Ocean Engineering 158 (2018) 221–231

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

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 R T I C L E I N F O

Keywords:
Prediction
Tidal currents
Machine learning
Gaussian process

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 over-fitting 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.

* Corresponding author.
E-mail address: dripta.academic@gmail.com (D. Sarkar).

https://doi.org/10.1016/j.oceaneng.2018.03.007
Received 27 November 2016; Received in revised form 4 January 2018; Accepted 3 March 2018

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: $y(t)$, $t = t_1, t_2, \ldots, t_M$, 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:

$$\begin{bmatrix} e^{i\phi_1^{(1)}} & \cdot & \cdot & e^{i\phi_1^{(N)}} & e^{-i\phi_1^{(1)}} & \cdot & \cdot & e^{-i\phi_1^{(N)}} & 1 & (t_1-t_0) \\ e^{i\phi_2^{(1)}} & \cdot & \cdot & e^{i\phi_2^{(N)}} & e^{-i\phi_2^{(1)}} & \cdot & \cdot & e^{-i\phi_2^{(N)}} & 1 & (t_2-t_0) \\ e^{i\phi_3^{(1)}} & \cdot & \cdot & e^{i\phi_3^{(N)}} & e^{-i\phi_3^{(1)}} & \cdot & \cdot & e^{-i\phi_3^{(N)}} & 1 & (t_3-t_0) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ e^{i\phi_M^{(1)}} & \cdot & \cdot & e^{i\phi_M^{(N)}} & e^{-i\phi_M^{(1)}} & \cdot & \cdot & e^{-i\phi_M^{(N)}} & 1 & (t_M-t_0) \end{bmatrix} \begin{bmatrix} a_1^+ \\ \cdot \\ \cdot \\ a_N^+ \\ a_1^- \\ \cdot \\ \cdot \\ a_N^- \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} u_1 + iv_1 \\ u_2 + iv_2 \\ u_3 + iv_3 \\ \cdot \\ \cdot \\ \cdot \\ u_M + iv_M \end{bmatrix}$$

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 $(\boldsymbol{Ta} - \boldsymbol{y})$, where $\boldsymbol{y} = [y(t_1), y(t_2), \ldots, y(t_m)]'$, $\boldsymbol{a} = [a_1^+, a_2^+, \ldots, a_N^+, a_1^-, a_2^-, \ldots, a_N^-, c_0, c_1]'$ and $\boldsymbol{T}$ 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 $\left|\left|\boldsymbol{Ta} - \boldsymbol{y}\right|\right|^2$, and the solutions are determined as $\boldsymbol{a} = (\boldsymbol{T}^*\boldsymbol{T})^{-1}\boldsymbol{T}^*\boldsymbol{y}$ 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 $\boldsymbol{a} = (\boldsymbol{T}^*\boldsymbol{WT})^{-1}\boldsymbol{T}^*\boldsymbol{Wy}$, where $\boldsymbol{W}$ is some weighting matrix which is

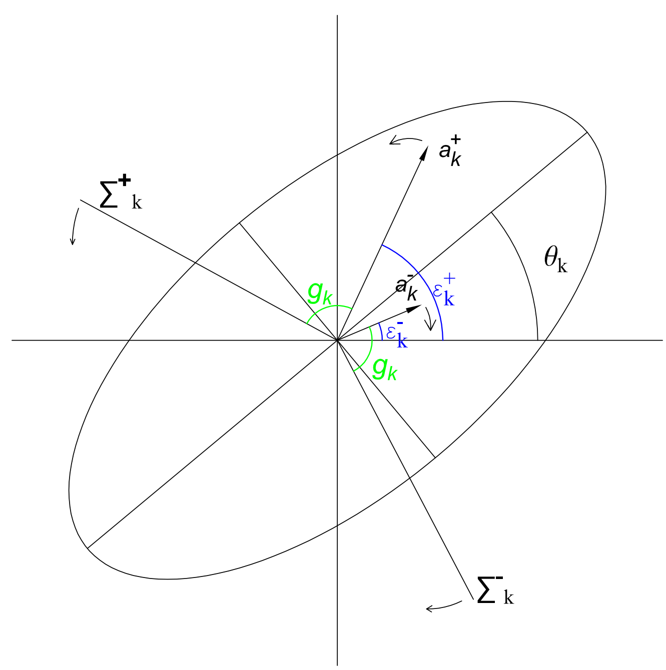

Fig. 1. For a particular tidal constituent k, the rotating vectors with amplitude a_k^+ and a_k^- are considered to be generated by two different fictitious stars $\sum_k^+$ and $\sum_k^-$, rotating in the counterclockwise and clockwise direction respectively, at a speed same as that of the constituents.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