Dear Dr Chandra,

thank you very much for your feedback. Our responses to the reviewer comments are organised in a point-by-point fashion following each comment in italics. The changes made to the manuscript are indicated in blue.

Reviewer 1

Reviewer 1 did not make any comments.

Reviewer 2

Congratulations on a very interesting paper! The following should be corrected prior to publication (line numbers refer to the track changes version of the paper)

Figs 4,5,7, 8 and appendix 2: Change "strongly regular" instead of "strong regular", "weakly regular" instead of "weak regular", "weak clustered" to "weakly clustered", "strong clustered" to "strongly clustered"

line 36: "only cover" instead of "only covering"

line 43 " this has even been..." instead of "this has been even..."

line 56 "reference data randomly distributed within" instead of "randomly distributed reference data within"

Line 106 "in the presence" instead of "in presence"

line 121 "coordinates of the training points" instead of "training points' coordinates"

line 123 "q clusters into k folds" instead of "q into k folds"

line 138 " in Euclidean space" instead of " in the Euclidean space"

line 139 "as the coefficient " instead of "as coefficient "

line 176 either: " a poor estimate" or "poor estimation " instead of "a poor estimation"

line 270 "does not" instead of "doesn't"

line 286 "in the absence" instead of "in absence"

We thank professor Mueller for the acknowledgement; we have implemented all suggested edits.

Finally the figures in appendix 2 need to be improved and titles/ explanations positioned on the same page as the relevant figure

We have improved the appendix as suggested by the reviewer.

Reviewer 3

This paper is an improvement on the efficiency of the authors' previous work NNDM LOO CV, which is significant from the results (Fig. 8). However, some necessary modifications need to be made before the paper can be accepted. In fact, I don't think there are too many technical problems in this paper, but there is a lot of room for improvement in the writing of the paper. In order to let readers better understand their work and expand the influence of the paper, I suggest that the author consider my opinions.

Thank you for the acknowledgement and suggestions.

1) The title of the paper is suggested to be changed to kNNDM CV: k-fold... Or simply remove kNNDM, because k-fold NNDM is more accurate, but this abbreviation is too long.

We have changed the title to "kNNDM CV: k-fold Nearest Neighbour Distance Matching Cross-Validation for map accuracy estimation" as suggested.

2) Since the biggest contribution of the paper is to improve the efficiency of NNDM LOO CV method, the abstract needs to explain the specific efficiency improvement shown by the experiment (For example: how many times improved).

We have added a sentence in the abstract quantifying the efficiency improvement case of 4000 strongly clustered training points (lines 15-18):

We found that kNNDM CV performed similarly to NNDM LOO CV and produced reasonably reliable map accuracy estimates across sampling patterns. However, compared to NNDM LOO CV, kNNDM resulted in significantly reduced computation times. In an experiment using 4,000 strongly clustered training points, kNNDM CV reduced the time spent on fold assignment and model training from 4.8 days to 1.2 minutes.

Furthermore, we have added a sentence in section 3.4 explicitly quantifying the efficiency improvement shown in the experiment (lines 207-209):

For a sample size of 4,000, kNNDM CV reduced the time spent on fold assignment and model training from 3.2 hours to 30 seconds for random samples and from 4.8 days to 1.2 minutes for clustered samples, as compared to NNDM LOO CV.

3) I noticed that Mila et al., 2022 had been cited 20 times in the paper, which generally only needs to be cited once, not every time it is mentioned.

We have revised our manuscript and removed instances of the citation that could be omitted. However, there are still multiple citations in the paper that are needed to support certain statements. The total number of references has been reduced to 12.

4) In Part 2 I noticed that the premise of understanding this paper seems to be reading (Mila et al., 2022), on the basis that I think each paper should be independent. Therefore, I suggest authors supplement the introduction of NNDM LOO method and related concepts in this paper?

We have reviewed section 2 and added all necessary background information so that reference to Milà et al. (2022) is no longer required to follow the text. Namely, we have:

- Included equations of nearest neighbour distances ECDFs rather than referring to them in the NNDM paper (lines 89, 92 & 98).
- Deleted unnecessary mentions to the original algorithm.
- Modified figure 1 to show an example of k-fold cross-validation rather than a LOO CV.

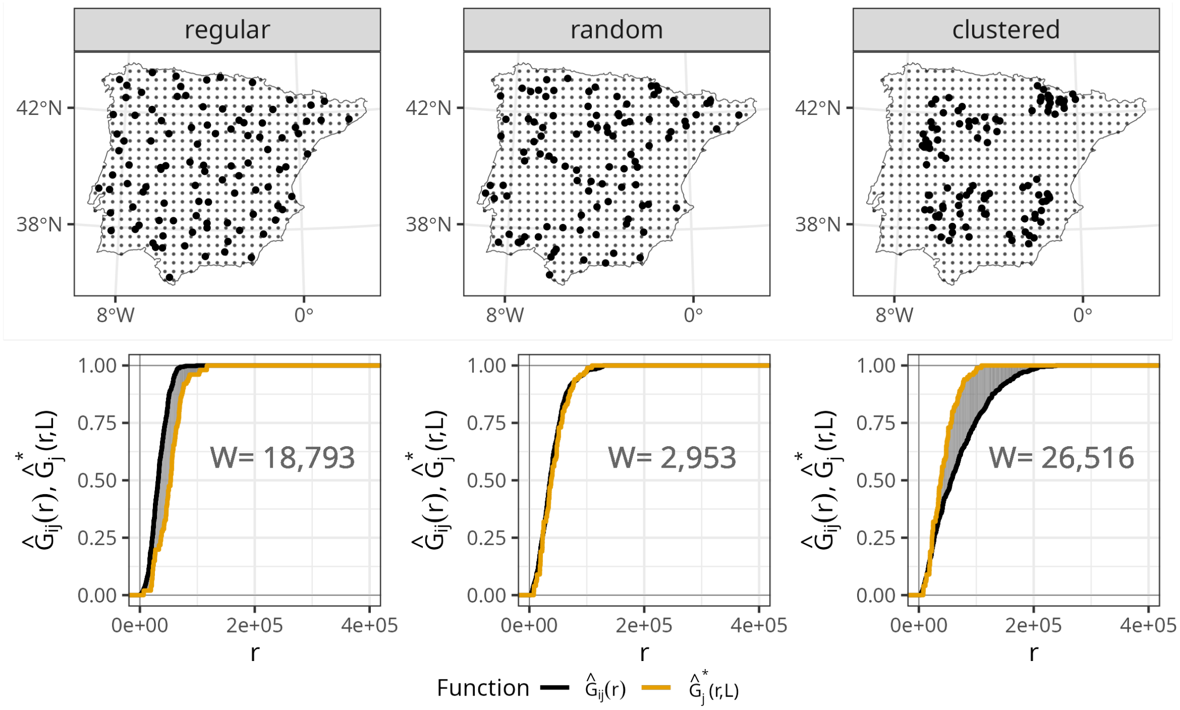

Figure 1: Top row: prediction points (regular grid) and training points with different spatial distributions (bold), simulated for visualization purposes only. Bottom row: NND ECDF between training and test locations during 10-fold random CV ($\hat{G}^*_j(r, \mathbf{L})$, orange) and NND ECDF between prediction and training locations ($\hat{G}_{ij}(r)$, black) corresponding to each of the sampling distributions in the top row. The shaded grey area corresponds to the W statistic, whose value is printed in the plots.

5) How does the paper implement k-fold with w statistic? Authors seem to want the reader to understand their thoughts from the algorithm, which is often difficult. I suggest they devote some content to addressing their idea directly. Figure 2 also looks too complicated. The opinions of the given Community comments, I also went to see Wang's paper (https://doi.org/10.1016/j.jag.2023.103364), the paper compared easier to understand, which can be refered.

We have added a new first paragraph to section 2 which presents, in non-technical language, the objective of the algorithm and the main ideas behind its implementation. With it, we equip the reader with an overall understanding of what we want to achieve before going to the finer and more technical details (lines 77-83):

The objective of kNNDM is to find a k-fold configuration such that the distribution of NND between test and train locations during CV matches as close as possible the distribution of NND between prediction and train locations. In other words, kNNDM aims to create predictive conditions in terms of geographical distances that resemble those found when using the model to predict a certain area. To do so, we use a clustering approach to create a set of candidate fold configurations with different degree of spatial clustering, of which we choose the one that offers the best match between the two distributions. Before explaining the algorithm in detail, we define the different NND distribution functions used in kNNDM, as well as the statistic used to evaluate the different fold candidates.

Regarding Figure 2, we have carefully considered it and tried to identify elements to simplify according to the reviewer's comment. While we cannot omit any of the included steps as these are the elements of the algorithm, we have reduced the text that was referring to the principal components part. The updated figure looks as follows:

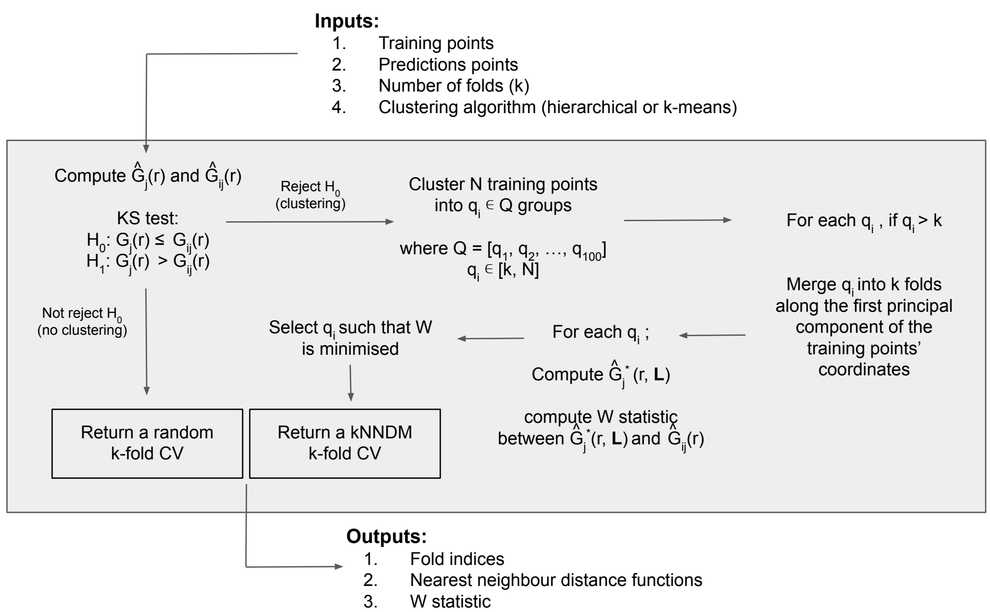

Figure 2: Workflow of the kNNDM algorithm.

In addition to this change, we would like to note that the main part of the algorithm, including the clustering and choosing the best fold configuration by minimizing the Wasserstein statistic, is graphically illustrated in Figure 3. Finally, we refer to Wang's paper in the discussion section (lines 269-271):

For example, Wang et al. (2023) recently developed a CV method that considers both the geographic and feature space, although it does not consider the prediction domain and predictive conditions of the model.

6) In my opinion, Figure 1 and Figure 3 are also unnecessary. These conclusions can be explained through demonstration or reference. If necessary, they should be included in the experimental part instead of here. The authors should note that since the data is presented in a later section, the reader will have more confusion about how these results were obtained.

In order to remedy the potential confusion regarding the origin of the data in Figures 1 and 3, we have improved their figure labels to highlight that their use is for illustrative purposes only. The figure captions now read:

Figure 1. Top row: prediction points (regular grid) and training points with different spatial distributions (bold), **simulated for visualization purposes only**. Bottom row: NND ECDF between training locations found during 10-fold random CV ($\hat{G}^*_j(r, \mathbf{L})$, orange) and NND ECDF between prediction and training locations ($\hat{G}_{ij}(r)$, black) corresponding to each of the sampling distributions in the top row. The shaded grey area corresponds to the W statistic, whose value is printed in the plots.

Figure 3. Top row: kNNDM with k=2 (red and blue points) for several number of clusters q. Prediction points consist of a regular grid (not shown) spanning the whole polygon. **Points were simulated for visualization purposes only.** Bottom row: NND ECDF between training locations during LOO CV ($\hat{G}_j(r)$, blue), between test and train locations during kNNDM CV ($\hat{G}^*_j(r, \mathbf{L})$, orange), and between prediction and training locations ($\hat{G}_{ij}(r)$, black) corresponding to each CV configuration in the top row.

Regarding moving Figures 1 and 3 to the appendix, we would like the reviewer and editor to consider leaving them in their current placement in section 2. We believe figures 1 and 3 are key visual aides necessary to understand the algorithm being described in that section. Figure 1 illustrates nearest neighbour distance functions as well as the W statistic that have just been described, which may not be clear from the included equations for some readers. Figure 3 complements the text describing the algorithm in that section by showing how the optimal fold configuration is chosen based on the W statistic. We hope that our justification has been convincing enough so as not to alter their placement.

7) Figure 7. A line chart might be better.

We have changed Figure 7 and Figure A4 to visualize the data as a line chart, where lines connect median values across different number of folds.

8) Additional simulation should be placed in the paper, not in a supplement. Figures 1 and 3 can instead be added to the supplement file.

We now include additional simulations as appendices in the main article. Unlike supplementary material, appendices are included in the main file and thus are accessible to the reader without having to download an additional file. Regarding the suggestion to move Figures 1 and 3 to the supplement, we would like to refer the reviewer to our response to point number *6)*.

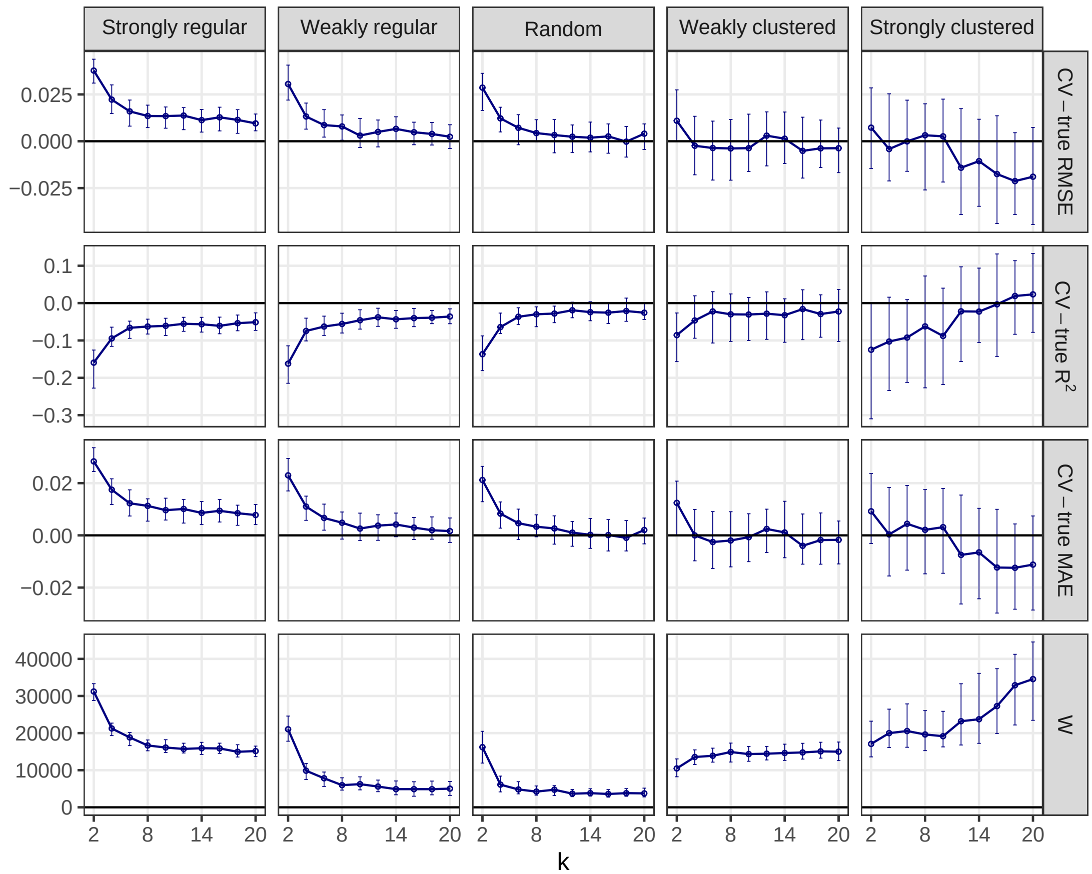

Figure 3: CV error estimates for kNNDM CV with different numbers of k (first three rows). The respective W statistic is shown in the fourth row. Points indicate median values, while error bars show the first and third quartile.

References

Milà, C., Mateu, J., Pebesma, E. & Meyer, H. (2022), 'Nearest neighbour distance matching leave-one-out cross-validation for map validation', *Methods in Ecology and Evolution* **13**(6), 1304–1316.
URL: *https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13851*

Wang, Y., Khodadadzadeh, M. & Zurita-Milla, R. (2023), 'Spatial+: A new cross-validation method to evaluate geospatial machine learning models', *International Journal of Applied Earth Observation and Geoinformation* **121**, 103364.
URL: *https://www.sciencedirect.com/science/article/pii/S1569843223001887*