Hidden in the Spectrum: Geographic Transfer of an Offshore Wave Index for Coastal Water-Level Oscillations

Abstract

Bulk wave height and peak period omit how energy is distributed within an offshore spectrum. We tested whether a low-frequency-weighted spectral index transferred more effectively than a conventional bulk wave-height–period scale from five Pacific Northwest buoy–gauge pairs to six California pairs. Models fitted to 49,786 observations from 2017–2022 were checked in 2023 and evaluated in 2024–2025. California data supplied 2016 response and predictor means for local normalization but not coefficient estimation. Across 24,878 external observations, the spectral model reduced normalized root-mean-square error by 11.7% (95% paired block-bootstrap interval 10.4–13.0%) for preliminary one-minute water levels, with favorable estimates at all six sites. The reduction was 4.6% (3.5–5.6%) for verified six-minute levels, with weak absolute skill ($R^2 = -0.066$). Against wave height alone, however, the index showed no clear preliminary advantage (3.0%, −0.5 to 6.6%) and was slightly worse for the verified response (−1.6%, −3.4 to 0.2%). Thus, the fixed relationship transferred relative to the prespecified height–period comparator after local normalization, but its advantage was product- and comparator-dependent. The index is a candidate covariate for fixed-point coastal oscillations requiring local reference data, not a calibration-free or operational predictor.

Keywords: coastal observations, external validation, ocean wave spectra, spectral moments, water-level oscillations

1 Introduction

Significant wave height and a representative period summarize an offshore sea state in quantities that are easy to observe, communicate, and use in empirical models. Their compactness is also a limitation. Sea states with similar height and peak period can differ in bandwidth, spectral shape, and the fraction of energy toward lower frequencies. Those differences matter when wave groups generate slowly varying radiation stresses and when nonlinear interactions transfer energy to motions longer than the incident-wave period (Bertin et al. 2018; Herbers et al. 1995; Longuet-Higgins and Stewart 1962). The resulting low-frequency field is modified as it crosses the shelf and nearshore zone, where shoaling, reflection, breaking, and dissipation depend on local setting (Smit et al. 2018; van Dongeren et al. 2007). A bulk height–period pair cannot retain all of the spectral information relevant to that sequence.

The observing problem spans two different environments. A buoy measures the incident wave field offshore, whereas a coastal gauge records the response after propagation and transformation through shelf, nearshore, harbor, or embayment geometry. A statistical association between those records need not identify a single generation mechanism. It can nevertheless show whether one description of the offshore sea state carries more transferable information than another. This is a stricter question than fitting a relationship at one instrument pair: a common equation must preserve its ordering despite changes in exposure, local response, and wave climate. Geographic evaluation is consequently central to deciding whether a spectral index describes a repeatable feature of the coupled observations or only the conditions of its development network.

Negative spectral moments provide one way to emphasize the low-frequency part of an offshore spectrum without reconstructing the complete nearshore transformation. For energy density $E(f)$ at frequency f, the moment $m_n = \int f^n E(f)\, df$ gives progressively greater weight to lower frequencies as n becomes more negative. Such a moment is not a direct measure of infragravity energy at the coast, nor does it specify how an individual shoreline will respond. It is instead a compact descriptor of an incident sea state whose association with coastal variability can be tested. A useful relationship at one coast may not transfer elsewhere because the incident climate, shelf, harbor or embayment geometry, sensor setting, and processing of the coastal observation all change alongshore.

This weighting also creates a practical validation requirement. Low-frequency bins exert increasing influence on negative moments, so archive conventions, frequency spacing, and the lowest valid frequencies can affect the predictor. A credible comparison must therefore integrate each observed frequency grid explicitly, apply the same offshore rows to the competing predictors, and test the chosen moment without selecting a new exponent from the evaluation sites. The question is not whether one can optimize a flexible spectral summary for California, but whether a relationship fixed from an antecedent hypothesis retains an advantage there.

The present study follows a spectral hypothesis proposed by Li et al. (2023). Using observations associated with unusually large runup on the US Pacific Northwest coast, that work identified the square root of the negative fourth spectral moment, $\sqrt{m_{-4}}$, as an offshore indicator and examined site-specific relationships at five buoy–gauge pairs during 2016–2018. It did not estimate or test the pooled relationships evaluated here. The development phase of the present study used 2017–2022 observations from the same five pairs to fit matched one-predictor regressions for $\sqrt{m_{-4}}$ and a conventional bulk wave-height–wavelength scale. The first two fitting years overlap the observation period in Li et al. (2023); 2019–2022 extend the record, and the pooled coefficients were estimated as part of the present study. A geographically separated evaluation is needed because temporal performance within the development network does not establish that regression coefficients will remain useful at a new coast (Roberts et al. 2017).

The coastal response also requires careful definition. Preliminary CO-OPS water levels are available at one-minute cadence and preserve short-period fluctuations, but they have not undergone all processing associated with the verified product. Verified water levels are quality controlled but reported at six-minute cadence. A response reconstructed from the two products can differ because of both product status and temporal processing. The sampling interval controls which fluctuations remain observable, while the moving-mean and RMS windows determine how those fluctuations are summarized. Agreement in model ranking across the two constructions would support a broader measurement-domain claim; attenuation would show that an apparent spectral advantage depends on how coastal variability is observed. Neither product measures maximum beach runup. Both describe filtered oscillation magnitude at a fixed tide gauge, which is distinct from moving-shoreline swash and from an individual hazardous incident.

We asked two related questions. First, do fixed relationships fitted at five Pacific Northwest pairs retain an advantage for $\sqrt{m_{-4}}$ over a matched bulk comparator at six California pairs in 2024–2025? Second, how much of that advantage persists when the response is reconstructed from verified six-minute rather than preliminary one-minute water levels? California observations did not enter coefficient estimation or predictor choice, but each site supplied a 2016 mean for scaling its predictors and responses. The design is therefore a test of fixed-coefficient geographic transfer after local reference normalization, not a test of deployment without local coastal observations. We also used the longer 2017–2025 California record, dependence-aware resampling, processing sensitivities, and a targeted product-harmonization diagnostic to establish where the result is stable and where its interpretation must narrow.

2 Methods

2.1 Study design and temporal roles

All analyses used the winter-to-early-spring interval from 1 January through 15 April, when energetic wave conditions and spectral coverage were most consistent across the network. The study separated model development, internal temporal checking, local reference scaling, and geographic evaluation. Five Pacific Northwest buoy–gauge pairs supplied model-fitting observations from 2017–2022. Their 2023 observations were held out from fitting and used only to check temporal ordering of the two models. The principal geographic test used qualifying observations from six California pairs in 2024–2025. A separate 2016 season at every site supplied arithmetic means for local, dimensionless normalization. The broader 2017–2025 California score was calculated after the principal evaluation as a temporal sensitivity rather than as part of the original held-out test.

The station rules, predictor definitions, response filters, normalization, regression forms, and 2024–2025 evaluation were recorded before the complete California outcomes were acquired and inspected. California data did not determine the spectral moment, comparator, regression form, practical reference, or model coefficients. Product harmonization, the height-only comparison, and the normalization-window analysis were post hoc sensitivities used to examine the boundaries of the primary comparison; none changed its specification or estimates.

2.2 Station pairs and public observations

National Data Buoy Center (NDBC) historical archives supplied nondirectional energy-density spectra (National Data Buoy Center n.d.a). Spectral field definitions and station metadata followed the corresponding NDBC product documentation (National Data Buoy Center n.d.b). The NOAA Center for Operational Oceanographic Products and Services (CO-OPS) Data API supplied preliminary one-minute and verified six-minute water levels and gauge metadata (NOAA Center for Operational Oceanographic Products and Services n.d.a). Product status and water-level definitions followed the CO-OPS documentation (NOAA Center for Operational Oceanographic Products and Services n.d.b). Timestamps were converted to Coordinated Universal Time before matching. When an NDBC archive contained more than one spectrum in a nominal hour, records were assigned to the nearest hour and the closest observation was retained; the earlier observation broke an exact tie. The original observation time was retained for subsequent matching and auditing.

The development network comprised La Push and Westport, Washington; South Beach and Charleston, Oregon; and Crescent City, California (Fig. 1; Table 1). These are the five geographic pairs connected to the original hypothesis. The 2017–2018 portion of the present fitting record overlaps the period considered by Li et al. (2023), while 2019–2022 extends it; the pooled regressions and all later evaluations belong to the present study. The 2017–2022 fitting set contained 49,786 paired observations. Four sites supplied 2023 internal-check observations; Charleston lacked a qualifying 2023 spectrum record under the common rules, leaving 7,779 observations for that check.

California candidates were screened without reference to model error. Eligibility considered station identity, buoy–gauge distance, spectral coverage, paired-hour completeness, verified-water-level availability, and representation in the late test period. A candidate required a 2016 reference season with at least 70% of nominal paired hours after quality control, at least five evaluation seasons meeting the same coverage threshold, and a qualifying 2024 or 2025 season. Candidate buoy–gauge separation

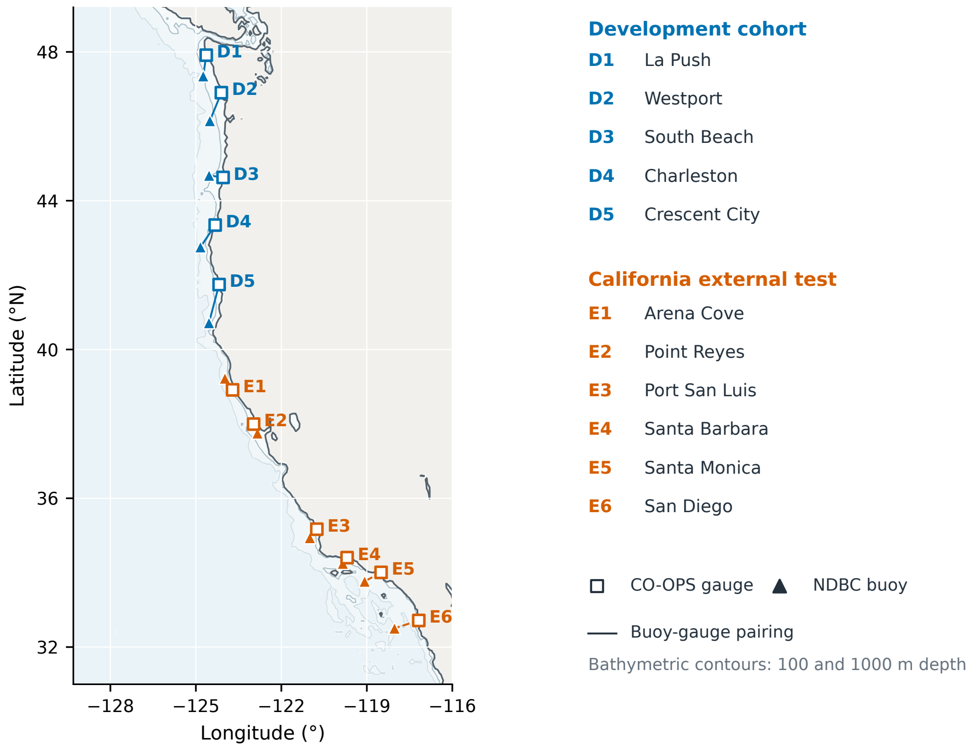

Fig. 1 NDBC buoy–CO-OPS gauge pairs used for model development (D1–D5) and the California geographic evaluation (E1–E6). Squares mark gauges, triangles mark buoys, and connecting lines identify statistical pairings rather than propagation paths. Pale relief and the 100- and 1000-m isobaths are from ETOPO 2022 (NOAA National Centers for Environmental Information 2022)

was limited to 100 km, and two primary gauges could not share one buoy. When eligible gauges shared a buoy, the gauge with more qualifying seasonal coverage was selected, with distance used to break a tie. Six pairs met the rules: Arena Cove–46014, Point Reyes–46026, Port San Luis–46011, Santa Barbara–46053, Santa Monica–46025, and San Diego–46086. Point Reyes and Santa Barbara contributed only 2024 to the principal score; the other pairs contributed both test years.

Table 1 Study cohorts, station pairs, and retained observations. Development counts are model-fitting observations from 2017–2022 followed, where available, by 2023 internal-check observations. External counts are the 2024–2025 evaluation rows.

Cohort	Gauge site (CO-OPS)	NDBC buoy	Distance (km)	Retained observations
Development	La Push (9442396)	46041	63.0	12,687 fit; 2,234 check
Development	Westport (9441102)	46029	89.6	10,561 fit; 2,515 check
Development	South Beach (9435380)	46050	39.2	11,065 fit; 2,379 check
Development	Charleston (9432780)	46015	78.0	6,665 fit
Development	Crescent City (9419750)	46022	118.3	8,808 fit; 651 check
External	Arena Cove (9416841)	46014	41.6	5,043 test (2024, 2025)
External	Point Reyes (9415020)	46026	29.6	2,302 test (2024)
External	Port San Luis (9412110)	46011	34.1	4,948 test (2024, 2025)
External	Santa Barbara (9411340)	46053	22.3	2,532 test (2024)
External	Santa Monica (9410840)	46025	59.7	5,016 test (2024, 2025)
External	San Diego (9410170)	46086	83.2	5,037 test (2024, 2025)

2.3 Spectral predictor and bulk comparator

Archive sentinels, negative energy densities, and nonpositive frequencies were removed. A spectrum was retained only when at least 95% of its bins were valid, frequency increased strictly, integrated energy was positive, calculated significant wave height was between 0.05 and 20 m, and peak period was between 2 and 40 s. The spectra were not interpolated to a common grid. Instead, each moment was integrated by the trapezoidal rule on the observed frequency coordinates,

$$m_n = \int f^n E(f)\, df, \tag{1}$$

where $E(f)$ has units of $\mathrm{m}^2\ \mathrm{Hz}^{-1}$. The primary moment used every valid positive-frequency bin supplied by the historical product. Separate calculations imposed lower limits of 0.04 and 0.05 Hz to test sensitivity to the lowest bins.

The spectral predictor was $X_s = \sqrt{m_{-4}}$. The bulk comparator was $X_b = \sqrt{H_s L_0}$, where $H_s = 4\sqrt{m_0}$, T_p was the reciprocal of the maximum-energy frequency, and $L_0 = gT_p^2/(2\pi)$ is deep-water wavelength. This height–wavelength combination is a conventional scale in empirical coastal-response parameterizations (Stockdon et al. 2006); here it served only as a physically interpretable bulk summary. It was not used as a runup prediction. Both predictors came from the same retained spectrum, were paired to the same response rows, and entered identical one-predictor linear forms. The comparison consequently changes the offshore scalar but not model flexibility or data availability.

2.4 Coastal water-level responses

The primary response was constructed from preliminary one-minute water levels. Values outside -10 to 15 m, isolated changes greater than 1 m min^{-1}, and constant runs of at least 120 min were excluded. A centered 61-min moving mean was subtracted from the quality-controlled series to remove variations longer than the target band. The RMS of the residual was then calculated in a centered 31-sample window. Both operations required at least 90% valid observations in the relevant window. The result, denoted η_{RMS}, is an amplitude measure of sub-hourly variability around a local moving mean.

Verified six-minute observations were processed independently because simply subsampling the preliminary response would not reproduce the verified product. An 11-point centered mean, spanning 60 min from the first to the last observation, was removed from the verified series. RMS was then calculated over five centered samples. The preliminary and verified responses therefore target related, but not identical, filtered variability. Model ranking and relative error were calculated within each response. Absolute RMSE values were not compared across the two products as if they measured an identical variable.

Each hourly offshore record was paired to the nearest eligible coastal response within half of the effective spectral sampling interval for that site–year. No temporal interpolation was used. A paired row was retained only when both predictors and the relevant response passed quality control. The preliminary and verified models were scored on common 2024–2025 hours for the principal product comparison.

2.5 Local normalization and model development

Response and predictor magnitudes differ substantially among sites because of wave climate, gauge setting, and local scaling. Each variable was divided by its arithmetic mean over the retained 2016 season at the same site. Separate 2016 means were calculated for the preliminary and verified responses, while each offshore predictor had its own site mean. This transformation produced dimensionless values and allowed one relationship to be carried among pairs without estimating site-specific regression coefficients.

The local scaling is an important boundary on the design. California observations contributed to the 2016 response and predictor means, so the evaluation was not calibration free. They did not enter coefficient fitting, predictor or moment selection, model selection, station choice based on performance, or definition of the 5% practical reference. The analysis asks whether coefficients estimated in the Pacific Northwest preserve their relative ranking after local reference scales are known at a new gauge.

Ordinary least squares was used to fit one model for each predictor to the 49,786 development observations. The fitted spectral relation was

$$\widehat{\eta}_{\text{norm}} = 0.34456 + 0.65749 X_{s,\text{norm}}, \tag{2}$$

and the bulk relation was

$$\widehat{\eta}_{\text{norm}} = 0.00305 + 0.98859 X_{b,\text{norm}}. \tag{3}$$

The choice of m_{-4} followed Li et al. (2023); the California outcomes were not searched for a better moment. After fitting, both intercepts and slopes were carried unchanged into the 2023 internal check and the California evaluations.

Two post hoc analyses examined specific limitations without altering the primary comparison. First, a height-only model was fitted to the same 49,786 development rows and applied to the same California evaluation rows. Because $H_s = 4\sqrt{m_0}$ and normalization removes the constant factor, normalized significant wave height is equivalent to normalized $\sqrt{m_0}$. The fitted relation was $\widehat{\eta}_{\text{norm}} = 0.23538 + 0.70421 H_{s,\text{norm}}$. Comparing $\sqrt{m_{-4}}$ directly with this model tested whether its advantage over $\sqrt{H_s L_0}$ also extended beyond wave height alone.

Second, sensitivity to the local reference period was evaluated with 7-, 14-, 30-, and 60-day windows fully contained within 1 January–15 April 2016 and beginning on each eligible day. A window was retained only when every California site had at least 70% of nominal hours jointly complete for both predictors and the response under evaluation. Predictor and response means were recalculated from those common hours, whereas the development coefficients and the 24,878 evaluation rows remained unchanged. Preliminary and verified response means were evaluated separately.

2.6 Performance measures and uncertainty

Performance was summarized by RMSE, mean absolute error (MAE), mean signed error (bias), Pearson correlation, and R^2. For a held-out response y, $R^2 = 1 - \text{SSE}/\text{SST}$, with SST calculated around the held-out mean. A negative value means that predictions were worse in squared error than using that mean. The principal contrast was relative RMSE reduction,

$$\Delta_{\text{RMSE}} = 100 \frac{\text{RMSE}_b - \text{RMSE}_s}{\text{RMSE}_b}, \tag{4}$$

where positive values favor the spectral model. A 5% reduction was recorded as a practical reference before external scoring. It is not a statistical significance threshold.

The pooled metrics were observation weighted: every retained site–hour contributed once. Uncertainty was estimated from 2,000 paired bootstrap replicates. Seven-day blocks were sampled within each site–year stratum, and the same draw was applied to errors from both models. This kept storm sequences together, preserved the observed number of rows contributed by each site–year, and compared models on identical resamples. The 2.5th and 97.5th percentiles formed the reported 95% intervals. These intervals describe variation among temporal blocks conditional on the observed stations; they are not population intervals for all coastlines. To examine shared regional forcing, a post hoc sensitivity resampled calendar-aligned seven-day blocks synchronously across California sites within each year.

2.7 Product harmonization and processing sensitivities

A post hoc diagnostic placed the preliminary observations on the verified product's temporal basis. The quality-controlled preliminary one-minute series was sampled at timestamps on the verified six-minute grid. The verified temporal construction was then applied: an 11-point centered mean required at least 10 valid samples, and the centered five-sample RMS required all five values. A separate 2016 site mean normalized this harmonized preliminary response. Existing offshore matches, normalized predictors, and model coefficients were unchanged.

The original preliminary, harmonized preliminary, and verified responses were evaluated on common 2024–2025 rows. The diagnostic changes cadence and filtering together and does not reproduce NOAA's verification workflow. It can therefore locate the harmonized result between the original products, but cannot attribute the separation to individual processing steps or product status.

Other analyses varied one processing choice at a time while retaining the California station set and matched model forms. They tested moment orders from 0 through -6, lower-frequency cutoffs of none, 0.04, and 0.05 Hz, preliminary RMS windows of 21, 31, and 45 min, and fixed predictor lags of 0–3 h. These calculations describe robustness and possible heterogeneity. They did not replace the prespecified m_{-4}, zero-lag, 31-min result.

Analyses were conducted in Python 3.12 using recorded package versions.

3 Results

3.1 Development and internal temporal check

The five development pairs contributed 49,786 observations from 2017–2022, with site totals ranging from 6,665 at Charleston to 12,687 at La Push (Table 1). The later 2023 check contained 7,779 observations from four pairs. Without refitting, spectral-model RMSE was 0.358 compared with 0.382 for the bulk model, a 6.3% reduction. MAE was 0.260 for the spectral relation and 0.279 for the bulk relation, while R^2 was 0.563 and 0.503, respectively. The temporal ordering supported carrying both equations into the geographic evaluation; it did not establish geographic transfer because the gauges remained within the development network.

The six California pairs retained 24,878 of 25,344 nominal observations across ten qualifying 2024–2025 site-seasons, or 98.2% completeness. Four pairs contributed both test years; Point Reyes and Santa Barbara contributed 2024 only. Site contributions ranged from 2,302 observations at Point Reyes to 5,043 at Arena Cove. The retained fractions were high enough that the principal score represented nearly complete qualifying seasons rather than isolated storm windows. Every scored row contained the same offshore spectrum and coastal response for the two competing models.

3.2 Geographic transfer for the preliminary response

The spectral relation retained a clear relative advantage for the preliminary one-minute response. Pooled normalized RMSE was 0.346 for $\sqrt{m_{-4}}$ and 0.391 for the bulk

a. *Error and relative improvement*

Table 2 Pooled performance in the Pacific Northwest internal temporal check and at the six California external sites after local 2016 normalization. Intervals use 2,000 paired seven-day site–year block-bootstrap replicates.

Response	Years	n	Spectral RMSE	Bulk RMSE	RMSE reduction (95% interval)
Pacific Northwest internal check	2023	7,779	0.358	0.382	6.3%
Preliminary 1-min	2024–2025	24,878	0.346	0.391	11.7% (10.4–13.0%)
Verified 6-min	2024–2025	24,878	0.445	0.466	4.6% (3.5–5.6%)
Preliminary 1-min	2017–2025	108,090	0.449	0.485	7.4% (5.7–9.7%)
Verified 6-min	2017–2025	108,090	0.584	0.599	2.4% (1.6–3.5%)

b. *Absolute error, association, and spectral-model skill at the California sites*

Response	Years	Spectral MAE	Bulk MAE	Spectral r	Bulk r	Spectral R^2
Preliminary 1-min	2024–2025	0.246	0.285	0.616	0.470	0.378
Verified 6-min	2024–2025	0.340	0.363	0.256	0.176	-0.066
Preliminary 1-min	2017–2025	0.262	0.298	0.534	0.416	0.283
Verified 6-min	2017–2025	0.352	0.372	0.270	0.203	0.015

comparator, giving an 11.7% reduction (95% paired block-bootstrap interval 10.4–13.0%; Table 2). MAE decreased from 0.285 to 0.246. Correlation increased from 0.470 to 0.616, and R^2 increased from 0.201 to 0.378. Spectral-model bias was 0.016 compared with 0.037 for the bulk model. Thus, the spectral advantage appeared in error and association measures within this response, not only in the percentage contrast.

All six site point estimates favored the spectral index (Fig. 2). Relative reductions were 15.6% at Arena Cove, 17.4% at Point Reyes, 15.8% at Port San Luis, 7.2% at Santa Barbara, 13.8% at Santa Monica, and 4.5% at San Diego. The paired bootstrap interval was positive at every site, including 1.2–7.3% at San Diego, where the point estimate was smallest. Intervals were wider at the two sites with only one test season, but the pooled ordering was not created by a single favorable station. The magnitude nevertheless varied substantially: the Point Reyes point estimate was almost four times the San Diego estimate. The contrast therefore supports transfer of model ranking more strongly than a coast-invariant effect size.

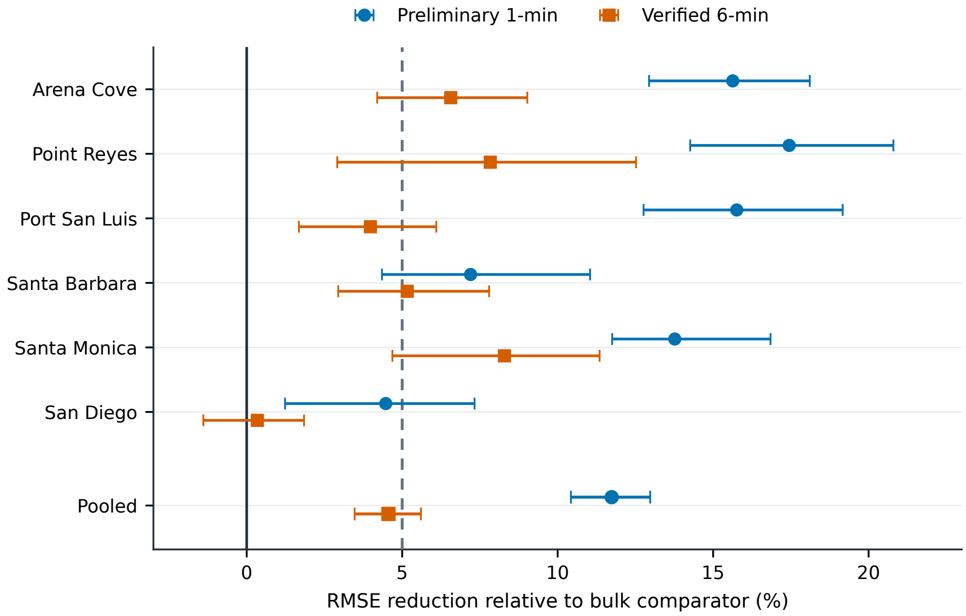

Fig. 2 Site-specific and pooled RMSE reductions for the fixed development spectral model relative to the matched bulk comparator in the 2024–2025 California evaluation after local 2016 normalization. Error bars are 95% intervals from 2,000 paired seven-day site–year block-bootstrap replicates. Positive values favor the spectral index; the dashed line at 5% is a practical reference, not a significance threshold

3.3 Attenuation with verified six-minute water levels

The verified response preserved the pooled ordering but reduced its magnitude. On the same 24,878 hours, spectral and bulk RMSE were 0.445 and 0.466, a 4.6% reduction (3.5–5.6%). The point estimate fell below the 5% practical reference. Spectral MAE was 0.340 rather than 0.363, bias was 0.050 rather than 0.071, and correlation was 0.256 rather than 0.176. Absolute performance remained weak: spectral R^2 was -0.066 and bulk R^2 was -0.170. The spectral model was therefore better than the bulk model on the verified response but worse than the held-out response mean in absolute squared error.

Site-level verified contrasts were also smaller. San Diego showed essentially no difference at 0.3% (-1.4 to 1.8%), while Port San Luis was 4.0%, Santa Barbara 5.2%, Arena Cove 6.6%, Point Reyes 7.8%, and Santa Monica 8.3%. Five point estimates remained positive, but only three exceeded the practical reference. Apart from San Diego, the paired intervals remained above zero; none, however, established a common effect size across sites. Because preliminary and verified responses were constructed differently, their absolute RMSE values are not directly comparable; the defensible comparison is the within-product ranking of models.

3.4 Longer record and dependence sensitivity

The expanded 2017–2025 California record contained 108,090 observations. The spectral reduction for the preliminary response was 7.4% (5.7–9.7%), with RMSE values of 0.449 and 0.485. All six site-level preliminary contrasts remained positive, ranging from 1.2% at San Diego to 14.9% at Point Reyes. For the verified response, the pooled reduction was 2.4% (1.6–3.5%), with RMSE values of 0.584 and 0.599. Five site contrasts remained positive, while San Diego favored the bulk comparator by 0.8%. The longer record therefore retained the pooled ordering seen in 2024–2025 but showed that its magnitude, and at one site its direction for the verified product, varied with the evaluated years. The principal two-year estimates should not be interpreted as fixed properties of the six coasts.

Synchronous resampling of calendar-aligned storm weeks widened the primary intervals only modestly. The preliminary interval changed from 10.4–13.0% to 10.3–13.5%, and the verified interval changed from 3.5–5.6% to 3.0–6.1%. Shared storm timing among neighboring sites therefore did not reverse either pooled contrast. This sensitivity addresses temporal covariance in the observed network, but it does not turn six regional stations into a representative sample of all coasts.

3.5 Product-harmonization diagnostic

All six sites retained valid 2016 reference means for the harmonized preliminary response, and 24,817 of the 24,878 principal rows were common to the three response constructions (99.75%). Spectral and bulk errors were evaluated on identical rows.

On those common rows, the original preliminary response gave an 11.8% reduction (10.4–13.0%). When preliminary water levels were sampled and filtered on the verified six-minute temporal basis, the reduction was 8.1% (7.2–8.9%). The verified response gave 4.5% (3.5–5.6%) on the same subset. At every site, the harmonized point estimate lay between its original preliminary and verified counterparts; harmonized reductions ranged from 3.2% at San Diego to 11.2% at Point Reyes. The intermediate pattern is consistent with the linked temporal construction contributing to attenuation, but the diagnostic does not isolate the effects of cadence, filtering, corrections, quality control, or other differences between NOAA products.

3.6 Processing and spectral sensitivities

The main model ordering did not depend on a single low-frequency cutoff, response window, or short lag (Table 3). Preliminary reductions were 11.7%, 11.8%, and 11.4% with no lower cutoff, 0.04 Hz, and 0.05 Hz, respectively. Changing the preliminary RMS window from 31 min to 21 or 45 min produced reductions of 10.8% and 9.0%. Fixed predictor lags from 0 to 3 h yielded 11.2–12.0%, with no material gain over the zero-lag definition.

Every tested moment order from 0 through -6 outperformed the prespecified height–period comparator, with reductions from 8.6% to 13.3%. The largest post hoc point estimate occurred at order -2, not at the prespecified order -4. The family therefore does not identify the negative fourth moment as a uniquely optimal exponent.

The more demanding height-only comparison narrowed the interpretation. For the preliminary response, $\sqrt{m_{-4}}$ had RMSE 0.346 and the height-only model had RMSE 0.356, a 3.0% reduction (-0.5 to 6.6%). For the verified response, the corresponding RMSE values were 0.445 and 0.438, a -1.6% reduction (-3.4 to 0.2%): height alone was slightly better. The primary result therefore establishes an advantage over the specified $\sqrt{H_s L_0}$ comparator, not a product-independent advantage over wave height itself.

Normalization-window sensitivity depended strongly on duration and water-level product. Across eligible 7-day windows, preliminary reductions ranged from 6.8% to 15.9% (99 start dates), whereas verified reductions ranged from -26.9% to 10.2% (100 start dates). The ranges narrowed with longer reference periods. For 60-day windows, preliminary reductions were 11.3–12.8% and verified reductions were 3.1–5.5% across 47 eligible starts. These values are ranges across eligible reference-window start dates, not confidence intervals. Month-scale reference periods thus reproduced the primary ordering more consistently than short calibration windows, especially for the verified response.

4 Discussion

4.1 What transferred across geography

The principal result concerns relative model performance across geography. Coefficients estimated from five Pacific Northwest pairs were applied unchanged to six California pairs, and the spectral model outperformed the matched height–period model at every site for the preliminary response. The 2023 check established temporal ordering within the development network; the California evaluation then exposed the fixed coefficients to new gauge settings, shelf environments, and regional wave climates.

Table 3 Sensitivity analyses for the pooled California comparison. Except where both products are named, processing, cutoff, lag, and moment analyses use the preliminary response. Height-only parentheses give 95% paired block-bootstrap intervals; normalization entries give ranges across eligible 2016 reference-window start dates, not confidence intervals.

Analysis	Variants and retained observations	Relative RMSE reduction	Interpretation
Moment order	0 (normalized wave height) to −6; 24,878 each	8.6–13.3%	All favored spectral; −2 largest post hoc
Height-only comparator	preliminary / verified; 24,878 each	3.0% (−0.5 to 6.6%) / −1.6% (−3.4 to 0.2%)	Advantage small for preliminary; reversed for verified
Lower-frequency cutoff	none / 0.04 / 0.05 Hz; 24,878 each	11.7% / 11.8% / 11.4%	Ordering unchanged
Preliminary RMS width	21 / 31 / 45 min; 24,858 / 24,878 / 24,853	10.8% / 11.7% / 9.0%	Ordering unchanged
Fixed predictor lag	0 / 1 / 2 / 3 h; 24,878 / 24,651 / 24,555 / 24,467	11.7% / 12.0% / 11.9% / 11.2%	No material lag gain
Synchronous regional bootstrap	preliminary / verified; 24,878 each	11.7% (10.3–13.5%) / 4.6% (3.0–6.1%)	Cross-site timing preserved
Product harmonization	preliminary on 6-min basis; 24,817	8.1% (7.2–8.9%)	Between original preliminary and verified
Normalization window	7 / 14 / 30 / 60 d; 99–47 preliminary; 100–47 verified starts	7 d: 6.8 to 15.9% / −26.9 to 10.2%; 60 d: 11.3 to 12.8% / 3.1 to 5.5%	Short verified windows unstable; longer windows narrower

This was not calibration-free transfer. Each California site supplied 2016 predictor and response means, which removed much of the between-site difference in scale before the common coefficients were applied. California outcomes did not tune the slopes or intercepts or determine the stations and predictors. The supported result is therefore fixed-coefficient geographic transfer after local reference normalization.

Normalization alone does not guarantee a favorable comparison because it does not make predictor–response relationships identical among sites. The preliminary ordering at all six gauges shows that the low-frequency-weighted index transferred better than the prespecified $\sqrt{H_s L_0}$ scale. The 4.5–17.4% site range, however, argues against a coast-invariant effect. The height-only sensitivity narrows the claim further: its preliminary advantage was small and uncertain and reversed for the verified response. The evidence supports a comparator-specific ranking, not a general benefit of low-frequency weighting.

4.2 Why the water-level product changes the conclusion

The verified analysis places the main limit on the claim. Its 4.6% reduction was less than half the preliminary value and below the 5% practical reference. The spectral model's verified R^2 was also negative: it was better than the bulk comparator, but both fixed models were worse in squared error than the held-out response mean. Relative improvement should not therefore be interpreted as useful absolute prediction.

Product harmonization located part of the attenuation without identifying its cause. Processing preliminary observations on the six-minute temporal basis moved the reduction from 11.8% to 8.1% on common rows, between the original preliminary and verified values. The intermediate result is consistent with temporal construction contributing to attenuation. It cannot separate cadence, filtering, corrections, quality control, or other product differences because those features change together and the diagnostic does not recreate NOAA's verification workflow.

The products represent different observational compromises. One-minute preliminary data retain rapid fluctuations but may include uncorrected sensor or processing effects. Verified data have undergone quality control, yet six-minute sampling and subsequent filtering can attenuate variability near the analyzed time scale. Thus, "verified" denotes product status rather than an error-free representation of the target response. The common-row result rules out different evaluation hours as the sole explanation but supports no stronger causal attribution.

4.3 Physical interpretation without mechanistic overreach

The primary comparison is compatible with established low-frequency wave physics. Significant height and peak period discard bandwidth and much of the incident spectral shape. Negative moments retain information about energy toward lower frequencies and may covary with wave-group forcing or conditions favorable to long-period motions (Bertin et al. 2018; Herbers et al. 1995; Longuet-Higgins and Stewart

1962). Shelf and nearshore processes subsequently reshape those motions (Smit et al. 2018; van Dongeren et al. 2007). This rationale makes a negative moment a plausible covariate, but the height-only result shows that the present data do not demonstrate a product-independent gain from the additional spectral weighting.

The analysis does not identify the pathway that produced the observed ranking. Tide gauges occupy harbors, embayments, or protected settings and respond to local resonance, depth, and geometry. The offshore buoy and gauge are separated by tens of kilometers, and no propagation model was fitted. The failure of uniform 0–3 h lags to improve the comparison does not prove instantaneous forcing, and the range of favorable moment orders does not isolate spectral shape as the mechanism. The -4 order remained the primary choice because it followed the antecedent hypothesis, not because the California data established it as uniquely optimal.

The endpoint is fixed-point tide-gauge variability after local filtering. Maximum runup is a moving-shoreline extreme shaped by breaking, slope, morphology, and swash processes absent from the model. No individual hazardous-event threshold or classifier was evaluated.

4.4 Statistical and geographic limits

The large number of hourly rows provides precise conditional estimates for the observed network, but it does not create thousands of independent geographic replications. Storm persistence motivated seven-day block resampling, while synchronous calendar blocks preserved shared regional storm timing. Both approaches gave the same qualitative conclusion. Neither, however, measures uncertainty over unobserved coasts; the intervals remain conditional on these six sites, their observed years, and the specified models.

Observation-weighted pooling gives each retained hour equal influence, not each coastline. Sites with two qualifying seasons contributed roughly twice as many rows as Point Reyes or Santa Barbara. The pooled result therefore represents the observed hours rather than an average California coast, while the site estimates and forest plot show the accompanying geographic heterogeneity.

The common 2016 scaling season is another source of conditionality. It offered a consistent pre-fitting reference and prevented test years from setting their own scales. However, 2016 followed exceptionally strong El Niño forcing and coastal response on the US West Coast (Barnard et al. 2017). The post hoc window analysis showed that 30- and 60-day reference periods gave comparatively narrow ranges around the primary contrasts, whereas 7-day windows were unstable, especially for the verified response. Local normalization therefore requires a sufficiently long and complete reference record; the present study does not establish transfer from a brief calibration sample or without local response observations.

Coverage criteria favored stations with long, complete public records and available verified observations. That selection improved comparability but may underrepresent difficult gauge environments or poorly monitored regions. Two external sites supplied only one test year, and all stations lie on the US West Coast. The expanded 2017–2025 analysis partly addresses sensitivity to the principal test years, but its smaller reductions show that performance varies with time. Application to other basins, seasons, and observing systems requires new external tests rather than extrapolation from the present intervals.

4.5 Implications for observation and model use

Within these limits, the low-frequency-weighted index is best viewed as a compact covariate for fixed-point coastal oscillations. Its preliminary advantage over the prespecified height–period scale survived geographic transfer, longer-record scoring, alternate cutoffs, filter widths, short lags, and dependence-aware resampling. The verified and height-only results make that consistency measurement- and comparator-specific rather than evidence of general superiority over simpler offshore measures.

The comparison also clarifies what can be transported to another network. Predictor definitions and fitted coefficients can be carried without tuning, but response processing and local normalization must remain explicit. A change in sampling cadence or product status changes the empirical target and requires renewed evaluation.

The next decisive observations should combine higher coastal sampling rates with stable quality control, reducing the present confounding between product status and temporal resolution. Networks outside the US West Coast would permit tests across coasts rather than only across time blocks within six sites. Independent calibration seasons would show whether the month-scale stability observed within 2016 persists under other wave climates.

Taken together, the preliminary, verified, and height-only comparisons support a bounded result: the index transferred better than the prespecified height–period scale for these fixed-point responses, but not consistently across products or against wave height alone.

5 Conclusion

Fixed coefficients based on the low-frequency-weighted index outperformed the prespecified height–period model at six California gauges after local 2016 normalization. The preliminary one-minute response showed an 11.7% RMSE reduction in 2024–2025. The reduction fell to 4.6% for verified six-minute observations, for which absolute skill remained weak. Against wave height alone, the corresponding contrasts were 3.0% for the preliminary response and −1.6% for the verified response, so a general advantage over wave height was not established.

These results support the index as a candidate covariate for sub-hourly oscillations measured at fixed coastal gauges when a sufficiently long local reference record is available. They do not establish calibration-free transfer or product-independent superiority over simpler offshore measures. Further evaluation should use higher-cadence quality-controlled coastal observations and independent networks outside the US West Coast.

References

Barnard PL, Hoover D, Hubbard DM, et al (2017) Extreme oceanographic forcing and coastal response due to the 2015–2016 El Niño. Nature Communications 8:14365. https://doi.org/10.1038/ncomms14365

Bertin X, de Bakker A, van Dongeren A, et al (2018) Infragravity waves: From driving mechanisms to impacts. Earth-Science Reviews 177:774–799. https://doi.org/10.1016/j.earscirev.2018.01.002

Herbers THC, Elgar S, Guza RT (1995) Generation and propagation of infragravity waves. Journal of Geophysical Research: Oceans 100(C12):24863–24872. https://doi.org/10.1029/95JC02680

Li C, Özkan-Haller HT, García Medina G, et al (2023) Observations of extreme wave runup events on the US Pacific Northwest coast. Natural Hazards and Earth System Sciences 23:107–126. https://doi.org/10.5194/nhess-23-107-2023

Longuet-Higgins MS, Stewart RW (1962) Radiation stress and mass transport in gravity waves, with application to surf beats. Journal of Fluid Mechanics 13(4):481–504. https://doi.org/10.1017/S0022112062000877

National Data Buoy Center (n.d.a) Historical NDBC data. URL https://www.ndbc.noaa.gov/historical_data.shtml

National Data Buoy Center (n.d.b) Nondirectional and directional wave data. URL https://www.ndbc.noaa.gov/measdes.shtml

NOAA Center for Operational Oceanographic Products and Services (n.d.a) Data retrieval API. URL https://api.tidesandcurrents.noaa.gov/api/prod/

NOAA Center for Operational Oceanographic Products and Services (n.d.b) Preliminary water level data. URL https://tidesandcurrents.noaa.gov/waterlevels.html

NOAA National Centers for Environmental Information (2022) ETOPO 2022 15 arc-second global relief model. https://doi.org/10.25921/fd45-gt74

Roberts DR, Bahn V, Ciuti S, et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40(8):913–929. https://doi.org/10.1111/ecog.02881

Smit PB, Janssen TT, Herbers THC, et al (2018) Infragravity wave radiation across the shelf break. Journal of Geophysical Research: Oceans 123(7):4483–4490. https://doi.org/10.1029/2018JC013986

Stockdon HF, Holman RA, Howd PA, et al (2006) Empirical parameterization of setup, swash, and runup. Coastal Engineering 53(7):573–588. https://doi.org/10.1016/j.coastaleng.2005.12.005

van Dongeren A, Battjes J, Janssen T, et al (2007) Shoaling and shoreline dissipation of low-frequency waves. Journal of Geophysical Research: Oceans 112(C2):C02011. https://doi.org/10.1029/2006JC003701