Supplementary Material for AISTATS paper "Multiphase MCMC Sampling for Parameter Inference in Nonlinear Ordinary Differential Equations"

1 Plots

1.1 Signal Transduction Cascade

Figure 1 presents the data from the signal transduction cascade simulated over time interval 0 to 100. The systems enters equilibrium and this causes problems in the inference problem as the lack of information provided by the data leads to a flattening of the likelihood surface.

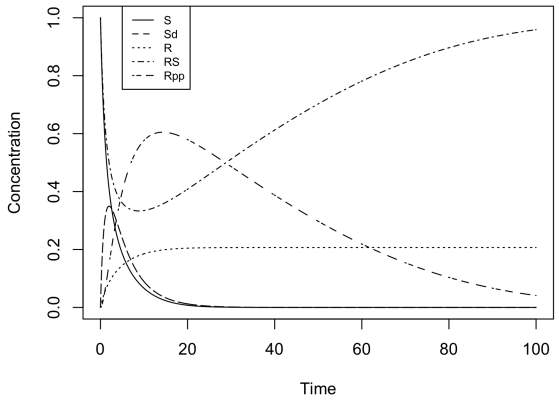

Figure 1: Data from the signal transduction cascade simulated from times 0 to 100. The system enters equilibrium and this then causes difficulties in the inference procedure as the likelihood surface flattens.

1.2 Periodicity Leads to Multimodality

Intuition would tell us that given a greater quantity of data, we would be able to better infer the parameters of ODEs since we have more data from which we gain more information for the parameter learning problem. With this in mind, we consider the Lotka-Volterra model with different amounts of data. In the first case, we will observe 16 data points and in the second, 100. Figure 2 allows consideration of a single period of data from the Lotka-Volterra model and the corresponding unimodal likelihood surface. In this case, MCMC methods would have no problem converging as there are no loca1 optima on the likelihood surface. Figure 5

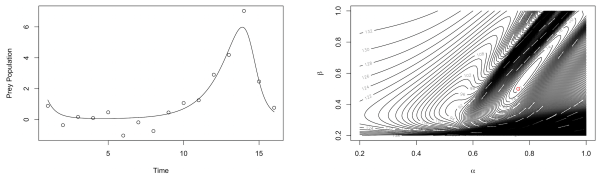

Figure 2: One period of data from the Lotka-Volterra model and the corresponding unimodal log-likelihood surface. It appears that without periodicity in the data, the learning problem becomes fairly easy.

presents the corresponding data and likelihood surface for the same model over time points that include multiple oscillations. It appears that the presence of these periods in the data introduces the problem of local optima to the inference problem. In Figure 5 we sample from the region of parameter space at which a local optimum is observed and plot the corresponding signal along with the true signal. This provides an understanding of the formation of these modes which result from a sort of signal aliasing where peaks of the estimated signal align with equally spaced peaks of the true signal. Considering the way in which the MCMC methods move through the parameter space, where we accept or reject proposal moves by assessment of their fit in function space to the observed data, it becomes apparent that our sampler will be vulnerable to these local optima. In the main paper, we observed the tendency for a DRAM sampler to be attracted into the local optima and fails to converge. Given the computational complexity involved in numerically solving the ODEs, we are not able to take a sufficient number of steps to enable the sampler to escape these local optima. Considering the Goodwin Oscillator, we observe a similar property of multimodality in Figure 3. The signal transduction cascade has non-periodic signal and so the resultant likelihood surface is smoother than in the other ODE systems. We plot this in two dimensions in Figure 4, but for this model the dimension of the parameter space is far greater than 2 dimensions.

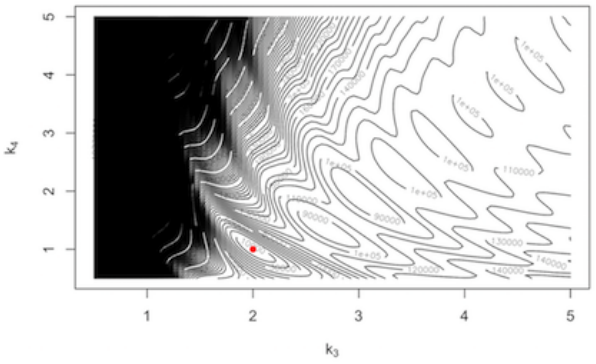

Figure 3: Negative log likelihood surface of the Goodwin Oscillator model. The red point is at the true parameter value. The different modes of the likelihood surface make this a difficult learning problem.

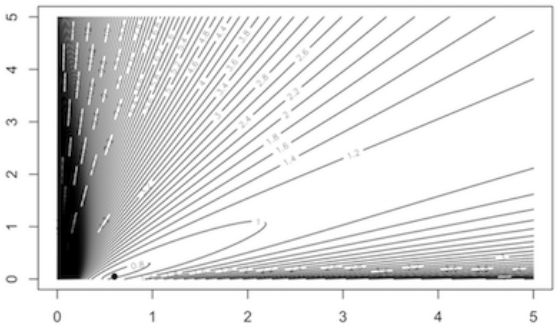

Figure 4: Signal transduction cascade likelihood surface in two dimensions. The true space is 6 dimensional and so it cannot be visualised.

1.3 Accuracy

Figure 6 considers the bias in the posterior samples obtained using each of the four alternative MCMC sampling techniques across each of the four models. The three phase proposed scheme and population MCMC perform similarly across each of the four different models with the two phase DAMH approach performing slightly worse. DRAM struggles in the first three models since these ODEs produce periodic signals. We must also consider performance of the methods in function space. This is provided in Figure 7 where we plot functional RMS for each of the posterior samples. The conclusions drawn are similar to the parameter space performance as the three phase method and population MCMC perform similarly while DRAM struggles due to convergence to local optima.

1.4 Efficiency

Figure 8 provides a more detailed exploration of the computational cost of each of the methods across the different benchmark data. Given the unsubstantial computational cost of the surrogate burnin phase, this is measured by summing the number of numerical integrations of the ODEs. We wish to compare computational cost reduction when adopting the three phase proposed scheme and so we plot the difference in number of numerical integration steps.

2 DAMH $\beta = 0$

In the original version of the paper, we presented DAMH inference results having used the parameter value $\beta = 0$. These are presented in Figures 9 and 10, where we observe the slight deterioration of performance compared with the updated results in the final paper draft.

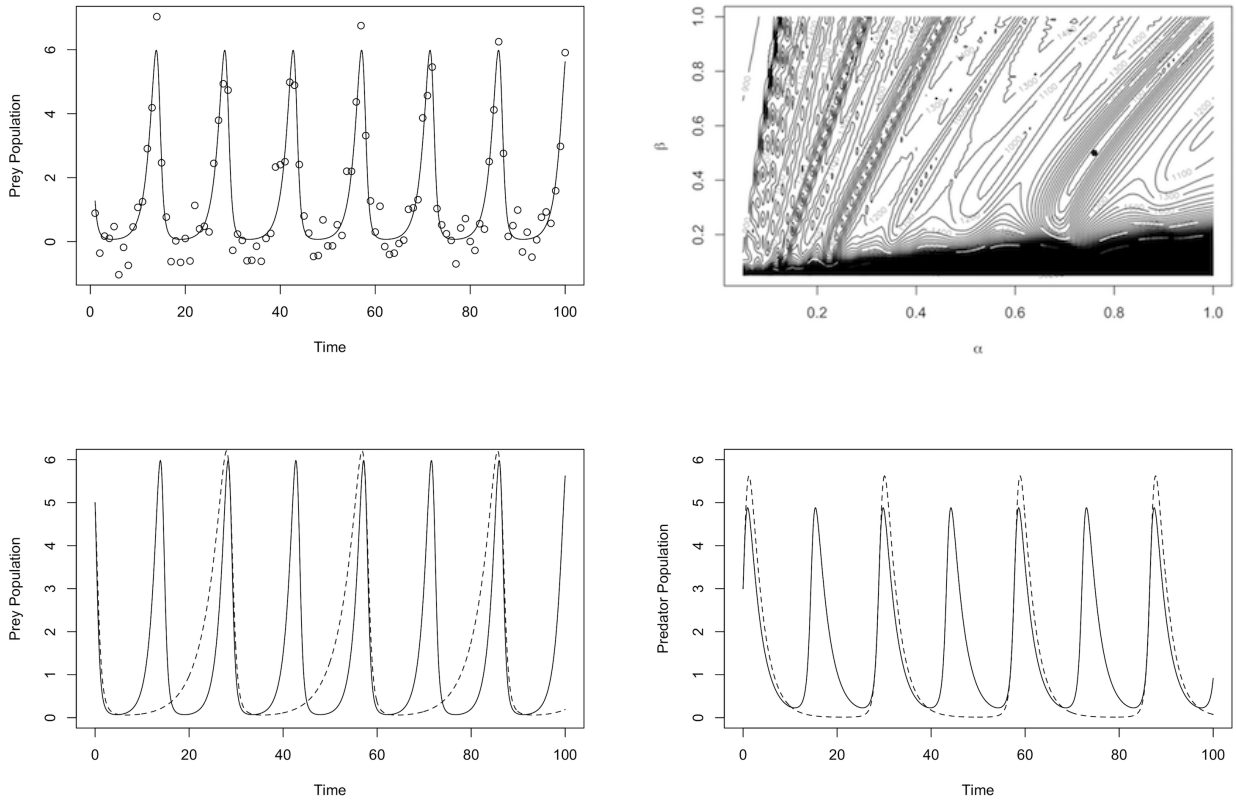

Figure 5: Top row: Multiple periods of data from the Lotka-Volterra system and the multimodal likelihood surface resulting from the data. Bottom Row: Signals obtained at a parameter sampled at a local optimum on the likelihood surface of the Lotka-Volterra system where the solid line is the true signal and the dashed line is the estimated signal.

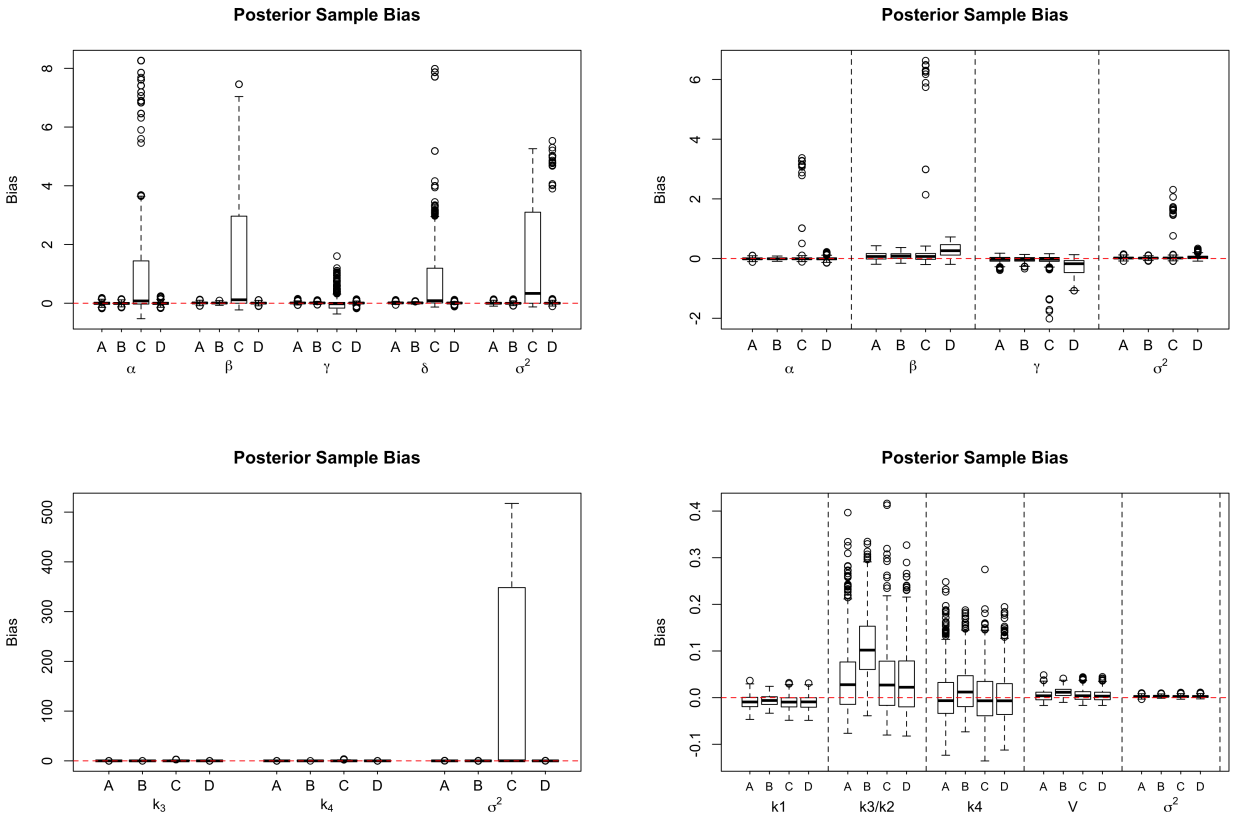

Figure 6: Bias plots for parameter posterior samples from each of the four different ODEs. The layout is as follows: topleft=Lotka-Volterra, topright=FitzHugh-Nagumo, bottomleft=Goodwin Oscillator and bottomright=signal transduction cascade. A=PropDRAM, B=PopMCMC, C=DRAM, D=DAMH.

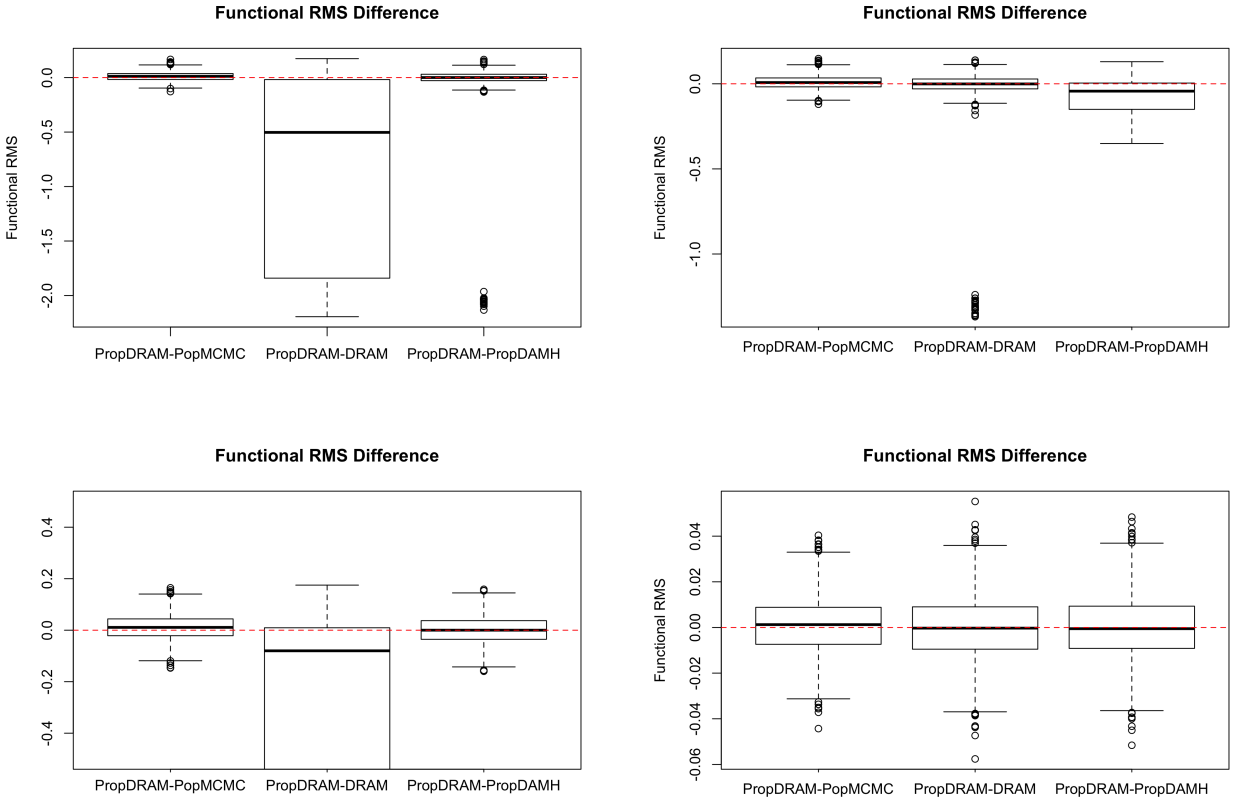

Figure 7: Difference between function space performance of the three phase proposed method and each of the other 3 benchmark algorithms. The layout is as follows: topleft=Lotka-Volterra, topright=FitzHugh-Nagumo, bottomleft=Goodwin Oscillator and bottomright=signal transduction cascade. In each case, we notice very similar performance of the three phase proposed scheme and population MCMC.

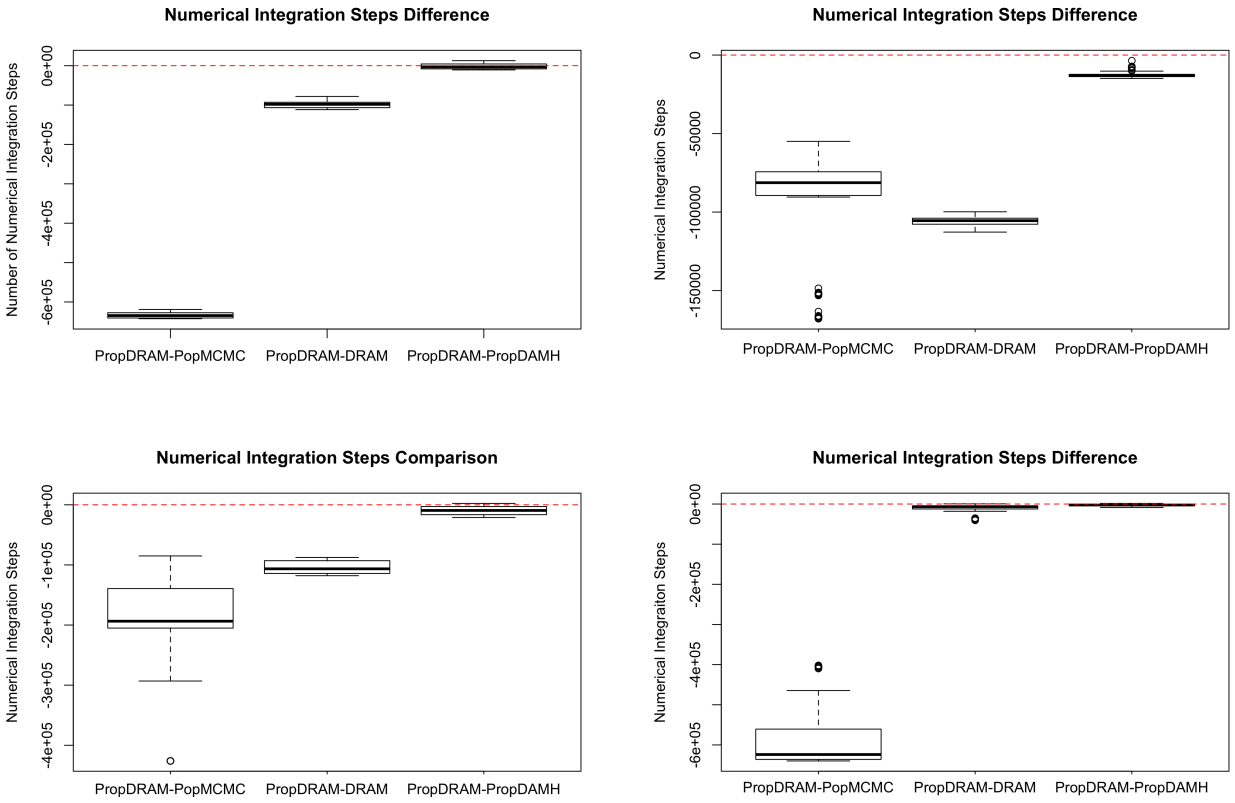

Figure 8: Number of numerical integration steps required to achieve convergence in the parameter inference in each of the different models. Values below the dashed line correspond to a lower number of numerical integrations required in the three phase proposed scheme.

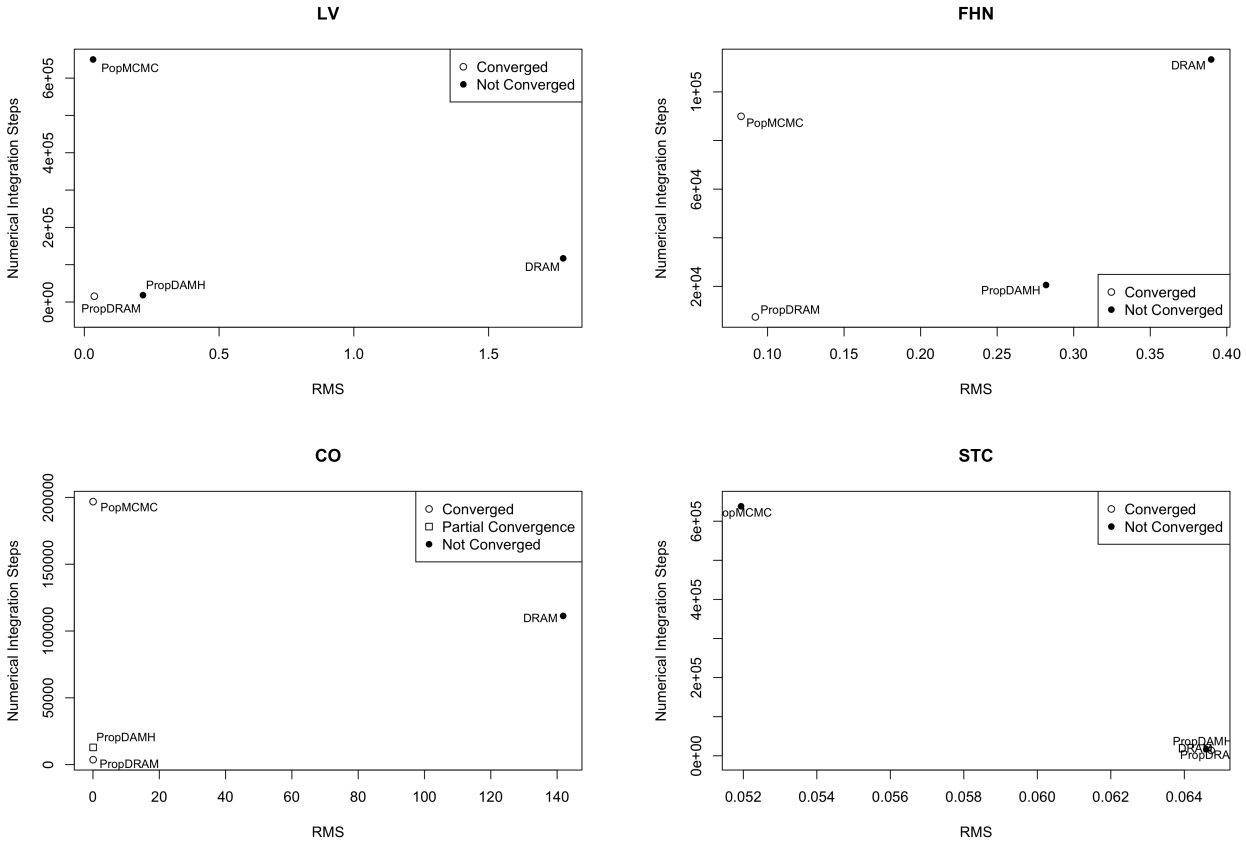

Figure 9: The figure shows the RMS scores obtained by combining the posterior distributions from 10 independent data instantiations versus number of numerical integration steps. This allows consideration of accuracy (horizontal axis) relative to computational complexity (vertical axis) for each of the four methods. Good performance is signified by a method appearing in the bottom left corner (with the exception of the STC model since RMS differences are so small). The three phase proposed scheme is the only method consistently appearing in the bottom left corner

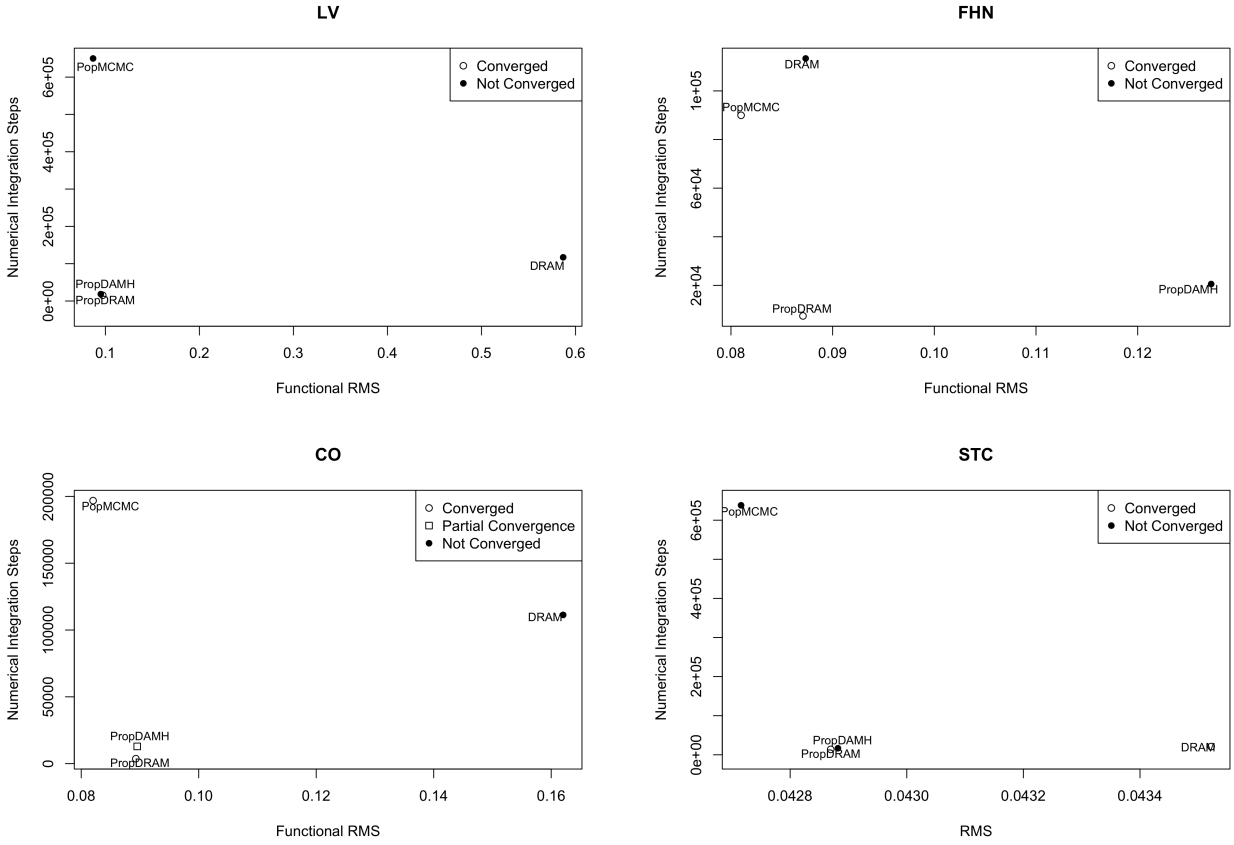

Figure 10: The figure shows the Functional RMS scores obtained by combining the posterior distributions from 10 independent data instantiations versus number of numerical integration steps. This allows consideration of accuracy (horizontal axis) relative to computational complexity (vertical axis) for each of the four methods. Good performance is signified by a method appearing in the bottom left corner. The three phase proposed scheme is the only method consistently appearing in the bottom left corner