

Evidence review · Land and freshwater

What an unclosed catchment water budget can tell us

Elín Bjarnadóttir · 3 October 2026

Fictional author and institute · Editorial demonstration, not peer reviewed

ABSTRACT

An imbalance in a catchment water budget does not identify a missing pathway, and exact closure does not validate every component. This evidence review synthesises two multi-basin closure studies and measurement-uncertainty guidance to develop a diagnostic approach for northern catchments. It examines boundary definitions, compatible units and periods, snow carryover, storage, and dependence between products. A fictional annual account shows why a residual can be small relative to combined uncertainty and why replacing an independent evapotranspiration estimate with a remainder is not validation. Seasonal residual patterns and targeted independent observations are considered as ways to discriminate explanations. The review supplies no universal closure threshold and performs no local assessment. Its conclusion is that retaining the residual and component histories makes a budget more informative than forcing an apparently complete account.

Keywords: water balance · budget closure · storage change · uncertainty · catchment boundaries

A source-based editorial text, not a report of original Still North research. Examples and imagery are illustrative.

A residual is a question, not a missing process

Catchment water budgets are often presented as tidy accounts: precipitation arrives, water leaves through streams and evapotranspiration, and storage changes by the remainder. When independently estimated terms fail to balance, it is tempting to name the residual as ground-water loss or adjust evapotranspiration until the account closes. Either step can create a seemingly complete explanation without evidence identifying the missing process. Conservation of water is certain; the adequacy of the measurements and boundary assumptions is not.

This evidence review asks how a northern catchment budget can guide investigation without disguising uncertainty. It draws on two published multi-basin analyses of water-budget closure and NIST guidance on combining uncertainties. The approach is a targeted methodological synthesis, not a systematic review or a new catchment assessment. A worked budget uses invented values to show which conclusions the arithmetic supports and which require independent observations.

Write the physical boundary before the equation

For a defined period, a useful account is precipitation plus external inflows, minus stream outflow and atmospheric losses, equals storage change. Atmospheric losses can include evaporation, plant transpiration, and snow sublimation. The reporting convention should make clear whether all are included in a combined term. Subsurface flow across the mapped catchment boundary, water abstraction, and deliberate transfers must be included where relevant. A topographic divide is not automatically a groundwater divide.

Expressing every term as millimetres over the same catchment area helps expose unit mistakes. Stream discharge must first be integrated through time and divided by that area. A water-level change in a lake is not directly an equivalent catchment-wide storage change: lake geometry and its area relative to the catchment are required. Similarly, soil-water measurements at one depth and point cannot be subtracted directly from basin precipitation.

The start and end dates matter. A snowmelt-season budget can draw on snow accumulated before the period begins. Treating that release as new precipitation would double count the water if snow storage were also represented. A hydrological year may reduce some seasonal carryover, but does not guarantee zero storage change. Groundwater, lakes, snow, and ground ice can retain a multi-year memory, particularly during persistent change.

Closure does not validate each component

Lehmann, Vishwakarma, and Bamber compared many combinations of precipitation, evapotranspiration, and runoff data with satellite-derived storage change across major river basins. Some combinations produced good closure partly through cancellation of errors in poorly estimated components. This is a consequential result: matching the total does not prove that each term is correct. Two wrong numbers can be wrong in compensating directions.

Michailovsky and colleagues examined water-balance-derived runoff across a larger catchment sample and found that product performance depended on catchment characteristics. Increasing snow cover, latitude, and altitude were associated with poorer closure in their analysis. Their comparison also showed why differing geograph-

ical coverage can distort apparent product rankings. A product assessed mainly in easier conditions should not automatically be declared the best option for a snowy northern basin.

These studies evaluate large-scale products, not the quality of a particular local gauge or sensor. Their direct value here is the diagnostic logic: assess components and support, not closure alone. A coarse satellite storage product cannot serve as a precise independent check for an arbitrarily small catchment simply because a map provides values over that location. Effective observational resolution, leakage, and shared model inputs must be considered.

Worked example: a positive residual with several explanations

Fictional worked example. Suppose a one-year budget over a fixed catchment estimates precipitation at 500 millimetres, stream outflow at 260 millimetres, atmospheric losses at 190 millimetres, and storage increase at 20 millimetres. The residual, defined as precipitation minus those three terms, is 30 millimetres. Under this convention, a positive residual could represent an omitted net outward pathway, underestimated losses or storage, overestimated precipitation, or a combination.

Assign illustrative standard uncertainties of 40, 15, 30, and 20 millimetres to precipitation, outflow, atmospheric losses, and storage change respectively. If these uncertainties are independent, the combined standard uncertainty of the residual is the square root of $40^2 + 15^2 + 30^2 + 20^2$, approximately 56 millimetres. A residual of 30 millimetres is small relative to that uncertainty. The calculation does not identify a groundwater leak, nor establish that groundwater exchange is zero.

It would be misleading to replace the atmospheric-loss estimate with 220 millimetres and then celebrate exact closure as validation. That number is a residual estimate conditional on the other terms and assumptions. It might be useful if labelled accordingly, but it is not independent evidence confirming evapotranspiration. The original 190-millimetre estimate should remain available so readers can see the disagreement rather than receiving only the reconciled result.

Consider a different pattern: the annual residual remains modest, but each spring shows a large negative residual followed by a compensating positive summer residual. This could point towards misaligned snow-storage accounting, delayed runoff routing, or seasonal bias in a product. Annual aggregation would conceal the diagnostic pattern. The sign and timing of residuals can guide the next measurement even when their annual total cannot establish a specific missing mechanism.

Do not assume independent errors

The illustrative 56-millimetre uncertainty is valid only under its stated assumptions. A modelled evapotranspiration product may use the same precipitation forcing as another budget term. Catchment area error affects several conversions together. Shared snow models or assimilated gauge records can make nominally separate products dependent. In the general uncertainty calculation, covari-

ance terms must be included with the correct signs from the budget equation.

NIST guidance distinguishes uncertainty from error and allows uncertainty in corrections for recognised systematic effects. This matters for precipitation undercatch and rating-curve corrections. Applying a correction does not make its uncertainty vanish. Conversely, an uncertainty interval should not replace a correction when evidence establishes a significant systematic effect. Identify the effect, apply the justified correction, and preserve the remaining uncertainty.

Long averaging periods reduce some random variability, but do not necessarily reduce common biases. Repeating a precipitation underestimate for ten years can create a persistent apparent excess of runoff. Averaging also changes the question: a long-term mean may be appropriate for climate-scale partitioning but inadequate for a short flood or drought. The analysis period should follow the decision, rather than being lengthened solely until the residual becomes small.

Use the residual to choose a discriminating observation

Start with a support audit. Confirm the same dates, area, unit conventions, and treatment of snow and ice. Check whether observed outflow is missing important winter or high-flow periods. Examine the elevation distribution of precipitation observations and whether the product covers all the terrain. These checks are often more informative than fitting another parameter, because a boundary or calendar mismatch can invalidate an otherwise sophisticated calculation.

Next identify competing explanations and ask which measurement would separate them. An apparent storage increase might be tested with lake levels and geometry, groundwater heads, or appropriate gravimetry at larger scale. Suspected cross-boundary groundwater exchange needs hydrogeological and hydraulic evidence. An uncertain atmospheric-loss estimate might benefit from independent energy-balance or flux observations, interpreted at their actual footprint. No single extra instrument resolves every budget term.

Report sensitivity to plausible alternatives. For example, compare corrected and uncorrected precipitation using a justified correction range, or show budgets with alternative storage assumptions. These are conditional scenarios, not independent estimates to average into confidence. If several products share inputs, their spread is not automatically an uncertainty distribution. Agreement can reflect common ancestry as readily as independent corroboration.

Limits and the value of an honest imbalance

This synthesis does not provide a universal percentage threshold for acceptable closure. A small absolute residual can matter greatly in a low-runoff catchment, while a larger one may be unresolved within measurement uncertainty elsewhere. Nonlinear processes, non-normal error distributions, and strongly correlated inputs can require simulation or more detailed uncertainty analysis than the linear worked example. Such methods still depend on credible input assumptions.

A useful catchment account publishes measured and modelled terms separately, retains the residual with its sign convention, and explains which pathways remain unconstrained. Perfect closure is not the aim if it is achieved by hiding disagreement. The stronger result is knowing whether the disagreement challenges the boundary definition, the observations, or a proposed mechanism. An unclosed budget can then become a well-posed next question rather than a deficit to erase.

References

1. Lehmann, F., Vishwakarma, B. D., and Bamber, J. (2022). How well are we able to close the water budget at the global scale? *Hydrology and Earth System Sciences*, 26, 35–54.
2. Michailovsky, C. I., Coerver, B., Mul, M., and Jewitt, G. (2023). Investigating sources of variability in closing the terrestrial water balance with remote sensing. *Hydrology and Earth System Sciences*, 27, 4335–4354.
3. National Institute of Standards and Technology. Technical Note 1297, section 5: Combined standard uncertainty.