Approximate Bayesian Inference for Individual-based Models with Emergent Dynamics

Jennifer Gaskell[1], **Nazareno Campioni**[1], **Juan M. Morales**[2], **Dirk Husmeier**[1], **and Colin J. Torney**[1]

[1]School of Mathematics and Statistics, University of Glasgow,
Glasgow, G12 8SQ, UK

[2] Grupo de Ecología Cuantitativa, INIBIOMA-CONICET, Universidad Nacional del Comahue, Argentina

Abstract - Individual-based models are used in a variety of scientific domains to study systems composed of multiple agents that interact with one another and lead to complex emergent dynamics at the macroscale. A standard approach in the analysis of these systems is to specify the microscale interaction rules in a simulation model, run simulations, and then qualitatively compare outputs to empirical observations. Recently, more robust methods for inference for these types of models have been introduced, notably approximate Bayesian computation, however major challenges remain due to the computational cost of simulations and the nonlinear nature of many complex systems. Here, we compare two methods of approximate inference in a classic individual-based model of group dynamics with well-studied nonlinear macroscale behaviour; we employ a Gaussian process accelerated ABC method with an approximated likelihood and with a synthetic likelihood. We compare the accuracy of results when re-inferring parameters using a measure of macro-scale disorder (the order parameter) as a summary statistic. Our findings reveal that for a canonical simple model of animal collective movement, parameter inference is accurate and computationally efficient, even when the model is poised at the critical transition between order and disorder.

Keywords: Bayesian inference, Gaussian processes, collective movement, emergent phenomena, statistical emulation, individual-based modelling.

1. Introduction

Current approaches for inferring the properties of complex interacting systems often either fit models at the macroscale via continuum approximations to the underlying dynamics [1–3], or fit to the fine-scale dynamics [4, 5] and then qualitatively compare with observed large-scale phenomena through individual-based simulation [6–9]. The first approach requires simplifying approximations when transitioning from a high-dimensional microscale system to a low dimensional macroscale model, while the second approach sacrifices formal uncertainty quantification. An alternate approach is to integrate inference with the simulation algorithm [10], however, running individual-based simulations of large populations over long time periods, such as a migratory season for an animal movement model, becomes very computationally expensive and this is exasperated by models that display sharp transitions in behavioural states due to small changes in parameter values. Here, we investigate the performance of an accelerated inference method based on Approximate Bayesian computation (ABC) using Gaussian processes and apply it to a model of collective animal movement.

Approximate Bayesian computation (ABC) is typically used for inference of model parameters, but this can quickly become computationally expensive with increasing model complexity, number of parameters and individuals, and time-frames. The performance of simulated runs compared to the data (or simulated 'data' to check the re-inference accuracy and certainty, in this case) is compared via summary statistic(s) which project the model output onto a lower dimension. The distance ,ρ, of the summary statistic for the model run with proposed parameters ψ^* to the summary statistic from the data with target parameters ψ is then compared to a tolerance level, T, and accepted if suitably close, i.e. $\rho = |S(\psi) - S(\psi*)| < T$, where $S(\cdot)$ is the projection of the data/model output onto a summary statistics. ABC proposes many ψ^* and sums up the acceptances to give an approximated posterior distribution. It is possible to accelerate this method by using sequential Monte Carlo (SMC)-ABC [11, 12], where rounds of ABC are performed sequentially, updating the prior of the parameters and reducing the tolerance for accepting. This allows a brief scan of the initial prior space, which may be uniform or even unbounded, before fine-tuning the range and weighting of proposed ψ^* in later rounds.

Gaussian processes (GP) are a powerful and flexible inference tool [13] that can be also used to accelerate the ABC process [14] by emulating the likelihood to reduce the number of forward simulations required (in this case, the individual-based model (IBM)). The GP-ABC method requires an acceptance kernel, which can be modified from the traditional accept/reject threshold in ABC, in order to approximate an initial likelihood function over a small number of points, which
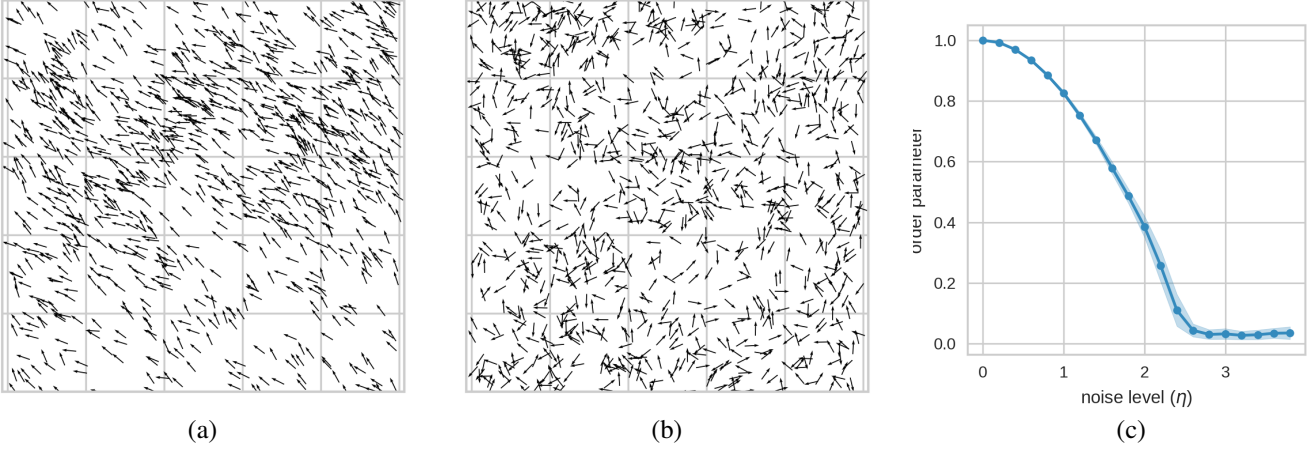
(a)
(b)
(c)

Fig. 1: a) Snapshot of the modelled particles (arrows indicating direction of travel) in 2D with noise $\eta = 0.5$, b) particles with $\eta = 2.0$, c) order parameter and standard deviation as a function of noise level, calculated over $M = 50$ repeats. We see the order of the particles decreases with increasing noise from an ordered state with order parameter close to 1, to a disordered state with order parameter close to 0.

then inform the emulator. The likelihood is only approximate as we have a stochastic system here. We investigate taking a Gaussian acceptance kernel to approximate the likelihood, and we compare this method to calculating the Gaussian synthetic likelihood function [15]. Both approximate likelihood approaches are accelerated using Gaussian process acceleration and sequential history matching [16] to efficiently rule out regions of parameter space.

2. The Vicsek Model

In this example, we will use the well-known Vicsek model [17] as the forward IBM we want to infer parameters for. The model contains particles that align to neighbours within a given radius, l_r, with the position of an individual, r_i, and its orientation, ϕ_i, given by:

$$\phi_i(t + 1) = \langle \phi_j \rangle_{|r_i - r_j| < l_r} + \Delta \phi_i, \tag{1}$$

$$r_i(t + 1) = r_i(t) + v \begin{pmatrix} \cos \phi_i(t) \\ \sin \phi_i(t) \end{pmatrix}, \tag{2}$$

where $\Delta \phi_i$ is a uniform random number drawn from the interval $[-\eta, \eta]$ and is the noise applied to each individual, i, while v is the constant speed of each particle. The system is governed by only two parameters; the density of particles (which is interchangeable with the interaction radius, l_r) and the amplitude of the noise, η. In a classic study of collective movement, Vicsek et al. [17] showed that this simple model displayed a phase transition between a disordered state, characterised by an average velocity of the individuals close to zero, and an ordered state in which the average velocity is close to v. This transition occurs as the noise parameter η and effective density of the particles is varied. Examples of this behaviour for 1000 particles are shown in Fig. 1.

For likelihood-free inference methods, we project the state of the system at the macro scale onto at least one summary statistic, which is used to compare the output of the model at the proposed parameters ψ^* to the data. For inference of the noise η in the 2D Vicsek model, we could take the summary statistic to be the order parameter,

$$\frac{1}{vN} \left| \sum_{i=1}^{N} \mathbf{v}_i \right| \tag{3}$$
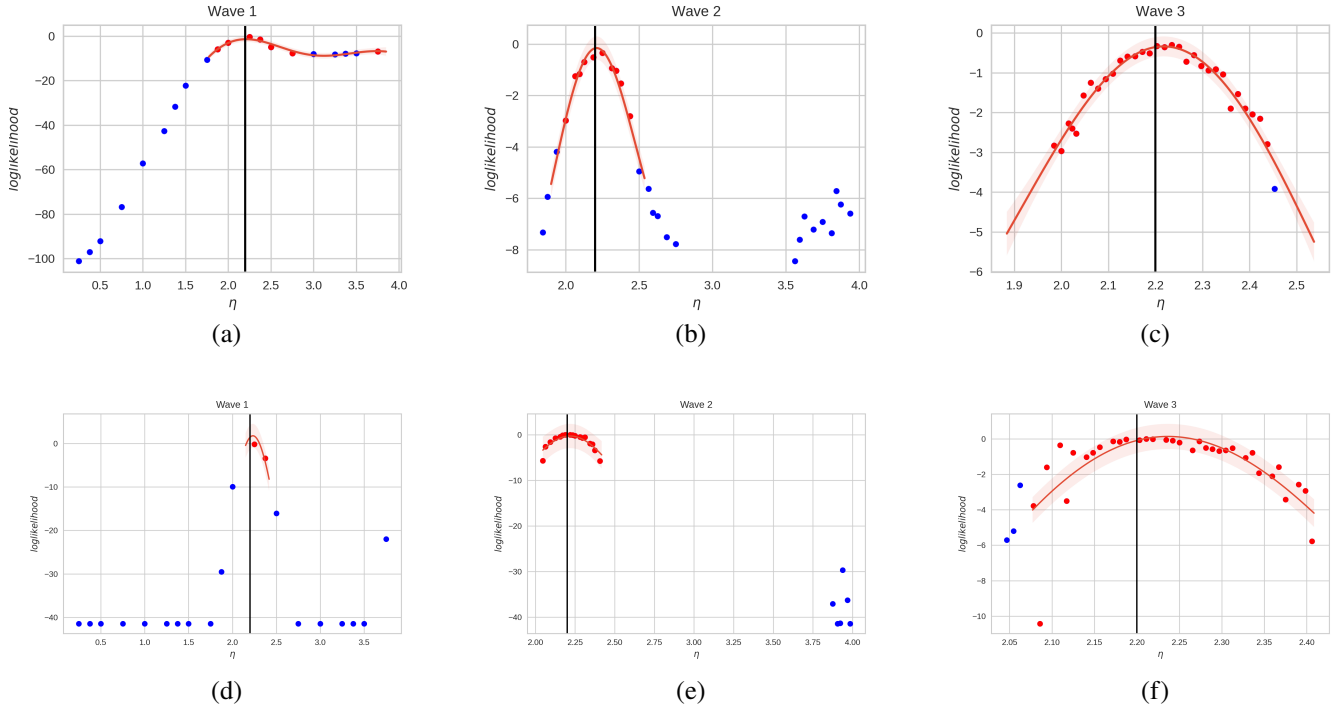
Fig. 2: Likelihood vs the noise parameter η for three GP-ABC waves, with simulated plausible (red) and implausible (blue) Sobol points shown, and the 'data' value $\eta = 2.2$ represented by the black line. Only the plausible regions are passed to the next wave, so note the change in x-axis. For each wave, another 20 Sobol points are added to the collection of plausible points, and the forward model is simulated at those points, with $M = 20$ repeats. (a-c) are GP-ABC waves using the approximate likelihood, (d-f) are GP-ABC waves using the synthetic likelihood.

where $\mathbf{v}_i$ is $[v\cos\phi, v\sin\phi]$, i.e. the observable state of individual i. The order parameter is equal to 1 for a fully-ordered system and 0 for a disordered system. Fig.1 shows the system in a) an ordered state with low η and b) the system in a disordered state with high η, with c) showing the order parameter after an initial burn-in phase, averaged over multiple repeats. We can see that the order parameter is a good candidate for a summary statistic within the given range shown in Fig.1(c) for the Vicsek model. Summary statistics are model dependent, but for collective animal movement models, candidates can include the degree of alignment [8], the nearest-neighbour distance [18], alignment and angular velocity [19], or position and velocity of the group centroid [20].

3. GP-ABC Method

Accelerated ABC methods reduce the number of microscale simulations required by replacing the approximation of the posterior with an emulation of the unknown likelihood function, assuming continuity and smoothness across small changes in the parameter space. Many simulators and acceptance kernels [14] may be chosen and, in this paper, we will use a Gaussian process-accelerated ABC (GP-ABC). The GP-ABC process occurs in sequential rounds, allowing a brief exploration of the parameter space before refining.

Initially, N points are sampled across the parameter space in a Sobol sequence to efficiently sample with a space-filling design. This first set of parameters make up ϵ_1 the first ensemble of Sobol points in the training set, and this ensemble is passed into the first wave of GP-ABC, where the microscale model is simulated for the chosen parameters, the summary statistic(s) are calculated and then the approximate log likelihood model is constructed. During this phase, we must choose a likelihood model, which will be discussed below. The implausibility of the points in parameter space is determined through the implausibility criterion:

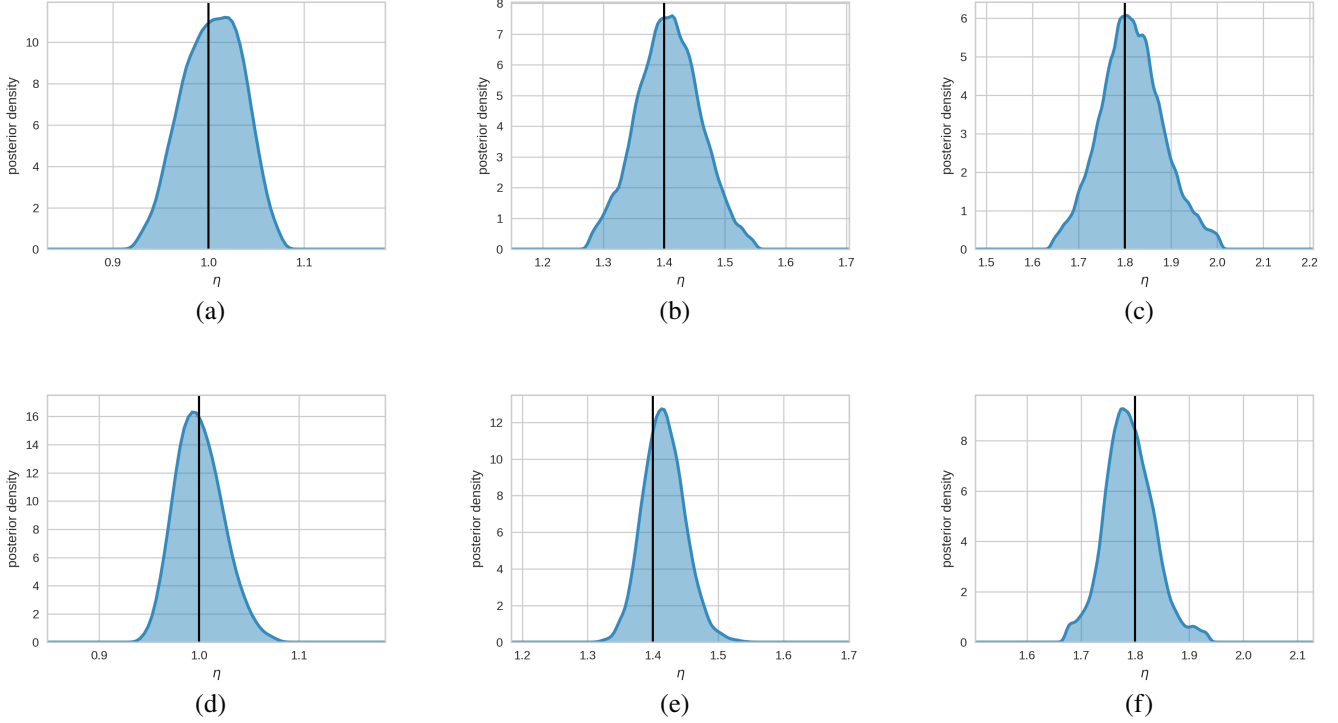$$m + 3\sigma < \max_{\psi \in \epsilon} \hat{l}_M(\psi) - T, \tag{4}$$

Fig. 3: Comparison of posterior distributions for the GP-ABC (top) with the synthetic likelihood (bottom) for three values of η used for the 'data' model. The true data parameter is indicated by the solid line.

where m and σ^2 represent the mean and the variance of the prediction of the modelled likelihood, $l(\psi)$, respectively, and T is a chosen threshold, which, in this case, we keep constant throughout GP-ABC rounds, as the decreasing GP uncertainty in sequential rounds will rule out further regions of the parameter space. For each GP-ABC round, we generate Sobol points, testing the implausibility using the mean and covariance function from the previous round, until we have collected a further N plausible points. This set of plausible Sobol points is then added to the surviving set from the previous round to make the next ensemble of points, ϵ_i for round i, and these are then used to construct the next GP model. From this GP model, we then update the mean and covariance in Eq.4, which is used to rule out any implausible points in the ensemble, as well as assessing new points for the next GP-ABC round. It is important to use the full ensemble of points from the entire sequence of GP rounds, as later rounds only consider regions deemed plausible by previous rounds, and therefore do not aid in prediction outside of these regions.

An advantage of the GP-ABC method is it's parallelizability and we use TensorFlow to run the repeats at each Sobol point in parallel. For the results below, we use the RBF kernel [21] for the GP model, with a small additional white noise kernel, to represent any measurement uncertainty.

3.1. Likelihood Selection
3.1.1. Approximate Likelihood

The GP-ABC likelihood function [14] $\hat{\pi}_{GP-ABC}(S(D)|\psi) = \int \pi(S(D)|X)\pi(X|\psi)dX$ can be approximated by the unbiased Monte Carlo sum:

$$\hat{\pi}_{GP-ABC}(S(D)|\psi) = \frac{1}{M}\sum_{I=1}^{M}\pi(S(D)|X_i), \tag{5}$$

where $\pi(S(D)|X_i)$ is an acceptance kernel, $S(D)$ is the projection of the true data onto a summary statistic, and X_i are simulations of the model at proposed points in the parameter space. The acceptance kernel can be chosen such that the hard
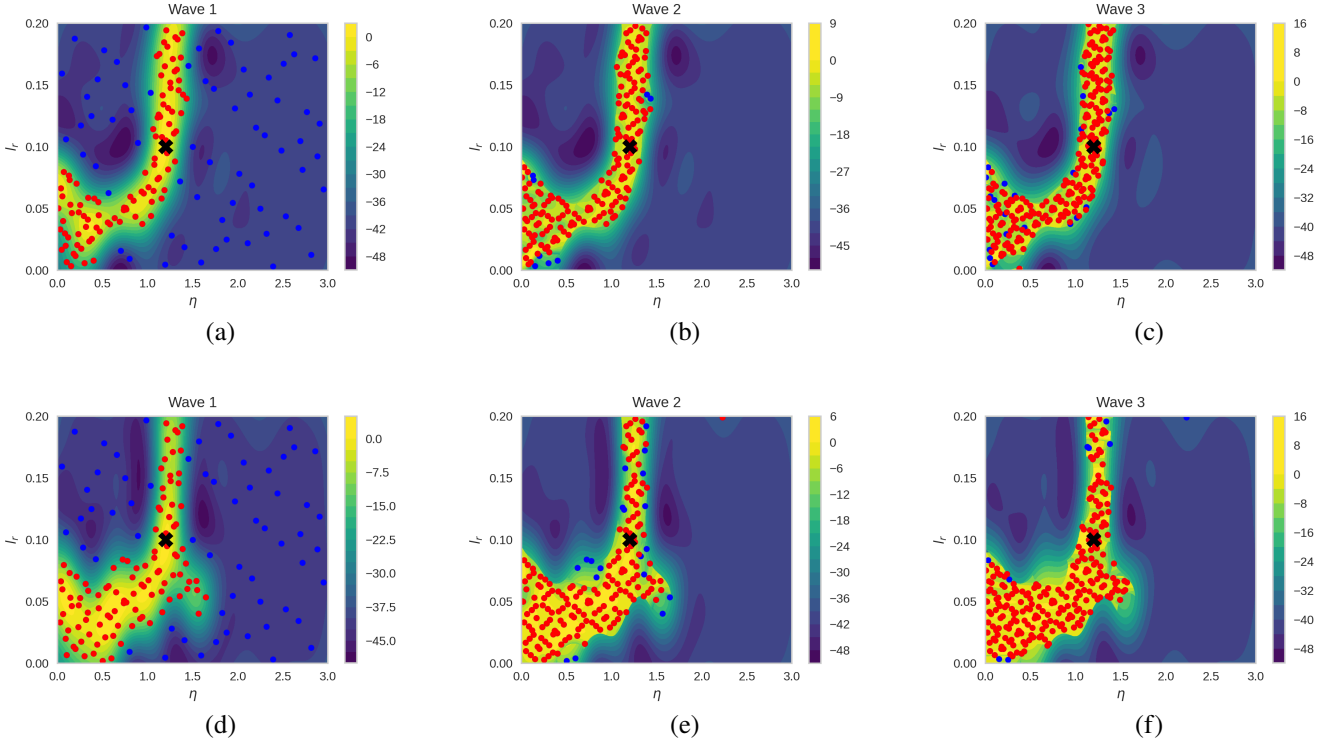
Fig. 4: Approximate loglikelihood(η, l_r) for a true data value of $\eta = 1.2$ and $l_r = 0.1$ (black cross), Sobol points shown, red indicating the points deemed plausible. Waves are generated through the sequential history matching method, using the Gaussian acceptance kernel to generate the likelihood (a-c) and the synthetic likelihood (d-f) with one summary statistic, the order parameter.

reject/accept step of the standard ABC algorithm, i.e. $\pi(S(D)|X_i) \propto \mathbb{I}_{\rho(D,X) \leq \epsilon}$, where ϵ is the threshold for acceptance, can be replaced by an alternative choice. Here, we will use a Gaussian acceptance kernel with $\mathcal{N}(0, T)$, where the tolerance T is taken to be the standard deviation of the simulation repeats at the 'data' point. This gives an approximation to the likelihood which will be fed into the GP emulator.

3.1.2. Synthetic Likelihood

Alternatively, we can consider the Gaussian synthetic likelihood,

$$\hat{\pi}_{syn}(S(D)|\psi) = \mathcal{N}(S(D); \hat{\mu}_\psi, \hat{\Sigma}_\psi), \tag{6}$$

where $\hat{\mu}_\psi$ and $\hat{\Sigma}_\psi$ are the mean and covariance of $S(X)$ estimated from M simulator evaluations at ψ, and $\mathcal{N}$ is a Gaussian density function.

Here, we will compare the performance of the approximate likelihood in Eqn.(5) when passed through GP-ABC waves, and the synthetic likelihood in Eqn. (6), also passed through GP-ABC waves, with the same total number of forward simulations of the model. For each method, we construct a model of the approximate/synthetic loglikelihood and pass this, and the uncertainty, into an MCMC-MH sampler. The loglikelihood model takes into account all waves from the GP-ABC, using the most-refined wave where possible. Any points deemed implausible are assigned a value of minus infinity so that they are rejected by the MCMC sampler and prevented from visiting implausible regions. We use a random realization, rather than the GP mean, to account for the error in the likelihood prediction, as it is only approximate for a stochastic model. Here, we compare the results of this final step for both of the tested methods and assess the accuracy whilst keeping the computational costs fixed by using the same number of forward simulations.

4. Results

4.1. 1D Inference

For the simple case of inferring one parameter, with a unimodal dependence of summary statistic on the inference parameter, both methods performed well by approximating the true data value, with low uncertainty. In Fig.2, we show the approximate loglikelihood through three waves of GP-ABC, using the approximate likelihood (a-c) and the synthetic likelihood (d-f). The synthetic likelihood method is quick to dismiss a large range of implausible values but does result in a slight off-set from the true data value, due to the inherent bias. Fig.3 shows the MCMC-MH posteriors after 20000 MCMC steps, calculated after three waves of GP-ABC for three different true data values, for the approximate likelihood method (top row) and the synthetic likelihood (bottom row). We note a slight advantage for the approximate likelihood method in terms of accuracy, but with a larger uncertainty. Both methods were computationally more efficient than running the same number of MCMC-MH steps with the forward model, as we have greatly reduced the number of simulations required [GPABC: 20 Sobol points x 20 repeats x 3 waves = 1200 simulations, compared to forward model and MCMC-MH: 20000 simulations] and can benefit from the parallelisation. The computational savings become more pronounced with increasing model complexity.

True η value	Approximate Likelihood	Synthetic Likelihood
0.6	0.015 / 1.000 / 8345	0.037 / 1.000 / 11640
1.4	0.031 / 1.000 / 2286	0.016 / 1.000 / 4477
2.2	0.076/ 1.001 / 628	0.019 / 1.001 / 1277

Table 1: MCMC convergence diagnostics after 20000 steps: The magnitude of the final Geweke chain value / Gelmen-Ruben diagnostic criterion / effective sample size for three 'data' values for the two likelihood approaches, show to 3 decimal places.

The MCMC chains used to construct Fig.3 are tested for convergence using the Geweke diagnostic, which takes the first 10% of the chain and performs a z test to compare to segments in the latter 50% of the chain. We show the last value from each chain in the first part of each cell in Table. 1 for the respective conditions, values between -2 and 2 show convergence. The latter part represents the Gelmen-Rubin diagnostic, which compares chains of MCMC with different starting conditions, and if the diagnostic is sufficiently close to 1, say below, 1.1, then the burn-in phase chosen is long enough to have allowed sufficient mixing. The burn-in phase used for all MCMC-MH results presented here is 1000 steps, out of a total 20000 steps. Both methods sufficiently converge for the tested parameters.

4.2. 2D Inference

Multiple parameters can be inferred simultaneously using either method presented here, or more traditional methods, such as SMC-ABC. Here, we investigate changing the noise, η, alongside the radius of interaction, l_r. Fig.4 shows the approximate loglikelihood through three waves of GP-ABC using the Gaussian acceptance kernel likelihood method and the synthetic likelihood method. Both methods arrive at approximate loglikelihoods focused around the true data value but would benefit from an additional summary statistic. In Fig.5(a), the order parameter summary statistic is shown, and we see the GP-ABC waves follow the available gradient towards the true value. Further summary statistics could be added to enhance the inference, but here we perform MCMC with the one summary statistic to reach the posteriors shown in Fig.5(b-c) for the two likelihood methods. The peaks do focus towards the true parameter values, but don't distinguish them fully due to η and l_r having a similar effect along the contour.

5. Conclusion

Inference of the underlying dynamics of complex systems from coarse observables is a fundamental challenge that spans multiple scientific domains. Approximate inference methods can be very beneficial for the study of these systems as the provide a principled way to link simulation models with empirical data. A major challenge for these approaches is coping with the complex emergent dynamics of interacting systems within reasonable computational limits. In this work we have shown how Gaussian processes can be used to accelerate approximate inference in complex systems. By taking advantage of
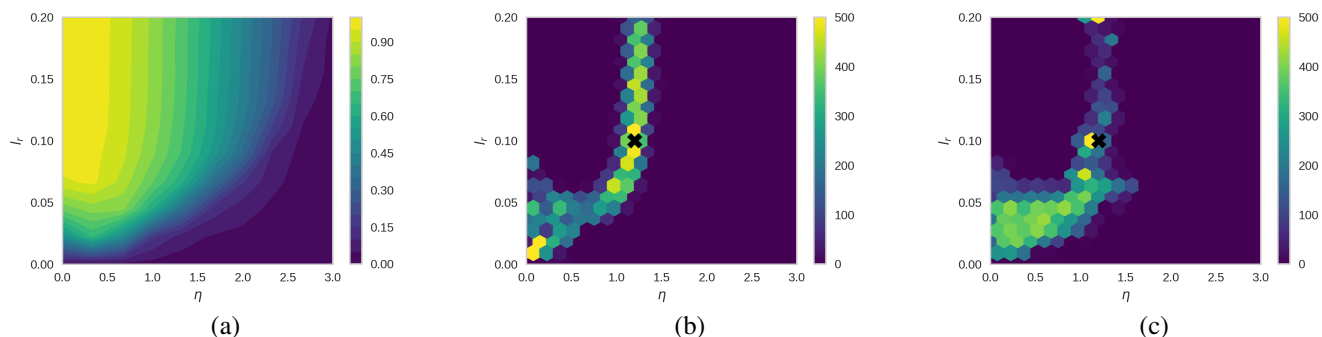
Fig. 5: a) The order parameter as a function of η and l_r, b) the MCMC posterior after three waves of GP-ABC using the Gaussian acceptance kernel likelihood, c) the MCMC posterior after three waves of GP-ABC using the synthetic likelihood.

the smoothness of the approximate likelihood surface and employing sequential history matching, our approach allows us to focus on plausible regions of parameter space and sample from the posterior without running expensive forward simulations at each step. By allowing the macroscale behaviour, which could depend on multiple parameters, to be condensed,via summary statistics, to an approximate likelihood function, we can infer parameters efficiently even when the system is poised at a critical transition between order and disorder.

After comparing two approximations to the likelihood, we note that the method proposed in [14] leads to accurate results in both 1 and 2 dimensional parameter spaces. Using the synthetic likelihood proposed in [15] within the GP-accelerated framework of [14] introduced some bias, and this is unsurprising as the method is based on a Gaussian approximation to the likelihood which is unlikely to hold given the nonlinear relationship between the microscale parameters and the summary statistic. However, further work is required to formally quantify the degree of bias and its dependence on the dimensionality of the parameter space.

The accelerated methods we employed were more computationally efficient than typical ABC methods, such as SMC-ABC, as the number of sample points in parameter space required for good results is much smaller, and the repeat simulations for each point can be parallelized. In a naive comparison of forward simulation evaluations, 1200 and 4800 [80 Sobol points x 20 repeats x 3 waves] forward simulations were performed for the 1 and 2 dimensional parameter inference respectively, and then 20000 MCMC steps were taken in both cases. Hence, using an approach that required a forward simulation at each MCMC step would require two orders of magnitude more forward simulations. With more complex models over longer time periods, with large numbers of individuals,such as modelling migratory populations, it becomes crucial to reduce the number of forward simulations required, and here,the emulation offered by methods such as Gaussian processes can be well utilised. A more thorough investigation into method comparisons in higher-dimensional inference will follow.

We note that our summary statistic is unable to distinguish between regions of the parameter space that in combination result in equivalent macroscale dynamics (see Fig 5a). Additional summary statistics could be used with consideration of weighting [22, 23] in these cases and will likely be essential for more complex models. Further, a promising avenue for future research is to integrate approximate inference based on coarse observables, such as undertaken here, with fine-scale inference of microscale interactions. In combination, multiscale inference of this form will be able to distinguish between regions of parameter space that are equivalent at the macroscale but are driven by different individual-level interactions.

References

[1] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," *Science Advances*, vol. 3, no. 4, p. e1602614, 2017.

[2] E. A. Ferguson, J. Matthiopoulos, R. H. Insall, and D. Husmeier, "Inference of the drivers of collective movement in two cell types: *Dictyostelium* and melanoma," *Journal of the Royal Society Interface*, vol. 112, no. 15, 2016.

[3] S.-N. Chow, X. Ye, H. Zha, and H. Zhou, "Influence prediction for continuous-time information propagation on networks," *arXiv:1512.05417*, 2016.

[4] A. Eriksson, M. Nilsson Jacobi, J. Nyström, and K. Tunstrøm, "Determining interaction rules in animal swarms," *Behavioral Ecology*, vol. 21, no. 5, pp. 1106–1111, 2010.

[5] C. J. Torney, M. Lamont, L. Debell, R. J. Angohiatok, L.-M. Leclerc, and A. M. Berdahl, "Inferring the rules of social interaction in migrating caribou," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1746, p. 20170385, 2018.

[6] A. Berdahl, C. J. Torney, C. C. Ioannou, J. J. Faria, and I. D. Couzin, "Emergent sensing of complex environments by mobile animal groups," *Science*, vol. 339, no. 6119, pp. 574–576, 2013.

[7] J. R. Potts, K. Mokross, and M. A. Lewis, "A unifying framework for quantifying the nature of animal interactions," *Journal of the Royal Society Interface*, vol. 11, no. 96, p. 20140333, 2014.

[8] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, J. Buhl, I. G. Kevrekidis, P. K. Maini, and D. J. T. Sumpter, "Inherent noise can facilitate coherence in collective swarm motion," *PNAS*, vol. 106, no. 14, pp. 5464–5469, 2009.

[9] S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, and I. D. Couzin, "Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion," *PNAS*, vol. 112, no. 15, pp. 4690–4695, 2015.

[10] C. Bretó, D. He, E. L. Ionides, and A. A. King, "Time series analysis via mechanistic models," *The Annals of Applied Statistics*, vol. 3, no. 1, pp. 319–348, 2009.

[11] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, vol. 6, no. 31, pp. 187–202, 2009.

[12] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate Bayesian computation," *Biometrika*, vol. 31, no. 6, pp. 983–990, 2009.

[13] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[14] R. Wilkinson, "Accelerating ABC methods using Gaussian processes," *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS), Reykjavik, Iceland*, 2014.

[15] S. N. Wood, "Statistical inference for noisy nonlinear ecological dynamic systems," *Nature*, vol. 466, no. 7310, pp. 1102–1104, 2010.

[16] P. S. Craig, M. Goldstein, A. H. Seheult, and J. A. Smith, "Pressure matching for hydrocarbon reservoirs: a case study in the use of bayes linear strategies for large computer experiments," in *Case studies in Bayesian statistics*, pp. 37–93, Springer, 1997.

[17] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Physical review letters*, vol. 75, no. 6, p. 1226, 1995.

[18] A. Kolpas, J. Moehlis, and I. G. Kevrekidis, "Coarse-grained analysis of stochasticity-induced switching between collective motion states," *PNAS*, vol. 104, no. 14, pp. 5931–5935, 2007.

[19] K. Tunstrøm, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin, "Collective states, multistability and transitional behavior in schooling fish," *PLoS Computational Biology*, vol. 9, no. 2, 2013.

[20] M. Raghib, S. Levin, and I. Kevrekidis, "Multiscale analysis of collective motion and decision-making in swarms: an advectiondiffusion equation with memory approach," *Journal of Theoretical Biology*, vol. 264, no. 3, pp. 893–913, 2010.

[21] C. E. Rasmussen and C. K. I. Williams, "Gaussian processes for machine learning," *MIT Press, ISBN 026218253X*, 2006.

[22] D. Prangle, "Adapting the ABC Distance Function," *Bayesian Analysis*, vol. 12, no. 1, pp. 289–309, 2017.

[23] P. Fearnhead and D. Prangle, "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 74, no. 3, pp. 419–474, 2012.