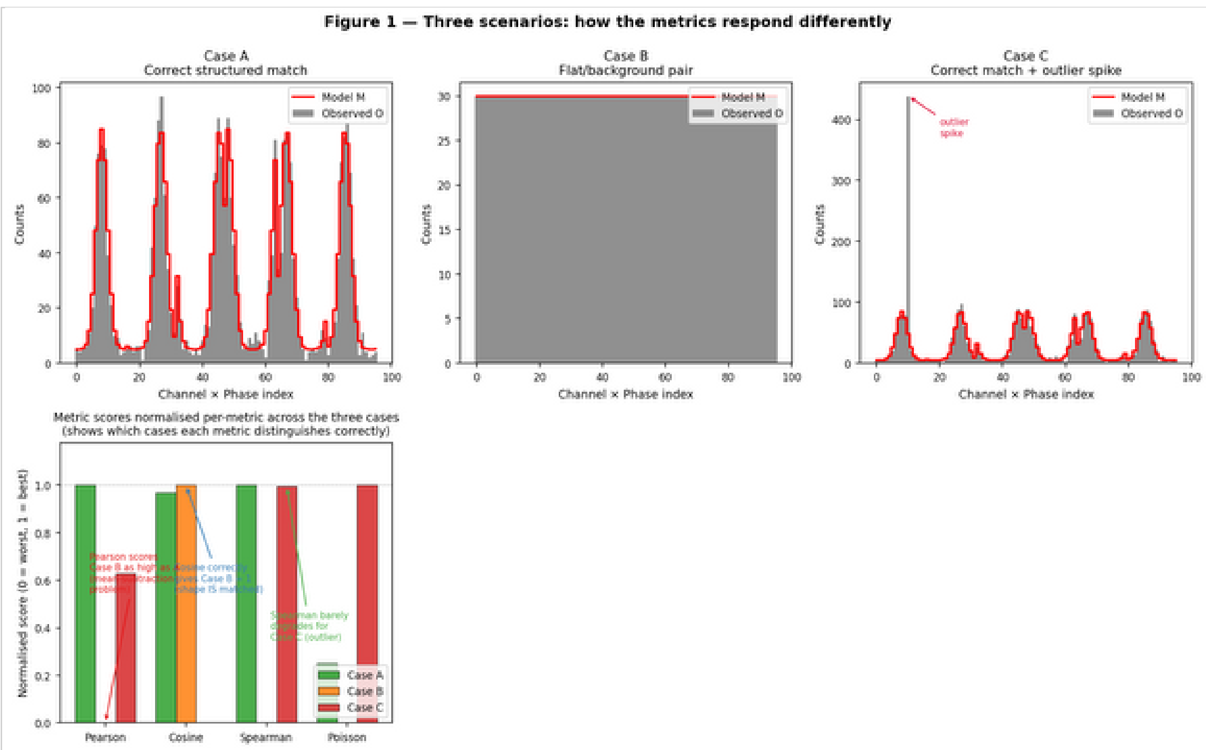
Gamma Reconstruction — Figure-of-Merit Metric Rationale

> [Note] Visualisations — run python visualise_metrics.py to regenerate the figures referenced throughout this document.

Background

The model-correlation reconstruction strategy works by scoring each candidate source position in a 2D spatial grid against the accumulated detector counts. For each candidate position, a model prediction vector is formed by concatenating, across all active coded-aperture phases, the expected detector response for that position and phase. This is compared against the corresponding observed count vector (the measured detector channel sums for each phase, integrated over the selected energy bin range).

The pixel whose model best matches the observation gets the highest score, and the resulting score map — after bicubic up-sampling — forms the reconstructed image. The choice of figure-of-merit (FOM) used to score that agreement is therefore central to reconstruction quality.



Three-scenario comparison

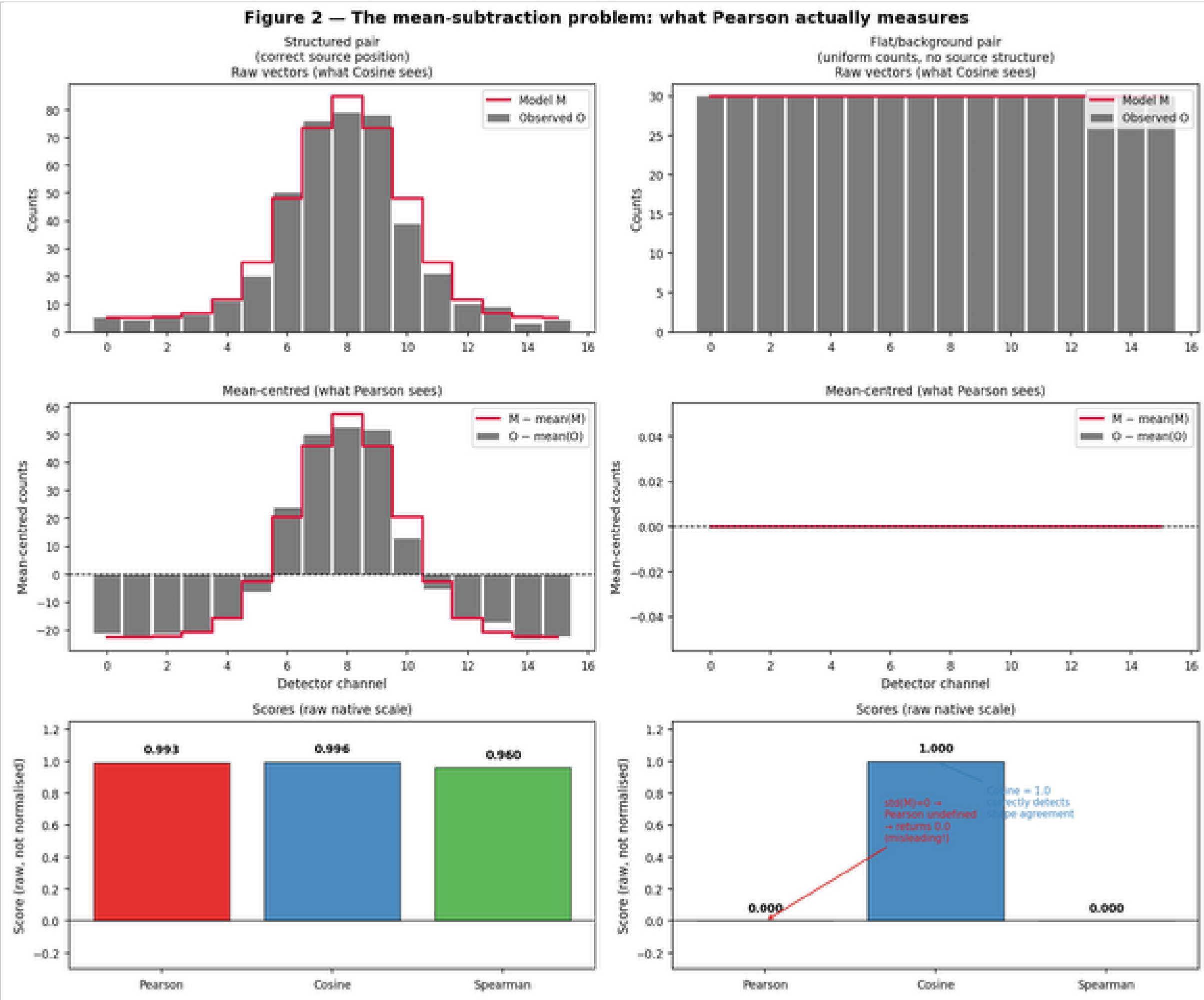
Figure 1 — Three scenarios showing how the four metrics respond differently. Case A: correct structured match. Case B: flat background pair (Pearson's mean-subtraction problem). Case C: correct match with a noise spike (Spearman's outlier robustness).

Why Pearson Correlation Is Not Ideal

The Pearson correlation coefficient was the original metric. It is defined as:

```
r = cov(O, M) / (σ_O · σ_M)
```

where O is the observed vector and M is the model prediction vector. While familiar and easy to compute, it has several properties that make it a poor choice for this application.



Mean-subtraction problem

Figure 2 — What Pearson actually operates on vs what Cosine operates on. For a flat/uniform pair (right column), mean-subtraction leaves only noise, making std(M) = 0 so Pearson returns 0 even though the shapes agree perfectly. Cosine correctly returns 1.0.

1. Mean-subtraction distorts sparse or imbalanced phase data

Pearson centres both vectors around their respective means before computing the inner product. In the context of coded-aperture gamma imaging, the observed counts per phase vary substantially — some rotation angles place the coded-aperture mask over high-sensitivity detector elements and others do not. Mean-subtraction can flip the sign of low-count phases (making a phase with genuinely few counts appear as a negative contribution) and can inflate the apparent agreement between a model that predicts near-zero counts for some phases and an observation that also has near-zero counts, when what has actually been measured is just background noise.

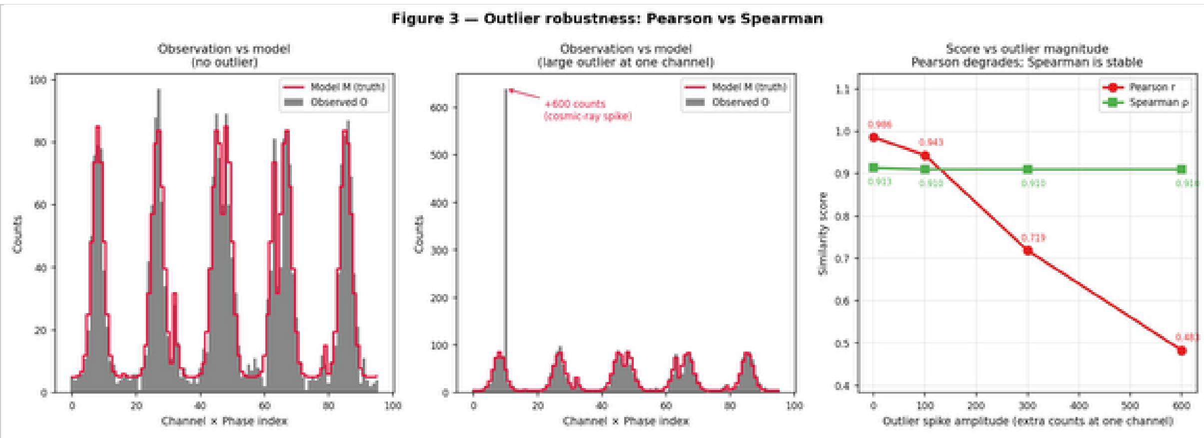
2. Incorrect noise model

Photon detection is a counting process governed by Poisson statistics: the variance of a count is equal to its expected value, not a fixed constant. Pearson correlation implicitly assumes Gaussian errors with uniform variance (i.e., ordinary least-squares geometry in the normalised space). This mismatch means that high-count bins and low-count bins are treated as equally reliable, even though high-count bins carry proportionally more statistical information. The result is that rare but large fluctuations in low-count channels can disproportionately affect the score.

3. Scale invariance is inappropriate here

Pearson is invariant to multiplicative rescaling of either vector. In principle this seems convenient (the absolute source intensity should not matter for localisation), but in practice it means that a model predicting a uniform, featureless count distribution can score highly against an observed vector that is also approximately uniform, even though no useful shape information has been matched. Put differently, Pearson can mistake flat agreement for structured agreement.

4. Sensitivity to outliers



Outlier robustness

Figure 3 — Pearson score (red) degrades rapidly as a single-channel noise spike grows. Spearman (green) is almost unaffected because the spike only displaces one rank rather than inflating the variance term.

A single detector channel with an unusually high count — due to, for example, a cosmic-ray event or electronic noise — has a large squared deviation from the mean and therefore drives the standard deviation term in the denominator. This shrinks the Pearson r value for otherwise well-matched vectors and degrades the score of the correct source position.

Alternative Metrics

Three alternative figures of merit are implemented, selectable via --reconstruction-metric.

Cosine Similarity (--reconstruction-metric cosine)

```
cos(O, M) = (O · M) / (‖O‖ · ‖M‖)
```

Cosine similarity measures the angle between the two vectors in the raw count space, without subtracting the mean. For strictly non-negative count data (as produced by a detector) the result lies in [0, 1], making it straightforward to interpret.

Why it improves on Pearson:

- No mean-subtraction. Phases with genuinely high counts contribute a large dot-product term; phases with near-zero counts contribute near zero. The metric therefore directly rewards matching the flux distribution shape across phases, which is precisely what localises the source.
- Naturally scale-invariant (via ℓ_2 normalisation) in a way that preserves the relative phase structure, unlike Pearson's mean-centred normalisation.
- Robust against overall count-rate offsets (e.g., changes in source activity between acquisition segments).

Limitation: Like Pearson, it does not respect the Poisson variance structure — all channels are weighted equally in the dot product regardless of their statistical uncertainty.

Spearman Rank Correlation (--reconstruction-metric spearman)

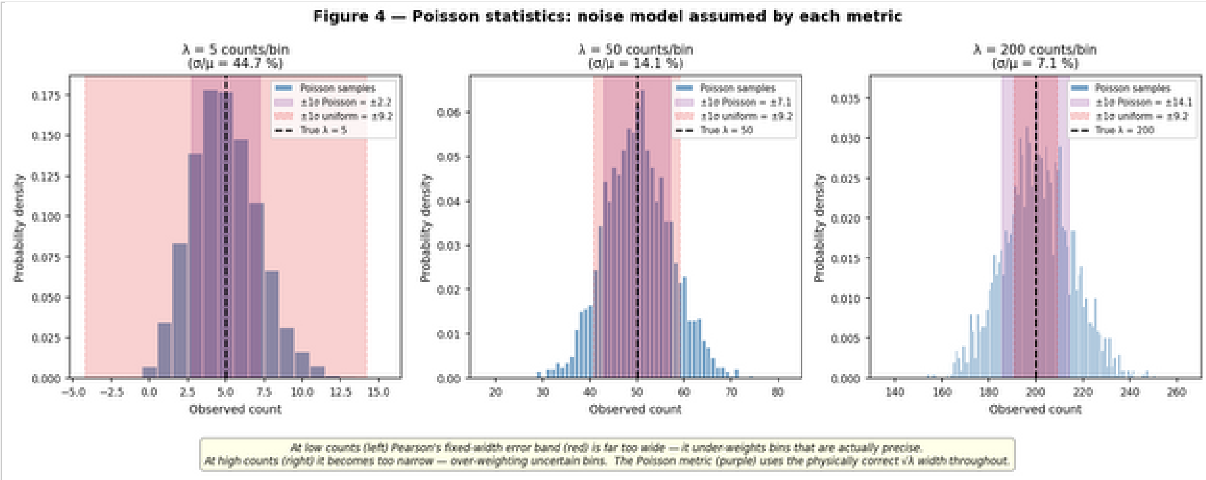
Spearman rank correlation \rho is computed by replacing each element of both vectors with its rank within that vector, and then computing Pearson correlation on the rank-transformed values.

Why it improves on Pearson:

- Outlier robustness. A single anomalously large channel value is mapped to a high rank, but its numerical magnitude no longer dominates the calculation. The reconstruction image is therefore much less sensitive to sporadic detector noise.
- Monotone rather than strictly linear agreement. The reconstruction quality relies on the phases being ordered by count rate in a way that mirrors the model prediction; Spearman captures this ordering faithfully even when the relationship between observation and model is non-linear (e.g., due to detector saturation or non-uniform flat-field corrections).
- Retains a familiar [-1, 1] range and zero-under-independence interpretation.

Limitation: Rank transformation loses absolute count magnitude information. Two candidate positions whose models predict different amounts of flux modulation across phases but the same rank ordering will receive identical Spearman scores.

Profile Poisson Log-Likelihood (--reconstruction-metric poisson)



Poisson noise model

Figure 4 — Poisson count distributions at three intensity levels. The Poisson-correct ±1σ band (purple) scales with √λ. Pearson's implicit uniform error band (red) is far too wide at low counts (under-weighting precise measurements) and too narrow at high counts (over-weighting noisy ones).

This is the statistically optimal figure of merit for photon-counting data.

The Poisson likelihood for observing counts O_i when the expected count is λ_i is:

```
L = ∏_i  exp(−λ_i) · λ_i^{O_i} / O_i!
```

The model prediction for pixel position p gives expected counts proportional to M_i. Because the absolute source intensity is unknown, a single scale factor α is introduced: λ_i = α M_i. The maximum-likelihood value of α is the solution to ∂ \ln L / ∂ α = 0, giving:

```
αˆ = Σ O_i / Σ M_i
```

Substituting this back into the log-likelihood yields the **profile log-likelihood**:

```
ll_profile = Σ_i [ O_i · log(αˆ M_i) − αˆ M_i ]
```

which depends only on the shape of the model vector, not its overall normalisation. The score stored in the reconstruction image is this quantity divided by the number of valid terms (so that images accumulated over different numbers of active phases remain comparable).

Why this is more optimal than Pearson:

- Correct noise model. The score is derived directly from the Poisson

probability of the observed data given the model shape. Bins with high expected counts contribute proportionally more to the score, correctly reflecting their greater statistical weight.

- Scale-free by construction. The unknown source intensity is profiled

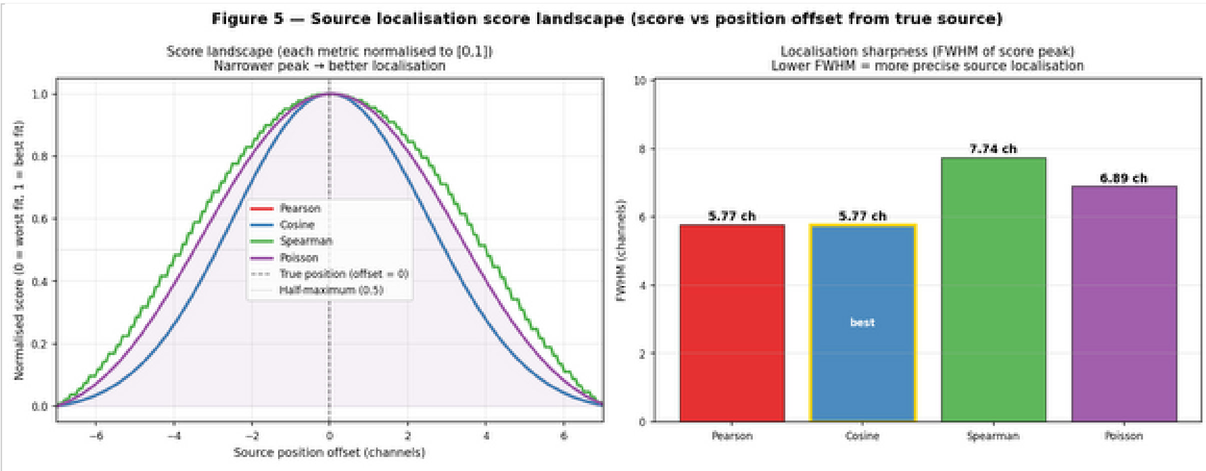
out analytically, with no need to normalise both vectors post-hoc.

- Sensitivity to shape, not offset. The log term O_i · log(λ_i)

rewards having the model concentrate predicted counts where observed counts are high; the −λ_i penalty discourages over-predicting empty channels.

- Asymptotically efficient. Among all unbiased statistics, the likelihood-ratio test (of which this is a component) achieves the minimum possible variance in the large-sample limit (Cramér–Rao bound).

Limitation: The score is in log-likelihood units, which are not bounded to [0, 1] or [-1, 1], so the reconstructed image values are less immediately interpretable without normalisation. The metric also requires $\lambda_i > 0$ (channels where the model predicts zero counts are excluded from the sum).



Score landscape

Figure 5 — Score vs source-position offset from the true location (toy 1D model, Poisson-noisy observation). The narrower the peak the more precisely the metric can localise the source. FWHM bars (right) give a quantitative summary; the metric with the lowest FWHM is highlighted.

Metric Comparison Summary

Property	Pearson	Cosine	Spearman	Poisson
Mean-subtracted	Yes	No	Yes (ranks)	No
Correct noise model (Poisson)	No	No	No	Yes
Outlier robust	No	Moderate	Yes	Moderate
Scale-invariant	Yes	Yes	Yes	Yes (profiled)
Bounded output	[-1,1]	[0,1]	[-1,1]	Unbounded
Optimal for count data	No	No	No	Yes

The current default is cosine. For data with well-calibrated flat-field corrections and sufficient statistics per phase, poisson is expected to give the best source localisation because it is the only metric derived from the true generative model of the data. spearman is recommended when detector noise or occasional high-count outliers are suspected. pearson is retained for reproducibility comparisons with prior results only.