Dear editor,

we are very pleased about the opportunity to submit a minor revision of our paper 'Sequential double cross-validation for assessment of added predictive ability in high-dimensional omic applications' for publication in Annals of Applied Statistics. Please, find below our point-to-point answer to the comments (if any) raised by you and two reviewers' reports.

Following your suggestions, we have modified our manuscript. All the changes are highlighted in bold type. We hope that we have answered all the concerns adequately and have improved the quality of the manuscript so that you will consider it for publication in Annals of Applied Statistics.

Sincerely,

Mar Rodríguez-Girondo on behalf of the authors

1 Conditional vs. joint fitting

1.1 Comment of the editor:

The AE and the reviewers agree that this revision is appropriate for publication with a few notation and wording changes in the document to clarify the relationship of this approach with other available approaches. The significance of the work is brought up again by two reviewers. I would like to see the notation issues described by reviewer 2 and the discussion points raised by reviewers 2 and 3 about significance addressed, directly and honestly (pros, cons, limitations) in a final version.

1.2 Comment of reviewer 1 (report AOAS1607-015R2R4):

Ultimately I am still not convinced by the implied conclusion in the final section of the manuscript that the 'two-stage' approach may be superior to any 'joint' approach. However, I think the authors' can argue that their two-stage approach is robust to differences in scaling and dimensionality between the two data sources, which must be carefully considered in any joint approach.

1.3 Response:

We agree with the editor and the reviewer that our discussion still needed to be improved to be clear about the strengths and limitations of our sequential approach and the potential promising lines of future research involving 'joint' approaches. We have modified the discussion in order to make these points clear. Namely, in the new version of our manuscript, we explicitly mention the asymmetric nature of our approach as a limitation and its robustness for dealing with omic sources which largely differ in scale and dimension as a strength.

Moreover, we indicate as promising lines of research approaches based on omic-specific penalization using, for example, group penalization.

2 Response to reviewer 2 (report AOAS1607-015R2R5):

2.1 Minor comment:

The notation PRESS, CVSS, res and the like is somewhat distracting and makes equations more difficult to read. It would be better if the authors defined more compact notation, e.g. replacing res with r, or, barring that, at least make these math operators so they display like PRESS instead of *PRESS*

2.2 Response:

We have followed the suggestion of the reviewer and replaced **res** by **r**. We have kept the notation PRESS and CVSS, but following the display suggestion of the reviewer.

2.3 Minor comment:

The description of procedure on page 5-6 is hard to read, in part because of the excessive indentation used in the itemize environment. This may be easier to follow if presented in an algorithm environment.

2.4 Response:

We have followed the suggestion of the reviewer and we have rewritten the description of the double cross-validation procedure in an algorithm environment.

2.5 Minor comment:

The use of multiple superscripting, e.g. $(S^{(j)})^{(-k)}$, is quite the mess. I have more often seen $S[j]$ to indicate the elements not including element j. This may not be the best choice here, but certainly something better than the current notation should be easy to achieve.

2.6 Response:

Following the suggestion of the reviewer, we have reformulated our notation in Section 2.2, avoiding the use of multiple superscripting. E.g.: $(S^{(-j)})^{(-k)}$ has been replaced by $S_{[-j;-k]}$ and $(S^{(-j)})^{(k)}$ has been replaced by $S_{[-j;k]}$.

2.7 Minor comment:

Tables 1 and 2 are difficult to read because so much space is taken up by the first two columns, which just serve to identify the scenario and its properties. Perhaps removing n = everywhere in the second column, which is already called n, will alleviate this a bit.

2.8 Response:

We have followed the suggestion of the reviewer and we have removed 'n=' everywhere in the second column of Tables 1 and 2.

2.9 Minor comment:

In several places awkward wording remains, e.g. at the top of page 2 among which the evaluation of the ability (perhaps replace among which with including) and near bottom of page 2 partially common underlying biological information (perhaps replace with shared underlying biological factors or common biological factors).

2.10 Response:

The indicated changes have been introduced in the new version of our manuscript (highlighted in bold).