We thank the referee sincerely for the time and effort expended on reviewing our manuscript. This reply addresses all points raised by the referee, considering also comments from the other two reviews.

In the following, the referee's comments are shown in blue.

This paper presents a comparison of the model BSHcmod and TRIM with six nearsurface drifters with little or no windage. The paper is well written (albeit too lengthy), the English is good and the figures are readable. I would recommend publication after minor corrections.

We suppose that concerns about the length of the paper relate to Section 3.2. Please see our response to the last comment below.

Comments

- The work shows that the BSHcmod needs the addition of either about 50% of the surface Stokes drift or about 0.6% windage. The authors concede that this is probably a reflection of the poor vertical resolution of the model as much as it reflects missing windage and/or Stokes drift. I would recommend clarifying that the Stokes drift really *is* missing whereas the windage is probably negligible as the drifter is subsurface save for the antenna.

  We are not sure whether we got the referee's comment right. What is meant by "Stokes drift really *is* missing"? Indeed Stokes drift wasn't taken into account when running BSHcmod and TRIM. For BSHcmod (for which a coupled setup exists) this is now explicitly mentioned at the end of the first paragraph of Sections 2.2.1: "*The option to include Stokes drift from surface wave models (as described in Dick et al., 2001) is not activated operationally so that effects of Stokes drift are also not included in archived surface current data.*"

  Another question is to which extent Stokes drift affects surface drifters in the experiment. Simulations with wave model WAM provide an estimate of the strength of Stokes drift. For more details see our response to the next comment.

- The TRIM results should be studied a little further. Please consider adding Stokes drift to these as well and report which percentage works best. This would help answer the question of how much the Stokes drift really should contribute to an object which sits in the upper metre or so of the water column. Ideally the Stokes drift should be vertically averaged over the upper metre (see Li et al, 2017), but a Stokes drift representative of the midpoint (say 0.5 m) will probably be close enough.

  Of course we also did experiments with variable strengths of windage and Stokes drift in combination with TRIM. The problem is that (as with BSHcmod) a real calibration of factors α and β in Eq. (1) is impossible as errors are very different during different periods of time (with different wind conditions). Following recommendations of referee #2, we introduced a new paragraph at the end of Section 2.2.3 which explains the way strengths of windage and Stokes drift, respectively, were specified: *"The assumed strengths of either wind forcing or Stokes drift resulted from trying to achieve an overall eastward displacement of simulated drifters that roughly agreed with observations. This approach must not be confused with sound model calibration, which seems impossible based on the very limited data available. Models perform differently during different periods and it is hard to distinguish, for instance, between deficiencies in the hydrodynamic model and implications of imperfect atmospheric forcing. Also independent data needed for model validation are not available. However, already the simple approach enables an appraisal of how successful drifter simulations will depend on a distinction between wind drag and Stokes drift."*

  If we would add a description of "calibration" experiments for TRIM, the same would have to be shown for BSHcmod. In our opinion, this would definitely overload the paper. And, what is more, additional effects of windage or Stokes drift for TRIM are identical with those for BSHcmod, as in both cases the same fields (calculated offline) are just superimposed upon Eulerian currents from the respective model.

  Our study shows that effects of windage and Stokes drift can hardly be disentangled based on an experiment like the one we describe. Combining both effects would add another degree of freedom uncontrolled by data. Even if windage for the drifter is negligible (which seems a reasonable assumption, as the referee mentions), the poor vertical resolution of archived BSHcmod data may be remedied by kind of windage for the 1 m depth surface layer. Adding Stokes drift is an alternative option having the same effect. At the end of the first paragraph of the conclusions we added a remark regarding a similar problem when using HF radar currents: *"In a similar way, Ullman et al. (2006) attributed a bias of trajectories predicted based on HF radar currents not to a drifter leeway but rather to the fact that effective depth of HF radar measurements exceeded that of surface layer drifters."*

Although we cannot really answer the question of how much Stokes drift really contributes, it is nevertheless interesting to note that the 50% factor we chose for a reduction of surface Stokes drift is of the order of magnitude that should be expected for an object drifting in a 1 m depth surface layer. In the discussion we had already referred to a paper by Röhrs and Christensen regarding the decrease of Stokes drift with depth (last paragraph on page 23 of the original manuscript). We now added another sentence (highlighted below) stating explicitly that these values are roughly consistent with the assumed 50% factor: *"Based on these formulas, Röhrs and Christensen (2015) calculated in the context of a drifter experiment in the Barents and Norwegian Sea that an average Stokes drift of 8.9 cm/s at the surface contrasted with an average of 3.7 cm/s at 1 m depth. For the present study we neither applied theoretical profiles nor conducted an in depth model calibration.* ***However, in the light of the above numbers, the 50% factor α in Eq. (1) we chose for BSHcmod+S seems a reasonable value for drifters representing a surface layer of about 1 m depth.****"*

We added the reference Li et al. (2017) the referee provided. In their Eq. (23), Li et al. show Stokes drift being proportional to 10 m winds. The factor of 1.6 % is consistent with the data we provide in Fig 9(b) of the original manuscript (Fig. 10(b) in the revised manuscript).

- Please mention in the text after Eq (1) that the full windage is actually a rotation (called the leeway divergence) and not simply a factor *β*.

  We added an explanation of why a drift component perpendicular to the downwind direction was not included: *"Eq. (1) describes windage (or leeway) as a drag in downwind direction, neglecting any crosswind lift component. Such lift component depending on the specific overwater structure of a drifting object is crucial for search and rescue (Breivik and Allen, 2008}. For surface drifters used in experiments, however, these effects should be negligible."*

- Section 3.2 is too lengthy. Please consider moving some of this verbiage to an appendix.

  This comment agrees with a comment of referee #1 who suggested an abbreviation of Section 3.2 by 50%. After a removal of too many details we actually achieved this. We believe that moving part of the listing to an appendix would not improve readability as it would destroy the clearly structured chronological listing, enabling a quick scan for specific events while skipping others. It would also imply a shift of corresponding figures. These figures, however, provide relevant information, which referee #2 even suggested to expand. We followed his advice and complemented Figs. 7 and 8 by a third figure (Fig. 9) so that now 12 instead of 8 example days can be shown (we compensated for that by a removal of Fig. 6, which just combined panels from different figures in the appendix). As a result, the reader is now less often forced to switch to supplementary material.

References

- Allen, A. and J. V. Plourde, 1999: Review of Leeway: Field Experiments and Implementation. Tech. Rep. CG-D-08-99, US Coast Guard Research and Development Center, 1082 Shennecossett Road, Groton, CT, USA, available through http://www.ntis.gov.

  This report is cited in Breivik and Allen (2008). We did not add the reference as our specific study does not deal with search and rescue problems.

- Li, Q., B. Fox-Kemper, Ø. Breivik, and A. Webb, 2017: Statistical Models of Global Langmuir Mixing. *Ocean Model*, **113**, 95–114, 10.1016/j.ocemod.2017.03.016.

  This article has been added as another reference for a constant ratio of surface Stokes drift and 10 m winds.