Volume Title
*ASP Conference Series, Vol. **Volume Number***
Author

$\mathcal{L}IRA$, the Low-counts Image Restoration and Analysis Package: a Teaching Version via R

A. Connors[1], Nathan M. Stein[2], David van Dyk[3], Vinay Kashyap[4], and Aneta Siemiginowska[4]

[1]*Eureka Scientific, 2452 Delmer St. Suite 100, Oakland CA 94602*

[2]*Harvard Statistics, 1 Oxford St., Cambridge MA 02138*

[3]*UCI Statistics, Bren Hall 2019, Irvine CA 92697*

[4]*Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge MA 02138*

Abstract. In low-count discrete photon imaging systems, such as in high energy astrophysics, the spatial distribution of a very few (or no!) photons per pixel can indeed carry important information about the shape of interesting emission. Our Low-counts Image Restoration and Analysis package, $\mathcal{L}IRA$, was designed to: 'deconvolve' any unknown sky components; give a fully Poisson 'goodness-of-fit' for any best-fit model; and quantify uncertainties on the existence and shape of unknown sky components. $\mathcal{L}IRA$ does this without resorting to χ^2 or rebinning, which can lose high-resolution information. However, since it combines a Poisson-specific multi-scale model for the sky with a full instrument response, within a (Bayesian) probablility framework, sampled via MCMC — running it thoughtfully requires understanding several key areas.

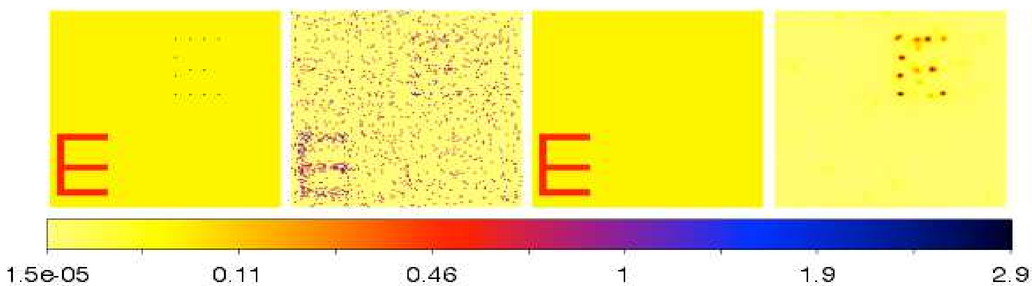
To this end, we have created and are releasing a 'teaching' version of $\mathcal{L}IRA$. It is implemented in R (cran.r-project.org). The accompanying tutorial and R-scripts step through all the basic analysis steps, from simple multi-scale representation and deconvolution; to model-testing; setting quantitative limits; and even simple ways of incorporating uncertainties in the instrument response.

1. Intro: Wonder, Glee, Skepticism, and $\mathcal{L}IRA$

As one confronts beautiful, beautifully processed, astronomical images – such as many in these proceedings – who does not feel the pull of wonder? As well, when one recognizes that a newly visible feature appears to match one's theory, isn't there a sharp pull of glee? Yet, in this paper, we advocate doubt: "where are the error bars?".

$\mathcal{L}IRA$ was developed precisely to quantify this doubt, for low-count Poisson data. To do this, $\mathcal{L}IRA$ brings together several different kinds of machinery, from Multi-scale (MS) models to Markov chain Monte Carlo (McMC) in a Bayesian framework [1,2,3]. Although made for Poisson counts, our schema of: a flexible or non- or semi-parametric model; with a background or Null model; within a full likelihood framework; can serve as a model for more general data. The combination can at first feel un-intutive for even seasoned researchers. Hence, we have created a 'teaching' version, with many examples, within the easy-to-use public statistical package 'R'. $\mathcal{L}IRA$ is available from: `nathanmstein at gmail.com` or `aconnors at eurekabayes.com`

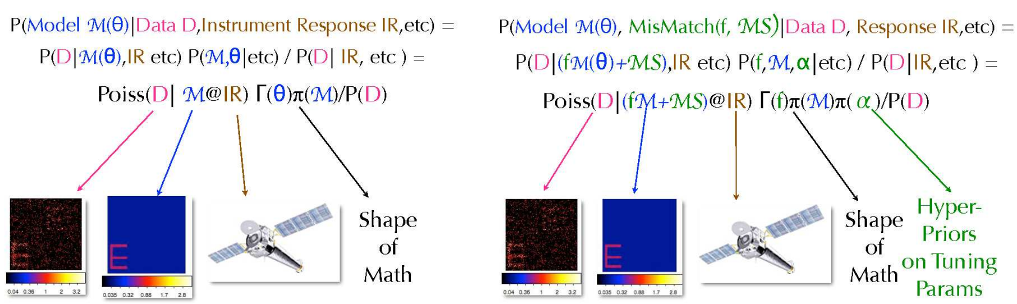
Here, we briefly exhibit parts of one of the teaching examples. It is based on a hypothetical 'skytruth' of a diffuse component (a broad letter E) and a cluster of point sources (also forming a letter E) on a flat background, in 128x128 bins, as shown in the first figure. The instrument smearing, or point-spread function (PSF), is assumed to be a circular Gauss-Normal distribution with $\sigma = 1.5$ bins. Simulated Poisson data D based on these is shown in the 2nd panel. We display a 'Null Model' of the diffuse emission based on hypothetical measurements and theory: a broad 'E' – in the 3rd panel. The simulated data, PSF, and Null model, are inputs to $\mathcal{L}IRA$; one of the outputs is the mean 'mismatch' between the data and theory, shown in the last panel of the 1st figure.



SkyTruth(unconvolved); Simulated Data; Null Model(unconvolved); $\mathcal{L}IRA$ result

2. $\mathcal{L}IRA$ Mechanics:

$\mathcal{L}IRA$ can be termed a 'forward-fitting' likelihood-based method, built under a 'Bayesian umbrellla'. That is, we use a Bayesian framework to successively add 'spokes' to the total likelihood: Poisson likelihood of the data D (red); given a Null Model with parameters θ, be designated by $M(\theta)$ (blue); and the Instrument Response by IR (brown). Then, using Bayes' theorem, the posterior probability can be written as in the first panel of the second figure, where @ designates a convolution.

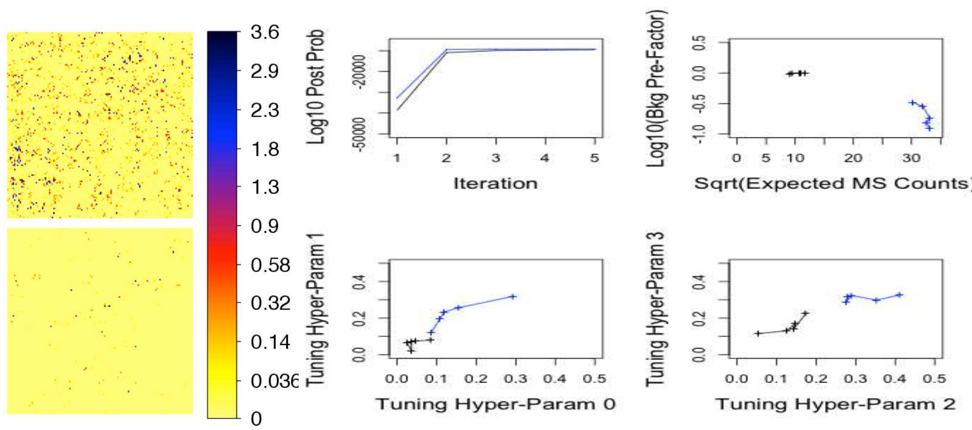


Bayes Umbrellas: Adding a apoke (right panel; green).

In this form, it is easy for us to add a *model/data mis-match* 'spoke' (green) to our Bayesian umbrella. In our low-count regime, we formulate the *mode/data mis-match* term to be a *prefactor* times the null model, plus a Poisson-tailored *multi-scale model* [1,2,4] that will handle both fine details and broad features. But now there are a great many parameters: rates at each successively finer multi-scale level, given the previous level; tuning (or smoothing or regluarization) hyper-parmeters, for each level; the Null Model prefactor. Hence rather than e.g. a Powell or Levenberg-Marquardt method for finding a mode, we use Markov chain Monte Carlo to map out the full probability space. This allows us to get both a 'best fit', and a way to express uncertainties on any feature from the data/model mis-match.

3. Running *ℒIRA*

In the next several figures, we illustrate McMC in action, mapping out the shape of our posterior likelihood (or Bayesian Umbrella from the second figure). It shows both a 'burn-in' phase and a converged phase. Finally we illustrate that, in order to get full quantitative limits, we must perform the same *ℒIRA* analysis on a handful of simulated data sets based on the Null Model (convolved with the instrument response). We then use a small subset of the parameters — in this case, the total counts infered to be in the multi-scale (*MS*) component — as a *summary statistic* of the 'distance' between the data and the null of the summary statistic give the upper and lower bounds on the *shape* of the Data/Null-Model mis-match.



Iteration 005 Two start values: high (top image; '+'); low (bottom image; '+').

Acknowledgments. This work was supported by NSF Awards DMS 04-06085, DMS-09-07522, and DMS-09-07185. A.C. and N.M.S. were supported by NASA AISRP grant NNG06GF17G. *ℒIRA* is based on EMC2, and is an out-growth of CHASC, the California Harvard Astrostatistics Collaboration.

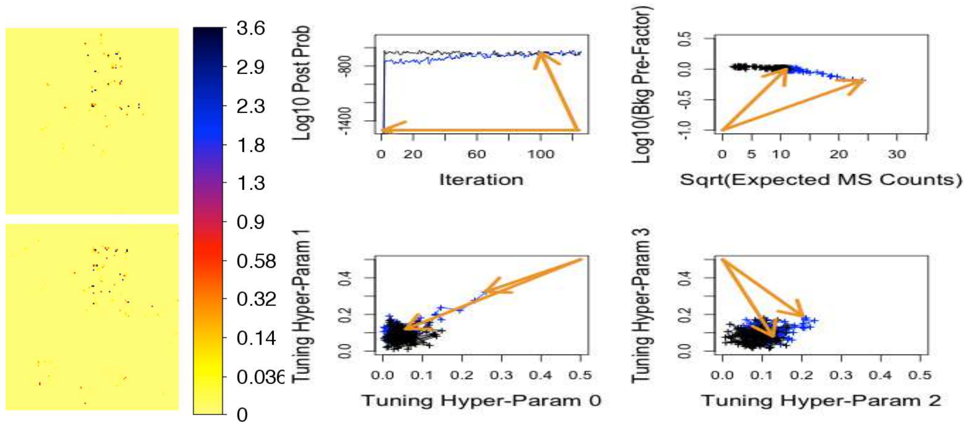
References

[1] Esch, D. N., et al., (2004), *ApJ*, 610, 1213.

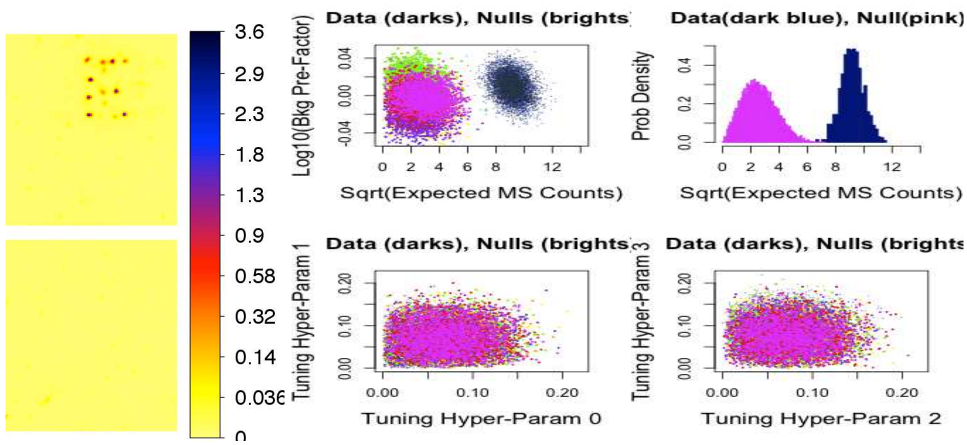
[2] Connors, A. and van Dyk, D. A. (2007), In *Statistical Challenges in Modern Astronomy IV* (Editors: G. J. Babu and E. D. Feigelson), *ASP* (San Francisco, CA), CS371, 101.

[3] van Dyk, D. A., et al., (2006), *Bayesian Analysis*, 1, 189.

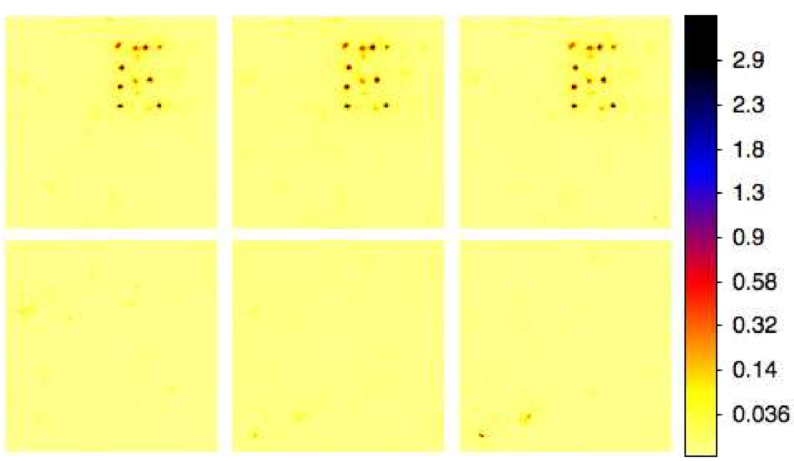
[4] Kolaczyk, E.D. and Nowak, R.D. (2005), *Biometrika*, 92:1, 119.



Iteration 125. Two start values: high (top image, '+'); low (bottom image, '+'). Orange arrows roughly indicate burn-in range for high starting values.



ℒIRA **Results, after burn-in.** Left: Mean Images from Data (top) vs. Simulated Nulls (bottom). Right: Distributions of Data (dark colors) vs Simulated Nulls (bright colors).



ℒIRA **Results, limits on shape.** Data (top) vs Simulated Null (bottom): Left: lower 5% limit; Middle: Mean; Right: upper 95% limits