

Key Points:

- This paper introduces a new Bayesian nonparametric machine learning technique for spatiotemporal tidal current prediction
- It provides better spatiotemporal description with data from one source than classical techniques using continuous data from multiple sources
- Method can be applied to spatiotemporal tidal current data from any region

Correspondence to:

D. Sarkar,
dripta.academic@gmail.com

Citation:

Sarkar, D., Osborne, M. A., & Adcock, T. A. A. (2019). Spatiotemporal prediction of tidal currents using Gaussian processes. *Journal of Geophysical Research: Oceans*, 124. <https://doi.org/10.1029/2018JC014471>

Received 14 AUG 2018

Accepted 13 MAR 2019

Accepted article online 19 MAR 2019

Spatiotemporal Prediction of Tidal Currents Using Gaussian Processes

Dripta Sarkar^{1,2} , Michael A. Osborne¹ , and Thomas A. A. Adcock¹ 

¹Department of Engineering Science, University of Oxford, Oxford, UK, ²Now at Department of Mathematics, Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, Howrah, India

Abstract Predicting fast tidal currents can be a challenging task. Unlike tidal water levels, currents can vary sharply over short distances. The classical approach of harmonic analysis can analyze data at point locations and there is a need for a method that can handle spatiotemporal data, as well as be robust to the uncertainty and noise inevitable in real-world measurements. In this work, we present a Bayesian machine learning (ML) approach to tackle the problem. The method is based on Gaussian processes, a nonparametric ML technique that uses a kernel function to capture structures in the data. A case study is performed using data from a validated numerical model simulating the tidal dynamics in the Pentland Firth region, UK. Several sampling strategies are explored and the case where measurement location is changed after every sampling is found to produce the lowest average error in the predictions. We show that the presented methodology using data from just a single moving data source can provide a better spatiotemporal description than traditional techniques using continuous data from a large number of locations. The work can be useful to developers of tidal energy farms, navigation, and other purposes.

1. Introduction

In deep waters, tidal dynamics are essentially linear. However, as tides move into shallow waters, they are distorted due to nonlinear interactions that can make the dynamics extremely complex. Such interactions are strongly dependent on the local bathymetry and topography features and can produce tidal asymmetry (dissimilarity between flood and ebb tides) resulting from the interactions with overtides and compound tides (see, e.g., Neill et al., 2014). While signal processing techniques like the Fourier transform and wavelet transform are commonly used for investigating time series data, the classical approach of harmonic analysis (Foreman, 1977; Godin, 1972) is popular in tidal analysis and codes based on it having been used for decades (see, e.g., Codiga, 2011). These methods can only analyze and make point predictions, and one needs to resort to numerical hydrodynamic models to understand spatial variability. Such approaches are computationally expensive, are often unable to model all the relevant physics, and require real-world data at predicted locations for validation (Parker, 2007). Unlike tidal heights that vary smoothly over hundreds of kilometers, tidal currents can be distinctively different at two locations separated by a few hundred meters (Parker, 1991). Tidal currents will vary not just spatially (in horizontal coordinates) but also vertically with depth. However, for simplicity, and lack of data, in this paper we ignore the depth variation of tidal currents. In general, tides (tidal heights) have been analyzed far more extensively than that of currents, with the tidal height record data available in abundance. Godin (1983) showed that tidal currents cannot be predicted with the same level of precision as that of heights.

Some studies have focused on understanding the spatial variation of tidal currents at regions that could be important from a navigational or economic point of view, demanding massive in situ data measurement operations. For example, in the recent work of Dusek et al. (2016), 48 acoustic Doppler current profilers (ADCPs) recordings were analyzed to estimate the spatial variability of tidal currents in Puget Sound, Washington. The study is a part of a bigger project in which National Oceanic and Atmospheric Administration (NOAA), USA, is deploying ADCPs at more than a hundred locations over a 3-year period to cover regions in Washington State (see Dusek et al., 2016). In the context of renewable energy extraction, tidal turbine arrays are planned at locations around the world. The tidal energy community has considerable interest on the spatial variability of currents at promising sites because the power output from energy extracting devices is directly proportional to the velocity cubed. Thus, errors in the current predictions

can have considerable effect on the operationally relevant parameters (Polagye et al., 2010); for example, small errors in the current predictions may lead to large errors in the power outputs. Tidal energy sites could vary from flood dominant to ebb dominant over small spatial scales. It is important to have a good understanding of the spatiotemporal variability of the resource prior to selecting a site for a tidal energy project. However, there are issues that must be addressed including the need for a method that could process such data and capture the dependencies in space, time, and between the variables, and quantify the associated uncertainty in a principled way. Vessel mounted ADCPs are often used for data collection of oceanographic parameters, with particular locations being sampled for long durations of time. In many practical applications, the interest is in knowing the general variability in a region, and making predictions at unsampled locations. In the case of tidal currents, knowledge of the spatial variability will enable in making better energy estimates, reduce risk to the investors, and also help in the optimization of the tidal energy farm layouts (Robins et al., 2015). From a data collection perspective, the goal is to formulate a sampling strategy that can extract as much information as possible from limited amount of data, both in space and in time. Several studies have analyzed spatial variability of oceanographic data from ship-based ADCP surveys (see, e.g., Dowd & Thompson, 1996; Gomis et al., 2001; Kincaid et al., 1996; Vennell & Beatson, 2009). The majority of such investigations aimed at achieving a parametric fit with spatial interpolation of oceanographic variables.

In this work we use ML techniques in the framework of Bayesian Gaussian processes (GPs) to analyze tidal current data over a spatial domain and then perform predictions. GP is a widely used method for performing inference in Bayesian ML (GP description provided in the next section). For a given data set, GP can learn a distribution of nonlinear functions that explains it. The characteristics of such distribution of functions are governed by the kernel function (also known as the covariance function) that defines the similarity between pairs of data points, thus discovering the structures in the data. As a probabilistic approach, it enables incorporation of uncertainty in all forms, be it noise, parameters, or the model (Ghahramani, 2015). The framework provides tools for handling data from dynamic, noisy, and observation costly (expensive to sample) environments (Roberts et al., 2013). We present a spatiotemporal GP approach to model sparse and irregularly sampled tidal current data sets that could resemble observations made by autonomous vehicles or vessels. The model can make predictions at any arbitrary location (sampled and unsampled) in and around the sampled spatial domain for all times, along with a quantification of uncertainties for such predictions. The next sections describe the ML approach, the tidal model, and the sampling strategies. This is followed by a case study of the Pentland Firth region in the United Kingdom with data from a numerical model, and discussions.

2. Gaussian Process

GPs belong to the Bayesian class of ML techniques and have been widely used in modeling of a variety of regression problems (Ghahramani, 2015; Rasmussen & Williams, 2006) including some in the oceanic setting (see, e.g., Hollinger et al., 2012). Among the other statistical ML techniques, support vector machines (Steinwart & Christmann, 2008) have been applied to several classification problems on oceanographic data (see, e.g., Li & Castagna, 2004). Other types of techniques used in wave and tidal modeling include Local Models inspired by chaos theory, which after combining observations and results from numerical model, updates output variables to forecast and correct errors generated by numerical model (see Babovic et al., 2004; Sannasiraj et al., 2004). Statistical approaches used in the analysis of spatiotemporal data sets include inverse methods that infer model (parameters) from measurements and empirical orthogonal functions that identify orthogonal basis components explaining variance in the data (Wikle et al., 2013). Spline-based interpolation techniques have also been used on different oceanographic problems (see, e.g., McIntosh, 1990; Trossman et al., 2009). Most of the known prediction techniques are parametric in nature; in contrast GPs represent a nonparametric approach to modeling unknown functions.

GP describes a probability distribution, but in the space of functions (see Rasmussen & Williams, 2006). Just like a Gaussian distribution is completely defined in terms of a mean vector and a covariance matrix, a GP is described in terms of its mean and covariance function (MacKay, 2003). A visual illustration as to

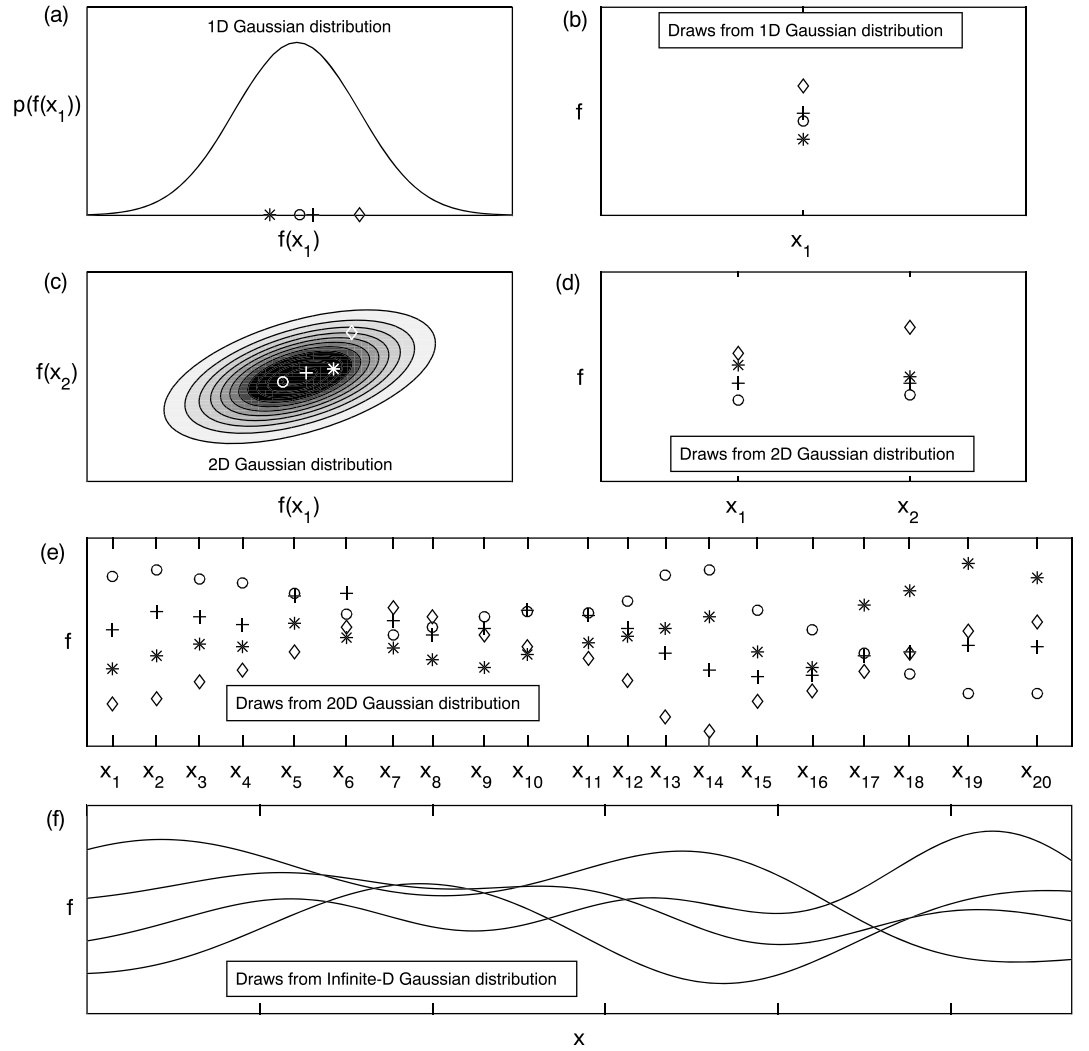


Figure 1. The plot in (a) shows a 1-D Gaussian distribution, and some random samples drawn from this distribution are shown in (b). Similarly, (c) shows the contour plot of the joint Gaussian distribution of the outputs $f(x_1)$ and $f(x_2)$ at two input locations x_1 and x_2 . Four random draws from this 2-D distribution have an equivalent representation in the input-output space as shown in the right plot (d). It is difficult to visualize higher dimensional distributions, but one could plot the draws from such distributions. Panel (e) shows some random draws from a 20-D Gaussian distribution—which is the joint distribution of 20 random variables. The bottom plot (f) shows a cartoon of draws from an infinite dimensional multivariate Gaussian distribution functions.

how one can move from a simple 1-D Gaussian distribution to a GP is shown in Figure 1. The covariance (kernel) function discovers structure in the data. Let $D = (x_i, y_i), i = 1, \dots, N$ be the set of given data points. The output values (y) are modeled as a sum of a latent function (f) and Gaussian white noise (ϵ) such that

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_n^2), \quad (1)$$

where the symbol “ \sim ” implies distributed according to, and σ_n^2 denotes the noise variance. The latent function f is modeled using a GP such that $f \sim \mathcal{GP}(m(x), K(x, x'))$ where $m(x)$ is the mean function and $K(x, x')$ is the covariance function defined as

$$m(x) = \mathbb{E}[f(x)], \quad (2)$$

$$K(x, x') = \text{cov}(f(x), f(x')). \quad (3)$$

The unknown parameters of the kernel function (discussed later) are known as hyperparameters θ . GP can be considered as a multivariate Gaussian distribution of infinite dimensions, and marginalization

(integrating out the uncertainty) over points not associated with the data results in an N -dimensional Gaussian distribution $p(\mathbf{f}|\theta) = \mathcal{N}(\mathbf{f}; \mathbf{m}, \mathbf{K})$, where $\mathbf{m} = [m(x_1), m(x_2), \dots, m(x_N)]$ is the mean vector and

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_N) \\ k(x_3, x_1) & k(x_3, x_2) & \cdots & k(x_3, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & \cdots & k(x_N, x_N) \end{bmatrix}$$

is the covariance matrix describing the correlations between the input data points. We are interested in deriving $p(\mathbf{y}^*|\mathbf{y})$ —the conditional distribution of the outputs at the test locations $\mathbf{x}^* = [x_1^*, x_2^*, \dots, x_L^*]$ given the input data points, also known as the posterior distribution of the outputs. The mean of such distribution describes the expected value of the outputs for the given data, while the variance expresses the associated uncertainties. First, using Bayes' rule, the conditional distribution of latent function values given the data points— $p(\mathbf{f}|\mathbf{y}, \theta)$, which is a function of hyperparameters θ —is derived:

$$p(\mathbf{f}|\mathbf{y}, \theta) = \frac{p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}|\theta)}{\int p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}|\theta)d\mathbf{f}}, \quad (4)$$

where $p(\mathbf{y}|\mathbf{f}, \theta) = \mathcal{N}(\mathbf{y}; \mathbf{f}, \sigma_n^2 \mathbf{I})$ is the likelihood function modeling the noise in the data with σ_n^2 as the noise variance, and $p(\mathbf{f}|\theta)$ is the prior distribution for the function θ . Note, in (4) the denominator is an expansion of $p(\mathbf{y}|\theta)$. The distribution of the latent function values at the training and test data points together can be expressed as

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & \mathbf{K}_{**} \end{bmatrix} \right) \quad (5)$$

where

$$\mathbf{k} = \begin{bmatrix} k(x_1, x_1^*) & k(x_1, x_2^*) & \cdots & k(x_1, x_L^*) \\ k(x_2, x_1^*) & k(x_2, x_2^*) & \cdots & k(x_2, x_L^*) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1^*) & k(x_N, x_2^*) & \cdots & k(x_N, x_L^*) \end{bmatrix}, \quad \mathbf{K}_{**} = \begin{bmatrix} k(x_1^*, x_1^*) & \cdots & k(x_1^*, x_L^*) \\ k(x_2^*, x_1^*) & \cdots & k(x_2^*, x_L^*) \\ \vdots & \ddots & \vdots \\ k(x_L^*, x_1^*) & \cdots & k(x_L^*, x_L^*) \end{bmatrix} \quad (6)$$

with \mathbf{k} being then $N \times L$ matrix of covariance between the training points \mathbf{x} and test locations \mathbf{x}^* , while \mathbf{K}_{**} is $L \times L$ covariance matrix of the test data points and \mathbf{m}_* is the mean vector at the test locations. Algebraic manipulations using Gaussian identities then yields the distribution $p(\mathbf{f}_*|\mathbf{f}, \theta)$ with mean $\mathbf{m}_* + \mathbf{k}^T \mathbf{K}^{-1}(\mathbf{f} - \mathbf{m})$ and variance $\mathbf{K}_{**} - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k}$. Marginalization of this distribution with respect to latent function at training points \mathbf{f} gives $p(\mathbf{f}_*|\mathbf{y}, \theta)$.

The distribution $p(\mathbf{f}_*|\mathbf{y}, \theta)$ can then be used along with the likelihood function for the test data points $p(\mathbf{y}^*|\mathbf{f}_*)$ to yield

$$p(\mathbf{y}^*|\mathbf{y}, \theta) = \int p(\mathbf{y}^*|\mathbf{f}_*)p(\mathbf{f}_*|\mathbf{y}, \theta)d\mathbf{f}_* = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*) \quad (7)$$

where the expressions for the mean $\boldsymbol{\mu}_*$ and variance $\boldsymbol{\sigma}_*$ are given by

$$\boldsymbol{\mu}_* = \mathbf{m}_* + \mathbf{k}^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m}) \quad (8)$$

$$\boldsymbol{\sigma}_* = \mathbf{K}_{**} + \sigma_n^2 \mathbf{I} - \mathbf{k}^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}, \quad (9)$$

(see Rasmussen & Williams, 2006), where \mathbf{I} is the identity matrix. Note, the test locations are the points at which the predictions are sought. The posterior distribution $p(\mathbf{y}^*|\mathbf{y}, \theta)$ is a function of unknown hyperparameters θ that needs to be marginalized as

$$p(\mathbf{y}^*|\mathbf{y}) = \int p(\mathbf{y}^*|\mathbf{y}, \theta)p(\theta|\mathbf{y})d\theta, \quad (10)$$

where $p(\theta|\mathbf{y})$ is the posterior distribution of the hyperparameters. We adopt the maximum a posteriori (MAP) approach in estimating the integral in (10) where the hyperparameter posterior $p(\theta|\mathbf{y})$ is considered to be a delta function at its maximum (see, e.g., Watanabe & Chien, 2015). The mean function \mathbf{m} (see (2)) is considered to be constant in the present application.

2.1. Kernel Function

One of the key aspects of the GP approach is the kernel functions that expresses the structures in the data. Kernel-based ML methods offer a rigorous approach in modeling functions; such methods are, in some respects, more interpretable than alternatives, including neural networks (Hofmann et al., 2008). Many of the classical statistical methods are based on linear models; however, most real-world problems demand nonlinear techniques to capture structures in the data. Kernel-based methods exploit the best of both these approaches by performing a nonlinear transformation of the data to a high dimensional space and then applying linear methods of statistics in that space. Some of the popular kernel functions include the exponentiated quadratic, *Matérn*, periodic (MacKay, 2003), spectral-mixture (Wilson & Adams, 2013). The kernel function comprises hyperparameters that decide the properties of functions chosen from the GP prior. For example, the length scale hyperparameter in the exponentiated quadratic regulates the dependency horizon with the long length scales appropriate for modeling gradual changes with respect to the input variables (Topa & Honkela, 2015). Simplistically, the kernel function is an index of the similarity between data points, and two input data points are highly correlated if they are close to each other.

2.2. Choice of Kernel Function

Selection of an appropriate kernel function is crucial for GP-based algorithms. Generic kernel functions like the exponentiated quadratic can work with arbitrary data sets; however, they fail to perform predictions at locations that are more than a length scale (a kernel function hyperparameter) away from the data set. Various combinations of the kernel functions are possible as the sums and products (including tensor product) of positive semi-definite kernels are also positive semi-definite. Such generalizations enable us to model our complex problem that demands a kernel function that can account for both the space and time input correlations, as well as that between the outputs: the two horizontal components of velocity u and v . Until now the correlations between input data points have been discussed. The correlation between the outputs— u and v —can be accommodated with the aid of a separate kernel component Y . The resultant kernel K can be represented as the kronecker product \otimes of two kernel components, modeling the correlations between the inputs and outputs, respectively. In such setting, the output data vector \mathbf{y} would comprise both the velocity components, that is, $\mathbf{y} = [u_1, u_2, \dots, u_N, v_1, v_2, \dots, v_N]$. Given the nature of tidal data, summation of periodic kernel components are considered to model contributions from disparate frequencies (Sarkar et al., 2018). The final form of the covariance function can be expressed as

$$K = \sum_{q=1}^Q Y_q \otimes K_q \quad (11)$$

and the matrix Y_q is assumed to be of the spherical parameterization kind (Osborne et al., 2012; Pinheiro & Bates, 1996):

$$Y_q = \begin{pmatrix} \sigma_{u,q}^2 & \sigma_{u,q}\sigma_{v,q}\cos\chi_q \\ \sigma_{u,q}\sigma_{v,q}\cos\chi_q & \sigma_{v,q}^2 \end{pmatrix}, \quad (12)$$

while

$$K_q(i, j) = \exp\left(-\frac{2}{l_q^2} \sin^2\left(\frac{\pi|t_i - t_j|}{p_q}\right)\right) \exp\left(-\frac{1}{2} \begin{bmatrix} (x_i - x_j) \\ (y_i - y_j) \end{bmatrix}^T \mathbf{M}_q^{-1} \begin{bmatrix} (x_i - x_j) \\ (y_i - y_j) \end{bmatrix}\right) \quad (13)$$

where

$$\mathbf{M}_q = \begin{pmatrix} l_{x,q}^2 & \rho_q l_{x,q} l_{y,q} \\ \rho_q l_{x,q} l_{y,q} & l_{y,q}^2 \end{pmatrix} \quad (14)$$

is the Mahalanobis covariance matrix. Note in (12)–(14), $\sigma_{u,q}^2$ and $\sigma_{v,q}^2$ are the variance of the outputs, χ_q is associated with spherical parameterization and models the correlation between outputs, p_q is the period, l_q is the length scale associated with the temporal kernel component, $l_{x,q}$ and $l_{y,q}$ are the spatial length scales in the x and y directions, respectively; and $\rho_q l_{x,q} l_{y,q}$ is the covariance between the two length scales; thus, there are eight hyperparameters associated with the q th kernel subcomponent. The second exponential term in

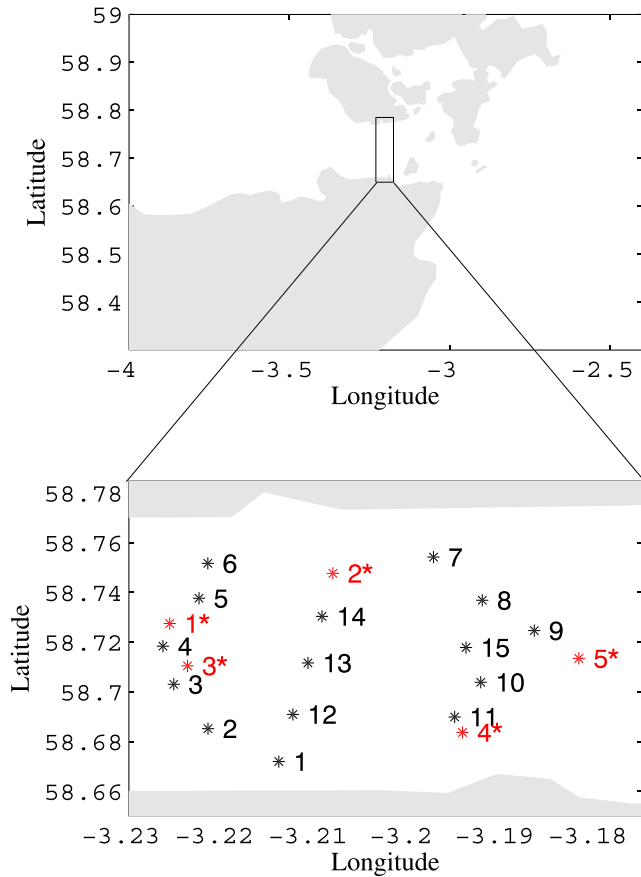


Figure 2. Data sampled from locations marked in black (1 to 15) are used to train the machine learning algorithm, while locations marked in red (1* to 5*) are taken to be unsampled. The latitude and longitude of the sampled points (shown in the figure) are given in Table 1. The regions in gray indicate land. The figure is just for illustration.

(13) can be written as $\exp(-d_M^2/2)$, where d_M is the Mahalanobis distance (De Maesschalck et al., 2000) that generalizes the standard Euclidean distance by enabling linear scalings and rotations of the feature space (Davis et al., 2007). The periodic kernel component accounts for the periodicity in time, while the exponentiated quadratic models the spatial variations. The structure of the kernel component capturing the time harmonicity is similar to that used by Sarkar et al. (2018) —with the capability of capturing nonsinusoidal variations and nonharmonic effects (in the short term). The effect of any predominant nontidal frequency can also be modeled using an additional kernel component with all the hyperparameters inferred from the optimization scheme.

2.3. Hyperparameter Estimation

In a full Bayesian approach, the hyperparameters θ ($\sigma_{u,q}^2$, $\sigma_{v,q}^2$, χ_q , p_q , l_q , $l_{x,q}$, $l_{y,q}$, ρ , with $q = 1, 2, \dots, Q$) are marginalized (integrated over all possible values, see 10) using stochastic numerical techniques like Markov chain Monte Carlo (MCMC). However, such techniques are computationally demanding—convenient for problems with small data sets and hyperparameters. The MAP approach, a widely adopted technique in the ML community, is pursued here for hyperparameter estimation. It belongs to the class of approximate deterministic inference techniques. In this approach, the hyperparameter posterior $p(\theta|\mathbf{y})$ is approximated as a delta function δ at its maximum $\hat{\theta}$, that is, $p(\theta|\mathbf{y}) = \delta(\theta - \hat{\theta})$. By virtue of this simplification the posterior distribution $p(\mathbf{y}^*|\mathbf{y})$ (see 10) can then be approximated as $p(\mathbf{y}^*|\mathbf{y}, \hat{\theta})$ with

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \\ &= \arg \max_{\theta} p(\mathbf{y}|\theta)p(\theta)\end{aligned}\quad (15)$$

where $p(\theta)$ is the hyperparameter prior distribution. The expression in (15) can be simplified and the optimization is eventually performed on

$$\log p(\hat{\theta}|\mathbf{y}) = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2} |\mathbf{K} + \sigma_n^2 \mathbf{I}| - \frac{N}{2} \log(2\pi) + \log p(\hat{\theta}). \quad (16)$$

The first term in (16) penalizes the difference between the data and the model output, the second term penalizes the model complexity, and the fourth term tries to maximize the probability of the hyperparameter prior distribution at $\hat{\theta}$; the third term can be ignored from the optimization process as it is a constant.

The hyperparameters of kernel functions dictate the higher-level characteristics of the model—the properties of functions chosen from the GP distribution. They do not necessarily have correspondence to variables in physics-based models, although some of them have interpretations, for example, the frequency, length scale hyperparameters, which can be related to the physical properties. The periodic hyperparameters of the kernel function — $p_q s$ — are set at periods determined from the automated selection algorithm of Foreman (1977), with the capability of capturing more nonlinearities than traditional techniques. In general, for any arbitrary ML problem, all the hyperparameters are inferred from the data; however, for tidal data sets, prior knowledge of the (precise) frequencies assists in achieving robust and quicker solutions. The length scale hyperparameters in (14) governs the range of spatial dependencies—a longer length scale suggests a slower spatial variation of currents associated with the particular kernel component.

Ideally, the kernel function is intended to represent all the structures in the data, and the mean function is taken to be constant (usually zero). The mean function in its usual form has a parametric description, making it more prone to overfitting. However, in certain instances, it can be efficacious to use

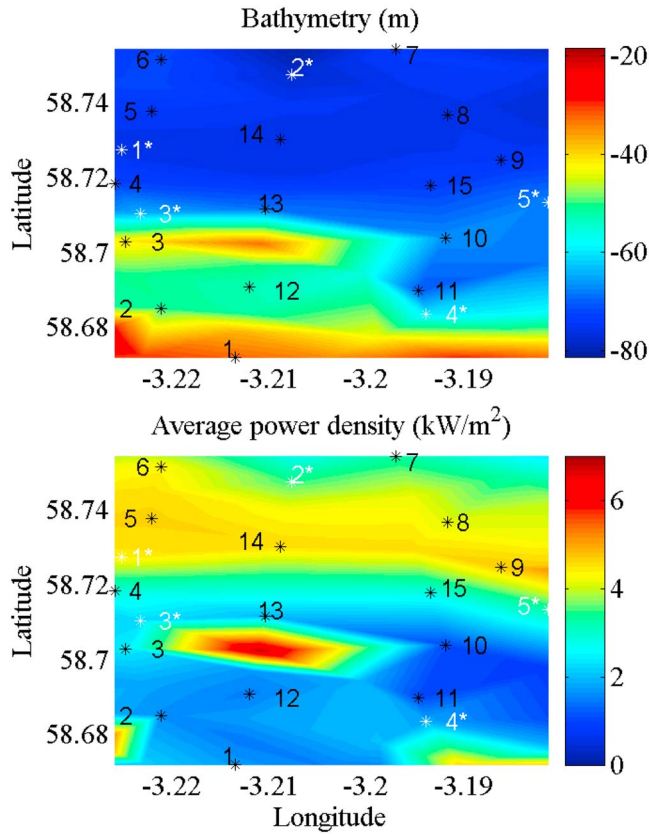


Figure 3. Spatial variation of bathymetry (upper plot) and average power density (lower plot) of 91 days of training data (from 1 January 2014 to 1 April 2014) around the sampled region. The black markers indicate the points used for training (locations 1 to 15).

the mean function, for example, in capturing nontidal contributions for short-term prediction of currents (see Sarkar et al., 2018).

3. Tidal Model

The analysis is performed with data obtained from numerical simulation of depth-averaged tides in the Pentland Firth region (see Figure 2). This region in Scotland is considered to be a prime location for installation of large arrays of tidal turbines. The model was tuned and comparisons were made with limited field data measurements in Adcock et al. (2013). The numerical model, based on an irregular grid, was forced with eight tidal constituents—principal lunar semidiurnal (M_2), principal solar semidiurnal (S_2), larger lunar elliptic semidiurnal (N_2), lunisolar semidiurnal (K_2), variational (MU_2), larger lunar evictional (NU_2), lunar diurnal (O_1), lunar diurnal (K_1), and the domain extends to the continental shelf to the west and an approximately equal distance to the east. In the relevant region, the numerical model had a mesh resolution of around 600 m. The governing equations of the depth-averaged model are

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \mathcal{U} \hat{\mathbf{z}} \times \mathbf{u} = -g \nabla \zeta - \frac{\gamma_b |\mathbf{u}| \mathbf{u}}{H + \zeta} + A_h \nabla^2 \mathbf{u}, \quad (17)$$

$$\frac{\partial \zeta}{\partial t} + \nabla \cdot \{(H + \zeta) \mathbf{u}\} = 0, \quad (18)$$

where \mathbf{u} is the depth-averaged horizontal velocity, t indicates time, $\hat{\mathbf{z}}$ is the vertical unit vector, g is the acceleration due to gravity, ζ is the free surface elevation, H is depth below the mean sea level, γ_b is the coefficient of bottom friction, A_h is the coefficient of the horizontal eddy viscosity, \mathcal{U} is the Coriolis parameter, and $\nabla = (\partial/\partial x, \partial/\partial y)$ is the horizontal differential operator. The bathymetry around the sampled locations used in the tidal model is shown in Figure 3. The figure also depicts the average power density in the region over 91 days. There are significant changes in both the variables around the sampled locations.

Because the ML method is ideally intended to be applied to real-world data, we consider the current data (from numerical simulations) at a particular location for a certain period of time, equivalent to using an ADCP at the location for the same period. We refer to them as virtual ADCPs in this work.

Table 1
Location of the Tidal Observation Sites

Location	Latitude	Longitude	Depth (m)
1	58.6720°N	−3.2139°W	32.44
2	58.6852°N	−3.2215°W	50.86
3	58.7030°N	−3.2252°W	41.49
4	58.7184°N	−3.2264°W	72.28
5	58.7377°N	−3.2225°W	74.38
6	58.7517°N	−3.2215°W	71.94
7	58.7542°N	−3.1974°W	73.98
8	58.7368°N	−3.1921°W	75.83
9	58.7247°N	−3.1865°W	74.48
10	58.7039°N	−3.1923°W	66.80
11	58.6899°N	−3.1951°W	70.37
12	58.6909°N	−3.2124°W	52.85
13	58.7116°N	−3.2108°W	68.86
14	58.7304°N	−3.2093°W	76.08
15	58.7179°N	−3.1939°W	73.07

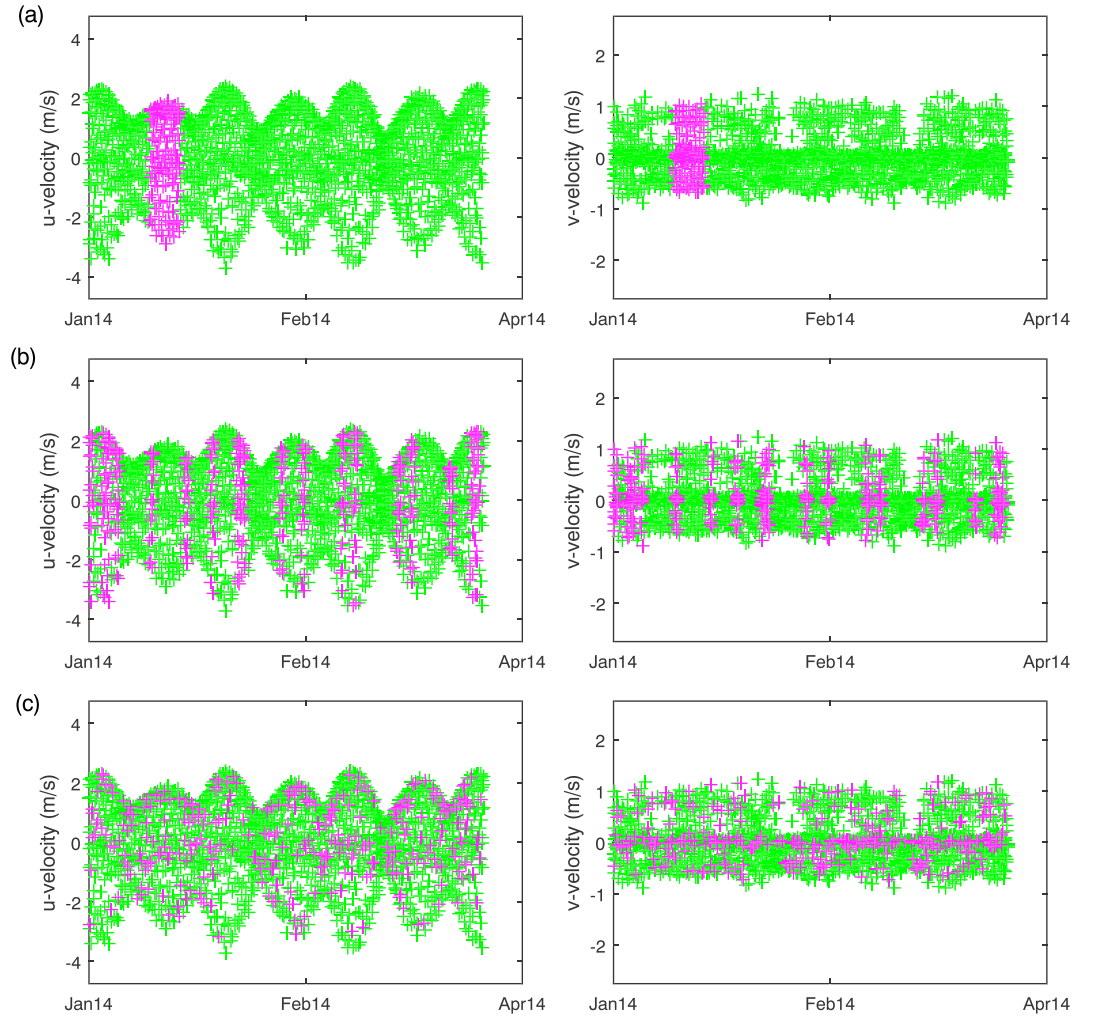


Figure 4. Training data and predictions at location 13 using three different sampling strategies: (a) GPweek, (b) GPday, and (c) GPhour. The data points in magenta “+” indicate the sampled time instances at location 13 using the respective sampling strategies. The green “+” markers indicate the complete data set available from the numerical model and is shown only for convenience. The GP algorithm is trained using the data points in magenta color only. Data collection is made over a period of 91 days starting from 1 January 2014. Two virtual acoustic Doppler current profilers were used in sampling in all the three cases.

3.1. Harmonic Analysis

We provide a brief description of the harmonic analysis approach. The tidal current data are modeled as a complex vector: $y(t) = u(t) + iv(t)$, where $u(t)$ and $v(t)$ are horizontal components of the current velocity. The time series outputs are expressed as the sum of a finite number of sinusoidal basis with some offset and trend as

$$y(t) = \sum_{k=1}^n \left(a_k^+ e^{i\omega_k(t-t_0)+iV_k} + a_k^- e^{-i\omega_k(t-t_0)-iV_k} \right) + c_0 + c_1(t - t_0), \quad (19)$$

where ω_k is the angular frequency of the k th constituent, a_k^+ and a_k^- are the unknown complex amplitudes, V_k is some astronomical argument, while c_0 and c_1 are the mean and trend slope, respectively. The standard least squares approach is adopted in this work to determine the solution to the coefficients. The constituents are selected from a basis of 69 tidal constituents (45 astronomical and 24 shallow water) using the automated selection algorithm of Foreman (1977). The selection is made using the Rayleigh criterion that requires two neighboring tidal constituents with frequencies f_i and f_j to satisfy the condition

$$|f_i - f_j| \geq \frac{R_{min}}{T_0}, \quad (20)$$

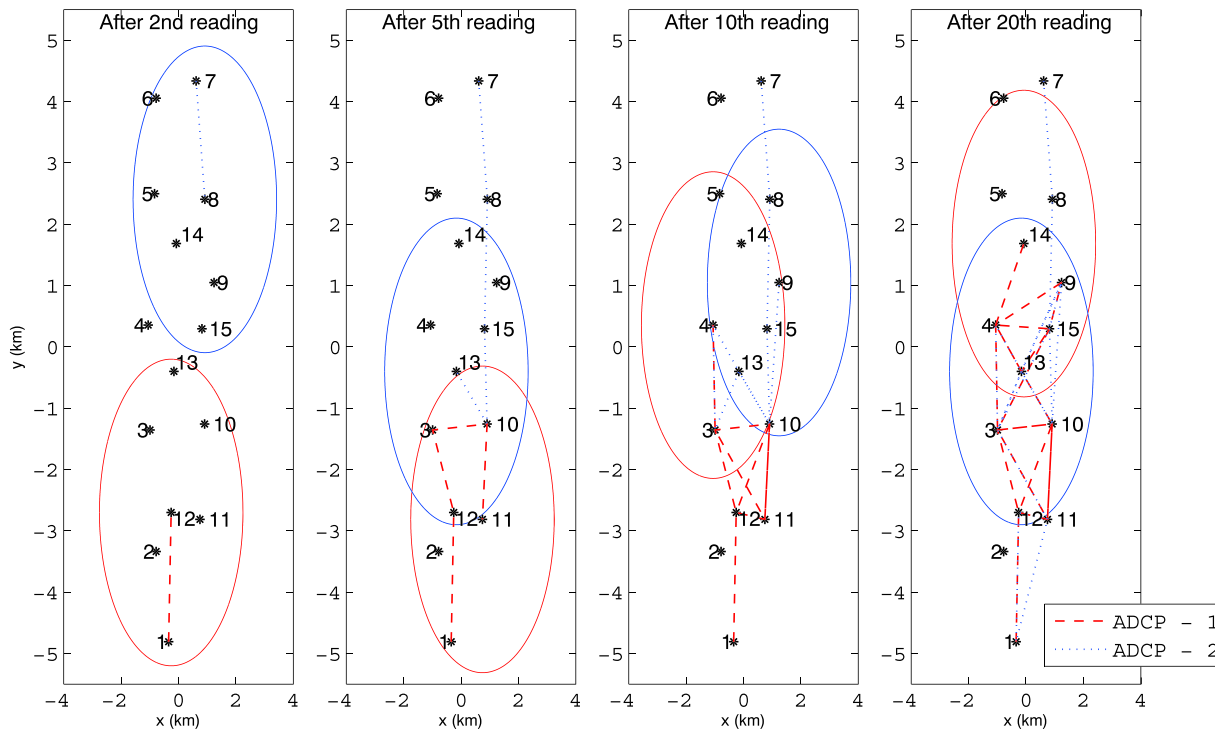


Figure 5. The track of the two vessels (with ADCPs) during the second, fifth, tenth, and twentieth reading. After the measurements are made, the vessels move to another location in the region encompassed by the circle (of radius 2.5 km) centered at the present location. The vessel with ADCP-1 starts at location 1 while the other starts at location 7. Next, they move to locations 12 and 8. ADCPs = acoustic Doppler current profilers.

where T_0 is the effective time duration of the whole record and R_{\min} is the Rayleigh threshold which is usually taken to be one—this will be taken as the default value in this work, unless and otherwise indicated. Nodal corrections with exact times are implemented.

4. Sampling

We perform the analysis with disparate sampling strategies. Locations are considered in the Pentland Firth region and are sampled on an hourly basis with the aid of two virtual ADCPs—so at any point in time the two virtual devices are capturing data at two different locations, simultaneously. Different cases of data sampling are considered depending on the time spent continuously (stoppage time) at a particular location—the virtual ADCPs are considered to be positioned at a particular location over periods of 1 week (GPweek), 1 day (GPday), and 1 hr (GPhour). The last case (GPhour) corresponds to a situation in which the virtual ADCPs are making discrete movement from one location to another every 1 hr. Data are sampled at 15 different locations and Table 1 shows the latitude and longitude of these locations (see Figure 2). The specific choices of spatiotemporal sampling (using three different strategies at 15 locations) are made for simplicity. The sequences of sampling the 15 chosen locations, for all the strategies discussed above, are selected randomly. The number of times a particular point is sampled at nonconsecutive instances of time (sampling made at other locations in the intermediate time) vary, depending on the randomized selection and also on the sampling strategy adopted, for example, in GPhour scheme the sampling locations change frequently (after every observation) in contrast to GPweek where locations change after 1 week. In a realistic situation, sampling could be less random in both space and time, and the ML approach would be equally capable of handling such data. Examples of the sampled data (used for training the ML algorithm) at one of the locations (no. 13) using the three different sampling strategies are shown in Figure 4—(a) GPweek, (b) GPday, and (c) GPhour. Note, in all the three cases shown in the plot, data collection was made using two virtual ADCPs in the region illustrated in Figure 2. The top figure shows the case in which sampling locations are changed after every 1 week, and location 3 was sampled for a cumulative time of 2 weeks (data points in magenta) within the 91 days of the data-sampling campaign.

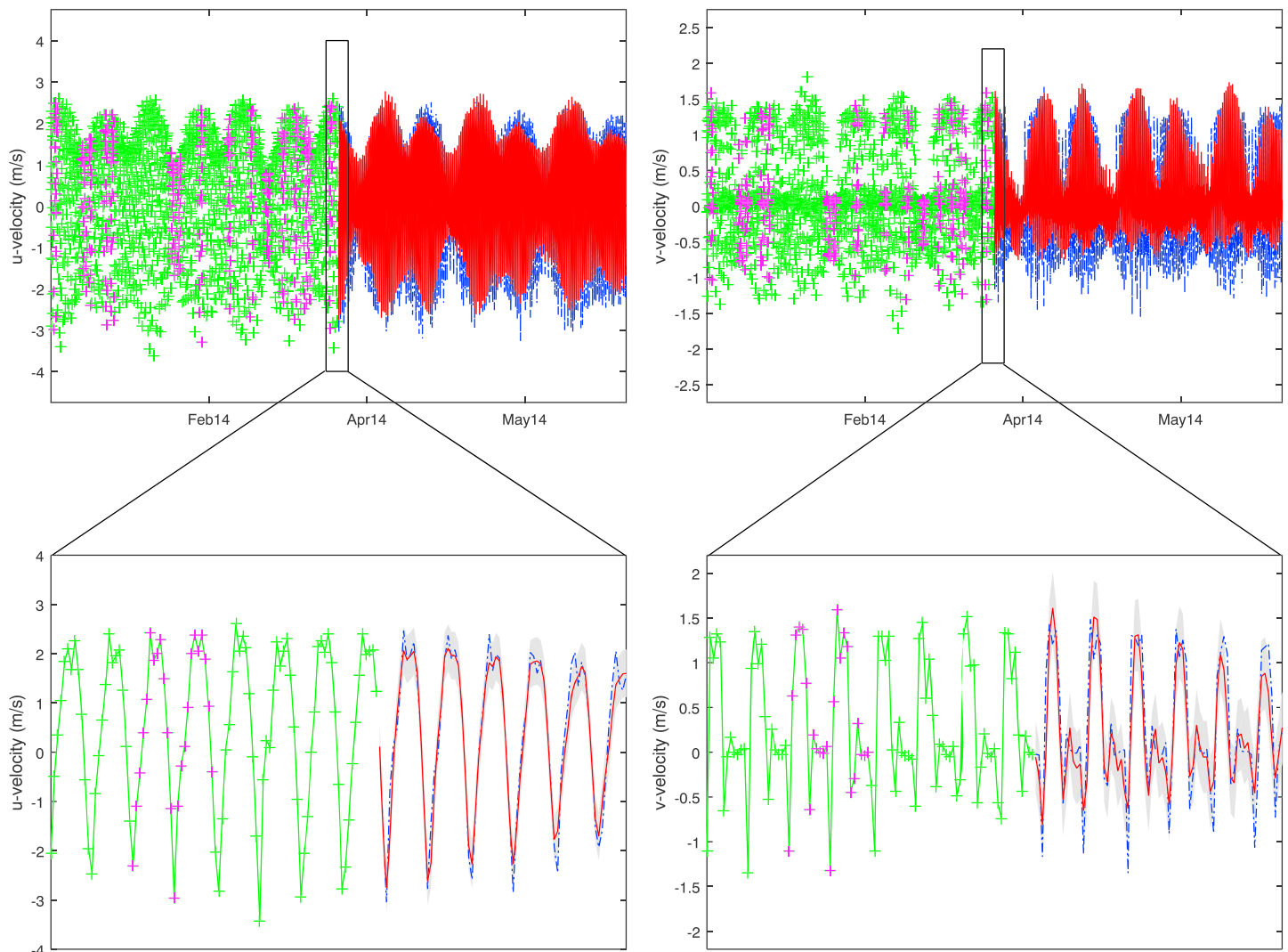


Figure 6. The top figures show the horizontal components of the current velocity at a particular location (location 3 in Table 1). The bottom figures show a zoomed portion of the top plots. The position of the rectangular box is chosen such that it covers a portion of the training and test (prediction) period simultaneously. The data points in magenta “+” indicate the sampled points at the location used to train the algorithm using the GPday strategy. The red curve indicates the mean of the Gaussian processes predictions (see (2)), while the blue dashed curve shows the actual values (from numerical simulations). The region in gray indicates the 95% confidence level. The data points in green are only shown for convenience and only the points in magenta are used for training the machine learning algorithm. Such a presentation is possible as prior knowledge of tidal current variations at all times and locations are known from the numerical model. Note, green traces are drawn through the first 91 days of data (shown in green “+”) in the lower plots for convenience.

A short comparison is also performed with a more realistic sampling scheme that could closely resemble the tracks of vessel-based surveys; we call it GPreal. Note, the term “vessel” is used generically to include crew-borne vessels as well as autonomous vehicles. The data sampling is done as follows: the vessel starts its motion from a particular location. The next location is then selected randomly from among the few nodes lying within a certain distance (we consider a boundary with radius 2.5 km from the present location). The last step is repeated until the end, with a restriction that the vessel cannot go back to its previous node (from where it came to its present location). There is a time interval of 1 hr between each node, which makes this strategy (GPreal) similar to GPhour except that the vessel is much constrained in its movement. We consider two vessels starting from opposite sides from the sampled region—one from location 1 and the other from location 7 (see Figure 5). The locations of the vessels during second, fifth, tenth, and twentieth reading are shown in Figure 5. The ML algorithm for modeling the spatiotemporal tidal current data was implemented in MATLAB, making use of the basic functionalities of the existing GPML toolbox. However, the complexity of the problem demanded several modifications and additions to the toolbox, for example, multioutput,

Table 2
Comparison of the RMSE (in m/s) of the Prediction Using the HA and GP Algorithm

Loc.	<i>u</i>					<i>v</i>				
	GPweek	GPday	GPhour	GPreal	HA	GPweek	GPday	GPhour	GPreal	HA
1	0.605	0.551	0.495	0.506	0.450	0.250	0.199	0.194	0.194	0.173
2	0.348	0.330	0.329	0.364	0.321	0.218	0.194	0.193	0.208	0.166
3	0.545	0.411	0.402	0.398	0.382	0.469	0.411	0.420	0.412	0.386
4	0.379	0.264	0.271	0.271	0.299	0.316	0.242	0.257	0.257	0.240
5	0.409	0.276	0.271	0.283	0.318	0.199	0.159	0.167	0.168	0.150
6	0.349	0.281	0.258	0.302	0.299	0.159	0.130	0.136	0.133	0.121
7	0.261	0.249	0.240	0.264	0.279	0.152	0.149	0.147	0.149	0.138
8	0.303	0.258	0.241	0.249	0.297	0.204	0.176	0.179	0.173	0.161
9	0.338	0.256	0.259	0.266	0.317	0.268	0.229	0.233	0.224	0.209
10	0.578	0.454	0.450	0.436	0.406	0.431	0.325	0.320	0.310	0.301
11	0.354	0.338	0.322	0.332	0.309	0.405	0.393	0.370	0.372	0.336
12	0.280	0.284	0.283	0.292	0.271	0.303	0.297	0.295	0.284	0.261
13	0.437	0.349	0.363	0.346	0.340	0.327	0.285	0.285	0.281	0.265
14	0.340	0.267	0.277	0.276	0.323	0.232	0.193	0.192	0.196	0.177
15	0.363	0.274	0.272	0.271	0.302	0.284	0.257	0.252	0.249	0.227
Avg.										
RMSE	0.393	0.323	0.316	0.324	0.328	0.281	0.243	0.243	0.241	0.221

Note. The root-mean-square error (RMSE) from the HA are computed with complete 91 days of data at each location, separately. The background cell colors used in the table indicate different regimes of RMSE: $0.10 < RMSE \leq 0.20$ (dark-blue), $0.20 < RMSE \leq 0.30$ (light-blue), $0.30 < RMSE \leq 0.40$ (yellow), $0.40 < RMSE \leq 0.50$ (green), $0.50 < RMSE \leq 0.60$ (light-red), and $0.60 < RMSE \leq 0.70$ (dark-red). The first column indicates the sampled locations, columns two to five show the RMSEs of the *u* velocity component, while the remaining columns are for the *v* velocity component corresponding to different cases of sampling and HA.

spatiotemporal kernel components. Comparisons are made with HA, however, with a different data set—the complete continuous time series data at all the locations, with a temporal sampling frequency of once every hour (same as the sparse spatiotemporal data set). The HA implementation has been made using the UTide MATLAB toolbox (Codiga, 2011).

We will make comparisons of the spatiotemporal GP approach employed on data sets from different sampling strategies with HA performed on 91 days of data at all the sampled locations—1 to 15 (see Figure 2). The latter data (used for HA) corresponds to the situation where fixed ADCPs collect data at all the sampled locations over the whole sampling duration. We will also check the quality of predictions with the GP model at unsampled locations—1* to 5* (see again Figure 2). Because HA is implemented using data from 15 fixed ADCPs (at locations 1 to 15), the input training data for the unsampled locations (1* to 5*) are derived using linear interpolation of hourly sampled 91 days of data from all the 15 fixed ADCPs. The linear interpolation is performed using kernel function of the following form:

$$K(i, j) = (1 + b - 2b|x_i - x_j|)(1 + b - 2b|y_i - y_j|) \quad (21)$$

where *b* is an unknown hyperparameter (Minka, 2000).

5. Results

The data from the different sampling schemes were modeled using the GP formulation. A visual demonstration of the data and results is shown in Figure 6 which depicts the variation of the horizontal components of velocity *u* and *v* versus time at location 3. The sampling campaign spanned 91 days, starting from 1 January 2014, and the data collected were used to train the GP algorithm. This model was then used to make predictions for the following 91 days, starting from 2 April 2014. Because the figure plots the variations over a period of 6 months (91 days of training + 91 days of predictions), the exact details are not properly visible; however, the gist is clear. A detailed presentation of the RMSE of the future predictions (91 days) at the 15 locations using the three approaches is made in Table 2. The model captures

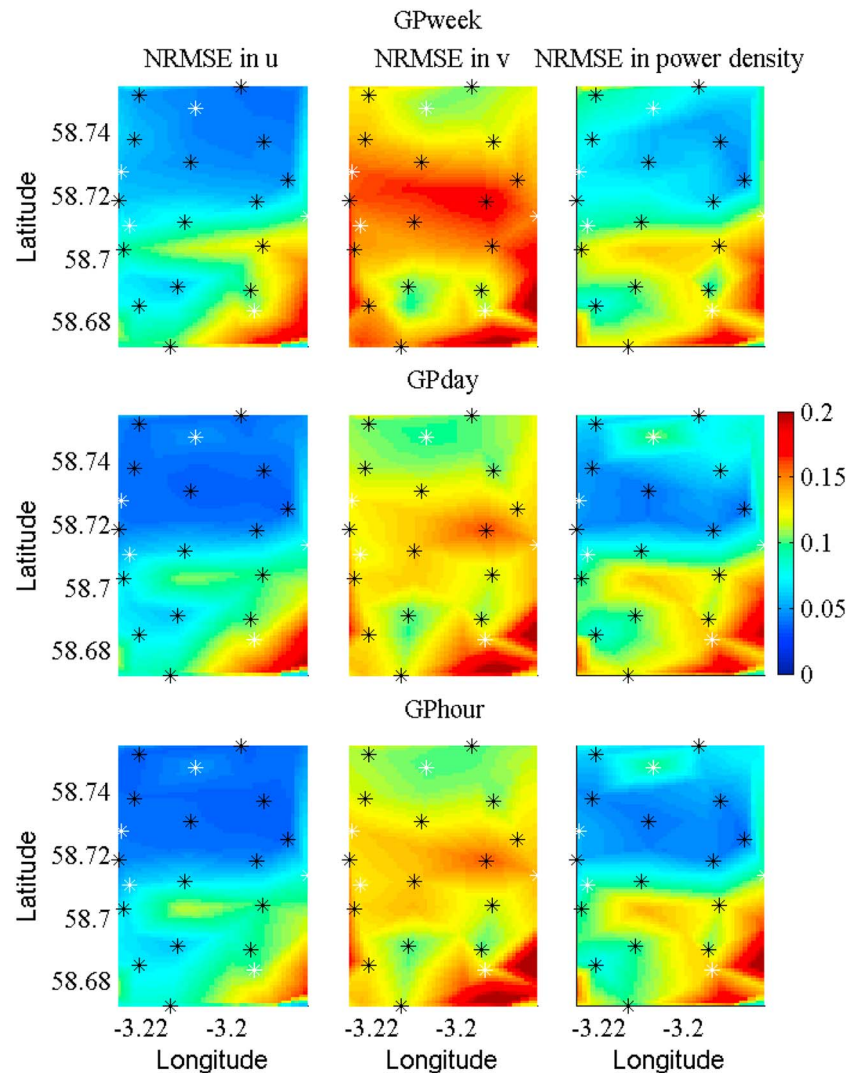


Figure 7. Spatial variation of the NRMSEs, from the different sampling approaches: (top plots) GPweek, (middle plots) GPday, and (bottom plots) GPhour. The black markers indicate the points used for training (locations 1 to 15), and white markers indicate unsampled locations. NRMSE = normalized root-mean-square error; GP = Gaussian process.

the general spatial variation of the currents quite well, except for a few cases. Notably, location 1 (see Figure 2), which is isolated from the other points and is situated near to the coast, unlike other points, exhibits the highest combined RMSE among the sampled locations in the predictions of u and v (see Table 2). The last row shows the average RMSE of all the locations, and clearly GPhour yields the least average error in the predictions. Such a result could be due to aliasing, and one way of overcoming that is by random sampling. By construction, GPweek strategy has more regularity in sampling than GPhour, which could partly explain the behavior. The last columns of u and v shows the RMSE of the predictions from the HA approach, but it uses the full hourly sampled 91 days of data at each of the locations to make predictions for the next 91 days. An R_{\min} value of 1 led to the selection of a total of 35 tidal constituents using 91 days of data. It is used for benchmarking the quality of the spatiotemporal predictions. In Figure 7, we depict the spatial variation of errors of variables in terms of normalized-RMSE (NRMSE)—defined as RMSE divided by the difference between maximum and minimum of the variable.

Table 3
Comparison of the Average RMSE (in m/s) Between GPhour and HA for Different Rayleigh Thresholds

R_{\min}	n_c	u		v	
		GPhour	HA	GPhour	HA
0.25	59	0.298	0.290	0.245	0.259
1.00	35	0.316	0.335	0.243	0.237
3.00	29	0.365	0.357	0.245	0.242
4.00	17	0.429	0.435	0.245	0.251

Note. n_c indicates the number of tidal constituents selected with the particular R_{\min} value.

Figure 7 shows the NRMSE of the horizontal components of velocity and power density around the sampled region for the three sampling strategies (top plot = GPweek; middle plot = GPday; bottom plot = GPhour). Note the higher levels of NRMSE in the v velocity component (second column) compared to the u component—this could be due to the (marginally) lower values of RMSE of v compared to that of u , and to the triggering of more complex flow structures in the cross-channel direction as the incoming current interacts with the topography and bathymetry features around the Pentland Firth region (see Figure 2). Relatively large NRMSEs can be observed at the bottom right corner region of all the subplots which lies outside of the sampled domain; however, the errors decrease from GPweek to GPhour strategies. Also, higher NRMSEs were between locations 12 and 13 across all the three sampling strategies; such discrepancies could be attributed to sharp changes in bathymetry in the specific region (see Figure 3) in addition to being unsam-

pled (none of the 15 locations used to train the ML algorithm lie between locations 12 and 13). The results demonstrate the importance of sampling in regions with large gradients in bathymetry.

The analysis is also performed using GPreal sampling strategy for which the average RMSEs for future predictions of u and v velocities were 0.324 and 0.241 m/s, respectively. In general, the results from GPreal strategy are similar to those from GPhour (see Table 2), demonstrating the robustness of the approach to different sampling strategies.

5.1. Effect of Rayleigh Threshold R_{\min}

GP, unlike HA, is a nonparametric approach as the number of parameters grows with data. However, in the presented GP formulation the number of kernel function components (Q in (11)) is dependent on the Rayleigh threshold R_{\min} . This factor (R_{\min}) indicates the number of tidal constituents n_c in the HA. Most of the results and comparisons in this work are performed with R_{\min} value of 1, usually taken as default in tidal analysis. Results for comparisons between the HA and GPhour strategy for different values of R_{\min} are shown in Table 3. There is an increase in the overall average RMSE with the larger values of R_{\min} which leads to the selection of fewer number of tidal constituents. A R_{\min} value of 1 is usually considered to be conservative for less noisy data sets, and indeed we observe that a lower R_{\min} value of 0.25 slightly reduces the combined average RMSE. Overall the GPhour approach consistently performs similar to the HA (using continuous data at all locations), highlighting its robustness with respect to model complexity. Note, a lower R_{\min} value entails consideration of additional kernel components in (11), and concurrently inferring a larger number of hyperparameters. For example, in case of $R_{\min} = 0.25$, a total of 474 hyperparameters (kernel - $8 \times 59 + \text{likelihood/noise} - 2$) were needed to be inferred by the algorithm.

5.2. Sampling With Different Numbers of ADCPs

The analysis until now was made with data collected with two virtual ADCPs simultaneously. One would expect better predictions on using more ADCPs (simultaneously) to sample data. To explore this issue, two other sampling cases are considered: one in which just a single virtual ADCP is used for data collection, and the other with three virtual ADCPs. The results are presented in Table 4. In general, there is a decrease in the combined RMSE of the predictions with increase in number of virtual devices, which can be attributed primarily to the availability of a greater number of training data points. Remarkably, predictions with data

Table 4
Comparison of the Average RMSE (in m/s) in Three Different Cases With GPhour (the Locations of the Virtual ADCPs Changed Every Hour)

Devices	u	v
1 ADCP	0.323	0.246
2 ADCPs	0.316	0.243
3 ADCPs	0.315	0.243

Note. RMSE = root-mean-square error; ADCPs = acoustic Doppler current profilers.

sampled from a single moving virtual ADCP yields similar overall accuracy (average RMSE) as that from the HA. This result suggests that using the GP approach, a single moving data source can provide a description as good as large number of fixed ADCPs covering the region under consideration. It is interesting to note that the ML approach using the GPhour strategy essentially uses the same number of data points as a fixed ADCP would collect at just one particular location over a continuous period of time. However, the HA employed on the latter data can make predictions at only the discretely sampled locations, while the spatiotemporal GP can make predictions throughout the sampled region with just a sparsely sampled data set.

Table 5
Location of the Tidal Prediction Sites With No Observations

Location	Latitude	Longitude	Depth (m)
1*	58.7275°N	−3.2256°W	76.41
2*	58.7477°N	−3.2082°W	77.76
3*	58.7107°N	−3.2237°W	64.32
4*	58.6837°N	−3.1943°W	60.36
5*	58.7135°N	−3.1818°W	67.70

5.3. Predictions at Unsampld Locations

One of the key features of this method is its ability to generate predictions at arbitrary locations across the sampled region—something not possible with HA. The quality of such predictions at a few unsampled locations (see Table 5) is now briefly analyzed. Figure 8 plots the time variation of the velocity components at those locations. The limits of the ordinate of all the u velocity plots, corresponding to different spatial locations, are maintained. The same is done for the v component. This is done to achieve a visual understanding of the spatial variability. Table 6 shows the RMSE for those cases and comparisons are made with predictions from HA. The training data for the HA were generated

using linear interpolation of complete temporal data sets (hourly sampled data over 91 days) at all 15 sampled locations. The GP results show reasonable predictive accuracy, significantly better than those from HA, demonstrating the efficacy of the methodology.

5.4. Periodic Versus Cosine Kernel Components

As discussed in section 2.1, the kernel function of GP plays the important role of capturing structures in the data. For modeling of the spatiotemporal tidal current data set, the kernel was carefully selected to express

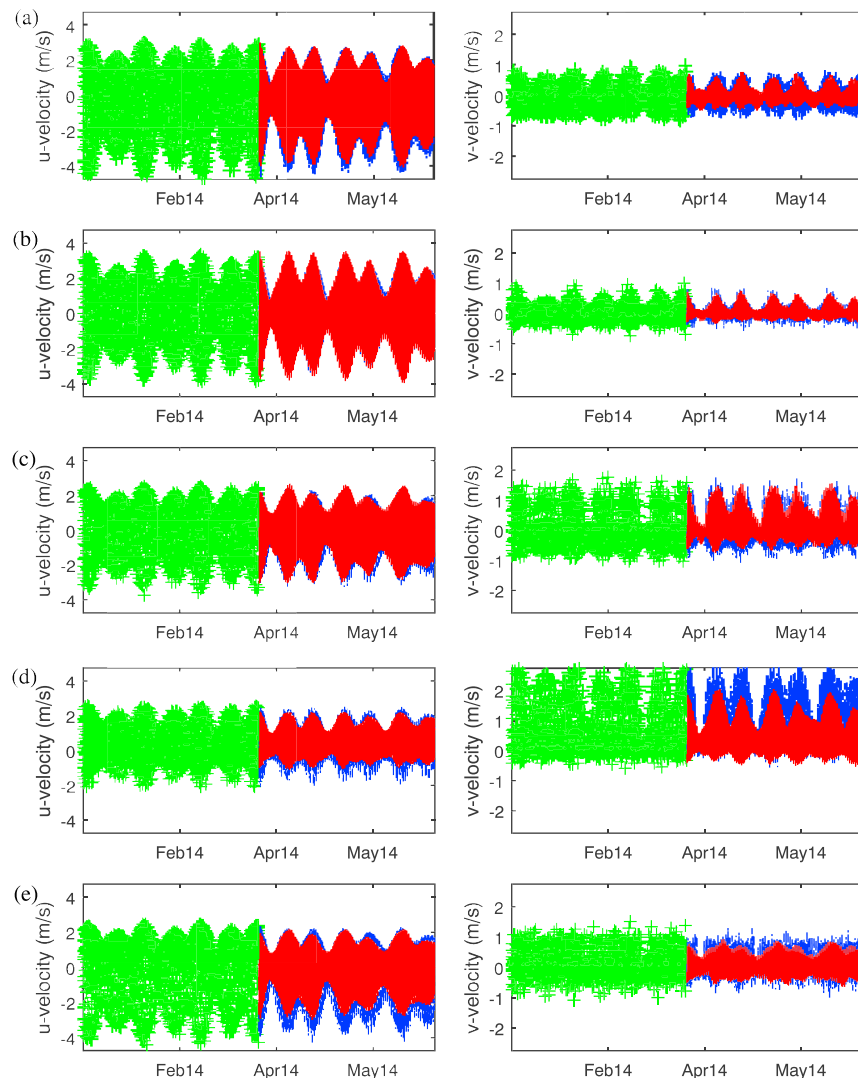


Figure 8. Tidal current velocities at five unsampled locations: (a) 1*, (b) 2*, (c) 3*, (d) 4*, and (e) 5*, (see Table 5). The green data points for the first 3 months are only shown for convenience and they are not used at all by the machine learning approach. The predictions from the Gaussian processes model are indicated by the red curve, while the actual values at those instances are in blue.

Table 6

Comparison of the RMSE (in m/s) of the Prediction Using the HA and GP Algorithms at Unsamped Locations

Location	u		v	
	GPhour	HA(interp)	GPhour	HA(interp)
1*	0.292	1.011	0.216	0.236
2*	0.311	0.907	0.145	0.161
3*	0.345	0.887	0.366	0.395
4*	0.449	0.669	0.446	0.793
5*	0.538	1.195	0.324	0.322

Note. RMSE = root-mean-square error; HA = harmonic analysis; GP = Gaussian processes.

correlations in space, time, and between outputs (see section 2.2). To assess the effectiveness of this kernel function, we analyze another kernel component which could have been a first choice:

$$K_q(i, j) = \cos(2\pi\mu_q|t_i - t_j|) \exp\left(-\frac{1}{2} \begin{bmatrix} (x_i - x_j) \\ (y_i - y_j) \end{bmatrix}^T \mathbf{M}_q^{-1} \begin{bmatrix} (x_i - x_j) \\ (y_i - y_j) \end{bmatrix}\right), \quad (22)$$

where μ_q is the frequency hyperparameter of the q th kernel component. The key difference between the two kernels lies in modeling of the correlations in time. In the original formulation (see (13)), the temporal dependencies were modeled with periodic kernel components, while in (22) the temporal kernel components have a cosine form. The latter could be a natural choice as cosine functions form the basis in the HA formulation.

One useful measure that can be used to compare different models of the same data is the negative mean log probability (NMLP; Carney et al., 2005), which penalizes predictions that are overconfident or underconfident and is expressed as

$$\text{NMLP} = \frac{1}{L} \sum_{i=1}^L -\log(p(\mathbf{y}^*(i)|\mathbf{y})) = \frac{1}{L} \sum_{i=1}^{n_s} \frac{(\mathbf{y}^*(i) - \boldsymbol{\mu}(\mathbf{x}^*(i)))^2}{2\sigma_*^2(i)} + \frac{1}{2} \log(2\pi\sigma_*^2(i)). \quad (23)$$

This is the mean of the negative logarithm of the predictive probability, and unlike RMSE, is less sensitive to the overall scale of the target values (Rasmussen & Williams, 2006). Comparisons of the kernels for different sampling strategies using two ADCPs are listed in Table 7. The periodic kernel clearly outperforms the cosine kernel form, both in average RMSE and NMLP, for all the three sampling strategies. The former kernel is capable of capturing the nonlinearities better than the simple cosine form—this feature becomes more important at locations of fast tidal current where tidal distortions due to interactions with bathymetry could be significant. Even in the analysis of data from point locations, the GP approach fares better than the HA due to its ability to capture nonsinusoidal effects (Sarkar et al., 2018).

5.5. Measurement Noise

The results from data-driven models will be affected by noise in measurements. A short analysis is performed with a case where the current data are contaminated with a Gaussian noise having standard deviations of

Table 7

Comparison of Average of the RMSE and NMLP for Different Sampling Strategies of Two Different Kernel Components (Cosine Kernel and Periodic Kernel) Modeling the Time Variations

Strategy	u				v			
	RMSE (m/s)		NMLP		RMSE (m/s)		NMLP	
	Cos	Periodic	Cos	Periodic	Cos	Periodic	Cos	Periodic
GPweek	0.706	0.393	3.366	0.781	0.334	0.281	0.877	0.616
GPday	0.478	0.323	1.001	0.453	0.266	0.243	0.433	0.129
GPhour	0.467	0.316	0.705	0.420	0.260	0.243	0.135	0.102

Note. RMSE = root-mean-square error; NMLP = negative mean log probability.

Table 8*Comparison of Average of the RMSE and NMLP for Different Sampling Strategies When Modeling Flux and Current Data*

Strategy	<i>u</i>				<i>v</i>			
	RMSE (m/s)		NMLP		RMSE (m/s)		NMLP	
	Flux	Current	Flux	Current	Flux	Current	Flux	Current
GPweek	0.481	0.393	1.796	0.781	0.257	0.281	0.247	0.616
GPday	0.342	0.323	0.386	0.453	0.234	0.243	0.040	0.129
GPhour	0.312	0.316	0.281	0.420	0.227	0.243	−0.005	0.102

Note. RMSE = root-mean-square error; NMLP = negative mean log probability.

0.35 m/s in *u* and 0.15 m/s in *v* velocity components. Employing the GPhour strategy for sampling, the predictions of the GP model using the noisy data set yielded average RMSEs of 0.483 and 0.287 m/s in *u* and *v* velocity components, respectively. On the other hand, HA employed on the complete data set (sampled hourly over 91 days at all 15 locations) produced average RMSEs of 0.480 and 0.271 m/s in predictions of *u* and *v* components of velocity, respectively. The results of the two approaches are similar, with HA predictions being marginally better. Note, the sparsely sampled GP approach uses significantly fewer training points compared to HA, and as discussed earlier, the spatiotemporal predictions (at arbitrary locations) of the GP approach are substantially better than interpolations made using HA.

5.6. Current Versus Flux Modeling

Tidal currents can vary rapidly within short distances in shallow waters where it is influenced by bathymetry. Comparatively, volumetric flux fluctuates less in space and could be pursued for modeling. However, in a realistic scenario, determination of flux can be intricate as it demands knowledge of the vertical profile of currents. The present analysis is based on a depth-averaged numerical model, and the flux at point locations can be easily computed. The numerical flux is now modeled using the same GP approach. The current data can be straightforwardly recovered by dividing the flux predictions at a particular location by corresponding depth. Comparisons of the average RMSEs and NMLPs of the current predictions from independent modeling of flux and current data, with different sampling approaches, are presented in Table 8. The results are interesting: modeling the flux through GPhour yields better results in both measures, while that using GPweek is complete contrary with large RMSEs and NMLPs of the *u* velocity indicating that the model is not confident. The relative performance using GPday lies somewhere in between GPweek and GPhour. In short, modeling of flux improves the results of the best case (GPhour) and degrades worst case (GPweek).

One of the key takeaways from the results is the importance of planning an appropriate sampling strategy. The ML approach can also “learn” nonharmonic variations from the data, which could be useful in making short-term predictions (see Sarkar et al., 2018). The Bayesian approach captures structures in randomly sampled data, while at the same time accommodating noise in a principled manner. Such an analysis is not possible using standard techniques widely adopted in investigating engineering problems including the existing methods in tidal analysis.

6. Discussion

The tidal current data used in this work were generated from a depth-averaged numerical model, where the vertical profile of the flow is taken to be constant. However, in reality, substantial distortion from the assumed vertical structure occurs in several geographic contexts. For example, headlands have a significant influence on the spatial structure of tidal currents as counter rotating eddies are generated on either side of sharp headlands, and the residual velocity could be a significant portion of the total velocity (Geyer & Signell, 1990). Note, accounting for the depth variations of current would be an extension of the model to three dimensions. The presented ML approach can be augmented to account for the vertical profile of the current using, for example, additional exponentiated quadratic kernel components. In the latter case, new hyperparameters—length scales in the vertical direction—would have to be determined or inferred. Note, modeling such a problem entails a huge amount of training data.

One of the foremost work on modeling spatial variability of currents (from data) was that by Candela et al. (1992); however, it focused on the separation of tidal and subtidal components based on data from moving vessels. The spatial variability was investigated using two different interpolating functions—polynomial

and Green's functions solutions of the biharmonic equation. Such parametric approaches are intrinsically less flexible than a nonparametric Bayesian GP (in a nonparametric approach the number of parameters grows with data to increase the amount of information the hyperparameters can capture). Furthermore, the method (of; Candela et al., 1992) models the two horizontal components of velocity separately and using only a few tidal constituents (semidiurnal and diurnal). The generality of the GP approach presented in this manuscript admits arbitrary tidal current data sets.

In this work, a single set of spatial length scale hyperparameters (for each frequency p_q^{-1} , see (13)) was considered for modeling. However, in reality, there are cases where the natural length scales of spatial variations fluctuate considerably in a region, depending on the complexities of the bathymetric features. Such cases can be potentially modeled using mixture-of-GP-experts (Rasmussen & Ghahramani, 2002), product-of-GP-experts (Deisenroth & Ng, 2015), or warping of the GP (Snelson et al., 2004). This would be an interesting topic for future research.

The variation of tidal currents in shallow waters can be complex, and achieving robust spatiotemporal predictions across a region with limited data can be a challenging task. The quality of predictions of the GP model is influenced by the number of samples from and near the location of interest, as well as the surrounding bathymetry profile; the latter is a critical variable in the analysis of waves and currents (Holman et al., 2013). As tidal currents can vary sharply within short distances in shallow waters, reasonable predictions can be expected at locations that have sampled data points in their neighborhood (ideally locations within one spatial length scale of kernel components). Caution is advised when making predictions in regions with sharp bathymetric variations and few sampled data points. A detailed survey of the bathymetric features in the region of interest is recommended prior to conducting data collection campaigns. Regions exhibiting greater geographical variability need to be sampled more. Incorporating the depth of the sampled locations in the GP model as another input variable (in addition to time and horizontal spatial coordinates) may yield better results; however, such a consideration would increase the complexity of the problem (dimension of the optimization).

The nonparametric ML approach does not directly provide information of the individual tidal constituents (their amplitude and phase), which could be of interest in certain applications. The details could be ascertained by employing HA on the GP predictions at any location in the neighborhood of the sampled points. Note that the sparsity and irregularity in the data sets (with the described sampling strategies) prohibits the direct application of HA even at sampled locations.

One of the shortcomings of the naive GP implementation is its high computational cost $O(n^3)$ and memory requirement $O(n^2)$, which can prohibit its application to problems with large number of data points (see Quiñero-Candela & Rasmussen, 2005). In the GP literature, several methods have been proposed to overcome the computational complexity while still retaining the predictive accuracy of the method (see, e.g., Snelson & Ghahramani, 2005; Walder et al., 2008). Such approaches could facilitate modeling of large data sets, generated when data are sampled simultaneously by several ADCPs and/or sampled over long duration of time and/or sampled with high frequency. Computations for the case studies in this work were made in a reasonable time (approximately days) on a single CPU, the majority of time being expended in determination of hyperparameters through the optimization scheme.

The data collection campaign strategies presented in section 4 may appear slightly arduous with manned vessels. However, in recent years, autonomous robotic technologies have witnessed significant advancements, and it is envisaged that such systems, equipped with appropriate sensors, could be used for current measurements. The presented method can complement such data by aptly extracting the spatiotemporal dependencies. Some autonomous vessels, for example, AutoNaut and C-Enduro, are powered by renewable sources of energy which can enable them to sustain and allow them to remain in the marine environment for longer times. Using the described spatiotemporal GP model, a single autonomous vehicle-based survey can provide predictions as good as a large number of fixed (bottom mounted) ADCPs covering a wide range of space at the sampled locations (15 in our study). Moreover, the GP method can make predictions and quantify uncertainties at any point in the neighborhood of the sampled region. Studies such as that by Palodichuk et al. (2013) have illustrated shortcomings of the fixed ADCPs, which involve deployment and recovery operations during suitable times (slack water), appropriate vessels, and manpower. Although a proper quantification of the cost benefits is beyond the scope of this work, expenses associated with a single instrumented autonomous vehicle would be less than the cost of using a large number of stationary ADCPs

over the same period of time. Note, the approach can be used to improve understanding and insights into existing spatiotemporal tidal current data sets. The study also compares different sampling strategies and sets out an approach for new data collection campaigns.

7. Conclusion

A Bayesian ML methodology for spatiotemporal modeling of tidal currents is presented in this work. Such a task is beyond the capabilities of the classical HA approach widely used in tidal analysis. The statistical ML approach uses a kernel function comprising combinations of periodic kernel components to capture the harmonic correlations in time and exponentiated quadratic kernel components to model the spatial correlations. A case study is performed using data generated from a numerical model. The best results (lowest average RMSE) were obtained when ADCPs changed location after every sampling (every 1 hr in our case). The GP-based ML approach can produce predictions from sparsely sampled spatiotemporal data sets, that is, data sampled at few spatial locations with gaps in time. The overall accuracy of the GP predictions at the sampled locations, using data from just a single moving vessel-based (virtual) survey, are similar to those from HA that uses continuous data (hourly sampled over 91 days) from all locations. In addition, the method provides a spatiotemporal description over the whole region along with uncertainty estimates.

The GP-based approach has a wide range of flexibilities to handle additional complexities with minor modifications, for example, capturing unknown frequencies not considered by the HA and dealing with nonharmonic effects. This could be useful in modeling data from low-energy sites, where nontidal effects can have energy comparable to that of the tidal. Short-term nondeterministic components can be modeled using additional kernel components or by representing the mean function of the GP with another GP as in Sarkar et al. (2018). Such analysis is not possible using parametric approaches, which suffer from limited modeling capacity and are prone to overfitting/underfitting.

How exactly the sampling locations should be selected for the best predictions in the region could be a topic of a future study. In general, it is envisaged that more sampling needs to be made in regions exhibiting sharp changes in flow, primarily induced by bathymetric variations. The length scales of the kernel function can be made to vary spatially to accommodate abrupt changes in bathymetric features. Integrating such effects (into the ML algorithm) would require construction of a more complex covariance function. Also, in a realistic scenario, sampling can be made more gradually across the spatial domain with a temporal sampling frequency higher than that in this work. Such a sampling procedure would generate more data that would assist in better training of the GP hyperparameters to improve model predictions. Large data sets could be tackled with better computational resources and/or with sparser GP algorithms. Future works will deal with such cases.

Analysis and prediction of tidal currents is complex and as noted by Godin (1983) three decades back— it should be dealt with as a research problem rather than just some routine data processing tasks. There has been significant progress in the field of ML since then, and this work makes use of the advances to provide a direction in approaching the intricate problem of spatiotemporal tidal current prediction.

Acknowledgments

This work is supported by the Engineering and Physical Sciences Research Council, United Kingdom (grant EP/M021394/1) under the research project: Learning tidal currents. The authors would like to thank the developers of UTide and GPML, and Mark McAllister for his time in helping to run code. They are also grateful to the five anonymous reviewers for their comments which led to the improvement of the manuscript. The numerical data can be found at <https://figshare.com/articles/spatio-temporal-tidal-current-GP-zip/4275047>.

References

- Adcock, T. A. A., Draper, S., Houlby, G. T., Borthwick, A. G. L., & Serhadlioglu, S. (2013). The available power from tidal stream turbines in the Pentland Firth. *Proceedings of the Royal Society A*, 469, 21.
- Babovic, V., Sannasiraj, S. A., & Chan, E. S. (2004). Error correction of a predictive ocean wave model using local model approximation. *Journal of Marine Systems*, 53, 1–17.
- Candela, J., Beardsley, R. C., & Limeburner, R. (1992). Separation of tidal and subtidal currents in ship-mounted acoustic Doppler current profiler observations. *Journal of Geophysical Research*, 97(769–788).
- Carney, M., Cunningham, P., Dowling, J., & Lee, C. (2005). Predicting probability distributions for surf height using an ensemble of mixture density networks. In *Proceedings of the 22th International Conference on Machine Learning* (pp. 113–120). Bonn, Germany.
- Codiga, D. L. (2011). Unified tidal analysis and prediction using the UTide Matlab functions (Technical Report 2011-01) (59 pp.). Narragansett, RI: Graduate School of Oceanography, University of Rhode Island.
- Davis, J. V., Kulis, B., Jain, P., Sra, S., & Dhillon, I. S. (2007). Information-theoretic metric learning. *Proceedings of the 24th International Conference on Machine Learning*, 22, 209–216.
- De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50, 1–18.
- Deisenroth, M., & Ng, J. W. (2015). Distributed Gaussian processes. *Proceedings of the 32nd International Conference on Machine Learning*, 37, 1481–1490.

- Dowd, M., & Thompson, K. R. (1996). Extraction of tidal streams from a ship-borne acoustic Doppler current profiler using a statistical-dynamical model. *Journal of Geophysical Research*, 101, 8943–8956.
- Dusek, G., Pico, C., Paternostro, C., & Fanelli, P. (2016). Spatial variability of tidal currents in Puget Sound, Washington. In *96th American Meteorological Society Annual Meeting* (8 pp.). New Orleans, LA, January 10–14, 2016.
- Foreman, M. G. G. (1977). Manual of tidal heights analysis and prediction (Pacific Marine sciences Report 77-10). Patricia Bay: Institute of Ocean Sciences.
- Geyer, W. R., & Signell, R. (1990). Measurements of tidal flow around a headland with a shipboard acoustic Doppler current profiler. *Journal of Geophysical Research*, 95, 3189–3197.
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521, 452–459.
- Godin, G. (1972). *The analysis of tides*. Toronto: University of Toronto Press.
- Godin, G. (1983). The spectra of point measurements of currents: Their features and their interpretation. *Atmosphere-Ocean*, 21, 263–284.
- Gomis, D., Ruiz, S., & Pedder, M. A. (2001). Diagnostic analysis of the 3D ageostrophic circulation from a multivariate spatial interpolation of CTD and ADCP data. *Deep-Sea Research Part, 1*, 48, 269–295.
- Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 36, 1171–1220.
- Hollinger, G., Pereira, A., Ortenzi, V., & Sukhatme, G. (2012). Towards improved prediction of ocean processes using statistical machine learning. Robotics: Science and Systems Workshop on Robotics for Environmental Monitoring.
- Holman, R., Plant, N., & Holland, T. (2013). cBathy: A robust algorithm for estimating nearshore bathymetry. *Journal of Geophysical Research: Oceans*, 118, 2595–2609. <https://doi.org/10.1002/jgrc.20199>
- Kincaid, C., Pockalny, R. A., & Huzzey, L. M. (1996). Spatial and temporal variability in flow at the mouth of Narragansett Bay. *Journal of Geophysical Research*, 108(C7), 3218.
- Li, J., & Castagna, J. (2004). Support vector machine (SVM) pattern recognition to AVO classification. *Geophysical Research Letters*, 31, L02609. <https://doi.org/10.1029/2003GL018299>
- MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge: Cambridge University Press.
- McIntosh, P. C. (1990). Oceanographic data interpolation: Objective analysis and splines. *Journal of Geophysical Research*, 95(C8), 13,529–13,541.
- Minka, T. P. (2000). Deriving quadrature rules from Gaussian processes. Pittsburgh: Statistics Department, Carnegie Mellon University, Technical Report.
- Neill, S. P., Hashemi, M. R., & Lewis, M. J. (2014). The role of tidal asymmetry in characterizing the tidal energy resource of Orkney. *Renewable Energy*, 68, 337–350.
- 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 Transactions on Sensor Networks (TOSN)*, 9(1), 1–32.
- Palodichuk, M., Polagye, B., & Thomson, J. (2013). Resource mapping at tidal energy sites. *IEEE Journal of Ocean Engineering*, 38(3), 433–446.
- Parker, B. B. (2007). Tidal analysis and prediction. US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Center for Operational Oceanographic Products and Services.
- Parker, B. B. (1991). *Tidal hydrodynamics*. New York: John Wiley.
- Pinheiro, J. C., & Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6, 289–296.
- Polagye, B. L., Epler, J., & Thomson, J. (2010). Limits to the predictability of tidal current energy. *Oceans*, 2010.
- Quiñonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6, 1939–1959.
- Rasmussen, C. E., & Ghahramani, Z. (2002). Infinite mixtures of Gaussian process experts. *Advances in Neural Information Processing Systems*, 14, 881–888.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*: MIT Press.
- Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984).
- Robins, P. E., Neill, P., Lewis, M. J., & Ward, S. L. (2015). Characterising the spatial and temporal variability of the tidal-stream energy resource over the northwest European shelf seas. *Applied Energy*, 147, 510–522.
- Sannasiraj, S. A., Zhang, H., Babovic, V., & Chan, E. S. (2004). Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Advances in Water Resources*, 27(7), 761–772.
- Sarkar, D., Osborne, M. A., & Adcock, T. A. A. (2018). Prediction of tidal currents using Bayesian machine learning. *Ocean Engineering*, 158, 221–231.
- Snelson, E., & Ghahramani, Z. (2005). Sparse Gaussian processes using pseudo-inputs. *Advances in Neural Information Processing Systems*, 17, 1257–1264.
- Snelson, E., Rasmussen, C. E., & Ghahramani, Z. (2004). Warped Gaussian processes. *Advances in Neural Information Processing Systems*, 16, 337–344.
- Steinwart, I., & Christmann, A. (2008). *Support vector machines*. New York: Springer.
- Topa, H., & Honkela, A. (2015). Gaussian process modelling of multiple short time series. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 83–88). Bruges, Belgium.
- Trossman, D. S., Thompson, L., & Hautala, S. L. (2009). Application of thin-plate splines in two dimensions to oceanographic tracer data. *Journal of Atmospheric and Oceanic Technology*, 28(11), 1522–1538.
- Vennell, R., & Beatson, R. (2009). A divergence-free spatial interpolator for large sparse velocity data sets. *Journal of Geophysical Research*, 114, C10024. <https://doi.org/10.1029/2008JC004973>
- Walder, C., Kim, K. I., & Schölkopf, B. (2008). Sparse multiscale Gaussian process regression. In *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, pp. 1112–1119.
- Watanabe, S., & Chien, J. T. (2015). *Bayesian speech and language processing*. New York, NY: Cambridge University Press.
- Wikle, C. K., Milliff, R. F., Herbei, R., & Leeds, W. B. (2013). Modern statistical methods in oceanography: A hierarchical perspective. *Statistical Science*, 28(4), 466–486.
- Wilson, A. G., & Adams, R. P. (2013). Gaussian process kernels for pattern discovery and extrapolation. *Proceedings of the 30th International Conf. on Machine Learning*, 28, 1067–1075.