

<https://doi.org/10.1038/s43247-025-02550-4>

Benchmarking shoreline prediction models over multi-decadal timescales



Yongjing Mao¹✉, Giovanni Coco², Sean Vitousek³, Jose A. A. Antolinez⁴, Georgios Azorakos⁵, Masayuki Banno⁶, Clément Bouvier⁷, Karin R. Bryan², Laura Cagigal⁸, Kit Calcraft¹, Bruno Castelle⁵, Xinyu Chen^{6,9}, Maurizio D'Anna^{2,10}, Lucas de Freitas Pereira¹¹, Iñaki de Santiago¹², Aditya N. Deshmukh¹, Bixuan Dong¹, Ahmed Elghandour^{13,14}, Amirmahdi Gohari¹⁵, Eduardo Gomez-de la Peña², Mitchell D. Harley¹, Michael Ibrahim¹⁵, Déborah Idier¹⁶, Camilo Jaramillo Cardona¹¹, Changbin Lim¹¹, Ivana Mingo⁵, Julian O'Grady¹⁷, Daniel Pais^{18,19}, Oxana Repina²⁰, Arthur Robinet⁷, Dano Roelvink^{4,14,21}, Joshua Simmons²², Erdinc Sogut²³, Katie Wilson¹ & Kristen D. Splinter¹

Robust predictions of shoreline change are critical for sustainable coastal management. Despite advancements in shoreline models, objective benchmarking remains limited. Here we present results from ShoreShop2.0, an international collaborative benchmarking workshop, where 34 groups submitted shoreline change predictions in a blind competition. Subsets of shoreline observations at an undisclosed site (BeachX) over short (5-year) and medium (50-year) periods were withheld from modelers and used for model benchmarking. Using satellite-derived shoreline datasets for calibration and evaluation, the best performing models achieved prediction accuracies on the order of 10 m, comparable to the accuracy of the satellite shoreline data, indicating that certain beaches can be modelled nearly as well as they can be remotely observed. The outcomes from this collaborative benchmarking competition critically review the present state-of-the-art in shoreline change prediction as well as reveal model limitations, facilitate improvements, and offer insights for advancing shoreline-prediction capabilities.

Sandy beaches provide critical protection to inland areas, support biodiversity, and offer substantial recreational and economic value. However, these dynamic landscapes can evolve rapidly in response to environmental forces that can be exacerbated by changing wave climates and rising sea levels^{1,2}. As such, reliable prediction of key coastal indicators, such as the shoreline position³, across a broad range of timescales from individual storms to multi-decadal climate patterns is essential for the sustainable management of coastal landscapes⁴.

State-of-the-art shoreline prediction goes beyond fitting simple linear regression to historical data⁵. Predictive models help explain dynamic shoreline behavior and guide coastal management decisions^{6,7}. Over the past several decades, dozens of shoreline models have been developed, ranging from physics-based and statistical models to machine learning methods^{8–18}. These models have become more complex and accurate, integrating advanced computational techniques with increasing volumes of data to better simulate shoreline dynamics due to processes such as wave-driven longshore and cross-shore sediment transport and sea-level rise, which are essential for shoreline evolution at event and engineering time scales (i.e.,

days to decades)^{19,20}. Despite model advancements, objective inter-comparisons of model performance under standardized conditions, such as using the same datasets, calibration methods, and evaluation metrics, remain rare in coastal science.

Benchmarking model performance plays a constructive role in the broader process of model selection and confidence building in predictions^{21,22}. By providing a standardized and transparent comparison of predictive accuracy across models, benchmarking helps identify strengths and limitations in model behavior under consistent testing conditions²². While predictive performance is not the only consideration, as factors such as model complexity relative to the modeling scale, representation of physical processes, and data availability are also critical^{23,24}, benchmarking offers an objective starting point that supports informed judgment. When used alongside complementary criteria, benchmarking improves the ability to inform decision-making in dynamic and uncertain environments such as the coastal zone.

Several benchmarking studies within the field of Earth and environmental sciences have been reported in recent years, focusing on model and

A full list of affiliations appears at the end of the paper. ✉e-mail: Yongjing.mao@unsw.edu.au

methodological evaluation and comparison^{21,22,25,26}. To objectively evaluate shoreline model performance, a model competition called Shoreshop was first held in 2018²⁵. In this competition, 20 models were calibrated and trained using 15 years of video-derived shoreline data from Tairua Beach, New Zealand. Participants were then asked to submit their predictions for an additional 3 years of withheld shoreline data to conduct a blind test of model performance. The success of ShoreShop1.0 highlighted the value of model benchmarking in fostering collaboration within the shoreline modeling community. It was also the first initiative to provide an unbiased assessment of the applicability of various shoreline models to a wave-dominated beach.

The lessons learned from ShoreShop1.0 stimulated discussions around advancements in shoreline modeling^{7,20,27}. Probabilistic approaches, rather than deterministic ones, became increasingly adopted^{28,29}, the consideration of model non-stationarity gained prominence^{18,30–32}, and the realization occurred that truly blind model benchmarks are the objective means of ensuring model accuracy while avoiding overtuning. The rapid development and availability of satellite-derived shorelines (SDS)²¹, since the first ShoreShop has also addressed issues related to data scarcity. Despite their larger uncertainties (accuracy ≈ 8.9 m³³) compared to traditional sub-meter-scale surveys, satellite datasets have proven to be robust alternatives for calibration and validation of shoreline models³⁴ especially when high temporal resolution is needed to capture dynamic changes. Satellite observations available over large spatio-temporal scales have also enabled the development of data-driven shoreline models^{17,35} and data-assimilation in hybrid models^{30,34,36–38}.

In this paper, we summarize the outcomes of ShoreShop2.0, an international collaborative benchmarking competition that solicited model submissions from experts across the globe over a six-month period. Building on approaches established in ShoreShop1.0, ShoreShop2.0 advanced the assessment of the state-of-the-art of shoreline models at a natural embayed beach (Fig. 1) by evaluating model's ability to incorporate spatio-temporal dynamics, leverage open-source datasets, and make predictions across short-term (5-year) to medium-term (50-year) timescales using an open-submission system on GitHub. Unique to this second competition was a truly 'blind' test, where participants were not provided with the physical location of the target study site, which was anonymized as BeachX. This benchmarking competition demonstrated the applicability of shoreline models across varying time scales, offering valuable insights into future advancements, establishing a standard for model intercomparison studies, and promoting open science within the coastal research community.

Results

Model submissions

As a benchmarking exercise conducted at an anonymized site (BeachX), with full details provided in the Benchmarking Setup section, ShoreShop2.0 solicited submissions from all types of shoreline models, including physics-based, hybrid, and data-driven models. However, only models defined as data-driven models (DDM) and hybrid models (HM) were submitted. DDMs, including regression, machine learning, and statistical models, rely entirely on data to establish relationships between wave characteristics and shoreline positions. In contrast, HMs include physical constraints through defined mathematical relationships and use data to calibrate free parameters. In ShoreShop2.0, 34 models, including 12 DDMs and 22 HMs, were evaluated and compared as part of the blind competition. Nearly all models were transect-based, with free parameters that were independently associated with and calibrated for each transect, except for four non-transect-based models that used a single set of free parameters for all transects. All submitted models completed the short-term (2019–2023) prediction task, while 29 provided medium-term (1951–1998) predictions, and 20 extended projections for the long-term period (2019–2100). Seven additional DDMs and five HMs were submitted after ShoreShop2.0 and are included here as references for potential model improvements, informed by lessons learned during the workshop and additional insights into the shoreline data; however, they are not considered blind tests because the

initially withheld data were made available immediately following the workshop. For HMs, such as *COCOONED*³⁹, *CoSMoS-COAST*³⁴, *ShoreFor*¹¹, *LX-Shore*¹³ and *ShorelineS*⁴⁰, different versions from various modelers were also evaluated. While most of these models have been validated and applied across different beach types, this benchmarking tested their ability to transfer to an unstudied site. The characteristics of each model submission are provided in Supplementary Table S1, and a detailed description of each model is available in the GitHub and archived repository⁴¹ as individual README files. Previous validation and application practices of the models are summarized in Supplementary Table S2.

Short-term model comparison

With agglomerative-hierarchical clustering⁴², blind model predictions for the short-term period (2019–2023) can be grouped into six distinct clusters based on the dissimilarity of temporal patterns (Fig. 2a). Details of the clustering process are described in the Model Clustering section. Cluster 1 & 2 (Fig. 2b–d) consist of HMs, most of which rely on the MD04⁴³ or Y09⁹ empirical shoreline models to quantify cross-shore sediment transport. These two clusters are characterized by sharp shoreline retreat in response to storms, followed by gentle recovery, which is evident in the ensemble of Cluster 1 & 2. The main distinction between Cluster 1 and 2 is their approach to incorporating longshore sediment transport. Most models in Cluster 1 either do not explicitly model longshore sediment transport (e.g., *Y09_LFP*, *SLRM_LIM*, and *EqShoreB_MB*) or incorporate it using beach rotation models (e.g., *IH_MOOSE_LFP*¹⁴), while models in Cluster 2 adopt CERC-like equations⁴⁴ to quantify shoreline change related to gradients in longshore sediment transport.

Cluster 3 & 4 (Fig. 2e–g) consist of a mixture of HMs and DDMs. Models in these clusters have relatively low-frequency variation and smooth trends. Cluster 3 includes the three best-performing models for the short-term period (i.e., *GAT-LSTM_YM*, *iTransformer* and *CoSMoS-COAST-CONV_SV*, ranked in Supplementary Fig. S1) with coherent variability independent of model type. All the HMs in Cluster 4 incorporate longshore sediment transport with CERC-like equations. Although some of them (e.g., *CoSMoS-COAST* models) use MD04⁴³ or Y09⁹ model for cross-shore sediment transport, the models in Cluster 4 are less responsive to storms compared to the models in Cluster 1 & 2.

Cluster 5 & 6 (Fig. 2h–j) consist of DDMs that struggle to predict shoreline positions (based on the results in Supplementary Fig. S1). Among these models, *SARIMAX_AG*, *XGBoost_AG*, and *Catboost_MI* in Cluster 5 are characterized by high-frequency fluctuations that correspond closely to daily wave characteristics. In contrast, models like *SPADS_AG*, *ConvLSTM2D_LFP* and *wNOISE_JAAA* in Cluster 6 exhibit less noise but struggle to accurately capture shoreline variability. As a result, the ensemble of models in Clusters 5 and 6 exhibits the highest noise and the lowest accuracy. Across all clusters, transects 2 and 8, which represent the ends of the beach and experience larger shoreline variations, are predicted more accurately, whereas transect 5, with its smaller and more irregular variations, presents a greater prediction challenge.

Medium-term model comparison

With the timescale of analysis increasing from short-term (5 years) to medium-term (50 years), the clustering of model predictions changes (Fig. 3a). The first cluster of medium-term predictions is the same as Cluster 6 of short-term and includes noisy DDMs. Despite their daily-scale variations, the inter-annual variability of these models is comparable to those smoother models in Cluster 2, which mostly overlaps with short-term Cluster 3 and represents the best-performing models. Models in Cluster 3 & 4 of the medium-term predictions have large overlap with Clusters 2 & 1, respectively, of the short-term predictions. These model predictions feature large and quick responses to storms, which became more evident in the medium-term with more severe storm events (e.g., in 1972 and 1974) observed. Model ensembles in Cluster 3 & 4 tend to predict larger shoreline erosion in response to these events than other clusters.

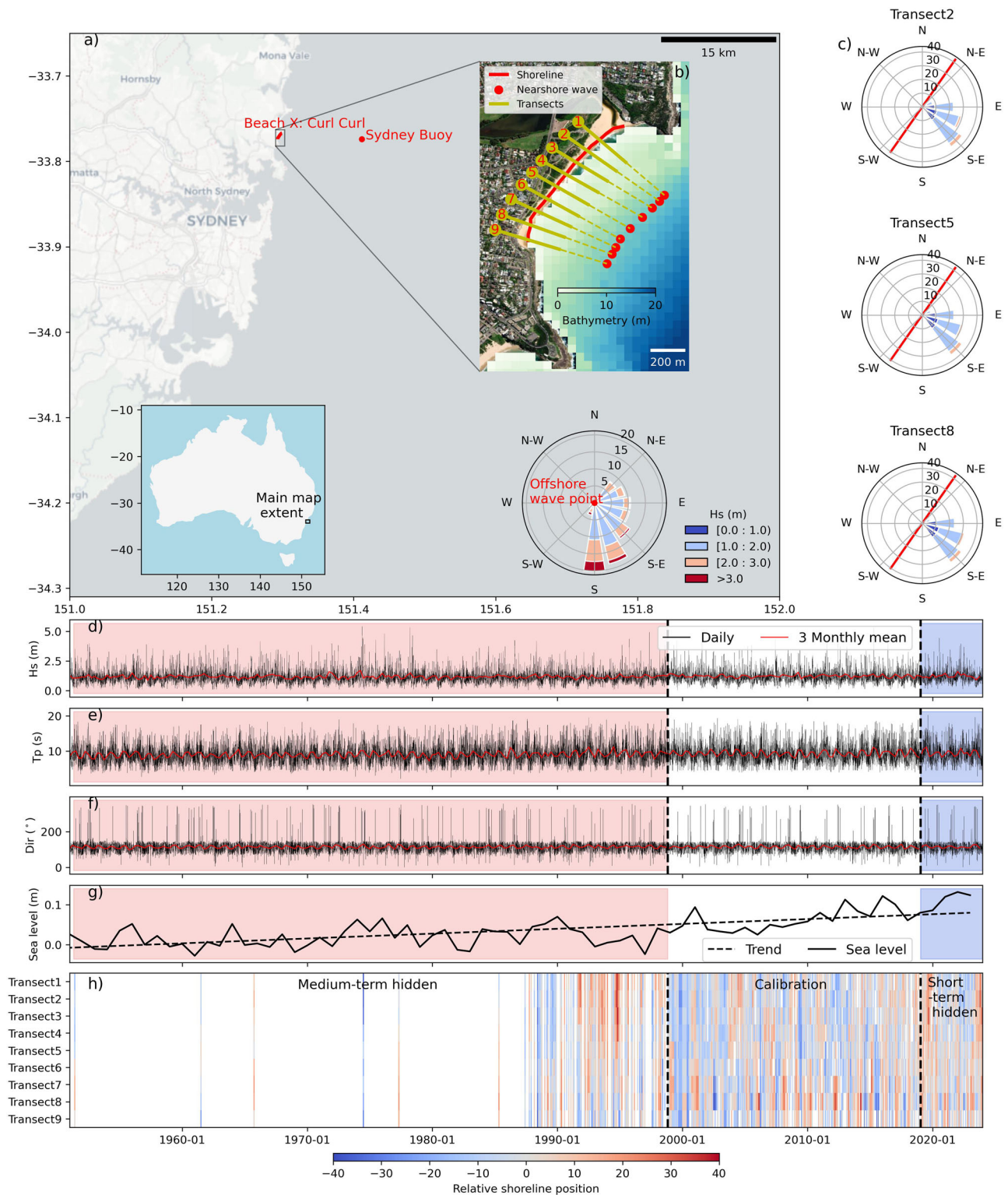


Fig. 1 | Study site and input data. **a** Location of BeachX. Inset wave rose shows the location and distribution of offshore waves from ERA5. **b** The detailed map of BeachX (Curl Curl Beach, New South Wales, Australia). The color gradient at the seaside indicates bathymetry. Yellow lines represent transects. Red dots show the location of nearshore waves. **c** Nearshore (depth = 10 m) wave roses for transects 2, 5 and 8. The red solid line represents the mean beach orientation. **d** Significant wave height H_s (**e**) peak wave period (**f**) mean wave direction (**g**) annual mean sea level

and its trend. **h** Spatio-temporal distribution of relative (de-measured) shoreline position (blue/negative values indicating erosion and red/positive values indicating accretion). Basemaps: CartoDB Positron (©OpenStreetMap contributors) in panel **a**; Esri World Imagery (© Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP and the GIS User Community) in panel **b**. Basemap tiles are accessed via the Contextily Python package.

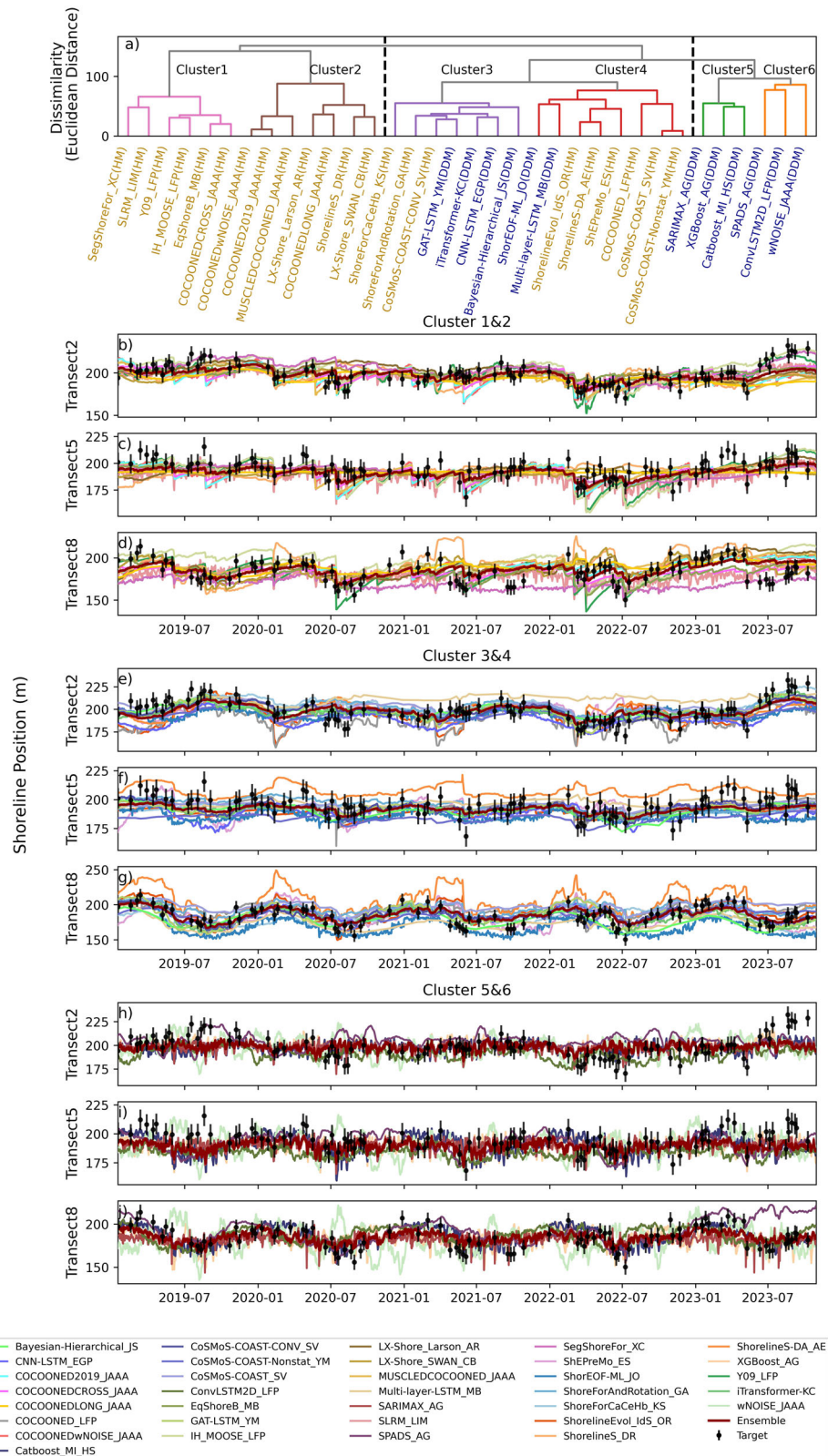


Fig. 2 | Clustering of short-term model predictions from the ensemble of blind-test submissions. **a** Dendrogram resulting from Euclidean distance-based Ward's minimum variance clustering³¹. Blue and brown colors of tick labels represent DDM and HM, respectively. **b–j** Short-term prediction of shoreline positions from different clusters of models for different transects. The deep red line is the ensemble

mean (interval mean between 5th and 95th percentiles) of models within each cluster. Black scatters with error bars are SDS shoreline positions with 8.9 m RMSE. The predictions by each individual model can be visualized using the online, interactive version of this plot (https://shoreshop.github.io/ShoreModel_Benchmark/plots.html).

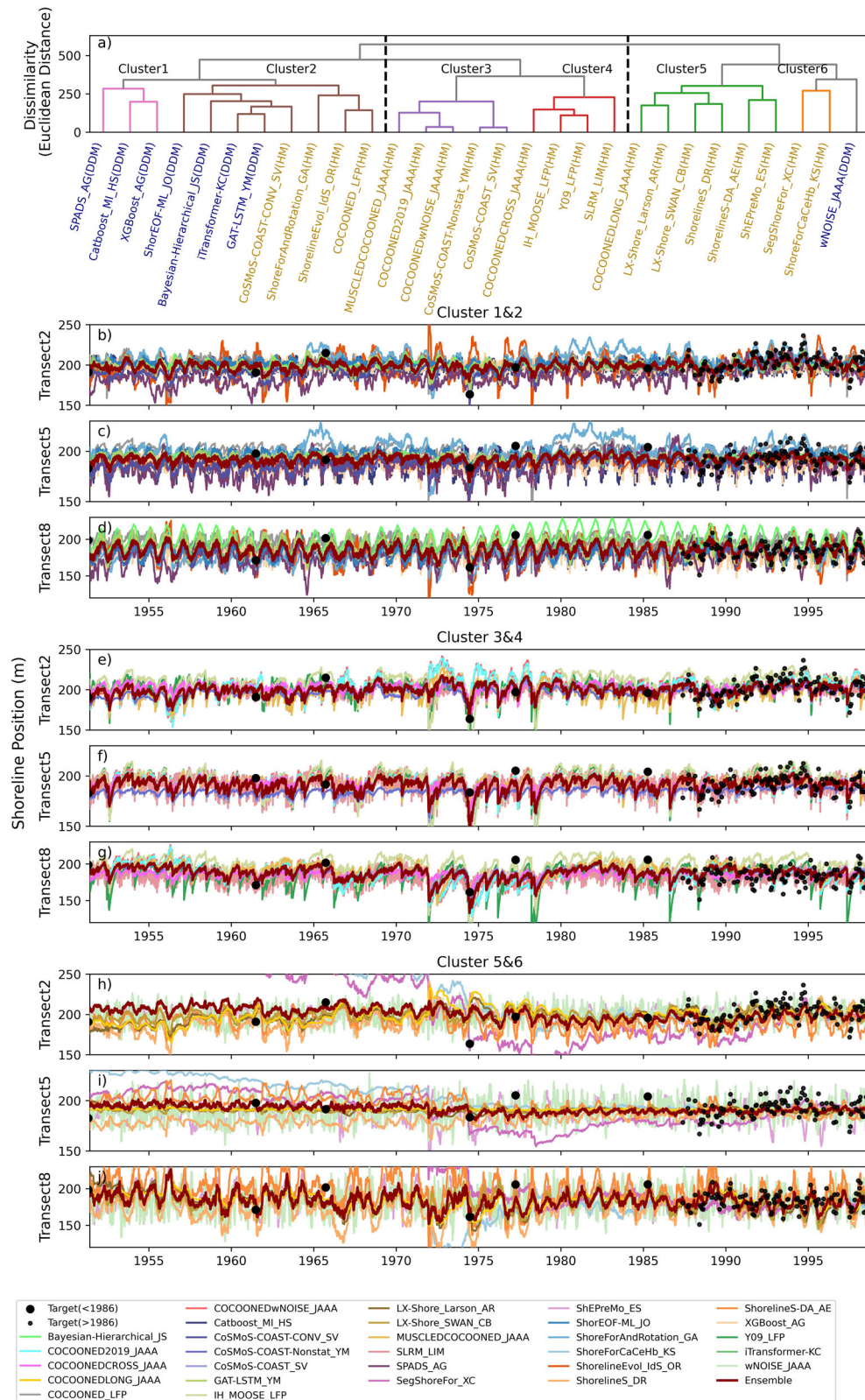


Fig. 3 | Clustering and ensemble of medium-term model predictions.

a Dendrogram resulting from Euclidean distance-based Ward's minimum variance clustering³¹. Blue and brown colors of tick labels represent DDM and HM, respectively. **b–j** clusters of medium-term prediction of shoreline positions for different transects. Deep red line is the ensemble mean (interval mean between 5th and 95th

percentiles) of models within each cluster. Black circles and dots represent the target shoreline position pre and post 1986 respectively for better visualization. The predictions by each individual model can be visualized using the online, interactive version of this plot (https://shoreshop.github.io/ShoreModel_Benchmark/plots.html).

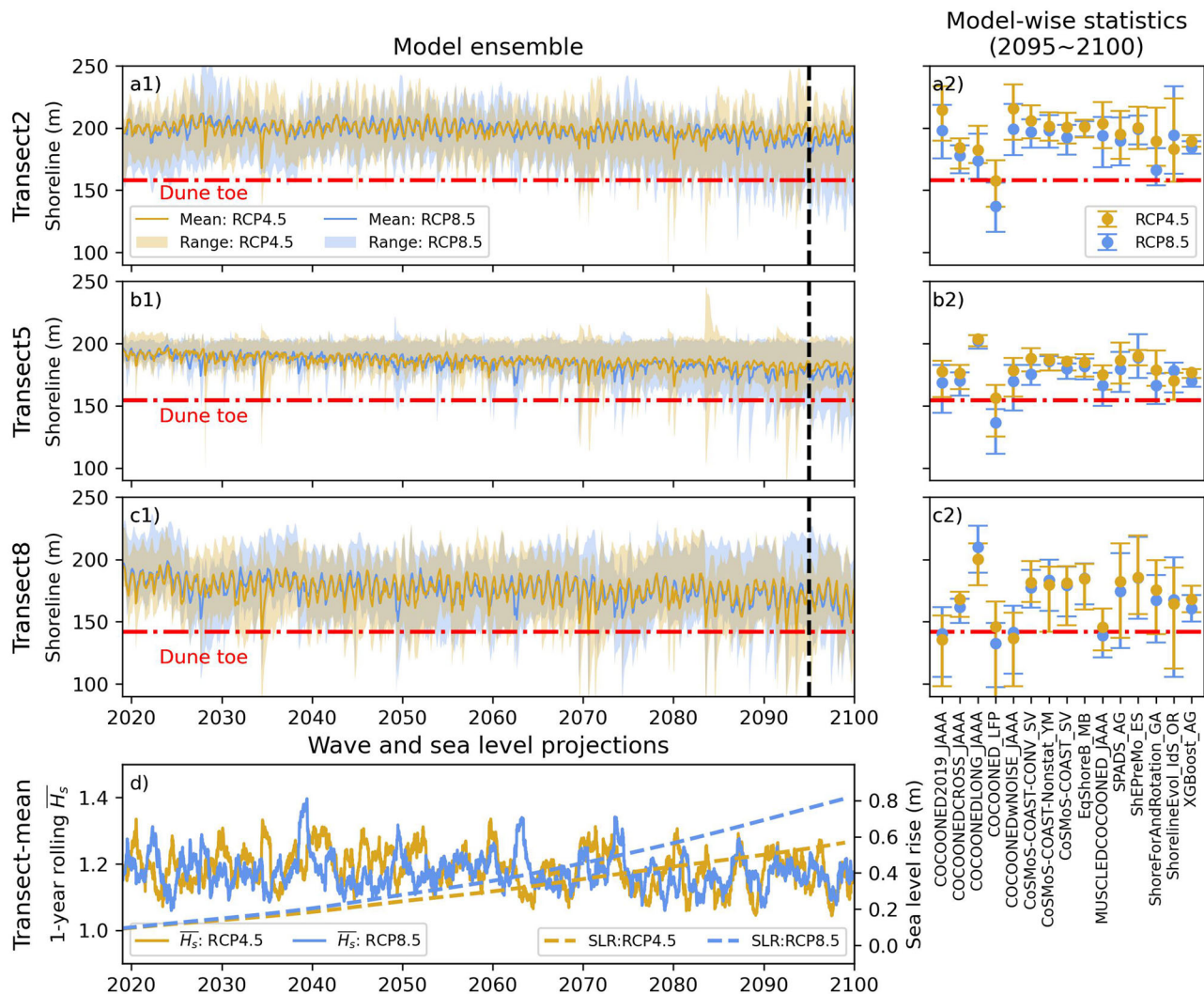


Fig. 4 | Long-term shoreline projections in response to waves and sea-level rise. **a1–c1** Ensemble of monthly long-term shoreline projections in RCP4.5 and RCP8.5 scenarios, including only models that account for sea-level impacts. Solid lines are ensemble means while the shaded areas represent the range between minimum and maximum projections. The red dash-dot line marks the position of the present-day dune toe. **a2–c2** Model-wise statistics of shoreline projections between 2095 and 2100. Circles represent means while caps indicate the range

between temporal minimum and maximum. **d** Wave and sea-level projections. Solid lines are the 1-year running backwards mean of significant wave H_s , while dashed lines are yearly sea-level rise with respect to mean sea-level recorded between 1995 and 2014. Projection of each individual model can be visualized in the online, interactive version of this plot (https://shoreshop.github.io/ShoreModel_Benchmark/plots.html).

Cluster 5 & 6 consist of model predictions with larger medium-term variations for different reasons. For the models in Cluster 5, shoreline change is primarily driven by gradients in longshore sediment transport, resulting in planform response and redistribution of sediment, contrary to the episodic beach erosion caused by cross-shore sediment transport^{43,45,46}. The large variation of model performance in Cluster 6 is attributed to the extreme sensitivity of the ShoreFor model to shifts in wave climate^{11,30,47}. As the hindcast wave data uses different observations for data assimilation pre and post 1979⁴⁸, the wave climate changes slightly around 1979. This minor change in the distribution of waves leads to the large long-term divergence of the ShoreFor-based models (e.g., *SegShoreFor_XC* and *ShoreForCaCeHb_KS*) unless additional modeling techniques to address this issue are included (*ShoreForAndRotation_GA*).

Long-term model comparison

Although prediction of future state is a common goal among modeling applications, the accuracy of long-term (2019–2100) model projections cannot be critically evaluated due to the absence of observational data. Instead, the ensemble and variability of these projections can be used for

statistical analysis of long-term coastal erosion risks (Fig. 4). Here, the 15 models incorporating sea-level rise (Supplementary Table S1) are included in the analysis. The ensemble projections (Fig. 4a1–c1) in both future climate scenarios exhibit strong seasonal and interannual variability driven by the variation of wave climates (Fig. 4d). This variability is more pronounced than the long-term trend of shoreline retreat caused by sea-level rise (Fig. 4d), particularly at transects 2 and 8. With the combined impacts of changing wave climates and sea-level rise over time, the frequency of shoreline erosion reaching the cross-shore location of the present-day dune toe increases with time. Similar to the first five years evaluated in the short-term comparison, the final five years of the 21st century (2095 ~ 2100, Fig. 4a2–c2) show that most models continue to provide consistent shoreline prediction statistics. Only a few models (one for transects 2 and 5, and four for transect 8) project that the average shoreline position will reach the present-day cross-shore location of the dune toe. However, when wave-driven shoreline erosion and seasonal effects are considered (i.e., the temporal variation of the predictions), the dune-erosion risk increases, particularly at transect 8, where 7 out of 15 models project maximum seasonal shoreline erosion to reach the present-day dune toe in both RCP 4.5 and

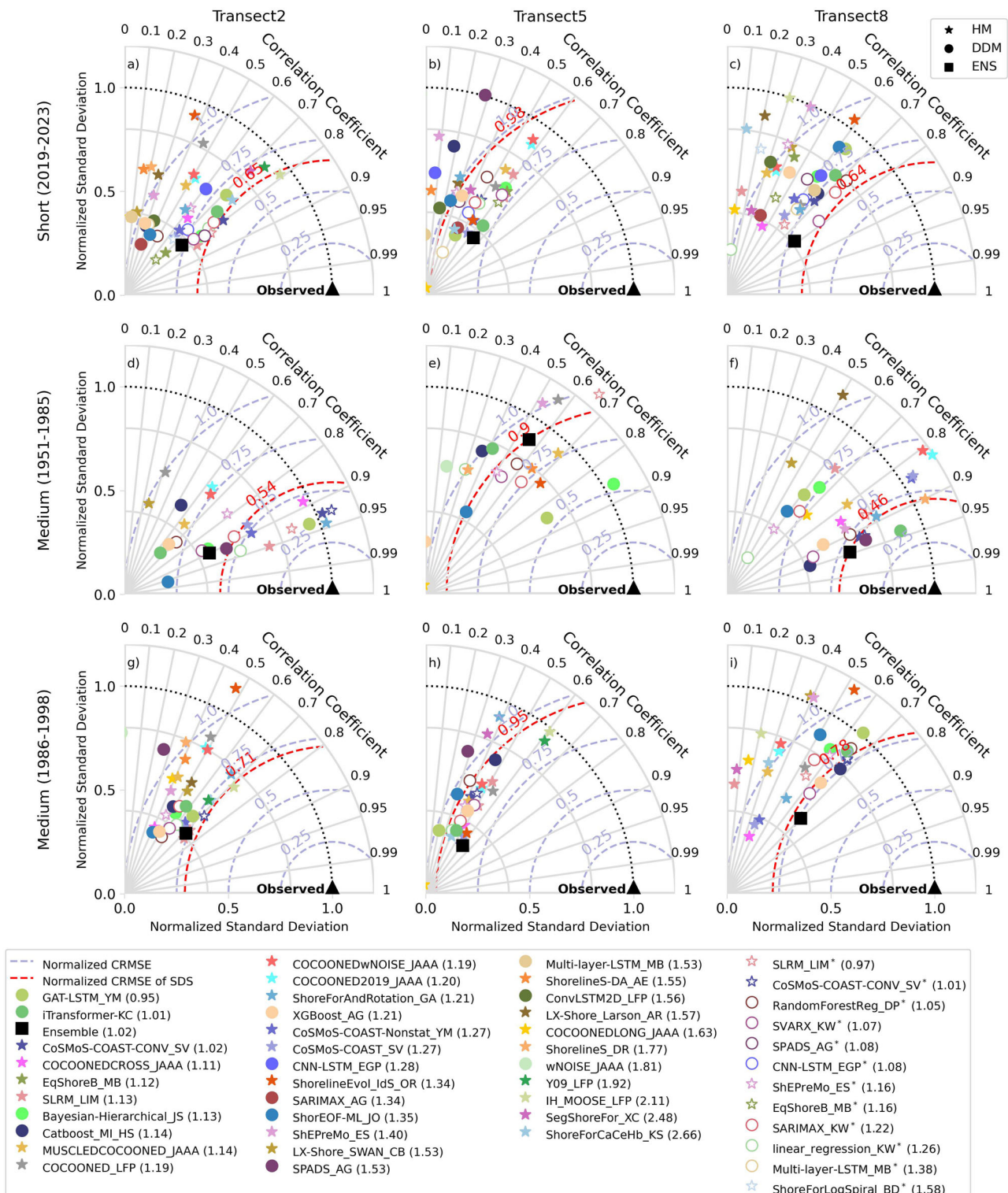


Fig. 5 | Model performance in Taylor diagrams. a–i Taylor diagram for different transects and timescales. The diagrams show the normalized standard deviation (radial - x and y-axis), correlation coefficient (curved axes along the circumference of the circle), and normalized centered root mean square error (CRMSE, concentric dashed arcs). Stars, circles, and squares represent HM, DDM, and ensemble mean respectively. Solid and hollow markers distinguish models submitted before (blind) and after (non-blind) ShoreShop2.0, respectively. The black triangle (Observed) shows the observed data in a Taylor diagram with zero error. The model

performance is indicated by the distance of scatter points of model predictions to the observed. The red dashed arc indicates the normalized RMSE of SDS (8.9 m) with respect to the observed shoreline standard deviation (STD) for that time period. The legends are sorted based on the average loss \bar{L} (displayed within the bracket) for all transects and timescales where predictions are available. The superscript * after a model name indicates non-blind models submitted after ShoreShop2.0. The Taylor diagrams and model ranking for each timescale can be found on (https://github.com/ShoreShop/ShoreModel_Benchmark).

RCP 8.5 scenarios. For transect 8, most models project similar shoreline positions during the 2095–2100 period for both scenarios in terms of temporal minimum, maximum, and mean. However, the difference between the RCP scenarios is substantially larger for transects 2 and 5, with most models projecting greater erosion in the RCP 8.5 scenario.

Model metrics

The Taylor diagram⁴⁹ and related loss function (\mathcal{L} , refer to Eq. 2 in Methods) are used to benchmark model performance in ShoreShop2.0. Models are ranked based on the average loss $\bar{\mathcal{L}}$ across all the different transects and for each timescale (Fig. 5). The evaluation for the medium-term task is separated into pre-1986 (1951–1985) and post-1986 (1986–1998) periods due to differences in the density and source of target data (i.e., photogrammetry versus satellite). In the majority of the Taylor diagrams on Fig. 5, the centered root mean square error (CRMSE) of models reaches the intrinsic accuracy (8.9 m) of SDS as reported for the adjacent Narrabeen Beach²¹, suggesting that the model accuracy starts to be limited by the accuracy of shoreline data used to train and validate the models. Examining Fig. 5 in more detail, the general model performance is comparable for the two ends of the beach, transects 2 (left column) and 8 (right column), across all periods and is substantially better than for transect 5 representing the center of the embayment (center column). This is because the ends of the embayed beach oscillate with the seasonal directional wave climate, whereas the center of the embayment may be more influenced by contrasting cross-shore and alongshore processes or the alongshore propagation of sand waves and sandbars through the middle of the beach. The model performance for medium-term prediction (Fig. 5d–i) is comparable if not better than for the short-term period, demonstrating the potential of the suite of shoreline models available for this benchmarking competition to reliably predict up to 50 years of coastal variability and shoreline change. The better skill metrics of Medium (1951–1985) (Fig. 5d–f) compared to other periods can be attributed to two factors. First, there are only six data points available pre-1986 for validation using the available photogrammetry data compared to more than 100 data points available in other periods (refer to Fig. 3), which will undoubtedly influence the error statistics. The aerial photogrammetry dataset also generates a full beach profile above mean sea level from which a specific mean sea level (MSL) shoreline contour can be extracted. Shoreline data based on MSL contours are less susceptible to noise than SDS data containing errors associated with tides, wave setup and runup²¹. The limitation of SDS data is further described in the discussion section.

Comparing the average loss across all three periods, the top 3 performing models were the *GAT-LSTM_YM*, *iTransformer-KC*, and *CoSMoS-COAST-CONV_SV*, two of which are DDM. The *GAT-LSTM_YM* model was the top-performing Medium (1951–1985) model and *CoSMoS-COAST-CONV_SV* was the top-performing model for both Short (2019–2023) and Medium (1986–1998) tasks. The median $\bar{\mathcal{L}}$ of HMs (1.27) was marginally better than DDMs (1.28). In contrast to ShoreShop1.0 in 2018 with the model ensemble recognized as the top-performing prediction, several individual models outperformed the ensemble in ShoreShop2.0. The predictions from most models are highly correlated, with only a few model pairs showing statistical non-correlation (P value > 0.01 in Pearson's non-correlation test; Supplementary Fig. S2). With availability of previously hidden shoreline data and input-data pre-processing methods learned from the ShoreShop2.0 in-person workshop held in October 2024, all the non-blind model submissions except for *EqShoreB_MB* improved their accuracy. The detailed loss scores for each model and for different transects and tasks can be found in Supplementary Fig. S1.

The model performance was further evaluated using quantile-quantile plots and metrics used in ShoreShop1.0 for the short-term and medium-term (1986–1998) tasks with abundant target data (Fig. 6a–f). Although most models have high quantile-quantile correlations with the target data, biases are evident in several models (Fig. 6a–f). Notably, the underestimation of extreme shoreline positions is a recurring issue for many models, particularly for transects 2 and 5, a limitation that was also identified in ShoreShop1.0²⁵.

Following ShoreShop1.0, the Mielke's modification index λ ⁵⁰ that accounts for both bias and dispersion is also used to evaluate model performance. λ values range from 0 to 1, with $\lambda = 1$ representing perfect agreement, and $\lambda = 0$ representing no agreement between observation and prediction. Compared to ShoreShop1.0 that benchmarked models over a 3-year period at Tairua Beach, NZ, the λ value in ShoreShop2.0 shows slight improvements in some instances (Fig. 6g) despite the use of less accurate and less frequent shoreline data available for training. However, model performance is also not necessarily consistent across all transects and timescales as indicated by the range of λ for each model. Some of the best short-term models are also the worst medium term performers (e.g., *IH_MOOSE_LFP* and *SegShoreFor_XC*) whereas some other models exhibit more consistent metrics (e.g., *CoSMoS-COAST-CONV_SV*, and *GAT-LSTM_YM*) across different timescales for transects 2 and 8. This is attributed to the different governing physics and architectures of the models used here in ShoreShop2.0. Most non-blind models substantially improve their score for individual transects and tasks; however, the consistency of performance shows less improvement in the non-blind models submitted after the workshop.

Discussion

ShoreShop2.0 highlighted substantial advances in shoreline modeling over the past 6 years since ShoreShop1.0, owing to advancements in data availability from satellite derived shorelines as well as algorithmic improvements, particularly in data-driven modeling. Competition participants were provided with information for an unnamed embayed beach, BeachX, including timeseries of observed shorelines, representative bathymetry, water level timeseries, and inshore directional wave climate. While the individual shoreline models that were submitted to ShoreShop2.0 include a variety of different processes, including onshore/offshore sediment transport, gradients in longshore transport, and shoreline change induced by sea-level rise, they all exhibit strong predictive capability across a range of different timescales relevant to decision making and planning. In general, the models successfully capture both the overall temporal variability and trends as well as the response to storms in not only short-term (e.g., 5-year) but also medium-term (e.g., 50-year) predictions. The top-performing models as part of the blind competition, such as *CoSMoS-COAST-CONV_SV*, *GAT-LSTM_YM*, and *iTransformer-KC*, outperform the ensemble and deliver accurate and similar shoreline predictions (Supplementary Fig. S2) across all timescales. Among these models, *CoSMoS-COAST-CONV_SV* is a hybrid model that explicitly integrates longshore and cross-shore sediment transport, sea-level rise, and long-term residual trends with discrete convolution operations to generate predictions. *GAT-LSTM_YM* is a data-driven model using Graph Attention Network (GAT)⁵¹, and Long Short-Term Memory (LSTM)^{17,52}, network to model spatial and temporal variation of shorelines, respectively. *iTransformer* is a different data-driven model leveraging the transformer architecture and the self-attention mechanism to model multivariate time series of shoreline positions across transects. These models, despite their substantially different model architectures, can capture the observed shoreline evolution accurately. In addition, many models now demonstrate performance approaching the internal error of SDS (8.9 m²¹, Fig. 5). This suggests that the accuracy of shoreline data used for calibration and validation has likely become one of the primary constraints limiting further improvement in shoreline model performance.

With advances in machine-learning methods over the past 10 years, competitors hypothesized prior to ShoreShop2.0 that data-driven models (DDMs) would likely outperform more constrained hybrid models that were the focus of ShoreShop1.0. This, however, was not necessarily the case. Both the best-performing and median models from the DDM and HM groups achieved similar accuracy. This could be due to several factors related to the shoreline data, including the plateauing accuracy, as noted above, as well as the low and irregular (~weekly) temporal resolution that can further complicate the development of accurate DDMs. It is anticipated that as more and better satellite data become available with the ever-increasing suite of CubeSats⁵³, DDMs may continue to improve and eventually outpace performance of traditional models.

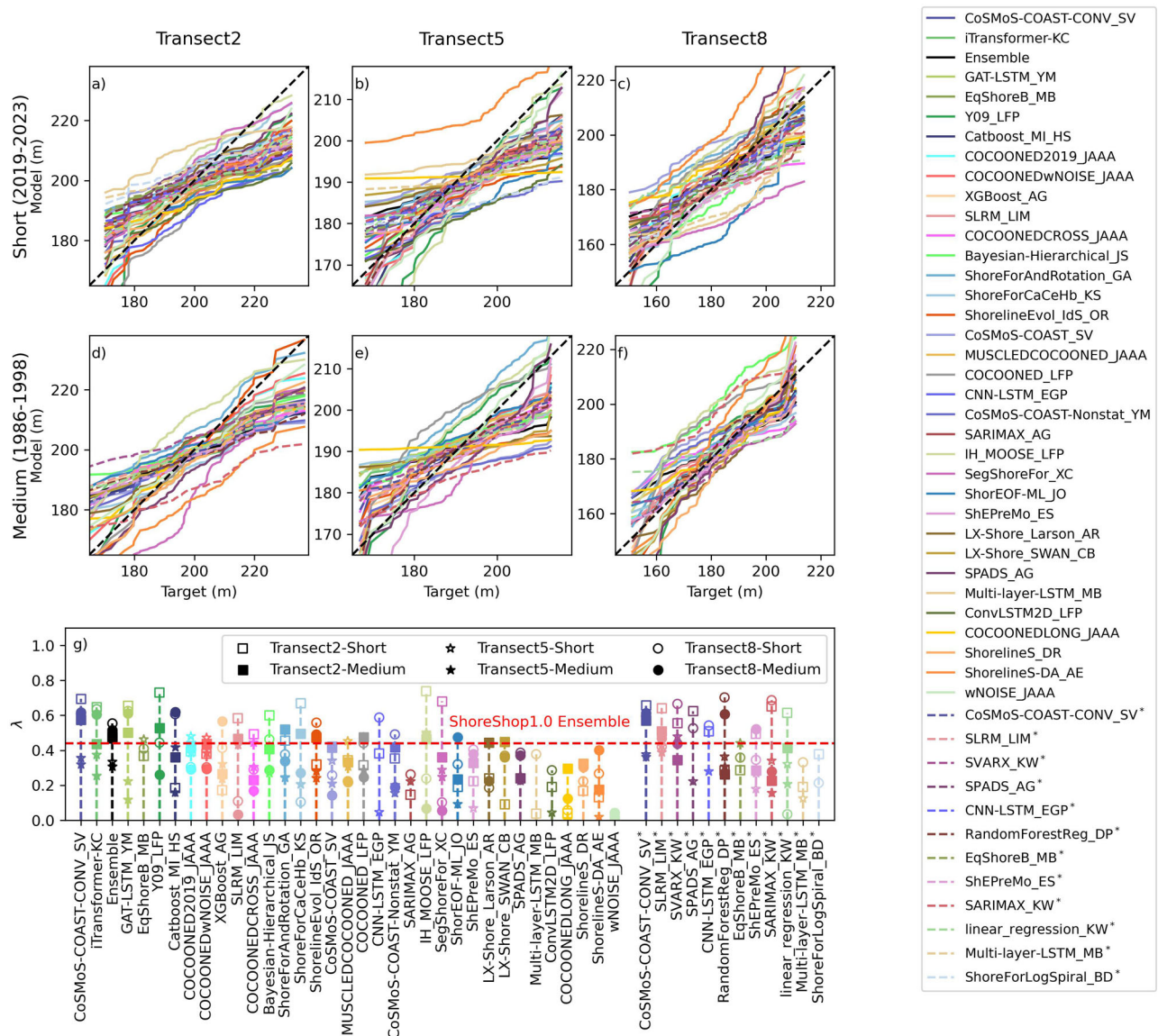


Fig. 6 | Blind model performance for short and medium-term model predictions. **a–f** Quantile-Quantile plots for the three target transects across short-term (2019–2023) and medium-term (1986–1998) timescales. **g** Mielke's modification (λ). Squares, stars, and circles correspond to transects 2, 5, and 8, respectively, while hollow and solid markers distinguish short-term (2019–2023) and medium-term (1986–1998) results. The horizontal dashed red line indicates the ensemble model

metrics reported in ShoreShop1.0. Models are arranged based on the average loss function $\bar{\mathcal{L}}$ across target transects for the short-term prediction. The superscript * after a model name indicates non-blind models submitted after ShoreShop2.0. The Quantile-Quantile correlation of each individual model can be found in the online, interactive version of this plot (https://shoreshop.github.io/ShoreModel_Benchmark/plots.html).

ShoreShop2.0 highlights several opportunities for future advancements. In general, coastal monitoring is data poor in many regions with limited in-situ long-term datasets available^{54–57}. The availability of satellite-derived products has vastly increased the number of coastal observations available to develop and test models against. However, when compared to other remote sensing methods, such as fixed cameras⁵⁸, the satellite-derived shoreline data has relatively low temporal frequency (weekly to monthly) and moderate accuracy ($RMSE \cong 8.9$ m) arising from geo-referencing issues, satellite pixel footprint, as well as the temporal variability of the instantaneous water line, including wave setup and runoff²¹. This noise has a particularly pronounced impact on Transect 5, where the RMSE of the satellite-derived shorelines is comparable to the standard deviation of the observed shoreline positions. As a result, it becomes challenging for models to distinguish genuine shoreline variability from noise in the training data, leading to substantially poorer model performance at this transect compared to the other two. A key lesson learned during the

workshop discussions was that data preprocessing became a critical factor influencing shoreline prediction skill beyond the models themselves. For instance, two of the best-performing models, *CoSMoS-COAST-CONV_SV* and *iTransformer-KC*, applied spatio-temporal smoothing and interpolation techniques on the shoreline data used for calibration, practices not widely adopted by most other modelers in this blind competition. Given the same smoothed data (provided using the robust 2D smoothing method^{59,60} as used by *CoSMoS-COAST-CONV_SV*) the modelers who chose to resubmit post-workshop also achieved improved model skill compared to their original submissions (Fig. 5).

High-quality inshore local wave data is also very hard to achieve due to inaccuracies in offshore wave hindcasts, as well as complex wave transformation processes across partially unresolved bathymetry. Moreover, using daily mean wave conditions instead of peak values can underestimate wave energy and shoreline retreat driven by extreme events. A further source of uncertainty arises from extracting wave data at the 10 m depth contour

rather than at the breaking point. Not accurately refracting the waves into the coast has been shown to induce spurious alongshore transport gradients in models that simulate these processes⁶¹, and may partly explain the underperformance of longshore-only models in the medium-term predictions. To account for errors in the modeled wave direction, some models (e.g., *CoMoS-COAST*) in *ShoreShop2.0* applied a directional bias correction to the inshore wave data, aligning mean wave directions to shore-normal to reduce spurious alongshore transport gradients and improve long-term model stability, an approach that followed from similar lessons learned at the nearby Narrabeen Beach^{61,62}.

Beyond data quality, data requirements also impact models differently. While most DDMs and HMs perform well with only wave, water level, and shoreline datasets to train on, models such as *LX-Shore*¹³ and *ShorelineS*⁴⁰, which are not transect-based, require detailed information on headland contours and accurate bathymetry in order to be skillful. Both datasets were provided during the competition at the request of *ShorelineS* and *LX-Shore* modelers and were found to be critical for improving these models' performance. While headlands can be retrieved from satellite imagery, obtaining a reliable nearshore bathymetry is often not feasible along much of the world's coastline. Bathymetry estimation using remote sensing techniques is a promising area of active research^{63–68}.

For long-term predictions, HMs in *ShoreShop2.0* heavily rely on the 60-year-old Bruun rule⁶⁹ to simulate sea-level-driven shoreline retreat. The limitations of this simple model have led to questions over the reliability of its use for long-term predictions⁷⁰. Developing alternative approaches that better account for complex shoreline responses to sea-level rise^{15,71,72} will be crucial for improving these models. The benchmarking of these alternative methods to model the shoreline retreat due to sea level rise will likely be possible in the near future with the availability of global-scale long-term datasets of SDS^{21,73}.

Notably absent from the submissions of models were commercial (e.g., GENESIS⁸) and physics-based models (e.g., Delft3D, MIKE21). The authors acknowledge that the absence of physics-based models may be due to their complexity and reliance on extensive data, which were not available under the data-poor conditions of this study. Similarly, to the authors' knowledge, there was no model that explicitly attempted to model the cross-shelf exchange of sediment (e.g., *ShoreTrans*¹⁵) and very few models (e.g., *LX-Shore*⁷⁴) incorporated non-erodible features and sediment budgets. While most shoreline models assumed unlimited sediment supply and relied on wave and shoreline data for model training and validation, as these are the most readily available data over large spatial and temporal scales¹⁵, other coastal processes such as cross-shelf movement of sediment and subsurface sediment availability can play an important role in longer-term sediment budgets⁷⁵. Many of the models submitted to the *ShoreShop2.0* competition implicitly accounted for these processes through the calibration of model free parameters that relate observed shoreline variability to the combined effects of waves, water levels, and unresolved processes^{12,47}. Inclusion of models that explicitly account for more coastal processes, such as cross-shelf exchange of sediment and the impact of human interference (i.e., structures or nourishment) in future comparative studies would enhance the comprehensiveness of the evaluation framework.

Building on the results of *ShoreShop2.0*, future shoreline benchmarking efforts can be improved in several key areas. Both *ShoreShop1.0* and *ShoreShop2.0* focused on natural embayed beaches, but with different data availability. The relatively regular shoreline time series, along with the absence of sediment sources, sinks, and engineering activities, make these beaches suitable for modeling by most approaches (Supplementary Table S2)—particularly data-driven models (DDMs) and transect-based hybrid models (HMs) that rely predominantly on relationships between wave forcing and shoreline position to predict daily to multi-decadal shoreline change. However, the benchmarking results from these beach types may not be directly transferable to other coastal settings. To better evaluate model capabilities in other environments of interest, future benchmarking might focus on more challenging sites, including those with complex geomorphology, human interventions, and changing sediment

budgets. More accurate shoreline and wave data, along with a wider range of supporting datasets—such as sediment budgets and engineering histories—might be provided to support more comprehensive modeling approaches. This would also enable the inclusion of physics-based models. Finally, while *ShoreShop1.0* and *2.0* used metrics that tended to reward models with smooth and stable predictions, future benchmarking efforts may benefit from exploring additional metrics, including those that evaluate performance at event scales or under extreme conditions, which are particularly relevant in the context of coastal management and planning.

It is important to clarify that this research does not aim to serve as a prescriptive guide for selecting shoreline models in operational decision-making contexts. While our benchmarking exercise provides insight into the relative predictive skill of different modeling approaches, model selection in practice should also consider a range of other factors, including model complexity, physical process representation, alignment with stakeholder needs, etc.^{23,24}. To support informed application, we provide supplementary Table S1 and README files describing each model's structure and process representation. These details, when combined with performance benchmarking, allow practitioners to assess which models may be suitable for their specific management or research objectives.

Methods

Benchmarking setup

The *ShoreShop2.0* benchmarking exercise was conducted at Curl Curl Beach, New South Wales, Australia (Fig. 1a, b). To ensure a blind testing environment, the site was anonymized and referred to as BeachX during the competition. While the planform shape and beach orientation were preserved, all geospatial references were removed to prevent participants from identifying the actual location. Participants were informed only that BeachX was an embayed sandy beach.

The dataset provided for the competition included several open-source inputs used for model calibration and prediction. Daily mean directional wave characteristics at the 10 m depth contour were supplied from 1940 to 2100 (Fig. 1d–f), with hindcast and forecast data derived from global wave models^{76,77} and downscaled using the BinWaves approach⁷⁸. Shoreline positions were derived from satellite imagery using the CoastSat toolkit³³ and were available at approximately fortnightly intervals from 1999 to 2018 (Fig. 1h) for model calibration. Shorelines for validation were primarily sourced from the same dataset as calibration, with additional data derived from open-source photogrammetry⁷⁹ to extend coverage into the pre-satellite era (1951–1985). Validation shoreline positions were unavailable to participants. In addition to wave and shoreline data, tidal data from the FES2014 global model⁸⁰, as well as historical and projected sea-level data based on buoy measurements⁸¹ and regional projections⁸², were also provided.

To capture site-specific coastal characteristics, participants were also given geomorphic parameters including mean grain size, depth of closure estimated following Hallermeier equation⁸³, and beach face slope⁸⁴. Shoreline data were extracted along nine shore-normal transects spaced 100 m apart; however, model evaluation focused only on transects 2, 5, and 8, representing the northern, central, and southern sections of the beach, respectively. To support non-transect-based area models, such as *LX-Shore*¹³ and *ShorelineS*⁴⁰, representative bathymetry and headland contours were also provided as essential inputs. Detailed information about the characteristics of the target site and the processing of input data can be found in the Site description and Data collection and preprocessing sections.

Unlike *ShoreShop1.0*²⁵, which focused on short-term prediction (up to 3 years), *ShoreShop2.0* required participants to provide daily predictions of shoreline position for two periods: the short-term (2019–2023) and the medium-term (1951–1998). Only the model predictions for these two periods were used to evaluate model performance. Although future wave and sea-level projections were also provided to facilitate long-term forecasts extending to 2100, these were not included in the evaluation due to the lack of observed shoreline data.

Site description

Curl Curl Beach is a 1-km long embayed beach, situated within Sydney's Northern Beaches region in southeast Australia (Fig. 1a). The beach is characterized by fine to medium quartz sand with grain size $D_{50} \cong 0.3$ mm, estimated based on the adjacent Narrabeen Beach⁵⁴. The depth of closure is ~11 m with the slope of the active beach profile being 0.022, estimated with Hallermeier equation⁸³, whereas the inter-tidal beach face slope is about 0.07⁸⁵. The northern end of the beach is backed by an intermittent open and closed lagoon (ICOL). To minimize interference from the lagoon, shore-normal transects were defined starting 100 m south of the inlet. (Fig. 1b).

The deepwater wave climate in the Sydney region is characterized by moderate to high wave energy ($H_s \cong 1.6$ m and $T_p \cong 10$ s) with distinct seasonal and inter-annual variations. It is dominated by persistent, long-period swell waves from the SSE direction, as well as high-energy wind waves from the south (Fig. 1a)⁵⁴. As waves propagate toward the nearshore, processes such as shoaling and refraction alter their direction and magnitude (Fig. 1c). At the 10-m depth contour near Curl Curl Beach, the average H_s reduces to 1.2 m (Fig. 1d). As waves refract and shoal, the dominant wave directions shift to SE and ESE, with an average direction of 114° (Fig. 1e), ~85° relative to the shoreline. Curl Curl Beach is located in a micro-tidal environment with a mean spring tidal range about 1.3 m⁵⁴.

In addition to shoreline oscillations related to cross-shore beach processes, Curl Curl Beach also has a prominent rotational signal (Fig. 1g) as evidenced in many other embayed beaches in New South Wales, Australia, due to the trapped longshore sediment transport within individual embayment and the alongshore variability in cross-shore processes^{86,87}.

Data collection and preprocessing

The shoreline data used in ShoreShop2.0 were derived from freely available satellite images of Landsat 5, 7, 8, and 9 extracted with the open-source CoastSat toolbox³³. Shoreline position was defined as the distance from the landward end of a transect to the point of intersection with the shoreline. All shoreline positions were corrected for tidal effects to represent instantaneous positions at MSL³³. Validation against limited photogrammetry data⁷⁹ for Curl Curl Beach demonstrated high accuracy of SDS, with RMSE values below 7 m for transects 1–8 and up to 15 m for transect 9 near the headland (Supplementary Fig. S3). These metrics are generally better than the 8.9 m RMSE reported for the nearby Narrabeen Beach site when compared to ground-truth data²¹.

The hindcast (1940–2023) wave data used in ShoreShop2.0 was obtained by downscaling offshore directional wave spectra to nearshore areas. The offshore wave hindcast was from the ECMWF Reanalysis v5 (ERA5)⁷⁶. The hourly wave data were resampled to daily averages, using the mean value for significant wave height (H_s), peak wave period (T_p), and mean wave direction (Dir). The BinWaves approach⁷⁸ was applied to transform the offshore wave data to the nearshore. The H_s , T_p and Dir were extracted along each shore-normal transect at the 10 m bathymetry contour with a daily interval from 1940 through 2023. The projected wave data were made using a nested WAVEWATCH III wave model with surface wind projections from the Australian Community Climate and Earth System Simulator (ACCESS)⁷⁷ as inputs. Wave projections have been performed for Representative Concentration Pathway (RCP) with medium (RCP 4.5) and high (RCP 8.5) range carbon emission scenarios throughout 2006–2100. The offshore wave forecast was transformed to nearshore following the same approach applied to hindcast wave data⁷⁸. For each transect, to ensure the consistency between the hindcast and forecast wave data, the forecast wave climates were calibrated based on the joint distribution of H_s , T_p and Dir over the overlapping time period of 2006–2023 with the multivariate bias correction algorithm (MBCn)⁸⁸.

The observational annual sea-level data for 1950–2023 were obtained from the tidal gauge at Sydney, Fort Denison^{81,89}. Sea-level projections for 2019–2100 were sourced from the regional projections included in the Intergovernmental Panel on Climate Change (IPCC) 6th Assessment Report (AR6)⁸². Both observational and projection datasets were calibrated relative to the baseline average sea-level recorded between 1995 and 2014. A

uniform sea-level dataset was applied consistently across all transects. Tidal data were extracted from the FES2014 global tidal model⁸⁰ and resampled to daily mean values.

The representative bathymetry was obtained from the New South Wales Marine LiDAR Topo-Bathy dataset⁹⁰, which has a spatial resolution of 5 meters. The headland contour was extracted from the Topo-Bathy data at the 0-meter depth contour.

Evaluation methodology

As a graphical summary of model performance, Taylor diagrams⁴⁹ have been used to benchmark shoreline models¹⁶. Typically, a Taylor diagram evaluates and visualizes model performance using three metrics: the correlation coefficient (Corr), the standard deviation (STD), and the CRMSE. However, recognizing that these metrics do not account for bias in model predictions, we modified the loss function by replacing CRMSE with the root mean square error (RMSE). While CRMSE remains a component of the Taylor diagram for visualizing model performance, RMSE was specifically employed in the loss function to better capture prediction bias during model evaluation. To ensure comparability across transects with different shoreline variations, both RMSE and predicted STD were normalized by the STD of the observed (obs) shoreline data:

$$RMSE_{norm} = \frac{RMSE_{model}}{STD_{obs}}, STD_{norm} = \frac{STD_{model}}{STD_{obs}} \quad (1)$$

The loss function \mathcal{L} is defined to reflect the distance between the model predictions and the observed data ($RMSE_{norm}=0$, $Corr=1$, and $STD_{norm}=1$) in a Taylor diagram by incorporating multiple metrics:

$$\mathcal{L} = \sqrt{(0 - RMSE_{norm})^2 + (1 - Corr)^2 + (1 - STD_{norm})^2} \quad (2)$$

Due to the difference between CRMSE and RMSE, the loss \mathcal{L} is indicative of, but not identical to, the distance between the model points and the observation point in the Taylor diagram shown above in Fig. 5.

In addition to the metrics derived from the Taylor diagram, Mielke's modification λ ⁵⁰ used in ShoreShop1.0 was also included for comparison purposes²⁵:

$$\lambda = 1 - \frac{N^{-1} \sum_{i=1}^N (X_i - Y_i)^2}{\sigma_X^2 + \sigma_Y^2 + (\bar{X} - \bar{Y})^2} \quad (3)$$

where X and Y denote the target and predicted shoreline positions respectively, and N is the number of records in X and Y .

Model clustering

For model clustering, the shoreline predictions were standardized per transect by removing the mean and scaling to unit variance, which was then concatenated into a single time series of size N . The predictions from M models were stacked to construct an $M \times N$ array X . Pairwise Euclidean distances (D_{ij}) among M models in the N -dimensional space were calculated. Agglomerative-hierarchical clustering⁴² was then performed on D_{ij} to cluster M models into 6 classes, with the similarity criterion defined by Ward's variance minimization algorithm⁹¹.

Data availability

All the data that support the findings of this study are available at (https://github.com/ShoreShop/ShoreModel_Benchmark/tree/main) and archived at (<https://doi.org/10.5281/zenodo.15259391>). This repository includes input datasets as well as the shoreline predictions generated by each of the shoreline models.

Code availability

The code used to perform the model comparisons and generate the figures in this paper is publicly available at (https://github.com/ShoreShop/ShoreModel_Benchmark/tree/main) and archived at (<https://doi.org/10.5281/zenodo.15259391>). The repository also includes README files that briefly describe each modeling approach. Participants were encouraged to submit their model code alongside their blind shoreline predictions. Codes for ShoreFor, CoSMoS-COAST-CONV and ShoreOF-ML_JO are available under the /algorithms directory. However, as this was an open competition involving both open-source and proprietary models, codes were only made available at the discretion of the modelers providing submissions. Nevertheless, many of the model codes used for the ShoreShop2.0 submissions, described here, are publicly available in separate, model-specific repositories, as noted in the corresponding README files.

Received: 17 February 2025; Accepted: 7 July 2025;

Published online: 24 July 2025

References

- Nicholls, R. J. & Cazenave, A. Sea-level rise and its impact on coastal zones. *Science* **328**, 1517 (2010).
- Ranasinghe, R. impacts on open sandy coasts: a review. *Earth-Sci. Rev.* **160**, 320–332 (2016).
- Boak, E. H. & Turner, I. L. Shoreline definition and detection: a review. *J. Coast. Res.* 688–703 (2005).
- Nicholls, R. J., French, J. R. & van Maanen, B. Simulating decadal coastal morphodynamics. *Geomorphology* **256**, 1–2 (2016).
- Frazer, L. N., Anderson, T. R. & Fletcher, C. H. Modeling storms improves estimates of long-term shoreline change. *Geophys. Res. Lett.* <https://doi.org/10.1029/2009GL040061> (2009).
- Woodroffe, C. et al. A framework for modelling the risks of climate-change impacts on Australian coasts. In *Applied Studies in Climate Adaptation*. 181–189 (2014).
- Ranasinghe, R. et al. On the need for a new generation of coastal change models for the 21st century. *Scientific Reports* (2020).
- Hans, H. Genesis: a generalized shoreline change numerical model. *J. Coast. Res.* **5**, 1–27 (1989).
- Yates, M. L., Guza, R. T. & O'Reilly, W. C. Equilibrium shoreline response: observations and modeling. *J. Geophys. Res.: Oceans* <https://doi.org/10.1029/2009JC005359> (2009).
- Roelvink, D. J. & A. Reniers, *A Guide To Modeling Coastal Morphology*. **12**, World Scientific. (2011).
- Davidson, M. A., Splinter, K. D. & Turner, I. L. A simple equilibrium model for predicting shoreline change. *Coast. Eng.* **73**, 191–202 (2013).
- Vitousek, S. et al. A model integrating longshore and cross-shore processes for predicting long-term shoreline response to climate change. *J. Geophys. Res.: Earth Surf.* **122**, 782–806 (2017).
- Robinet, A. et al. A reduced-complexity shoreline change model combining longshore and cross-shore processes: The LX-Shore model. *Environ. Model. Softw.* **109**, 1–16 (2018).
- Jaramillo, C. et al. A shoreline evolution model for embayed beaches based on cross-shore, planform and rotation equilibrium models. *Coast. Eng.* **169**, 103983 (2021).
- McCarroll, R. J. et al. A rules-based shoreface translation and sediment budgeting tool for estimating coastal change: ShoreTrans. *Mar. Geol.* **435**, 106466 (2021).
- Gomez-de la Peña, E. et al. On the use of convolutional deep learning to predict shoreline change. *Earth Surf. Dynam.* **11**, 1145–1160 (2023).
- Calcraft, K. et al. Do LSTM memory states reflect the relationships in reduced-complexity sandy shoreline models. *Environ. Model. Softw.* **183**, 106236 (2025).
- Montaño, J. et al. A multiscale approach to shoreline prediction. *Geophys. Res. Lett.* **48**, e2020GL090587 (2021).
- Cowell, P. & B. Thom, *Morphodynamics of Coastal Evolution*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (1994).
- Hunt, E. et al. Shoreline modelling on timescales of days to decades. *Camb. Prisms: Coast. Futures* **1**, e16 (2023).
- Vos, K. et al. Benchmarking satellite-derived shoreline mapping algorithms. *Commun. Earth Environ.* **4**, 345 (2023).
- Collenteur, R. A. et al. Data-driven modelling of hydraulic-head time series: results and lessons learned from the 2022 groundwater time series modelling challenge. *Hydrol. Earth Syst. Sci.* **28**, 5193–5208 (2024).
- Blanco, B. et al. *Coastal Morphological Modelling For Decision-makers*. Bristol, UK: Environment Agency (2019).
- French, J. et al. Appropriate complexity for the prediction of coastal and estuarine geomorphic behaviour at decadal to centennial scales. *Geomorphology* **256**, 3–16 (2016).
- Montaño, J. et al. Blind testing of shoreline evolution models. *Sci. Rep.* **10**, 2137 (2020).
- Kollet, S. et al. The integrated hydrologic model intercomparison project, IH-MIP2: a second set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resour. Res.* **53**, 867–890 (2017).
- Splinter, K. D. & Coco, G. Challenges and opportunities in coastal shoreline prediction. *Front. Mar. Sci.* **8**, 788657 (2021).
- D'Anna, M. et al. Uncertainties in shoreline projections to 2100 at Truc Vert Beach (France): role of sea-level rise and equilibrium model assumptions. *J. Geophys. Res.: Earth Surf.* **126**, e2021JF006160 (2021).
- Toimil, A. et al. Visualising the uncertainty cascade in multi-ensemble probabilistic coastal erosion projections. *Front. Mar. Sci.* **8**, 683535 (2021).
- Ibaceta, R. et al. Improving multi-decadal coastal shoreline change predictions by including model parameter non-stationarity. *Front. Mar. Sci.* **9**, 1012041 (2022).
- Ibaceta, R. et al. Enhanced coastal shoreline modeling using an ensemble Kalman filter to include nonstationarity in future wave climates. *Geophys. Res. Lett.* **47**, e2020GL090724 (2020).
- Schepper, R. et al. Modelling cross-shore shoreline change on multiple timescales and their interactions. *J. Mar. Sci. Eng.* **9**, 582 (2021).
- Vos, K. et al. CoastSat: a Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery. *Environ. Model. Softw.* **122**, 104528 (2019).
- Vitousek, S. et al. A model integrating satellite-derived shoreline observations for predicting fine-scale shoreline response to waves and sea-level rise across large coastal regions. *J. Geophys. Res.: Earth Surf.* **128**, e2022JF006936 (2023).
- Mao, Y. et al. Determining the shoreline retreat rate of Australia using discrete and hybrid Bayesian networks. *J. Geophys. Res.: Earth Surf.* **126**, e2021JF006112 (2021).
- Vitousek, S. et al. Scalable, data-assimilated models predict large-scale shoreline response to waves and sea-level rise. *Sci. Rep.* **14**, 28029 (2024).
- Azorakos, G. et al. Satellite-derived equilibrium shoreline modelling at a high-energy meso-macrotidal beach. *Coast. Eng.* **191**, 104536 (2024).
- Alvarez-Cuesta, M., Toimil, A. & Losada, I. J. Which data assimilation method to use and when: unlocking the potential of observations in shoreline modelling. *Environ. Res. Lett.* **19**, 044023 (2024).
- Antolínez, J. A. A. et al. Predicting climate-driven coastlines with a simple and efficient multiscale model. *J. Geophys. Res.: Earth Surf.* **124**, 1596–1624 (2019).
- Roelvink, D. et al. Efficient modeling of complex sandy coastal evolution at monthly to century time scales. *Front. Mar. Sci.* **7**, 535 (2020).

41. Mao, Y. et al. *ShoreShop 2.0: Advancements in Shoreline Change Prediction Models*. <https://doi.org/10.5281/zenodo.15259391> (2025).
42. Müllner, D. et al. Modern hierarchical, agglomerative clustering algorithms. arXiv preprint <https://doi.org/10.48550/arXiv.1109.2378> (2011).
43. Miller, J. K. & Dean, R. G. A simple new shoreline change model. *Coast. Eng.* **51**, 531–556 (2004).
44. Coastal Engineering Research Center (CERC), Shore protection manual. US Army Corps of Engineers, Washington DC, Vol. I, 597, II, 603, 37–53 (1984).
45. Wright, L. D. & Short, A. D. Morphodynamic variability of surf zones and beaches: a synthesis. *Mar. Geol.* **56**, 93–118 (1984).
46. Lim, C., González, M. & Lee, J.-L. Estimating cross-shore and longshore sediment transport from shoreline observation data. *Appl. Ocean Res.* **153**, 104288 (2024).
47. Splinter, K. D. et al. A generalized equilibrium model for predicting daily to interannual shoreline response. *J. Geophys. Res.: Earth Surf.* **119**, 1936–1958 (2014).
48. Bell, B. et al. The ERA5 global reanalysis: Preliminary extension to 1950. *Q. J. R. Meteorological Soc.* **147**, 4186–4227 (2021).
49. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmospheres* **106**, 7183–7192 (2001).
50. Duveiller, G., Fasbender, D. & Meroni, M. Revisiting the concept of a symmetric index of agreement for continuous datasets. *Sci. Rep.* **6**, 19401 (2016).
51. Veličković, P. et al. Graph attention networks. In International Conference on Learning Representations (2018)
52. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
53. Doherty, Y. et al. A Python toolkit to monitor sandy shoreline change using high-resolution PlanetScope cubesats. *Environ. Model. Softw.* **157**, 105512 (2022).
54. Turner, I. L. et al. A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen, Australia. *Sci. Data* **3**, 160024 (2016).
55. Castelle, B. et al. 16 years of topographic surveys of rip-channelled high-energy meso-macrotidal sandy beach. *Sci. Data* **7**, 410 (2020).
56. McCarroll, R. J. et al. Coastal survey data for Perranporth beach and Start Bay in southwest England (2006–2021). *Scientific. Data* **10**, 258 (2023).
57. Banno, M. et al. Long-term observations of beach variability at Hasaki. *Jpn. J. Mar. Sci. Eng.* **8**, 871 (2020).
58. Holman, R. A. & Stanley, J. The history and technical capabilities of Argus. *Coast. Eng.* **54**, 477–491 (2007).
59. Garcia, D. Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.* **54**, 1167–1178 (2010).
60. Garcia, D. et al. smoothn, MATLAB Central File Exchange. <https://www.mathworks.com/matlabcentral/fileexchange/25634-smoothn>. Retrieved July 21, 2025.
61. Repina, O. et al. Evaluating five shoreline change models against 40 years of field survey data at an embayed sandy beach. *Coast. Eng.* **199**, 104738 (2025).
62. Chataigner, T. et al. Sensitivity of a one-line longshore shoreline change model to the mean wave direction. *Coast. Eng.* **172**, 104025 (2022).
63. Klotz, A. N. et al. Nearshore satellite-derived bathymetry from a single-pass satellite video: Improvements from adaptive correlation window size and modulation transfer function. *Remote Sens. Environ.* **315**, 114411 (2024).
64. Daly, C. et al. The new era of regional coastal bathymetry from space: a showcase for West Africa using optical Sentinel-2 imagery. *Remote Sens. Environ.* **278**, 113084 (2022).
65. Bergsma, E. W. J. et al. Coastal morphology from space: a showcase of monitoring the topography-bathymetry continuum. *Remote Sens. Environ.* **261**, 112469 (2021).
66. Ma, Y. et al. Satellite-derived bathymetry using the ICESat-2 lidar and Sentinel-2 imagery datasets. *Remote Sens. Environ.* **250**, 112047 (2020).
67. Laignel, B. et al. Observation of the coastal areas, estuaries and deltas from space. *Surv. Geophys.* **44**, 1309–1356 (2023).
68. de Michele, M. et al. Shallow bathymetry from multiple Sentinel 2 images via the joint estimation of wave celerity and wavelength. *Remote Sens.* **13**, 2149 (2021).
69. Bruun, P. Sea-level rise as a cause of shore erosion. *J. Waterw. Harb. Div.* **88**, 117–132 (1962).
70. Cooper, J. A. G. & Pilkey, O. H. Sea-level rise and shoreline retreat: time to abandon the Bruun Rule. *Glob. Planet. change* **43**, 157–171 (2004).
71. Davidson-Arnott, R. G. D. & Bauer, B. O. Controls on the geomorphic response of beach-dune systems to water level rise. *J. Gt. Lakes Res.* **47**, 1594–1612 (2021).
72. D’Anna, M. et al. Reinterpreting the Bruun rule in the context of equilibrium shoreline models. *J. Marine Sci. Eng.* <https://doi.org/10.3390/jmse9090974> (2021).
73. Calkoen, F. R. et al. Enabling coastal analytics at planetary scale. *Environ. Model. Softw.* **183**, 106257 (2025).
74. Robinet, A. et al. Controls of local geology and cross-shore/longshore processes on embayed beach shoreline variability. *Mar. Geol.* **422**, 106118 (2020).
75. Harley, M. D. et al. Single extreme storm sequence can offset decades of shoreline retreat projected to result from sea-level rise. *Commun. Earth Environ.* **3**, 112 (2022).
76. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorological Soc.* **146**, 1999–2049 (2020).
77. Lewis, S. et al. ACCESS1-0 climate model output prepared for the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical experiment, r21p1 ensemble., ARC Centre of Excellence for Climate System Science. <https://doi.org/10.1594/WDCC/CMIP5.CSA0hi> (2013).
78. Cagigal, L. et al. BinWaves: An additive hybrid method to downscale directional wave spectra to nearshore areas. *Ocean Model.* **189**, 102346 (2024).
79. Harrison, A. et al. *NSW beach photogrammetry: A new online database and toolbox*, in *Australasian Coasts & Ports 2017: Working with Nature: Working with Nature. Engineers Australia, PIANC Australia and Institute of Professional Engineers*. 565–571 (2017).
80. Carrere, L. et al. *FES2014, a new tidal model-Validation results and perspectives for improvements*, presentation to ESA Living Planet Conference. Prague. (2016).
81. Permanent Service for Mean Sea Level (PSMSL), *Tide Gauge Data*. <http://www.psmsl.org/data/obtaining/> (2024).
82. Garner, G. G. et al. *IPCC AR6 Sea Level Projections*. <https://doi.org/10.5281/zenodo.6382554>. (2021).
83. Hallermeier, R. J. A profile zonation for seasonal sand beaches from wave climate. *Coast. Eng.* **4**, 253–277 (1980).
84. Vos, K. et al. Beach Slopes From Satellite-Derived Shorelines. *Geophys. Res. Lett.* **47**, e2020GL088365 (2020).
85. Vos, K. et al. Beach-face slope dataset for Australia. *Earth Syst. Sci. Data* **14**, 1345–1357 (2022).
86. Harley, M. D. et al. A reevaluation of coastal embayment rotation: The dominance of cross-shore versus alongshore sediment transport processes, Collaroy-Narrabeen Beach, southeast Australia. *J. Geophys. Res.: Earth Surf.* **116**. <https://doi.org/10.1029/2011JF001989> (2011).
87. Harley, M. D., Turner, I. L. & Short, A. D. New insights into embayed beach rotation: the importance of wave exposure and cross-shore processes. *J. Geophys. Res.: Earth Surf.* **120**, 1470–1484 (2015).

88. Cannon, A. J. Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim. Dyn.* **50**, 31–49 (2018).
89. Holgate, S. J. et al. New data systems and products at the permanent service for mean sea level. *J. Coast. Res.* **29**, 493–504 (2013).
90. State Government of NSW and NSW Department of Climate Change, Energy, the Environment and Water. NSW Marine LiDAR Topo-Bathy 2018 Geotif, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/45089194-912d-4ecf-8200-969e0796afee>], date accessed 2025-07-21 (2019).
91. Ward Jr, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).

Acknowledgements

Y.M. and K.S. are supported by ARC Future Fellowship FT220100009 and the US Geological Survey Research Co-op (G21AC10672). G.C., K.B., E.G.P., and M.G. are funded by the “Our Changing Coast Project” (MBIE-NZ RTVU2206). S.V. and E.S. acknowledges support from the USGS Coastal and Marine Hazards and Resources Program. L.F.P., C.J.C., and C.L. acknowledge the support of the ThinkInAzul program, supported by MCIN/Ministerio de Ciencia e Innovación with funding from the European Union NextGeneration EU (PRTR-C17.11) and by Comunidad de Cantabria. C.J.C. has been supported by a Margarita Salas post-doctoral fellowship funded by the European Union-NextGenerationEU, Ministry of Universities, and the Recovery and Resilience Facility, through a call from the University of Cantabria. L.C. acknowledges support from the Government of Cantabria, and the European Union NextGenerationEU/PRTR under projects Perfect-Storm (2023/TCN/003) and CE4Wind (CPP2022-010118). A.R., C.B., G.A., B.C., and D.I. acknowledge the support of BRGM and the SHORMOSAT project, funded by Agence Nationale de la Recherche (ANR-21-CE01-0015). M.D. is funded by the European Union HORIZON-MSCA-2022-PF-01 under the project PhySeaCS (101107336). D.P. acknowledges the Portuguese Fundação para a Ciência e Tecnologia (FCT) support, under the 2022.13776.BDANA PhD research fellowship. I.D.S acknowledges support from KOSTARISK joint laboratory and Programa de Movilidad del Personal Investigador Doctor del Gobierno Vasco. M.B. acknowledges support from JSPS KAKENHI (24K00996) and the Advanced Studies of Climate Change Projection (JPMXD0722678534) supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions

Y.M., K.D.S., and G.C. designed the study, and invited all participants. Y.M. prepared the testbed with wave data provided by L.C. and A.N.D. Y.M. performed the accuracy assessment, prepared the figures, and drafted the manuscript with major contribution from K.D.S., G.C., and S.V. Each modeler ran their model at the testbed: *COCOONED* models by J.A.A.A., *ShoreForAndRotation_GA* by G.A. and B.C., *Multi-layer-LSTM_MB* and *EqShoreB_MB* by M.B., *LX-Shore_SWAN_CB* by C.B., *iTransformer-KC* by K.C., *SegShoreFor_XC* by X.C., *COCOONED_LFP*, *ConvLSTM2D_LFP*, *IH_MOOSE_LFP*, and *Y09_LFP* by L.F.P. and C.J.C., *ShorelineEvol_IdS_OR*

by I.D.S and O.R., *ShoreForLogSpiral_BD* by B.D., *ShorelineS-DA_AE* by A.E., *SARIMAX_AG*, *SPADS_AG*, and *XGBoost_AG* by A.G., *CNN-LSTM_EGP* by E.G.P., *Catboost_MI_HS* by M.I., *SLRM_LIM* by C.L., *GAT-LSTM_YM* and *CoSMoS-COAST-Nonstat_YM* by Y.M., *ShorEOF-ML_JO* by J.O., *RandomForestReg_DP* by D.P., *LX-Shore_Larson_AR* by A.R. and D.I., *ShorelineS_DR* by D.R., *Bayesian-Hierarchical_JS* by J.S., *ShEPReMo_ES* by E.S., *ShoreForCaCeHb_KS* by K.D.S., *CoSMoS-COAST-CONV_SV* and *CoSMoS-COAST_SV* by S.V., and *linear_regression_KW*, *SARIMAX_KW* and *SVARX_KW* by K.W. K.B., M.D., M.D.H., I.M., and all other authors contributed to discussing the results and writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-025-02550-4>.

Correspondence and requests for materials should be addressed to Yongjing Mao.

Peer review information *Communications Earth & Environment* thanks Andres Payo, Rodrigo Mikosz Gonçalves and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Olusegun Dada and Alireza Bahadori. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Water Research Laboratory, School of Civil and Environmental Engineering, UNSW Sydney, Sydney, Australia. ²School of Environment, Faculty of Science, University of Auckland, Auckland, New Zealand. ³U.S. Geological Survey, Santa Cruz, CA, USA. ⁴Department of Hydraulic Engineering, Delft University of Technology, Delft, Netherlands. ⁵Université de Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, Allée Geoffroy Saint-Hilaire, Pessac, Nouvelle Aquitaine, France. ⁶Port and Airport Research Institute, Yokosuka, Japan. ⁷BRGM (French Geological Survey), Parc technologique Europarc, Pessac, France. ⁸Geomatics and Ocean Engineering Group, Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Cantabria, Spain. ⁹Department of Civil and Earth Resources Engineering, Kyoto University, Kyoto, Japan. ¹⁰Universitat Politècnica de Catalunya, Physics Dep., Carrer Jordi Girona, Barcelona, Spain. ¹¹IHCantabria – Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Cantabria, Spain. ¹²AZTI Marine Research, Basque Research and Technology Alliance (BRTA), Herrera Kaia. Portualdea z/g, Pasaia, Gipuzkoa, Spain. ¹³Port Said University, Port Said, Egypt. ¹⁴IHE Delft Institute for Water Education, Delft, Netherlands. ¹⁵Computer Engineering Department, Cairo University, Cairo, Egypt. ¹⁶BRGM (French Geological Survey), Orléans, France. ¹⁷CSIRO

Environment, Aspendale, VIC, Australia. ¹⁸CoLAB+ATLANTIC LVT, Museu das Comunicações, Rua do Instituto Industrial, Lisbon, Portugal. ¹⁹Departamento de Geologia & Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Lisbon, Portugal. ²⁰Coastal and Marine Research Centre, Griffith University, Southport, Gold Coast, QLD, Australia. ²¹Deltares, Delft, Netherlands. ²²ARC Training Centre in Data Analytics for Resources and Environments (DARE), University of Sydney, Sydney, Australia. ²³Cherokee Nation System Solutions, Contractor to the U.S. Geological Survey, St. Petersburg Coastal and Marine Science Center, St. Petersburg, FL, USA. ✉e-mail: Yongjing.mao@unsw.edu.au