Responses to Comments on the Manuscript:

"Estimating subpixel turbulent heat flux over leads from MODIS thermal infrared imagery with deep learning"

(ID: tc-2020-363)

We sincerely thank the anonymous referee #2 for his/her detailed and useful comments and suggestions during the whole review process. We have carefully studied these comments, and made corrections or changes according to the comments and suggestions to improve our paper, and we are now resubmitting a revised manuscript which we hope will meet with your approval. The major revised portions are marked in green in the revised manuscript. The item-by-item responses to the reviewers' comments are listed as follows:

This study evaluates the use of "convolutional neural networks" (CNN) to improve calculations of turbulent heat flux (THF) in Arctic leads from 1km resolution MODIS Terra infrared satellite imagery. It represents a novel application of machine learning to both MODIS imagery, which is widely used and invaluable in studies of polar regions, and THF in leads, which is an important parameter in our understanding of climate change in Arctic regions. Though the discussion section is lacking and editing is needed with regards to the English language, this is a exciting paper that should be published following revision.

Overall comments:

Comment 1: **I agree with the first reviewer in that "Ice Surface Temperature (IST)" would be a more appropriate term than SST. "IST" is used in studies of polynyas using MODIS data, even when describing open water pixels. One of many literature examples is: https://www.mdpi.com/2072-4292/10/3/366. It is also the term used by NASA for their level-2 product.**

Response: Thanks very much for your suggestion. Indeed, "ice surface temperature" is more appropriate, we have changed all "sea surface temperature" to "ice surface temperature" in the revised manuscript. Additionally, in order to demonstrate that it is appropriate to use the term "ice surface temperature", the related literature "A New Approach for Monitoring the Terra Nova Bay Polynya through MODIS Ice Surface Temperature Imagery and Its Validation during 2010 and 2011 Winter Seasons" on "https://www.mdpi.com/2072-4292/10/3/366" has also been cited.

Comment 2: **The discussion section would be suitable if this were submitted to a journal purely**

focused on machine learning techniques, but because this is The Cryosphere and readers will be interested in the DeepSTHF methodology in the context of polar science, this section should be expanded to include discussion of:

- **The significance of the improvement in calculated THF using the DeepSTHF vs other methods, eg. is the magnitude of improvement in W/m^2 significant relative to the overall heat budget in the Beaufort Sea or similar study area? Is it worth the additional computing resources to utilize DeepSTHF over CubicSTHF?**

  Response: Thanks for the valuable suggestion. The real experiments on three dates, namely, 25 April 2008, 5 May 2009, and 31 March 2020, demonstrated that the THF estimated by the DeepSTHF was more accurate than those of the OriTHF and CubicSTHF, which was mainly because it correctly identified more lead pixels, as well as obtained higher IST in SR, especially along the leads boundaries. Specifically, when the area included the leads of various widths, such as on 31 March 2020, the DeepSTHF estimated approximately 30% more THF than the other two methods. Even when the study area consisted mainly of a large lead network, such as on 9 May 2019, the THF calculated by the DeepSTHF was 11% larger than those of the OriTHF and CubicSTHF methods. It should be noted that regardless of the leads cover only 1%–2% of the sea surface in the Arctic region during winter, they contributed to more than 70% of the upward THF (Marcq and Weiss, 2012; Maykut, 1978). Therefore, it can be inferred that compared to the OriTHF and CubicSTHF method, the DeepSTHF method can calculate considerably more THF for the Arctic region, which is significantly relative to the overall heat budget. These significance of DeepSTHF over CubicSTHF has been added in the discussion part (Lines 538-547).

- **How applicable do the authors think DeepSTHF would be to other areas outside the Beaufort Sea in the Arctic, or even the Antarctic? Does the fact that DeepSTHF was unable to capture very narrow leads mean it may be less applicable in certain areas or times of year, when narrow leads are more common? A discussion of the nature of leads in the study area/broader Arctic would add some helpful context**

  Response: Thanks for the advice. To test the generalization ability of the proposed method, we also validated the performance of DeepSTHF in the Barents Sea of the Arctic. The visual performance and corresponding quantitative evaluation were shown in Fig.1 and Table 1, from which we can see that the model has a good generalization ability and performed well in the other

Arctic sea ice region besides the Beaufort Sea. The detailed experiment result has been demonstrated in the discussion part (Lines 548-571).

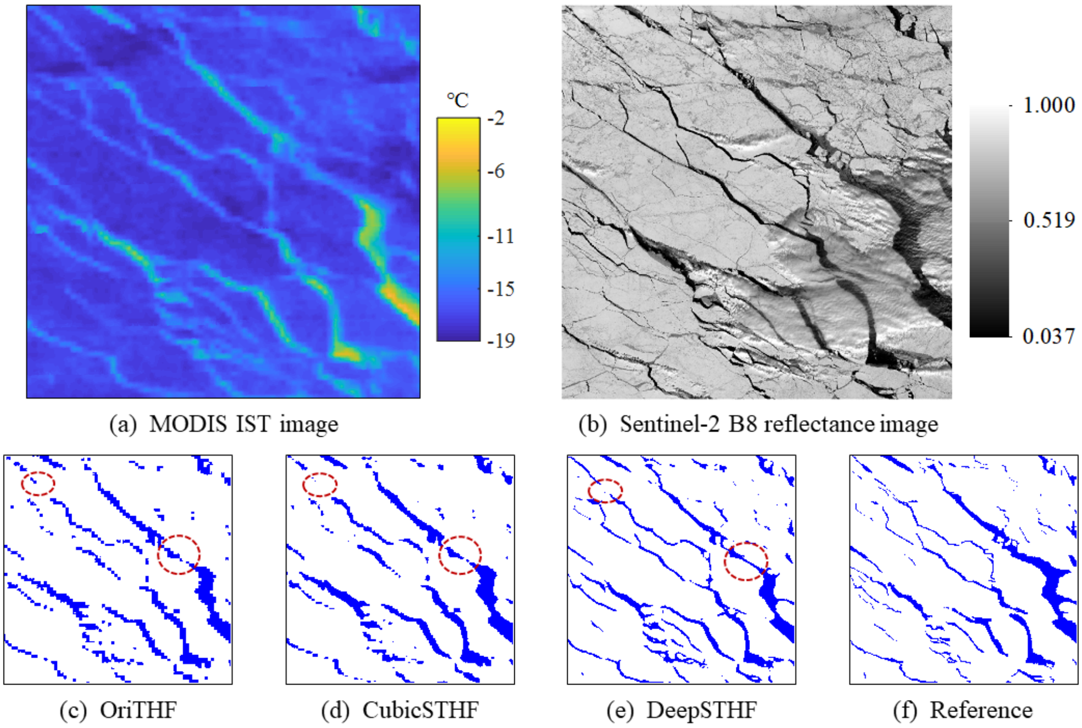

Figure 1. (a) The MODIS IST image of a subarea in the Barents Sea; (a) the Sentinel-2 B8 reflectance image; (c) the lead map obtained from the MODIS IST image by the OriTHF method; (c) the lead map obtained from the MODIS IST image by the CubicSTHF method; (c) the lead map obtained from the MODIS IST image by the DeepSTHF method; (f) the reference lead map extracted from the Sentinel-2 image. The red ellipse in (e) represents the area impacted by the drifting snow.

Table 1. The lead mapping results of the OriTHF, CubicSTHF, and DeepSTHF methods.

Method	Overall accuracy	Commission error	Omission error	MIOU
OriTHF	0.918	0.049	0.350	0.686
CubicSTHF	0.916	0.054	0.333	0.684
DeepSTHF	**0.941**	**0.035**	**0.265**	**0.753**

Note: MIOU stands for the mean intersection over union. The most accurate results are highlighted in bold text.

The reliability of the proposed DeepSTHF would decrease as the amount of very narrow leads increases. For leads, they vary with time and areas, it has been revealed that lead fractions are minimal in the period from February to early March (in the winter season) when the surface temperature decreases to near its minimum value in a year and then increases quickly in April; the lead fractions then reach more than 10% ice area in June since the temperature rises largely in the summer season (Qu et al., 2021). Therefore, the DeepSTHF might not perform well in the central Arctic region during the winter season. These features of leads and the applicability of DeepSTHF have been discussed in the revised manuscript (Lines 616-626).

- **DeepSTHF is promising, but exactly how far is it, or neural networks in general, from being widely implemented in studies of THF in leads? What are the next steps?**

  Response: Thanks for the suggestion. In practical applications, as the number of very narrow lead networks that with width narrower than five pixels increases, the uncertainty of the DeepSTHF will increase as well. In addition, several factors might also influence the performance of DeepSTHF. They mainly include the large spatial resolution gap between input MODIS and Landsat-8 data, the adverse atmospheric conditions such as cloud, and the low spatial resolution of meteorological data. The specific related discussion has been added in the revised manuscript. (Lines 613-635).

  The proposed DeepSTHF method can be further improved for future use. First, in this study, the integrated framework is applied with the MODIS thermal images, which have a spatial resolution of 1 km. However, there are other spectral bands with finer resolution in the MODIS product, including the first and second bands with a spatial resolution of 250 m. These finer-resolution images contain more spatial texture than the thermal infrared images, so the accuracy of an SR analysis may be increased by combining them (Li et al., 2013). Second, so as to make sure a lead indicated by the MODIS images is consistent with that of the Landsat-8 images, a change detection method can be first employed to detect abrupt lead change area during the overpass time of the two satellites. Third, the proposed method can achieve higher efficiency for long-time large-area analysis since the fine SST image and lead map can be generated with high efficiency by the proposed CNNs once the model training is completed. Therefore, the proposed method could be applied to produce accurate long-time series THF products using the MODIS images in the Arctic region. These contents have been added in the revised manuscript (Lines 642-651).

Reference:

Marcq, S., and Weiss, J.: Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere, The Cryosphere, 6, 143-156, 10.5194/tc-6-143-2012, 2012.

Maykut, G. A.: Energy exchange over young sea ice in the central Arctic, Journal of Geophysical Research: Oceans, 83, 3646-3658, 10.1029/JC083iC07p03646, 1978.

Qu, M., Pang, X., Zhao, X., Lei, R., Ji, Q., Liu, Y., and Chen, Y.: Spring leads in the Beaufort Sea and its interannual trend using Terra/MODIS thermal imagery, Remote Sensing of Environment, 256, 112342, 10.1016/j.rse.2021.112342, 2021.

Li, X., Ling, F., Du, Y., and Zhang, Y.: Spatially adaptive superresolution land cover mapping with multispectral and panchromatic images, IEEE transactions on geoscience and remote sensing, 52, 2810-2823, 10.1109/TGRS.2013.2266345, 2013.

Comment 3: Also agreed with the first reviewer in that the paper would benefit from professional English editing.

Response: Thanks for kindly suggestion. According to your suggestion, the revised manuscript has been polished by a professional English editing service. The certificate of English language editing is shown in Fig.2.

Certificate of English Language Editing

Manuscript Title:

Estimating subpixel turbulent heat flux over leads from MODIS thermal infrared imagery with deep learning

Date of Revision

April 16, 2021

Abstract:

The turbulent heat flux (THF) over leads is an important parameter for climate change monitoring in the Arctic region. Currently, the THF over leads is often calculated from satellite images, but the accuracy of the estimated THF is low for images consisting of mixed pixels that include both ice and leads because the existence of mixed pixels along lead boundaries decreases the measuring accuracy of the surface temperature over leads and the corresponding lead map. To address this problem, this paper proposes a deep residual convolutional neural network (CNN)-based framework to estimate THF over leads at the subpixel scale (DeepSTHF) based on remotely sensed images. The proposed DeepSTHF provides an ice surface temperature (IST) image and the corresponding lead map with a finer spatial resolution than the two-CNN model so that the subpixel scale THF can be estimated from them. The proposed approach is verified using...

This document certifies that the manuscript listed above was copy edited for proper English language at LetPub. All of our language editors are native English speakers with long-term experience in editing scientific and technical manuscripts. We are committed to leveling the playing field for researchers whose native language is not English.

- Neither the research content nor the authors' intended meaning were altered in any way during the editing process.
- Documents receiving this certification should be considered ready for publication where language issues are concerned. *However, the authors may accept or reject LetPub's suggestions and changes at their own discretion.*
- If you have any questions or concerns about this edited document, please contact us at support@letpub.com

LetPub
Your Manuscript Editor

LetPub is an author service brand owned and operated by Accdon LLC. Headquartered in the Boston area, we are a full-spectrum author services company with a large team of US-based certified language and scientific editors, ISO 17001 accredited translators, and professional scientific illustrators and animators. We advocate ethical publication practices and are an official member of the Committee on Publication Ethics (COPE).

For more information about our company, services, and partnership programs, please visit www.letpub.com.

Figure 2. The certificate of English language editing

- - - - - - - - - - - - - - - - - -

Line/Section/Figure Notes:

Comment 1: Line 50: Would be helpful to list out specific (satellite-related?) examples instead of just citing the papers.

Response: Thanks very much for your kindly reminding. The specific examples have been listed in the revised manuscript (Lines 50-53).

Comment 2: Line 53: Is there a way to explain how CNN is used to produce super-resolution imagery to those of us without a extensive technical understanding of its mechanics or machine learning in general?

Response: Thanks for the question. In order to make readers understand how does the CNN achieve super-resolution fully, we modified the original sentence "Among them, convolutional neural network (CNN)-based methods have provided significantly improved performance in producing SR imagery, because they have a powerful ability to model the latent nonlinear relationship between fine spatial resolution image and the corresponding coarse spatial resolution one through a large amount of training data (Dong et al., 2014;Ledig et al., 2017;Ling et al., 2019;Jia et al., 2019;Ling and Foody, 2019)" to "Among them, convolutional neural network (CNN) -based methods have provided significantly improved performance in producing SR images due to their ability to model a nonlinear relationship between the input and output data (Dong et al., 2014; Ledig et al., 2017). Specifically, in these methods, first, the relationship between an image with a fine spatial resolution and the corresponding image with a coarse spatial resolution is established through the training process with a large amount of training data, and then the trained model is used to super-resolve the testing coarse spatial resolution image".

Reference:

Dong, C., Loy, C. C., He, K., and Tang, X.: Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 10.1109/TPAMI.2015.2439281, 2014.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 105-114, 2017.

Ling, F., Boyd, D., Ge, Y., Foody, G. M., Li, X., Wang, L., Zhang, Y., Shi, L., Shang, C., Li, X., and

Du, Y.: Measuring River Wetted Width From Remotely Sensed Imagery at the Subpixel Scale With a Deep Convolutional Neural Network, Water Resources Research, 55, 5631-5649, 10.1029/2018wr024136, 2019.

Jia, Y., Ge, Y., Chen, Y., Li, S., Heuvelink, G. B. M., and Ling, F.: Super-Resolution Land Cover Mapping Based on the Convolutional Neural Network, Remote Sensing, 11, 1815, 10.3390/rs11151815, 2019.

Ling, F., and Foody, G. M.: Super-resolution land cover mapping by deep learning, Remote Sensing Letters, 10, 598-606, 10.1080/2150704X.2019.1587196, 2019.

Comment 3: Line 60: It would be nice to explain these names so they are easier to remember- eg. DeepSTHF is based on deep learning and THF, but I am unsure what the S is for. Likewise, does the "Ori" in OriSTHF stand for "original" image?

Response: Thanks very much for your kindly suggestion. In the original manuscript, DeepSTHF denotes Deep learning-based Subpixel THF estimation; CubicSTHF represents Cubic convolution interpolation-based Subpixel THF estimation; and OriTHF stands for Original image-based THF estimation. These explanations have been added in the revised manuscript (Lines 258-260).

Comment 4: Line 93: Why is SST calculated manually instead of using the Level 2 MODIS IST product with the same 1km resolution?

Response: Thanks very much for your suggestion. In the Level 2 MODIS IST product (NSIDC MOD29), pixels labeled as cloud (according to a cloud mask from MOD35) are removed. However, from visual inspection (especially by comparing with the corresponding Landsat-8 imagery), some lead areas with ocean fog or plume are mistakenly marked as cloud in the Level 2 MODIS IST product, which would influence the experiment. Therefore, to preserve potential leads, we calculated it from MOD021KM product. Additionally, the cloud in the MOD021KM product was determined by using MOD35 and visual inspection.

Comment 5: Line 93: Should also cite original work of Key et al. (1994, 1997), which Hall and Riggs (2001) followed for IST calculations.

Response: Thanks for the advice. The original work of Key et al. (1994, 1997) has been cited in the revised manuscript.

Comment 6: Line 95: Because clear sky is so important to the split-window algorithm, can you explain how you chose the 10% cloud cutoff?

Response: Yes. For MODIS imagery, it has a corresponding cloud mask product MOD35, we used it as the cloud indicator. Since some lead areas with ocean fog or plume are mistakenly marked as cloud in MOD35 product, which would influence the experiment, we inspect the cloud marked area visually to exclude pixels that in practice have been misclassified. Then we calculated the ratio of cloud cover for the study area and chose the 10% cutoff.

Comment 7: Line 120: You introduce these formulae here but don't show them until lines 221-222. To avoid confusion, perhaps just describe the meterological variables collected and don't describe the formulae yet.

Response: Thanks for the suggestion. We didn't describe the bulk formulae in this part in the revised manuscript.

Comment 8: Line 167-168: Move to beginning of section.

Response: Thanks for the suggestion. These lines have been moved to the lines 156-158 where was more appropriate than the place in the original manuscript, thanks again for your suggestion.

Comment 9: Line 169: Move "The following subsection explains the two CNNs more fully" to Line 162, after the sentence "Therefore, we used two CNNs... to achieve generation of a fine resolution SST image and lead map."

Response: Thanks for the suggestion. It has been moved according to your suggestion.

Comment 10: Line 225: Though the Goosse paper does describe everything in detail, because the latent and sensible flux calculations are so essential to this study, it would be good to write out more detail in this section, especially regarding the calculation of transfer coefficients, which can be parameterized many ways.

Response: Thanks for your advice. According to your suggestion, the latent and sensible flux calculations have been described more detailly in the revised manuscript (Lines 231-256).

Comment 11: Line 342: DeepSTHF method achieved the most accurate THF, but there still appears to be a considerable amount of scatter in Figure 10c. Related to my overall comment regarding the discussion section, further discussion of the magnitude of the discrepancy between reference and estimated THF and what the nature of this discrepancy means for a potential broader application of DeepSTHF is important

Response: Thanks for the suggestion. Although the proposed DeepSTHF achieved the most accurate THF among all the methods in the experiments, there was still a large discrepancy between the estimated and reference THF data. This discrepancy was mainly due to the errors of IST imagery SR reconstruction and SR lead mapping, whose major part originated from the very narrow lead network that the DeepSTHF could not identify, especially those with a width of less than five pixels. Therefore, in practical applications, as the number of very narrow lead networks increases, the uncertainty of the DeepSTHF will increase as well. Related discussion has been added in the revised manuscript (Lines 636-641).

Comment 12: Line 400: Why were these specific dates chosen?

Response: Thanks for the question. In practice, leads can be meters to kilometers wide (Qu et al., 2021), and the widths of them vary with time. It has been revealed that leads fraction became minimum in February–early March (in winter) when surface temperature decreases to near minimum in a year, then it increased quickly in April, and reaches more than 10% in June as the temperature rises largely in the Summer (Qu et al., 2021). Therefore, satellite images during late March-May had been chosen to test the performance of the proposed DeepSTHF as leads during the time were abundant with a variety of sizes and shapes. Meanwhile, Landsat-8 imagery had been used as the reference data in the experiment, it has a long revisit circle (16 days) together with frequent cloud cover and other poor atmospheric conditions, thus the available data was limited. Because of above mentioned factors, data on these specific dates had been chosen to validate DeepSTHF comprehensively. On one hand, they were observed in different months, the surface temperature of leads and ice might be considerable different, which was essential to test the performance of DeepSTHF in IST SR. On the other hand, the leads in different dates were of various sizes and shapes (Figs. 13d, 13h, and 13l). For example, image on 9 May 2019 mainly contained a large lead network and several very narrow leads, while image on 31 March 2020 comprised lots of narrow and scattered leads. The leads with multiple widths and shapes were key to test the performance of DeepSTHF in SR lead mapping.

Reference:

Qu, M., Pang, X., Zhao, X., Lei, R., Ji, Q., Liu, Y., and Chen, Y.: Spring leads in the Beaufort Sea and its interannual trend using Terra/MODIS thermal imagery, Remote Sensing of Environment, 256, 112342, 10.1016/j.rse.2021.112342, 2021.

Comment 13: Line 408: Please describe this additional layer of error caused by temporal correction

Response: Thanks for the advice. At present, the diurnal temperature cycle (DTC) method is frequently used to correct the temporal difference of surface temperature (Van Doninck et al., 2011). For MODIS data, because of the limited daily observations (≤ 4 times a day), the diurnal temperature cycle was empirically approximated as a sinusoid. However, surface temperature could have complex variations, especially for mixed pixels containing lead and ice, using sinusoidal method might cause considerable error, which would lead to additional layer of error in the THF calculation. We explained this in the revised manuscript (Lines425-428).

Reference:

Van doninck, J., Peters, J., De Baets, B., De Clercq, E. M., Ducheyne, E., and Verhoest, N. E. C.: The potential of multitemporal Aqua and Terra MODIS apparent thermal inertia as a soil moisture indicator, International Journal of Applied Earth Observation and Geoinformation, 13, 934-941, 10.1016/j.jag.2011.07.003, 2011.

Comment 14: Section 3.3: I don't think the second paragraph (Lines 237-243) is necessary, as you describe these statistics in the results section. Instead, it would be more helpful for the reader if, after describing the three methods DeepSTHF, CubicSTHF, and OriSTHF in Lines 230-236, you explain that you will run two experiments: One with simulated MODIS imagery and one without, and why you do this. I was unaware at start that there are two separate experiments, which led to confusion while reading the long results section.

Response: Thanks very much for the suggestion. Lines 237-243 in the original manuscript has been deleted in the revised manuscript. In our work, to assess the performance of the proposed DeepSTHF method comprehensively, two experiments using the simulated and real MODIS images were performed. The experiment with the simulated MODIS images was conducted to explore the performances of the DeepSTHF model as well as to avoid the uncertainty due to co-registration and temperature differences between Landsat-8 and MODIS data. The experiment with the real MODIS images was conducted to assess the performance of the DeepSTHF model in practical applications. We have added these contents in the revised manuscript (Lines 266-270).

Comment 15: Section 4.2.2: Why are the THF algorithms compared to Landsat for the lead maps but not the overall THF calculation as they are in Section 4.1.2 with the simulated MODIS imagery? Quantifying how well DeepSTHF performs relative to the reference is important, especially when applied to real MODIS imagery.

Response: Thanks very much for the question. In practice, large temperature differences could be observed between MODIS and Landsat-8 SST images, which had mainly resulted from the different overpass times of the MODIS and Landsat-8 satellites. Meanwhile, as mentioned above, there was few appropriate methods to eliminate the temporal difference of surface temperature. Therefore, the results of IST SR as well as the overall calculated THF of the experiment with real MODIS imagery were not compared due to a lack of true fine resolution SST reference imagery. Because of this, a simulated experiment was applied to explore the strengths of DeepSTHF model as well as to avoid the uncertainty of temperature differences between Landsat-8 and MODIS data.

Comment 16: Figure 6: The caption suggests it is a simulated super resolution image when it is instead a coarse resolution MODIS image simulated from Landsat imagery (I think). Please clarify this

Response: Yes, Fig. 6a is a coarse resolution MODIS image simulated from Landsat imagery. Thanks very much for pointing out this inappropriate caption. It has been modified in the revised manuscript.

Comment 17: Figure 14 - Why is THF calculated from the reference Landsat imagery not shown as well?

Response: Similar to the comment 15. As large temperature differences could be observed between MODIS and Landsat-8 SST images, the results of IST SR as well as the overall calculated THF of the experiment with real MODIS imagery were not compared due to a lack of true fine resolution SST reference imagery. Therefore, THF calculated from the reference Landsat imagery has not been shown.