supplement S6C for further discussion).

### 7.5. Adaptation of techniques

Each of the two new techniques should be adaptable to water balance time-steps that are shorter or longer than one year. Each technique may be of value to adapt for use in distributed parameter models of watershed precipitation and runoff.

We suggest the approach of residual-redistribution could have broad applicability not only within hydrology, but in other areas of science, engineering, economics, etc. The approach yields reduced random error in estimates of the distribution of a conserved quantity (e.g. flow of mass, energy, money, etc.). The major requirement for application of this technique is an estimate of relative magnitudes of random error variances among the various components (or bins) within which the conserved quantity is distributed.

### 8. Summary and conclusions

We have introduced a new approach of identifying statistical relationships that exist between a series of water mass balance component estimates and the consequent water balance residual error series, and using such relationships to modify and improve the component estimates. Two novel specific statistical techniques are developed, which have been corroborated and validated using water balances for the Tahoe Basin.

A criterion of de-correlation of multi-annual water balance residual error series from the precipitation component estimate series has been proposed. Upon application to a Lake Tahoe water balance series, a ‘precipitation-decorrelation’ estimate of WY 1968–2000 mean annual Lake Tahoe evaporation of 93.8 cm/yr results, which closely matches independent measurement-based estimates by Trask (2007), Huntington and McEvoy (2011), and Sahoo et al. (2013). Uncertainty of this statistical estimate is just ±8.9% (>90% confidence), a level of accuracy typically achievable only by high quality energy balance investigations (Winter, 1981). Since precipitation is generally the main driving force in natural water balances, it is likely that application of precipitation-decorrelation to water balances in many other regions may (as for Tahoe) result in water balances for which the residual error series have no statistically significant correlation with any of the component series, and for which the residual error variance is small.

Redistribution of each random residual error over all water balance components, in accord with a specific error variance minimization criterion, yields revised component estimates that are predicted to have smaller associated random errors. This prediction has been validated using Tahoe water balance series, whereby random error variances of revised Tahoe areal precipitation series are shown to be reduced by ~50–75% to a c.v less than 4%. Such large reductions are otherwise typically achievable only by a substantial increase in precipitation gage density. Revised estimates of inter-annual variations in Tahoe areal precipitation have very little associated random error ‘noise’, which enables: (a) resolution of the relationship of streamflow to precipitation from prior years, and thus the contribution of low (one to several years) modes of groundwater discharge to streamflow, (b) clarification of the extent to which inter-annual variations in atmospheric loss from the watershed parallel those of nearby lake evaporation, and (c) more reliable separation of inter-annual changes in watershed subsurface water storage from inter-annual variations in atmospheric loss, which is otherwise generally difficult to achieve.

The new techniques can be especially valuable for application to multi-annual water balances spanning a decade or more, since associated uncertainties decrease with an increasing number of water balance periods. It is our hope that this article provokes further investigation of statistical relationships between water balance residual error series and water balance component estimate series, which have potential to be applied to improve the accuracy of component estimates.

### Acknowledgements

We thank Nick Newcomb for preparing the Tahoe map (Fig. 1). Tahoe hydrologic data used in this article are listed and/or referenced in Trask (2007). The work presented in this article was funded in part by the United States Environmental Protection Agency, Grants R819658 and R825433.

### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhydrol.2016.12.029.

### References


