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

I thank the authors for their efforts. With the section about betweenness centrality removed, I think the manuscript is almost ready for publication.

One change that is still necessary is the inclusion of analysis code in the repository, to make their works easier to reproduce and to provide more transparency into the exact methods that were used to come to their computations (a point where, in my opinion, the authors have remained somewhat obscure).

Next to this, the authors give two clarifications in their rebuttal that I think could be included in the main manuscript.

To more easily reply to individual comments, I include my replies to the author's comments in red. I left out the portions for which no reply was warranted on my side.

I wish the authors good luck with the final implementations and thank them for their perseverance.

We thank referee #3 for his important and helpful comments. The revised text has been improved to address most of them. Our specific responses to these comments are as follows:

Comment: 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 seethe merit of investigating betweenness centrality when the information in the flow field is reduced to only 16 nodes connected by varying integration times.

Reply to two main issues related to **network construction** and **betweenness centrality**:
We have **removed** the sections on **betweenness centrality**, as suggested by the referee, but it is necessary to add some additional information about the analysis of betweenness we developed in our study:

1. We used **a fixed time step of dt = 1 hour** for all simulations. This approach ensures that each simulation runs at a consistent pace, allowing for accurate comparisons and analysis of results.

My concern was not about the time step, but about the integration time. The total amount of time for which a particle is advected will influence the connectivity. If transit times are studied, then this integration time may vary, but if betweenness centrality is computed, then the integration time should be the same for all particles and be clearly reported.

However, **the initial deployment time** of the numerical particles in our simulations **varied**. This variation is consistent with the basic principles of Lagrangian studies, where particles are tracked from their initial positions and advected by the flow for the duration of the study. In our case, we set a maximum advection time of 5 years for the particles. It should also be taken into account that in this study we are addressing the **transit time** and **not the residence time** of the numerical particles.

2. Our connectivity analysis is based on the concept of **minimum transit time**. This means that when a digital particle, let us call it particle A, leaves site "i" and arrives at site "j", the first arrival time is recorded and used for our analysis. However, if particle A were to return to site "j" at a later time, that transit time would not be used for our analysis. (See: Jönsson, B., Watson, J.: The timescales of global surface-ocean connectivity, Nat Commun. 7, 11239, https://doi.org/10.1038/ncomms1123, 2016).

To calculate the transit time between sites "i" and "j", we take the average of the minimum arrival times of all numerical particles that traveled from site "i" to site "j". By using the minimum arrival time, we can ensure that the transit time is calculated based on the fastest possible route between the two sites.

I agree with this approach.

3. The reviewer mentioned that "the authors compute Lagrangian trajectories, but they do not indicate how exactly the information from these trajectories is translated into a network" and also asked how transfer probabilities are computed? This is quite clear and we followed the approach described in our seminal paper for betweenness studies by Costa et al. (2017) (Costa A, Petrenko AA, Guizien K, Doglioli AM. On the calculation of betweenness centrality in marine connectivity studies using transfer probabilities, PLoS ONE, 12(12): e0189021, https://doi.org/10.1371/journal.pone.0189021, 2017). Specifically, transfer probability refers to the probability that a particle will move from site "i" to site "j". These probabilities (a ij) were then used as weights in our network. However, because these probabilities (a ij) tend to be very small, Costa et al. (2017) suggested taking the logarithmic inverse of the transfer probability a ij, which we also applied in our study. This approach allowed us to better analyze and visualize the resulting network. But again, it should be noted that the time step for all simulations is fixed.

I agree with the approach taken here, and agree that it is an obvious approach in general. However, I disagree that the manuscript reflected clearly that this was indeed the approach that was used. The manuscript would benefit from a simple (formulaic) explanation of how your network is defined, in section 2. The authors could add a brief paragraph under a 'network construction' subsection. Moreover, the fact that the analysis code still is not available adds to the obscurity. Either way, since the section on betweenness centrality has now been removed, this issue is resolved.

4. It is important to note that the distribution of sites in our study was carefully considered and based on a number of different analyses. The referee should consider, however, that we have two types of connectivity: a) one-way transport connectivity (i.e. movement of particles from coastal areas or river mouths to the open ocean or oceans, this is a way to study the movement of larvae or individuals from different marine populations) b) exchange connectivity which is the case of our study, and we are interested here in tracking the exchange of information between different sites distributed in the basin.

However, we would like to respond to the referee's comment about dividing the pool into smaller bins and calculating the interdependence values in each bin. This is completely contrary to the definition of betweenness centrality, since the latter attempts to identify nodes that play an important role in the exchange of information between different sites or parts of the basin.

Since the section on betweenness centrality has been removed, this issue is resolved.

To nonetheless reply to the authors' comment here on betweenness centrality: betweenness centrality indeed tries to identify those nodes within a network that are most important in the exchange of information between other nodes. This can only be done if ALL regions of the domain correspond to a bin. Otherwise, if a particle does not move from one of the specific 16 sites to a location outside of the sites, this information is lost. Having large parts of the domain *not* be linked to a site/node gives rise to biases in the representation of information flow.

5. We would like to clarify that we used the betweenness centrality approach. We did develop it. The mathematical definition and application is provided in the following references:

I am well-acquainted with the definition of betweenness centrality. However, I think that the way in which the authors applied it on the limited network (only representing a very specific regions of the

entire fluid domain, and using trajectories with different integration times) was applicable for the research question. For this, I refer back to my comments in the previous review round.

Ser-Giacomi, E., Ruggero Vasile, Emilio Hernández-García, and Cristóbal López.: Most probable paths in temporal weighted networks: An application to ocean transport, Phys. Rev. E 92, 012818, https://doi.org/10.1103/PhysRevE.92.012818, 2015.
Lindner, M., Donner, R.V.: Spatio-temporal organization of dynamics in a two-dimensional periodically driven vortex flow: A Lagrangian flow network perspective. Chaos 27.3 , https://doi.org/10.1063/1.4975126, 2017.

Reply to Comment about Open Science:
We have added this paragraph as a

Data availability
Major parts of the data and codes used in this study are available upon request by contacting the corresponding author at saeed hariri@io-warnemuende.de. Some sample data and parts of Lagrangian tools are accessible at https://doi.org/10.5281/zenodo.7954707. We encourage the use and sharing of our data and code for further research and scientific advancement. Please note that access to the codes may be subject to restrictions due to privacy or confidentiality concerns.

Unfortunately, the analysis code for creating the figures and obtaining the statistics is still missing. I do not understand why this has not been uploaded in the Zenodo-repository. The authors cite privacy or confidentiality restrictions. Can the authors elaborate which parts of the analysis code could possibly violate anybody's privacy? Since Ocean Science is a journal committed to Open Science, I think that being transparent about how the analysis has been conducted should be a crucial requirement for publication.

Reply to the other comments
Comment: 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.
We added this paragraph to discussion part of the paper
In particular, in coarse resolution simulations, the dispersion of particles is degraded. This results in longer transit times. It also limits the connection between water particles at different depths. A possible solution to overcome this problem when integrating Lagrangian trajectories using the velocity calculated in coarse resolution simulations is to parameterize the missing dispersion. Some methods have been proposed in the literature. The simplest parameterization consists in adding a random walk to the successive position of each particle, which is compatible with an advection-diffusion equation and is equivalent to a stochastic "Markovian" parameterization (Berloff & McWilliams, 2002). However, this stochastic parameterization does not reproduce adequately the small-scale ocean dynamics that involves consistency in advection (Klocker et al., 2012; Veneziani et al., 2004). Different Markov parameterizations of higher order have been proposed in an attempt to better reproduce the effect of the small-scale ocean dynamics (Berloff & McWilliams, 2002; Griffa, 1996; Rodean, 1996; Sawford, 1991). Other improved parameterizations include particle looping due to eddy coherence (Reynolds, 2002; Veneziani et al, 2004), as well as relative dispersion between different particles (Piterbarg, 2002). While these methods have been developed and applied to horizontal flows, recent developments include an isopycnal Markov-0 (Spivakovskaya et al, 2007) or shear-dependent formulation (Le Sommer, 2011) and, more recently, an isoneutral Markov-1 formulation (Reijnders et al., 2022). The latter appears to better mimic the coherent behavior of the 3D ocean dispersion at small scales. It would be interesting in future work to evaluate how such methods, applied in a Lagrangian framework, might improve the results we obtained with a coarse resolution field.

I think this is a useful discussion point and thank the authors for including it.

Comment, 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).
It is important to note that in our study, the Lagrangian time step was set for all simulations to dt= 1 hour

(dx=U.dt). In addition, we did not use a variable time step in our analysis. However, we chose to deploy the numerical particles at different initial times and the particles continue their motion for the rest of the period of the 5-years long simulation. This is an arbitrary choice, that provide more robustness in terms of the variable initial conditions. As mentioned earlier, we based our connectivity time on the minimum transit time. For example, if particle A started at site "i" and arrived at site "j", we recorded the first arrival time. However, if this particle continued to move and returned to site "j" after a certain amount of time, we did not include this time in our connectivity time estimates. This approach has been well established in previous studies such as Jönsson, B., Watson (2016), as cited in our paper. Therefore, deploying particles together at the same initial time for connectivity analysis is not correct and is not necessary.

My concern was not about the time step, but about the integration time. The total amount of time for which a particle is advected will influence the connectivity. If transit times are studied, then this integration time may vary, but if betweenness centrality is computed, then the integration time should be the same for all particles and be clearly reported.

Since the betweenness section has been removed, this issue is resolved.

Comment, 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?
These values are based on all site combinations; We simply divided the difference between the transit time (i.e. "(Transit timeHR3D)-(Transit timeCR3D)" or "(Transit timeHR3D) - (Transit timeHR2D)") by the transit timeHR3D.
I thank the authors for elaborating on this. I encourage them to include this explanation in the manuscript.

Comment, 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.
Thank you for your comment. The CR (coarse resolution) is less smooth than the HR due to the dispersion process. In the HR (high resolution) case, the simulated ocean dynamics disperses the particles more than in the CR case and the numerical particle concentration in the HR case is smoother.
In the HR (high resolution) case, the flow field is more turbulent and contains more small-scale dynamical structures than in the CR (coarse resolution) case. These small-scale features can trap and release particles in batches or form blocking patterns, resulting in high particle concentrations in some regions. However, due to the chaotic nature of the flow field, these concentrations are not maintained and the particles are eventually dispersed throughout the domain, resulting in a smoother concentration distribution.
In contrast, the CR simulation has a smoother and more predictable flow field, resulting in a more uniform dispersion of particles and a less fluctuating concentration distribution. This may result in a less smooth concentration distribution than in the HR simulation.

I thank the authors for elaborating on this. I encourage them to include this explanation in the manuscript.