

Response to Referees

A physically grounded input distance resolves longitude-encoding ambiguity but not specification dependence in ensemble persistent homology of cyclone trajectories — R. Dai

We thank the three referees again for reports that shaped this manuscript far beyond the points they raised directly. Each item below is given in three parts: the referee’s point, our response, and the specific place in the revised manuscript where the change appears. Section numbers refer to the revised manuscript, and each change is located by page and line number of the revised manuscript PDF submitted with this response.

Two things should be read before the point-by-point replies. The first is a list of values reported in the preprint that no longer stand. The second is a list of statements we made in the interactive discussion that no longer stand. We give both in full, with the reasons, because several of the referees’ requests are what led us to find them.

Corrections to values reported in the preprint

This is the first revised manuscript we have uploaded, so every change described in this letter is a change relative to the discussion paper of 23 June 2026, and the marked-up version is a comparison against that same file. The table below lists every value or claim in the discussion paper that the audit does not sustain, what replaces it, and why.

Reported in the discussion paper	What we report now	Why
Title: “Encoding-dependent verdicts and H_1 miscalibration...”	The miscalibration half is withdrawn; the title now states what the physical distance does and does not fix	The calibration verdict does not survive the reproducibility audit below, so the claim it asserted cannot be carried into a revision.
D1 pool H_1 $p = 0.130 / 0.009 / 0.214$	0.037 (archived longitude convention) or 0.108 (single convention) / 0.021 / 0.156	The single-run implementation built its control array from an unordered set, so it drew a different control sample in every process even with the random seed fixed, and its interpreter hash seed was not recorded. Five independent orderings of the sphere give 0.007–0.018 and five of the cylinder 0.145–0.212, bracketing the published values. The linear value is reproducible under no ordering, because that run’s linear normalizer was overwritten in place and cannot be audited.

Reported in the discussion paper	What we report now	Why
D1 pool H_0 $p = 0.088 / 0.082 / 0.190$	0.052 / 0.059 / 0.067 on the deterministic pipeline; the revised manuscript does not carry them, reporting the H_0 channel through the placebo calibration instead	Same defect, same correction. The corrected values lie closer together and remain encoding-dependent, but a single-run H_0 p-value is the quantity §3.5 argues should not be read alone, so §4.1 gives the H_0 event distance and leaves the channel's verdict to §4.2.
Event H_1 distances 0.538 / 0.646 / 0.529	0.353 / 0.656 / 0.546	The sphere and cylinder normalizers survive in the archived code and reproduce to within a few percent. The linear normalizer does not: 0.538 differs from what the §3.3.1 encoding produces by 52 %, which is how the overwritten implementation was detected.
Placebo H_1 false-positive rates 8.2 % / 10.2 % / 14.3 % (4, 5 and 7 of 49), “inflated by a factor 1.6× to 2.9× regardless of encoding choice” Placebo H_0 false-positive rates 8.2 % / 4.1 % / 6.1 %	1, 4 and 4 of 49 on the deterministic ordering; pooled 9 to 12 of 147 across six auditable orderings, every one nominal-compatible 2, 1 and 2 of 49 on the deterministic ordering; pooled 4 to 16 of 147 across the same six orderings	Every cell came from the process-randomized order. Under the reproducible pipeline the inflation does not appear, and we withdraw the factor claim entirely. Same defect. Reporting H_0 as a distribution rather than a single ordering also reverses the paper's dimension-asymmetry reading: the H_0 spread is the wider of the two, and its upper ordering crosses the nominal rate.
“an intrinsic H_1 inflation that is not curable by lon-encoding choice alone”; “dimension-asymmetric in calibration”	Withdrawn. We claim no general inflation of H_1 over H_0	Under the linear encoding H_1 (1/49) sits below H_0 (2/49). What survives is narrower: the H_1 verdict sits near α at every ordering while the H_0 verdict is far from it at most (§4.2, §5.2).
Stratified attribution, cylinder column 0.058 / 0.488 / 0.400 / 0.044 / 0.736, and the subsample-size table	Regenerated in full (§4.3, §4.4, §S0)	The scripts behind both tables carried the same unordered-set defect and depend on an unrecorded hash seed; re-running under two hash seeds gave two different sets. Both are corrected and now give bit-identical output under both seeds.

Reported in the discussion paper	What we report now	Why
A minimum protocol of three requirements	Five requirements (§5.6), with a pre-registration template as §S2	The two additions are the ones this revision's own findings forced: register the exponent of the matching cost, and register the reference set.
The diagram distance described as Wasserstein-2	The degree-1 Wasserstein distance W_1	<code>persim.wasserstein</code> carries no exponent argument and returns the plain sum of matched costs under a Euclidean ground metric. The description misnamed the implementation while leaving the implementation itself correct; the true degree-2 distance and the exact bottleneck are now reported alongside it (§3.2, §4.5).
The event-subsample seed, treated as a fixed implementation detail	A specification axis, across which the verdict is not stable	Five further subsample seeds, with identical sampled point indices across representations, show that the combination of verdicts reported in the discussion paper occurs in one draw of six (§4.1, §S5).
The exponent of the matching cost, not identified as a choice	The largest single source of dependence we find	On this dataset it moves the event-level p-value from 0.003 to 0.087 while the event-to-control ratio moves by 2.5 % (§4.6).
Linear-encoding results computed on the archived longitude field without comment	The archived field mixes two longitude conventions	2.19 % of fixes are recorded east of +180° while others carry the negative representation of the same meridians, so the linear normalization places two representations of one meridian 1.0 apart. Reducing to a single convention moves the linear p-value from 0.037 to 0.108, and the branch cut spans 0.022 to 0.108 (§3.3.1, §S6).
The reference set of control diagrams, not identified as a choice	A third unregistered choice, of the same weight as the resampling resolution	Five alternative reference sets drawn under the same rule move the linear p-value over 0.036 to 0.047, a spread that drawing more null draws does not reduce (§3.6).

Statements in the interactive discussion that we withdraw

Six statements we made in the discussion phase do not survive the audit above. We quote each and say what replaces it.

1. *“The pooled H_1 false-positive count over 147 placebos ranges from 9 to 16 across implementations, with median 10 ...; the published value, 16, is the single failing tail draw.”* (AC1) — Withdrawn in its attribution. Two elements fail. Ascribing the discrepancy to control-order randomization alone is wrong for the linear row, whose normalizer was overwritten and cannot be audited at all; and the count of 16 cannot be placed among the auditable orderings, so it is not a draw from the distribution we now report. What we can report is 9 to 12 over six auditable orderings. Replaced by §4.2 and §S0. Separately, the H_0 channel does reach 16 of 147 at one of those six orderings; that is a different quantity that happens to share the count, and we flag the coincidence explicitly so that it is not read as a rediscovery of the withdrawn claim.
2. *“The corrected, deterministic pipeline at its median implementation is the first documented nominally-calibrated configuration of this class.”* (AC1) — Withdrawn entirely. It is a priority claim resting on the same seven-implementation campaign, and the audit leaves us unable to place that campaign’s pooled count among the auditable orderings. The revision reports that the 49-placebo calibration is nominal in H_1 at every auditable ordering and makes no claim of precedence.
3. *“The revision adds a fourth requirement to the minimum protocol: calibration claims must be reported as a distribution over control-resampling implementations with a pre-registered summary statistic.”* (AC1; restated as binding in AC2, reply to RC2.5) — Superseded, and in one respect reversed. The protocol now carries five requirements rather than the promised four, but reporting a distribution over orderings is not among them. Varying the control-draw ordering is a diagnostic check on requirement 5, not a reporting format, because reordering a uniformly resampled control pool selects a different realisation while the estimand does not move; what should be reported alongside the estimate is its Monte-Carlo interval. We set this out at §3.5 and §5.6 rather than leaving the earlier commitment to stand unexplained.
4. *“A small robust false-positive core (two placebo dates under two encodings) persists across implementations and is reported as such.”* (AC1) — Not carried into the revision. We can point to no auditable run that reproduces that core, and under the standard this revision applies to itself a number that cannot be pointed at is not reported. The threshold-straddling population quoted in the same passage is carried: §4.2 reports it at 11.6 % of the 147 combinations, alongside the dispersion asymmetry measured on the reproducible cache, where the coefficient of variation of the event-to-null distance is 2.0 to 2.2 times larger in H_1 than in H_0 .
5. *“Under this distance the verdict is single-valued and stable across seven control-draw orderings (per-ordering 2–4/49).”* (AC3, reply to RC3.1) — Withdrawn as stated. It combined two quantities. The verdict was tested on three control-draw orderings, where it stays within [0.006, 0.017]; the seven orderings and the 2–4/49 counts are the placebo calibration, which is a different quantity. And “single-valued” does not survive at all once the event-subsample axis is tested: the same verdict crosses α in one subsample draw of six. Replaced by §4.5 and §4.7.1, and the manuscript now says “stable across control-draw orderings” wherever it previously said “stable”.
6. *“A full multi-seed distribution with encoding-matched sample indices is left to a larger study.”* (AC3, reply to RC3.5) — Superseded. The scan was within reach and has been carried out, with identical sampled point indices across representations, exactly as Referee 3 asked. It is reported in §4.1 and §S5, and it is the reason §4.7.1 replaces the single realisation by an average over ten.

Referee #1

RC1.1

Referee’s point. *The manuscript diagnoses defects without proposing a fix. Can the author propose a method to overcome the flaws?*

Author’s response. This was the central weakness, and we have addressed it in two stages rather than one, because the first stage failed and we think the failure is informative.

The first stage followed the route the preprint itself proposed in its own closing section: repair the calibration at the level of the statistic. We pre-registered the criteria — a pooled exact binomial test of the 147-placebo H_1 false-positive count against nominal, with the decision rule fixed and deposited before the runs — and then tested the two candidates a follow-up would try first: a z-score normalization of the event-to-null distance, and an L^2 -persistence-landscape summary in place of the diagram distance. Neither restored calibration: the landscape statistic gives 12/147 and the z-score 14/147, against 9/147 for the baseline. We report this as a negative result rather than omitting it, since it is what pointed us at the input geometry instead.

The second stage is the fix. We replace the arbitrary coordinate encodings with a physically interpretable common distance — great-circle geometry on the sphere the storms live on, with an intensity weight — fed directly to the persistence computation. Under it the encoding ambiguity disappears and the verdict becomes single-valued across control-draw orderings and nominally calibrated. We then tested what that repair leaves untouched, and found two further unregistered choices that move the same verdict by comparable amounts, so the fix is reported as partial and is accompanied by a test that selects neither the exponent nor the subsample, together with a pre-registration protocol that would prevent the situation arising again.

Changes in the manuscript. New §3.3.6 (p. 6, l. 210) (physical distance) and §4.5 (p. 11, l. 385). New §4.7 (p. 14, l. 510), the test that selects no specification. New §5.6 (p. 19, l. 673), the five-requirement protocol, with the template as Supplementary §S2. The pre-registered remedy experiments and their decision rule are documented in Supplementary §S8.1; their outcome is in §4.2 (p. 9, l. 323).

RC1.2

Referee’s point. *The manuscript is too technical. Provide physical interpretation, and explain why the encodings disagree.*

Author’s response. We add a worked example in plain language at the start of §1.2 — six storms whose tracks ring a quiet patch of ocean, the ball around each point touching its neighbours before the middle fills — so that the object the method keeps is concrete before any formalism. And we give the reason the encodings disagree explicitly: re-encoding a periodic coordinate changes the pairwise distance and hence the input metric to the filtration, and nothing forces two embeddings of the same data to give the same diagram. An effect visible under one embedding and not others is therefore ambiguous between a coupling legible in one geometry and a metric-induced amplification, and topology alone cannot separate them. We state the consequence as a principle: representation invariance is necessary but not sufficient for physical attribution. Where the difference actually sits is now measured rather than argued: it is at intermediate persistence, and the single largest feature is at the 82nd percentile of its null.

Changes in the manuscript. §1.2 (p. 2, l. 40), worked example. §5.1 (p. 17, l. 584), second paragraph. §4.6.1 (p. 12, l. 437), the persistence profile.

RC1.3

Referee’s point. *Condense the physical introduction if the paper is methodological.*

Author’s response. Accepted. The solar–cyclone background is now a single paragraph of roughly 220 words, framed as motivation, and the introduction leads with the methodological problem. The whole manuscript is shorter than the discussion paper.

Changes in the manuscript. §1.1 (p. 1, l. 25), condensed; §1.2 (p. 2, l. 40), now the main line of the introduction.

Referee #2

RC2.1

Referee’s point. *The statistical, representational and causal levels of robustness are conflated; an effect could be encoding-specific rather than a failure of the tool. A synthetic injection of a known signal would calibrate detection power.*

Author’s response. We agree, and the three levels are now kept apart throughout: statistical robustness in §4.2, representational in §4.1 and §5.1, and causal in §4.5, §5.4 and §5.5. We no longer equate representation-inconsistency with tool failure.

On the synthetic injection: **we have not run it, and we say so plainly.** We attempted a synthetic-diagram benchmark twice and abandoned both attempts, for a reason worth recording. A generator that samples births and persistences independently does not reproduce the archived diagrams: in the real diagrams birth and persistence are strongly associated (Spearman 0.39 to 0.45), and a Gamma fit to the persistence distribution is rejected for every one of the six representations. Output produced by a generator that fails those checks cannot bear on the behaviour of the statistic on real diagrams, so we discarded it rather than reporting it. A benchmark that would be informative has to inject the perturbation at the level of the point cloud and run the full filtration, which is a study in its own right. We have added it to the manuscript as the priority follow-up, and we note that the present revision establishes the precondition — a physically interpretable, calibrated pipeline whose specification dependence is mapped — that would make such an injection interpretable.

Changes in the manuscript. §4.2 (p. 9, l. 323), §4.1 (p. 8, l. 303) and §5.1 (p. 17, l. 584), §4.5 (p. 11, l. 385) and §5.4 (p. 19, l. 648) for the three levels; §5.4, last sentence, naming the injection design as the priority follow-up and stating that it is beyond the present scope.

RC2.2

Referee’s point. *The H_1 inflation is described rather than explained. Is it an artefact of the Wasserstein-2 metric? Add a bottleneck comparison, persistence landscapes, and an H_0 -versus- H_1 morphology.*

Author’s response. All three are done, and answering the referee’s first question is what uncovered the naming error listed in the corrections table above. The statistic is the degree-1 Wasserstein distance W_1 , not W_2 ; the implementation was always the sum of matched costs under a Euclidean ground metric. Having established that, we did what the referee asked in the form the referee intended, and computed the true degree-2 distance and the exact bottleneck as separate metrics.

The outcome is now in one table. Under W_1 the event–control difference survives all four matched-null designs; under the true W_2 and under the exact bottleneck none of the twelve cells reaches α , though the

event sits between the 68th and the 87th percentile of its control distribution in every bottleneck cell. We therefore describe the residual as specification-conditional and not metric-robust, which is a weaker claim than the discussion paper made.

On the ordering sensitivity specifically, the exact bottleneck is if anything the more order-sensitive of the two: a sign test on the paired per-placebo ranges gives $p = 1.5 \times 10^{-6}$. On morphology, the H_1 channel carries a coefficient-of-variation excess over H_0 of 2.0 to 2.2 in the event-to-null distance and 3.9 to 4.4 in the null mean. We are careful about what those numbers license: they are measured across placebos, so they establish that the H_1 summaries move more from placebo to placebo, and they do not establish that the H_1 null is narrower within any one placebo. The discussion paper drew the stronger inference; we have withdrawn it and rewritten the passage as a co-occurrence.

Changes in the manuscript. §3.2 (p. 4, l. 148), naming and the two additional metrics. Table 5 and its discussion. §4.2 (p. 9, l. 323), the two candidate explanations. §5.2 (p. 17, l. 610), rewritten.

RC2.3

Referee's point. *Placebo ecological validity: ENSO or QBO state could confound the false-positive rate.*

Author's response. Three lines address this, and the substance is as we set it out in the interactive discussion. First, the same placebo dates — hence an identical climate state — change false-positive status purely when the control-draw order is permuted, so the variability across orderings cannot be climatic. Second, we do not treat the placebo count as a pure methodological false-positive rate: it is the pipeline's empirical rejection frequency under realistic historical variability, which folds in climate, era and basin variability, and the matched-null designs probe how much of the event-side signal that variability can carry. Third, the ENSO and solar state of the event and placebo windows are reported: the three events sit in the middle of the placebo ENSO distribution, so the event set is not ENSO-selected.

Changes in the manuscript. §4.2 (p. 9, l. 323), final paragraph; §4.5 (p. 11, l. 385); Supplementary §S5.

RC2.4

Referee's point. *Ambient dimension is confounded with periodicity; add a 4-D linear control.*

Author's response. Done, with the two controls kept apart because they answer different questions. A 4-D linear encoding with a constant fourth coordinate is a code-path identity check and reproduces the 3-D result exactly. A 4-D linear encoding with a Gaussian fourth coordinate is the real control: it leaves the placebo calibration nominal, so the calibration pattern is not a dimensional artefact, while it does move the event verdict from 0.037 to 0.072. We therefore concede the referee's point at the level of the §4.1 comparison and retain it only for the §4.2 diagnosis.

Changes in the manuscript. §3.3.5 (p. 5, l. 199).

RC2.5

Referee's point. *The proposed minimum protocol is not itself followed. Operationalize the agreement criterion.*

Author's response. The discussion paper proposed three requirements and our reply in the interactive discussion promised a fourth. The protocol now has five, and they are not the earlier three plus one. Two are additions this revision's own findings forced: register the exponent of the matching cost or report a statistic that selects none, and register the reference set. The fourth we promised — report calibration as a distribution

over control-draw orderings — is not among them, for the reason given in the corrections above and derived at §3.5: reordering a uniformly resampled control pool selects a different realisation while the estimand does not move, so the ordering spread is a diagnostic check on requirement 5 rather than the quantity to report. What requirement 5 asks to be reported alongside a p-value is its Monte-Carlo interval, separately for H_0 and H_1 . The agreement criterion is operational: representations agree when all of them place the verdict on the same side of α with overlapping Wilson intervals, and a physically interpretable distance overrides a disagreement only if its verdict is stable across the control-draw ordering, the subsample realisation *and* the reference set. We also state where the present analysis falls short of its own protocol: 49 placebos against the recommended 200, and 100 null draws per placebo.

Changes in the manuscript. §5.6 (p. 19, l. 673), five requirements and the shortfall paragraph; Supplementary §S2, the template, with §S2.9 the agreement rule.

RC2.6

Referee’s point. *The physical background is disproportionate to a non-physical conclusion.*

Author’s response. Accepted; see RC1.3.

Changes in the manuscript. §1.1 (p. 1, l. 25), §1.2 (p. 2, l. 40).

Referee #3

RC3.1

Referee’s point. *Use a common, physically defined distance — for example great-circle plus an intensity weight — rather than three geometric models.*

Author’s response. This became the revision’s main addition, and the present revision reports what it does and does not achieve. The distance is $d(i,j) = \sqrt{[(\gamma(i,j)/\pi)^2 + (\lambda \cdot |\Delta_{\text{wind}}|/200)^2]}$. It is not another arbitrary embedding: the great-circle term is the intrinsic geometry of the sphere the storms live on, and the intensity term is a dimensional balance whose single weight is sensitivity-tested at $\{0, 0.5, 1\}$.

What it achieves: on the axis on which the three encodings disagreed, it returns one value, and that value is stable across control-draw orderings, within $[0.006, 0.017]$.

What it does not achieve, and here we withdraw the description we gave in the interactive discussion: the verdict is **not** single-valued in the wider sense we then claimed. It moves with the exponent of the matching cost (0.003 to 0.087), with the event subsample (0.012 to 0.069) and with the reference set (0.036 to 0.047), and it is absent under the true W_2 and the exact bottleneck. The title of the manuscript now states this asymmetry directly, and the sensitivity test the referee’s distance made possible remains informative: under a single subsample the difference is carried by the intensity channel, although after averaging over subsamples the gradient across the weight compresses to a factor of 1.75 and the two weighted settings are no longer separable.

Changes in the manuscript. Title. §3.3.6 (p. 6, l. 210), §4.5 (p. 11, l. 385), §4.6 (p. 12, l. 433), §4.7 (p. 14, l. 510), §5.3 (p. 18, l. 622).

RC3.2

Referee’s point. *A matched null preserving per-window count and basin. Clarify which fixes enter the ensemble.*

Author’s response. Done, and extended beyond what the referee asked. The control draw preserves the event ensemble’s per-window storm counts (11/9/5), its genesis basin, its genesis era in 5-year bins, and the joint basin-by-era composition. All four are now reported in one table under three metrics rather than one, which is what shows the residual to be metric-conditional. The four are non-monotonic and informatively so: matching basin alone sharpens the difference, matching era alone weakens it most, which points to era as the principal confounder. On fix inclusion, §2 states that the ensemble uses all lifetime fixes of each selected storm, that the storm rather than the fix is the unit of exposure and of resampling, that this includes out-of-window fixes and weights longer-lived storms more heavily, and that a window-cropped variant is a natural sensitivity for a larger study.

Changes in the manuscript. §2 (p. 3, l. 106), event windows; Table 5 and its discussion.

RC3.3

Referee’s point. *Per-proportion Wilson intervals; a paired H_0 -versus- H_1 comparison; more placebos and more permutations.*

Author’s response. The Wilson intervals and the paired comparison were added at the first revision and are retained, with the qualification kept: the H_1 excess over H_0 is modest and representation-specific, and under the linear encoding H_1 sits below H_0 , so we claim no general inflation of H_1 over H_0 . The present revision goes further on the referee’s third point in a direction we did not anticipate. Reporting the H_0 channel as a distribution over the same six orderings, rather than as the single value the discussion paper quoted, shows the H_0 spread to be the wider of the two, running 4 to 16 of 147 against 9 to 12 for H_1 . That withdraws the reading, carried in the discussion paper, that only the H_1 swing crosses the threshold. What distinguishes the channels is that the H_1 verdict sits near α at every ordering while the H_0 verdict is far from it at most. We agree that 100 null draws per placebo is small, and we now report where that matters: a per-placebo decomposition attributes roughly half to two-thirds of the cross-ordering movement to the finite-resampling term, so raising it would compress the spread. Separately, the five load-bearing event results are recomputed at 10,000 null draws, which tightens their Monte-Carlo intervals by about a factor of three, and none of the tightened intervals contains 0.05.

Changes in the manuscript. §4.2 (p. 9, l. 323) and Table 2, both channels; §5.2 (p. 17, l. 610); §3.5 (p. 6, l. 227), Monte-Carlo uncertainty and the 10,000-draw recomputation; Supplementary §S1.

RC3.4

Referee’s point. *Multiplying by a ratio is not a calibrated p -value; use an empirical null, and apply the finite-sample correction.*

Author’s response. Accepted in full. The ratio scaling was removed; the calibration is the placebo empirical null, 49 placebo tests per representation compared to nominal with the exact binomial. On the finite-sample correction, our reply in the interactive discussion stated that it “shifts every value by < 0.001 ”, which is true only at 1000 null draws; the correction’s size depends on the number of draws, and we now state it by regime — below 0.001 at 1000 draws, 0.002 at 500, and 0.0099 at the 100 used per placebo. It changes no qualitative verdict and no placebo classification.

Changes in the manuscript. §3.5 (p. 6, l. 227).

RC3.5

Referee’s point. *The subsample sizes are single-seed. Repeat over seeds and report distributions.*

Author’s response. We agree that the non-monotonic values reflect sampling variability and presented the table as one seeded realization per cell, and deferred the multi-seed distribution. We have now done it, in the exact form the referee specified: five further event subsample seeds with **identical sampled point indices across encodings**, everything else held fixed, and a positive control that reproduces the archived rows bit for bit.

The outcome is more consequential than we expected, and it is the reason the manuscript’s central claim is now qualified. The verdicts are not stable across this axis: the great-circle verdict crosses α in one draw of six, the linear encoding in one, the sphere in three, the cylinder in none, and the frozen seed is the outlier of the six at 2.91 standard deviations against 1.16 to 1.75. The relative order of the three encodings, by contrast, holds under every seed. Because a subsample is a Monte-Carlo realisation rather than an analyst’s choice, we do not test over it; we average over it, recomputing both sides symmetrically over ten independent subsamples. Doing so moves the great-circle basin-matched p-value from 0.003 to 0.044, and the unmatched from 0.012 to 0.069.

Changes in the manuscript. §4.1 (p. 8, l. 303) and Table 1, six-seed range column; §4.4 (p. 10, l. 374); §4.7.1 (p. 15, l. 522) and Table 8; Supplementary §S5; Supplementary §S2.4, the corresponding pre-registration requirement.

Data and code

All analysis code, run logs and derived tables reproducing every value in the manuscript are archived at Zenodo under the MIT licence, with the DOI given in the Code and data availability section. The archive includes a verification script that re-derives all 87 registered specifications from the archived data and compares each with the value printed in the manuscript, together with the frozen outputs of the superseded runs listed in the corrections table above, so that every correction in this letter can be inspected rather than taken on trust.