Reply on RC3

General comments

This manuscript by Wang et al. presents NZ-BeachTopo30, a 30 m beach topography dataset for New Zealand that fills intertidal data voids in DeltaDTM by fusing ICESat-2 photon-counting altimetry with Sentinel-2 multispectral composites via XGBoost. The problem is well-motivated — intertidal voids in global coastal DEMs are a genuine and widely recognised obstacle for coastal hazard assessment — and the dataset itself is a welcome contribution. The production workflow is clearly described, the validation framework is multi-layered, and the sea-level rise demonstration nicely illustrates the practical value of improved intertidal coverage. I have only a few minor comments.

Reply:

We sincerely thank the reviewer for the thoughtful and constructive comments on our manuscript. We appreciate the recognition that the problem is well motivated, the dataset is a welcome contribution, and the production workflow and validation framework are clearly described. The reviewer's suggestions about framing the novelty more precisely around the data product and fusion design, clarifying the transferability assessment, emphasizing the intertidal validation metrics, discussing the Lidar resampling effect, refining the SHAP interpretation, and providing details on the vertical datum conversion are all highly valuable. Incorporating these comments has improved the clarity, accuracy, and completeness of our manuscript.

Below, we provide a point-by-point response to your comments. For your convenience, our direct responses are formatted in blue text, while the specific new content incorporated into the revised manuscript is highlighted in *blue italics*.

Comment 1:

The manuscript positions NZ-BeachTopo30 as "the first spatially continuous and full-coverage beach topography dataset for New Zealand." The national-scale application and the specific integration strategy with DeltaDTM are clearly valuable. That said, the ICESat-2 + Sentinel-2 + ML approach for coastal elevation estimation has been demonstrated in several publications, and XGBoost for structured regression is well established. The authors might consider framing the novelty more precisely around the data product and the fusion design rather than the methodology itself — this would make the contribution stand out more clearly on its own merits.

Reply 1:

We thank the reviewer for this constructive suggestion. We agree that the methodological components (ICESat-2, Sentinel-2, and XGBoost) have been individually applied in previous coastal studies. Therefore, we have revised the manuscript to reframe the contribution more clearly around the data product itself and the specific fusion design that combines a high-accuracy baseline (DeltaDTM) with machine-learning-predicted intertidal elevations to achieve seamless spatial continuity.

Specifically, we have made the following modifications:

1. In the abstract, we have refined the wording to emphasize the data product and the fusion design. The revised text now states that the dataset is constructed by fusing ICESat-2 and Sentinel-2 within a framework that uses DeltaDTM as the backshore baseline, and that the key contribution is a national-scale dataset that bridges the intertidal data gap.

2. In the introduction (Section 1), we have revised the sentence that previously claimed the first spatially continuous dataset to more precisely highlight that the full-coverage reconstruction is achieved by the specific fusion design. The revised text now reads: Unlike existing global DEMs that contain systematic voids in intertidal zones, NZ-BeachTopo30 provides seamless spatial continuity from backshore to foreshore by integrating a high-accuracy baseline (DeltaDTM) with machine-learning-predicted elevations for previously missing intertidal areas. To our knowledge, this dataset represents the first national-scale product that achieves such complete topographic coverage for New Zealand’s sandy beaches through this targeted fusion strategy.

3. In the conclusion (Section 7), we have added a sentence that states the contribution in terms of the data product and fusion framework without using negative phrasing. The revised text now reads: NZ-BeachTopo30 is the first national-scale beach topography dataset for New Zealand that achieves full spatial coverage from backshore to foreshore. This is enabled by a fusion framework that combines a high-accuracy backshore baseline (DeltaDTM) with machine-learning-based intertidal elevation reconstruction using ICESat-2 and Sentinel-2 data. The resulting product provides seamless topographic continuity across the land-sea interface and fills a critical data gap for coastal applications in New Zealand.

These revisions ensure that the contribution is framed around the data product and the integration strategy, which are the genuine merits of this work.

Modifications in the revised manuscript:

Abstract (modified text):

To bridge this fundamental data gap, we present NZ-BeachTopo30 which is a national-scale and full-coverage 30 m beach topography dataset for New Zealand constructed by fusing ICESat-2 photon-counting altimetry with Sentinel-2 multispectral time series. Using DeltaDTM as a high-precision baseline for the stable backshore, we trained an XGBoost model on ICESat-2 control points and Sentinel-2 spectral-geometric features to reconstruct the missing intertidal topography specifically.

Introduction (Section 1, revised sentence):

To our knowledge, NZ-BeachTopo30 represents the first spatially continuous and high-resolution national-scale beach topography dataset for New Zealand. Unlike existing global DEMs that contain systematic voids in intertidal zones, NZ-BeachTopo30 provides seamless spatial continuity from backshore to foreshore by integrating a high-accuracy baseline (DeltaDTM) with machine-learning-predicted elevations for previously missing intertidal areas. It provides a full-coverage description of the beach surface and offers a new foundation for further applications.

Conclusion (Section 7, added sentence):

NZ-BeachTopo30 is the first national-scale beach topography dataset for New Zealand that *achieves full spatial coverage from backshore to foreshore. This is enabled by a fusion framework that combines a high-accuracy backshore baseline (DeltaDTM) with machine-learning-based intertidal elevation reconstruction using ICESat-2 and Sentinel-2 data. The resulting product provides seamless topographic continuity across the land-sea interface and fills a critical data gap for coastal applications in New Zealand.*

Comment 2:

The 4,392 valid ICESat-2 pixels span 340 of 1,576 beach units (~22%). The transferability analysis in Section 4.4 is a real strength of the paper, and the modest degradation in accuracy (RMSE 0.93 → 1.02 m) is reassuring. It may be worth noting that this assessment can only be carried out where airborne Lidar is available, which itself tends to be concentrated in more accessible or higher-priority coastal segments. A brief acknowledgement of whether performance on remote, unsampled beaches might differ would give readers a more complete picture.

Reply 2:

We thank the reviewer for this thoughtful observation. We agree that the transferability analysis, while a strength of the paper, is necessarily constrained by the availability of airborne Lidar validation data. The reviewer correctly notes that Lidar coverage tends to be concentrated in more accessible or higher-priority coastal segments, and performance on remote, unsampled beaches could differ from the reported metrics. We have therefore added a brief acknowledgement of this limitation in the discussion section to provide readers with a more complete picture of the dataset's expected performance across different beach types.

Specifically, we have added a new paragraph in Section 5 (Discussion) that addresses this point. The added text acknowledges that the validation beaches with Lidar coverage may not be fully representative of all beach environments in New Zealand, and that caution is warranted when extrapolating the reported accuracy to remote or poorly sampled beaches. We also note that the transferability analysis in Section 4.4 already shows a modest degradation in accuracy from beaches with ICESat-2 training samples (RMSE = 0.93 m) to those without (RMSE = 1.02 m), which provides some indication of the model's robustness. However, we explicitly state that without independent validation data for those remote beaches, the exact performance remains uncertain.

Modifications in the revised manuscript (Section 5, added paragraph):

One limitation of the transferability analysis presented in Section 4.4 is that the validation using airborne Lidar is only possible where such data are available. In New Zealand, airborne Lidar coverage is not uniformly distributed across all coastal segments; it is more frequently available in accessible areas or those with higher priority for coastal management. Consequently, the reported accuracy metrics for beaches without ICESat-2 training samples (mean RMSE = 1.02 m) are derived from a set of validation beaches that may not fully represent the diversity of all unsampled beach environments across the country. The performance of the model on remote, poorly accessible, or low-priority beaches could differ from these reported values. While the modest degradation in accuracy from 0.93 m to 1.02 m (Section 4.4) suggests that the model generalizes reasonably well to beaches without direct training data, the absence of independent high-precision validation for those specific locations means that the exact accuracy remains unknown. Users of the NZ-BeachTopo30 *dataset should be aware of this limitation when applying the product to remote or unsampled coastal segments.*

Comment 3:

The internal model evaluation (Fig. 3) against ICESat-2 train/val/test splits usefully demonstrates model consistency, while the independent Lidar comparison (Fig. 4) provides the more informative accuracy diagnostic. It might be worth giving slightly more prominence to Fig. 4(f) — the accuracy specifically within the newly predicted intertidal voids ($R^2 = 0.61$, RMSE = 0.94 m) — since this is the metric most directly relevant to the paper's core contribution. The combined-product metrics ($R^2 = 0.75$, RMSE = 1.17 m) include the high-quality DeltaDTM backshore pixels and could, on their own, give readers a somewhat optimistic impression of the ML prediction component. Relatedly, the Lidar DEM was resampled from 1 m to 30 m via mean aggregation; on steeper beach faces this could introduce representativeness differences that inflate apparent disagreement. A brief note on this point would be helpful.

Reply 3:

We thank the reviewer for these constructive suggestions. We agree that Figure 4(f), which shows the accuracy specifically within the newly predicted intertidal voids ($R^2 = 0.61$, RMSE = 0.94 m), is the metric most directly relevant to the core contribution of the paper. We also agree that the combined product metrics ($R^2 = 0.75$, RMSE = 1.17 m) include the high quality DeltaDTM backshore pixels and could, if considered alone, give readers an overly optimistic impression of the machine learning prediction component. Furthermore, the reviewer raises an important point about the Lidar resampling procedure. The mean aggregation from 1 m to 30 m could indeed introduce representativeness differences on steeper beach faces, where elevation varies more rapidly across space, potentially inflating the apparent disagreement between the predicted and reference elevations.

We have therefore made the following modifications to address these points:

1. To give more prominence to Figure 4(f), we have revised the text in Section 4.2 (Accuracy assessment using Lidar DEM) to explicitly state that this metric is the most direct indicator of the model's performance in the intertidal zones that were originally missing from DeltaDTM. We have also added a cautionary note that the combined product metrics should be interpreted with the understanding that they benefit from the inclusion of the high accuracy DeltaDTM backshore pixels.

2. Regarding the Lidar resampling issue, we have added a brief note in Section 3.1.1 (Public topographic data preprocessing) where the resampling procedure is described. The added sentence acknowledges that mean aggregation on steep beach faces may smooth local topographic variations and introduce representativeness differences, which could contribute to the observed error metrics.

Modifications in the revised manuscript:

Section 4.2 (modified text, added sentences after presenting Figure 4(f)):

Among these validation results, the accuracy within the newly predicted intertidal voids (Figure 4(f); $R^2 = 0.61$, RMSE = 0.94 m, MAE = 0.70 m) directly indicates the model's ability to reconstruct the topography originally missing from the DeltaDTM baseline. This sub-meter *accuracy demonstrates the effectiveness of the gap filling approach. The combined product metrics for the entire study area (Figure 4(e);* $R^2 = 0.75$*, RMSE = 1.17 m, MAE = 0.82 m) include the high quality DeltaDTM backshore pixels and therefore reflect the performance of the integrated product, not solely the machine learning prediction component.*

Section 3.1.1 (modified text, added sentence after describing the mean aggregation resampling):

The Lidar DEM was then resampled to the 30 m resolution of DeltaDTM. A mean aggregation method was applied during resampling to preserve area averaged elevation values. It should be noted that on steeper beach faces, where elevation varies rapidly across short distances, this mean aggregation may smooth local topographic variations and introduce representativeness differences between the original 1 m resolution data and the resampled 30 m pixels. Such differences could contribute to the observed disagreement between the predicted elevations and the Lidar reference, particularly in areas with sharp elevation gradients.

Comment 4:

The SHAP analysis is well executed and adds useful transparency to the modelling. The main findings — NIR reflectance correlating positively with elevation, distance-to-coast tracking the cross-shore gradient — align with established coastal remote sensing understanding. The analysis is valuable precisely as confirmation that the model has learned physically sensible relationships. The authors might consider framing it in those terms rather than as a discovery of "physical driving mechanisms," which could slightly overstate what the interpretability exercise reveals.

Reply 4:

We thank the reviewer for this positive and constructive comment. We agree that the SHAP analysis primarily serves to confirm that the model has learned physically sensible relationships, rather than to discover new physical driving mechanisms. The reviewer correctly notes that the main findings, such as the positive correlation between NIR reflectance and elevation and the role of distance to coastline in tracking the cross-shore gradient, align with established coastal remote sensing understanding. Therefore, we have revised the manuscript to reframe the interpretation of the SHAP analysis in more accurate terms, avoiding overstatement of what the interpretability exercise reveals.

Specifically, we have made the following modifications in Section 4.5:

We revised the opening sentence of the first paragraph to state that the visualizations help examine how the model uses input features, rather than interpreting the model from different perspectives in a way that might imply discovery of new mechanisms.

We revised several sentences throughout the section that previously used phrases such as indicates that the model effectively leverages the physical principle, confirms the dominant mechanisms driving the intertidal reconstruction, and clarifies how the model resolves the intertidal interface. These have been replaced with more neutral phrasing that emphasizes confirmation of expected physical relationships rather than discovery.

We revised the final sentence of the section to avoid claiming that the model captures regional-scale variations as a novel finding, instead stating that the patterns align with expectations.

Modifications in the revised manuscript (Section 4.5, full revised text):

Based on the computed SHAP values, three types of visualizations were generated to examine how the model uses input features for elevation prediction. The first is the SHAP summary plot, which conveys feature importance and the direction of impact. Features are ranked in descending order of global significance, represented by their mean absolute SHAP values. Each point in the plot corresponds to a single sample; the color indicates the magnitude of the feature value, while the horizontal position denotes the direction (positive or negative) and strength of its contribution to the prediction. The second visualization is the SHAP feature-interaction heatmap, which illustrates the average interaction intensity between pairs of important features. Finally, the SHAP dependence plot was used to investigate the effects of individual features further. For the six most important features identified in the summary plot, dependence graphs were generated to show the relationship between each feature's actual value and its corresponding SHAP value.

High near-infrared (NIR) reflectance is associated with higher predicted beach elevation (Fig. 12(a)). Among all predictors, B8_20p (the 20th percentile of the near-infrared band) contributes most positively to the model output. This is consistent with the physical principle of strong water absorption in the NIR spectrum, where high NIR values correspond to dry, exposed supratidal sands and low NIR values are characteristic of frequently inundated intertidal zones. Conversely, B3_20p (the 20th percentile of the green band) has the strongest negative effect. High green band reflectance is typically associated with shallow water bodies *or moist sediments in low lying areas, which the model correctly interprets as indicators of lower elevation. Geometric and positional features such as In_dis, Coast_dis, Co_ratio, X, and Y also rank among the most influential variables, indicating that the model uses spatial context to enforce a realistic topographic structure where predicted elevations follow a plausible cross-shore geomorphological gradient.*

The distribution of SHAP values across all samples (Fig. 12(b)) shows that B8_20p has a clear positive contribution trend, while B3_20p exhibits variable and opposing effects. This confirms that the model relies heavily on the moisture sensitive NIR band to resolve the primary elevation gradient from the sea to the land. Spatial and geometric features consistently show positive SHAP values, reinforcing their supportive role in refining the spatial continuity of the beach profile. The variability in SHAP values arises from the heterogeneity of coastal conditions, reflecting that the model captures distinct spectral-topographic relationships across the backshore, foreshore, and intertidal zones.

Nonlinear feature responses (Fig. 12(c)) further illustrate how the model handles the intertidal interface. B8_20p exhibits a sharp monotonic increase in SHAP values at lower reflectance ranges before stabilizing at high levels. This steep rising segment corresponds to the transition from the wet intertidal zone to the dry backshore, demonstrating that the model is sensitive to moisture changes in this critical zone. In contrast, B3_20p displays a pronounced negative trend that levels off at moderate reflectance, consistent with its association with low lying wet areas or shallow water. In_dis shows a nonlinear response with substantial variability, reflecting the complex morphological transition between landward and seaward beach zones. B8_50p follows a positive trend similar to B8_20p but *with earlier saturation, suggesting that using multi-percentile statistics allows the model to capture temporal variations in tidal exposure. Geographic coordinates (X, Y) reveal broad spatial gradients in predicted elevation that are further modulated by spectral variables, indicating regional scale variations in coastal morphology. These findings align with established coastal remote sensing understanding and confirm that the model's predictions are based on interpretable, physically meaningful features.*

Comment 5:

Finally, the conversion from NZVD2016 to EGM2008 orthometric heights (lines 188–190) is an important step but is described only briefly. The difference between these two vertical reference surfaces varies spatially across New Zealand. It would be helpful if the authors could indicate the approximate magnitude and spatial pattern of this offset, so that readers can judge whether residual datum inconsistencies contribute meaningfully to the reported error budget.

Reply 5:

We thank the reviewer for raising this important technical point. The conversion from NZVD2016 to EGM2008 orthometric heights is indeed a critical step for ensuring vertical consistency across datasets. The reviewer correctly notes that the difference between these two vertical reference surfaces varies spatially across New Zealand, and understanding this variation is essential for interpreting the reported error budget.

To address this comment, we performed a quantitative analysis of the offset between NZVD2016 and EGM2008 across New Zealand. Specifically, we computed the difference between the EGM2008 geoid height and the NZGeoid2016 quasigeoid height ($N_{EGM2008}$–N) on a common grid. The resulting spatial pattern is shown in Fig. 1 (which will be included as Fig. S2 in the Supplementary Material of the revised manuscript). As illustrated in the figure, the offset ranges from approximately -1.365 m to +0.015 m across the country. The largest negative offsets (down to -1.365 m) occur in the South Island mountainous regions, where the EGM2008 geoid is substantially lower than the NZGeoid2016 quasigeoid. In coastal and beach areas, which are the primary focus of this study, the offset values are consistently close to zero, with a maximum of 0.015 m.

We have added the following information to the revised manuscript in Section 3.1.1, and we have prepared a new supplementary figure (Fig. S2) showing the offset map, which is reproduced below as Fig. 1 in this reply. Importantly, the added text in Section 3.1.1 only describes the spatial pattern of the offset and refers readers to the supplementary figure, without introducing results or making claims about the error budget.

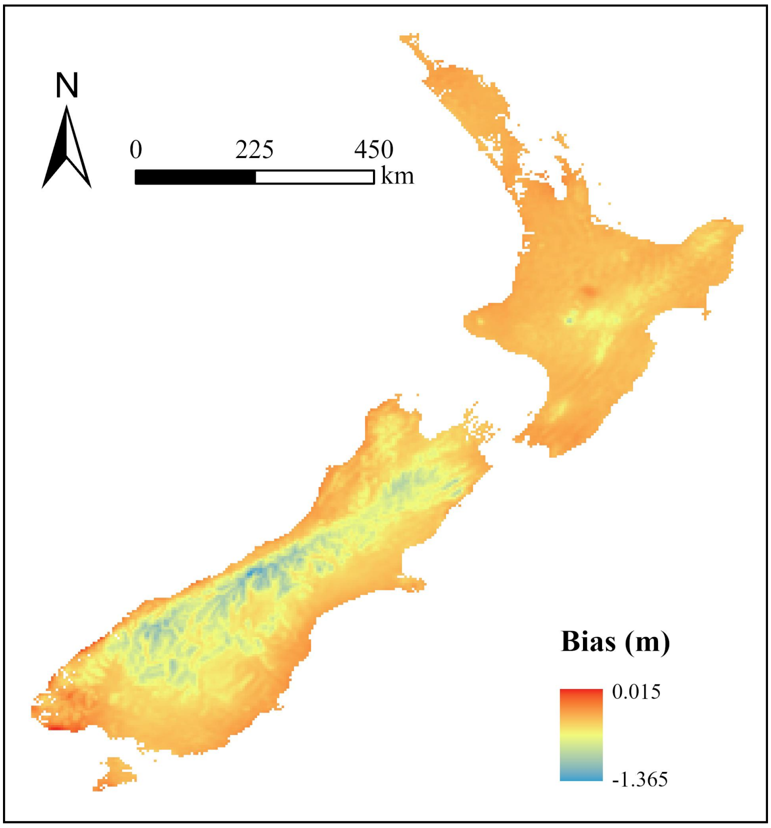

Fig. 1 Spatial distribution of the offset ($N_{EGM2008}$–N) between EGM2008 and NZGeoid2016 across New Zealand. The offset ranges from -1.365 m to 0.015 m. Coastal beach areas show offsets close to zero.

Modifications in the revised manuscript (Section 3.1.1, added after the datum transformation description):

To quantify the spatial variation of the offset between NZVD2016 and EGM2008, we computed the difference between the EGM2008 geoid height and the NZGeoid2016 quasigeoid height ($N_{EGM2008}$*–*N*) on a common grid. The resulting spatial pattern is shown in Fig. S2. The offset ranges from -1.365 m to 0.015 m across New Zealand. Large negative offsets (down to -1.365 m) occur in the South Island mountainous regions, while coastal and beach areas exhibit offsets consistently close to zero (up to 0.015 m).*

Thank you very much for all the comments on this manuscript, which have significantly improved it. If any issues still need to be addressed in the revised version, we would greatly appreciate the opportunity to further revise and improve the manuscript.

Best regards.