Uncertainty in Model Predictions: Does it Preclude Effective Decision Support?

Peter Reichert and **Mark E. Borsuk**

Swiss Federal Institute for Environmental Science and Technology (EAWAG), 8600 Dübendorf, Switzerland (reichert@eawag.ch)

Abstract: The uncertainty in the predictions of models for the development of environmental systems is usually very large. In many cases the width of predictive probability distributions for variables of interest is significantly larger than the difference between the expected values of the results for different policy alternatives. This seems to lead to a serious problem for model-based decision support, as policy actions appear to have an "insignificant" effect on decision variables relative to the predictive uncertainty. However, in some cases it is evident that some of the alternatives at least lead to changes with the desired trend. A formal analysis of this argument is made based on the dependence of the probability distribution of the variables of interest for different policy alternatives. This analysis leads to the conclusion that the uncertainty in the difference of model results based on different policy alternatives may be significantly smaller than the uncertainty in the results themselves. The knowledge about the width of this distribution is useful to assess the effectiveness of the alternative especially in the presence of high uncertainty of the predictions. This general argument is illustrated for phosphorus reduction scenarios in a simple, didactical model of phosphorus loading to a lake.

Keywords: Prediction uncertainty; Rational choice; Policy alternatives; Decision theory

1. INTRODUCTION

Mathematical models are useful tools to quantitatively summarize the state of knowledge about a system and to predict its future development. To provide a basis for choosing among alternatives, such predictions can be calculated for different policy scenarios. This makes model predictions an extremely useful tool for decision support.

As an example in integrative assessment of climate change, Fig. 1 shows globally averaged mean radiative forcing projections for three different tax scenarios without consideration of aerosol effects according to Morgan and Dowlatabadi [1996]. An evaluation of the implications of such projections is useful information for decisions regarding tax policy.

However, it has been realized that model predictions without accompanying estimates of uncertainty are not very useful for decision support [Morgan and Henrion, 1990; Reckhow, 1994; Morgan and Dowlatabadi, 1996].

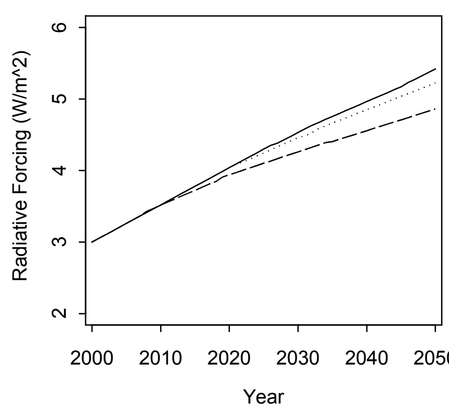

Figure 1. Globally averaged mean radiative forcing projections for three different tax scenarios without consideration of aerosol effects according to Morgan and Dowlatabadi [1996].

Many techniques are available to perform such uncertainty analyses [Beck, 1987; Morgan and Henrion, 1990]. For environmental systems, such analyses typically lead to the result that the uncertainty in model predictions is rather large [Morgan and Dowlatabadi, 1996; Reckhow and Chapra, 1999; Omlin et al., 2001; Reichert and Vanrolleghem, 2001]. If the expected values of model predictions for relevant system variables for different policy alternatives are closer together than the uncertainty range of the predictions, this raises questions regarding the usefulness of the predictions to support the choice between the alternatives. Such situations seem to imply that the current state of knowledge represented by the predictions does not allow the scientist to guarantee that the policies will have a significant impact.

This result of formal analysis can be in disagreement with the argument that a policy alternative is useful because it at least leads to a change in the desired direction. In situations in which the uncertainty in the actual value of a prediction of a system variable of interest is much larger than the uncertainty of the change resulting from a policy alternative, the above-mentioned argument is certainly true. This situation occurs naturally, if the major sources of uncertainty, e.g. lack of scientific knowledge or natural variability of external influence factors, affect the predictions for different policy alternatives in a similar way. It will not occur if the estimation of the effects of the policies (including the base scenario) is the major source of uncertainty. This raises interest in calculating the probability distribution of the difference in the predictions of the system variables for different policy alternatives as an additional piece of information relevant for the decision. A comparison of the distribution of the predicted differences with the distribution of the actual predictions would allow the decision maker to distinguish between those two cases.

To return to the example of integrated assessment of climate change mentioned above, the uncertainty in the three predictions shown in Fig. 1 can be assumed to be significantly larger than the difference in the predicted mean values. (In fact, the uncertainty of the predictions was calculated by the authors, but they did not report its value in the summarizing paper from which Fig. 1 was derived.) One example of a cause of uncertainty in such predictions is the effect of aerosols on radiative forcing. Figure 2 shows the effect of considering the effect of aerosols on the mean predictions [Morgan and Dowlatabadi, 1996].

Although, again, only mean effects are reported, based on an understanding of the underlying mechanisms, we can expect that the larger part of the uncertainty resulting from the estimation of the effect of aerosols will be the same for all three tax policy alternatives. Only a minor part of this uncertainty can be expected to lead to different effects for different alternatives. The situation can be expected to be similar for sources of uncertainty other than the role of aerosols.

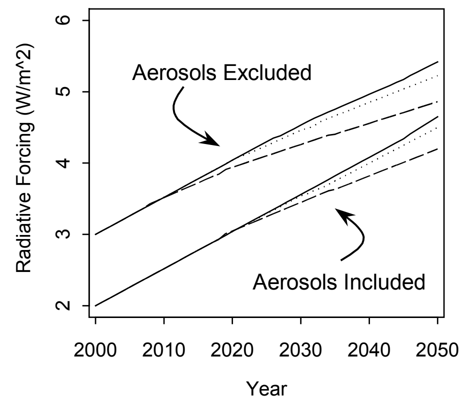

Figure 2. Globally averaged mean radiative forcing projections for three different tax scenarios without and with consideration of aerosol effects according to Morgan and Dowlatabadi [1996].

The discussion in this section leads to the insight that, in addition to the probability distribution of model predictions of variables for different policy alternatives, the distribution of the difference in these variables for different alternatives is of interest. Arguments support the hypothesis that, depending on the underlying mechanisms, the distributions of the differences can be (but are not necessarily) significantly narrower than the distributions of the actual variables. The knowledge about the width of the distribution of differences relative to the width of the distribution of predictions can be of importance for the decision maker, but is not usually considered in decision analysis. Therefore, the topic of this paper is to formally describe and illustrate how this information can be gained and used for decision making.

2. FORMAL DESCRIPTION OF THE PROBLEM

The results of a model prediction for given system variables at given locations in time and space, $\mathbf{y} = (y_1,...,y_n)^{\mathrm{T}}$, and for a given policy alternative, a, may be described by random variables,

$$\mathbf{Y}(a) = \left(Y_1(a),...,Y_n(a)\right)^{\mathrm{T}},$$

with a probability density,

$$f_{\mathbf{Y}}(\mathbf{y}, a).$$

Decisions are then usually based on the expected values of the variables of interest for the different alternatives,

$$\mathrm{E}[\mathbf{Y}(a)]$$

Because of the linearity of the operation of taking expectations, the expected values of the difference in the variables of interest is just the difference of the expectations between the results for the two policy options:

$$\mathrm{E}[\mathbf{Y}(a_1)-\mathbf{Y}(a_0)]=\mathrm{E}[\mathbf{Y}(a_1)]-\mathrm{E}[\mathbf{Y}(a_0)]$$

This simple equation involving expectations may disguise the fact that for the distribution of the difference,

$$\mathbf{Y}(a_1)-\mathbf{Y}(a_0)$$

there is no such simple formula. This distribution depends in a nontrivial way on the dependence between the two (vectors of) random variables $\mathbf{Y}(a_1)$ and $\mathbf{Y}(a_0)$. Because the relative effectiveness of any policy option depends on the value of this difference as compared to baseline conditions, it is the distribution of this difference that is often the quantity of most interest to the decision maker.

3. METHODS OF IMPLEMENTATION

It can be complicated to calculate the distribution of $\mathbf{Y}(a_1)-\mathbf{Y}(a_0)$ even when it is not difficult to calculate the distribution of $\mathbf{Y}(a_1)$ and $\mathbf{Y}(a_0)$ individually. Under certain circumstances, however, there are two relatively simple options.

The first option, proposed by Reckhow [1980], is to develop a model for the difference $\mathbf{Y}(a_1)-\mathbf{Y}(a_0)$ directly without building a model for the individual results, $\mathbf{Y}(a_1)$ and $\mathbf{Y}(a_0)$. This is an elegant solution to the problem if modelling differences is easier than modelling actual values of the system variables and if the individual results are not needed. This can be the case if the policy options are designed to improve an actual situation for which the variables of interest can be measured. This technique is more difficult to apply for situations that require projections into the future (as e.g. the integrated assessment example in section 1), because despite the interest in the difference that is stressed in this paper, the expected level of the base case is certainly also of relevance for the decision.

A second option can be realized if the uncertainty in some components of the model can be assumed to be either perfectly independent or perfectly dependent across policy scenarios. Such a dependence structure can be easily implemented using Monte Carlo simulation by holding some of the random realizations of parameters constant across policy scenarios and allowing others to be resampled.

4. ILLUSTRATION WITH A PHOSPHORUS LOADING MODEL

In this section, the second solution discussed above is illustrated using a simple, didactical model for phosphorus loading to a lake.

4.1 Model Description

It is assumed that there are three independent sources of phosphorus input to a lake, so that total yearly inputs can be calculated as,

$$\mathbf{Y}=\mathbf{Y}_\mathrm{B}+\mathbf{Y}_\mathrm{W}+\mathbf{Y}_\mathrm{A},$$

where $\mathbf{Y}$ is the total input, $\mathbf{Y}_\mathrm{B}$ is the base input associated with discharge from the catchment during dry weather, $\mathbf{Y}_\mathrm{W}$ is input from wastewater treatment plants, and $\mathbf{Y}_\mathrm{A}$ is agricultural input associated with high rainfall events. We assume that base input is distributed lognormally with mean μ_B and standard deviation σ_B. Similarly, treated wastewater input is also distributed lognormally with mean μ_W and standard deviation σ_W. Finally, we assume high rainfall events are Poisson-distributed with mean λ. Each high rainfall event leads to flushing of phosphorus from agricultural areas with a loading distributed lognormally with mean μ_A and standard deviation σ_A.

4.2 Policy Alternatives

We investigate two policy alternatives.

The first alternative consists of the implementation of a phosphorus removal procedure in the wastewater treatment plants. It is assumed that this leads to reduced phosphorus loading such that the new contribution is still lognormally distributed with a reduced mean value and the original standard deviation.

The second scenario consists of a new agricultural policy that reduces the average agricultural input of phosphorus per high rainfall event. Again, it is assumed that the new input is lognormally distributed with a reduced mean and the same standard deviation.

For both policy alternatives we assume that the modified distributions are independent of the distribution for the base case. This assumption is a simplification of reality made to keep the example as simple as possible. For the wastewater treatment plant this means that the implemented process modifications lead to a new behavior, the uncertainties of which are independent of those of the old plant. For the agricultural policy alternative this means that the modification not only consists of a reduced phosphorus application, but that the fertilization schedule is also modified. Otherwise there would be a correlation between the distributions of the two alternatives due to the occurrence of flood events.

4.3 Parameter Values

The values of the seven parameters of the model described in section 4.1 are given in Table 1.

Parameter	Value (in tons of phosphorus)		
	baseline	P-rem.	agri. pol.
μ_B	**4**	4	4
σ_B	**1**	1	1
μ_W	**4**	**2**	4
σ_W	**0.5**	0.5	0.5
μ_A	**4**	4	**3**
σ_A	**1**	1	1
λ	**0.5**	0.5	0.5

Table 1. Values of model parameters for the baseline scenario and two policy alternatives

Note that the expected value of total phosphorus load is equal to the sum of expected values of each component, or 16 t (4 t from basic discharge, 4 t from wastewater treatment plants and 8 t due to agricultrual inputs during high rainfalls). Both policy options reduce this expected value to 14 t. However, calculating the uncertainty in this reduction is not straightforward, as will be discussed in the next section.

4.4 Results

The model described above was implemented in the statistics and graphics package R [http://www.r-project.org] and 100000 Monte Carlo simulations were performed to approximate the probability distributions.

Figure 3 shows the probability distributions for total phosphorus loading for all three alternatives.

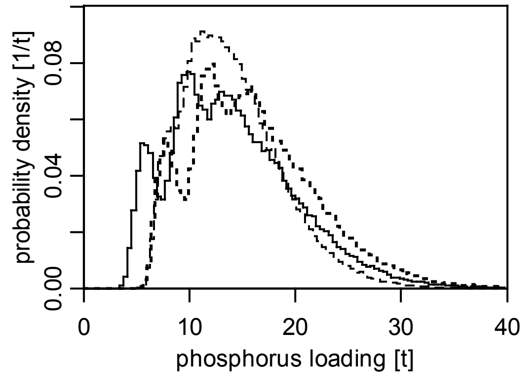

Figure 3. Phosphorus loading for the baseline scenario (dotted), for the wastewater phosphorus removal alternative (solid) and for the agricultural policy alternative (dashed).

The prediction uncertainty is very large for all three scenarios. Most of this uncertainty is due to the hydrologic variability associated with the number of high rainfall events during the year. Peaks at the lower end of the distributions are caused by the occurrence of zero, one, or two high rainfall events per year. For larger numbers of events, the peaks overlap more strongly, so that they are no longer visible in the density function. The distributions shown in Fig. 1 demonstrate that the reduction by 2 t achieved by both policies is overwhelmed by the width of the distributions. This is especially the case for the agricultural policy scenario because this scenario does not lead to a significant change in the lower tail of the distribution. The lower tail of the distribution consists primarily of cases in which no high rainfalls occur. An alternative that only reduces phosphorus flushing by high rainfalls, therefore, does not lead to a change in loading in these cases.

If, in addition to the independence assumptions stated in Section 4.2, we were to assume independence between the distributions for the baseline scenario and the policy options, the resulting distributions of the difference in phosphorus loading (Figure 3) would show that the widths of the distributions are much larger than the mean value of –2 t. (Standard deviations are 8.4 t and 7.5 t for the phosphorus removal alternative and for the agricultural policy, respectively).

A more refined and reasonable assumption in this case, however, would be to assume some degree of dependence among the predictions for the three scenarios.

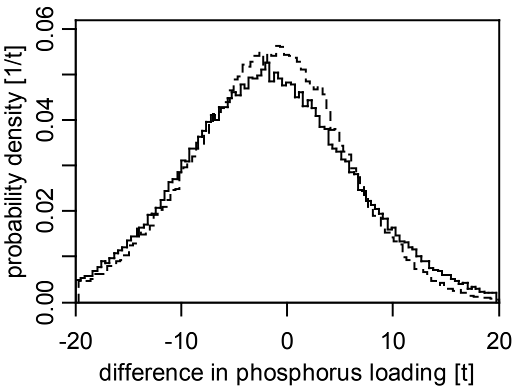

Figure 4. Difference in phosphorus loading for the wastewater phosphorus removal scenario and the baseline scenario (solid) and for the agricultural policy scenario and the baseline scenario (dashed) under the assumption of independence.

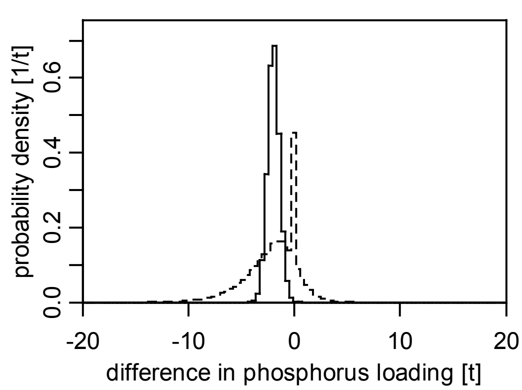

Figure 5. Difference in phosphorus loading for the wastewater phosphorus removal scenario and the baseline scenario (solid) and for the agricultural policy scenario and the baseline scenario (dashed) considering the dependence of the corresponding predictive distributions shown in Fig. 3.

For example, we would not expect that either the base input or the agricultural input would be affected by the proposed policy to increase phosphorus removal in wastewater treatment plants. Therefore, the sources of uncertainty for these two inputs can be assumed to be perfectly dependent for the baseline scenario and the wastewater removal scenario. This is accomplished by using the same Monte Carlo realizations for the parameters associated with these inputs for both policy scenarios, while allowing the parameters associated with the treated wastewater inputs to change. Similarly, we would not expect a change in agricultural policy to affect the sources of uncertainty in predicting the base or the treated wastewater inputs. Therefore, the distributions for these two sources can be assumed completely dependent across the baseline and agricultural policy. In addition, both policies do not change the number of high rainfall events occurring in a particular year. This makes the rainfall event distribution also perfectly dependent among the scenarios. However, as stated in section 4.2, the modified distributions of phosphorus loading per event and of loading from wastewater treatment plants are assumed to be independent from those of the base case.

These refinements lead to the distributions of differences across scenarios shown in Fig. 5. The distributions are now much tighter than predicted when all sources of uncertainty were assumed to be independent (Fig. 4). The standard deviations are reduced from 8.4 to 0.6 t and from 7.5 to 2.4 t for the wastewater treatment and agricultural policy scenarios, respectively.

The peak in the histogram approximating the density function of the difference at a value of zero for the agricultural alternative is caused by the probability of 27% that no high rainfall event occurs during a particular year. As, in reality, this is a probability for a discrete value, the results in Figure 5 might be more appropriately represented as a density function without this peak, representing 73% of the cases, and then a probability of 27% assigned to a value of zero. However, for ease of graphical presentation, we maintain the form of Figure 5.

5. DISCUSSION

The main reasons why the distributions of the differences in phosphorus loading in our example were much tighter than those of the predictions were: (1) the importance of natural variation affecting all alternatives in the same way, and (2) the composition of total load of three independent sources of which only one was affected by the policy option. The first reason leads to the same number of high rainfall events in each year over which the differences were calculated. The second reason implied that only the differences for the affected source had to be considered (the other sources would lead to the same loadings for the baseline scenario and for the policy alternative for any particular year).

In order to use the distributions discussed above for decision support, it is important to know the source of uncertainty. If the uncertainty in the distributions is dominated by natural variability, it can be argued that, in the long run and for systems that integrate the load, there is not much difference between the distributions shown in Fig. 4 and Fig. 5. In this case, only the expected value is of relevance because the influence of fluctuations decreases by integration. On the other hand, if scientific uncertainty were the major cause of uncertainty, a prediction such as that shown in Fig. 4 would provide little assurance that either policy will lead to visible reductions in phosphorus inputs and would suggest that there is no basis for distinguishing between the two options.

Figure 5, however, gives a significantly different picture. The distributions for both options lie largely in negative territory. Thus, decision makers can be more confident that these policies will have some beneficial impact on phosphorus loading. This is particularly the case for the wastewater treatment option, which has a virtually negligible risk of non-reductions. Thus, if knowledge uncertainty is the dominant source of uncertainty in the predictions shown in Figure 5, and we assume that public decision-makers would prefer to avoid the risk of an ineffective policy, then the wastewater treatment option should be preferred to the agricultural policy.

If, on the other hand, natural variability is the dominant cause of uncertainty, then both options would be similarly attractive for a system for which the long term load is of primary interest. An increase in the load for one year would then be compensated by a higher decrease in another other year leading to the same average load over many years. Confidence in the mean value is then more important than the width of the distribution.

It should be noted that the narrowness of the distributions shown in Fig. 5 does not imply that the effects can be easily detected upon implementation of the policy. The difference between any two alternatives in the same year is not a measurable quantity (a policy is either implemented or not). This means that the effect is covered by the wide distribution of the actual value of thevariable of interest, even when the distribution of the difference relative to the baseline scenario is very narrow.

Of course, for our simple example, the results may be rather obvious. Clearly, reductions in wastewater phosphorus inputs will lead to a reduction in total loads under nearly all conceivable situations, while reductions in inputs associated with high rainfall events may be less certain, depending on the frequency of those events. Our final results support this intuition, but employing more naïve assumptions (such as those used to generate the results shown in Figure 4) would have obscured this fact. Unfortunately, such assumptions of independence in the sources of uncertainty across scenarios underlie the interpretation of many predictive analyses. The implications of these analyses for decision support may be quite misleading, depending on the criteria used for the decision.

6. CONCLUSIONS

The theoretical argumentation as well as the simple, didactical example demonstrate that it may be worth studying the distribution of the difference in a variable of relevance for different policy alternatives. This is usually not done in decision theory, but may lead to additional insight. Although our example is very simple, the basic technique presented in this paper can easily be applied to more complex models.

REFERENCES

Beck, M.B., Water quality modeling: A review of the analysis of uncertainty, *Water Resources Research*, 23(8), 1393-1442, 1987.

Morgan, M.G. and Dowlatabadi, H. (1996), Learning from integrated assessment of climate change, *Climatic Change* 34, 337-368,

Morgan, M.G. and Henrion, M., *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York, 1990.

Omlin, M., Brun, R. and Reichert, P., Biogeochemical model of lake Zürich: Sensitivity, identifiability and uncertainty analysis, *Ecological Modelling* 141(1-3), 105-123, 2001.

Reckhow, K. H. 1980. An incremental phosphorus loading change approach for prediction error reduction. Pages 325-328 *in* International Symposium on Inland Waters and Lake Restoration. Office of Water Regulations and Standards, USEPA, Portland, ME.

Reckhow, K. H. 1994. Importance of Scientific Uncertainty in Decision-Making. *Environmental Management* 18:161-166.

Reckhow, K.H. and Chapra, S.C., Modeling excessive nutrient loading in the environment, *Environmental Pollution* 100, 197-207, 1999.

Reichert, P. and Vanrolleghem, P., Identifiability and uncertainty analysis of the river water quality model no. 1, *Water Science and Technology*, 43(7), 329-338, 2001.