Manuscript Review

Referee #1

The chosen (ROMS) model is well documented in the literature and already implemented for the experiment area. The choice of assimilation schemes draws from a much larger and less well tested set of options.
The authors refer to several previous assimilation tests with HF radar-derived surface currents in paragraph 3 of the introduction. They may want to consider adding a reference to the recent review paper by Paduan and Washburn(2013), which includes a longer discussion of recent assimilation attempts:
Paduan, J. D., and L. Washburn, 2013: High Frequency radar observations of ocean surface currents. Annu. Rev. Mar. Sci., 5, 115-136.

Thank you for bringing this paper to our attention, a reference to this paper has been added to the manuscript.

The linear adjoint model assimilation technique used here is not common and the paper would benefit from a more in-depth discussion of how it compares to other techniques, as well as the, apparently, sever limitations that are imposed by the high computational costs.

Although 4DVAR is a fairly new method within ocean data assimilation, it is much used in NWP and is thoroughly documented in the literature. A comparison with other assimilation techniques is beyond the scope of this paper. However, some more details on the methods have been added, as well as references to papers in which the IS4DVAR driver of ROMS is thoroughly described.
We have also added a reference to a discussion paper on the advantages of EnKF and 4DVAR, which are the two main directions of sophisticated data assimilation methods today.
Both of these assimilation techniques requires substantial computational resources. The TOPAZ4 ocean forecast system for example, runs an ensemble of 100 members. As an example, one 4DVAR assimilation cycle (24h) with our application is about 133 000 CPU seconds, while a forecast run of same length is 4860 CPU seconds. Integrating 100 ensemble members forward one day, as a 100 member EnKF would require, would thus cost ~486 000 CPU seconds, substantially more than 4DVAR.

In fact, the critical results of this work, as shown in the summary Figures 9 and 10, might be considered marginal or untrustworthy given that they are based on a single realization. Please speak to what would be required to run the assimilation test for the entire 3-month data availability period and do so if at all possible (i.e. with regular restarts, not for a single 90 day prediction) If its not possible, it seems that one conclusion of this assessment is that the linear adjoint assimilation method is likely to be restricted to real-time applications and is not practical for use for multi-year reanalysis studies of the type that would be beneficial to, for example biological connectively studies.

Section 4.3.2, where a comparison of model results with ADCP measurements is presented has been altered, such that instead of results of a single realization, we now compare results from 10 sequential realizations (The analysis at the end of one 24-hours assimilation window provides the background for a new cycle). For each realization a 5-day forecast is initiated, and the results now shown in this section

is based on the statistics of these forecast simulations.
Our reason for not running the assimilation for the entire 3-month period, is the lack of complementary observations to validate the results against. It is, however, certainly computationally feasible.

In the recently accepted paper,

Oke, P. R., G. Larnicol, E. M. Jones, V. Kourafalou, A. K. Sperrevik, F. Carse, C. A. S. Tanajura, B. Mourre, M.Tonani, G. B. Brassington, M. Le Henaff, G. R. Halliwell, Jr., R. Atlas, A. M. Moore, C. A. Edwards, M. J. Martin, A. A. Stellar, A. Alvarez, P. De Mey, M. Iskandarani, Assessing the impact of observations on ocean forecasts and re-analyses: Part 2, Regional applications. Journal of Operational Oceanography,

results from a re-analysis conducted with a ROMS-4DVAR application of the US West Coast with 10 km resolutions for the period 1980-2012, are presented, demonstrating that the data assimilation methods implemented in ROMS are well-suited for re-analysis studies as well as real-time applications.

The entire section using the idealized channel model is of questionable added value, in my view. This may be because the description of the freely forced model is inadequate as is its justification as a viable ground truth in the situation being studied.

The motivation for the idealized study was that we wanted to isolate the nonlinear dynamics of an unstable slope current, i.e. without any influence of tides, storm surge etc. We expected that the performance of the assimilation system would decrease as the assimilation window increased (adjoint and tangent linear models not being valid), and also that the error correlations length scale would need to be rather small due to eddies and narrow fronts. This was more or less confirmed in our tests, both in the idealized and realistic setup. We have added a more detailed description of the idealized model and the motivation for its inclusion to the paper (see Section 3).

The data locations diagram in Figure 5 is confusing with regard to the pseudo HF radar data locations. The caption refers to the total vector estimate locations from a two cite HF radar grid, but the data locations appear to follow the radial geometry of a single HF radar data set. Please clarify.

This is a good point: the observation locations are taken at positions where the beams from the two idealized radars intersect, but they are limited by the fixed range of the radars. We do agree that it looks like a single radar and have added some text to clarify.

Also, please clarify the domain used for the standard deviation computations in Tables 1 and 2. I assume that the results represent the grid point across the entire domain, but it is not explained. Is that correct?

As stated in the beginning of Section 3.5, we only consider a limited region of interest similar to the HF radar coverage area , and also restrict the evaluation to the two uppermost vertical levels of the model. The region is indicated in Figure 4.

In both the idealized assimilation tests and those on the realistic model grid the surface current assimilation appears to provide better results of CTD profiles of temperature and salinity. The authors state in both cases that this is evidence that the density of the CTD observations is too

sparse. The CTD data locations are shown for the idealized case but the locations are not shown for the realistic case. Are they the same? If not it should/must be shown. A broader discussion of this result is warranted beyond implicating the profile resolution, which appears to be fairly dense but limited in area in at least the idealized model case. It is just as likely that the results implicate ageostrophic forcing dominating the surface currents. Given the fact that the history of data assimilation over many decades in numerical weather prediction and more recently in numerical ocean prediction is one that favors integral data from density profiles over interface velocities.

The positions of the CTD profiles in the realistic case are shown in Figure 1. However, as these are taken through the course of the cruise, only six CTD profiles slightly upstream of the HF coverage area fall within the one-cycle experiment.
In the idealized setup, one profile is assimilated every hour, simulating a ship moving through the domain. We agree that our results imply ageostrophic forcing and have added some comments to this effect to Section 4.3.1. We are of course also aware of the fact that hydrographic profiles usually have a high positive impact in ocean DA (e.g. Moore et al., 2011), and that we get detrimental results here indicate that we do not have sufficient hydrographic data to constrain the currents. We do see, however, that adding a few hydrographic profiles to the HF data gives significantly better results during the first part of the forecast period (Section 4.3.1) and the discussion in Section 5 have been expanded somewhat to clarify this point.

The error covariance functions relating surface current and the model state variables of any assimilation scheme. The authors here refer to the different error correlation length scales that were tested but they give no details. I assume that these were isotropic error covariance functions and were the same for velocity and temperature, but these details should be specified.

The referee is correct: the correlation lengths are isotropic and the same for all variables. We have specified this in Section 3.

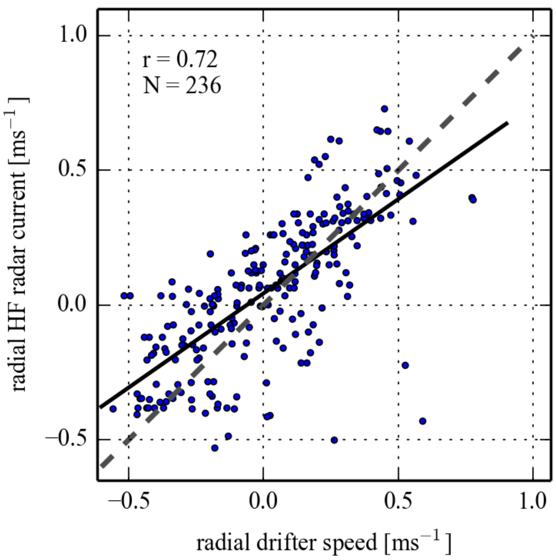
In the comparison of surface drifter trajectories with model-derived trajectories (Section 4.3.1) it is important that an additional result is computed and presented that compares the surface drifter velocities with the HF radar observations. This is necessary to bound the expectation of the assimilation tests and to establish the error levels in the HF radar observations beyond the simple geometric dilution of precision. Both Eulerian and Lagrangian comparisons of available drifter- and HF-derived vector velocities into the radial velocity components facing the individual radar sites is the preferred approach and, if the angles occupied span a wide range, may provide some insight into azimuthal error biases in the individual radars as was shown in the error study of Paduan et al. (2006)
Paduan, J. D., K. C. Kim, M. S. Cook, and F. P. Chavez, 2006: Calibration and validation of direction-finding high frequency radar ocean surface current observations. IEEE J. Oceanic Engin., 10.1109/JOE.2006.886195, 862-875

To evaluate the error of the assimilated HF radar observations we compare radial HF radar currents with the corresponding component of drifter speeds, shown in the new Figure 10. This direct comparison should be suitable to establish a first guess of what can be expected from HF radar assimilation into the ocean model. However, as argued in the manuscript, the most value of the HF radar assimilation lies within providing a picture of spatial variability. The benefit of using HF radar currents for assimilation can therefore be expected to exceed the quality of the HF radar measurements.

For a more detailed comparison of HF radar currents with both Eulerian and Lagrangian observations we would like to draw you attention to another manuscript that we are preparing for Ocean Science Discussion:

Röhrs et al. : "Comparison of HF radar measurements with Eulerian and Lagrangian surface currents". In preparation

In the new manuscript, the same data set is used as in this manuscript. The analysis shows that the SLDMB drifters are well suited for comparison with HF radar currents at this radar frequency, as the depth is sampled in similar ways. The analysis also showed that errors in radial HF radar currents are not correlated to the fraction between radial and azimuthal components in the current.



Minor comments:
Page 15, line 24: the the is repeated
Fixed
Page 16, line 26: too should be to
Fixed
Page 17, line 22: thee should be the
Fixed
Figures 9 and 10: Adding horizontal "zero" lines to the right-hand bias panels will make them more easy to be interpreted.
Zero-lines have been added.

Referee #2

One question that should be addressed for instance is the assimilation of radials (that is, the raw data from each radar station) instead of the "totals", that is the current velocity field. The radial-to-vector mapping procedure, presumably the standard unweighted least-squares mapping, is prone to errors associated with unbalanced radial distributions, which is known to bias the current vector magnitude and direction. Using radials would also increase the spatial coverage – as GDOP (that is, the geometry of the intersecting radar beams) reduces the actual radar coverage to a smaller area. Also, it would reduce costs and timing if a rapid-deployable HF radar system is to be used.

This is a good point and certainly the direction in which we are heading. The code base we used for this study did not allow for direct assimilation of radials, however, and hence we cannot easily test it in the context of this paper. A comment on the potential benefits of using radials is added in Section 5.

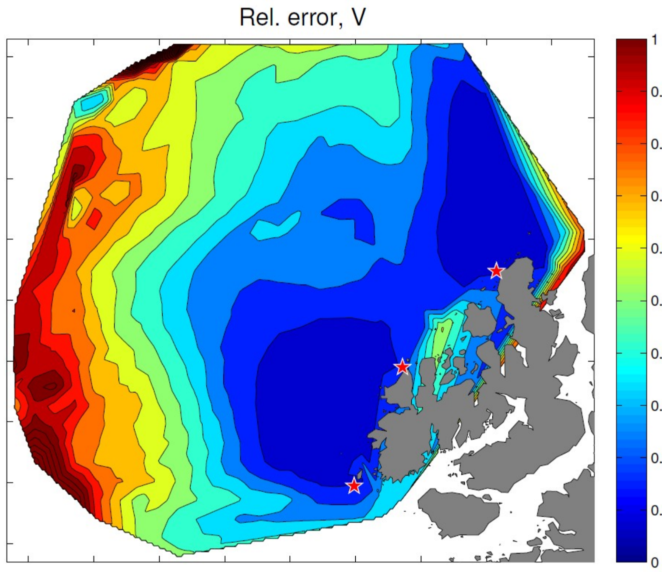
Section 1 provides some brief description of the circulation in the area, but it mentions only a strong northward slope current and high eddy kinetic energy with significant tidal signal without providing any references nor showing results from the actual radar deployment. This is probably not the aim of the present publication but it might have provided some more information for readers not familiar with the area.

We have added a more detailed description, as well as some additional references.

.

Section 2 introduces data and radar errors, mostly associated with GDOP. Chapman et al (1997) is provided as a reference for this error however this formulation is valid for the zonal – meridional components of velocity for two backscatter radars and is probably not fit to the case presented here of three SeaSonde stations. Also, some details and clarifications should be provided as for the so called "observation errors" in the data files. Is the author referring to radial errors? Or current errors? And how are they derived? More details are required as there is some concern and debate within the radar community as for how errors are computed in the data and if those errors should be considered errors.

The CODAR system provided measurement errors that, according to the CODAR documentation, are a combination of signal to noise ratios, velocity variances within range cells, and geometric dilution of precision for two or more systems. Being a test of rapidly deployable systems, we decided to test using the errors provided by the CODAR system. We investigated the average relative error (error divided by observation value), see example for northward velocity on top of next page, and found it similar to what we would expect if the errors were mainly due to GDOP. We would much prefer to put our efforts into assimilating radials directly in future work, which would allow for more accurate treatment of the observation error, and have therefore not considered other ways of modeling these. A brief discussion on this issue is added in Section 2.



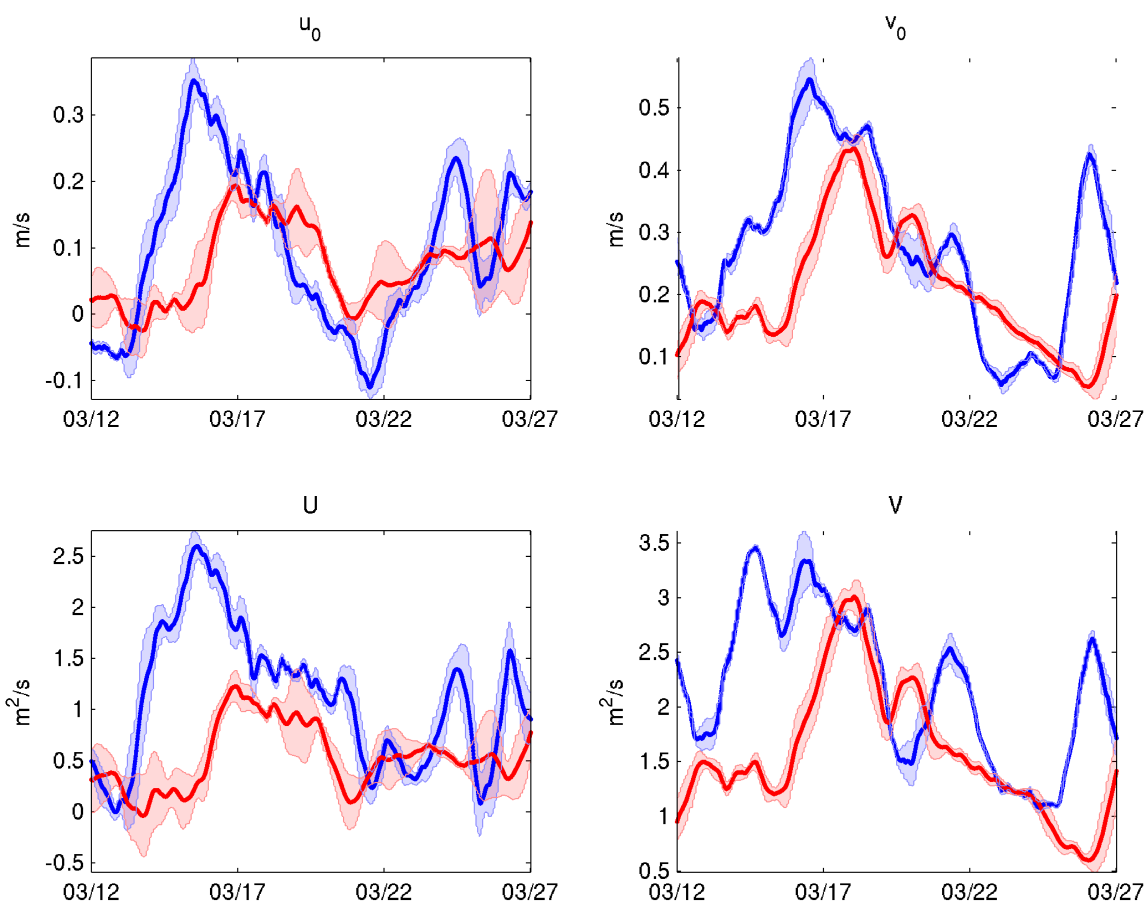
Further concern are related to the choice of the drifters. At the operating frequency of 13.525MHz, the depth of the measured currents is ~1m assuming that the lambda/4pi approximation made in Stewart and Joy (1974) holds; so, the choice of the iSLDMB drifters with a drogue centered at 65 cm below the surface is consistent with the radar vertical scale. Concerns arise for the choice of the iSphere type drifters (spherical surface floats that are half submerged), for which the authors themselves state: "Previous studies have shown that the behavior of these types of surface drifters can be markedly different, mainly depending on the wind and wave conditions (Röhrs et al, (2012)". Limiting the comparison to the iSLDMB would probably make more sense.

The iSpheres were originally only used in the skill score calculations, where the results were reported separately for each drifter type. Only the iSLDMB drifters are in fact used in the vector correlation calculations. We do however see that the inclusion of the iSphere type drifters may be a source of misunderstandings, and do not add any value to this study. We have therefore removed all references to the iSphere type drifters from the manuscript.

The use of virtual drifters is probably a weak point of this work, as other comparison metrics should be used to determine assimilation performance. the presumably large data set from the radar deployment should easily allow for a determination of the variance levels, the average currents and the typical current patterns in the area, which should be compared to model results before/after data assimilation
(see for instance:
Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro, Assimilation of surface velocity data into a primitive equation coastal ocean model, J. Geophys. Res.,107(C9), 3122, doi:10.1029/2000JC000511, 2002.;
Oke, P. R., J. S. Allen, R. N. Miller,G. D. Egbert, J. A. Austin, J. A. Barth, T. J. Boyd, P. M. Kosro, and M. D. Levine, A modeling study of the three-dimensional continental shelf circulation off Oregon, 1,Model-data comparisons, J. Phys. Oceanogr., 32, 1360–1382, 2002a.)
This approach would also help identifying errors and systematic biases in the model itself, related for instance to model set-up, bathymetry (and bathymetric gradients), or model forcing.

We do not fully understand what is meant with "virtual drifters". It was important to us that we could have independent data to compare with the HF observations, hence the deployment of the ADCP and the drifters. We have deployed drifters in this region before, and they are usually carried northwards within a few days, so we have a limited amount of co-located drifters and HF currents. Co-located HF observations and ADCP measurements we of course only have in a single point. There were intermittent technical issues with the HF radars so that unfortunately we do not have a large, complete data set covering such long period as in Oke et al, which could e.g. allow for an EOF analysis of low pass filtered currents. Loss of data from one station resulted in significant loss of “total vectors” during the entire period the station was inoperable (the southernmost station was the most robust). We do have a continuous time series from the ADCP (21 days) and we did some model (free run)/ADCP comparisons. The upper ocean transport and surface velocities (and their variance) are comparable in size, but the timing is not correct. A very similar model setup, albeit with higher resolution, is presented in Röhrs et al, Limnol. Oceangr., 59(4) (2014), with a more thorough validation, and a reference to this paper is added along with the above comments.

The figure below shows 24h running averages of surface velocities (upper left/right) and transport in the upper 7 meters (lower left/right). Here ADCP data is in blue, model data in red. The standard deviations during the 24h periods are indicated. Our apologies for the low quality of the plot, which is due to the use of transparency in presenting the standard deviations.



Referee#3

The authors choose an idealized model configuration of a frontal system to test some parameters of the analysis. However the significance of these results for a realistic system is unclear and this part is not well integrated with the rest of the manuscript. It would have made much more sense to me if the initial test would have been done in the realistic model configuration used later with a classical twin experiment where pseudo-observations from a "truth" run of the model would be assimilated in a perturbed model run.

Similar comments were given by Ref #1. As we stated in our reply, we wanted to isolate the dynamical features we expected would be most problematic in the realistic case and to test various parameters in the data assimilation system using a simpler setup. We refer to our reply to Ref #1 and the added discussion in Section 3..

The assimilation window length of 4DVAR is an important parameter which defines the propagation in time of the information contained in the observations. The authors note in the manuscript that a "slight improvement is obtained when reducing the window length from 72 to 24 h, but there is essentially no difference when the window length is further reduced to 6 h". As this is a central parameter in a 4DVAR scheme, I think that more discussion is necessary. Please show the RMS error (compared to the independent data sets) for different window lengths starting from 72 h and going down to 0 h (which makes the assimilation effectively a 3DVAR scheme). This experiment can be done either with the idealized setup or the realistic case (which would be the most relevant). Authors should also mention the typical time scale of the system. As the manuscript is written now, it seems that a windows length of 6 h is essentially as good as a window length of 24 h (the optimal choice) which leads to the question how appropriate the linearized error propagation under the assumption of a perfect model is (i.e.strong constrained 4DVar)

Additional experiments have been performed with the realistic model, where the assimilation window is gradually reduced from 72 hours down to 3 hours, and a discussion on the outcome of these experiments is added to Section 4.2.

The assimilation of the CTD profiles seems to deteriorate the current forecast. This aspect should be analyzed more detail. In general, the CTD observations should be presented. In the idealized experiment only temperature was assimilated. For the realistic experiment, the authors do not mention if also salinity was assimilated. The authors should also include which RMS error variance was used during the assimilation, how it was determined and how sensitive the validation results with ADCP and drifter are to changes in this parameter.

Salinity was indeed assimilated in the realistic case, thanks for pointing this out. In the idealized case we used a linear EOS and only considered temperature. We have clarified this in Section. 4.2. A more detailed description of the CTD measurements, and the associated errors used in the tests is also presented.

The model simulation for the idealized and realistic configurations is very short (1 day of analysis and 4 day of forecast, with a single assimilation cycle). How statistically robust are the results?

See our reply to similar comment from Ref#1.

Section 2: "...correction of the tidal signal before assimilation, as described in Zhang et al. (2010). As our time series is too short to provide a good estimate of the observed tidal signal, no such corrections have been made": Can you be more specific if tides are included or not in the model and observations?

The model includes tides through its boundary conditions. Due to discrepancies in the bottom topography, resolution etc. the modeled tide may however deviate somewhat from the true tidal signal. The observations also includes tides. To avoid problems caused by different tidal signals, the tidal signal is often removed from the observations, and replaced by the corresponding model tidal signal. To do so, however, requires time series of at least 3 months. That would have left us with very few, if any, current observations to assimilate. More details are added to the manuscript.

Section 3.3: "The 4DVAR schemes implemented in ROMS also has options for multivariate background error correlations, but since the underlying theories are dubious for high latitudes and eddy resolving models, we do not make use of any such options here.": Please provide more information about what approximation are dubious.

The multivariate correlations are based on the assumption of geostrophic balances and the baroclinic contribution to sea surface height as described in Moore et al. (2011). In our case, the Rossby radius is small and the flow is highly unstable and energetic, and in addition the main current is close to the coast and runs along very steep bathymetry, hence we chose not to test the explicit multivariate error correlations. We have added some comments motivating this choice in Section 3.3

Section 3.3: Instrumental error is one error component. Another one (and in general the largest) is the error of representativity. How has this error been dealt with in idealized experiments? The manuscript explains how the error variance that has been used for the assimilation was derived. However, it is not clear if this error has been actually added to the pseudo-observations (coming from the truth model). And if so, which spatial correlation length of the observation error(generally noted the R matrix) has been used?

The resolution of the (idealized) observation and the idealized model is about the same, and the errors were simply assumed to be proportional to range in addition to GDOP. The "range error" values were then adjusted so that the total errors were similar to errors reported from a medium-range HF radar system that was operated in Norway for the period 2007-2010 (Fedje), please also see our reply to Ref #2 w.r.t. to observation errors. We have added some comments in Section 3.4 clarifying this issue.

Section 4: How is the model initialized?
This information is found in the last paragraph in Section 4.1

Section 4.1: "For tracers and baroclinic [the last i is missing in the manuscript] velocities, boundary conditions as described in Marchesiello et al. (2001) are used. During assimilation, however, clamped boundary conditions with a sponge layer are applied." Why did the authors choose a different boundary condition for the data assimilation?

Boundary conditions as described by Marchesiello et al. (2001) are well-tested for the nonlinear model application. These options are however not available in the tangent linear and its adjoint, and it is therefore not possible to use this configuration during assimilation.

Table 3: Why is there a large difference between iSLDMB and iSPHERE drifters during analysis? For the forecast column, does the sum in the skill score also include the time instance from the analysis ? Assume that this is the case, but it could be stated a bit more clearly in the manuscript.

The data provided in the forecast column are derived from the first 48h of the forecast, starting at the end of the analysis. This is stated more clearly in the reviewed manuscript in Section 4.3.1
The iSphere results are removed from the manuscript, as they have limited added value.

Section 4.3.1: "Two examples of modeled vs. observed trajectories are shown in Fig. 8.": A bit more description would be useful.

This was indeed a bit short. We have changed this to
"As an example of the forecast model's ability to predict pathways of drifting objected before and after assimilation, two examples of modelled versus observed trajectories are shown in Fig 9."

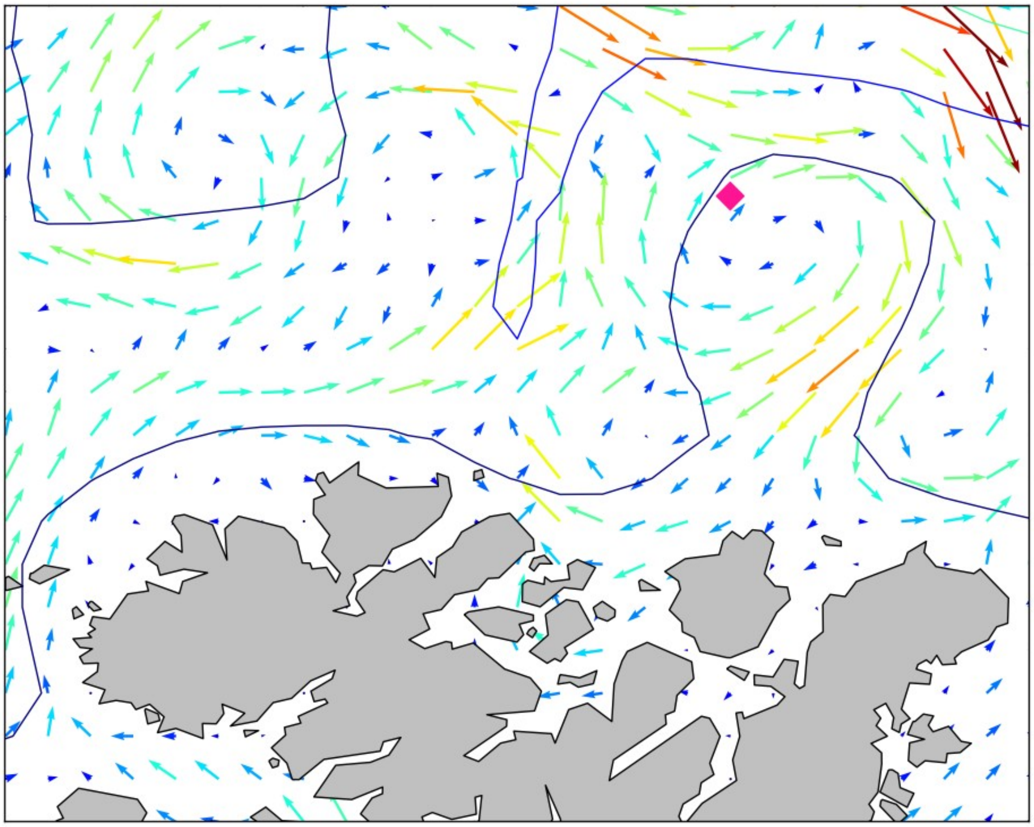
(A new figure has been added to the manuscript prior to the Figure in question)

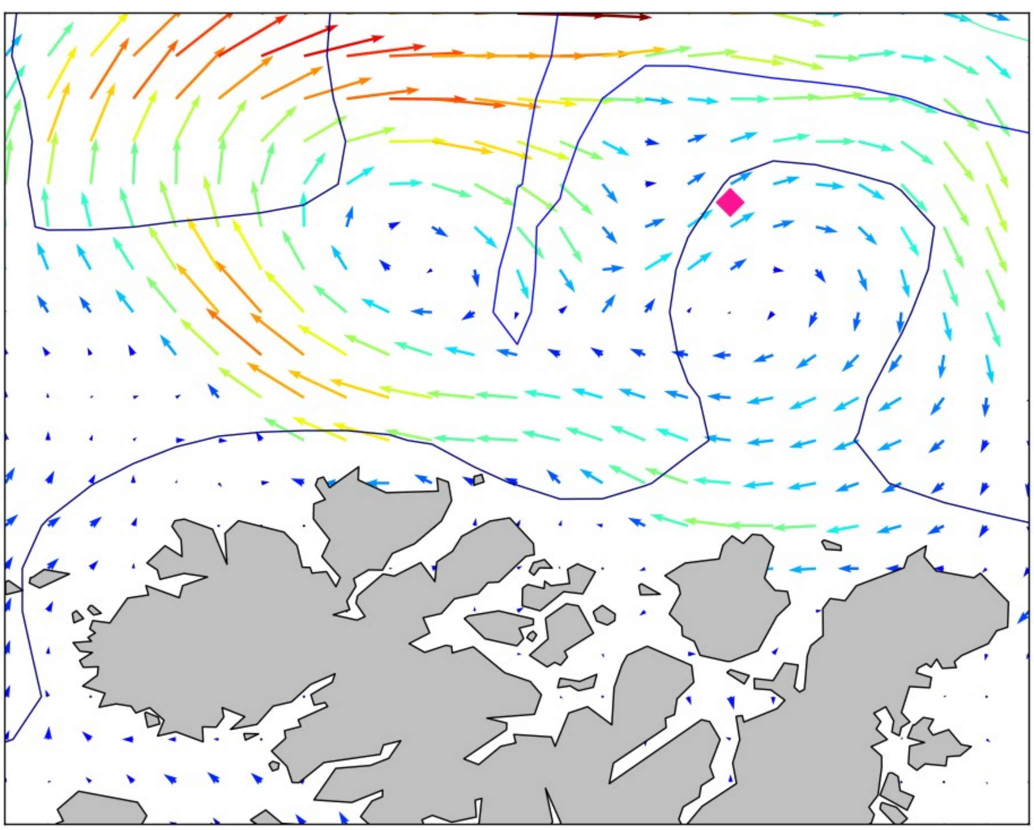
Section 4.3: What is the RMSE of CTD observation in the analysis?

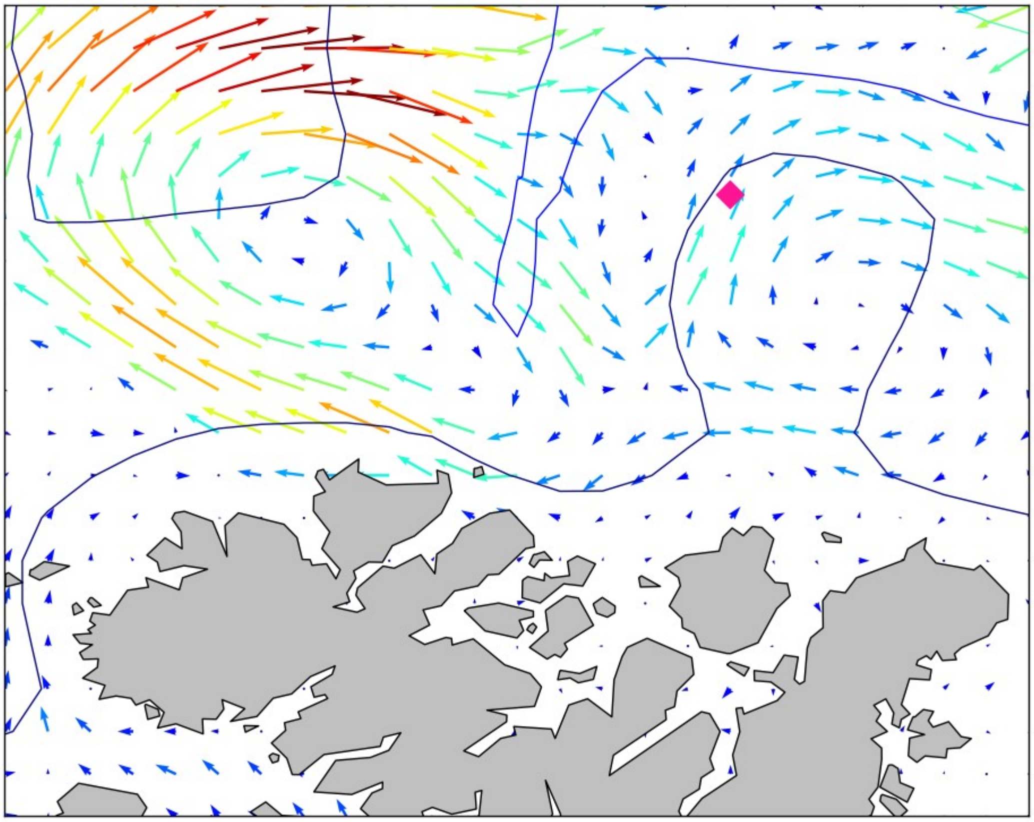
RMS values have been calculated and added to the discussion in Section 5.

Section 4.3.2: "Due to a displacement of an eddy in ALL, this simulation performs poorer for the ADCP location than HF and CTRL during the last days of the forecast." Please show this.

The ADCP comparison has been altered to feature data from 10 sequential assimilation + forecast simulations. The quoted text is thus no longer a part of the manuscript. Figures showing the current field at transition from forecast day +3 to +4 is however shown here for CTRL, HF and ALL. The pink diamond marks the ADCP location. CTRL and HF shows an eddy with approximate the same location, resulting in similar current direction in the ADCP location in the two simulations. In the ALL simulation the eddy is more to the right in the figure, causing the current direction in the ADCP location to be directed more outwards than in the other simulations.


CTRL


HF



ALL

Please include in caption of the Figures 7, 9 and 10 to what the model is com-pared with (drifter or ADCP). It is in the text, but it would be easier for the reader if this is also included in the caption

Captions will be altered to state more clearly which data have been used.