

Dear Editor,

Thank you for giving me the opportunity to review the manuscript by Reg W. Dowse, Shane R. Keating, and Moninya Roughan, entitled "*A Geometric Framework for Mapping Ocean Eddies Using Elliptical Streamfunction Parameterisation*".

The manuscript introduces the Elliptical Streamfunction Parametrisation (ESP) and develops three implementations - SOLO, DOPPIO, and LATTE - intended to provide a physically interpretable, low-dimensional framework for single transects, orthogonal transects, gridded velocity fields, and unstructured velocity observations. The central idea is clear. The constraint $\det(Q) = 1$ helps separate geometric deformation from radial scale, and the three sampling configurations have potential practical value. The archived code and data, the detailed derivations, and the multi-platform applications are important strengths of the study.

However, the current manuscript presents the demonstrations as sufficient validation of robustness, generality, threshold-free operation, and three-dimensional reconstruction. The idealized experiments mainly use velocity fields generated by the same model family assumed in the inversion. The real-data applications rely largely on in-sample goodness of fit and visual overlays. Fixed spatial and temporal criteria, background flow, observation weighting across platforms, parameter identifiability, and temporal aliasing have not been quantified. Most importantly, the claimed capability to reconstruct three-dimensional structure and vertical tilt is not demonstrated in the results. These issues affect the main scientific conclusions and cannot be resolved through wording or presentation changes alone.

I therefore recommend major revision. The manuscript would be substantially strengthened by systematic validation under realistic departures from the assumed model, a clear definition of the method's operating limits and failure conditions, quantitative uncertainty assessment for the real-data applications, and a narrower treatment of the three-dimensional claims unless new supporting results are added.

Major Comments:

1. The Validation Design Does Not Support the Claims of Robustness and Generality
The idealized experiments in Sections 3.1.1, 3.2.1, and 3.3.1 vary transect position, transect length, and sampling geometry, but the control velocity fields are generated from the same elliptical Gaussian streamfunction used in the inversion. The

manuscript does not state that these experiments include observational noise, spatially correlated errors, or model misspecification. The results therefore show that the derivations and implementation are self-consistent when the model assumptions are exactly satisfied; they do not establish the repeatedly stated claims of robustness, generality, or applicability to realistic non-ideal vortices.

The real-data cases do not close this validation gap because the true parameters are unknown and the reported goodness of fit is calculated from the same observations used to estimate the parameters. The existing evidence is better described as a proof of concept than as an assessment of the method's performance envelope.

I recommend that the authors:

- Add ensemble or Monte Carlo experiments that include realistic observation noise, spatially correlated errors, and departures from the elliptical Gaussian model, rather than relying only on single deterministic examples.
- Separate any experiments used to tune algorithm settings from independent validation experiments.
- If such validation cannot be added, narrow statements such as "robust," "general," and "applicable to arbitrary sampling patterns" to claims that are explicitly limited to the tested conditions.

2. The Threshold-Free Description Conflicts with Fixed Algorithmic Criteria

The abstract, introduction, and conclusion describe ESPresso as a "threshold-free" framework, yet the implementation repeatedly uses fixed criteria. SOLO, DOPPIO, and LATTE use a 30 km inner-core search radius and a 100 km outer-core radius. Temporal tracking stops when the daily displacement exceeds 100 km or when the sign of vorticity changes. The drifter analysis uses a plus or minus 1 day time window. SOLO also fixes a cubic polynomial representation and restricts the fit to observations near the point of closest approach. Section 2.2 notes that the inner-core sample can be adjusted after examining goodness of fit, but it does not provide a reproducible decision rule.

These choices determine which observations enter the objective function, whether the peak velocity is sampled, and whether tracking continues. They are therefore substantive algorithmic settings, not incidental implementation details. Fixed kilometre scales are also inconsistent with the range of eddy sizes discussed in the manuscript: R_c is below 65 km at some times and above 100 km at others, while the

idealized control value is 85 km. For different resolutions, regions, or smaller eddies, 30 km and 100 km may represent very different dynamical regimes. Without a sensitivity analysis, the term "threshold-free" understates the method's dependence on user-selected settings.

I recommend that the authors:

- Remove or carefully qualify the term "threshold-free," and list all user-defined settings, default values, and stopping conditions in one place.
- Where possible, replace fixed dimensional criteria with nondimensional or adaptive criteria tied to an initial scale estimate, grid spacing, or deformation radius.
- Provide sensitivity tests, failure flags for insufficient data or unstable estimates, and recommended operating ranges.

3. Unmodelled Background Flow and Shear May Systematically Bias the Eddy Parameters

Equations (6) and (8) attribute the observed velocity entirely to a single nondivergent elliptical Gaussian vortex. The model does not include a uniform background flow, linear shear, or another large-scale component, and the manuscript does not clearly state whether such components are removed from the real data before fitting. This issue is particularly important in the East Australian Current region, where satellite, numerical-model, and S-ADCP velocities may contain background current, nearby structures, tides, inertial motions, and near-surface ageostrophic components. Section 3.2.2 attributes a local contour displacement to external flow but does not quantify the resulting parameter bias.

The potential bias is evident from the SOLO relation in Equation (16), $y_0 = U(x_0)/V'(x_0)$. Any unremoved uniform along-track velocity directly changes $U(x_0)$ and therefore changes the inferred cross-track distance y_0 . Linear shear enters the velocity-gradient matrix and may be interpreted as part of Ω or Q . DOPPIO and LATTE may likewise misinterpret background shear as eddy deformation or vorticity. A high in-sample R^2 does not exclude this bias because a low-dimensional elliptical model may still fit the dominant part of the velocity field.

I recommend that the authors:

- Jointly estimate at least a constant background velocity in the forward model and assess whether linear background shear is also required, or provide a physically justified and reproducible background-removal procedure.
- Add idealized tests with background flows, shear, and radial components of different magnitudes and directions, and report the resulting parameter biases.
- Show vector residual maps and the spatial structure of residuals for the real-data cases. If some components are not identifiable, state the identifiability conditions and limit the physical interpretation accordingly.

4. Multi-Platform Fusion Lacks Observation-Error Weighting and Scale Harmonization

LATTE places satellite-derived geostrophic velocities at 1/5 degree resolution, S-ADCP velocities from approximately 37 m depth, and drifter velocities from approximately 15 m depth in the same least-squares objective function. The manuscript does not describe platform-specific weights, observation-error covariance, spatial or temporal autocorrelation, or effective sample size. The three platforms also do not measure exactly the same quantity: the satellite product mainly represents surface geostrophic flow, whereas the in situ observations may include ageostrophic motion and vertical shear.

If all points receive equal weight, the platform with more observations or denser sampling will exert greater influence on the objective function, even when those samples do not provide an equivalent amount of independent information. Adjacent grid cells, consecutive ADCP records, and successive observations from the same drifter are correlated. The smoother multi-platform estimates may therefore reflect the numerical or spatial dominance of one platform rather than genuine noise reduction or stabilization. Without a treatment of systematic differences among platforms, the fused estimates and their physical meaning are difficult to assess.

I recommend that the authors:

- Report the sample count, time span, representative depth, and uncertainty estimate for each platform in every fit.
- Use weighting or generalized least squares based on observation uncertainty and correlation scales, or at minimum spatially and temporally thin each platform so that raw point count does not determine its influence.

- Add sensitivity tests for platform weights, leave-one-platform-out validation, and tests for inter-platform bias. Explain how satellite geostrophic velocity and in situ total velocity are placed in a physically comparable framework.

5. Treating Asynchronous Observations as an Instantaneous Flow Field Causes Temporal Aliasing

The SOLO S-ADCP transect spans approximately 24 h, the S-ADCP and satellite LATTE case combines observations collected over approximately 4.5 d, and the drifter and satellite LATTE case uses a plus or minus 1 d window. These observations are then treated as spatial samples of one static ellipse at a single time. Interpolating the satellite field to the mean observation time supplies a reference snapshot, but it does not remove the fact that measurements along the ship track were collected at different times.

If the eddy translates, rotates, deforms, or changes intensity within the sampling window, temporal evolution is aliased into spatial structure. Differences between the early and late parts of a transect may be interpreted as deformation, scale, or centre displacement. If profiles at different depths are not sampled simultaneously, temporal evolution may also be mistaken for vertical tilt. The discussion acknowledges this assumption, but the manuscript does not use the satellite time series to quantify displacement or parameter changes over 1 d, 2 d, or 4.5 d, and it does not provide a time-window sensitivity analysis.

I recommend that the authors:

- Use an independent daily eddy-centre trajectory to transform observations into a co-moving coordinate system, and where possible interpolate or correct velocities to a common reference time.
- Compare several time windows, for example plus or minus 6 h, 12 h, 24 h, and 48 h, and report the trade-off between parameter stability and available sample size.
- For long S-ADCP combinations, consider time-dependent parameters or segmented fits, and propagate uncertainty from temporal aliasing into the final parameter estimates.

6. The Main Claims About Three-Dimensional Reconstruction and Vertical Tilt Are Not Supported by the Results

The abstract and introduction frame the method as a tool for mapping three-dimensional eddy structure. The discussion states that the study has demonstrated ESPresso's ability to estimate eddy tilt, and the conclusion lists three-dimensional evolution and vertical tilt among the properties that can be estimated consistently. However, the results show surface or near-surface two-dimensional velocity fields and their time series. They do not present depth-resolved ESP parameters, displacement of the eddy centre with depth, tilt angle, vertical-continuity tests, or a synthetic experiment with a known tilt.

This issue should be addressed in one of two ways:

- If three-dimensional reconstruction and tilt are retained as major contributions, add a three-dimensional synthetic experiment with known vertical tilt and a depth-resolved observational or model case. Report $x_0(z)$, $y_0(z)$, $R_c(z)$, $\alpha(z)$, principal-axis orientation, and uncertainty, and assess the effects of depth-dependent sampling and temporal offsets.
- If these results cannot be added, describe three-dimensional reconstruction and tilt consistently as potential extensions, and remove statements that claim the capability has been demonstrated in this study.

Minor Comments:

- Page 15, Equation (25): $U_2(y_2)$ appears although y_2 is not defined. Please correct the symbol so that it is consistent with the derivation, and check Equations (25) and (26) term by term rather than retaining a known error in the text.
- Page 15 states that the horizontal position errors are below 1.1 km and the R_c errors are below 3.2 km. In Table 2, however, the absolute R_c error for EP3 is 5.457802 km. If "horizontal position error" means Euclidean distance, the EP3 error is approximately 1.38 km and also exceeds 1.1 km. Please state whether the reported quantity is a coordinate-wise error or Euclidean distance, and make the text and table consistent.
- Page 24 cites "Figure 7, Table 4" when discussing the DOPPIO idealized validation. This appears to require Table 2; Table 4 is the later comparison of relative errors between DOPPIO and LATTE.

- Page 19 states that the errors in x_0 , y_0 , ω , and α are small for Ex1-Ex4, but Table 3 reports an aspect-ratio error of 0.171844 for Ex4, approximately 11.2% of the control value. Please define a quantitative criterion for "small" and moderate the statement accordingly.
- Please explain how the 95% prediction intervals in Figures 6 and 12 were calculated. State whether they are confidence intervals for the mean or prediction intervals for new observations, specify the assumed error distribution, and explain whether spatial and temporal correlation among velocity samples was considered.
- Section 2.3.1 gives one approximately 24 h transect and a complete S-ADCP record from 17 October to 2 November 2023, whereas Section 3.3.2 states that the S-ADCP observations span approximately 4.5 d. For each application, report the retained start and end times, cumulative observation duration, number of data points, and selection rules. Please also explain how 2023-10-19T16:30 was calculated as the mean observation time.
- Please define the transformation from longitude and latitude to the local Cartesian coordinates X and Y , including the projection, reference latitude, Earth radius, and positive directions. Errors in eddy centre and radius expressed in kilometres depend on this transformation; code availability does not replace a basic definition in the manuscript.
- R_c is the Gaussian characteristic radius, whereas the maximum normalized tangential velocity occurs at $R_c/\sqrt{2}$. The text and captions sometimes use R_c , core radius, eddy scale, and the contour of maximum tangential velocity interchangeably. Please define these terms at first use and apply them consistently.
- Tables 1-3 report six decimal places and retain floating-point residuals of order 10^{-14} km, far beyond the resolution and precision supported by the input data. Please round values according to observational resolution and parameter uncertainty, and replace numerical round-off noise with zero.
- Please state the sign conventions for the local coordinates and vorticity, including the positive x and y directions, the orientation of J , and the signs of Ω and ω for a Southern Hemisphere cyclone. These definitions are necessary for readers to reproduce and interpret the results in other regions.

- Q contains both aspect ratio and principal-axis orientation, but the result time series mainly report alpha. If orientation is part of the claimed "complete geometry," report its error or temporal evolution. When alpha is close to 1 and orientation is poorly defined, use a non-reporting or low-confidence criterion.
- Page 25 states that DOPPIO can in principle be generalized to transects intersecting at arbitrary angles, but no derivation, algorithm, or validation is provided. Please identify this as a future extension or add the equations and numerical tests for a general intersection angle.
- The code and data availability section should provide the fixed version, release tag, or commit hash corresponding to this manuscript, together with an environment file and a reproducible entry point from scripts to figures. A permanent DOI is helpful, but readers must still be able to identify the exact version used to generate the reported results.
- Several references contain a duplicated DOI prefix, for example <https://doi.org/https://doi.org>, and some article numbers contain inserted spaces. Please check DOI links, journal names, volume and page information, and article-number formatting throughout the reference list.

In summary, I recommend major revision and re-evaluation. The highest priorities are to strengthen the validation design and uncertainty assessment for the real-data applications; address background flow, multi-platform weighting, temporal aliasing, and identifiability; and revise the claims about threshold-free operation, general robustness, three-dimensional reconstruction, and vertical tilt in line with the evidence. Once these issues have been addressed substantively, the methodological contribution of the manuscript can be assessed more reliably. I hope these comments will help the authors improve the manuscript.

Sincerely,

Reviewer