

Potential questions for discussing Ray et al. and Rivollat et al. papers.

Both studies use ABC to perform model selection and parameters inference. Both of them give important practical insights on how to best proceed for model estimation.

Ray et al.

Choice of summary statistics.

1) How do you choose the best set of summary statistics (e.g. based on previous knowledge or exploratory analyses)? Can you comment on this sentence *“However, because we modeled a very simple scenario for Asia, Asian-based summary statistics are likely to dominate the overall Euclidean distances computed between observed and simulated data and prevent a proper discrimination between settlement scenarios for the Americas.”*?

2) Can you explain the reasoning behind the following approach *“Parameters were estimated under a conventional ABC framework (Beaumont et al. 2002) after transformation of the original summary statistics using a partial least-squares (PLS) approach.”*?

Inference.

1) What is the preferred algorithm for obtaining the posterior distribution (e.g. rejection, local linear regression, weighted multinomial logistic regression, ...)?

2) Can you explain the following method *“The quality of the estimated parameters (i.e., the potential for a parameter to be correctly estimated) was assessed using the coefficient of determination R^2 (i.e., the proportion of parameter variance explained by the summary statistics) computed across all simulations.”*?

3) How do you evaluate posterior distribution for our parameters? Comment on *“Our modal values have a relatively wide 95% credible interval (400–956 generations for the Asian2S/AllAmericas data set), reflecting the difficulty in estimating this parameter from summary statistics that explain only 8% of its overall variability. However, the examination of the posterior distribution suggests that there is information about this parameter”*.

Model selection.

1) How can you use ABC to assess the relative probability of alternative scenarios? How do you deal with non-nested model in practise? How do you take into account the different number of free parameters each model can have? Comment on *“The penalty for models with additional parameters is simple to understand: for a fixed number of simulations, the parameter space of models with more parameters will be less well explored than for simpler models. If the additional parameters are not informative, there will be fewer simulations close to the observations, and the model with extra parameters will have a lower posterior probability.”*

2) Comment on *“Note that the better fit for the RGF model is not due to a bad choice of priors for the other two models because each model is able to reproduce each observed summary statistics.”*

3) Why do you get very different relative model probability using a different set of summary statistics? Comment on *“These less clear-cut results confirm that the addition of summary statistics computed on Asian samples make it more difficult to distinguish between models of the settlement of the Americas,”*

Prior probability.

How can you make informed decisions on which prior distribution is more suitable for a specific model? Comment on *“When data are not very informative, priors may dominate the outcome. One should, therefore, check that the final results do not overly depend on the chosen priors, for instance, if models are based on very different sets of priors. “*

Rivollat et al.

Summary statistics.

Why did you use only one summary statistics? How confident are you on your results based on the assessment of a single summary statistics? Comment on *“We used the rejection algorithm of the ‘abc’ package available in R to retain the parameter combinations that generated simulated pairwise F_{ST} the closest to the six observed values. “*

Why didn't you use modern genomes? Comment on *“However, we were not constrained to simulate population expansion, since we did not consider modern DNA data in our analyses. “*

Inference and model choice.

Comment on Figure 3.

Software.

Which software would you recommend to perform an ABC analysis (R package, ABCtoolbox, DIY-ABC, ...)?