Faculty of Engeneering and Natural Sciences

Spectral density-based and measure-preserving ABC for SDEs

Research seminar: Institute of Applied Statistics, JKU Linz

Irene Tubikanec
Institute for Stochastics, JKU Linz

Joint work with
Massimiliano Tamborrino (University of Warwick)
Evelyn Buckwar (JKU Linz)

Reference

Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs.

- Buckwar, E., Tamborrino, M. & Tubikanec, I.
- Statistics and Computing 30, 627-648 (2020).
- Available open access: `https://doi.org/10.1007/s11222-019-09909-6`

Setting of interest

Stochastic differential equations (SDEs)

1. We consider the n-dim SDE with parameter vector $\theta = (\theta_1, \ldots, \theta_k)$

$$dX(t) = f(t, X(t); \theta)\ dt + \mathscr{G}(t, X(t); \theta)\ dW(t), \quad t \geq 0$$
$$X(0) = X_0.$$

Stochastic solution process: $\mathbf{X} = (X(t))_{t \geq 0} \in \mathbb{R}^n$

Partially observed SDEs

1. We consider the n-dim SDE with parameter vector $\theta = (\theta_1, \dots, \theta_k)$

$$dX(t) = f(t, X(t); \theta)\ dt + \mathscr{G}(t, X(t); \theta)\ dW(t), \quad t \geq 0$$
$$X(0) = X_0.$$

Stochastic solution process: $\mathbf{X} = (X(t))_{t \geq 0} \in \mathbb{R}^n$

2. The n-dimensional solution process $\mathbf{X}$ is partially observed through the one-dimensional output process

$$\mathbf{Y}_\theta = (Y_\theta(t))_{t \geq 0} = g(\mathbf{X}), \quad g : \mathbb{R}^n \to \mathbb{R}.$$

Partially observed SDEs with an invariant distribution

1. We consider the n-dim SDE with parameter vector $\theta = (\theta_1, ..., \theta_k)$

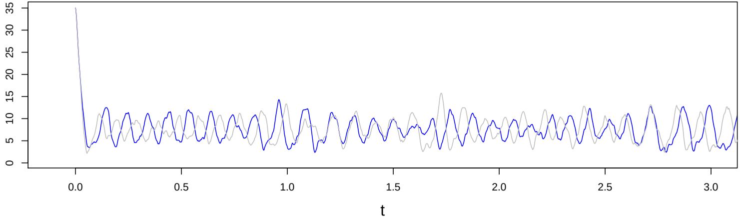
$$dX(t) = f(t, X(t); \theta)\ dt + \mathscr{G}(t, X(t); \theta)\ dW(t), \quad t \geq 0$$
$$X(0) = X_0.$$

Stochastic solution process: $\mathbf{X} = (X(t))_{t \geq 0} \in \mathbb{R}^n$

2. The n-dimensional solution process $\mathbf{X}$ is partially observed through the one-dimensional output process

$$\mathbf{Y}_\theta = (Y_\theta(t))_{t \geq 0} = g(\mathbf{X}), \quad g : \mathbb{R}^n \to \mathbb{R}.$$

3. The output process $\mathbf{Y}_\theta$ admits an invariant distribution $\eta_{\mathbf{Y}_\theta}$.



Parameter inference for partially observed SDEs with an invariant distribution

1. We consider the n-dim SDE with parameter vector $\theta = (\theta_1, ..., \theta_k)$

$$dX(t) = f(t, X(t); \theta)\, dt + \mathscr{G}(t, X(t); \theta)\, dW(t), \quad t \geq 0$$
$$X(0) = X_0.$$

   Stochastic solution process: $\mathbf{X} = (X(t))_{t\geq 0} \in \mathbb{R}^n$

2. The n-dimensional solution process $\mathbf{X}$ is partially observed through the one-dimensional output process

$$\mathbf{Y}_\theta = (Y_\theta(t))_{t\geq 0} = g(\mathbf{X}), \quad g : \mathbb{R}^n \to \mathbb{R}.$$

3. The output process $\mathbf{Y}_\theta$ admits an invariant distribution $\eta_{\mathbf{Y}_\theta}$.

4. **Our goal:** Inference of θ (via ABC) based on observations of the output process $\mathbf{Y}_\theta$ and using $\eta_{\mathbf{Y}_\theta}$.

Motivating example

Stochastic Jansen and Rit Neural Mass Model (JR-NMM)[1]

Model: $n = 6$-dimensional stochastic JR-NMM

$$d \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} P(t) \\ -\Gamma^2 Q(t) - 2\Gamma P(t) + G(Q(t);\theta) \end{pmatrix} dt \; + \; \begin{pmatrix} \mathbb{O}_3 \\ \Sigma_\theta \end{pmatrix} dW(t),$$

with parameters $\theta = (\sigma, \mu, C)$ and non-linear $G\colon \mathbb{R}^3 \to \mathbb{R}^3$

Solution process: $\mathbf{X} = (\mathbf{Q}, \mathbf{P})^T$ with (unobserved) components $\mathbf{Q} = (\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3})$ and $\mathbf{P} = (\mathbf{X_4}, \mathbf{X_5}, \mathbf{X_6})$

Output process: The process $\mathbf{X} = (\mathbf{Q}, \mathbf{P})^T$ is observed through

$$\mathbf{Y}_\theta = \mathbf{X_2} - \mathbf{X_3} \quad \text{(EEG)}$$

Property: The process $\mathbf{Y}_\theta$ admits an invariant distribution $\eta_{\mathbf{Y}_\theta}$

[1] M. Ableidinger, E. Buckwar, and H. Hinterleitner.
"A Stochastic Version of the Jansen and Rit Neural Mass Model: Analysis and Numerics."
Journal of Mathematical Neuroscience 7(8) (2017)

EEG data[2]

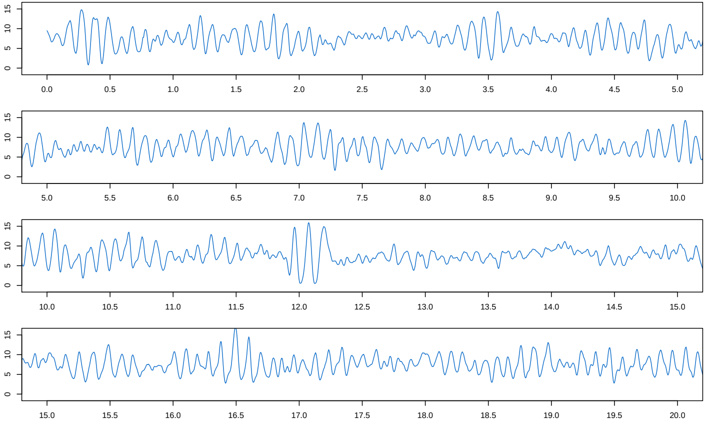


Figure: $T = 20$ seconds of an α-rhythmic EEG segment recorded with a sampling rate of 173.61 Hz.

[2]Data available at:
http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3

ABC Algorithm

Notation

- Observed reference data: $y = (y(t_i))$
- Simulated synthetic data: $y_\theta = (y_\theta(t_i))$
- Prior: $\pi(\theta)$
- Posterior: $\pi(\theta|y)$
- ABC posterior: $\pi(\theta|y) \approx \pi_{\text{ABC}}(\theta|y)$

Algorithm

Reference table acceptance-rejection ABC

Input: Observed data y
Output: Samples from the posterior $\pi_{\mathrm{ABC}}(\theta|y)$

1: Precompute a vector of summary statistics $s(y)$
2: Choose a prior distribution $\pi(\theta)$ and a percentile p
3: **for** $i = 1$ to N **do**
4: Draw $\theta^i = (\theta_1^i, ..., \theta_k^i)$ from the prior $\pi(\theta)$
5: Conditionally on θ^i, simulate synthetic data y_{θ^i} from the output process $\mathbf{Y}_\theta$
6: Compute the summaries $s(y_{\theta^i})$
7: Calculate the distance $D_i = d(s(y), s(y_{\theta^i}))$
8: **end for**
9: Compute ε as the percentile p of the calculated distances
10: If $D_i < \varepsilon$, keep θ^i as a sample from the posterior, for $i = 1, \ldots, N$

Key ingredients

1. How to choose the summaries s?
2. How to simulate synthetic data y_θ?

Summaries

Challenge: Internal randomness of the model

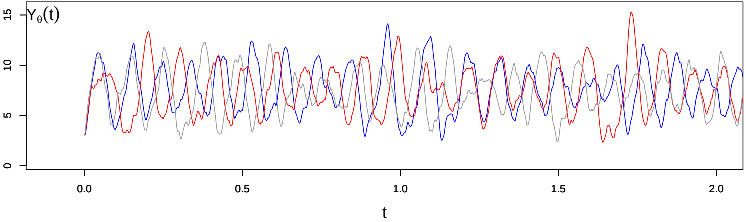


Figure: 3 realisations of the output process $\mathbf{Y}_\theta$.

Observed dataset: blue trajectory (simulated), $\theta_{\text{observed}} = 135$
Synthetic datasets: grey and red trajectories, $\theta_{\text{synthetic}} = 135/139$

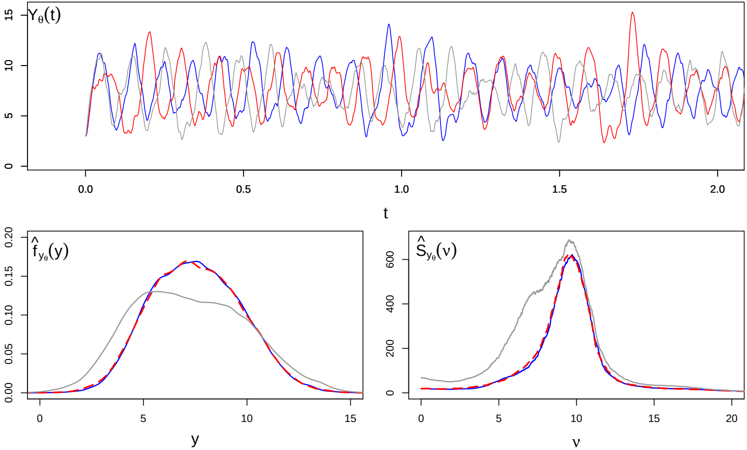
Question: Which distance is smaller, d(blue,red) or d(blue,grey)?

How to choose the summaries?

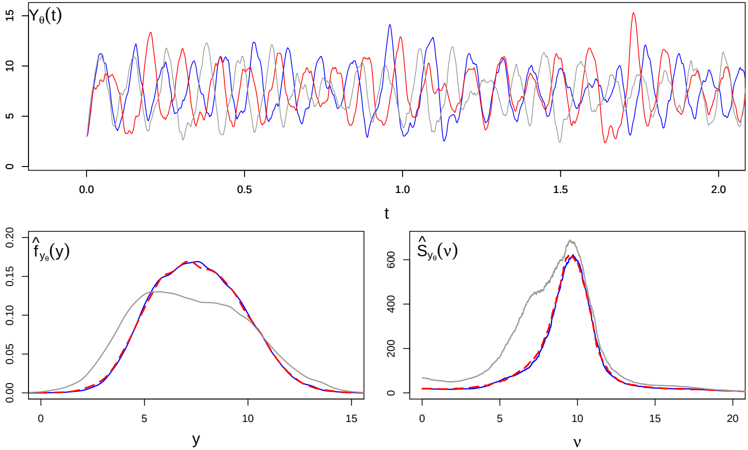
Proposal 1: Use the property of an invariant distribution $\eta_{\mathbf{Y}_\theta}$ and map the realisation y_θ of the output process $\mathbf{Y}_\theta$ to its

1) Invariant density $f_{\mathbf{Y}_\theta}$ (kernel estimator $\hat{f}_{y_\theta}$)
2) Invariant spectral density $S_{\mathbf{Y}_\theta}$ (periodogram estimator $\hat{S}_{y_\theta}$)

Summaries: Invariant density and spectral density

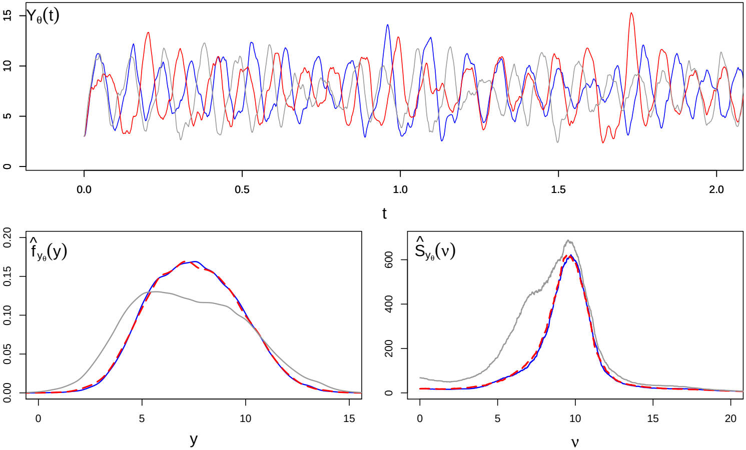


Summaries: Invariant density and spectral density



Question: Which distance is smaller, d(blue,red) or d(blue,grey)?

Summaries: Invariant density and spectral density



Question: Which distance is smaller, d(blue,red) or d(blue,grey)?

Parameter values: $\theta_{\text{observed}} = 135$, $\theta_{\text{synthetic}} = 135$, $\theta_{\text{synthetic}} = 139$

ABC distance

Data: Observed dataset y and synthetic dataset y_θ

Summaries: Invariant densities and spectral densities

$$s(y) := (\hat{S}_y, \hat{f}_y), \quad s(y_\theta) := (\hat{S}_{y_\theta}, \hat{f}_{y_\theta})$$

Distance: Weighted sum of the areas between the densities

$$D = d(s(y), s(y_\theta)) := \mathrm{IAE}(\hat{S}_y, \hat{S}_{y_\theta}) + w \cdot \mathrm{IAE}(\hat{f}_y, \hat{f}_{y_\theta})$$

Integrated absolute error:

$$\mathrm{IAE}(g_1, g_2) := \int_{\mathbb{R}} \left| g_1(x) - g_2(x) \right| \, dx$$

Spectral density-based ABC

Reference table acceptance-rejection ABC

Input: Observed data y resulting from M datasets $y_1, \ldots, y_M$
Output: Samples from the posterior $\pi_{\text{ABC}}(\theta|y)$

1. Precompute the summaries $s(y_j) = (\hat{S}_{y_j}, \hat{f}_{y_j})$, $j = 1, \ldots, M$
2. Choose a prior distribution $\pi(\theta)$ and a percentile p
3. **for** $i = 1$ to N **do**
4. Draw $\theta^i = (\theta_1^i, ..., \theta_k^i)$ from the prior $\pi(\theta)$
5. Conditionally on θ^i, simulate synthetic data y_{θ^i} from the output process $\mathbf{Y}_\theta$
6. Compute $s(y_{\theta^i}) = (\hat{S}_{y_{\theta^i}}, \hat{f}_{y_{\theta^i}})$
7. $D_i = \text{median}\left\{\text{IAE}(\hat{S}_{y_j}, \hat{S}_{y_{\theta^i}}) + w \cdot \text{IAE}(\hat{f}_{y_j}, \hat{f}_{y_{\theta^i}})\right\}_{j=1}^{M}$
8. **end for**
9. Compute ε as the percentile p of the calculated distances
10. If $D_i < \varepsilon$, keep θ^i as a sample from the posterior, for $i = 1, \ldots, N$

JOHANNES KEPLER UNIVERSITY LINZ

Simulation from the model

Numerical simulation methods for SDEs

Time discretisation:

- Time interval: $[0, T]$
- Discrete points: $t_i,\ i = 0, \ldots, n,\ t_0 = 0,\ t_n = T$
- Time step: $\Delta = t_i - t_{i-1}$

1. **Exact simulation of the process at** t_i: $y_\theta = (Y_\theta(t_i))$

$$\pi(\theta|y) \approx \pi_{\text{ABC}}(\theta|y)$$

2. **Approximation of the process at** t_i: $\tilde{y}_\theta = (\widetilde{Y}_\theta(t_i)) \approx (Y_\theta(t_i))$

$$\pi(\theta|y) \approx \pi_{\text{ABC}}(\theta|y) \approx \pi_{\text{ABC}}^{\text{num}}(\theta|y)$$

Numerical simulation methods for SDEs

Time discretisation:

- Time interval: $[0, T]$
- Discrete points: $t_i,\ i = 0, \ldots, n,\ t_0 = 0,\ t_n = T$
- Time step: $\Delta = t_i - t_{i-1}$

1. **Exact simulation of the process at** t_i: $y_\theta = (Y_\theta(t_i))$

$$\pi(\theta|y) \approx \pi_{\text{ABC}}(\theta|y)$$

2. **Approximation of the process at** t_i: $\tilde{y}_\theta = (\widetilde{Y}_\theta(t_i)) \approx (Y_\theta(t_i))$

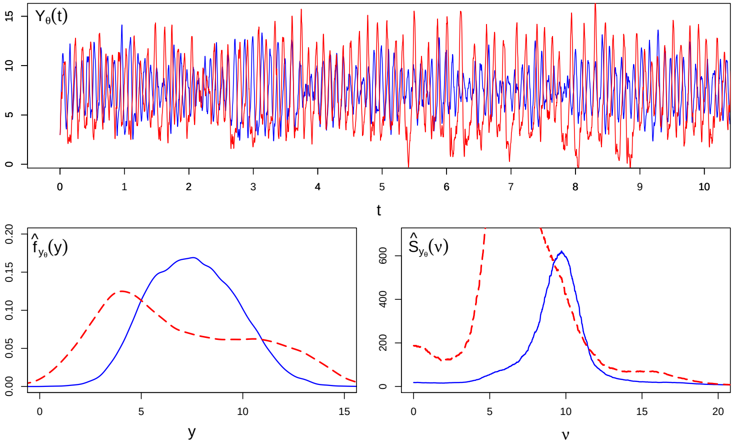
$$\pi(\theta|y) \approx \pi_{\text{ABC}}(\theta|y) \approx \pi_{\text{ABC}}^{\text{num}}(\theta|y)$$

2.1 Measure-preserving method: $Y_\theta(t_i) \approx \widetilde{Y}_\theta(t_i) \sim \eta_{\mathbf{Y}_\theta}$

2.2 Non-preserving method: $Y_\theta(t_i) \approx \widetilde{Y}_\theta(t_i) \nsim \eta_{\mathbf{Y}_\theta}$

Challenge: Standard methods (Euler-Maruyama) may be non-preserving

$$\widetilde{X}(t_{i+1}) = \widetilde{X}(t_i) + f(t_i, \widetilde{X}(t_i); \theta)\Delta + \mathscr{G}(t_i, \widetilde{X}(t_i); \theta)\Delta W$$



Toy Model

Model: $n = 2$-dimensional damped stochastic harmonic oscillator

$$d\begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} P(t) \\ -\lambda^2 Q(t) - 2\gamma P(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} dW(t),$$

with $\theta = (\lambda, \gamma, \sigma)$ and $\lambda^2 - \gamma^2 > 0$ (weak damping)

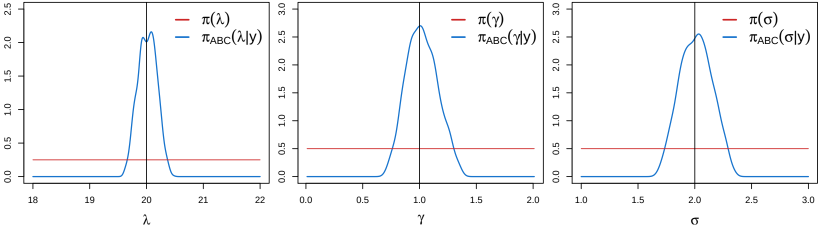
Output process: The process $\mathbf{X} = (\mathbf{Q}, \mathbf{P})^T$ is observed through $\mathbf{Y}_\theta = \mathbf{Q}$

Property: The output process admits an invariant distribution $\eta_{\mathbf{Y}_\theta}$

Simulation: Exact

Spectral density-based ABC: Toy Model

ABC Results: $\theta = (\lambda, \gamma, \sigma)$,
Exact simulation, Time step $\Delta = 10^{-2}$

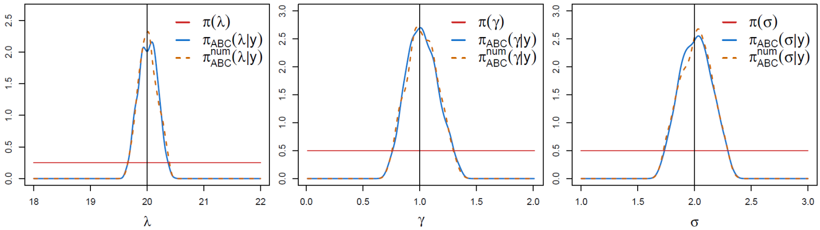


ABC Setup:

- Uniform priors: $\lambda \sim U(18, 22)$, $\gamma \sim U(0.01, 2.01)$, $\sigma \sim U(1, 3)$
- Observed data: $M = 10$ paths, using $\Delta = 10^{-2}$ and $T = 10^3$
- Synthetic data: $N = 2 \cdot 10^6$ paths, using the same Δ and T
- Threshold level: $\epsilon = 0.05^{th}$ percentile

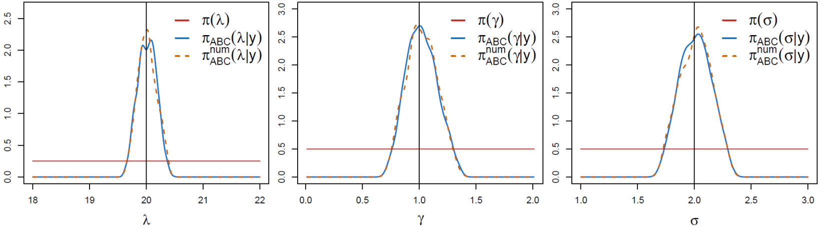
Spectral density-based and measure-preserving ABC: Toy Model

ABC Results: $\theta = (\lambda, \gamma, \sigma)$,
Measure-preserving simulation, Time step $\Delta = 10^{-2}$



Can we use Euler-Maruyama?

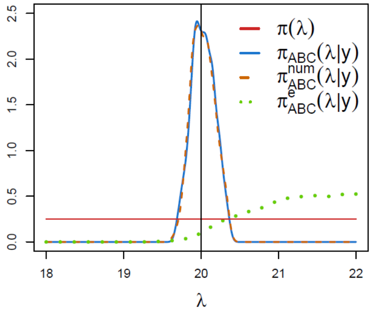
ABC Results: $\theta = (\lambda, \gamma, \sigma)$,
Measure-preserving simulation, Time step $\Delta = 10^{-2}$



Euler-Maruyama is NOT APPLICABLE for $\Delta = 10^{-2}$
$\widetilde{Y}_\theta(t_i) \approx \infty$ (Computer overflow)

Simplest task: Inferring only one parameter

ABC Results: $\theta = \lambda$,
Different numerical methods, Smaller time step $\Delta = 10^{-3}$

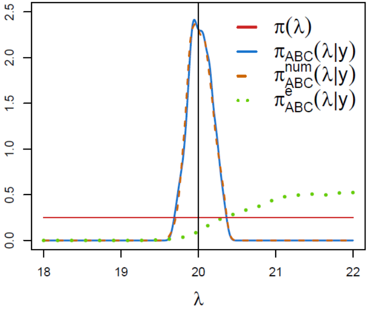


ABC Setup:

- Uniform priors: $\lambda \sim U(10, 30)$
- Observed data: Same as before
- Synthetic data: $N = 10^5$ paths, using a smaller $\Delta = 10^{-3}$
- Threshold level: $\varepsilon = 1^{st}$ percentile

Simplest task: Inferring only one parameter

ABC Results: $\theta = \lambda$,
Different numerical methods, Smaller time step $\Delta = 10^{-3}$



Even smaller Δ required for Euler-Maruyama
$\Longrightarrow$ Highly inefficient
$\Longrightarrow$ ABC: computationally infeasible

How to simulate from the model?

Proposal 2: Use a measure-preserving numerical method.
$\Longrightarrow$ Splitting method

Measure-preserving splitting for the stochastic JR-NMM[1]

Model:

$$d\begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} P(t) \\ -\Gamma^2 Q(t) - 2\Gamma P(t) + \underbrace{G(Q(t);\theta)}_{\text{nonlinear}} \end{pmatrix} dt + \begin{pmatrix} \mathbb{O}_3 \\ \Sigma_\theta \end{pmatrix} dW(t)$$

Splitting:

1. **Equation 1**: linear SDE

$$d\begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} P(t) \\ -\Gamma^2 Q(t) - 2\Gamma P(t) \end{pmatrix} dt + \begin{pmatrix} \mathbb{O}_3 \\ \Sigma_\theta \end{pmatrix} dW(t)$$

2. **Equation 2**: non-linear (but simple) ODE

$$d\begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} 0_3 \\ G(Q(t);\theta) \end{pmatrix} dt$$

[1] M. Ableidinger, E. Buckwar, and H. Hinterleitner.
"A Stochastic Version of the Jansen and Rit Neural Mass Model: Analysis and Numerics."
In: Journal of Mathematical Neuroscience 7(8) (2017)

Measure-preserving splitting for the stochastic JR-NMM

1. **Equation 1**: The linear SDE can be written as

$$dX(t) = AX(t)dt + BdW(t)$$

Explicit solution: Exact paths are obtained through

$$X(t_{i+1}) = e^{A\Delta}X(t_i) + \xi_i,$$

where ξ_i are 6-dimensional Gaussian vectors with mean 0_6 and variance $C(\Delta)$, where $\dot{C}(t) = AC(t) + C(t)A^T + BB^T$, $C(0) = \mathbb{O}_6$.

2. **Equation 2**: non-linear (but simple) ODE

$$d\begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} 0_3 \\ G(Q(t);\theta) \end{pmatrix} dt$$

Explicit solution: Exact paths are obtained through

$$X(t_{i+1}) = X(t_i) + \begin{pmatrix} 0_3 \\ \Delta G(Q(t_i);\theta) \end{pmatrix}.$$

JOHANNES KEPLER UNIVERSITY LINZ

Measure-preserving splitting for the stochastic JR-NMM

Splitting:

1. **Explicit solution of Equation 1:**

$$X(t_{i+1}) = e^{A\Delta} X(t_i) + \xi_i$$

2. **Explicit solution of Equation 2:**

$$X(t_{i+1}) = X(t_i) + \begin{pmatrix} 0_3 \\ \Delta G(Q(t_i); \theta) \end{pmatrix}$$

Composition (Strang approach):
Given $\widetilde{X}(t_i)$, how to obtain $\widetilde{X}(t_{i+1})$?

1: $X_b = \widetilde{X}(t_i) + \begin{pmatrix} 0_3 \\ \frac{\Delta}{2} G(Q(t_i); \theta) \end{pmatrix}$

2: $X_a = e^{A\Delta} X_b + \xi_i$

3: $\widetilde{X}(t_{i+1}) = X_a + \begin{pmatrix} 0_3 \\ \frac{\Delta}{2} G(Q_a; \theta) \end{pmatrix}$

Application: Spectral density-based and measure-preserving ABC

Spectral density-based and measure-preserving ABC

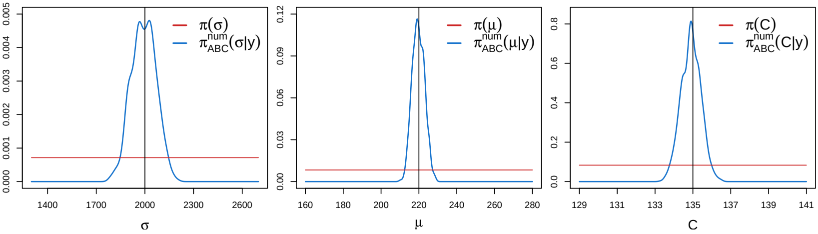
Reference table acceptance-rejection ABC

Input: Observed data y resulting from M datasets $y_1, \dots, y_M$
Output: Samples from the posterior $\pi_{\text{ABC}}^{\text{num}}(\theta|y)$

1: Precompute the summaries $s(y_j) = (\hat{S}_{y_j}, \hat{f}_{y_j})$, $j = 1, \dots, M$
2: Choose a prior distribution $\pi(\theta)$ and a percentile p
3: **for** $i = 1$ to N **do**
4: Draw $\theta^i = (\theta_1^i, \dots, \theta_k^i)$ from the prior $\pi(\theta)$
5: Conditionally on θ^i, simulate synthetic data $\tilde{y}_{\theta^i}$ using a measure-preserving numerical method (Splitting)
6: Compute $s(\tilde{y}_{\theta^i}) = (\hat{S}_{\tilde{y}_{\theta^i}}, \hat{f}_{\tilde{y}_{\theta^i}})$
7: $D_i = \text{median}\left\{\text{IAE}(\hat{S}_{y_j}, \hat{S}_{\tilde{y}_{\theta^i}}) + w \cdot \text{IAE}(\hat{f}_{y_j}, \hat{f}_{\tilde{y}_{\theta^i}})\right\}_{j=1}^{M}$
8: **end for**
9: Compute ε as the percentile p of the calculated distances
10: If $D_i < \varepsilon$, keep θ^i as a sample from the posterior, for $i = 1, \dots, N$

Parameter inference of the JR-NMM via the proposed ABC

ABC results: $\theta = (\sigma, \mu, C)$,
Using the measure-preserving splitting method

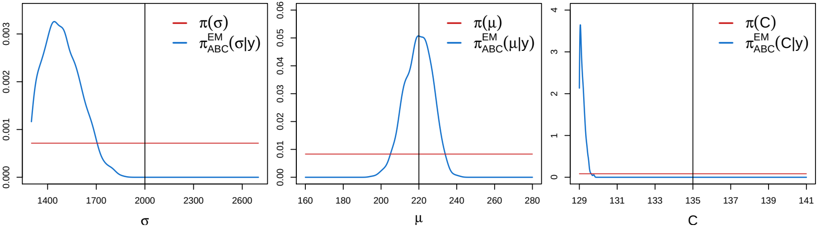


ABC Setup:

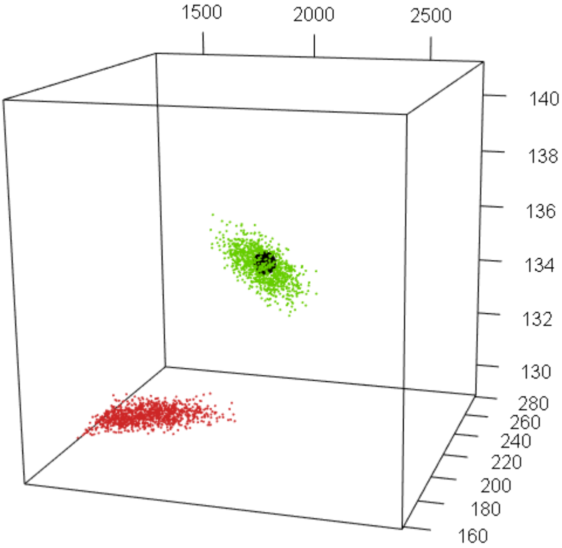
- Priors: $\sigma \sim U(1300, 2700)$, $\mu \sim U(160, 280)$, $\sigma \sim U(129, 141)$
- Observed data: $M = 30$ paths, using $\Delta = 2 \cdot 10^{-3}$, $T = 200$
- Synthetic data: $N = 2.5 \cdot 10^6$ paths, using the same Δ and T
- Threshold level: $\varepsilon = 0.05^{th}$ percentile

Parameter inference based on the non-preserving Euler-Maruyama method

ABC results: $\theta = (\sigma, \mu, C)$,
Using the non-preserving Euler-Maruyama method



ABC results: A comparison of splitting and Euler-Maruyama



Parameter inference from real EEG data

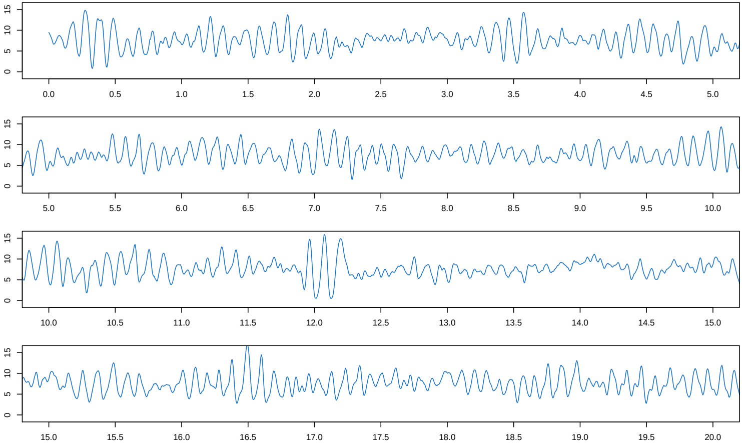
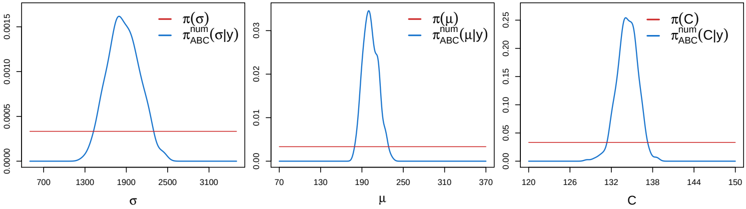


Figure: $T = 20$ seconds of an α-rhythmic EEG segment recorded with a sampling rate of 173.61 Hz.

Parameter inference from real EEG data

ABC Results: $\theta = (\sigma, \mu, C)$,
Using the measure-preserving splitting method



ABC Setup:

- Priors: $\sigma \sim U(500, 3500)$, $\mu \sim U(70, 370)$, $\sigma \sim U(120, 150)$
- Observed data: $M = 3$ α-rhythmic EEG recordings, sampled with $\Delta = 173.61^{-1} \approx 5.76 \cdot 10^{-3}$ and $T = 23.6$ seconds
- Synthetic data: $N = 5 \cdot 10^6$ paths, using $\Delta = 2 \cdot 10^{-3}$ and same T
- Threshold level: $\varepsilon = 0.02^{nd}$ percentile

Conclusions

1. The proposed ABC approach yields successful inference when combining:
   - invariant measure-based summaries (density and spectral density)
   - efficient and measure-preserving numerical methods (splitting)
2. The inference returned using standard non-preserving numerical methods (Euler-Maruyama) fails. Its performance may improve for „small enough“ time steps $\Longrightarrow$ Computationally infeasible.
3. Successful results under the basic acceptance-rejection ABC. $\Longrightarrow$ The proposed techniques can be applied to more advanced algorithms.

Thank you for your interest