Imperial College London

MSc in Applied Computational Science and Engineering

Independent Research Project

A Machine Learning Approach to the Prediction of Tidal elevation

by

Chen Qian

Supervisor: Professor Matthew Piggott

Imperial College London

Content

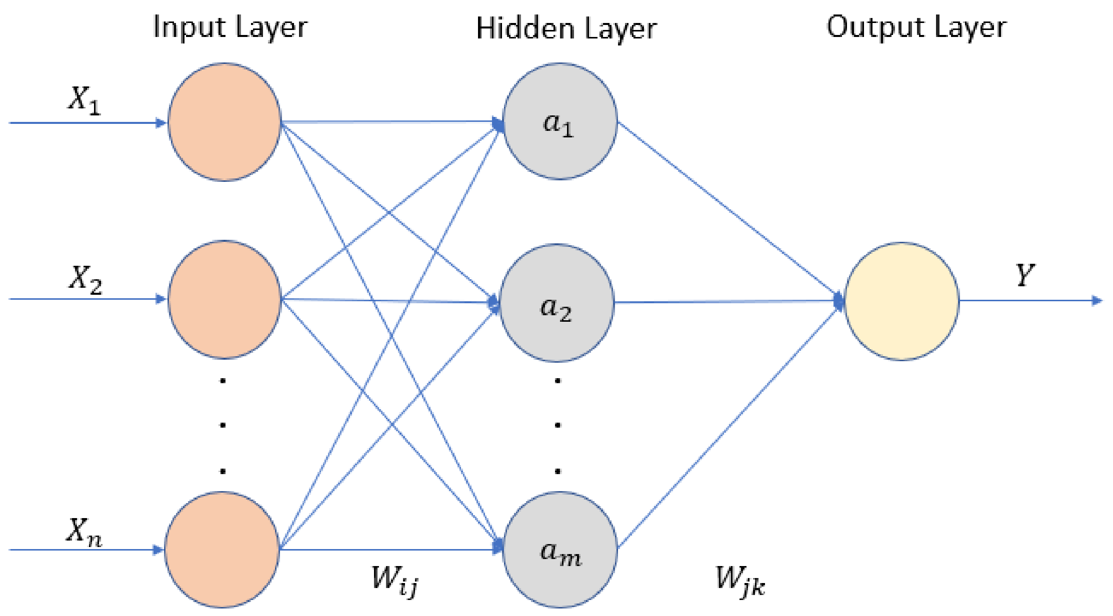
Imperial College London

Introduction

Accurate tidal prediction is of great significance for the effective generation of energy.

This project proposes one prediction model combining periodic analysis and back propagation (BP) neural network.

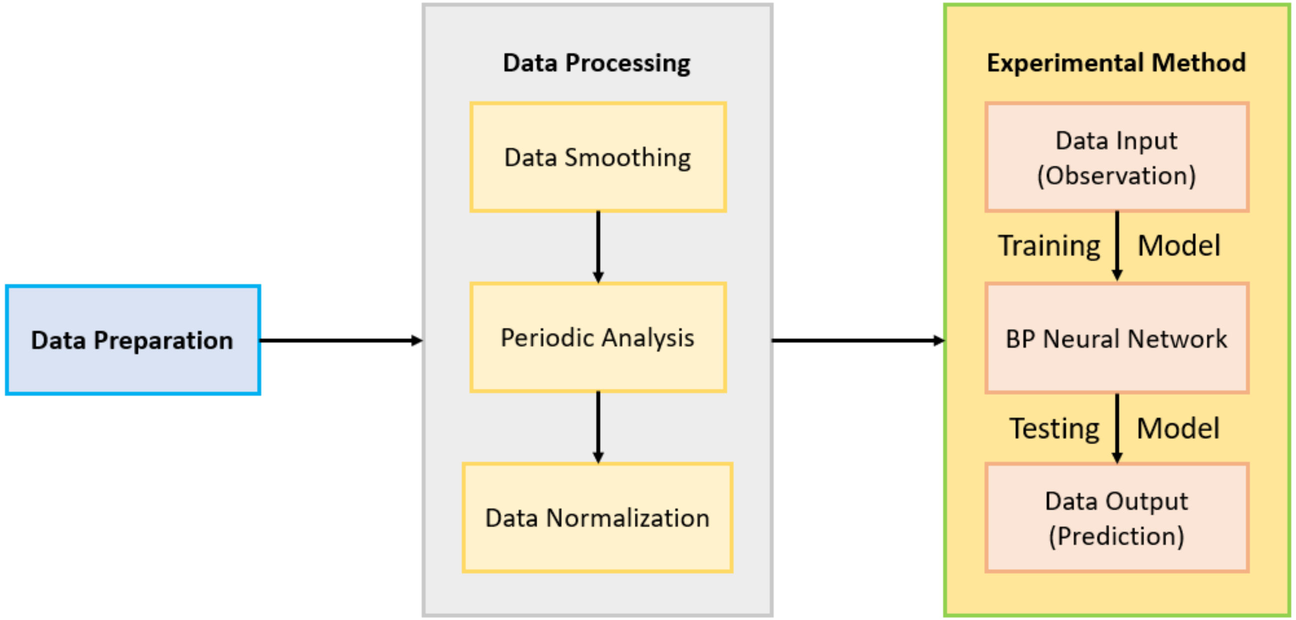
By learning the correlation between training samples and training labels, the BP neural network can determine the fitting weight coefficients of the prediction curve.



A classical three-layer BP neural network model, which includes n input, m hidden and *1* output neurons.

Imperial College London

Software & Description



The software consists of several parts, which can be divided into three types: data preparation, data processing, and experimental method.

Platform and Libraries

Imperial College London

Software & Description

Prediction Method	Historical Data Sample	Prediction data
One-step iterative prediction	$\{x_n, x_{n-1}, x_{n-2}, \dots, x_{n-m+1}\}$ (totally m data)	$\{x_{n+k} \mid k = 1\}$
Multi-step iterative prediction	$\{x_n, x_{n-1}, x_{n-2}, \dots, x_{n-m+1}\}$ (totally m data)	$\{x_{n+k} \mid k \geq 2\}$

Two types of traditional prediction methods, which are One-step and Multi-step iterative prediction. Select appropriate historical data to forecast data $\{x_{n+k} \mid k \geq 1\}$

Concerning one-step iterative prediction, one new improved prediction method is introduced based on it, which is autoregressive moving average model (ARMA)

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots + \alpha_q \varepsilon_{t-q}$$

x_t / ε_t: Observation / Residual value

β_t / α_t : Weight parameters of observation / residual value

Imperial College London

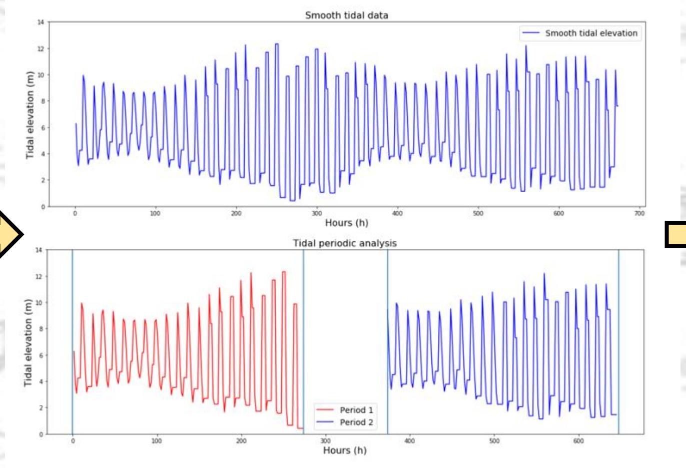
Implementation & Code

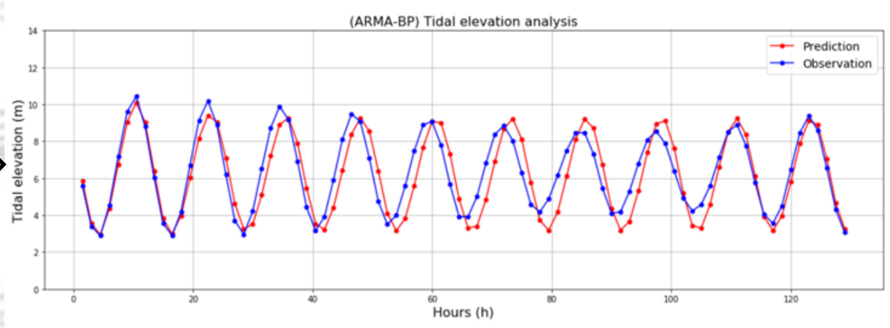
	Cycle	Hour	Elevation
0	7.0	1.5	6.266
1	13.0	3.0	3.809
2	19.0	4.5	3.072
3	25.0	6.0	4.245
4	31.0	7.5	6.426
...	...	...	...
444	2671.0	667.5	9.001
445	2677.0	669.0	11.012
446	2683.0	670.5	10.344
447	2689.0	672.0	7.608
448	2695.0	673.5	4.457

449 rows × 3 columns

	Cycle	Hour	Elevation
0	7.0	1.5	5.584
1	13.0	3.0	3.400
2	19.0	4.5	2.885
3	25.0	6.0	4.534
4	31.0	7.5	7.186
...	...	...	...
444	2671.0	667.5	9.273
445	2677.0	669.0	10.749
446	2683.0	670.5	9.163
447	2689.0	672.0	5.945
448	2695.0	673.5	2.883

449 rows × 3 columns





Data Preparation

Select timescale and then read the sample data.

Data Processing

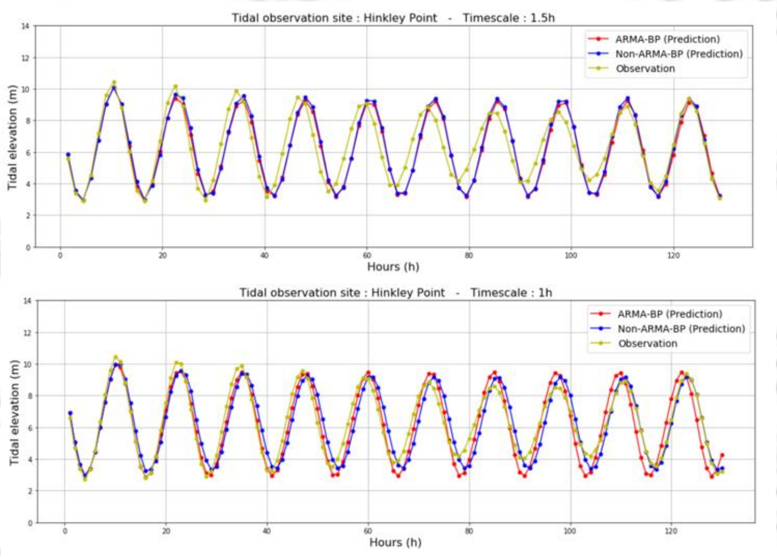
Estimate the periodicity of sample data.

Experimental Method

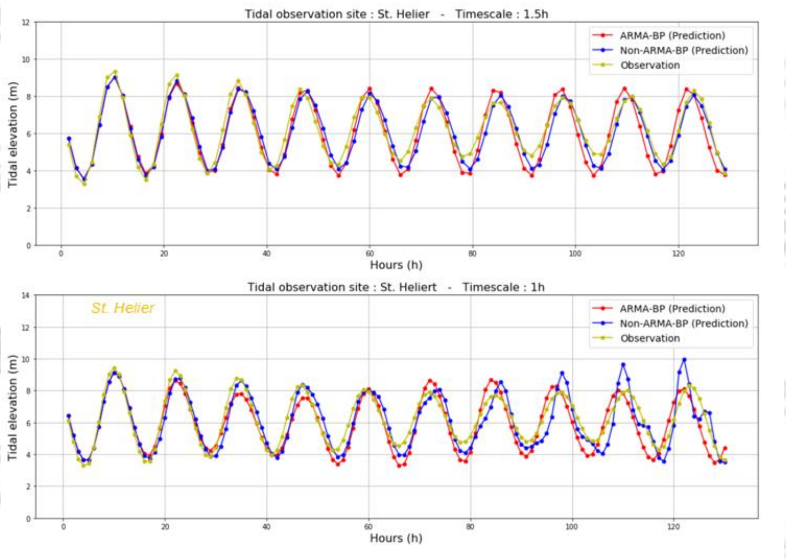
Train BP neural network and use it to make predictions

Imperial College London

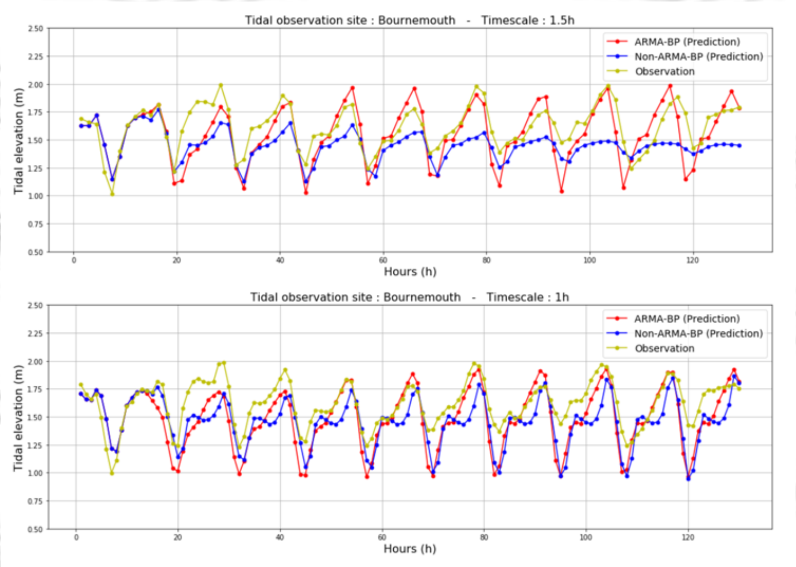
Performance Analysis



Tidal observation site
(**Hinkley Point**)
Timescales (1.5h / 1h)
Prediction methods
(ARMA / Non-ARMA-BP)



Tidal observation site
(**St. Helier**)
Timescales (1.5h / 1h)
Prediction methods
(ARMA / Non-ARMA-BP)



Tidal observation site
(**Bournemouth**)
Timescales (1.5h / 1h)
Prediction methods
(ARMA / Non-ARMA-BP)

Imperial College London

Performance Analysis

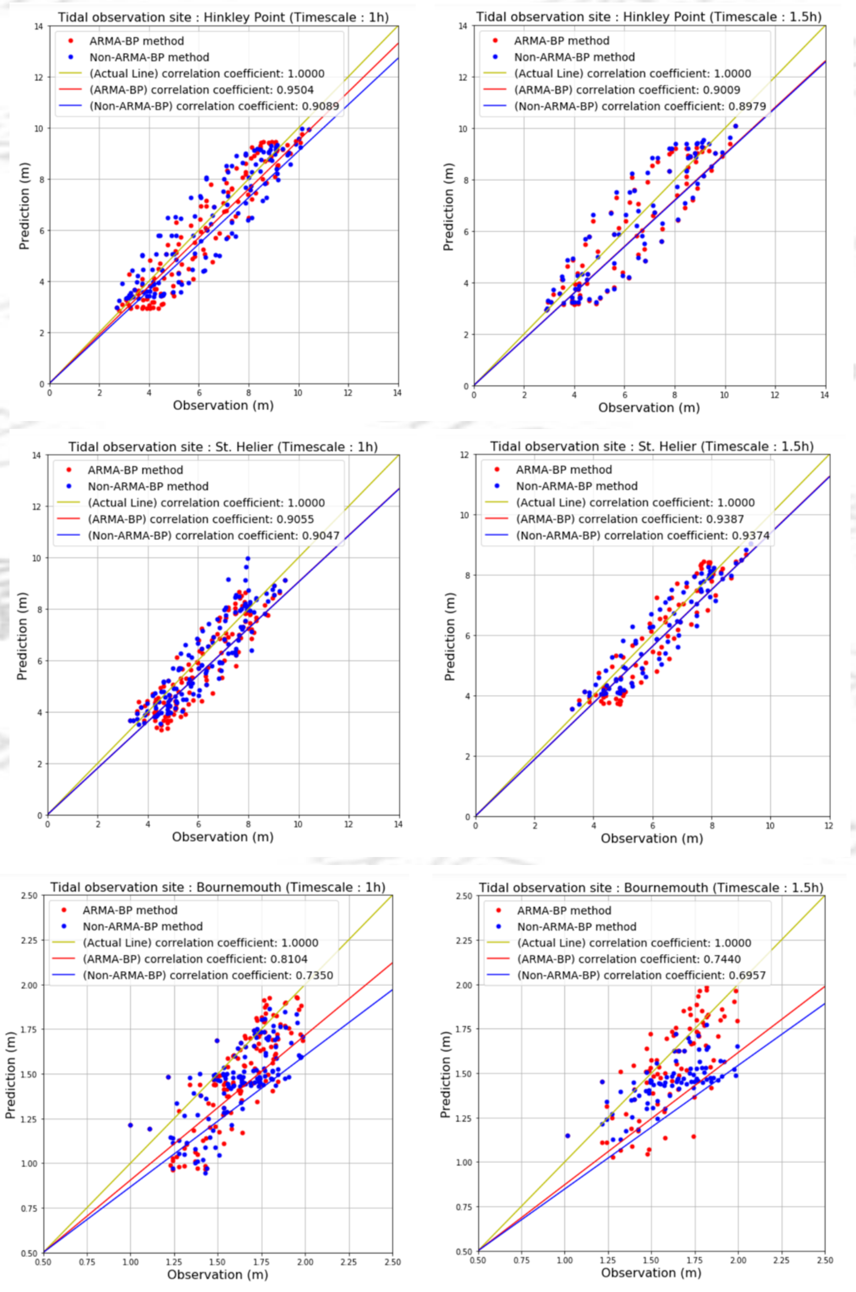
To further quantifying performance, one new measurement indicator is introduced, calling the **correlation coefficient**.

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]\,Var[Y]}} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})\cdot\sum_{i=1}^{n}(Y_i-\bar{Y})}}$$

X_i : Prediction Y_i : Observation

Comparison of correlation coefficients using different timescales (1.5h / 1h) and different prediction methods (ARMA / Non- ARMA-BP).

The tidal observation sites from top to bottom are Hinkley Point, St. Helier and Bournemouth respectively.



Imperial College London

Discussion & Conclusion

In general, *ARMA-BP* (one-step iterative prediction) method dominates over the *Non-ARMA-BP* (multi-step iterative prediction) method, either on different tidal observation sites or timescales.

For observation points with regular fluctuations, the trained BP neural network can make a relatively accurate prediction. On the contrary, for observation points with relatively disordered fluctuations, the prediction accuracy is usually very poor.

Furthermore, the amplitude of tidal elevation is also a significant point. Larger amplitude means that the observation points are more likely to be uniformly distributed, neither too dense nor too sparse. Uniform data distribution plays a crucial role on improving prediction accuracy.

Imperial College London

Discussion & Conclusion

Future Work

Long-term prediction should be focused on. Meanwhile, it is essential to trade-off between prediction accuracy and training time. In order to solve this problem, there are three preliminary ideas: Optimize the raw neural network code structure; Propose a new type of neural network algorithm; Apply tensor computation (Pytorch) to achieve GPU acceleration.

In addition, there is no good prediction method for data with irregular fluctuations at present. If someone would like to continue research in this area, it is better to consider using other prediction methods or combining neural network with other methods.

Q & A

Thanks for your attendance