We appreciate the Referees' valuable remarks and recommendations and carefully addressed them in the new version of the manuscript.

On behalf of all authors,
Elena Zakharova

Reply to Referee 1

Our answers below are marked ">>"

Interactive comment on "River ice phenology and thickness from satellite altimetry. Potential for ice bridge road operation" by Elena Zakharova et al.

Anonymous Referee #1
Received and published: 4 February 2021
The paper titled "River ice phenology and thickness from satellite altimetry. Potential for ice bridge road operation" by lead author Elena Zakharova and coauthors explored using radar altimetry data to infer river ice phenology and ice thickness. By conducting the study over the lower Ob river in Russia, the authors reported accurate retrieval of river ice phenology and ice thickness by comparing ice phenology/thickness estimation from altimetry data at virtual stations to those obtained manually and those from the in situ gauge records. The authors have done an excellent job of describing the details and nuances of the ice processes, and how it can complicate the radar backscatter signals. The authors thoroughly described the uncertainties of the studies and provided valuable recommendations for future work and an assessment of the social impact of the conducted research.

My major concerns with the paper is the lack of clarity in the methods section. The authors reported many interesting results however, as I detailed below, not all of their methods were well described. Please see below for my comments. I would recommend the authors make the methods clearer and make the figures more informative and easier to read. Overall, I think the paper is well written and the implication and uncertainty of the study thoroughly discussed.

>> The manuscript was significantly edited according to recommendations of Referees. The section of Methods was extended. The sections Results and Discussion were reshaped. All figures were revised. Several figures were removed or combined after revision of corresponding paragraphs as recommended by Referee 2.

Major comments
Figure 4: are the dates in the format of dd-mm? I suggest to make the dates more explicit and move the surface types to a more prominent places (e.g. using a.b.c and refer to the surface type in the caption)
>> Figure 4 is redone.

On line 373: the authors argue that decrease in backscatter is proportional to gain in ice thickness. If this is the argument, would it make sense to plot the changes in ice thickness (Hice) against cumsum(dsig0/dt)?
>> Yes, the relationship Hice vs CumSum(dsig0/dt) is shown in the Figure 5b.

On line 375: the authors mention cumsum(dSig0/dt), which should be negative for the freeze-up period. However, in Figure 5b all the values are positive along the x-axis.Am I missing something here?
>> Yes, the cumsum is negative. The X-axis title should be abs(CumSum(dSig0/dt)). The Fig.5b was removed from the text as it was considered to be redundant by the Referee 2.

Lines 387–393: calibration and validation using the eight VSs were mentioned in this paragraph, however, no detailed methods were provided in text nor in figure 6. I would highly recommend providing how the calibration and validation were carried out.
>> We added a phrase describing our calibration approach and moved part of text from the Results Section to the Methods to facilitate the reading.

Figure 5: labels for the subfigures should be placed at more prominent locations. The legends should be placed at a consistent location of the figures.
>>Figure 5 is redone.

Figure 6: it is nice to have a flowchart to guide the readers through the processing steps. However, I found the one presented here hard to follow: data and procedure are better separated and represented using different boxes.
>>Figure 6 is redone.

Line 400: shouldn't phenology estimations be compared to gauge records closest to the VSs?
>> Yes, we did exactly this validation. Necessary details are provided in the new version of manuscript.

Line 413: please clarify how "close to zero" was defined.
>> The phrase was corrected for "In 56% of the cases this difference is equal to zero".

Figure 7: the author should discuss why for melt end, the results from the manual algorithm have a much bigger bias than that from the automated algorithm.
>> The algorithm was developed for detection of the melt start. The manual implementation of the algorithm is more accurate than automated implementation (what we can naturally expect during an algorithm development, as not all Sig0 behaviour cases can be coded correctly). As the manual implementation tuned for the melt start detection, logically, it should be not good for the melt end detection. Probably, the dedicated paragraph was not clear. This paragraph was re-written. "Comparing the dates of altimetry-derived melt onset with the ice state records provided by gauging stations, we conclude that manual implementation of our algorithm detects well the start of ice thermal degradation. In 88% of the cases, the difference between manually-retrieved melt dates and in situ observations of first water appearance is less than ±10 days (Figure 7, b). The automated melt date retrievals demonstrate worth accuracy for detection of the melt start, comparing to the manual ones. Only in 54 % of the cases the difference with in situ melt start observations is less than ±10 days. The automated approach is more efficient for the detection of the melt end as ±10 days accuracy was reached in 67% of cases (Figure 7 b,c)."

Figure 8, 9, and 10: the authors need to justify why for the gauge data the mean was used and for the VS data the median.
>> We were not specific enough – the mean did not mean "arithmetic mean". Changed to "median".

Line 444: "significant variability"â˘A˘ Tdoes this refer to the difference between the manually determined and the gauge mean, or does it refer to the variability amongst the gauge data. Need clarification.
>> Thank you for catching this imprecision. We meant the difference between satellite and in situ observations. The text is corrected for " Significant difference between gauging and virtual stations (order of 20 days) is observed only for melt start dates in 2014."

Line 440–451: it is easy to attribute years lacking north-south detected difference to local effect. However, such explanation is not satisfying without any evidence backing up the claim, especially given that so many factors (e.g. uncertainties in the percentage of pixels of different surface features) can affect the detected dates.
>> We agree that the uncertainties in estimation of the ice phenology dates from altimetric measurements can also be a reason of lacking of the North-south gradient for certain years. As it was

recommended by the Reviewer 2, we removed these results from the manuscript as this question would take a separate subsection to address all remarks and will result in manuscript extension.

Figure 10 and 11: the authors need to clarify or show the location of Tr187 in the x-axis label.
>>Figure and caption are modified

Figure 11: highly recommend the authors using color to represent data from different years or use a better way to differentiate the data.
>> The figure was deleted as in the new version of the manuscript the phrases referred to the figure were removed.

Lines 494–500: the active melting period (melt end) is highly dynamic and presents a challenge, as the authors noted, to any automated algorithm. I think the patterns presented in Figure 11b–c and the interpretation given in the paper is very interesting. However, it will make a much stronger argument if similar patterns contrasting the similarity at the melt start and variability at the melt end can be found in the in situ gauge data.
>> On the secondary branches, there is only one gauging station. We could not assess the big/small channel ice phenology difference from the in situ observations. By discussing this difference observed in our altimetry retrievals we wanted to demonstrate the value of the remote sensing methods. As the topic stirred a lot of remarks and questions and their addressing would extend the manuscript, we decided to delete it from the new version of the manuscript. The theme of spatial variability of ice phenology dates will be a subject of a new more detailed study based on multi-sensor approach with more solid proof base.

Figure 12: it is rather hard to see the rivers when everything is frozen. I suggest the authors add some labels on the images for a few key locations discussed in the paper to orient the readers.
>> The figure is no more present in the new version of the manuscript as there is no more reference in the text.

Lines 507–508: the authors need to clarify whether the correlation and RMSE were calculated based on the gauge that was left out of the parameter estimation step.
>> We added necessary details in the text in the Method Section. When designing the Cal/Val exercise, we initially separated our data on the Cal/Val periods (1:1 split) and found that the results of validation depend on selection of the Cal/Val years. We consider that the use of the leave-one-year-out method allows to avoid the effect of subjectivity, when separating the short time series (10 winters in our case) into Cal/Val sets. Calibrated by leave-one-year-out method parameters, calculated as average from parameters received from 10 leave-one-year-out fitting runs, better account for interannual variability. Using this method, the uncertainties can be estimated for all period and not for one left-out year. Corresponding reference on the method is added.

Lines 573–576: the authors mentioned both approaches of building relations were evaluated at 11 VSs nearest to gauges. However, only two sets of values were presented for VS 109 and 12. I think it will be really helpful for the authors to explain Table 2 in detail since it is pivotal to the understanding of the algorithm performance.
>> The table 2 was edited. The title was changed, the pairs of VS-gauging stations were grouped for better visual representation, the additional pair of stations used for intermediate tests was removed resulting in 10 pairs (one gauging station surrounded by 2 virtual stations).

Line 592: please explain the term “ridging flags”
>> we referred to qualitative records describing the quality of the ice near the gauging station. The "flags" was replaced by "events".

Figure 14: instead of using decimal year in x-axis labels, it will help readability by

converting it to month.
>> Figure 14 is redone based on the suggestions.

Line 613: could optical remote sensing provide information on the ice onset on rivers?
>> Yes. This was demonstrated by several works cited in the Introduction. The problem of the optical images is the cloudiness.

The authors should add scale bar and north arrow for all the maps presented in the paper.
>>Done

Figures in the paper are of various styles and should be made in a consistent style with consistent places for subfigure labels and the legend
>> Figures were redrawn.

Minor comments

Line 33: “erosion of channels and banks”: need citation here.
>> The reference on Ettema (2002) was added
Line 38: “catastrophic flooding”: need citation here.
>> The reference on Beltaos et al., (2013) was added
Line 75: “state and regime” is there a difference between the two?
>> We mean solid or liquid state (is the water frozen or not), which is different from the regime (temporal variability). The phrase was re-written.
The paper needs some language editing.
>> We carefully edited the language in the new version of the manuscript using professional Elsevier Service.
Line 126: proposes should be “proposed”
>>Done
Line 198: delete “in” before later freezing, and “in the” thinning of the ice cover.
>> Done
Many cases of unnecessary “an” and “the”
(e.g. Line 187: “The Matplotlib Basemap Toolkit”
>>No longer relevant. Other source was used.
; Line 573: delete the “the” before “both approaches”;
>> Done
Line 18: for “an” estimation (should be “the”);
>>Done

==========

Reply to Referee 2

Our answers below are marked ">>"

General comment

This study analyses the potential of radar altimetry to detect the seasonal freeze and melt events as well as the ice thickness on the Ob River, Russia. The study employs the backscatter (sigma nought) signal from Jason-2 and Jason-3 missions, as well as in situ observations from several gauging stations.

While the idea behind the study is clear and definitely shows a good potential of the radar altimetry backscatter to monitor river ice, the presentation of the methods and results lacks structure and scientific rigour. The quality of figures and captions is in my opinion poor. I also recommend to concise the paper by focusing on fewer aspects and removing redundant parts. I also insist on a professional English proofreading to simplify and smooth the style and improve the clarity. Please check the terminology: **reflect vs scatter**, bias **vs difference vs delay**, **algorithm vs method vs approach**, etc. Please also check the comparatives: you very often use "higher", "more", "less", etc. without mentioning what do you compare it with.

Please find my comments below. There are both specific and more general ones, but I found it easier to comment the parts at their place.

>> The manuscript was significantly edited according to recommendations. The section of Methods was extended. The sections Results and Discussion were restructured. All figures were revised. Several figures were removed or combined after revision of corresponding paragraphs as recommended. Many paragraphs were deleted or shortened. We hope that modifications done allowed for significant improvement of the text clarity.

Introduction:

In my opinion too long and needs concision. Think about subsections and narrow down your story from general introduction to your specific objectives. For example: The importance of river ice monitoring remote sensing observations so far altimetry principles and observations of fresh ice (1-phenology, 2-ice thickness) Jason mission knowledge gap your objectives.

>> The introduction was modified and shorten.

49-50: I think you show with your own study that the word "excellent" is a little bit of a stretch.

>> the word was deleted

51: please specify: you mean here the temporal resolution and not the frequency of the EM wave, correct?

>> we meant the temporal resolution. The phrase was changed "... for characterization of river ice at temporal resolution suitable..."

61: "Active sensors operating in the microwave region are weather independent and provide higher spatial resolution" – higher than what?

Phrase changed " Active sensors operating in the microwave region are weather independent and provide the spatial resolution higher than MODIS and AVHRR instruments."

What about high- and medium-resolution optical sensors? Please include a sentence or two for the consistency.

>> In first sentence of the paragraph we already mentioned the optical sensors of medium resolution (MODIS and AVHRR) with corresponding citations. We did not find studies, which used high resolution (Landsat or SPOT) optical sensors for **monitoring** freshwater ice phenology and thickness. However, we do not exclude that the HR optical images were used for study of some specific ice process (and not for multi-annual monitoring).

63: to my knowledge, “largely” is an exaggeration for such studies.
> The word is deleted

75: please explain what means “water state and regime”?
>> We mean the water solid or liquid state (frozen or not), which is different from the regime (temporal variability). The paragraph was re-written.

74-79: there are repetitions, please compact
>> The phrase was deleted.

81, 82: what are the “materials”?
83: please specify “reflected” or “scattered”?
84-101: I propose not to go deep into the SAR studies, as they are not very relevant to your study. Instead, formulate briefly the main principles of the altimetry, how the data are collected, what are the data obtained (backscatter, waveform, what else?)
You can also combine some of the information with the paragraph 54-79.
97: Higher than what?
98-99: is the penetration depth important here considering that the typical ice thickness < 2 m? Would C-band have insufficient penetration? I would focus on the properties of Jason mission later when you describe it for your study.
103: you haven’t mentioned before, that the altimetry missions are accompanied by radiometers
103-112: is it only about resolution or the different nature of measurements plays a role too?
>> for remarks on lines 81-112 : the paragraph was deleted as not relevant to the Introduction.

113-117: is it already a description of the method you are using?
>> The paragraph was moved as not relevant to the Introduction and re-written.

118-132: this reads rather specific and too long. Please consider revising, shortening or moving to the Methods section
>> The text was moved and shortened as recommended.

132-142: this part is a presentation of your own results and should not appear in the Introduction
>> The paragraph was deleted as not relevant to the Introduction

143-156: this is a place for your objectives, please state them instead of listing the sections
>>The objectives are formulated and the listing was deleted.
151: please introduce and explain what is a virtual station before
>> The text was re-written as recommended above. The definition is done in the section 2.2.2 **Altimetry**

2. Regional setup
There is not even one reference to all the information you provide for your study area. Please provide them.
>> The references are provided. The information about socio-economical situation in the region is taken from personal observations and discussions with local population.
Please work on the study area figure. Make it colorful. There is no way to understand where is the study area on the globe, especially in black and white version. Is it topography or is it an optical imagery? The scale of the globe is too coarse and of the zoom-in is too fine. The zoom-in area lacks geographical coordinates and the tracks are strangely projected. The circles (gauging stations) are invisible. Please think about the reader who has no idea about your research.
>>Figure 1 is completely redone.
173: please show these cities on the map.
>>Done
175-205: If this is part of your results, please move it to the corresponding section. You also introduce your in situ data only later.

>> The text was removed and put after in situ data description. As this text represents just a short description of environmental situation based on gauge data available, we do not consider it can be a part of our results.

195-199: please show and explain the trends, I do not see any.
>> As we mentioned in the first sentence of the paragraph, we did not find a significant trend for all period of observations on gauging stations. Only during last several years some tendencies could be noted. We deleted the phrase about the tendencies for last years as doubtful.
Figure 2: please improve the quality of the figure. Is that mean on figure a)? why only three stations of five are used?
>>Figure 2 and its caption are redone.
3.1 Altimetry
Please provide more information on the data selection: tracks, cycles, virtual stations, repeat overpass etc.
>> We added information on sampling frequency, distance between measurements etc.

231-233: please provide more information on how did you extracted the data via portal (programs used, codes?) and how did you use Landsat images for the selection. Is it a manual / visual selection?
>> Details are added: " High-resolution optical Landsat 8 images https://earthexplorer.usgs.gov/) were used for the geographical selection of altimetric radar and radiometer measurements over river channels using our own Python code that overlaps along-track Jason measurements and Landsat images. "

What exactly are the track numbers in relation to cycle and pass numbers?
>> We added phrase " One satellite cycle consisted of 127 revolutions and 254 tracks (odd numbers for ascending and even numbers for descending orbits)."

3.2 Optical imagery
235-243: I do not think you need to describe the optical imagery in a separate data section. The first purpose you already mentioned in the previous section. The second purpose is not represented later in the paper substantial enough. You can describe the images in place.
>> As no more optical images are used for illustration of spatial heterogeneity of the freezing/melting processes, the description was deleted as recommended.

I think the radiometric measurements deserve to be described in a separate subsection.
>> The measurements from the radiometer were used only as auxiliary information. We tried to avoid completely their usage thinking about extension of our algorithm to other altimetric missions. Among the altimetric missions, the Cryosat 2 is not equipped by the radiometer. However, in the complex cases (we suppose when freezing is long or oxbow lakes contaminate the radar echo), the radiometric measurements are useful. That why we did not enter into detailed description of the radiometer and added only description relevant to our study : band and sampling frequency, footprint size and how altimetric measurements were joint with the radiometric. Please note, that in the altimetric practice , there are two standard ways to joint 20Hz and 1Hz measurements : 1) spatial interpolation of 1Hz measurements, and 2) use of closest 1Hz measurement. In all our previous publications and here as well, we used the first way. Relevant phrase was added.

Please explain what do you refer to as an altimetric measurement, i.e. you talk about a precise selection over the river channel but at the same time mention multiple times that the signal is contaminated by the surrounding land. I encountered later the 400 m band oscillation – please refer to it early enough.
>> The corresponding phrases are added. " The altimetric radar aboard Jason-2 provided measurements at Ku (13.6 GHz) and C (5.3 GHz) bands with 20Hz sampling frequency allowing for 375 m distance between adjacent measurements. The ground track repeatability of the mission was maintained within a ±1 km cross-track at the equator. At the latitudes of our study region (63-66°N), the cross-track oscillation band was approximately 400 m wide. "

3.3 In situ data

254-255: please clarify what is measured daily and what is measured 3-6 times per month.

>> the phrase is rewritten " The standard protocol of river ice monitoring includes 1) daily visual observations of ice presence/absence and ice type; and 2) 3-6 times per months measurements of ice thickness and on-ice snow depth."

256: do they drill a new hole every time for a new measurement?

>> Yes, the protocol insists on this.

How do you credit the data from gauging stations? Are they publicly available?

>> The data are provided by Hydrometeorological Service and available after request.

4. Methods

4.1 This does not look like methods to me. You describe the environmental processes and the response of altimetric backscatter on them. Create a separate (sub-)section and provide references to all the processes you describe.

>> the new Section was created " **Temporal variability of radar altimetry signal over frozen rivers** "

Exclude your own results and move them to the Results section.

>> This is not our result. This is a description of behaviour of the altimetric backscatter in relation to ice formation/break up mechanisms. We consider that this information is important for understanding of proposed algorithms. Moreover, this is first article dedicated to river ice processes in relation to altimetric measurements and we want to provide the details, which will be useful not only for glaciological community, but also for scientists working on improvement of altimetric water level algorithms for the Arctic rivers.

I might disagree that the presence of ice increases the specularity. The ice-water interface should be rough, shouldn't it?

>> This section serves exactly to introduce the process of formation of river ice (which differs in some way from that in lakes and seas). However, here we speak about the first ice. The first ice is usually thin and wet with smooth bottom (nilas type). This ice type provides very specular returns for nadir radars (many publications from sea ice altimetric community address this fact). An important ice bottom roughness during the freezing can be found in areas of ice bridging (floes accumulation). Floating objects usually calm the water, thus, reducing the water surface roughness (we speak about this in the text). In mixed "rough ice/calm water" radar footprint, the main signal will come from the calmed specular water and will mask the weak return from rough bridging ice. We address this question in the Discussion section. The thick river ice may be rough, we agree. But the effect of the water current, which can smooth the ice bottom in scales of radar wave length, can't be excluded.

Mention the difference between the nadir-looking altimeter and side-looking SAR instruments (i.e. low and high backscatter from rough and smooth surfaces).

>> The phrase is added " In opposite to slant-looking SAR instruments, for nadir-looking altimetric radar the smooth surface produces higher return echo than the rough one."

Figure 3: This figure would be more informative if you show a mean backscatter with the range of values across all the stations for one typical year. Then you would guide the reader through the seasonal evolution of backscatter and attribute the backscatter change to environmental processes. Adding the in situ dates of freeze up and melt begin/end onto figure would help. The plot in its current version does not add any value in my opinion.

>> We highlighted the ice period on the figure. However, we preferred to keep the figure in the current form of the time series. In this form the figure demonstrates that 1) the seasonal variability is common for northern and southern reaches, 2) there is an interannual and spatial variability in the Sig0 in the beginning of the freezing (why we used CumSum(dSig0/dt) and not the absolute Sig0), 3) there are intermediate winter peaks (why we applied a smoothing filter), 4) during certain years, the detection of

the first winter peak is not a trivial task (two adjacent high peaks separated by important Sig0 dropdown), etc. We discuss all these issues in different sections of the manuscript and provide now the reference to the figure 3.

304-305: I am not sure that wet snow leads to an increase of the backscatter. To my understanding, wet snow attenuates the signal stronger than the dry snow. Please provide a proof.
>> Yes, the volumetric scattering increases as the water content of snow increases for slant-looking instruments (even at low angles). However, for nadir-looking instruments , the situation differs. The behaviour of backscatter for snow covered surface in nadir angle is presented in Ulaby et al. "Snow cover Influence on Backscattering from Terrain , IEEE, 1984" (we based on his fig. 11). Moreover, the Sig0 increase of order of several dB can be seen on fig 3 of our manuscript during several years before and after the spring Sig0 drop (due to ice metamorphism described also in the text). The proofs for the increase of backscattering during melting episodes can be found in Slatter et al.2019 (see list of manuscript references). We suppose that the increasing water content on the snow surface (due to melting or rain) decreases the signal penetration depth (Ulaby, 1984, Slatter, 2019) and, consequently, the volumetric scattering. Moreover, **at nadir angle the surface reflection increases** as the water liquid content increases. The total effect is the increase in the backscatter.

312-313: the sentence is not clear.
>> The sentence was deleted.

314-315: I do not agree that the waves are increasing with the water level. Should depend on the wind?
>> the sentence is changed for : " As the water becomes free of ice , the backscatter decreases due to increased surface roughness induced by wind."

319: Provide more specific title for the section
>> The title is changed on " **Ice onset and break-up algorithm"**

Was this algorithm ever applied already or it is brand new?
>> This is a new algorithm

320: referring to an example figure as described above would help to follow.
>> The reference was added

320: not sure about the word “annual” in this context
>> Changed on "last high peak of each year"

320: how do you define a peak?
>> We wrote a simple code looking for peaks in time series for the automated routine and did it visually analysing year by year for the manual routine.

321-322: “In the case of a multi-peaky recession limb, this peak should be of order of spring and summer peaks”. Why? What is spring and summer peaks?
>>Spring peak is the peak after the ice melt (see fig 3). The summer peaks are also frequent and occur when the water surface roughness is low (J3 is equipped by nadir looking instrument) due to the calm weather. All these peaks can be seen on Fig 3. As we insisted many times in the manuscript, depending on configuration of the virtual station, the land part in the footprint is different for different VS. We were not able to setup a definitive Sig0 threshold to select between several autumnal peaks. So, we elaborated several assumption helping us to define the freezing moment. One of them is the peak magnitude, which has to be of order of summer peaks at a given virtual station. The magnitude of summer peaks can differ from VS to VS.

322: what is recession limb?

>>we introduced this word in the text when describing the Sig0 time series in the figure 3. " This peak is followed by a progressive winter decrease, which forms a recession limb on backscatter time series." The "recession limb" term is used widely in river hydrology when describing the water level time series, meaning the progressive decrease of a parameter.

322: “If the selection of peak is not straight forward...” – how does the algorithm select a peak?
>> we explain this in the next phrase: '... an additional criterion based on the brightness temperature difference (dTb) between 34.0 and 18.7 GHz frequencies is introduced."

324: Explain why and how do use the dTb and add dTb on the figure described above.
>> We explain this in next phrase: " ...for example, two high peaks within one month or prominence of peak is low..." . How do we use the dTb is explained in the reply for the Rem.322

326: why exactly t-1, t+2? Why 2K? Can you refer to previous studies or explain this choice?
>> This is a part of our algorithm design. This have not been published before. The criterion was setup experimentally basing on verification of results : the part of successful retrievals comparing to in situ observations (described in the Results Section). The window [t-1, t+2] means that 1 satelite cycle before (==10 days) the freezing has started on the land (already mentioned fact) and that this freezing progresses further for next 20 days (t+2cycles). This allows us for rejection the early cold episodes or for better handling the cases of long freezing (alternation of cold and warm episodes, which do not lead to river freezing) . Several phrases are added.

Please provide a conceptual scheme for the automated algorithm.
>> The conceptual scheme is similar to that of manual and described already in the text and presented in the figure 5.

351-360: if you do not use this approach in the end, I do not see a reason to include it in the paper. Or “guided over river ice by the main waveform peak” means that you use this information? Please clarify.
>> We prefer to keep the phrase, as this behaviour of the waveforms has never been shown before **for river ice**. However, we explain, why we did not use this approach. The fig 4a aimed to demonstrate that our statistical approach has a physical background. And we return to the figure in the section dedicated to the potential improvement of the Hice algorithm. However, following the Referee recommendation, we deleted the figures 5b and c as they are not discussed in the text.

You never refer to the Fig. 5b. The information on a) and b) seems to me redundant. Is a) just one example? Why did you choose only two virtual stations on b)? Why not to include all gauging stations? Improve the quality of the figures, avoid work-in-progress axis titles and legend entries.
>> The figure 5b,c were deleted

388: how did you create a virtual station?
>> We explain this earlier in the sec. 2.2.2 " The cross-section of an altimetric track with a water body is called the virtual station (VS)." Similar to the in situ gauging station, the virtual station concept was introduced in scientific literature in 1990th for inland water studies using altimetry, see for example [Birkett et al., 1998].

Did you use the same stations for the dates of ice freeze-up and ice break extraction?
>> Yes, the same stations.

390: what is “ice thickness relations”?
>> We meant the "ice thickness-backscatter relations". The phrase was edited.

391: by “extrapolated’ you mean “applied”?
>> “extrapolated’ was changed for “applied”

391: what is “other “main set””? the rest 40 stations?
>> We meant stations not used for calibration. Phrase changed for "The established relations were applied to other 38 virtual stations..."

393: 2) not clear what do you correlate with what
>>The text was re-written " For each virtual station (VS_i) we used the coefficients a and b of those virtual station from the training set (VS_{jt}), which expressed the best correlation between $Sig0_i$ and $Sig0_{jt}$."

395: the scheme needs to explain better what are you doing. What do white and grey color mean? What are the grey outlined circles?
Figure 6 is redone.
Results:
Maybe it would make sense to make subsections and give them titles. It is not easy to understand when the topic is suddenly changing. For example, algorithm evaluation, interannual variability, spatial variability, etc…
>> The subsections were made
400-403: this paragraph seems redundant to what you already described in the Methods.
>> the paragraph was re-written
5.1 Ice phenology
Please start with an accurate description of your results.
>> Done. The introduction phrase to the section was deleted.

405-408: this seems to me a repetition of your methods section, at the same time rather a hypothesis of what you think is captured by the altimetric measurements.
407: would open river water always appear rougher than the young ice? Even if the flow is calm and there is no wind?
408: decrease compared to open water?
>> The part of the text referred to rem.405-408 was removed.

411: I think “bias” is not correct term for what you describe since bias has a direction. Difference would do better?
>> Changed for "difference"

411: “accuracy for Jason” – not for the altimeter but for your algorithm, for the date extraction?
>> Changed for "... altimetry-derived ice phenology dates."

411: how many retrievals are there? 8 stations by 12 years? 48 stations?
>> 10 VS from the training set by 10 winters (2008-2018). Only VS near 5 gauging stations are selected for verification.

413: what is close to zero and if it is not exact how did you end up with 56%?
>> Changed for "is equal to zero".

417: less accurate than what?
>> Added "... than using manual routine"

491: what results and what do you refer to here?
>> The phrase was deleted

424: better than what? Do you mean the algorithm or manual approach?
>> The phrase was changed for " ... manual routine of our algorithm detects well the start of ice thermal degradation"

426-428: what means “least” here and what do 54% and 67% refer to?

>> The phrase was changed " The automatically derived melt date estimations demonstrate worth accuracy for detection of the melt start, comparing to the manually derived estimations. Only in 54 % of the cases the difference with in situ melt start observations is ±10 days."

430: what do you mean by outliers? Please explain. Is it possible to adapt the algorithm? What about radiometric measurements? Do they not help? Can the manual peak extraction be affected if the person knows in advance the true date of freeze up and melt from the in situ measurements?
>> "outliers" changed on "unrealistic retrievals". Yes, it is possible to improve the algorithm and we speak about it in the discussion. The radiometric measurements helped considerably, but unfortunately, not in 100% of the cases. Probably, a better radiometric criteria can be developed. The algorithm improvement is the subject of our future studies.
Yes, of course, if one looks firstly on the in situ measurements, it will be difficult to exclude the subjectivism in the retrievals. However, after algorithm setup, we did the manual date retrieving without using the gauging data.

433: melting and freezing before 10 April and after 10 June? How did you define these dates?
>> This is an expert knowledge. We have being working in the region since 2004 and know quite well the hydrological situation. We looked at our previous publications dedicate to the altimetry and the Ob River discharge and added 3 weeks for earliest (considering potential climate change or extreme years) and 2 weeks for latest water level rise start. We did not speak about this as 1)we considered such details is not an important information and 2) the manuscript was already long.

Figure 7: please use the total number instead of norm pdf. Please mark the 90% cases on the graph.
>> Figure 7 presents the histograms normalised on total number of observations and not cumulative probability function, e.g. we cannot show 90% of cases. One can obtain all percentage that we discuss in the text by selecting the range (for example +-10 days) and summarising the percentage taken from the y axis for -10d, 0d, +10d.

How do you explain
• Bimodality and bias towards earlier freeze up of the automatic algorithm for freeze up?
>> In the automatic implementation some early Sig0 high peaks are detected as the ice onset and produce the second mode on the plot 7a. Probably, these peaks are followed by the high peak 2 cycles later. When analysing these cases visually, the progressive decreasing of the Sig0 after the 2nd peak argues for its selection. So, the choice of the second peak is evident for manual retrieving. However, we could not tune the automatic routing for handling this situation. We did several tests, but lost efficiency in the cases of multi-peaky recession Sig0 limb (see for example fig. 3, VS 12, winters 2018, 2019). We hope to improve the automatic routine in the future.

Do I understand correctly that negative values correspond to the earlier date?
>> Yes.

• Bias towards the later melt start for both automatic and manual approach?
>> Sorry, but we do not observe the bias toward the late melt start for manual approach on the fig 7b. It is true for the automated routine. As we explained in the text, with 10 days Jason repeat cycle we conceder +-10 days as acceptable accuracy for the dates retrievals. And as we mentioned in the text, the automatic routine is better for the melt end detection (which occurs ~2-4 weeks later). So, it is logically to see the second high mode toward later detection on the fig 7b.

• Poor results for the melt end for both approaches, especially for the manual approach?
>> The melt algorithm was designed to detect the melt start. So, this is logically true, that manual routine is not suitable for the melt end (fig.7c). The visual selection of the Sig0 rise start on the Sig0 time series plot is straightforward in most cases. However, to code all possible situations that can occur in the nature (small intermediate peaks at the end of winter, small peaks in the rising Sig0 limb etc.) was not possible. As a result, the automated routine only partly handled the melt start (fig 7b). The other

part of detected automatically melt dates corresponds to the melt end dates as they area observed on the gauging stations (fig 7c). This is probably due to these intermediate peaks during the Sig0 rise. Obviously, the radiometric dTb criterion is less effective during the spring than during the autumn. Its adjustment is the matter of our future studies.

436: as mentioned, this is generally not a bias.
>> Changed for "difference"

437: why 11 stations? You mentioned 48 in the text. Why not to include all of them?
>> The verification of ice phenology and ice thickness algorithms is done for the virtual stations from training set. As we explained earlier, we selected those located north and south from the 5 gauging stations + one additional located only in several km from the one from the training set. Now we keep only 10 VSs in the training set to facilitate the understanding.

438: I thought that training set applies to ice thickness retrieval? Why do you need a training set for the ice dates?
>> We do not need a training set for phenology algorithm. We did a verification of the algorithm only for VS from this set as these stations are located close to gauging stations. Doing the verification of the dates on the training set allowed us to avoid potential uncertainties related 1) to difference in the freezing/melting dates between main and secondary branches (now this section is removed as recommended); and 2) to effect of longitudinal gradient, which can be observed in the altimetric dates for certain years (now this part is also removed to shorten the manuscript).

439: how do you evaluate that? How do you define what is a good sensitivity? I think you need to find a rigid approach to evaluate your algorithm performance. Maybe you do not even need to go on with the automatic algorithm in describing multiyear variability and spatial variability? You can choose manual and show the results only for it as it is clearly better.
>> The paragraph was re-written according to recommendations.

440: not the altimeter but your algorithm?
>> We agree. The text was modified.

441: what do you mean by “noisy”? Noisier than what?
>> The paragraph was re-written.

442: “Nevertheless, a clear coherence exists between the corresponding time series”. Please clarify this sentence. What do you mean by coherence, what are the corresponding time series, and why is it clear. Not clear to me.
>> The paragraph was re-written.

439-445. Please find more consistent and smooth way to describe your results, than by picking some years and saying that some events are noticeable.
Why on Fig. 9 you present the different events (melt start and melt end) AND different approaches (manual and automatic)? Please use the same approach for consistency.
446: why important?
>> The paragraph was re-written and the figure 9 was removed as the text was shortened.

447: Why Salekhard station is not included from the very beginning? “Adding” to where?
Describe this gradient, i.e. earlier freeze up at the northern stations? If you talk about trends, provide the parameters, i.e. days per degree (or km).
448: average what? Why “calculated”? Are you talking about the algorithm?
449-451: either show what you mean or remove. Very difficult to follow. What years, what half?
452: the same comment as before, please describe these gradients more detailed.

452-453: what do you mean by “gradient in the order of 20 days”? What is denominator of the gradient? Km?
454: the same here.
>> The paragraph was removed to shorten and simplify the text. All above remarks are no longer relevant.

Figure 8: why do you use median for the altimetric data and mean for the gauging stations?
Please use legend to display the information in the caption, it is very difficult to follow.
>> We were not specific enough – mean did not mean "arithmetic mean". Changed to "median". Legend added to the figure

458: does it mean 4 gauging stations average? Why 20 virtual stations? How did you decide which ones to include?
>> The details are now provided. The median is compared to median. Only four gauging stations (all located on the main river branch) are selected, as the secondary branches can demonstrate some difference in ice phenology dates. Only 20 VS located on the main river branch were selected for consistency with in situ observations.

Figure 9: as already mentioned, make it consistent with the approaches. Comments for Fug. 8 applies here as well.
>>Fig 8 and 9 were simplified and combined
Figure 10: what is the x-axis? Please use more reader-friendly labels. Use figure titles in addition to a,b,c
>>Fig 10 is now deleted as no more text related to this figure exists.

Why do you show manual approach for a and b and the automatic one for c?
Please be consistent. Please indicate stations on the red line. Use legend to explain the lines and shaded areas.
474: what do you mean by “different years”?
>>The remarks are no more relevant. The text was modified.

479-500: this whole section is a little hard to follow. How did you select the shown years? How certain you can be with the algorithm results? What about in situ data? I am not sure that the way of presenting those results as in Fig. 11 is optimal. Would it be possible to use mean values for the entire period of observations instead of only three years?
486-487: “This is shown clearly on a Sentinel-2 optical image…” I do not see anything clearly on the Sentinel image. Please use color version, and mark on the image what is what (land, ice, water, main channel, narrow channels, etc). The same comment applies to the Fig. 12b.
489: branches, not brunches
489-491: Please show the image. If there is a polynya until March, does it mean it is not freezing at all during some years?
492: “between the branches” – do you mean main and secondary branches?
494-496: “At the beginning of ice degradation local morphological controls only play a small role (Figure 11b). Their role amplifies during mechanical break-up, which is better captured by our automated algorithm (Figure 11c).”. How do Figures 11b and 11c illustrate both statements?
497: uncertainty in what?
497: higher that what?
500: please explain Fig. 12b in more details and support the statement.
>> The whole section dedicated to main/secondary brunches phenology comparison was deleted from the new version. This allowed us to shorten significantly the manuscript.

5.2 Ice thickness
Would it make sense to show the Fig. 5b here? To support the equation 1?
>> We moved the equation to the Method section. So there is no needs to move the Fig 5b

501: what means different gauging stations? All of them? Some of them?

>> " Coefficients a and b of the equations were estimated for each gauging - virtual station pair from training set. Using leave-one-year-out method for each pair we obtained a set of a and b coefficients. The average values from each set were used for ice thickness estimation for a given VS. The accuracy of the ice thickness retrievals was evaluated using correlation coefficient and Root Mean Square Error calculated between retrieved and observed Hice for all 2008-2018 period."

501-503: how do you come up with 9 runs? Please explain the approach more detailed.
>> For 10 winters of 2008-2018, the leave-one-year-out test gives 10 runs. The text was corrected

How to read Table 2? How did you sort rows in the Table 2? Why there are two Pitlar, two Gorki, and two Muzhi entries?
>> The Table 2 was reshaped. The rows are now sorted by the virtual station. The additional (11th) station was removed for simplification. As it can be seen on the fig 1 , one in situ station is surrounded by two virtual stations, that is why each gauging station name is provided 2 times.

Why did you decide to work with an individual relationship for each station and not with one universal relationship?
>> Our previous study dedicated to the lake ice (the paper in preparation) demonstrated that Sig0 - Hice relations vary from lake to lake. We expected the same for the river ice. We introduced some modification to the lake algorithm and obtained quite interesting results : except for 1 VS, the coefficient a and b very insignificantly from VS to VS from training set**.** We discussed this fact in the manuscript. Further research is needed to investigate this problem in details.

How would Fig. 5b look if you include all the gauging stations?
>>The fig 5b will look messy with all gauging stations.

What do the correlation coefficients and RMSE describe? The relationship between backscatter and ice thickness? Then RMSE should be not in m but m per dB? Why correlation coefficient and not coefficient of determination, especially considering that the relationship is not linear? Or R and RMSE describe the relationship between in situ and altimetry ice thicknesses?
>> The text was modified :" The accuracy of the ice thickness retrievals was evaluated using correlation coefficient and Root Mean Square Error calculated between retrieved and observed Hice for all 2008-2018 period." In this case the RMSE is in meter units.

509-510: How do I see on the Fig. 13 that those stations are northern? I'm confused here, I thought that the stations 138 and 161 are used in the training set? Please include all the stations in the Fig. 13 and use gradient color scale for the timing of measurements. The same for the spatial distribution of the stations – maybe different point style?
>> Here should be the reference to the Table 2. The text is edited and significant modifications were introduced. The VS numbers for northern and southern stations was added into the text. We preferred to not colour the timing, as we suppose that, logically, low ice thickness occurs in the beginning of ice period and high ice thickness occurs before its end.

bigger in size and better resolved, use different line colors and styles (if you keep individual years). In the left part of the graphs, especially on c) nothing is visible. Explain x-axis, what is Tr187? Give the direction of the north. An additional map with the location of the shown stations would be useful here.
>> The figure 11 was deleted from the new version of the manuscript

Fig.12. Please see the comments before. What band do you use or is it a color composite? Also please include coordinates, scale. Please put the labels on the in situ and virtual stations. What is the difference between open circles and squares for the virtual stations?
>> The figure 12 was deleted from the new version of the manuscript

540-542: where could I see it myself? In the table 2? Then please organize the Table in a clearer way. Table 2: please organize it in a more understandable way and provide more explanations on how to read it.
>> The Table 2 was reshaped and re-organised and details were added to the corresponding section of the text.

555-608: Please include these sections into the Results. This is clearly a continuation of your results, and not discussion.
>> The section was moved to the Results.

557: please indicate which parameters did you use, there are many different ones.
>> Details are added: "... coefficients a and b of the Equation (1)..."

562: please include a short explanation how did you choose the window size
>> The details were added: "The size of applied window allowed for preserving the magnitudes and spatial heterogeneity of ice thickness in spatial domain, as well as for reducing the residual noise in temporal domain, which is left after smoothing of backscatter time series with Loess filter". We did not enter into technical details how we adapt our window as it can be found in different corresponding manuals of filters' design and explanation could result in one more paragraph of important size. Instead, we provided the criteria, which were important for selection of the window.

Fig. 13: please see comments to this figure above.
>> comments were addressed and reply provided above.

569: please change the word “analog”, it is unclear what you mean.
>> The text was modified to avoid this word.

569-576: I think it would make sense to move this paragraph to the beginning of the section, and to describe the ice thickness product after that. Do I understand correctly that you use the second approach to create the ice thickness product?
>> The paragraph was moved

571: why only four stations if you have five? Please mark on Fig. 1 the clusters of virtual stations attributed to the in situ stations based on their proximity.
>> Only 4 gauging stations are located on the main river branch. Here, we discuss estimations of ice thickness only for the main branch of the entire river reach. Corresponding phrase is introduced.

571-573: please reformulate sentence, it is unclear. What is main VS? What is time shift?
>> The text was modified.

573: “The performance of the both approaches was evaluated at 11 virtual stations nearest to the location of the gauging stations”. Now I am again confused. These 11 are in the training set? But why?
>> The text was modified and simplified. We provide the results only for one (best) approach and do not speak anymore about other one (which was tested, but not used)

574-575: how can I see the results of the first approach?
>> The text was modified. We deleted information about first tested approach as it was not used and hope this simplified the text.

584: “The interannual variability in maximum ice thickness retrieved from altimetric measurements at many virtual stations indicates a clear decrease from 2008 to 2012.” Is this something shown in Fig. 15? Then refer to the figure 15 right there. What means many? Why did you include only 2 of in situ and 2 of virtual stations in the Figure? Could you show a plot similar to the Fig. 8-9,, which would include all of the stations?

>> The reference to the Figure 15a was placed as recommended. The figure was modified.

592: please explain what is ridging flag
>> the word "flag" was changed for "event"

Figure 14: I suggest to combine the yearly plots into one multiyear plot, and add some vertical lines to mark the timing - to show the interannual variability. Please indicate the north-south direction, add unit on the color scale.
Please describe results shown in Figure 14 in the text. Right now, you only mention that you created the product. What is the area we see in this product? What is the extent of it?
>> Figure 14 is redone using above-mentioned suggestions. However , the combination into one multiyear plot lead to figure becoming unreadable, so we keep the initial layout.

600: more disagreement than what? Is this statement shown in Fig. 15? Refer to the Figure. 601: "This does not strongly contradict expectations as for most virtual stations this disagreement lies within estimated RMSE values (0.07-0.18 m).". Please reformulate and explain what you mean. I cannot follow logics here.
602: "Besides the reasons noted earlier..." – where, what do you refer to?
>> We addressed all remarks from the lines 600-602 in the new modified phrase. " The interannual variability of altimetric ice thickness on 1 December differs from those, observed on gauging stations. However, this difference is not high and lies within algorithm uncertainties 0.07-0.18 m. Besides the geophysical reasons and algorithm simplicity noted above, the degradation in quality of the in situ time series and the low representativeness of the one-hole sampling protocol can be evoked. "

603: why there is a degradation in the in situ time series? Lower representativeness – lower than what?
>> The reasons are general for many Arctic areas: not sufficient financial support, shrinkage of the ground network, absence of well trained stuff etc. The word "lower" changed on "low".

604-608: Please explain what are you trying to point out here, it is not clear.
>> This section was deleted

609-610: how do you demonstrate that they are accurate enough?
>> The agreement in interannual variability of our retrievals of Hice_max with the observations on the gauging stations (fig 14 and 15a) is the basis for this statement.

611-612: do you present this result? Is it something that I can see in Fig. 15?
>> Yes, first time we noted this fact in the text in relation to the Figure 13 (see low ice thickness points corresponding to the beginning of the freezing). This is also can be visible from the Figure 15 b, where the plots of H_alti lie above the plots of H_insitu.

614: how could the date of first consolidated ice be detected and why then you did not use this approach?
>> We did not use this approach because of the algorithm has not yet been developed.
In this section we speak about future potential improvement and indicate several possible ways.

614-616: please reformulate the sentence to support your suggestion
>> The sentence was reformulated. " In our algorithm the ice thickness estimation starts from the date of first ice (bank ice or frazil floes) appearance. Usually, the river reach in area of virtual station at this moment is not fully frozen. The detection of the date of the first consolidated ice (e.g. fully frozen reach) may help to reduce the errors in the retrievals in a low range of ice thickness."

618-623: is this something recommended in the previous studies (then please include references) or this is your own hypothesis?
>> This is new.

Figure 15: please see above the comments to this Figure. Why there are two plots for the b)?
>>Figure is redone, the caption is changed

The Discussion section has only one subsection which actually discusses the results in the context of the physics of the radar signal return and potential errors related to that. Another subsection considers a study case of an ice road and applies the developed methods to this study case, which again reads more like results. I was missing a discussion of your results in the context of the relevant studies (which are not only methodologically relevant).
>> Sorry, but we did not understand what do you mean exactly under " discussion of your results in the context of the relevant studies". However, we significantly modified the Discussion section.

I have an impression that the whole section 6.2 would benefit from shortening and compacting. Some of the discussed issues are mere speculations and raise more questions than give answers, and some other points are not really relevant to your study (e.g. layering as you pointed out in 682).
>> The discussion section was restructured and now consist of two parts:
6.1 Factors affecting ice thickness retrievals from altimetry
6.2 Potential improvement of algorithms
The subsection 6.1 was significantly shortened as recommended.

646: "…grows gradually from January until April." – is there no wind redistribution?
649: Please include the figure 4 here if needed
652-653: please explain better what you mean with the ratios and 40%, and 25%.
>> the section was re-written and the comments taking into consideration rem. 646 - 653

660: you mean the power relationship becomes weaker? What correlation do you mean?
>> Yes, thank you for suggestion how to ameliorate the phrase.

665: "Further congelation of inter-floes volume as well as ice growth lead to leveling of the ice lower boundary.". Can you support this hypothesis? Do your observations of the decreasing backscatter (669-671) contradict this statement?
668: please explain what do you mean by the first two cycles
669: please explain what do you mean by the note "due to the waveform peak power"?
>> The text was edited to avoid the contradiction and shortened as recommended. The remarks 646 - 669 are no longer relevant.

703: I do not think that we have seen any clear tendencies in the referred section.
705: can changes be robust?
706-707: please provide a reference for this statement
>> The part of the text relevant to Rem.703-707 was removed

708-709: how do you define an outlier? These are the valid observations, right?
>> The word is changed for "... measurements untypical for a given month... " . We detected some strange records in the data provided by one of the gauging stations (sharp temporal decrease or increase of Hice in the middle of winter). We could not find any geophysical process that could explain such a seasonal variability of Hice.

712-721: this part sounds like an outlook to me and should belong to the Conclusions section.
>> This part was modified and moved to Conclusion

746: please expand on what you mean by a delay. Delay to what? Is it the best term in this case? Please check it further in the text as well.
>> word "delay" is replaced by "ahead" or "lead time".

746-748: this sentence is hard to follow. Exception from what? Why this is an exception? Where is the Salekhard river reach?
>> The sentence was deleted for simplification.

751, 754: what is circulation?
>> The word was changed for "traffic"

754: earlier than what?
>> The phrase was modified

762: perhaps, “could be adjusted” instead of “has to be adjusted”?
>> Thank you for suggestion. The phrase was modified

760-771: this also sounds like an outlook and can be combined with 712-721
>> This part was modified and moved to Conclusion

Figure 16: You never mention Figure 16b in the text.
>> The reference on the fig 16b was added.

778-780: please reformulate the sentence: what is the second record of melt onset?
>> reformulated, see reply to rem. 780-783:

780: correlations between which dates? Is the information on Fig. 17a similar to the Fig. 16b?
>> No, Fig 16b shows the interannual changes of the predicted date, while the fig 17a shows the relationship between "second earliest melt onset" (see reply on the rem.780-783) and observed dates.

780-783: please explain better what are you doing here. How did you produce a forecast? Is forecast is just the melt onset day derived from altimetry (whatever the second record means)? This is a bit confusing.
>> The sentence was re-written " Using altimetric retrievals of the melt start for the entire set of 48 VSs, for each year, we searched the date when at least two altimetric melt onsets (AMO2) were detected within the entire 400 km river reach".
The approach is based on the fact that the melting starts in the south of the region and progress to the north. By detecting the melt start in southern reaches, we logically suppose to be able to predict the melt (and road closure) in the northern reaches.

789-794: again outlook, can be combined with previous ones and moved to Conclusions
>> This part was moved to Conclusion

Figure 17: Please explain what ROI are you referring to? In Fig. b) please correct the label of y-axis – the current label is not clear – what is ice road closing delay?
>> The title is modified.
Conclusions: please use this section not only for a dry summary of the results but also for a more general wrap-up (reinforcing the problem importance, filling the knowledge gap, outlook and recommendations etc).
>> The section was modified and extended.