Review of "Sensitivity of Gyrescale Marine Connectivity Estimates to Fine-scale Circulation" by Saeed Hariri et al. – second iteration

The authors have made significant improvements to their first submission, mainly coming from rewriting the section related to betweenness centrality. However, there are still three major issues that need to be addressed before the manuscript is suitable for publication, next to some smaller corrections and minor revisions.

Two of the main issues can in my opinion be addressed by removing the parts about betweenness centrality, since the definition of the network is vague and potentially flawed due to the use of different integration times. Moreover, with regards to the research question, I currently do not see the merit of investigating betweenness centrality when the information in the flow field is reduced to only 16 nodes connected by varying integration times. The other issue is related to open science. I elaborate on these and the smaller issues below.

Network construction

The manuscript remains vague about how the network is constructed for which betweenness centrality is computed. The authors compute Lagrangian trajectories, but they do not state how information from these trajectories is exactly translated into a network.

L180-181 states "By representing portions of the sea as nodes and the transfer probabilities between them as edges allows us to apply graph theory to the study of marine connectivity." However, it is unclear how the transfer probabilities are computed. Normally, a domain is divided into discrete bins, and transfer probabilities between bins are computed by integrating particle trajectories for a fixed timestep. The transfer probability is then the probability that a particle moves from bin *i* to bin *j* during that fixed timestep. Here, however, particle trajectories are computed with various integration times. The authors do not make clear how this is accounted for in the network construction, nor do they discuss whether this may introduce biases in the connectivity. After all, one can only make a statement on whether parts of the basin are connected <u>over a certain time period</u>. Rather than using variable integration times, the authors should use a single integration time for computing the betweenness. Moreover, the authors only treat a few sites within the basin and do not discuss what happens to particle trajectories if particles do not reach other sites. Are they discarded, or are they still accounted for in the computation of a_{ij}?

To clear up these confusions, the authors need to give an in-depth mathematical treatment of their network construction. Alternatively, the section about betweenness centrality may be omitted, as the main conclusions from the paper are supported by the transit time analysis.

Betweenness Centrality

Betweenness centrality can be a useful metric for inspecting the importance of specific sites within a network. However, I am not yet convinced that in the way it is applied in this study it is useful for investigating the sensitivity of marine connectivity to flow field resolution.

L401-403: "We utilized a site-to-site (node-to-node) metric to calculate the shortest paths, using transfer probabilities obtained from Lagrangian simulations. This matrix provides valuable information for understanding the structure of the network and can be used to inform future simulations and analyses".

This is highly dubious. Betweenness centrality gives information about how many trajectories in a flow field would pass through a certain site, thus giving information about the importance of that site, with respect to transport in the whole domain. However, here the authors only construct a network using 16 nodes, and information about flow within the domain is severely reduced to a transition matrix that only includes these sites. I do not believe the betweenness centrality computed from that can inform us about the importance of that node in the entire flow field. For example, in the hypothetical situation that many particle trajectories pass from a hypothetical site 17 through site 10 to site 11, then site 10 should have a high betweenness centrality. However, if site 17 is removed, this will decrease the betweenness centrality of site 10 significantly. To properly get information about betweenness centrality, the entire flow field should be taken into account. This is usually done in the Lagrangian flow network approach, where the entire domain is divided into bins, so the only information reduction of the flow field occurs in the time dimension.

In line with the previous major issue, I suggest to remove the section on betweenness centrality, or to switch to an integral view of the flow field, by focusing on all regions in the domain, rather than specific sites.

Apart from this: please make sure to differentiate *betweenness centrality* from *betweenness* and use the proper name throughout the manuscript (also in figure labels, e.g. Figure 10a)

Open Science

The authors now include some code for which it is entirely unclear what it does, provided as raw text (seemingly some installation script for ARIANE and one ARIANE script), while important analysis code is lacking.

I would like to ask the authors to take open science seriously. Currently, the authors do not provide readers with crucial insights into their analyses. In line with the requirements of open science, please properly include the following in a persistent repository (such as Zenodo):

- Lagrangian analysis configuration code (the ARIANE scripts)
- Hydrodynamic model configuration code (not just linking to nemo-ocean.eu, but providing more insight into this specific configuration. If this can be found in another paper, specifically mention this, so that the reader will know where to look)
- Analysis scripts that were used for calculations (for instance, where do the 39% and 8.4% in the conclusion come from?) and for plotting figures.

I specifically asked for these in the previous round of review, as the authors leave out important details that could be checked by looking at the underlying code. For example, the previous points about the network definition could be partially cleared up by including the code that was used for network construction (although the most important aspects should still be covered in the main text).

The data policy for Ocean Science is found at: https://www.ocean-science.net/policies/data_policy.html.

Other

- In my previous review, I asked the authors to briefly discuss whether parameterizing the missing dispersion in the coarse-resolution simulations may remedy the issue of the dispersion being too low, leading to longer transit times (see comment 3 from initial review). This is still missing from the discussion.
- L66: "relatively simple": here the authors minimize the contribution of previous sophisticated methods. For example, the 'hydrodynamic provinces' approach in Rossi et al. 2014 is, in my opinion, more sophisticated than computing betweenness centrality and transit times. I suggest removing these two words, to stay neutral.
- L111: "the vertical velocity is one to two orders of magnitude smaller": I don't see this from the image. Please include the standard deviation in HR and CR in order to quantify this.
- L127-128: Why is the integration time varying? For constructing a Lagrangian flow network, it is important that integration times are all the same. Otherwise, one introduces a bias into the connectivity matrix that favors some connections over others. Connectivity should be defined with respect to a certain, fixed, timescale (see earlier comment about network definition).
- L216: It is unclear to me what an improbable trajectory would be. Please elaborate.
- L496-500: This entire paragraph seems redundant. The bulk of this paragraph is in between brackets. Why? Which interdisciplinary methods are meant? I do not think ecology is always necessary for connectivity measures; it only is if ecological connectivity is studied (rather than, say, water mass connectivity).
- L507: The 39% reduction: where does it come from? It's not mentioned previously. Is this computed using all site combinations, or only using specific sites as start and end locations?
- L510: The 8.4% increase: again, where does it come from? Please show how this is computed. Is this computed using all site combinations, or only using specific sites as start and end locations?
- Figure 5: The authors should elaborate on why the CR case is less smooth than HR (in 1 to 15 and 10 to 12)? I would expect HR includes coherent structures that can trap and release particles in batches, or form blocking patterns, whereas I would instead expect these features to be smoothed out in CR, leading to a smoother spreading of travel times.
- Figure 10b: This figure is illegible. Please use the adjacency matrix representation of the network instead.
- Figure 11b: indicates the differences, but it is not clear enough which quantity is subtracted from which. Please mention this.
- In the supplementary information authors added PDF fields for particle deployments at select stations, only for HR-3D. Since the authors compare HR-3D and CR, it is important

to also show some PDF fields for CR, in order for the reader to be able to compare the cases.

Technical corrections

Line by line:

- L61: “high resolution velocity fields” → give a spatial scale
- L65: “litterarure” → literature
- L73: “relevant amount of transfers across a graph (a specific location in the domain) passes through”: please clarify this vague wording
- 519: Sabrina Speich should be abbreviated as SS instead of not abbreviated as SP