Hierarchical Bayesian Inference – Its Application to Dust Emission Modelling

Frédéric GALLIANO

AIM, CEA/Saclay, France

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

2 Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

2 Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

Conclusion & Prospectives

Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

2 Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

4 Conclusion & Prospectives











• Dust models \Rightarrow highly non-linear ($L_{\nu} \propto T_{dust}^{6}$).

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.

- Dust models \Rightarrow highly non-linear ($L_{\nu} \propto T_{dust}^{6}$).
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

- Dust models \Rightarrow highly non-linear ($L_{
 u} \propto T_{dust}^{6}$).
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

 Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

Uniqueness of HerBIE (HiERarchical Bayesian Inference for dust Emission; Galliano, 2018)

Wide database of realistic optical properties;

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

- Wide database of realistic optical properties;
- Size distributions with stochastically heated grains;

- Dust models \Rightarrow highly non-linear ($L_{
 u} \propto T_{dust}^{6}$).
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

- Wide database of realistic optical properties;
- Size distributions with stochastically heated grains;
- Large variety of models ⇒ numerous dust components, stellar & radio continua, AGN radiative transfer models;

- Dust models \Rightarrow highly non-linear $(L_{\nu} \propto T_{dust}^6)$.
- Instrumental effects \Rightarrow poor spectral sampling & correlated calibration uncertainties.
- Least-squares \Rightarrow noise-induced false correlations & biases (Shetty *et al.*, 2009).

The Hierarchical Bayesian (HB) Solution

- Kelly et al. (2012) ⇒ hierarchical Bayesian inference could solve these problems (case of a single Modified Black Body; MBB).
- Other HB models: Juvela *et al.* (2013); Veneziani *et al.* (2013); Shetty *et al.* (2013) \Rightarrow all restricted to simple analytical models (combinations of MBBs).

- Wide database of realistic optical properties;
- Size distributions with stochastically heated grains;
- Large variety of models \Rightarrow numerous dust components, stellar & radio continua, AGN radiative transfer models;
- Large library of photometric filters & their partially correlated calibration uncertainties.

The likelihood of an SED:

The likelihood of an SED:

Effect of Noise: $L_{
u}^{\text{obs}}(\lambda) =$

The likelihood of an SED:

Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) +$

The likelihood of an SED:

Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda)$

The likelihood of an SED:

 $\text{Effect of Noise: } L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda;\vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1).$

The likelihood of an SED:

$$\begin{split} \text{Effect of Noise:} \ \ L_{\nu}^{\text{obs}}(\lambda) &= L_{\nu}^{\text{mod}}(\lambda;\vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1). \\ &\Leftrightarrow \epsilon(\lambda;\vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda;\vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}. \end{split}$$

The likelihood of an SED: Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1).$ $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$ Likelihood: $\rho\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda})|\vec{x}\right)$

The likelihood of an SED:

Effect of Noise:
$$L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda)$$
 $(\langle \epsilon \rangle = 0, \sigma(\epsilon) = 1).$
 $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$
Likelihood: $p\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i}; \vec{x})).$

The likelihood of an SED:

$$\begin{split} \text{Effect of Noise:} \ & L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda;\vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1). \\ & \Leftrightarrow \epsilon(\lambda;\vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda;\vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}. \\ & \text{Likelihood:} \ & p\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i};\vec{x})). \end{split}$$

The Bayesian Thing:

The likelihood of an SED: Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda)$ $(\langle \epsilon \rangle = 0, \sigma(\epsilon) = 1)$. $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$ Likelihood: $p\left(\vec{L}_{\nu}^{obs}(\vec{\lambda})|\vec{x}\right) = \prod p(\epsilon(\lambda_i;\vec{x})).$

The Bayesian Thing:



posterior distribution

The likelihood of an SED: Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1).$ $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$ Likelihood: $p\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i}; \vec{x})).$



 $\cap -(-)$

The likelihood of an SED: Effect of Noise: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty$

$$\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) + \ell(\lambda) + \ell(\lambda) + \delta_{\nu}(\lambda)}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$$

$$\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$$

Likelihood: $p\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i}; \vec{x})).$



The likelihood of an SED: Effect of Noise: $L_{\nu}^{\text{obs}}(\lambda) = L_{\nu}^{\text{mod}}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{\text{obs}}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1).$ $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x})}{\sigma_{\nu}^{\text{obs}}(\lambda)}.$ Likelihood: $p\left(\vec{L}_{\nu}^{\text{obs}}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i}; \vec{x})).$



The likelihood of an SED: Effect of Noise: $L_{\nu}^{obs}(\lambda) = L_{\nu}^{mod}(\lambda; \vec{x}) + \epsilon(\lambda) \times \sigma_{\nu}^{obs}(\lambda) \qquad (\langle \epsilon \rangle = 0, \ \sigma(\epsilon) = 1).$ $\Leftrightarrow \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{obs}(\lambda) - L_{\nu}^{mod}(\lambda; \vec{x})}{\sigma_{\nu}^{obs}(\lambda)}.$ Likelihood: $p\left(\vec{L}_{\nu}^{obs}(\vec{\lambda}) | \vec{x}\right) = \prod_{i} p(\epsilon(\lambda_{i}; \vec{x})).$



Sampling the distribution of parameters:

The most common way: Markov Chain Monte-Carlo (MCMC) \Rightarrow randomly drawing parameter values from the posterior.
Demonstration with an MBB: $L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



Demonstration with an MBB:
$$L_{\nu} = M_{\text{dust}} \times 4\pi\kappa_0 \left(\frac{\nu}{\nu_0}\right)^{\beta} \times B_{\nu}(T).$$



The Principle of Hierarchical Bayesian Inference

Posterior: $\underbrace{p(x_1 \dots x_N | L_{\nu}^{\text{obs},1} \dots L_{\nu}^{\text{obs},N})}_{\nu}$

posterior of the whole image

$$\begin{array}{l} \textbf{Posterior:} \quad \underbrace{p\left(x_{1} \ldots x_{N} \middle| \ \mathcal{L}_{\nu}^{\text{obs},1} \ldots \mathcal{L}_{\nu}^{\text{obs},N}\right)}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p\left(\mathcal{L}_{\nu}^{\text{obs},i} \middle| x_{i}\right)}_{i^{\text{th pixel likelihood}}} \times \end{array}$$

osterior:
$$\underbrace{p(x_1 \dots x_N | L_{\nu}^{obs,1} \dots L_{\nu}^{obs,N})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{obs,i} | x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

The Principle of Hierarchical Bayesian Inference

Modelling an ensemble of sources:

Posterior:
$$\underbrace{p(x_1 \dots x_N | L_{\nu}^{\text{obs},1} \dots L_{\nu}^{\text{obs},N})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{\text{obs},i} | x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ_i ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

The Principle of Hierarchical Bayesian Inference

Modelling an ensemble of sources:

Posterior:
$$\underbrace{p(x_1 \dots x_N \mid L_{\nu}^{\text{obs,1}} \dots L_{\nu}^{\text{obs,N}})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{\text{obs,i}} \mid x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

Hierarchical Bayesian sampling:

At each iteration, sampling:

- every parameter of every pixel;
- every hyperparameter;

Posterior:
$$\underbrace{p(x_1 \dots x_N | L_{\nu}^{obs,1} \dots L_{\nu}^{obs,N})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{obs,i} | x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

Hierarchical Bayesian sampling:

At each iteration, sampling:

- every parameter of every pixel;
- every hyperparameter;
- \Rightarrow sampling a parameter space of dimensions:

 $N_{dim} =$

Posterior:
$$\underbrace{p(x_1 \dots x_N | L_{\nu}^{obs,1} \dots L_{\nu}^{obs,N})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{obs,i} | x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

Hierarchical Bayesian sampling:

At each iteration, sampling:

- every parameter of every pixel;
- every hyperparameter;
- \Rightarrow sampling a parameter space of dimensions:

$$N_{\rm dim} = N_{\rm pixel} \times N_{\rm par} +$$

individual parameters

The Principle of Hierarchical Bayesian Inference

Modelling an ensemble of sources:

Posterior:
$$\underbrace{p(x_{1} \dots x_{N} \mid L_{\nu}^{obs,1} \dots L_{\nu}^{obs,N})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{obs,i} \mid x_{i})}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_{1} \dots x_{N})}_{\text{global prior}}.$$
Prior: $p(x_{1} \dots x_{N}) = \prod_{i=1}^{N} p(x_{i}, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

Hierarchical Bayesian sampling:

At each iteration, sampling:

- every parameter of every pixel;
- every hyperparameter;
- \Rightarrow sampling a parameter space of dimensions:

$$N_{\rm dim} = \underbrace{N_{\rm pixel} \times N_{\rm par}}_{N_{\rm par}} + \underbrace{2 \times N_{\rm par}}_{2 \times N_{\rm par}} +$$

individual parameters hyper-mean & std-dev

The Principle of Hierarchical Bayesian Inference

Modelling an ensemble of sources:

Posterior:
$$\underbrace{p(x_1 \dots x_N \mid L_{\nu}^{\text{obs,1}} \dots L_{\nu}^{\text{obs,N}})}_{\text{posterior of the whole image}} \propto \prod_{i=1}^{N} \underbrace{p(L_{\nu}^{\text{obs,i}} \mid x_i)}_{i^{\text{th pixel likelihood}}} \times \underbrace{p(x_1 \dots x_N)}_{\text{global prior}}.$$
Prior: $p(x_1 \dots x_N) = \prod_{i=1}^{N} p(x_i, \mu, \Sigma)$ is controlled by hyperparameters:
• average μ ;
• covariance matrix Σ .

Hyperparameters are sampled too \Rightarrow the shape of the prior is fit to the data.

Hierarchical Bayesian sampling:

At each iteration, sampling:

- every parameter of every pixel;
- every hyperparameter;
- \Rightarrow sampling a parameter space of dimensions:

$$N_{\rm dim} = \underbrace{N_{\rm pixel} \times N_{\rm par}}_{N_{\rm pixel}} + \underbrace{2 \times N_{\rm par}}_{2 \times N_{\rm par}} + \underbrace{C_{N_{\rm par}}^{N_{\rm par}-2}}_{N_{\rm par}}.$$

correlations

individual parameters hyper-mean &

F. Galliano (AIM)

Demonstration of the Hierarchical Bayesian SED Fitting


































Treatment of the Calibration Uncertainties

Likelihood: $\epsilon(\lambda; \vec{x}) = rac{L_{
u}^{
m obs}(\lambda) - L_{
u}^{
m mod}(\lambda; \vec{x})}{\sigma_{
u}^{
m obs}(\lambda)}$

Likelihood:
$$\epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{abs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x}) \times (1 + \delta(\lambda))}{\sigma_{\nu}^{\text{obs}}(\lambda)}$$

$$\text{Likelihood:} \ \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{bbs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x}) \times (1 + \delta(\lambda))}{\sigma_{\nu}^{\text{bbs}}(\lambda)} \qquad \text{where } \langle \delta \rangle = 0.$$

$$\text{Likelihood:} \ \epsilon(\lambda;\vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda;\vec{x}) \times (1 + \delta(\lambda))}{\sigma_{\nu}^{\text{obs}}(\lambda)} \quad \text{ where } \langle \delta \rangle = 0.$$

Correlations:

- Full correlation between pixels;
- Partial correlation between wavelengths.

$$\text{Likelihood:} \ \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x}) \times (1 + \delta(\lambda))}{\sigma_{\nu}^{\text{obs}}(\lambda)} \quad \text{ where } \langle \delta \rangle = 0.$$

Correlations:

- Full correlation between pixels;
- Partial correlation between wavelengths.
- $\Rightarrow \delta$ treated as a nuisance variable.

$$\text{Likelihood:} \ \epsilon(\lambda; \vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda; \vec{x}) \times (1 + \delta(\lambda))}{\sigma_{\nu}^{\text{obs}}(\lambda)} \quad \text{ where } \langle \delta \rangle = 0.$$

Correlations:

- Full correlation between pixels;
- Partial correlation between wavelengths.
- $\Rightarrow \delta$ treated as a nuisance variable.

Numerical treatment:

Problem: known to create very long chain correlations (Kelly, 2011).

$$\begin{array}{ll} \text{Likelihood:} \ \epsilon(\lambda;\vec{x}) = \frac{L_{\nu}^{\text{obs}}(\lambda) - L_{\nu}^{\text{mod}}(\lambda;\vec{x}) \times (1+\delta(\lambda))}{\sigma_{\nu}^{\text{obs}}(\lambda)} & \quad \text{where } \langle \delta \rangle = 0. \end{array}$$

Correlations:

- Full correlation between pixels;
- Partial correlation between wavelengths.
- $\Rightarrow \delta$ treated as a nuisance variable.

Numerical treatment:

Problem: known to create very long chain correlations (Kelly, 2011).

Solution: implement Ancillarity-Sufficiency Interweaving Strategy (ASIS; Yu & Meng, 2011) on every parameter.













Diagnostics:

Integrated autoccorrelation time: t_{int} (Sokal, 1995).



Diagnostics:

Integrated autoccorrelation time: t_{int} (Sokal, 1995).

Effective sample size: $N_{\rm eff} = t_{\rm MCMC}/t_{\rm int} \gtrsim 100$.

Outline of the Talk

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

4 Conclusion & Prospectives

A Large Grid of Simulated SEDs

Application to the most useful model:

Dust mixture: distribution of starlight intensities (U): $dM \propto U^{-\alpha} dU$.

Application to the most useful model:

Dust mixture: distribution of starlight intensities (U): $dM \propto U^{-\alpha} dU$. **Stellar continuum:** black body (only near-IR).

Application to the most useful model:

Dust mixture: distribution of starlight intensities (U): $dM \propto U^{-\alpha} dU$. **Stellar continuum:** black body (only near-IR).

Varying SED parameters:

SED shape: cold, warm & hot.














SED sample:

SED sample:



SED sample:



SED sample:



SED sample:



SED sample:



SED sample:







F. Galliano (AIM)



F. Galliano (AIM)





Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y-y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \le N_{\sigma} \Rightarrow \text{better than}$ $N_{\sigma};$
- Direct comparison of χ^2 and HB.



Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y - y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \leq N_{\sigma} \Rightarrow$ better than N_{σ} ;
- Direct comparison of χ^2 and HB.



Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y-y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \leq N_{\sigma} \Rightarrow$ better than N_{σ} ;
- Direct comparison of χ^2 and HB.

Quality control:

 Recovered values in good agreement with true values;



Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y-y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \leq N_{\sigma} \Rightarrow$ better than N_{σ} ;
- Direct comparison of χ^2 and HB.

Quality control:

- Recovered values in good agreement with true values;
- Uncertainties properly estimated;



Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y - y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \leq N_{\sigma} \Rightarrow$ better than N_{σ} ;
- Direct comparison of χ^2 and HB.

Quality control:

- Recovered values in good agreement with true values;
- Uncertainties properly estimated;
- 3 Few outliers \Rightarrow most degenerate parameters;



Quantifying the goodness of the recovery:

For any parameter y:

$$\mathcal{D}[y] = rac{y - y^{\mathsf{true}}}{\sigma_y^{\mathsf{HB}}}.$$

- $|D[y]| \leq N_{\sigma} \Rightarrow$ better than N_{σ} ;
- Direct comparison of χ^2 and HB.

Quality control:

- Recovered values in good agreement with true values;
- Uncertainties properly estimated;
- 3 Few outliers \Rightarrow most degenerate parameters;
- 4 Systematically better than χ^2 .

Integrated Autocorrelation Times:

Integrated Autocorrelation Times:



Integrated Autocorrelation Times:



Integrated Autocorrelation Times:



Integrated Autocorrelation Times:



Integrated Autocorrelation Times:

 \Rightarrow quantifies the length of independent draws in the chain.



F. Galliano (AIM)

Presence of Intrinsic Correlations

Presence of Intrinsic Correlations

Efficiency to recover correlation coefficients:

Presence of Intrinsic Correlations

Efficiency to recover correlation coefficients:













Effect of the Far-IR Wavelength Coverage

Effect of the Far-IR Wavelength Coverage

Capacity of the prior to compensate the lack of coverage:

 \Rightarrow tests removing SPIRE bands.
Effect of the Far-IR Wavelength Coverage

Capacity of the prior to compensate the lack of coverage:



Effect of the Far-IR Wavelength Coverage

Capacity of the prior to compensate the lack of coverage:



Effect of the Far-IR Wavelength Coverage

Capacity of the prior to compensate the lack of coverage:



Capacity of the prior to compensate the lack of coverage:



Capacity of the prior to compensate the lack of coverage:



Capacity of the prior to compensate the lack of coverage:



Extending the prior distribution to other parameters (e.g. M_{gas}, Z, etc.)

Extending the prior distribution to other parameters (e.g. M_{gas} , Z, etc.)



Extending the prior distribution to other parameters (e.g. M_{gas} , Z, etc.)



Extending the prior distribution to other parameters (e.g. M_{gas} , Z, etc.)



Extending the prior distribution to other parameters (e.g. M_{gas}, Z, etc.)



Outline of the Talk

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

2 Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

4 Conclusion & Prospectives

The Anomalous Microwave Emissions (AME)

The Anomalous Microwave Emissions (AME)

















(Bell et al., in press)







• $p(\rho_{\text{dust}} < \rho_{\text{PAH}}) = 1.000$



⁽Bell et al., in press)

- $p(
 ho_{
 m dust} <
 ho_{
 m PAH}) = 1.000$
- $p(
 ho_{ ext{dust}} <
 ho_{ ext{PAH}+}) = 1.000$
- $p(\rho_{\text{PAH}} < \rho_{\text{PAH}+}) = 0.934$



⁽Bell et al., in press)

- $p(
 ho_{
 m dust} <
 ho_{
 m PAH}) = 1.000$
- $p(
 ho_{\mathsf{dust}} <
 ho_{\mathsf{PAH}+}) = 1.000$
- $p(\rho_{\text{PAH}} < \rho_{\text{PAH}+}) = 0.934 \Rightarrow$ better correlation with PAH⁺.




























Outline of the Talk

1 Motivations & Principles of Bayesian Inference

- The Trickiness of Dust SED Fitting
- Principles of Bayesian Sampling
- Formalism & Numerical Implementation

2 Demonstration of the Code's Performances

- The Reference Simulation Grid
- Variations on the Reference Grid

3 Examples of Applications to Real Astrophysical Data

- Spinning Grains in λ -Orionis
- Dust Evolution in Galaxies

Conclusion & Prospectives

Benchmarking Summary

Benchmarking Summary

Simulations to test the performances varying S/N, sample size & SED shape:

- All parameters well recovered;
- Optimal uncertainties;
- 3 Systematically better than χ^2 .

Benchmarking Summary

Simulations to test the performances varying S/N, sample size & SED shape:

- All parameters well recovered;
- 2 Optimal uncertainties;
- 3 Systematically better than χ^2 .

Variations on the reference grid:

- 1 Very efficient at recovering accurate correlation coefficients;
- 2 Can compensate the lack of spectral coverage;
- 3 External parameters in the prior \Rightarrow accuracy increase.

Benchmarking Summary

Simulations to test the performances varying S/N, sample size & SED shape:

- All parameters well recovered;
- 2 Optimal uncertainties;
- 3 Systematically better than χ^2 .

Variations on the reference grid:

- Very efficient at recovering accurate correlation coefficients;
- 2 Can compensate the lack of spectral coverage;
- 3 External parameters in the prior \Rightarrow accuracy increase.

Prospectives: Further Development & Applications

1 Include complex radiative transfer models.

Benchmarking Summary

Simulations to test the performances varying S/N, sample size & SED shape:

- All parameters well recovered;
- 2 Optimal uncertainties;
- 3 Systematically better than χ^2 .

Variations on the reference grid:

- Very efficient at recovering accurate correlation coefficients;
- 2 Can compensate the lack of spectral coverage;
- 3 External parameters in the prior \Rightarrow accuracy increase.

Prospectives: Further Development & Applications

- Include complex radiative transfer models.
- **2** Same approach to decomposing mid-IR spectra \Rightarrow JWST (Dangning Hu's thesis).

Benchmarking Summary

Simulations to test the performances varying S/N, sample size & SED shape:

- All parameters well recovered;
- 2 Optimal uncertainties;
- 3 Systematically better than χ^2 .

Variations on the reference grid:

- Very efficient at recovering accurate correlation coefficients;
- 2 Can compensate the lack of spectral coverage;
- 3 External parameters in the prior \Rightarrow accuracy increase.

Prospectives: Further Development & Applications

- Include complex radiative transfer models.
- 2 Same approach to decomposing mid-IR spectra \Rightarrow JWST (Dangning Hu's thesis).
- 3 Application to SPICA (2030s) observations.