Dear Dennis and Michael,

Thank you very much for the time invested by both of you in reviewing our paper and for your useful comments. In the light of your remarks, we greatly improved our manuscript. Please find below our response to your different suggestions. We took the spelling or typographical remarks into account and omitted to comment them.

Response to Dennis Prangle

Major comments:

1. *I suggest a discussion of Papamakarios and Murray (2016). This recent method also performs ABC-like inference using a machine learning regression method – neural networks – in place of nearest neighbours. It would be interesting to discuss, at least briefly, the relative pros and cons of random forests and neural networks in this setting. Two potential advantage of neural networks are that they can produce an approximation of a multivariate posterior, and that they can, in theory, work with raw data without requiring features to be proposed. (Although this seems hard to implement in the genetics application.)*

   Thank you very much for this remark. The work of Papamakarios and Murray (2016) is very interesting and promising. We now discuss/comment this method in Section 4 of the manuscript (main text).

2. *I'd like to see some discussion of the time required to fit the random forests compared to ABC regression-adjustment methods.*

   Such a comparison has been added in Section 8 (supplementary material).

3. *Section 3.3 discusses two tuning choices: number of simulations and number of trees. Another choice is the minimum leaf size - do you have any comments on how this might affect the method? One might think this in some sense controls the level of approximation.*

   We now discuss this point in the Section 3.3 ("Practical recommendation") of the main text. The new Table 6 in the main text highlights that $N_{\min} = 5$ is a good choice.

4. *The paper focuses on approximating univariate posteriors. Is there any prospect of achieving multivariate posteriors – e.g. using multiple objective random forests (Kocev et al., 2007)?*

   We now discuss this point in Section 4 (main text) and propose several perspectives on how to use RF to possibly recover the joint posterior.

5. *The paper nicely illustrates that the random forest posterior variance estimates tend to be biased upwards. I wonder if this is due to the random selection of features for decision trees. That is, some trees will select less informative features and so produce estimates biased towards the prior. There might be some scope to avoid this using alternatives to random feature selection e.g. Bayesian additive regression trees, or boosted regression*

trees. This could be worth discussing (I'll leave this to the discretion of the authors.)

We might tentatively and verbally discuss that point but we prefer avoiding speculative comments on an item which, although of interest for specialists, is in our opinion a bit under the scope of our work.

Minor comments:

1. *Section 2.3.4: "This representation remains valid since the weights are equal to zero when $\eta(y) \neq \eta(y^{(t)})$ in the limiting case of exact ABC, namely when only accepting parameter values for which the summary statistics of simulated data are identical to the summary statistics of the observed data."*
   I didn't understand this sentence - can you elaborate? In particular:

   - *What does "valid" refer to? (consistency as $N \to \infty$?)*
   - *What's the relevance of exact ABC to the random forest method?*

   This part was indeed not clear at all. We reformulated it completely. Thank you for pointing it out.

2. *In Section 3, could you briefly mention the definition of normalised mean absolute error. I'm not completely sure what normalisation would be used.*

   Precisions on the definition of normalised mean absolute error were added in the new version of the manuscript.

3. *In Section 3.2 you describe the demographic model. I suggest briefly mentioning the genetic model as well.*

   Thank you for this remark. We took it into account by adding a brief text devoted to the algorithm of Hudson (2002) in the legend of Figure 4 (where information regarding the demographic model was already provided).

4. *Section 4. "The performances for covariance approximation are quite encouraging as well..." I suggest emphasising that all the details of this are in the supplementary material.*

   Done.

5. *Pg 18 - "quantile estimation is not uniformly optimal". Does this mean that regression correction is sometimes better? Which table/figure is this referring to?*

   This part has been reformulated. By "uniformly" we meant depending on the quantile probabilities, as we can see on Figure 2.

6. *Supplementary material pg 1: I didn't understand how "method 2" would work. I suggest adding a brief description of it.*

   We developed the description of this part, hoping that it is now clearer.

7. *Supplementary material pg 4: "maximum node size equal to 10". I think this should say "minimum node size". Also on page 7 of the main paper minimum node size was 5 why change to 10 here?*

This were mistakes that we have corrected in the new version of the manuscript.

8. *Supplementary material pg 4: Is multivariate ABC regression correction used here, or is a single scalar parameter targeted as for the random forest approach?*

For local linear or ridge regression the corrections are univariate. That is not the case for neural networks which, by construction, perform multivariate correction. This info has been included in the new version of the manuscript.

Response to Michael Blum

Major comments:

1. *According to my reading of the manuscript, there are 3 main advantages when using random forest (RF) for parameter inference with ABC. The first advantage is that RF can handle many summary statistics and that dimension reduction is not needed because of attractive properties of RF. I perfectly understand this argument. However, I think that it is purely theoretical at the moment because the results do not provide support for this. Simulation results where uninformative summary statistics are added and where the properties of different ABC approaches (without dimension reduction) are compared may be worthwhile to report.*

Thank you very much for that judicious comment. On the human population genetics example we added 20 independent uniform noisy variables and found (as expected) that it does not impact the results in terms of NMAE, coverage and parameter estimation. This new result is now reported at the end of Section 3.2 (main text) as well as in several results Tables and Figures.

2. *The main drawback of the proposed approach concerns joint parameter inference. There are many settings where the joint parameter distribution is of interest and the proposed approach cannot provide joint posterior distribution. Estimating the covariance is not enough information because the joint posterior distribution may not be Gaussian.*

Thank you for this comment also made by the other referee (Dennis Sprangle). We now discuss multidimensional posterior estimation in Section 4 (main text). In this new discussion paragraph, we introduce a way using RF methodology to recover the joint posterior; this is still preliminary work.

Minor comments:

1. *Please indicate, if you consider homoscedastic or heteroscedastic adjustments when considering regression adjustment approaches*

Heteroscedastic adjustments have been used; now clarify this issue in Section 3 (main text).

2. *It would nice in the R package to provide histograms or density distributions to visualize posterior distribution. Visualization routines are useful when performing data analysis.*

   The `densityPlot` function provides a density estimation of the marginal posterior distributions. We now explain how to use this function at the end of the R script detailed in Section 3 (supplementary material).

3. *The paper "Saulnier, Emma, Olivier Gascuel, and Samuel Alizon. Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study. PLoS computational biology 13.3 (2017): e1005416." contains nice comparison results about accuracy of ABC estimators with regression adjustment. It could be discussed or cited in the present manuscript.*

   Thank you for that remark. We now cite and briefly discuss this paper in Section 2.1 (main text).

4. *Introduction*
   I do not know what is the calibration side of RF and readers might not know either.

   The calibration of RF refers to the choice of the tuning forest parameters: the number of trees, the number of covariates sampled at each node, the maximum leaf size. It is now further detailed and discussed in Section 3 (main text).

5. *ABC for parameter inference Section "calibration of the ABC algorithm" should be defined. It seems to me but I am not sure that the word calibration here refers to something different than the calibration of the posterior distribution. The same word should not be used with different meanings to avoid confusion.*

   In this paper, we did not mention the notion of calibration of posterior distribution and by calibration we mean the tuning of the different algorithm used. "calibration of the ABC algorithm" hence refers to the choice of the tolerance ϵ. This is specified all over the manuscript, especially in the third paragraph of Section 2.1 (main text).

6. *Alternative variance approximation I do not understand the new approximation because I do not know what τ_{oob} is.*

   This is a commonly used notation in RF but there is no τ_{oob} to estimate, $\hat{\tau}_{oob}$ designs the estimated posterior expectation using the out-of-bag observations of the different trees of the forest. This is already and clearly specified in the manuscript.

7. *Figure S3 and corresponding text. Plotting cumulative distribution function is not suited to visualize differences between distributions. Density distribution functions should be preferred.*

   We now represented the density distribution functions as you suggested, thank you.

8. *Figure 3: The caption should contain more details to understand what boxplots represent.*

9. *Discussion I would not mention the covariance approximation because it seems that it is still preliminary work.*

   As suggested by Dennis Prangle, we mention that this point is now detailed in the supplementary materials only.

10. *Uniformly optimal is difficult to understand without being defined. What are ancillary covariates?*

    We rephrased this part to make it easier to understand.

11. *References*
    Please check the references
    Csillèry should be Csilléry, Blum M should be BLUM MGB, Maples is not the only author of his paper, Wright and Ziegler paper has now been published in Journal of Stat Software...

    The list of reference has been checked and improved, thank you.

References

Hudson, R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18(2):337-338.

Kocev, D. et al. (2007) Ensembles of Multi-Objective Decision Trees. In Machine Learning: ECML 2007. Lecture Notes in Computer Science, vol 4701. pages 624–631. Eds. Kok J. N., Koronacki J., Mantaras R. L., Matwin S., Mladenič D. and Skowron A. *Springer, Berlin, Heidelberg*, 2007.

Papamakarios, G. and Murray, I. (2016) Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. In Advances in Neural Information Processing Systems 29, pages 1028–1036. Eds. Lee D. D., Sugiyama M., Luxburg U. V., Guyon I. and Garnett R. *Curran Associates, Inc.*, 2016.