Comments from Reviewer #2
I got very excited by this work reporting the possibility to simulate self-recruitment (SR) without the need of information on larval production, thereby solving a long-time reported issue for comparing models and data of larval connectivity: specifically, biophysical models are appropriate tools to simulate local retention (LR) whereas field studies estimate SR.

Response: We sincerely thank the Reviewer for reading and commenting on our work. The comments have been addressed in our point-by-point responses below.

We agree that local retention (LR) is typically more challenging to evaluate empirically, compared to self-recruitment (SR) (Lett et al., 2015). However, note that Almany et al. (2017) estimated LR using field-based parentage analysis.

1. Unfortunately, I believe that the proposed method implicitly assumes spatially homogeneous larval production and is therefore no more appropriate for simulating SR than the usual method. The assumption of homogeneous larval production is here implicit, because of the use of backward-in-time tracking, as opposed to being explicit when using forward-in-time tracking.

Response: The reviewer does not explain why they believe homogeneous larval production would occur in an implicit tracking scheme or with an implicit assumption, so it is, therefore, difficult to respond to this concern.

Specifically, it is unclear whether the reviewer is referring to an implicit numerical scheme, or an implicit assumption. An implicit numerical scheme computes the state of a system using information from both the present and the next time step. An explicit scheme computes the state of a system using only information from the present time step. Our forward/backward tracking scheme is an explicit scheme (see Shi et al 2024; Eq. 1 & 2), using only present time step velocities computed with an independent hydrodynamic model, saved as snapshots every 12 time steps. The only difference is that we reverse the sign of the velocity vector. Therefore, our model is explicit, not implicit.

An implicit assumption means something is understood or implied but not directly stated. Backtracking models, ours and in general, do not assume spatially homogeneous larval production. On the contrary, the source location and larval production are quantities that backtracking models aim to simulate, which was presented in Eq. 11 in this paper.

From a mathematical standpoint, there is no implicit requirement that the probability density distribution, $P(t_i, X; t_f, Y)$, representing the distribution of particles (larvae, in this context), be spatially homogeneous. The temporal evolution of P is governed by the Kolmogorov forward and backward equations, first introduced by Kolmogorov

(1931; English translation in Kolmogorov, 1992). These equations mathematically describe how particle distributions evolve under stochastic processes involving advection and diffusion. Specifically, the backward-in-time tracking methodology involves solving the Kolmogorov backward equation from a known final distribution at time t_f (larval settlement time) backward toward the unknown initial distribution at time t_i (larval production time). The transport processes (advection and diffusion) employed in these equations are independently determined from hydrodynamic model outputs. Therefore, the backward-in-time tracking approach does not inherently assume that the initial probability density distribution at t_i must be spatially homogeneous (i.e., a uniform distribution). Instead, the primary goal of backward tracking is explicitly to reconstruct the initial probability density distribution at time t_i based on the known distribution at time t_f.

In this research, we released larval particles at known nursery sites and track them backward in time, using an explicit scheme, to determine how many particles may have been advected from the source locations using independently evaluated hydrodynamics. There are no assumptions on how many were produced at these locations.

If backtracking models assumed that each potential hatching location produced a consistent number of larvae, then the modeled density of larval particles that settle at each location should be consistent, which means backtracking models are meaningless since the modeled results are known to be consistent. However, as presented by many studies that used backtracking models to estimate spawning/hatching grounds of fish species, larval particles settled at each location with different densities (e.g., Figure 3b in Shi et al., 2024; Rowe et al., 2022; Torrado et al., 2021; Gargano et al., 2022). This is like larval settlement not being implicitly assumed to be spatially homogeneous in forward tracking models.

In this research, we estimated SR using forward tracking simulations and assuming larval production was consistent at four source regions. The SR at regions C and D was estimated as 0.68 and 0.32, respectively. If backtracking simulations used an implicit scheme that assumed spatially homogeneous larval production, the SR computed using backtracking model should equal to the SR above. However, using the backtracking model, SR at regions C and was estimated as 0.97 and 0.18, respectively. This demonstrates that backtracking simulations have not implicitly assumed spatially homogeneous larval production.

Indeed, imagine that a given zone is particularly suitable for larval production in the reality. In the model there is no reason why more particles would be advected back to that zone "hydrodynamically" . To represent this "biological" reality of enhanced larval production the number of particles advected back to the given zone should be weighted by larval production from that zone in order to

calculate SR correctly. If not, then homogeneous larval production is implicitly assumed.

Response: We disagree with the reviewer. The hydrodynamics change spatially and temporally in the lake on a 5 min time step. Based on observations, we released particles at known first-feeding times in known nursery locations.   This represents the biological reality of the larvae. It is then up to the hydrodynamic model to track them backward in time to potential spawning locations. The hydrodynamics could also transport some or all the particles backward in time to unsuitable spawning locations, which would result in no larval production in those locations. In this study, we do not simulate biology, but rather how the flow field regulates where the particles are advected backward to, how many particles are advected there and if they are known spawning sites.

Note that the number of larvae being transported from a source zone to a settlement zone is not controlled by the larval production of the source zone, it depends on both larval production and dispersal rate. For example, source zones A, B, and C produce 200, 150, 100 larvae, respectively, but if the dispersal rate from these source zones to a settlement zone Z is 0.1, 0.6, 0.9, then the number of larvae transported from source zones to the settlement zone Z are 20, 90, 90. When performing backtracking simulations and releasing 1000 particles from zone Z, 1000×20/(20+90+90)=100 particles will be transported to zone A. But in addition, 450 and 450 particles will be transported to zones B and C. Source zone A has a highest larval production but the lowest number of particles advected back to it in backtracking simulations.

2. It is true that the quantity simulated using back-tracking will not depend on the number of particles released in a settlement area, as long as this number is high enough to obtain a statistically meaningful value. However that simulated quantity will be comparable to SR as assessed in the field only if larval production is spatially homogeneous.

Response: We disagree with the reviewer. In forward tracking simulations, SR does depend on the larval production and dispersal rates, as presented in Eq. 3. However, in backtracking simulations, SR depends only on the settlement rate, with no relationship to larval production, with larval settlement as presented in Eq. 7. The settlement rate is terminology specific to backtracking simulations, defined as the ratio of particles that settle locally at a location divided by the total number of particles released from that location. This is similar to forward tracking simulations, where LR only depends on the dispersal rate, while in backtracking simulations, LR depends on both larval settlement and the settlement rate. However, we cannot deduce that spatially homogeneous larval settlement has been implicitly assumed in forward tracking simulations. The factor that regulates both the dispersal rate and settlement rate is the hydrodynamics, which controls the movement trajectories of larval particles.

Moreover, the hydrodynamic model (AEM3D) has been shown to accurately simulate the flow field in Lake Erie. The AEM3D model, and its non-parallel predecessor the Estuary and Lake Computer Model (ELCOM), have been applied to hindcast the thermal structure (León et al., 2005), internal wave dynamics (Valipour et al., 2015), surface wave / sediment transport (Lin et al., 2021), nutrient and chlorophyll-a distributions (Leon et al., 2011), seasonal succession of phytoplankton groups (Wang et al., 2024); and to forecast storm surge and upwelling/downwelling events (Lin et al., 2022). In our prior work, we applied the model, to Lake Erie, to backtrack the present Lake Whitefish larval observations and determine hatching locations (Shi et al., 2024). As we have discussed, the biggest drawback with using hydrodynamic models for backtracking, is that they are diffusive backward-in-time, rather than being convergent. Therefore, comparisons with results from parentage analysis should be undertaken to further verify the validity of SR when computed using backtracking models.

3. In conclusion I sadly do not see how the proposed method based on back-tracking is an improvement from the one based on forward-tracking. Back-tracking is an approach that may be more efficient to use computationally than forward-tracking when focusing on particles origin rather than destination but both approaches should give the same results.

Response: In the present research, we show that, using both theoretical arguments and numerical simulations of Lake Whitefish (*Coregonus clupeaformis*) larval observations in Lake Erie, forward tracking cannot compute SR accurately unless realistic larval production is released from all potential source locations. However, backtracking simulations can easily compute SR with no knowledge of larval production. This is shown theoretically in the comparison of Eq. (3) and Eq. (7).

We employed various strategies for releasing larval particles in the forward tracking simulations, including releasing a random number, a constant number, or a number of particles proportional to the location area or the larval production. In all cases, the numerical simulations of Lake Whitefish (*Coregonus clupeaformis*) in forward tracking cannot compute SR unless realistic larval production values were released from all potential source locations. While in contrast, tracking larval particles backward from the settlement location was shown to be a straightforward approach for computing SR. Similarly, backtracking models cannot compute LR unless a realistic number of larval recruits are released from all settlement locations.

4. lines 40-48 I found this paragraph confusing as it looks like the authors decided to define local retention (LR) and theoretical local retention (TLR) oppositely to what was previously used (Hogan et al. 2012, Burgess et al. 2014, Lett et al. 2015).

Response: Yes, Hogan et al. (2012) and Lett et al. (2015) proposed a different measure to assess local retention.

- Botsford et al. (2009) defined LR as: $\dfrac{\text{locally produced settlement}}{\text{total number of larvae locally released}}$

- Hogan et al. (2012) defined LR as: $\dfrac{\text{locally produced settlement}}{\text{total number of settlers locally released}}$

which was termed relative local retention (RLR) by Lett et al. (2015).

- More recently, Almany et al. (2017) defined LR as:
$$\dfrac{\text{number of larvae that left reef X and survived and settled on reef X}}{\text{number of larvae that left reef X and survived and settled somewhere else}}$$

which corresponds to the RLR, defined by Lett et al. (2015).

Consequently, there are two definitions of local retention in the literature. The difference between these two terms lies in the denominator. Hogan and Almany's LR, includes only successfully settled larvae, Botsford and Lett's LR, encompasses both successful and unsuccessful settlers (those settling in unsuitable nursery locations). In field studies, sampled fish are typically survivors. To ensure comparability with empirical results, we adopt Hogan's LR, as local retention, while Botsford's LR, is termed "theoretical" local retention (TLR). TLR provides a minimum value of LR as stated by Shi et al. (2024), as the denominator contains both successfully and unsuccessfully settlers.

5. 49: Parentage analysis and/or larval tagging is widely used to estimate SR, not the other quantities, to my knowledge.

Response: We agree. Almany et al. (2017) used parentage analysis to estimate local retention. We have rewritten this sentence as:

Parentage analysis and/or larval tagging have been widely used to estimate LR and SR (Almany et al., 2017; Jones et al., 1999; Pinsky et al., 2012; D'aloia et al., 2013; Lett et al., 2015; Planes et al., 2009).

6. 69-72: True, but as explained above I believe this issue remains when using back-tracking.

Response: As we explained above, we have not no backtracking models implicitly assumed spatially homogeneous larval production.

7. 87-91: Same here.

Response: As we explained above, we have not implicitly assumed spatially homogeneous larval production.

Response: As we explained above, no backtracking models implicitly assumed spatially homogeneous larval production.

8. 103-4: It has already been shown that LR and TLR are independent on larval production (Burgess et al. 2014, Lett et al. 2015) and there is no need of simulations to prove this.

Response: That LR and TLR are independent of larval production is not the conclusion of this research. We used the mathematical methods to derive LR, TLR and SR. The derived LR and TLR agree with previous research (Lett et al., 2015), which verifies the accuracy of our derived results, that SR depends on larval production but is independent on the number of recruits at a settlement location.

9. Equ. (1) this quantity is defined as LR in Lett et al. (2015), why change?

Response: As we explained above in comment #4, there are two definitions of local retention; the difference between these two definitions lies in the denominator. Hogan and Almany's LR, includes only successfully settled larvae, Botsford and Lett's LR, encompasses both successful and unsuccessful settlers (those settling in unsuitable nursery locations). In field studies, sampled fish are typically survivors. To ensure comparability with empirical results, we adopt Hogan's LR, as local retention, while Botsford's LR, is termed "theoretical" local retention (TLR). TLR provides a minimum value of LR as stated by Shi et al. (2024), as the denominator contains both successfully and unsuccessfully settlers.

10. Equ. (2) this quantity is defined as RLR in Lett et al. (2015), why change?

Response: As indicated in response (9) we have chosen to pursue alternate terminology.

11. 167-8 This is true, however as explained above that simulated quantity will be SR as assessed in the field only if larval production is spatially homogeneous.

Response: The question about spatial homogeneity was addressed above in responses (1) to (3).

291-2 I don't understand why SR values estimated from back-tracking and forward-tracking would be different. I don't agree that one is "correct" and the other "erroneous". They should be similar as both are based on the same assumption of homogeneous larval production, as explained above. There may be numerical reasons for the reported differences, or other technical reasons that could be explored.

Response: This was addressed at length above in responses (1) to (3). That SR values estimated from backtracking and forward-tracking are different is the key finding of this research. We have shown this outcome both theoretically and numerically. Therefore, it is not for technical reasons that SR from forward tracking is different from backtracking.

When estimating SR using forward tracking models, SR varies with the number of larval particles released from each source location. Notably, the value of SR can be artificially manipulated to be any value from 0 to 1, by adjusting the quantity of larvae released. The correct SR from forward tracking simulations can only be obtained when realistic larval production is used. Consequently, SR from a backtracking model will only match that from forward tracking if realistic larval production numbers are used in the forward tracking simulations.

References

Almany, G. R., Planes, S., Thorrold, S. R., Berumen, M. L., Bode, M., Saenz-Agudelo, P., Bonin, M. C., Frisch, A. J., Harrison, H. B., Messmer, V., Nanninga, G. B., Priest, M. A., Srinivasan, M., Sinclair-Taylor, T., Williamson, D. H., and Jones, G. P.: Larval fish dispersal in a coral-reef seascape, *Nat Ecol Evol*, 1, 148, 10.1038/s41559-017-0148, 2017.

Botsford, L. W., White, J. W., Coffroth, M. A., Paris, C. B., Planes, S., Shearer, T. L., Thorrold, S. R., and Jones, G. P.: Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs, *Coral Reefs,* 28, 327-337, 10.1007/s00338-009-0466-z, 2009.

Burgess, S. C., Nickols, K. J., Griesemer, C. D., Barnett, L. A., Dedrick, A. G., Satterthwaite, E. V., Yamane, L., Morgan, S. G., White, J. W., and Botsford, L. W.: Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design, *Ecol Appl,* 24, 257-270, 10.1890/13-0710.1, 2014.

Hogan, J. D., Thiessen, R. J., Sale, P. F., and Heath, D. D.: Local retention, dispersal and fluctuating connectivity among populations of a coral reef fish*, Oecologia*, 168, 61-71, 10.1007/s00442-011-2058-1, 2012.

Kolmogorov, A. N. (1992). On analytical methods in probability theory. In A. N. Shiryayev (Ed.), Selected works of A. N. Kolmogorov: Volume II-Probability theory and mathematical statistics (pp. 62–108). Springer.

Lett, C., Nguyen-Huu, T., Cuif, M., Saenz-Agudelo, P., and Kaplan, D. M.: Linking local retention, self-recruitment, and persistence in marine metapopulations, *Ecology,* 96, 2236-2244, 10.1890/14-1305.1, 2015.

Rowe, M. D., Prendergast, S. E., Alofs, K. M., Bunnell, D. B., Rutherford, E. S., and Anderson, E. J.: Predicting larval alewife transport in Lake Michigan using hydrodynamic and Lagrangian particle dispersion models, *Limnology and Oceanography,* 67, 2042-2058, 10.1002/lno.12186, 2022.

Shi, W., Boegman, L., Shan, S., Zhao, Y., Ackerman, J. D., Amidon, Z., Jabbari, A.,

and Roseman, E.: A Larval "Recruitment Kernel" to Predict Hatching Locations and Quantify Recruitment Patterns, *Water Resources Research,* 60, 10.1029/2023wr036099, 2024.

Torrado, H., Mourre, B., Raventos, N., Carreras, C., Tintoré, J., Pascual, M., and Macpherson, E.: Impact of individual early life traits in larval dispersal: A multispecies approach using backtracking models, *Progress in Oceanography*, 192, 10.1016/j.pocean.2021.102518, 2021.

Gargano, F., Garofalo, G., Quattrocchi, F., and Fiorentino, F.: Where do recruits come from? Backward Lagrangian simulation for the deep water rose shrimps in the Central Mediterranean Sea, *Fisheries Oceanography*, 31, 369-383, 10.1111/fog.12582, 2022.

Leon, L. F., Smith, R. E. H., Hipsey, M. R., Bocaniov, S. A., Higgins, S. N., Hecky, R. E., Antenucci, J. P., Imberger, J. A., and Guildford, S. J.: Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie, *Journal of Great Lakes Research,* 37, 41-53, 10.1016/j.jglr.2010.12.007, 2011.

León, L. F., Imberger, J., Smith, R. E. H., Hecky, R. E., Lam, D. C. L., and Schertzer, W. M.: Modeling as a tool for nutrient management in Lake Erie: a hydrodynamics study, *Journal of Great Lakes Research,* 31, 309-318 2005.

Lin, S. Q., Boegman, L., Shan, S. L., and Mulligan, R.: An automatic lake-model application using near-real-time data forcing: development of an operational forecast workflow (COASTLINES) for Lake Erie, *Geoscientific Model Development,* 15, 1331-1353, 10.5194/gmd-15-1331-2022, 2022.

Lin, S. Q., Boegman, L., Valipour, R., Bouffard, D., Ackerman, J. D., and Zhao, Y.: Three-dimensional modeling of sediment resuspension in a large shallow lake, *Journal of Great Lakes Research,* 47, 970-984, 10.1016/j.jglr.2021.04.014, 2021.

Valipour, R., Bouffard, D., Boegman, L., and Rao, Y. R.: Near-inertial waves in Lake Erie, Limnology and Oceanography, 60, 1522-1535, 10.1002/lno.10114, 2015.

Wang, Q., Boegman, L., Nakhaei, N., and Ackerman, J. D.: Multi-year three-dimensional simulation of seasonal variation in phytoplankton species composition in a large shallow lake, *Ocean Modelling,* 189, 10.1016/j.ocemod.2024.102374, 2024.