Parameter estimation in stock assessment modelling: Caveats with gradient-based algorithms

Article *in* ICES Journal of Marine Science · April 2018

DOI: 10.1093/icesjms/fsy044

CITATION
1

READS
83

1 author:

Sam Subbey
Institute of Marine Research in Norway
77 PUBLICATIONS **402** CITATIONS

SEE PROFILE

ICES Journal of Marine Science

ICES Journal of Marine Science (2018), doi:10.1093/icesjms/fsy044

Food for Thought

Parameter estimation in stock assessment modelling: caveats with gradient-based algorithms

Sam Subbey[1,2]*

[1]Institute of Marine Research, PO Box 1870, N-5817 Bergen, Norway
[2]Cornell University, Room 120 Fernow Hall, Ithaca, NY 14853, USA

*Corresponding author: e-mail: samuels@imr.no.

Subbey, S. Parameter estimation in stock assessment modelling: caveats with gradient-based algorithms. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsy044.

Received 12 November 2017; revised 19 March 2018; accepted 20 March 2018.

Using simple illustrative examples, this note highlights some of the caveats with gradient-based algorithms. This class of algorithms underpins the state-of-the-art modelling platform in fisheries science. The goal is to sound a cautionary note about an increasing trend in fisheries science, where blind faith is being invested in results obtained from algorithms that are *fast*, and proven to have *machine precision*.

Keywords: automatic differentiation (AD), global optimum, Hessian, likelihood, local minima, objective function, optimization, parameters, positive definite, uncertainty.

Introduction

In fisheries science and ecology, models have become standard tools for the study and understanding of complex ecosystem phenomena, and to aid in management decisions (see e.g. Munch and Kottas, 2009; Subbey *et al.*, 2014; Collie *et al.*, 2016; Storch *et al.*, 2017). The Automatic Differentiation Model Builder (ADMB)/Template Model Builder (TMB) platform (Fournier *et al.*, 2012; Kristensen *et al.*, 2015) has made it possible to estimate the parameters of high-dimensional complex models, and to quantify model and parameter uncertainties. The underlying parameter estimation algorithm is fast, stable, and accurate, and has been described as capable of handling several hundreds of parameters. In fisheries science, the majority of end-users (of complex models and sophisticated algorithms) of the ADMB/TB platform consist of scientists with varying computational backgrounds. It is therefore unfair to expect that there will be uniformity across the scientific community in understanding the verbiage used (mathematical/statistical terminologies), and the functioning (including caveats) of the underpinning computational algorithms. The goal of this article is to contribute to reducing this non-uniformity. The approach is to present some of the general characteristics of the optimization algorithm used by ADMB and TMB, in a language that is understandable to fisheries scientists, independent of quantitative background. Mathematical details will therefore be kept to the barest essentials, in order to effectively communicate key issues.

This article has been motivated in part, by the author's feeling that very often, explanations for inconsistent parameter estimation results have been inadequate, or in some extreme cases, grossly misleading. A commonly reported experience is that small changes in initial guess values of parameters; result in disproportionate variation in the model results. As this article will later show, ignoring such inconsistencies may result in the acceptance of wrongly calibrated models. This also implies that model predictions and any inference based on the calibrated model will be uncertain, and uncertainty envelopes (or confidence intervals) generated on the basis of the calibrated model may fail to encapsulate true values. For stock assessment models, this may lead to erroneous inference about population size, and wrong estimates of parameters that are central to management. As a consequence, management decisions may result in non-sustainable fisheries, or misguided recovery plans.

A growing trend in fisheries science is the increasing use of the ADMB/TMB platform to develop stock assessment models of varying functional complexity and degrees of parameterization. The attraction with ADMB/TMB platform is the ease of use (flexibility and modularity) that enables a scientist, irrespective of

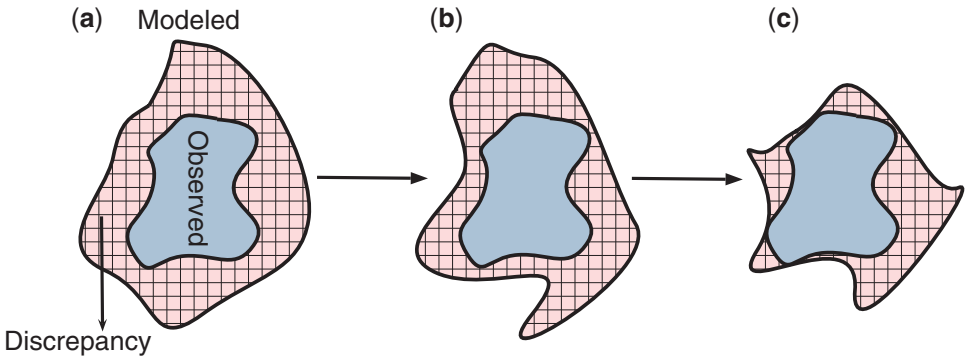

Figure 1. The goal of optimization is to find the correct set of parameters that minimize the size (area) of the hatched space (the discrepancy) between observations (inner figure) and their corresponding values from the model (outer figure). The process could be iterative (from a to c), and the optimal parameter set is that associated with the least possible discrepancy computed by the procedure.

quantitative background, to develop a (complex and computationally demanding) model within days, which several years ago, would have taken a statistician/mathematician several months or even years to develop. Unfortunately, this same input-output ease of use enables the platform to be treated as a modelling black-box, where the criteria for validation of results is limited to two expressions—"*iteration has converged*", and "*the Hessian is positive definite*". Fisheries scientists with background in mathematics/statistics may recognize that these expressions are necessary, but not sufficient conditions for accepting model results. Other scientists, on the other hand, may need to be made aware of the inherent limitations of the computational algorithm that supports the ADMB/TMB platform, so that caution will be applied in interpreting model results. This knowledge is also fundamental to understanding the extended consequences of inconsistent model results.

The caveats discussed in this article are far from exhaustive; that has not been the goal. The aim is to provide information that facilitates admission of biologists and other non-quantitative fisheries scientists into the fold, when discussing the credibility of parameter estimation and uncertainty quantification results. The author will consider his quest fulfilled, if this article lowers the threshold for such inclusive discussion.

Mathematical and statistical models

Mathematical/statistical concepts and language are usually employed to describe our conception of complex systems and how their dynamics evolve. The concepts may describe relationships (e.g. linking biological processes such as age and weight), while the language may consist of abstract codes, which represent antecedent or future events and concepts, Carnie (2013). The collection of the concepts and language define what is referred to as a mathematical/statistical model. When the language is mathematical/statistical, the model is referred to as a mathematical/statistical model. Whether mathematical or statistical, models are defined in terms of some observations, and unknown variables of the system. The variables are referred to as parameters.

Application of a model, either to understand the system being modelled or for predicting its future behaviour, requires knowledge of the model parameters. For ecological and fisheries models, the parameters are estimated on the basis of observed data. When, as is customary, the data is sampled over some discrete time, it is referred to as *time series*. The goal of optimization is to find appropriate parameter values of the model, so that it is able to replicate the observations as closely as possible. Figure 1 illustrates the process. Note that the discrepancy (hatched area in Figure 1) can represent differences in time series, sets of parameters, or even matrices (e.g. survey indices or commercial catch statistics).

If we assume, for the sake of discussion, that our conceptual model and its parameters are based on perfect knowledge, then the discrepancy will simply be the observation errors. Unfortunately, these assumptions are unrealistic. In practice, we usually have imprecise observations, while our models, because they are built on partial knowledge, are usually uncertain. This is particularly true because functions used to represent complex biological and ecological phenomena may be oversimplified and inappropriate (structural uncertainty). Furthermore, large uncertainties may be associated with the model parameters (parametric uncertainty). These (data and model) uncertainties imply that there will always exist a discrepancy (hatched region in Figure 1) between our observations and those produced by the model.

In mathematics, one uses a function to define a *merit number* that represents the discrepancy between observations, and their corresponding modelled values. This function is referred to as the *objective function*. In statistics, this *merit number* is usually expressed in terms of probability, and referred to as the *likelihood* value. In the simplest approach, we can define the (least-squares) objective function by the sum of squared deviations of the observations from their corresponding modelled values. There are however, other possible ways of expressing the objective function, such as the sum of absolute deviations (see Uncini, 2015; Berry *et al.*, 2016). Irrespective of the definition form of the objective/likelihood functions, the optimization (parameter estimation) problem involves using a set of computational (mathematical/statistical) rules and procedures (collectively referred to as optimization *algorithms*), to find the best parameter set, so that the objective function value (merit number) is almost zero. In simple terms, operations of optimization algorithms can be likened to *searching a mountainous terrain to identify the lowest lying valley, and the co-ordinate of the lowest point in this valley*. This lowest point is called the *global minimum*.

There are two main classes of algorithms for identifying the global minimum. Stochastic algorithms (SAs) iteratively try new candidate solutions in the region that defines the parameter

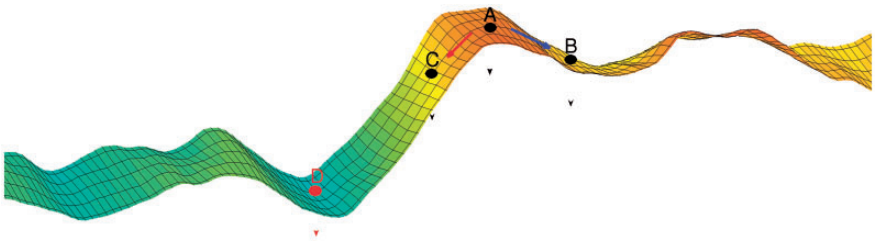

Figure 2. An illustrative misfit surface.

space, based on a probability rule. This class of algorithms includes Monte Carlo sampling, Genetic, and Simulated annealing algorithms (see e.g. Li and Jiang, 2000). Because they attempt to sample the whole parameter space, SAs are in general very slow and computationally ineffective for problems involving high number of model parameters. A viable alternative to SAs are Gradient Descent Algorithms.

Gradient descent algorithms

When implemented as computer procedures, this class of algorithms need input information (see Figure 2) about (i) start point, A, (ii) the number of steps, n, to perform, (iii) a procedure (rule) for how to transition from one point to the another (e.g. $A \rightarrow B$), and (iv) a criterion to determine whether the global minimum point (the point D in Figure 2) has been identified.

In most practical applications (especially for high-dimensional parameter problems), the choice of starting point and number of steps, is usually arbitrary. The guiding rule is to ensure that the starting point lies within the region defined by the parameter ranges, and large enough number of steps so that the algorithm does not terminate prematurely. The rule for moving from one point to another (A to B, say) is critical to the performance of Gradient Descent algorithms. Recall that the algorithm moves on a surface (referred to as the *misfit surface*) defined by the horizontal plane (parameters), and altitude (hills and troughs), whose values are determined by the objective function. Movement from one point to another on this surface involves a two-step decision namely, determining the direction in which to move ($A \rightarrow B$ or $A \rightarrow C$), and the step size in the chosen direction. The direction in which to move is motivated by the need to move along a descent path to the lowest point (global minimum) on the misfit surface. The following analogy illustrates the search philosophy. Suppose you are located on top of a mountain, in a poorly conversant terrain. However, you know there is a lake down the valley, and being extremely thirsty, you need to navigate your way down to the lake for a drink. The task come with the caveat that it is pitch dark, and visibility is practically zero. A practical solution will be to take a step at a time, and feel the ground near you, to identify which direction the terrain tends to descend. If this is done one step at a time (iteratively), you will eventually reach the valley. If the starting point is A in Figure 2, there will be a bigger propensity to move from A in the direction of C, than towards B. This is because the steepness from A to C is bigger than from A to B. Gradient algorithms move in the direction of steepest descent. The speed and accuracy in arriving at the global minimum is dependent therefore, on precision of the steepness (gradient) information. Automatic differentiation (AD), (also known as algorithmic-, or computational differentiation) is a technique for numerically determining the steepness (degree of descent or gradient) of a function specified by a computer programme. The AD technique allows for exact calculation of the gradient at any point in the misfit surface, to machine precision. For most effective implementations, the step size is adaptive; getting smaller as the algorithm approaches the global minimum.

The algorithm is considered to have converged, when the difference in the objective function (heights) of two consecutive points is less than a user-defined precision level. This value is known as the *convergence criterion*, and in practice, it is usually set close to machine precision. A stronger convergence criterion can be set by requiring that in addition, the surface in the vicinity of the identified optimal point also satisfies some curvature conditions. This curvature determines whether the point identified as global optimum lies in a trough (as would be expected) or otherwise (see Thacker, 1989). The curvature is calculated using a metric that is based on the *Hessian matrix*. In mathematics, the Hessian matrix is used to calculate the local curvature of a function (e.g. objective function) of many variables (parameters). When the curvature is concave-up (i.e. in a trough, such as at point D in Figure 2), the surface curvature is referred to as having a *positive-definite* Hessian. On the other hand, the surface curvature in the vicinity of e.g. point A, is said to have a *negative-definite* Hessian.

Fast, stable, and accurate algorithms

Though implementation details may differ (e.g. adaptive, rather than fixed step size) the principal rule for moving from one point P_i to the next, P_{i+1}, is given by

$$P_{i+1} = P_i - S \times G, \tag{1}$$

where S is the step size, and G is a measure of the gradient. The most challenging part of this implementation is obtaining accurate values for G. When information about G is available, the algorithms behave much the same way as descending down a hill where, at every step along the way, there is precise information about the shortest (steepest) path to take. Most algorithms are slow, unstable or inaccurate because the value of G is obtained by approximation. Thus the algorithms perform a large number of "*detours*" that result in prolonged computation time, and round-off (approximation) errors pile up to make the algorithms unstable and inaccurate. AD (Corliss *et al.*, 2002) uses computer algebra (see e.g. Von Zur Gathen and Gerhard, 2013) to derive computer codes from which values of G can be obtained with machine precision. The resulting gradient-based algorithm is therefore usually fast, stable and accurate.

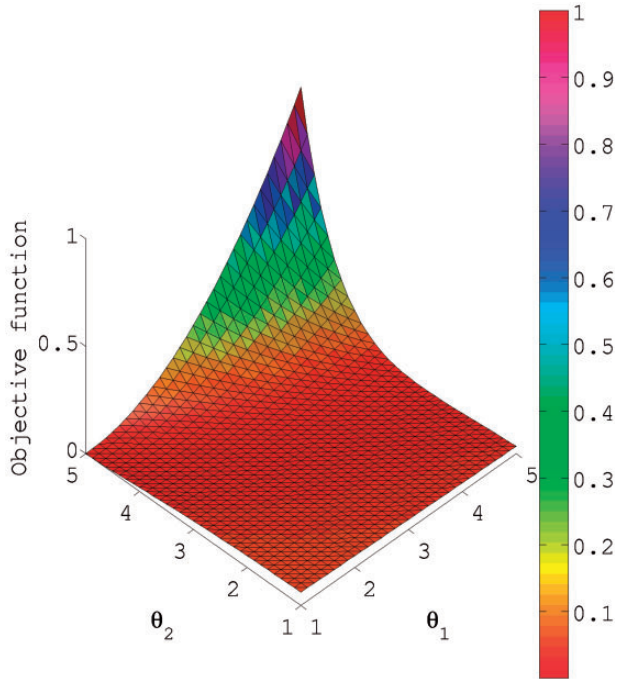

Figure 3. Sum of squared deviations.

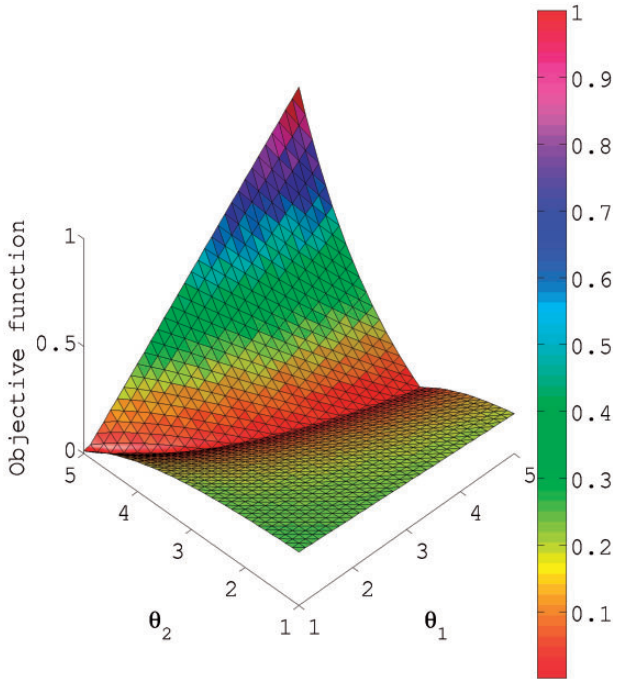

Figure 4. Sum of absolute deviations.

The gradient algorithm underpinning the ADMB/TMB platform uses the AD technique to obtain precise gradient information, and the global minimum is assumed to have been identified when both conditions of minimum convergence criterion, and positive-definiteness of the surface around the identified global minimum are fulfilled.

Caveats

In this section, I discuss six caveats associated with the class of algorithms discussed in the previous section. When they are deemed to simplify the discussion, graphical illustrations will be used to emphasize key points.

Caveat 1. The misfit surface is non-unique

The surface depends on our assumptions of the error structure (correlated or uncorrelated) in the observations and their assumed statistical distribution. In the literature, the functional description of the error structure is often referred to as the *error model*, see Fuller (2009). The error model is termed Gaussian, when errors are assumed to be independently and normally distributed about an average value Fuller (2009). On the other hand when the errors are assumed to be distributed about a median value, the error model is termed Laplacian, Norton (1984). Figure 3 (Gaussian) and Figure 4 (Laplacian) show how the likelihood surfaces differ, depending on the error model assumed. The volume of a sphere $\left(V = \frac{4\pi}{3} r^3\right)$ with radius $r = 2$ represents the truth. Next, I assumed that the constant $\left(\frac{4\pi}{3}\right)$ and exponent of the radius (3) are unknown. This exemplifies a model with two unknowns, θ_1 and θ_2, whose true values are respectively, $\frac{4\pi}{3}$ and 3. Figures 3 and 4 show the misfit surfaces for different combinations of θ_1 and θ_2, and for different error distribution assumptions. For data that has long tailed distributions or outliers, defining the objective function in terms of absolute deviations is known to be superior to least squares (see e.g. Alarcon-Aquino *et al.*, 2005). However, partly because of mathematical convenience and computational ease, and also because least squares approaches (and generalizations) have well known properties, this definition of the objective function has enjoyed unparalleled popularity in statistics and fisheries science (Subbey, 2017).

In the rest of the article, two benchmark (objective) functions will be used to better highlight the caveats to be discussed. These functions are the Branin (Jones *et al.*, 1998) and the Three-hump camel function (Thcf) (Liu, 2002). Detailed mathematical information about these functions can be found in the literature cited, though such details have been intentionally omitted in this article, given the target audience, and also to focus the discussion.

Caveat 2. An appropriate starting point may be crucial

There is no uniquely established method for choosing an appropriate starting point that guarantees convergence of the algorithm to a global minimum. The rule of the thumb is to ensure that the starting point is within the space defined by the parameter ranges. For convergence to the global minimum irrespective of starting point, the misfit surface must obey specific mathematical conditions that seldom apply in practice. We use the Thcf in Figure 5 to illustrate the disproportionate consequence of an insignificant variation in the starting point. The only difference between the starting points $A(-0.85, 2.00)$ and $B(-0.80, 2.00)$ lies in the x-axis coordinates. If one starts from point B, the algorithm converges to the desired optimal point. However, a 6.25% perturbation in the x-axis coordinate of B (resulting in A) causes convergence to the wrong point $A^*(-1.75, 0.87)$.

Caveat 3. Satisfaction of convergence criteria is no guarantee for global minimum

The Thcf (as in Figure 6) is used here to illustrate this caveat. The algorithm was run with identical convergence criteria

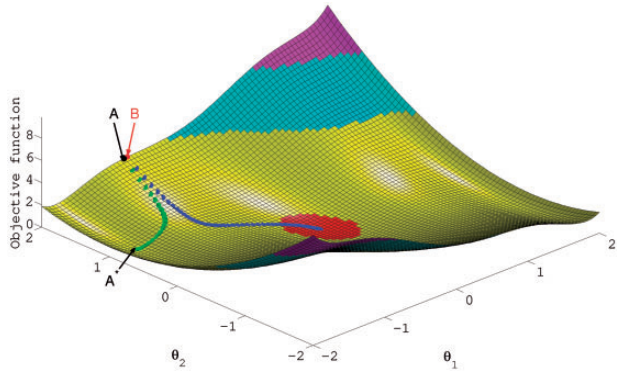

Figure 5. Trajectory of the search algorithm starting from two close, but different points (*A* and *B*) on the surface. The circular patch in the surface marks the absolute minimum point, located at the origin (0, 0), and its neighbourhood.

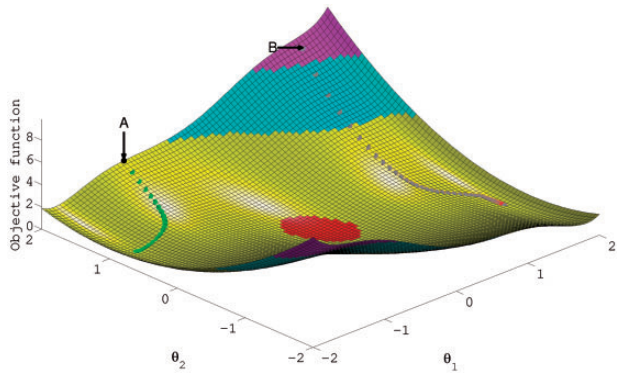

Figure 6. Trajectories of the search algorithm starting from two different points in the surface. The circular patch marks the absolute minimum point, the origin (0, 0), and its neighbourhood. The identical convergence criteria are fulfilled for each search result. However point $A(-0.85, 2.00)$ converged to $(-1.75, 0.87)$, while $B(1.55, 1.90)$ converged to $(1.75, -0.87)$, none of which is the global minimum.

(maximum convergence error of 10^{-6}, and surface curvature—positive definite—conditions) for both points. Observe however, that though gradient values to machine precision were used, and despite meeting the convergence criteria, none of the points converged to the correct global minimum at the origin.

Caveat 4. Multi-start approaches have limited application in practice

Multi-start approaches (see e.g. Martí *et al.*, 2013) involve initiating the optimization algorithm from different points within the search region, thus increasing the probability of hitting the absolute minimum. Unfortunately, this approach still suffers from observations made under Caveat 3. For high-dimensional problems, the approach can be time and computationally expensive, while offering no guarantee of success. Depending on the nature of the misfit surface, multi-start points that are thought to be wide apart may in fact, lie in the same neighbourhood, but converge to points that are widely separated in the parameter space. The set of all converged points may also fail to capture the global optimum point, and the point with the least objective function

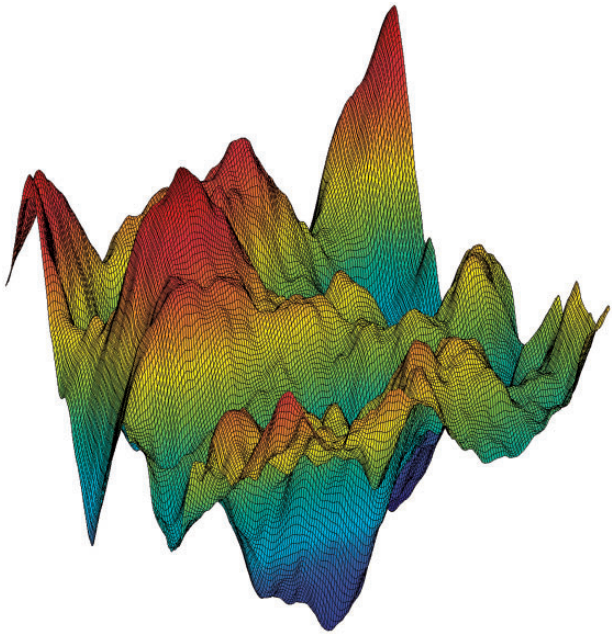

Figure 7. An arbitrary misfit surface with multiple local minima. It is non-trivial and inconceivable, without a priori knowledge of the surface, how a multi-start search may be developed for such a misfit surface.

value therefore, may not represent the global minimum. In summary, even for low-dimensional problems (see Figure 6) a multi-start approach offers no guaranteed of identifying the optimal parameter set. This consideration becomes even poignant, when one considers that the arbitrary surface presented by Figure 7, may be representative of misfit surfaces for high-dimensional problems, and entrapment in a local minimum point is a realistic scenario.

Caveat 5. Limitations posed by single optimum identification

According to the literature (see e.g. Hagstrom and Levin, 2017), complex adaptive systems provide a unified framework for explaining marine ecosystem phenomena. The inherent high complexity is expressed by the exhibition of both randomness and regularity, resulting from organized and structured interactions among ecosystem components. Thus ecosystems have the possibility of existing in multiple alternative states of equilibrium (Petraitis and Dudgeon, 2004; Schröder *et al.*, 2005). It is therefore natural to expect that observations (e.g. time series) from ecosystem observations may contain information from different states. The proportion of time the system spends in each of the state may be used to assign probabilities to the states. A measure of complexity of the system may then be defined, based on these probabilities. For marine systems to have high statistical complexity, they must have a large number of approximately equiprobable possible states, Ladyman *et al.* (2013). In terms of modelling, this translates to the possibility of different sets of model parameters, with approximately identical likelihood/objective function values. Using a misfit surface defined by the Branin function, Figure 8 illustrates the drawback with gradient-based algorithms, when

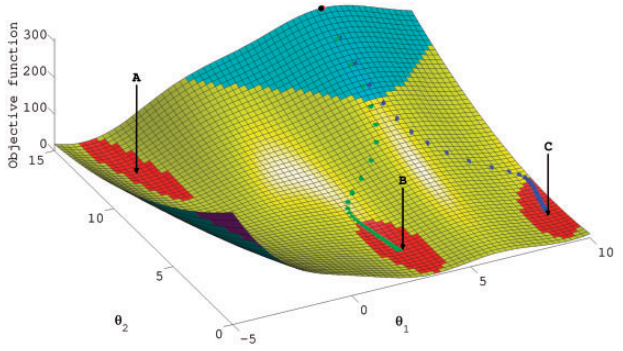

Figure 8. The Branin function has three global minima [at $A(-3.142; 12.275)$; $B(3.142; 2.275)$; and $C(9.425; 2.425)$], with the same value (0.398) for the objective function. The neighbourhood of the global minima are represented by circular patches. Initiating the search algorithm from two almost identical points $(6.20, 15)$ and $(6.25, 15)$, result in two notably different results, $B(3.1416, 2.2750)$ and $C(9.4248, 2.4750)$, respectively.

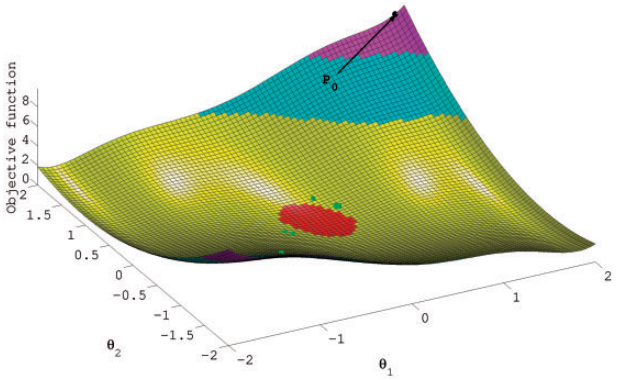

Figure 9. Result for the Thcf with large (10-fold) step-size, compared with. The algorithm was initiate at $(1.9, 2.0)$, and the dots around the circular patch show the trajectory of the algorithm, which fails to converge to the absolute minimum value. The algorithm got stuck between two points (0.3502, 0.1751) and $(-0.3502, -0.1751)$.

faced with multiple, equiprobable solutions, such as are assumed to characterize marine ecosystems.

Caveat 6. Non-convergence of the algorithm may be due to control parameters

We use Figure 9 to illustrate an example of this caveat, which involves the effect of step-size on the performance of the algorithm. The algorithm was executed using the same parameters as previously (for the Thcf), except that the step size (from one point to the next) was multiplied by a factor of 10. This particular example shows that too large time-steps lead to non-convergence of the algorithm. On the other hand, choosing too small time-steps means the algorithm will take a long time to converge, making it impractical for models with large throughput and number of parameters.

Discussion

This article has demonstrated some of the drawbacks with gradient-based algorithms, using analytical functions to represent the misfit surfaces. For these surfaces, gradients can be calculated (as with AD) to machine precision. It becomes poignant to think that for real problems, Figure 7 may be more representative of the misfit surfaces that have to be explored by the optimization algorithms. The caveats discussed for the 2D problems become exacerbated since such surfaces are usually characterized by multiple optima, existence of infeasible regions, flat or nearly flat spots, discontinuities, and infinitesimal area containing the global minimum/minima, compared with total misfit surface. A full discussion and illustration of how gradient-based algorithms might fare on surfaces with such characteristics is infeasible. However, examples presented in this note enable us to conjecture the possible outcomes. It is worth noting that some of the caveats raised in this article have been reported in different stock assessment and modelling fora. Unfortuntely, the discussion has never lasted longer than a usual stock assessment meeting. When optimal parameters are missed or wrongly identified, model performance predictions and associated uncertainties become exercises in futility. It is therefore hoped that the issues raised in this article will serve as caution, in an ever growing trend of almost blind faith in models, optimization algorithms, and in modelling frameworks that are based on such algorithms.

A sequel of (quantitative) papers will examine an alternative, and more robust approach to the parameter estimation problem. The approach integrates parameter space partitioning, a metaheuristics strategy (Blum and Roli, 2003) to guide a global exploration of the misfit surface, and a local (gradient-based) search of the parameter space. The sequel papers will also discuss choices of optimization algorithms for specific classes of problems in fisheries science, highlight issues connected to parametric complexity, and limitations in approaches (e.g. likelihood profiling) to escape entrapment in local minima.

Acknowledgements

This work has been partly funded by the Reduced Uncertainty in Stock Assessments (REDUS, Project no. 14809-01) at IMR. Research at Cornell University has been made possible through a Fulbright Foundation grant, for which the author is grateful.

References

Alarcon-Aquino, V., Garcia-Treviño, E., Rosas-Romero, R., Ramirez-Cruz, J., Guerrero-Ojeda, L., and Rodriguez-Asomoza, J. 2005. Wavelet-network based on l1-norm minimisation for learning chaotic time series. Journal of Applied Research and Technology, 3: 211–219.

Berry, K. J., Mielke, Jr, P. W., and Johnston, J. E. 2016. Permutation Statistical Methods. Springer, Cham.

Blum, C., and Roli, A. 2003. Metaheuristics in combinatorial optimization: overview and conceptual comparison. ACM Computing Surveys (CSUR), 35: 268–308.

Carnie, A. 2013. Syntax: A Generative Introduction. John Wiley & Sons, UK.

Collie, J. S., Botsford, L. W., Hastings, A., Kaplan, I. C., Largier, J. L., Livingston, P. A., Plagányi, É. *et al.* 2016. Ecosystem models for fisheries management: finding the sweet spot. Fish and Fisheries, 17: 101–125.

Corliss, G., Faure, C., Griewank, A., Hascoet, L., and Naumann, U. 2002. Automatic Differentiation of Algorithms: From Simulation to Optimization, vol. 1. Springer Science & Business Media, New York, USA.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A. *et al.* 2012. "AD Model Builder": using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 233–249.

Fuller, W. A. 2009. Measurement Error Models, vol. 305. John Wiley & Sons, New York, USA.

Hagstrom, G. I., and Levin, S. A. 2017. Marine ecosystems as complex adaptive systems: emergent patterns, critical transitions, and public goods. Ecosystems, 20: 458–476.

Jones, D. R., Schonlau, M., and Welch, W. J. 1998. Efficient global optimization of expensive black-box functions. Journal of Global Optimization, 13: 455–492.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. 2015. TMB: automatic differentiation and Laplace approximation. arXiv preprint arXiv: 1509.00660.

Ladyman, J., Lambert, J., and Wiesner, K. 2013. What is a complex system? European Journal for Philosophy of Science, 3: 33–67.

Li, B., and Jiang, W. 2000. A novel stochastic optimization algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 30: 193–198.

Liu, X. 2002. Several filled functions with mitigators. Applied Mathematics and Computation, 133: 375–387.

Martí, R., Resende, M. G. C., and Ribeiro, C. C. 2013. Multi-start methods for combinatorial optimization. European Journal of Operational Research, 226: 1–8.

Munch, S., and Kottas, A. 2009. A Bayesian modeling approach for determining productivity regimes and their characteristics. Ecological Applications, 19: 527–537.

Norton, R. M. 1984. The double exponential distribution: using calculus to find a maximum likelihood estimator. The American Statistician, 38: 135–136.

Petraitis, P. S., and Dudgeon, S. R. 2004. Detection of alternative stable states in marine communities. Journal of Experimental Marine Biology and Ecology, 300: 343–371.

Schröder, A., Persson, L., and De Roos, A. M. 2005. Direct experimental evidence for alternative stable states: a review. Oikos, 110: 3–19.

Storch, L. S., Glaser, S. M., Ye, H., and Rosenberg, A. A. 2017. Stock assessment and end-to-end ecosystem models alter dynamics of fisheries data. PloS One, 12: e0171644.

Subbey, S. 2017. Regularization of a parameter estimation problem using monotonicity and convexity constraints. ESAIM: Proceedings and Surveys, 57: 86–96.

Subbey, S., Devine, J. A., Schaarschmidt, U., and Nash, R. D. M. 2014. Modelling and forecasting stock–recruitment: current and future perspectives. ICES Journal of Marine Science, 71: 2307–2322.

Thacker, W. C. 1989. The role of the Hessian matrix in fitting models to measurements. Journal of Geophysical Research: Oceans, 94: 6177–6196.

Uncini, A. 2015. Least squares method. *In* Fundamentals of Adaptive Signal Processing, pp. 143–204. Springer.

Von Zur Gathen, J., and Gerhard, J. 2013. Modern Computer Algebra. Cambridge University Press, Cambridge, UK.

Handling editor: Ernesto Jardim