Time series prediction using Gaussian Processes on the Sotonmet data set

Jonathan Carter

Abstract

A GPU compatible library for Gaussian Process regression[1] was implemented in PyTorch, using its automatic differentiation suite to optimise the marginal likelihood via gradient descent. Using this library, a co-variance kernel was designed for inference on the Sotonmet dataset and is shown to give good performance in both smoothing and forecasting applications.

1. Introduction

Gaussian Processes (GPs, [1]) are a class of non-parametric models that give rise to distributions over functions, from which the joint distribution over any subset of outputs is multivariate Gaussian. GPs are entirely defined by mean and co-variance kernels, $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$, which define a prior over the underlying function of interest, f. Formally, we can write:

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x'})) \tag{1}$$

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \tag{2}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))] \tag{3}$$

Given observed inputs $\boldsymbol{X}$, corresponding outputs $\boldsymbol{y}$ and new unobserved data points $\boldsymbol{X_*}$, the posterior distribution over the unobserved function outputs is also Gaussian:

$$f_* | X_*, y, X, \theta \sim \mathcal{N}(\bar{\boldsymbol{f}}_*, cov(\boldsymbol{f_*})) \tag{4}$$

$$\bar{\boldsymbol{f}}_* \coloneqq m(\boldsymbol{X_*}) + k(\boldsymbol{X_*}, \boldsymbol{X})\Sigma_X^{-1}(\boldsymbol{y} - m(\boldsymbol{X})) \tag{5}$$

$$cov(\boldsymbol{f_*}) \coloneqq k(\boldsymbol{X_*}, \boldsymbol{X_*}) - k(\boldsymbol{X}_*, \boldsymbol{X})\Sigma_X^{-1} k(\boldsymbol{X}, \boldsymbol{X_*}) \tag{6}$$

where $\Sigma_X^{-1} \coloneqq k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 I$ is the co-variance matrix of the data. Using a zero mean function, the log-marginal likelihood of the training data is given by:

$$\log p(\boldsymbol{y}|\boldsymbol{X}, \theta) = -\frac{1}{2}\boldsymbol{y}^T \Sigma_X^{-1} \boldsymbol{y} - \frac{1}{2}\log|\Sigma_X| - \frac{n}{2}\log 2\pi \tag{7}$$

which has tractable gradients with respect to the co-variance function hyper-parameters, collectively denoted as θ.

[1] See Appendix A for code

2. Sotonmet data set

Sotonmet provides weather data for the port of Southampton. This usually arrives at 5 minute intervals but under adverse conditions some readings can be lost. In this report, GP regression is used to infer the tide height during a stormy period in 2007. The recorded and ground truth tide heights can be seen in Figure 1 of the Results section. From this plot a few observations about the nature of the data are:

- The mean tide height is roughly constant.
- The data has a clear periodic trend, with amplitude variation on a day-day basis.
- The tide height shows smoothness over a range of length scales, the distinct shapes at high tide and from low-mid tide being clear examples of variation at short and long length scales respectively.

These observations are directly incorporated into the GP model developed in the subsequent section.

3. Kernel selection

Since the data is roughly stationary the mean was subtracted and a zero mean function used. After investigating a combination of standard kernels (8-10), the PRQ kernel (11) was found to give a good fit to the data in terms of both root-mean squared error (RMSE) and log-likelihood on the test data. This kernel directly incorporates the observations detailed in Section 2 into the GP prior; it enforces a base periodic trend whilst permitting amplitude variations of varying smoothness through the rational quadratic component.

$$k_{SQE}(x, x') = \sigma^2 \exp\left(\frac{-|x - x'|^2}{2l^2}\right) \tag{8}$$

$$k_{Per}(x, x') = \sigma_1^2 \exp\left(\frac{-2\sin(\frac{\pi|x-x'|}{p})}{l_1^2}\right) \tag{9}$$

$$k_{RQ}(x, x') = \sigma_2^2 \left(1 + \frac{(x - x')^2}{2\alpha l_2^2}\right)^{-\alpha} \tag{10}$$

$$k_{PRQ}(x, x') = k_{Per}(x, x')(1 + k_{RQ}(x, x')) \tag{11}$$

4. Results

Table 4 gives the RMSE and average log likelihood of the test data for the designed (PRQ) kernel along with the squared exponentiated quadratic (SQE) and periodic kernels for comparison. The hyper-parameters of each kernel were optimised using Adam for 50 epochs with a learning rate of 0.001 in each case.

The optimised period parameter p for the Periodic kernel was 742, equivalent[2] to 12h22m, three minutes lower than the true tidal period.

A plot of the predictive distribution for the PRQ kernel is shown in Figure 1. Plots using the SQE and periodic kernels can be found in the appendix.

Kernel	$\log p(\boldsymbol{y}\|\boldsymbol{X},\theta^*)$	RMSE	$\log p(\boldsymbol{f}_*\|\boldsymbol{X}_*,\boldsymbol{X},\theta)$
k_{Iso}	1273	0.156	4.00
k_{Per}	221	0.240	-204
k_{PRQ}	1420	0.042	4.04

Table 1. Sotonmet tide height regression results varying kernels.

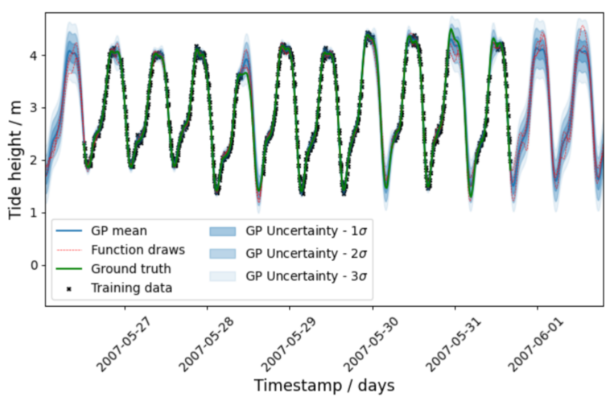


Figure 1. Sotonmet conditional distribution using k_{PRQ}.

A lookahead plot, showing the joint predictive distribution over inputs given training data up to an hour before, is shown in Figure 2. A similar plot using the SQE kernel is given in the appendix, showing much poorer predictive accuracy.

5. Discussion

The poor results using a single periodic kernel can be attributed to the strong prior over functions that it enforces: strictly periodic functions of fixed amplitude. This leads to both a poor RMSE and poor uncertainty quantification; given this prior and enough data the model becomes highly certain of the underlying latent function from this restrictive (and incorrect) prior.

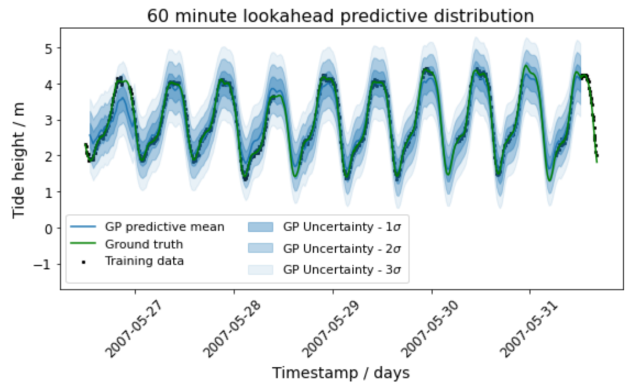


Figure 2. Sotonmet lookahead inference using k_{PRQ}.

Convergence to the period parameter was achieved for a wide range of initial values several hours above and below the true value. Setting the initial value even higher resulted in convergence to double the true value. In general when incorporating a periodic kernel the marginal likelihood will have local minima at multiples of any underlying periods in the data, emphasising the importance of a good initialisation.

The SQE kernel gives comparable performance to the PRQ kernel when used for smoothing close to the training data. However, as the SQE kernel is a convex, decreasing function of the Euclidean distance between inputs, moving away from the data means the GP reverts to its inexpressive prior distribution. Using the kernel for forecasting therefore leads to much poorer performance.

The inclusion of a periodic component in the kernel introduces long-term dependencies and allow the training data to condition phase and amplitude information of test data far from it. This leads to substantially better performance in forecasting using the PRQ kernel.

6. Conclusions

A GP model was fitted to Sotonmet tidal height data, using observations of the data to design a suitable co-variance function. This was used for smoothing and forecasting and shown to give good performance in both applications in terms of RMSE and predictive uncertainty.

When used for forecasting the improved performance of the PRQ kernel in modelling the data is much clearer than metrics such as the log marginal likelihood and the RMSE indicate. This suggests that for such applications it may be better to select and optimise kernels based on lookahead metrics. Further work could look at developing such metrics, for example, maximising the posterior distribution of lookahead data conditioned on data up to a given point.

[2] See Appendix C for more details on data processing

References

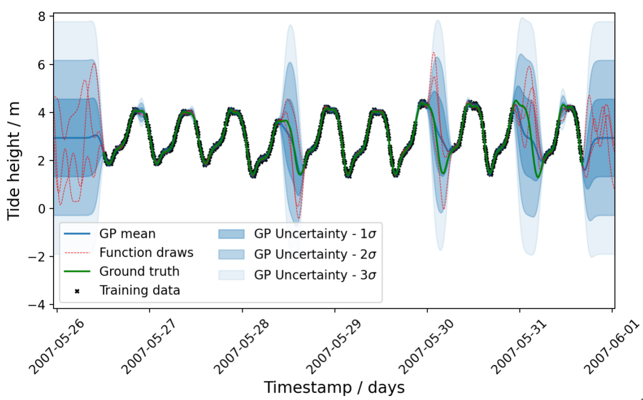
[1] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

A. Code

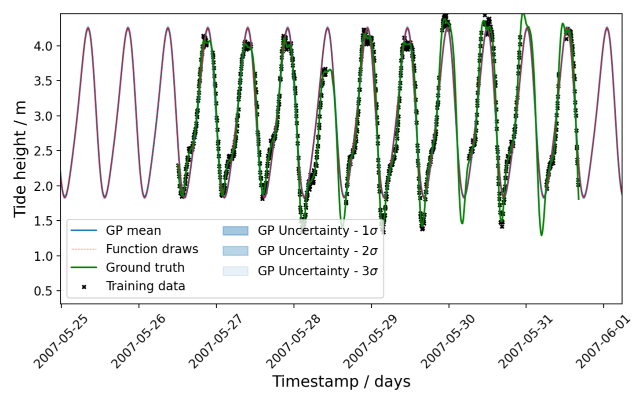
The plots in this report can be reproduced by running the following Colab notebook: https://colab.research.google.com/github/joncarter1/GP-Regression/blob/master/Analysis.ipynb

All code used can be found in the repository here: https://github.com/joncarter1/GP-Regression

B. Additional plots



(a) Using k_{SQE}.



(b) Using k_{Per}.

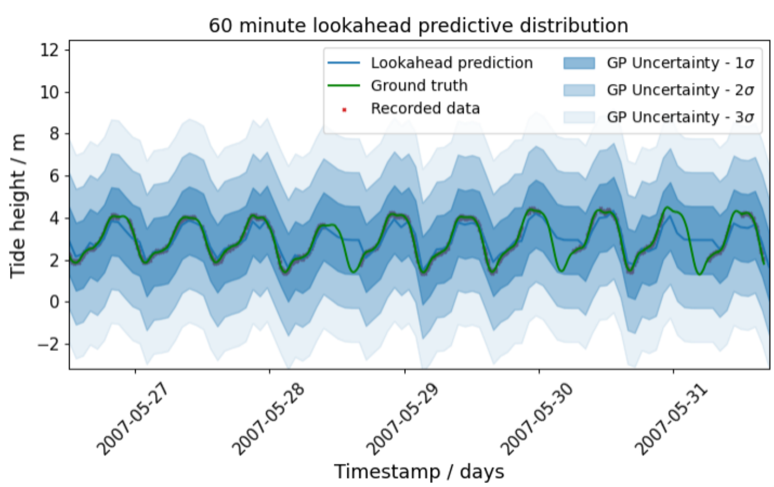


Figure 4. Sotonmet lookahead inference using k_{SQE}.

C. Data pre and post-processing

Time readings were converted to minutes and tide heights were normalised to zero mean and unit variance to construct the GP. When calculating the RMSE and test data log-likelihood the output means and co-variance matrices were re-scaled back to the original units.