LSST-like Lensing Dataset: Generator & Metrics

(for advisor meeting)

October 9, 2025

Setup & Notation (simple view)

- We simulate three LSST bands: g, r, i with PSF and 10-year coadd.
- For each sample we render:
  - $\mathbf{I}_{\text{subA}}$: with subhalo, with noise.
  - $\mathbf{I}_{\text{cleanA}}$: no subhalo, with the *same* noise seed.
  - $\mathbf{I}_{\text{cleanB}}$: no subhalo, with an *independent* noise seed.
- Residuals per pixel (vector over the 3 bands):

$$\mathbf{r}_{\text{sub}} = \mathbf{I}_{\text{subA}} - \mathbf{I}_{\text{cleanA}}, \qquad \mathbf{r}_{\text{null}} = \mathbf{I}_{\text{cleanB}} - \mathbf{I}_{\text{cleanA}}.$$

$$R_{\text{sub}}(x, y) = \|\mathbf{r}_{\text{sub}}(x, y)\|_2, \quad R_{\text{null}}(x, y) = \|\mathbf{r}_{\text{null}}(x, y)\|_2.$$

What we save as `source_rgb`

- **Source-only** rendering (no lens), same grid and PSF as images.
- LSST-like color composite with a simple $\sqrt{\cdot}$ stretch at the 95th percentile.
- Channel map: RGB $\leftarrow (i, r, g)$.

$$\mathrm{RGB} = \left(\sqrt{i/i_{95}},\ \sqrt{r/r_{95}},\ \sqrt{g/g_{95}}\right).$$

Old vs New Generator (at a glance)

Old

- Used a **ring mask** near θ_E.
- SNR computed on the ring only.
- Sometimes no source-only visualization.

New (this work)

- **No mask**: we use the **full image**.
- SNR is global (more stable).
- We save `source_rgb` for quick checks.
- Noise pairing kept (same band seeds in A; different in B).

SNR proxy (old)

- Previously, SNR was measured on an annulus around the Einstein ring:

$$\mathrm{SNR}_{\mathsf{old}} = \frac{\mathsf{RMS}_{\mathsf{ring}}(R_{\mathsf{sub}})}{\mathsf{RMS}_{\mathsf{ring}}(R_{\mathsf{null}})}.$$

- This depends on the mask size and location.

SNR proxy (now, full image)

- We compute SNR on the **entire frame**:

$$\mathrm{SNR} = \frac{\mathsf{RMS}_{\mathsf{full}}(R_{\mathsf{sub}})}{\mathsf{RMS}_{\mathsf{full}}(R_{\mathsf{null}})}, \quad \mathsf{RMS}_{\mathsf{full}}(A) = \sqrt{\frac{1}{HW}\sum_{x,y} A(x,y)^2}.$$

- Simple idea: "how different subA is from cleanA" compared to typical noise (cleanB - cleanA).
- Threshold for detectability: $y = \mathbb{I}\{\mathrm{SNR} \geq T_{\mathsf{snr}}\}$ (e.g., $T_{\mathsf{snr}} = 20$).

How SNR relates to χ^2 here

- If we use the paired null to estimate variance, a global chi-square-like score is:

$$\chi^2_{\mathsf{signal}} \;\propto\; \sum_{x,y} \frac{R_{\mathsf{sub}}(x,y)^2}{R_{\mathsf{null}}(x,y)^2}.$$

- Our RMS-based SNR collapses this to a single ratio:

$$\boxed{\chi^2_{\mathsf{signal}} \equiv \mathrm{SNR}^2}.$$

- This χ^2_{signal} is a **global detection statistic** (not divided by DoF).

How we compute χ^2 (this code)

$$\boxed{\chi^2_{\text{signal}} = \mathrm{SNR}^2 = \left(\frac{\mathrm{RMS}_{\text{full}}(R_{\text{sub}})}{\mathrm{RMS}_{\text{full}}(R_{\text{null}})}\right)^2}.$$

- No DoF normalization here (χ^2_{reduced} would be $\chi^2_{\text{signal}}/\mathrm{DoF}$).
- Goal: a robust, simple **one-number** "is there a signal?" indicator.

Scores (optional)

- **Binary**: $y = \mathbb{I}\{\mathrm{SNR} \geq T_{\mathsf{snr}}\}$.
- **Continuous (from SNR)**:

$$\mathsf{score}_{\mathrm{SNR}} = \sigma(a\,[\log_{10}(\mathrm{SNR}) - b])\,, \quad \sigma(z) = \frac{1}{1+e^{-z}}.$$

- **Continuous (from $\chi^2_{\textbf{signal}}$)**: use $\chi^2_{\mathsf{signal}} = \mathrm{SNR}^2$ in a similar mapping if desired.

Practical details

- **Noise pairing**: same seed per band for A; independent seed for B.
- **Saved arrays**: $\mathbf{I}_{\text{subA}}$ (vis), $\mathbf{I}_{\text{cleanA}}$ (vis), residual stacks in linear space.
- **Mass sampling**: log-uniform (configurable).
- **Source preview**: `source_rgb` helps spot issues quickly.

Takeaways

- SNR now uses the **whole image** $\Rightarrow$ more stable than ring-based.
- We define a consistent global statistic: $\chi^2_{\text{signal}} = \mathrm{SNR}^2$.
- We store `source_rgb` (source-only, LSST-like) for easy visual checks.
- Threshold T_{snr} controls positive rate; can also use continuous scores.