Portability and Limits of an Interpretable Polynomial Framework: A Global Significant-Wave-Height Case Study

Vahram A. Martirosyan
Dept. of Computer and Information Science
Higher Colleges of Technology
Abu Dhabi, United Arab Emirates
vmartirosyan@hct.ac.ae

***Abstract*—Global wave climatologies are produced by numerical wave models and stored as multi-gigabyte gridded fields, which are impractical for on-board or embedded use and are not directly auditable. We show that the monthly climatology of significant wave height (Hs) can be represented by a single, human-readable polynomial. The study doubles as a portability test of our interpretable SRPR framework, extending it from subsurface scalar fields (temperature, salinity, oxygen) to a boundary-sensitive surface field. Using the Ifremer WAVEWATCH III CFSR hindcast (global 0.5°, 2005–2014), we fit a stagewise-regularized polynomial regression (SRPR) on six surface features—Cartesian position on the unit sphere, absolute latitude, and the seasonal cycle—and obtain a degree-8 model of 3002 explicit terms with cross-validated $R^2 = 0.945$ and RMSE 0.262 m. Under strict evaluation the model retains $R^2 = 0.933$ when whole grid cells are held out, drops to 0.73 ± 0.18 when whole 30° regions are held out, and reaches anomaly $R^2 = 0.875$ relative to the zonal/seasonal mean. Independent validation against in-situ buoys shows the model emulates the *open-ocean* climatology well (offshore RMSE ≈ 0.5 m) but, being a deep-water climatology, cannot represent fetch-limited coastal seas. We also report a negative result: adding a distance-to-coast feature degrades performance because coastal validation points fall outside the strided training distribution. Mixing the in-situ buoy observations into training instead raises the held-out coastal R^2 from 0.34 to 0.52 and removes the bias, at a small open-ocean cost. A comparison with lightweight MLP surrogates shows that tiny neural networks are competitive in accuracy-per-kilobyte, so the polynomial's distinctive value is not raw accuracy but its *closed-form, auditable* representation. We therefore position it as the transparent backbone of a *gray-box* model, a physics-readable macro-scale layer to which a compact neural residual corrector can be added, rather than as a stand-alone wave tool.**

***Index Terms*—ocean waves, significant wave height, polynomial regression, elastic net, interpretable machine learning, edge deployment, WAVEWATCH III.**

I. Introduction

The significant wave height H_s—the average height of the highest third of waves—is the primary integrated quantity in wave climatology, and underpins wave-energy siting, ship routing, offshore design, and coastal hazard assessment. State-of-the-art global wave information comes from third-generation spectral models (WAVEWATCH III, WAM) forced by reanalysis winds [1], [3], or from products such as ERA5 [5] and satellite altimetry. These datasets are physically consistent and accurate, but they are large (gigabytes per field), require substantial storage and compute, and their internal behaviour cannot be read off directly.

This paper is part of an ongoing programme that validates a single interpretable engine—*stagewise-regularized polynomial regression* (SRPR) with an analytic solver—as a general-purpose framework for geophysical fields [10]. Prior targets were subsurface *scalar* fields (ocean temperature, salinity, dissolved oxygen). Here we deliberately stress-test the framework's *portability and boundary conditions* on a qualitatively different target: H_s, a *vector-integrated, boundary-sensitive* surface field whose climatology is organized by storm tracks, fetch, and coastal sheltering rather than by a depth-stratified column.

The contribution is therefore two-fold. First, a *method-level validation* of SRPR outside its original regime: we show what transfers unchanged, and where the framework breaks. Second, a *closed-form, auditable* representation of the global wave field suitable for edge deployment. We chart the full operational envelope—successes (open-ocean climatology, anomaly skill, and a buoy-mixing correction of the coastal bias), limits (coastal blindness and a failed distance-to-coast feature), and the parameter regime of the fix—rather than reporting accuracy alone.

II. Related Work

A. Wave climatologies

Global wave data are produced by third-generation spectral models such as WAVEWATCH III and WAM, either in forecast/hindcast configuration [1], [3] or embedded in atmospheric reanalyses; ERA5 [5] provides hourly H_s at 0.5°. These are physically complete but heavy and opaque. Machine learning has been applied to wave forecasting and downscaling, but the resulting models are typically neural networks or tree ensembles [12] and do not provide closed-form, auditable fields.

B. Compact surrogates for edge deployment

Deploying learned models on microcontrollers and embedded hardware [11] demands kilobyte-scale footprints and microsecond inference. This motivates closed-form surrogates that need no runtime or framework and whose entire behaviour can be inspected on-device.

C. Interpretable regression

Inherently interpretable models expose their decision rule rather than approximating it post-hoc [9], and sparse regression can recover governing equations from data [8]. Elastic net [6], [7] yields sparse polynomial coefficients compact enough to inspect and embed. The present work carries this line to the global wave-height climatology.

III. Data and Features

A. Target: a hindcast wave climatology

We use the Ifremer WAVEWATCH III hindcast forced by CFSR winds [1], [2], [4], distributed as per-variable monthly files (global $0.5°$ grid, 317×720 cells, 3-hourly). We use the H_s variable for 2005–2014 (ten years). Each monthly file is averaged in time, then averaged across years, giving a *climatological monthly-mean* H_s per grid cell. The resulting field (12 months against lat/lon) is the regression target; its annual mean is shown in Fig. 1.

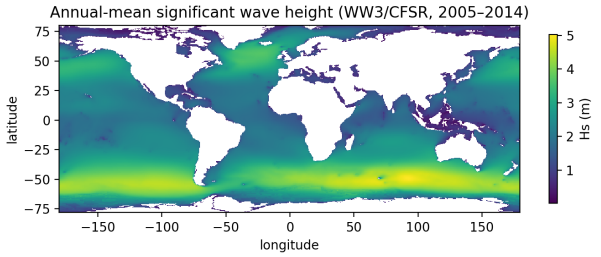


Fig. 1. Annual-mean significant wave height of the target climatology, with the mid-latitude storm belts and the Southern Ocean maximum clearly visible.

B. Features

Each sample is a (location, month) pair encoded by six surface features: Cartesian position on the unit sphere (x, y, z), absolute latitude $|\phi|$ (equator–pole contrast), and the seasonal harmonics $\sin(2\pi M/12), \cos(2\pi M/12)$. The Cartesian mapping removes the longitude seam; no depth feature is needed for a surface field. All features are standardized to zero mean and unit variance (Table I).

TABLE I
SIX-FEATURE SURFACE ENCODING.

Feature	Quantity
x, y, z	Cartesian position (unit sphere)
$\lvert\phi\rvert$	absolute latitude
$\sin(2\pi M/12)$	seasonal sine
$\cos(2\pi M/12)$	seasonal cosine

C. Independent observations

For validation we use in-situ moored buoys from three networks: NOAA NDBC [13] (CoastWatch ERDDAP), the Irish Marine Institute wave buoy network, and SECOORA—none of which were used in training. For each buoy we build a monthly-mean H_s climatology (2006–2015) and compare it with the model evaluated at the buoy location and month.

IV. Method

We fit the elastic-net objective

$$\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_i \big(y_i - \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\beta}\big)^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{\beta}\|_2^2, \qquad (1)$$

with an analytic two-stage solver (ridge preconditioning followed by coordinate descent on an active set) [6], [7]. A full degree-K polynomial in six variables has $\binom{6+K}{K}$ terms (923 at K=6, 3002 at K=8).

Normalization matters. The production SRPR path applies a per-sample min/max rescale to $[0, 2]$ *after* z-scoring. For this global field that transform is destructive: with it, degree-6 reached only $R^2 = 0.832$. Using z-score-only features (the `prenormalized` path) raised degree-6 to $R^2 = 0.920$. All results below use the corrected path.

V. Results

A. Degree ablation

Table II and Fig. 2 report 5-fold cross-validated accuracy versus polynomial degree. Accuracy rises monotonically and has not fully saturated at degree 8 ($R^2 = 0.945$, RMSE 0.262 m); degree 4 already reaches $R^2 = 0.874$ with only 209 terms.

TABLE II
5-FOLD CV OF THE GLOBAL H_s CLIMATOLOGY VS POLYNOMIAL DEGREE.

Degree	Terms	RMSE (m)	R^2
1 (ridge)	6	0.874	0.393
2	27	0.622	0.692
3	83	0.493	0.807
4	209	0.398	0.874
5	461	0.348	0.904
6	923	0.317	0.920
7	1715	0.283	0.936
8	**3002**	**0.262**	**0.945**

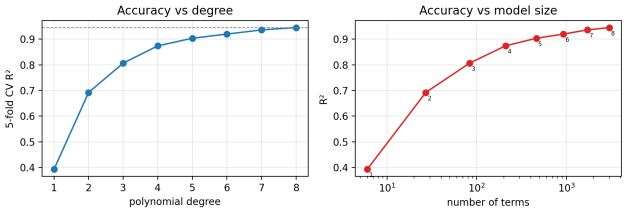


Fig. 2. Accuracy versus polynomial degree (left) and model size (right).

B. Strict evaluation and anomaly skill

Random-sample cross-validation is optimistic for spatially correlated fields. Table III reports strict protocols at degree 6. Holding out whole grid cells leaves the score essentially unchanged (0.933), so the model interpolates to unseen locations; holding out whole 30° regions reduces it to 0.73 ± 0.18 (mean $\pm$ std across five folds), with high fold-to-fold variance because some held-out regions are much harder than others. The anomaly R^2 of 0.875 shows the skill is not merely the latitude/season mean (which alone explains only ∼45% of the variance).

TABLE III
STRICT EVALUATION PROTOCOLS AT DEGREE 6.

Protocol	R^2
Random-sample CV	0.935
Cell-block CV (unseen cells)	0.933
Basin-block CV (unseen 30°)	0.73 ± 0.18
Anomaly (vs zonal/seasonal mean)	0.875

C. Comparison with lightweight neural baselines

We benchmark against small MLP surrogates on identical features and folds (Table IV), including a *gray-box* variant in which a small MLP is fit to the polynomial's residual. On random splits the degree-6 polynomial (3.6 KB, $R^2 = 0.935$) is competitive, and the gray-box reaches 0.950–0.953. On the harder basin-block split, tiny MLPs generalize to unseen regions *better* than the global polynomial—MLP$(32, 32)$ (5.1 KB) attains 0.791 ± 0.117 versus 0.732 ± 0.176—and with lower variance. We report this as a deliberate *boundary result*, not a hidden weakness: on a smooth field a tiny black-box is competitive per kilobyte, so the polynomial's case cannot rest on accuracy or size. What it adds is an *auditable* decision rule—and, as the gray-box rows show, the transparent backbone can be coupled to a small residual network to match the best black-box accuracy (0.950–0.953) while keeping the macro-scale field readable and physics-checkable.

TABLE IV
POLYNOMIAL VS LIGHTWEIGHT MLPS (5-FOLD, MEAN $\pm$ STD R^2).

Model	**KB**	**Random**	**Basin** 30°
SRPR polynomial (deg 6)	3.6	0.935±.001	0.732±.176
MLP (16)	0.5	0.884±.005	0.768±.064
MLP (32,32)	5.1	0.953±.002	0.791±.117
poly + MLP(32,32)	8.7	0.950±.002	0.709±.196

D. Validation against in-situ buoys

The model, trained only on the hindcast, reproduces observed buoy climatologies. At *offshore* NDBC sites the agreement is good ($R^2 = 0.60$, RMSE 0.50 m, bias +0.11 m). Adding sheltered/coastal/lake stations (Irish Marine Institute, SECOORA) sharply increases error: e.g. SmartBay (Galway Bay) is observed at 0.70 m but predicted at 2.93 m, and Lake Michigan station `0Y2W3` at 0.44 m vs 1.43 m. The multi-network pooled R^2 therefore falls to 0.34 (Fig. 3, Table V). This is expected: the model is a deep-water climatology and cannot represent fetch-limited coastal wave heights.

TABLE V
PER-NETWORK BUOY VALIDATION (2006–2015).

Network	**Stations**	**RMSE (m)**	R^2	**Bias (m)**
NDBC (offshore)	16	0.503	0.602	+0.112
Ireland (MI)	4	1.454	−0.140	+0.901
SECOORA	3	0.657	−3.464	+0.543
Pooled	23	0.763	0.337	+0.299

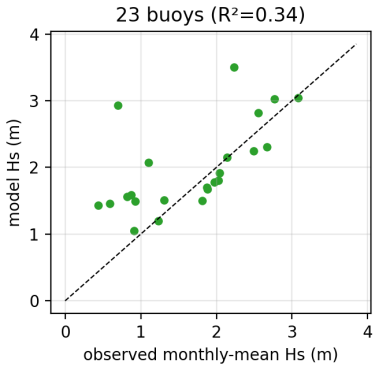


Fig. 3. Model vs observed monthly-mean H_s at 23 buoys from three networks. Offshore stations cluster near the 1:1 line; sheltered/coastal/lake stations are overpredicted.

E. Mixing in-situ observations into training

The hindcast-only model overpredicts coastal stations (Section V-D). Since these stations carry information absent from the smooth grid, we test adding them to training with an upweighting factor W (grid weight $= 1$) and evaluating by *leave-one-station-out* (Table VI). The coastal bias is largely removable: at $W{=}100$ the held-out-buoy R^2 rises from 0.337 to 0.522 and the bias falls from +0.30 to +0.05 m, at a small cost to the open-ocean field ($0.935 \rightarrow 0.922$). Excessive weight collapses both, so an optimum exists near $W{=}100$. This is a reweighting-based bias correction, not new coastal physics; the mixed model is stored separately.

Does the correction transfer across networks? Leave-one-network-out shows it transfers *partially*: with NDBC held out, training on grid plus Ireland/SECOORA buoys raises the NDBC R^2 from 0.59 to 0.72 and cuts the bias from +0.14 to +0.02 m; with SECOORA held out, RMSE falls from 0.66 to 0.46 m. The Irish stations, however, do not transfer (their +0.9 m bias persists), because their sheltered, tidally mixed regime is absent from the other networks. The correction is thus partially generic but regime-dependent: stations resembling the training set benefit, while a distinct regime needs its own observations.

TABLE VI
MIXING BUOYS INTO TRAINING (DEGREE 6, LEAVE-ONE-STATION-OUT).

Buoy weight W	**Buoy R^2**	**Buoy RMSE (m)**	**Grid R^2**
0 (hindcast only)	0.337	0.763	0.935
10	0.475	0.679	0.933
100	**0.522**	**0.648**	0.922
1000	0.427	0.710	0.861

F. Interpretability

Because every coefficient is explicit, the field decomposes into readable terms. The leading terms are $|\phi|$ (poleward intensification, i.e. storm tracks), $z = \sin\phi$ (hemispheric contrast, the Southern Ocean storm belt), the interaction $z \cdot \sin(2\pi M/12)$ (seasonal winter swell), and basin-scale combinations of x, y (fetch and swell propagation geometry). The model recovers the two dominant physical drivers—latitude and season—as its largest coefficients, without being told them.

G. Computational footprint

A degree-8 model stores 3002 float32 coefficients ($\approx$ 12 KB); the degree-6 model stores 923 (≈ 3.7 KB). Inference is a single polynomial evaluation—microseconds per point on a scalar CPU, with no GPU, matrix operation, or external dataset. This is the same edge profile as the SRPR models we have reported for ocean temperature, salinity and dissolved oxygen, and it is what distinguishes the approach from gridded climatologies and neural emulators.

H. A negative result: distance-to-coast

Motivated by the coastal failure, we tested adding a fetch proxy, $\log(1 + d_{\text{coast}})$, computed from the model's land mask. It *degraded* performance (buoy station-mean RMSE $0.67 \rightarrow 1.49$ m) and produced wild extrapolations near the coast, because the strided training grid (2°) contains few cells close enough to the coast, so coastal validation points are an extrapolation of the new feature. We report this to discourage the obvious but ineffective fix; a coastal model requires higher-resolution, region-specific treatment.

VI. DISCUSSION

What this model is for. It is a *closed-form, auditable* surrogate for an open-ocean monthly wave climatology: a single explicit polynomial of a few thousand coefficients that evaluates in microseconds on a scalar CPU, versus gigabytes and GPU-scale inference for the gridded fields it emulates. Crucially, the comparison in Section IV-C shows that a tiny MLP is competitive in accuracy-per-kilobyte on the same data, so the polynomial's case does not rest on accuracy or size alone—it rests on *transparency*: every term is readable, the field can be audited, and the model can serve as the physics-readable backbone of a *gray-box* system in which a compact neural residual corrector refines the macro-scale field for coasts and storms. The gray-box rows of Table IV are the recommended operating point: a transparent polynomial plus a small residual network.

What it is not. It is not a forecast or nowcast and it does not resolve individual storms or swell events. It is not valid in sheltered/coastal/lake waters *unless* in-situ observations are mixed into training, which recovers much of the coastal skill but is a bias correction rather than coastal physics. It is an emulator of a hindcast climatology, so its accuracy is bounded by that product.

Framework validation and the operational envelope. For the SRPR programme, the value of this study is the map it draws of the framework's *operational envelope*. The same engine, unchanged, transfers from depth-stratified scalar columns (temperature, salinity, dissolved oxygen) to a boundary-sensitive surface field (H_s), reaching $R^2 = 0.945$ under interpolation and 0.73 across unseen regions, and recovers the dominant physical drivers as readable coefficients at kilobyte scale. It also breaks predictably—a global smooth basis cannot represent fetch-limited coastal seas, and a naive distance-to-coast feature makes this worse by extrapolating beyond its training support—and the break is repairable in a principled way by upweighting in-situ observations ($0.34 \rightarrow 0.52$), which partially transfers to unseen stations. Charting successes, failures, and the fix is precisely what distinguishes a *framework validation* from an application study, and is the intended contribution of this work.

Modularity of the engine. None of this required changing the solver. A boundary-sensitive field is ingested purely through the feature map: here a seasonal sine/cosine pair and absolute latitude are appended to the Cartesian position, and the analytic two-stage SRPR solver—normalization, stagewise selection, elastic-net shrinkage, and the normal-equation solve—is used verbatim. The buoy-mixing extension is likewise transparent to the solver, entering only as upweighted design-matrix rows. This clean separation of a fixed, general-purpose analytic core from a field-specific feature map is what makes the framework portable across temperature, salinity, dissolved oxygen, and now significant wave height, and it isolates where a new field needs care (its feature support and boundary conditions) from where the engine is already general.

VII. CONCLUSION

A single degree-8 polynomial in six geometric/seasonal features represents the global open-ocean monthly significant-wave-height climatology with cross-validated $R^2 = 0.945$ (RMSE 0.262 m), independent offshore buoy RMSE of ≈ 0.5 m, and $R^2 = 0.73 \pm 0.18$ under whole-region holdout, while remaining fully readable and edge-deployable. Mixing the in-situ buoy observations into training (leave-one-station-out) raises the held-out coastal R^2 from 0.34 to 0.52 and removes the coastal bias at a small open-ocean cost. Because a tiny MLP is competitive in accuracy-per-kilobyte on the same data, we do *not* claim accuracy or size superiority: the polynomial's contribution is a *closed-form, auditable* representation, best used as the transparent backbone of a gray-box model with a compact neural residual corrector. We have deliberately

documented its boundaries—coastal inapplicability and the failure of a naive distance-to-coast feature—so that it is used where it is valid.

The resulting artifact is a 3.7–12 KB closed-form expression that evaluates in microseconds on a scalar CPU with no runtime, framework, or GPU. This edge footprint is a distinctive property of the framework: unlike neural surrogates of comparable accuracy, whose deployed weights are as opaque as they are compact, every coefficient here can be read, signed, and audited against physics, so the model can be shipped and trusted on constrained or safety- relevant platforms where an uninterpretable network would not be.

ACKNOWLEDGMENT

The author thanks Ifremer for the WAVEWATCH III hindcast, NOAA NDBC, the Irish Marine Institute and SECOORA for the buoy observations used in validation.

REFERENCES

[1] F. Ardhuin *et al.*, "Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation," *J. Phys. Oceanogr.*, vol. 40, no. 9, pp. 1917–1941, 2010. https://doi.org/10.1175/2010JPO4324.1

[2] S. Saha *et al.*, "The NCEP Climate Forecast System Reanalysis," *Bull. Amer. Meteor. Soc.*, vol. 91, no. 8, pp. 1015–1057, 2010. https://doi.org/10.1175/2010BAMS3001.1

[3] H. L. Tolman *et al.*, "User manual and system documentation of WAVEWATCH III R," NOAA/NWS/NCEP Tech. Note, 2014. https://github.com/NOAA-EMC/WW3/wiki/Files

[4] N. Rascle and F. Ardhuin, "A global wave parameter database for geophysical applications. Part 2: Model validation with improved source term parameterization," *Ocean Model.*, vol. 70, pp. 174–188, 2013. https://doi.org/10.1016/j.ocemod.2012.12.001

[5] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020. https://doi.org/10.1002/qj.3803

[6] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. B*, vol. 67, pp. 301–320, 2005. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. B*, vol. 58, pp. 267–288, 1996. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[8] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci.*, vol. 113, no. 15, pp. 3932–3937, 2016. https://doi.org/10.1073/pnas.1517384113

[9] C. Rudin, "Stop explaining black box machine learning for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, pp. 206–215, 2019. https://doi.org/10.1038/s42256-019-0048-x

[10] V. A. Martirosyan, "On-device TinyML for multi-variable ocean profiling: Sub-millisecond interpretable polynomial inference on resource-constrained marine sensor platforms," under review, 2027.

[11] P. Warden and D. Situnayake, *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. Sebastopol, CA, USA: O'Reilly Media, 2019. https://www.oreilly.com/library/view/tinyml/9781492052036/

[12] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[13] National Data Buoy Center, "NDBC standard meteorological data." [Online]. Available: https://www.ndbc.noaa.gov