Author Response to CC1.

The comments in CC1 by Jennifer Frederick are in font color black. The authors' responses are in green.

We thank Jennifer Frederick for going through our model results in detail. We are very much aware of the current model development efforts in which she is involved and value her opinion in improving our model. She has proposed, with her own analysis, that: ''ArcticBeachv1.0 cannot predict the annual erosion rate any better than a random number generator can.'' Seeing this as a rather crippling statement to our work, we have carefully gone through her code and analysis, and respectfully assert that this statement is not supported nor reflected by what she has provided in her analysis.

We assert:

- Contrary to what Jennifer Frederick has claimed, she did not compare our model results with random numbers. Instead, she used an unrealistically-constrained array of values having limits of 10% within observed retreat rates.
- We would like to point out that *any model* or hindcast would be out-performed by a random number generator as long as the random numbers being generated were constrained to a close enough value to the observations.
- Given that observed retreat rates themselves can have a much greater error than 10% and are difficult to assess (Lantuit et al., 2008, Jones et al. (2018)), we find it an unfair and misleading claim to assert that the results of our model are no better than choosing random numbers when the 'random' numbers were chosen closer to observations than the uncertain error bar on the observations themselves. Further, Jennifer Frederick does not provide any reasoning for choosing a 10% threshold of deviations from observed retreat rates as her definition for her 'random numbers'. It is not mentioned in the main body of her comment that her 'random' numbers are constrained by this unrealistic threshold, well within the error of observations. It is instead left to the reader to go through the methodology of the supplemental section in her community comment, or find the line in her model code. This tactic is potentially damaging in the sense that it could lead to misconceptions of our paper if the reader is unwilling to go through the methodology supplement of community comments in a discussion forum.
- Our model includes essential physics that cannot be produced by a random number generator. For example, water levels are essential in driving coastline retreat (Barnhart et al., (2014), Kobayashi et al. (1999) and mentioned in Section 1, line 43-46 of the manuscript). ArcticBeachv1.0 is able to calculate relative water levels using wind speed, wind direction, bathymetry, and coastline angle. In light of declining sea ice cover and lengthened open water season, especially in those locations where freeze-up is being delayed further into the windy fall storm season, the importance of including a physical representation of changes in relative water levels due to wind forcing is paramount. Water levels have long been known to be a driving factor in erosion, especially erosion of partially-frozen coastlines present in the Arctic (Aré (1988), Casas-Prat and Wang (2020)).
- ArcticBeachv1.0 is the coupling of two widely-known and well-cited physics-based numerical models (Kobayashi et al. (1999) and Freeman et al. (1957)). It is the first time a water level model has been coupled to a simplified Arctic erosion model with the aim of developing a computationally efficient physics-based parameterization of arctic erosion.

- It is the first time such an approach is used that does not focus on one segment of coastline, such as the highly specialized processes occurring at Drew Point, Alaska. This is mentioned in the manuscript in Section 1, lines 27-38.
- To say this coupled model is 'under-developed' suggests that a state-of-the-art similar approach already exists, and ours is less developed than state-of-the-art.
- Since it is the first time an arctic erosion model has been developed such that it can simulate retreat on diverse types of coastlines in a computationally-efficient manner, we argue that ArcticBeach v1.0 sets the state-of-the-art in developing a parameterization of arctic shoreline erosion based on physical principles.

Review of ArcticBeach v1.0: A physics-based parameterization of pan-Arctic coastline erosion
by Rebecca Rolph et al.
Reviewed by Jennifer M. Frederick, Sandia National Laboratories, jennifer.frederick@sandia.gov

Summary
This manuscript describes a model for Arctic coastal erosion that is based on a simplified physical erosion model of a partially frozen cliff and beach, coupled to a storm surge model. It is presented as a first step toward a parameterization of pan-Arctic shoreline erosion at a coarse spatial scale for capturing erosion rates on the order of years to decades. It uses physical data as boundary conditions, such as wind speeds and directions, wave period and height, and sea surface temperature, as well as accounting for sea ice cover. The authors claim the new model provides a promising starting point to project the retreat of Arctic shorelines, or to evaluate historical retreat in places that have had few observations.

General Summary of Comments
I do not recommend publication in its current form. The model presented (ArcticBeachv1.0) is under-developed and the authors have not shown that this model has any predictive skill that outperforms a random number generator (proof described in detail in my review). For transparency, I have also included the Python script which performs this analysis.

We have gone through Jennifer Frederick's random number generator in her script mentioned above (see the line of her code pasted below in green) and it is not stated in the text of her detailed review (including in her description explaining what a null hypothesis is to the authors) why she tuned the 'random' numbers to generate random values within 10% of the observed retreat. The 10% used in her script refer to the error reported by authors in calculating erosion rates based on remote sensing imagery. The error of observations calculated for remote-sensing based computations of retreat are generally based on equations attempting to encapsulate all errors associated with the manipulations of the operator and of the georeferencing process (see Lantuit et al., 2008, 2011; Günther et al., 2013, Jones et al. 2009). The total error or dilution of accuracy is the root of the sum of all squared error factors (RMSE from georeferencing, digitizing error, pixel resolution) divided by the number of observation years. The individual error factors remain the same independently of the period of observation, while the number of years obviously varies. The total error will therefore vary drastically based on the observation period. In other words, the error reported in these studies relates to longer periods of observations and not to individual years. Previous studies have reported on errors varying between 4 to greater than 25% (see Lantuit et al., 2008, 2011;

Günther et al., 2013, Jones et al. 2008, 2009)., but these errors could be much greater for shorter periods of observations. At Drew Point the total error varied between 4 and 9% (Jones et al. 2009) but would be greater if the observation periods were shortened. This shows that the 10% error chosen here does not reflect the error actually observed for individual years and actually does not relate to one specific period of observation. In fact, Jones et al. (2018) indicate that "it is difficult to accurately assess errors in erosion rate measurements associated with this type of analysis." Finally, the error for observations at Drew Point should be considered as among the lowest along the whole arctic coastline, and is an inaccurate depiction of observational error for average arctic coastal erosion. The selection of a 10% value for error is therefore highly subjective and not systematic.

We would like to re-iterate any model would be deemed out-performed by a random number generator as long as the random number generator was constrained to a close enough value to the observations. While ArcticBeachv1.0 is calibrated with observed values, it is not a random number generator. It is based on a previously published, and well-cited, physical model of Arctic shoreline erosion (Kobayashi et al., 1999). Thus, unlike a random number generator, it provides the opportunity to add more physical processes not yet captured in this study. Additionally, as mentioned in Section 4.3 lines 405-413, the possibility to avoid water level calibration in the future could be implemented in future work.

The line in the code provided by Jennifer Frederick, which constrains the random number generator within 10% of the observed retreat rates, is copy-pasted below:

```
randomlist_DP.append(random.uniform(min(obs_DP)*0.90,max(obs_DP)*1.10))
```

As with all numerical physics-based models, the opportunity exists to add more physical processes with the goal of reproducing model output that more closely resembles reality. One of the primary uses of numerical models is to develop our understanding of the real world, and to predict what might happen in the future. A random number generator that is constrained to past observations is not used in place of numerical models because we do not have observations of the future. While data assimilation is a great tool, it is not possible to solely rely on observations to predict future processes --- we need to understand the physics involved. Or, indeed, if one would like to look at the sensitivity of a certain process, we would have to use a model with multiple processes in order to (for example) effectively 'turn off' that one process we are interested in and see what impact that has on the results. Such an understanding is not possible by using only a random number chosen well within the error of observations. Sensitivity studies to different coastline properties has been conducted and those results explained in Section 3.5 'Sensitivity to critical model parameters'.

One such process that is important for erosion is water level, and we are able to calculate the relative water levels at a very low computational cost, at any point along the Arctic coastline, from globally-available reanalysis wind speed and direction, or, for example, CMIP projected winds. This is done with a well-known and well-cited storm surge modelling approach, explained in Section 2.2, lines 139-157. We argue that a random number generator is in no position to calculate physically-relevant variables that drive Arctic coastal erosion, and our model provides establishes the state-of-the-art of computationally-efficient coupled storm-surge Arctic erosion processes.

My suggestion to the authors is further development of the model and resubmission for publication at a later date and after further collaboration and consultation with peers in this research field.

We view the assumption above that we have not collaborated with peers in this research field as presumptuous. Further, this comment does not refer to any part of the scientific scope of the manuscript.

One benefit of the model presented is its low computational cost. If the low computational cost can be maintained while improving its ability to robustly predict coastal retreat rates, this would represent a ground-breaking advance in the field!

We appreciate that Jennifer Frederick sees our approach as one that is worth developing further, and would like to add that ArcticBeach v1.0 represents a way forward not only in the field of arctic erosion, but also those fields involved in nearshore carbon cycling and biogeochemistry (referenced in Section 4.3, lines 387-393 of the manuscript).

The results summarized in Figure 4 show the modeled annual and cumulative retreat at Mamontovy Khayata (MK) and Drew Point (DP) vs observations at each site. At first glance, the modeled retreat looks poor, but an error analysis was not provided to quantify model performance. For any predictive model, a thorough analysis of model predictive skill is required to evaluate its performance and ability to make reliable, robust predictions. One of the simplest routines is to test model predictions against a random prediction. If the model has good predictive skill, it should outperform a prediction generated at random within a plausible range of possible outcomes. This is essentially like posing the null hypothesis and showing that the model can disprove the null hypothesis. In this case, the null hypothesis states that, 'ArcticBeachv1.0 cannot predict the annual erosion rate any better than a random number generator can.' This null hypothesis would be disproved for any possible model as long as the random number generator is constrained to a close enough range to observations. If all models could be constrained so closely by observations at the same time as knowing what values will be observed in the future, any numerical model hindcast or prediction is irrelevant. Given that Jennifer Frederick unrealistically constrained her random number generator closer to observed values than error of the observations themselves, we do not find her analysis relevant in a scientific context. If the ArcticBeachv1.0 model can predict annual erosion rate statistically significantly better than a random number generator, then it can rightfully claim predictive skill. My concern here for both locations is that, while there are a few years where modeled erosion matched observed erosion fairly well, there are also many years in this time series where the erosion is far outside of the running average. In these years, a model with high predictive skill should be able to reproduce the trend, if it has captured the correct physics. However, the ArcticBeachv1.0 model predictions end up under- or over-estimating the retreat, in the OPPOSITE direction just as many times as they estimate the retreat in the CORRECT direction (above or below the mean retreat).

The conclusion from the analysis for predictive skill (described in full detail below) shows that the ArcticBeachv1.0 model has no predictive skill at the DP location, and has inverse predictive skill at the MK location. Based on the error analysis, I disagree with the authors, as stated in the abstract, that the ArcticBeachv1.0 model provides a promising starting point to project the retreat of Arctic shorelines, or to evaluate historical retreat in places that have had few observations. The results of this analysis at both locations indicate that the model in its current form is under-developed, and cannot be relied upon to provide robust and skillful predictions for coastal retreat rates in the Arctic more than a randomly generated number can (in the case of the DP location) nor can be relied to provide a prediction in the correct trend direction (in the case of the MK location). These statements are misleading and do not provide the full scope of the reviewer's case for 'random number generator'. The argument that has been made multiple times by the reviewer that the ArcticBeach v1.0 is no more useful at evaluating erosion than choosing random numbers could be applied to any model, not just ArcticBeach v1.0.

We would like to re-iterate to the point that if all models could be constrained so closely by reality (or, as done in Jennifer Frederick's analysis, even closer to reality than the error of what we are able to observe) any numerical model hindcast is irrelevant.

Detailed Analysis

I performed an analysis on the modeled retreat vs the observed retreat to quantify the error. I used the mean squared error (MSE) of the annual retreat predictions as the performance metric. The MSE will penalize large differences between predicted and observed values more so than small differences, which is appropriate in this case because swings in retreat far outside the normal or average erosion behavior signify major disruptions in erosion drivers, which is what we want to capture with a robust, skillful predictive model. It is also thought that future conditions will become more extreme as climate changes in the Arctic, and thus erosion may continue to behave erratically. The mean squared error is defined as

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(M_j - O_j)^2$$

where N is the number of retreat predictions with available retreat observations, M is a modeled retreat prediction, and O is a corresponding retreat observation for the model prediction. For this test, the MSEs for each location for ArcticBeachv1.0 vs observations are shown in Table 1. For the MK location, the MSE of the annual retreat between 1995 - 2018 was 125.48 m2, and for the Drew Point location, the MSE of the annual retreat between 2007 - 2016 was 61.55 m2. A perfect prediction for every year would yield an MSE of 0 m2 at both locations.

[Table 1 is given in original CC1].

Next, the model predictions are tested against a random number generator to judge predictive skill and give meaning to the MSE values calculated for ArcticBeachv1.0. For the MK location, a random number from within a plausible range of retreat was generated for each year using Python's random package. The range in retreat was calculated as the minimum and maximum of the observed retreat data with a 10% envelope (e.g. 1.18 m – 12.04 m). The MSE for the randomly chosen annual retreat was calculated against the observations. This numerical experiment was performed 5,000 times, and a histogram of results was created to obtain statistical behavior (shown in Figure 1). As reported in Table 1, the average MSE of the annual retreat from the randomly generated model was 16.36 m2, as compared to the ArcticBeachv1.0 model value of 125.48 m2 (shown as the red line superimposed on the histogram). The error is much larger for the ArcticBeachv1.0 model than the randomly generated model, while also lying *significantly* outside of the 1st standard deviation of the randomly generated model's "predictions" (predictions in quotations because they are not truly predictions but random numbers). This suggests that the ArcticBeachv1.0 model has predictive skill, but its predictive skill is *opposite* of the observations (in the direction of larger error or in the opposite direction from mean annual retreat). This is clearly seen by inspection of Figure 4a in the manuscript, where large mismatches in the opposite direction from the mean annual retreat rates are predicted by the ArcticBeachv1.0 model, especially between years 2002 - 2018.

[Figure 1 is given in original CC1].

The test was repeated for the Drew Point location. For the DP location, a random number between a plausible range of retreat was generated for each year using Python's random package. The range in retreat was calculated as the minimum and maximum of the observed retreat data with a 10% envelope (e.g. 5.94 m – 24.83 m). The MSE for the randomly chosen annual retreat was calculated against the observations. This numerical experiment was performed 5,000 times, and a histogram of results was created to obtain statistical behavior (shown in Figure 2). As reported in Table 1, the average MSE of the annual retreat from the randomly generated model was 53.85 m2, as compared to the ArcticBeachv1.0 model value of 61.55 m2 (shown as the red line superimposed on the histogram). In this case, the ArcticBeachv1.0 model performed slightly worse than the randomly generated model (since the MSE for the ArcticBeachv1.0 model was higher than the mean MSE for the randomly generated model). Additionally, the MSE for the ArcticBeachv1.0 model sits within the 1st standard deviation of the MSE for the randomly generated model. This suggests that the ArcticBeachv1.0 model does not predict erosion rates *significantly* different than a randomly generated number. If it did, then the MSE would be well below the 1st standard deviation of the randomly generated model. The performance can also be seen by inspection of Figure 4b, where the ArcticBeachv1.0 model predictions end up under- or over-estimating the retreat at DP, in the OPPOSITE direction just as many times as they estimate the retreat in the CORRECT direction (above or below the mean retreat over the time period).

[Figure 2 is given in original CC1].

Furthermore, the analysis was extended to quantify the error in the cumulative erosion. The cumulative erosion error was calculated as the difference between the sum of the observed annual retreat values and the sum of the modeled annual retreat values. As reported in Table 1, the cumulative retreat error was 48.77 m (reported as "roughly 40 m" in the manuscript text, line 215) for the MK location, and 3.42 m (reported as "within a few meters" in the manuscript text, line 215) for the DP location.
Similarly to the random model numerical experiments presented for the annual retreat predictions, the same procedure is repeated for the cumulative erosion error. For each year, using the same set of random numbers that were generated for annual retreat, the cumulative retreat was calculated by summing the random annual retreat values for each numerical experiment. A histogram was created for each location, shown in Figure 3 (MK) and Figure 4 (DP).
For the MK location, the mean cumulative erosion error for randomly generated model was 39.19 m (see Table 1), as compared to the ArcticBeachv1.0 model value of 48.77 m2 (shown as the red line superimposed on the histogram). In this case, the ArcticBeachv1.0 model performed slightly worse than the randomly generated model (since the cumulative erosion error for the ArcticBeachv1.0 model was higher than the error in the randomly generated model). Moreover, the mean cumulative erosion error for the ArcticBeachv1.0 model sits within the first standard deviation of the cumulative erosion error for the randomly generated model. This suggests that the ArcticBeachv1.0 model does not predict cumulative erosion

significantly different than a randomly generated number at the MK location. Interestingly, while the annual retreat predictions were skillful (albeit in the opposite direction), the cumulative retreat might as well have been generated at random.

[Figure 3 is given in original CC1].

At the DP location, the mean cumulative erosion error for randomly generated model was 20.28 m (see Table 1), as compared to the ArcticBeachv1.0 model value of 3.42 m2 (shown as the red line superimposed on the histogram). In this case, the ArcticBeachv1.0 model performed significantly better than the randomly generated model (since the cumulative erosion error for the ArcticBeachv1.0 model was lower than the error in the randomly generated model and it was positioned outside of the 1st standard deviation of the randomly generated model error). This makes sense because the ArcticBeachv1.0 model did a decent job predicting the erosion rate at Drew Point for years in which the erosion was relatively average, but happened to over- or under- estimate the erosion for anomalous years at roughly equal magnitudes, and as a result summing to roughly zero, thus providing little contribution to the cumulative retreat error metric.

[Figure 4 is given in original CC1].

The conclusion from the analysis for predictive skill shows that the ArcticBeachv1.0 model has no predictive skill at the DP location, and has inverse predictive skill at the MK location. Based on the error analysis, I disagree with the authors, as stated in the abstract, that the ArcticBeachv1.0 model provides a promising starting point to project the retreat of Arctic shorelines, or to evaluate historical retreat in places that have had few observations. The results of this analysis at both locations indicate that the model in its current form is under-developed, and cannot be relied upon to provide robust and skillful predictions for coastal retreat rates in the Arctic more than a randomly generated number can (in the case of the DP location) nor can be relied to provide a prediction in the correct trend direction (in the case of the MK location). As disheartening as this error analysis seems, the MK location does show promise because of its ability to capture opposite trends. I suggest to the authors to investigate this behavior more closely, as it probably indicates some physical behavior captured in the model that may be relevant for erosion rates, but in the opposite sense.
Table 1: From what I understand, the two study locations have identical material properties, but they differ in geometry only with cliff height. Is this an adequate demonstration of the model's ability to provide a "physics-based numerical model that can be applied across all partially frozen shorelines"? (Quote from lines 36-37) I was expecting more diversity between demonstration sites.

We thank Jennifer Frederick for her perspective, and realize we needed to add more description in the manuscript that make clear how unambiguously different our validation sites are. We have now added a new subsection called 'Validation Sites' that describes them in more detail. However, we would also like to point out that one should not look at the simplified material properties that represent study sites in a model framework in order to get an idea of what that study site is like in real life. When there is no effort to examine the differences between the real-life sites themselves, one will have a misconception of each site because simplifications must be taken when representing a real-life site in a model. This misconception is exemplified by the approach taken by Jennifer Frederick in her comment above when she states that the 'two study locations have identical material properties, but they differ in geometry only with cliff height' and there is not enough 'diversity between demonstration sites.' We note that it is impossible to capture all of the material diversity at the spatial and temporal scale of real-life into a model, that our values come from published literature (see references in Table 1) and we would like to highlight that one should not look at model representations of study sites to judge what the real-life site is like.

When we want to examine if physical segments of Arctic coastlines are different, we go directly to the coast itself during fieldwork (as has been done for many years by co-authors of this work, with a decades-long strong German-Russian-Canadian collaboration). We also examine published observational literature (naturally, including those papers the authors have not written themselves, because these sites have been the focus of co-authors work), and also examine historical and satellite data.

In addition, not only are the real-life physical coastline properties just a part of the story of what determines differences at a coast, but what also must be included are the differences in the climate variables at the coast, such as sea ice coverage and wind speed and direction. All of these parts of the story are vital to understand if one would like to compare whether or not two coastlines are similar. Our study sites, for example, Mamontovy Khayata, on Bykovsky Peninsula, Siberia, and Drew Point, Alaska, USA, are starkly different in their main erosional features (e.g. dominance of block erosion at Drew Point, mentioned in the manuscript on lines 32-34 and also in newly added Methods subsection 2.1.3, see response to Reviewer #1) and Mamontovy Khayata (e.g. dominance of thermodenudation, now highlighted in the new Methods subsection 2.1.3 in response to Reviewer #1).

We would also like to point out that our two study sites are located on roughly opposite sides of the Arctic Ocean, they have different open water season lengths (Figures 6-8), wind speed and direction (Figures 6-7) and bathymetry. These differences are all taken into account in our coupled storm surge arctic erosion model, and consequently in our coupled erosion model. This is mentioned in Section 2, lines 58-67.

References:

Are, F. E.(1988). Thermal abrasion of sea coasts, *Polar Geography and Geology*, 12, 1–86, from: Termoabraziya morskikh beregov, Nauka, Moscow, 1980, 158 pp. doi:10.1080/10889378809377343

Casas-Prat, M., & Wang, X. L. (2020). Projections of extreme ocean waves in the Arctic and potential implications for coastal inundation and erosion. *Journal of Geophysical Research: Oceans*, *125*(8), e2019JC015745.

Günther, F., Overduin, P. P., Sandakov, A. V., Grosse, G., & Grigoriev, M. N. (2013). Short- and long-term thermo-erosion of ice-rich permafrost coasts in the Laptev Sea region. *Biogeosciences*, *10*(6), 4297-4318.

Kobayashi, N., Vidrine, J. C., Nairn, R. B., & Soloman, S. M. (1999). Erosion of frozen cliffs due to storm surge on Beaufort Sea Coast. *Journal of coastal research*, 332-344.

Jones, B. M., Hinkel, K. M., Arp, C. D., & Eisner, W. R. (2008). Modern erosion rates and loss of coastal features and sites, Beaufort Sea coastline, Alaska. *Arctic*, 361-372.

Jones, B. M., Arp, C. D., Jorgenson, M. T., Hinkel, K. M., Schmutz, J. A., & Flint, P. L. (2009). Increase in the rate and uniformity of coastline erosion in Arctic Alaska. Geophysical Research Letters, 36(3).

Jones, B. M., Farquharson, L. M., Baughman, C. A., Buzard, R. M., Arp, C. D., Grosse, G., ... & Romanovsky, V. E. (2018). A decade of remotely sensed observations highlight complex processes linked to coastal permafrost bluff erosion in the Arctic. *Environmental Research Letters*, *13*(11), 115001.

Lantuit, H., & Pollard, W. H. (2008). Fifty years of coastal erosion and retrogressive thaw slump activity on Herschel Island, southern Beaufort Sea, Yukon Territory, Canada. *Geomorphology*, *95*(1-2), 84-102.

Lantuit, H., Atkinson, D., Paul Overduin, P., Grigoriev, M., Rachold, V., Grosse, G., & Hubberten, H. W. (2011). Coastal erosion dynamics on the permafrost-dominated Bykovsky Peninsula, north Siberia, 1951–2006. *Polar Research*, *30*(1), 7341.

Lantuit, H., Overduin, P. P., Couture, N., Wetterich, S., Aré, F., Atkinson, D., ... & Vasiliev, A. (2012). The Arctic coastal dynamics database: a new classification scheme and statistics on Arctic permafrost coastlines. *Estuaries and Coasts*, *35*(2), 383-400.