This manuscript introduces a new numerical method to solve the cyclogeostrophic balance equation, which is useful for estimating velocity from SSH beyond geostrophic balance. The numerical method phrases the PDE problem as a minimization of the residual and solves it using gradient-based optimization methods, leveraging the capability of `JAX`. The method is shown to solve the cyclogeostrophic balance equation robustly with minimal noisy artifacts, and the result matches physical expectations. The authors also compared with drifter velocities and showed that the cyclogeostrophic velocity reduces bias for the most energetic regions of the ocean. However, they also note that the resolution of the SSH product also influences the results.

I have used the authors' package `jaxparrow`, which has existed for a year now on GitHub. From my experience with the package, I agree with the authors' claim that it is a superior way to solve the cyclogeostrophic balance equation. The implementation using `JAX` is high quality (though I am not an expert in open-source software) and should be the new standard for applying the cyclogeostrophic balance diagnostic of velocity from SSH. Cyclogeostrophic balance has received a revival of interest due to the capability of high-res SSH observation of SWOT [Archer et al., 2025, Tranchant et al., 2025, Tchonang et al., 2025] (additionally, Zhang and Callies [2025, (7)] is essentially the first step of (3)). Unfortunately, `jaxparrow` was not used in any of these papers. With this paper, `jaxparrow` will be more widely known to the community. This will improve standardization between implementations of cyclogeostrophic balance and promote cross-validation of results. For this, I recommend the paper for publication. I also recommend that the authors submit another paper to JOSS, where they have a culture of working with the authors to improve the code quality of an open-source package.

With the praise in mind, here are some suggestions that I hope will improve the presentation of the paper.

1. The authors name their method "variational", based on calling (4) the "variational form" of the cyclogeostrophic balance equation. I would recommend changing this.
   - "Variational" is a loaded mathematical word. For a nonlinear PDE like (2), the variational formulation does not straightforwardly lead to the minimization of residuals (4). Therefore, naming the method "variational" is not very informative of the underlying principle and can be confusing.
   - I would recommend a different name, say "least-square" or "minimization/optimization based".
2. I would also recommend that the author call the traditional method written in (3) "fixed-point". The minimization procedure is also iterative. Therefore, the name "iterative" can be applied to both methods, and can thus be confusing.
   Additionally, the numerical divergence behavior of the traditional method can be understood through classic theory about the behavior of the fixed-point method and contraction mapping.
3. Before (1), I think a quick derivation of the cyclogeostrophic balance equation is helpful.
   - It would show that (1) is based on ignoring the time partial in the horizontal momentum equation.
   - This is not a very well-grounded approximation, especially in the submesoscale, where the time partial term can be equal in magnitude to the advective and the Coriolis terms.
   - I view this as not an issue when the focus is on solving the cyclogeostrophic balance equation. But this is important when comparing with drifter data. Cyclogeostrophic

balance is not obviously better than geostrophic balance in the submesoscale, since it includes one but ignores another term of equal asymptotic order.

4. §2.3: I cannot find a description of the minimization procedure.

   - I do note that the authors state more exploration can be done (Line 271), but I want to know what is currently applied.
   - From looking at the code, the authors used SGD in `Optax`, which I am not familiar with the details. What is the difference from GD in this case, since there is no minibatching?
   - Some math formula that writes out the gradient descent iteration might be helpful when comparing to (3). It is also an opportunity to elucidate how `JAX` automates the evaluation of the gradient (adjoint) needed in this iteration. This is helpful for an oceanographic audience that is not familiar with these techniques.

5. From my point of view, there are two goals this paper tries to achieve. They are both referred to as "accurate" velocity, which is confusing.

   (a) Showcase a new numerical method for solving the cyclogeostrophic balance equation. It is better than the traditional method

   (b) Study if cyclogeostrophic balance improves upon geostrophic balance for velocity inversion from SSH.

   Then I think the paper has shown (a) convincingly. And I think this should be the main point of the paper, as also stated in the paragraph starting at Line 57.

   - One recommendation is to show the SWOT figures in Appendix B in the main text to make the point that cyclogeostrophic balance only depends on local SSH (cf. SQG-based method). Therefore, it can be applied to nontrivial but 2D observation geometry like SWOT swath. It will make mentioning of SWOT appear less gratuitous.

6. Point (b) is very complex, and I think it is perpendicular to goal (a). I would like the authors to consider removing goal (b) entirely from this paper and addressing it in a separate piece. But I understand this requires major changes and delays. I leave the authors to decide. With that in mind, I have some comments and questions

   - In recent works Tranchant et al. [2025], Tchonang et al. [2025] have shown in the SWOT context that cyclogeostrophic balance's improvement depends on the region and season. This is consistent with the result in Figure 5.
     The authors should take these papers into account when making the statement "These results suggest that cyclogeostrophic corrections will become increasingly relevant as SSH products achieve higher effective resolution" in Line 257.
   - Figure 5 uses 6-hourly interpolated surface current velocity. This might include NIW signal that the cyclogeostrophic balance will not be able to get (cf. comment 3 above). In general, internal waves and tidal signals might contribute to the error statistics. This fact needs to be mentioned, and better, quantified.
   - When answering this question, model output is your friend. I would like to see Figure A1 plotted as the error of the diagnosed velocity.
     Question: What is the "observed" SSH here? Is there a low-pass filter and added noise to match observations?

A bigger task is to do this for a bigger region and get error statistics like in Figure 4 and 5. The authors can decide if they want to take on this task in the revisions.

7. §3.2.2. The error metric can be improved. I agree with Referee #1 on having concerns here.
    - A better metric that is commonly used is mean squared error.
    - Then one can talk about the estimation error of the MSE, where the concept of the standard error can be applied. Right now the term "standard error" is applied erroneously. There should be a division by $\sqrt{N}$ if the author meant the standard deviation of the sample mean.

Minor comments

1. Title of §2. Resolution is a loaded word. "The solution to" might be better.
2. In (2), $\boldsymbol{u}_g$ is not defined. I would recommend not using $\boldsymbol{u}_g$ here, instead use the rotated gradient of η.
3. The norm in (5) is better called the little ℓ_2 norm, which applies to vectors. The big L_2 norm is usually applied to functions. (4) is an L_2 norm.
4. Line 100, 149: make sure to state the *numerical* divergence problem. "Divergence" can be confused with velocity divergence.
5. Line 102: "the minimum of J yields a smooth velocity field that does not exhibit the unrealistic features". It is too early in the paper to make this statement. The smoothness of the solution is empirical and can only be inferred from looking at the solution fields in Figure 2. It probably also depends on the SSH data's effective resolution. At this point, there is no reason from the setup that $\boldsymbol{u}_{cg}$ should be smooth. One can achieve this by using some smoothness regularization, but this is not done here.
6. Line 196: "geostrophy can be a *coarse* approximation". Coarse is a loaded word. It could mean it works better for course resolution data. Do you mean that? I don't think Figure 1 is enough to show this.

References

M. Archer, J. Wang, P. Klein, G. Dibarboure, and L.-L. Fu. Wide-swath satellite altimetry unveils global submesoscale ocean dynamics. *Nature*, 640(8059):691–696, Apr. 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-08722-8.

B. Tchonang, J. Wang, A. F. Waterhouse, A. Lucas, C. G. Griffin, M. R. Archer, L. Kachelein, M. Lankhorst, J. Sevadjian, and L.-L. Fu. SWOT Geostrophic Velocity Validation against in-situ measurements in the California Current, Apr. 2025.

Y.-T. Tranchant, B. Legresy, A. Foppert, B. Pena-Molino, and H. Phillips. SWOT Reveals Fine-Scale Balanced Motions Driving Near-Surface Currents and Dispersion in the Antarctic Circumpolar Current. *Earth and Space Science*, 12(8):e2025EA004248, 2025. ISSN 2333-5084. doi: 10.1029/2025EA004248.

X. Zhang and J. Callies. Assessing submesoscale sea surface height signals from the SWOT mission, May 2025.