Parameter stability and consistency in an alongshore-current model determined with Markov chain Monte Carlo

B. G. Ruessink

ABSTRACT

When a numerical model is to be used as a practical tool, its parameters should preferably be stable and consistent, that is, possess a small uncertainty and be time-invariant. Using data and predictions of alongshore mean currents flowing on a beach as a case study, this paper illustrates how parameter stability and consistency can be assessed using Markov chain Monte Carlo. Within a single calibration run, Markov chain Monte Carlo estimates the parameter posterior probability density function, its mode being the best-fit parameter set. Parameter stability is investigated by stepwise adding new data to a calibration run, while consistency is examined by calibrating the model on different datasets of equal length. The results for the present case study indicate that various tidal cycles with strong (say, >0.5 m/s) currents are required to obtain stable parameter estimates, and that the best-fit model parameters and the underlying posterior distribution are strongly time-varying. This inconsistent parameter behavior may reflect unresolved variability of the processes represented by the parameters, or may represent compensational behavior for temporal violations in specific model assumptions.

Key words | Markov chain Monte Carlo, model calibration, uncertainty

B. G. Ruessink
Department of Physical Geography,
Faculty of Geosciences,
Institute for Marine and Atmospheric Research Utrecht,
Utrecht University,
PO Box 80.115,
3508 TC Utrecht,
Netherlands
Tel.: +31 30 253 2405
Fax: +31 30 253 1145
E-mail: *g.ruessink@geo.uu.nl*

NOTATION

The following symbols are used in this paper:

CV	coefficient of variation
L	number of model simulations (iterations)
M	model–data misfit
MCMC	Markov chain Monte Carlo
N	number of observations
RF	potential scale reduction factor
i, j, k	integers
k_a	apparent roughness
$2n$	chain length
$p(\theta\|y)$	posterior probability density function
q	number of complexes
s	number of parameter sets in initial population
t	time
$\bar{v}$	alongshore current velocity
x	cross-shore distance
y	data
$\hat{y}$	model predictions
α	acceptance probability
β	roller slope
γ	kurtosis parameter
θ	parameter set
$\hat{\theta}$	best-fit parameter set
μ	mean value
ν	eddy viscosity
σ	standard deviation

doi: 10.2166/hydro.2008.016

INTRODUCTION

Inverse problems are ubiquitous in coastal civil engineering. Datasets are used to condition the value of parameters in mathematical models of the system of interest. For example,

Battjes & Stive (1985) used wave height measurements obtained along various cross-shore coastal profiles to infer values of a breaker index inherent to their formulation of the wave energy balance. Parameters are inherent to our mathematical models, as they represent unresolved model complexity or nondescribed spatial–temporal process variability. Because parameters rarely represent measurable attributes of the system under study, their values have to be obtained inversely (also referred to as "model calibration") using observed input/output relationships in either a trial-and-error ("manual") or automated approach. A calibrated model can then be used to simulate the behavior of the system of interest outside the calibration period or at other locations, provided that the model input is not drastically different from that in the calibration dataset.

In its simplest form, model calibration aims to find a unique and feasible parameter set that minimizes some measure of misfit between the data and the model predictions. However, the search for a single "best-fit" parameter set does not treat parameter uncertainty. Parameters, and hence model predictions, are inherently uncertain because of issues related to, for instance, model conceptualizations, measurement errors in the inputs and errors in the data to calibrate the model. Consequently, model calibration should result in an estimate of the most likely parameter set and of the underlying parameter posterior distribution. When the relation between the data and the parameters is nonlinear, one has to resort to iterative samplers, often Markov chain Monte Carlo methods. Instead of determining a best-fit parameter set $\hat{\theta}$ only, such methods generate a series of parameter sets that describe the posterior distribution $p(\theta|y)$. This distribution provides the probability p of the parameters θ conditioned on the data y and given the imposed model structure. The mode of $p(\theta|y)$ is $\hat{\theta}$.

When a model is to be used as a practical tool, parameters should preferably have a small uncertainty and $p(\theta|y)$ should be time-invariant. Vrugt *et al.* (2006) referred to these prerequisites as parameter stability and consistency (or stationarity). An essential question with respect to parameter stability is the amount of data that is actually required to obtain stable estimates. Clearly, a very limited dataset will result in unstable estimates but, conversely, it is not necessarily true that adding more data to the calibration procedure increases parameter stability. Likely, the information content of the data is more relevant than the sheer abundance of the data; adding more data may mean "more of the same", increasing the computational effort unnecessarily. Consistency implies that different calibration sets result in the same $\hat{\theta}$ and $p(\theta|y)$. Inconsistent values may highlight flaws in the model structure or in the input data that have different relevance between the various calibration sets. The purpose of the present paper is to illustrate how parameter stability and consistency can be addressed with Markov chain Monte Carlo using data and predictions of alongshore mean currents in the nearshore as a case study.

MARKOV CHAIN MONTE CARLO

In Markov chain Monte Carlo (MCMC) the form of the posterior probability distribution, $p(\theta|y)$, also referred to as the target distribution, is approximated by sampling iteratively from this distribution rather than by computing $p(\theta|y)$ explicitly as done in first-order approximations. To this end, parameter sets are drawn sequentially as a Markov process to create one (or more) chain(s) of parameter values whose stationary distribution converges to $p(\theta|y)$ after a sufficient number of draws. The Markov property implies that each chain is a sequence of parameter sets, $\{\theta_1, \theta_2, \theta_3, \ldots\}$, such that at each draw (or iteration) i the next parameter set θ_{i+1} is sampled from a distribution that depends on the present parameter set θ_i only. Here, the MCMC technique introduced by Vrugt *et al.* (2003) is adopted. This technique, known as the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm, is a multi-chain algorithm based on the Metropolis sampler (Metropolis *et al.* 1953) and can be considered as a mixture of the well-known global search Shuffled Complex Evolution algorithm of Duan *et al.* (1993) and the simulated annealing technique of Metropolis *et al.* (1953). The reader is referred to Gelman & Rubin (1996), Gilks *et al.* (1996) and Gelman *et al.* (2004) for extensive introductions to Markov chain Monte Carlo methods.

SCEM-UA begins with an initial population of s parameter sets θ distributed randomly in feasible parameter space, for each parameter defined by a realistic lower and

upper bound. Here, the multi-variate uniform distribution reflects that no *a priori* knowledge about the parameters exists. Clearly, if such knowledge is available from earlier model experience or physical constraints, another prior distribution, such as an overdispersed multi-variate normal distribution, may be more appropriate. The posterior density $p(\theta|y)$ of each parameter set is computed using the Bayesian inference scheme of Box & Tiao (1973):

$$p(\theta|y) \propto [M(\theta)]^{-\frac{1}{2}N(1+\gamma)} \tag{1}$$

where

$$M(\theta) = \sum_{j=1}^{N} \left| y_j - \hat{y}_j \right|^{2/(1+\gamma)} \tag{2}$$

Here, N is the total number of observations (in time and space) and $\gamma \in \langle -1, 1]$ determines the shape of the distribution of the differences between the observations y and model predictions $\hat{y}$. For $\gamma = 0$, $M(\theta)$ is the sum of squared differences, with the distribution of these differences (or errors) being normal. When $\gamma = 1$, the distribution is the double exponential, while when γ tends to -1, the distribution becomes rectangular (Box & Tiao 1973). By varying γ it can be investigated how sensitive $p(\theta|y)$ is to the distribution of the errors.

After each set in the initial population has been assigned its posterior density, the initial population is partitioned into q complexes. In each complex, a parallel chain is launched from the point with the highest posterior density. This launch implies the selection of a new candidate point using the Metropolis sampler. Using a random number between 0 and 1 a new parameter set, θ_{i+1}, is selected from a multi-variate normal distribution centered at the point with the highest posterior density and with the covariance of the complex. The posterior density of θ_{i+1} is computed using Equation (1). The Metropolis-annealing (Metropolis *et al.* 1953) criterion is used to decide whether θ_{i+1} should be added to the current chain. The acceptance probability α of this criterion is

$$\alpha(\theta_i, \theta_{i+1}) = \min\left[1, \frac{p(\theta_{i+1}|y)}{p(\theta_i|y)}\right] \tag{3}$$

Thus, a new parameter set that outperforms the best parameter set so far is always accepted ($\alpha = 1$); should the new parameter set perform worse than the best set, there is still a chance $p(\theta_{i+1}|y)/p(\theta_i|y)$ that the new set is accepted. In other words, Equation (3) always accepts better points and, occasionally, a worse point. An accepted candidate point (or the best existing point, in case the candidate point was not accepted) is added to the complex by randomly replacing a member in this complex using a trapezoidal weight distribution. After each complex has evolved for a user-specified number of iterations L, the complexes are combined, shuffled and broken up into new complexes, after which the procedure repeats itself. This shuffling ensures a rapid sharing of information between the complexes. Should, for instance, one particular complex have found a very promising part of parameter space (i.e. that part with a very high posterior distribution), the shuffling ensures that other complexes know of this part of parameter space quickly.

An important aspect of MCMC is to decide when the samples in the chains are representative of the posterior density distribution of the model's free parameters: in other words, when it is safe to conclude convergence. An extensive review of convergence diagnostics can be found in Cowles & Carlin (1996). Here, the "potential scale reduction factor", referred to as *RF* in the following, introduced by Gelman & Rubin (1992) and later on modified by Brooks & Gelman (1998), is used as the convergence diagnostic. Essentially, *RF* is the square root of the ratio between the variance of the values of the kth parameter in all chains and the mean of the variances of the kth parameter in each chain. When the various chains occupy different portions of parameter space, the former, pooled variance will exceed the mean of the individual chain variances and, accordingly, *RF* is well above 1. A large *RF* thus signifies that drawing more samples will improve the inference about $p(\theta|y)$. When the parallel chains are overlapping, both variances are about the same and *RF* approaches 1. In practice, it is very hard to achieve $RF = 1$ and convergence is typically declared once for each parameter RF is below 1.2 (Gelman & Rubin 1996; Brooks & Gelman 1998; Vrugt *et al.* 2003). Note that the computation of *RF* involves the last n iterations (where $2n$ is the chain length) only to avoid the effect of the initial population on *RF*.

Finally, the draws in the converged part of the Markov chains are used to summarize various statistics of the

posterior distribution of the model parameters. The draw with the largest density is the mode of the parameter distribution and reflects the best-fit parameter set. Other statistics include the mean and standard deviation of the kth parameter (μ_k and σ_k, respectively) and their ratio, the coefficient of variation ($CV_k = \sigma_k/\mu_k$), where μ_k and σ_k are given by

$$\mu_k = \sum p(\theta_k|y)\theta_k \tag{4}$$

and

$$\sigma_k = \sqrt{\sum(\theta_k - \mu_k)^2 p(\theta_k|y)} \tag{5}$$

The 2.5% and 97.5% percentiles of the posterior distribution of each parameter provide the central 95% posterior interval. The linear interdependence of the parameters is estimated in a straightforward manner from the parameter covariance matrix. In the present work, inferences about $p(\theta|y)$ were based on 250 unique parameter sets. If, for the moment, it is assumed that a parameter has an approximately normal posterior distribution, then the 2.5% and 97.5% percentiles are estimated to an accuracy of $\sqrt{[0.025(1 - 0.025)/250]} = 0.01$, which is good enough for the present work.

CASE STUDY

Model

In the absence of alongshore nonuniformities in bathymetry, the cross-shore (x) structure of the depth-averaged alongshore current $\bar{v}$ in the nearshore can be predicted accurately from the one-dimensional (1D), time- and depth-averaged alongshore momentum balance between wave, wind and tidal forcing, and bottom stress and lateral mixing (e.g. Longuet-Higgins 1970; Thornton & Guza 1986; Reniers & Battjes 1997; Ruessink *et al.* 2001). On (near) planar beaches the alongshore current is predicted to peak in the mid-surfzone and to reduce to near-zero values on the beach and seaward of the surf zone. On barred beaches and with the presence of rollers in the wave forcing, broadly distributed (in the cross-shore) current jets are predicted, with maximum values near the bar crest and smaller but nonzero values in the bar troughs (e.g. Figure 1).

The 1D alongshore current model used throughout this study was formulated in Ruessink *et al.* (2001) and is therefore not reiterated here. The model contains three free parameters (apparent bed roughness k_a, wave-front slope

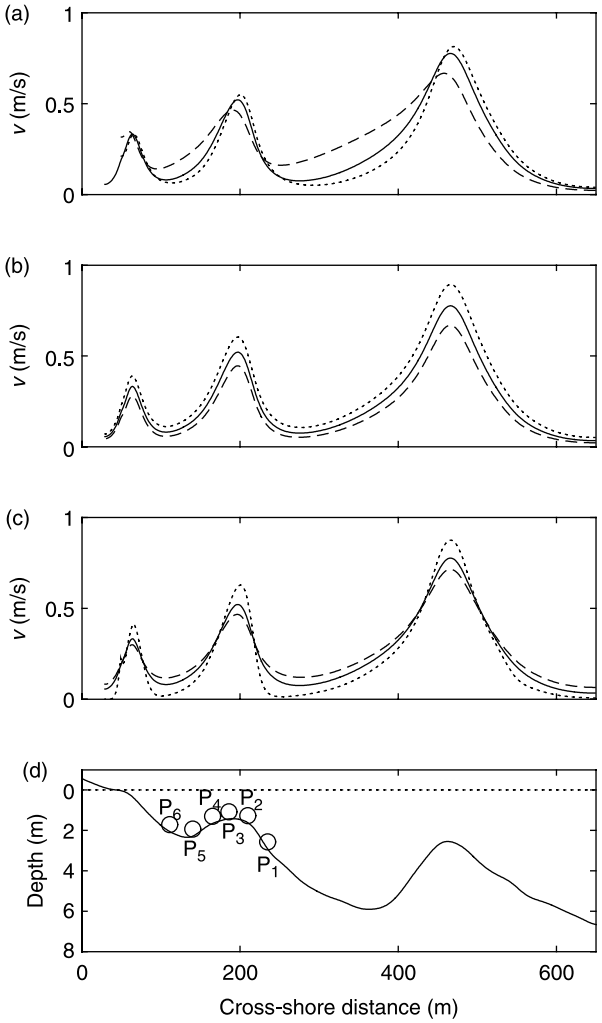

Figure 1 | Alongshore current $\bar{v}$ versus cross-shore distance x showing the effect of changing the wave-front slope β, apparent bed roughness k_a and eddy viscosity ν. In all panels, the offshore conditions are: offshore root-mean-square wave height = 1.5 m, wave period = 8 s and wave angle relative to shore normal = 30°, with no wind and tidal forcing. The solid curve in panels (a)–(c) is the standard run ($\beta = 0.1$, $k_a = 0.1$ m and $\nu = 0.5$ m^2/s). Other curves correspond to changes in one parameter with the others held constant: (a) $\beta = 0.15$ (dotted line) and $\beta = 0.05$ (dashed line), (b) $k_a = 0.05$ m (dotted line) and $k_a = 0.2$ m (dashed line), (c) $\nu = 0$ m^2/s (dotted line) and $\nu = 1$ m^2/s (dashed line). (d) The depth profile was measured at Egmond on 15 October 1998. The shallow parts of the profile near cross-shore distances of 200 and 460 m are morphological features known as nearshore sandbars. The circles in (d) are the instrumented positions used in the present work.

β and eddy viscosity ν) that control the magnitude, location and width of the current jets (Figure 1). The feasible parameter space was specified by lower and upper bounds for each parameter, [0.02, 0.15] for β, [0.001, 0.075] m for k_a and [0, 6] m^2/s for ν. These values were based on physical reasoning and earlier experience with the model.

Observations

The data used for calibration and validation were collected during the Coast3D experiment near Egmond aan Zee, The Netherlands, at six positions on a cross-shore transect across the inner of the two subtidal sandbars (Figure 1 (d)). Details on data acquisition and processing are given by Ruessink *et al.* (2001). The Egmond data comprises a total of 500 h of data, during which the sandbar morphology was sufficiently alongshore-uniform to warrant the use of the 1D model. During these 500 h, the absolute maximum alongshore velocity reached values of about 1.2 m/s, typically near or just shoreward of the bar crest (P2–P4 in Figure 1(d)), see Figure 2(a). Missing data in Figure 2(a) (e.g. P4, $t = 196$–246 h) reflect situations when the bidirectional current meter used to measure $\bar{v}$ was estimated to be within 0.2 m of the bed. Time series of the offshore wave and waterlevel forcing used as input into the model are provided in Figures 2(b)–(e). Other inputs include time series of offshore wind speed and direction, and the tide-induced large-scale alongshore surface slope, see Figure 4 in Ruessink *et al.* (2001).

Probability distributions

The implemented sampler was applied to the first 80 h of data ($t = 1$–80 h), when all six sensors were operational, to illustrate its overall performance. The evolution of the convergence diagnostic RF is provided in Figure 3, based on $\gamma = 0$, $s = 100$, $q = 5$ and $L = 4$. Because of the random initialization of the population, each chain initially occupies different parts of the feasible parameter space, resulting in relatively large RF values. From about 4000 simulations, the RF of each parameter remains below 1.2, suggesting that the five parallel chains have converged to a stationary posterior distribution.

Table 1 presents various characteristics of the parameter posterior distribution, based on the observations at $t = 1$–80 h. The best-fit parameter values for this calibration set amounts to $\beta = 0.047$, $k_a = 0.028$ m and $\nu = 0.76\ m^2/s$. The relative low values for the standard deviation and the

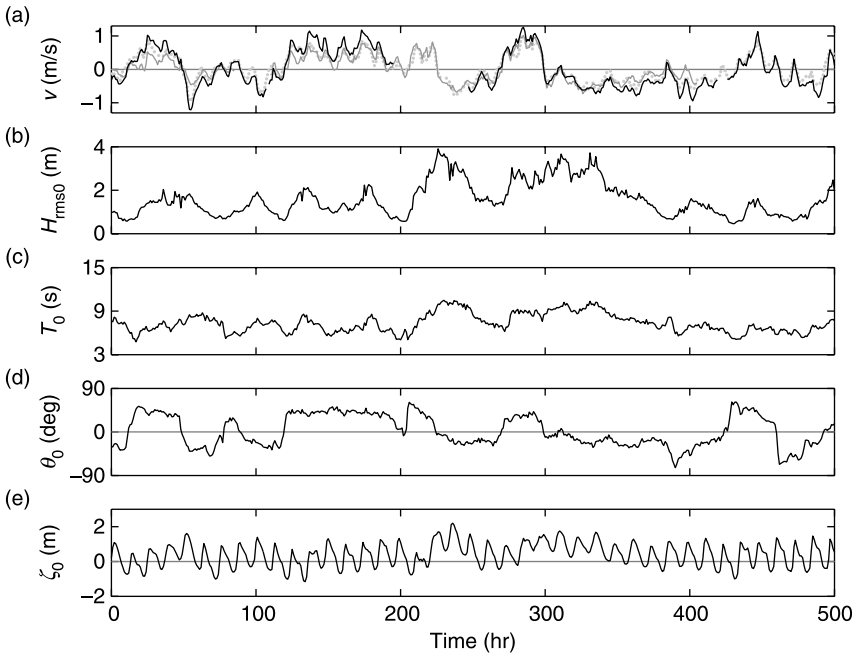

Figure 2 | (a) Alongshore current at P1 (dots), P4 (solid line) and P5 (gray solid line), (b) offshore root-mean-square wave height H_{rms0}, (c) offshore wave period T_0, (d) offshore wave angle θ_0$, and (e) offshore water level ζ_0 versus time at Egmond. Time = 1 h corresponds to 16 October 1998, 00:00 MET.

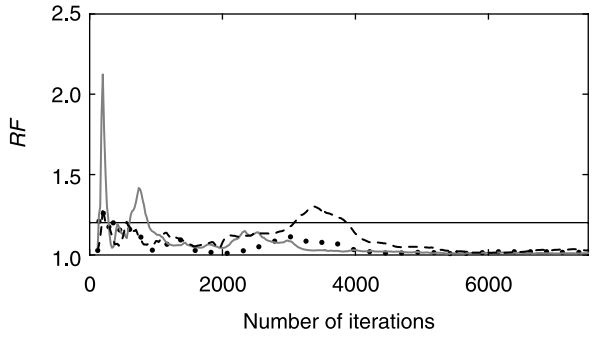

Figure 3 | The Brooks–Gelman potential scale reduction factor *RF* for β (dots), k_a (solid line) and ν (dashed line) versus the number of model simulations (iterations). The horizontal line indicates $RF = 1.2$.

Table 1 | Posterior best-fit, mean, standard deviation, coefficient of variation (%) and correlation coefficients between the parameters in the alongshore current model, based on $t = 1$–80 h

Parameter	Best	μ	σ	CV	β	k_a	ν
β	0.047	0.048	0.0038	8.0	1.00	−0.15	0.67
k_a	0.028	0.029	0.0016	5.7		1.00	−0.11
ν	0.76	0.75	0.17	22.2			1.00

coefficient of variation for β and k_a demonstrate that both parameters occupy a relatively narrow part of the feasible parameter space, see also Figures 4(a, b); in other words, the observations were sufficiently informative to enable to identify the values of both parameters well. The larger standard deviation and coefficient of variation for ν illustrate that the mixing coefficient could be defined less well from the observations. Figure 4 further shows that the shape of the marginal posterior parameter distributions was approximately normal, with a small tendency to a positively skewed distribution for β and ν.

Interestingly, ν was found to be correlated significantly with β (Table 1, see also Figure 5), possibly resulting in the poorer identifiability of ν. The β–ν interdependence is not surprising because both parameters affect the width of the current jet across the inner bar, where an increase in width with ν is counteracted by a decrease in width with β. Because β also determines the location of the maximum velocity in the jet, it is possible that the cross-shore spacing between consecutive instruments was too coarse for the calibration to determine β primarily on this maximum velocity location and ν on the jet width. Thus, the β–ν interdependence may be an artifact of the measurement layout rather than a structural model error.

The results in Figures 3–5 are based on the assumption of normality for the differences between observed and predicted $\bar{v}$, that is, $\gamma = 0$. Figure 6 indicates that the calibration results essentially remain unchanged when the error distribution is assumed to be double exponential ($\gamma = 1$); however, when the error distribution tends to rectangular ($\gamma \to -1$) the best-fit values of all three parameters increase somewhat, as does their uncertainty. There is no reason to suppose here that the errors follow the rectangular distribution. The differences are likely to originate from a number of sources (errors in the flow measurements, offshore forcing, bathymetric surveys, model

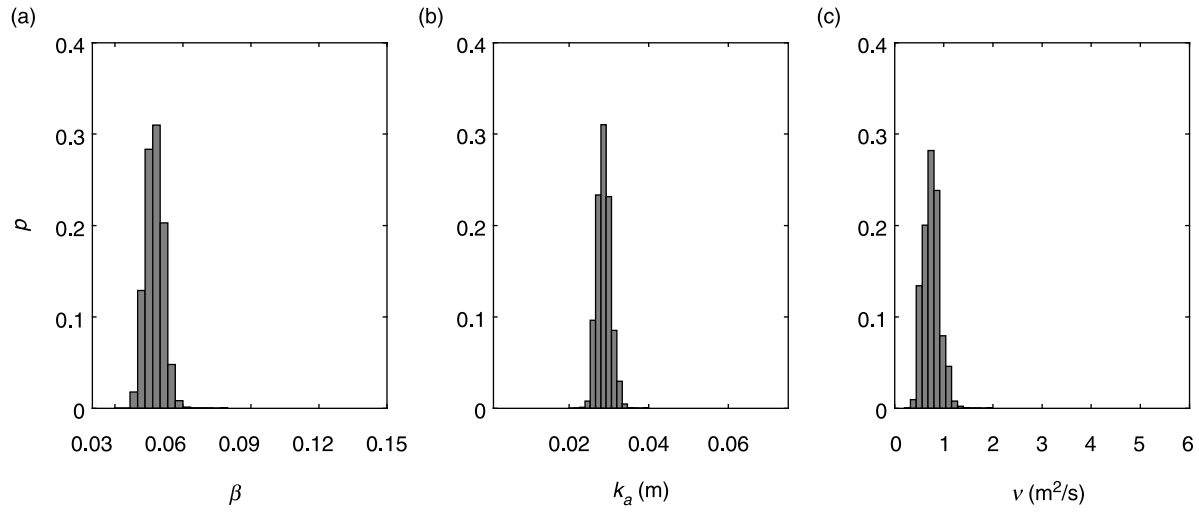

Figure 4 | Histograms of the marginal posterior density distribution of (a) β, (b) k_a and (c) ν, based on $t = 1$–80 h. In each plot the x axis spans the entire search space.

structure, and so forth), none of which is clearly dominant. This will tend to produce a central limit effect (Box & Tiao 1973), implying the distribution of the errors to approximately normal. Figure 6 shows that slight deviations from normality (say, $-0.5 < \gamma < 0.5$) do not change the calibration results relative to the assumption of normality in any noteworthy fashion.

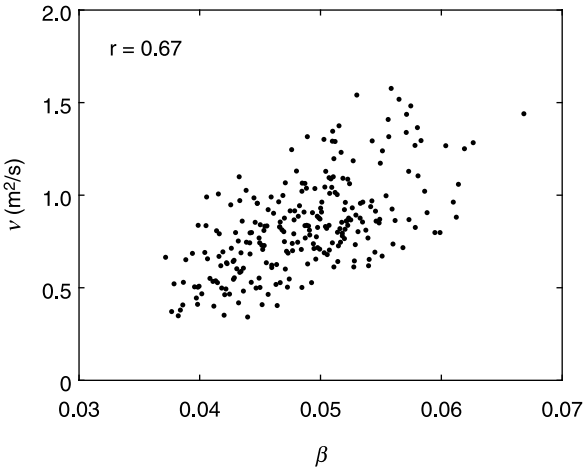

Figure 5 | Scatterplot of the 250 β–ν samples of the converged part of the Markov chains based on $t = 1$–80 h, showing the positive dependence of ν on β.

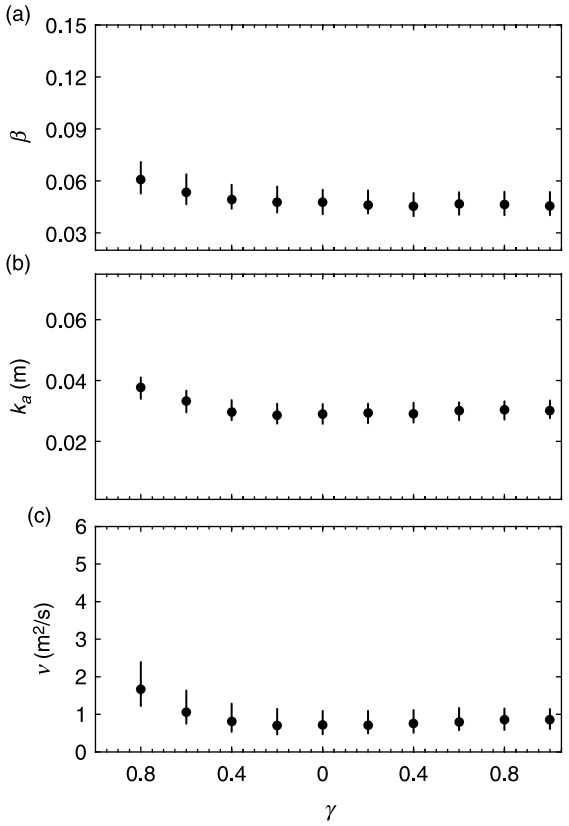

Figure 6 | Best-fit value (symbol) and central 95% confidence interval (vertical line) of (a) β, (b) k_a and (c) ν versus γ, based on $t = 1$–80 h.

Stability

The same 80 h of data were used to study the effect of the length of calibration set on parameter stability. To this end, 80 optimization runs were performed, the first containing the observations at $t = 1$ h only, the second using the observations at $t = 1$–2 h, and so forth, with the last one equaling the calibration ($t = 1$–80 h) presented in the previous subsection. For each calibration run the best-fit parameter set as well as the central 95% confidence levels were determined.

As can be seen in Figures 7(a–c), the best-fit values for all three parameters vary widely until the large $\bar{v}$ observed

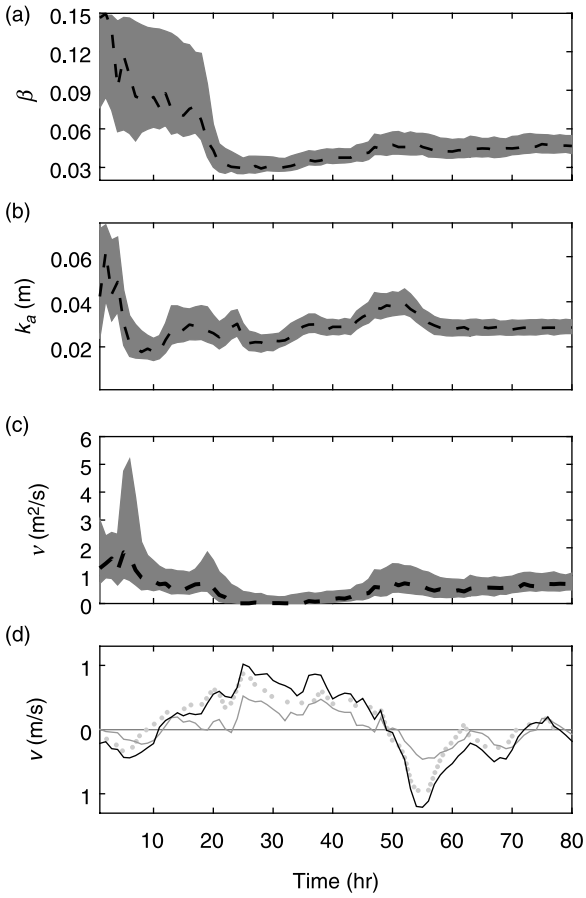

Figure 7 | Evolution of the (dashed line) best-fit parameter set and its central 95% confidence interval (shaded area) using the observations up to time t for (a) β, (b) k_a and (c) ν. For reference, panel (d) shows time series of alongshore current $\bar{v}$ measured at P1 (dots), P4 (solid line) and P5 (gray solid line).

after $t \sim 25$ h are part of the calibration dataset; also, the confidence band encompass a considerable portion of the search space. Apparently, the relatively low velocities during the first 20–25 h do not contain much information to define the parameters, leading to unstable parameter estimates. However, even after $t \sim 25$ h there are still some noteworthy changes in the best-fit parameter values, for instance for ν from $\sim 0\,m^2/s$ at $t \sim 30$ h to $\sim 1\,m^2/s$ at $t \sim 50$ h. Note that, after $t \sim 65$ h, adding more data neither changes the best-fit values nor reduces the uncertainty associated with the parameters. Not surprisingly it appears to be the middle section of the calibration set with the largest alongshore velocities (broadly, $t = 25$–65 h) that has the highest information content. On the whole, Figures 7(a–c) signify that, at least for this particular dataset, various tidal cycles with strong (say, $|\bar{v}| > 0.5$ m/s) currents are required to obtain stable parameter estimates and that the results of an optimization run depend on the amount of data (in time) considered.

Validation

To examine the effect of the length of the calibration set on model performance for unseen data, the model was run for $t = 81$–500 h with the 250 parameter sets determining the $p(\theta|y)$ based on the first 5, 10, 20, 40, 65 and (all) 80 h. In each case, the validation thus resulted in an ensemble of 250 predictions for $t = 81$–500 h, from which the 2.5%, 50% and 97.5% exceedance bias and root-mean-square (rms) errors were computed.

As can be seen in Figure 8(a), the use of a longer calibration dataset initially reduces the bias from (50% value) about 0.03 to 0.02 m/s, but does not reduce it any further for lengths in excess of 20 h. Perhaps surprisingly, the effect on the median rms error appears to be marginal (Figure 8(b)). Apparently, the use of only 10 h of calibration data results in model predictions that are fairly similar to that based on the 80 h calibration dataset. The use of a longer calibration set primarily reduces the dispersion around the median rms error value, in particular the 50–97.5% range (Figure 8(b)). This is due to the stabilization of the parameter posterior density function for calibration lengths in excess of 40–65 h (Figure 7). Consequently, the 250 parameter sets result in a narrower $\bar{v}$ range during the evaluation period.

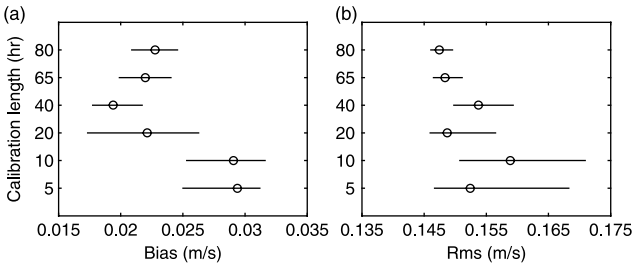

Figure 8 | Effect of the length of the calibration data set on the (a) bias and (b) root-mean-square error of model predictions for the validation period ($t = 81$–500 h). The error statistics are calculated for all six positions simultaneously. The circles are the median (50%) values, while the horizontal lines are the central 95% range.

Consistency

The nonzero bias for model predictions in the validation period already highlights that the $p(\theta|y)$ would have been different if the model had been calibrated on data in the validation period. To examine the consistency (or stationarity) of $p(\theta|y)$ further, a split sample test was conducted by dividing the entire 500 h of data into 43 consecutive blocks of 80 h data blocks with a step size of 10 h. For each block (thus $t = 1$–80, 11–91, 21–101 up to and including 421–500 h) the parameter posterior probability density function was determined with the SCEM-UA algorithm. A block length of 80 h was chosen because the results in Figure 7 signify that 80 h is sufficient to obtain stable parameter estimates.

The temporal evolution of the central 95% confidence interval and of the best-fit value for each of the model parameters is illustrated in Figure 9. While the apparent roughness k_a shows little temporal variability in both its optimum value as well in the width of the central 95% confidence interval, both other parameters occupy remarkably different portions of the feasible parameter space with time. This implies that there is considerable temporal variability and uncertainty in the two parameters that determine the maximum velocity and width of the current jet across the inner bar.

Interestingly, the temporal variability in optimum β and ν, as well as the associated uncertainty, appears quite systematic. The increased uncertainty for blocks encompassing data from $t \sim 175$–325 h may be due to the reduced number of observations, see Figure 9(d). Observations were

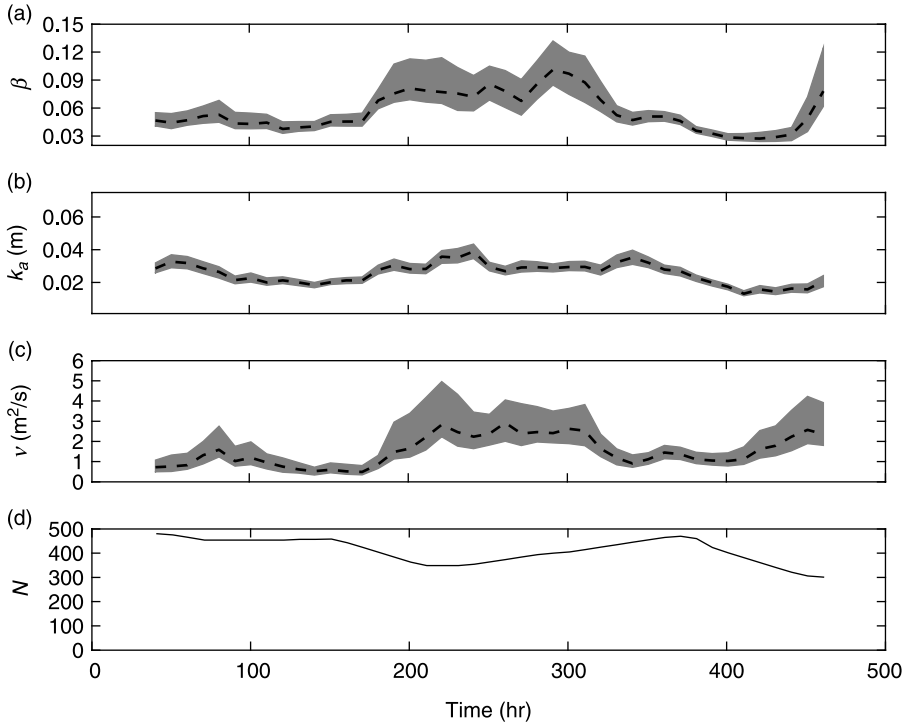

Figure 9 | Evolution of the (dashed line) best-fit parameter set and its central 95% confidence interval (shaded area) using the observations up to time t for (a) β, (b) k_a and (c) ν. Results are values for windows of 80 h. Values for each window are plotted at the middle (in time) of the windows. Panel (d) shows the number of observations in each window.

missing from particularly P3 and P4 (because of instrument burial), where the peak of the current jet is expected to be located (Figure 1). The absence of observations that are most informative for β and ν may have further increased the width of the 95% confidence interval. The increase in ν during the most energetic conditions of the campaign (Figure 2(b)) is not inconsistent with literature suggestions that ν is not a time-invariant and cross-shore constant parameter (as assumed here) but varies as a function of local wave height and roller dissipation (e.g. Reniers & Battjes 1997), both of which are larger under storm conditions. Figures 9(a) and (c) suggest that β and ν best-fit values calibrated for relatively energetic conditions ($t \sim 175$–325 h) result in sub-optimal values for low and intermediate energetic conditions. Finally, we cannot rule out the possibility that the obtained variability in best-fit values represents parameter compensations of temporal violations in specific model assumptions, such as the neglect of alongshore variability in morphology and waves. If these violations indeed have a time-varying magnitude, then parameter adjustments will try to compensate for the associated temporal variability in model-data misfit, resulting in time-varying best-fit parameter values.

A warning that arises from Figure 9 is that one has to be careful to compare best-fit values obtained using datasets from different field sites. Ruessink *et al.* (2001), for instance, noted how best-fit model values at Egmond differed from values obtained with a similar dataset at Duck, NC. They ascribed these differences to potential differences in the directional characteristics of the offshore wave climate. In hindsight, the different length of the datasets, the different offshore wave heights and, as such, the different information content of the dataset may have been more likely causes for the intersite differences in best-fit values.

SUMMARY

Using data and predictions of alongshore mean currents in the nearshore as a case study we have, following Vrugt *et al.* (2006), illustrated how parameter stability and consistency can be assessed using Markov chain Monte Carlo. When a model is to be used as a practical tool, then stable

(that is, small uncertainty) and consistent (that is, time-invariant) parameters are a prerequisite. Parameter stability was investigated by stepwise adding new data to a calibration run, while consistency was examined by calibrating the model on different, equally long datasets. The results for the present alongshore current model and Egmond dataset suggest that several tidal cycles (40–60 h) are required to obtain stable parameter estimates and that the characteristics of the posterior probability density function can change dramatically when different datasets of equal length are used for calibration. The temporal parameter variability may reflect unresolved variability of the processes represented by the parameters, or may represent compensational behavior for temporal violations in specific model assumptions.

ACKNOWLEDGEMENTS

This work was sponsored by the Netherlands Organization for Scientific Research (NWO) under contract 864.04.007 and by the Stichting Nationale Computerfaciliteiten (National Computing Facilities Foundation, NCF) for the use of supercomputer facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research, NWO). Special thanks to Jasper Vrugt (Los Alamos National Laboratory) for providing the SCEM-UA code and to Willem Bouten (Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam) for stimulating discussions on model calibration.

REFERENCES

Battjes, J. A. & Stive, M. J. F. 1985 Calibration and verification of a dissipation model for random breaking waves. *J. Geophys. Res.* **90**, 9159–9167.

Box, G. E. P. & Tiao, G. C. 1973 *Bayesian Inference in Statistical Analyses*. Addison-Wesley-Longman, Reading, MA.

Brooks, S. P. & Gelman, A. 1998 General methods for monitoring convergence of iterative simulations. *J. Comput. Graphical Stat.* **7**, 434–455.

Cowles, M. K. & Carlin, B. P. 1996 Markov Chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Assoc.* **91**, 883–904.

Duan, Q., Gupta, V. K. & Sorooshian, S. 1993 A shuffled complex evolution approach for effective and efficient optimization. *J. Optimiz. Theor. Appl.* **76**, 501–521.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 2004 *Bayesian Data Analysis*, 2nd edn. Chapman & Hall/CRC, New York/Boca Raton, FL.

Gelman, A. & Rubin, D. B. 1992 Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457–511.

Gelman, A. & Rubin, D. B. 1996 Markov Chain Monte Carlo methods in biostatistics. *Statist. Methods Med. Res.* **5**, 339–355.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. 1996 *Markov Chain Monte Carlo In Practice*. Chapman & Hall/CRC, New York/Boca Raton, FL.

Longuet-Higgins, M. S. 1970 Longshore currents generated by obliquely incident sea waves. *J. Geophys. Res.* **75**, 6778–6789.

Metropolis, N., Rosenbluth, A. W., Rusenbluth, N. M., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091.

Reniers, A. J. H. M. & Battjes, J. A. 1997 A laboratory study of longshore currents over barred and non-barred beaches. *Coastal Engng.* **30**, 1–22.

Ruessink, B. G., Miles, J. R., Feddersen, F., Guza, R. T. & Elgar, S. 2001 Modeling the alongshore current on barred beaches. *J. Geophys. Res.* **106**, 22451–22463.

Thornton, E. B. & Guza, R. T. 1986 Surf zone longshore currents and random waves: field data and models. *J. Phys. Oceanogr.* **16**, 1165–1178.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003 A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. *Wat. Res. Res.* **39**, 1201.

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T., Bouten, W. & Verstraten, J. M. 2006 Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model. *J. Hydrol.* **39**, 288–307.

First received 18 January 2007; accepted in revised form 4 November 2007