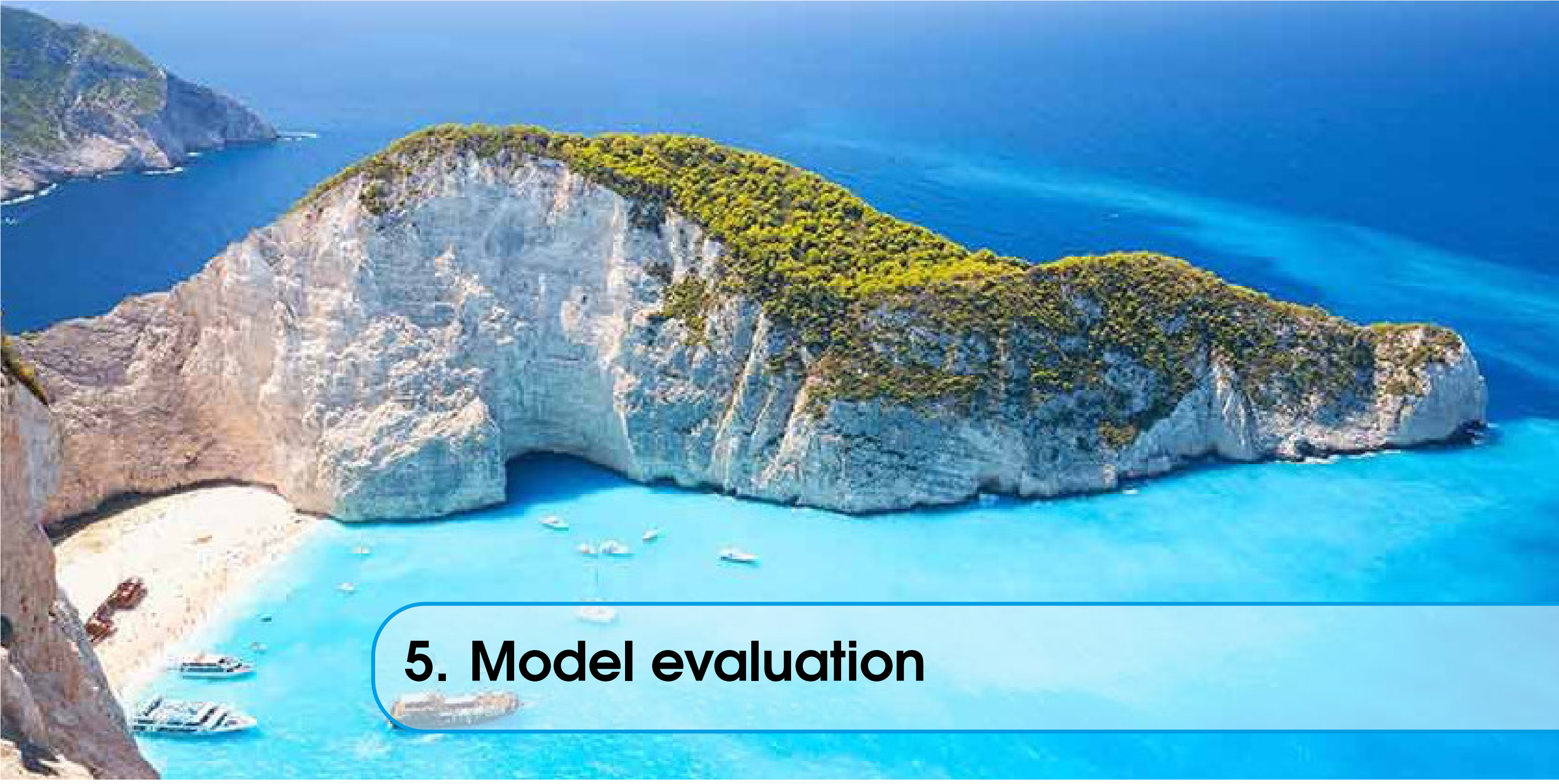


5. Model evaluation

5.1 Introduction

In this section, we address model evaluation and its related topics. How to evaluate a model is a crucial question for those who develop or use models, for it is the way to assess how well (or bad) the model performs when trying to simulate reality. But it's not enough to know that model evaluation is a necessity; one must also know what type of evaluation is need, and this is also addresses in this section. Finally, the importance of plotting the results is tackled, because it can (1) reveal the accuracy of the simulation, (2) illustrate the behavior of the model components, and (3) highlight to which degree model components are important.

5.2 How to evaluate a model

First, let's start by stating that model evaluation provides a methodology that helps us realize how different from reality model simulations can be. Also, it allows for us to be certain if the model behaves in the way we expect or think it should. Finally, this process is useful to determine which parts of the model are most likely to affect the results. We can use models without evaluating it first. However, this is not recommended, as it will seriously limit our understanding on the way they functions and on their limitations. As such, the lack of a proper model evaluation may compromise our modeling effort.

There are two main reason to evaluate a model:

- To check for human error in its construction or parametrization;
- To determine the circumstances/conditions where the model is likely to fail (because a model is a simplification of reality)

It is important to remember that, while not being possible to confirm if that a model will work, they will work, it is possible to check under what circumstances it is likely to fail, such like hypothesis that cannot be proved, only disproved. Ultimately, if a model fails after being evaluated, this can still teach us something about the system, and progress can be made even so.

5.3 What type of evaluation is needed?

There are a number of consequences we must expect from the exercise of model evolution. Among other outcomes, this process can lead to determine the accuracy of the simulation. Hopefully, this outcome will help to acknowledge how confident can we be in the results. Another outcome is an analysis of the behavior of the model. In this respect, it is possible to check if the model respond in the expected way to changes in the conditions of the simulation. Finally, it is also possible to resolve which components of the model are most important in determining the results. But before a quantitative model evaluation is performed, we should assess the results graphically, by plotting the results and have a quick, yet critical, look at them and to answer these two simple questions: (i) do they look approximately correct?, i.e., are the results within the limits and/or according to patterns observed in the "real world", and if not, (ii) what are the likely sources of error?

If the model fails this simple analysis, then we mat save a lot of time by not performing a thorough evaluating of a model that is obviously wrong. But if after plotting the results, the model appears to provide a reasonable fit to measured data (when available), then it should be evaluated quantitatively.

5.3.1 Quantitative analysis

This analysis, needed in almost all applications, determines the accuracy of the simulation. It consists of a collection of statistical tests that allows the simulated values to be compared to measured values, and the outcome informs us how confident we can be in the simulated results.

5.3.2 Sensitivity analysis

The main objective of a sensitivity analysis is to evaluate the behavior of the model. This method is used to determine whether the model responds in the expected way to changes in the conditions of the simulation, and its output can be a plot of the changes in the simulated values against changes in the model components.

5.3.3 Uncertainty analysis

This analysis relies on similar techniques and software as the sensitivity analysis, and resolves the importance of the model components. Overall, it determines how much uncertainty is introduced into the model output by each component of the model. The most important outcome of this analysis is the information on how much effort should be focused into improving measurements of the different input values.

5.4 Plotting the results

Plotting the results is the simplest and, frequently, the fastest way to evaluate model performance. Plots may reveal the accuracy of the simulation, illustrate the behavior of the model components and eventually lead to the identification of important model components. The type of plot and the information on it depends on which of the model evaluation is being considered (accuracy of the simulation, behavior of the model components, or importance of model components).

5.4.1 Plots to reveal the accuracy of the simulation

These plots should include the measured and simulated values, usually presented on the same plot, thus allowing for differences to be clearly highlighted and not hide by changes in the axes, etc. There are two main ways in which results are commonly presented: (i) simulated values plotted against measured values, and (ii) simulated and measured values plotted against some other variable that is used in the model.

Simulated values can be plotted on the y-axis against measured values on the x-axis; the accuracy of the simulation is revealed by the proximity of the points to the 1.1 line. Such plots (i) highlight outliers, (ii) reveal any systematic shift in the simulated values with respect to measures values, and (iii) show differences in the trends in simulated and measured values.

Simulated and measured values can both be plotted on the y-axis against some other variable on the x-axis. This representation allows patterns in errors to be identified (e.g., shifts in time with respect to measures values, possibly indicating the rate of some process to be too slow). This type of plot (e.g., Fig. 5.1) can help to identify why a model is performing badly.

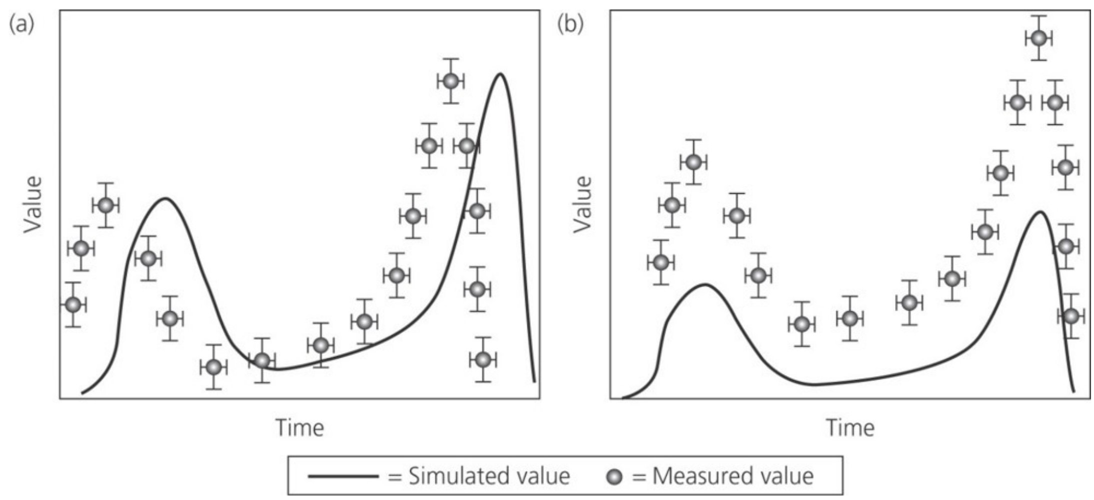


Figure 5.1: Examples on how plots can reveal patterns in errors: (a) simulated values showing a forward shift in time; (b) simulated values showing an upward shift with respect to measured values.

If error in the measurements are known, then these should be represented on the plots as error bars (usually showing the standard error or confidence interval), allowing for the accuracy of the simulations to be estimated directly just by looking at the plot. If the model is stochastic, variations in the simulated values should be represented as error bars or a band of potential results

5.5 Verification and validation

A model can be validated by examining how it performs over a range of conditions compared to observations of target system. A way to achieve this is by evaluating the "goodness-of-fit" between the model and observed values (e.g., coefficient of determination, R^2, or chi-squared, X^2); the closer the fit, the more accurate the model.

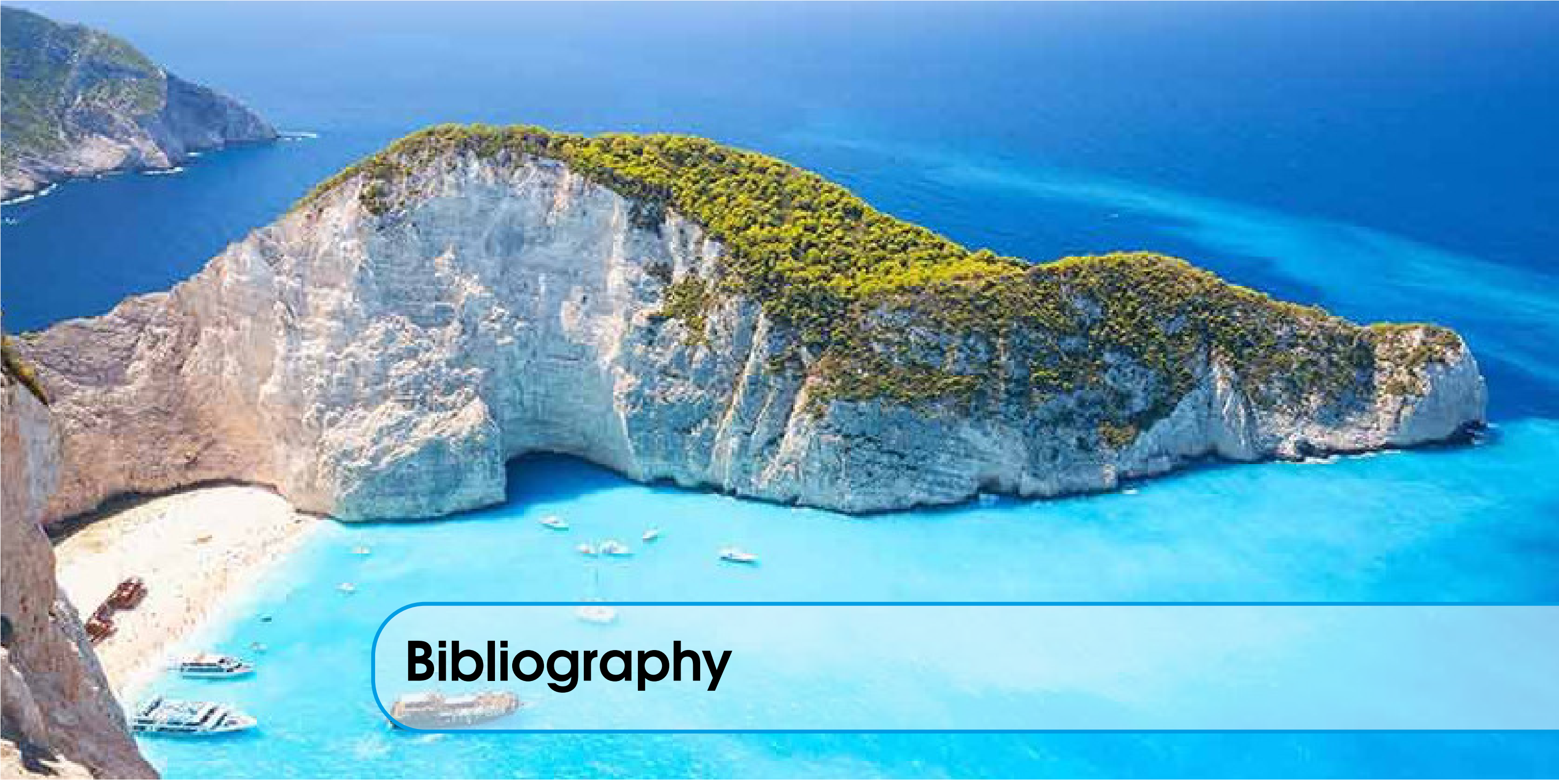
There are various reasons why the model output may differ from the observations. Among the main reasons are:

- Incorrect assumptions made in the conceptual model;
- Errors or inappropriate methods employed in the mathematical model;
- Mistakes or incorrect techniques in the implementation of the computational model;
- Inexact arithmetic (i.e., rounding errors) performed in the computational model;
- Uncertainty or errors in the data used to parameterize the model;
- Errors in the data used to test the model (i.e., measurement error).

5.6 Sensitivity analysis

Sensitivity analysis can be defined as the process of establishing the sensitivity of a model to its parameters. This analysis focuses attention on the critical parts of a model and the environmental system that it purports to represent. It is usually performed by perturbing the values of model parameters by known amounts, measuring the effect these have on the model outputs.

The variation can be made one parameter at a time (OAT); also known as univariate sensitivity analysis, or instead by multivariate sensitivity analysis techniques (e.g., Monte Carlo simulation with simple random sampling).



Bibliography

[1] J. Gunawardena, "Biology is more theoretical than physics," *Mol Biol Cell*, vol. 24, 2013.

[2] T. R. Anderson, "Progress in marine ecosystem modelling and the "unreasonable effectiveness of mathematics"," *Journal of Marine Systems*, vol. 81, no. 1–2, pp. 4–11, 2010.

[3] K. Davidson, "Modelling microbial food webs," *Marine Ecology-Progress Series*, vol. 145, no. 1-3, pp. 279–296, 1996.

[4] E. Hofmann, *Models for marine ecosystems*, pp. 409–432. International Thomson Publishing, 1997.

[5] S. L. Peck, "Simulation as experiment: a philosophical reassessment for biological modeling," *Trends in Ecology & Evolution*, vol. 19, no. 10, pp. 530–534, 2004.

[6] R. M. Nisbet, E. B. Muller, K. Lika, and S. A. L. M. Kooijman, "From molecules to ecosystems through dynamic energy budget models," *Journal of Animal Ecology*, vol. 69, no. 6, pp. 913–926, 2000.

[7] J. Blackford, J. I. Allen, T. R. Anderson, and K. A. Rose, "Challenges for a new generation of marine ecosystem models: Overview of the advances in marine ecosystem modelling research (amemr) symposium, 23–26 june 2008, plymouth uk," *Journal of Marine Systems*, vol. 81, no. 1–2, pp. 1–3, 2010.

[8] J. Harte, *Consider a Spherical Cow - A course in environmental problem solving*. Sausalito, CA: University Science Books, 1988.

[9] R. H. Austin and S. S. Chan, "Of spherical cows, cloudy crystal balls, and proteins," *Biochemical and Biophysical Research Communications*, vol. 312, no. 1, pp. 215–221, 2003.

[10] S. J. Cartwright, K. M. Bowgen, C. Collop, K. Hyder, J. Nabe-Nielsen, R. Stafford, R. A. Stillman, R. B. Thorpe, and R. M. Sibly, "Communicating complex ecological models to non-scientist end users," *Ecological Modelling*, vol. 338, pp. 51–59, 2016.

[11] M. Weijerman, E. A. Fulton, A. B. G. Janssen, J. J. Kuiper, R. Leemans, B. J. Robson, I. A. van de Leemput, and W. M. Mooij, "How models can support ecosystem-based management of coral reefs," *Progress in Oceanography*, vol. 138, pp. 559–570, 2015.

[12] M. Hantush and A. Chaudhary, "Bayesian framework for water quality model uncertainty estimation and risk management," *Journal of Hydrologic Engineering*, vol. 19, no. 9, p. 04014015, 2014.

[13] G. Bal, E. Rivot, J.-L. Baglinière, J. White, and E. Prévost, "A hierarchical bayesian model to quantify uncertainty of stream water temperature forecasts," *PLoS One*, vol. 9, no. 12, p. e115659, 2014.

[14] M. Svensson, P.-E. Jansson, D. Gustafsson, D. B. Kleja, O. Langvall, and A. Lindroth, "Bayesian calibration of a model describing carbon, water and heat fluxes for a swedish boreal forest stand," *Ecological Modelling*, vol. 213, no. 3–4, pp. 331–344, 2008.

[15] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling & Software*, vol. 15, no. 1, pp. 101–124, 2000.

[16] G. J. Bowden, G. C. Dandy, and H. R. Maier, "Input determination for neural network models in water resources applications. part 1—background and methodology," *Journal of Hydrology*, vol. 301, no. 1–4, pp. 75–92, 2005.

[17] L.-h. Chen and X.-y. Zhang, *Application of Artificial Neural Networks to Classify Water Quality of the Yellow River*, pp. 15–23. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[18] D. D'Ambrosio, S. Di Gregorio, S. Gabriele, and R. Gaudio, "A cellular automata model for soil erosion by water," *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, vol. 26, no. 1, pp. 33–39, 2001.