Hyperparameter selection comes for free with Sequential Monte Carlo samplers

Alberto Sorrentino

Dipartimento di Matematica, Università di Genova

Bayesian Estimation for Engineering Solutions (BEES)

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

with

- y measured data

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

with

- y measured data
- S unknown number of *sources*

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

with

- y measured data
- S unknown number of *sources*
- $f(\cdot)$ known *forward* model

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

with

- y measured data
- S unknown number of *sources*
- $f(\cdot)$ known *forward* model
- N is noise
- p_s, x_s unknown *parameters* identifying a single source

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

with

- y measured data
- S unknown number of *sources*
- $f(\cdot)$ known *forward* model
- N is noise
- p_s, x_s unknown *parameters* identifying a single source

- non-linear part p: location, shape of the source

Application context

A multi-source problem: many inverse/inference problems can be written as

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

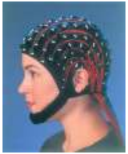
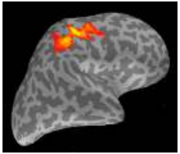
with

- y measured data
- S unknown number of *sources*
- $f(\cdot)$ known *forward* model
- N is noise
- p_s, x_s unknown *parameters* identifying a single source

- non-linear part p: location, shape of the source
- linear part x: strength of the source

Two examples

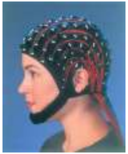
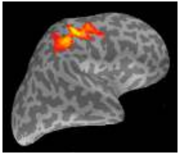
Brain imaging M/EEG

- y electric/magnetic field

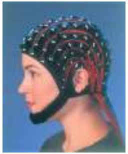
Two examples

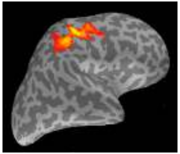
Brain imaging M/EEG

- y electric/magnetic field
- $f(\cdot)$ Biot Savart

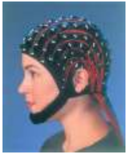
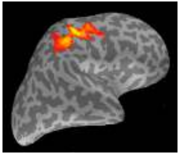
Brain imaging M/EEG





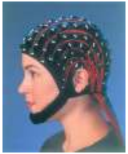
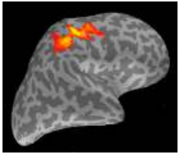
- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents

Brain imaging M/EEG

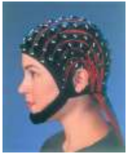
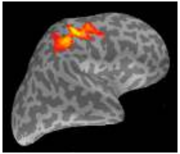
- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location

Brain imaging M/EEG

- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location
- x neural current intensity

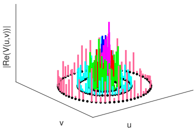
Brain imaging M/EEG

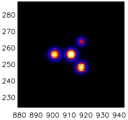



- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location
- x neural current intensity

S. et al. **Inverse Problems** 2014

Solar (flare) imaging

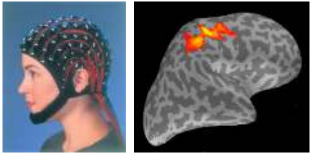




- y visibilities

Two examples

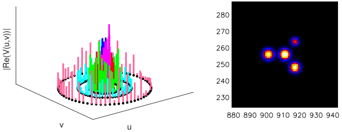
Brain imaging M/EEG



- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location
- x neural current intensity

S. et al. **Inverse Problems** 2014

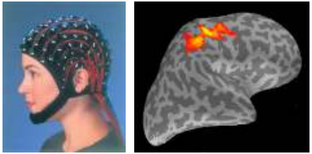
Solar (flare) imaging



- y visibilities
- $f(\cdot)$ Fourier Transform

Two examples

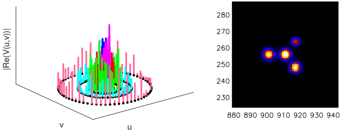
Brain imaging M/EEG



- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location
- x neural current intensity

S. et al. **Inverse Problems** 2014

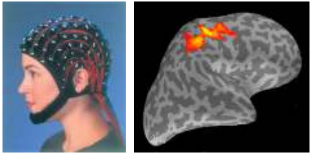
Solar (flare) imaging



- y visibilities
- $f(\cdot)$ Fourier Transform
- parametric model: geometric shapes
- p flare location, shape

Two examples

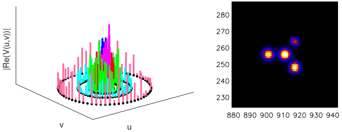
Brain imaging M/EEG



- y electric/magnetic field
- $f(\cdot)$ Biot Savart
- parametric model: point–like (dipolar) currents
- p neural current location
- x neural current intensity

S. et al. **Inverse Problems** 2014

Solar (flare) imaging



- y visibilities
- $f(\cdot)$ Fourier Transform
- parametric model: geometric shapes
- p flare location, shape
- x flare intensity

Sciacchitano, Lugaro and S. **SIAM Journal of Imaging Science** 2019

Parallel with linear inverse problems (Austin et al. 2010)

NB: the multi-source model

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

is just an uncommon, explicit representation of the linear inverse model on the right:

Parallel with linear inverse problems (Austin et al. 2010)

NB: the multi-source model

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

is just an uncommon, explicit representation of the linear inverse model on the right:

$$y = \sum_{b=1}^{B} f(p_b) \cdot x_b + N = F \cdot X + N$$

with B (large) number of *basis elements* (voxels/pixels/other) for which a *sparse solution* is sought

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(p_1) & f_1(p_2) & \dots & f_1(p_B) \\ f_2(p_1) & f_2(p_2) & \dots & f_2(p_B) \\ \dots & \dots & \dots & \dots \\ f_n(p_1) & f_n(p_2) & \dots & f_n(p_B) \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_B \end{pmatrix}$$

Parallel with linear inverse problems (Austin et al. 2010)

NB: the multi-source model

$$y = \sum_{s=1}^{S} f(p_s) \cdot x_s + N$$

is just an uncommon, explicit representation of the linear inverse model on the right:

(in red: unknowns)

$$y = \sum_{b=1}^{B} f(p_b) \cdot x_b + N = F \cdot X + N$$

with B (large) number of *basis elements* (voxels/pixels/other) for which a *sparse solution* is sought

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(p_1) & f_1(p_2) & \dots & f_1(p_B) \\ f_2(p_1) & f_2(p_2) & \dots & f_2(p_B) \\ \dots & \dots & \dots & \dots \\ f_n(p_1) & f_n(p_2) & \dots & f_n(p_B) \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_B \end{pmatrix}$$

Multi–source model

Aim: to estimate S and
$\{p_1, x_1, \ldots, p_S, x_S\}$.

Multi–source model

Aim: to estimate S and
$\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g. $\mathbb{R}^n$);
then configuration space:

$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit representation?

Multi–source model

Aim: to estimate S and $\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g. $\mathbb{R}^n$);
then configuration space:

$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit representation?
things get more involved

Multi–source model

Aim: to estimate S and
$\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g.
$\mathbb{R}^n$);
then configuration space:

$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit
representation?
things get more involved
non-linearity is bad

Multi–source model

Aim: to estimate S and $\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g. $\mathbb{R}^n$);
then configuration space:

$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit representation?
things get more involved
non-linearity is bad
interpretation is straightforward

Multi–source model

Aim: to estimate S and
$\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g. $\mathbb{R}^n$);
then configuration space:

$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit representation?
things get more involved
non-linearity is bad
interpretation is straightforward
we can go bayesian

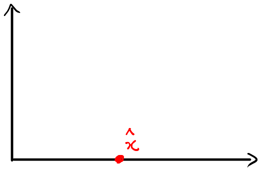
Aim: to estimate S and $\{p_1, x_1, \ldots, p_S, x_S\}$.

$\mathcal{S}$ single source parameter space (e.g. $\mathbb{R}^n$);
then configuration space:

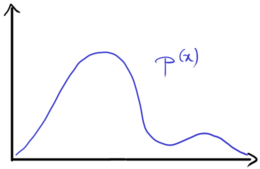
$$\mathcal{X} = \bigcup_k \mathcal{S}^k$$

Why stick with the explicit representation?
things get more involved
non-linearity is bad
interpretation is straightforward
we can go bayesian

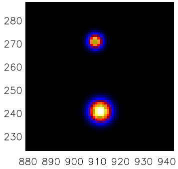
Change of perspective: from point estimate $\hat{x}$



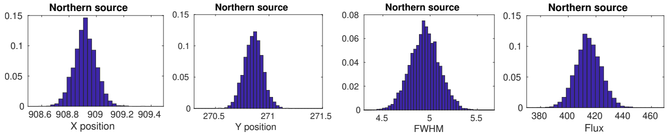
to whole probability distribution $\pi(x) \forall x \in \mathcal{X}$

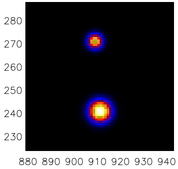


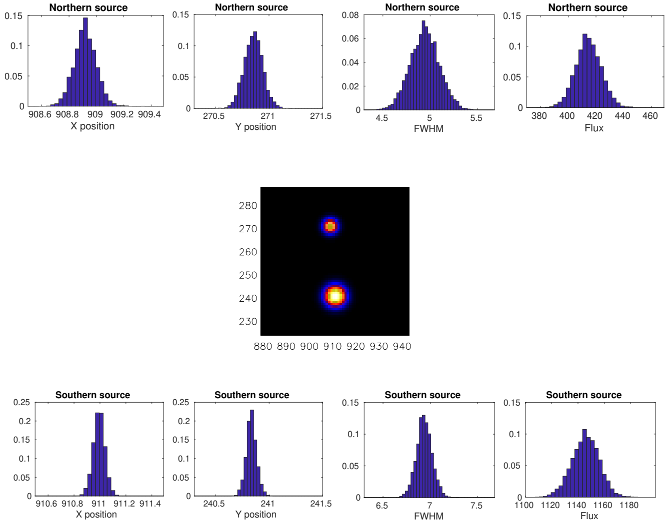
More than imaging



More than imaging






Northern source
X position
Northern source
Y position
Northern source
FWHM
Northern source
Flux
Southern source
X position
Southern source
Y position
Southern source
FWHM
Southern source
Flux

Bayesian multi-source model

Choose a prior distribution on $\mathcal{X}$
(from here on, x is the whole set of unkonwns)

$$\pi(x) = \mathbb{P}(s)\pi(p_1, x_1, \ldots, p_s, x_s)$$

with

- $\mathbb{P}(s)$ probability of s sources
- $\pi(p_1, x_1, \ldots, p_s, x_s)$ prior of specific configuration

Bayesian multi-source model

Choose a prior distribution on $\mathcal{X}$
(from here on, x is the whole set of unkonwns)

$$\pi(x) = \mathbb{P}(s)\pi(p_1, x_1, \dots, p_s, x_s)$$

with

- $\mathbb{P}(s)$ probability of s sources
- $\pi(p_1, x_1, \dots, p_s, x_s)$ prior of specific configuration

Choose a likelihood

$$\pi(y|x) \propto \exp\left(-\frac{\|y - \sum_s f(p_s)x_s\|^2}{\sigma^2}\right)$$

- σ^2 noise variance

Bayesian multi-source model

Choose a prior distribution on $\mathcal{X}$
(from here on, x is the whole set of unkonwns)

$$\pi(x) = \mathbb{P}(s)\pi(p_1, x_1, \ldots, p_s, x_s)$$

with

- $\mathbb{P}(s)$ probability of s sources
- $\pi(p_1, x_1, \ldots, p_s, x_s)$ prior of specific configuration

Choose a likelihood

$$\pi(y|x) \propto \exp\left(-\frac{\|y - \sum_s f(p_s)x_s\|^2}{\sigma^2}\right)$$

- σ^2 noise variance

Compute/Characterize the posterior

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

Bayesian multi-source model

Choose a prior distribution on $\mathcal{X}$
(from here on, x is the whole set of unkonwns)

$$\pi(x) = \mathbb{P}(s)\pi(p_1, x_1, \ldots, p_s, x_s)$$

with

- $\mathbb{P}(s)$ probability of s sources
- $\pi(p_1, x_1, \ldots, p_s, x_s)$ prior of specific configuration

Choose a likelihood

$$\pi(y|x) \propto \exp\left(-\frac{\|y - \sum_s f(p_s)x_s\|^2}{\sigma^2}\right)$$

- σ^2 noise variance

Compute/Characterize the posterior

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

NB: it depends on (at least) one (hyper)parameter

- sparse linear ip

$$\arg\min \|y - \sum_{b=1}^{B} f(p_b) \cdot x_b\|^2 + \lambda \sum_{b=1}^{B} |x_b|$$

Parallel with sparse linear inverse problems

- sparse linear ip

$$\arg\min \|y - \sum_{b=1}^{B} f(p_b) \cdot x_b\|^2 \; + \; \lambda \sum_{b=1}^{B} |x_b|$$

- Bayesian multi–source model

$$\pi(x|y) \propto \exp\left(-\frac{\|y - \sum_{s=1}^{S} f(p_s) \cdot x_s\|^2}{\sigma^2}\right) \quad \mathbb{P}(s)\pi(x_1, \ldots, x_s)$$

Parallel with sparse linear inverse problems

- sparse linear ip

$$\arg\min \|y - \sum_{b=1}^{B} f(p_b) \cdot x_b\|^2 \; + \; \lambda \sum_{b=1}^{B} |x_b|$$

- Bayesian multi–source model

$$\pi(x|y) \propto \exp\left(- \frac{\|y - \sum_{s=1}^{S} f(p_s) \cdot x_s\|^2}{\sigma^2} \right) \quad \mathbb{P}(s)\pi(x_1, \ldots, x_s)$$

$\mathbb{P}(s)$ can penalize larger models, e.g. $\mathbb{P}(s) = Poiss(\gamma)$ with $\gamma < 1$.

Monte Carlo approximation

We aim at *characterizing*

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

Monte Carlo approximation

We aim at *characterizing*

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

Goal (of Monte Carlo methods): obtain a set of weighted points $\{x^{(p)}\}_{p=1,\dots,P}$ (particles) that represent the posterior distribution, i.e. such that

$$\sum_p h(x^{(p)})w^{(p)} \to \int h(x)\pi(x|y)dx$$

for any (well behaved) function $h(\cdot)$

Monte Carlo approximation

We aim at *characterizing*

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

Goal (of Monte Carlo methods): obtain a set of weighted points $\{x^{(p)}\}_{p=1,\ldots,P}$ (particles) that represent the posterior distribution, i.e. such that

$$\sum_p h(x^{(p)})w^{(p)} \rightarrow \int h(x)\pi(x|y)dx$$

for any (well behaved) function $h(\cdot)$

$$\pi(x|y) \simeq \sum_p w^{(p)}\delta(x - x^{(p)})$$

sampling $\longleftrightarrow$ **approximating**

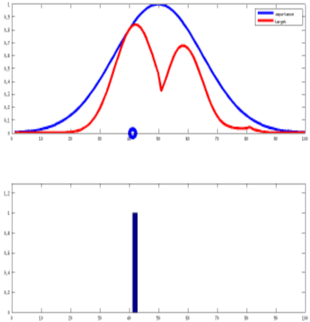
$$\int h(x)\pi(x|y)dx = \int h(x)\frac{\pi(x|y)}{\eta(x)}\eta(x)dx$$

$\eta(x)$ importance density s.t.
$\pi(x|y) > 0 \rightarrow \eta(x) > 0$;

if $\{x^{(p)}\}_{p=1,\dots,P}$ i.i.d. from $\eta(x)$
LLN guarantees

$$\sum_p \frac{1}{P}\frac{\pi(x^{(p)}|y)}{\eta(x^{(p)})}h(x^{(p)}) \rightarrow \int h(x)\pi(x|y)dx$$

- **global** and **parallel**



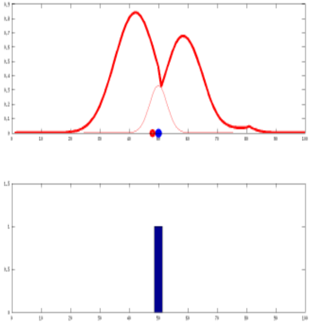
An *irreducible*, *aperiodic* Markov Chain has *invariant* distribution.

Build a $\pi(x|y)$-invariant kernel $K(x'|x)$, then sample $x^{(p+1)}$ from $K(\cdot|x^{(p)})$ ergodic theorem

$$\sum_p \frac{1}{P}\delta(x - x^{(p)}) \simeq \pi(x|y)$$

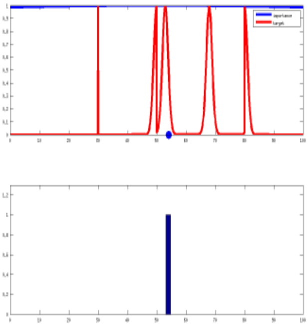
Example: Metropolis-Hastings proposal+acceptance/rejection.

- **local** and **serial**

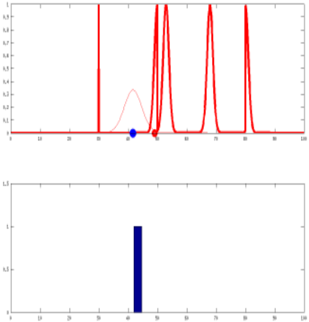


Limitations of basic MC

Importance Sampling

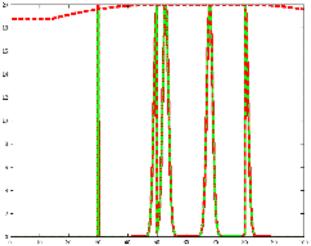


Markov Chain Monte Carlo



Sequential Monte Carlo (Del Moral et al. 2006) construct a sequence of distributions $\pi_1(x), \ldots, \pi_I(x)$ such that

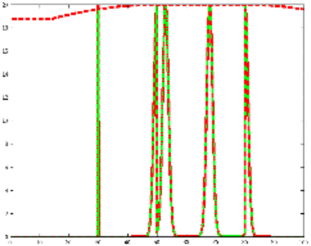
- $\pi_1(x) = \pi(x)$ is the prior distribution
- $\pi_I(x) = \pi(x|y)$ is the posterior distribution
- $\pi_i(x) \sim \pi_{i-1}(x)$



Sequential Monte Carlo samplers

Sequential Monte Carlo (Del Moral et al. 2006) construct a sequence of distributions $\pi_1(x), \ldots, \pi_I(x)$ such that

- $\pi_1(x) = \pi(x)$ is the prior distribution
- $\pi_I(x) = \pi(x|y)$ is the posterior distribution
- $\pi_i(x) \sim \pi_{i-1}(x)$



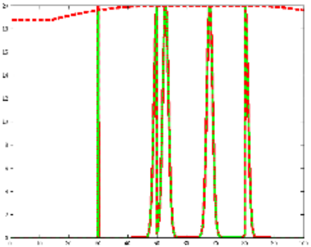
Our choice:

$$\pi_i(x) \propto \pi(y|x)^{\alpha_i} \pi(x)$$

$\alpha_i \in [0, 1]$, $\alpha_i < \alpha_{i+1}$

Sequential Monte Carlo (Del Moral et al. 2006) construct a sequence of distributions $\pi_1(x), \dots, \pi_I(x)$ such that

- $\pi_1(x) = \pi(x)$ is the prior distribution
- $\pi_I(x) = \pi(x|y)$ is the posterior distribution
- $\pi_i(x) \sim \pi_{i-1}(x)$



Our choice:

$$\pi_i(x) \propto \pi(y|x)^{\alpha_i} \pi(x)$$

$\alpha_i \in [0, 1]$, $\alpha_i < \alpha_{i+1}$
($\sim$ simulated annealing, tempering,...)

Sequential Monte Carlo samplers

General scheme:

0 Initialization: sample P *particles* $\{x_1^{(p)}\}_{p=1,\dots,P}$ from the prior distribution $\pi_1(x)$;

General scheme:

0 Initialization: sample P *particles* $\{x_1^{(p)}\}_{p=1,\ldots,P}$ from the prior distribution $\pi_1(x)$;

then iterate

i MCMC step: for every sample (*particle*) propose and accept/reject

Sequential Monte Carlo samplers

General scheme:

0 Initialization: sample P *particles* $\{x_1^{(p)}\}_{p=1,\dots,P}$ from the prior distribution $\pi_1(x)$;

then iterate

i MCMC step: for every sample (*particle*) propose and accept/reject

ii importance sampling step: update the particle weight

$$w_i^{(p)} = w_{i-1}^{(p)} \frac{\pi_i(x_{i-1}^{(p)})}{\pi_{i-1}(x_{i-1}^{(p)})}$$

Sequential Monte Carlo samplers

General scheme:

0 Initialization: sample P *particles* $\{x_1^{(p)}\}_{p=1,\dots,P}$ from the prior distribution $\pi_1(x)$;

then iterate

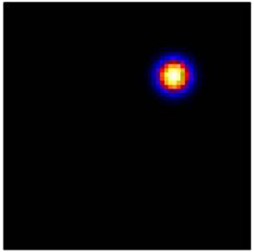
i MCMC step: for every sample (*particle*) propose and accept/reject

ii importance sampling step: update the particle weight

$$w_i^{(p)} = w_{i-1}^{(p)} \frac{\pi_i(x_{i-1}^{(p)})}{\pi_{i-1}(x_{i-1}^{(p)})}$$

N.B.: at iteration i, $\{x_i^{(p)}, w_i^{(p)}\}_{p=1,\dots,P}$ approximate

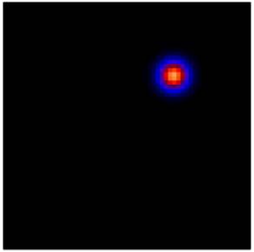
$$\pi_i(x) \propto \pi(y|x)^{\gamma_i}\pi(x)$$

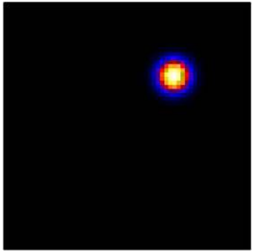


Classic parameter update

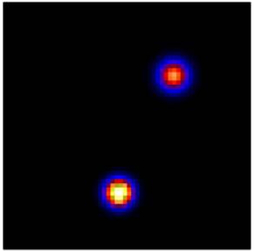
$\Longrightarrow$

$\Longleftarrow$

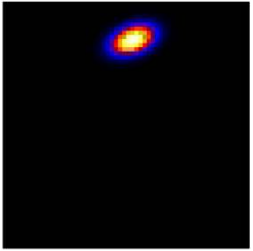




Birth
$\Longrightarrow$
Death
$\Longleftarrow$



MCMC moves
Exploring the configuration space

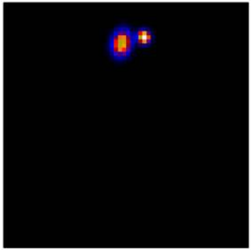


Split

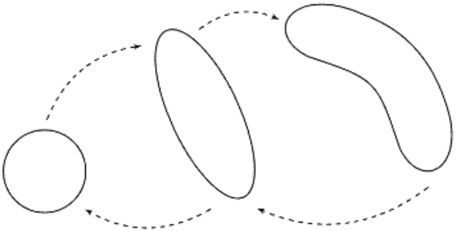
$\Longrightarrow$

Merge

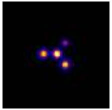
$\Longleftarrow$



Type change

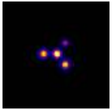


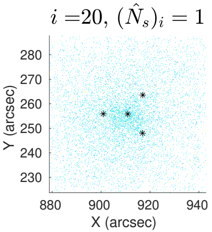
True
image

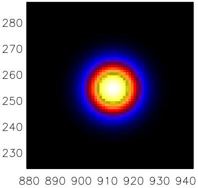


An insight into the iterative procedure (Sciacchitano et al. 2019)

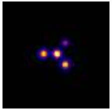
True image

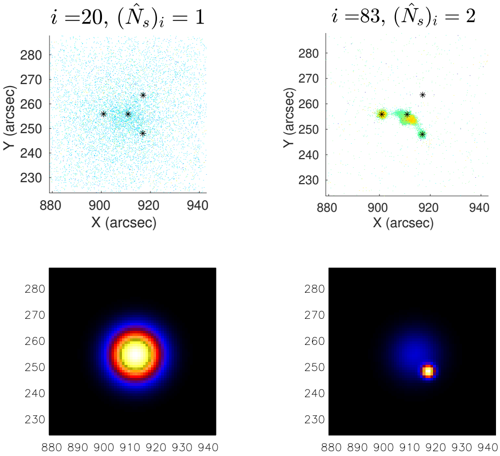






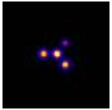
True image

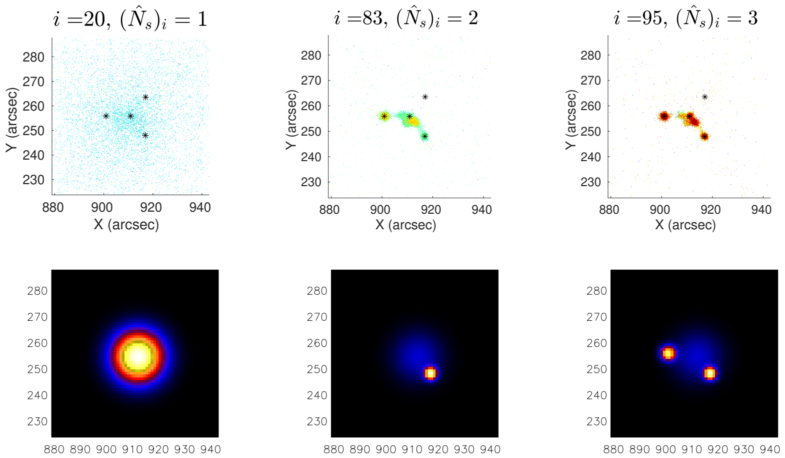



i =20, $(\hat{N}_s)_i = 1$
i =83, $(\hat{N}_s)_i = 2$
Y (arcsec)
X (arcsec)
280
270
260
250
240
230
880
900
920
940
880 890 900 910 920 930 940

An insight into the iterative procedure
(Sciacchitano et al. 2019)

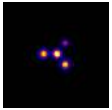
True image



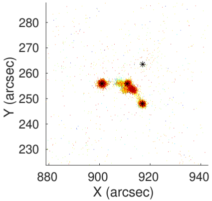


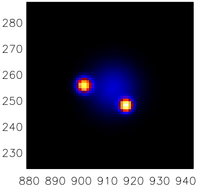
An insight into the iterative procedure
(Sciacchitano et al. 2019)

True source



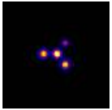
$i = 95,\ (\hat{N}_s)_i = 3$

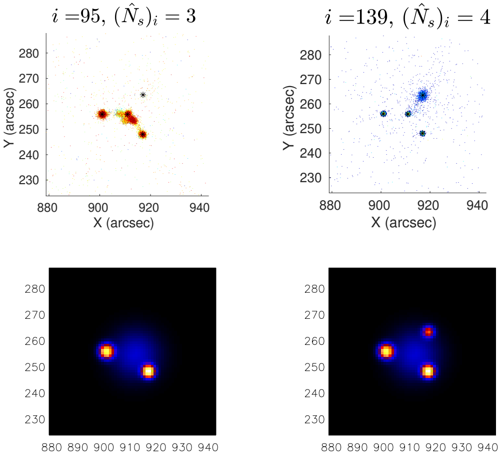




An insight into the iterative procedure
(Sciacchitano et al. 2019)

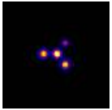
True source

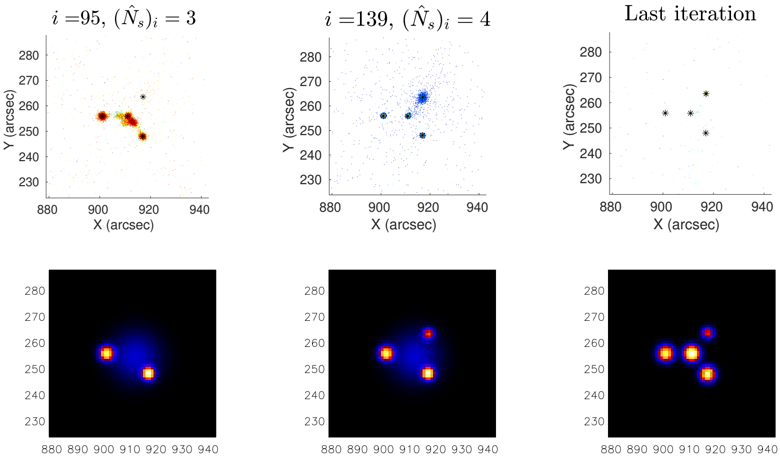




An insight into the iterative procedure (Sciacchitano et al. 2019)

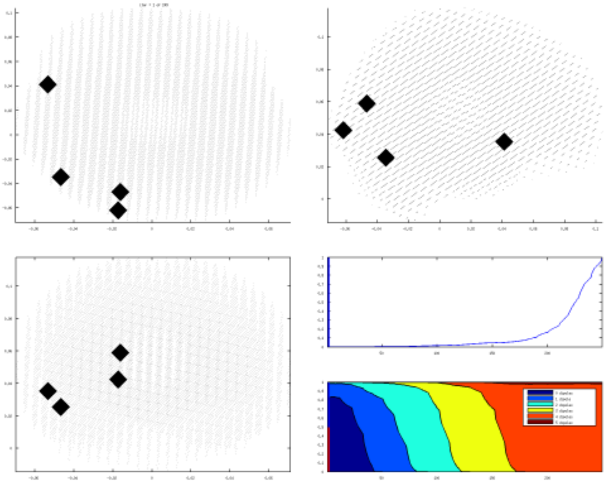
True source





Example on brain imaging
(Sorrentino et al. 2014)

True locations. **Density $\geq 1/1000$.** **Point estimates.**



Better exploiting SMC samplers
Viani and S., in preparation

Problem/opportunity: Intermediate distributions are typically discarded
But they look nice! Can we use them?

Problem/opportunity: Intermediate distributions are typically discarded
But they look nice! Can we use them?
Typical construction of the sequence is

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

with

- $\alpha_1 = 0$ - start from the prior $\pi(x)$
- $\alpha_I = 1$ - last distribution is the posterior $\pi(x|y)$

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

For a likelihood in the natural exponential family with natural parameter θ

$$\pi_\theta(y|x)^{\alpha_i} \propto \pi_{\alpha_i\theta}(y|x)$$

raising to a power corresponds to rescaling the parameter!

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

For a likelihood in the natural exponential family with natural parameter θ

$$\pi_\theta(y|x)^{\alpha_i} \propto \pi_{\alpha_i \theta}(y|x)$$

raising to a power corresponds to rescaling the parameter!
Example: for a Gaussian likelihood, $\theta = 1/\sigma^2$ and

$$\left(\exp - \frac{(y - f(x))^2}{2\sigma^2}\right)^{\alpha_i} = \exp\left(-\frac{(y - f(x))^2}{2\left(\frac{\sigma}{\sqrt{\alpha_i}}\right)^2}\right)$$

Intermediate distributions are posterior distributions with a different value of the hyperparameter!

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

For a likelihood in the natural exponential family with natural parameter θ

$$\pi_\theta(y|x)^{\alpha_i} \propto \pi_{\alpha_i\theta}(y|x)$$

raising to a power corresponds to rescaling the parameter!
Example: for a Gaussian likelihood, $\theta = 1/\sigma^2$ and

$$\left(\exp -\frac{(y-f(x))^2}{2\sigma^2}\right)^{\alpha_i} = \exp\left(-\frac{(y-f(x))^2}{2\left(\frac{\sigma}{\sqrt{\alpha_i}}\right)^2}\right)$$

Intermediate distributions are posterior distributions with a different value of the hyperparameter!
For Gaussian likelihood, we have shrinking *effective* variance $\sigma_i = \frac{\sigma}{\sqrt{\alpha_i}}$

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$

$$
\begin{aligned}
\pi_i(x) &= \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)} \\
&= \frac{\pi(x)\pi(y|x,\alpha_i\theta)}{\pi_i(y|\alpha_i\theta)}
\end{aligned}
$$

Two key observations
Observation 2

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$
$$= \frac{\pi(x)\pi(y|x, \alpha_i\theta)}{\pi_i(y|\alpha_i\theta)}$$

the normalization constant (denominator) becomes

$$\pi_i(y) = \pi(y|\theta_i) \qquad \theta_i = \theta\alpha_i$$

and SMC samplers provide an estimate of the normalization constant!

$$\pi_i(x) = \frac{\pi(x)\pi(y|x)^{\alpha_i}}{\pi_i(y)}$$
$$= \frac{\pi(x)\pi(y|x, \alpha_i\theta)}{\pi_i(y|\alpha_i\theta)}$$

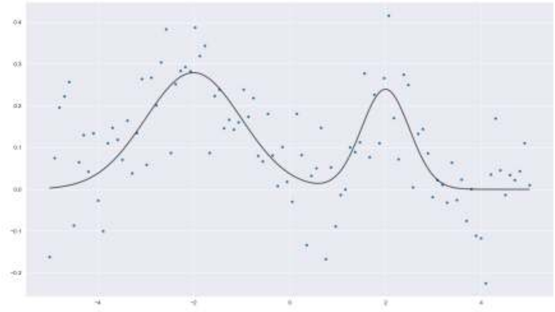
the normalization constant (denominator) becomes

$$\pi_i(y) = \pi(y|\theta_i) \qquad \theta_i = \theta\alpha_i$$

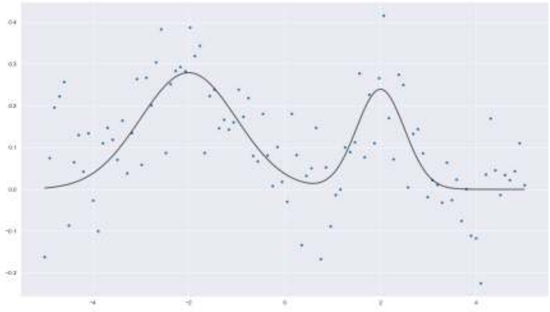
and SMC samplers provide an estimate of the normalization constant!
We therefore have the *evidence* for θ for free (evaluated at $\{\theta_i = \alpha_i\theta\}_{i=1,\dots,I}$)

$$\pi(\theta_i|y) \propto \pi(y|\theta_i)\pi(\theta_i)$$

Given a hyperprior $\pi(\theta)$, we can select the hyperparameter, or average across it! We can embody uncertainty on the hyperparameter.



Uknown parameters:
number of Gaussian functions;
mean, variance and height of each Gaussian function;



Uknown parameters:
number of Gaussian functions;
mean, variance and height of each Gaussian function;

measured data:
perturbed samples of the mixture, Gaussian noise

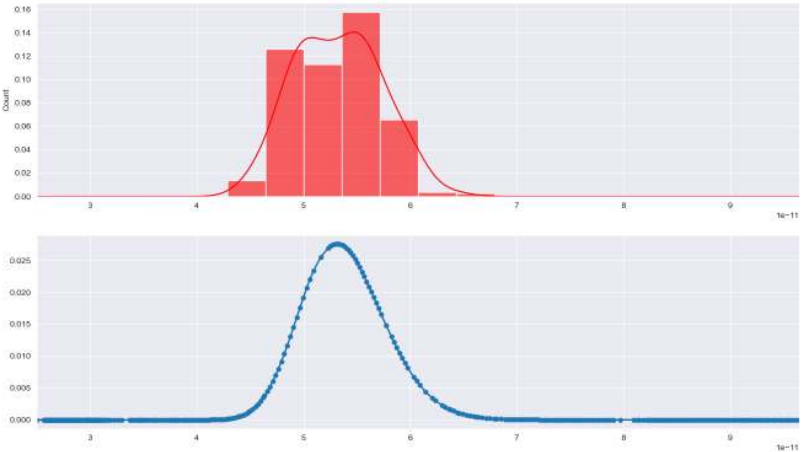
We compare our method with a standard approach:

- augment state space with σ
- use Monte Carlo to sample from

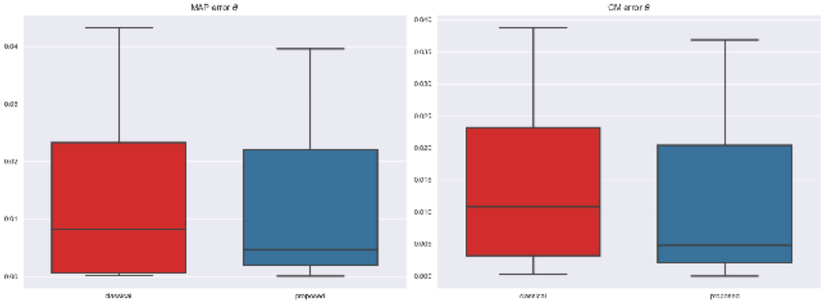
$$\pi(s, x_1, \ldots, x_s, \sigma | y)$$

Computationally (somewhat) heavier

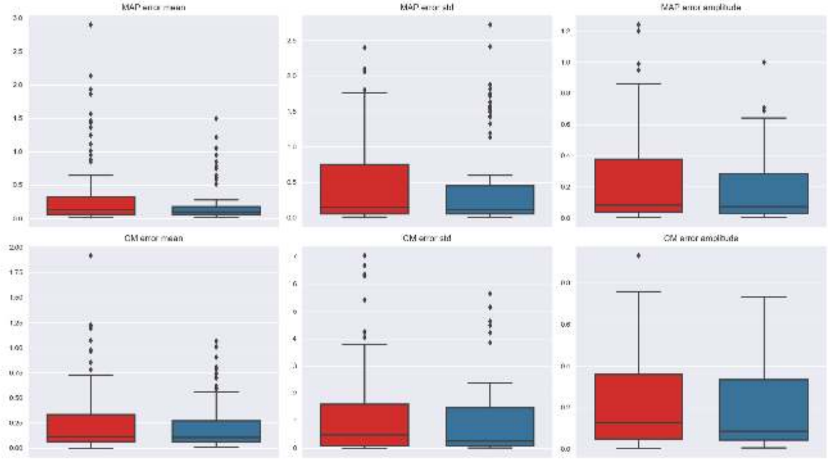
The posterior distribution of the hyperparameter [top: standard; bottom: new]



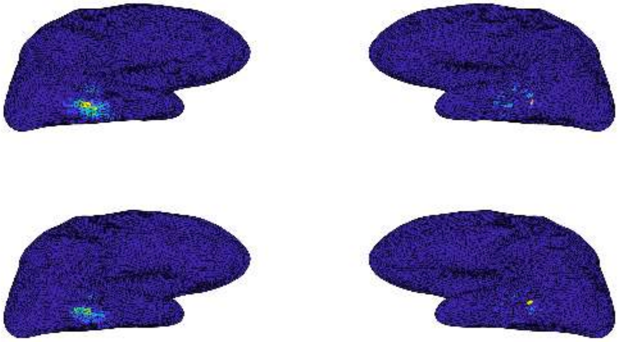
Error on hyperparameter estimate [red: standard; blue: new]



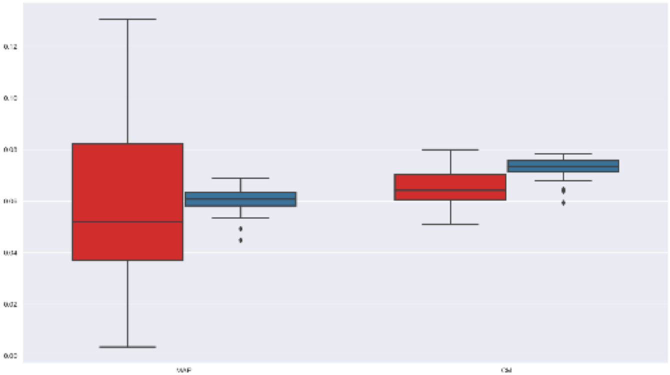
Error on parameter estimates [red: standard; blue: new]



More accurate probability maps [top: new; bottom: standard;



Standard approach provides higher variance, often better estimates of the hyperparameter [red: standard; blue: new]



Perspectives

- understand imaging application (...)

Summarizing

Perspectives

- understand imaging application (...)
- work on the adaptive choice of $\{\alpha_i\}_{i=1,\dots,I}$

Summarizing

Perspectives

- understand imaging application (...)
- work on the adaptive choice of $\{\alpha_i\}_{i=1,\ldots,I}$
- find reasonable hyperpriors

Perspectives

- understand imaging application (...)
- work on the adaptive choice of $\{\alpha_i\}_{i=1,\dots,I}$
- find reasonable hyperpriors

Conclusions

- regularization parameter selected for free

Summarizing

Perspectives

- understand imaging application (...)
- work on the adaptive choice of $\{\alpha_i\}_{i=1,\ldots,I}$
- find reasonable hyperpriors

Conclusions

- regularization parameter selected for free
- averaging across different values embodies uncertainty

- growing need for sound statistical models

- growing need for sound statistical models
- widening gap btw theory and applications

- growing need for sound statistical models
- widening gap btw theory and applications
- different problems, same methods

Bayesian Estimation for Engineering Solutions s.r.l.

- growing need for sound statistical models
- widening gap btw theory and applications
- different problems, same methods
- startup, born June 2021 as a spinoff of UNIGE

Bayesian Estimation for Engineering Solutions s.r.l.

- growing need for sound statistical models
- widening gap btw theory and applications
- different problems, same methods
- startup, born June 2021 as a spinoff of UNIGE
- explore commercial potential of Bayesian models and Monte Carlo algorithms

Bayesian Estimation for Engineering Solutions s.r.l.

- growing need for sound statistical models
- widening gap btw theory and applications
- different problems, same methods
- startup, born June 2021 as a spinoff of UNIGE
- explore commercial potential of Bayesian models and Monte Carlo algorithms
- nothing to do with actual bees...

Acknowledgements

Riccardo Aramini
Silvio Lugaro
Gianvittorio Luria
Michele Piana
Federica Sciacchitano
Sara Sommariva
Alessandro Viani

Acknowledgements

Riccardo Aramini
Silvio Lugaro
Gianvittorio Luria
Michele Piana
Federica Sciacchitano
Sara Sommariva
Alessandro Viani

That's all!

Thank you