In: *Progress in Belgian Oceanographic Research*, National Committee of Oceanology, Royal Academy of Belgium, pp 43–46 (1996)

On model validation, sensitivity, and intercomparison

Eric Deleersnijder

"Science is now a tripartite endeavour with *simulation* added to the two classical components, *experiment* and *theory*" (Robinson, 1987). The routine use of numerical "simulation in scientific research (numerical experimentation, sensitivity and process studies, etc.) is thought by many to represent the first major step forward in the basic scientific method since the seventeenth century" (Robinson, 1987).

Marine sciences, as many other fields of scientific research, now intensively use mathematical and numerical models. There might even be a tendency, among some oceanographers, to overly rely on the predictions of the models, which are, in fact, no more than artefacts, *i.e.* man-made objects. The models, however complex they may be, provide, at best, a reasonable estimate of the behaviour of the system under study.

Numerical models are now so appealing that one might progressively forget that they are no substitute to reality. To avoid such "alienation", it is essential that every model be subject to activities such as validation, sensitivity analysis, and intercomparison — with other models. Validation, sensitivity analysis and intercomparison are closely related subject matters, which are briefly described below.

Validation. To assess a model, a series of numerical simulations may be carried out and discussed with appropriate reference to a reliable data set. Speculations may also be produced about the quality of the model results in conditions that are reasonably similar to those in which the model has actually been tested. This does not imply that, in a situation that has not yet been examined or envisaged, the model will prove successful. Therefore, since words such as *verification* or *validation* refer, in a certain sense, to the *truth* or the *legitimacy*, a model cannot be *verified* or *validated* (Oreskes *et al.*, 1994).

We should say that a model has been *confirmed* when arguments have been put forward showing that the model is capable, in certain circumstances, of satisfactorily representing

Keywords: validation, sensitivity analysis, intercomparison.

the behaviour of the natural system under study. *Validation* and *verification* are often used in the scientific literature in place of *confirmation*. Although this is somewhat irrelevant, we will continue to resort to the usual terminology, *i.e.* to use *validation*, and adopt Dee's (1995) definition: the "validation of a computational model is the process of formulating and substantiating explicit claims about the applicability and accuracy of computational results, with reference to the intended purposes of the model as well as to the natural system it represents". Since Dee (1995) distinguishes three stages in the construction of a model, there must be three validation stages, namely the conceptual, algorithmic and software validation.

As regards software validation, it must be pointed out that a Belgian model, MU-ROFI of the Management Unit of the North Sea Mathematical Models, includes a series a standard test cases applied every time the model is updated. This procedure does not guarantee that the software implementation is correct, but "at least checks many of the model components and reduces considerably the risk of program modifications contaminating previously-tested components" (Ruddick, 1995).

Another approach to model validation, complementary to Dee's, may be considered. According to Nihoul (1994), there are various types of models, *i.e.* the *test-*, *process-*, and *system-oriented* models. The validation process implies different requirements for each type of model:

— The test-oriented models are needed for "testing mathematical and numerical methods and their software implementation" (Nihoul, 1994). These models are based on a reduced set of equations and, hence, are generally not intended to realistically simulate the reality. Validating such models can be achieved without detailed reference to the natural system, as exemplified in Deleersnijder (1996).

— The process-oriented models focus on a few dominant or important processes. It is necessary that the key processes be well represented, at least from a qualitative point of view. In Deleersnijder *et al.* (1992), for instance, the model was to represent eddies in the lee of an island, as well as upwelling motion in the core of these eddies. Those phenomena were indeed simulated, but it was not certain that the magnitude of the vertical velocity was correct. Nevertheless, the model was deemed valid.

— The system-oriented models are "called upon to tackle a practical situation" (Nihoul, 1994). The results of these models must be as realistic as possible, which usually requires an appropriate data assimilation procedure to operate together with the prognostic program. Quantitative comparison between *in situ* measurements and model results must be made in order to validate such models (*e.g.* Heemink and Metzelaar, 1995).

Sensitivity analysis. A model cannot be claimed to be validated if no sensitivity analysis has been performed. The latter is the process of comparing the results obtained from a given model when modifying one or several features of the model — *i.e.* boundary conditions, initial conditions, parameterizations, numerical schemes, etc —, and producing interpretations thereof.

Sensitivity experiments are needed to understand how a model is working, to determine what aspects of a model are crucial and deserve most of the attention, and to help evaluating what part of the model results may or may not be trusted.

James (1986) performed a sensitivity analysis that is worth singling out. The advective part of a hydrodynamic model was dealt with. Model runs were carried out in identical conditions. An eddy was expected to form in the flow under consideration. Both the first-order upwind and the centered second-order discretizations produced unacceptable results.

With a TVD-like scheme, however, the development of an eddy was satisfactorily simulated.

Another example is the sensitivity study of a Rhine plume model to the turbulence closure scheme (Ruddick *et al.*, 1995). It was shown that the simplest parameterizations did not take into account the appropriate turbulent processes, leading to unacceptable salinity fields. On the other hand, complex turbulence models provided results that were qualitatively in agreement with the available observations. Beyond a certain level of sophistication, it proved difficult to determine whether or not it was desirable to further increase the complexity of the turbulence closure scheme.

Model intercomparison. Comparing the results produced by various models — addressing the same natural system — is generally thought to be a useful way of identifying the strengths and weaknesses of the models involved.

Over recent years the intercomparison of model results has been performed in various shallow sea domains (see Deleersnijder *et al.* (1995) for a short list of references). All of these exercises followed relatively similar procedures. First, the model runs to be carried out were defined so that the models had sufficiently similar computational domains and forcings. Second, the distance between the model results — and, in most cases, between the model reults and the observations — was evaluated according to various norms. Finally, attempts were made to identify the reasons for the discrepancies between the model results.

It seems that the first and second tasks were generally carried out successfully. The third stage, however, did not always lead to detailed or useful conclusions. One of the reasons underlying this partial failure was obviously the difficulty to link cause and effect when dealing with complex, non-linear models. In view of these difficulties, we have decided to seek techniques that could facilitate the understanding of the discrepancies between model results (Deleersnijder *et al.*, 1995). The potential of Stein and Alpert (1993) factor separation method is being investigated in this perspective.

Other problems are posed when trying to set up a model intercomparison programme. For example, if the models are to be used in an idealized situation, a sufficient knowledge of the properties of the exact solution is needed. This seems trivial... However, most well-known analytic solutions pertain to simplistic cases, which are thus virtually useless. This kind of difficulty is well illustrated in Roed *et al.* (1995). It is also worth pointing out that Baptista *et al.* (1995) suggested a number of error measures which may help circumvent some of the obstacles faced by Roed *et al.* (1995).

In realistic situations, the following cases may arise:

— All model results are in agreement with the data set, which is the ideal case. For example, in the Tidal Flow Forum (briefly reviewed by Werner, 1995) the tidal signal was very well represented by all models.

— The models agree with each other, but not with the data (*e.g.* de Vries *et al.*, 1995). In such a case, either the models are plagued by similar deficiencies, or the forcing is inappropriate — a problem which data assimilation may help to address.

— The models disagree with each other, and with the data (*e.g.* Hackett and Roed, 1995). It is then necessary to look for shortcomings in the models and in the forcing procedures.

Overall, most of the intercomparison exercises carried out so far in shallow sea domains failed to meet their declared objectives. However, these failures have generally been useful: something was always learned about the models, the test cases, or the intercomparison procedures.

References

Baptista, A.M., E.E. Adams and P. Gresho, 1995, Benchmarks for the transport equation: the convective-diffusion forum and beyond, *Quantitative Skill Assessment for Coastal Ocean Models* (D.R. Lynch and A.M. Davies), American Geophysical Union, pp. 241–268.

Dee, D.P., 1995, A pragmatic approach to model validation, *Quantitative Skill Assessment for Coastal Ocean Models* (D.R. Lynch and A.M. Davies), American Geophysical Union, pp. 1–13.

Deleersnijder, E., 1996, On the numerical treatment of a lateral boundary layer in a shallow sea model, *J. Mar. Syst.* (in press).

Deleersnijder, E., A. Norro and E. Wolanski, 1992, A three-dimensional model of the water circulation around an island in shallow water, *Cont. Shelf Res.*, **12**, 891–906.

Deleersnijder, E., J. Ozer and B. Tartinville, 1995, A methodology for model intercomparison: preliminary results, *Ocean Modelling*, **107**, 6-9 (+figures).

de Vries, H., M. Breton, T. de Mulder, Y. Krestenitis, J. Ozer, R. Proctor, K. Ruddick, J.-C. Salomon and A. Voorrips, 1995, A comparison of 2D storm surge models applied to three shallow European seas, *Environ. Soft.*, **10**, 23–42.

Hackett, B. and L.P. Roed, 1994, Numerical modeling of the Halten Bank area: a validation study, *Tellus*, **46A**, 113-133.

Heemink, A.W. and I.D.M. Metzelaar, 1995, Data assimlation into a numerical shallow water fow model: a stochastic optimal control approach, *J. Mar. Syst.*, **6**, 145–158.

James, I.D., 1986, A front resolving sigma coordinate sea model with a simple hybrid advection scheme, *Appl. Math. Model.*, **10**, 87–92.

Nihoul, J.C.J., 1994, Do not use a simple model when a complex one will do, *J. Mar. Syst.*, **5**, 401–406.

Oreskes, N., K. Shrader-Frechette and K. Belitz, 1994, Verification, validation, and confirmation of numerical models in the Earth sciences, *Science*, **263**, 641–646.

Robinson, A.R., 1987, Predicting open ocean currents, fronts and eddies, *Three-Dimensional Models of Marine and Estuarine Dynamics* (J.C.J. Nihoul and B.M. Jamart, Eds.), Elsevier, pp. 89-111.

Roed, L.P., B. Hackett, B. Gjevik and L.I. Eide, 1995, A review of the Metocean Modeling Project (MOMOP) Part 1: Model comparison study, *Quantitative Skill Assessment for Coastal Ocean Models* (D.R. Lynch and A.M. Davies), American Geophysical Union, pp. 185–305.

Ruddick, K.G., 1995, *Modelling of Coastal Processes Influenced by the Freshwater Discharge of the Rhine*, PhD Thesis, University of Liège, 247 pp.

Ruddick, K.G., E. Deleersnijder, P.J. Luyten and J. Ozer, 1995, Haline stratification in the Rhine-Meuse freshwater plume: a three-dimensional model sensitivity analysis, *Cont. Shelf Res.*, **15**, 1597–1630.

Stein, U. and P. Alpert, 1993, Factor separation in numerical methods, *J. Atmos. Sci.*, **50**, 2107–2115.

Werner, F.E., 1995, A field test case for tidally forced flows: a review of the Tidal Flow Forum, *Quantitative Skill Assessment for Coastal Ocean Models* (D.R. Lynch and A.M. Davies), American Geophysical Union, pp. 269–283.

Acknowledgements. *The author is a Research Associate with the National Fund for Scientific Research of Belgium. Thanks are due to Professor Nihoul for allowing the author to make an oral presentation at the 1996 workshop "Progress in Belgian Oceanographic Research".*

Eric Deleersnijder, Institut d'Astronomie et de Géophysique G. Lemaître, Université Catholique de Louvain, 2 Chemin du Cyclotron, B-1348 Louvain-la-Neuve, Belgium