<General Comments>

The proposed research compared the different surface object removal method on Arctic DEM, and analyzed how DEM error correction impacts flood inundation simulation. Considering increasing availability of high-resolution DEM, I think the suggestions from this research (which correction method is better, which parameters are feasible, how correction impact flood inundation simulation) is very useful. It is good to see that the DEM error correction using SMRF method is robust using wide range of parameters. One major suggestion I’d like to provide to enhance the manuscript is to include more discussion on the transferability of the proposed method (SMRF algorithm with optimum parameter range) to other regions. Readers must be interested in whether the optimum SMRF parameters detected by this study can be safely used to other regions or not. Please include some discussions about the parameter transferability (detailed suggestions are in the Specific Comments.)

Other than the above concern, the manuscript is I think very well organized. And it can be accepted after minor revision.

<Specific Comments>

Line 41: “exponentially increased computational costs”.
“Exponentially” is not precise. The calculation cost of 2D flow simulation follows approximately \((1/dx)^3\) where dx is the special resolution. If grid size becomes half, the computational cost is almost 8 \((=2^3)\) times, it’s not exponential.

Line 78: ASTER GDEM

Why not including AW3D DEM in reference here, which is more precise and now being widely used as high-accuracy stereo-view DEM?

L84: SETSM

Please explain what “SETSM” stands for? When it appears first time.

P257: Table 1

I suggest to put a line to distinguish PMF and SMRF, as the boundary is not clear.

P270: Replaced with the LIDAR DTM values.

Please describe the situation of the Arctic DEM original value here. Are they “missing data”, or there are large error?

P278: ensuring that the difference between the simulations was distinguishable.
The logic here is unclear. If large-magnitude flood is used as a test case, I assume flood extent is more confined by large-scale topography. Focusing on smaller-magnitude flood might be better to discuss the impact of topography improvement on flood risk estimation.

P286: small isolated wet areas

Please explain the mechanism of how these are caused?

P347: Figure 4

I don’t think the cross marks for man error (right column) are meaningful. The optimum points exist as a “line” in white-color area, rather than as a point in case of the mean error. Putting one cross mark could be misleading.

L361: More than 40% of the parameter combinations can reduce the RMSE by greater than a half.

This is important, but there must be something more to discuss for ensuring the robustness of the method. [1] The optimum combinations are almost same for three different land covers, suggesting the robustness of the parameter for various-time land-surface characteristics. [2] The skill-score does not significantly drop when parameter combination is slightly changed from the optimum location, suggesting the robustness of estimated parameters. There must suggest the transferability of the method to another region?

L379: The error distribution of...
Please connect this sentence to the following paragraph. One sentence paragraph is not recommended.

L382: Figure 5

The blue color overlaid on satellite map is very difficult to see. Please adjust colors.

L405: Figure 6:

Can you add one more “cross symbol” which represent the result of the best optimum parameter combination (which are common in all land covers). Readers must be interested in how “best-corrected DEM performs” simultaneously for all skill scores for all land covers.

L446: ArcticDEM-SMRF with larger error.

What do you mean by “error” is not clear here. Do you mean “larger elevation error”?

L450: shown at the spike areas in Fig6

I cannot find where is “the spike” in Figure 6. Please provide better explanation.
L458: Figure 7

Please check the color map of the bottom figure. It seems there are many “green” colors which is not in the color bar.

L517: similarly good flood simulation

I agree that the flood extents are almost similar, but how about flood depth? Can we say “similar”? 

L543: ICESAT2

It should be “ICESat-2”.

L589: which resulted in an optimal window size of 30 m and slope threshold of 0.07 in the city of Helsinki.

Please make some discussions on the possibility of transferring this parameter to other regions, or possibility of estimating best parameter for other regions (without Lidar DEM coverage). Readers must be interested in this. If the parameter has relationship to land object characteristic (such as typical building size), there might be a chance to find good parameters for other regions.