Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gaussian process reconstruction and seasonal change in tidal non-linearity in a monsoon-influenced estuary

Pattharaporn Thongnim

pattharaporn@buu.ac.th

Burapha University

Research Article

Keywords: tidal asymmetry, estuarine tides, shallow-water constituents, seasonal modulation, Gaussian process, covariance kernel selection

Posted Date: September 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11024647/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Ocean Dynamics on October 19th, 2026.
See the published version at https://doi.org/10.1007/s10236-026-01869-6.

Gaussian process reconstruction and seasonal change in tidal non-linearity in a monsoon-influenced estuary

Pattharaporn Thongnim[1*]

[1*]Department of Mathematics, Faculty of Science, Burapha University, Chonburi, 20131, Thailand.

Corresponding author(s). E-mail(s): pattharaporn@buu.ac.th;

Abstract

Shallow estuaries distort the incoming tide as it propagates into shallowing water. How strong that distortion becomes controls the speed at which water levels rise, the direction of net sediment transport and the landward limit of salt intrusion. This study quantifies that distortion at Bang Pakong, a monsoon-influenced estuary on the inner Gulf of Thailand, using 353,644 water-level readings recorded at about ten-minute intervals between May 2019 and June 2026. A least-squares harmonic analysis with 33 constituents and nodal corrections explains 94.7 % of the hourly variance and gives a form factor of 1.28, placing the station in the mixed, mainly semidiurnal regime. The record is strongly flood dominant with a mean rise of 6.26 h against a mean fall of 8.12 h and a duration asymmetry of 1.86 h measured directly from 7,889 tidal cycles. Allowing annual modulation of the non-degenerate constituents shows that the shallow-water constituents vary seasonally two to three times more than the principal ones, with relative swings of 22–37 % for $MS_{\mathbf{4}}$, $M_{\mathbf{4}}$, $MO_{\mathbf{3}}$, $MN_{\mathbf{4}}$ and $M_{\mathbf{6}}$ against 13 % for $M_{\mathbf{2}}$. All of them peak between June and August, when the mean water level is 14.8 cm lower and river discharge is highest. The $M_{\mathbf{2}}$ amplitude declines by $0.55 \pm 0.09\,\mathrm{cm\,yr^{-1}}$, a trend reproduced independently in both seasons and not attributable to nodal mis-correction. Annual modulation of $K_{\mathbf{1}}$, $S_{\mathbf{2}}$ and $P_{\mathbf{1}}$ proves unidentifiable when $S_{\mathbf{1}}$ and $T_{\mathbf{2}}$ are retained, a degeneracy that yields physically meaningless amplitudes if ignored. Because 6.4 % of the hourly record is missing, seven Gaussian process covariance structures are compared by closed-form leave-one-out residuals, by the Akaike information criterion and by withheld-data experiments. A two-scale Matérn 3/2 kernel with a semiannual mean term is selected and recovers 11 % more tidal cycles for the duration analysis without changing the harmonic amplitudes.

Keywords: tidal asymmetry, estuarine tides, shallow-water constituents, seasonal modulation, Gaussian process, covariance kernel selection

1 Introduction

The tide entering a shallow estuary does not keep the shape it had offshore. Friction acting on a water column whose depth varies over the tidal cycle, together with the changing width and cross-section of the channel, transfers energy from the astronomical constituents into higher harmonics (Friedrichs and Aubrey 1988). The resulting wave is asymmetric. The rise and the fall of the water surface take different lengths of time, and the associated currents differ in strength. This asymmetry is not a detail of tidal bookkeeping. A shorter rise concentrates the flood current, favours landward transport of fine sediment and shortens the warning time available before a high-water peak.

The standard diagnostics of this distortion are the amplitude ratio of the first overtide to its parent constituent, M_4/M_2 and the relative phase $2\phi_{M2} - \phi_{M4}$ (Friedrichs and Aubrey 1988). Both are convenient but both depend on the phase convention adopted and the second becomes ambiguous in mixed regimes where diurnal constituents are comparable in size to the semidiurnal ones. A duration asymmetry measured directly from the record avoids these difficulties and the measure has been applied successfully in mixed semidiurnal systems (Gong et al. 2016).

Tidal distortion is not constant in time. Seasonal changes in stratification, in river discharge and in mean water level all modify the frictional balance and seasonal variation of M_2 of 5–10 % has been reported for coastal and polar regions worldwide (Müller et al. 2014). In monsoon-influenced settings the effect can be larger and can extend to the overtides. Seasonal variation of M_2 and M_4 currents has been observed in the Gulf of Khambhat (Mandal 2024) and modelling of large estuaries indicates that river discharge enhances overtide generation through increased bottom friction with the strongest effect when river and tidal discharge are comparable (Guo et al. 2023).

The inner Gulf of Thailand has received far less attention in this respect. The regional tide itself is understood. The gulf is strongly diurnal and Cui et al. (2019) show that this character is a passive response to the quarter-wavelength resonance of the South China Sea rather than a resonance of the gulf. Regional analyses of the tide-gauge records have then focused on mean sea level and on trends in the principal constituents (Pongsiri et al. 2020) rather than on non-linear distortion and the estuaries that discharge into the inner Gulf remain largely undescribed. This is a gap worth closing. The Bang Pakong estuary supplies water for agriculture, aquaculture and municipal use in Chachoengsao Province, is subject to seasonal salt intrusion, and drains a province that experiences monsoon flooding.

This paper describes the tidal regime of the Bang Pakong estuary from seven years of ten-minute observations and asks three questions of the tide itself. How asymmetric is the tide, measured without recourse to phase conventions? Does the distortion vary with the monsoon, and if it does, how large is the change and which way does it go?

And is there evidence of change over the seven years of record that cannot be explained by the nodal cycle?

Answering the first of these questions from the record as it stands is not straightforward. A rise or a fall can be timed only where the water level is continuous from one turning point to the next, and 6.4 % of the hourly values are missing. Gaps in tide-gauge records are usually closed in one of two ways. Harmonic prediction restores the astronomical tide but not the weather-driven and river-driven part of the signal and interpolation in time treats the missing value as a function of the clock alone and returns a single number that carries no measure of its own uncertainty. Neither uses what the record itself says about how the non-tidal level is correlated from one day to the next. The reconstruction used here treats that non-tidal part as a Gaussian process and the filled value and its uncertainty then follow together from a covariance structure estimated from the observed days. That structure is an assumption in its own right and seven candidates are therefore compared rather than one being adopted. A fourth question follows. How much does the choice of covariance structure change what is recovered and does the reconstruction disturb the tidal quantities it is meant to support?

2 Data and methods

The analysis runs in eight steps, shown in Fig. 1. The ten-minute record is screened for instrument faults and placed on an hourly grid. A harmonic model with 33 constituents is then fitted to the hourly values and subtracted which separates the deterministic tide from the non-tidal residual. That residual is averaged to a daily mean water level and modelled as a Gaussian process, whose covariance structure is chosen among seven candidates rather than assumed. Adding the harmonic prediction back to the interpolated daily level completes the hourly record and the completed record supplies the rises and falls used for the duration analysis. Seasonal modulation and interannual change are estimated from the harmonic model itself and do not use the reconstruction. The subsections below follow these steps in order.

2.1 The station and the record

The Marine Department of Thailand operates a water-level station in Tha Kham Subdistrict, Bang Pakong District, Chachoengsao Province, near the mouth of the Bang Pakong River on the eastern side of the inner Gulf of Thailand. The record analysed here contains 353,644 readings at a nominal interval of ten minutes from 7 May 2019 to 23 June 2026.

2.2 Screening the record and building an hourly series

Five screening rules were applied in order of decreasing certainty. Duplicated timestamps were merged and exact zeros were removed. A plateau at 5.67–5.69 m in late 2022, during which the sensor was evidently stuck, was discarded, as were runs in which the value did not change for at least three hours. Isolated spikes were removed

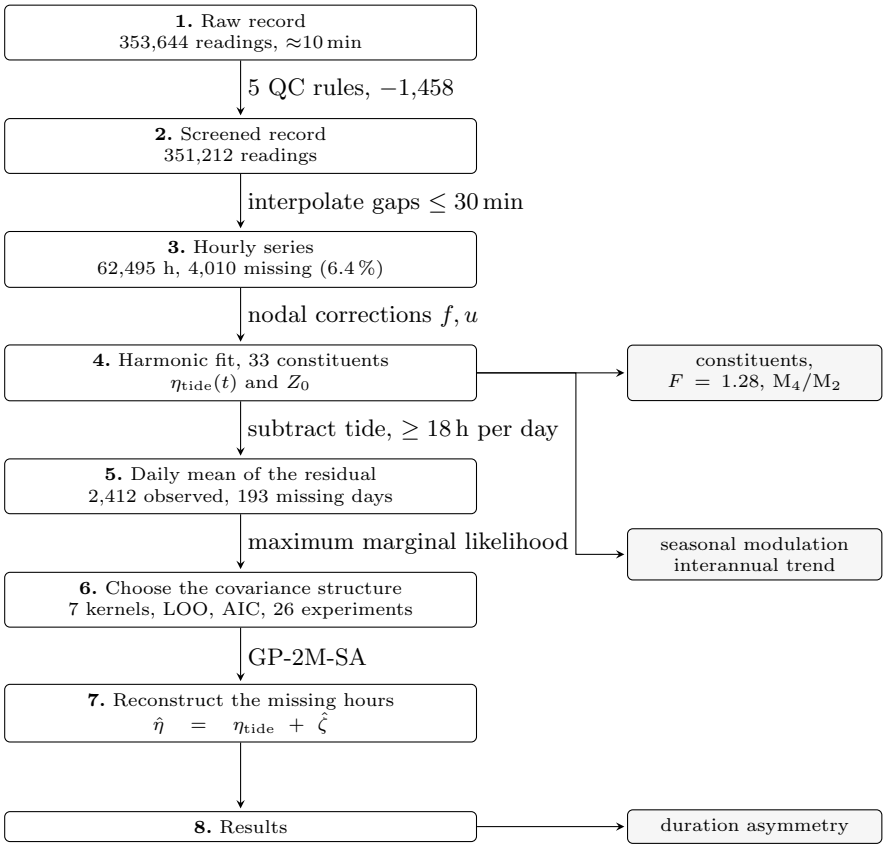


Fig. 1 The analysis pipeline. Boxes on the left are the successive states of the data. Labels on the arrows give the operation applied and the numbers involved. Shaded boxes on the right show which results come from which stage. Only the duration analysis uses the reconstructed record, while the harmonic quantities are estimated from the observed hours alone

with a Hampel filter that rejected values deviating by more than 0.25 m from a centred 70-minute running median, applied in two passes. The threshold was chosen from the distribution of the deviations themselves rather than from the residuals of a tidal model, so that observations the tidal model cannot explain are not discarded at this stage. In total 1,458 readings were removed, leaving 351,212.

Hourly values were obtained by interpolating between screened readings no more than 30 minutes apart which avoids drawing a straight line across an interval in which the tide has completed part of a cycle. This produced 62,495 hourly values of which 4,010 (6.4 %) are missing, the longest outage lasting 75 days between 25 June and 7 September 2021.

Fig. 2 shows the resulting coverage month by month. Of 2,605 calendar days, 2,412 reach the 18-hour completeness threshold used later for the daily means and 193 do not. The three episodes that dominate the loss are a 25-day outage in November 2019, a 39-day outage from 26 November 2020 to 5 January 2021 and a 75-day outage from

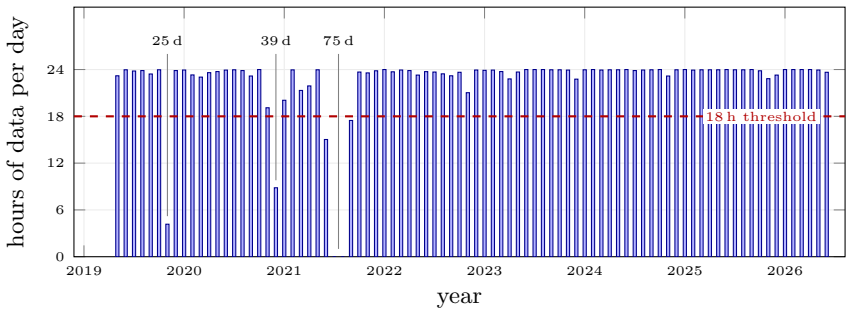


Fig. 2 Coverage of the ten-minute record, shown as the mean number of hours of data per day in each month

24 June to 8 September 2021. The last of these removes the whole 2021 wet season which is why no wet-season estimate is available for that year.

2.3 Separating the tide from the residual

Water level was represented as

$$\eta(t) = Z_0 + \sum_{k=1}^{33} f_k(t)\, A_k \cos(\omega_k t + u_k(t) - \phi_k)\,, \tag{1}$$

where ω_k are fixed astronomical angular speeds obtained from the harmonic development of the tide-generating potential (Doodson 1921) and f_k, u_k are the nodal amplitude factor and phase correction (Schureman 1958), evaluated at every observation time rather than held constant within a year. Expanding the cosine makes (1) linear in the coefficients $a_k = A_k \cos\phi_k$ and $b_k = A_k \sin\phi_k$. The $1 + 2 \times 33 = 67$ unknowns, a constant plus an amplitude and a phase coefficient for each constituent, were obtained by ordinary least squares on the $62{,}495 - 4{,}010 = 58{,}485$ valid hourly values. Amplitudes and phases follow as $A_k = (a_k^2 + b_k^2)^{1/2}$ and $\phi_k = \arctan(b_k/a_k)$. Table 1 sets out how the 33 are made up. Twenty-one of them are the standard astronomical constituents of the diurnal and semidiurnal bands, whose frequencies come directly from the orbital motion of the Moon and the Sun. The remaining twelve belong to species three and above. These carry no astronomical forcing of their own and are generated within the estuary by the non-linear processes described in the introduction, which makes them the part of the set that measures the distortion. The set follows the Rayleigh comparison of Foreman (1977), in which a constituent is retained only when its frequency differs from that of every other by at least one cycle over the record and which is the rule implemented in standard harmonic analysis packages (Pawlowicz et al. 2002). For a record of this length that limit is 0.006° per hour and the closest separations in the set are the 0.041° per hour between S_2 and T_2 and between K_1 and S_1 which one year of data would already resolve. The long-period constituents are not included, which leaves the annual and semiannual variation of mean water level to the mean function of the Gaussian process.

Table 1 Composition of the constituent set of (1). The species number is the number of cycles per day, with species 1 diurnal and species 2 semidiurnal. Species 3 and above are shallow-water compounds produced by non-linear interaction rather than by astronomical forcing

Species	Band	n	Constituents
1	Diurnal	10	$2Q_1$, Q_1, ρ_1, O_1, NO_1, P_1, S_1, K_1, J_1, OO_1
2	Semidiurnal	11	$2N_2$, μ_2, N_2, ν_2, M_2, λ_2, L_2, T_2, S_2, K_2, $2SM_2$
3	Terdiurnal	4	MO_3, M_3, MK_3, SK_3
4	Quarter-diurnal	5	MN_4, M_4, MS_4, MK_4, S_4
6	Sixth-diurnal	2	M_6, $2MS_6$
8	Eighth-diurnal	1	M_8
	Total	33	

2.4 Measuring rise and fall durations

High and low waters were located as local extrema of the hourly series, requiring five consecutive hourly values and enforcing alternation between highs and lows. The rise duration is the interval from a low water to the following high water and the fall duration the interval from a high water to the following low water. Intervals shorter than 2 h or longer than 20 h were discarded as artefacts of gaps. Of the 8,038 intervals between successive turning points, 79 fall below the lower bound and 70 above the upper which leaves the 7,889 cycles used in what follows. The width of the window sets how prominent a candidate has to be before it counts. A three-hour window admits 1,372 further extrema whose median prominence is 3.4 cm against a median range of 1.81 m between successive turning points which is wind and instrument noise on the flank of the tide rather than a turning point. A window as wide as nine hours instead approaches the mean fall duration and begins to merge genuine half cycles. Widening the window from three to seven hours moves the mean asymmetry from 1.40 to 2.06 h without changing its sign or the absence of an interannual trend. This gives a duration asymmetry that is independent of any phase convention.

2.5 Testing the constituents for seasonal change

To test for seasonal variation, each modulated constituent was allowed an amplitude and phase that vary annually, which is equivalent to multiplying its two design columns by $\sin 2\pi t_y$ and $\cos 2\pi t_y$ with t_y the time in years and adding the four resulting columns to the design matrix.

This cannot be done for every constituent. Modulating a constituent at one cycle per year creates sidebands displaced by exactly the rate of the solar mean longitude, $0.0410686^\circ\,\mathrm{h}^{-1}$. For three constituents in the set those sidebands coincide with constituents that are already present, namely

$$\mathrm{K}_1 - 1\,\mathrm{cpy} = \mathrm{S}_1, \qquad \mathrm{S}_2 - 1\,\mathrm{cpy} = \mathrm{T}_2, \qquad \mathrm{P}_1 + 1\,\mathrm{cpy} = \mathrm{S}_1. \tag{2}$$

Including both the modulation terms and the coincident constituents makes the design matrix numerically singular. When this was done the fitted S_2 amplitude reached 9.9 m, an obviously meaningless value produced by two nearly collinear columns cancelling one another. Modulation was therefore applied only to M_2, N_2, O_1, Q_1, K_2, M_4, MS_4, MN_4, MK_3, MO_3, M_6 and $2MS_6$, for which the condition number of the design matrix is 2.3.

Confidence intervals for the seasonal swing were obtained by a moving-block bootstrap with 30-day blocks and 200 replicates which preserves the short-range correlation of the residuals.

2.6 Estimating change from year to year

Constituent amplitudes were estimated separately for each complete calendar year from 2020 to 2025 and an ordinary least-squares line was fitted to the resulting six annual values with the standard error of the slope taken from the scatter of those values about the line. Because the record is not evenly distributed across seasons in every year, and because the constituents themselves vary seasonally, the same analysis was repeated using only the dry season (December to March) and only the wet season (June to October). May and November are left out of both so that the two seasons do not overlap at the turn of the monsoon. Every season is complete apart from the 2021 wet season which the 75-day outage removes.

2.7 Filling the gaps with a Gaussian process

Least squares does not require a complete series and the harmonic analysis therefore uses the 58,485 valid hours directly. The duration analysis, however, needs five consecutive hourly values and every gap therefore removes the tidal cycles around it. To test whether this loss biases the result, the missing hours were reconstructed and the analysis repeated on the completed series.

The daily series that the process is fitted to is formed from the residual rather than from the water level itself and the distinction matters. On a day with all twenty-four hours present the two are identical but on an incomplete day a plain average is biased because the hours that are missing are not spread evenly over the tidal cycle. At this station that bias exceeds 0.3 m on the worst days. The harmonic prediction is therefore subtracted first, the residual is averaged over the day and the mean level Z_0 is added back. A day is kept only if at least eighteen hours survive quality control which leaves 2,412 of the 2,605 calendar days.

The non-tidal part of the record was modelled as a Gaussian process with an explicit mean function (Rasmussen and Williams 2006; Xu et al. 2024). Writing ζ_i for the daily mean water level on day t_i,

$$\zeta_i = \mathbf{h}(t_i)^{\top}\boldsymbol{\beta} + g(t_i) + \varepsilon_i, \qquad g \sim \mathcal{GP}\big(0, k(t,t')\big), \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2), \tag{3}$$

with the basis

$$\mathbf{h}(t) = \big[\,1,\; t,\; \sin 2\pi t,\; \cos 2\pi t,\; \sin 4\pi t,\; \cos 4\pi t\,\big]^{\top}, \tag{4}$$

which carries the trend and the annual and semiannual cycles. Collecting the basis vectors in $\mathbf{H} = [\mathbf{h}(t_1), \dots, \mathbf{h}(t_n)]^\top$ and writing

$$\mathbf{K}_y = \mathbf{K} + \sigma_n^2 \mathbf{I}, \qquad [\mathbf{K}]_{ij} = k(t_i, t_j), \tag{5}$$

the coefficients are obtained by generalised least squares,

$$\hat{\boldsymbol{\beta}} = \mathbf{B}\,\mathbf{H}^\top \mathbf{K}_y^{-1} \boldsymbol{\zeta}, \qquad \mathbf{B} = \left(\mathbf{H}^\top \mathbf{K}_y^{-1} \mathbf{H}\right)^{-1}, \tag{6}$$

where $\mathbf{B}$ is also the covariance of $\hat{\boldsymbol{\beta}}$. At a day t_* that was not observed the predictive mean and variance are

$$\mu_* = \mathbf{h}_*^\top \hat{\boldsymbol{\beta}} + \mathbf{k}_*^\top \mathbf{K}_y^{-1} \left(\boldsymbol{\zeta} - \mathbf{H}\hat{\boldsymbol{\beta}}\right), \tag{7}$$

$$\sigma_*^2 = k_{**} - \mathbf{k}_*^\top \mathbf{K}_y^{-1} \mathbf{k}_* + \mathbf{r}^\top \mathbf{B}\,\mathbf{r}, \tag{8}$$

with $\mathbf{k}_*$ the vector of covariances between t_* and the observed days, $k_{**} = k(t_*, t_*)$ and $\mathbf{r} = \mathbf{h}_* - \mathbf{H}^\top \mathbf{K}_y^{-1} \mathbf{k}_*$. The last term of (8) adds the uncertainty of $\hat{\boldsymbol{\beta}}$. Far from any observation $\mathbf{k}_*$ tends to zero and (7) returns to the mean function. The reconstruction therefore degrades gracefully into the deterministic seasonal curve rather than into an arbitrary value.

The residual autocorrelation of the daily series requires two time scales. The covariance is accordingly the sum of two Matérn 3/2 kernels,

$$k(\tau) = \sum_{m=1}^{2} \sigma_m^2 \left(1 + \frac{\sqrt{3}\,\tau}{\ell_m}\right) \exp\!\left(-\frac{\sqrt{3}\,\tau}{\ell_m}\right), \qquad \tau = |t - t'|. \tag{9}$$

The hyperparameters $\boldsymbol{\theta} = (\sigma_1^2, \ell_1, \sigma_2^2, \ell_2, \sigma_n^2)$ were estimated by maximising the log marginal likelihood with $\boldsymbol{\beta}$ profiled out,

$$\log p(\boldsymbol{\zeta} \mid \boldsymbol{\theta}) = -\tfrac{1}{2} \mathbf{e}^\top \mathbf{K}_y^{-1} \mathbf{e} - \tfrac{1}{2} \log |\mathbf{K}_y| - \tfrac{n}{2} \log 2\pi, \qquad \mathbf{e} = \boldsymbol{\zeta} - \mathbf{H}\hat{\boldsymbol{\beta}}, \tag{10}$$

which gave $\ell_1 = 0.86\,\mathrm{d}$ and $\ell_2 = 5.6\,\mathrm{d}$ with $\sigma_1 = 7.8\,\mathrm{cm}$, $\sigma_2 = 5.5\,\mathrm{cm}$ and $\sigma_n = 0.9\,\mathrm{mm}$. The two length scales correspond to day-to-day variability and to weather-driven anomalies lasting several days.

Missing hours were then reconstructed as the harmonic prediction plus the interpolated daily level,

$$\hat{\eta}(t) = Z_0 + \sum_{k} f_k(t) A_k \cos(\omega_k t + u_k(t) - \phi_k) + \hat{\zeta}(t). \tag{11}$$

The reconstruction is used only as a robustness check and all values reported in this paper come from the observed data unless stated otherwise.

2.8 Comparing seven covariance structures

The kernel of (9) was not assumed. Seven covariance structures were compared, spanning three design choices, namely the roughness of the sample paths, the number of time scales, and whether the seasonal shape is allowed to evolve. Besides the Matérn 3/2 form of (9) with one and with two components, the candidates were the Matérn 1/2 kernel $k(\tau) = \sigma_f^2 \exp(-\tau/\ell)$ which is the continuous first-order autoregressive process, the smoother Matérn 5/2 kernel

$$k_{5/2}(\tau) = \sigma_f^2 \left(1 + \frac{\sqrt{5}\,\tau}{\ell} + \frac{5\tau^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5}\,\tau}{\ell}\right), \tag{12}$$

the rational quadratic kernel

$$k_{\mathrm{RQ}}(\tau) = \sigma_f^2 \left(1 + \frac{\tau^2}{2\alpha\ell^2}\right)^{-\alpha}, \tag{13}$$

which is a scale mixture of squared-exponential kernels and therefore acts on a continuum of time scales, and a quasi-periodic term

$$k_{\mathrm{QP}}(\tau) = \sigma_f^2 \exp\left(-\frac{2\sin^2(\pi\tau/P)}{\ell_p^2} - \frac{\tau^2}{2\ell_d^2}\right), \tag{14}$$

added to a Matérn 3/2 component, with P fixed at one year. A decay length ℓ_d much shorter than the record would indicate a seasonal cycle whose shape changes from year to year.

The three values of the Matérn smoothness parameter ν differ in how rough the sample paths are allowed to be. A process with $\nu = 1/2$ is continuous but nowhere differentiable, $\nu = 3/2$ is once differentiable in mean square and $\nu = 5/2$ is twice differentiable while the rational quadratic and the quasi-periodic forms inherit the infinite smoothness of the squared-exponential kernel from which they are built. A larger ν commits the reconstruction to a more slowly varying curve where no observation is near and narrows the predictive interval accordingly which is the reason the roughness is selected from the data rather than fixed in advance.

Table 2 sets out the seven structures with the short names used for them in what follows. Every structure except GP-2M-SA carries the four-term basis of trend and annual cycle alone, and the comparison between Two-scale Matérn 3/2 and GP-2M-SA therefore isolates the semiannual pair as the only difference between them.

Each structure was compared in three ways. The first is the Akaike information criterion (Akaike 1974), taken as twice the negative log marginal likelihood at the optimum plus twice the number of free parameters, the latter counted as the covariance hyperparameters together with the terms of the mean basis, both of which appear in Table 6. It rewards a structure that accounts for the whole record and penalises one that needs many parameters to do it. The second is the leave-one-out (LOO) residual

Table 2 The seven covariance structures compared in this section. The basis column gives the number of terms of (4) retained in $\mathbf{h}(t)$, four for the trend and the annual cycle and six when the semiannual pair is added

Structure	Basis	Covariance and what it represents
Matérn 1/2	4	$\sigma_f^2 \exp(-\tau/\ell)$, the continuous first-order autoregressive process. Paths are continuous but nowhere differentiable and one time scale is available
Matérn 3/2	4	A single component of (9). Paths are once differentiable, the roughness usually adopted for environmental series
Matérn 5/2	4	Equation (12). Paths are twice differentiable, the smoothest of the three Matérn forms tried
Rational quadratic	4	Equation (13), a scale mixture of squared-exponential kernels. It acts on a continuum of time scales and carries the longest memory of the seven
Two-scale Matérn 3/2	4	Both components of (9) with ℓ_1 and ℓ_2 free to separate into a short and a long memory
GP-2M-SA	6	The same two-scale kernel with the semiannual pair of (4) added to the mean function which moves a known deterministic cycle out of the covariance
Matérn 3/2 + quasi-periodic	4	One Matérn 3/2 component plus (14) with P fixed at one year which allows the shape of the seasonal cycle to change from year to year

which for a Gaussian process with a vague prior on $\boldsymbol{\beta}$ has the closed form

$$r_i = \frac{\left[\tilde{\mathbf{K}}^{-1}\boldsymbol{\zeta}\right]_i}{\left[\tilde{\mathbf{K}}^{-1}\right]_{ii}}, \qquad \tilde{\mathbf{K}} = \mathbf{K}_y + \mathbf{H}\boldsymbol{\Sigma}_\beta\mathbf{H}^\top, \tag{15}$$

so that all n residuals follow from a single matrix inversion rather than n refits (Rasmussen and Williams 2006). It measures the easiest prediction there is, a single day recovered while both of its neighbours remain in the fit. Neither of the first two describes a gap of several weeks, which is the situation this record presents. The third is therefore a set of 26 withheld-data experiments. Days were removed from the fit in blocks, each structure was refitted on what remained and the removed days were predicted. Fig. 3 sets out the design, which varies the size, the position and the shape of what is withheld. Groups A, B and D reuse the same four withheld fractions of 10, 20, 30 and 40 %. Group A withholds one interior block, centred at 35 % and again at 65 % of the record which gives eight experiments. Group B withholds a block at the end, so that each structure has to predict forward rather than interpolate which gives four. Group D withholds runs of one to ten days scattered through the record which gives four more. Groups C1 and C2 copy the shape of the two real outages instead of a fraction. C1 withholds 25 June to 7 September, the 75-day summer window, in each of the five years for which that window is complete and C2 withholds 27 November to 4 January, the 39-day winter window, likewise. The total of 26 follows from the design rather than from a target.

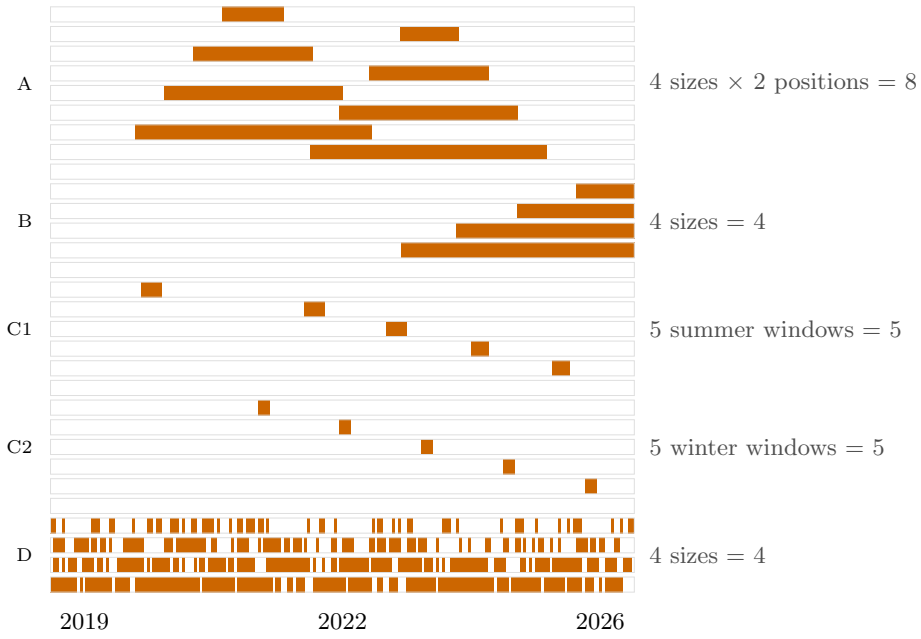


Fig. 3 The 26 withheld-data experiments. Each row is one experiment and spans the whole record from May 2019 to June 2026 and the shaded interval is the set of days withheld from the fit

Differences in leave-one-out error were tested by pairing the squared residuals day by day. The differences are serially correlated. The autocorrelation of the paired difference reaches 0.31 at a lag of two days for the Matérn 3/2 comparison and 0.42 for the Matérn 1/2 one, a remnant of the spring-neap cycle. An ordinary paired t test therefore overstates significance. Standard errors were therefore computed with a Newey-West estimator using a lag of eight days and the tests were repeated with a moving-block bootstrap.

3 Results

3.1 Tidal regime and constituents

The harmonic model explains 94.7 % of the variance of the hourly record. The largest constituents are K_1 (0.652 m), M_2 (0.563 m), O_1 (0.416 m), S_2 (0.272 m) and P_1 (0.181 m), giving a form factor

$$F = \frac{A_{K1} + A_{O1}}{A_{M2} + A_{S2}} = \frac{0.652 + 0.416}{0.563 + 0.272} = 1.28, \tag{16}$$

which places the station in the mixed, mainly semidiurnal class. The mean level is $Z_0 = 2.819$ m above station datum. The first overtide reaches 0.029 m, an $\mathrm{M}_4/\mathrm{M}_2$ ratio of 0.052.

Table 3 The thirty-three constituents fitted to the hourly record, ordered by amplitude. Phases are lags relative to 1 January 2020 and uncertainties are one standard error, and a phase is left out when its bootstrap spread exceeds 30°, which happens only for M_3

Const.	Speed (°/h)	Amp. (cm)	Phase (°)	Const.	Speed (°/h)	Amp. (cm)	Phase (°)
K_1	15.0411	65.24 ± 0.22	160.6 ± 0.2	μ_2	27.9682	2.22 ± 0.28	103.1 ± 7.8
M_2	28.9841	56.30 ± 0.43	280.9 ± 0.4	λ_2	29.4556	2.17 ± 0.28	145.8 ± 7.2
O_1	13.9430	41.62 ± 0.22	262.2 ± 0.3	$2SM_2$	31.0159	1.74 ± 0.18	319.0 ± 5.3
S_2	30.0000	27.16 ± 0.22	221.0 ± 0.4	SK_3	45.0411	1.66 ± 0.07	328.7 ± 2.5
P_1	14.9589	18.14 ± 0.22	183.6 ± 0.6	ν_2	28.5126	1.66 ± 0.28	203.8 ± 9.8
K_2	30.0821	10.81 ± 0.18	352.2 ± 0.9	T_2	29.9589	1.56 ± 0.22	154.7 ± 7.9
N_2	28.4397	8.83 ± 0.27	62.8 ± 1.9	ρ_1	13.4715	1.45 ± 0.13	202.0 ± 5.0
Q_1	13.3987	7.46 ± 0.14	50.1 ± 1.0	$2N_2$	27.8954	1.28 ± 0.28	195.3 ± 13.0
NO_1	14.4967	3.65 ± 0.21	333.1 ± 3.4	MN_4	57.4238	1.04 ± 0.05	246.7 ± 3.1
L_2	29.5285	3.09 ± 0.23	306.8 ± 4.2	$2MS_6$	87.9682	0.93 ± 0.05	187.3 ± 2.9
S_1	15.0000	3.06 ± 0.21	263.8 ± 3.9	$2Q_1$	12.8543	0.80 ± 0.13	196.8 ± 9.7
OO_1	16.1391	2.93 ± 0.22	232.7 ± 4.1	M_6	86.9523	0.80 ± 0.05	264.9 ± 3.4
J_1	15.5854	2.92 ± 0.15	58.4 ± 3.6	S_4	60.0000	0.65 ± 0.04	354.6 ± 4.1
M_4	57.9682	2.91 ± 0.06	98.7 ± 1.2	MK_4	59.0662	0.31 ± 0.05	190.6 ± 9.8
MS_4	58.9841	2.79 ± 0.05	36.3 ± 1.1	M_3	43.4762	0.11 ± 0.09	—
MK_3	44.0252	2.77 ± 0.13	12.6 ± 2.7	M_8	115.9364	0.10 ± 0.02	346.5 ± 13.0
MO_3	42.9271	2.29 ± 0.08	151.9 ± 2.1				

Table 3 lists all thirty-three constituents. Standard errors come from a moving-block bootstrap of the regression residuals with thirty-day blocks which is the same device used for the seasonal intervals and keeps the strong serial correlation of the residuals in the resampled series.

3.2 Flood dominance and duration asymmetry

Across 3,931 rises and 3,958 falls the mean rise duration is 6.26 h and the mean fall duration 8.12 h. The estuary is therefore strongly flood dominant, with a duration asymmetry of 1.86 h, or 13 % of the mean tidal period. Medians of 6.00 and 7.00 h give the same sign with a smaller difference of 1.00 h, which reflects the right-skewed distribution of fall durations rather than a change of direction. Fig. 4 shows a representative four-day interval and Fig. 5 the distributions of the two durations which overlap but are clearly offset.

The same asymmetry can be read from the overtide. Over the whole record the M_4/M_2 amplitude ratio is 0.052 and the relative phase $2\phi_{M2} - \phi_{M4}$ is 103.2°, which places the station in the flood-dominant half of the diagram of Friedrichs and Aubrey (1988) and agrees with the duration measurement made without any phase convention. Fig. 6 shows how the two quantities move. The ratio grows from 0.051 in 2020 to 0.055 in 2024 and falls back to 0.050 in 2025, while the relative phase stays within four degrees of its mean. The distortion therefore changes in strength rather than in character. The seasonal contrast is the larger of the two. Pooled over the record the wet season has a ratio of 0.054 against 0.050 in the dry season and a relative phase of 100.8° against 106.5°, and the wet season is the more distorted in every year for which both seasons survive quality control and the more flood dominant in four of those five years.

Table 4 Seasonal modulation of tidal constituents at Bang Pakong. The swing is the difference between the largest and smallest amplitude over the year in the modulated harmonic model. Intervals are 95 % moving-block bootstrap intervals

Constituent	Mean (cm)	Swing (cm)	95 % interval	Peak
M_2	56.44	7.36	5.64–9.41	Jul
N_2	8.85	1.31	0.51–2.71	Aug
O_1	41.63	2.75	1.82–3.96	Dec
K_2	10.93	3.30	2.63–4.23	Sep
M_4	2.96	0.70	0.46–1.01	Jul
MS_4	2.84	0.64	0.40–0.87	Jun
MN_4	1.06	0.38	0.19–0.62	Aug
MK_3	2.74	0.50	0.15–1.28	Apr
MO_3	2.37	0.57	0.30–1.05	Jul
M_6	0.83	0.31	0.11–0.60	Aug

3.3 Seasonal change in the distortion

Allowing annual modulation reduces the residual standard deviation from 17.98 to 17.62 cm. The seasonal swings with bootstrap intervals are given in Table 4 and the fitted annual envelopes in Fig. 7. Every interval excludes zero. The pattern is systematic, in that the shallow-water constituents swing by 22–37 % of their mean amplitude, two to three times the relative swing of M_2, and all of them reach their maximum between June and August.

The mean water level is 2.756 m in the wet season (June to October) against 2.904 m in the dry season (December to March), a difference of 14.8 cm. The season of strongest non-linear distortion is therefore also the season of shallowest water.

3.4 Change over the seven years

The annual M_2 amplitude declines from 58.10 cm in 2020 to 54.91 cm in 2025. Table 5 gives the annual values of six constituents together with the fitted rates. Fitting a straight line to the six annual values gives $-0.545 \pm 0.088\,\mathrm{cm\,yr^{-1}}$ (Fig. 8). Restricting the analysis to the dry season gives $-0.497 \pm 0.161\,\mathrm{cm\,yr^{-1}}$ and to the wet season $-0.488 \pm 0.160\,\mathrm{cm\,yr^{-1}}$. The three estimates agree to within their uncertainties, and uneven seasonal sampling can be ruled out as the cause of the decline. No other constituent shows a trend that is consistent in sign between the two seasons.

The obvious alternative explanation is imperfect nodal correction. Over 2020–2025 the lunar node moved through 263° of its 18.61-year cycle and the nodal factor f_{M2} changed by −3.57 %, in the same direction as the observed decline. However, the correction was applied in the fit and an error of 10 % in the nodal formulation would leave a spurious trend of only $-0.034\,\mathrm{cm\,yr^{-1}}$, sixteen times smaller than the observed value.

The duration asymmetry itself does not follow the amplitude down. Fig. 9 gives the mean rise and fall duration for each year, the fall exceeding the rise in every one of them. Both drift by about a fifth of an hour over the record and the difference

Table 5 Annual amplitudes (cm) of selected constituents and the fitted linear trend. Uncertainties are standard errors of the ordinary least-squares slope over the six annual estimates

	2020	2021	2022	2023	2024	2025	Trend (cm yr^{-1})
M_2	58.10	56.81	56.17	56.17	55.76	54.91	-0.545 ± 0.088
S_2	27.05	27.71	27.19	27.08	27.44	26.65	-0.084 ± 0.088
K_1	65.62	65.31	65.61	65.28	65.12	65.50	-0.042 ± 0.049
O_1	40.93	41.35	42.21	41.52	41.82	41.81	$+0.147 \pm 0.093$
M_4	2.94	2.90	2.92	3.01	3.05	2.76	-0.010 ± 0.026
M_6	0.78	0.79	0.72	0.82	0.78	0.98	$+0.030 \pm 0.019$

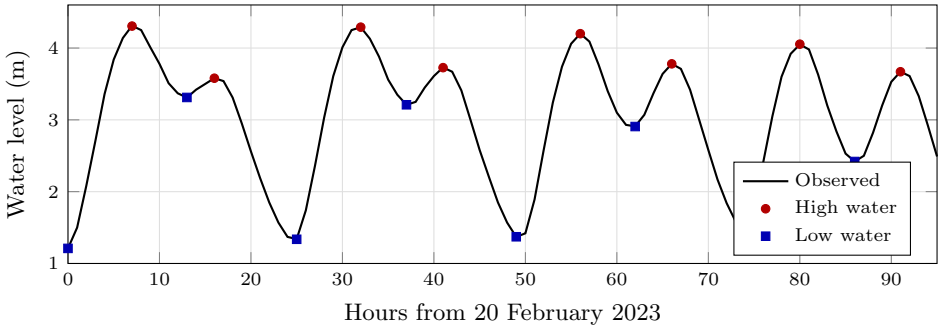


Fig. 4 A representative four-day interval of the hourly record with the identified high and low waters. The rise from low to high is visibly shorter than the following fall

between them moves between 1.39 and 2.24 h without any ordered progression and a weighted straight line through the six complete years gives 0.04 ± 0.05 h yr^{-1} which is not distinguishable from no change ($t = 0.9$). Nor is the asymmetry a seasonal quantity in the way the amplitudes are. Pooled over the record the wet season gives 1.80 ± 0.10 h and the dry season 1.83 ± 0.13 h. The overtide ratio in Fig. 6 and the duration asymmetry therefore carry different information and the second is the more stable of the two.

3.5 The best covariance structure

Table 6 compares the seven structures. Three results stand out. Path roughness barely matters for prediction but is penalised by the likelihood. Moving from Matérn 1/2 to 3/2 to 5/2 changes the leave-one-out error by less than 0.5 % (6.807, 6.833 and 6.836 cm) while AIC deteriorates steadily from -5415.6 to -5371.7. The daily level is driven by wind and pressure forcing that changes abruptly and the data prefer the rougher process.

The number of time scales matters. Two Matérn 3/2 components reduce the leave-one-out error to 6.670 cm and improve AIC by 79 units relative to the single component. The rational quadratic, which mixes scales continuously with one extra parameter, achieves almost the same fit. Its drawback appears in the trend, because

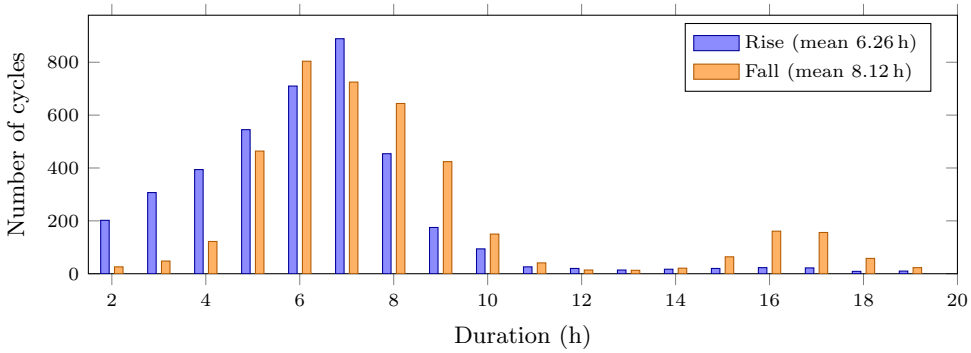


Fig. 5 Distribution of rise and fall durations over 3,931 and 3,958 tidal cycles. The two distributions overlap but the fall is systematically longer

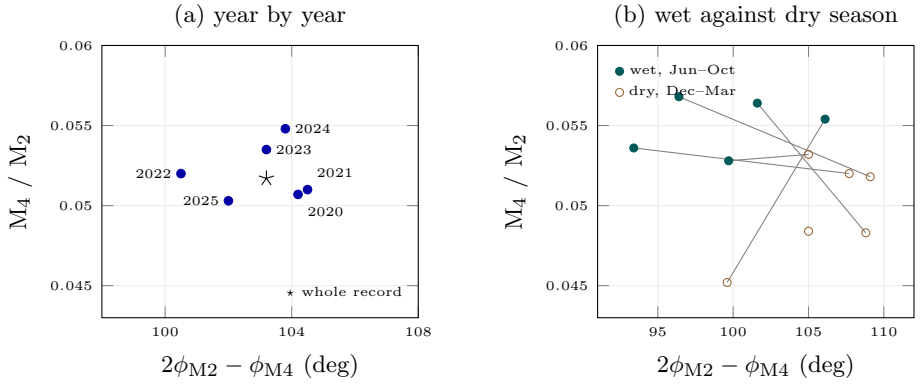


Fig. 6 Non-linear distortion in the coordinates of Friedrichs and Aubrey (1988). The horizontal axis is the relative phase of the first overtide, which is flood dominant between 0° and 180°, and the vertical axis is the M_4/M_2 amplitude ratio. (a) Separate harmonic fits to each complete year, with the whole-record value as a star. (b) The same fits split into wet and dry season and joined in pairs. The 2021 wet season is missing because of the 75-day outage in Fig. 2

the long tail of (13) absorbs low-frequency variance that would otherwise be attributed to the linear term and the standard error of the rate inflates to $\pm 3.41\,\mathrm{mm\,yr^{-1}}$, the largest of the seven.

No evidence was found for an evolving seasonal shape. The maximum-likelihood decay length of the quasi-periodic term ran to its upper bound of more than fifty years which describes a fixed annual cycle and the structure is worse than the two-scale kernel by both criteria.

Paired tests on the squared leave-one-out residuals are given in the last two columns of Table 6. After the Newey–West correction the two-scale kernel with a semiannual mean term is better than the four single-component structures and than the quasi-periodic structure, all with $p < 10^{-3}$. The correction reduces the test statistics by 13 to 26 % relative to the naive paired t test but overturns none of the conclusions. The moving-block bootstrap of the last column reduces them a little further and agrees

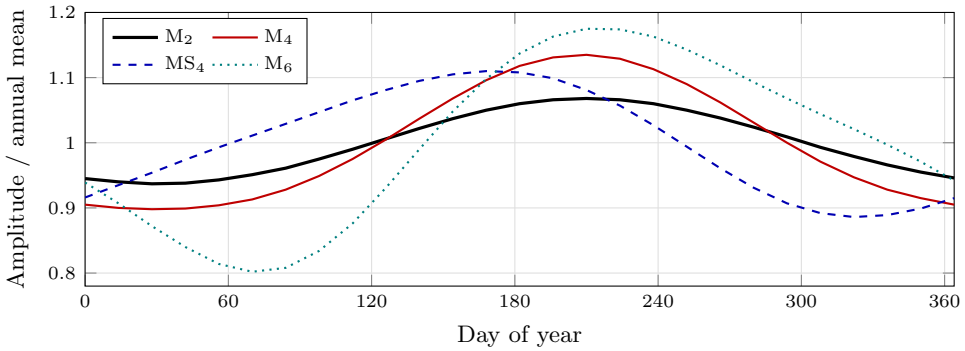


Fig. 7 Fitted annual envelopes of the amplitude of M_2 and of three shallow-water constituents, each normalised by its own annual mean. The shallow-water constituents vary two to three times more than M_2 and peak between June and August (days 150–240)

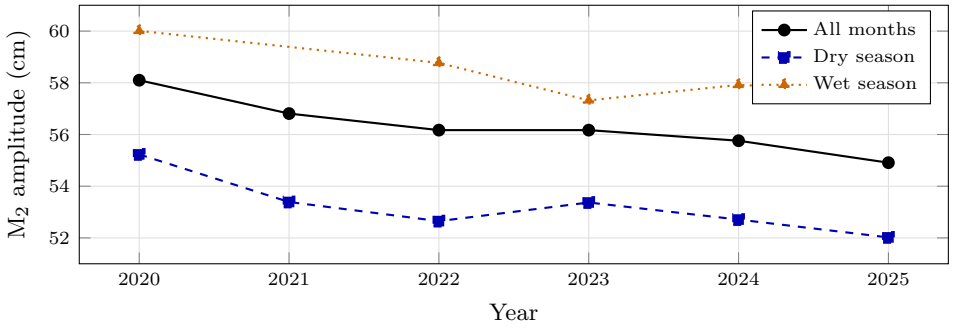


Fig. 8 Annual M_2 amplitude estimated from all months and from the dry and wet seasons separately. The three series decline at indistinguishable rates, which rules out uneven seasonal sampling as the cause

with the Newey–West result in every row. The single comparison that falls short of significance is against the same two-scale kernel without the semiannual basis term ($p = 0.78$). The semiannual term therefore does not help prediction close to the data but it improves AIC by 16 units by moving a deterministic cycle out of the covariance and into the mean function which matters where the kernel has no neighbouring observations to work with.

Fig. 10 shows the seven reconstructions of the same 39-day gap drawn on one axis, where the mean curves lie on top of one another and the largest separation across the gap is about the thickness of the plotted lines. The choice of kernel therefore changes the uncertainty attached to a filled value far more than it changes the value itself. Fig. 11 separates the structures in four ways. Panel (a) shows that at a lag of ten days the three single Matérn kernels have forgotten the record almost completely ($k/k(0) < 0.01$) while the rational quadratic, the two-scale and the quasi-periodic structures still retain 6–10 %, and that at thirty days only the rational quadratic and the quasi-periodic term keep any memory at all. Panel (b) shows that every structure saturates within about six days of the edge of the gap, at a plateau between 8.38

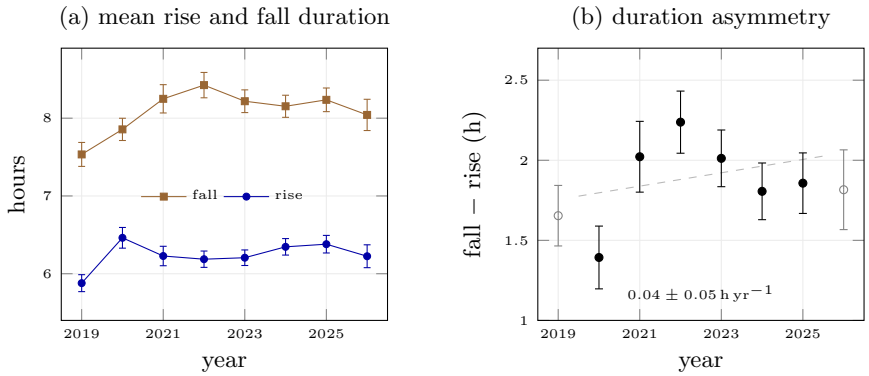


Fig. 9 Duration asymmetry through the record, from high and low waters identified directly in the hourly series. Error bars are one standard error of the mean. (a) The mean rise and fall durations for each year. (b) The difference between them, with the weighted straight line fitted to the six complete years. Open circles mark 2019 and 2026, which cover only part of a year and are excluded from the fit.

Table 6 Seven covariance structures fitted to the daily mean water level. The last three columns test the paired difference of squared leave-one-out residuals against GP-2M-SA. t_{HAC} and p use Newey–West standard errors at a lag of eight days, and t_{boot} repeats the test with a moving-block bootstrap of 30-day blocks. GP-2M-SA is the reference of that comparison. Its own row therefore carries no test statistic and no p value and a dash marks this rather than a missing result

Kernel	$\lvert\boldsymbol{\theta}\rvert$	basis	LOO (cm)	AIC	rate (mm yr^{-1})	t_{HAC}	p	t_{boot}
Matérn 1/2	3	4	6.807	−5415.6	2.58 ± 1.97	4.90	$< 10^{-3}$	4.20
Matérn 3/2	3	4	6.833	−5385.6	2.59 ± 1.75	5.47	$< 10^{-3}$	4.66
Matérn 5/2	3	4	6.836	−5371.7	2.60 ± 1.68	5.65	$< 10^{-3}$	4.82
Rational quadratic	4	4	6.705	−5457.1	2.30 ± 3.41	3.81	10^{-4}	3.26
Two-scale Matérn 3/2	5	4	**6.670**	−5464.7	2.52 ± 2.46	−0.28	0.78	−0.29
GP-2M-SA	5	6	6.671	**−5480.9**	2.44 ± 2.20	—	—	—
Matérn 3/2 + quasi-periodic	6	4	6.770	−5417.0	2.49 ± 1.60	5.17	$< 10^{-3}$	4.47

The basis column gives the number of terms in $\mathbf{h}(t)$, four for $[1, t, \sin 2\pi t, \cos 2\pi t]$ and six when the semiannual pair is added. GP-2M-SA is the structure used for the reconstruction.

and 9.82 cm. Panel (c) shows that the six alternatives stay within 2.7 cm of the GP-2M-SA mean, and that the quasi-periodic structure is the only one that departs in the opposite direction, because its seasonal term is still turning while the others have relaxed to the deterministic mean. Panel (d) places the two-scale structures in the lower left corner on both the leave-one-out and the Akaike criterion.

The full-record criteria say nothing directly about behaviour when data are actually missing. Table 7 gives the root-mean-square error over the withheld days in each of the 26 experiments of Fig. 3, averaged within group, and Table 8 the paired comparison of each structure with GP-2M-SA across the 26 cases. A positive Δ there means the alternative structure has the larger error, and GP-2M-SA is the reference of the comparison rather than a row of its own.

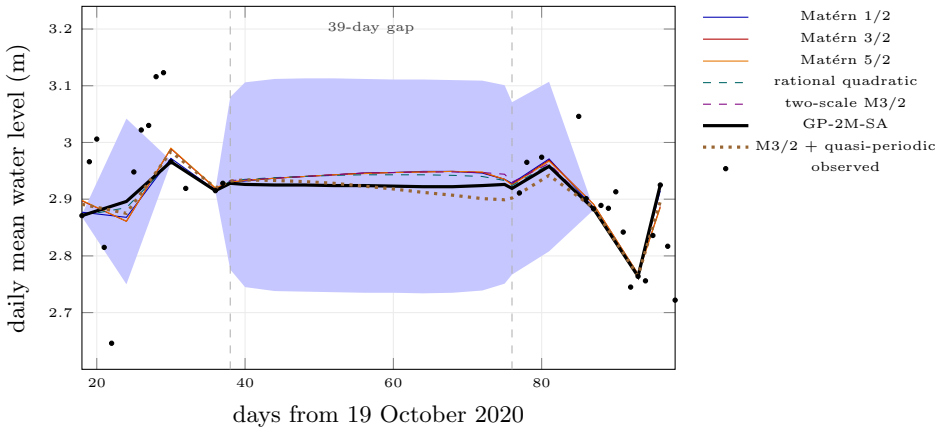


Fig. 10 The seven covariance structures reconstructing the same gap, marked by the dashed lines, with the observed daily levels as dots. All seven mean curves are drawn on one axis and they lie on top of one another. The shaded band is the 95 % predictive interval of GP-2M-SA alone, the other six being omitted so that the means remain visible

GP-2M-SA has the lowest error overall, 9.40 cm against 9.48 to 9.62 cm for the rest, and it is the best structure in three of the five groups and in 11 of the 26 individual experiments, and it is the better of the two in 20 or 21 of the 26 paired comparisons with each single Matérn kernel. The paired t test nevertheless rejects equality for none of the six competitors (p between 0.07 and 0.41), while the Wilcoxon signed-rank test rejects it for all three single Matérn kernels ($p = 0.027$, 0.009 and 0.009). The two tests disagree because the differences are not symmetric. In twenty of the experiments GP-2M-SA is ahead by 0.1 to 0.5 cm, but in three experiments of group C1 it is behind by 0.8 to 1.3 cm, and those three are heavy enough in both the mean and the variance to absorb the rest. A rank test, which ignores the size of a difference, still sees the majority.

The three exceptions are informative rather than accidental. Group C1 withholds the 75-day summer window, the longest of the gaps, and it is the one place where a seasonal term has to be carried across a large part of a season with no data to anchor it. The two structures that hold such a term explicitly, GP-2M-SA through its semiannual basis and the quasi-periodic kernel through its one-year component, are the two worst in that group at 8.59 and 8.68 cm while the rational quadratic, which has no seasonal term but the longest memory of the seven (Fig. 11a), is the best at 8.11 cm. For the shorter winter gaps of group C2 the ordering returns. GP-2M-SA is first at 9.54 cm, the quasi-periodic structure second at 9.61 cm and the rest at 9.90 cm or more.

The practical conclusion is that the covariance structure matters least for the quantity this paper needs from it. Across the 26 experiments the spread of mean error between the best and the worst of the seven is 0.22 cm against a typical error near 9.5 cm, and no structure is significantly worse than GP-2M-SA on a paired t test. GP-2M-SA is kept for the reconstruction because it is the best on the full-record criteria

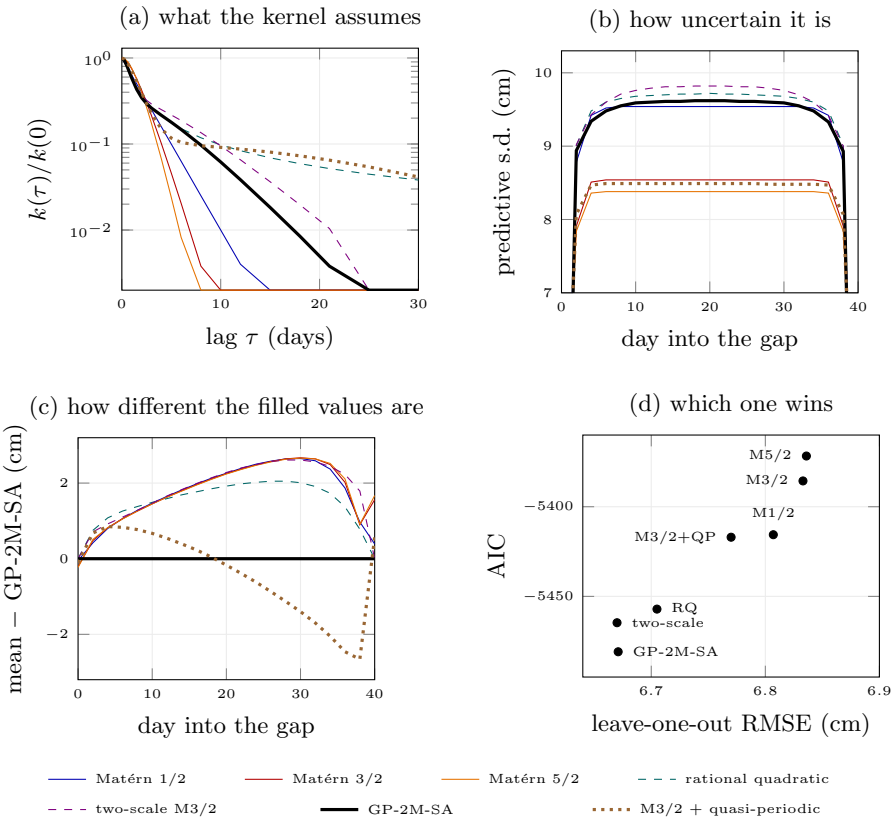


Fig. 11 Where the seven covariance structures actually differ. (a) The fitted correlation function on a logarithmic scale. (b) The predictive standard deviation through the 39-day gap of Fig. 10, with the vertical axis starting at 7 cm so that the plateaus separate. (c) The reconstructed mean minus the GP-2M-SA mean. (d) Leave-one-out RMSE against AIC, lower being better on both axes

and on average here, but none of the tidal quantities reported here would change materially under any of the other six.

3.6 Effect of the reconstruction

Repeating the harmonic analysis on the reconstructed series changes no amplitude by more than 0.12 cm. M_2 moves from 56.30 to 56.31 cm, K_1 from 65.24 to 65.12 cm, S_2 from 27.16 to 27.23 cm and M_4 from 2.913 to 2.929 cm. This is the expected behaviour because least squares already uses the observed hours optimally and the reconstruction adds no independent information. Filling gaps before a harmonic analysis is therefore unnecessary and would give imputed values the same weight as measurements.

The duration analysis benefits differently. Reconstruction recovers 449 rises and 467 falls that were previously lost to gaps, an increase of 11 % in the sample. The mean durations shorten slightly to 6.12 and 7.96 h because the reconstructed series is smoother than the observations but the asymmetry is essentially unchanged at 1.84 h

Table 7 Withheld-data experiments. Mean root-mean-square error in centimetres over the withheld days, by group of Fig. 3 and over all 26 experiments. The number of experiments in each group is given under the heading and bold marks the lowest mean in each column

Kernel	A (8)	B (4)	C1 (5)	C2 (5)	D (4)	all (26)
Matérn 1/2	9.99	10.28	8.31	9.95	9.13	9.570
Matérn 3/2	10.00	10.28	8.31	10.01	9.25	9.605
Matérn 5/2	10.01	10.28	8.31	10.03	9.31	9.620
Rational quadratic	9.93	10.32	**8.11**	9.90	**8.96**	9.484
Two-scale Matérn 3/2	9.95	10.30	8.28	9.90	9.00	9.525
GP-2M-SA	**9.76**	**9.99**	8.59	**9.54**	**8.96**	**9.404**
Matérn 3/2 + quasi-periodic	9.81	10.06	8.68	9.61	9.21	9.500

Table 8 Paired comparison of each covariance structure with GP-2M-SA across the 26 experiments of Table 7. Δ is the mean difference in root-mean-square error, t and p_t the paired t test, p_W the Wilcoxon signed-rank test and the last column the number of experiments GP-2M-SA wins

Kernel	Δ (cm)	t	p_t	p_W	GP-2M-SA better in
Matérn 1/2	+0.167	1.49	0.148	**0.027**	21 of 26
Matérn 3/2	+0.201	1.77	0.089	**0.009**	20 of 26
Matérn 5/2	+0.216	1.88	0.071	**0.009**	20 of 26
Rational quadratic	+0.080	0.83	0.414	0.105	17 of 26
Two-scale Matérn 3/2	+0.121	1.06	0.299	0.063	18 of 26
Matérn 3/2 + quasi-periodic	+0.096	1.06	0.302	0.075	17 of 26

against 1.86 h. The flood dominance reported above is therefore not an artefact of which cycles happen to be complete.

4 Discussion

The seasonal contrasts in the shallow-water constituents and in mean water level are consistent with a single mechanism. Frictional distortion of a tidal wave scales with the ratio of tidal amplitude to water depth and a lower mean level increases the relative importance of friction and transfers more energy into the overtides (Speer and Aubrey 1985). At Bang Pakong the mean level is lowest in the wet season, when the northeast monsoon is not piling water into the inner Gulf and it is then that M_4, MN_4, MO_3 and M_6 are largest. Increased river discharge during the same months acts in the same direction by raising the effective friction (Guo et al. 2023; Sassi and Hoitink 2013).

The magnitude is worth noting. Müller et al. (2014) report seasonal variation of M_2 of 5–10 % for coastal regions worldwide, attributed mainly to stratification. The 13 % found here is at the upper end of that range, and the 22–37 % swings of the shallow-water constituents are considerably larger. This is consistent with the expectation that overtides, being generated by the non-linear terms, respond more sensitively than the

constituents that force them. Seasonal movement of the mean level in the inner Gulf is itself well documented and is driven by the monsoon (Saramul and Ezer 2014) which allows the two halves of this explanation to be checked independently.

The regime at this station complicates the usual reading of those numbers. With $F = 1.28$ the diurnal constituents are as large as the semidiurnal ones, and in such a mixture asymmetry is generated not only by the overtides of M_2 but also by triads that couple the diurnal and the semidiurnal bands, so that the M_2/M_4 pair no longer describes the whole of it (Song et al. 2011). Nidzieko (2010) met this by taking the sampling skewness of the time derivative of water level, a measure that needs no phase convention and does not care how many constituents are significant. The duration asymmetry used here is in the same spirit and the contrast between Figs. 6 and 9 is what that reasoning would lead one to expect. The overtide ratio and the relative phase answer to the semidiurnal non-linearity alone and move with the season while the duration asymmetry aggregates every contribution and stays where it is. The consequences of mixing can be sharper still. In a creek of the same broad regime in the Persian Gulf, Hosseini et al. (2016) found the K_1–O_1–M_2 triad reversing the sense of the asymmetry within the fortnightly cycle, ebb dominant at springs and flood dominant at neaps. A single whole-record figure can therefore conceal structure and at a mixed station the overtide diagram and a convention-free measure are best reported together rather than one standing in for the other.

Flood dominance has three consequences at this site. First, the time from low water to high water is shorter than the time from high to low by nearly two hours, which shortens the interval available for warning before a high-water peak. Second, flood dominance favours net landward transport of fine sediment, the link between asymmetry and estuarine form set out by Dronkers (1986), a tendency towards infilling that interacts with upstream sediment supply, channel dredging and sand extraction. Third, a stronger, shorter flood promotes salt intrusion which matters directly for irrigation and municipal supply in Chachoengsao Province.

The seasonal result sharpens the first of these. Non-linear distortion is strongest between June and August which is also the season of highest river discharge and of monsoon rainfall. The tidal and fluvial contributions to high water are therefore not independent in time and compound flood risk in this estuary should be assessed with a seasonally varying tidal model rather than a constant one.

Table 9 places the station beside the studies this paper draws on. The entries are deliberately qualitative, because none of those papers reports a form factor, an M_4/M_2 ratio or a duration asymmetry for its own site in a form that can be set against the numbers here. A column of matched figures would be an invention rather than a comparison. What the table does show is that the three mechanisms invoked above (stratification, river discharge and channel geometry) are each supported by an independent study elsewhere, and that the combination of a mixed, mainly semidiurnal regime with strong flood dominance and a seasonally varying overtide has not previously been documented for a Thai estuary.

Seven years is long enough to measure the M_2 trend and too short to interpret it. The decline of $0.55 \pm 0.09\,\mathrm{cm\,yr^{-1}}$ is close to one per cent of the mean amplitude in each year, and it is the same within uncertainty in the wet and the dry season.

Table 9 The Bang Pakong station set beside the studies used in the discussion. The third column reproduces what each study reports for its own system. The studies use different variables and different platforms and the entries are not directly comparable quantities

Study	System	Reported findings
This paper	Bang Pakong estuary, Gulf of Thailand	Mixed, mainly semidiurnal, $F = 1.28$. Flood dominant by 1.86 h. $M_4/M_2 = 0.052$. M_2 falling 0.55 cm yr^{-1}. Seasonal M_2 swing 13 %, shallow-water constituents 22–37 %
Gong et al. (2016)	Huangmaohai estuary, Pearl River Delta	Funnel-shaped estuary with mixed semidiurnal tides. Interaction among the residual flow and the tidal constituents produces prevailing ebb dominance, which river flow enhances
Hosseini et al. (2016)	Tidal creek, Bushehr Port, Persian Gulf	Mixed, mainly semidiurnal. The K_1–O_1–M_2 triad makes the creek ebb dominant at spring tides and flood dominant at neaps
Guo et al. (2023)	Schematized large estuary, motivated by the Changjiang and the Amazon	River discharge lowers the M_4 amplitude landward and raises it seaward. M_4 energy reaches a transitional maximum when river discharge and tidal mean discharge at the mouth are comparable
Mandal (2024)	Gulf of Khambhat, Arabian Sea	M_2 currents strongest in September under strong stratification. M_4 currents between 3.15 and 16.50 cm s^{-1}
Müller et al. (2014)	Global coastal and polar ocean	Seasonal change of the M_2 amplitude of 5–10 % in coastal and polar regions, attributed to stratification on the shelf and to sea ice
Pongsiri et al. (2020)	Eleven gauges, Gulf of Thailand	Increasing trends in almost all tidal levels and in mean sea level
Saramul and Ezer (2014)	Coast of Thailand	Sea level along the coast varies with the monsoon and is further affected by extreme land subsidence and by earthquakes

No plausible error in the nodal formulation accounts for more than a sixteenth of it, which makes it a property of the record rather than of the processing. Whether it is a property of the estuary is a different question. Analyses of the longer gauge records in the Gulf of Thailand report rising tidal levels and rising mean sea level (Pongsiri et al. 2020), which does not contradict a falling M_2, since those levels are dominated by the diurnal constituents and by the mean level while what falls here is a single semidiurnal constituent. It does show, though, that the present result cannot be carried over to the inner Gulf on seven years at one gauge.

Where such a change would originate is equally open. Cui et al. (2019) show that the large diurnal tide of the Gulf of Thailand is not a resonance of the gulf itself but a passive response to the quarter-wavelength resonance of the South China Sea, and the tide arriving at the mouth of this estuary is set largely outside it. That result

concerns the diurnal band rather than M_2, and the semidiurnal response of the gulf has not been described in the same terms. A regional origin for the decline found here can therefore be neither supported nor excluded. The local alternatives are the more tractable. Dredging, sand extraction and channel change all alter the depth and the friction that set the amplitude at a station this far inside an estuary, and a bathymetric record covering the same years would test them directly. Of everything that could be added to this analysis, a water-level record longer than the 18.6-year nodal cycle would be worth the most.

Two points of method deserve emphasis before those limitations. Equation (2) matters because it is a property of the frequency set rather than of this particular record. Any study that fits annual modulation of K_1, S_2 or P_1 while retaining S_1 and T_2 is fitting a rank-deficient model, and the resulting amplitudes are arbitrary. The symptom is not always as obvious as the 9.9 m amplitude obtained here. With slightly different sampling the two collinear columns may produce plausible-looking but meaningless values. Either the modulation terms or the coincident constituents must be dropped and which choice is appropriate depends on whether the seasonal variation or the separate constituent is of interest.

The covariance comparison carries a message of its own. Across the 39-day winter gap the seven structures place the mean level within 2.7 cm of one another, which at the scale of Fig. 10 is about the thickness of the plotted lines, while the predictive standard deviation they attach to that level differs by 17 %, from 8.38 to 9.82 cm. The choice of covariance function is therefore mostly a statement about how much is claimed to be known rather than about what is claimed. For a record used to count tidal cycles this is a convenient outcome, because a turning point is located from the mean curve and not from the interval around it, which is why the duration statistics are insensitive to a decision that the model-selection criteria of Table 6 separate cleanly. A record used for storm surge or for extreme levels would weigh the two the other way and there the 17 % spread would matter more than the 2.7 cm. The comparison also shows what a reconstruction of this kind cannot do. Every structure relaxes to the deterministic mean function within about six days of the edge of a gap and beyond that distance the filled values carry the seasonal curve rather than any information from the record itself.

Three limitations should be stated plainly. First, the record covers seven years, which is less than half a nodal cycle and the apparent M_2 decline cannot be separated from an oscillation of decadal period. A sustained decline at the observed rate is not physically plausible over long periods. Changes of this size in the main constituents have been reported at gauges worldwide and are not always monotonic (Woodworth 2010) which is a further reason to treat seven years as a snapshot. Second, an instrumental cause cannot be dismissed, although a uniform loss of sensor gain is not consistent with the record. Such a loss would scale every constituent alike and a rate matching the M_2 decline would require K_1 to fall by about $0.63\,\mathrm{cm\,yr^{-1}}$ against the observed -0.042 ± 0.049 of Table 5. A frequency-dependent change such as progressive restriction of the stilling well, would damp the semidiurnal band more than the diurnal one and remains possible. Comparison with a neighbouring gauge is the natural test, the more so because parts of the inner Gulf coast are subsiding fast enough to

move a gauge relative to the water it measures (Saramul and Ezer 2014). Finally, the analysis uses water level alone. Duration asymmetry of the surface elevation is related to, but not identical with, the velocity asymmetry that controls sediment transport (Gong et al. 2016). Current measurements would be required to close that step.

5 Conclusions

Seven years of ten-minute observations show that the Bang Pakong estuary is strongly flood dominant with a rise that is on average 1.86 h shorter than the fall. The nonlinear distortion responsible for this asymmetry is not constant. The shallow-water constituents vary seasonally by 22–37 %, two to three times more than M_2, and peak between June and August, when the mean water level is lowest and river discharge highest. Over the seven years the M_2 amplitude declined by $0.55 \pm 0.09\,\mathrm{cm\,yr^{-1}}$, consistently in both seasons and by an amount that nodal mis-correction cannot explain, although a longer record and a neighbouring station are needed before this can be interpreted as a change in the estuary itself.

Because 6.4 % of the hourly record is missing, the non-tidal level was reconstructed with a Gaussian process whose covariance structure was selected from seven candidates rather than assumed. The reconstruction recovers 11 % more complete tidal cycles for the duration analysis while changing no harmonic amplitude by more than 0.12 cm and the seven structures agree on the filled level to within 2.7 cm while differing by 17 % in the uncertainty they attach to it. A methodological consequence of general interest is that annual modulation of K_1, S_2 and P_1 is not identifiable alongside S_1 and T_2, and must be excluded from such analyses.

Acknowledgements. The author thanks the Marine Department of Thailand for the water-level records used in this study, and the members of the Eastern Tech Fusion Research Laboratory, Department of Mathematics, Faculty of Science, Burapha University, for their support.

Statements and Declarations

Funding. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests. The author declares no competing interests.

Data availability. The water-level records were provided by the Marine Department of Thailand which should be credited as the source of the observations. The record itself, the analysis code and every table reported here are openly available at https://github.com/Pattharaporn-Aon/BangPakong_Tidal_Asymmetry, together with a notebook that reproduces the analysis from the raw file.

Clinical trial number. Not applicable.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705

Cui X, Fang G, Wu D (2019) Tidal resonance in the Gulf of Thailand. Ocean Science 15(2):321–331. https://doi.org/10.5194/os-15-321-2019

Doodson AT (1921) The harmonic development of the tide-generating potential. Proceedings of the Royal Society of London Series A 100(704):305–329. https://doi.org/10.1098/rspa.1921.0088

Dronkers J (1986) Tidal asymmetry and estuarine morphology. Netherlands Journal of Sea Research 20(2-3):117–131. https://doi.org/10.1016/0077-7579(86)90036-0

Foreman MGG (1977) Manual for tidal heights analysis and prediction. Pacific Marine Science Report 77-10, Institute of Ocean Sciences, Patricia Bay, Sidney, British Columbia, Canada

Friedrichs CT, Aubrey DG (1988) Non-linear tidal distortion in shallow well-mixed estuaries: a synthesis. Estuarine, Coastal and Shelf Science 27(5):521–545. https://doi.org/10.1016/0272-7714(88)90082-0

Gong W, Schuttelaars H, Zhang H (2016) Tidal asymmetry in a funnel-shaped estuary with mixed semidiurnal tides. Ocean Dynamics 66(5):637–658. https://doi.org/10.1007/s10236-016-0943-1

Guo L, Zhu C, Cai H, et al (2023) River flow induced nonlinear modulation of M4 overtide in large estuaries. Estuaries and Coasts 46(4):925–940. https://doi.org/10.1007/s12237-023-01183-0

Hosseini ST, Chegini V, Sadrinasab M, et al (2016) Tidal asymmetry in a tidal creek with mixed mainly semidiurnal tide, Bushehr Port, Persian Gulf. Ocean Science Journal 51(2):195–208. https://doi.org/10.1007/s12601-016-0017-9

Mandal S (2024) Observed seasonality of M2 and M4 tidal currents in the Gulf of Khambhat using high-frequency radars. Ocean Dynamics 74(3):251–265. https://doi.org/10.1007/s10236-024-01601-2

Müller M, Cherniawsky JY, Foreman MGG, et al (2014) Seasonal variation of the M_2 tide. Ocean Dynamics 64(2):159–177. https://doi.org/10.1007/s10236-013-0679-0

Nidzieko NJ (2010) Tidal asymmetry in estuaries with mixed semidiurnal/diurnal tides. Journal of Geophysical Research: Oceans 115(C8):C08006. https://doi.org/10.1029/2009JC005864

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

Pongsiri N, McNeil R, Chuai-Aree S (2020) Trends in tidal levels and mean sea level in the Gulf of Thailand. Ocean Science Journal 55(4):495–503. https://doi.org/10.1007/s12601-020-0039-1

Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA

Saramul S, Ezer T (2014) Spatial variations of sea level along the coast of Thailand: Impacts of extreme land subsidence, earthquakes and the seasonal monsoon. Global and Planetary Change 122:70–81. https://doi.org/10.1016/j.gloplacha.2014.08.012

Sassi MG, Hoitink AJF (2013) River flow controls on tides and tide-mean water level profiles in a tidal freshwater river. Journal of Geophysical Research: Oceans 118(9):4139–4151. https://doi.org/10.1002/jgrc.20297

Schureman P (1958) Manual of Harmonic Analysis and Prediction of Tides. Special Publication No. 98, U.S. Coast and Geodetic Survey, Washington, DC, USA

Song D, Wang XH, Kiss AE, et al (2011) The contribution to tidal asymmetry by different combinations of tidal constituents. Journal of Geophysical Research: Oceans 116(C12):C12007. https://doi.org/10.1029/2011JC007270

Speer PE, Aubrey DG (1985) A study of non-linear tidal propagation in shallow inlet/estuarine systems Part II: Theory. Estuarine, Coastal and Shelf Science 21(2):207–224. https://doi.org/10.1016/0272-7714(85)90097-6

Woodworth PL (2010) A survey of recent changes in the main components of the ocean tide. Continental Shelf Research 30(15):1680–1691. https://doi.org/10.1016/j.csr.2010.07.002

Xu K, Hu S, Jin S, et al (2024) Reconstruction of geodetic time series with missing data and time-varying seasonal signals using Gaussian process for machine learning. GPS Solutions 28(2):79. https://doi.org/10.1007/s10291-024-01616-8