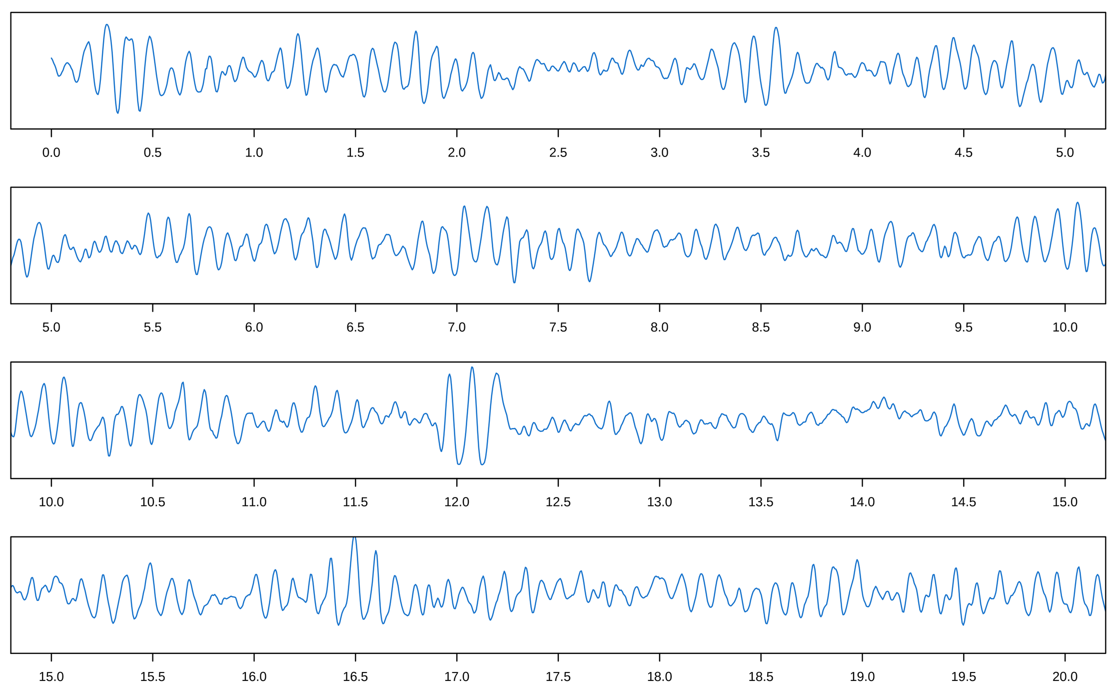
Structure-preserving Approximate Bayesian Computation for complex stochastic models

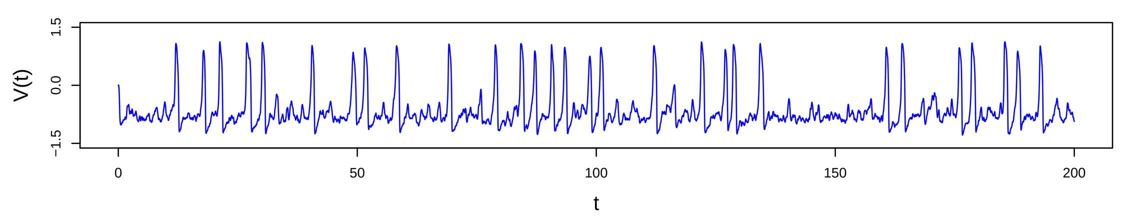
Massimiliano Tamborrino, Department of Statistics, University of Warwick.
massimiliano.tamborrino@warwick.ac.uk www.warwick.ac.uk/tamborrino

WARWICK
THE UNIVERSITY OF WARWICK

Considered neural recordings



20s α-rhytmic EEG data, sampling rate 173.61 Hz, $\Delta \approx 5.76ms$ [1].



200s recording of membrane voltage (simulated data)

Stochastic models and setting of interest

X: d-dimensional stochastic process depending on $\theta \in \Theta \subseteq \mathbb{R}^k$,

$$dX_t = F(X_t;\theta)dt + \Sigma(X_t;\theta)dW_t, \qquad t\in[0,T], \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

X, F and W d-dimensional, Σ: $d\times d$ matrix. State space: $D \subseteq \mathbb{R}^d$.

Model properties

- It exists an invariant distribution.
- X partially observed via $Y_\theta = g(X), g: \mathbb{R}^d \to \mathbb{R}^m$.
- The noise may not enter in all components (Σ_{ii} may be 0).

1. Stochastic Harmonic Oscillator

$$d\underbrace{\begin{pmatrix} Q_t \\ P_t \end{pmatrix}}_{:=X_t} = \underbrace{\begin{pmatrix} P_t \\ -\lambda^2 Q_t - 2\gamma P_t \end{pmatrix}}_{:=F(X_t;\theta)} dt + \underbrace{\begin{pmatrix} 0 \\ \sigma \end{pmatrix}}_{:=\Sigma(\theta)} dW_t, \quad X_0 = x_0.$$

with $\lambda^2-\gamma^2>0$ (weakly damped case), $\theta=(\lambda,\gamma,\sigma)$ and $Y_\theta = Q$.

2. Stochastic Jensen and Rit neural mass model (JRNMM)

$$d\underbrace{\begin{pmatrix} X_{1;t} \\ X_{2;t} \\ X_{3;t} \\ X_{4;t} \\ X_{5;t} \\ X_{6;t} \end{pmatrix}}_{:=X_t} = \underbrace{\begin{pmatrix} X_{4;t} \\ X_{5;t} \\ X_{6;t} \\ Aa\mathcal{S}\left(X_{2;t}-X_{3;t}\right) - 2aX_{4;t} - a^2X_{1;t} \\ Aa(\mu + C_2\mathcal{S}(C_1X_{1;t})) - 2aX_{5;t} - a^2X_{2;t} \\ BbC_4\mathcal{S}\left(C_3X_{1;t}\right) - 2bX_{6;t} - b^2X_{3;t} \end{pmatrix}}_{:=F(X_t;\theta)} dt + \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ \tau \\ \sigma \\ \tau \end{pmatrix}}_{:=\Sigma(\theta)} dW_t$$

with $X_0=x_0, C_1=C, C_2=0.8C, C_3=C_4=0.25C$, $\theta=(\sigma,\mu,C)$ and $Y_\theta = X_2 - X_3$.

3. Stochastic FitzHugh-Nagumo (FHN)

$$d\underbrace{\begin{pmatrix} V_t \\ U_t \end{pmatrix}}_{:=X_t} = \underbrace{\begin{pmatrix} \frac{1}{\epsilon}\left(V_t - V_t^3 - U_t\right) \\ \gamma V_t - U_t + \beta \end{pmatrix}}_{:=b(X_t)} dt + \underbrace{\begin{pmatrix} 0 \\ \sigma \end{pmatrix}}_{:=\Sigma} dW_t, \quad X_0 = x_0.$$

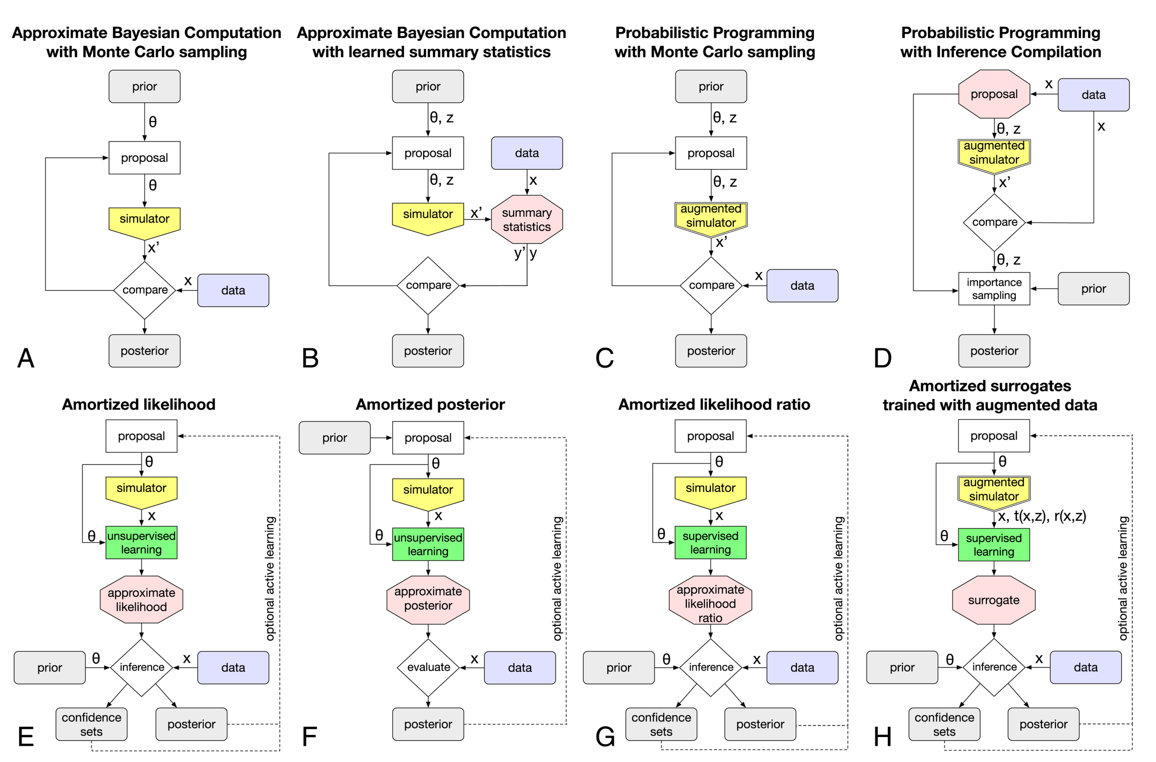
$\theta = (\epsilon,\gamma,\beta,\sigma)$ and $Y_\theta = V$.

Simulation-based inference

Goal: Estimate θ based on the available partial observations Y_θ.
Challenge: The underlying likelihood is intractable!

$$\underbrace{\pi(\theta|y)}_{posterior} \propto \underbrace{\pi(y|\theta)}_{\substack{likelihood \\ (intractable)}} \underbrace{\pi(\theta)}_{prior}$$

$\Rightarrow$ Likelihood-free approaches, here simulation-based inference, in particular Approximate Bayesian Computation (ABC) [2].



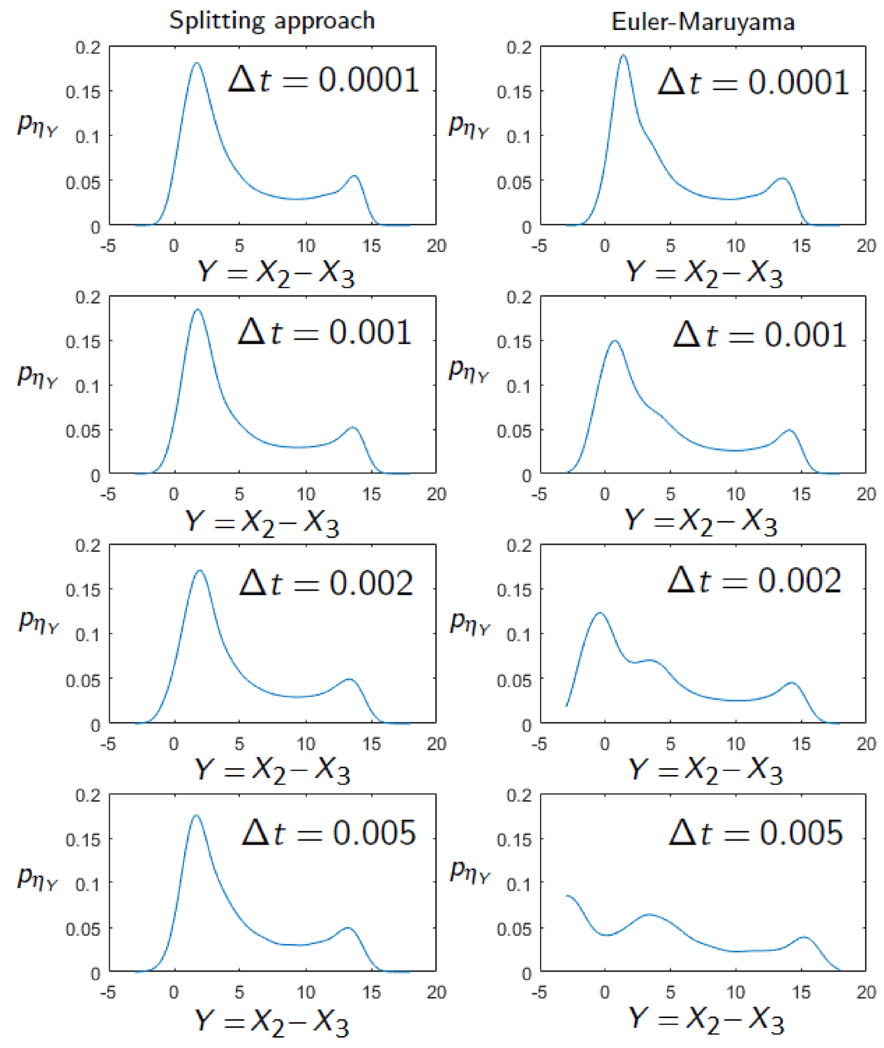
Overview of different approaches to simulation-based inference from [3].

Key ingredients in simulation-based inference

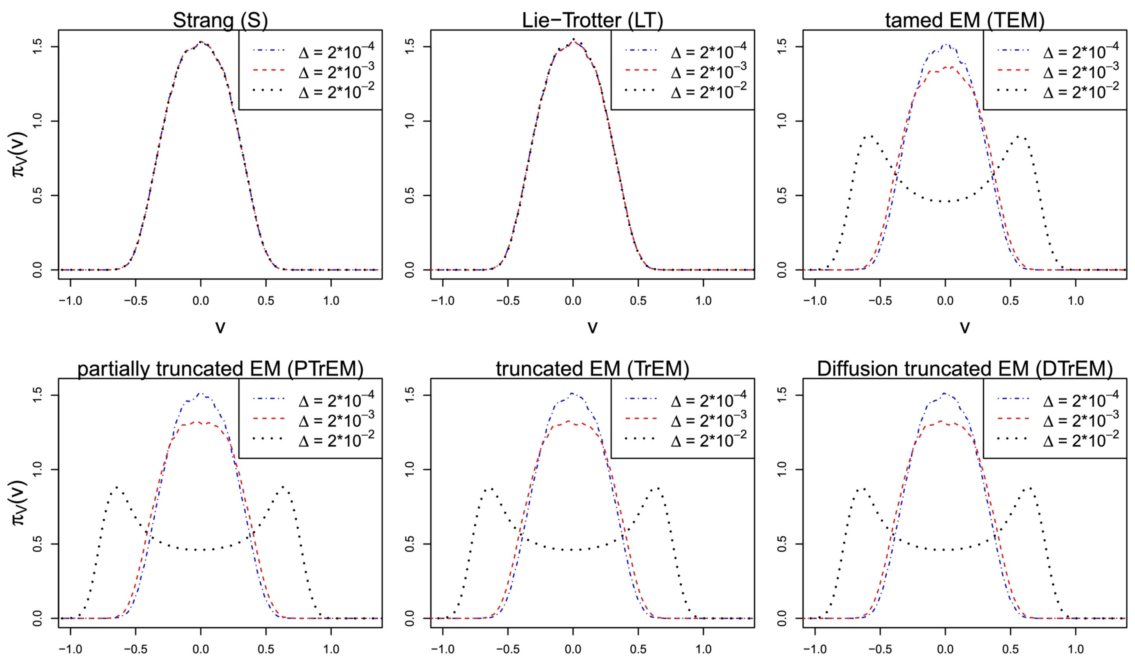
1. Ability to simulate from the model (simulator).
2. Choice of the summary statistics.

Ability to simulate from the model

Conditionally on θ^* from a proposal, we need to simulate a new realisation y_{θ^*} from the model (simulator).
Challenge: Exact simulation schemes are rarely available for SDEs.



Invariant density of the JRNMM estimated from simulated data [4]



Invariant density of the FHN estimated from simulated data [5].

1st Take home message

- Be sceptic with Taylor schemes (e.g. Euler-Maruyama and Milstein).
- Use reliable (converngent AND property-preserving) numerical schemes, here splitting schemes.

Numerical Splitting schemes in a nutshell

Consider $\widetilde{X}_t \approx X_t$. How to simulate $\widetilde{X}_{t_i}$ given $\widetilde{X}_{t_{i-1}}$?
Step 1: Split the SDE into explicitly solvable sub-equations.

$$F(X_t;\theta) = \sum_{j=1}^{d} F^{[j]}(X_t;\theta), \qquad \Sigma(X_t;\theta) = \sum_{j=1}^{d} \Sigma^{[j]}(X_t;\theta),$$

Step 2: Derive the explicit solutions of the sub-equations.

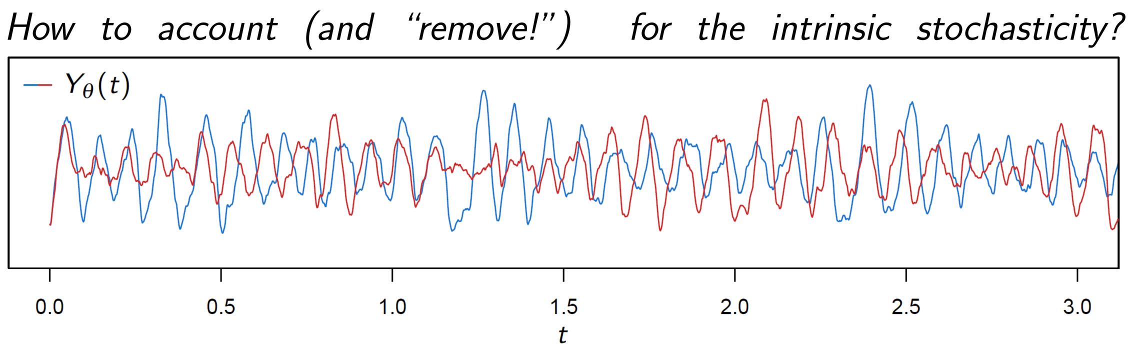
$$dX_t = F^{[j]}(X_t;\theta)dt + \Sigma^{[j]}(X_t;\theta)dW_t, j\in\{1,\ldots,d\}.$$

Step 3: Compose the derived explicit solutions $X_t^{[j]} = \phi_t^{[j]}(x_0)$ (Strang approach)

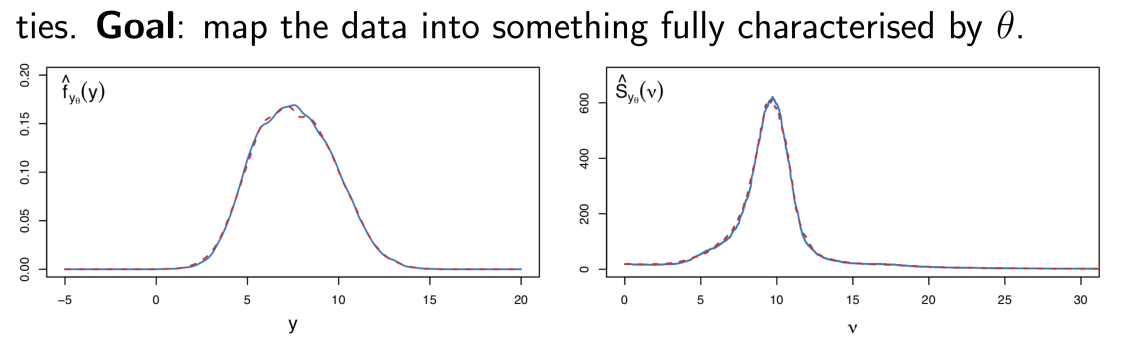
$$\widetilde{X}_{t_i} = \left(\phi^{[1]}_{\Delta/2} \circ \cdots \phi^{[d-1]}_{\Delta/2} \circ \phi^{[d]}_{\Delta} \circ \phi^{[d-1]}_{\Delta/2} \circ \cdots \phi^{[1]}_{\Delta/2}\right)(\widetilde{X}_{t_{i-1}}),$$

Choice of the summary statistics

How to account (and "remove!") for the intrinsic stochasticity?



Two realisations of Y_θ for the JRNMM with the same θ.
Proposal: Derive summaries based on the charactersing model properties. **Goal**: map the data into something fully characterised by θ.



Estimated invariant density (left) and invariant spectral density (right) for the data above.

2nd Take home message: Incorporate SDE dynamics and structural properties to obtain summaries less sensitive to the intrinsic stochasticity of the model.

Structure-preserving ABC[1]

Reference table acceptance-rejection ABC Algorithm
Input: Observed data y from Y_θ.
Output: Samples from the posterior $\pi^{\epsilon}_{\text{ABC}}(\theta|s_y)$.

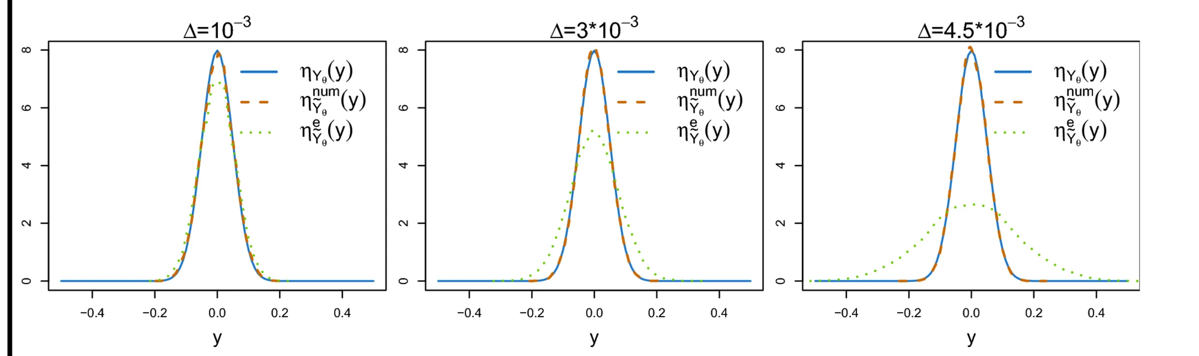
Choose a prior distribution $\pi(\theta)$ and a *percentile* p.
for $i=1:N$ **do**
1. Draw θ^* from the prior $\pi(\theta)$.
2. Conditionally on θ^*, simulate a new realisation y_{θ^i} from the model using the measure (property) preserving numerical splitting method.
3. Calculate the distance $D_i = \text{IAE}(\hat{S}_y, \hat{S}_{y^*_\theta}) + w \cdot \text{IAE}(\hat{f}_y, \hat{f}_{y^*_\theta}), w \geq 0$.

end for
Compute ϵ as the percentage p of the calculated distances.
For $i=1,\ldots,N$, keep θ^i as a sample from the ABC posterior if $D_i<\epsilon$.

Key features: inefficient but parallelisable

$$\Rightarrow \pi(\theta|y) \overset{?}{=} \pi(\theta|s_y) \approx \pi^{\epsilon}_{\text{ABC}}(\theta|s_y) = \pi(\ \theta \mid d(s_y, s_{y^*_\theta}) < \epsilon\ )$$

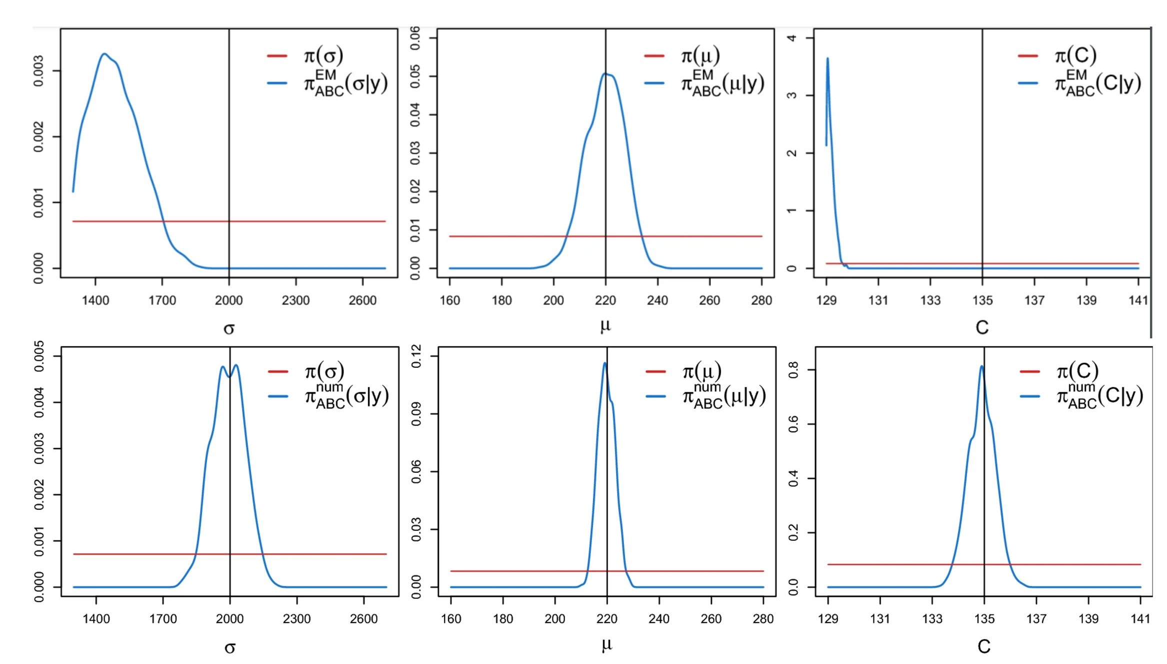
Stochastic harmonic oscillator with $\theta = \lambda$



- π_{ABC}: ABC posterior obtained with exact simulation;
- $\pi^{\text{num}}_{\text{ABC}}$: ABC posterior obtained with Strang splitting scheme;
- $\pi^{\text{e}}_{\text{ABC}}$: ABC posterior obtained with Euler-Maruyama scheme.

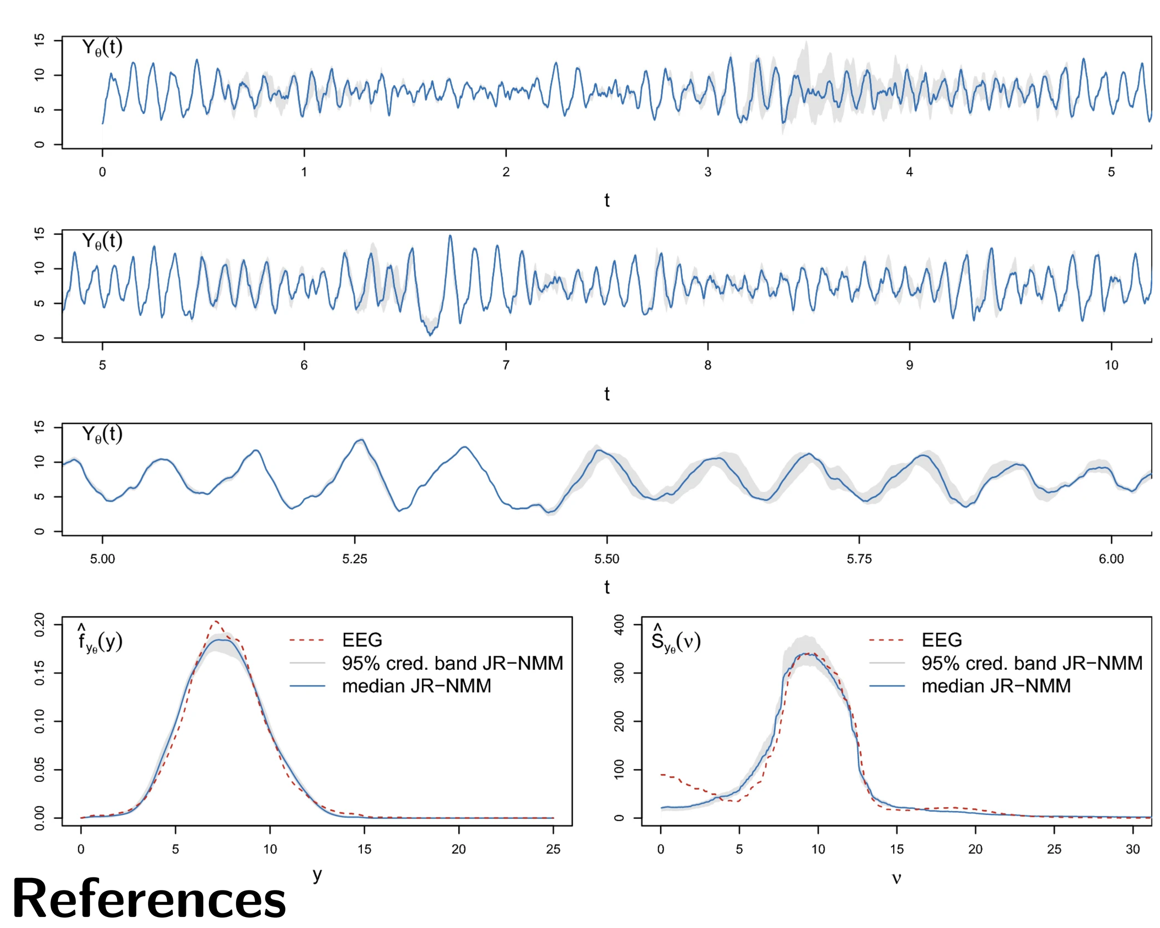
Stochastic JRNMM

Illustration on simulated data



ABC posteriors obtained with Euler-Maruyama scheme (top) vs splitting scheme (bottom).

Illustration on real EEG data



References

[1] E. Buckwar, M. Tamborrino, I. Tubikanec. Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Stat. Comput.*, 30, 627–648, 2020.

[2] S. A. Sisson, Y. Fan, M.A. Beaumont. *Handbook of Approximate Bayesian Computation*, CRC, 2018.

[3] K. Cranmer, J. Brehmer, G. Louppe. The frontier of simulation-based inference. *PNAS*, 117 (48) 30055-30062, 2020.

[4] M. Ableidinger, E. Buckwar, H. Hinterleitner. A Stochastic Version of the Jansen and Rit Neural Mass Model: Analysis and Numerics. *J. Math. Neurosc.*, 7 (8), 2017.

[5] E. Buckwar, A. Samson, M. Tamborrino, I. Tubikanec. A splitting method for SDEs with locally Lipschitz drift: Illustration on the FitzHugh-Nagumo model. *Appl. Num. Math*, 179, 191-220, 2022.

Scan Me