A Bayesian estimation of parameter-induced uncertainty in a nearshore alongshore current model

B. G. Ruessink

ABSTRACT

In many process-based models, parameters have to be estimated from data. It is important to obtain not only the optimum value of the parameters, but also to assess the uncertainty in the parameters and, hence, in the models’ output. In this paper, the Bayesian Monte Carlo technique known as Generalised Likelihood Uncertainty Estimation (GLUE) is used to evaluate the parameter-induced predictive uncertainty of a three-parameter model that predicts alongshore currents over a nearshore barred profile. GLUE performs a fully random sampling of feasible-parameters space, assigning non-zero likelihoods to those model simulations that outperform a user-defined threshold. Based on data gathered at six cross-shore position across an inner bar at Egmond aan Zee, The Netherlands, non-zero likelihoods were found for a rather wide range of parameter values, largely induced by an interdependence between two parameters that affect the width of current jets across the bar. The width of the 95% uncertainty interval was found empirically to increase linearly with the predicted magnitude of the alongshore current, from about 0.02–0.06 m/s when the current magnitude is near zero to about 0.2 m/s when it is near its maximum of about 1.1 m/s. These widths are approximately equal to a rough estimate of the errors in the data. In many cases the 95% uncertainty interval brackets the observations, although there are also various instances where this is not the case and apparently model structural errors dominate over parameter-induced errors. Model non-linearity and parameter interdependence cause the marginal parameter posterior distributions to differ remarkably from those obtained from traditional first-order approximations.

Key words | calibration, GLUE, nearshore, physically based simulation modelling, uncertainty analysis

NOTATION

\( C \)  
\( F \)  
\( i, p \)  
\( J \)  
\( J_m \)  
\( k_a \)  
\( L \)  
\( L(\theta|\bar{y}_{obs}) \)  
\( N_a \)  
\( N_p \)  
\( N_t \)  
\( p(\theta) \)  
\( p(\theta|\bar{y}_{obs}) \)  
\( q \)  
\( r \)  
\( R \)  
\( R_M \)  
\( t \)

\( \text{covariance matrix} \)
\( \text{residual quadratic error at } \theta_{\text{opt}} \)
\( \text{integers} \)
\( \text{cost function} \)
\( \text{Jacobian matrix} \)
\( \text{apparent bed roughness} \)
\( \text{number of neurons in hidden layer} \)
\( \text{likelihood of parameter set } \theta \)
\( \text{number of behavioural parameter sets} \)
\( \text{number of free model parameters} \)
\( \text{number of temporal observations} \)
\( \text{prior distribution of parameter set } \theta \)
\( \text{posterior distribution of parameter set } \theta \)
\( \text{quantile} \)
\( \text{correlation coefficient} \)
\( \text{Nash & Sutcliffe (1970) efficiency factor} \)
\( \text{threshold efficiency factor } R \)
\( \text{time} \)

doi: 10.2166/jh.2006.009
INTRODUCTION

Process or physically based models are simplified representations of natural systems, containing equations that express scientifically accepted principles, for instance, continuity and energy and/or momentum conservation. Process-based modelling is directly linked to the advancement of system understanding as modelling provides a framework for testing new hypotheses (e.g. by comparing model predictions to data) and a means to study how processes interact. However, even the most elaborate process-based model is imperfect, implying that model results are inherently uncertain and that the quantification of this uncertainty, for instance as a prediction uncertainty interval, may be as important as the modelling effort itself.

Different sources of uncertainty can be distinguished in process-based modelling. One source of prediction uncertainty is due to model-structure errors, induced by crudely described, incorrectly omitted or even unknown processes. For instance, many coastal modellers have long struggled to correctly predict the onshore migration of nearshore sandbars under low-energy, non-breaking conditions (e.g. Thornton et al. 1996; Gallagher et al. 1998). Recently, Hoefel & Elgar (2003) have argued that this is related to the neglect of fluid accelerations in the sediment transport computations. Observational errors induce another source of uncertainty, related both to the data required to drive the model (input errors) and to the data necessary to check or optimize model performance (output errors). For instance, Guinot & Gourbesville (2003) showed how a crudely schematized catchment geometry could result in a total mismatch of the height and timing of the peak river discharge at the downstream end of the catchment. Another problem of process-based models (and the focal point of the present work) is the limited knowledge of the parameters that govern the model equations. The presence of parameters is a direct consequence of the need to simplify reality. Essentially, parameters characterize lacking complexity or non-described spatial–temporal variability of the processes considered. Parameters do not always represent measurable attributes of the modelled system and, therefore, their values have to be determined through calibration using observed input/output relationships.

Within coastal modelling, which is the field of modelling considered in this paper, manual calibration is by far the most adopted calibration method. This involves numerous trial-and-error computations, in which parameters are varied one at a time, and visual inspections between model results and observations. Although the latter is occasionally supported by the quantifications of some measure of ‘goodness-of-fit’, manual calibration is subjective and time-consuming; in addition, there is no knowledge on whether the appropriate parameter values have indeed been found and information on their accuracy is lacking altogether. Examples of automated calibration algorithms have just begun to appear in the coastal literature. Plant et al. (2004), for example, adopted the local-search, gradient-based Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) to find the optimum parameter settings of a nearshore bed evolution model and imposed a multi-variate normal distribution to estimate confidence intervals for the optimum parameter values. This so-called first-order approximation to assess parameter uncertainties assumes the parameters to be independent and the model to be linear, at least in the vicinity of the optimum. Both assumptions are unlikely to be met in nearshore process models (e.g. Ruessink 2004). Later on, Ruessink (2004)
presented a hybrid genetic algorithm, comprising a global population-evolution-based search strategy and a local Nelder–Mead (Nelder & Mead 1965) simplex search, to calibrate nearshore process models. To obtain uncertainty information about the optimal solution, Ruessink (2004) used multiple calibration periods. The standard deviation in all obtained optimum values was then considered as a measure of parameter uncertainty. Because in this example the number of calibration periods was limited, the inferred parameter uncertainty is likely to have been inaccurate.

In this paper the Bayesian Monte Carlo technique known as Generalized Likelihood Uncertainty Estimation (GLUE) (Binley & Beven 1991; Beven & Binley 1992) is adopted to show how parameter uncertainty in coastal (nearshore) process-based models can be estimated and how this uncertainty is reflected in the estimation of parameter-induced predictive uncertainty. GLUE evaluates multiple model runs resulting from different parameter sets within a given model structure by assigning a likelihood value to each parameter set. Based on these likelihoods, predictive uncertainty can be quantified. Here, GLUE is applied to the simulation of alongshore currents in the nearshore of Egmond aan Zee (The Netherlands) using the process-based model developed by Ruessink et al. (2001). In the Discussion section, various subjective choices in the likelihood assignment are examined and the estimated marginal posterior parameter distributions are compared to those based on first-order approximation; also, the relative importance of parameter-induced uncertainty and measurement errors is discussed.

**MODEL**

**Description**

The alongshore current is a time-averaged (over ∼3600 s) current that flows parallel to the coast and, in the absence of alongshore variability in bathymetry, is forced primarily by obliquely incident breaking waves, the alongshore component of the wind stress and 10–100 km scale alongshore surface slopes owing to tides. Its cross-shore distribution over an arbitrary cross-shore depth profile can be computed by solving the one-dimensional, time- and depth-averaged alongshore momentum balance, in which the three forcing terms balance with bottom stress and lateral mixing. Here, the alongshore current model as proposed by Ruessink et al. (2001) is adopted. The wave-forcing term in the model follows from the model’s wave module, which requires a cross-shore depth profile and values of the offshore water level, root-mean-square wave height, peak period and angle of incidence as input. The wind and tidally induced surface slope are (measured) inputs to the current model. Alongshore variations in morphology, waves and currents are neglected, as are Earth rotation, variation of water density and fluid acceleration.

On a barred profile the model produces current (v) jets that are located at or on the shoreward side of each bar and near the shoreline (Figure 1). The magnitude of v, the location of the maximum current of each jet (v_{max}) and the cross-shore width of each jet are determined by three parameters. The wave-front slope β influences the breaking-wave forcing; a decrease in β shifts the v_{max} location onshore and broadens the current jet by increasing v in the trough (Figure 1(a)). The apparent bed roughness k_{a}, a parameter within the bottom stress formulation, affects the magnitude of v but does not alters the cross-shore shape of v (Figure 1(b)). The magnitude of the depth-averaged eddy viscosity ν determines the degree of lateral mixing, which smooths the cross-shore distribution of v without shifting the location of v_{max} (Figure 1(c)). For more details, see Ruessink et al. (2001).

**Earlier experience**

The model has previously been applied to mean alongshore currents observed on the barred beach at Egmond, The Netherlands (Ruessink et al. 2001). The observations span 500 h at six cross-shore locations, labelled E1–E6 from offshore to onshore (Figure 1(d)), in October–November 1998 when alongshore non-uniformities in the bathymetry, such as rip channels, were sufficiently small to warrant the use of a one-dimensional model. Details on data acquisition and processing are given in Ruessink et al. (2001). The present work also focuses on the available Egmond data.

A manual calibration with the 500 h Egmond data resulted in good agreement between measured and modelled v for β = 0.05, k_{a} = 0.022 m and ν = 0.5 m^2/s (Ruessink et al. 2001). With these settings the Nash &
The Generalized Likelihood Uncertainty Estimation (GLUE), introduced by Binley & Beven (1991) and Beven & Binley (1992), is a Bayesian Monte Carlo simulation based technique to compute posterior likelihood distributions of the free model parameters and to estimate uncertainty bounds on model predictions. Updating of these distributions and uncertainty bounds is possible when new data becomes available (e.g. Beven & Binley 1992; Freer et al. 1996) using the Bayes equation.

An important step in GLUE is the definition of a prior distribution of the parameter set \( \theta \), denoted \( p(\theta) \), which reflects the initial knowledge of feasible parameter space. In practice, the prior distribution is often defined by assigning lower and upper bounds to each parameter, and by using a uniform distribution between the bounds. In this way, GLUE is essentially an extension of Uniform Random Sampling (URS), one of the earliest (e.g. Brooks 1958) global optimization techniques. URS samples a user-defined number of parameter sets from feasible space using a non-informative uniform sampling strategy. Once the parameter sets are selected, the model is run for each set. The model output for each run is then quantified into some measure of ‘goodness-of-fit’, for which the Nash & Sutcliffe (1970) efficiency factor \( R \) is usually taken. The parameter set that performs best in terms of \( R \) is then assumed to be the global optimum. In GLUE, the URS concept of an ‘optimal’ set of parameter values for a given model structure and observed input–output is abandoned. Instead, all parameter sets that perform better that some user-specified threshold, \( R_M \), are considered as acceptable (or behavioural) sets. In the present work \( R_M \) was set to 0.8937, the \( R \) based on

\[
R = 1 - \frac{\sum_{x,t} \left( \bar{w}_{\text{obs}}(x,t) - \bar{w}_{\text{mod}}(x,t) \right)^2}{\sum_{x,t} \left( \bar{w}_{\text{obs}}(x,t) - \bar{w}_{\text{obs}}(x,t) \right)}
\]  

in which the subscripts ‘obs’ and ‘mod’ refer to an observed and modelled value, respectively, and the double overbar represents a temporal (dataset) average value at a specific cross-shore location. Later on, Ruessink (2004) applied a hybrid genetic algorithm to the same data and found that the optimum settings were \( \beta = 0.055, k_a = 0.026 \text{ m} \) and \( v = 1.50 \text{ m}^2/\text{s} \). Although these settings are notably different from their manually determined values, in particular \( v \), \( R \) improved marginally to 0.8996. This signifies that virtually the same model skill can be obtained for rather different parameter settings.

**METHODOLOGY**

The Generalized Likelihood Uncertainty Estimation (GLUE), introduced by Binley & Beven (1991) and Beven & Binley (1992), is a Bayesian Monte Carlo simulation based technique to compute posterior likelihood distributions of the free model parameters and to estimate uncertainty bounds on model predictions. Updating of these distributions and uncertainty bounds is possible when new data becomes available (e.g. Beven & Binley 1992; Freer et al. 1996) using the Bayes equation.

An important step in GLUE is the definition of a prior distribution of the parameter set \( \theta \), denoted \( p(\theta) \), which reflects the initial knowledge of feasible parameter space. In practice, the prior distribution is often defined by assigning lower and upper bounds to each parameter, and by using a uniform distribution between the bounds. In this way, GLUE is essentially an extension of Uniform Random Sampling (URS), one of the earliest (e.g. Brooks 1958) global optimization techniques. URS samples a user-defined number of parameter sets from feasible space using a non-informative uniform sampling strategy. Once the parameter sets are selected, the model is run for each set. The model output for each run is then quantified into some measure of ‘goodness-of-fit’, for which the Nash & Sutcliffe (1970) efficiency factor \( R \) is usually taken. The parameter set that performs best in terms of \( R \) is then assumed to be the global optimum. In GLUE, the URS concept of an ‘optimal’ set of parameter values for a given model structure and observed input–output is abandoned. Instead, all parameter sets that perform better that some user-specified threshold, \( R_M \), are considered as acceptable (or behavioural) sets. In the present work \( R_M \) was set to 0.8937, the \( R \) based on
Ruessink et al. (2001)’s manual calibration. This implies that all parameter sets that outperform the manual calibration results were considered behavioural, whereas all sets that did worse than the manually determined results were considered non-behavioural. Because $R_M$ is close to the maximum $R$ (based on the optimum parameter settings), this threshold choice can be considered as very strict. The chosen lower and upper parameter bounds in the present GLUE application are listed in Table 1 and were based on earlier experience with the model.

The use of a uniform prior distribution often leads to a very inefficient way of estimating the posterior parameter distributions (e.g. Khu & Werner 2003; Hossain et al. 2004; Uhlenbrook & Sieber 2005). The main inefficiency comes from the often computationally expensive model calls to compute $R$. Because those parts of feasible space that turn out to be most promising are sampled equally densely as non-behavioural parts, a considerable number of simulations are, in the end, performed in vain. In the present work, a single simulation took about 6 min on a standard 2.53 GHz PC. It would take, for instance, more than 24 d (longer than the duration of the available Egmond data!) to complete 5000 model simulations of which, say, only 10–20% might turn out to be behavioural.

To avoid these excessive computational efforts, the alongshore current model was replaced by an artificial neural network (ANN) after 1000 model simulations. An ANN is a universal approximator for non-linear function mappings between multi-dimensional spaces, in which the mappings take the form of simple processing units (neurons) with adjustable biases and weights. Here, the mapping capability of a standard multi-layer perceptron (MLP) was used to reconstruct $R$ based on the 1000 parameter sets used in the alongshore current model. The applied MLP comprised an input layer with 5 neurons (one for each parameter), a hidden layer with $L$ neurons and an output layer with a single neuron for $R$. The hidden layer had hyperbolic tangent transfer functions, whereas those in the output layer were linear. Training of the network was performed with the backpropagation algorithm, using the Levenberg–Marquardt method to optimize the adjustable parameters.

One of the problems in implementing neural networks is the dependence of the output on the initialization of the ANN biases and weights (Gardner & Dorling 1998; Hsieh & Tang 1998). Because of local minima in the cost function (that is, the mean squared difference $J$ between $R$ predicted by the alongshore current model and the ANN), a different initialization will cause the network to end up in a different local minimum. A second problem is overfitting (e.g. Weigend et al. 1990). When an ANN is overfit, it performs very well on the training data but poorly on separate cross-validation or test data. This means that the neural network tends to fit the noise in the training data at the cost of its skill to generalize. Here, following Hsieh (2001), the local minima problem was handled by training 30 individual neural networks, initialized with different random biases and weights. During the training of each ANN, 20% of the input data (200 parameter sets) was randomly selected for cross-validation and thus withheld from training. ANNs having a mean squared difference for the cross-validation data that was 20% larger than the mean squared difference for the training data were considered overfit and discarded. After training, each individual non-overfit ANN possessed a different value of the cost function because it had ended up in a different local minimum. The neural network solution

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Central tendency and dispersion of prior and posterior parameter distributions at Egmond</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.03</td>
</tr>
<tr>
<td>$k_a$</td>
<td>0.005</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0</td>
</tr>
</tbody>
</table>
used below was taken as the non-overfit individual with the lowest \( J \). With the best ANN replacing the alongshore current model, the sampling of feasible parameter and the computation of \( R \) was continued with negligible computational effort.

Sampling of parameter space was terminated when a total \( N_a \) of 500 acceptable parameter sets was obtained. Each acceptable parameter set \( \theta_i \) was then assigned a likelihood that is proportional to \( R \). All parameter sets with \( R \) less than \( R_M \) have a zero likelihood. Thus

\[
L(\theta_i | \bar{v}_{obs}) = \begin{cases} 
0 & \text{if } R < R_M \\
\lambda R & \text{if } R \geq R_M 
\end{cases}
\]

where \( L(\theta_i | \bar{v}_{obs}) \) is the likelihood measure for the \( i \)th parameter set \( \theta \) conditioned on the observations \( \bar{v}_{obs} \) and \( \lambda \) is a normalization factor such that the cumulative sum of all likelihood measures equals 1. Finally, the likelihoods were applied to their respective model predictions to give a cumulative probability distribution of model output. Note that this required that the alongshore current model was run with all acceptable parameter sets produced by the ANN, as the ANN estimates \( R \) but not the associated \( \bar{v}_{mod}(x,t) \). From this cumulative probability distribution, the 0.5\% and (1–0.5\%) quantiles \( q \) can be calculated to represent the 100\% parameter-induced model uncertainty. In the present work, \( \alpha \) was set to 0.05. If, for the moment, it is assumed that a parameter has an approximately normal posterior distribution, then the quantiles \( q = 0.025 \) and 0.975 are estimated to an accuracy of \( \sqrt{q(1-q)/N_a} = \sqrt{0.025(1-0.025)/500} = 0.007 \), which is good enough for the present work.

**SIMULATION RESULTS**

After a heuristic search over the range \( L = [5\ldots20] \), \( L = 15 \) was selected as the ‘optimum’ number of neurons in the hidden layer. With this number, the best ensemble member reproduced the \( R \) values of the training and verification data very well (Figure 2). The best-fit linear regression line between the \( \bar{v} \) model and the ANN-estimated \( R \) in the verification data (Figure 2(b)) has a slope of \( \sim 1 \), an intercept of \( \sim 0 \) and skill \( r^2 \) in excess of 0.99, showing the suitability of this ANN to replace the alongshore current model to estimate \( R \). With an increase in \( R \) the performance of the best ensemble ANN member did not improve relative to \( L = 15 \), whereas it deteriorated somewhat for ANNs with less than 15 neurons in the hidden layer.

In total, about 12,000 parameter sets were evaluated to obtain 500 acceptable simulations, implying a ratio of acceptable to total number of simulations of about 4.2\%. In hindsight, this low ratio clearly highlights the need for the use of the ANN as a surrogate to the alongshore current model. Figure 3 demonstrates \( R \) for each of the three free parameters. Each dot represents one of the approximately 12,000 parameter sets. Figure 4 shows the likelihood values \( L(\theta_i | \bar{v}_{obs}) \) as a function of the free model parameters. Note that, because all behavioural \( R \) vary in a very small range (0.8937–0.8996), the likelihood of each behavioural
parameter set is approximately equal to $1/500 = 0.002$. Intriguingly, behavioural simulations can be found for a wide range of parameter values. For instance, behavioural $\nu$ ranges from 0.3 – 4.2 m$^2$/s (Figure 4(c), Table 1), almost the entire range of the prior distribution, and behavioural $\beta$ can vary between 0.032 – 0.108 (Figure 4(a), Table 1). Of course, every parameter value that results in an $R$ equal to or larger than $R_M$ can also result in poor simulations owing to the values of the other parameters.

The rather wide posterior ranges for, in particular, $\beta$ and $\nu$ may be due to either parameter insensitivity or parameter interdependence (or both). Sensitivity tests, like those in Figure 1, showed model $\tilde{\nu}$ to be sensitive to both parameters in the ranges adopted here. In more detail, the sensitivity to $\beta$ and $\nu$ was found to diminish with increasing $\beta$ and $\nu$, respectively. This resulted in the positively skewed shape of the likelihood distributions (Figure 4), with skewness values of 0.4 for $\beta$ and 0.2 for $\nu$ (Table 1). However, the most important reason that $\beta$ and $\nu$ can be identified less well than $k_a$ is their interdependence. Figure 5(b) depicts that the behavioural $\beta$–$\nu$ parameters were correlated with $r = 0.51$: a parameter set with a low (high) $\beta$ that resulted in $R > R_M$ always possessed a low (high) $\nu$ as well. A statistically significant correlation at the 5% level was, surprisingly, also found for the parameters $\beta$–$k_a$ (Figure 5(a)), while behavioural $k_a$–$\nu$ parameters were not correlated (Figure 5(c)). The correlation between behavioural $\beta$ and $\nu$ is understandable as both parameters affect the width of the current jets. For instance, an increase in $\beta$ causes the current jets to become more pronounced, which can be undone by a simultaneous increase in $\nu$. There is, however, one main difference between $\beta$ and $\nu$. The parameter $\beta$ also

![Figure 3](https://iwaponline.com/jh/article-pdf/8/1/37/392771/37.pdf)  
**Figure 3** | Scatter plots of model performance $R$ as a function of (a) $\beta$, (b) $k_a$ and (c) $\nu$ at Egmond. Each panel contains 11,929 dots, of which 1000 (10,929) were computed with the alongshore current model (ANN). The asterisk and square represent the manually determined (Ruessink et al. 2001) and optimum $R$ (Ruessink 2004), respectively.

![Figure 4](https://iwaponline.com/jh/article-pdf/8/1/37/392771/37.pdf)  
**Figure 4** | Scatter plots of likelihood $L$ as a function of (a) $\beta$, (b) $k_a$ and (c) $\nu$ at Egmond. Each panel contains 500 dots, representing the behavioural parameter sets.
changes the location of $\sigma_{\text{max}}$. Apparently, the Egmond measurement layout was too coarse (in a cross-shore sense) to accurately determine the location of $\tilde{\sigma}_{\text{max}}$; had there been more instrumented positions across the inner bar, $\beta$ might have been identified better based on the location of $\tilde{\sigma}_{\text{max}}$, making $\nu$ the main parameter to affect the width of the current jets. It is, therefore, possible that the observed $\beta-\nu$ interdependence is, at least partly, invoked by the cross-shore measurement layout.

Figure 6 presents the parameter-induced 95% prediction uncertainty interval of $\tilde{\sigma}$ at all six positions for a representative 115 h part of the data. Results for the remaining parts of the 500 h are not shown for visibility reasons. As can be seen more clearly in Figure 7, the width of the 95% interval depends linearly on the mean $|\tilde{\sigma}|$ of all behavioural simulations (henceforth denoted $\sigma_b$), from about 0.02–0.06 m/s when $|\tilde{\sigma}_b|$ is near-zero to about 0.2 m/s when $|\tilde{\sigma}_b|$ is near its maximum of about 1.1 m/s. Because currents are typically largest near the bar crest (Figures 1 and 6), the width of uncertainty interval is largest here as well. Intriguingly, there are two situations (near $t = 170$ and 205 h at E2 and E3) with near-zero $|\tilde{\sigma}_b|$ but with 95% uncertainty intervals of 0.1–0.2 m/s. Why precisely these two events possess a large parameter-induced uncertainty is not understood.

Also plotted in Figure 6 are the measured $\tilde{\sigma}$. In many cases, $\tilde{\sigma}_{\text{obs}}$ are bracketed by the parameter-induced prediction uncertainty ranges, showing the suitability of the present model formulations to hindcast the Egmond data. One should, however, keep in mind that the width of the 95% prediction range is based on a subjective $R_M$ choice and that measurement errors have not been accounted for. There are, admittedly, also various instances where the measurements are well outside the parameter-induced prediction uncertainty range (e.g. $t = 150–165$ h at E3 and E4, and $t = 200–215$ h at E5 and E6). During these moments, inaccurate depth profiles (surveys were only carried out once every 2–3 days) or model structural errors, such as the temporal violation of alongshore uniformity or the assumption of the cross-shore constant and time-independent nature of the free parameters, may dominate over parameter-induced and measurement ($\tilde{\sigma}$) errors.

**DISCUSSION**

**Subjectivity**

The GLUE methodology contains a number of subjective choices that may profoundly affect the parameter posterior ranges and the associated width of the central uncertainty band. These include the functional form of the goodness-of-fit function and the critical value of this function to determine whether a specific parameter set is behavioural. Freer et al. (1996), for instance, used $R^p$ (with $p \geq 1$) as the goodness-of-fit measure and found that for an increase in $p$ the parameter posterior ranges narrowed, leading to less and less observations falling inside the predicted uncertainty bounds. The critical value of $R$, $R_M$, obviously affects the number of behavioural parameters and, as such, the width of uncertainty bounds. If $R_M$ is set very high, very few parameter sets will be
behavioural and, consequently, the prediction bands will be unrealistically small. Ultimately, $R_M$ can be chosen so high that only the best simulation is accepted, thereby reducing GLUE to URS without an assessment of parameter uncertainty. In contrast, if $R_M$ is chosen very low, virtually all parameter sets will be behavioural, resulting in uncertainty bounds that are so wide that one would be inclined to dismiss the model altogether.

In the present work, $R_M$ was simply set to the $R$ obtained with an earlier manual calibration. Because manual calibration is subjective, this $R_M$ is subjective as well. Given that $R_M$ is very close to the optimum $R$ (0.8937 and 0.8996, respectively), relatively few parameter sets were accepted (about 4%) and the width of the 95% prediction range was perhaps more narrow than one would a priori have expected. A second cause for the narrowness of

Figure 6  95% prediction uncertainty ranges (light gray) from offshore (E1) to onshore (E6) versus time at Egmond. The solid line within each range is the mean $\bar{v}$ of all behavioural simulations (denoted $\bar{v}_b$) versus time. The dots are the measured $\bar{v}$ and the vertical lines represent the 95% measurement error estimated with the noise model of Feddersen & Guza (2003), see the Discussion section.
the prediction range is the interaction between two of the model’s free parameters ($\beta$ and $\nu$), which cause virtually the same model output for rather different parameter combinations and, consequently, relatively narrow intervals. One should not forget, however, that the present work considered uncertainty in the model output owing to uncertainty in the parameter estimates only. Uncertainty in the data forcing the model, in the alongshore current observations, and in the model itself have not been considered. It is likely that the explicit consideration of all these sources of uncertainty, as for instance by Vrugt et al. (2005) in hydrological modelling, broadens the width of the prediction ranges considerably with respect to the present GLUE-based ranges.

The dependence of the results presented here on $R_M$ is illustrated by computing the 95% central confidence interval for each of the three parameters as a function of $R_M$. The 95% confidence interval based on $R_M = 0.8937$ are $[0.04…0.10]$ for $\beta$, $[0.020…0.035]$ m for $k_a$ and $[0.58…3.97]$ m$^2$/s for $\nu$. As is obvious from Figure 8, these ranges are rather sensitive to the choice of $R_M$, in particular for $\beta$ and $\nu$. For both these parameters, $R_M = 0.89$ results in confidence intervals that encompass almost the entire prior parameter interval. Despite the arbitrariness and non-objective elements in GLUE, its application has been informative for an insight into the model’s sensitivity for the free parameters, for insight into parameter interdependence and for providing estimates of parameter-induced uncertainty in a field of science usually devoid of uncertainty quantification.

First-order approximations

Although the use of an artificial neural network has greatly reduced the computational burden relative to the original GLUE procedure, the methodology remains computationally demanding when compared to the number of computations required when using a first-order approximation (FOA) to estimate posterior distribution functions. FOA approximates the posterior distribution as a multi-variate normal distribution centred at the optimum parameter set $\theta_{\text{opt}}$ and a spread determined by the covariance matrix at $\theta_{\text{opt}}$. In contrast to GLUE, FOA does not require knowledge of the prior parameter distributions. Although there are
various examples in the literature that show multi-variate normal distributions to be poor approximations of actual posterior distributions (e.g. Kuczera & Parent 1998; Vrugt & Bouten 2002), there are also numerous examples where FOA results in the same distributions as a Monte Carlo approach (see Tyagi & Haan (2001) for an overview).

Assumptions in the FOA approach are (Melching 1995) linear model performance, at least in the vicinity of \(u_{\text{opt}}\) and small coefficients of variations (typically less than 0.2 (Tyagi & Haan 2001)) of the most uncertain parameters. In the present work, the model performance in the vicinity of \(u_{\text{opt}}\) is non-linear, as displayed by the positive skewness values (Table 1) and the coefficient of variations of the posterior distributions estimated by GLUE are around 0.2.

To compare the GLUE and the first-order approximations of the posterior distributions for the alongshore current model, the multi-variate normal distributions were computed by using the local-search, gradient-based Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) to find \(\theta_{\text{opt}}\) and the Jacobian matrix \(J_m\) at \(\theta_{\text{opt}}\). This matrix can be transformed into the covariance matrix \(C\) by

\[
C = \left(J_m^T J_m\right)^{-1} F(N_t - N_p)
\]

where the superscript T is the matrix transpose operator, \(F\) is the residual quadratic error at \(\theta_{\text{opt}}\) and \(N_t\) and \(N_p\) are the number of temporal observations and parameters, respectively.

Starting from an arbitrary parameter set in feasible space, the Levenberg–Marquardt algorithm found the known (Ruessink 2004) global optimum (\(\beta = 0.055\), \(k_a = 0.026\) m and \(v = 1.50\) m\(^2\)/s), with \(F = 55.66\) m\(^2\)/s\(^2\). The FOA-based standard deviation of the marginal posterior distributions, computed as the square root of the diagonal elements of \(C\), was 0.0042 for \(\beta\), 0.0025 m for \(k_a\) and 0.5251 m\(^2\)/s for \(v\), about half the standard deviations suggested by GLUE using \(R_M = 0.8937\) (Table 1). The smaller standard deviations cause the FOA marginal posterior distributions to be substantially less wide and, accordingly, more peaked than the GLUE based distributions (Figure 9). Two-sided Kolmogorov–Smirnov tests (e.g. Blalock 1981) showed the GLUE- and FOA-based marginal distributions to indeed be significantly different at the 99% confidence level. Also note that the FOA marginal distribution for \(v\) extends into non-feasible space (\(v < 0\), Figure 9(c)). Interestingly, FOA also suggested the interdependence of \(\beta\) and \(v\), of \(\beta\) and \(k_a\), and of \(\beta\) and \(k_a\), computed using the off-diagonal elements of \(C\), to be considerably smaller than those based on GLUE. For \(\beta\) and \(v\), for instance, FOA suggested \(r = 0.24\), as opposed to \(r = 0.51\) suggested by GLUE (Figure 9(b)).

The differences between the GLUE and FOA results may, to some extent, be due to the subjective choice for \(R_M\). Obviously, if \(R_M\) had been chosen more strict, the GLUE-based marginal posterior distributions would have been less wide and more peaked, similar to the FOA-based distributions. Additional computations with a higher \(R_M\) show that indeed the skewness values decrease with an \(R_M\) increase, that is, the GLUE-based posterior distributions approach a more normal shape, albeit that for \(\beta\) the distribution remains positively skewed up to
about $R_M = 0.897$. However, $R_M$ does not appear to influence the correlation coefficient between $\beta$ and $\kappa$, which remains around 0.5 for $R_M$ up to 0.898. For even larger $R_M$ the number of acceptable simulation sets becomes too small to reliably estimate characteristics of the marginal posterior distributions. On the whole it is safe to say that, for the present model, the approach to quantify parameter uncertainty through a multi-normal approximation may not be valid.

**Measurement errors**

The errors associated with $v_{\text{obs}}$ are not well understood and no tests were carried out during the Egmond measurement campaign to assess the error magnitudes. A rough estimate of the 95% measurement uncertainty interval can, however, be obtained using the noise model proposed by Feddersen & Guza (2003). In this model, the current meter noise is assumed to be composed of two Gaussian zero-mean variables representing offset errors (velocity measured with no fluid motion) and gain errors. The standard deviation of the offset error is given by deviation $\sigma_{\text{off}}$ and that of the gain error by $\sigma_{\beta} = \kappa \delta_{\text{obs}}$, where $\kappa$ is a fractional gain error. The gain error is thus assumed to be linear in $\delta_{\text{obs}}$. The total current meter noise, the sum of both errors, is thus a Gaussian zero-mean variable with standard deviation

$$\sigma_t = \left( \sigma_{\text{off}}^2 + (\kappa \delta_{\text{obs}})^2 \right)^{1/2}.$$  \hfill (4)

The 95% interval is measured as $\delta_{\text{obs}} \pm 1.96 \sigma_t$. It is unknown whether this model is applicable to the Egmond sensors; should it be, values for $\sigma_{\text{off}}$ and $\kappa$ can only be guessed. Taking the values obtained for similar measurements at Duck, North Carolina ($\sigma_{\text{off}} = 0.03$ m/s and $\kappa = 0.04$) by Feddersen & Guza (2003) yields 95% measurement errors of about 0.1 m/s when $v_{\text{obs}}$ is 0, increasing to about 0.2 m/s when $v_{\text{obs}}$ is about 1.1 m/s (Figure 6). On the whole, the estimated 95% measurement errors are thus about the same as the estimated 95% parameter-induced errors (based on $R_M = 0.8937$) when $v_{\text{obs}}$ is large and about twice as large when $v_{\text{obs}}$ is near zero.

**CONCLUSIONS**

With the application of the Bayesian Monte Carlo method GLUE to the process-based modelling of alongshore currents at Egmond aan Zee it has been shown that a relatively wide range of parameter sets results in acceptable simulations. The primary cause for these wide ranges is parameter interdependence, in particular for the roller slope and the lateral mixing coefficient, which both affect the width of the current jets. The 95% uncertainty prediction interval increases with the magnitude of the alongshore current, from about 0.02–0.06 m/s when the current magnitude is near zero to about 0.2 m/s when it is near its maximum of about 1.1 m/s. These widths are approximately equal to a rough estimate of the errors in the data. In many cases the uncertainty interval brackets the observations, although there are also various instances
where this is not the case and apparently model structural errors dominate over parameter-induced errors. The GLUE-based marginal posterior parameter distribution are non-linear, reflected by non-zero skewness values. This, in combination with the observed parameter inter-dependence, violate the use of a first-order approximation to obtain posterior parameter distributions.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Organisation for Scientific Research under project 864.04.007.

REFERENCES


