SUPPLEMENTARY SECTION FOR “BAYESIAN NONPARAMETRIC LEARNING OF MILKY WAY MODEL PARAMETERS USING A NEW GAUSSIAN PROCESS-BASED METHOD”

BY DALIA CHAKRABARTY*†‡§, MUNMUN BISWAS†**, SOURABH BHATTACHARYA¶§††, 
University of Warwick¶ and University of Leicester∥, Indian statistical Institute***, Indian Statistical Institute†††

Throughout, we refer to our main manuscript as CBB.

1. Details of dynamical simulations of astrophysical models. In Chakrabarty (2007), the modelling involves the following. A sample of phase space coordinates \( \{w_0\} \), is drawn from a chosen (to mimic real disc galaxies’) phase space density \( g(w) \) at \( t = 0 \), and is evolved in a (chosen) parametric Galactic gravitational potential \( \Psi : \mathbb{R}^3 \times T \to \mathbb{R} \), where the time variable \( T \in T \subset \mathbb{R}_{\geq 0} \). \( \Psi(X, t) \) is chosen to emulate a realistic background Galactic potential \( \Psi_0(x) \in \mathbb{R} \), perturbed by the chosen parametric forms (rigidly rotating, quadrupolar) of the gravitational potential of the bar \( \Psi_b(x) \in \mathbb{R} \) and the (logarithmic spiral) gravitational potential of the spiral pattern \( \Psi_s(x) \in \mathbb{R} \) in the Milky Way. The perturbation strengths of these features \( (\varepsilon_b : T \to \mathbb{R} \ and \ \varepsilon_s : T \to \mathbb{R}) \) are slowly increased, to saturation values over time \( T_s \in \mathbb{R} \), \( T_s > 0 \) which is chosen such that \( T_s \ll T_{\text{sim}} \), where \( T_{\text{sim}} \in \mathbb{R} \) is the total simulation time. Thus, in the \( q^{th} \) astrophysical model, 
\[
\Psi^{(q)}(X, t) = \Psi^{(q)}_0(X) + \varepsilon_b(t)\Psi^{(q)}_b(X, t) + \varepsilon_s(t)\Psi^{(q)}_s(X, t),
\]
for each \( q = 1, \ldots, 4 \). At \( T = T_{\text{sim}} \), orbits are sampled in phase and recorded stroboscopically in the rotating frame of the bar at times when \( (\Omega - \Omega_s)t = 0 \).

In order to sort the orbits by the value of the solar location in a galactocentric coordinate frame, chosen ranges of \( R \in [1.7, 2.3] \) in simulation units and \( \Phi \in [0, 90] \) in degrees, are used to define a regular, 2-D polar grid with \( N_R \times N_\phi \) cells of radial and azimuthal widths of \( \delta_R \) and \( \delta_\phi \), \( \delta_R, \delta_\phi \in \mathbb{R} \). The \( i^{th} \) cell in this grid represents the \( i^{th} \) value of the model parameter vector, \( S(i) \), with \( n = N_R \times N_\phi \), \( i = 1, \ldots, N_R \times N_\phi \). In order to find stars that end up in the simulations with spatial coordinates within the interval around \( s(i) \) that defines the \( i^{th} \)-cell, we identify stars with radius \( \in [s_1(i), s_1(i) + \delta_1) \) and azimuth \( \in [s_2(i), s_2(i) + \delta_2) \); these stars comprise the simulated velocity information \( V_i \) in the \( i^{th} \)-cell. This is performed for each \( i, i = 1, \ldots, n \). The same 2-D polar grid is used for each of the Milky Way astrophysical models.

2. Posteriors of \( B \) and \( C \) given training data. Here, we discuss the achievement of the posteriors of the process matrices \( B \) and \( C \), given the training data. To write the posteriors, we use uniform prior on \( B \) and a simple non-informative prior on \( C \), namely, \( \pi(C) \propto \|C\|^{-(j+1)/2} \).

Recalling that the distribution of the training data \( D_s \) is matrix normal, we write the posterior of \( B \) as

\[
[B \mid \Sigma, C, Q, D_s] \sim MN_{m,jk}(\hat{B}_{GLS}, (H_D^T A_D^{-1} H_D^{-1}, \Omega)),
\]

\[2.1\]
where \( \hat{\mathbf{B}}_{GLS} := (\mathbf{H}_D^T \mathbf{A}_D^{-1} \mathbf{H}_D)^{-1}(\mathbf{H}_D^T \mathbf{A}_D^{-1} \mathbf{D}_s) \); note that \( \mathbf{H}_D \hat{\mathbf{B}}_{GLS} \) can be interpreted as the generalised least square (GLS) solution to the equation that given the chosen design vectors, the mean matrix \( \sum \) parameters in batches as follows:

\[
\text{Given } \hat{\mathbf{C}}_{GLS} := \sum_{i=1}^k \sum_{t=1}^k \sigma_{it}^{-1} \mathbf{M}_{it}^* \text{. Then it can be shown that for this given choice of the design vectors, the posterior distribution } \mathbb{C} \sim \mathcal{W} \text{ is inverse Wishart with parameters } (n - m)k \hat{\mathbf{C}}_{GLS} \text{ and } (n - m)k \text{ (degrees of freedom)}.
\]

3. Positive definiteness of \( (n-m)k \hat{\mathbf{C}}_{GLS} \). Now marginalisation of the posterior \( [\mathbf{s}(test), \mathbf{C}, \mathbf{B}, \mathbf{C}] \) are sensitive to the posterior \( [\mathbb{C}, \mathbf{Q}, \mathbf{D}_s] \). Thus, we need to ensure that \( (n-m)k \hat{\mathbf{C}}_{GLS} \) is positive definite in order for the posterior of \( \mathbb{C} \) to be well-defined. We now present the proof that \( (n-m)k \hat{\mathbf{C}}_{GLS} \) is positive definite.

**Proof.** Since \( \mathbf{D}_s^T \mathbf{M} \mathbf{D}_s = [\mathbf{M}_{it}^*; i, t = 1, \ldots, k] \) is positive definite, \( \mathbf{B}^T \mathbf{M} \mathbf{D}_s = \mathbf{P} \mathbf{P}' \), where \( \mathbf{P} \) is a lower triangular matrix with strictly positive diagonal entries. Writing \( \mathbf{P} = (\mathbf{P}_1; \mathbf{P}_2; \ldots; \mathbf{P}_k)^T \), where each \( \mathbf{P}_i \) is of order \( j \times j \), it then follows that \( \mathbf{M}_{it}^* = \mathbf{P}_i \mathbf{P}_i \). Hence, for any non-null \( j \)-component vector \( \mathbf{s} \),

\[
\sum_{i=1}^k \sum_{t=1}^k \sigma_{it}^{-1} \mathbf{s}^T \mathbf{M}_{it}^* \mathbf{s} = \sum_{i=1}^k \sum_{t=1}^k \sigma_{it}^{-1} (\mathbf{P}_i \mathbf{s})^T (\mathbf{P}_i \mathbf{s}).
\]

Here \( \sigma_{it} \) is the \( it \)-th element of the matrix \( \mathbf{\sigma}_{\mathbf{D}}(\cdot) := [a(\cdot, s_1), \ldots, a(\cdot, s_n)]^T \). The right hand side of the last equation is \( \mathbf{y}^T (\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_j) \mathbf{y} \), where \( \mathbf{y} := (\mathbf{y}_1^T, \ldots, \mathbf{y}_k^T)^T \), with \( \mathbf{y}_t = \mathbf{P}_i \mathbf{s} \), and \( \mathbf{I}_j \) is the \( j \)-th order identity matrix. Since Kronecker product of positive definite matrices yields positive definite matrices, it follows that the last expression is positive, implying that \( (n-m)k \hat{\mathbf{C}}_{GLS} \) is positive definite.

4. Batchwise TMCMC based on additive transformations. Using TMCMC we can update the parameters in batches as follows:

(i) Initialise the unknown quantities by fixing arbitrarily initial values \( \left( \mathbf{s}^{(test,0)}, \mathbf{Q}^{(0)}, \mathbf{\Sigma}^{(0)} \right) \). In our case, \( \mathbf{s}^{(test,0)} = (s_1^{(test,0)}, \ldots, s_d^{(test,0)}) \), \( \mathbf{Q}^{(0)} \) is characterised by the initial values of the \( d \) smoothness parameters, which we denote by \( (b_1^{(0)}, \ldots, b_d^{(0)}) \) and \( \mathbf{\Sigma}^{(0)} \) denotes the initial choice of the \( k \times k \) matrix \( \mathbf{\Sigma} \).

(ii) Assume that at iteration \( t \), the state of the unknown parameters is \( (\mathbf{s}^{(test,t)}, \mathbf{Q}^{(t)}, \mathbf{\Sigma}^{(t)}) \).

(iii) (a) Propose \( \epsilon \sim g(\cdot) \mathbf{I}_{\epsilon > 0} \), where \( g(\cdot) \) is some arbitrary distribution, and \( \mathbf{I} \) denotes the indicator function. In our applications, we shall choose \( g(\cdot) = N(0, 1) \), so that, \( \epsilon > 0 \) is drawn from a truncated normal distribution. Update the \( d \) components of location \( \mathbf{s}^{(test)} \) by setting, with probabilities \( \pi_j \) and \( (1 - \pi_j) \), \( s_j^{(test,t+1)} = s_j^{(test,t)} + c_j \epsilon \) (forward transformation) and \( s_j^{(test,t+1)} = s_j^{(test,t)} - c_j \epsilon \) (backward transformation), respectively, where, for \( j = 1, \ldots, d \),
\( \pi_j \) are appropriately chosen probabilities and \( c_j \) are appropriately chosen scaling factors. Assume that for \( j_1 \in \mathcal{S}, s_{j_1}^{(\text{test},t)} \) gets the positive transformation, while for \( j_2 \in \mathcal{S}^c, s_{j_2}^{(\text{test},t)} \) gets the backward transformation. Here \( \mathcal{S} \cup \mathcal{S}^c = \{1, \ldots, d\} \).

(b) Accept the new proposal \( s_{(\text{test},t+1)} \) with acceptance probability

\[
\alpha_s = \min \left\{ 1, \frac{\prod_{j_1 \in \mathcal{S}} (1 - \pi_{j_1}) \prod_{j_2 \in \mathcal{S}^c} \pi_{j_2} \prod_{j_1 \in \mathcal{S}} \pi_{j_1} \prod_{j_2 \in \mathcal{S}^c} (1 - \pi_{j_2}) \times r_s}{\prod_{j_1 \in \mathcal{S}} \pi_{j_1} \prod_{j_2 \in \mathcal{S}^c} (1 - \pi_{j_2}) \times r_s} \right\}
\]

where \( r_s \) denotes the ratio of \( |A_{D_{\text{aug}}}^{-\frac{1}{2}} H'_{D_{\text{aug}}} A_{D_{\text{aug}}}^{-1} H_{D_{\text{aug}}}^{-\frac{1}{2}}| \) evaluated at the new value \( (s_{\text{test},t+1}) \) and the current value \( (s_{\text{test},t}) \) of \( s_{\text{test}} \) respectively, with \( b \) and \( \Sigma \) remaining fixed at the current values \( b^{(t)} \) and \( \Sigma^{(t)} \), respectively. In our applications, we shall choose \( \pi_j = \frac{1}{2} \forall j \), which simplifies the acceptance ratio by eliminating the first factor altogether.

(c) If the new proposal \( s_{(\text{test},t+1)} \) is not accepted, we set \( s_{(\text{test},t+1)} = s_{(\text{test},t)} \).

(iv) Using the same TMCMC strategy, update \( b \). Note that the probability of acceptance is zero if some components of the proposed value of \( b \) is negative.

(v) Update \( \Sigma \) by first decomposing it into \( LL' \), where \( L \) is the appropriate lower-triangular matrix. Then update all the non-zero elements of \( L \) in a single block using the same TMCMC strategy. The form of the acceptance ratio remains same. Again, reject the moves that yield negative diagonal elements.

(vi) Cycle over steps (ii)–(v) for \( K_1 + K_2 \) iterations, assuming that the Markov chain converges after \( K_1 \) iterations. Store the realisations

\[ \left\{ (s^{(\text{test},t)}, Q^{(t)}, \Sigma^{(t)}): t = K_1 + 1, \ldots, K_2 \right\} \]

as samples obtained (approximately) from \( [s^{(\text{test})}, Q, \Sigma | D_{\text{aug}}] \). In particular, the realisations \( \{ s^{(\text{test},t)} : t = K_1 + 1, \ldots, K_2 \} \) are samples (approximately) from the marginal distribution \( [s^{(\text{test})} | V^{(\text{test})}, D_{\text{aug}}] \).
References.


DEPARTMENT OF STATISTICS
University of Warwick
COVENTRY CV4 7AL, U.K.
d.chakrabarty@warwick.ac.uk

AND DEPARTMENT OF MATHEMATICS
University of Leicester
LEICESTER LE1 7RH, U.K.
dc252@le.ac.uk

AND STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
KOLKATA 700108, INDIA
munmun.biswas08@gmail.com

AND BAYESIAN AND INTERDISCIPLINARY RESEARCH UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
KOLKATA 700108, INDIA
sourabh@isical.ac.in