

Supplementary methods and numerical results

Calibrating chemical and kinematic dependence in small stellar catalogues

Nick Barua and Koji Oda

Companion to Online Resource 1 for Astrophysics and Space Science

Corresponding author: Nick Barua · s.nick.barua@gmail.com

AN Holdings Co., Nishinomiya, Japan; Department of Legal Medicine, Shiga University of Medical Science, Otsu, Japan

This document specifies the current simulation, permutation and sensitivity analyses. Its PDF is the publication copy; the matching Word file is supplied as an editable source. ESM_1.zip and Supplementary_Data_S1.zip contain identical bytes; paths below are relative to the extracted ESM_1 folder. Section S6 distinguishes an earlier study retained for provenance. Its different experiments are not pooled with the current results.

S1 Current simulation design

Catalogues have four chemical and three kinematic coordinates. Calibration uses $N = 78, 156, 500$ and 2000 ; $\rho = 0, 0.2, 0.4$ and 0.6 ; two geometries; and 400 catalogues per cell, totalling $12,800$. The Gaussian channel and invertible velocity transformation are defined in manuscript Eqs. (6)–(7), with $\beta = 0$ or 1 . The transformation preserves information. GC and KSG means, bias, standard deviations and RMSE are evaluated against the analytic information.

The original testing grid contains $2,200$ distinct catalogues. Independent, linear ($\rho = 0.35$) and quadratic cases each use 200 catalogues at $N = 78, 156$ and 500 . Six-group cases add 200 at $N = 78$ and 200 at $N = 156$. Each grouped catalogue receives global and within-group tests; two null schemes do not create additional independent catalogues.

The quadratic generator in manuscript Eq. (8) drives only the first velocity coordinate: a centred, unit-variance squared first chemical coordinate has coefficient 0.8 , with independent Gaussian noise coefficient 0.6 . Other velocity coordinates remain independent standard Gaussians. The alternatives are not information-matched. The intersection requires both GC and KSG to reject the same null at 0.05 .

The calibration seed sequence begins at 20261003 and the original testing sequence begins at 21261003 . Every catalogue seed is stored in its replicate row. For the original synthetic testing grid and quadratic follow-up, the permutation seed is the catalogue seed plus 100000000 . A separate follow-up uses $5,000$ quadratic catalogues at $N = 78$, with seeds 202610070000 through 202610074999 and the same generator, preprocessing, estimators and 199 permutations. The replicate count and seed sequence were fixed before inspecting these new outcomes. These analyses followed the original observational results and were not preregistered. There are $7,200$ current testing catalogues in total, separate from the $12,800$ calibration catalogues.

Table S1. Quadratic follow-up at $N = 78$. Intervals are 95% Wilson intervals for Monte Carlo rejection fractions, not uncertainty intervals for stellar measurements. The original 200-catalogue estimates remain archived.

Method	Original rate 200 catalogues	Follow-up count	Follow-up rate 95% Wilson interval
GC	5.0%	285 / 5,000	5.70% (5.09–6.38%)
KSG $k = 5$	31.0%	1,685 / 5,000	33.70% (32.40–35.02%)
Distance correlation	32.5%	1,499 / 5,000	29.98% (28.73–31.26%)
GC and KSG intersection	3.5%	170 / 5,000	3.40% (2.93–3.94%)

Figure 2 uses the 5,000-catalogue follow-up for all four methods at quadratic $N = 78$ and the original 200-catalogue results at every other point. The files `results/testing_summary.csv` and `results/testing_replicates.csv` remain the original experiment; the plotting script reads the replacement point from `quadratic_power_audit/new_seed_summary.csv`.

S2 Statistics and permutation counts

GC is the uncorrected plug-in covariance determinant score after marginal mid-rank normalisation. It is reported in nats and is not a universal information bound. KSG uses estimator 1 with sample-standardised coordinates, maximum-norm distances, strict marginal neighbour counts and $k = 5$ unless stated otherwise. Negative estimates are retained. No undocumented random jitter is applied to the baseline. Transformations are cached; rounded observations can be sensitive to floating-point decisions at tied neighbour boundaries.

The current manuscript and archive report distance correlation itself: the non-negative square root of the squared statistic in manuscript Eq. (4). For the 78-star sample, distance correlation is 0.4504546771 and its square is 0.2029094161. Both are dimensionless. The earlier archive in Section S6 stores the squared statistic under `dCor2`; its scale must be read accordingly. Squaring or taking the square root preserves the ordering of these non-negative statistics and therefore the permutation rejection event for identical samples and shuffles.

Each permutation moves an intact three-coordinate velocity row. The p value is one plus the number of permuted statistics at least as large as the observed statistic, divided by $B + 1$. Tests reject at $p \leq 0.05$. Methods share permutations within an analysis. Global permutations test marginal independence; restricted permutations require exchangeability within the declared groups. Centring by the permutation mean is descriptive and does not estimate conditional information or remove measurement error.

Table S2. Exact permutation counts in the current manuscript. B counts random shuffles; the observed statistic supplies the additional one in the denominator.

Analysis	B	p value denominator
Original synthetic testing and quadratic follow-up	199	200
Observed global comparison	1,000	1,001
Principal instrument by metallicity tertiles	2,000	2,001
KSG neighbour grid $k = 3, 4, 5, 6, 8, 10$	2,000	2,001
Each alternative null design and removal of [Fe/H]	1,999	2,000
Each binary-pair exclusion	1,999	2,000
Each rounding or added-noise draw	499	500

Alternative designs are instrument-only, metallicity-tertiles-only, instrument by halves, and instrument by quartiles. Removing [Fe/H] retains the baseline instrument-by-tertiles labels. Holm adjustment applies only to the six-value neighbour-order family. Complete null arrays are stored for the seven observational schemes in `results/observed_null_*.npz`; other experiments retain observed scores, p values and reproducible seeds rather than every shuffle statistic.

S3 Catalogue and observational sensitivities

The source is Nissen and Schuster (2010), CDS/VizieR J/A+A/511/L10, DOI 10.26093/cds/vizier.35119010. The 100 instrument observations include six duplicate stars. Preferring UVES leaves 94 unique objects. All 658 adopted numerical cells and instrument labels match the included CDS export. The speed cut above 180 km/s retains 78 stars. The chemical coordinates are [Fe/H], [Mg/Fe], [Si/Fe] and [Ni/Fe]; the velocity coordinates are U, V and W. No chemical population label is used as an analysis target.

The principal strata cross instrument with pandas qcut metallicity tertiles on the 78 retained stars. FIES cell sizes are 12, 14 and 15; UVES sizes are 16, 11 and 10. The current analysis table contains only identifiers, instrument and the seven adopted variables. Extra metadata from earlier transcriptions are not used. There are no duplicate seven-variable rows, although individual coordinates contain repeated rounded values.

Rounding sensitivity uses 100 independent uniform perturbations within ± 0.005 dex for each chemical coordinate and ± 0.5 km/s for each velocity coordinate. Added-noise sensitivity separately uses 100 independent Gaussian perturbations with chemical standard deviations (0.035, 0.035, 0.035, 0.010) dex and velocity standard deviations (12, 16, 9) km/s. Each draw recomputes its statistic and 499-permutation null. Original sample membership and stratum labels remain fixed. These perturbations do not represent new independent stellar observations, propagate star-specific covariances, or provide sampling confidence intervals.

Table S3. Current perturbation results. Rejection counts are out of 100 draws per exercise; p value ranges are empirical 5th–95th percentiles across the dependent sensitivity analyses.

Sensitivity	Method	Rejections	p value percentile range
Publication rounding	GC	0 / 100	0.4878–0.6401
Publication rounding	KSG $k = 5$	100 / 100	0.0039–0.0180
Added Gaussian noise	GC	2 / 100	0.1152–0.6292
Added Gaussian noise	KSG $k = 5$	99 / 100	0.0020–0.0200

Each binary sensitivity excludes G112-43, G112-44 or both and preserves the original metallicity-bin assignments of the remaining stars. It uses 1,999 restricted permutations. Bins are not recalculated. The differences from the earlier rebinned, 2,000-permutation exercise are described in Section S6. Removing the known pair does not establish independence of every remaining star.

Table S4. Current binary exclusions using the fixed baseline strata.

Excluded	N	GC p	KSG p	Distance correlation p
G112-43	77	0.6435	0.0170	0.0020
G112-44	77	0.5925	0.0150	0.0010
Both components	76	0.6450	0.0025	0.0030

S4 Selection and Monte Carlo interpretation

The selection illustration starts from independent scalar standard Gaussians and applies the joint logistic inclusion probability in manuscript Eq. (9). It is not a model of the stellar speed cut. Direct selected-density integration is evaluated at five sharpnesses, seven thresholds and Gauss–Hermite orders 80, 160, 320 and 640 per dimension. At sharpness 8 and threshold 1.5, order 640 gives selected fraction 0.1475, mutual information 0.3764 nats and raw-covariance Gaussian score 0.2714 nats. That covariance score does not include rank normalisation and must not be labelled GC or substituted for the information integral.

Wilson intervals describe finite simulation precision. For the original $N = 78$ within-group null, GC, KSG and distance correlation reject in 14, 16 and 15 of 200 catalogues. Their respective intervals are 4.22–11.41%, 4.98–12.60% and 4.60–12.00%. These results are consistent with 5%, with substantial uncertainty at 200 replicates; they do not establish exact size at every setting. Global rejection in the group-shift simulation is detection of genuine marginal dependence, not a false positive for marginal independence.

S5 Execution and output provenance

The current scripts require Python 3.12 with NumPy 2.3.5, SciPy 1.17.0, pandas 2.2.3 and Matplotlib 3.10.8, as recorded in requirements.txt. All included data can be analysed offline after dependencies are installed. Run commands from the extracted ESM_1 directory. Preserve an unmodified copy: reproduction writes over result files in the working copy.

```
python verify_checksums.py
python verify_catalogue.py
python check_implementation.py
python make_figures.py
```

These commands check archive integrity, source-data agreement and implementation behaviour, then regenerate the six current figures from recorded results. They do not rerun the complete simulation suite. For complete reproduction, use run_all.py with --workers 6 and set OPENBLAS_NUM_THREADS=1 and OMP_NUM_THREADS=1. The separate follow-up can be rerun with quadratic_power_audit/run_new_seed_check.py --workers 6 under the same thread settings. README.md maps each script to its outputs. Seeds and environment records accompany the calculations; runtimes are environment-specific.

The recorded full rerun reproduced all 11 original result CSVs and seven observational permutation arrays. The separate follow-up retains 20,000 method-level rows from 5,000 independent catalogues. Figure 2 explicitly combines the follow-up point with the original grid; all other current figures use the corresponding archived results. The earlier archive described below is not invoked by run_all.py and is not included in the current 12,800 or 7,200 catalogue totals.

OpenAI ChatGPT/Codex assisted analysis, code generation and revision, execution, drafting, and figure preparation. A critique produced with Claude prompted further checking and the larger quadratic experiment. This assistance extended beyond copy editing. The current manuscript gives the full disclosure and author contributions; the human authors remain responsible for interpretation and submitted content.

S6 Earlier validation retained separately

The provenance directory preserves the supplied earlier 69-file archive without changing its bytes, together with the separately supplied original Word companion. The earlier study is internally consistent but uses a different design. Its own README, figure labels and Supplementary_Methods.docx describe that earlier revision. They are historical records, not the active methods or figures for the current manuscript. The active companion is this document; the active figure-generation script is `make_figures.py` in the `ESM_1` root.

The earlier simulation uses seven scenarios with 500 catalogues each at $N = 78$: independence; Gaussian channels with correlations 0.3, 0.6 and 0.9; a triangular chemical transformation with coefficient 1.2; a three-coordinate quadratic alternative; and a different six-stratum mean-shift construction. The shared-stratum case is tested under two nulls, giving 3,500 generated catalogues, 4,000 catalogue–scheme combinations and 12,000 method-level rows. It uses seed 20261003 and 199 permutations. Its tables and exact generator definitions remain in the preserved archive.

The earlier quadratic generator places a centred, unit-variance squared chemical term in each of the three velocity coordinates, with signal coefficient 1 and independent noise coefficient 0.75. Its KSG rejection rate is 98.8% (494/500). The current generator places signal in only one coordinate and gives 33.70% (1,685/5,000) in the larger follow-up. These rates concern different distributions and must not be treated as a reproduction discrepancy or pooled into one estimate. No intersection result was recorded in the earlier three-method study.

The earlier rounding checks use two sets of 200 perturbations and 199 shared restricted shuffles; the current rounding and added-noise checks each use 100 draws and a separately seeded 499-permutation null per draw. The earlier binary exclusions recalculate metallicity bins and use 2,000 permutations; the current exclusions preserve the original bins and use 1,999. Earlier `dCor2` outputs use the squared statistic. These choices explain why similarly named tables contain different values.

The earlier record also retains raw CDS tables, checks on four unused HD193901 metadata fields, the tied-neighbour audit, backend comparisons, timing records, and its five historical article figures. These supplementary diagnostics are preserved for traceability and do not replace the current source table, six figures or sensitivity outputs. Its lightweight verification and all 68 listed checksums passed when the record was incorporated. No earlier Monte Carlo values were changed.

To inspect or reproduce that earlier study, extract `provenance/earlier_validation_20261003.zip` into a separate working directory and follow its own README.txt. The current archive manifest verifies the preserved ZIP as a single file; its internal SHA256SUMS.txt verifies the 68 contents listed within it. The original standalone Word file is `provenance/Supplementary_Methods_original.docx`.