Applying emulators for improved flood risk analysis

Sajni Malde[1,a], David Wyncoll[1] , Jeremy Oakley[2], Nigel Tozer[1] and Ben Gouldby[1]

[1]*HR Wallingford, Howbery Business Park, Wallingford, Oxfordshire OX10 8BA, UK*
[2]*The University of Sheffield, School of Mathematics and Statistics, Sheffield, South Yorkshire S10 2TN, UK*

Abstract. Flood risk analysis often involves the integration of multivariate probability distributions over a domain defined by a consequence function. Often, solutions of this risk integral involves Monte-Carlo sampling techniques, whereby 1000's of potential flood events are generated. It is necessary to evaluate the consequence of flooding for each sampled event. A significant computational time is required in running flood related physical process models, making it computationally impractical to evaluate flood risk using this approach. To overcome the computational challenges, this paper focusses on the Gaussian Process Emulator (GPE) meta-modelling approach. Traditionally, a "look-up table" method is used when a large number of simulations from a numerical model are required. This approach typically involves simulating conditions defined across a regular matrix, and then linearly interpolating intermediate conditions. In this paper we compare a traditional "look-up table" approach to the GPE and analyse their performance in approximating SWAN wave transformation model. In both cases, selecting an appropriate training design set is important and is taken into consideration in the analysis. The analysis shows that the GPE approach requires significantly fewer SWAN runs to obtain similar (or better) accuracies, enabling a substantial reduction in computation time, hence aiding the practicality of Monte-Carlo sampling techniques in advanced flood risk modelling.

1 Introduction and background

Flood risk is generally recognized as the product of probability and consequence where the probability relates to probabilities of flood hazards e.g. extreme rainfall, river flows, coastal waves and sea levels or multivariate combinations of these variables. The probability can also relate to the performance of flood defence infrastructure and likelihood of failure. By definition, flood risk analysis involves the integration of multivariate probability distributions over a domain defined by a consequence function. Often, solutions of this risk integral involve the use of Monte-Carlo sampling techniques, whereby tens of 1000's of potential flood events are generated through statistical sampling techniques. It is, in principal, necessary to evaluate the consequence of flooding for each of these sampled events. There is, however, typically a significant computational time involved in running physical process models that are capable of simulating the consequences of flooding. It can therefore become computationally impractical to evaluate flood risk using this approach.

Coastal flood risk in the UK is recognized to relate to both extreme offshore waves, winds (local wind-generation) and sea levels. To evaluate coastal flood risk it is therefore necessary to extrapolate these variables to extreme values, whilst accounting for the dependency between the variables. There are various methods that have been employed to do this. A robust multivariate

extreme value approach is described by [1], and this has been implemented in coastal flood risk analysis by [2,3,4]. The output of the method is a Monte-Carlo simulation comprising 1000's of extreme events with the potential to cause coastal flooding. For each event it is required to model the transformation of waves from offshore to nearshore, wave overtopping, flood inundation and impact. Various models exist for undertaking this analysis. It is however, computationally challenging and impractical to execute all the models for each event.

To overcome the computational challenges in practice, a few representative training events are simulated and the rest of the events are evaluated using various approximation techniques. Traditionally, within coastal modelling, a "look up table" (LUT) approach based on a number of training events selected from a regular grid is used. These training events conditions are simulated and used to approximate the SWAN model using linear interpolation to evaluate intermediate conditions. More recently, advances in research show that more efficient and accurate meta-modelling approaches can be applied. These approaches include: Piecewise Polynomials, Neural Networks, and Gaussian Process Emulators (GPE). This paper focusses on the GPE meta-modelling approach which has been shown to have advantages over other approaches [5].

Suppose, there exists a large set of pre-selected events, *D*, that a model needs to be evaluated at. Each

---

[a] Corresponding author: S.Malde@hrwallingford.com

model run is however, computationally intensive, and thus time consuming to evaluate. The objective of this analysis is to minimize the number of model runs, n, required in order to estimate the model evaluations at all events in D to an appropriate degree of approximation.

The case study used in this paper uses data generated from the SWAN model. This model is a third generation spectral wave model for obtaining realistic estimates of wave parameters in coastal areas, lakes and estuaries from given wind, bottom and current conditions, [6]. The model can compute how waves transform across the model domain by taking into consideration different tidal, wave and wind boundary conditions. The output of SWAN includes several parameters describing the properties of the wave at a given near shore location. Each model run can take up to few hours to evaluate.

This paper extends analysis undertaken by Camus et al [7-8] and presents an analysis of the benefits of using a GPE of the SWAN spectral wave transformation model over the traditional "look-up table" approach within the context of a coastal flood risk analysis modelling chain. Additionally, in both cases, selecting appropriate training events is important, hence the efficiency of selecting the training events and the span of the training events are taken into consideration in presenting the analysis.

2 Approximation techniques

Both, the traditional LUT method and the proposed GPE method, follow a two-step approach to approximate the SWAN model. First, appropriate training events are selected from the full set of events, D and the model is evaluated at these events. Next, using the training events and the respective model evaluations as training data, we estimate the model output for all the other events in, D.

2.1 The Traditional LUT Approach

The LUT approach uses a regular grid technique to select the representative training events. Typically, a user selects a number of values from a rectilinear grid across each input dimension and chooses to run the model at all combinations. Occasionally, some combinations are excluded based on prior knowledge of e.g. the geography of the site and boundary conditions. It is important to note that the training events selected are not well spread out in terms of coverage of each marginal input. If, for instance, each input has only 4 distinct values and the model is only sensitive to one input, there are essentially only 4 unique training data. Additionally, the number of training events required to get suitable predictions increases exponentially with the number of input dimensions.

Once the training events are selected, the computer model is run at those conditions and model evaluations are obtained. Linear interpolation is then carried out to evaluate the model evaluations at all the remaining events in D.

2.2 The Proposed GPE Approach

A GPE is defined as a statistical process that approximates a deterministic input - output computer model. Here we review the principles of GPE's without too much mathematical detail; further details and discussions are given by Kennedy and O'Hagan [9-10].

Let $y = f(x)$ represent a real valued computer model evaluation for a single event at inputs, $x = (x_1, ..., x_p)$. GPE's are applied in a Bayesian framework, where the posterior distribution is given as the product of a prior distribution and training data expressed as a likelihood function. The prior distribution represents initial beliefs of the computer model before any data is available. These initial beliefs can be represented by a Gaussian Process distribution as follows:

$$f(.)| \beta, \sigma^2 \sim GP(m_0, V_0(.,.)) \qquad (1)$$

where m_0 is a mean function described as

$$m_0(x) = h(x)^T \beta, \qquad (2)$$

where $h(x)$ is typically a linear function of input variables and β is the vector of unknown coefficients. The covariance function $V_0(.,.)$ satisfies:

$$V_0(f(x), f(x')|\sigma^2) = \sigma^2 c(x, x'), \qquad (3)$$

where σ^2 is an unknown correlation parameter and $c(x, x')$ is some covariance function. For the purpose of this analysis, we use the Gaussian covariance function defined as:

$$c(x, x') = \exp\left(\sum_{i=1}^{p} \left(\frac{x_i - x_i'}{\delta_i} \right)^2 \right) \qquad (4)$$

for roughness parameters δ_i.

The GPE is used to approximate the computer model given a set of training events, $X = (x_1, ..., x_n)$, and their true model evaluations, $y_1 = f(x_1), ..., y_n = f(x_n)$. In order for the GPE to give appropriate predictions, it is important that the training events are selected in a way that they are well spread out and as far as possible from each other. Such a design is said to have space filling properties.

One way of selecting the training events for a GPE approach is to use the Maximum Dissimilarity Algorithm (**MDA**), described by [11], and as applied in the context of coastal analysis by [7]. This is a sampling technique that is based on the distance in the multidimensional space from each point. Given an initial value, the MDA choses the next point based on the point that is the furthest away in Euclidean distance in the multidimensional space from the point it started with, having standardized the variables. This method outputs a subset of points which efficiently represents all the events in D.

Other alternative training point selection methods include; Optimised Latin hypercube designs – where a set of randomly chosen points are selected subject to a constraint that ensures that across each input, values are evenly spread. However, this method does not guarantee that the training events are space filling, and further optimisation needs to be carried out. More details (and other methods) can be found in [12 au-dessous]. One reason for choosing the MDA over the other techniques is that the event set is known a priori and it also selects a

subset of the events in D, hence, some of the final evaluations will be exact.

The training events and their model evaluations are used as training data for the emulator. The emulator is then fitted to this data, and used to approximate the computer model at a larger set of events. The GPE outputs a posterior distribution where the mean of the distribution represents the best estimate of the model and the variance of the distribution represents the uncertainty around this estimate.

The posterior mean, given a new set of inputs x, is represented mathematically as follows:

$$\boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta} + \boldsymbol{t}(\boldsymbol{x})^T A^{-1}(\boldsymbol{y} - H\boldsymbol{\beta}) \tag{5}$$

whereas, the posterior variance is evaluated as:

$$c(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{t}(\boldsymbol{x})^T A^{-1} \boldsymbol{t}(\boldsymbol{x}')$$
$$+ (\boldsymbol{h}(\boldsymbol{x})^T$$
$$- \boldsymbol{t}(\boldsymbol{x})^T A^{-1} H)(H^T A^{-1} H)^{-1}(\boldsymbol{h}(\boldsymbol{x})^T$$
$$- \boldsymbol{t}(\boldsymbol{x})^T A^{-1} H)^T \tag{6}$$

where

$$\boldsymbol{t}(\boldsymbol{x})^T = c(\boldsymbol{x}, \boldsymbol{x}_1), \dots, c(\boldsymbol{x}, \boldsymbol{x}_n)) \tag{7}$$
$$A = \left[c(\boldsymbol{x}_i, \boldsymbol{x}_j) \right]_{i,j=1}^n, \tag{8}$$
$$H^T = (\boldsymbol{h}(\boldsymbol{x}_1), \dots, \boldsymbol{h}(\boldsymbol{x}_n)) \tag{9}$$

The GPE is used to predict the model evaluation for all the remaining events in D.

2.3 Discussion

In this section, we have explained the details of two methods used to approximate some model evaluations. In both approaches the ideal way of selecting training events for the estimation used is explained, and the estimation technique is also described.

For the LUT approach, training events do not cover the marginal input space efficiently, however it allows for some incorporation of prior knowledge. Additionally, since points need to be in a regular grid format, the training events selected are not necessarily a subset of D, and the number of training events required increases exponentially with the number of input dimensions, making this approach, as a whole, computationally inefficient. However, the second step for the LUT approach is very easy to understand, and although it requires training data on a regular grid, the linear interpolation itself is very efficient.

In contrast, for the GPE approach, the MDA method not only efficiently selects events that are both marginally and multi-dimensionally well spread out, but also selects events that are a subset of D. Moreover MDA is an automated process, and even though, it is computationally time consuming to apply where D is extremely large, in reality, D is seldom significantly large enough.

The output from a GPE is a distribution, which means that not only do we obtain a prediction but we also obtain the value for the uncertainty around this prediction. Moreover, the GPE doesn't require the training events to be in a regular grid format, which is beneficial in that some of the predictions of events in D will be exact, and as the training events are more spread out the GPE can estimate appropriately with fewer

training events. This is illustrated in the next section using a case study.

3 Case study

In this section, the methods described above are compared to find an appropriate prediction of the SWAN model. This case study is based in the Red Sea at Jizan, Saudi Arabia, and is particularly difficult to model as the nearshore area is partly sheltered by shallow reefs (as illustrated in Figure 1). The strong variations in fetch lengths and bathymetry make it a relatively complex area to accurately predict wave conditions.

For this case study, we chose $p = 5$ uncertain parameters to serve as inputs to the SWAN model which describe the boundary conditions. We denote the 5 independent input parameters $\boldsymbol{x} = (x_1, \dots, x_5)$ and are, namely; significant wave height ($H_{s.off}$, measured in meters), peak wave period (T_p, seconds), wind speed (C, meters per second), wave direction (θ_{Wave}, °N) and wind direction (θ_{Wind}, °N). The remaining parameters, e.g. water level are fixed at values chosen by the modelers. The model predicts the corresponding nearshore wave conditions of significant wave height (H_s), mean wave period (T_{m-10}) and mean wave direction (θ) at all points across the model domain.
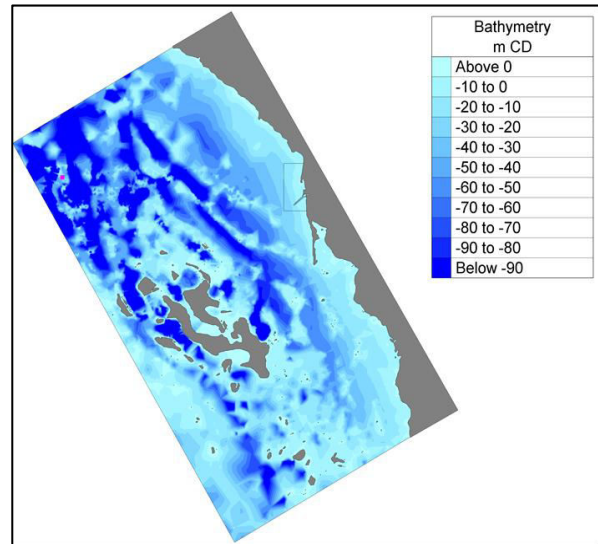

Figure 1. Bathymetry plot of offshore area

The full set of events, D, at which SWAN needs to be evaluated at are a series of boundary wave and wind conditions at 3 hourly time steps starting from 1st December 1983 to 31st December 2009 which total 64949 time steps or events. In both cases the training events were selected based on this entire set of events. Due to time constraints, the model could not be run for all these events, hence SWAN was run for a subset of these events (1st January to 31st December, 1984) which contained 1757 events. These model evaluations are used to assess the accuracy of each approximation technique. No measured data was available to validate the accuracy of the SWAN model itself

We use the MDA method to select up to 1000 training events for the GPE approach. For the LUT approach the time series of offshore waves and wind conditions are discretized into a four dimension look-up table comprising 6 discrete wave height, 10 wave direction, 8 wind speed and 13 wind direction bins. Any combination of these parameters that does not occur in the boundary time series is then ruled out, leaving a total of approximately 1200 training events.

In Figure 2, we show a plot of the selected training events for the two approaches. For the LUT approach we used approximately 1200 training events (asterisks) and for the GPE approach we plot only the first 100 training events (diamonds). The full set of events, *D* are plotted in light grey circular points. From these plots we can see that the points selected for the LUT approach

inefficiently cover the sample space, whereas the MDA selected points for the GPE, are well spread out. In some cases, for instance, the plot of wave direction against the peak period it looks like the points for the LUT approach have a lot fewer events. This is because they are in the form of a recti-linear grid, and many points lie on top of each other.

We used the respective selected training events to approximate the model evaluations at the events generated by SWAN model runs for 1984. We compared the predictions using a Root Mean Square Error (RMSE) which is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(x_i) - f(x_i)\right)^2},\qquad(10)$$
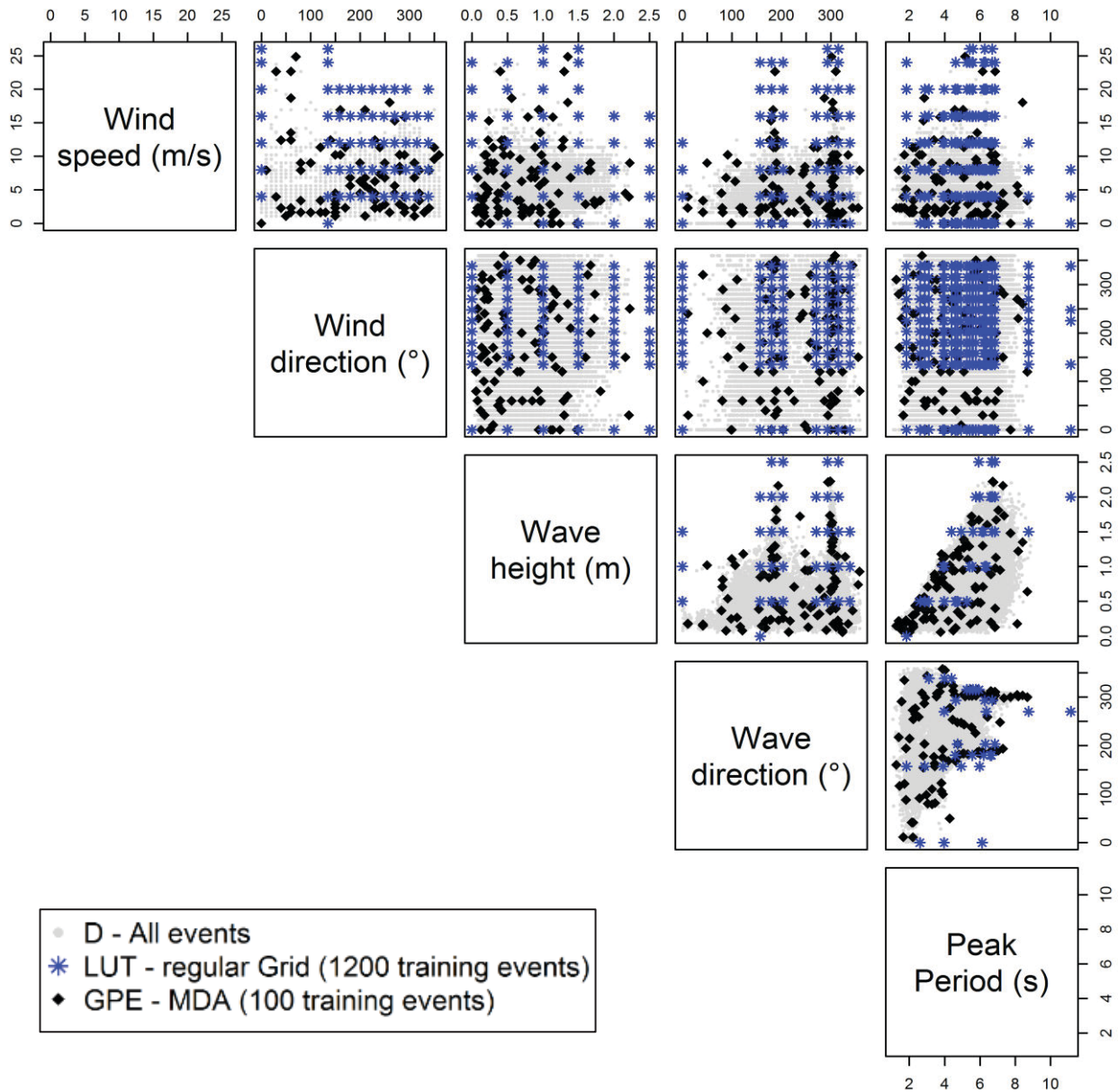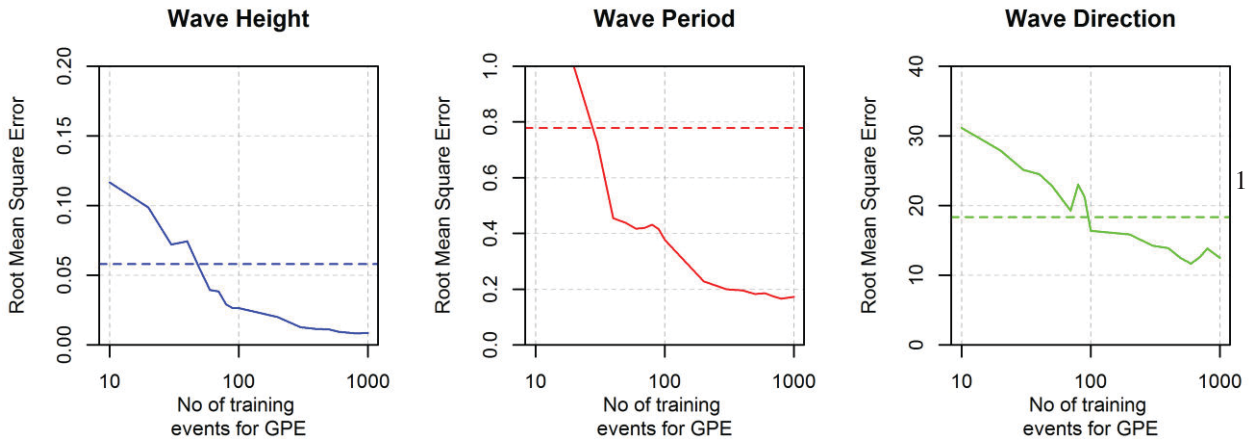

Figure 2. Training events selected using the Regular grid and MDA approach

Figure 3. RMSE for Wave height, period and direction using LUT(dashed line – with 1200 training events) and GPE (solid line – using between 10 to 1000 training events).

where $\hat{f}(x_i)$ represents an approximation to $f(x_i)$ at x_i. The lower the RMSE the better the approximation.

We used 1200 training events to predict the model evaluations (at each outputs) for the year 1984 using the LUT approach, and calculated the RMSE of the predictions against true model evaluations from the validation data. This is represented by the dashed line in Figure 3. For the GPE approach we varied the number of training events from 10 to 1000 to predict the model evaluations to find the minimum number of training events required to match the RMSE of the LUT approach. We also show that if we use nearly as many training events as the LUT approach we can achieve much better accuracies.

Figure 3 compares the RMSEs of the predictions for the three different outputs. It is evident that the GPE approach can achieve similar accuracy using only 100 training events. Moreover if more training events are used, further gains in accuracy can be achieved. It is noticeable that there are a few inconsistencies in the GPE approach, however on average RMSE decreases with an increase in training events. All further analysis assumes that 100 training events were used for the GPE approach.

It is also important to assess how well the predictions are on an individual basis. Figure 4 compares the true SWAN output against the GPE (red) and LUT (blue) predictions. The black solid line represents perfect predictions, the points that are closest to the black line are better predictors.

From Figure 4 it is evident that all the predictions using the LUT approach are more widely spread out, which suggests that the predictions are far from the true SWAN output. The predictions from the GPE approach, on the other hand are less widely spread out, and from this we can conclude that the GPE approach is a much better way of predicting the model.

Figure 5 shows a time series plot of the predicted wave period to illustrate how well the prediction methods compare with the SWAN model results. We chose the wave period plot at this particular time period to illustrate the credible intervals, and to demonstrate the accuracy of a prediction at a training point. The solid (black) line represents the true SWAN output. The long dashed line (blue) represents the predictions form the LUT approach and the dotted and dashed lines (red) represent the GPE approach. This shows that the GPE approach provides a better estimate of SWAN output and the dotted red line represents the 95% credible interval as produced by the GPE. Note, only 20 consecutive events are shown in this figure.
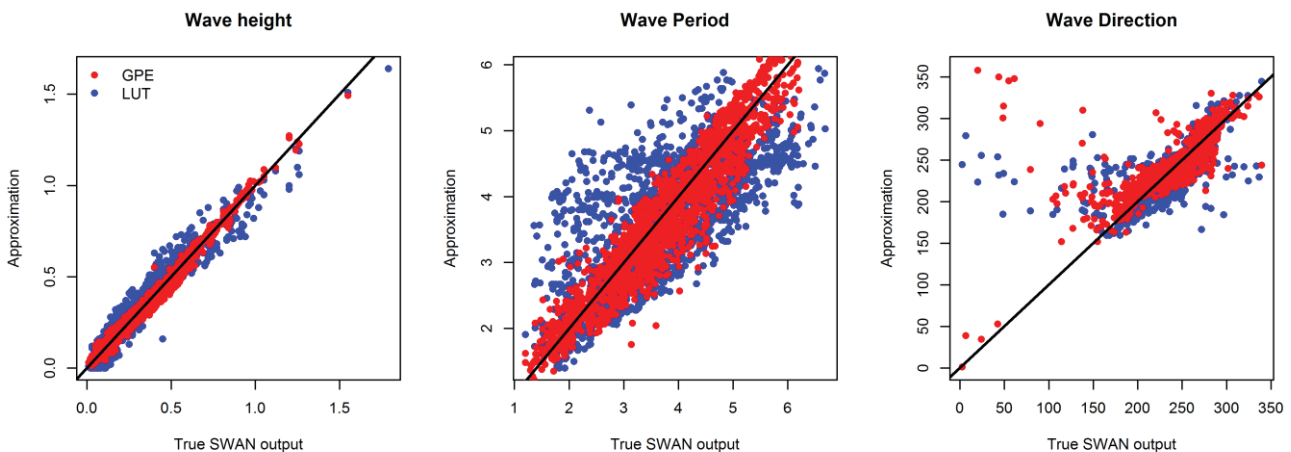
Figure 4. Comparison of approximations for LUT and GPE with the true SWAN output
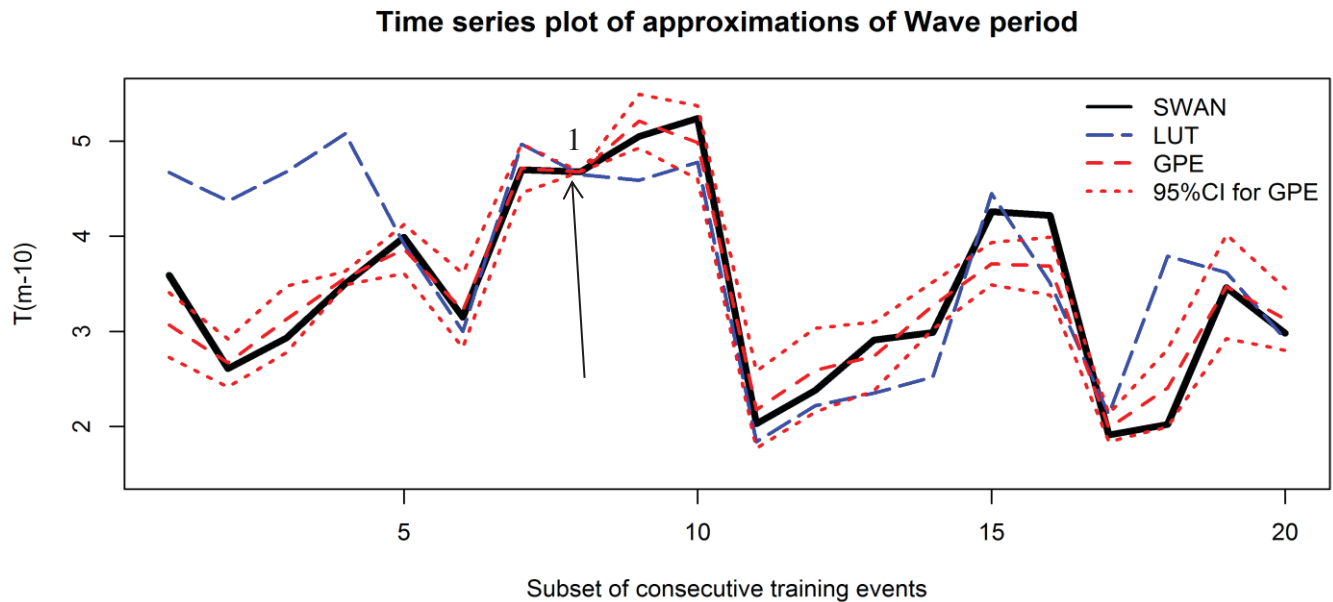
Figure 5. Comparisons of LUT and GPE approximations of Wave Period

Figure 5 shows that the true SWAN output appears between the 95% credible intervals for a majority of time. It is also visible that the credible intervals widens and narrows depending on how uncertain the prediction may be. Moreover, at point 1 (indicated by the arrow in Figure 5), which was one of the training events for the GPE approach, it is evident that the credible interval collapses to zero and the prediction for the GPE approach is equal to its true value. This property holds for all training events in the GPE approach.

4 Discussion and Conclusion

In this paper, we have used a basic univariate GPE approach and shown that, with only a few training events, the GPE approach can give more accurate predictions compared to the a traditional LUT approach. Further gains are expected to be made using multivariate GPE approaches, and perhaps using different training event selection methods.

Moreover, the use of the GPE approach allows us to potentially increase the spatial resolution or complexity of the model to allow more detailed and accurate analysis. The GPE approach also provides a basis for more accurate and efficient uncertainty and sensitivity analysis to be carried out.

This method has proven to be effective in practice. A recent national scale analysis undertaken in England has used the technique to generate a large nearshore wave dataset. This data set has the potential to be used for a wide-range of purposes including a coastal flood risk assessment, and has great potential for use in climate change impact assessments and coastal flood forecasting.

5 References

1. Heffernan, J.E., and Tawn, J.A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **66(3)**, 497–546
2. Lamb, R. . Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P. and Batstone, C, (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, **3(4)**: 323–336
3. Wyncoll D. and Gouldby B. (2014). Integrating a multivariate extreme value method within a system flood risk model, *Journal of Flood Risk Management*, **8(2)**, 145–160
4. Gouldby, B., Mendez, F., Guanche, Y., Rueda, A. and Minguez, R. (2014) A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis, *Coastal Engineering, Vol. 88,* June.
5. Buhmann, M. D. (2004). Radial basis functions: theory and implementations.*Cambridge monographs on applied and computational mathematics*, *12*, 147-165.
6. Booij, N et al. (2004). SWAN cycle III version 40.41 user manual. *Delft University of Technology* 115.
7. Camus, P., Mendez, F.J. and Medina, R. (2011). A hybrid efficient method to downscale wave climate to coastal areas. *Coastal Engineering*, **58(9):** 851-862.
8. Camus, P., Mendez, F.J., Medina, R. and Cofiño, A.S. (2011). Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, **58(6):** 453-462.
9. Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal*

Statistical Society: Series B (Statistical Methodology), *63*(3), 425-464.

10. Kennedy, M. C., Anderson, C. W., Conti, S., & O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, *91*(10), 1301-1309.

11. Santner, T. J., Williams, B. J., & Notz, W.I. (2013). The design and analysis of computer experiments. *Springer Science & Business Media*. 121-161.

12. Snarey, M., Terrett, N. K., Willett, P., & Wilton, D.J. (1997). Comparison of algorithms for dissimilarity-based compound selection. *Journal of Molecular Graphics and Modelling,* **15(6),** 372-385.