The manuscript “Sea ice classification of TerraSAR-X ScanSAR images for theMOSAiC expedition incorporating per-class incidence angle dependency of image texture” presents methodology and results of sea ice type classification of TerraSAR-X imagery obtained during the MOSAiC expedition. Despite very interesting findings, due to large diversity of methods, results and analysis and a large size of the manuscript, it is recommended to split the manuscript in two parts, improve the order of the presentation and resubmit the manuscript(s) after a major revision.

Major comments

Although only two objectives are formulated in the introduction, the impression is that the manuscript attempts to fulfil at least four: 1. Investigate per-class AI dependence; 2. Optimize parameters of texture features; 3. Train and evaluate classifier; 4. Analyse time series. In my opinion such variety of objectives does not allow to focus well. That makes the manuscript too long to read and the story too difficult to follow. I would suggest to completely remove section 3.3 and correspondingly reduce section 3.4. I’m confident that results shown in these sections deserve a separate paper. I will therefore focus my review on the first, methodological part.

What is GIA classifier? Authors refer to that term in many places, but it is never defined or explained. I guess, that’s one of the central blocks in the classification algorithm: apparently, the backscatter, the texture features, the IA are passed into the mysterious “GIA classifier” for doing the actual classification. But how?! I’m very curious to know. GIA classifier needs to be clearly explained.

Order of presentation needs to be revised in order to correspond to the selected logic (Intro, Data and methods, Results and Discussion): Lines 133 – 144 and Fig. 4 should come in Section 2.1 Data; Lines 151 – 164 with Table 1 and lines 271 – 275 with Table 2 belong to Introduction as they describe state-of-the-art; Section 2.3.2 belongs to Results as it describes WHAT is achieved and not HOW it is achieved.

Analysis of IA dependence for various ice types need to be increased as it is an important result of this work. What is error-bars of the slopes (it can be computed, e.g. by bootstrapping) and what is significance? What is the reason for large positive bias of the slopes – speculation on stronger volume scattering needs to be expanded. What is physical reasoning behind positive slope for leads?

The suggested method and parametrisations seem to be difficult to use in other conditions (C-band, other IA, other ice types, summer). Although it is mentioned as a limitation in the end, I believe it is important to also underline in the Introduction – the goal is to study a specific TSX SC timeseries and for analysis of another dataset a similar full-scale analysis needs to be performed.

Image size and number of texture features are undoubtfully important hyperparameters of the Haralick algorithm. However, neglecting quantisation level and distance to neighbour pixels can lead to significantly worse results. Sensitivity to these two parameters should also be studied, for example in this respect: how does despeckling boxcar filter impact the GLCM? In theory, if a 3x3

filter is applied and then GLCM is computed with 2 pix distance, there should be almost no elements in GLCM off the main diagonal. On another note, Haralick (1973) suggested using adjacent pixels (d=1) so the choice of authors d=2 should be tested and explained better.

Minor comments

L7. Phrases in parenthesis make the sentence very unclear. Please split into two sentences.

L12. Unfortunately the GIA classifier and class probabilities are never explained in the manuscript.

L24. Please provide reference to prove the "largest expedition in history"?

L71. Objective 1 is actually two objectives: 1 . to investigate and demonstrate per-class IA dependencies of TSX SC HH intensity and GLCM textures; 2. to determine the feasibility and optimal parameterization of including texture measures as input features to the GIA classifier.

Figure 1 shall be removed as it does not explain anything.

L96. "and shown in details in " -> ", dates shown in"

L107. Why were these ice categories chosen? It should be written that other categories were not present in the dataset and the method cannot be extrapolated.

L129. Polygons == rectangles? This is unclear.

L133. Maybe "evolution of young ice" ?

L135. Please rewrite "wide-spread lead openings of open water or nilas" as "wide opening of leads with water or nilas"

Figure 3. The smallest sub-images seem to be very blurred. Is it the effect of the despeckling filter or just visualisation?

L208. Cannot agree here. Other authors also studied distance and number of grey levels (e.g. Clausi 2002). Sensitivity to these two parameters need also to be studied (see major comments).

L221. "...and thus is a relatively..."

L236 and 237. Is that already results of parameter optimization? Then it is better to move to the Results section.

L268. Whay volume scattering is presumed to be stronger?

Figure 6. Is positive slope for leads even physical? How the strong positive bias of the slopes can be explained?

L274. "This is given that" can be removed.

L276 – 280. This seems to be logical after the results, in the Conclusions section.

L321. A reference to unpublished work just supports my concern that it is too early to include this section in the manuscript.

Figure 10. It is impossible to see shades of blue on the roughness transects.

Figure 11 and Lines 389 - 392. Why the 10% sudden drop of the polygon area on ~15 December is not reflected in a similar change of MY or young ice? Why does the lead ice increases ~3 times on 1 March and this is not reflected in the polygon area? Where are the plots of “other mosaic studies” that are easy to compare with fractions of different ice types? I’m afraid it is too early to write “that the classified time series is valuable as indicator of ice openings” as I cannot see a proof of that. Instead, the variations of ice fractions seem to be rather spontaneous and connected to uncertainty of the algorithm.

L396. “The leads class are mostly fully represented in the classification map” is it really a limitation? Can be removed from that section.

L425. Convolutional neural networks also deserve being mentioned as a potential tool.