Sensitivity analysis approaches to investigate uncertainty in process-based models, with application to aquaculture

Michael Currie[1], Claire Miller[1], Marian Scott[1], Alan Hills[2]

[1] School of Mathematics & Statistics, University of Glasgow, Scotland
[2] Scottish Environment Protection Agency, Scotland

E-mail for correspondence: `m.currie.1@research.gla.ac.uk`

Abstract: Sensitivity and uncertainty analyses are effective tools for assessing uncertainty around parameter estimation for complex computer models, and hence increase confidence in model predictions. NewDEPOMOD is a particle tracking model used for monitoring the environmental impacts of aquaculture, and will be used as an application to extend sensitivity analysis methods to consider models with a multivariate response.

Keywords: Sensitivity Analysis; Shape Analysis; Aquaculture.

1 Introduction & Background

There remain environmental challenges which can only acccurately be assessed by process-based modelling. An example of this is monitoring the environmental impacts of aquaculture, where the difficulty and cost of collecting data over large areas make modelling the more effective approach. Such modelling approaches are computationally intensive and do not account for uncertainty. Therefore, sensitivity and uncertainty analyses of these models provide approaches to quantify uncertainty in model responses and attribute them to variations in the model parameters. NewDEPOMOD is a development of DEPOMOD (Cromey *et al.* 2002) that was created to estimate and predict the transportation of waste particles from fish farm cages to the seabed. NewDEPOMOD produces a number of scalar outputs as well as a map of the impacted area. Scalar outputs such as the Total Area Impacted, 99th Percentile and Mass balance will be considered as a starting point for the sensitivity analysis, before extending this to consider the shape of the impact as the response.

This paper was published as a part of the proceedings of the 35th International Workshop on Statistical Modelling (IWSM), Bilbao, Spain, 19–24 July 2020.

2 Sensitivity Analysis

A sensitivity analysis can be used to address uncertainties within a model to increase confidence in the predictions (Saltelli *et al.* 2000). Given a process-based model that maps the n inputs, $\mathbf{x} = [x_1, \ldots, x_n]$, to the output, $\mathbf{y}$, using the function, f,

$$\mathbf{y} = f(\mathbf{x}) = f(x_1, \ldots, x_n).$$

A sensitivity analysis considers how variations in the output, $\mathbf{y}$ can be associated with variations in the inputs, $\mathbf{x} = [x_1, \ldots, x_n]$. Saltelli *et al.* (2000) described a typical sensitivity analysis workflow: 1) Determine the questions relating to the model that should be answered and identify the inputs required, 2) Establish suitable ranges of variation for each input, 3) Identify an appropriate design to generate the input matrix, 4) Complete model runs to create the required outputs, and 5) Analyse the effect of each input on the output. Traditionally, a sensitivity analysis is applied to a model with a scalar output, but in this work we extend this to develop a sensitvity analysis approach that uses area and shape as the response which will be illustrated using output from NewDEPOMOD.

2.1 Sensitivity Analysis for Scalar Outputs

The scalar outputs for NewDEPOMOD, mentioned previously, were considered as the outputs of the sensitivity analysis. In collaboration with the Scottish Environment Agency (SEPA), a set of inputs were identified as being of most importance, which included Critical Shear Stress for Erosion and Settling Velocity of Faeces. Suitable ranges for the inputs were established using the literature, where possible, and the experience and knowledge of SEPA in cases where no literature was available. A correlated Latin Hypercube Sampling (LHS) was used to account for the relationships between inputs and capture the sample space effectively. It relies on a restricted pairing procedure (Iman & Conover 1982), where a target correlation matrix, $\mathbf{C}^*$, is identified at the outset. Following this, an initial LHS, $\mathbf{L}$, is calculated with sample correlation, $\mathbf{T}$. A Cholesky Decomposition of $\mathbf{T}$ and using other variance reduction techniques, allows a matrix, $\mathbf{S}$, to be found such that the correlated LHS is given as follows:

$$\mathbf{L}^*_{\mathbf{B}} = \mathbf{L}\mathbf{S}^{\mathrm{T}}$$

where the correlated LHS, $\mathbf{L}^*_{\mathbf{B}}$ has a sample correlation matrix $\mathbf{M}_{\mathbf{B}}$, approximately equal to $\mathbf{C}^*$. The input matrix allowed 10,000 model runs to be completed in order to calculate the scalar summaries and determine the impact of uncertainties in the inputs. Random forests were used as a ranking method as they are able to deal with non-linear relationships, interactions

TABLE 1. Top 3 ranked inputs using Total Area Impacted as the output.

Inputs	Importance Value
Settling Velocity of Faeces	127.24
Critical Shear Stress for Erosion	75.99
Rate of Erosion	50.81

between inputs and also the interpretability of the results. Table 1 shows the 3 highest ranked inputs for the scalar output, Total Area Impacted. The random forest model identified Settling Velocity of Faeces as having the biggest impact on the Total Area Impacted, which was expected as this determines the time particles spend settling from the cages to the seabed.

2.2 Sensitivity Analysis of the Shape of the Impact

As NewDEPOMOD produces a map of the predicted shape and size of the impacted area on the seabed, it was important to extend the traditional sensitivity analysis of scalar outputs and consider the influence of uncertainty in the inputs on the predicted shape and size. A landmark approach (Dryden & Mardia 2016) was used to identify the main shape of the impact (example seen in Figure 1), by considering transects from the farm.

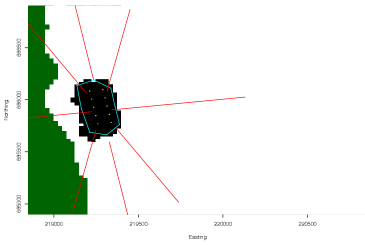

FIGURE 1. Plot illustrating the landmarks calculated to analyse shape.

Landmarks were calculated for each map of the predicted impact using this approach, before using a Generalised Procrustes Analysis (GPA) approach to identify variations in the shapes. GPA is defined as the translation, rescaling and rotation of the shape configurations $(X_1, X_2, \ldots, X_n)$ relative to each other, to minimize a total sum of squares (Dryden & Mardia 2016):

$$G(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} \parallel (\beta_i X_i \Gamma_i + 1_k \gamma_i^{\mathrm{T}}) - \mu \parallel^2$$

with respect to $\beta_i, \Gamma_i, \gamma_i$, for $i = 1, \ldots, n$ and μ, subject to an overall size constraint that is chosen. $\beta_i > 0$ refers to a scale parameter, Γ_i is a rotation matrix, γ_i is a location vector and μ is the population mean shape. A Principal Components Analysis was then applied to the landmarks data to identify the areas of variability in the shapes.

TABLE 2. Table of the Principal Component percentages.

Principal Component	% of Variability Captured
PC 1	59.0%
PC 2	22.0%
PC 3	7.5%
Total	88.5%

Table 2 shows the variability described by the first 3 principal components (PCs). Settling Velocity of Sediment and Release Height were identified, with the 3 inputs from Table 1, as having an influence on the variations described by the first 3 PCs.

3 Conclusion

Traditional sensitivity analysis methods were extended to multivariate response models and applied to NewDEPOMOD to identify parameters that influenced the variation in the shape of the impacted area on the seabed. Further work aims to develop a spatio-temporal emulator of NewDEPOMOD to allow predictions to be made without the computational cost.

Acknowledgments: Special thanks to EPSRC (Award Ref: 1953182) and SEPA for funding the PhD and Andrew Berkeley (formerly SEPA) for his support in the completion of this work.

References

Cromey, C.J., Nickell, T.D., and Black, K.D. (2002). DEPOMOD - modelling the deposition and biological effects of waste solids from marine cage farms. *Aquaculture*, **214**, 211 – 239.

Dryden, I. and Mardia, K. (2016). *Statistical shape analysis: with applications in R, 2nd Edition* John Wiley & Sons Ltd.

Iman, R. and Conover, W. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, **11**, 311 – 334.

Saltelli, A. and Chan, K. and Scott, M. (2000). *Sensitivity Analysis.* John Wiley & Sons Ltd.