# Model uncertainty in risk analysis

Scott Ferson[1]
*Centre de Recherches de Royallieu, Université de Technologie de Compiègne*
*F-60205 Compiègne Cedex, France, sandp8@gmail.com*

**Abstract:** How can we project uncertainty about model variable(s) $X$ through a function $f$ to characterize the uncertainty about $Y = f(X)$ when $f$ itself has not been precisely characterized? Although the general problem of how to quantitatively express and project model uncertainty though mathematical calculations in a risk analysis can be addressed by only a few strategies, all of which seem either dubious or quite crude, there are a variety of special cases where methods to handle model uncertainty are rather well developed and available solutions are both comprehensive and subtle. For instance, uncertainty about the shapes of probability distributions can be captured as credal sets or p-boxes. Likewise, uncertainty about the stochastic dependencies between distributions can be projected using Kolmogorov–Fréchet bounding. Numerical experiments suggest that there is also another special case of model uncertainty that can be addressed fairly robustly: when evidence of the statistical relationship between variables has been condensed into regression analyses. In this case, simple computational strategies can be employed to obtain conservative bounds on $Y$.

**Keywords:** model uncertainty; model-form uncertainty; enveloping; bounding; regression; polynomial regression; Bayesian model averaging; Monte Carlo model averaging; sensitivity analysis; risk analysis

## 1. Introduction

In some areas of science, the appropriate model to use for a particular situation is well established. However, in many if not most areas, there is still quite a bit of controversy among scientists about how to best describe the physical interactions in a system. This is especially often the case in new areas of science and in scientific fields where empirical investigation is difficult or expensive. It is also true in domains of study involving performance of new materials or new system designs, and behavior of systems under abnormal or extreme conditions. These domains are very commonly encountered in risk assessments.

The phrase 'model uncertainty' refers to that incertitude about the correct form that the model should take. Are the mathematical expressions right? Are the dependencies and interactions among physical components reasonably and accurately represented? Are the submodels appropriate for the situation and do they mesh together coherently? Have the relevant and essential underlying physical or biological processes been appropriately represented? The model in a risk assessment includes all the structural decisions made by the analyst or modeler that govern how the parameters interact. Each of these decisions is in principle subject to some degree of doubt. Model uncertainty is about whether or not those parameters are combined together in the right way. In most cases, model uncertainty is a form of epistemic uncertainty because we are unsure whether their constructions are reasonable and complete. (It would be aleatory uncertainty only

---

[1] Present affiliation: Applied Biomathematics, Setauket, New York, USA.

if the structure of the governing model were itself to change over time, across space, or among components in some population.)

Model uncertainty has various names in the literature, including model structure uncertainty (Walker et al. 2003), model-form uncertainty (Oberkampf and Roy 2010), and structural uncertainty (IPCC 2007). Model uncertainty is distinguished from parametric uncertainty, which is the uncertainty about the value or values of a particular constant or variable. Risk analysts have many computational tools available to them to assess the consequences of parametric uncertainty. But analyses consist of statements about both parameter values and the model relationships that tie the parameters together. An important source of uncertainty in risk assessments is doubt about the structure of the model or the form of the risk expression. In fact, it is sometimes quantitatively more important than all other sources of uncertainty (e.g., Linkov and Burmistrov 2003; Refsgaard et al. 2006; inter alia). Yet this uncertainty is rarely even acknowledged, much less accounted for in a comprehensive way. Almost all risk analyses and, indeed, statistical analyses in general neglect this source of uncertainty entirely. Ignoring model uncertainty could lead to over-confident inferences and decisions that are more risky than one thinks they are. A risk analyst who constructs a single model for use in an assessment and then uses it to make forecasts is behaving as though the chosen model is actually correct. Draper (1995) argued compellingly that model uncertainty should be taken very seriously in computing forecasts and calculating parameter estimates. But just how should this be done?

Section 2 briefly reviews several general schemes that intend to account for model uncertainty, including sensitivity analysis, model averaging into stochastic mixtures, Bayesian model averaging, and enveloping. Section 3 describes some important special cases where model uncertainty can be comprehensively characterized using the methods of probability bounds analysis. Section 4 focuses on issues that arise when models have been characterized using statistical regression analyses.

## 2. General Approaches

There are four general approaches for handling model uncertainty, which are briefly reviewed in turn.

*Sensitivity analysis.* Perhaps the simplest way to address model uncertainty is to repeat the analysis or calculation for each possible model and present the collection of outputs as the results. This approach has been used by the Intergovernmental Panel on Climate Change in its forecasts about future climate (cf. IPCC 2001). The approach is mathematically trivial and computationally straightforward, and it is easy to explain to others as a 'what if' analysis. Its main drawback is that it has combinatorial complexity. For example, in analyzing climate change scenarios, the number of times the analysis must be repeated might be the number of possible models for carbon emissions multiplied by the number of possible models for population trends, itself multiplied by the number of possible models about mitigation plans. Every issue in a complex assessment might in principle unfold into a variety of possible models, and the number of analyses that must be undertaken is the product of the counts of such models across all model issues in doubt. The computational costs of this approach may be tolerable on modern computers, but the effort needed to summarize and document the range of possible outcomes can quickly become burdensome. The approach is also limited to situations in which the possible models can be comprehensively enumerated by the analyst.

*Monte Carlo model averaging.* A traditional Monte Carlo analysis might handle model uncertainty by creating a new parameter, say *m*, to represent which model to use and varying the model used in simulations according to random values of this value over many replications. This approach is in fact widely used and recommended in practice (e.g., Apostolakis 1995; Morgan and Henrion 1990; cf. Cullen and Frey 1999). For example, if there are two possible models, the *m* parameter would be represented as a Bernoulli random variable taking on both possible values with even probabilities, or perhaps with probabilities determined by the relative likelihoods that either model is the right one. These likelihoods can represent the differential credibilities that the analyst would assign to the two models. Whenever these likelihoods are unknown and the analyst does not think either model is more credible than the other, the traditional approach is to assume both models are equiprobable. If there are several possible models, then the parameter *m* would be a more general discrete variable, whose values would again be equiprobable unless the relative probabilities of the different models were known or assumed by the modeler. Finally, if there are infinitely many models possible, but they can be parameterized in a single-dimensional family, then a continuous version of the parameter *m* can be used. In all cases, values for this variable are randomly generated in the Monte Carlo simulation, and which model is to be used in one of the replications is determined by the random value. Typically, the model selection would happen in the outer loop of a two-dimensional simulation (in which the inner loop simulated variability), but this is not essential. The result of the Monte Carlo simulation depends then on a randomly varying model structure. This approach requires that the analyst know, and be able to enumerate or at least continuously parameterize, all the possible models.

*Bayesian model averaging.* The Bayesian approach to handling model uncertainty, which is called Bayesian model averaging (Raftery et al. 1997; Hoeting et al. 1999), has essential similarities to the Monte Carlo model averaging approach, and it will typically produce similar if not identical results. Until very recently, Bayesian analysts chose a single model and then acted as though it had generated the data. Bayesian model averaging recognizes that conditioning on a single selected model ignores model uncertainty, and therefore can lead to underestimation of uncertainty in forecasts. The Bayesian strategy to overcome the problem involves averaging over all possible models when making inferences about quantities of interest. Draper (1995) suggested employing standard techniques of data analysis, but when a good model is found, embedding it in a richer family of models. By assigning prior probabilities for the parameters of this family of models and treating model selection like other Bayesian parameter estimation problems, this approach produces a weighted average of the predictive distributions from each model, where the weights are given by the posterior probabilities for each model. The prior probabilities assigned to each model can represent the differential credibility an analyst ascribes to it. By averaging over many different competing models, this approach incorporates model uncertainty into conclusions about parameters and predictions. In practice, however, this approach is often not computationally feasible because it can be difficult to enumerate all possible models for problems with a large number of variables. However, a variety of methods for implementing the approach for specific kinds of statistical models have been developed. The approach has been applied to many classes of statistical models including several kinds of regression models (Hoeting et al. 1999).

Bayesian model averaging and the Monte Carlo model averaging strategy are similar in that they both use what is essentially a stochastic mixture of the competing models. Aside from the technical burden of parameterizing the space of possible models and assigning a probability to each, there is a far greater problem with the approach that these strategies use. In representing model uncertainty as a stochastic mixture of the possible models, this approach effectively *averages together incompatible theories* (Finkel

1995). It is equivalent in this respect to the approach to modeling what is fundamentally incertitude as an equiprobable stochastic mixture (the uniform distribution). This approach is due originally to Laplace, but when it is applied in risk analysis to the study of distributions (rather than estimating point values), it can underestimate the true tail risks in an assessment. The potential results are distributions that no theories for any of the models would consider reasonable.

*Enveloping.* A more reasonable strategy might be to use an envelope of the models rather than an average or mixture of models. Because model uncertainty typically has the form of doubt about which of a series of possible models is actually the right one, such an approach would propagate precisely this doubt through subsequent calculations. In this sense an enveloping approach would clearly be more comprehensive than the traditional approach based on model averaging, in the same way an interval bounding a set of real values is a better characterization of the uncertainty of the set than its average is. Averaging is a way to handle uncertainty, but not a way to account for or propagate it. An average erases uncertainty rather than really accounts for it.

Note that an enveloping approach would also be able to handle non-stationarity of distributions, which is another important source of uncertainty that is usually ignored in traditional assessments for lack of a reasonable strategy to address it. Unlike sensitivity analysis and the Monte Carlo and Bayesian model averaging strategies, an enveloping approach can work even if the list of possible models cannot be enumerated or parameterized. So long as we can somehow *bound* the regions in any output or intermediate variables that depend on the choice of the model, we can represent and propagate the uncertainty about the model in a comprehensive way. The next section gives some examples of how such bounding can work to characterize model uncertainty.

The limitations of the enveloping approach are (1) it cannot account for differential model credibilities as the model averaging approaches can, (2) it cannot make use of any available data as Bayesian model averaging does, and (3) it does not account for 'holes' in the set of possible models as model averaging and sensitivity analysis do.

*Numerical example.* Let us contrast the four general approaches to model uncertainty using a small numerical example. Suppose that the function $f$ is one of two possibilities. Either $f(A,B) = f_{Plus}(A,B) = A + B$, or $f(A,B) = f_{Times}(A,B) = A \times B$ is the correct model, but the analyst does not know which it is. Thus the model uncertainty is that we are not sure whether we should be adding or multiplying to combine to two inputs together. To make the example concrete and simple, let $A \sim N(0, 1)$ and $B \sim N(5,1)$, normal random deviates. To see the effect of accounting for prior differential credibilities of the models, let us suppose the analyst thinks that $f_{Plus}$ is twice as likely as $f_{Times}$. To see the effect of Bayesian model averaging using data, we can suppose there is a single observation from the distribution of $f(A,B)$ which is, say, 2.1.

Figure 1 shows the two possible distributions for $f(A,B)$ which are $f_{Plus}(A,B) = N(5, \sqrt{2})$ and $f_{Times}(A,B) \approx N(0, \sqrt{26})$. The Monte Carlo model average, computed as the (vertical) average of the $f_{Plus}$ and $f_{Times}$ functions with weights $p=2/3$ and $(1-p)=1/3$ respectively, is shown as a gray cumulative distribution on the right graph. The Bayesian model average instead uses the weights $pL_P/(pL_P+(1-p)L_T) = 0.489$ and $(1-p)L_T/(pL_P+(1-p)L_T) = 0.511$, where $L_P$ and $L_T$ are the respective likelihoods (probability densities) associated with the datum 2.1 from $f_{Plus}$ and $f_{Times}$. It is shown as the inner black distribution function on the right graph. The sensitivity analysis approach uses the pair $f_{Plus}$ and $f_{Times}$ which are the outer black distributions that the analysis started with. Their envelope is depicted as the shaded region on the right graph.
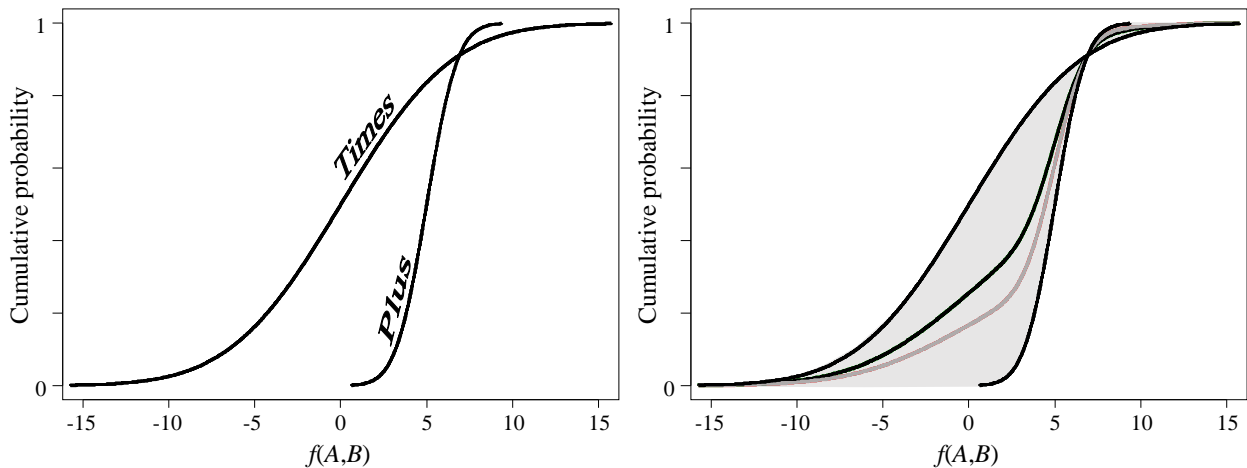
*Figure 1.* Uncertainty about two possible models (Times, Plus in left graph), characterized by enveloping (shaded region), Monte Carlo model averaging (gray curve), Bayesian model averaging (interior black curve), or sensitivity analysis (outer curves).

## 3. Special Cases

How can one bound a class of models without enumerating or parameterizing them? There are several examples of how this can be done. The manifestations of model uncertainty are extremely numerous, but there are some particular forms for which useful bounding approaches have been developed. These include uncertainty about distribution family, intervariable dependence, choice among specific competing theories, and choice among unknown theories when consequences are bounded. This section briefly reviews enveloping strategies for handling these situations.

*Uncertainty about distribution shape.* Model uncertainty about distribution family is the focus motivating the development of both robust Bayes methods (Berger 1984; 1985; Wasserman 1992) and the analytical techniques for probability boxes (Ferson et al. 2003; Berleant 1996). It is fair to say that an analyst using these techniques could conduct a distribution-free risk analysis that makes no assumptions about the families of statistical distributions from which variables are drawn. Obviously *some* assumptions or evidence is necessary (such as constraints on the range or moments, or empirical sampling data), but no assumption about the statistical *family* of distributions need be required. These techniques allow an analyst to fully explore the consequences of model uncertainty about distribution shape, and they therefore represent a scheme for distribution-free risk analysis.

Figure 2 depicts several example p-boxes for different sets of constraints. The ordinate for each graph is cumulative probability from zero to one. The abscissas are the possible values of different random variables. For instance, in the upper, leftmost graph, if an analyst only knows the possible range for a particular random variable, the rectangular p-box which includes all distributions that have supports inside that range can be used to represent the analyst's state of knowledge. If the analyst additionally knows the mean in addition to the range, the p-box will be tighter, as depicted in the leftmost graph on the second row in the figure. Knowing the median instead of the mode implies a still tighter p-box that is pinched to a point

at the mode. The pinching happens because all matching distributions must go through this point at the probability level of 0.5. Various other assumptions about modality, symmetry or dispersion result in the other p-boxes depicted in the figure, which can generally be computed by straightforward constraint analyses from any given set of specifications.
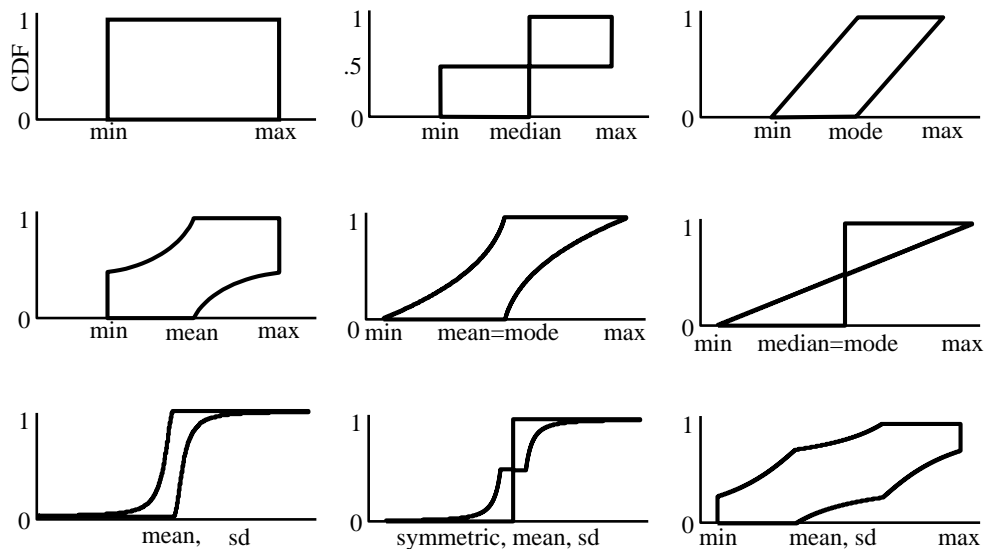


*Figure 2.* Example probability boxes based on constraints about certain statistical characteristics without any assumption about the distribution shape or family.

Notice that none of these p-boxes makes any assumption about the family or shape of the distribution (such as normal, Weibull, etc.). Each depicted p-box encloses *all* distribution functions that have these respective statistical characteristics, so any analysis based on such p-boxes can be comprehensive about the model uncertainty respecting distribution shape per se (Ferson et al. 2003). On the other hand, these p-boxes also enclose distributions that do not meet their respective specifications. For example, consider the leftmost distribution that can be inscribed in the {min, mean, max} p-box on the left graph of the second row. Unless specified otherwise, monotonicity is the only constraint governing distribution functions inside p-boxes. A distribution that traces the left edge of this p-box would clearly not have the prescribed mean. In most applications in risk analysis where analysts want to compute tail risks, the existence of such unmatching distributions in the p-boxes makes no difference at all to the calculation of how likely extreme outcomes might occur. If the analysis does need to exclude unmatching distributions, it may need to be based on robust Bayes methods (Berger 1985) or perhaps credal sets (Levi 1980; Walley 1991) rather than p-boxes, although these structures are usually not nearly as easy to compute with as p-boxes are.

*Uncertainty about dependence.* Techniques in probability bounds analysis also comprehensively account for model uncertainty about the dependencies among random variables. Although independence is widely assumed in risk assessment as a matter of convenience, in fact there is often considerable doubt about the nature of the dependence among variables, and sometimes that uncertainty is close to total in that analysts know essentially nothing about the possible dependencies. Because dependence among variables can very strongly influence the distributions that arise as arithmetic functions of them, ignoring the uncertainty about

dependence can lead to serious misestimation in risk assessments (cf. wishful thinking by Smith et al. 1992), and to underestimated tail risks that claim extreme events are less likely than they really are.

In general, dependence among random variables is captured in a copula which is a real-valued function of two (or more) variables (the inverses of the distribution functions for the random variables). Solving a problem originally posed by Kolmogorov, Makarov (1981), Rüschendorf (1982) and Frank et al. (1987) developed the mathematical tools to compute bounds on sums of random variables about which only marginal information is available. Williamson and Downs (1990) generalized these tools for other arithmetic functions of random variables (i.e., generalized convolutions) and extended them to the case when only bounds on the marginals are known. The bounds are called Kolmogorov-Fréchet bounds because they make no assumption at all about the dependence between the random variables and yet strictly enclose the possible distributions that arise as arithmetic functions of them. The resulting bounds are known to be best possible (Frank et al. 1987; Williamson and Downs 1990) in the sense that they could not be narrower without precluding some possible convolutions under some possible dependence.

Such Kolmogorov-Fréchet bounds are p-boxes as they are bounds on distribution functions (Ferson et al. 2004). This makes them easy to use in risk assessments based on probability bounds analysis. In practice, the effect of relaxing unwarranted independence assumptions all the way to making no assumption about dependence sometimes makes a big difference in the final result of the analysis, but sometimes it makes only a negligible one. The outcomes dependence on the function combining the inputs together, the breadths of uncertainties of the inputs, and the other dependence assumptions in the model.

Figure 3 shows two examples of addition. The ordinate of each graph in the figure is cumulative probability, and the abscissas are values of the indicated random variable. In the left column of graphs, the addends are precise probability distributions. Their sum appears below them as a p-box whose breadth arises from the uncertainty about the dependence between the random variables *a* and *b*. In the right column of graphs, the addends start out as p-boxes, so their sum contains both the parametric uncertainties about the addends as well as the uncertainty about their dependence.

Note that the Kolmogorov-Fréchet bounds on the answers are *not* equivalent to those obtained by merely computing the results under all possible correlation coefficients from −1 to +1; they are in fact often considerably wider. In fact, no sensitivity analysis could possibly comprehensively characterized model uncertainty about the dependence relations between variables. The variety of copulas that define them make such a problem infinite-dimensional which effectively precludes any simulation-based approach. Yet, because they are based on bounding, the Kolmogorov-Fréchet p-boxes are sure to contain the sum (or product or whatever) of the distributions.

*Choice among specific competing theories.* In a variety of cases, there are two or a few models that have appeared in the relevant scientific literature as descriptions of an incompletely understood phenomenon. Toxicologists, for instance, routinely argue about whether the extrapolation of toxicant responses to very low doses should be modeled linearly or nonlinearly. Likewise, ecologists often debate about whether a biological population should be modeled with the Ricker or Beverton−Holt function for density dependence, and about whether trophic interactions should be modeled with the Lotka−Volterra prey-dependent equation or the Arditi−Ginzburg ratio-dependent equation. Climatologists collectively are not sure about whether the increase in anthropogenic greenhouse gases in the atmosphere will cause the earth's climate to get progressively warmer, or whether released fresh water from melting Arctic ice sheets might affect salinity gradients and thereby precipitate a new ice age by disrupting the global ocean currents that

today keep the northern hemisphere warm. Engineers might argue about whether an untested system design will produce a laminar or turbulent flow.
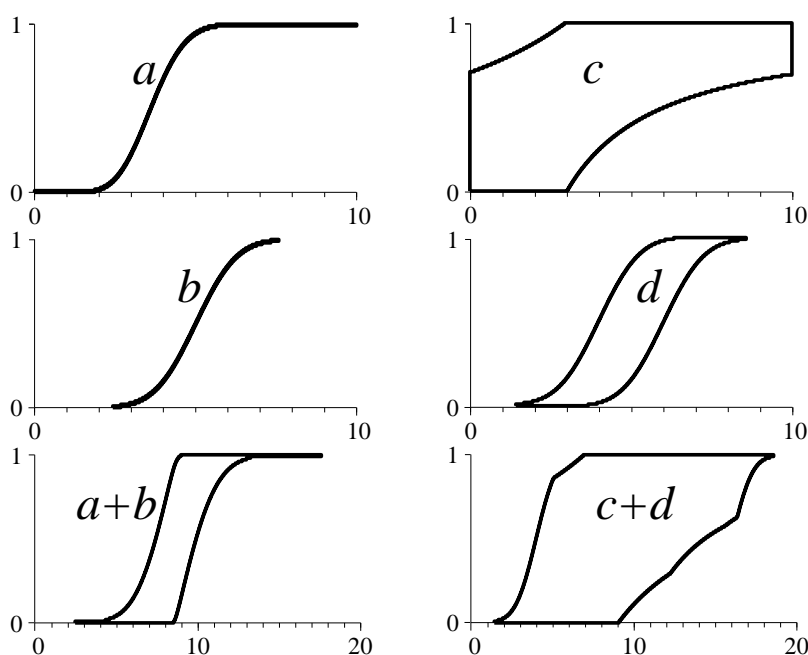


*Figure 3.* Examples of Kolmogorov-Fréchet sum convolutions of distributions (left) and of p-boxes (right).

If model uncertainty can be narrowed to a controversy among a finite list of specific competing theories, then in principle any of the four general methods for handling model uncertainty could be employed. Monte Carlo model averaging or Bayesian model averaging might be reasonable strategies if different mechanisms appear to be acting in different circumstances, so that the system is a stochastic mixture of different outcomes. On the other hand, if there is no particular evidence of actual variability like this, then the preferred strategy would treat the uncertainty as epistemic. In this case, disparate outcomes from the various possible models often cannot sensibly be averaged. What would it mean to average a warmer climate with an ice age? Even if small differences in initial conditions could cause the system to switch between the domains, stochastic mixtures would seem to be an inappropriate characterization of the model uncertainty if the expected eventuality is that the system lands in one domain and not the other. Treating uncertainty in such a system as though it were necessarily variability is to conflate epistemic and aleatory uncertainties. Just as not knowing the value of a quantity does not imply the quantity is varying, not knowing which model is the right one to use does not justify assuming each model governs the system part of the time.

To account for epistemic uncertainty about which of a list of competing models is the correct model, one can instead translate the forecasts from each model into separate distributions of possible model outcomes, just as would be done in a mixture approach, but then form the *envelope* of these distributions. Such envelopes are of course p-boxes, the breadths of which represent model uncertainty. These p-boxes are used as single objects in subsequent calculations so they escape the combinatorial complexity of a sensitivity analysis approach. Note that this approach will yield correct bounds which should still enclose

the true results even if there actually was unrecognized variability that the analyst treated as epistemic uncertainty.

Many analysts reflexively make the criticism that bounding approaches may lead to very wide uncertainties. Surely it is true that bounding can yield large uncertainties, but the width of the uncertainty may be irrelevant if those bounds can be shown to be best possible or are otherwise small enough to lead to useful risk analyses that support effective decision making. In any case, it is better to have a correct analysis that honestly distinguishes between variability and incertitude than an analysis that depends on unjustified assumptions and wishful thinking. A correct analysis does not mistake ignorance on the part of the analyst for variability in a natural system. Of course it is always easy to get tight estimates of uncertainty if we're not constrained to account for what is known and what is not known. If the price of a correct assessment is broad uncertainty as an admission of limitations in our scientific knowledge, then we must pay that price.

Yet certainly not all distributions inside an envelope are necessarily possible. Just as there may be distributions inside a p-box optimally constrained by given statistical characteristics that do not match those characteristics, these model uncertainty envelopes are likely to include outcome distributions that do not arise from any theory considered possible. Indeed, the envelope will include the mixture distribution. Because there are only limited ways to practically exclude distributions from the set represented by a p-box, it may be necessary to use Robust Bayes methods or more generally imprecise probability methods with credal sets rather than p-boxes if it turns out to be important to exclude some distributions from the envelope of possible distribution. This probably will not be a concern if the focus of the analysis is tail risks.

*Choice among an infinite variety of models with boundable outcomes.* Even in situations where we cannot list the possible models because there are infinitely many of them, it may still be possible to bound the consequences of model uncertainty. The methods discussed above that account for uncertainty about distribution shape and intervariable dependence are examples. Another case is a model that determines a probability value. We know that the probability is constrained to the interval [0,1], no matter what the model is. Other nontrivial cases may depend on specific knowledge of the physics or engineering of the system under study. Note that it is often possible to bound the results, even when we cannot form a mixture distribution. In many circumstances where this approach can be used, the models are extreme cases of a continuum that captures the possible ways that a phenomenon could work.

Of course competing models are not necessarily like poles of a continuum if they are the idiosyncratic products from individual scientists or labs, so an arbitrary collection of models cannot generally be claimed to bound possible outcomes. Furthermore, although there are several important forms of model uncertainty that are amenable to a complete assessment by enumeration or bounding, there are, of course, other forms of model uncertainties that remain difficult to address in comprehensive way, such as choice of what parameters to use and choice about the level of abstraction and depth of detail to incorporate into the model. For such uncertainties, the family of possible models may be infinite-dimensional and the analyst may lack any natural way to bound the parameters that depend on the model selection.

## 4. Regression Analysis

The general approaches to handle model uncertainty considered in Section 2 are methods of aggregating disparate models any of which might represent the true structural relationships that govern the modeled system. This section considers model uncertainty in the context of regression analyses in which there are not obvious competing models that could be aggregated (Hoeting et al. 1999; Loquin and Dubois 2010).

Risk analysts often must cobble together assessments consisting of several models, typically originally developed by others. This effort sometimes involves making use of previously published statistical models that relate variables of interest in the assessment. In such cases, it is common for the risk analyst not to have access to the original data, but only to a summary regression analysis that made it into the published paper. How should risk analysts use such regressions in their assessments?

Consider the simplest case of a univariate least-squares linear regression which is based on a small sample size. The traditional regression approach assumes that the form of the function is well known—in this case, a straight line—but that its parameters are not known and need to be estimated from data. Suppose that there are twenty sample $X,Y$-points, as depicted in Figure 4. We could subject those points to regression analysis and obtain a best-fit line, as also depicted. We could pretend that this line is a perfect description of the relationship between $X$ and $Y$. We might therefore use it to translate every possible value of $X$ into the corresponding value of $Y$. The risk assessment then only requires projecting the uncertainty about $X$ through the precisely specified regression expression

$$Y = a + bX$$

for each possible value of $X$ where the regression coefficients $a$ and $b$ are real-valued constants fitted to the data. Some analysts in the past have acted as though the regression line in such cases sufficiently captures the relationship between the variables, and have used the linear transformation it represents to predict a value of $Y$ for any value of $X$. Yet we intuitively recognize that even if $X$ were a precise, real value, there should be *some* uncertainty about the value of $Y$ that comes simply from the fact that the linear regression is estimated from a small collection of points scattered loosely around the line.
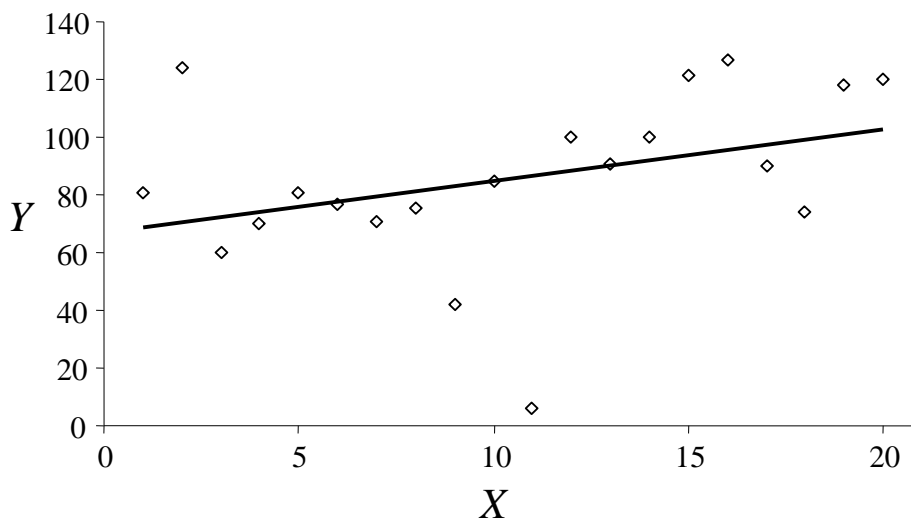


*Figure 4.* A regression line cleaves through the scatter of hypothetical data points (open diamonds).

Some risk analysts (e.g., EPA 2005) have suggested that the uncertainty about the regression should be incorporated into the prediction, so that the uncertainty about $Y$ would include the uncertainty about $X$, plus some additional uncertainty that arises from the regression analysis itself, to take account of the fact that the analysis selected a single regression line to describe the relationship between the variables about which there is some doubt. They suggested projecting the statistical uncertainty about the regression coefficients $a$ and $b$ through the regression equation. The uncertainty about $a$ and $b$ could be characterized as confidence intervals, or better as normal distributions from which those confidence intervals were derived. The calculation would then be

$$Y = \text{N}(a, s_a) + \text{N}(b, s_b) \times X$$

where $s_a$ and $s_b$ are standard errors of the regression coefficients. This entails a multiplicative convolution between the distributions of $b$ and $X$, and another ordinary additive convolution between their product and the distribution of $a$. In practice, assumptions of independence were used between all the inputs in this calculation, even though the sampling distributions of the slope and intercept are not at all independent.

It is easy to demonstrate that the approach that merely projects coefficient uncertainty understates the uncertainty about $Y$. It does not reconstruct the scatter of the original data points on which the regression was based, which is what a risk assessment projection ought to do. Figure 5 illustrates the problem. The open diamonds depict the population from which the observations used to compute the regression were randomly drawn. Many points from this population are displayed here so that their true dispersion is obvious. The filled diamonds are the scatter one gets from projecting the uncertainty about the coefficients. The mismatch is clear and should be significant to risk analyses because this approach grossly underestimates the true scatter. The underestimation actually gets worse as the sample size increases. Notice also that the predicted points seem to be heteroscedastic over different values of $X$ which is inconsistent with the assumptions of linear regression.
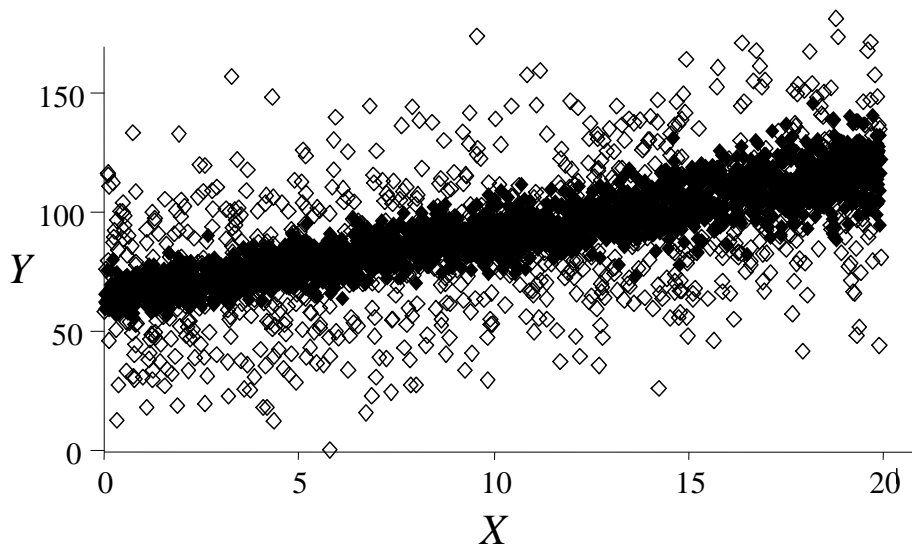


*Figure 5.* Original scatter of data (open diamonds) compared to predictions with only coefficients' uncertainty (filled diamonds).

A much more reasonable way to proceed is to use the actual linear regression equation

$$Y = a + bX + N(0, \sigma),$$

with point values for $a$, $b$ and $\sigma$, and an explicit normal distributed error term, for the predictions. As depicted in Figure 6, this equation *does* reconstruct the scatter originally present in the data used to develop the regression. That is, if we apply it to the original $X$-values used in the regression, we obtain a collection of corresponding $Y$-values that jointly form a scattergram that is statistically similar to the original $X,Y$-data. Instead of two convolutions as in the old approach, this reconstruction entails one convolution, between the uncertain number $a+bX$ (which is just a rescaling of $X$ deviates) and the error term $N(0,\sigma)$. In this case, the independence assumption is clearly justified as it was one of the explicit assumptions in the original regression analysis.
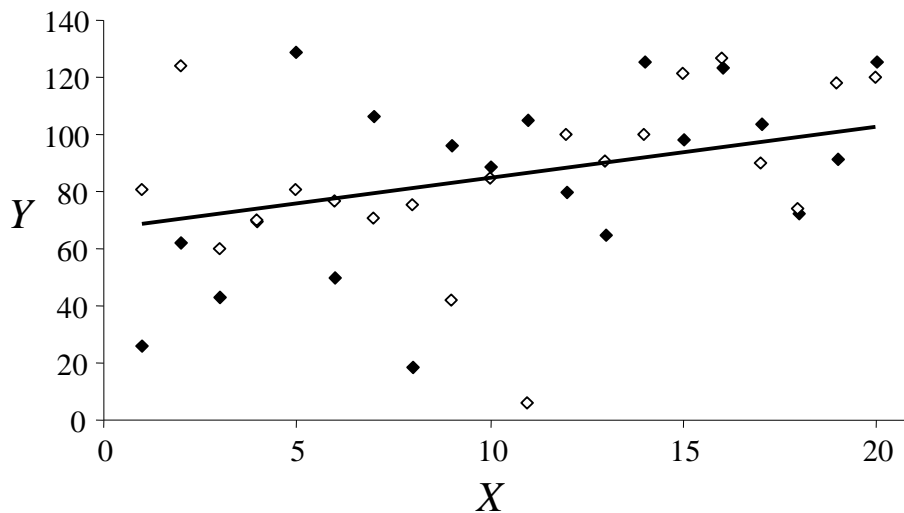


*Figure 6.* Reconstructed scatter of points (filled diamonds) about a regression line through original data points (open diamonds).

Adding the estimated error term is not the only possible approach to reconstruct the scatter of the data informing a regression analysis. Computing a prediction distribution (from which prediction intervals are derived) for each $X$-value would be another method. In fact, prediction distributions can more reliably account for the uncertainty associated with small sample sizes in the original regression. They would also be wider when the $X$-values to be projected are outside of the original range of $X$-values used in the regression so as to account for greater uncertainty attending to extrapolations.

The simple approach of adding back on the estimated error term to regression expressions seems to extend simply to linear multivariate and polynomial regressions. In each case, the regression model expresses the dependent variable as a real-valued function of the independent variable(s) plus a residual error term which is a random distribution. Including the error term would seem likewise to be an essential step whenever surrogate models such as response surface methods are used to simplify calculations. Without the uncertainty analysis enabled by attending to the $\sigma$ error term, which represents the uncertainty that has been left out of the regression model, the use of such reduced models could be highly misleading. The essential parameter $\sigma$, known as the standard error of the regression, is not always reported in a

published summary of a regression analysis. When it is missing, it can be computed from any of the following sets of numbers:

- – regression parameters $a$ and $b$ and their standard errors, and the sample size $n$,
- – the coefficient of determination $R^2$, $a$, $b$, $n$,
- – mean sum of squares explained, $p$ value,
- – mean sum of squares explained, $F$ value,
- – sum of squares unexplained, $n$,
- – mean sum of squares unexplained,
- – standard error of the regression,
- – the six regression quantities ($n$, $\sum X$, $\sum X^2$, $\sum Y$, $\sum Y^2$, $\sum XY$), or
- – the raw data $X_i$ and $Y_i$.

Thus the parameter $\sigma$ will be recoverable from all but the most cursory summaries of a regression analysis.

*Uncertainty about the order of the regression equation.* As already mentioned, traditional regression assumes that the form of the function relating $Y$ to $X$ is well known and only its parameters need to be found. This is arguably the zeroth approach to handling uncertainty about the function, because it totally ignores this uncertainty. Yet there is a question about the regression's choice of a linear model for the functional relationship. Would a better model have been a quadratic rather than a linear model? Cubic? It is well known that regression analysis does not necessarily reveal a model structure that actually reflects how data were generated (Adams 1991). What if we do not know which order equation should have been used in the regression analysis? Could a mistake in the form of the dimensionality or form of the model have a substantial consequence for the results we infer when we use the regression analysis in a risk assessment? In general, how should we handle ignorance about the correct regression model to use (linear, quadratic, cubic, etc.) when we want to project uncertainty in an $X$-variable through a function about which our only information comes from sample data? Can we make conservative or otherwise reasonable estimates given such uncertainty? Given that analysts typically do not know the correct order model to use in a regression analysis, is there a conservative strategy they might use that accounts for this uncertainty? Can such a strategy be applied when the regression analysis is obtained from published summaries and cannot be changed by the analyst?

Simple numerical experiments reveal a conservative strategy for this problem. In these numerical experiments, we assume that $Y$ really is a function of $X$ according to some high-order model

$$Y \sim a + bX + cX^2 + dX^3 + ... + N(0, \varepsilon)$$

where $a, b, c, d, ...$ are particular real numbers and $N(0, \varepsilon)$ denotes a noise term, i.e., a normally distributed random value with zero mean and some positive real standard deviation $\varepsilon$. Thus, we posit a true polynomial model, about which we know both its order and the precise values of its coefficients, including its scedasticity. We sampled $n$ pairs $(X,Y)$ from this hypothetical relationship with error. We then apply polynomial regression to those sampled points to construct $n+1$ regression models of orders 0, 1, ..., $n$. The zeroth-order model is a fitted constant, i.e., the average, and the first-order model is the linear model. The $n$th-order model goes through every sample point. These constructed regression models each have estimated epsilon terms expressing residual uncertainties, which get small or zero as the order of the model approaches the number of sample points.

We can now project uncertainty about the $X$ variable, which might be expressed as a probability distribution, through these constructed stochastic models of the function between $X$ and $Y$. Because we know the true model, we can also project the uncertainty about $X$ through the true model directly and

compare the answer it yields (which is the correct characterization of uncertainty about *Y*) to the estimates produced by reconstructions of various orders. Several interesting and practically important observations arise from these numerical simulations: (1) models of *all* orders yield conservative characterizations of the variance of *Y*, (2) models of *all* orders yield reasonably conservative characterizations of the tail risks of *Y*, and (3) the envelope of resulting predicted *Y* distributions expressed as a p-box is conservative in all respects (i.e., it is effectively sure to enclose the real distribution of *Y*). This p-box seems to represent the model uncertainty induced in *Y* owing to the underlying uncertainty about *X and* the model uncertainty about which degree polynomial is appropriate, contingent on the presumption that a polynomial model of some order is appropriate. These observations suggest a very simple and inexpensive strategy for computing conservative bounds on *Y*. Moreover, these expressions of uncertainty do not seem overly conservative. It does not matter which order a prior regression analysis may have used. It is possible to obtain appropriately conservative estimates of tail risks for $Y = f(X)$ whatever model was used.

*Other causes of model uncertainty in regressions.* There are, of course, a host of sundry reasons that the function embodying a model might be uncertain, including a lack of mechanistic knowledge about the process or system being modeled, small sample size, stochastic variation or imprecision in the data to which the function is fitted, or a mismatch between the model's predictions and relevant empirical evidence. This section has already discussed sample-size-invariant strategies to redress the effect of stochastic variation in the relationship being modeled, and at least one form of deficiency of mechanistic knowledge relating to whether it should be linear, quadratic, cubic, or some other shape. But there may also be uncertainty about a model because the model is fitted to or otherwise based on data that themselves have substantial imprecision. Such imprecision can come from poor measurements, censoring or missing values. This can occur in regression analyses, but it can also appear in likelihood functions in Bayesian inferences. Data imprecision is very often neglected in statistical analysis, but sometimes it is so large that it cannot be reasonably ignored. We have discussed elsewhere bounding strategies to account for imprecision in data that has the form of intervals (Ferson et al. 2007). There may also be model uncertainty implied by a mismatch between a model's predictions and available empirical observations from validation or calibration studies. It seems clear that the degree to which data contradicts a model's predictions should be used to characterize the model's predictive capability, but exactly how to define this connection is the subject of on-going discussions in engineering (Oberkampf and Roy 2010).

## 5.  Conclusions

The central elementary question in the consideration of how to account for model uncertainty is how uncertainty about a (possibly multivariate) variable *X* can be projected though an imperfectly known function *f* to obtain a reasonable characterization of $Y = f(X)$. The uncertainty about *X* might include either aleatory or epistemic uncertainty or both, and it might be expressed simply as a collection of real values or a precise probability distribution, or perhaps as an interval, p-box, or some more complex uncertainty structure (Ferson et al. 2003). The form of *Y* will depend on that of *X* and the definition of the function *f*.

There are four general strategies to handle model uncertainty when it is manifested as a series of competing models. The most flexible is a sensitivity or 'what if' analysis that simply keeps track of the predictions made by all of the possible models. Although this approach is straightforward and does not

conflate different kinds of uncertainties, it can only be applied when the list of possible models can be explicitly and finitely enumerated, and it has combinatorial complexity as the number of submodels in an assessment grows.

Monte Carlo model averaging and Bayesian model averaging are similar approaches in that they both produce a synthesis of the competing models by forming a stochastic mixture of the probability distributions of their outcome variables. These approaches are convenient because they condense the bestiary of possible models to a single probability distribution. They can take account of differential credibilities an analyst might ascribe to the different models, and Bayesian model averaging can even use validation data to improve the specification of the mixture distribution. A limitation of these approaches is that the competing models must be explicitly enumerated. A more serious disadvantage of the approaches is that they average together incompatible theories producing outcome distributions that are supported by no theory. This means that they can often potentially underestimate the probabilities of tail events which are usually the focus of risk assessments.

Enveloping outcome distributions to form p-boxes is the fourth general approach to handling model uncertainty. Like the model averaging approaches, the enveloping approach yields a single object for computation so it does not suffer combinatorial complexity, and it does not conflate the different kinds of uncertainties, or understate tail risks. However, it cannot presently account for relevant validation data or differential credibility of the different competing models. The enveloping approach can comprehensively address model uncertainty if its effects on variables in an assessment can be bounded somehow. It offers particularly convenient solutions for several special cases of particular importance in risk analysis, including the problem of not being sure about the shape or family of a probability distribution, and not knowing the statistical dependence between stochastic variables.

Guidance about how risk analysts should address model uncertainty depends on whether it is possible to enumerate and spell out what all the competing models are. When the models can be enumerated, sensitivity analysis is preferable for its simplicity when it can handle all the questions that arise, but it may not be feasible in big or complex problems. Monte Carlo averaging, although widely employed, seems to have little justification, and is, at best, an ad hoc method. When data are abundant, a Bayesian approach may be best, but otherwise, it may just be wishful thinking. Enveloping is always reliable, but may yield results that are too uncertain. When the models cannot be enumerated, enveloping may be the only tenable general strategy.

Because of the interdisciplinary nature of their charge, risk analysts often must cobble together assessments by assembling several models together. They commonly use previously published statistical regression models that relate variables of interest in the assessment. Simple and inexpensive computational strategies enable risk analysts to make use of summary regression analyses that are common in the literature without requiring any access to their original data. These strategies can be used to yield conservative characterizations with respect to output variances and even tail risks.

In many respects, accounting for model uncertainty depends on an analyst being able to describe what models are possible. It appears that no systematic approach to model uncertainly will be helpful if one has no idea how a system works, or is unaware of the important processes and mechanisms that govern it. In some risk assessment problems such as those focused on abstract notions like 'ecosystem health', it may not even be clear what measurable and latent variables would be relevant to the analysis. In these cases, the need is not so much for ways to account for model uncertainty, but rather for exploratory tools for theory building and hypothesis generation.

## Acknowledgments

## References

Adams, J.L. (1991). A computer experiment to evaluate regression strategies. Proceedings of the American Statistical Association Section on Statistical Computing, 55−62.

Apostolakis, G.E. (1995). A commentary on model uncertainty. *Proceedings of Workshop on Model Uncertainty*, A. Mosleh, N. Siu, C. Smidts and C. Lui (eds.), Center for Reliability Engineering, University of Maryland, College Park, Maryland.

Arendt, J.S., and D.K. Lorenzo (2000). Evaluating process safety in the chemical industry: a user's guide to quantitative risk analysis. American Chemistry Council, Arlington VA.

Berger, J.O. (1984). The robust Bayesian viewpoint (with discussion). In J. B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63−144. North-Holland, Amsterdam.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Berleant, D. (1996). Automatically verified arithmetic on probability distributions and intervals. *Applications of Interval Computations*, B. Kearfott and V. Kreinovich (eds.), Kluwer Academic Publishers, pages 227−244.

Cullen, A.C. and H.C. Frey (1999). *Probabilistic Techniques in Exposure Assessment*. Plenum Press.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B* 57: 45−97.

EPA [U.S. Environmental Protection Agency] (2005). GE/Housatonic River Site in New England, http://www.epa.gov/ne/ge/.

Ferson, S., V. Kreinovich, L. Ginzburg, K. Sentz and D.S. Myers (2003). Constructing probability boxes and Dempster−Shafer structures. SAND2002-4015, Sandia National Laboratories, Albuquerque, New Mexico. http://www.ramas.com/unabridged.zip

Ferson, S., R. Nelsen, J. Hajagos, D. Berleant, J. Zhang, W.T. Tucker, L. Ginzburg and W.L. Oberkampf (2004). *Dependence in Probabilistic Modeling, Dempster−Shafer Theory, and Probability Bounds Analysis*. SAND2004-3072, Sandia National Laboratories, Albuquerque, New Mexico. www.ramas.com/depend.pdf

Ferson, S., V. Kreinovich, J. Hajagos, W.L. Oberkampf and L. Ginzburg (2007). *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. SAND2007-0939, Sandia National Laboratories, Albuquerque, New Mexico. http://www.ramas.com/intstats.pdf

Finkel, A.M. (1995). A second opinion on an environmental misdiagnosis: the risky prescriptions of *Breaking the Vicious Circle*. *New York University Environmental Law Journal* 3: 295−381.

Frank M.J., R.B Nelsen, B. Schweizer (1987). Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* 74: 199−211.

Gray, G.M., and J.D. Graham (1991). Risk assessment and clear air policy. *Journal of Policy Analysis and Management* 10: 286−295.

Hansen, L.P., and T.J. Sargent (2001). Robust control and model uncertainty. *American Economic Review* 91: 60-66.https://files.nyu.edu/ts43/public/research/.svn/text-base/aea06.pdf.svn-base

Hoeting, J., D. Madigan, A. Raftery and C. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical Science* 14: 382−417.

Model uncertainty in risk analysis

IPCC [Intergovernmental Panel on Climate Change] (2001). Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, http://www.ipcc-nggip.iges.or.jp/public/gp/english/, especially Annex 1 "Conceptual Basis for Uncertainty Analysis", http://www.ipcc-nggip.iges.or.jp/public/gp/english/A1_Conceptual.pdf.

IPCC [Intergovernmental Panel on Climate Change] (2007). Guidance notes for lead authors of the IPCC fourth assessment report on addressing uncertainties, http://www.ipcc.ch/meetings/ar4-workshops-express-meetings/uncertainty-guidance-note.pdf.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts.

Linkov, I., and D. Burmistrov (2003). Model uncertainty and choices made by modelers: lessons learned from the International Atomic Energy Agency model intercomparisons. *Risk Analysis* 23: 1297−1308. http://www.blackwell-synergy.com/toc/RISK/23/6

Loquin, K., and D. Dubois (2010). Kriging and epistemic uncertainty: a critical discussion. Pages 269−305 in *Methods for Handling Imperfect Spatial Information*, Studies in Fuzziness and Soft Computing, volume 256, edited by R. Jeansoulin, O. Papini, H. Prade and S. Schockaert, Springer. http://www.irit.fr/~Didier.Dubois/Papers1208/fuzzy_Kriging-livre1.pdf

Makarov, G.D. (1981). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and Its Applications* 26: 803–806.

Miura, J., and Y. Shirai (2000). Modeling motion uncertainty of moving obstacles for robot motion planning. Pages 2258−2263 in Proceedings of the IEEE International Conference on Robotics and Automation, San Francisco, California. http://ir.library.osaka-u.ac.jp/metadb/up/LIBTEK001/00846363.pdf

Morgan, M.G. and M. Henrion (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge.

Nassreddine, G., F. Abdallah, and T. Denoeux (2009). A state estimation method for multiple model systems using belief function theory. *Proceedings of the 12th International Conference on Information Fusion, Seattle, WA, USA, July 6−9, 2009*. http://isif.org/fusion/proceedings/fusion09CD/data/papers/0142.pdf

NRC [National Research Council, Committee on Risk Assessment of Hazardous Air Pollutants] (1994). *Science and Judgment in Risk Assessment*. The National Academies Press, Washington, DC.

Oberkampf, W.L., and C.J. Roy (2010). *Verification and Validation in Scientific Computing*. Cambridge: Cambridge University Press.

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety* 91: 1290–1300. http://www.tonyohagan.co.uk/academic/pdf/BACCO-tutorial.pdf

Raftery, A.E., D. Madigan, and J.A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92: 179−191.

Refsgaard, J.C., J.P. van der Sluijs, J. Brown, P. van der Keur (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* 29: 1586–1597.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability* 14: 623–632.

Smith, A.E., P.B. Ryan and J.S. Evans. (1992). The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Analysis* 12: 467–474.

Walker, W.E., P. Harremoes, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, M.P. Krayer von Krauss (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* 4: 5–17.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Wasserman, L. A. (1992). Recent methodological advances in robust Bayesian inference (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 483−502. Oxford University Press, Oxford.

Williamson, R.C., and T. Downs (1990). Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4: 89–158.