Bayesian Hierarchical Models for Stellar Evolution

Shijing Si*, David van Dyk*, Ted von Hippel†

*Imperial College London

†Embry-Riddle Aeronautical University

March 22, 2016

Overview

Shijing Si*, David van Dyk*, Ted von Hippel†

Pros & Cons

Pros:

- able to combine multiple data sources and lead to comprehensive analysis;
- shrinkage estimates have smaller MSE than case-by-case analyses;

Cons:

- maybe computationally intensive, especially when the likelihood is complicated.

An Example of Shrinkage Estimates

- Consider the simple model:

$$Y_i \sim N(\theta_i, \sigma^2), i = 1, 2, \cdots, k, \tag{1}$$

- σ^2s are known.
- The ML estimates are $\mu_i^{\text{ind}} = Y_i$ and their mean squared errors $\mathrm{E}(\sum_{i=1}^{k}(\mu_i^{\text{ind}} - \mu_i)^2|\boldsymbol{\mu}) = k\sigma^2$.
- With homogeneous population, $\theta_1 = \theta_2 = \cdots = \theta_k$ and the pooled estimate, $\hat{\theta}_i^{\text{pool}} = \bar{Y} = \frac{1}{k}\sum Y_i$.

James-Stein Estimator

The James-Stein estimator of $\theta_i, i = 1, 2, \cdots, k$,

$$\hat{\theta}_i^{\mathrm{JS}} = (1 - \hat{B})\hat{\theta}_i^{\mathrm{ind}} + \hat{B}\hat{\theta}_i^{\mathrm{pool}} \tag{2}$$

where

$$S^2 = \sum(Y_i - \bar{Y})^2/(k-1), \hat{B} = (k-3)\sigma^2/(k-1)S^2$$

.

James-Stein estimators outperform MLE in terms of MSE. James-Stein estimates reduce MSE:

$$\begin{aligned} \mathrm{E}\Big[\sum_{i=1}^{k}(\hat{\theta}_i^{\mathrm{JS}} - \theta_i)^2|\boldsymbol{\theta}\Big] &= k\sigma^2 - \sigma^2(k-3)\mathrm{E}(\hat{B}) \\ &< k\sigma^2 = \mathrm{E}\Big[\sum_{i=1}^{k}(\hat{\theta}_i^{\mathrm{ind}} - \theta_i)^2|\boldsymbol{\theta}\Big]. \end{aligned}$$

James-Stein Estimators & Hierarchical model

Further assume θ_i from the same population, then we have a hierarchical model

$$Y_i \sim N(\theta_i, \sigma^2), i = 1, 2, \cdots, k; \quad (3)$$

$$\theta_i \sim N(\gamma, \tau^2). \quad (4)$$

- Approaches: Fully Bayesian (FB), Empirical Bayes (EB)
- FB infers all parameters from their joint posterior, usually via MCMC
- EB optimizes part of parameters then infers others from their conditional posterior
- James-Stein estimators can be obtained from EB (Efron 1972) in simple models
- James-Stein requires data have same variance, however FB and EB handle all cases

Application I: Distance Modulus to LMC

Method	Result on Distance Modulus	References
Cepheids: trig. paral.	18.70 ± 0.16	feast 1997
Cepheids: MS fitting	18.55 ± 0.06	laney 1994
Cepheids: B-W	18.55 ± 0.10	gieren 1998
Cepheids: P/L relation	18.575 ± 0.2	groenewegen 2000
Eclipsing binaries	18.4 ± 0.1	Fitzpatrick 2002
Clump	18.42 ± 0.07	Clementini 2003
Clump	18.45 ± 0.07	Clementini 2003
Clump	18.59 ± 0.09	Romaniello 2000
Clump	18.471 ± 0.12	Pietrzynski 2002
Clump	18.54 ± 0.10	Sarajedini 2002
Miras	18.54 ± 0.18	van 97
Miras	18.54 ± 0.14	Feast 2000
SN 1987a	18.54 ± 0.05	Panagia 1998

Empirical Bayes Estimation

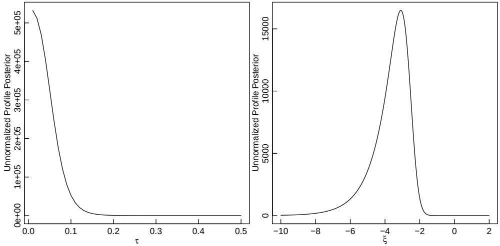
- Statistical model:

$$D_i \sim N(\mu_i, \sigma_i^2), \ i = 1, \cdots, 13, \quad (5)$$
$$\mu_i \sim N(\gamma, \tau^2), \quad (6)$$

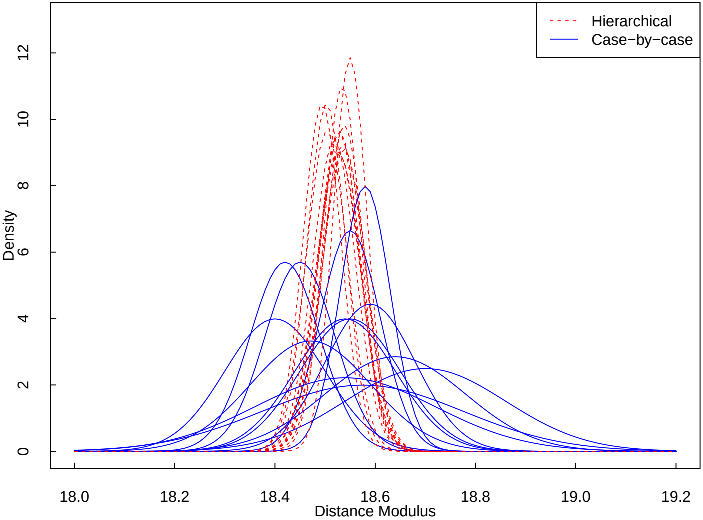
- object-level parameter μ_i, the real estimate of the distance modulus based on method/dataset i,
- D_i, the actual estimated distance modulus based on the method/dataset i,
- σ_i the known standard deviation of the statistical error,
- γ the true distance modulus of the LMC, and τ is the standard deviation of systematic errors of various methods.

MAP for Population-level parameters

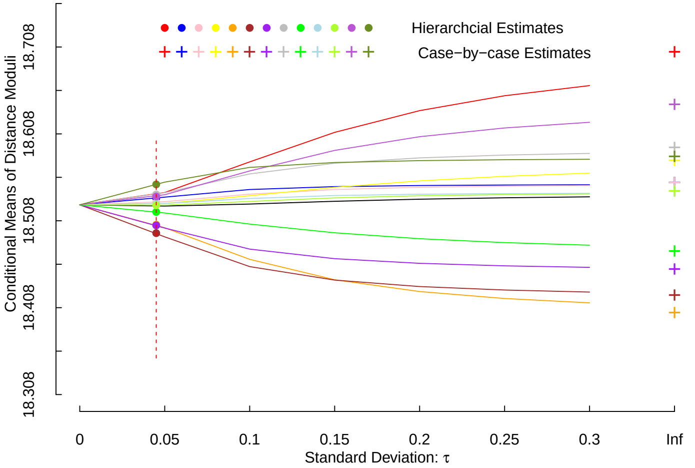
- non-informative prior, $p(\gamma, \tau) \propto 1$
- marginal posterior of τ and $\xi = \log \tau$ in the figure below
- τ peaks at 0, ruins its modal estimate



Shrinkage Versus Case-by-case Analysis



Empirical Bayes Screening



Application II: A Group of Galactic Halo WDs

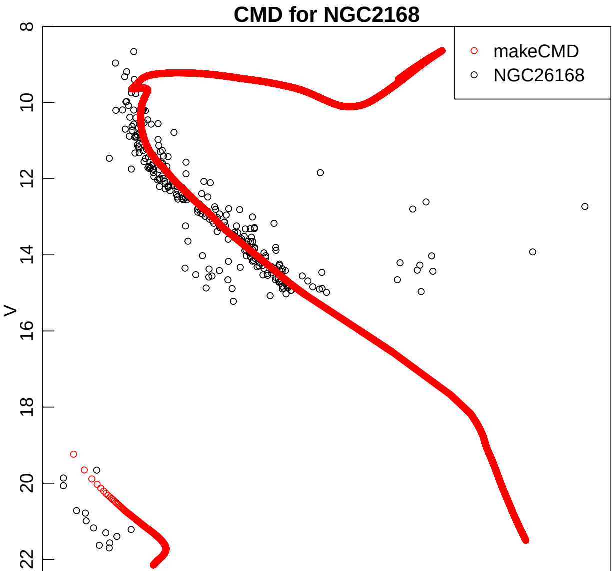
- The statistical model underlying BASE-9 relates a WD's photometry to its parameters,

$$\mathbf{X}_i|A_i, \mathbf{\Theta}_i \sim N_K\Big(G(A_i, \mathbf{\Theta}_i), \mathbf{\Sigma}_i\Big), \tag{7}$$

where,

- N_K represents a K-variate Gaussian dsitribution
- $\mathbf{\Theta}_i = (D_i, M_i, T_i)$ is the i-th WD's distance modulus, mass, metallicity
- $\mathbf{\Theta}_i$ is the i-th WD's log base 10 age
- $G(\cdot)$, astrophysical models based on Color-Magnitudes Diagrams(CMD), connecting photometry to its parameters

CMD plot



Prior Distributions

- independent prior on $(A_i, \mathbf{\Theta}_i)$ for i-th WD:

$$p(A_i, \mathbf{\Theta}_i) = p(A_i|\mu_{A_i}, \sigma_{A_i}) \\ p(D_i|\mu_{D_i}, \sigma_{D_i})p(Z_i|\mu_{Z_i}, \sigma_{Z_i})p(M_i), \tag{8}$$

- $p(A_i|\mu_{A_i}, \sigma^2_{A_i})$, $p(D_i|\mu_{D_i}, \sigma^2_{D_i})$, $p(Z_i|\mu_{Z_i}, \sigma^2_{Z_i})$ are normal densities each with its own prior mean (i.e., μ_{A_i}, μ_{D_i}, and μ_{Z_i}) and standard deviations (i.e., $\sigma_{A_i}, \sigma_{D_i}$, and σ_{Z_i}).
- log normal prior on mass M_i
  $\log_{10}(M_i) \sim N(-1.02, 0.67729^2)$ (Miller 1972).

Case-by-case Analysis

The available software package, BASE-9, can analyze each WD with its own photometry.
The joint posterior density is

$$p(A_i, \mathbf{\Theta}_i | \mathbf{X}_i) \propto p(\mathbf{X}_i | A_i, \mathbf{\Theta}_i) p(A_i | \mu_{A_i}, \sigma_{A_i})$$
$$p(D_i | \mu_{D_i}, \sigma_{D_i}) p(Z_i | \mu_{Z_i}, \sigma_{Z_i}) p(M_i). \quad (9)$$

Hierarchical Modelling

we model the A_i via

$$A_i \sim N(\gamma, \tau^2). \tag{10}$$

Denote $A = (A_1, \cdots, A_n)$ and $\mathbf{\Theta} = (\mathbf{\Theta}_1, \cdots, \mathbf{\Theta}_n)$. The joint posterior for all parameters in the hirarchical model is

$$p(\gamma, \tau, A, \mathbf{\Theta} | \mathbf{X}) \propto p(\gamma, \tau) \times \prod_{i=1}^{n} p(\mathbf{X}_i | A_i, \mathbf{\Theta}_i) p(A_i | \gamma, \tau) p(D_i | \mu_{D_i}, \sigma_{D_i}) p(Z_i | \mu_{Z_i}, \sigma_{Z_i}) p(M_i). \tag{11}$$

Idea

- hierarchical modelling leads to shrinkage estimates with smaller MSE
- take advantage of existing packages for case-by-case analysis to obtain hierarchical results
- no need to rewrite hierarchical model fitting code
- save human time investment

Setup

Suppose we are interested in this hierarchical model:

$$Y_i \sim p(y_i|\theta_i), i = 1, 2, \cdots, I; \quad (12)$$

$$\theta_i \sim p(\theta_i|\gamma); \quad (13)$$

prior distribution on γ, $p(\gamma)$;

$\theta_i, i = 1, 2, \cdots, I$ are object-level parameters;

γ is the population-level parameter.

We have an available toolkit which is able to fit the first level model 12 with a prior distribution $p_0(\theta_i)$. In other words, we can obtain good samples from the case-by-case analysis

$$p(\theta_i|Y_i) \propto p(Y_i|\theta_i)p_0(\theta_i) \quad (14)$$

via the existing toolkit.

Fully Bayesian (FB) Analysis

From Equations 12–13, we can have the joint posterior distribution, i.e.,

$$p(\gamma, \theta_1, \cdots, \theta_I | Y_1, \cdots, Y_I) \propto p(\gamma) \prod_{i=1}^{I} p(\theta_i|\gamma) p(Y_i|\theta_i). \quad (15)$$

Clearly, given $\theta_1, \cdots, \theta_I$, we can update γ as the common Gibbs sampler, i.e.,

$$p(\gamma|\theta_1, \cdots, \theta_I) \propto p(\gamma) \prod p(\theta_i|\gamma). \quad (16)$$

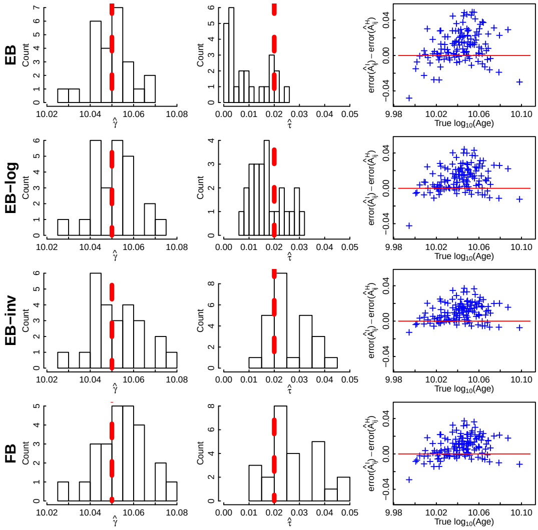
Update θ_i

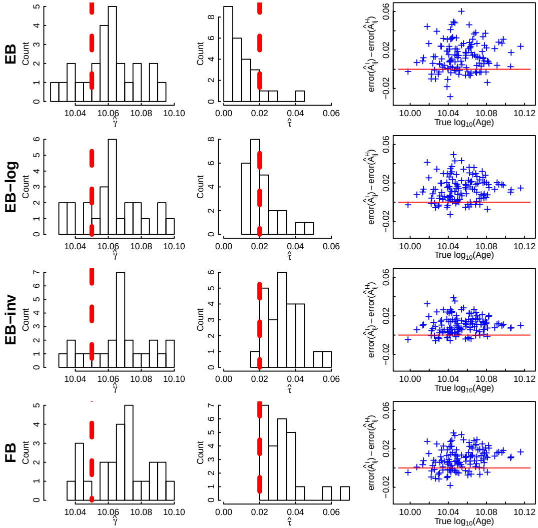
- Given γ, FB updates $\theta_i, i = 1, 2, \cdots, I$ from

$$p(\theta_i|\gamma, Y_i) \propto p(Y_i|\theta_i)p(\theta_i|\gamma). \quad (17)$$

- We have the case-by-case sample $\theta_i^{[1]}, \cdots, \theta_i^{[S]}$ from Equation 14
- We use case-by-case samples as proposals and metropolis-hastings rule to accept draws from $p(\theta_i|Y_i, \gamma)$

Fitting the Distributions of Age of Halo WDs




EB
EB−log
EB−inv
FB
Count
$\hat{\gamma}$
$\hat{\tau}$
True $\log_{10}$(Age)
error($\hat{A}_{ij}^{L}$) − error($\hat{A}_{ij}^{H}$)
10.04
10.06
10.08
10.10
0.00
0.02
0.04
0.06
10.00
10.12

Summary

Table 1 illustrates numerical comparison of the shrinkage and case-by-case estimates. Specifically it presents the average bias and the average root of mean sqaure error (RMSE) from each method, i.e.,

$$\mathrm{Bias}(A) = \frac{1}{500}\sum_{j=1}^{25}\sum_{i=1}^{20}(\hat{A}_{ij} - A_{ij}),$$

$$\mathrm{RMSE}(A) = \sqrt{\frac{1}{500}\sum_{j=1}^{25}\sum_{i=1}^{20}(\hat{A}_{ij} - A_{ij})^2}.$$

Summary Table

Table 1 : Summary of Estimations of the $\log_{10}(\text{Age})$ of simulated WDs

Simulation Cases		EB		EB-log		FB		Case-by-case	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\tau = 0.02$	$\sigma = 0.03$	6.96e-3	1.62e-2	7.02e-3	1.56e-2	7.98e-3	1.63e-2	1.86e-2	2.75e-2
	$\sigma = 0.05$	7.16e-3	1.72e-2	7.43e-3	1.65e-2	8.57e-3	1.76e-2	2.34e-2	3.20e-2
$\tau = 0.04$	$\sigma = 0.03$	1.38e-2	2.49e-2	1.39e-2	2.46e-2	1.46e-2	2.50e-2	1.88e-2	2.76e-2
	$\sigma = 0.05$	1.77e-2	3.33e-2	1.76e-2	3.25e-2	1.91e-2	3.36e-2	2.37e-2	3.40e-2

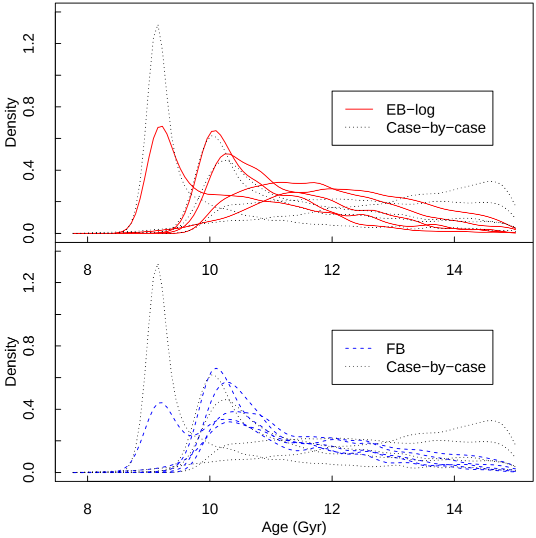
Five White Dwarfs

Shijing Si*, David van Dyk*, Ted von Hippel†

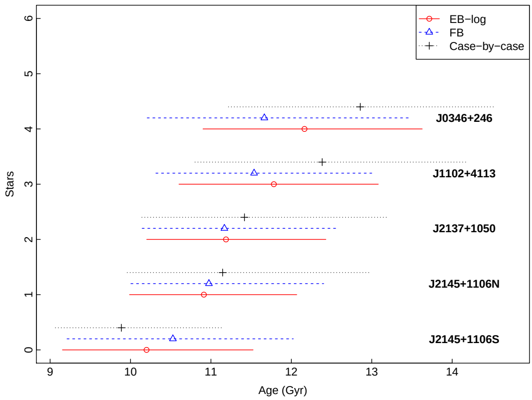
Table 2 : Prior distributions for distance moduli of five white dwarfs

White Dwarf	Distance Modulus	Reference
J0346+246	$\mathrm{N}(3.8, 2.5^2)$	Kilic et al. 2012
J1102+4113	$\mathrm{N}(2.64, 0.13^2)$	Kilic et al. 2012
J2137+1050	$\mathrm{N}(4.0, 2.4^2)$	Kilic et al. 2010
J2145+1106N	$\mathrm{N}(4.0, 2.4^2)$	Kilic et al. 2010
J2145+1106S	$\mathrm{N}(4.0, 2.4^2)$	Kilic et al. 2010

Result


Density
0.0
0.4
0.8
1.2
EB−log
Case−by−case
8
10
12
14
Density
FB
Case−by−case
Age (Gyr)

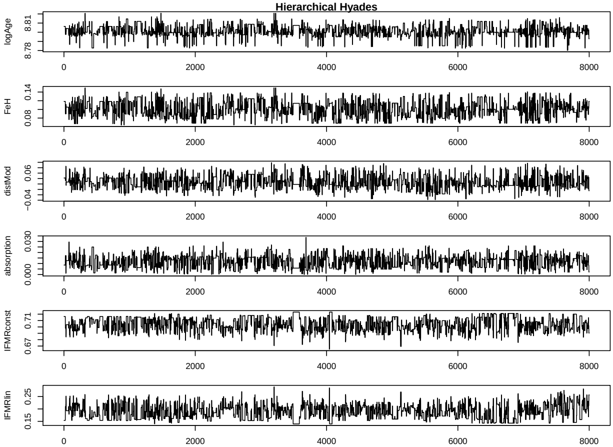
Estimates of Each WD

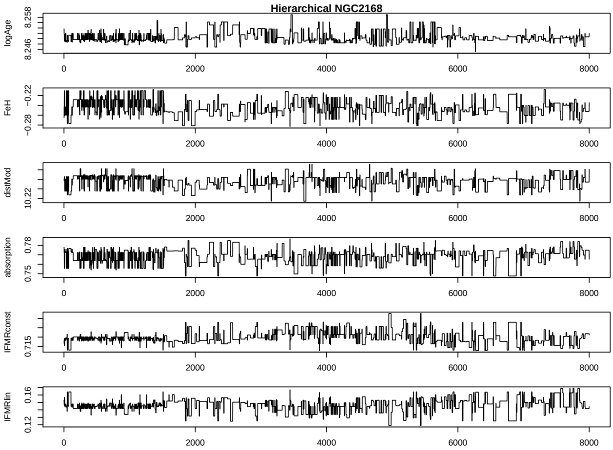


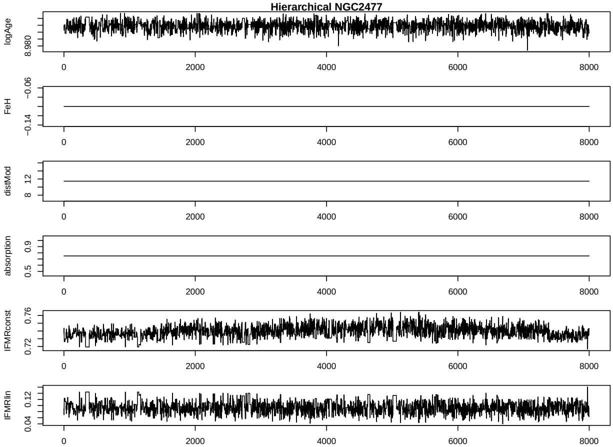
IFMR hierarchical model

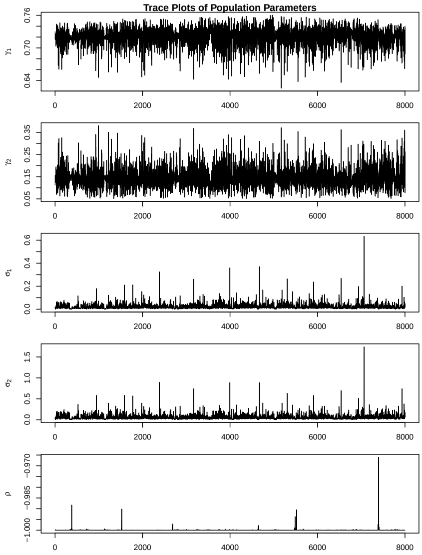
$$
\begin{aligned}
\mathbf{X}_i &\sim N(G(\mathbf{\Theta}_i, \mathbf{M}_i, a_i, b_i), \Sigma_i); \\
M_{ij}^{WD} &= b_i + a_i(M_{ij} - 3.0); \\
\begin{pmatrix} a_i \\ b_i \end{pmatrix} &\sim N(\boldsymbol{\gamma}, \boldsymbol{\Sigma}),
\end{aligned}
\tag{18}
$$

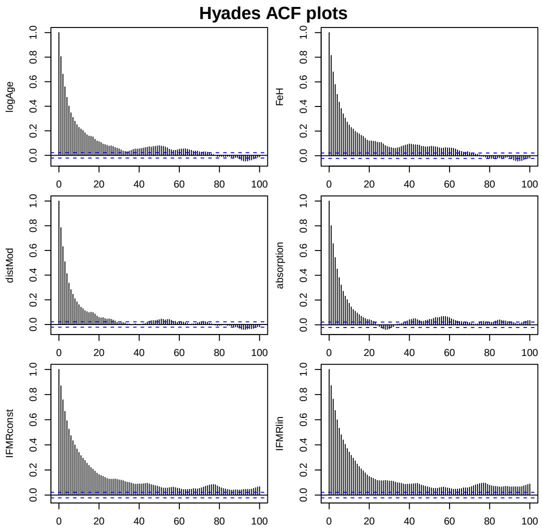
$\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ is a bivariate vector and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is the covariance matrix.

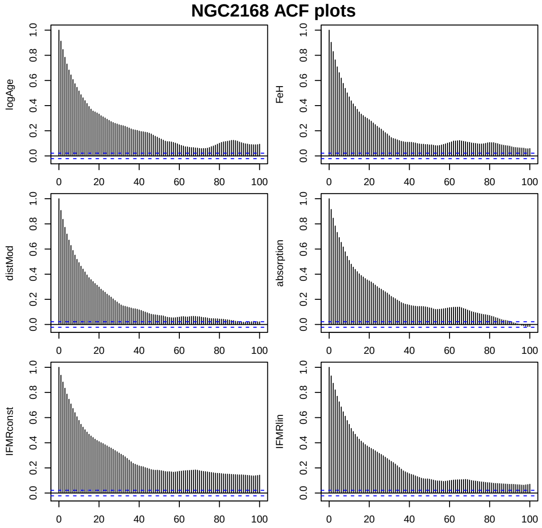

Hierarchical Hyades
logAge
8.78
8.81
FeH
0.08
0.14
distMod
−0.04
0.06
absorption
0.000
0.030
IFMRconst
0.67
0.71
IFMRlin
0.15
0.25
0
2000
4000
6000
8000

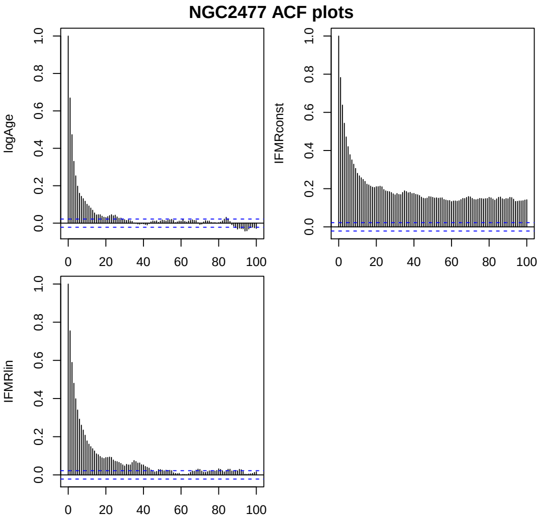

Hierarchical NGC2168
logAge
8.258
8.246
FeH
−0.22
−0.28
distMod
10.22
absorption
0.78
0.75
IFMRconst
0.715
IFMRlin
0.16
0.12
0
2000
4000
6000
8000

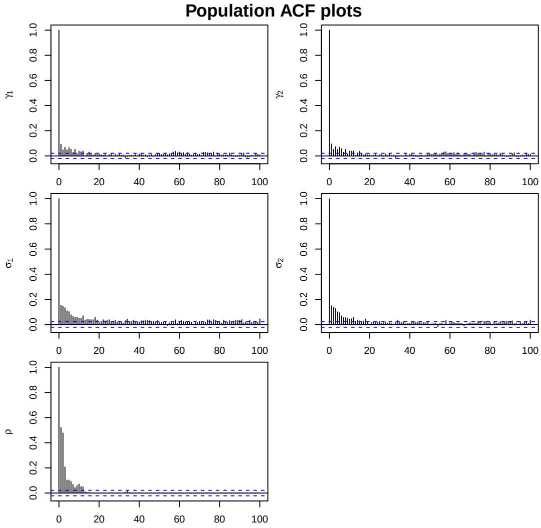

Hierarchical NGC2477
logAge
8.980
FeH
−0.06
−0.14
distMod
12
8
absorption
0.9
0.5
IFMRconst
0.76
0.72
IFMRlin
0.12
0.04
0
2000
4000
6000
8000


Trace Plots of Population Parameters
γ_1
γ_2
σ_1
σ_2
ρ
0
2000
4000
6000
8000


Hyades ACF plots
logAge
FeH
distMod
absorption
IFMRconst
IFMRlin
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


NGC2168 ACF plots
logAge
FeH
distMod
absorption
IFMRconst
IFMRlin
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


NGC2477 ACF plots
logAge
IFMRconst
IFMRlin
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


Population ACF plots
γ_1
γ_2
σ_1
σ_2
ρ
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100