Supplement of

Ocean Model Analysis and Prediction System version 4.1i (OceanMAPSv4p1i)

Divakaran et al.

Correspondence to: Prasanth Divakaran (prasanth.divakaran@bom.gov.au)

1 Eddy verification

The section presents information on eddy-tracking verification of 24-hr forecasts from OceanMAPSv4p1i and OceanMAPSv4p0i, in accordance with the methods and region specified in the Smith and Fortin (2022) [hereafter SF 2022]. As in SF 2022, we use the py-eddy-tracker (Mason et al., 2014) to identify and track eddies. Py-eddy-tracker is freely available open-source Python software under the GNU license (https://github.com/AntSimi/py-eddy-tracker). This approach is also adopted in version 3 of the AVISO Eddy Atlas (AVISO, 2016). The AVISO Ssalto/Duacs altimeter analysis dataset used in the study is the latest high-resolution (0.125°) gridded Level 4 product produced and distributed by Copernicus Marine and Environment Monitoring Service (CMEMS) (https://doi.org/10.48670/moi-00148). The main difference in eddy identification criteria between SF 2022 and the current study is in the number of minimum grids, which has increased from 5 to 20 to align with the higher resolution in the datasets. The verification domain covers the Gulf Stream region in the northwest Atlantic, for the period 20 August 2023 to 31 January 2024.

The identification of anticyclonic and cyclonic eddies from two consecutive days (26 August 2023 and 27 August 2023) is shown in Fig. S1 and Fig. S2 for AVISO, OceanMAPSv4p0i, OceanMAPSv4p1i and OceanMAPSv4p1i run001 products. Qualitatively, there is a high correspondence between eddy contours and filtered SSH as evident in the observational analysis and in all three OceanMAPS forecasts. However, visually, the number of eddies detected in the forecasts is fewer than in the observational analysis. Generally, average forecasts have eddy numbers that are lower than those of individual members, as ensemble averaging smooths out random errors that increase with the wavenumber. However, we note that OceanMAPSv4p0i forecasts fail to adequately represent the majority of smaller eddies north of 50° N in the study domain. The persistence of eddies and mesoscale features can be done by comparing daily snapshots from consecutive days (Fig. S1 and Fig. S2). Mesoscale features in the observational analysis and forecast products from OceanMAPSv4p1i look consistent and show minimal differences between 26th August and 27th August. However, OceanMAPSv4p0i forecast differs greatly between Fig. S1 and Fig. S2. For example, the anti-cyclonic eddy located at 60° W, 37.5° N, surrounded by a ring of smaller cyclonic eddies, is a prominent feature in all products. This feature has shown greater persistence in AVISO observational analyses and in OceanMAPSv4p1i forecasts; however, it appears to vary substantially across consecutive forecasts from OceanMAPSv4p0i. As noted in the paper, this is due to an inconsistency in the forecast system layout in OceanMAPSv4p0i, which has been rectified in OceanMAPSv4p1i.

Eddies in forecasts generally exhibit qualitatively similar properties in terms of radius and amplitude to those observed in AVISO analyses (Fig. S3). However, the total number of detected eddies in the forecasts over the study period, relative to the number in the observed product, is ~68 % and ~65 % in the average forecasts from OceanMAPSv4p1i and OceanMAPSv4p0i, respectively. The single-member forecasts from OceanMAPSv4p1i produced nearly 80 % of the total observed eddy population. Though the total number of eddies in the average

forecasts from OceanMAPSv4p0i and OceanMAPSv4p1i are more or less similar, OceanMAPSv4p1i forecasts have improved representation of smaller eddies with radii close to 20 km adequately. As noted in the system description, OceanMAPSv4p0i average forecasts are initialised with the average of ensemble restart files at the end of the hindcast run. The inherent randomness of individual members increases over the three days of the hindcast run, and averaging restarts at the end of the hindcasts cancels out many small-scale features. To address the forecast initialisation issue noted in OceanMAPSv4p0i and to account for the importance of randomness in generating spread in an EnKF system, we introduced a near-real-time (NRT) analysis in OceanMAPSv4p1i forecast system. This intermediate step leverages the 2-day observation latency relative to real time, and works without affecting the spread of individual members. The average NRT analysis (best estimates) is used as initialisation for the forecast forward run to represent the mesoscale ocean state better in OceanMAPSv4p1i forecasts.

Matching of eddies in the observation and forecasts are done using the cost function (C) minimisation described in equation 1 in SF 2022. Here, C is a function of the difference in amplitude (A), radius (R) and distance (D) between eddy centres, and is used to identify the most likely match as shown below in the equation, where the subscripts 1 and 2 refer to the two eddies being considered. In our study, subscript 1 denotes the eddy from observation, and 2 indicates the eddy from forecasts matched against 1.

$$C = \sqrt{\left(\frac{A1-A2}{A1}\right)^2 + \left(\frac{R1-R2}{R1}\right)^2 + \left(\frac{D}{125km}\right)^2} \qquad (1)$$

Following SF 2022, we examine the fraction of observed eddies captured by ocean forecasts using the probability of detection (POD). The fraction of predicted eddies in the forecast that did not have a match in the observation, using the False Alarm Ratio (FAR). The domain-average time series of POD and FAR are shown in Fig. S4. Both statistics vary daily, with minimal seasonal variation. The individual member forecast from OceanMAPSv4p1i has the highest POD (~0.6), with the run004 average forecasts from v4p1i (~0.55) and v4p0i (~0.53) ranking second and last, respectively. This indicates that run004 forecasts from OceanMAPSv4p1i exhibit a 5.4% higher POD than those from OceanMAPSv4p0i. FAR is greater for individual member forecasts from OceanMAPSv4p1i, with values close to 0.23. This may be due to chance/randomness in the large number of predicted eddies in the run001 forecasts, which in turn yields higher POD and FAR values. The false-alarm rate in the OceanMAPSv4p1i run004 forecast (~0.178) is the lowest among the three, with the OceanMAPSv4p0i run004 forecast (~0.188) coming in the middle. That means a difference of -5.5 % on FAR for OceanMAPSv4p1i over OceanMAPSv4p0i in the average forecast product.

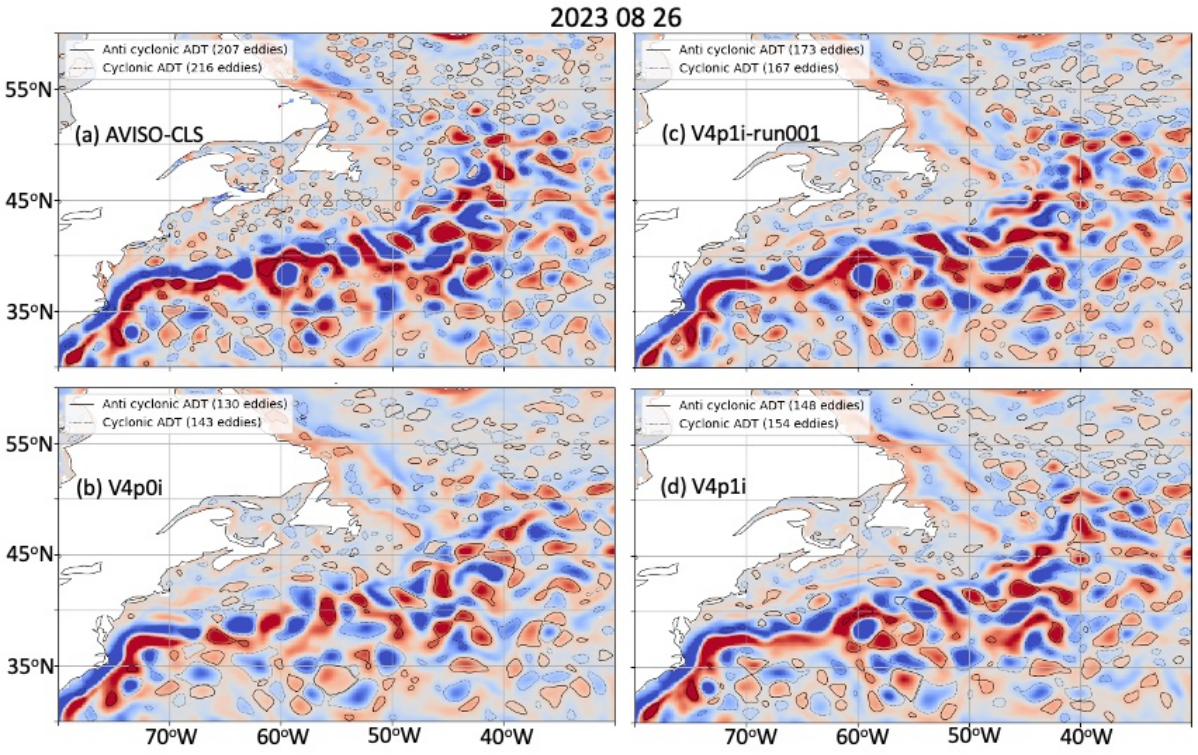

Figure S 1: Filtered SSH field valid for 26 August 2023 for (a) AVISO gridded satellite observations, 24-hour forecast fields from (b) v4p0i, (c) v4p1i-run001 and (d) v4p1i.

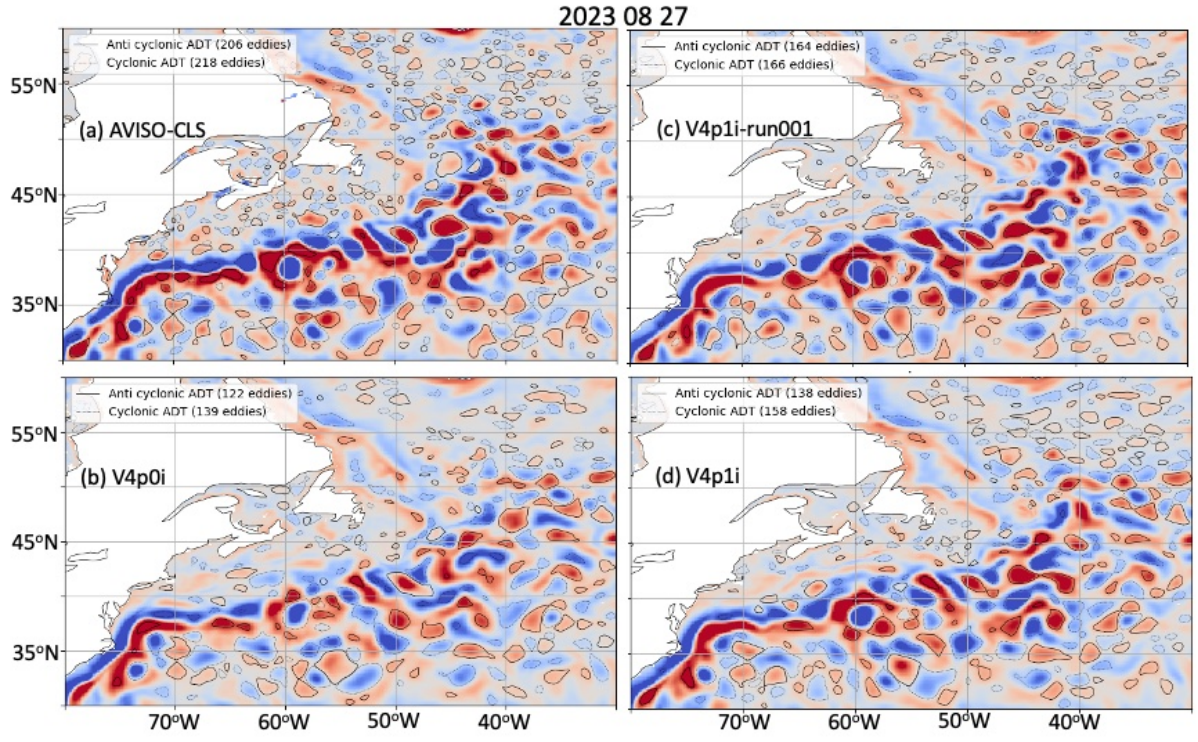

Figure S 2: Filtered SSH field valid for 27 August 2023 for (a) AVISO gridded satellite observations, 24-hour forecast fields from (b) v4p0i, (c) v4p1i-run001 and (d) v4p1i.

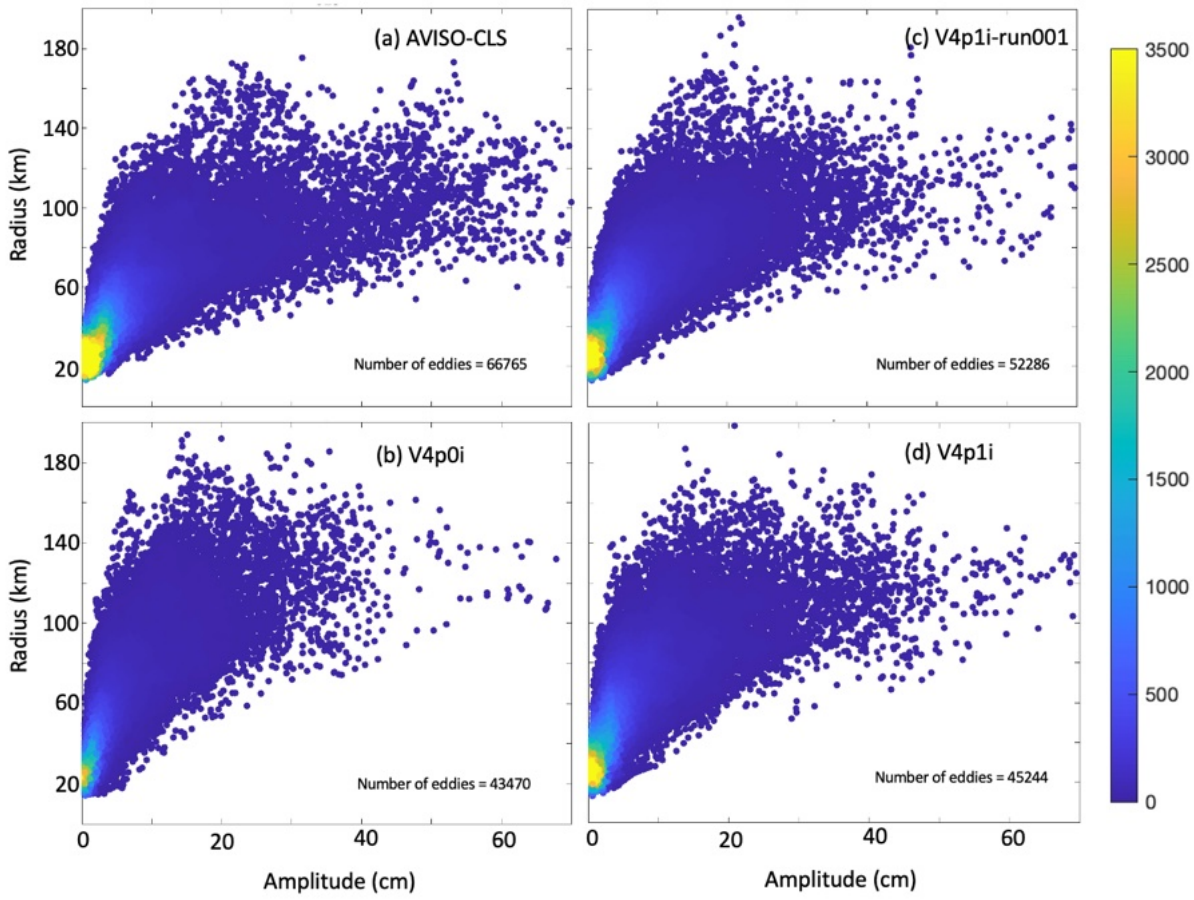

Figure S 3: Density scatter plot showing characteristics of eddies identified over 20 August 2023 to 31 January 2024 for (a) AVISO, (b) V4p0i, (c) V4p1i-run001, and (d) V4p1i.

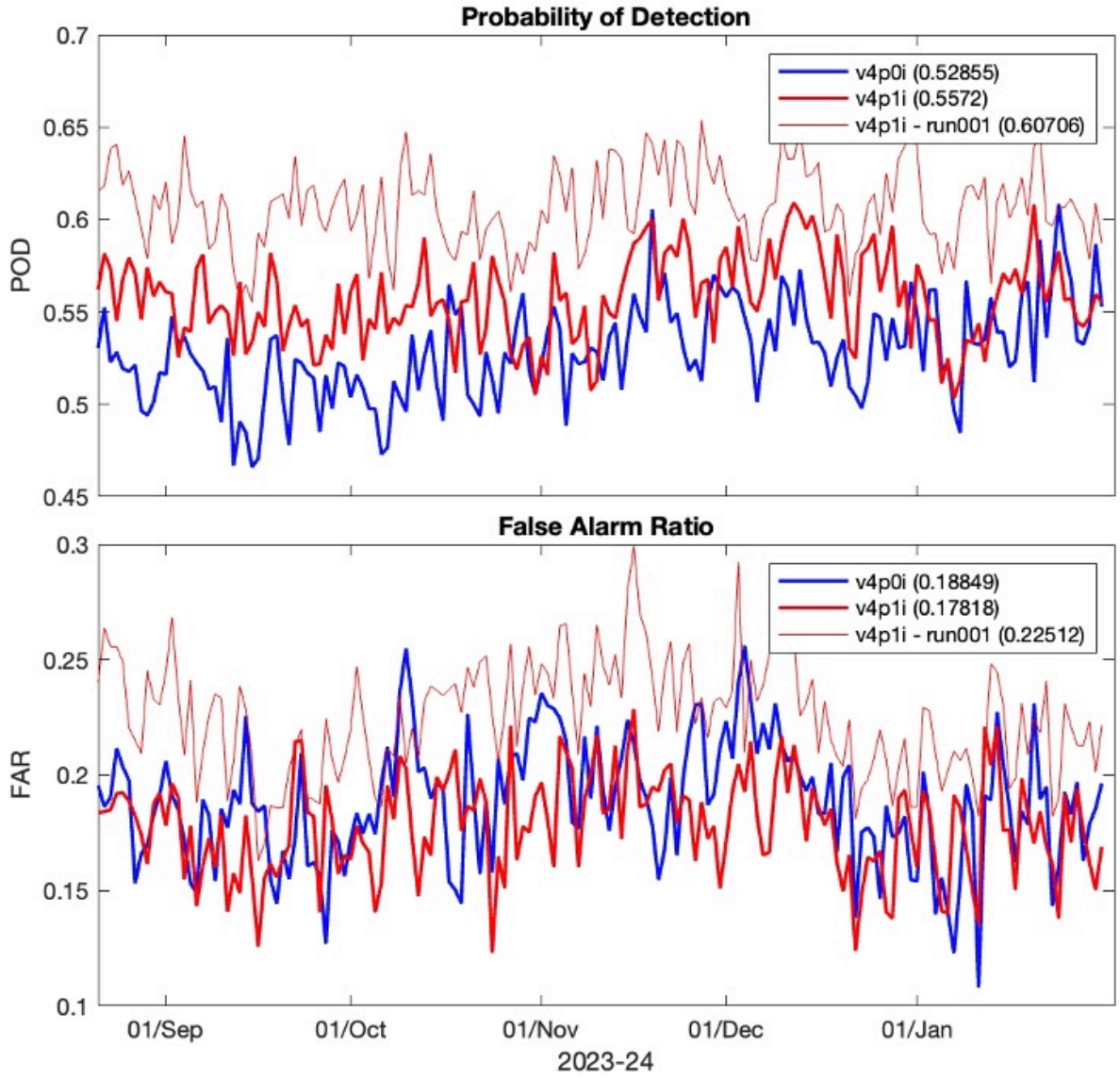

Figure S 4: Time series of matching statistics for v4p0i (blue), v4p1i (red) and v4p1i-run001 (Blue thin line) over the period 20 August 2023 to 31 January 2024. The upper panel shows the probability of detection (POD), and the false-alarm ratio (FAR) is shown in the lower panel.

2 Neighbourhood verification- Continuous Ranked Probability Score (CRPS)

Neighbourhood verification can account for the double-penalty effect (due to slight spatial and temporal errors in high-resolution datasets) by regridding forecasts from different resolutions into a common grid. Neighbourhood verification has been applied to SST forecasts using HiRA spatial methods (Crocker et al., 2020; Mittermaier, 2014), in which forecast grid points within neighbourhoods centred on an observing location are treated as pseudo-ensemble members. An assumption is made that the observation is not only valid at the measurement location but

also representative of the surrounding area (Barbosa et al., 2024). Here, we use CRPS in a similar but different way. In our case, we test the hypothesis that the ensemble average (run004 forecast) has lost spectral power (i.e., is smoother) and benefits from a reduction in the double penalty, even though it is at the exact grid resolution. If this is true, we expect the reduction in error for run001 (single-member forecast) to be substantially greater than that for the ensemble average restart (run004) forecasts. Furthermore, we expect the error curves to converge.

Leadtime - continuous ranked probability score (CRPS) is used in Fig. S5 and Fig. S6 for average forecasts and single-member forecasts from OceanMAPSv4.0i and OceanMAPSv4.1i. NB1 represents 3 x 3 nearest forecast points in a square around the observation, whereas NB2 represents 5 x 5 forecast points. Results suggest that there is negligible information lost in the average forecast relative to the single-member forecast in both OceanMAPSv4.0i and OceanMAPSv4.1i. CRPS results are consistent with the MAE results shown in the paper.

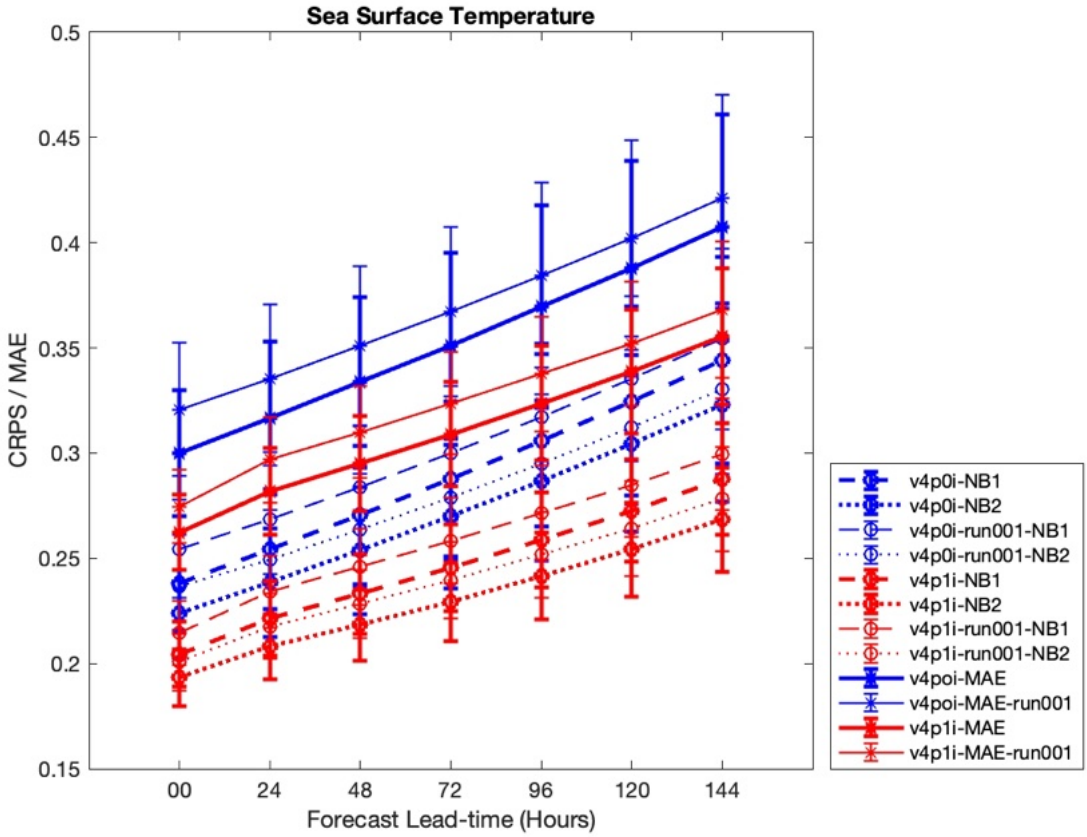

Figure S 5: Leadtime SST CRPS, using drifter observations for forecast products from OceanMAPSv4p1i and OceanMAPSv4p0i using grids NB1 and NB2. Overlayed by single-point forecast MAE.

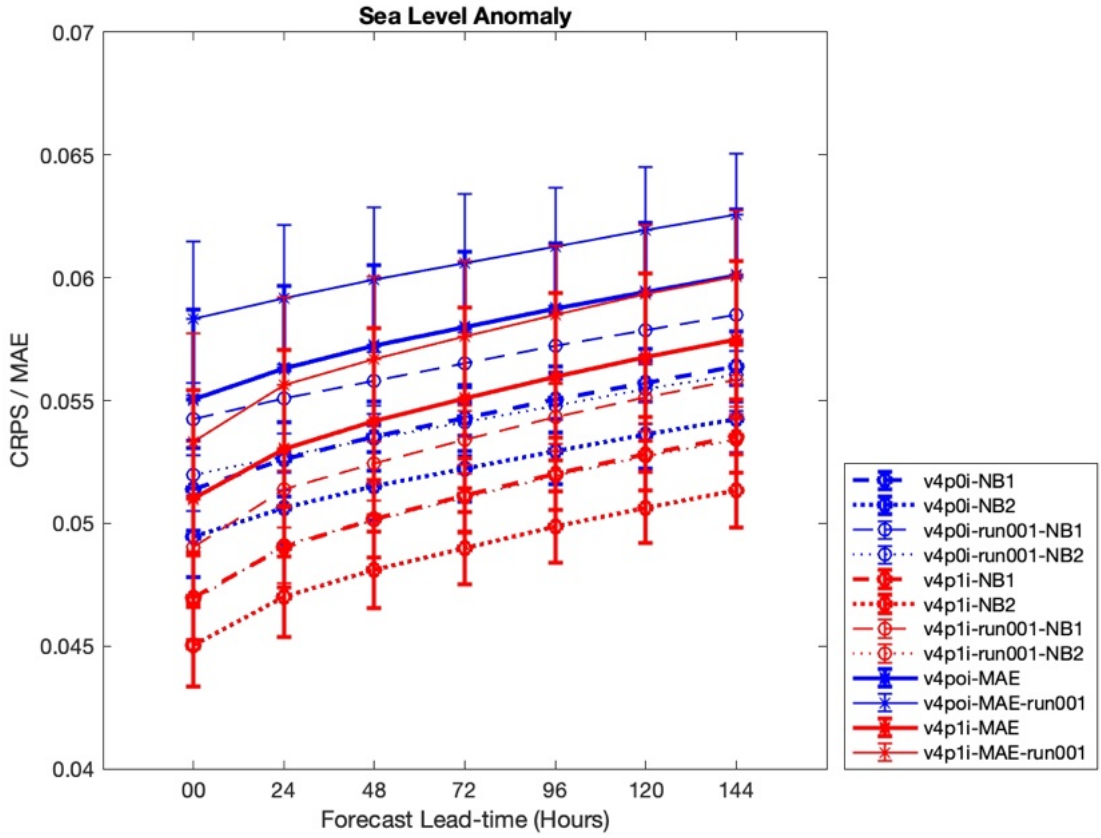

Figure S 6: Leadtime SLA CRPS, using Jason3 observations for forecast products from OceanMAPSv4p1i and OceanMAPSv4p0i using grids NB1 and NB2. This is overlayed by single-point forecast MAE as shown in Figure 11.

Reference

AVISO.: SSALTO/Duacs user handbook: (M)SLA and (M)ADT near-real time and delayed time products, CLS-DOS-NT-06-034, Issue 5rev0, Mesoscale eddy trajectory atlas product handbook, SALP-MU-P-EA-23126-CLS, Issue 3rev2, https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/hdbk_eddytrajectory_META2018.pdf , 2016.

Barbosa Aguiar, A., Bell, M.J., Blockley, E., Calvert, D., Crocker, R., Inverarity, G., King, R., Lea, D.J., Maksymczuk, J., Martin, M.J. and Price, M.R.: The Met Office Forecast Ocean Assimilation Model (FOAM) using a 1/12-degree grid for global forecasts., Quarterly Journal of the Royal Meteorological Society, 150(763), 3827-3852, 2024.

Crocker, R., Maksymczuk, J., Mittermaier, M., Tonani, M., and Pequignet, C.: An approach to the verification of high-resolution ocean models using spatial methods, Ocean Sci., 16, 831–845, https://doi.org/10.5194/os-16-831-2020, 2020.

Mason, E., Pascual, A., McWilliams. J. C.: A new sea surface height-based code for oceanic mesoscale eddy tracking, Journal of Atmospheric and Oceanic Technology., 31, 1181–1188, https://doi.org/10.1175/JTECH-D-14-00019.1, 2014.

Mittermaier, M. P.: A Strategy for Verifying Near-Convection-Resolving Model Forecasts at Observing Sites, Weather Forecast., 29,185- 204, https://doi.org/10.1175/WAF-D-12-00075.1, 2014.

Smith, G. C. and Fortin, A.S.: Verification of eddy properties in operational oceanographic analysis systems, Ocean Modelling., 172, 101982, 2022.