

A generic and practical wave overtopping model that includes uncertainty

T. Pullen¹, Y. Liu², P. Otinar Morillas³, D. Wyncoll⁴,
S. Malde⁵ B. Gouldby⁶

¹ Principal Engineer, Coastal Structures, HR Wallingford Ltd, ² Principal Statistician, Flood Risk, HR Wallingford Ltd. ³ Research Fellow, Coastal Structures, HR Wallingford Ltd ⁴ Senior Statistician, Smith Institute ⁵ Statistician, Flood Risk, HR Wallingford Ltd. ⁶ Chief Technical Director, Flood Risk, HR Wallingford Ltd
(Corresponding author: Pullen T (t.pullen@hrwallingford.com))

Published in Proceedings of the Institution of Civil Engineers – Maritime Engineering, 2018

Abstract

Mean wave overtopping discharge is generally accepted to be a primary design criterion for assessing the performance of coastal structures. It is a boundary condition for many coastal flood risk assessments. Modern methods for assessing wave overtopping discharges and their consequences are well documented and reported. Among the various tools available for assessing wave overtopping, the use of artificial neural networks has become increasingly popular. This paper introduces the next stage in the development of these models. Using the same source data, the new generic meta-modelling overtopping model reduces uncertainties and gives clear guidance on the range and validity of the outputs.

Keywords

coastal engineering/hydraulics & hydrodynamics/risk & probability analysis

Notation

D_M	Mahalanobis distance
f	function that relates p input variables x to the output y , such that $y = f(x)$
m	foreshore slope
Q_{pred}, Q_{meas}	discharges predicted and measured, respectively
R	set of all real numbers
R^p	set of all p -dimensional real number vectors
V_{ij}	(i, j) th entry for the covariance function of the Gaussian process
X	$n \times p$ matrix of input variables, with the i th row being the transpose of vector x_i
x	$p \times 1$ column vector representing one observation of the p input variables
x_1, x_2, \dots, x_n	a set of n indexed observation of the p input variables
$x_i^{(k)}, x_j^{(k)}$	k th component of vector x_i and x_j , respectively
y	output overtopping rate
β	common scale factor used in constructing the nuggets for the Gaussian process
δ	size of the 'nugget'
μ	$p \times 1$ column vector representing the column-wise mean of input variables X
Σ	$p \times p$ square matrix representing the variance-covariance matrix of input variables X

σ^2	a positive universal scale factor used in constructing the covariance matrix V for the Gaussian process
τ	minimum period of an periodic input variable
$\varphi_1, \varphi_2, \dots, \varphi_p$	correlation lengths used in constructing the covariance matrix V for the Gaussian process, one for each of the p input variables

1. Introduction

Given the relatively well-defended coastline of developed countries, wave overtopping of coastal structures features significantly in coastal-flood-related failures (both structural failure and excessive flooding). Sea levels are rising and coastal flood risk is likely to increase in the future. Methods for predicting wave overtopping rates have been applied for designing structures and estimating the related coastal flood risk over the last 30 years. There are, however, well-known complexities associated with the dynamic interaction of waves, tides and structures and significant uncertainties remain.

Although there has been an increase in the development and application of advanced computational fluid dynamics (CFD) models (Chen *et al.*, 2016; Dimakopoulos *et al.*, 2014; Higuera *et al.*, 2013), empirical techniques still dominate in practice (Pullen *et al.*, 2007). In general terms the empirical techniques, largely based on physical model experiments, fall into two categories: structure-specific, non-dimensional formulae of the type originally developed around 30 years ago (Owen, 1980); and generic (i.e. not structure-specific) approaches developed using meta-modelling methods. The former are well documented within the EurOtop Manual (Pullen *et al.*, 2007), while the latter have been increasingly applied in recent years and include those developed by a range of authors (Formentin *et al.*, 2017; Kingston *et al.*, 2009; Van Gent *et al.*, 2007; Zanuttigh *et al.*, 2016).

This paper describes the development of a new generic metamodeling overtopping model, based on the application of Gaussian process emulation (GPE) techniques. GPE offers several advantages over previous methods that used neural networks. The model is named Bayonet GPE (BGPE) as it is an evolution of a Bayesian overtopping (Bayonet) model developed by Kingston *et al.* (2009). It provides a more explicit and complete representation of uncertainty and also provides improved guidance to users on the range of validity of the model.

2. Background

2.1. Meta-modelling

Meta-models, sometimes known as surrogate models or emulators, are computationally efficient statistical models. Under this broad category there is a range of different techniques. They essentially act as interpolators between known input/output data points and are often applied to represent computationally intensive models, known as simulators (Sacks *et al.*, 1989). In that context the simulator is typically applied at points (known as design points) selected to cover an appropriate area of the input parameter space. The output of the model is then recorded and the results used to 'train' the metamodel. The meta-model is then applied in the place of the simulator. Given their efficiency they are often applied to computationally intensive procedures; uncertainty and sensitivity analysis and optimisation, for example. These methods can trace their origins in several fields. They closely relate to the response surface method (Box and Draper (1959) and Hill and Hunter (1966)), which is extensively applied in the field of structural reliability analysis. The

methods are also widely utilised in geospatial analysis (Oliver and Webster, 1990). The case study examples have been developed (Kennedy *et al.*, 2006), and a general framework for dynamic meta-modelling is described by Castelletti *et al.* (2012).

2.2. Meta-modelling applications within coastal engineering

In the context of coastal engineering, different meta-modelling techniques have been applied for different processes. For example, radial basis functions and Gaussian process emulators (GPEs) have been used to approximate the Swan wave model (Camus *et al.*, 2011; Malde *et al.*, 2016). A neural network has been used to assess the failure mechanisms of levees that played a critical aspect in the Hurricane Katrina flooding of New Orleans (Kingston *et al.*, 2011).

The technique that has been applied in specific relation to overtopping comprises a neural network applied to a database of physical model experimental data under the EU-funded Clash project (Van Gent *et al.*, 2007), hereinafter referred to as CNN. This model was further refined through the inclusion of a Bayesian approach to represent uncertainties around the fitted parameters and data (Kingston *et al.*, 2009). More recently, a refined neural network has been developed to predict transmission and reflection, as well as overtopping (Zanuttigh *et al.*, 2016). The refinements stem from the inclusion of new data and increased parameter representation in relation to the foreshore slope and structure permeability. This refined model is now available online (hereinafter this is referred to as ONN).

The approach that has been adopted here focuses specifically on wave overtopping and utilises an existing database of physical model data (hence these are the meta-model 'design points') but replaces the neural network with a GPE.

GPEs have a number of desirable properties that offer advantages over alternative approaches (Oakley and O'Hagan, 2002; O'Hagan, 2006; Van de Schoot *et al.*, 2014). They are applied in a Bayesian framework and thereby seek to represent uncertainties explicitly from different sources. In contrast, existing neural network wave overtopping models, applied in practice (CNN and ONN), do not take account of uncertainty within the physical model data themselves and provide confidence interval estimates based on mean estimates from a range of bootstrap samples. No account is taken of the residual errors associated with the fit from each bootstrap sample. This can lead to an underestimation of uncertainty (Heskes, 1997; Kingston *et al.*, 2009; Papadopoulos *et al.*, 2000).

2.3. Overtopping model parameterisation

The CNN comprised 15 parameters. The introduction of the ONN saw the number of input parameters extended to include, among other parameters, the foreshore slope (m). It is of note, however, that the foreshore slope is well known to influence wave conditions significantly at the toe of the structure. The wave conditions at the toe of the structure are an input to CNN and ONN. Hence, the foreshore slope is implicitly accounted for within both of these models. While it is conceivable that residual effects of foreshore slope, not accounted for by wave conditions at the structure toe, are present, it remains to be established that these residual effects are distinguishable from the general noise associated with the overtopping data.

Exploration of the ONN has shown the predicted results are more sensitive to m than would be expected. This may be indicative of over-parameterisation and the associated over-interpolation of noise present in the data, associated with the fitting process.

For reasons relating to economy, known issues with the over-parameterisation of neural network models (Gaume and Gosset, 2003) and insufficient evidence within the data to support the inclusion of additional parameters within BGPE, faith was maintained with the original CNN parameterisation that excludes the foreshore slope (m). It is also of note that recent developments relating to the ONN model have seen the removal of the m parameter from its configuration (Zanuttigh, 2017, personal communication).

2.4. Application limits of generic overtopping models

As with all empirical methods, it is important to consider the range of applicability of the model in terms of the input parameters used for predictions. Meta-modelling methods are developed to predict through the interpolation of known output results. Although the fitted functions can provide estimates beyond the range of the data, there is little theoretical basis for their use in this regard. The CNN, widely used in practice, specifies input parameter guidance based, primarily, on the maximum and minimum range of the data used to train the model, for each of the specific input parameters (Coeveld *et al.*, 2005). It is of note, however, that the practice of using the maximum and minimum values of each input parameter to define the range of applicability is questionable, particularly when parameters are correlated. There can be significant areas of the input parameter space unpopulated and hence predictions in these areas are generated by extrapolation not interpolation. This is illustrated in concept in Figure 1. This effect can be significantly exacerbated in a 15 dimension parameter space. Figure 2 shows an example of the areas of unpopulated parameter space for the existing underlying physical model data used to train the meta-models.

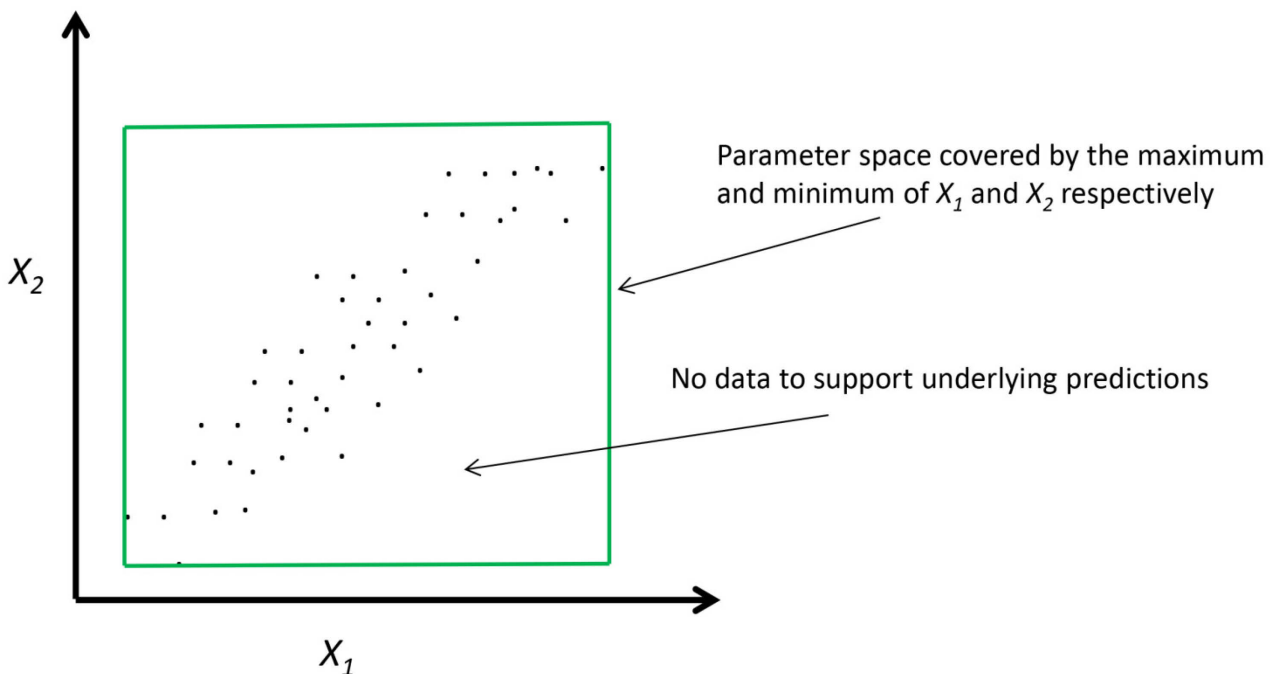


Figure 1: Conceptual diagram showing incomplete parameter space coverage for correlated parameters

The desire to extend the range of applicability of the model outside the range of the training data is perhaps understandable given the prevalence of existing structures in this parameter space and the additional

expense associated with constructing site-specific physical models or more sophisticated numerical models. It is, however, appropriate to recognise and acknowledge explicitly that predictions resulting from extrapolation should be treated accordingly. Within the approach described here, specific attention has been directed towards the provision of guidance relating to the area of applicability of the model. The Mahalanobis distance (MD) (Mahalanobis, 1930) provides a quantifiable measure that can guide users on regions of valid application. This aspect is described in detail below.

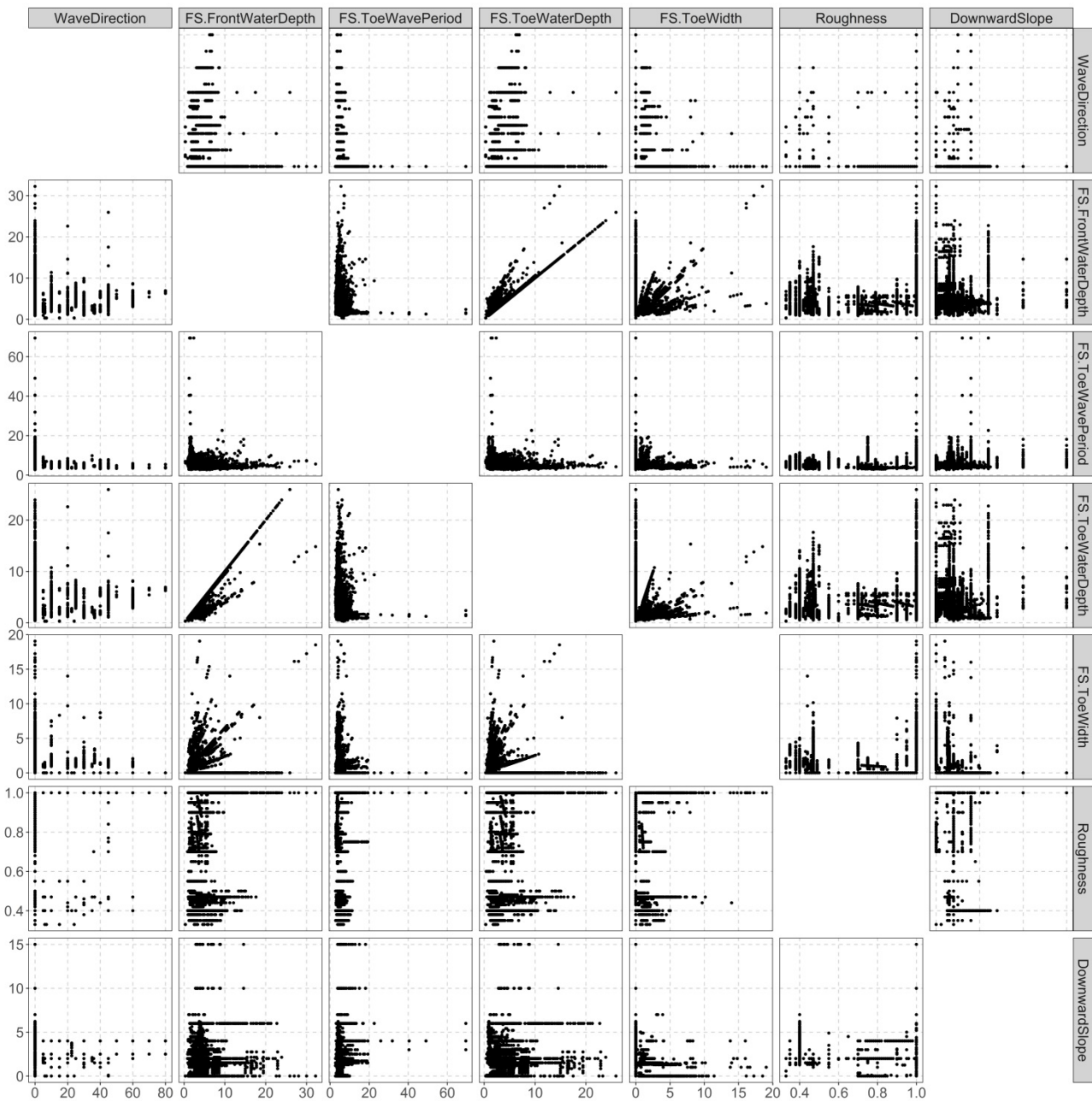


Figure 2: Sub-set of overtopping data showing incomplete parameter space coverage

3. BGPE generic overtopping model

3.1. The overtopping database

The overtopping database (OTD) is based on the database collated under the Clash project (Van der Meer *et al.*, 2009; Van Gent *et al.*, 2007) and extended to include tests completed at HR Wallingford and Ghent University since the conclusion of Clash. The original database comprises data collated from a range of laboratories including the Clash partners, the Danish Hydraulic Institute, Sogreah and the Port Airport and Research Institute (Japan). There are now 13 500 entries of wave overtopping test results, represented by parameters that describe the hydraulic and structural characteristics of the tests. It also contains information about the reliability of the tests and the complexity of the test structures, represented by a reliability factor (RF) and complexity factor (CF), respectively. The RF ranges from 1 (very reliable) to 4 (not reliable), while the CF also ranges from 1 (very simple) to 4 (very complex). In previous analyses these factors have been combined into a single weight factor (WF). The WF was calculated according to $WF=(4-RF)(4-CF)$, so high WF relates to data that are considered reliable.

The WF is used to calculate some of the performance metrics detailed below. In contrast to previous analyses, however, the WF, a somewhat subjective measure, is not an integral part of the fitting of BGPE. Uncertainty associated with the measurements is, however, explicitly included through the inclusion of an error term.

3.2. Emulation method

GPEs are applied in a Bayesian framework and thereby seek to represent uncertainties from different sources explicitly. GPEs return a perfect prediction, with zero variance, at the design points (assuming the training data set does not contain random noise) and hence are relatively flexible and not constrained to a specified regression form. Away from the design points, the variance is non-zero, reflecting uncertainty through consideration of distance from the design points.

In context here, the GPE forms the function (f), which relates a vector of input variables x comprising the structural properties (geometry and protection type) and the input hydraulic loading conditions (sea conditions) to the output overtopping rate (y). A prior distribution, which represents initial beliefs or knowledge about the function f , was specified and then updated to obtain a posterior distribution, with the data available within the database containing the measured overtopping rates, $y = \{y_1 = f(x_1), \dots, y_n = f(x_n)\}$. For any new set of input conditions, \tilde{x} , the posterior mean of the distribution of $f(\tilde{x})$ provides a point estimate (mean) of the overtopping rate, and the posterior variance of f represents the uncertainty around this estimate.

Where it is known the data themselves contain uncertainty (noise), this can be included within the model fitting through an error term known as a ‘nugget’ (Andrianakis and Challenor, 2012). It is well known that there is uncertainty associated with physical model experiments. Successive tests with the same conditions in the same laboratory can yield some variation. Moreover, equivalent tests in different laboratories can exhibit significant uncertainties. It is desirable to account for these uncertainties in predictive models that are based on these data. In the application here, the ‘nugget’ has been used to capture these uncertainties.

Therefore, in summary, the application of the GPE here provides output predictions of wave overtopping rates that include uncertainties associated with the physical model data themselves, as well as uncertainties associated with the model fit in relation to the distance from the design points.

The mathematical formulation of the GPE and ‘nugget’ used in the application here is described in detail in the Appendix.

3.3. Limits of application of BGPE

Like all empirical models, the input parameter range of validity is constrained by the limits of the OTD. Given the 15 dimension set of parameters, a number of which are correlated, it is not straightforward to deduce the range of valid application. To aid practitioners in the application of BGPE, the MD has therefore been used. MD is a measure of a point from a multivariate distribution. Unlike Euclidean distance (ED), the MD accounts for correlated parameters, an important factor in the overtopping data. MD between a specific set of observations x and a set of n -dimensional observations X is given by

$$D_M(x, X) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$$

where the column vector μ and the square matrix Σ are the mean and the variance-covariance matrix of set X respectively.

Figure 3 highlights one element of the set x , and shows the difference between ED and MD. The concentric circles (Euclidean) and ellipsoids (Mahalanobis) are equidistant in terms of their respective measures. The top panel (A) compares two distances to a specific point (the cross symbol) in the sample data set. For example, a point on the coordinates (5, 5) would have an MD of 2 but an ED of 2.6. The point is considered closer using the MD as it recognises the positively correlated nature of the sample data set. Conversely in the lower panel (B), considering the distance between point (2, 5, -2.5) and the cross symbol, the MD of 3 is higher than the ED of 2. Again, this difference arises as a result of the positive correlation. The critical value for the MD for a significance level depends on both the size of the training data set and the number of parameters (Ververidis and Kotropoulos, 2008). The MD is provided as an output from BGPE, alongside estimates of the wave overtopping rate and associated uncertainty.

4. Performance of BGPE

A range of tests have been undertaken to validate BGPE. These include comparison with the measured data, ONN and empirical formulae from the EurOtop manual (EurOtop, 2016). Graphical and quantitative analysis, using the root-mean-square error (RMSE), has been used for the comparison. These comparisons utilise $\log(Q)$, as opposed to Q . This is due to the large range of Q within the data and the form of the overtopping equations, which are based on the observed rates in the physical model experiments. To take into account the reliability of the test measurements and to maintain consistency with previous analyses, the WF was implemented. A weighted RMSE has therefore been calculated as well, which is defined as

$$RMSE_w = \sqrt{\frac{\sum_{i=1}^N WF * (\log(Q_{pred}) - \log(Q_{meas}))^2}{\sum_{i=1}^N WF}}$$

where Q_{pred} and Q_{meas} are the discharges predicted and measured, respectively. All the values of RMSE associated with each assessment carried out in this analysis are presented in Table 1.

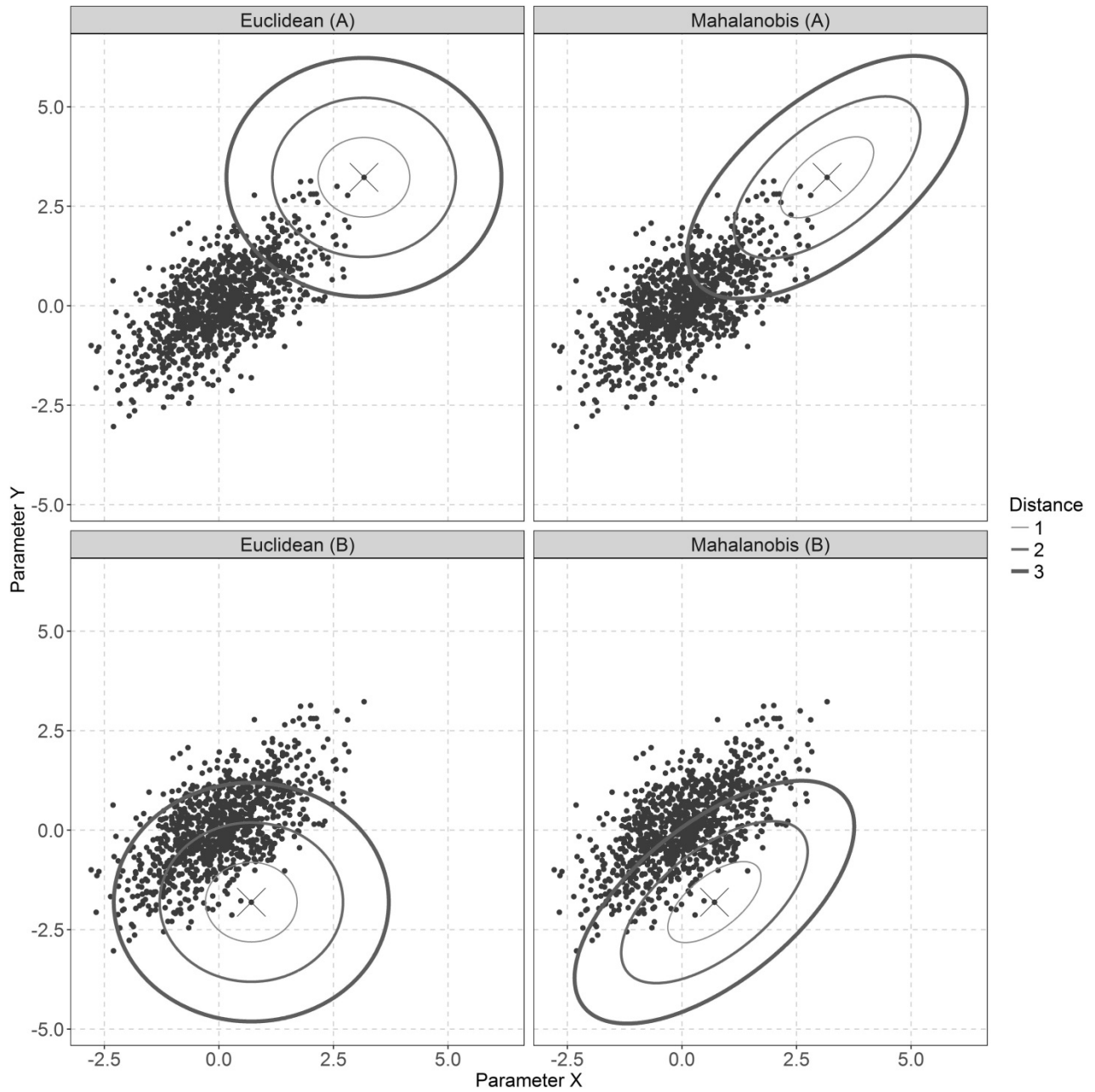


Figure 3: Comparison of the Mahalanobis distance (right column) to the Euclidean distance (left column) – based on the distance between the rest of the sample to point A (top row) and point B (bottom row)

Table 1: RMSE values for each assessment

Model/data	Description	RMSE	Weighted RMSE
BGPE (All data used in training)	BGPE performance when fitted to all of the data and predicting all of the data it has been trained on.	0.23	0.05
BGPE (Validation dataset)	BGPE cross-validation performance when trained on a subset of 90% of the data and used to predict the remaining 10% of data not used in the training.	0.30	0.08
BGPE (Smooth structures)	BGPE performance when trained on all of the data and used to predict a subset of results for simply sloping structures ($y_f = 1$). This facilitates direct comparison with the ONN (row below).	0.21	0.04
ONN (Smooth structures)	ONN performance when trained on all of the data and used to predict a subset of results for simply sloping structures ($y_f = 1$). This facilitates direct comparison with BGPE (row above)	0.28	0.08
BGPE (Smooth simple slope structures)	BGPE performance when trained on all data and used to predict a subset of results for smooth ($y_f = 1$) simply sloping structures. This facilitates comparison with structure specific empirical formulae from EurOtop (row below).	0.21	0.04
EurOtop Empirical Formulae (EFF) (Smooth simple slope structures)	Structure specific EurOtop empirical formulae performance used to predict a set of results for smooth ($y_f = 1$) simply sloping structures. This facilitates comparison with BGPE (row above).	0.52	0.28

4.1. BGPE against measured overtopping rates

In common with previous overtopping models (CNN, ONN), it is possible to compare the predictive performance of BGPE using the measured data from physical models on which it is based. It is, however, important to note that standard BGPEs, unlike ANNs, predict, with no variance, the design points (i.e. the measured data on which the BGPE is based). A resulting plot of measurements against predictions from BGPE would therefore show all data points lying on the perfect prediction line. Similar plots with ANN models tend to show considerable scatter as a result of the ANN smoothing the response function (Van Gent *et al.*, 2007; Zanuttigh *et al.*, 2016). In the case of BGPE, the so-called ‘nugget’ has been introduced to account for the known noise (measurement uncertainty) within the overtopping data (Appendix). Figure 4 shows point prediction outputs from BGPE (i.e. using the mean value of the ‘nugget’) for the measured overtopping data. An inspection of Figure 4 shows a good prediction with little uncertainty for relatively high discharges (>10 l/s/m). This is consistent with findings from comparison of physical model data over many years; this is discussed further in EurOtop (Pullen *et al.*, 2007). Uncertainty increases significantly with lower discharges. Below 1 (l/s)/m 90% of the data fall within + or – a factor of 5. The resulting scatter is directly related to the inclusion of the ‘nugget’ (i.e. the noise in the overtopping data). The implementation of the ‘nugget’ is described in the Appendix. The uncertainty associated with the ‘nugget’ itself can also be included within the predictions and this is shown in Figure 5. Inclusion of the ‘nugget’ uncertainty increases the upper limits of predictive uncertainty from BGPE by approximately an order of magnitude. It is of note that this source of measurement uncertainty is not explicitly included with equivalent ANN models, CNN and ONN. The estimates of uncertainty associated with these neural network models should therefore be treated with this in mind.

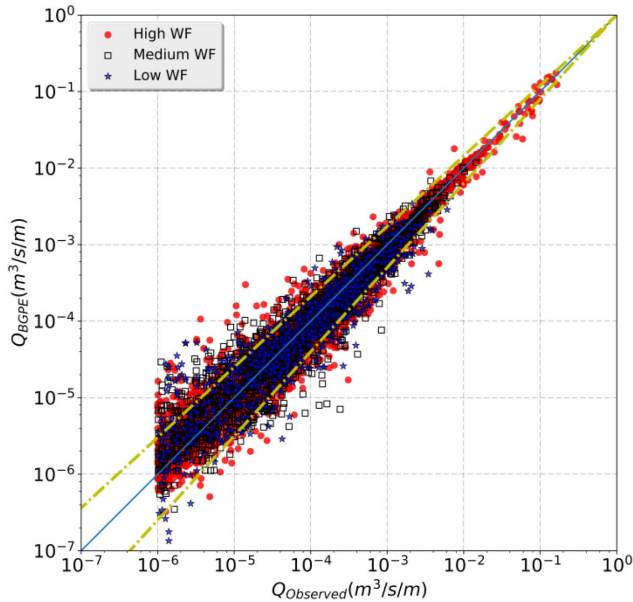


Figure 4: Performance of the BGPE based on the measured data – mean predictions

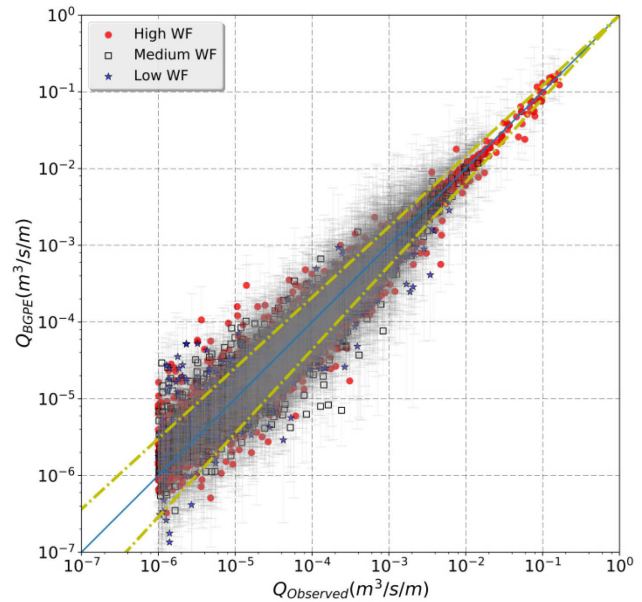


Figure 5: Performance of the BGPE based on the measured data – with 90% credible intervals

4.2. BGPE cross-validation

While it is insightful to understand the performance of BGPE against the measured data it has been trained with, for application in practice it is more important to understand its performance using data that were not available for training. To do this it was necessary to undertake a cross-validation analysis, whereby a subset of data was selected for validation purposes.

Approximately 10% of the data were chosen to form the validation data, and the BGPE was re-trained on the remaining 90% of the data (i.e. omitting the subset of validation data) and then used to predict the validation data. In a realistic situation, the prediction point should fall within a reasonable distance of the training data set, as measured by MD. Therefore, a hierarchical clustering analysis was used to ensure that both the training and validation subsets are evenly spread across the entire distribution of the full data set.

The results of this analysis for point (mean predictions) of BGPE are shown in Figure 6 and summarised in Table 1. It is evident, as would be expected, that the performance of BGPE on the validation data shows an increase in uncertainty in the predictions. This is a result of the emulator explicitly recognising, and incorporating, the predictive uncertainty that arises as a result of distance from the ‘known’ design points. This increase in uncertainty can be observed in both the scatter of the plot (Figure 6) and the value of the RMSE. In general terms, 90% of the data are within a factor of 10 when using the mean estimate of the ‘nugget’. When uncertainty in relation to the distance from the design points and the ‘nugget’ is included (Figure 7), the credible bounds that cover 90% of the data extend over more than four orders of magnitude in places.

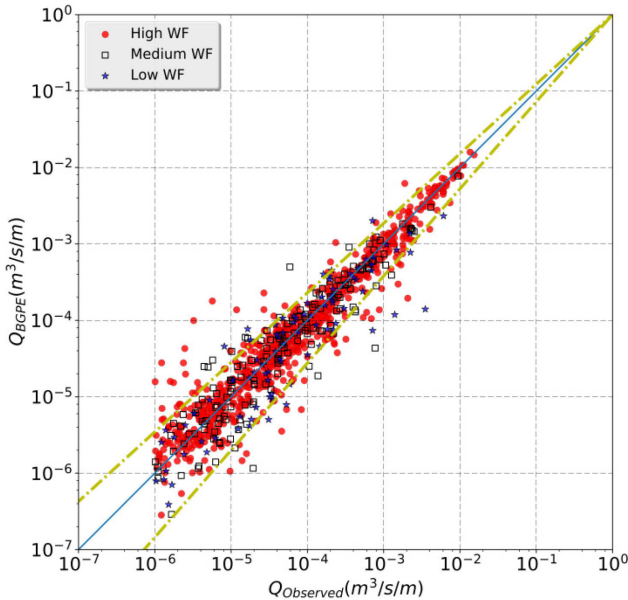


Figure 6: Cross-validation performance of the BGPE using non-training data – mean predictions

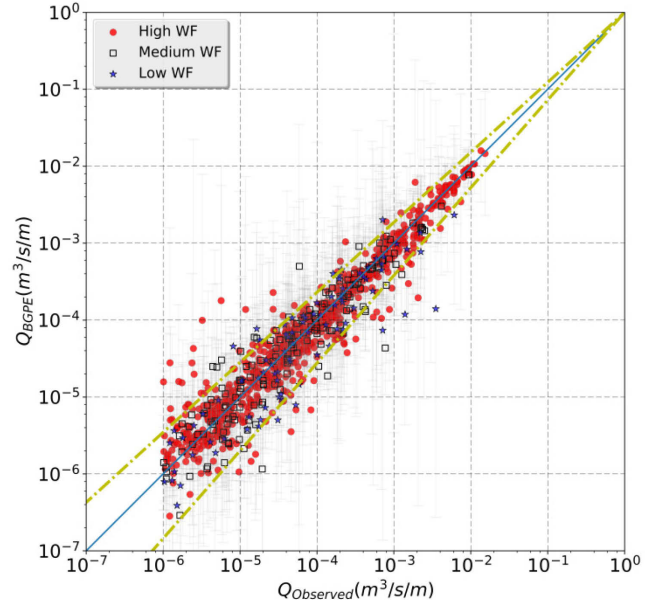


Figure 7: Cross-validation performance of the BGPE using non-training data – with 90% credible intervals

4.3. BGPE against ONN

When comparing BGPE with the ONN, it is necessary to consider the different input parameters used by the models. Given the additional parameters included within the ONN not in BGPE, any comparison between the models needs to define a subset of the measurement data set where there is no influence of the additional input parameters. A data set where the roughness factor (γ_f) was 1 (i.e. impermeable) and a shallow foreshore slope (m parameter in ONN) was present, was therefore used. The comparison of the two models is shown in Figure 8 where it is evident that the BGPE offers a marginal improvement in the prediction over ONN, and this is confirmed by slightly lower RMSE (Table 1).

As well as the point predictions it is also of relevance to compare the estimates of uncertainty (Figure 9). When viewing this comparison it is important to note that this is not a like-for-like comparison. Within the ONN uncertainty estimates, no account is taken of the residual errors associated when fitting the bootstrap samples that are used in the prediction. These residual errors are explicitly accounted for within BGPE through the ‘nugget’ and, moreover, this contributes a significant portion to the overall uncertainty estimate.

Figure 9 shows the noticeable differences in the uncertainty bands associated with the two models. At higher discharges, the ONN has close to two orders of magnitude uncertainty, even though a significant source of uncertainty is not included within the estimate, whereas BGPE is closer to one order of magnitude, when more sources of uncertainty are considered.

In conclusion, for this specific data set, BGPE provides uncertainty estimates that have lower predictive uncertainty than the ONN and also includes more sources of uncertainty within these predictions.

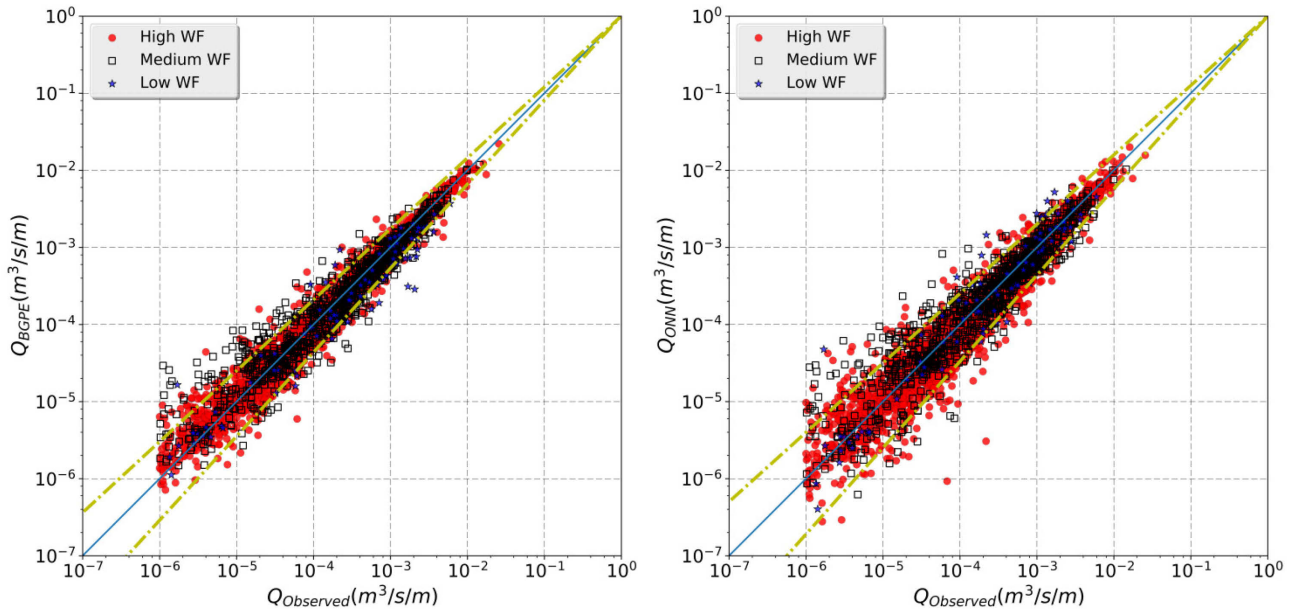


Figure 8: Comparison of BGPE (left) to ONN (right) using data with roughness = 1 – mean predictions

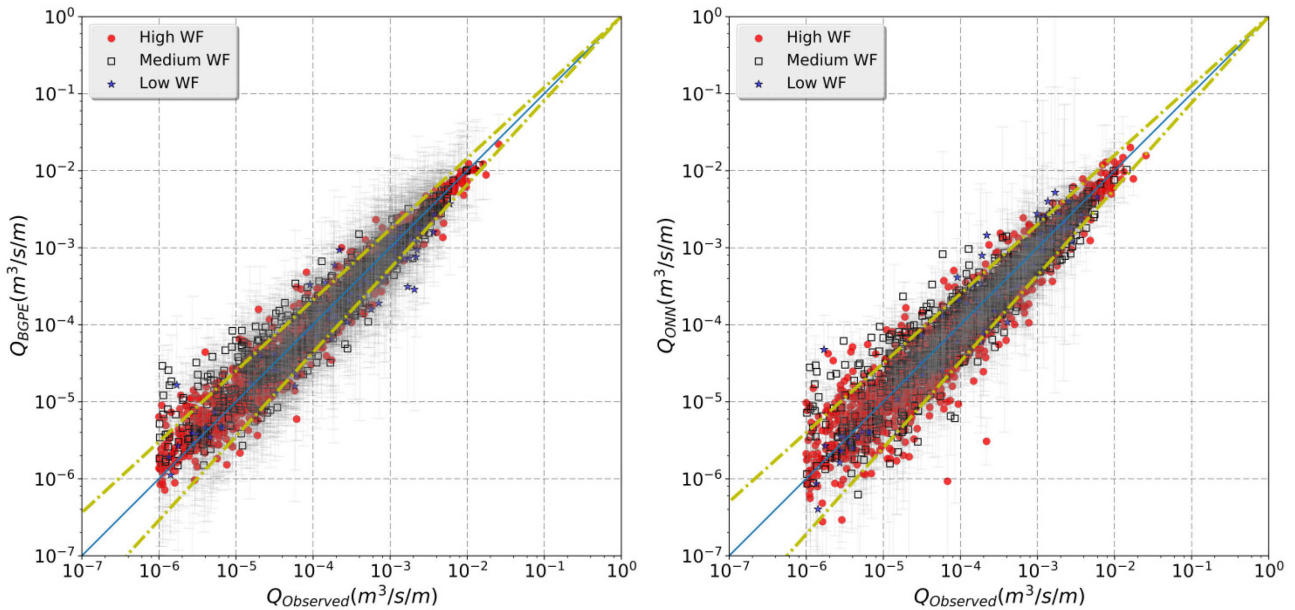


Figure 9: Comparison of BGPE (left) to ONN (right) using data with roughness = 1 – with respective uncertainty measures

4.4. Comparison with EurOtop empirical formulae

In this section, the performance of BGPE is compared with structure-specific non-dimensional formulae from EurOtop II. This manual provides empirical formulae for specific types of structures, but does not cover the full range of structures that appear in the overtopping database. For the purposes of the comparison, smooth simply sloping structures have been selected from the database, so all the formulae that have been applied

are described in chapter 5 of the EurOtop II manual. The comparison of the qualitative predictive performance of these two approaches using the mean estimates is shown in Figure 10. This shows a more robust performance of BGPE and this is confirmed in Table 1.

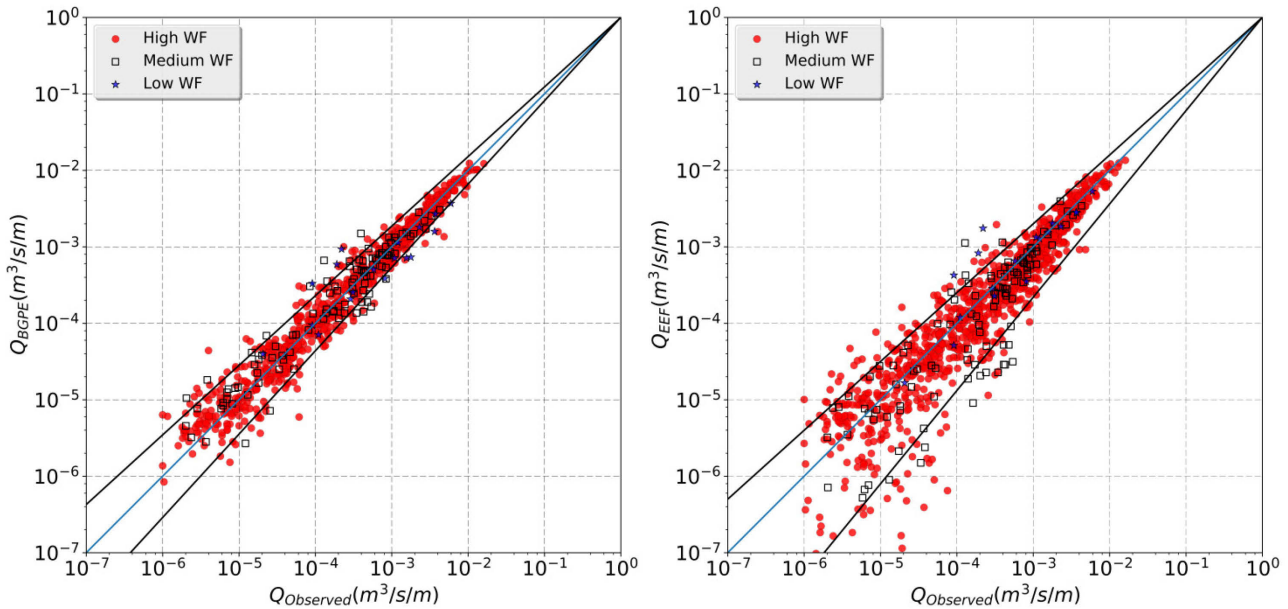


Figure 10: Qualitative prediction performance versus empirical methods

5. Discussion

Estimation of wave overtopping rates is a key aspect in the design of coastal structures and coastal flood risk analysis. It is well known that due to the complexities of the physical processes there are significant uncertainties associated with the estimation of overtopping discharges. Robust engineering design recognises and accounts for these uncertainties.

Existing empirical methods for estimating wave overtopping rates offer limited quantified information on uncertainties. This generally comprises a rule of thumb ‘order of magnitude’ estimate, or quantified estimates that exclude important sources of uncertainty and hence provide uncertainty underestimates. The model, BGPE, enables the estimation of wave overtopping rates with similar or improved performance when compared to existing approaches. In contrast to existing approaches, however, it facilitates a more thorough quantified estimate of uncertainty than has hitherto been possible. Consideration of this uncertainty offers the opportunity for developing a more robust basis for engineering design.

Like all empirical models, the predictive capability of models like the EurOtop formulae, CNN, ONN and BGPE is limited to the bounds of the range of physical model tests on which they are based. Given the multidimensional input parameter space and correlated nature of these input parameters, it is not necessarily apparent when specified structures, and their hydraulic loading conditions, are within the bounds of the physical model tests. Previous guidance in relation to the applicability bounds of the CNN model used minimum and maximum ranges of each input parameter, under the assumption that the input parameters were independent of one another. It has also been shown that this significantly overstates the range of applicability. It has also been shown that the MD, which explicitly includes correlation of input parameters,

offers a useful measure for assessing the proximity of a proposed set of input conditions to the data that underlie the empirical model. This measure can be used to make an assessment of the validity of predictions, thereby avoiding a situation that can arise with users unknowingly using estimates that are outside the range of the underlying data.

6. Conclusions

Practical empirical methods for calculating wave overtopping rates of structures have been applied for many years for coastal engineering design and flood risk analysis purposes. These methods are well known to contain significant sources of uncertainty. There are limitations with regard to the extent of quantitative uncertainty information that has been available to date. BGPE addresses this by implementing a comprehensive method that takes account of a wider range of uncertainties, while maintaining, or in some cases improving, the robustness of the mean predictions.

Current guidance for the applicable range of the CNN wave overtopping model relies on the extrapolation properties of the underlying neural network in some cases. There is little theoretical basis for extrapolating empirical models beyond the range of the data. MD offers a measure that provides useful guidance to users on the range of applicability (i.e. the bounds of the underlying multidimensional database). It is envisaged that further developments will involve the release of this model in a web-based interface. This is with a view to supporting its application in practice.

7. Acknowledgements

The authors express their gratitude to the Environment Agency for partial funding of the development under contract 9NKD-V8N1NW, and EA's project manager Dr Andy Tan. They are also grateful to Andrea Polidoro and Sotiria Eleni Zve, who as research fellows at HR Wallingford, carried out much of the work involved in extending the database and developing and testing early versions of BGPE.

8. Appendix: Mathematical formulation of the BGPE and the 'nugget' effect

A response function $y = f(x)$ for univariate responses $y \in R$ and p -dimensional inputs $x = \{x^{(1)}, \dots, x^{(p)}\} \in R^p$ is said to follow a Gaussian process distribution if for any finite number n , the responses $y = \{y_1 = f(x_1), \dots, y_n = f(x_n)\}$ follow a n -dimensional multivariate normal distribution, the mean and covariance matrix of which are both functions of inputs $\{x_1, \dots, x_n\}$.

The mean function can be any function of $x \in R^p$, but the covariance function must satisfy the property that the resulting covariance matrix is always non-negative-definite. In this paper, the mean of the Gaussian process was assumed to be constant whereas the (i, j) -th entry of the covariance matrix, V_{ij} , was given by

$$V_{ij} = \sigma^2 \exp \left\{ - \sum_{k=1}^p \frac{\|x_i^{(k)} - x_j^{(k)}\|^2}{\psi_k^2} \right\}$$

where variable $x_i^{(k)}$ and $x_j^{(k)}$ are the k th component of input x_i and x_j respectively; parameter σ^2 is a positive universal scale factor; parameters $\{\psi_1, \dots, \psi_p\}$ are called correlation lengths and are applied individually to

each of the p components of the inputs; operator $\|x_i^{(k)} - x_j^{(k)}\|$ denotes the absolute difference $|x_i^{(k)} - x_j^{(k)}|$ if the k -th component of the inputs x is non-periodic, or the circular distance $\min|x_i^{(k)} - x_j^{(k)}|, \tau - |x_i^{(k)} - x_j^{(k)}|$ with period τ if it is periodic.

Under this model, if two x_i and x_j are statistically similar, then the resulting outputs y_i and y_j should be highly correlated. The similarity between x_i and x_j is quantified by the distance between them after normalisation by the correlation lengths ψ . When fitting the Gaussian process to the training data set, parameters σ^2 , φ_k^2 and the constant mean are all estimated using a Bayesian inference method, as introduced in Oakley and O'Hagan (2002).

When accounting for the uncertainty in the output data y , the 'nugget' effect is added to the above formulation, such that the diagonal terms of the covariance matrix are increased by a specific amount (see Andrianakis and Challenor, 2012). Mathematically, the added 'nugget' also increases the stability of the parameter estimation for correlation lengths ψ as the resulting covariance matrix is more likely to be non-negative definite. In this paper, the size of the 'nugget' δ was assumed to be inversely proportional to the WF, that is $\delta = \beta/WF$, where the scale parameter β is estimated jointly with the other parameters for the GPE.

9. References

- Andrianakis I and Challenor PG (2012) The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics and Data Analysis* 56(12): 4215–4228, <https://doi.org/10.1016/j.csda.2012.04.020>.
- Box GEP and Draper R (1959) A basis for the selection of a response surface design. *Journal of the American Statistical Association* 54(287): 622–654.
- Camus P, Mendez FJ and Medina R (2011) A hybrid efficient method to downscale wave climate to coastal areas. *Coastal Engineering* 58(9): 851–862.
- Castelletti A, Galelli S, Ratto M, Soncini-Sessa R and Young PC (2012) A general framework for dynamic emulation modelling in environmental problems. *Environmental Modelling & Software* 34: 5–18.
- Chen Q, Kelly DM, Dimakopoulos AS and Zang J (2016) Validation of the PICIN solver for 2D coastal flows. *Coastal Engineering* 112: 87–98.
- Coeveld E, van Gent M and Pozueta B (2005) *Neural Network Manual*. WL – Delft Hydraulics, Delft, the Netherlands, CLASH WP8, EVKK_CT_2001-00058.
- Dimakopoulos A, Guercio A and Cuomo G (2014) Advanced numerical modelling of tsunami wave propagation, transformation and run-up. *Proceedings of the Institution of Civil Engineers – Engineering and Computational Mechanics* 167(3): 139–151, <https://doi.org/10.1680/eacm.13.00029>.
- EurOtop (2016) *Manual on Wave Overtopping of Sea Defences and Related Structures. An Overtopping Manual Largely Based on European Research, But for Worldwide Application*. EurOtop. See <http://www.overtoppingmanual.com> (accessed 18/09/2018).
- Formentin SM, Zanuttigh B and Van der Meer JW (2017) A neural network tool for predicting wave reflection, overtopping and transmission. *Coastal Engineering Journal* 59(2): 1750006.

- Gaume E and Gosset R (2003) Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology? *Hydrology and Earth System Sciences Discussions*, European Geosciences Union 7(5): 693–706.
- Heskes T (1997) Practical confidence and prediction intervals. In *Advances in Neural Information Processing Systems* (Mozer M, Jordan MI and Petsche T (eds)). MIT Press, Cambridge, MA, USA, vol. 9, pp. 176–182.
- Higuera P, Lara JL and Losada IJ (2013) Simulating coastal engineering processes with OpenFOAM®. *Coastal Engineering* 71(1): 118–134.
- Hill WJ and Hunter WG (1966) A review of response surface methodology: a literature survey. *Technometrics* 8(4): 571–590.
- Kennedy MC, Anderson CW, Conti S and O’Hagan A (2006) Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety* 91(10): 1301–1309.
- Kingston G, Robinson D, Gouldby B and Pullen T (2009) Reliable prediction of wave overtopping volumes using Bayesian neural networks. *Flood Risk Management: Research and Practice* (Samuels P, Huntington S, Allsop Wand Harrop J (eds)). CRC Press/Balkema, Leiden, the Netherlands, pp. 101–102.
- Kingston GB, Rajabalinejad M, Gouldby BP and Van Gelder PHAJM (2011) Computational intelligence methods for the efficient reliability analysis of complex flood defence structures. *Structural Safety* 33(1): 64–73.
- Mahalanobis P (1930) On tests and measures of group divergence I. Theoretical formulae. *Journal and Proceedings of the Asiatic Society of Bengal* 26: 541–588.
- Malde S, Wyncoll D, Oakley J, Tozer NP and Gouldby BP (2016) Applying emulators for improved flood risk analysis. In *E3S Web of Conferences*. EDP Sciences, Les Ulis, France, vol. 7, p. 04002.
- Oakley J and O’Hagan A (2002) Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika* 89(4): 769–784.
- O’Hagan A (2006) Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety* 91(10–11): 1290–1300.
- Oliver MA and Webster R (1990) Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System* 4(3): 313–332.
- Owen M (1980) *Design of Seawalls Allowing for Wave Overtopping*. Hydraulics Research, Wallingford, UK, HR Report EX 924.
- Papadopoulos G, Edwards PJ and Murray AF (2000) Confidence estimation methods for neural networks: a practical. *Proceedings of the ESAAN 200*, Bruges, Belgium, pp. 75–80.
- Pullen T, Allsop W, Bruce T et al. (2007) *EurOtop: Wave Overtopping of Sea Defences and Related Structures – Assessment Manual*. See <http://www.overtoppingmanual.com> (accessed 18/09/2018).
- Sacks J, Welch WJ, Toby JM and Wynn HP (1989) Design and analysis of computer experiments. *Statistical Science* 4(4): 409–423.
- Van der Meer JW, Verhaeghe H and Steendam GJ (2009) The new wave overtopping database for coastal structures. *Coastal Engineering* 56(2): 108–120.
- Van de Schoot R, Kaplan D, Denissen J et al. (2014) A gentle introduction to Bayesian analysis: applications to developmental research. *Child Development* 85: 842–860, <https://doi.org/10.1111/cdev.12169>.

Van Gent MRA, Van den Boogaard HFP, Pozueta B and Medina JR (2007) Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering* 54(8): 586–593.

Ververidis D and Kotropoulos C (2008) Gaussian mixture modelling by exploiting the Mahalanobis distance. *IEEE Transactions on Signal Processing* 56(7): 2797–2811, <https://doi.org/10.1109/TSP.2008.917350>.

Zanuttigh B, Formentin SM and Van der Meer JW (2016) Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network. *Ocean Engineering* 127: 7–22.