RESPONSE TO REVIEWERS

Editor

Editor:

Response:

Reviewer: 1

R1:

Response:

Reviewer: 2

R2:

Response:

Reviewer: 3

Main text

R3: I think the authors should add some comparison to other approaches of inference for OU processes (for example in the introduction), to allow the reader to understand how the work sits in the context of dynamical inference methods

ResponseDraft: *Link with comments of Reviewer 1: we can probably group the two answers and modify the intro in a way that satisfies both.*

R3: I think that statement of the problem is an important part that probably deserves to be a section in itself instead of a subsection of Methods.

Response: *"Statement of the problem" is now a section in itself.*

R3: The extreme cases discussed at lines 193-195 seem the same discussed at lines 144-147. If this is the case, the authors could re-think whether it is worth to repeat this point or to establish explicitly the connection to what has been discussed before.

Response: *We agree with reviewer 3 that this was redundant, and modified the text accordingly. The last paragraph of section II (former lines 144-147) now introduces timescales of the OU process without referring to the inference procedure. Former lines 193-195 then detail what these timescales imply for the inference procedure (the two extreme cases), referring to section II.*

R3: Is it necessary to introduce a new symbol, z, for the expression (13)? Note that z is used also in the Appendix for something with a different meaning (e.g. line 637)

Response: *The symbol z is only used in the couple of equations that follow its introduction, and we agree that it is not strictly necessary. However, we felt that it made the derivation clearer by separating the calculation of eigenvalues and eigenvectors of $\mathbb{G}$ from the specific functional form of eigenvalues of $\mathbf{\Lambda}$.*

R3: I think it is not immediately clear what the Results section will address. I think one could immediately state (like at line 316) that the goal is to assess the accuracy of Cmax compared to the real C and that such estimate will be compared to the alternativa estimate provided by the empirical C. The definition of empirical C could be also explicitly written down. Later on, at line 374, the authors mention the empirical coupling matrix but it is not clear how it is derived, is this simply then the inverse of the empirical C? Clarifying this would help better understand why the fact that C is close to singular (line 349) has an impact especially on the error on J (Fig. 3 right).

Response: *We agree with the reviewer that the Results section lacked explanation. We added an introductory paragraph that states the goal (evaluating the accuracy of the inference procedure) and describes how we proceed. The empirical covariance matrix as well as the different coupling matrices are also clearly defined in this new paragraph.*

R3: I thought also that it is not clear how the predictions for an i.i.d sample (lines 355, 376) are obtained: do they serve as a random control giving the random expectation in for the errors in e.g. Fig. 3? I think this should be stated more clearly, also in the caption, maybe along with the numerical value itself of the i.i.d. error (that is difficult to read from the plotss scale).

ResponseDraft: **Edwin: any way that you can get and add to the text the exact numerical values for the error of the i.i.d. of sizes 16 / 32 / 64? Those are the examples we give in the text.**

Response: *The predictions of the* i.i.d. *sample can indeed be used as a control to compute the minimal error that we can expect for a sample size N. Here, we use them to define an effective sample size: to each value of the reconstruction error $||\boldsymbol{C}_{max} - \boldsymbol{C}||_2$ corresponds an* i.i.d. *sample of size N for which the same l2-error is obtained when using the empirical estimation of the covariance matrix. This allows us to formulate the gains obtained by our method in terms of an equivalent increase in* i.i.d. *sample size.*

We have reformulated the corresponding paragraph to clarify this idea, and added more explanations to the caption of figure S5.

R3: Line 336: It is written that the sampling is repeated 100 times, it is not clear if also the inference is repeated 100 times or if this is done to increase N, the sample size. In relation to this, it would be important to write somewhere what N is considered in the computational tests and hence what is the computational complexity given this N (and $L = 4, 10$).

ResponseDraft: *What does he mean by computational complexity given values of N and L? I've only ever seen computational complexity expressed as an asymptotic quantity.*

Response: *The inference is indeed repeated a hundred times. We do this to average our results accross different realizations of the OU process. We now clearly state N corresponds to the number of leaves of the tree, in our case $2^9 = 512$.*

R3: Line 370: Please add some references for the problem of contacts in proteins.

R3: Line 421: Please add some references for the evidence that protein landscapes are well described by pairwise models.

ResponseDraft: *TODO*

R3: Regarding the results, do the authors believe that inferring γ could be interesting in terms of biological applications? Could it inform about typical evolutionary timescales?

Response: *The inference of γ indeed provides information about typicaly times of the underlying evolutionary process, provided that the eigenvalues of the covariance matrix $\boldsymbol{C}$ are also known (Eq. (7) and comments below it). If the evolution of, say, protein sequences is reasonably well described by the OU process, we could then be able to make statements about the expected change in an amino acid sequence as a function of time or number of generations.*

A current limitation of that approach is that γ will be inferred using distances separating leaves on a tree. In typical tree-inference software from biological sequences, the branch length is related to the expected number of mutations on that branch, but cannot be directly related to physical time or number of generations.

ResponseDraft: *Not sure if this is a good answer. Branch lengths on trees can be related to real time using a molecular clock hypothesis, but it's unclear to me if that's consistent with the OU process itself?*

R3: Pierre: Below: issues that I have addressed by changing the MS, but that perhaps do not require one response each.

- Line 191: I would state explicitly why the inference becomes impossible
- Line 240: should "we construct the direct product of vectors s_a and u_{ka}, defining vectors S_{ka} of dimensions $L \times N$ be we construct the direct product of vectors u_{ka} and s_a, defining vectors S_{ka} of dimensions $N \times L$"? I would also state explicitly that the superindex refers to the components of u.
- Line 278: double brackets around 17.
- Is there a minus in front of r.h.s of (19)?
- Line 332: I would add a comment on the choice of the range for gamma (why not even smaller or bigger) - this is probably due to the fact that all the variation concentrates in the range considered but I would state it.
- Line 346: As far as I see the l2-error is not defined anywhere - I think it would be important to have it defined somewhere (could be also in a caption or in the appendix).