

Methods note · Observation and data

Designing the overlap between instruments

Amina Okafor · 9 October 2026

Fictional author and institute · Editorial demonstration, not peer reviewed

ABSTRACT

Replacing an instrument can improve measurement while disrupting a long record. This methods note synthesises GCOS change-management principles, NIST traceability policy, the JCGM uncertainty guide, and Arctic observing-network context to examine what an overlap period must establish. A fictional temperature comparison demonstrates how a small mean offset can conceal a scale-dependent difference. The analysis treats condition coverage, measurand equivalence, time alignment, correlated pairs, uncertainty in both instruments, and reversible corrections as distinct design questions. It separates calibration, numerical continuity, and validation, and explains the additional limits when simultaneous comparison was impossible. No universal overlap duration, sample count, or calibration procedure is proposed. The aim is a bounded statement of comparability supported by preserved evidence before retirement makes that evidence unobtainable.

Keywords: instrument transitions · overlap periods · calibration · measurement uncertainty

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

The replacement introduces a comparison question

An observing record can outlive the instrument that began it. A replacement may be more stable, easier to maintain, and better characterised than its predecessor, yet still create a discontinuity in the published series. Differences can arise from response time, exposure, sampling interval, processing, or measurement principle. The transition therefore needs evidence about comparability, not simply a note that newer equipment was installed.

This methods note considers how to design and interpret an overlap period for a long environmental record. It synthesises the updated GCOS Climate Monitoring Principles, NIST's traceability policy, and the JCGM uncertainty guide, with the Arctic observing-network assessment as context. The numerical comparison is fictional and deliberately simple. It is not a calibration procedure, an equipment specification, or evidence about any actual replacement sensor.

Begin with the quantity whose continuity matters

Two devices can be physically adjacent without measuring equivalent quantities. A screen-height air temperature differs from surface temperature. A fast instantaneous reading differs from a ten-minute mean. A sensor moved a few centimetres in a stratified water column may sample a different layer. Before comparing numbers,

specify the measurand, location, exposure, interval, time-stamps, and processing that define the intended series.

The 2024 GCOS principles call for assessing system changes before implementation and maintaining a suitable overlap long enough to determine inter-instrument biases. They do not establish one duration for every transition. Adequacy depends on the conditions that influence the comparison and the uncertainty acceptable for the record's purpose.

A transition plan should consequently state what would count as a consequential discontinuity. A broad seasonal classification and a small long-term trend require different tolerances. This does not license choosing a tolerance after seeing an inconvenient difference. It requires deciding, before replacement, which properties the record needs to preserve and what evidence will support that decision.

Fictional case: a mean difference conceals a slope

Suppose an old thermometer reports x , while an invented replacement reports 1.04 times x plus 0.10°C under controlled comparison conditions. Treat x as exact only for this arithmetic demonstration. At old-instrument values of -10°C , 0°C , and 10°C , the new readings are -10.30°C , 0.10°C , and 10.50°C . The corresponding new-minus-old differences are -0.30°C , 0.10°C , and 0.50°C .

Old reading	New reading	New minus old	Residual after subtracting a constant 0.10°C
-10.00°C	-10.30°C	-0.30°C	-0.40°C
0.00°C	0.10°C	0.10°C	0.00°C
10.00°C	10.50°C	0.50°C	0.40°C

The average difference across those three equally weighted points is 0.10°C. Correcting every new value by that constant removes the average difference but leaves opposite residuals at the range ends. A short overlap near 0°C might therefore suggest an adequate offset adjustment while missing the scale-dependent relationship. The example demonstrates why coverage of relevant conditions matters more than the appearance of a small overall average.

It does not show that every replacement requires a linear correction. Real relationships may vary with wind, radiation, humidity, fouling, or response time. They may also reflect a fault in the old system rather than a defect in the new one. Preserving numerical continuity with an earlier bias is not the same goal as improving measurement accuracy. Both raw series and the chosen interpretation need to remain recoverable.

Design overlap around influences, not a calendar slogan

An overlap should span the important measurement range and conditions relevant to the intended analysis. For a seasonal record, that may require more than a convenient maintenance week. For a response-time difference, it may require periods of rapid change rather than more hours of steady conditions. For a satellite transition, surface types, viewing geometry, retrieval algorithms, and field validation can matter as much as simultaneous orbital coverage.

The 2025 Arctic observing-network assessment identifies recalibration, in situ measurements, expertise, and staff time as part of satellite transitions. This is a useful reminder that overlap is not free data produced by leaving two systems switched on. Planning, maintenance, comparison, and interpretation all consume capacity that should be funded before the older system is retired.

Mounting both instruments together controls some environmental differences but may introduce interference. They can shade one another, disturb flow, or compete for limited power. Moving one farther away can avoid that problem while increasing representativeness differences. Document the arrangement and the reason for it. A photograph may help describe exposure, but it cannot substitute for dimensions, configuration, and timing.

Time alignment deserves its own check. If one system timestamps the start of an interval and another the end, a nominally paired comparison can contain a systematic lag. Resampling can create apparent agreement by smoothing away the very response difference under investigation. Compare at a defined common interval, retain the original sampling information, and explain what temporal behaviour the comparison can no longer resolve.

Analyse paired evidence without inventing independence

Plot the paired differences against relevant influences and time, not only new readings against old. A high correlation is expected when both respond to a large seasonal cycle; it can coexist with a consequential offset or drift. Examine whether differences depend on magnitude, conditions, deployment changes, or maintenance.

An apparently random residual may also contain temporal structure that a single summary conceals.

Thousands of minute-level pairs are not necessarily thousands of independent tests. Persistent weather and slowly varying instrument effects can correlate successive differences. A naive standard error based on treating every pair as independent may make the estimated adjustment appear much more precise than the evidence warrants. Analysis should use an appropriate dependence model or aggregation strategy, and report the resulting effective evidential limits rather than merely the file's row count.

The fictional calculation treats the old instrument as exact for clarity. In practice, both systems carry uncertainty. Ordinary least-squares regression with an error-free predictor assumption may then misrepresent the relationship. The method should reflect uncertainties in both measurements, shared references, and other relevant inputs. The JCGM Guide provides the general framework for modelling inputs, propagating uncertainties, and retaining covariance; it does not choose a field comparison model on the analyst's behalf.

Calibration, continuity, and validation answer different questions

NIST's policy defines metrological traceability for a measurement result through a documented chain of calibrations, each contributing uncertainty, and explicitly warns that traceability alone does not guarantee fitness for purpose. A replacement with excellent calibration can still sample an altered environment. Conversely, agreement with an old instrument does not create an independent traceability chain.

Where a correction is justified, record its parameters, applicable range, uncertainty, and version. Explain whether published values are on the old scale, a new reference scale, or an explicitly reconstructed common scale. Avoid calling a field adjustment a universal calibration. A relationship established in one deployment may not transfer to another mounting, season, firmware version, or sensor of the same model.

Testing a chosen correction on data not used to fit it is useful where the evidence allows. It can reveal overfitting, but a random split of adjacent time points may provide little independent challenge. Holding out distinct conditions or periods may be more informative. No split can validate extrapolation into a regime absent from the overlap. Mark that boundary clearly rather than extending a fitted line because software permits it.

When overlap was impossible

Sometimes an older instrument fails before comparison begins. Nearby records, reference checks, and documented deployment history may help investigate a break. They do not recreate simultaneous paired evidence. Statistical homogenisation can be useful under stated assumptions, but it should not be presented as equivalent to a measured overlap. An unresolved discontinuity may require separate segments or additional uncertainty in a trend analysis.

A failure discovered after deployment also raises timing uncertainty. A post-deployment offset does not prove

that drift accumulated smoothly from the previous check. A sudden impact or configuration change may have occurred. Preserve the evidence, use independent records where appropriate, and distinguish an estimated correction history from an observed one.

Limits and implication

No universal overlap duration or sample count follows from this synthesis. The sources supply principles and metrological reasoning, not a substitute for variable-specific expertise. The worked relationship omits random noise, shared uncertainty, environmental response, and ageing. Its purpose is to expose a failure of averaging, not to characterise field hardware.

The durable outcome of an overlap is a defensible statement about comparability over a defined range, with enough evidence to revisit it. Plan that statement before procurement and retirement decisions make comparison impossible. A replacement then becomes a documented transition in knowledge rather than an unexplained step in a continuous-looking line.

References

1. Global Climate Observing System (2024 principles; page updated 2025). GCOS Climate Monitoring Principles, especially System Change Management and Calibration.
2. National Institute of Standards and Technology (updated 2024). NIST Policy on Metrological Traceability.
3. Joint Committee for Guides in Metrology (2008, corrected 2010). Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008.
4. Starkweather, S., Armstrong, M. B., and Shapiro, H. (2025). Assessing the State of the Arctic Observing Network. Arctic Report Card 2025.