

Independent validation of harmonic tidal predictions at Lagos Harbour Nigeria

Anyadiegwu P. C.

Department of Surveying and Geoinformatics, Imo State University, Owerri, Nigeria
Corresponding author email: hod.svgimsu.edu.ng@gmail.com

Abstract

Tidal predictions support sounding reduction and coastal operations, yet their errors require evaluation on observations excluded from model fitting. This retrospective study assessed hourly tidal predictions at the Lagos tide station, Nigeria, using the University of Hawaii Sea Level Center research quality hourly record. A fixed-frequency, ten-constituent harmonic regression was fitted to 13,897 valid hourly observations from 1993–1994 and tested against 7,034 valid observations from 1995. The archive is referenced to local tide staff zero; timestamps are reported in UTC. Prediction error was defined as predicted minus observed water level. Mean error, mean absolute error (MAE), root mean square error (RMSE), and monthly performance were computed; seven-day moving-block resampling quantified uncertainty. The mean error was -0.035 m (95% block-bootstrap interval -0.056 to -0.014 m), MAE was 0.071 m, and RMSE was 0.090 m (95% interval 0.074 – 0.104 m). Of the matched predictions, 73.6% were within 0.10 m, while the 95th percentile absolute error was 0.177 m. Monthly RMSE ranged from 0.051 m in April to 0.151 m in November among months with substantial coverage. An alternative model trained on 1993 alone achieved an RMSE of 0.084 m, indicating that more training data did not automatically improve this historical holdout. These results quantify performance for the archived 1995 observations, not present-day navigation or chart-datum operations. Datum verification and evaluation with recent station records are required before operational application.

Keywords: tide gauge; Lagos Harbour; harmonic regression; out-of-sample validation; prediction residual; hydrographic surveying.

Date of Submission: 09-10-2026

Date of acceptance: 20-10-2026

I. Introduction

Tidal prediction is central to hydrographic surveys because soundings measured at different stages of the tide must be related to a consistent vertical reference. Astronomical tides can be represented by constituents with distinct frequencies, but observed levels also include non-astronomical variation. Consequently, agreement between a harmonic prediction and observations is an empirical question at the station and period of interest (NOAA, 2007).

Badejo and Akintoye (2017) analysed 500 hourly Lagos Harbour observations using eleven tidal constituents and reported high- and low-water predictions. Their study establishes local feasibility. A longer, independently held-out test provides a separate estimate of performance and makes temporal variation in prediction error visible. Here, a historical research quality station record permits a reproducible evaluation, although its age limits the present operational implications.

The objectives were to estimate ten fixed-frequency harmonic components from the 1993–1994 record; quantify hourly predictive error in 1995; examine monthly error; and test sensitivity to training length and overtide inclusion. The contribution is a documented out-of-sample benchmark with explicit data coverage and uncertainty rather than an assertion of present-day Lagos tide-table accuracy.

II. Theory and related work

The harmonic representation expresses predicted water level as a mean offset plus sine and cosine terms at known tidal frequencies: $\hat{h}(t) = \beta_0 + \sum_j [a_j \cos(\omega_j t) + b_j \sin(\omega_j t)]$. The fitted coefficients represent amplitude and phase at each frequency. Least squares estimates β from available training observations. More elaborate packages apply astronomical phase and nodal corrections (Pawlowicz et al., 2002; NOAA, 2007). The deliberately simple fixed-frequency model used here omits those corrections, so its numerical results must not be identified with a national hydrographic authority's official predictions.

The observed-minus-predicted residual may contain wind and pressure effects, river influences, unmodelled tidal constituents, changes in mean sea level, instrumental error, or datum and timing mistakes. The decomposition is conceptual: this study has no independent meteorological and discharge measurements with

which to attribute a particular residual to any one mechanism. Serial dependence in hourly residuals also means that conventional independent-observation confidence intervals would be misleading.

III. Materials And Methods

3.1 Study station and source data

The Lagos station is listed near 6.42° N, 3.41° E in the IOC Sea Level Station Catalog. We used UHSLC station 233, research quality version c, hourly CSV h233c.csv, covering 2 August 1992 to 6 October 1996. The associated quality-assessment note states that observations are GMT-referenced and measured in millimetres relative to tide staff zero, with no obvious level shifts; it documents substantial gaps. The station catalog separately reports benchmark ties. Because no conversion to nautical chart datum was established for the downloaded series, all computations retain the archive's local vertical reference (IOC, n.d.; UHSLC, n.d.).

The downloaded CSV contains year, month, day, hour and sea level in millimetres without a header. Negative fill values (−32767 in the CSV), and any nonpositive or implausibly high entries outside 0–5 m, were excluded; no missing observations were interpolated. This operational filter removed the recorded fill values and retained 1–5 m values. Analysis was limited to complete calendar years: 1993–1994 for fitting and 1995 for independent validation. The archived quality note reports 86%, 72%, and 80% yearly completeness respectively, consistent with the extracted coverage after rounding.

Period	Expected hours	Valid hours	Coverage
1993–1994 training	17,520	13,897	79.3%
1995 validation	8,760	7,034	80.3%

3.2 Harmonic regression and prediction

The ten prespecified constituents were Q1 (13.398661° h^{−1}), O1 (13.943036), P1 (14.958931), K1 (15.041069), N2 (28.439730), M2 (28.984104), S2 (30.000000), K2 (30.082137), M4 (57.968208) and MS4 (58.984104). Their speeds follow standard tidal harmonic definitions (NOAA, n.d.). Time *t* was elapsed UTC hours from 2 August 1992 at 20:00 UTC. Ordinary least squares estimated the intercept and twenty sine/cosine coefficients on valid training rows only. Coefficients were then fixed and predictions produced at the timestamps of valid 1995 observations. This continuous-time construction can predict through gaps without inventing observations.

For matched validation observation h_i and prediction $\hat{h}_i$, error $e_i = \hat{h}_i - h_i$. Mean error ME = mean(e_i); MAE = mean(| e_i |); RMSE = $\sqrt{\text{mean}(e_i^2)}$. Positive ME denotes overprediction. We also calculated the median and 95th percentile of absolute error and the share with $|e_i| \leq 0.10$ m; this 0.10 m criterion is a descriptive threshold, not an adopted navigation safety standard.

3.3 Uncertainty and sensitivity

For ME and RMSE uncertainty, 2,000 circular moving-block bootstrap replicates of the ordered 1995 validation error sequence were drawn with blocks of 168 observed hourly entries and fixed random seed 20260927; percentile intervals are reported. Blocks follow the sequence of available observations, so a block can cross a missing calendar interval. The intervals therefore approximate serial dependence but do not model gap mechanisms. As sensitivity analyses, the same ten constituents were fitted on 1993 alone and an eight-constituent model excluding M4 and MS4 was fitted on 1993–1994; both were assessed against the identical 1995 holdout. No alternative was selected based on a second holdout.

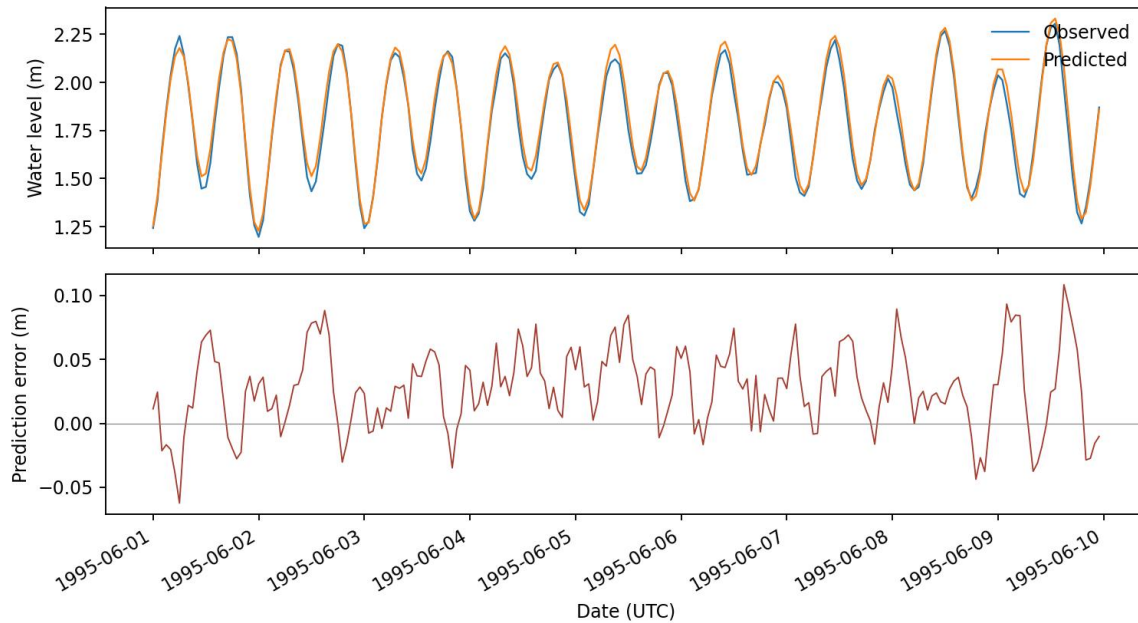
IV. Results

4.1 Overall predictive accuracy

Metric	1995 holdout result
Valid matched observations	7,034 / 8,760 (80.3%)
Mean error	−0.035 m (95% interval −0.056 to −0.014)
Mean absolute error	0.071 m
Root mean square error	0.090 m (95% interval 0.074 to 0.104)
Median absolute error	0.060 m
95th percentile absolute error	0.177 m
Within ±0.10 m	73.6%
Within ±0.20 m	97.6%

The negative ME indicates an average underprediction relative to the station's local gauge datum. The positive 95th percentile absolute error and larger RMSE than MAE indicate that occasional departures exceed typical discrepancies. The bootstrap interval for ME excludes zero under the specified resampling procedure; its substantive interpretation remains limited by the historical station reference, gaps, and the fixed-frequency model.

Figure 1. Observed and predicted station levels and prediction residuals for a nine-day illustration in June 1995.

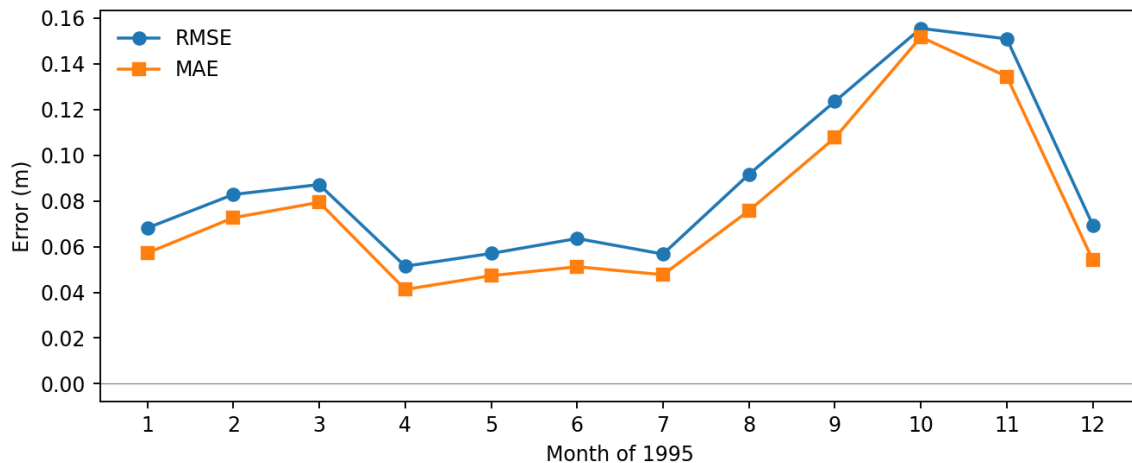


4.2 Monthly variation

Month	Valid hours	ME m	MAE m	RMSE m
Jan	744	0.050	0.057	0.068
Feb	383	-0.070	0.073	0.083
Mar	527	-0.079	0.079	0.087
Apr	720	-0.006	0.041	0.051
May	744	-0.028	0.047	0.057
Jun	720	0.013	0.051	0.064
Jul	359	0.021	0.048	0.057
Aug	744	-0.066	0.076	0.092
Sep	593	-0.106	0.108	0.124
Oct	128	-0.152	0.152	0.156
Nov	720	-0.133	0.134	0.151
Dec	652	0.026	0.054	0.069

Monthly performance was heterogeneous. April had the lowest RMSE (0.051 m; 720 valid hours), while November had the largest among months with at least 500 valid hours (0.151 m; 720 hours). October's RMSE was 0.156 m, but only 128 hours were available, so it cannot represent the month as a whole. November also showed an average underprediction of 0.133 m. These patterns establish temporal variation; they do not identify its physical cause.

Figure 2. Monthly MAE and RMSE in the independent 1995 validation year.



4.3 Sensitivity analyses

Model	Training observations	1995 ME m	1995 MAE m	1995 RMSE m
Ten constituents, 1993–1994	13,897	−0.035	0.071	0.090
Ten constituents, 1993 only	7,555	−0.015	0.068	0.084
Eight constituents, 1993–1994	13,897	−0.035	0.073	0.091

Excluding the two quarter-diurnal terms raised holdout RMSE by about 0.002 m. Restricting training to 1993 lowered RMSE by about 0.006 m and reduced bias. This is an empirical comparison on one held-out year; it does not prove that shorter training is generally better. A shifted training mean, unequal seasonal coverage, or interannual variability could contribute, but these mechanisms were not independently identified.

V. Discussion

The holdout assessment indicates that the specified harmonic regression captured much of the hourly oscillation while systematically underpredicting mean observed level in 1995. The result is station-, datum-, model- and period-specific. Because fixed constituent frequencies lack explicit nodal and astronomical phase adjustments, the comparison should be read as a transparent benchmark rather than the best achievable tide prediction. Badejo and Akintoye (2017) used a different observation period, model specification and evaluation design, so a direct numerical ranking of the two studies would be inappropriate.

Monthly residual shifts warrant examination against atmospheric forcing, freshwater discharge, datum records and instrument maintenance logs if those records can be obtained. A nearly complete November exhibited comparatively large errors, while the severe October gap prevents a confident monthly inference. Missingness may also be non-random: operational failures during extreme events would bias the apparent accuracy. The archive quality note says timing is good and identifies no obvious level shifts, but those observations do not establish chart datum equivalence or recent station stability.

The practical implication for hydrographic surveying is limited but useful: a 0.090 m historical hourly RMSE cannot be translated directly into an allowable sounding uncertainty or a present-day reduction correction. A sounding survey must use the applicable chart datum, contemporaneous water-level observations and the accuracy requirements of its survey class. Independently verifying the station's datum ties is a prerequisite to operational use.

VI. Conclusions and recommendations

A ten-constituent harmonic regression trained on archived 1993–1994 Lagos observations achieved 0.071 m MAE and 0.090 m RMSE on 7,034 independent observed hours in 1995. Its average prediction was 0.035 m below observed level; errors were uneven across months. Sensitivity testing showed modest differences under alternative constituent and training selections. These findings establish a reproducible historical benchmark, not a current tide prediction service.

Future work should obtain a recent, quality controlled multi-year Lagos record and documented vertical datum ties; compare a nodally corrected harmonic package with the fixed-frequency benchmark on predeclared holdouts; incorporate verified weather and discharge variables when attributing residuals; report high-water timing and height from a higher-frequency observed record; and assess whether missing observations cluster around extreme conditions. Until those checks are complete, use this study for methodological assessment rather than operational sounding reduction.

Data availability and reproducibility

The original hourly series is publicly archived as UHSLC research quality Lagos station 233c, file h233c.csv. The associated station quality note qa233c.dmt documents source, units, reference level and gaps. Reproduction requires reading the five CSV columns as year, month, day, UTC hour and millimetres; removing –32767 fill values; restricting training to 1993–1994 and validation to 1995; converting levels to metres; forming the 21-column intercept/sine/cosine design matrix at the ten stated fixed frequencies; solving least squares on training rows; and calculating the listed statistics only on valid 1995 rows. No field observations were collected for this analysis.

References

- [1]. Badejo, O. T., & Akintoye, S. O. (2017). High and low water prediction at Lagos Harbour, Nigeria. *Nigerian Journal of Technology*, 36(3), 944–952. <https://doi.org/10.4314/njt.v36i3.39>
- [2]. IOC Sea Level Station Catalog. (n.d.). Lagos station details, SSC-lagos. <https://www.ioc-sealevelmonitoring.org/ssc/stationdetails.php?id=SSC-lagos>
- [3]. National Oceanic and Atmospheric Administration. (2007). Tidal analysis and predictions. NOAA Special Publication NOS CO-OPS 3. https://tidesandcurrents.noaa.gov/publications/Tidal_Analysis_and_Predictions.pdf
- [4]. National Oceanic and Atmospheric Administration. (n.d.). Harmonic constituents and constituent speeds. https://tidesandcurrents.noaa.gov/about_harmonic_constituents
- [5]. Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers & Geosciences*, 28(8), 929–937. [https://doi.org/10.1016/S0098-3004\(02\)00013-4](https://doi.org/10.1016/S0098-3004(02)00013-4)
- [6]. University of Hawaii Sea Level Center. (n.d.). Lagos-C research quality hourly data, station 233c (h233c.csv), and quality assessment (qa233c.dmt). <https://uhsdc.soest.hawaii.edu/data/csv/rqds/atlantic/hourly/h233c.csv>; <https://uhsdc.soest.hawaii.edu/rqds/atlantic/doc/qa233c.dmt>