Final author response for the manuscript tc-2015-233 submitted on 24 Dec 2015 with the title:
"Sea ice diffusion in the Arctic ice pack: a comparison between observed buoy trajectories and the neXtSIM and TOPAZ-CICE sea ice models"

by Pierre Rampal, Sylvain Bouillon, Jon Bergh, and Einar Ólason

Anonymous Referee #1

Received and published: 25 February 2016

Dear referee,

First of all, we would like to thank you for your in-depth review, insightful comments and suggestions, which greatly helped us improving our manuscript. Like all the other reviewers you agreed on the interest of the sea ice diffusion analysis and suggested to emphasis that part of our work in the revised version, while reducing the weight given to the inter-comparison of the two sea ice models. We agree with this point and, therefore, largely restructured the manuscript following your recommendations and those of the other reviewers. Our manuscript is now entitled: "Arctic sea ice diffusion from observed and simulated Lagrangian trajectories". In addition to the restructuring, we tried to answer your questions and comments as carefully as possible.
Below, you will find your original **comments in bold**, our answers in red and the **added text** to our manuscript in **bold red**.
Please note that our new manuscript containing all the changes we made is attached as a supplementary material to the present document.

The manuscript describes a sea-ice drift field analysis of two different sea ice model in the context of buoy drift data. The paper is clearly written (with a few smaller problems that I marked in the annotated PDF, e.g., sometimes the language is a little sloppy and has a colloquial tone not appropriate for a scientific text), but the scientific focus of the paper is not very clear. On the one hand, the "main goal" is to evaluate the models (and there are many figures comparing these models), but on the other hand most of the conclusions section focusses on the "secondary objective" to illustrate how statistics of sea ice drift can be useful. Frankly I find the "secondary objective" more interesting and scientifically better handled (although not much of that is really new) than the "main goal". The model results of the neXtSIM model are remarkable, but the model comparison is biased and the conclusions that are drawn are likely to be either very specific or misleading. Therefore I recommend a major revision to redefine the focus of the paper properly (or change the weight given to the topics in the conclusions section) and change the nature of the model comparisons. I include an annotated PDF with notes and comments that I made while reading the text. They are meant as well-meaning suggestions (or can be found again below).

We agree that the objectives of our study were not correctly presented in the first version of the paper and that the structure was not appropriate.

The paper has been restructured with a stronger introduction presenting the context and the objectives of the study. In the first part of the paper we present the method used and its application to a reference dataset, whereas the second part of the paper presents the application of the diffusion analysis to simulated and observed trajectories. The highlight is not anymore on the comparison of the model results but on their comparisons to observations in the context of long-term trajectories modelling. The conclusion and the abstract have also been completely rewritten.

Details of my critique:

The model comparison is not very meaningful because the models start from different initial conditions, are driven by different atmospheric forcing and ocean conditions. In fact, all of these aspects tend to favor the neXtSIM simulation (initialization from observations rather than spinup, removed thickness bias, higher resolution forcing data, assimilated surface ocean) with respect to realism, so that I find the conclusion that the neXtSIM model performs better very much confounded by the totally different initial (important for a short integration period) and boundary conditions.

We agree that we cannot conclude that a model is better than the other one.
All the statements related to model comparisons have been removed.

At the same time the text seems to try to "sell" neXtSIM, which appears to inappropriate for a scientific paper.

Yes, we agree, it may seem like that.

The new version of the manuscript contains a longer analysis and discussion of the results, and also highlights the differences between the results of neXtSIM and the observations.

E.g. I can clearly see from Fig6+7 that neXtSIM statistics are better than those for TOPAZ, but at least in Fig7, the performance is not as great as the authors are trying to make me think (l382/Fig7: "follow an exponential distribution" may be true below 30cm/s, although it's hard to know if this significant, but clearly it is as bad as TOPAZ above 30cm/s which goes almost unmentioned in the text, or:

Yes, we agree, and this is a very good point as it indicates that the two model setups have similar deficiencies for the highest values of the fluctuating speed.

The range in which the exponential distribution is followed is now indicated and the missing high values of fluctuating speeds are discussed.

On page 13, ll388 the authors discuss the initial and boundary value issues and acknowledge them as the weak point in the comparison, but still state that they expect TOPAZ to be a reasonable reference for other models, which is not based on any evidence. From Bouillon et al (2013), we know that standard EVP is not getting it right.

As the quality of the simulated trajectories seems to greatly depend on the details of the simulation setup (initial conditions, parameters, forcings), it would be hazardous to generalise the conclusions coming from one specific setup to other simulations obtained with different setups and models.

We removed the statement of using TOPAZ as a reference for other models and now insist on the use of the diffusion analysis as a prerequisite before further analysing simulated trajectories.

neXtSIM is said to be tuned to fit observations but TOPAZ uses a drag coefficient that leads to too high drift speed? This adds to the non-comparability of the models. The first thing I would have tried is to reduce the TOPAZ drag to 1e-3 to reduce the drift speed bias.

The values of the parameters of the neXtSIM setup are those found to be optimal in Rampal et al. (2015, still under review) for the same forcing and a similar setup. The values of the parameters of the TOPAZ setup are those found to be optimal and currently used for the operational forecast platform (only documented in an internal report). As the overestimation of the drift in the TOPAZ setup is not homogeneous but mainly localised along the CAA where the ice is thick and almost immobile, playing on the air drag parameter would probably not help much.

The reference to the full description of the two setups are now added in the setup descriptions.

Or use the re-analysis/assimilated solution of TOPAZ to begin with.

We are not sure to understand this suggestion. If the reviewer means that we should analyse the reanalysis instead of a free-run of TOPAZ, we do not agree because it would require to rerun the reanalysis as the archived outputs do not have a high enough temporal resolution, and that would be too expensive and difficult to handle (100 members to analyse). If the reviewer means that we should start the free-run from the reanalysis, we agree that it would maybe be better but still not enough to have a proper comparison of the two models (we should use the same atmospheric and oceanic forcings, the same thermodynamical model, tune all the parameters of the models with the same approach, etc...). However, as the goal is not the comparison of the model setups, but their respective evaluation compared to observations, the two model setups do not need to be the same.

No changes

c_a = 0.0076 is a drag coefficient much higher than "standard" (although I acknowledge that c_a has a tuning range). And c_a has been measured or at least been inferred from observations, e.g., SHEBA observations have a mean of 1.7e-3, a general range is maybe between 1e-3 and 2e-3 (e.g. Nguyen et al 2011), maybe locally values of 5e-3 are OK, but such a high global value smells like a problem in the model (that is compensated by the high drag coefficient). I have had many discussions with meteorologists who work on atmosphere-ice-drag (both observations and atmospheric models) about the functional form and value of the drag coefficients and believe me, they would not accept drag coefficients outside the range of observations. It is surprising that neXtSIM still produces so slow drift. Should be discussed in a few sentences, what is compensated by this value (forcing?) and why nothing is done to adjust the TOPAZ value.

The value of the air drag parameter used for the neXtSIM setup is the one found to be optimal in Rampal et al. (2015, still under review) for the same forcing (ASR-i 30 km) and a similar setup. In that paper, we did the same exercise for the ERA-interim forcing and found an optimal value equal to ca = 0.0023, which is in the range of classical values. Those values have been optimised for free drift events only and are then independent to the mechanical parameters (ice strength,...). The high value found for ASR-i then does not compensate any "smelly" problem in the sea ice model but is directly linked to the low bias documented for ASR-i surface wind. This is discussed in details in Rampal et al. (2015, still under review). The value of the air drag coefficient for TOPAZ is the one optimised for the operational forecast system (see also the answer here above).

We now clearly indicate that the parameters used here for the two models are those found optimal in previous study/report for similar setups and forcings. The information asked by the reviewer can be found in Rampal et al. (2015, under review), whereas the information on the optimisation of the air drag coefficient for TOPAZ is not publicly available.

What's wrong in summer in the model(s) that everything is focussed on winter? Summer is notoriously harder to simulate with "classic" models. Does neXtSim offer new opportunities or new problems? Is this discussed anywhere (in the Rampal et al TCD paper, doi:10.5194/tcd-9-5885-2015 ?)?, otherwise it would be good to say a little about this somewhere in the text.

We do focus on winter because it has been identified as critical for pollutant tracking. We do not completely agree with the sentence: "Summer is notoriously harder to simulate with "classic" models.". It may be true for the thermodynamics (due to many processes, melt ponds, albedo feedback,...) but not for the dynamics. We suspect that the ice being thinner and less concentrated, the rheology will play a less important role in summer, and sea ice dynamics may then be more easily reproduced even with simple free-drift models. For that reason and also due to the availability of the data, we have so far only analysed the simulated sea ice drift and deformation for winter periods (Bouillon and Rampal, 2015; Rampal et al., 2015, still under review).

We now justify and document why we focus on the winter season in the context of pollutant transport by sea ice.

I suggest to either drop the comparison to TOPAZ, because it is unfairly couched, or repeat the comparison on “equal footings”.

We agree that the proposed analysis of model outputs cannot be used for model comparison and we then drop any statement on model comparison. As the definition “equal footings” would always be debatable and could be expanded to all the aspects influencing the simulations (initial conditions,external forcings, resolution, computational time, number of tuning parameters, number of degrees of freedom, tuning method,...), we prefer not to follow the second suggestion.

We now evaluate separately the results of the TOPAZ setup and those of the neXtSIM setup by comparing them to observations. All the sentences about comparing models have been removed.

The data analysis and suggestions for using ice drift statistics are interesting and appear useful, but it should be made more explicit, what is new compared to Rampal et al. (2009), and what it reproduced from Rampal et al. (2009).

We agree.

The similitude and differences compared to Rampal et al. (2009) are now clearly stated.

Sometimes, the buoy data analysis start unexpectedly, e.g. l397, and while most aspects are compared to model simulations, the result of Fig8 are not (why?).

We agree.

We now clearly split the analysis of the reference buoys data set and the analysis of model simulations. The presentation of the method is also better structured, starting with the presentation of the reference data set, the decomposition of the sea ice motion and then the application of the diffusion theory. This section ends with the presentation of the results of the analysis on the reference dataset and the discussion.

Fig8, compared to Fig14 of Rampal et al (2009) has 10 times larger values of r’. Why is that so? As far as I can see, if really r’ =sqrt(<r’ˆ2>) is plotted then, I would expect no larger differences of the geometrical sum than a factor of sqrt(2).

The magnitude of the displacement in Fig14 and variance in Fig15 of Rampal et al (2009) are wrong by a factor 10 and 100, respectively. This can be verified by looking at the inconsistency with the value given for the absolute diffusivity.

We now discuss this difference along with the explanation on how we checked the consistency between the computed and estimated values of the fluctuating displacements.

Please also note the supplement to this comment: http://www.the-cryosphere-discuss.net/tc-2015-233/tc-2015-233-RC1-supplement.pdf
Our responses to these additional remarks are made in the pdf attached as supplementary material

Final author response for the manuscript tc-2015-233 submitted on 24 Dec 2015 with the title: "Sea ice diffusion in the Arctic ice pack: a comparison between observed buoy trajectories and the neXtSIM and TOPAZ-CICE sea ice models"
by Pierre Rampal, Sylvain Bouillon, Jon Bergh, and Einar Ólason

Anonymous Referee #2

Received and published: 2 March 2016

Dear referee,

First of all, we would like to thank you for your in-depth review, insightful comments and suggestions, which greatly helped us improving our manuscript. Like all the other reviewers you agreed on the interest of the sea ice diffusion analysis and suggested to emphasis that part of our work in the revised version, while reducing the weight given to the inter-comparison of the two sea ice models. We agree with this point and, therefore, largely restructured the manuscript following your recommendations and those of the other reviewers. Our manuscript is now entitled: "Arctic sea ice diffusion from observed and simulated Lagrangian trajectories". In addition to the restructuring, we tried to answer your questions and comments as carefully as possible.
Below, you will find your original **comments in bold**, our answers in red and the **added text** to our manuscript in **bold red**.
Please note that our new manuscript containing all the changes we made is attached as a supplementary material to the present document.

The manuscript "Sea ice diffusion in the Arctic ice pack: a comparison between observed buoy trajectories and the neXtSIM and TOPAZ-CIEC sea ice models" by Rampal, Bouillon, Bergh, and Olason consist of two parts: (1) a model inter- comparison between TOPAZ-CICE and neXtSIM including validation with observed ice displacement, and (2) an analysis of the diffusive processes going along with sea ice motion. The paper is well written; figures are clear and support the arguments.

While the second part is very interesting and holds novel results useful beyond the scientific community, e.g. also for oil exploration, the model inter-comparison part suffers from the unfortunate experimental design (see details below). I thus suggest focusing on part 2 as the main message of the paper and reduce the extent of the pure model inter-comparison. I think that studying sea ice diffusion with two types of sea ice models – a "traditional" one represented by a version of CICE, and the "novel" neXtSIM model – is a great addition to earlier work on this topic by the first author (referenced as Rampal et al., 2009b in the manuscript), in particular as both are designed for sea ice forecasting systems. This should be emphasized more, possibly by changing the sequence of the presentation of results with the model inter-comparison becoming more of a model validation against the same buoy data in both cases.

We do agree and thank the reviewer for this excellent suggestion.

We have restructured the paper accordingly to this suggestion. The introduction has been rewritten to better justify the objectives and the context of the study. In the first part of the paper we present the method used, its application to a reference dataset and how the results could be used for pollutant transport modelling. In the second part of the paper, the diffusion analysis is applied to simulated and observed trajectories to evaluate the quality of the simulated trajectories. The conclusion and the abstract have also been completely rewritten. We do not put anymore the focus on the short-term forecast but on pollutant transport modelling because the results present here are mainly pertinent for long-term trajectories.

The manuscript may be acceptable for publication after major revisions. In particular, I strongly recommend shifting the focus of the manuscript from the model comparison to the diffusion analysis and provide additional information for commercial applications.

We do agree.

We shifted the focus on the diffusion analysis and removed all the statements about model comparison. We better present how the information retrieved from the diffusion analysis may be used for passive tracer modelling, response planning and for the evaluation of simulated sea ice trajectories. We also discuss the limitations of the different approaches and the possible future work.

I apologize for providing my review late.

Major comments:

From the beginning, incl. the title, the authors misleadingly address TOPAZ-CICE as “sea ice model” where it is in fact a coupled sea-ice/ocean model. This needs clarification. In this respect: I think the title does not need to include the model names but could be shortened to “Sea ice diffusion in the Arctic ice pack from observations and models”. The generalization is in order as TOPAZ-CICE represents a whole suite of “classic” sea ice models (as also stated in line 394f).

We agree and thanks the reviewer for this good suggestion.

The title has been changed and does not contain anymore the name of the models. We better present the TOPAZ system and clearly define what we call the “TOPAZ model”. The link with CICE is limited to using the implementation of the EVP rheology that was in CICE version 4. This is now indicated in the text, and the name “TOPAZ-CICE” is not used anymore.

The design of the experimental set up has certain deficiencies that I think inhibit a clean comparison of the two sea ice dynamics models:

We agree.

The presentation of the first manuscript as a model intercomparison was not appropriate. No more comparison between the two models are proposed.

1. TOPAZ-CICE is a coupled sea-ice/ocean model whereas neXtSIM is a sea ice model run in stand-alone mode (swamp ocean with nudging to TOPAZ reanalysis, which differs for the TOPAZ run used here). On short time scales of a few days the ocean forcing should not matter as the ice motion is mostly wind-driven. But this is not as clear for periods of 30 days and longer. This needs to be explained and the bias quantified if possible.

We do not fully agree. For short time scale the inertial oscillations are known to play an important role and their representation highly depends on the coupling between the ice and the ocean. For longer time scale, we agree that the mean circulation is influenced by the mean circulation in the ocean.

We now discuss the missing representation of the inertial oscillations as a factor for underestimating short term variability. Evaluating such process would however require specific data (with higher temporal resolution) and is left for future studies, along with the implementation of alternative ice-ocean coupling methods. We also better discuss the link between the mean circulation of sea ice and of the ocean. As no model inter-comparison is done, we do not need to discuss in details the differences between the ocean states of the TOPAZ free-run and re-analysis.

2. The atmospheric forcing of the two models differs. Wind forcing is a key driver of sea ice motion on short (daily) timescales however. This must have a major impact on the results, which briefly noted but not discussed.

Yes, and we can actually quantify the differences between using ASR and ERA interim. When using the optimal air drag parameter found in Rampal et al. (2015, still under review), we estimated the error between simulated and observed 3-days drift (using the GloBICE data set), and we found the error when using ERA interim being 50% larger than the error when using ASR interim. This information was not written in that paper, but it will be added.

We add the reference to Rampal et al. (2015) in the description of the neXtSIM setup so that the reader can find the needed information. As the model setups are not compared, we do not need to discuss the differences in the atmospheric forcings, but we state in the conclusion that looking at the impact of the atmospheric forcing resolution would be the first thing to do for further analysing the causes of the missing high velocities.

3. Sea ice drift tracks used to compare the two models with each other and with observations are computed during post-processing using hourly model output from the

TOPAZ system and during run-time in neXtSIM. It should be demonstrated that this has no major impact on the results.

We checked that by comparing a small set of trajectories simulated off-line with hourly sea ice velocities and simulated online with the tracer module of TOPAZ, and we found no significant differences. We do not use this online tracer module because it does not work properly near the Pole. Furthermore, as we only look at time scales larger than 12 hours, using hourly velocity fields should not impact our statistics. Using daily or even weekly value, as it is sometimes done in passive tracer modelling, would indeed have an non-negligible impact.

No changes, as we already explained in the text that we checked the validity of the offline tracking approach.

While I understand that both models are tuned to produce "best guesses", the authors need to present more convincing arguments that the uncertainties associated with above differences are smaller than the errors originating from the sea ice dynamics that are compared. Differences in the wind forcing, drag coefficients (0.0016 vs. 0.0076), ice strength and resulting ice thickness distribution can already explain some of the reported biases. It needs to be shown that this is not the case or that these biases have different characteristics and can thus be separated from ice dynamics issues.

We acknowledge that the presentation of the original manuscript suggested that we wanted to compare two sea ice models. Especially when expressions like: "neXtSIM performs better than TOPAZ-CICE", "the performances of the two models",... are used. Such sentences did not correctly reflect our intentions, which was to evaluate the long-term trajectories simulated by the standard free-run of the TOPAZ system and by a standalone simulation of the neXtSIM model. It is an abuse of language to say the "the performance of this model", we should have written "the performance of the simulations presented here and using this model". The term comparison was used to express the comparison of two different model setups, based on different model, initial condition, forcings,...

We now clearly describe the two model setups used to produce the simulations. We drop all the statement referring to model inter-comparison, as we just evaluate separately the simulated trajectories against observations. As suggested by the reviewer, we try to better discuss the causes of the identified biases. This is done in the results section of section 3. The consequences of the identified biases are now discussed in the discussion section of section 3.

Further, I strongly recommend restructuring Section 2. I would expect this section to feature three short sub-sections on "IABP buoy data", "The TOPAZ-CICE model", and "The neXtSIM model" just stating retrieval of sea ice velocities from the buoy data and the model set ups and experiments being used. For the buoy subsection the current texts at the beginning of section 2 and in section 2.1 should be merged. Finally, these the three sections could be followed by a section "Modeled trajectories" describing the

derivation of “float” trajectories from both models. Although I think such a subsection would rather belong to Section 3 Methods.

Yes, this section needed to be restructured along with the rest of the paper.

Due to the restructuration of the paper, the reference IABP data set is now described in section 2 where we present the analysis and its application to the full IABP data set. Section 3 starts with a subsection “Observed and simulated trajectories data sets” containing the 3 subsections suggested by the reviewer on IABP, TOPAZ and neXtSIM.

Some minor remarks (by line number):

line 1: “. . . activity in the Arctic . . .”
Corrected at 2 locations in the text

4: “. . . simulated by the ocean/sea-ice model TOPAZ-CICE and the stand-alone sea ice model neXtSIM. We compare . . .”
Corrected with a complete restructuration of the abstract.

16–20: This aspect of the paper including related results of time and length scales should be emphasized more.
The link with the advection/diffusion equation or with Lagrangian passive tracers is now better described.
See the changes in the introduction and the discussion in section 2 and 3.

64: “. . . as follows: data sets and models are presented . . .”
Corrected with a complete restructuration of the paper.

69: either “We use all 12-hourly buoy positions ...” or “we use the full data set of 12-hourly buoy positions . . .”
Corrected

73: “. . . we also generate virtual buoy trajectories by simulating “floats” in the TOPAZ and neXtSIM models. The float simulations are initialized . . . buoys and stopped when the IABP buoy . . .”
Corrected

76: please rephrase sentence for clarity: “After removing . . .”
Corrected. This description of the treatment of the non-overlapping periods was not necessary as it only happens near the ice edge, which is excluded from our domain of analysis.

86: from this statement it is not obvious that the 35-day trajectory chunks are independent since, for instance, the starting point of such a segment relates to the last

10 days of the previous segment. However, in line 406 you mention that only every 10th segment is used. These are independent estimates, I agree, but this must be stated near line 86 as well.
The condition of independence is not needed. We actually realised that it is better to have as many 35 days trajectories as possible, which is done by starting the 35 days periods every 12 hour. This is now explained in the text in the discussion of section 2.

97 & 99: change "were" to "are" if this is still the standard procedure for the buoy positions and does not refer to some method used in the past
Corrected

104: rewrite sentence for clarity, e.g. all buoy data north of 70◦N are used if at least 100 km from coasts, only between longitudes XXX (Greenland) and XXX (Severnaya Zemlya) the southern limit is 80◦N.
Corrected

110: I don't think "embark" is the right expression here, maybe "installed"?
Corrected

114: at this point it is not clear why Figure 1 shows all buoy data from 1979-2010 since it seems that only years 2007-2010 are considered for the analysis. This should be clarified, e.g. by adding the statement from line 294. (also see comment on restructur- ing Section 2 above).
Corrected

118: it would be nice to support this statement with a figure showing a single buoy track. In fact, Figure 4 shows just this. Please consider moving this figure to become Figure 2 to be referenced here. (You must not yet address the separation of mean and fluctuation also shown in the plot.)
Corrected

128: I guess it is not the sea ice strength set to a constant value but rather the ice strength parameter (P*)? Or is 27500 N a model mean?
Corrected

151: replace "kills"
Corrected

153: explain how the interpolation is done
We add "(bilinear interpolation)".

172: rephrase: ". . . and finishing on May 15th for three consecutive winters from 2007 to 2010.
Corrected

177: provide reference for low-biased TOPAZ ice thickness or show ice thickness distributions. In fact the latter would be a helpful additional plot to get a better idea of the behavior both models.
As suggested by the two reviewers, we will not present a detailed comparison of the two models. The suggestion of adding a figure comparing the sea ice thickness fields will then not be implemented. The plots comparing the results from TOPAZ to observations can be found in Sakov et al. (2012).
The low bias of TOPAZ thickness is discussed in the references now added to the text. The plots for neXtSIM can be found in the paper "neXtSIM: a new Lagrangian sea ice model" that is still under discussion and already cited several times in the text.

212: ". . . for the decomposition of the motion into a mean and a fluctuating part, $u = \bar{u} + u'$,wefollow..."
Corrected

220: "evolve" (remove "s")
Corrected

229 "... is referred to as ..."
Corrected

304 ". . . in the IABP data set, This is most pronounced in the winter 2007/2008, in which short IABP trajectories . . . Archipelago are the result . . ." Again, it would be helpful to see the modeled ice thickness distribution.
For the additional plots, see our response above on the remark on line 177.
Corrected

317: remove "and so we did"
Corrected

328: remove "and shall be"
Corrected

332: remove "Unsurprisingly,"
Corrected

334: I believe the correct term is Transpolar Drift Stream 339: replace "size" by "extent"
The term Transpolar Drift is also commonly used (e.g. http://oceans.taraexpeditions.org/wp-content/uploads/2014/08/Exploring-Arctic-Transpolar-Drift-During-Dramatic-Sea-Ice-Retreat.pdf)
Corrected

340f: remove "in the "lower" part of the Beaufort Gyre," It is not clear what "lower part" means and "along the Canadian Arctic Archipelago" is sufficient anyways.

Corrected

335–345: the whole paragraph relies on a rather visual, qualitative comparison. I would prefer to see a scatter plot of daily observed vs. modeled drift velocities and the associated regression for a more objective and quantitative comparison.
Looking at scatter plot of daily drift would not help in distinguishing the bias at representing the mean and the fluctuating parts of the sea ice motion. The visual inspection is useful to identify the regions where the mean circulation is not well reproduced and thus determine what processes could be missing. This would not be highlighted with a scatter plot.
When presenting the model, we add a reference to Rampal et al. 2015 (under review) where such a scatter plot is presented.

352: mean value of the mean drift: the PDF in Figure 6 somewhat emphasizes the extreme values at 20–25 cm/s. While this needs to be mentioned I think that the median should be used instead of the mean to limit the influence of these extremes.
For exponential distributions the mean and median are just related by a factor sqrt(2). Using one or the other is then equivalent.
No Changes.

383–396: This paragraph belongs in a discussion section, which is missing at all, by the way. Really, the model differences should be discussed in more detail with respect to the differences in the experimental set up. (also see major comment above).
Yes, we agree.
The results and discussion are now in two separate subsections. The model results are not directly compared, so that we do not need to discuss in details the differences in the set-up. We enhance the discussion on the causes of the biases.

397: I strongly recommend starting a new section here, possibly entitled "Sea ice diffusivity". In my opinion the results presented in lines 397–469 are the more interesting ones. Consider to show these first and present all other results as model validation thereafter.
We now dedicate section 2 to the application of the diffusion analysis on a reference data set. We present the model validation afterwards in section 3, as suggested by the reviewer.

417–426: This part should be expanded. These numbers are really what a community interested in applications would be interested in. Consider providing a look up table listing all numbers for 1, 2, and 3 standard deviations for 5, 10, 20, and 30 days (decision makers may view a 70% or 95% chance to find the polluted floe as high already). Possibly also check numbers from before 2000 from buoy data to demonstrate a temporal evolution.
To split the dataset for years before and after 2000 will lower the statistical significance, which highly depends on the number of data. As the data after 2000 are part of the data analysed here, they will be included in the searching area when using 3 standard deviation.

We add those numbers in an additional table as suggested.

Figures:

Consider to shift Figure 4 to be shown earlier (see comment above).
Done

Also add a marker at the starting point of the track.
We think it is clearer to add this sentence to the caption: "The starting point here defines the origin of the axes."

Figure 8: in upper panel dashed red line needs to be thicker. Examples of the "search area"' would gain a lot from being shown bigger and with less trajectories, also indicating 1 and 2 standard deviations. Demonstrating the search radius is a key figure of the paper. Please make an effort to improve the graphical presentation.
Corrected

Figure 9: labels of slopes "1" and "2" not clear; explanation missing in caption. I suggest using more intuitive labels, such as "Brownian" and "ballistic" or "Eq. 7" and Eq. 6".
We kept the labels but improved the caption.

Final author response for the manuscript tc-2015-233 submitted on 24 Dec 2015 with the title: "Sea ice diffusion in the Arctic ice pack: a comparison between observed buoy trajectories and the neXtSIM and TOPAZ-CICE sea ice models"
by Pierre Rampal, Sylvain Bouillon, Jon Bergh, and Einar Ólason

Anonymous Referee #3

Received and published: 17 March 2016

Dear referee,

First of all, we would like to thank you for your in-depth review, insightful comments and suggestions, which greatly helped us improving our manuscript. Like all the other reviewers you agreed on the interest of the sea ice diffusion analysis and suggested to emphasis that part of our work in the revised version, while reducing the weight given to the inter-comparison of the two sea ice models. We agree with this point and, therefore, largely restructured the manuscript following your recommendations and those of the other reviewers. Our manuscript is now entitled: "Arctic sea ice diffusion from observed and simulated Lagrangian trajectories". In addition to the restructuring, we tried to answer your questions and comments as carefully as possible.
Below, you will find your original **comments in bold**, our answers in red and the **added text** to our manuscript in **bold red**.
Please note that our new manuscript containing all the changes we made is attached as a supplementary material to the present document.

The study aims to evaluate the sea ice drift fields of two sea ice models with different ice rheology and coupling. They are mainly interested in sea ice diffusion accuracy of ice trajectory forecast as, e.g., needed for prediction of oil spill dispersion. In addition, the study derives useful metrics for ice diffusivity from buoy tracks.

In general, this is an interesting paper suitable for the Cryosphere. The model results and metrics derived from the buoys are, e.g., useful to guide future research on oil spill modeling in sea ice. The study provides, e.g., the time development of a potential oil spill search radius and which of the two model systems currently would be more suitable to predict the ice diffusivity more realistic.

The clear shortcoming of this article is that it is geared towards promoting the neXtSIM model system, which would not be a problem if the judgment always would be fair. This is not always the case. For example, the TOPAZ mean drift is too fast while the neXtSIM mean drift is too slow. While the too fast TOPAZ drift is mentioned and exposed several times the too slow neXtSIM drift is only mentioned at the margins. The study cannot solve the question if the difference in mean drift and diffusion observed between TOPAZ and neXtSIM is intrinsic to the models, i.e., ice rheologies, or the different initial

conditions and atmospheric forcings (TOPAS: thinner ice, lower resolution forcing -> can cause faster ice and less drift fluctuations as observed here). While mentioned in the study these points should be stated more clearly. In the end, however, it is clear and I agree that neXtSIM performs better in the current setup than TOPAZ.

We agree with the main remark of the reviewer, which is in line with the two other reviewers.
We have adapted the manuscript to implement their suggestions.

Find more details comments below. After these points are improved I recommend the article for publication in the Cryosphere.

Page, line

1,1: I would suggest to add more information to the abstract about the buoy only results obtained in this study (e.g. p14, Fig.8), which are relevant for e.g. oil spill forecast. If possible this could also be reflected in the article title
We add this sentence to the abstract:"We discuss how these values are linked to the evolution of the fluctuating displacements variance and how this information could be used to define the size of the searching area around the position predicted by the mean drift." and we change the title.

3, 59: the sentence sounds as if you are using the same buoy dataset as in Rampal, 2008, 2009. But this cannot be the case as the time series is longer. Do you mean you are using the same pre-processing? Otherwise why not reference the original IABP data?
Yes the sentence was not clear.
It is now clearly stated that it is the same method as in Rampal et al. (2009) that is used. The reference to Rampal et al. (2008) is not used anymore as it refers to another type of analysis of the same data.

6, 142: what is an "off-line float tracking system"? Mention that this will be described below.
This part of the text have been rearranged to be clearer.

6, 164: slab ocean: this is a strong difference to TOPAZ and the differences should be discussed more.
Yes, you are right, that could have been important if we were interested by doing model inter-comparison.
We dropped all the statement about model inter-comparison.

7, 176-179: if you are initializing with the TOPAZ ice thickness, why don't you use that for the comparison of the two models? That would be much fairer. By increasing the ice thickness in neXtSIM but not in TOPAZ it can be expected that the ice in TOPAZ is moving faster for the same forcing. If possible you should repeat the experiment with the original not adapted TOPAZ ice thickness. You are not interested in the accuracy of the

total sea ice volume here; you are interested in differences in the motion fields. The initial conditions therefore should be as equal as possible for the two model setups in my opinion. While I can understand that nothing can be done about that one is a couple ice-ocean model and the other is a ice model only, I cannot understand why you are introducing this artificial ice thickness difference here, which naturally will favor neXtSIM (because the ice thickness is pulled towards more realistic ice thicknesses every year).
The two setups have not been designed to be compared. The setup (initial conditions, forcings and parameters) of neXtSIM is the same as the one used in Rampal et al (2015, under review) and has been defined to provide the best possible representation of the sea ice drift, thickness and concentration with the current version of the neXtSIM model running in a stand-alone configuration. The setup of TOPAZ is the same as the one used for the TOPAZ forecast and for the TOPAZ reanalysis. We did not use the TOPAZ reanalysis directly because the available outputs are not appropriate for a diffusion analysis: only daily fields and for the ensemble mean. Moreover the reanalysis for the period 1979-2011 was produced with a too high value for the air drag parameter leading to a much too fast sea ice. This has been corrected by the developers of the TOPAZ system but only for the reanalysis after 2011 and in the forecast system. We use the corrected value that was found optimal by the developers of TOPAZ (Internal technical report, available on request).

As our goal is not to compare the two platforms but just to give two distincts examples of the application of the diffusion analysis, we do not think it is useful to degrade the results of neXtSIM by using a setup more similar to the one used for TOPAZ. It would indeed be interesting to try to improve the results of TOPAZ but the task seems to us too ambitious and we limit ourself to us the default setup. We understand that the present exercise should not be presented as a strict comparison and we hope that the new version of the manuscript has cleared that ambiguity.

See also the answer to the main comments of the other reviewers.

The text has been restructured and rewritten to avoid any “unfairness”.

7, 185-189: Also here, why are you using different atmospheric forcing for the two model setups? The ASR is a newer and mostly believed better reanalysis for the Arctic than ERAi. As ice drift strongly depends on the quality of the atmospheric forcing different results can be expected even if the same sea ice model would be used.

neXtSIM uses ASR for this Arctic configuration because we found in Rampal et al. (2015, under review) that it allows us to have a lower RMSE than with ERAinterim when comparing the sea ice drift simulated by neXtSIM against observed sea ice drift from the GLOBICE dataset. Such sensitivity study to the forcing is not available for the TOPAZ system. The TOPAZ setup is defined on a larger domain also covering the North Atlantic, which is not covered by ASR.

No changes

8,217: apart -> away (?)
Corrected

8,218: not a specialist on this but "steady and homogenous" sound like the wrong adjectives for a turbulent flow without mean flow to me.
It is a common expression used in turbulence but we agree that it could be misleading. It means "steady and homogeneous in a statistical sense", meaning that the characteristics of the fluctuating velocities (mean, variance,...) do not vary in space and in time.
Corrected: by adding the term "statistically" before "steady and homogeneous" as it is also commonly used.

8,227: you have not mentioned what tau is exactly so far.
We now add for clarity:
"tau is the time interval for which the autocorrelation is defined"

9,263: what is k? The different buoys?
k was the indice used to list "all the buoy velocities **u**_k$ recorded at a distance less than L/2 from **x** and within the time window [t-T/2;t+T/2]"
We modified that paragraph to avoid having to use this notation. We think it is much clearer now.

11,302-304: I actually cannot see that. I only see that for 2007. It is at least not as clear from the figures as you formulate it. Please adapt.
Yes we agree.
We shortened that paragraph and do not try to discuss the spaghetti plots anymore, but rather use these plots to justify the need of separating the mean and fluctuating motion.

11, 308: How is the artificial thicker ice in neXtSIM affecting this behavior? Would the difference be similar if neXtSIM would have used the TOPAZ ice thickness for initialization?
The correction of the initial conditions positively impacts the results of the neXtSIM model. However we have not tried to quantify this impact separately from the other differences between TOPAZ and neXtSIM. As explained here above, the configurations of neXtSIM and TOPAZ used here are the default ones. Their design, tuning and validation have been done independently and presented in other papers and technical reports.
We have not run neXtSIM with uncorrected/biased initial conditions because it would require to redo the tuning and validation, which is out of the scope of this study, and maybe not very useful. We then cannot answer firmly the question. Note that the goal here is not to perform a sensitivity study of TOPAZ or neXtSIM but to illustrate the interest of the diffusion analysis.
We add more details on the definition of the models setups.

12,337-345: Can you quantify these statements by numbers? I can see that TOPAZ is too fast. In e.g. 2007, however, it looks like neXtSIM is too slow and maybe also in 2008.

The differences between the simulated and observed mean drift are quantified in the next paragraph by looking at statistical distributions.
We changed the sentence " The mean ice drift simulated by the neXtSIM model is much more similar to the observations in that respect." Into "The mean ice drift simulated by the neXtSIM model reproduces well the mean circulation patterns, slightly underestimates the Beaufort Gyre but reproduces the almost immobile ice north of the Canadian Arctic Archipelago. "

13,376: Maybe point out here that the majority of the IABP buoys are deployed on MYI.
Yes, it is important.
We add this sentence in the description of the IABP data:
"The buoys are mainly deployed over multi-year ice and then do not represent the dynamics of weaker seasonal ice." and add "multi-year" to "...may indicate that multi-year sea ice dynamics are dominated...".

13,395: This paragraph gives a good summary of possible causes for the observed differences, which I agree with and which should be summarized in the conclusions again. Your goal, however, is to separate the differences caused by the two ice rheologies. How can you achieve that goal if the forcings, thickness etc. are so different that they affect the results?
The goal is not to compare two rheologies but to illustrate the interest of the diffusion analysis.
We modified the text in many places and the structure of the paper to emphasis on the diffusion analysis, and to better explain the differences between the two model setups.

15,445: The last two paragraphs actually give interesting and useful information derived from the buoy dataset only. This information is independent from the model comparison. Maybe it would be easier to find for the reader if summarized in an extra sub-section. Also from the abstract I would not have expected to find such information here. Mentioning it could widen the readership.
Yes, you are perfectly right.
We changed the structure so that these two paragraphs are now in the first section on the diffusion analysis. These results are also now mentioned in the abstract and conclusions.

16,472: Before you were always giving T in days
Corrected

16,471-502: you should clarify that you are talking about the buoy results not models here.
yes.
The analysis of the buoys is now discussed separately in the conclusion.

17,508: The TOPAZ circulation pattern looks right just the magnitude is a bit off.

We do not agree. The mean drift along the CAA is clearly problematic and could not be corrected by simply applying a multiplicative factor.
We changed the sentence to be more specific: "The mean velocities in the simulations using TOPAZ… do not represent correctly the circulation near the Canadian Arctic Archipelago."

17,509: Yes, but mention again that neXtSIM is too slow and that these differences for the mean fields well can be due to the different forcing and ice thickness and not the models themselves.
Corrected. We know mention the low bias and we change the formulation into: "The mean velocities in the simulations using TOPAZ..." and "The mean velocities in the simulations using neXtSIM" to avoid unsupported generalisation.

17,520: Doesn't look to me that TOPAZ was tuned the same way regarding the ice thickness as neXtSIM and using better atm. forcing also does not sound like finer tuning to me.
The two setups used here are the default ones defined for other studies. Both of them have been tuned differently and independently.
We know better describe those differences in the text and in the conclusion.

18,524: add a few more introductory sentences. After reading the paper it was not clear to me where the Appendix relates to the rest of the paper.
The content of the appendix has been transferred to the introduction.

19, 572: Research Council?
Corrected

25, Fig.4: I see three lines: thin black, thick black, red. You only describe two in the caption.
Corrected

26, Fig.5: mention what L is in caption or shorten whole caption.
Corrected

29, Fig. 8: Hard to see anything in the upper plot
Corrected

31, Fig10: color scale and annotations impossible to read.
Corrected