Inferring the Spatial Structure of the Pleiades

A Bayesian approach

Olivares, J.

Sarro, L. M., Moraux, E. Bouy, H., Berihuete, A.

∼

ETS Ingeniería Informática
UNED, Madrid

∼

Fost, IPAG, Grenoble

November 2014

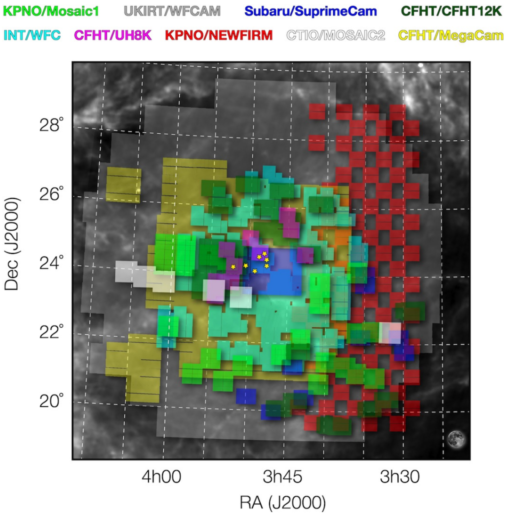
Motivation

- Learn statistics
- Spatial structure *per se*
- Will be used to infer membership probabilities.

Methodology

1. Analyse the data.
2. Select *a priori* the model(s) according to data.
3. Construct a probabilistic framework for the model.
4. Use Bayes theorem and MCMC to:
   - Check accuracy and precision with mock data.
   - Obtain the posterior for the parameters.
5. Analyse the posteriors.

DANCe Data


KPNO/Mosaic1
UKIRT/WFCAM
Subaru/SuprimeCam
CFHT/CFHT12K
INT/WFC
CFHT/UH8K
KPNO/NEWFIRM
CTIO/MOSAIC2
CFHT/MegaCam
28°
26°
24°
22°
20°
Dec (J2000)
4h00
3h45
3h30
RA (J2000)

Data from Sarro et al. 2014

Table: Number of stars at different R_{max}

1°	2°	3°	4°	5°	6°
496	1028	1354	1576	1735	1805

Table 4: True positive rates and contamination rates for different values of the membership threshold. The uncertainty intervals correspond to the range of values (maximum-minimum) observed in the five random samples.

p_{min}	0.50	0.7	0.8	0.90	0.95	0.96	0.97	0.98	0.99	0.9975
TPR (%)	98.4± 0.5	97.1±0.7	96.0±0.9	92.9±1.5	88.0±2.8	85.9±3.0	82.6±3.2	76.7±4.9	63.8±7.7	36.3±7.7
CR (%)	11.0±2.0	8.0 ±1.5	6.6 ±1.3	4.5±1.1	2.9±0.5	2.6±0.6	2.1 ±0.5	1.6 ±0.3	1.1 ±0.3	0.4±0.4

Data from Sarro et al. 2014: λ parameter

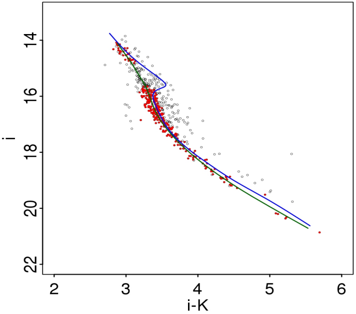


Fig. 2: Principal curve fits to the initial reference set (blue line) and to the subset of sources with all magnitudes fainter than its closest point in the first principal curve (green line). This subset of points is represented in red.

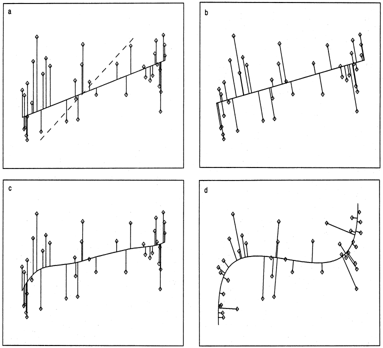


Figure 1. (a) The linear regression line minimizes the sum of squared deviations in the response variable. (b) The principal-component line minimizes the sum of squared deviations in all of the variables. (c) The smooth regression curve minimizes the sum of squared deviations in the response variable, subject to smoothness constraints. (d) The principal curve minimizes the sum of squared deviations in all of the variables, subject to smoothness constraints.

Hastie & Stuetzle, 1989

Probabilistic Framework

Bayes theorem.

$$p(\theta \mid D, I) = \frac{p(D \mid \theta, I)p(\theta \mid I)}{p(D \mid I)}$$

Evidence,

$$p(D \mid I) = Z = \int P(D \mid \theta, I)P(\theta \mid I)d\theta$$

The generative model, $p(D \mid \theta, I)$, is a pdf.

$$\int p(D \mid \theta, I)dD = 1. \quad (1)$$

"I will say that you have a *generative model* of data point n if you can write down or calculate a pdf $p(D_n \mid \theta, I)$ for the measurement D_n, conditional on a vector or list θ of parameters and a (possibly large) number of other things I (*prior information*) on which the D_n pdf depends, such as assumptions, or approximations, or knowledge about the noise process, or so on." Hogg 2012.

Number Density Profiles

We used the classical surface density families of models:

$$King = S_c\left(\frac{1}{\sqrt{1+\frac{r^2}{r_c^2}}} - \frac{1}{\sqrt{1+\frac{r_t^2}{r_c^2}}}\right)^2$$

$$Plummer = S_c\left(1+\frac{r^2}{r_c^2}\right)^{-2},$$

modified by:

- Field density S_f as a Contamination ratio

$$Cr = \frac{\pi R_{max}^2 S_f}{N}.$$

- r_c as a linear function of Sarro's et al. (2014) λ,

$$r_c = r_{c0} + r_{c1}\lambda.$$

Generative Model. Example of Plummer profile

$$p(r \mid r_c) = \frac{dN(r)}{N_{tot}} \frac{1}{dr}$$

$$= \frac{2\pi S_0 r \left(1 + \frac{r^2}{r_c^2}\right)^{-2} dr}{\pi S_0 r_c^2} \frac{1}{dr}$$

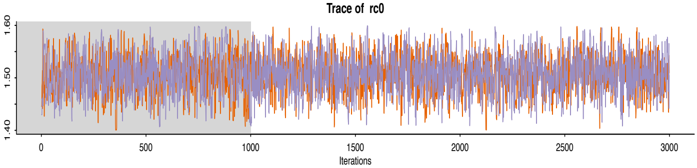
$$= 2\frac{r}{r_c^2}\left(1 + \frac{r^2}{r_c^2}\right)^{-2}.$$

If data are truncated, as in our case, the pdf in the interval $(0, R_{max})$ is

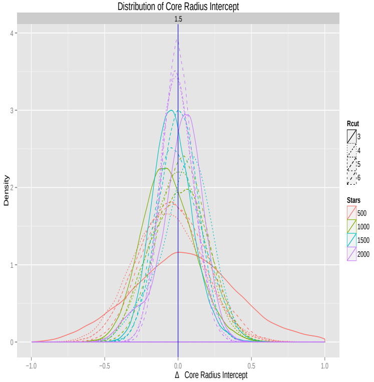
$$p(r \mid r_c) = 2\frac{r}{R_{max}^2} \frac{\left(1 + \frac{R_{max}^2}{r_c^2}\right)}{\left(1 + \frac{r^2}{r_c^2}\right)^2}.$$

Sampling the Posterior

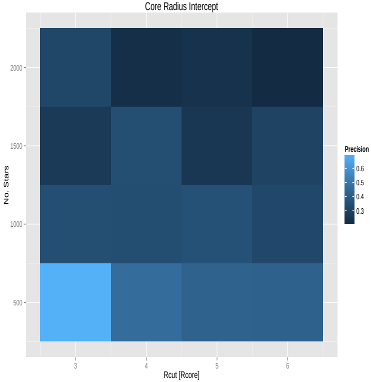
- MCMC Sampler: Stan (mc-stan.org, Hoffman-Gelman, 2011).
- Convergence, R-hat criterion (Gelman & Rubin, 1992).



Mock Data: Plummer v1 Contamination

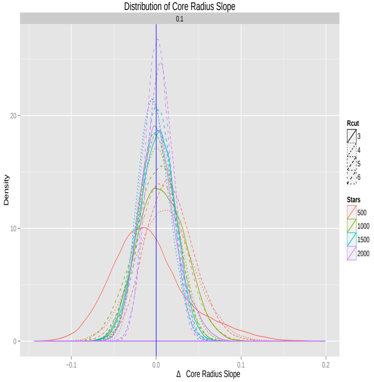


Accuracy

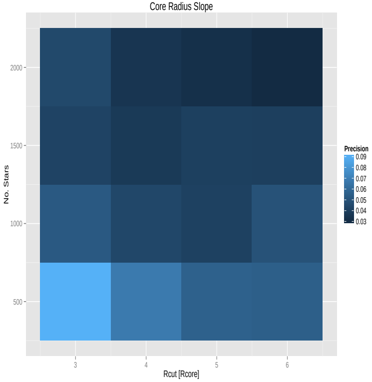


Precision

Mock Data: Plummer v1 Contamination

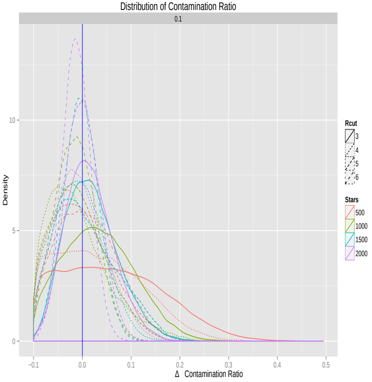


Accuracy

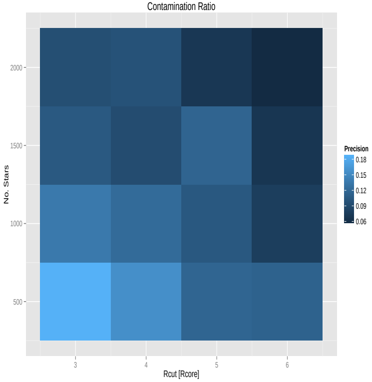


Precision

Mock Data: Plummer v1 Contamination

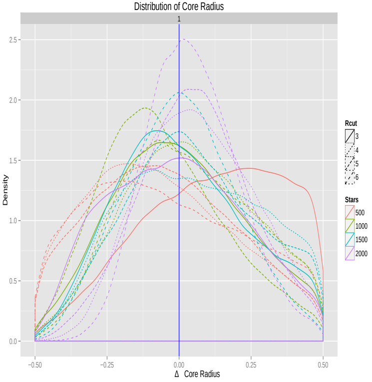


Accuracy

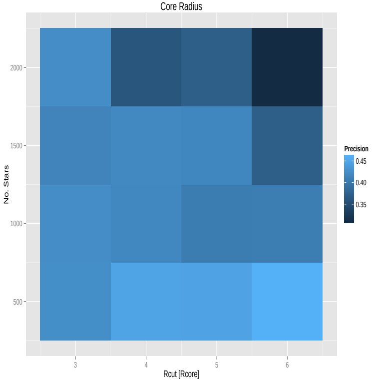


Precision

Mock Data: King v1 Contamination

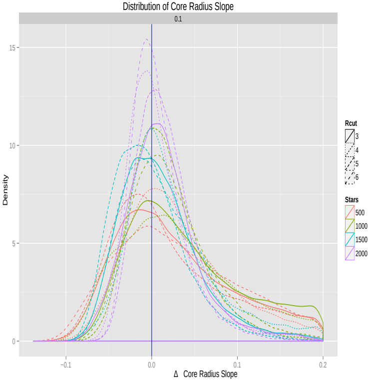


Accuracy

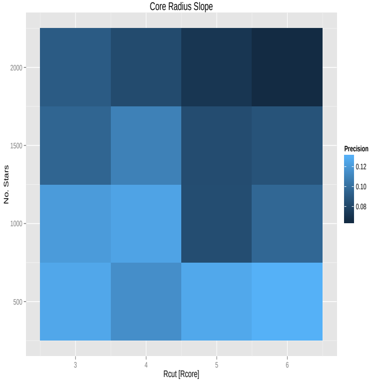


Precision

Mock Data: King v1 Contamination

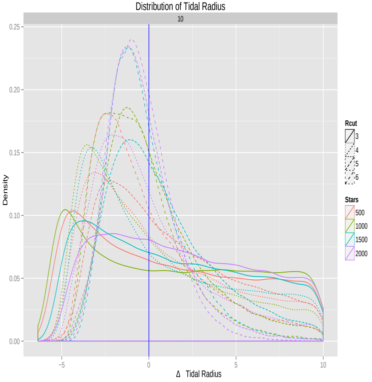


Accuracy

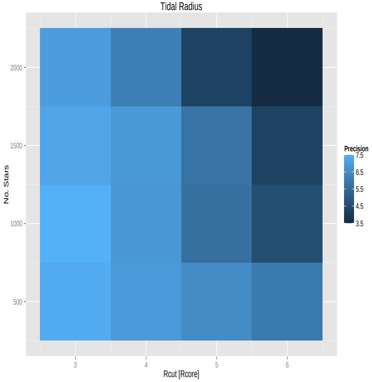


Precision

Mock Data: King v1 Contamination

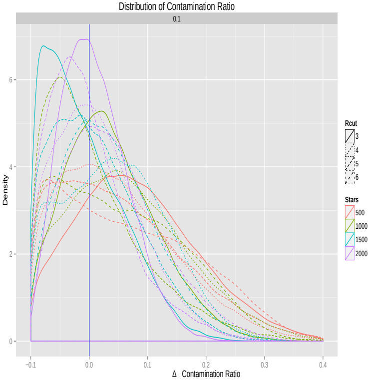


Accuracy

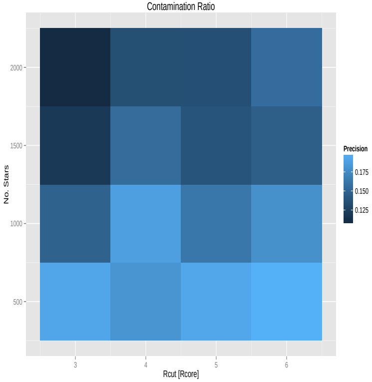


Precision

Mock Data: King v1 Contamination

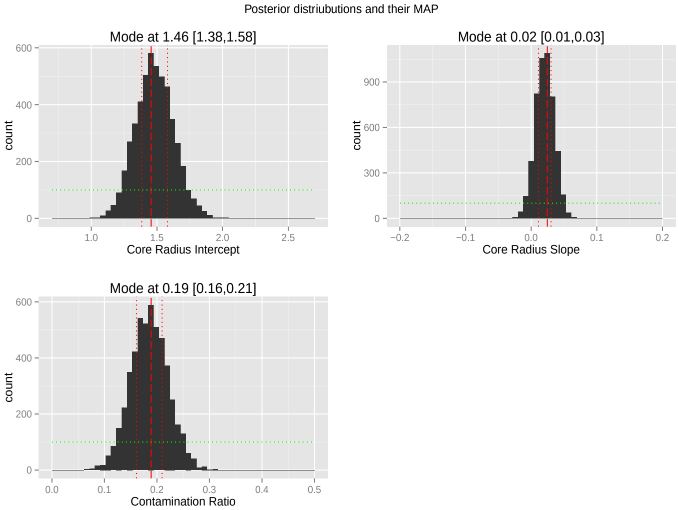


Accuracy

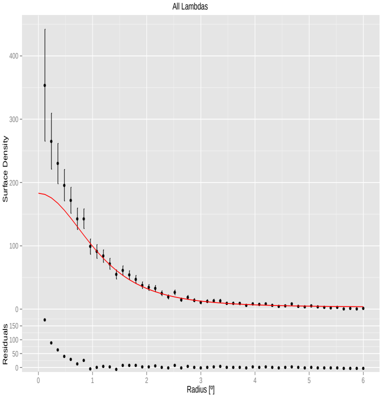


Precision

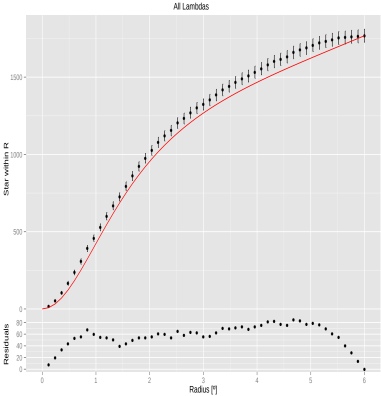
Real Data: Plummer v1 Contamination



Real Data: Plummer v1 Contamination

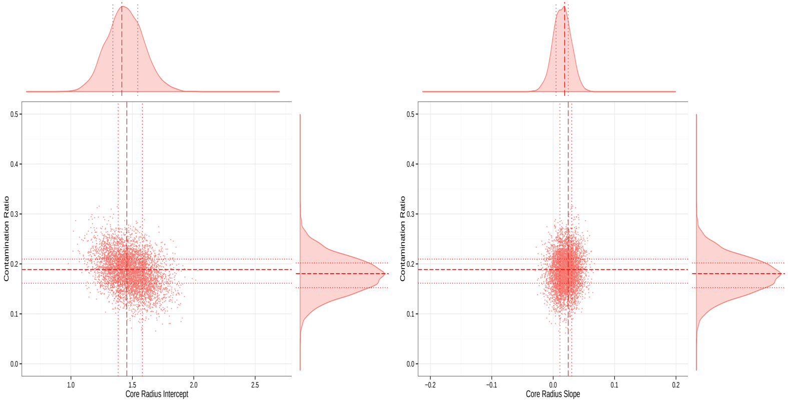


Number Density

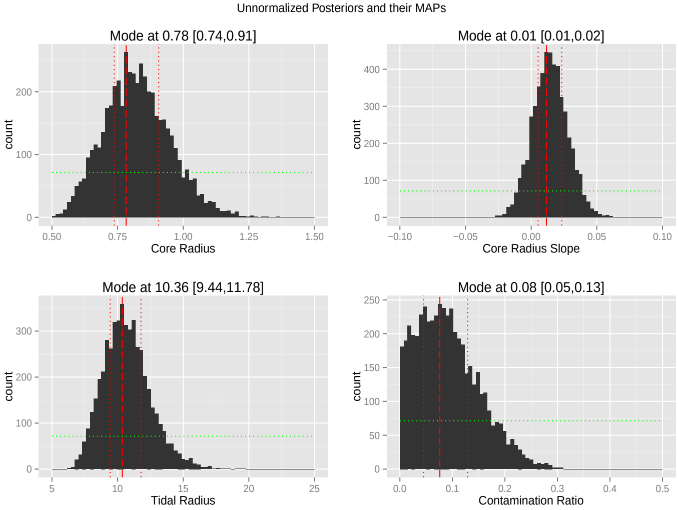


Number

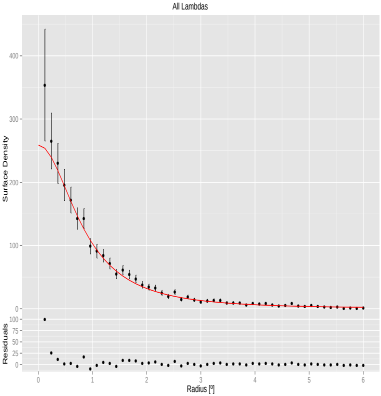
Real Data: Plummer v1 Contamination



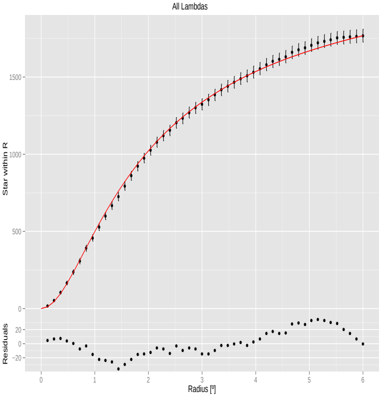
Real Data: King v1 Contamination



Real Data: King v1 Contamination

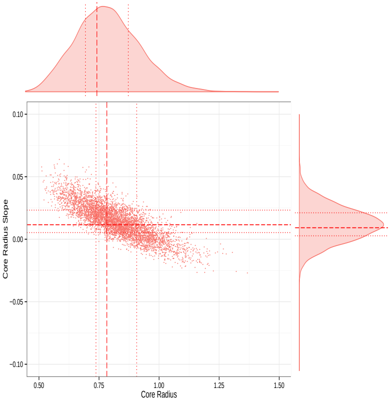


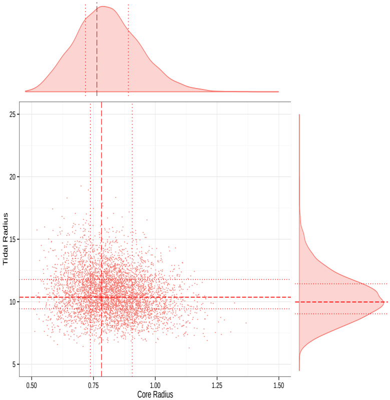
Number Density



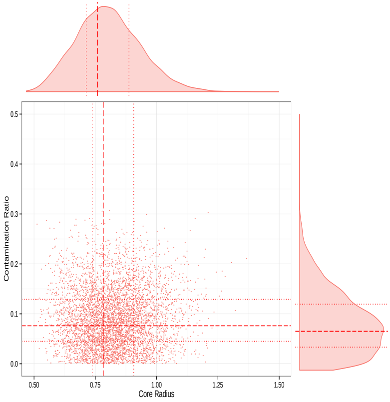
Number

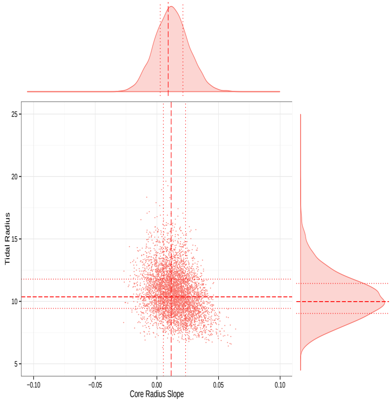
Real Data: King v1 Contamination



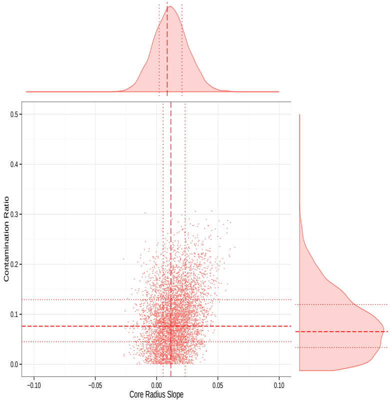


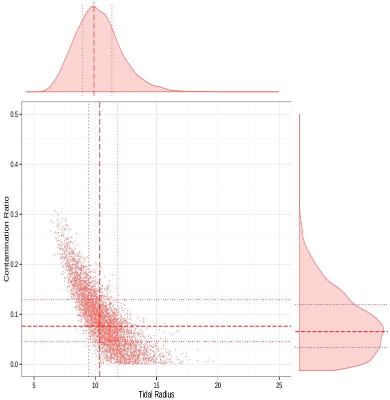
Real Data: King v1 Contamination





Real Data: King v1 Contamination





Model Selection: Bayes Factor

Use evidence to select appropriate model,

$$K_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} = \frac{\int P(D \mid \theta_1, I) P(\theta_1 \mid I) d\theta_1}{\int P(D \mid \theta_2, I) P(\theta_2 \mid I) d\theta_2} = \frac{Z_1}{Z_2}$$

Approximate Z by HMA (Newton and Raftery, 1994),

$$Z_{HMA} = \left(\frac{1}{m} \sum_{i}^{m} p(D \mid \theta^i)^{-1} \right)^{-1}.$$

Model Selection: Evidence

Table: $LogZ_{HMA}$. Plummer Models with varying R_{max}

Model	3°	4°	5°	6°
v_0	0.56	-1.29	-1.77	-1.95
v_0 Cr	-0.91	-1.44	-1.35	-0.57
v_1	0.30	0.57	-0.21	0.28
v_1 Cr	0.87	0.54	0.38	0.28

Model Selection: Evidence

Table: $LogZ_{HMA}$. King Models with varying R_{max}

Model	3°	4°	5°	6°
v_0	-0.18	-1.86	-1.43	-2.03
v_0 Cr	0.82	-1.55	-1.52	-1.11
v_1	-0.91	-1.44	-1.35	-0.57
v_1 Cr	0.87	0.57	0.38	0.23

Pinfield's et al. values

Using the distance to the Pleiades (136.2 pc, Melis et al. 2014)

Table: King v1 cr [68 % interval]

Parameter	6°
r_{c0} [pc]	1.84 [1.75-2.15]
r_{c1} [pc/ λ]	0.02 [0.0-0.05]
r_t [pc]	24.6 [22.4-27.9]
cr	0.08 [0.05,0.13]

Table 3. Pleiades King fit results.

Bin ($M_\odot$)	r_c (pc)	r_c limits (68 per cent confidence)	k	k limits (68 per cent confidence)	n	Mass
1	0.91	[0.50–1.51]	1.86	[0.87–3.74]	13	66
2	1.38	[1.15–1.66]	10.04	[7.86–12.69]	115	190
3	2.22	[1.98–2.49]	15.90	[14.17–17.81]	300	249
4	2.91	[2.63–3.23]	32.81	[30.51–35.37]	766	230

r_t = 13.1 pc, Cluster depth = $\pm$0.2 mag
M_c = 735 $M_\odot$, N_c= 1194, $\bar{m}_c$= 0.616 $M_\odot$

Figure: Pinfiled's values

Future Work

- Estimate de Credibility intervals.
- Determina the false positive rate of lambda segregation.
- Infer the number of stars.
- Try different Profiles (e.g. Elson, Fall & Freeman,1987)