Prediction of tidal currents using Bayesian machine learning

Dripta Sarkar [*], Michael A. Osborne, Thomas A.A. Adcock

Department of Engineering Science, University of Oxford, Oxford, OX13PJ, United Kingdom

A B S T R A C T

We propose the use of machine learning techniques in the Bayesian framework for the prediction of tidal currents. Computer algorithms based on the classical harmonic analysis approach have been used for several decades in tidal predictions, however the method has several limitations in terms of handling of noise, expressing uncertainty, capturing non-sinusoidal, non-harmonic variations. There is a need for principled approaches which can handle uncertainty and accommodate noise in the data. In this work, we use Gaussian processes, a Bayesian non-parametric machine learning technique, to predict tidal currents. The probabilistic and non-parametric nature of the approach enables it to represent uncertainties in modelling and deal with complexities of the problem. The method makes use of kernel functions to capture structures in the data. The overall objective is to take advantage of the recent progress in machine learning to construct a robust algorithm. Using several sets of field data, we show that the machine learning approach can achieve better results than the traditional approaches.

1. Introduction

Tidal waves are produced by changes in the gravitational forces of the sun and the moon. Prediction of tidal currents are necessitated by practical requirements like navigation, protection from flooding, coastal management to recent developments of energy extraction. Theoretical understanding of the tidal phenomenon began with Newton pioneering the gravitational theory and then later, Laplace deriving the expression for the tidal potential. There have been many advances in methodologies for tidal analysis since then. The most widely used method is that of the harmonic analysis (HA), where the observed tidal variations are considered as a resultant of various periodic components of known frequencies, with the amplitudes and phases determined using the least-squares fitting procedure. Computer codes based on HA have been used for decades for the prediction of tidal heights (1-D) and currents (2-D). Over the years various advances have been made to HA approach [see e.g. Pawlowicz et al., 2002; Foreman et al., 2009; Leffler and Jay, 2009]. Other techniques include tidal spectroscopy (Munk and Cartwright, 1966), and response method for unified tide and surge prediction (Cartwright, 1968), however they have not been widely adopted. HA has been extensively used in the analysis of stationary tidal (height) records, providing insights into the tidal dynamics. However, there are several shortcomings of this methodology. One of the challenging tasks in HA is the selection of tidal constituents, which if inaccurate can lead to overfitting of data or numerical issues (Jay and Flinchem, 1999). Appropriate

modelling of noise is another issue. In tidal analysis, signals which do not contribute to the tidal variations are classified as 'noise. In reality, there can be cases where the non-tidal signal is much stronger than the tidal e.g. the occurrence of a stormy event, and many of such effects are non-harmonic. It is difficult to incorporate such effects in the tidal HA formulation. In general, the technique is not suitable for application to non-stationary data (Jay and Flinchem, 1999). HA is also incapable of modelling the spatial variability of tides – this is not a big issue in modelling tidal heights which changes slowly in space, however tidal currents can vary sharply within short distances due to changes in bathymetry and topography. As tides move into shallow waters, they are distorted resulting in overtides (higher harmonics of principal constituents) and compound tides (interaction between different constituents). Such interactions can lead to asymmetry in the flood and ebb magnitudes of the current, depending on the phase relationship (Friedrichs and Aubrey, 1988). In HA, nonlinear characteristics are incorporated with the inclusion of shallow water constituents, some of which may need to be inferred, and such operations are often difficult. Even more complexities can result near headlands (Geyer, 1993), where complex flow structures can result in additional frequencies, which are not necessarily sinusoidal. In relation to the uncertainty estimation, HA generates confidence intervals for the current ellipse parameters (which are often large (Leffler and Jay, 2009)). However, in a lot of practical applications it is more useful to generate confidence interval estimates directly in the time domain.

In this work, we present a novel approach to predict tidal currents using probabilistic machine learning techniques in the Bayesian framework, which provide principled approaches for dealing with uncertainty, and can tackle the challenges of real world data (Roberts et al., 1984). Uncertainty could be introduced in many forms - ranging from measurement noise to uncertainty in the parameters of the model, and the mathematics of probability enables expressing the uncertainties (Ghahramani, 2015). Bayesian modelling approaches have been widely used in different disciplines e.g. geostatistics (Matheron, 1973), meteorology (Thompson, 1956), economics (Kim and Nelson, 1999), spatial statistics (Ripley, 2005, Rasmussen and Williams), machine learning (Rasmussen and Williams).

Gaussian process (GP), a Bayesian non-parametric approach, have been shown to be well-suited in solving a variety of time-series modelling problems (Roberts et al., 1984) and in this work we pursue this methodology to model tidal current data. We introduce the application of Bayesian machine learning to the tidal current prediction problem which can address some of the shortcomings associated with traditional techniques -

- modelling nonlinear interactions not captured in the HA especially at locations of fast tidal currents
- accommodating uncertainties of all forms e.g. noise is included directly in the mathematical formulation
- modelling non-harmonic variations in the short-term resulting from meteorological effects, barotropic to baroclinic conversion.
- generating confidence intervals directly in the time-domain.

The method can be used for the prediction from any generic tidal current time-series data. We show that the machine learning approach can produce better predictions than the HA even in cases where the latter is considered to be good (achieve good accuracy). An initial report on the novel machine learning approach to tidal currents was made in (Sarkar et al., 2016) where analysis was performed on tidal current data from a numerical model. In this work we provide a detailed description of the methodology with extensive discussions and analysis with real world datasets as well as present new approaches to model short and strongly contaminated datasets.

A brief overview of tides and HA is provided in the next section, followed by an introduction to GP regression. We then analyze long tidal current time series data using maximum a-posteriori and short tidal time-series data using Monte Carlo Markov chain techniques, and compare the results with classical HA approach. The work potentially opens up application of machine learning to other problems in tidal analysis which are otherwise not possible using traditional techniques.

2. Tides and harmonic analysis

Based on potential field theory, forces due to the sun and the moon produce hundreds of tidal constituents with distinct frequencies. Nonlinear interaction of the astronomical tidal components produces secondary tides known as overtides (higher harmonics) or compound tides (interaction between various tidal constituents). Let us consider a time series: $y(t)$, $t = t_1, t_2, ..., t_M$, where the observation times are regularly spaced at an interval Δt. The model equation with N constituents can be expressed as

$$y(t) = \sum_{k=1}^{N} \left(a_k^+ e^{i\omega_k(t-t_0)+iV_k} + a_k^- e^{-i\omega_k(t-t_0)-iV_k} \right) + c_0 + c_1(t-t_0) \tag{1}$$

where c_0 is some offset and c_1 indicate the trend, while the term inside the summation indicate the variation of the constituents with a_k^+ and a_k^- and being the unknown complex amplitudes, ω_k the angular frequency and V_k is some astronomical argument. Note, $y(t)$ is real if modelling tidal heights, while for tidal currents it is a complex variable: $y(t) = u(t) +$

$iv(t)$, and as such equation (1) can be expressed in matrix form as:

$$\begin{bmatrix} e^{i\phi_1^{(1)}} & . & . & e^{i\phi_1^{(N)}} & e^{-i\phi_1^{(1)}} & . & . & e^{-i\phi_1^{(N)}} & 1 & (t_1-t_0) \\ e^{i\phi_2^{(1)}} & . & . & e^{i\phi_2^{(N)}} & e^{-i\phi_2^{(1)}} & . & . & e^{-i\phi_2^{(N)}} & 1 & (t_2-t_0) \\ e^{i\phi_3^{(1)}} & . & . & e^{i\phi_3^{(N)}} & e^{-i\phi_3^{(1)}} & . & . & e^{-i\phi_3^{(N)}} & 1 & (t_3-t_0) \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ e^{i\phi_M^{(1)}} & . & . & e^{i\phi_M^{(N)}} & e^{-i\phi_M^{(1)}} & . & . & e^{-i\phi_M^{(N)}} & 1 & (t_M-t_0) \end{bmatrix} \begin{bmatrix} a_1^+ \\ . \\ . \\ a_N^+ \\ a_1^- \\ . \\ . \\ a_N^- \\ c_0 \\ c_1 \end{bmatrix}$$

$$= \begin{bmatrix} u_1 + iv_1 \\ u_2 + iv_2 \\ u_3 + iv_3 \\ . \\ . \\ . \\ u_M + iv_M \end{bmatrix}$$

where $\phi_m^{(k)} = \omega_k(t_m - t_0) + V_k(t_m)$. The solutions are determined by minimizing some function of the residual $(Ta - y)$, where $y = [y(t_1), y(t_2),...,y(t_m)]'$, $a = [a_1^+, a_2^+, ..., a_N^+, a_1^-, a_2^-, ..., a_N^-, c_0, c_1]'$ and T is a $M \times 2N + 2$ of linear and sinusoidal basis functions evaluated at the observation times. In case of the ordinary least squares (OLS) approach the objective function to be minimized can be expressed as $\left\| Ta - y \right\|^2$, and the solutions are determined as $a = (T^*T)^{-1}T^*y$ where superscript $*$ indicates the conjugate transpose of the matrix. However, a shortcoming of the OLS method is its sensitivity to non-tidal variations, as it can overfit such effects while trying to minimize the residual error (Leffler and Jay, 2009). The latest codes based on the HA uses the 'Iteratively Reweighted Least Squares' algorithm which reduces the influence of the non-tidal effects and the solution in this case is obtained as $a = (T^*WT)^{-1}T^*Wy$, where W is some weighting matrix which is
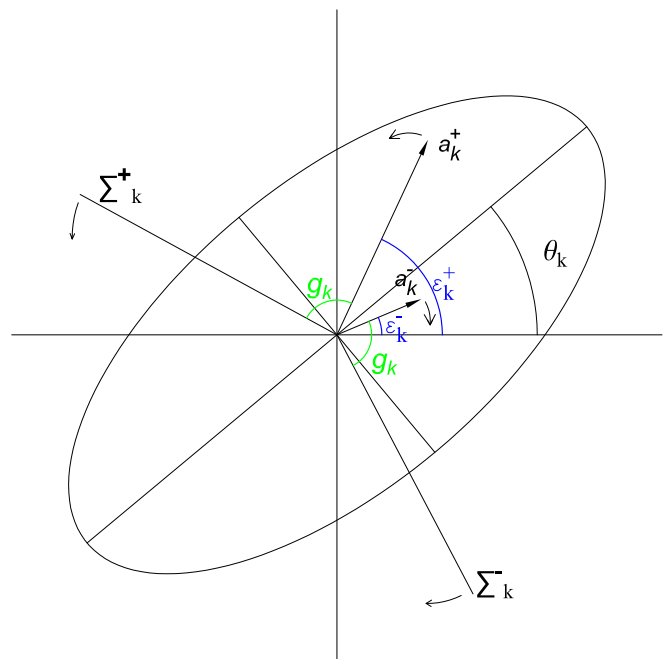

Fig. 1. For a particular tidal constituent k, the rotating vectors with amplitude a_k^+ and a_k^- are considered to be generated by two different fictitious stars $\sum_k^+$ and $\sum_k^-$, rotating in the counterclockwise and clockwise direction respectively, at a speed same as that of the constituents. The constant phase angle by which the rotating vectors lead or lag behind their respective stars is known as the greenwich phase g_k.

determined iteratively (see (Leffler and Jay, 2009) for details). The standard parameters of the particular constituent k are then computed as

$$L_k = \left| a_k^+ \right| + \left| a_k^- \right|, \quad l_k = \left| a_k^+ \right| - \left| a_k^- \right|$$

$$\theta_k = \left(\frac{\alpha_k^+ + \alpha_k^-}{2} \right), \quad g_k = v_k - \left(\frac{\alpha_k^+ - \alpha_k^-}{2} \right)$$

where L_k and l_k are the semi-major and semi-minor axis of the ellipse respectively, $\alpha_k^+ = ang(a_k^+)$, $\alpha_k^- = ang(a_k^-)$, θ_k is the angle of inclination of the northern semi-major axis counter-clockwise from due east, and g_k is the Greenwich phase (see (Foreman)). The parameters could be visualized as shown in Fig. 1.

2.1. Selection of constituents

There exists over 500 tidal constituents [see e.g. 2], considering all astronomical and shallow water constituents however, inclusion of all of them the HA demands almost 19 years of continuous data, which is hard to obtain in reality. In practice, the choice of constituents are made through the automated selection algorithm developed by (Foreman), where constituents are selected from a basis of 45 astronomical and 24 of the most important shallow water constituents. The Rayleigh criterion (Godin) is used for constituent selection, whereby a time series of minimum length T_0 can distinguish tidal constituents with a frequency difference of T_0^{-1}. If two constituents do not satisfy the Rayleigh criterion, then the one with a larger amplitude in the equilibrium analysis is considered in the HA, with the other constituent ignored. HA is not suitable for application to non-stationary processes unless the frequencies of the non-stationary processes, if harmonic with known frequencies, are included in the analysis (Jay and Flinchem, 1999).

3. Gaussian process regression

GPs represent a Bayesian non-parametric approach to modelling unknown functions, and have been widely used in solving a variety of regression problems (Ghahramani, 2015; Rasmussen and Williams). The modelling framework considers a prior distribution directly in the space of functions (see Fig. 2), with the properties of such functions governed by the mean and covariance function of the GP – similar to the description of a multivariate Gaussian distribution in terms of its mean and covariance matrix (MacKay, 2003). Consider a dataset with n observations $\mathcal{D} = \{(x_i, y_i), i = 1, 2, .., n\}$ and the objective is to make predictions for new inputs $\boldsymbol{x}_*$. The outputs $\boldsymbol{y}$ are generated by a latent function $f(x)$ with the addition of Gaussian white noise of constant variance (σ_n^2)

$$y_j = f\left(x_j\right) + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}\left(0, \sigma_n^2\right). \tag{2}$$

A prior distribution is considered on the latent functions such that for a given set of training points $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$, the corresponding vector of function evaluations is given by the n dimensional Gaussian distribution $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}_n, \boldsymbol{K}_n)$, and the properties of the prior are dictated by the choice of the mean vector $\mathbf{m}_n(i) = \mathbb{E}[f(x_i)]$ and the covariance (kernel) matrix $\boldsymbol{K}_n(i, j) = \mathbb{E}[(f(x_i) - m(x_i))(f(x_j) - m(x_j))]$. Note, the symbol $\sim$ means 'distributed according to'. The specification of the prior is important as it fixes the properties of the function considered for inference (Rasmussen and Williams). The model likelihood function represents the difference between noisy measurements and the underlying noise-free function (Rasmussen and Williams)

$$p(\boldsymbol{y}|\boldsymbol{f}) = \mathcal{N}\left(\boldsymbol{y}; \boldsymbol{f}, \sigma_n^2 \mathbf{I}\right),$$

where σ_n^2 is the noise variance and $\mathbf{I}$ is the identity matrix. Employing Bayes theorem, the distribution of latent function values given the data points can be expressed as

$$p(\boldsymbol{f}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{\int p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})\mathrm{d}\boldsymbol{f}}.$$

Since in the GP framework we have a multi-dimensional Gaussian distribution for any finite number of input points, the outputs of the latent function at the training locations $\mathbf{x}$ and test locations $\mathbf{x}_*$ can then be given by the distribution

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_n \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}_n & \boldsymbol{K}_{n*} \\ \boldsymbol{K}_{*n} & \boldsymbol{K}_{**} \end{bmatrix} \right)$$

where $\boldsymbol{K}_n$ is the covariance matrix expressing the correlations between all the training points, $\boldsymbol{K}_{n*}$ is the vector of covariance between the training points and the test target, $\boldsymbol{K}_{**}$ is the prior variance at the test locations, while $\mathbf{m}_n$ and $\mathbf{m}_*$ are the mean vectors evaluated at the training and test locations, respectively. $\sigma_n^2 \boldsymbol{y}_*$ Using Gaussian identities, we then have the conditional distribution

$$p(\boldsymbol{f}_*|\boldsymbol{f}) = \mathcal{N}\left(\boldsymbol{m}_* + \boldsymbol{K}_{*n}\boldsymbol{K}_n^{-1}(\boldsymbol{f} - \boldsymbol{m}), \boldsymbol{K}_{**} - \boldsymbol{K}_{*n}\boldsymbol{K}_n^{-1}\boldsymbol{K}_{n*}\right)$$

which is employed to determine the function values $\boldsymbol{f}_*$ at the unknown locations

$$p(\boldsymbol{f}_*|\boldsymbol{y}) = \int p(\boldsymbol{f}_*|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{y})\mathrm{d}\boldsymbol{f}.$$

Finally, the predictions of the output values $\boldsymbol{y}_*$ can be given by

$$p(\mathbf{y}_*|\mathbf{y}) = \int p(\mathbf{y}_*|\mathbf{f}_*)p(\mathbf{f}_*|\mathbf{y})\mathrm{d}\mathbf{f}_* = \mathcal{N}\left(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2\right)$$

where

$$\boldsymbol{\mu}_* = \mathbf{m}_* + \mathbf{K}_{*n}\left(\mathbf{K}_n + \sigma_n^2\mathbf{I}\right)^{-1}(\mathbf{y} - \mathbf{m}_n) \tag{3}$$

$$\boldsymbol{\sigma}_*^2 = \mathrm{K}_{**} + \sigma_n^2\mathbf{I} - \mathbf{K}_{*n}\left(\mathbf{K}_n + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{K}_{n*}, \tag{4}$$
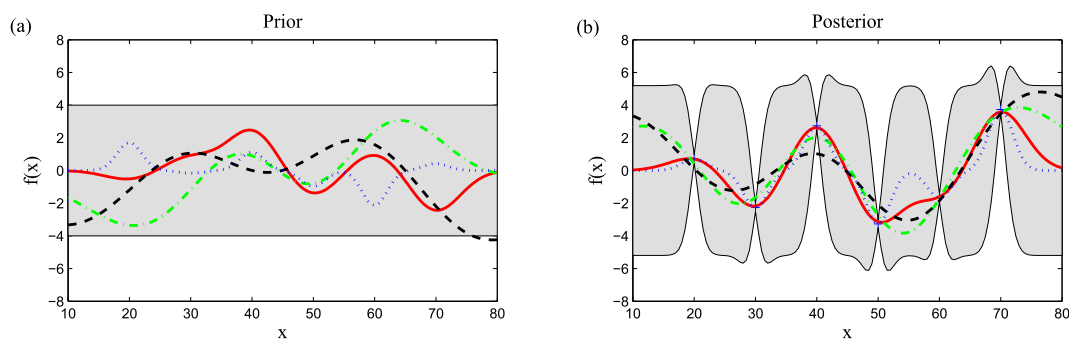

Fig. 2. GPs describe a prior distribution in the space of functions. The prior distribution expresses our belief in problem understanding before seeing the data (figure a). Once inference is performed, the posterior distribution can be described. Essentially this is the distribution of functions which with some added noise passes through the given data points. There is reduced uncertainty at locations near the observations.
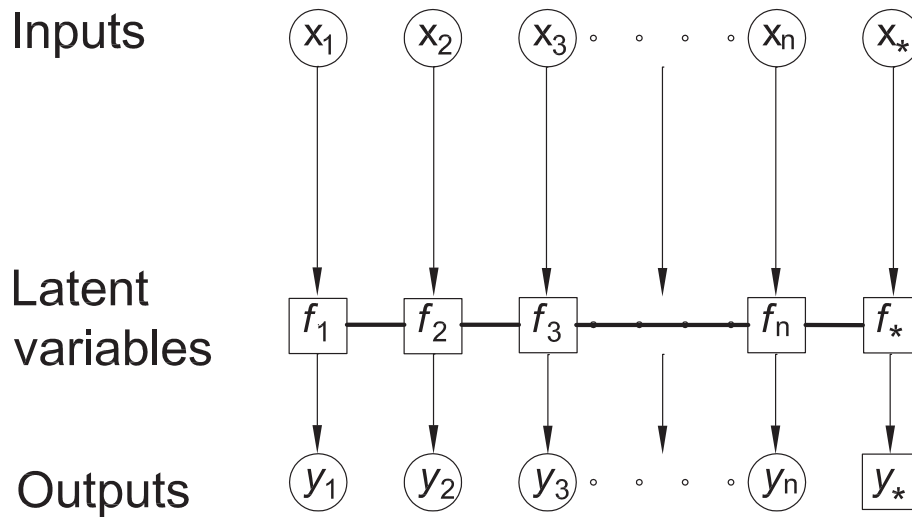
Fig. 3. Structure of Bayesian Gaussian process model for regression. The bold horizontal line represents a fully connected set of nodes. The variables within square boxes are unknown quantities. The variable y_* is the prediction at the unobserved input location x_*.

with $\mathbf{K}_{n*}$ being the matrix of covariance between the training points and test locations and $\mathbf{K}_{n*} = \mathbf{K}_{*n}^T$. The mean function in the present application is considered to be some constant plus a linear trend with respect to time. The trend term can be ignored if desired. Overall the structure of the Bayesian GP can be graphically summarized as in Fig. 3.

3.1. Kernel function

Various forms of the kernel function are known and they can be combined in numerous way (addition, multiplication), depending upon the problem, to model complex data, however the resultant matrix has to be positive semi-definite (Rasmussen and Williams). Note, in the machine learning literature as well as in this paper, the terms kernel function and covariance function are interchangeably used to denote the same function. The covariance matrix is generated by the kernel function. Designing an appropriate kernel function is an important aspect of a machine learning problem. The exponentiated quadratic is one kernel function which is ubiquitously used in solving a wide variety of problems

$$\mathbf{K}(i,j) = \sigma_f^2 \exp\left(\frac{|x_i - x_j|^2}{l^2}\right), \tag{5}$$

where σ_f^2 is the variance hyperparameter and l is the lengthscale hyperparameter of the kernel function. Another popular kernel function is the *Matérn* kernel function [see e.g. (Rasmussen and Williams), which is more appropriate for modelling rougher (non-smooth) variations. For the particular problem of tidal current prediction, given the harmonic nature of the variations, a good choice of would be a periodic kernel function:

$$\mathbf{K}(i,j) = \sigma_f^2 \exp\left(\frac{-2}{l^2}\sin^2\left(\frac{\pi|x_i - x_j|}{p}\right)\right), \tag{6}$$

where σ_f^2 is the variance hyperparameter, l is the lengthscale hyperparameter and p is the period hyperparameter of the kernel, and given the multiple number of distinct periodic tidal constituents, addition of finite number of periodic kernel functions could be appropriate for the present application. The variation of a periodic kernel function with same frequency and signal variance but different length-scales is illustrated in Fig. 4. One of the key strengths of the periodic kernel function is its potential in capturing non-sinusoidal variations.

The posterior distribution $p(\mathbf{y}*|\theta, \mathbf{y})$ is a function of unknown hyperparameters θ which needs to be marginalized (integrating out the uncertainty) as

$$p(\mathbf{y}_*|\mathbf{y}) = \int p(\mathbf{y}_*|\boldsymbol{\theta}, \mathbf{y})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}, \tag{7}$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution of the hyperparameters. Using the Bayes theorem to $p(\boldsymbol{\theta}|\mathbf{y})$, the above equation can be expressed as

$$p(\mathbf{y}_*|\mathbf{y}) = \frac{\int p(\mathbf{y}_*|\theta, \mathbf{y})p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{8}$$

where $p(\mathbf{y}|\boldsymbol{\theta})$ is known as the likelihood and $p(\boldsymbol{\theta})$ is the prior distribution on the unknown hyperparameters. One of the main challenges in the GP approach is finding a solution to the intractable integrals in (7) and (8). Two well known approaches of determining an approximate solution are the maximum likelihood estimation (MLE) also known as the type-II maximum likelihood, and the maximum a-posteriori (MAP) approach. Both these approach typically work well with large datasets where the likelihood function is typically highly peaked. The MLE approach approximates the likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ as a delta function at the location of the hyperparameters maximizing it:

$$\log p(\mathbf{y}|\boldsymbol{\theta}_{\mathbf{MLE}}) = \underbrace{-\frac{1}{2}(\mathbf{y} - \boldsymbol{m}_\mathbf{n})^T\left(\mathbf{K}_\mathbf{n} + \sigma_n^2\mathbf{I}_n\right)^{-1}(\mathbf{y} - \boldsymbol{m}_\mathbf{n})}_{\text{penalizes mismatch between data and prediction}} \overbrace{-\frac{1}{2}\left|\mathbf{K}_\mathbf{n} + \sigma_n^2\mathbf{I}_n\right|}^{\text{penalizes model complexity}}$$
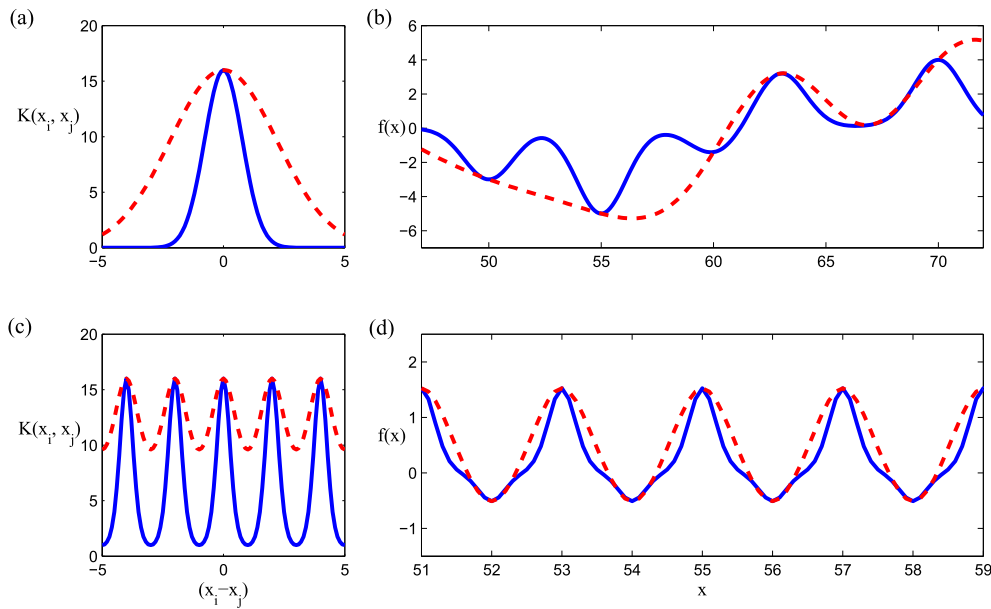$$-\frac{n}{2}\log(2\pi) \tag{9}$$

Fig. 4. Figure (a) shows an exponentiated quadratic kernel function with two different lengthscales, while figure (b) shows example draws from a GP prior with those length scales. The red line corresponds to a longer length-scale than the blue line. Figure (c) and (d) shows the same but for a periodic kernel function. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In the MAP method, the hyperparameter posterior $p(\boldsymbol{\theta}|\mathbf{y})$ in equation (7) is considered to be a Dirac delta function at its maximum, with the optimization performed on:

$$\log p(\boldsymbol{\theta}_{MAP}|\mathbf{y}) = -\frac{1}{2}(\mathbf{y} - \mathbf{m_n})^T (\mathbf{K_n} + \sigma_n^2 \mathbf{I}_n)^{-1} (\mathbf{y} - \mathbf{m_n}) - \frac{1}{2}|\mathbf{K_n} + \sigma_n^2 \mathbf{I}_n|$$
$$-\frac{n}{2}\log(2\pi) + \log p(\boldsymbol{\theta}_{MAP}) . \tag{10}$$

The prior $p(\boldsymbol{\theta}_{MAP})$ could either be approximated with a flat distribution or a probability distribution could be assigned to it. The optimization is performed using the method of conjugate gradients. For large tidal current datasets, we adopt the MAP approach for determining the hyperparameters as it incorporates additional information through hyperparameter prior distributions $p(\boldsymbol{\theta}_{MAP})$. For short time-series data, the approximation of hyperparameter posterior as a Dirac delta function can often become inappropriate. A Markov chain Monte Carlo (MCMC) in the framework of Hamltonian Monte Carlo can be used to determine a solution to (7) in such circumstances. The method simulates a fictitious Hamiltonian dynamical system to achieve better exploration of the posterior distribution of the hyperparameters (see (Rasmussen, 1996)).

3.2. Multi-output

The formulation described until now does not account for the correlations between the output variables, which in the case of tidal current prediction are the horizontal velocities u and v. In the GP framework, the problem of multiple output reduces to the specification of an appropriate covariance function, which while being positive semi-definite, captures the dependencies between the data points across all the outputs (Alvarez and Lawrence, 2011). The covariance function for multiple output can be expressed as

$$\mathbf{K} = \sum_{q=1}^{M} \boldsymbol{\Upsilon}_q \otimes \mathbf{K}_{,q} \tag{11}$$

where M indicates the finite number of frequencies considered from astronomical considerations,

$$\mathbf{K}_q(i,j) = \exp\left(\frac{-2}{l_q^2}\sin^2\left(\frac{\pi|x_i - x_j|}{p_q}\right)\right) \tag{12}$$

and the matrix $\boldsymbol{\Upsilon}_q$ is assumed to be of spherical parametrization kind (Pinheiro and Bates, 1996; Osborne et al., 2012) with

$$\boldsymbol{\Upsilon}_q = \text{diag}(\boldsymbol{e}_q)\boldsymbol{S}_q^T \boldsymbol{S}_q \text{diag}(\boldsymbol{e}_q). \tag{13}$$

For example, where we have only two outputs, in (13), e_q gives a description for the length scale of each output variable

$$\text{diag}(\boldsymbol{e}_q) = \begin{pmatrix} l_{u,q} & 0 \\ 0 & l_{v,q} \end{pmatrix} \tag{14}$$

and $\boldsymbol{S}$ is an upper triangular matrix, the i th column of which is associated with particular spherical coordinates of points of $\Re^i$:

$$\boldsymbol{S}_q = \begin{pmatrix} 1 & \cos\theta_q \\ 0 & \sin\theta_q \end{pmatrix} \tag{15}$$

and the final form of the coregionalization matrix is given by

$$\boldsymbol{\Upsilon}_q = \begin{pmatrix} l_{u,q}^2 & l_{u,q}l_{v,q}\cos\theta_q \\ l_{u,q}l_{v,q}\cos\theta_q & l_{v,q}^2 \end{pmatrix}. \tag{16}$$

4. Analysis

The data analyzed in this work are from public datasets available through National Oceanic and Atmospheric Administration (NOAA), USA (www.noaa.gov). The classical HA is performed using the state-of-the-art Unified Tidal analysis toolbox (UTide) developed by (Codiga, 2011–01) using the 'Iteratively Reweighted Least Squared' (IRLS) algorithm (Leffler and Jay, 2009; Foreman) which performs a weighted statistical fitting such that the effect of the outliers are reduced. An important point to note is that the GP is a non-parametric method, and the hyperparameters – which are unknown – are different from parameters in parametric models (HA). The hyperparameters (e.g. length-scale, variance, periodicity) govern the properties of functions chosen from the GP. They do not necessarily have physical interpretations, although the periodicity hyperparameters have some correspondence to the harmonic
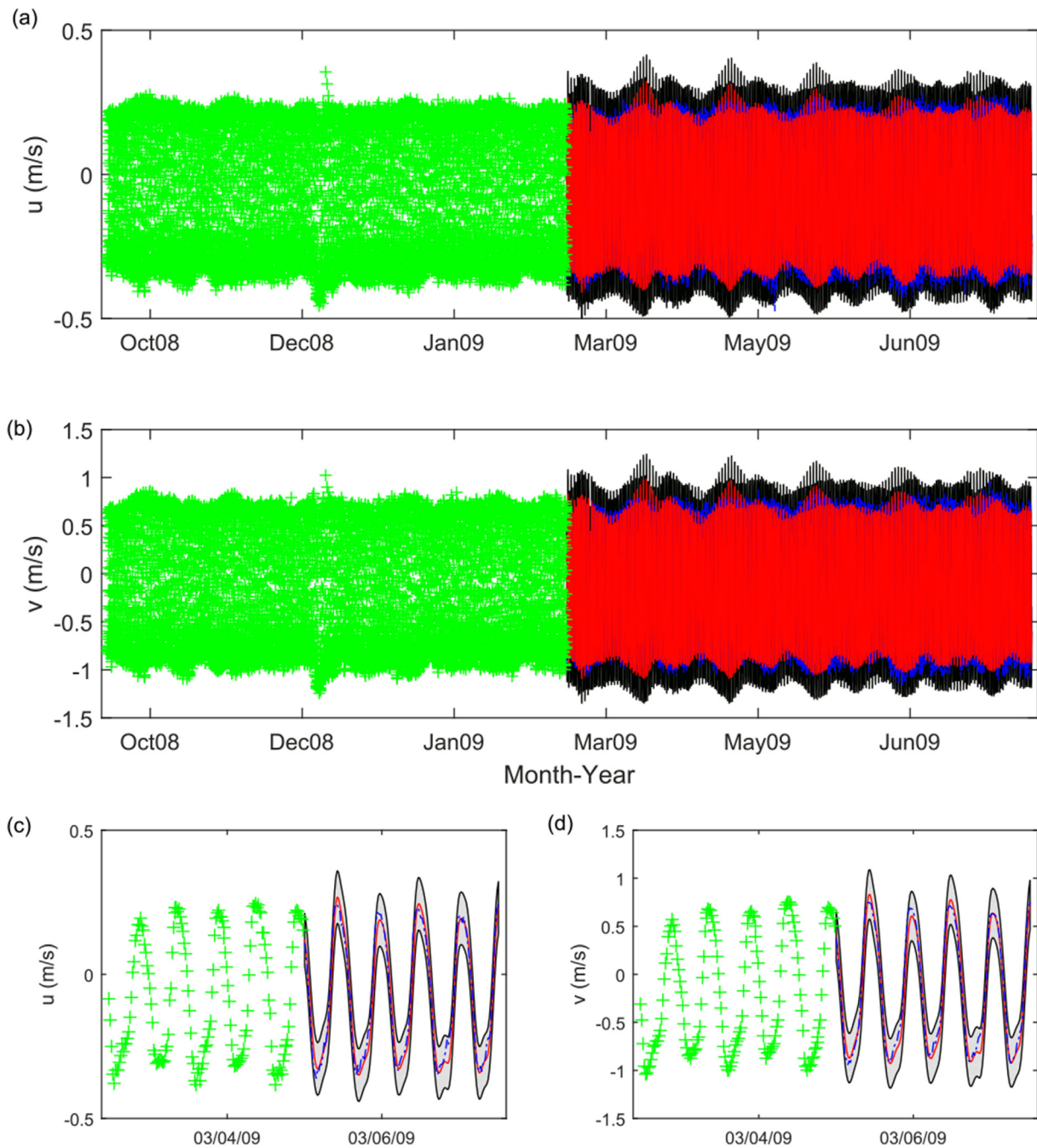
Fig. 5. The figure shows the variations of the horizontal velocities u and v with time at Sunshine Sky Bridge. Five months of data is used to predict the tidal current variations for the next five months. The green '+' markers indicate the training points, the red solid line indicates the mean of the predictions, the black solid line containing the region in grey indicates the 95% confidence interval and the blue dotted line indicates the actual values. The figures (a) and (b) provide a good macroscopic view of the tidal prediction, however the details are obscured due to the vast amount of information it contains. Figures (c) and (d) shows a zoomed portion of figures (a) and (b) respectively, including the transition from training to prediction zone in both cases. The last two figures clearly shows the predicted mean (red) and confidence interval (grey), along with the periodic nature of the variations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

nature of the tidal data. Computations for the two approaches (HA and GP) are performed with the same number of tidal constituents, chosen by the auto-selection algorithm developed by (Foreman). However note that the Rayleigh criterion decides the number of periodic kernel components of the GP; the form of the periodic kernel component enables the GP to capture more nonlinearities than HA. The frequencies of the periodic covariance functions in the GP formulation are specified to be the frequencies determined by the auto-selection procedure (Foreman). The locations of the longer datasets are presented in Table 1. The data used in

the analysis is sampled once every 24 min. Fig. 5 (a) and (b) plots the variation of the horizontal velocities (u and v) versus time obtained using GP for one of the locations - Sunshine Sky Bridge (see Table 1). The mean and confidence margins of the predictions agree well with the data. Due to the vast amount of information contained in Fig. 5, the short-term variations and the details are not properly visible. Fig. 5(c) and (d) presents a short section of Fig. 5(a) and (b) respectively, illustrating the details of the GP outputs and the velocity data. Another example is depicted in Fig. 6, showing the training and prediction data of the
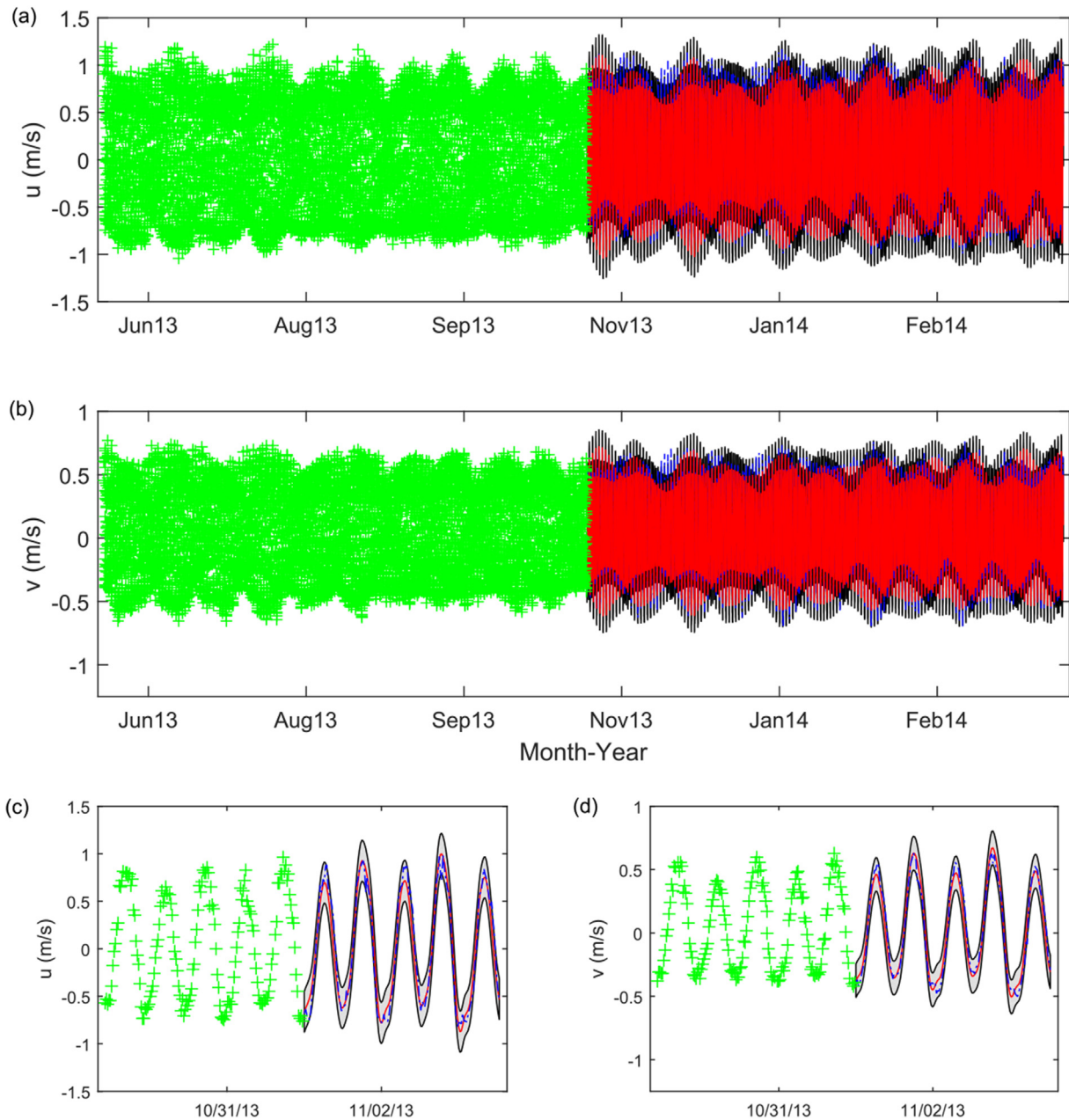
Fig. 6. The figure shows the variations of the horizontal velocities u and v with time at Martinez-AMORCO Pier. Five months of data is used to predict the tidal current variations for the next five months. The details are the same as the previous figure.

horizontal velocities at Martinez-AMORCO Pier (see Table 1).

Table 2 shows the root mean squared error (RMSE) of tidal current predictions using the GP and comparisons are made alongside with the HA approach. The first column shows the location, the second column shows the training duration, then the prediction/test duration followed by the errors in u and v velocities respectively. Four locations are considered and predictions are performed over 1 month, 2 months and 5 months for each of the cases. The auto-selection algorithm selects 35 tidal constituents for all the four cases presented in Table 2, with a minimum threshold factor of one in the Rayleigh criterion. The GP approach achieves better accuracy than the HA in each of the cases. On an average, the GP approach reduces the RMSE by around 5 percent from the HA, and in some cases doing better by as much as 10 percent. The observed fluctuations in the errors over different prediction horizons can be attributed to variations in the real-world data arising from non-tidal contributions, as well as from low frequency tidal components which are not separated

due to the lack of training data. Non-tidal effects can originate from sub-tidal variations, storm surges, etc. The purpose of the comparison is to show that the errors are comparable over different prediction horizons.

The training of the GP for the above cases is performed using an optimization algorithm, and the computational time depends on the number of function evaluations. On an average, the computational costs involved in using the GP are in the order of hours, while the HA takes

Table 1
Location of the tidal observation sites for long term predictions. The current measurements at all the four locations are made at near surface.

Location	Latitude	Longitude
Southampton Shoal Channel	37° 54.975′ N	122° 25.340′ W
Old Port Tampa	27° 51.772′ N	82° 33.224′ W
Sunshine Sky Bridge	27° 37.500′ N	82° 39.300′ W
Martinez-AMORCO Pier	38° 2.078′ N	122° 7.515′ W

Table 2

Comparison of the RMSE (in m/s) of the prediction using the HA and GP algorithm. The first column (left-most) indicate the locations, the second column shows the 5 month training period for the algorithms, while the third column present the prediction durations starting one day after the end of the training algorithm. All data for training and testing have been obtained from NOAA.

Location	Training Duration	Prediction	u		v	
			HA	GP	HA	GP
Southampton	25/03/13–24/08/13	1 month	0.0695	0.0668	0.1137	0.1083
Shoal Channel		2 months	0.0718	0.0687	0.1104	0.1050
		5 months	0.0738	0.0720	0.1042	0.1008
Old Port Tampa	15/01/13–14/06/13	1 month	0.0573	0.0532	0.1179	0.1092
		2 months	0.0562	0.0523	0.1174	0.1090
		5 months	0.0634	0.0586	0.1310	0.1158
Sunshine Sky	29/04/14–28/09/14	1 month	0.1340	0.1247	0.0916	0.0836
Bridge		2 months	0.1375	0.1337	0.0942	0.0904
		5 months	0.1252	0.1236	0.0841	0.0825
Martinez	29/04/14–28/09/14	1 month	0.1322	0.1250	0.0787	0.0756
		2 months	0.1297	0.1239	0.0773	0.0735
		5 months	0.1208	0.1180	0.0743	0.0713

seconds to evaluate. The HA is a parametric approach and is quasi-instantaneous; in contrast the Bayesian GP is non-parametric with the number of parameters growing with data, and as such takes more time for evaluation. This shortcoming can be addressed by using scalable approaches to GP inference (Hensman et al., 2013; Ding et al., 1708), and the results can be achieved in times comparable to that of HA.

Few short data-series are also analyzed, using the MCMC approach (see Table 3 for the locations). Some of the locations record quite strong tidal currents (Avatanak and Derbin Strait), which could make them suitable for tidal energy projects. Approximately half of the available data is used for training, while the rest for predictions. The auto-selection algorithm selects 17 tidal constituents for all these cases. Table 4 shows the RMSE of the predictions from the HA and GP approach for the short datasets. In this case again, the GP based approach outperforms the HA.

Some of the cases in Table 4 have much larger RMSE than those in Table 2. A general increase in error can be ascribed to the shorter length of input data in Table 4, and fewer tidal constituents are selected in the latter. However, the first two cases in Table 4 (Derbin and Avatanak Strait) have the largest magnitudes of tidal currents, and produces the largest values of RMSEs, although their percentage errors, with respect to overall current magnitudes, are similar to others. The results for one of the locations (Gibson Point) are shown in Fig. 8.

In many cases, we are interested in unified prediction of currents (tidal + non-tidal). The non-tidal effects are often non-harmonic and can have energy comparable to that of the tidal. Such perturbations could originate from a variety of sources – internal tides (Wunsch, 1975), river discharge, meteorological forcing, estuarine currents. Prediction becomes challenging in circumstances when the low frequency non-tidal variations are considerable. To accommodate such variations, the mean of the GP can be modelled with another GP – meanGP:

$$m \sim \mathscr{GP}(m_{mean}, k_{mean}) \tag{17}$$

In a simple construction, the meanGP can have a constant mean function m_{mean} with value corresponding to the average of the training dataset, while an exponentiated quadratic kernel function can be chosen as its covariance function (as such variations do not necessarily have to periodic). The generic kernel functions like the exponentiated quadratic

have the capability of expressing correlations to data points which are up to a length scale away from the domain of input data points, beyond which the predictions resorts to the output from the mean function. The length-scale hyperparameter in the exponentiated quadratic kernel function regulates the range of dependencies in the process. A short length-scale corresponds to rapidly varying functions with weak long-range dependencies, while a large length-scale correspond to slowly varying functions. The approach could ideally be useful for short-term predictions as the non-deterministic component modelled through meanGP can be hard to predict. We perform a short case study to show that GP can be used to reasonable short-term predictions in such cases. Data is considered from the location Port of Beaumont (Latitude: 30˚4.782′, Longitude: 94˚5.175′) through NOAA. A wavelet transform is performed on the current velocities to obtain some qualitative description of the characteristics of the tidal current data series. Wavelet transform (Daubechies et al, 1992) has been applied to a wide variety of signal processing problems (Mallat, 1999), and unlike Fourier transform, it provides a temporal description of the frequency spectrum, which entails a loss in frequency resolution, but gain in some temporal description. Fig. 7 shows the log of the wavelet power spectrum of the current velocity plotted versus the period of the tidal constituents (abscissa in $\log_2$ scale) and time (ordinate). As it can be noticed, the current data possess significant amount of non-tidal energy especially at low frequencies. It is not possible to capture such variations, appropriately, using the HA approach. Note, the wavelet transform can resolve different tidal species (diurnal, semi-diurnal) band, but not the tidal constituents within the species (Guo et al., 2015). Fig. 9 shows the current data at the

Table 3

Location of the tidal observation sites. - Short tidal series. The rightmost column shows the approximate depth at which the measurements are made.

Location	Latitude	Longitude	Approx. Depth (m)
Derbin Strait	54° 5.035′ N	165° 13.624′ W	67.18
Avatanak Strait	54° 6.767′ N	165° 28.544′ W	75.29
Hog Neck	41° 43.433′ N	70° 38.363′ W	8.53
Fire Island	61° 10.754′ N	150° 7.532′ W	8.44
Gibson Point	47° 13.148′ N	122° 35.320′ W	60.53

Table 4

Comparison of the RMSE (in m/s) of the prediction using the HA and GP algorithm for short-term data analysis.

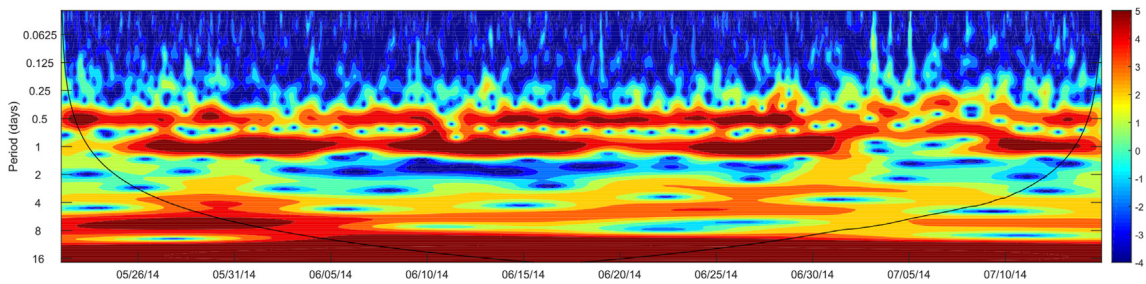| Location	Training Duration	Prediction	u		v	
			HA	GP	HA	GP
Derbin Strait	12/06/10–02/07/10	03/07/10–23/07/10	0.3470	0.3433	0.3631	0.3542
Avatanak Strait	11/06/10–02/07/10	03/07/10–23/07/10	0.2540	0.2442	0.2323	0.2267
Hog Neck	03/06/09–27/06/09	28/06/09–21/07/09	0.1873	0.1787	0.2431	0.2428
Fire Island	13/07/02 - 02/08/02	03/08/02–23/08/02	0.1657	0.1645	0.0895	0.0850
Gibson	29/05/15–22/06/15	23/06/15–17/07/15	0.1209	0.1196	0.2081	0.2000

Fig. 7. Wavelet transform showing the log of the wavelet power spectrum of velocity of short tidal current data series at Port of Beaumont. The analysis is performed using them methods of (Torrence and Compo, 1998) and (Liu and Miller, 1996). The Morlet wavelet has been used in the analysis and the black line encloses regions of greater than 95% confidence for a red noise process with a lag-1 coefficient of 0.72. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
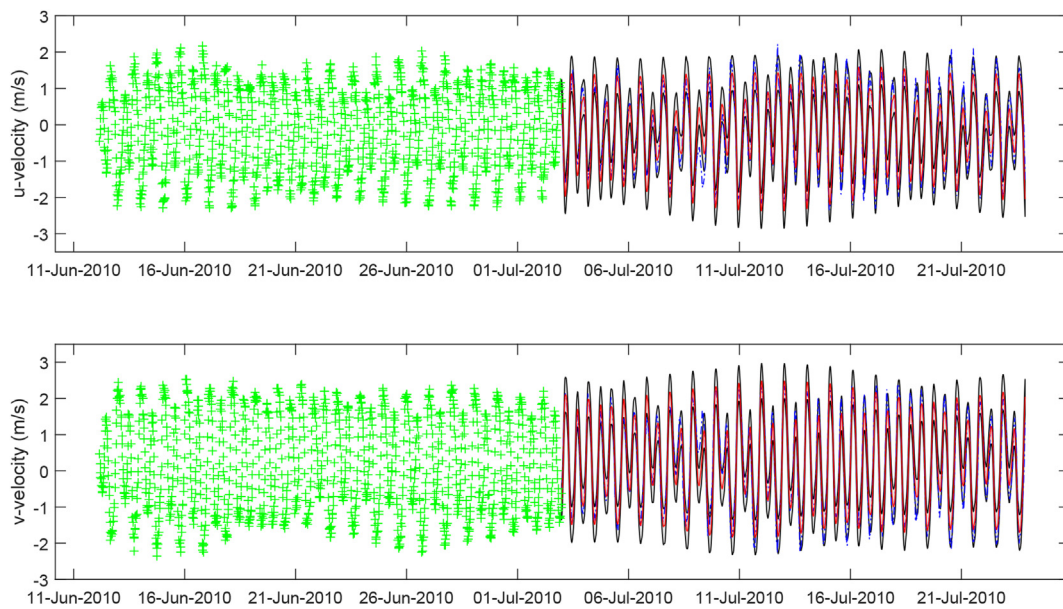

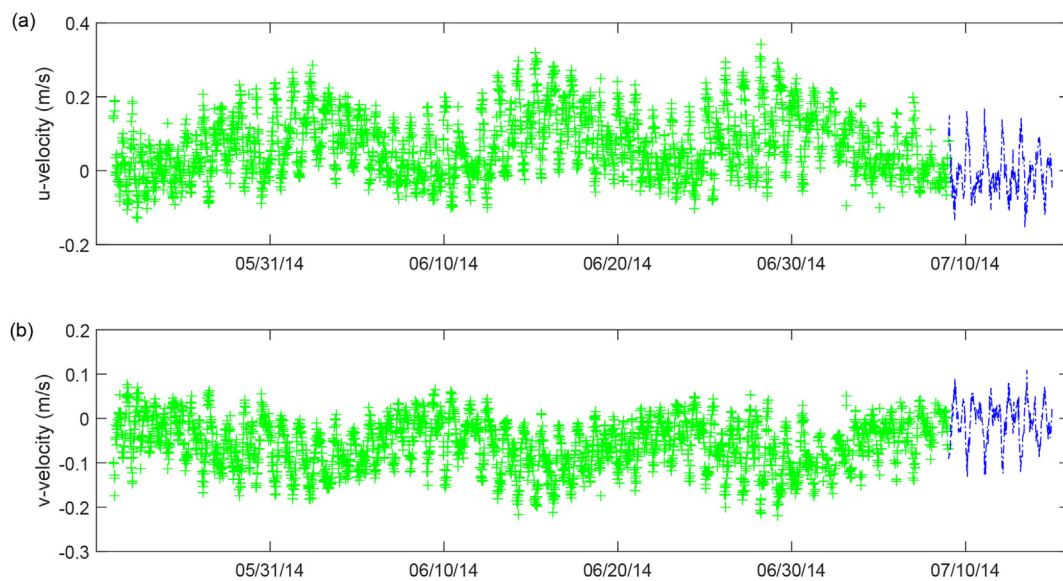**Fig. 8.** Training and prediction of a short tidal current data series at Gibson Point.

Fig. 9. Variation of horizontal current velocities (a) u and (b) v at Port of Beaumont (USA) between 22/May/2014 to 14/July/2014. The points in green indicate the initial training data while the blue dashed-dotted line indicate the data points at the test locations which we intend to predict half-a-day in advance. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5
RMSE of the half-a-day ahead predictions at Port of Beaumont. 1st half indicates the duration between midnight and noon and 2nd half indicates between noon till midnight on the same day. The interval between consecutive data readings is 24 min, although there do exist some minor gaps in the data.

Prediction	Duration	u		v	
		HA	GP	HA	GP
09/07/14	1st half	0.0693	0.0531	0.0444	0.0385
	2nd half	0.0512	0.0390	0.0346	0.0287
10/07/14	1st half	0.0456	0.0349	0.0346	0.0257
	2nd half	0.0771	0.0575	0.0496	0.0409
11/07/14	1st half	0.0679	0.0490	0.0380	0.0289
	2nd half	0.0819	0.0561	0.0534	0.0410
12/07/14	1st half	0.1068	0.0578	0.0632	0.0397
	2nd half	0.0890	0.0520	0.0564	0.0289
13/07/14	1st half	0.1172	0.0447	0.0693	0.0254
	2nd half	0.0945	0.0408	0.0619	0.0232
14/07/14	1st half	0.1128	0.0378	0.0712	0.0255
	2nd half	0.0987	0.0361	0.0577	0.0195

location from 22/May/2014 to 14/July/2014 - which is 54 days of data, and predictions are made over the last six days (shown in blue in the figure), half-a-day in advance. For this particular case, 24 hours moving average of the tidal current velocities is first computed and the meanGP formulation (described earlier) is employed on the data points (from moving average). Table 5 shows the RMSE of the half-a-day ahead predictions. Significant reductions in error have been achieved with an overall (considering 7 days) decrease of 43 percent in the u velocity and 40 percent in the v velocity. This demonstrates that even in the absence of additional information, ML can be used for enhanced current predictions. One instance of such predictions (half a day in advance) is shown in Fig. 10.

5. Discussion

The paper presents an alternate approach to the prediction of tidal currents, different from all known existing methods. Comparisons are made with the HA for some real tidal current datasets (see Tables 2 and 4) which are almost stationary (ideally suited for HA), and the GP approach is shown to outperform the HA. The advantages of the machine learning approach lies in its flexibility and applicability in modelling a wide variety of real world situations. In the GP framework, structures in the data are captured primarily through kernel functions which comprises of hyperparameters, and they are typically learned from the data. Unlike in

HA which strictly requires the frequencies of the constituents to be predetermined, the periodicity hyperparameter in the GP do not necessarily need to be known beforehand.

Additional information can be incorporated into the GP model which can enable it to make better short-term predictions e.g. meteorological forcing parameters like wind speed and direction which are known to influence the tidal current variations can be considered as extra inputs to the GP model. One can also be interested in knowing the tidal current variations at an unknown location given the data of some nearby locations. In fact, vessel mounted acoustic Doppler current profilers (ADCPs) are often used to gather data over a spatial domain within a certain timeframe. It is not possible to perform robust predictions using classical techniques for such datasets. Such cases would involve modelling the spatial variability in addition, and a GP model with an appropriate kernel function can be used for such purpose. With the problem being formulated in the machine learning framework, we have at our disposal a wide range of possibilities in modelling tidal datasets. Methods or algorithms can be chosen depending on the specifics of the dataset. Some other features of the presented methodology are:

- Non-stationarity could be introduced by some non-linear transformation (warping) of the inputs [see e.g. Sampson and Guttorp, 1992] or outputs [see e.g. Snelson et al., 2004], and then using the normal kernel functions (stationary) on the mapped space.
- There can be issues in modelling tidal current time series using HA when the size of the basis matrix becomes large. Sparse algorithms of the GP - sparse spectrum GP (Lázaro-Gredilla et al., 2010), fully independent training conditional (FITC) (Quinonero-Candela and Rasmussen, 2005) can be used in such circumstances. The algorithms have been shown to handle large datasets successfully.
- The likelihood function in GP models the noise in the data. Although the simple i. i.d. Gaussian distribution (white noise) is chosen in this work, other noise models - Laplace, Warped Gaussian can be considered within the same framework. Exact inference is not possible in such cases and approximate deterministic inference techniques (Variational Bayes, Expectation Propagation, Laplace)[see (Minka, 2001) or numerical sampling techniques (like MCMC) (Bishop, 2006) could be applied.
- The focus of the present paper is on tidal currents, however the method is easily applicable to the prediction of tidal heights. The latter involves just 1-D output (tidal heights) and hence is more straightforward.
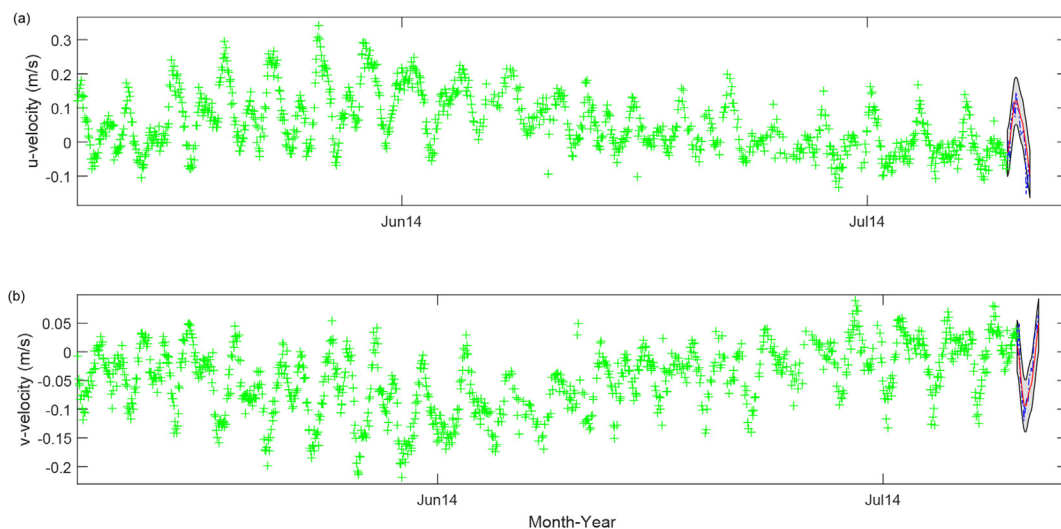

Fig. 10. Current velocity prediction half a day in advance. Note, the figure doesn't show the entire training data (in green) used for the predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- Effect of additional unknown frequencies in the tidal signal data could be incorporated by considering additional periodic kernel functions with 'free' periodicity hyperparameter.

The GP formulation is a non-parametric approach, and therefore has added flexibilities which are not provided by parametric methods like the HA. The periodic kernel function chooses from the space of harmonic functions and the output may not necessarily be exactly sinusoidal.

6. Conclusion

The problem of tidal current prediction is formulated in the framework of Bayesian GP. The goal of this work was develop robust prediction method which can address some of the shortcoming of the traditional approaches. The kernel function of the GP captures the structures in the data and amongst the wide choice of kernel functions, the periodic kernel function is in particular relevant for the present application due to its harmonic nature, without being restricted to just sinusoidal variations. The hyperparameters of the kernel functions are inferred either deterministically using MAP approach or stochastically using MCMC, with the former method being preferred for large datasets. The probabilistic nature of the machine learning approach enables it to handle uncertainty and noise, and the method generates confidence intervals directly in the time-domain (unlike HA), which would be useful in practical applications. The machine learning approach is also shown to capture non-harmonic variation using the exponentiated quadratic kernel function and can make better predictions in the short-term. We analyze and perform predictions on several sets of field data at various locations around USA. In general, the GP framework provide additional flexibilities not provided by the traditional techniques and can enhance tidal analysis. We have demonstrated that the GP based approach can achieve better accuracy than the traditional HA over a wide range of cases and varied signal characteristics. In general, enhanced prediction of currents requires better understanding of the physical processes such as the conversion of barotropic to baroclinic tides, decay of internal tides into turbulence (Egbert and Ray, 2000), meteorological effects, etc. Availability of data related to the variation of relevant parameters e.g. the meteorological forcing variables (wind stress intensity and direction) can aid in the construction of better predictive models and exploit the full potential of the machine learning algorithms.

Acknowledgements

References

Alvarez, M.A., Lawrence, N.D., 2011. Computationally efficient convolved multiple output Gaussian processes. J. Mach. Learn. Res. 12, 1459–1500.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.

Cartwright, D.E., 1968. A unified analysis of tides and surges round north and east britain. Philos. Trans. Royal Soc. London A: Math. Phys. Eng.Sci. 263 (1134), 1–55.

Codiga, D., 2011-01. Unified Tidal Analysis and Prediction Using the Utide Matlab Functions. technical report, p. 59. graduate school of oceanography, university of rhode island, narragansett, ri. ftp://www.po.gso.uri.edu/pub/downloads/codiga/pubs/2011Codiga-UTide-Report.pdf.

Daubechies, I., et al., 1992. Ten Lectures on Wavelets, vol. 61. SIAM.

Ding, Y., Kondor, R., Eskreis-Winkler, J.. Multiresolution Kernel Approximation for Gaussian Process Regression. arXiv:1708.02183 [stat]. http://arxiv.org/abs/1708.02183.

Egbert, G., Ray, R., 2000. Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. Nature 405 (6788), 775–778.

Foreman, M., Cherniawsky, J., Ballantyne, V., 2009. Versatile harmonic tidal analysis: improvements and applications. J. Atmos. Ocean. Technol. 26 (4), 806–817.

M. Foreman, Manual of Tidal Heights Analysis and Prediction., Pacific Marine sciences Report 77–10, Institute of Ocean Sciences, Patricia Bay.

Friedrichs, C.T., Aubrey, D.G., 1988. Non-linear tidal distortion in shallow well-mixed estuaries: a synthesis. Estuar. Coast Shelf Sci. 27 (5), 521–545.

Geyer, W.R., 1993. Three-dimensional tidal flow around headlands. J. Geophys. Res.: Oceans 98 (C1), 955–966.

Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. Nature 521 (7553), 452–459.

G. Godin, The Analysis of Tides, University of Toronto Press.

Guo, L., van der Wegen, M., Jay, D.A., Matte, P., Wang, Z.B., Roelvink, D., He, Q., 2015. River-tide dynamics: exploration of nonstationary and nonlinear tidal behavior in the yangtze river estuary. J. Geophys. Res.: Oceans 120 (5), 3499–3521.

Hensman, J., Fusi, N., Lawrence, N.D., 2013. Gaussian processes for big data. In: Conference on Uncertainty in Artificial Intelligence, pp. 282–290. http://arxiv.org/abs/1309.6835.

Jay, D.A., Flinchem, E.P., 1999. A comparison of methods for analysis of tidal records containing multi-scale non-tidal background energy. Continent. Shelf Res. 19 (13), 1695–1732.

Kim, C.-J., Nelson, C.R., 1999. Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. Rev. Econ. Stat. 81 (4), 608–616.

Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C.E., Figueiras-Vidal, A.R., 2010. Sparse spectrum Gaussian process regression. J. Mach. Learn. Res. 11, 1865–1881.

Leffler, K.E., Jay, D.A., 2009. Enhancing tidal harmonic analysis: robust (hybrid l1/l2) solutions. Continent. Shelf Res. 29 (1), 78–88.

Liu, P.C., Miller, G.S., 1996. Wavelet transforms and ocean current data analysis. J. Atmos. Ocean. Technol. 13 (5), 1090–1099.

MacKay, D., 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press.

Mallat, S., 1999. A Wavelet Tour of Signal Processing. Academic press.

Matheron, G., 1973. The intrinsic random functions and their applications. Adv. Appl. Probab. 439–468.

Minka, T.P., 2001. Expectation propagation for approximate bayesian inference. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp. 362–369.

Munk, W.H., Cartwright, D.E., 1966. Tidal spectroscopy and prediction. Philos. Trans. Royal Soc. London A: Math. Phys. Eng. Sci. 259 (1105), 533–581.

Osborne, M.A., Roberts, S.J., Rogers, A., Jennings, N.R., 2012. Real-time information processing of environmental sensor network data using bayesian Gaussian processes. ACM Trans. Sens. Netw. 9 (1), 1.

Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis including error estimates in matlab using t_tide. Comput. Geosci. 28 (8), 929–937.

Pinheiro, J.C., Bates, D.M., 1996. Unconstrained parametrizations for variance-covariance matrices. Stat. Comput. 6 (3), 289–296.

Quinonero-Candela, J., Rasmussen, C.E., 2005. A unifying view of sparse approximate Gaussian process regression. J. Mach. Learn. Res. 6, 1939–1959.

Rasmussen, C.E., 1996. Evaluation of Gaussian Processes and Other Methods for Nonlinear Regression. Ph.D. thesis. Citeseer.

C. E. Rasmussen, C. K. Williams, Gaussian processes for machine learning, the MIT Press.

Ripley, B.D., 2005. Spatial Statistics, vol. 575. John Wiley & Sons.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., Aigrain, S., 1984. Gaussian processes for time-series modelling. Philos. Trans. Royal Soc. London A: Math. Phys. Eng. Sci. 371 (2013) 20110550.

Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure. J. Am. Stat. Assoc. 87 (417), 108–119.

Sarkar, D., Osborne, M., Adcock, T., 2016. A machine learning approach to the prediction of tidal currents. In: The Proceedings of the 26th (2016) International Ocean and Polar Engineering Conference, Rhodes, Greece, vol. 2016, pp. 692–700.

Snelson, E., Rasmussen, C.E., Ghahramani, Z., 2004. Warped Gaussian processes. Adv. Neural Inf. Process. Syst. 16, 337–344.

Thompson, P.D., 1956. Optimum smoothing of two-dimensional fields. Tellus 8, 384–393.

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79 (1), 61–78.

Wunsch, C., 1975. Internal tides in the ocean. Rev. Geophys. 13 (1), 167–182.