Review of "Near-Real-Time Assimilation of Satellite-Derived Ocean Surface Currents Using a Multi-Model Ensemble Kalman Filter", by Baig et al.

This manuscript presents a technically sophisticated and timely framework for assimilating satellite-derived surface currents into a multi-model Ensemble Kalman Filter, addressing an important problem in ocean forecasting with clear potential operational relevance. The methodology is carefully constructed and supported by extensive sensitivity analyses. However, several aspects of the study raise substantive concerns that limit confidence in the results as currently presented. These include imprecise framing of the motivation, inconsistencies between the assimilated observations and the ensemble representation of the flow, the treatment of missing data, and the lack of independence in parts of the validation. In addition, the level of agreement with independent observations appears more moderate than implied, and the reported degradation of performance with increasing observational density, together with persistent biases, suggests that key assumptions underlying the data assimilation framework may not be satisfied in practice. For these reasons, significant clarification and revision are required before the manuscript can be considered for publication.

Detailed comments

1. A first issue relates to the motivation and the description of existing operational systems. At L2, the manuscript states: "Most operational systems primarily assimilate altimeter-derived geostrophic currents." I am not aware of operational ocean forecasting systems that directly assimilate surface geostrophic currents. In most cases, systems assimilate sea surface height and other scalar variables, from which geostrophic velocities may be inferred diagnostically, rather than assimilated explicitly.

   A similar point arises at L35-37, where the manuscript again suggests that operational systems assimilate geostrophic velocities derived from sea surface height, citing Le Traon et al. (2019). That reference does not present examples of such assimilation, but rather notes parenthetically that geostrophic velocities can be derived from sea level. As currently written, this aspect of the manuscript risks giving a misleading impression of operational practice. This does not undermine the broader motivation of the study. The authors can make a strong argument that direct assimilation of observed or derived velocity fields may complement existing approaches by representing ageostrophic contributions. Reframing the motivation along these lines would strengthen the manuscript.

Response: We agree with the reviewer that operational ocean forecasting systems generally assimilate sea surface height, sea surface temperature, and temperature and salinity observations, while geostrophic currents are implicitly contained in the analyzed sea surface height rather than assimilated directly as surface velocity vectors. The abstract and the corresponding text in the Introduction have therefore been revised to accurately describe current operational practice. The motivation has also been reframed to emphasize the actual gap addressed in this study, namely the limited direct observational constraint on the total surface velocity field, including both geostrophic and ageostrophic components.

Abstract (L2), revised*: "Most operational systems primarily assimilate SSH, SST, and T/S profiles, from which geostrophic currents are diagnostically inferred. Direct assimilation of satellite-derived ocean surface currents provides a complementary source of observation, with the potential to improve the dynamical consistency of NRT surface current estimates, particularly in regions of highly variable circulation where accurate knowledge of the evolving ocean state is critical for marine operations."*

Introduction (revised): "Direct assimilation of SST-derived total surface currents therefore complements, rather than replaces, the scalar observations traditionally assimilated in operational forecasting systems by providing a direct observational limitation on the surface velocity field. Because these observations contain signatures of both geostrophic and ageostrophic motions, they may improve the representation of near-surface circulation variability within a DA framework."

2. Related to this, the discussion of ageostrophic dynamics could be improved. At L40, the manuscript identifies "shelf seas and frontal zones" as regions where ageostrophic dynamics are significant. This is somewhat misleading. Ageostrophic contributions are important wherever the flow evolves in time, including much of the open ocean. In some frontal regions, geostrophic balance may in fact be particularly strong due to large density gradients. A more general and dynamically consistent description would be preferable.

Response: We agree that the original discussion of ageostrophic dynamics was too restrictive. The revised text now explains that ageostrophic motions occur across a wide range of oceanic environments. The discussion has also been revised to more clearly distinguish the information provided by sea level observations from the broader surface circulation represented by EO-derived surface current observations.

Edits: "*These motions occur throughout much of the global ocean wherever the surface circulation evolves in time (Wenegrat et al., 2017, McWilliams et al., 2016}. Consequently, geostrophic velocities inferred from sea surface height (SSH) provide only a partial description of the near-surface circulation because ageostrophic motions are not directly represented by sea level observations".*

3. The experimental design also raises some limitations that should be discussed more clearly. At L157, the study period is limited to one month (January 2021). The manuscript notes that the aim is not climatological generalization but rather a controlled assessment of the method. Even so, the short duration limits the robustness of the conclusions, particularly given the strong dependence of the results on observational density and environmental variability. This limitation should be more clearly reflected in the interpretation of the results.

Response: We agree that the one-month evaluation limits the extent to which the findings can be generalized. While the original manuscript acknowledged this in the experimental design, the limitation was not carried through consistently in the interpretation of the results or conclusions. The revised manuscript now makes the study scope explicit throughout by qualifying the findings in the Abstract, Results, and Conclusions, and by adding a dedicated discussion of how the January 2021 winter conditions influence observational availability, SST feature tracking, and the resulting performance of the proposed framework. A multi-season evaluation is identified as an important direction for future work.

4. Another important issue concerns the independence of the validation. At L207, the CMEMS reanalysis product is used both as part of the ensemble (Section 2.2.4) and as a dataset for evaluation (Section 2.2.3). This means that the comparison with CMEMS is not independent. At best, it demonstrates that the system is internally consistent and does not introduce major technical errors. The inclusion of HF radar observations as an independent validation dataset is a strength of the study, but the role of the CMEMS comparison should be more carefully described.

Response. We would like to clarify that the CMEMS reanalysis product is not used in the construction of the MEnKF ensemble. The ensemble members are generated solely from the multi-model forecast products described in Section 2.2.4, whereas the CMEMS NWS reanalysis is used only as a reference dataset for comparative evaluation. To avoid this ambiguity, we have revised the manuscript to more clearly distinguish the roles of the different datasets and updated Table 1 accordingly. We also clarify that the comparison with CMEMS is intended as an external consistency check against a widely used reanalysis product rather than an independent validation dataset.

Table 1. Summary of spatial resolution, latency, and temporal resolution of EO products and circulation models used in this study.

Dataset	Domain	Category	Role in This Study	Native DA Sys.	Spatial Res. (km)	Latency (h)	Temp. Res. (h)	Lead Time (day)
AVHRR	Regional	Satellite Data	Surface Current Retrieval	–	1.1	2–6	1–6	–
HFR	Coastal	In-Situ	Reference for Evaluation	–	5.5	–	1	–
HYCOM	Global	Forecast	Background Ensemble	NCODA (multivariate 3Dvar)	8.8	6	3	4
GRTOFS	Global	Forecast	Background Ensemble	RTOFS-DA (multivariate 3DVar)	8.8	12	3	8
MedSea Phys (006_013)	Regional	Forecast	Background Ensemble	OceanVar (3DVar)	4.2	6	1	4
CMEMS Glob Phys (001_024)	Global	Forecast	Background Ensemble	SAM2 (Kalman Filter)	9.25	12	3	9
NWS Reanalysis (004_009)	Regional	Reanalysis	Reference for Evaluation	NEMOVar (3DVar)	7	–	1	–

5. A key methodological concern relates to the consistency between the observations being assimilated and the ensemble fields. The MCC-derived surface currents represent total surface velocity, and therefore include contributions from tides and other high-frequency processes. In contrast, none of the ensemble models appear to resolve tides. At the same time, assimilation is performed every three hours using an eight-hour time window, and the observations used in the assimilation seem to include tidal signals. This creates a fundamental inconsistency: the innovations being assimilated include processes that are not represented within the ensemble. It is not clear how this is reconciled within the framework, and this issue should be addressed explicitly.

Response: The reviewer correctly notes that MCC-derived surface currents represent the total surface velocity and therefore may contain tidal and other high-frequency motions, whereas the ensemble forecast systems primarily represent mesoscale circulation and generally don’t include explicit astronomical tidal forcing. This introduces an inherent observation-model mismatch that cannot be eliminated within the present framework. To address this concern, we have revised the manuscript to explicitly acknowledge this limitation and clarify the assumptions underlying the assimilation strategy. Section 2.2.3

and Table 1 now describe the dynamical characteristics of the ensemble products, while Section 3.3.1 explains that the 8-hour observation window was selected to reduce temporal mismatch associated with irregular EO-derived surface current observations, and that any remaining mismatch arising from unresolved high-frequency variability is treated as part of the observation uncertainty rather than being explicitly represented by the ensemble. This limitation is also discussed in the discussion section.

6. The treatment of missing data is also a concern. At L269-276, missing values are assigned a near-zero velocity (0.001 m/s), justified as representing calm conditions. Although the authors provide a rationale for this approach, the assumption is problematic. Surface currents can remain strong in the absence of wind, for example in boundary currents or mesoscale eddies. More importantly, assigning artificial values to missing data alters the statistical structure of the observations and may influence covariance estimates in the assimilation. In a data assimilation framework, it is generally preferable to allow the model dynamics to fill observational gaps rather than imposing fixed values. This choice should be reconsidered or more carefully justified.

Response. We agree that missing observations should not be interpreted as physically calm ocean conditions. The manuscript has therefore been revised to clarify that the value of 0.001 m/s is used solely as a numerical placeholder to preserve matrix dimensional consistency during observation vector construction and is not treated as a physical observation during assimilation. Missing MCC retrievals remain identified by the original no-data mask and are assigned very large observation uncertainty, resulting in negligible Kalman gain and effectively excluding them from the analysis update. Consequently, observational gaps are not filled by the placeholder values; instead, they evolve according to the ensemble background and its forecast-error covariance, consistent with the standard practice of allowing the model dynamics to provide the estimate where observations are unavailable. We have also removed the previous reference to a "calm-weather value" and clarified that the original no-data mask is retained throughout the assimilation and evaluation, so that validation is performed only where valid EO-derived current observations are available. Finally, the potential use of advanced gap-filling techniques to reconstruct physically consistent surface current fields prior to assimilation has been identified as future work.

7. The temporal assumptions underlying the assimilation window also deserve clarification. At L305, the manuscript states that currents “remain approximately coherent over a period of up to three hours,” but the basis for this assumption is not clearly provided. Given the presence of tides, mesoscale variability, and rapidly evolving coastal processes, the validity of this assumption may vary significantly. Further justification would strengthen the methodological framework.

Response. We agree that the original wording overstated the physical basis for the selected assimilation interval. The manuscript has been revised to clarify that the 3-hour assimilation cycle and 8-hour observation window represent an operational trade-off between maximizing the availability of valid MCC-derived observations and limiting temporal mismatch with the model state, rather than assuming that the surface circulation remains unchanged over this period. We also acknowledge that the validity of this approximation depends on the dominant dynamical processes, with faster-evolving motions, such as tides and other high-frequency variability, being less well represented. These assumptions and their implications are now discussed explicitly in the revised manuscript.

8. The description of the CMEMS reanalysis as a "physically consistent benchmark" at L193 and L198 also warrants clarification. While reanalyses represent state-of-the-art products, they are not strictly dynamically consistent in the sense of being free solutions of the governing equations. Sequential data assimilation introduces imbalances at each analysis step, followed by model adjustment. This does not invalidate their use as a reference, but the terminology could be refined.

Response. We agree that the term "physically consistent benchmark" overstated the characteristics of the CMEMS NWS reanalysis. The manuscript has therefore been revised to describe it as a reference dataset rather than a physically consistent benchmark. We also acknowledge that, like other sequential reanalysis systems, CMEMS incorporates analysis increments that may temporarily introduce dynamical imbalances before model adjustment. The revised wording more accurately reflects its role as a widely used reference product for comparative evaluation.

Edit: "*In this study, the CMEMS reanalysis product is used as independent reference dataset for evaluating the performance of the MEnKF system*."

9. At L210, the manuscript states that "a summary of these models and their key characteristics is provided in Table 1." Table 1 lists spatial and temporal resolution but does not include key information about the models themselves, such as their dynamical formulation, forcing, or data assimilation approaches. Expanding this description would improve clarity.

Response. We revised both Table 1 and the accompanying text to provide a more complete description of the forecast and reanalysis systems used in this study. Table 1 now includes information on the dynamical core and native data assimilation system of each model, while the dataset description has been expanded to summarize the atmospheric forcing, forecast configuration, and initialization strategy. These additions provide a clearer description of the ensemble members and their structural differences within the proposed multi-model framework.

10. The quality of agreement between the MEnKF analysis and independent observations also deserves closer scrutiny. In the HFR comparisons (e.g., Figures 8-10), the reported RMSE values (~0.15-0.17 m/s) appear reasonable, but the visual comparisons suggest only moderate agreement in both magnitude and spatial structure. In particular, the vector fields and distribution plots indicate noticeable discrepancies in current strength and direction.

Response. We agree that the HFR comparison demonstrates moderate, rather than near-perfect, agreement in both current magnitude and spatial structure. The manuscript has therefore been revised to provide a more balanced interpretation by explicitly stating that the MEnKF analyses reproduce the dominant circulation patterns and overall current magnitude distribution, while localized differences in current strength, direction, and fine-scale features remain. We also clarify the likely causes of these discrepancies, including the irregular spatial coverage of the assimilated MCC-derived surface currents due to cloud contamination and satellite sampling (newly added Fig. 4), differences in the observation characteristics of the EO and HFR datasets, and unresolved fine-scale and high-frequency variability not fully represented by the ensemble forecasts. Consequently, the revised Results and Discussion emphasize that the reported RMSE values should be interpreted together with the vector-field comparisons, as the scalar metrics alone do not fully characterize the spatial agreement between the MEnKF analyses and the independent HFR observations.

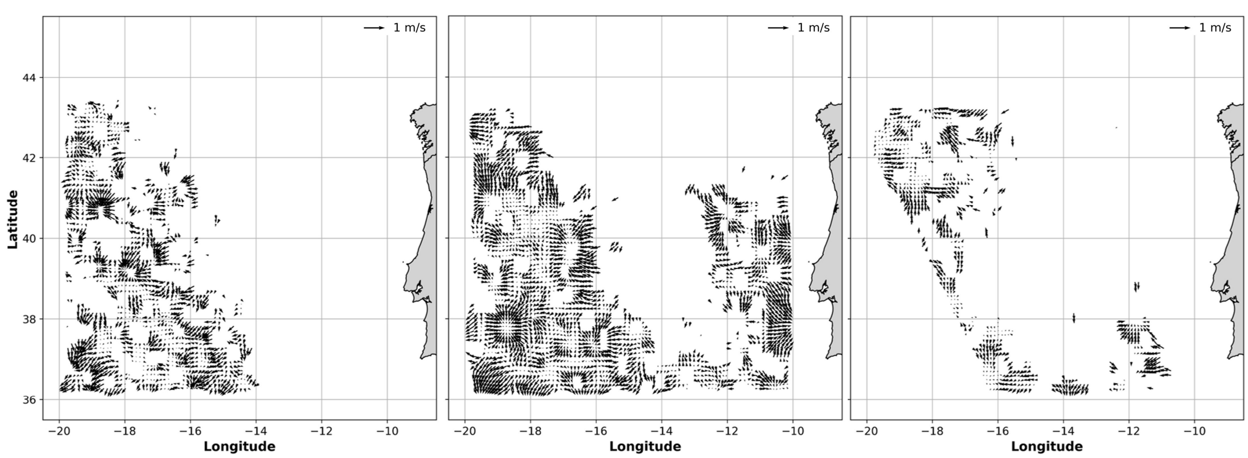

Figure 4. Representative examples of EO-derived surface current fields estimated using the MCC algorithm from AVHRR image pairs with short, moderate, and long temporal

separations ranging from approximately 1 to 24 h over 01, 02, 03 January 2021 (left to right). Grey-shaded regions along the right-hand side of each panel indicate the Spanish coastline.

11. Similarly, the comparison with the CMEMS reanalysis (Figures 11-16) shows systematic underestimation of current magnitudes and only moderate correlation (mean complex correlation ~0.48). This level of agreement suggests that the assimilation captures some aspects of the flow but does not fully reproduce the observed or reanalysis structures. While the manuscript acknowledges some of these limitations, the overall interpretation tends to emphasize success more strongly than the results appear to support. A more balanced discussion of the strengths and limitations of the method would improve the manuscript.

Response. We agree that the comparison with the CMEMS reanalysis demonstrates only moderate agreement and that the original discussion placed greater emphasis on the strengths of the proposed framework than the results warranted. The manuscript has therefore been revised to provide a more balanced interpretation by explicitly acknowledging the systematic underestimation of current magnitudes, the moderate complex correlation values, and the remaining discrepancies in the spatial structure of the surface circulation. The revised discussion now emphasizes that the assimilation captures the dominant circulation patterns and improves the overall representation of the flow, while recognizing that it does not fully reproduce all energetic and fine-scale features present in the reference reanalysis dataset.

12. The behavior of the bias in Figure 17 is somewhat unexpected and deserves further discussion. In an ensemble Kalman filter framework, both model and observation errors are typically assumed to be unbiased, such that the analysis should not exhibit large systematic biases if these assumptions hold. However, the manuscript reports persistent and sometimes relatively large biases across observational regimes.

This likely reflects the combined effects of several factors, including biases in the MCC-derived currents, the use of near-zero fill values over regions with missing data, and inconsistencies between the assimilated observations (which include high-frequency processes such as tides) and the ensemble models (which do not represent these processes). In addition, structural differences between the assimilated fields and the reanalysis used for comparison may contribute to the apparent bias.

These factors suggest that the unbiased-error assumption underlying the data assimilation framework is not satisfied in practice. A more explicit discussion of these issues, and their impact on the reported bias, would strengthen the interpretation of the results.

Response: In response, Figure 17 has been revised to provide a clearer representation of the bias behavior across the observational regimes. The updated results show relatively modest variations in bias, and the manuscript now includes a more explicit discussion of the factors contributing to this behavior. We clarify that the observed bias likely reflects the combined effects of uncertainties in the MCC-derived current retrievals, the numerical treatment of missing observations through placeholder values, observation–model mismatch associated with unresolved high-frequency variability, and structural differences between the MEnKF analyses and the NWS reanalysis used as the reference dataset. We also acknowledge that, although the classical EnKF formulation assumes unbiased forecast and observation errors, this assumption is only approximately satisfied in the present application. The revised discussion therefore emphasizes that the reported bias should be interpreted in the context of these practical limitations rather than as evidence of a deficiency in the EnKF formulation itself.

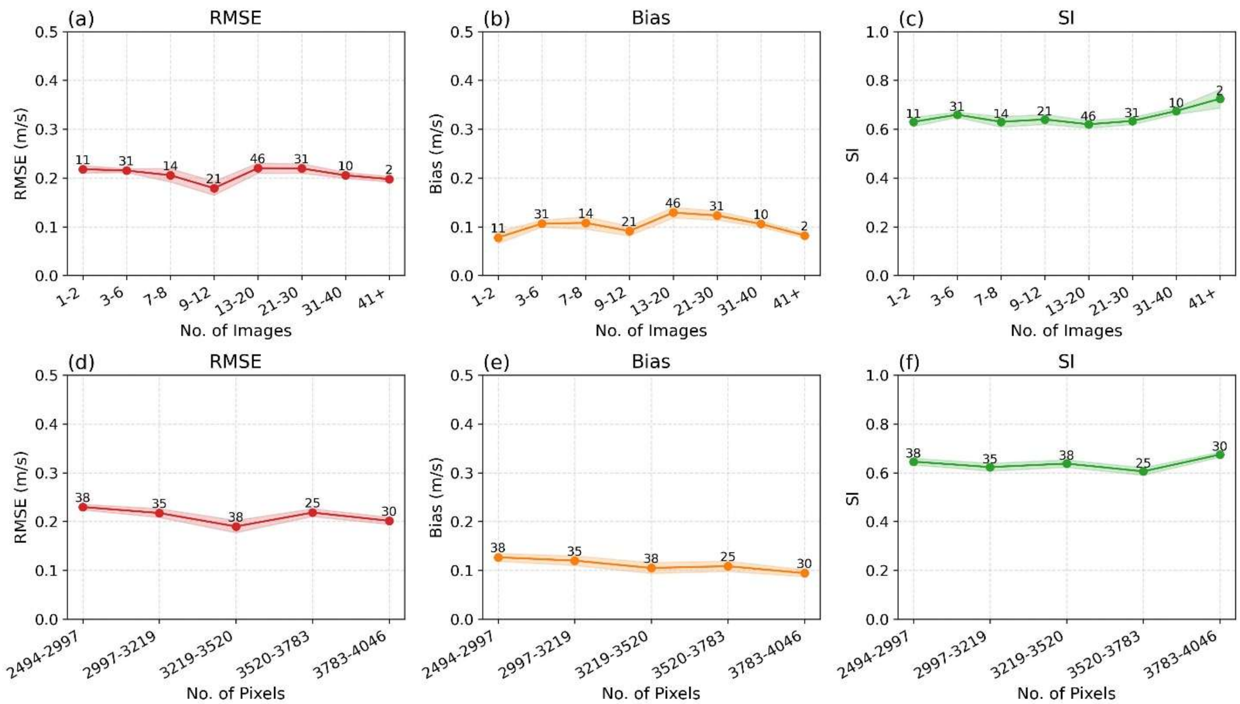

Figure 17: Sensitivity of MEnKF assimilation performance to EO-derived observation density and spatial coverage. Panels (a-c) show RMSE, bias, and Scatter Index (SI) for current speed as functions of observational density, expressed as the number of assimilated EO-derived current products per assimilation cycle. Panels (d-f) show the same metrics as functions of spatial coverage, expressed as the number of valid MCC-derived current vectors after quality-control filtering. Markers denote bin-averaged values, annotated with the corresponding sample counts (*n*), while shaded envelopes represent the ±1 standard error.

13. The discussion beginning at L697, where the authors note that both low and high observational densities degrade performance, is somewhat concerning and warrants further examination. While it is expected that insufficient observational coverage leads to poorer analyses, the degradation of performance at high observation densities is less intuitive and suggests more fundamental issues in the assimilation framework.

In a well-posed ensemble Kalman filter system, the addition of more independent and unbiased observations should not systematically degrade performance; at worst, one would expect diminishing returns. The behaviour reported here indicates that the effective information content of the observations is not increasing with observation count and may in fact be degrading the analysis.

The explanations provided – such as spatial clustering, temporal inconsistency, and ensemble limitations – are plausible contributing factors, but they point toward deeper issues, including correlated observation errors, inconsistency between observed and model-resolved dynamics (e.g., tides and high-frequency variability), and limitations in the representation of error covariance. The treatment of missing data may also play a role, as the introduction of artificial near-zero values could further reduce the effective quality of the assimilated observations at higher data volumes.

This behaviour deserves more careful analysis, as it suggests that the data assimilation system may not be optimally configured to handle the characteristics of the observational dataset. A more explicit examination of these mechanisms, or additional sensitivity tests, would strengthen the manuscript.

Response: We agree that the observed reduction in performance at high observational densities warrants further discussion. To address this concern, we have revised manuscript and discussion. First, Figure 18 has been updated to include sample counts and a weighted RMSE representation, allowing the influence of sampling frequency on the reported trends to be more clearly distinguished from sparsely populated observational regimes. Second, the discussion has been expanded to clarify that the observed behavior should not be interpreted as a consequence of assimilating more observations alone. Rather, it likely reflects the combined effects of temporally clustered AVHRR acquisitions, which reduce the independence of successive MCC-derived current fields, correlated observation errors, observation–model mismatch associated with unresolved high-frequency variability, and limitations imposed by the four-member ensemble in representing background error covariance. We also note that the numerical placeholder assigned to missing observations is excluded from the analysis update through its prescribed observation uncertainty and is therefore unlikely to be the primary cause of the observed degradation. Since these factors cannot be isolated with the present experiments, the manuscript now presents them as plausible mechanisms and identifies their quantitative assessment as an important direction for future work.

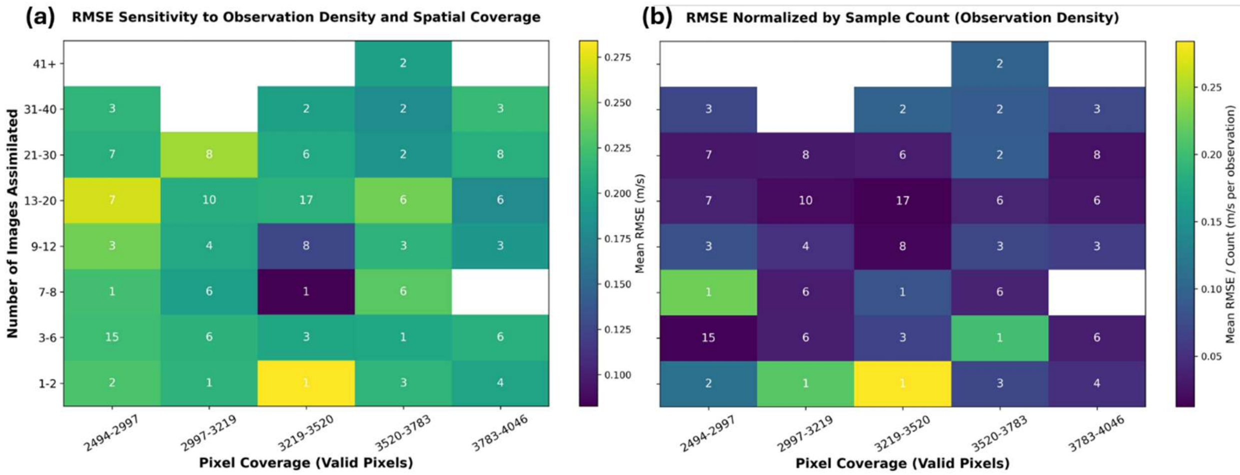

Figure 18: Joint sensitivity of MEnKF assimilation performance to observational density and spatial coverage. (a) Mean RMSE aggregated by binned number of assimilated images (y-axis) and valid observation coverage (x-axis), with color indicating average RMSE within each bin. (b) The values in (a) normalized by the number of assimilation cycles contributing to each bin, shown to illustrate how strongly each bin's RMSE estimate is influenced by its underlying sample size. White cells indicate bins with no available samples.

14. In summary, this manuscript presents an interesting and potentially valuable framework for assimilating satellite-derived surface currents into a multi-model ensemble system. The methodological approach and sensitivity analyses are strengths, and the study addresses an important problem in ocean forecasting. However, several aspects of the work require substantial clarification or revision. These include the framing of the motivation, the consistency between observations and ensemble fields, the handling of missing data, and the independence of the validation.

In addition, the level of agreement between the MEnKF analysis and independent datasets appears more moderate than implied by the overall narrative, with discrepancies in structure, magnitude, and correlation that warrant a more balanced interpretation. The presence of persistent bias and the reported degradation of performance at higher observational densities further suggest that key assumptions underlying the data assimilation framework – such as unbiased errors and consistent dynamical representation – may not be fully satisfied in practice. These behaviors point to potential structural limitations in the assimilation system that deserve more explicit examination.

Addressing these issues would significantly strengthen the manuscript and improve confidence in the reported results.

I recommend reconsideration after major revision.

Response: We appreciate the reviewer's comprehensive assessment and have revised the manuscript accordingly. The motivation has been reframed to more accurately reflect current operational data assimilation practice, the roles of the observational and reference datasets have been clarified, and the assumptions and limitations associated with the assimilation framework are now discussed more explicitly. We have also revised the interpretation of the results to provide a more balanced assessment of the assimilation performance by acknowledging the moderate agreement with the HF radar observations and NWS reanalysis, the remaining discrepancies in current magnitude and spatial structure, and the observed persistent biases under different observational regimes. The revised discussion further explains these behaviors in the context of the practical assumptions and limitations of the present framework, including representativeness differences, observation characteristics, and covariance specification. We believe these revisions provide a more transparent and balanced presentation of both the capabilities and the limitations of the proposed methodology.