iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Evaluating consistency of biosphere models: software tools for a web-based service

G. A. Alexandrov and T. Matsunaga

Office for Global Environmental Database, Center for Global Environmental Research, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki 305-8506, Japan (g.alexandrov@nies.go.jp)

Abstract: A biosphere model is a geographical extension of an ecosystem model, and so modelling ecosystems at global scale we are facing the same problem as at the local scale -- structural uncertainty. The structural uncertainty includes competing conceptual frameworks, lack of agreement on model structure, ambiguous definitions of system boundaries, inadequate description of significant processes. The typical approach to this problem is assessing the maturity of the underlying science through retrospection of modelling efforts. This displays either consensus building or paradigm shift. In this paper we present a general scheme and software tools for performing such analysis, discuss how this scheme can be used for benchmarking a newly developed model, and specify a web-based service relevant to this purpose.

Keywords: Biosphere; Modelling; Model Consistency; Benchmarking; Web-based service

1. INTRODUCTION

A number of biosphere models has been developed in connection to the International Geosphere-Biosphere Program that backs the work of the Intergovernmental Panel on Climate Change. They are essential for projecting climate change scenarios and assessing mitigation and adaption options, however no criterion exists to pick one model over another [Cramer and Field, 1999]. Instead, the models form an ensemble, which is assumed to be a consistent estimator -- that is, to be converging to the quantities being estimated as the number of models grows. Is this assumption close to reality? Are confidence intervals shrinking? To answer this question we need regular checks of model consistency based on well-agreed methodology.

A biosphere model is a geographical extension of an ecosystem model, and so modelling ecosystems at global scale we are facing the same problem as at the local scale -- structural uncertainty. The structural uncertainty includes competing conceptual frameworks, lack of agreement on model structure, ambiguous definitions of system boundaries, inadequate description of significant processes [Manning et al, 2004]. The typical approach to this problem is assessing the maturity of the underlying science through retrospection of modelling efforts [Oikawa, 2007]. This displays either consensus building or paradigm shift.

In this paper we present a general scheme and software tools for performing such analysis, discuss how this scheme can be used for benchmarking a newly developed model, and specify a web-based service suitable to this purpose.

2. GENERAL SCHEME

2. 1 Methodological background

The evolution of scientific theories is often considered as a Darwinian process of natural selection that determines which theory survives and drifts them toward consensus [Bradie, 1994]. The concept of natural selection explains well

1. why well-established beliefs normally remain stable
2. and why independent researches addressing the same question ultimately come up with the same answer.

This does not imply, however, that an ultimate consensus is a legitimate goal. Stability of well-established beliefs stems from stability of research methods, and therefore it could be temporal.

2.2 Normative data

The well-established beliefs are based on normative data -- that is, the data supporting a judgement about what ought to be. The normative data represent the bulk of estimates obtained in previous researches, and in an ideal case, they characterize the maturity of knowledge in statistical terms -- by the confidence intervals of the mean values for the measured quantities. The estimates of independent studies are expected (Figure 1)

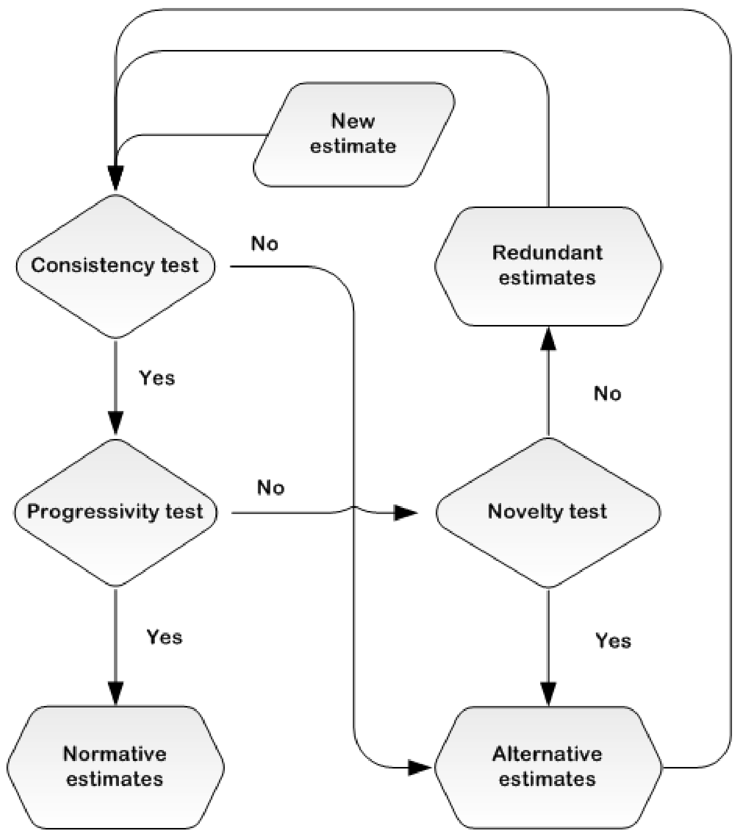

Figure 1. Flowchart for the process of building normative data

1. to fall within the bounds defined by normative data (consistency test),
2. to shift the mean value (novelty test),
3. and to narrow its confidence interval (progressivity test).

2.3 Alternative data

The process of "artificial selection" formalized above works against estimates suggesting too large shifts in mean values. These estimates form the pool of alternative data. With the passage of time, the bulk and consistency of alternative data may comprise to that of normative data, reflecting a paradigm shift.

Since the normative mean value is changing with every new estimate added to normative data, the pool of alternative data is re-processed and some estimates are moved to the pool of normative data. Therefore, the probability of sudden paradigm shift is quite low. This, however, does not completely remove the effect of order in which estimates appear, implying certain stability of beliefs originated from the pioneer studies.

2.4 Redundant data

The estimates that neither shift the normative mean value nor reduce its confidence interval form the pool of redundant data. This pool is also re-processed periodically and some estimates are moved to the pool of normative data.

3. SOFTWARE TOOLS

Program language

The software tools are written in *Mathematica* language [Wolfram, 1999] and arranged into a *Mathematica* package. The package contains the functions needed to perform tests (Figure 1) and to visualize the results (Figure 3).

Consistency test

Consistency test shows how far is the estimate (y) from the average ($\bar{x}$) in comparison to the lowest ($x_{\min}$) , or highest ($x_{\max}$) estimate:

$$z = \begin{cases} 100\dfrac{y-\bar{x}}{x_{\min}-\bar{x}}, & y<\bar{x} \\ 100\dfrac{y-\bar{x}}{x_{\max}-\bar{x}}, & y \geq \bar{x} \end{cases}$$

The test (z) is positive when estimate falls within the bounds defined by normative data ($x_{\min} \leq y \leq x_{\max}$), and negative otherwise. The estimate is said to be 100% consistent with the normative data if it coincides with the average estimate.

Progressivity test

Progressivity test is positive if inclusion of the estimate into normative data narrows the confidence interval of the average estimate. It returns the relative decrease in the width of the confidence interval:

$$z = 100\frac{x-y}{x}$$

where x is the original width of the confidence interval, y is the width of confidence interval changed due to inclusion of the estimate into the normative data.

Novelty test

Novelty test is positive if inclusion of the estimate into normative data shifts the average estimate. It returns the relative value of the shift with respect to the width of the confidence interval:

$$z = 100\frac{y-x}{u}$$

where x is the original mean value, y is the mean value changed due to inclusion of the estimate into the normative data, u is the half-width of the original confidence interval.

Case study

Net Primary Production (NPP) has been a focus of biosphere studies over the last three decades. First, the global pattern of NPP was characterized by the data collected during International Biological Program. Then, the data has been turned into empirical models that relate gradations in NPP to environmental factors of known geographic distribution such as mean annual temperature and precipitation (Miami model), actual evapotranspiration (Montreal model), and annually integrated NDVI. The ensemble (ToyBiMo) of three empirical models [Box et al., 1994] derived from the same NPP data shows high inconsistency of estimates in the case of larch forests and some other biomes (Figure 2). (Inconsistency is measured as half-width of the confidence interval of the normative mean value suggested by the model ensemble under concern.)

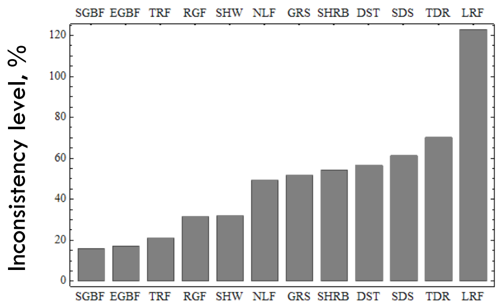

Figure 2. Inconsistency of the ToyBiMo ensemble. Legend: EGBF - evergreen broad-leaved forests, RGF - raingreen forests, TRF - tropical rainforests, SGBF - summer-green broad-leaved forests, SHW - subhumid woodlands, TDR - tundra, GRS - grasslands, NLF - needle-leaf forests, SDS - semi-desert scrubs, DST - deserts, SHRB - shrublands, LRF - larch forests.

A process-based model (TsuBiMo) calibrated with the same data [Alexandrov et al., 1992] reduces the ranges of uncertainty in the case of larch forest biome, but increase it in the case of the biome of evergreen broadleaf forest (Figure 3). Although the positive effects overweigh the negative effects, the case study illustrates well the fact that one can hardly expect an automatic reduction of uncertainties with every new model.

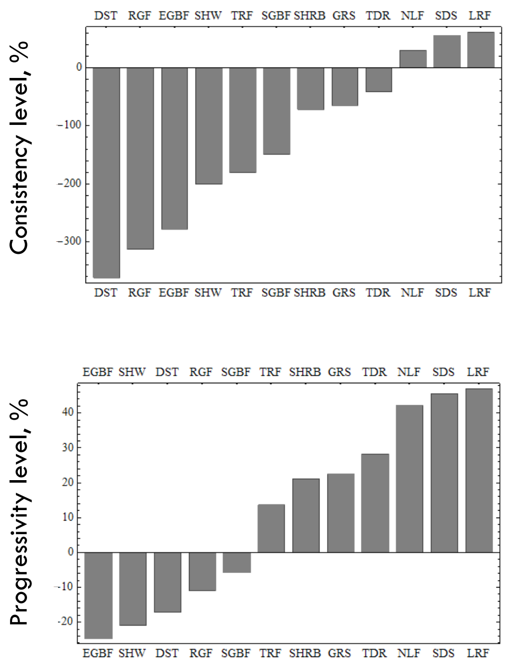

Figure 3. Consistency and progressivity charts: TsuBiMo vs ToyBiMo ensemble. Legend is the same as at the Fig 2.

The negative results of the consistency test stem from the fact that ToyBiMo suggests quite narrow range for NPP values, despite the wide confidence intervals. For example, in the case of evergreen broadleaf forests the highest and lowest values differ from average by 50 gC m^{-2} yr^{-1}. The width of confidence interval reflects the “sample size”, which is rather small.

Such a narrow range for NPP values makes it difficult for a model to pass this consistency test, and so many of current models would be considered as alternative models with respect to the ToyBiMo ensemble. Hence, one may anticipate that retrospection of modelling efforts would detect a paradigm shift in judgement about what the productivity of some biomes ought to be. The results of the tests are visualized in ascending order to display the “problem biomes”.

4. TOWARD A WEB-BASED SERVICE

4.1 Rationale

"In fact, the value of the software is proportional to the scale and dynamism of the data it helps to manage", [O'Reilly, 2005]. A retrospection of modelling efforts as well as benchmarking a new model requires a specialized database of model outputs. Without the data, the software tools are useless. The current version of the package includes some data. However, the database should be updated on regular basis, and the most efficient way of doing this is to employ the Web 2.0 "shiso" [Hayashi, 2007] that "aims to transform a society into an aggregated intelligence acting like a huge cyborg, by connecting people's individual intelligence (assumed as CPUs) through information and communications technology".

4.2 Technology

Since the software tools are written in *Mathematica* language, they can be used via the web interface [Wickham-Jones, 2006] to the *Mathematica* kernel. This interface, *webMathematica*, allows a web site to deliver JavaServer pages that call *Mathematica* commands. When the commands are evaluated, the computed result is placed in the page.

4.3 Specifications

In addition to the web interface mentioned above, the web-based service should include

1. a submission system for storing a model output in the specialized database
2. a reference system for getting information about the models available for retrospection analysis or benchmarking

4.4 Feasibility

The system for storing, retrieving and analysing 2-dimensional data as related to terrestrial carbon sink, recently developed by the Office for Global Environmental Database (OGED/CGER/NIES), includes all the components mentioned above. Therefore, feasibility of this web-based service depends mainly on the attitude of biosphere modellers, their inclination "to harness collective, net-enabled intelligence".

5. DISCUSSION

Modern computational and observational tools have caused an explosion of scientific data related to biosphere studies. This would improve eventually the consistency of biosphere models through creating multiple constraints for positioning 'true' values for model parameters. The limiting factor is thus the rate of building consensus on interpretation of the model outputs that will result from the new observations.

For example, terrestrial productivity is currently estimated at 60 GtC/yr. The consensus about this value was built in 1970s, although the estimates varied from 40 to 80 GtC/yr at that time, and re-analysis of the data (Alexandrov et al., 1999) revealed that estimates depend on how the data were classified with respect to the major regions of the world. In other words, this estimate is partly a social construction. Terrestrial productivity could be estimated at 50 or at 70 GtC/yr from the same observations.

The web-based service described above is to facilitate internalization of the new model results that may deviate from the existing consensus on biosphere characteristics. The scheme of building normative data simulates the natural process, but sets transparent criteria for distinguishing between normative and alternative data. It also suggests that an existing consensus should be re-considered when the bulk and consistency of alternative data comprise to that of normative data.

6. CONCLUSION

The biosphere models are essential for projecting climate change scenarios and assessing mitigation and adaption options. Many of them have been developed in connection to the International Geosphere-Biosphere Program (IGBP) that backs the work of the

Intergovernmental Panel on Climate Change (IPCC). Hence, there is certain demand for normative data on Net Primary Production (and its sensitivity to climate) as well as for normative data on other biogeochemical and eco-hydrological components of the climate system [Denman, et al., 2007].

This demand is partly satisfied by IPCC reports. Although one can hardly find the words "normative data" in IPCC reports, IPCC sets up certain norms by summarizing research results to produce policy-relevant recommendations. The recommendations are based on selected data, but the procedure of data selection is the Darwinian one: there are no explicit criteria of fitness.

There is nothing wrong in employing the method of "natural selection", with except to the risk of coming to an evolutionary deadlock. This risk can be significantly reduced through retrospection of modelling efforts, setting explicit criteria for distinguishing between normative and alternative data, and detecting paradigm shifts in timely fashion. OGED would spur these activities by providing a modern platform for processing the flows of data related to collaborative works of this sort.

ACKNOWLEDGEMENTS

The original manuscript was carefully revised in response to the valuable comments of Craig Aumann, Andrei Kirilenko, and Miquel Sanchèz-Marrè.

REFERENCES

Alexandrov, G.A., T. Oikawa and G. Esser, Estimating terrestrial NPP: what the data say and how they may be interpreted? *Ecological Modelling*, 117:361-369, 1999.

Alexandrov, G.A., T. Oikawa and Y. Yamagata, The scheme for globalization of a process-based model explaining gradations in terrestrial NPP and its application, *Ecological Modelling*, 148, 293-306, 2002.

Box, E.O., D.G. Dye, K. Fujiwara, R. Tateishi and X. Bai, Global environmental data sets from the Toyota Crown laboratory global engineering research project. Univ. of Tokyo, Tokyo, 1994.

Bradie, M., Epistemiology from an Evolutionary Point of View, In: Conceptual Issues in Evolutionary Biology (Ed.: E. Sober), MIT Press, 1994, 453-476.

Cramer, W. and C.B. Field, Comparing global models of terrestrial net primary productivity (NPP): introduction, *Global Change Biology*, 5, III-VII, 1999.

Denman, K.L., Brasseur, A., Chidthaisong, A., Ciais, P., Cox, R.E., Dickinson, D., Hauglustaine, D., Henze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P.L., Wofsy, S.C., Zhang, X., Couplings Between Changes in the Climate System and Biogeochemistry. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, M., Tignor, M., and Miller, H.L. (eds.)], Cambridge University Press, Cambridge, 2007.

Hayashi, S., "Information and Communications Technology and Shiso", *Science & Technology Trends-Quarterly Review*, 23, 11-22, 2007 [http://www.nistep.go.jp/achiev/ftx/eng/stfc/stt023e/qr23pdf/STTqr2301.pdf]

Manning, M.R., M. Petit, D. Easterling, J. Murphy, A. Patwardhan, H-H. Rogner, R. Swart, and G. Yohe (Eds), IPCC Workshop on Describing Scientific Uncertainties in Climate Change to Support Analysis of Risk and of Options: Workshop report. Intergovernmental Panel on Climate Change (IPCC), Geneva, 2004.

Oikawa, T., Private retrospection on ecosystem model development. Invited lecture presented at the OGED seminar, NIES, Tsukuba, 24 August, 2007.

O'Reilly, T., "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", 2005. [http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/ what-is-web-20.html]

Wickham-Jones, T., webMathematica: A User Guide, Wolfram Research, 2006.

Wolfram, S., Mathematica book, Cambridge University Press, Cambridge, 1999.