Statistics of Past Errors as a Source of Safety Factors for Current Models

Alexander I. Shlyakhter

Department of Physics and Northeast Regional Center for Global Environmental Change
Harvard University, Cambridge, MA 02138 USA

Abstract

Results of a comparative analysis of actual vs. estimated uncertainty in several data sets derived from natural and social sciences are presented. Data sets include: i) time trends in the sequential measurements of the same physical quantity; ii) environmental measurements of uranium in soil; iii) national population projections; iv) projections for the United States' energy sector. Probabilities of large deviations from the true values are parametrized by an exponential distribution with the slope determined by the data. One can hedge against unsuspected uncertainties by inflating reported uncertainty range by a default safety factor determined from the relevant historical data sets. This empirical approach can be used in the uncertainty analysis of the low probability/high consequence events, such as risk of global warming.

1. INTRODUCTION

Science policy often hinges on reliable assessment of the uncertainty in predictions derived from various models. Analysis of events with low probability but high consequences (such as probability of extreme global warming) is crucial for decision making. However it is well known that there is a strong tendency for researchers to underestimate uncertainties in results, thus decreasing their reliability and increasing the probability of "surprises".[1-6] The history of natural and social sciences contains a wealth of data about reliability of models and estimates of parametric uncertainty.

Empirical methods of building confidence intervals around point estimates are used in the weather, population, and economic forecasting.[7-11] They rely on the assumption that the distribution of errors in future forecasts is the same as the distribution of these errors in the past forecasts. In many other fields, however (such as physical measurements and environmental risk assessment), upper confidence limits for health risks are often calculated theoretically without empirical validation. In this paper, I present the results of a systematic analysis of actual errors in the datasets derived from nuclear physics, environmental measurements, energy and population projections.[12-19] The goal is to show how historic data on past overconfidence can be used to develop the empirical safety factors that can be applied to uncertainty estimates in current models.

Error in a physical measurement is an additive function of three types of errors: random errors, systematic errors, and blunders. Random errors come from statistical fluctuations of the mean values obtained with a finite number of trials. Systematic errors are errors that have the same sign in different trials, for example, errors caused by unknown thermal expansion of a measuring rod. Blunders are gross errors usually caused by trivial mistakes (such as errors in entering the data). Routine scientific data sets contain 5-10% of blunders.[25]

If the total uncertainty is dominated by random errors, then by the Central Limit Theorem (CLT) the distribution of the arithmetic mean of many observations around the true value is asymptotically normal. Uncertainty is usually reported as an average of many trial measurements (corrected for all *recognized* systematic effects), A, and an associated standard deviation, Δ. If the actual value of this quantity is a then the normalized deviation $x = (a - A)/\Delta$ follows the standard normal distribution. In that distribution, the range $A \pm 1.96\Delta$ has a 95 percent probability of including a.

The presence of systematic errors violates the assumptions necessary for use of the CLT. If most of the uncertainty comes from systematic errors, the usual justification for normal distribution does not apply. Despite this fact, the normal distribution remains implicit when researchers report measured values and the corresponding uncertainties.

In Sections 2 and 3, I present the distribution of the actual errors in several datasets derived from physical and environmental measurements. In Sections 4, 5, and 6, I present the methodology and the results of the analysis of the distribution of errors in population and energy projections. In all cases the observed distributions of errors have long tails that can be approximately described by the exponential distribution.

In Section 7, I show how underestimation of uncertainty can be quantified by the ratio of the total uncertainty to the reported standard deviation, Δ. Assuming that this ratio is a normally distributed random variable with the standard deviation u, I derive a compound distribution which is normal for $u = 0$ and is exponential for $u \geq 1$. In Section 8, I show how one can use the exponential distribution to hedge against unsuspected uncertainties in estimating the risk of global warming.

2. TRENDS IN PHYSICAL MEASUREMENTS

The first attempts to quantify overconfidence in physical measurements come from the work of Bukhvostov[20] and Henrion and Fischoff.[4] They compared elementary particle properties and fundamental constants in early compilations with much more accurate values taken from a more recent compilation. A convenient measure of the deviation of "new" values from the "old" values is the normalized deviation $x = (a - A)/\Delta$, with a the new value, A the old value, and Δ the old standard deviation.

Shlyakhter *et al.*[12-14] extended these original studies by following trends in data sets derived from nuclear and particle physics: masses and lifetimes of elementary particles, magnetic moments and lifetimes of the excited nuclear states, and neutron scattering lengths (see Figure 1). All data sets were first converted into a standard format. Successive measurements of the same quantity comprised a block of data; a data set typically consisted of several hundred such blocks. In order to limit the effects of "noise" in the data on final results, two selection criteria were applied: i) new stated uncertainty had to be much smaller than the old one: $\Delta_{old}/\Delta_{new} \geq 4$; and ii) only cases where deviation from the true value did not exceed ten standard deviations were included in the analysis (in this way most blunders were excluded). Criterion (i) ensures that the new measured value, a, is close to the unknown true value. Neglecting Δ_{new} introduces a small bias into the calculated x values. However, since the errors are combined in quadrature, this bias is only about 3%.

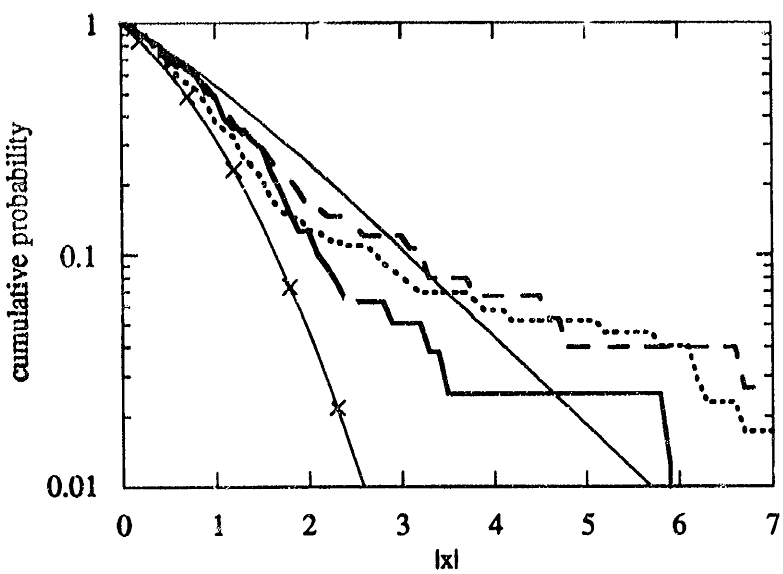

Fig. 1. Probability of unexpected results in physical measurements. The plots show the cumulative probability, $S(x) = \int_x^\infty p(t)dt$ that new measurements (a) will be at least |x| standard deviations (Δ) away from the old results (A); $x = (a - A)/\Delta$ as defined in the text. The cumulative probability distributions of |x| are shown for the three data sets: particle data (elementary particles[21]; heavy solid line); magnetic moments of excited nuclear states[22] (dotted line); neutron scattering lengths[23] (heavy dashed line). Also plotted is cumulative normal distribution, $\mathrm{erfc}(x/\sqrt{2})$, (thin solid line with markers), and compound exponential distribution with parameter $u = 1$ from Figure 5 (thin solid line).

The results confirm the earlier findings that a normal distribution grossly underestimates probability of large deviations from the expected values. A new finding is that the *pattern* of overconfidence is similar in different kinds of measurements.

3. TRENDS IN ENVIRONMENTAL MEASUREMENTS

Environmental measurements are rarely repeated with the same samples and it is hard to estimate how widespread are the unaccounted errors in routinely collected data. An opportunity to address this issue is provided by the data on the excess uranium in the soil around the former Feed Materials Production Center at Fernald, Ohio, which had been a key uranium metals fabrication facility for U.S. defense projects until it was closed in 1989.[24] The data set was requalified at the request of the legal counsel to establish its credibility for litigation purposes. About 200 pairs of measurements of the radioactivity of three uranium isotopes (^{234}U, ^{235}U, and ^{238}U) in the same samples have been provided to me by K.A. Stevenson (U.S. Department of Energy Environmental Measurements Laboratory).

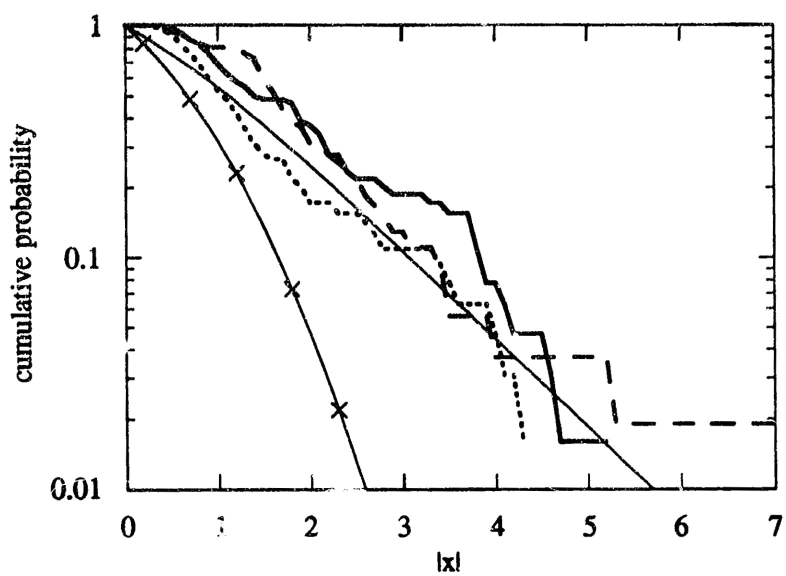

Fig. 2. Distribution of errors in measurement of excess uranium in soil. Presentation is similar to Figure 1. The cumulative probability distributions of |x| are shown for the three uranium isotopes (^{234}U, heavy dotted line; ^{235}U, heavy dashed line; and ^{238}U, heavy solid line). Also plotted are compound exponential distributions from Figure 5: u = 0 (Gaussian, thin solid line with markers) and *u* = 1 (thin solid line).

The distribution of deviations between the original and requalified measurements normalized by the reported standard error of the old measurement is shown in Figure 2. The distribution is similar to the distribution of errors in the measurements in nuclear physics shown in Figure 1. This similarity indicates that the fraction of unsuspected uncertainties may be similar in a wide variety of routine measurements.

4. UNCERTAINTY IN FUTURE PROJECTIONS

Uncertainty in future projections is defined less formally than uncertainty in physical measurements. In this section an algorithm for analysis of uncertainty in historical projections is presented. One can estimate the standard deviation Δ of an equivalent normal distribution and then draw the empirical probability distributions of the deviations of the old projections from the true values normalized by Δ. Experts may not necessarily imply the normally distributed error terms. However, the users of the results tend to base their decisions on the assumption that deviations exceeding several uncertainty ranges are improbable. Comparison of errors in historical data sets with those predicted by the normal distribution provides a useful measure of the credibility of current uncertainty estimates.

Uncertainty in the projections is usually presented in the form of "reference," "lower" and "upper" estimates (R, L, and U respectively) that are obtained by running a model with different sets of exogenous parameters (e.g. the annual rate of growth). The range of scatter around the reference value R does not formally define a Gaussian standard deviation because the fundamental uncertainties involved (e.g. the rate of future economic growth) are frequently not stochastic. However, it is reasonable to assume that the range of parameter variation presented by a forecaster represents a subjective judgment about the probability that the true value $T \in [L,U]$. Generally, lower and upper bounds present what is believed to be an "envelope" most likely to bracket the true value and include the majority of possible outcomes.

Note that bounded distributions (such as the triangular distribution) which are sometimes used to describe the probabilities of various scenarios assign zero probability to the outliers. This incorrectly implies that L and U are real bounds rather than the estimates of plausible range. Historical data presented below, however, suggests that deviations exceeding the expected uncertainty range are not uncommon. Therefore, using a normal (unbounded) distribution as a frame of reference *underestimates* true overconfidence.

The standard deviation of the equivalent normal distribution is calculated as follows:

a) Specify the subjective probability α that the true value will lie between the low (L) and high (U) estimates. I assume $\alpha = 68\%$; larger values of α increase the discrepancy between the Gaussian model and that calculated by this method.

b) Draw an equivalent normal distribution that would have a specified cumulative probability α between L and U. For $\alpha = 68\%$ the standard deviation of the equivalent normal distribution is $(U - L)/2$ so that $x = 2\cdot(T - R)/(U - L)$. Therefore this choice of α corresponds to the usual practice of splitting the uncertainty range in half and using it as a surrogate for the standard deviation.

c) If the reference value (R) is not in the middle of the (L,U) interval use two separate normal distributions truncated at zero: "left half" for (L,R) interval and "right half" for (R,U) interval each having $\alpha/2$ as the cumulative probability.

5. POPULATION PROJECTIONS

The history of population projections provides an opportunity to test the reliability of uncertainty estimates in demographic models. Shlyakhter and Kammen[12-14] analyzed United Nations population projections, made in 1972, for the year 1985 that can serve as the set of "exact" values, a.

The population data base includes projections from 164 nations with population exceeding 100,000 presented in the form of "high" and "medium" and "low" variants for each nation.[26] Data for 31 countries were excluded due to extreme errors resulting for example from unanticipated international migration; such cases can be considered as the use of a wrong model. Data for 133 nations satisfying the criteria $|x| < 10$ were included in the analysis.

The results are shown in Figure 3. Because all the population estimates come from an authoritative source, namely the United Nations, it might be expected that systematic errors would be small, representing a well-calibrated model. The unsuspected uncertainty, however is very large. Data for 37 industrialized countries (where data are generally more reliable) show little less surprise, but probability of large errors is still grossly underestimated by the normal distribution.

6. ENERGY PROJECTIONS

Forecasts of future energy consumption are a prerequisite for many major economic and policy decisions such as how best to reduce carbon dioxide emissions to alleviate global warming, or how best to stimulate the pace of development of alternate sources of energy. An analysis of credibility of uncertainty estimates was performed in Refs. 16 and 17 using the largest coherent set of US energy projections for the year 1990, the Annual Energy Outlook (AEO).[27]

AEO projections for 1990 made in 1983, 1985, and 1987 consist of 182, 185, and 177 energy producing or consuming sectors of the U.S. economy respectively. The variation in the number of sectors resulted because the low and high projections coincided in some cases, and no corresponding uncertainty range could be derived.

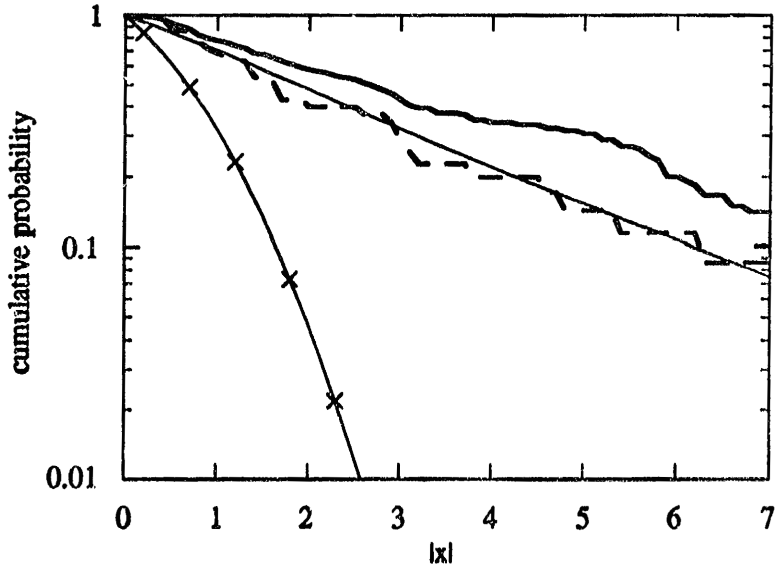

Fig. 3. Population projections. The plots depict the cumulative probability, $S(x) = \int_x^\infty p(t)dt$, that true values ($T$) will be at least $|x|$ standard deviations (Δ) away from the reference value of old projections (R). The population data base is described in the text. The cumulative probability distributions of $|x|$ are shown for the total dataset of 133 countries (heavy solid line) and for a subset of 37 industrialized countries (heavy dashed line). Also shown is the normal distribution (thin solid line with markers) and the compound distribution with $u = 3$ from Figure 5 (thin solid line).

In 47, 50, and 47 cases respectively, the values of $|x|$ (calculated as described in Section 3) exceeded 100; such cases were omitted as they were outside any reasonable range of parametric uncertainty of the AEO model. For all remaining cases the x values were calculated and the frequency distributions analyzed. The distribution of signed x values is approximately symmetric with respect to zero. There is no large systematic bias (e.g. a gross underestimation of energy consumption in all or many sectors) and no strong trends in the scattergrams of x values; this indicates that the projections are generally independent.

Figure 4 shows the cumulative probability distributions of $|x|$ for the projections made for 1990 in 1983, 1985, and 1987 together with the Gaussian and exponential distributions. The three empirical distributions are strikingly similar. Although the absolute error in projections made in 1987 for 1990 is somewhat smaller than made in 1983 for 1990, the range of uncertainty is also smaller so that probability of "large" deviations relative to the observed uncertainty is roughly the same as for the other two years. One would expect that energy projections for aggregated sectors of economy would be more reliable than projections for individual sectors. However, this appears not to be the case (Figure 4, heavy dashed line).

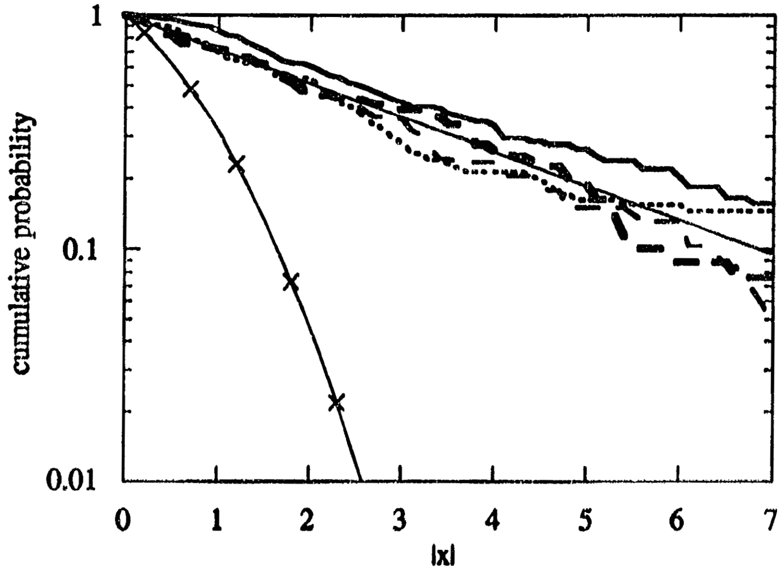

Fig. 4. Annual Energy Outlook projections. The presentation is as in Figure 3: 1983 to 1990 (heavy solid line); 1985 to 1990 (heavy dashed line); 1987 to 1990 (heavy dotted line), aggregated sectors of economy (very heavy dashed line); normal distribution (thin solid line with markers); compound distribution with $u = 3.4$ (thin solid line).

7. EXPONENTIAL PARAMETRIZATION OF OBSERVED DISTRIBUTIONS

Bukhvostov[20] and Shlyakhter and Kammen[14] suggested simple heuristic arguments to describe how an exponential distribution of errors might arise. Let us assume that the estimate of the mean, A, is unbiased but that the estimate of the true standard deviation, Δ' is randomly biased with a distribution $f(t)$ where $t = \Delta'/\Delta$. Here Δ is the estimated standard deviation. In other words, I assume that the deviations normalized by Δ', $x' = (a - A)/\Delta'$ follow the standard normal distribution while the deviations normalized by Δ, $x = (a - A)/\Delta$, follow a normal distribution with a randomly chosen standard deviation t:

$$p_t(x) = \frac{1}{\sqrt{2\pi}t}\exp\left\{-\frac{x^2}{2t^2}\right\} \qquad (1)$$

Integrating over all values of t gives a compound distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dt \frac{1}{t} f(t) \exp\left\{-\frac{x^2}{2t^2}\right\} \tag{2}$$

If f(t) has a sharp peak near $t = 1$, Eq. (2) reduces to the normal distribution. If f(t) is broad, however, the result is different. For simplicity, let us assume that for large t, f(t) follows the Gaussian distribution with standard deviation u: $f(t) \propto \exp[-t^2/(2u)]$. The main contribution to the integral in Eq.(2) comes from the vicinity of the saddle point where the exponential term reaches a maximum (for $t = t_{max}$: $t^2_{max} = u|x|$). It is straightforward to show that, for large values of x, the probability distribution p(x) is not ***Gaussian*** but ***exponential***: $p(x) \propto \exp(-|x|/u)$. In order to reflect the fact that experts are mostly overconfident ($\Delta' \geq \Delta$), I use a truncated normal form of f(t):

$$\begin{aligned} f(t) &= 0, \quad t \leq 1 \\ f(t) &= \sqrt{\frac{2}{\pi}} \frac{1}{u} \exp\left\{-\frac{(t-1)^2}{2u^2}\right\}, \quad t > 1 \end{aligned} \tag{3}$$

Integrating Eq. (2) with f(t) from Eq. (3) gives the cumulative probability S(x) of deviations exceeding $|x|$:

$$S(x) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{u} \int_1^\infty \exp\left\{-\frac{(t-1)^2}{2u^2}\right\} \operatorname{erfc}\left\{\frac{|x|}{t\sqrt{2}}\right\} dt \tag{4}$$

The normal ($u = 0$) and exponential distributions ($u > 1$) are members of a single-parameter family of curves (Figure 5). For quick estimates for $u \geq 1$, $x \geq 3$, one can use the approximation $e^{-|x|/(0.7u+0.5)}$. In this framework the parametric uncertainty can be quantified by analyzing the record of prior projections and estimating the value of u. Data presented in Figures 1-4 show that $u \approx 1$ for physical and environmental measurements and $u \approx 3$ for population and energy projections.

Parameterization with the one-parameter family of compound exponential distributions described above is not the only one possible. Trying to find the standard deviation that best fits the data while keeping the parametric form of the normal distribution produces a very poor fit. A goodness-of-fit analysis of the data sets presented in Figure 1 (Sakharov, private communication, February 1994) gave a good two-parameter fit when the distributions of negative and positive deviations were analyzed separately (with two different u values). Another good two-parameter fit was obtained with a sum of an inflated Gaussian distribution and a constant background term. Although increasing the number of free parameters improves the quality of fit to the data, it also complicates comparisons between datasets. Safety factors for uncertainty estimates should depend only on the general form of the distribution of errors, and for this purpose a simple one-parameter exponential parameterization is sufficient.

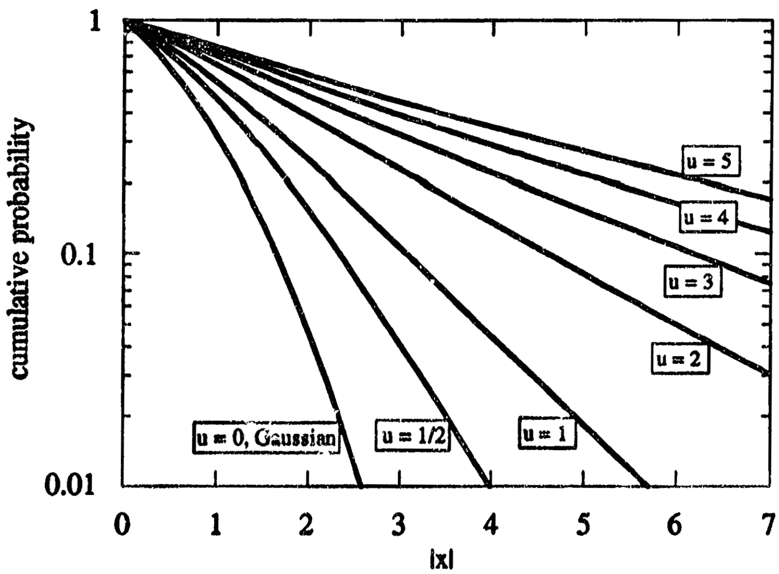

Figure 5. One-parameter family of compound distributions. Parameter u is defined in Eq. (3); it is a measure of uncertainty in the standard deviation Δ). The values of u are indicated in the figure. The curves demonstrate the continuum of probability distributions, from Gaussian (u=0) to exponential ($u > 1$).

There exists a broad spectrum of uncertainties extending from the negligible parametric uncertainties to a crucial uncertainty which of several models to choose. An analyst usually has a scale in mind for "important" uncertainties: smaller uncertainties are assumed to be negligible and are excluded from the detailed analysis. For example, in measurements of the excess uranium in soil, radioactive decay of ^{235}U introduces a negligible error because the characteristic time scale for this decay is billion years. "Important" uncertainties (such as detector efficiency) are carefully evaluated and combined to produce the published estimate of the combined standard uncertainty. However, errors of larger scales, particularly those arising from the use of a wrong model, are also possible. For example an important uncertainty that was overlooked in the old measurements of excess uranium was the uncertainty in the overall chemical yield of the procedure used to leach uranium from the soil.[24] This heuristic model is an attempt to describe the human thought processes that are responsible for the observed pattern of overconfidence. It can be also formulated as a fractal model with the distribution of errors described by Levy stable distribution (a long-tail generalization of the normal distribution).[25]

8. APPLICATION: RISK OF GLOBAL WARMING

Choosing appropriate safety factors as a hedge against unsuspected errors is particularly important in the uncertainty analysis of many situations in public policy. These describe events with low probability but high consequences that are determined by the tails of the probability distributions. I illustrate possible applications using the current debate about possible rise in global temperature in response to the emissions of greenhouse gases. For applications of the inflated confidence intervals to population and energy projections, sea-level rise, estimates of health risks, and epidemiologic studies see Refs. 14-19.

The causal sequence leading to sea-level rise is as follows: population → energy production → CO_2 emissions → greenhouse warming. One can present the temperature rise as a product of four roughly independent factors:

$$h = \text{Population} \cdot \left(\frac{\text{energy}}{\text{person}}\right) \cdot \left(\frac{CO_2}{\text{energy}}\right) \cdot \left(\frac{\Delta T}{CO_2}\right) \tag{5}$$

The first factor is the world population; the second factor is energy production *per capita*; the third factor is CO_2 emissions per unit energy production; the fourth factor is temperature increase ΔT *per* unit rise in CO_2. For each factor there are uncertainies in its respective model. In particular, the last factor includes a key scientific uncertainty of climate models, the global-mean surface temperature increase, ΔT_s, that would follow from a doubling of CO_2 concentrations.

If the Earth was dry, the global-mean surface temperature would increase by $\Delta T_d = 1.2°C$ and this estimate is quite reliable. However, numerous interactive feedbacks from water (most importantly, water vapor, snow-ice albedo, and clouds), introduce considerable uncertainties into ΔT_s estimates. The value of ΔT_s is roughly related to ΔT_d by the formula $\Delta T_s = \Delta T_d/(1 - f)$, where f denotes the sum of all feedbacks. The water vapor feedback is relatively simple: warmer atmosphere contains more water vapor which is itself a greenhouse gas. This results in a positive feedback as an increase in one greenhouse gas, CO_2, induces an increase in another greenhouse gas, water vapor. On the other hand, cloud feedback is the difference between the warming caused by the reduced emission of infrared radiation from the Earth into outer space and by the cooling through reduced absorption of solar radiation. The net effect is determined by cloud amount, altitude, and cloud water content. As a result, the values of ΔT_s from different models vary from $\Delta T_s = 1.9°C$ to $\Delta T_s = 5.2°C$.[28]

Note that two models with comparable ΔT_s values can have different strength of the feedback mechanisms. For example two models (labelled GFDL and GISS) show an unequal temperature increase as clouds are included (from 1.7°C and 2.0°C to 2.0°C and 3.2°C respectively).[28] The effects of ice albedo are different between models but oppositely, so that the results converge (4.0°C versus 4.2°C, respectively). Therefore, agreement between models may be spurious, and both could be wrong.

The causal sequence leading to sea-level rise is as follows: population → energy production → CO_2 emissions → greenhouse warming. One can present the temperature rise as a product of four roughly independent factors:

$$h = \text{Population} \cdot \left(\frac{\text{energy}}{\text{person}}\right) \cdot \left(\frac{CO_2}{\text{energy}}\right) \cdot \left(\frac{\Delta T}{CO_2}\right) \quad (5)$$

The first factor is the world population; the second factor is energy production *per capita*; the third factor is CO_2 emissions per unit energy production; the fourth factor is temperature increase ΔT *per* unit rise in CO_2. For each factor there are uncertainies in its respective model. In particular, the last factor includes a key scientific uncertainty of climate models, the global-mean surface temperature increase, ΔT_8, that would follow from a doubling of CO_2 concentrations.

If the Earth was dry, the global-mean surface temperature would increase by $\Delta T_d = 1.2°C$ and this estimate is quite reliable. However, numerous interactive feedbacks from water (most importantly, water vapor, snow-ice albedo, and clouds), introduce considerable uncertainties into ΔT_8 estimates. The value of ΔT_8 is roughly related to ΔT_d by the formula $\Delta T_8 = \Delta T_d/(1 - f)$, where f denotes the sum of all feedbacks. The water vapor feedback is relatively simple: warmer atmosphere contains more water vapor which is itself a greenhouse gas. This results in a positive feedback as an increase in one greenhouse gas, CO_2, induces an increase in another greenhouse gas, water vapor. On the other hand, cloud feedback is the difference between the warming caused by the reduced emission of infrared radiation from the Earth into outer space and by the cooling through reduced absorption of solar radiation. The net effect is determined by cloud amount, altitude, and cloud water content. As a result, the values of ΔT_8 from different models vary from $\Delta T_8 = 1.9°C$ to $\Delta T_8 = 5.2°C$.[28]

Note that two models with comparable ΔT_8 values can have different strength of the feedback mechanisms. For example two models (labelled GFDL and GISS) show an unequal temperature increase as clouds are included (from 1.7°C and 2.0°C to 2.0°C and 3.2°C respectively).[28] The effects of ice albedo are different between models but oppositely, so that the results converge (4.0°C versus 4.2°C, respectively). Therefore, agreement between models may be spurious, and both could be wrong.

The analyst has thus been placed in a position for which he or she has not been trained, does not desire, and should not be asked to take on. The opinion was expressed that because there is no way to know with certainty whether a regulatory requirement based on a probability has been met, such a regulation might be regarded as meaningless. Ideally, regulations are stated in terms of conditions in which both the regulator and the regulated individual or entity can know whether the regulated individual or entity is in compliance.

regulator ... have ... role to ... the ...tory

One can view the collection of all ΔT_s predictions as a random sample derived from the population of all possible models. We do not know whether current models cover all possible values of ΔT_s. I assume that with probability α the true value is within the range of reported values. Let us assume $\alpha = 99\%$; the standard deviation is then $2 \cdot 2.575$ times less than the reported range. Note that in estimating u values for measurements and projections I used $\alpha = 68\%$ for the reported uncertainty range. Therefore I assume that the collection of current climate models almost certainly covers the true value of ΔT_s. Had I assumed $\alpha = 99\%$ for the old forecasts, the derived standard deviations would be smaller and all x values would be larger. The resulting u values and the corresponding inflation factors would be also larger than the ones I used.

Since ΔT_s is determined by the value of the sum of all feedbacks, f, I assume the normal distribution for f and convert the range of ΔT_s values into the range of f values. For example, $\Delta T_s = 1.9°C$ gives $f = 0.37$ and $\Delta T_s = 5.2°C$ gives $f = 0.77$. This range of f values can be used to estimate the standard deviation of the equivalen normal distribution in the same way as for the population and energy projections above.

The corresponding distribution of ΔT_s values is shown in Figure 6 together with the exponential distribution for $u = 1$ and the distribution of ΔT_s from 21 global circulation models compiled in Ref. 28. By using the exponential distribution with $u = 1$, I assume that the fraction of unsuspected errors in climate models is similar to the fraction of unsuspected errors in physical measurements. With the normal distribution, there is a 1% chance that the true value of ΔT_s exceeds 5°C while with the exponential distribution this same probability corresponds to a catastrophic increase of more than 10°C. In a simple feedback description, $f \approx 1$ would result in a catastrophic runaway warming. Although the true picture will be much more complex, and negative feedbacks will ultimately limit the warming, the possibility of CO_2 atmospheric buildup that could lead to a runaway warming and a switch to a different climate equilibrium must be avoided by all means. In my view, prudent policy decisions should be based on the exponential as the "default" distribution, rather than the normal distribution. Any policy based on a more optimistic view of future temperature rise would have to be justified.

9. SUMMARY

Statistical analysis of the distributions of the actual parametric errors and frequency of model misspecification in natural and social sciences can provide valuable information about the credibility of the estimates of parametric uncertainty in the current models. Data sets for such analysis can be derived, for example, from time trends in sequential measurements of the same physical quantity or the same sample (for parametric uncertainty of the models used in natural sciences), or from the comparison of energy and population projections with actual values that became available later (for models of social parameters). For all data sets analyzed so far, distributions of errors show the same pattern: they have long tails that do not follow a normal distribution but can be pragmatically parametrized by an exponential distribution with the slope determined by the data.

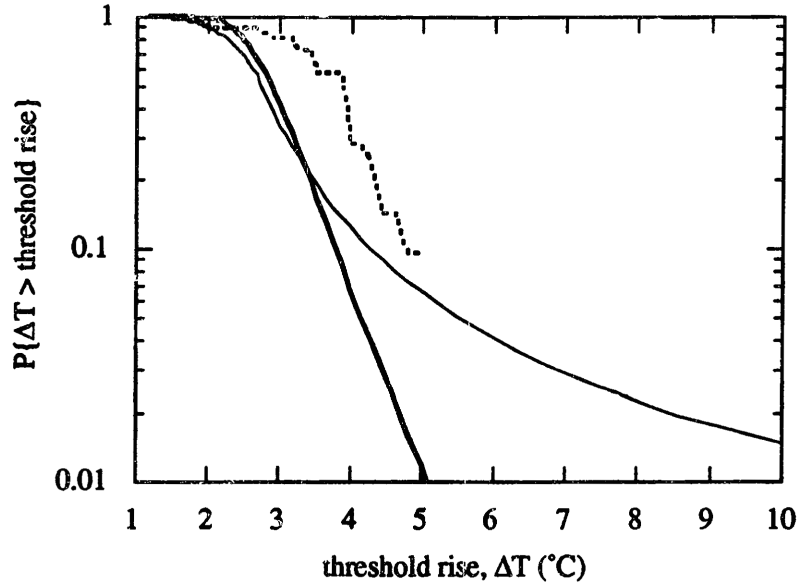

Fig. 6. Estimates of the mean-surface global temperature rise in response to doubling CO_2 concentration. The probability of a temperature rise, ΔT_s, greater than a given threshold is plotted for the normal distribution of uncertainty in the feedback factor *f* (heavy solid line); the exponential distribution with $u = 1$ (thin solid line); the distribution of ΔT_s values from 21 climate models[28] (heavy dotted line).

The additional component of uncertainty derived from such analyses can be viewed as a safety factor accounting for overconfidence of the experts. It therefore incorporates the possibility of human error into the framework of uncertainty analysis. Although data on past misunderstanding of a given situation cannot prevent our current misunderstanding of a significantly different situation, statistical analysis of the frequency of past underestimates of uncertainty can provide useful clues to the choice of the appropriate safety factors.

A legitimate concern about the use of the default inflation factors for the confidence intervals is that this procedure ignores the specifics of particular studies. Some of the studies may be of much higher quality than an average study in the data set from which the inflation coefficient was derived. Unfortunately, elicitation of expert opinions about all significant uncertainties is rarely feasible. The user can hedge against unsuspected uncertainties multiplying the reported uncertainty range by a safety factor. Such factor should be clearly specified and applied only *after* the uncertainty has been determined by a standard method, so that the operation may be easily reversed.[29]

Interestingly, the u values derived from our data sets cover a rather narrow interval: $u \approx 1$ for physical constants and $u \approx 3$ for current models of population growth and energy projections. Although there are many different scientific models with specific sources of uncertainty, it may be possible to combine them in several distinct groups according to reliability of past uncertainty estimates. This would allow to use default inflation factors when no historical data sets are available.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Department of Energy through the Northeast Center of the National Institute for Global Environmental Change (NIGEC) and by the Biological Effects Branch of the U.S. Department of Air Force through the contract F33615-92-C-0602. I am deeply grateful to D. Hattis, P. Stone, and R. Wilson for discussions and comments and to K.A. Stevenson for providing numerical data on excess uranium in the soil at Fernald site.
Electronic mail address: shlyakhter@huhepl.harvard.edu.

REFERENCES

1. L.G. Parrat, *Probability and Experimental Errors in Science*, John Wiley, New York, 1961.

2. J.R. Macdonald, Are the data worth owning?, *Science* , 176:1377 (1972).

3. S. Lichtenstein, B. Fischoff, and L.D. Phillips, Calibration of probabilities: the state of the art to 1980, in *Judgment Under Uncertainty*, D. Kahneman, P. Slovic, and A. Tversky, eds., Cambridge University Press, Cambridge, 1982.

4. M. Henrion and B. Fischoff, Assessing uncertainty in physical constants, *Amer. J. of Physics* , 54:791 (1986).

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

6. R.M.Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, 1991.

7. A.H. Murphy and R.L.Winkler, Reliability of subjective probability forecasts of precipitation and temperature, *J. Royal Stat. Soc.*, Ser.C, 26:41 (1977).

8. W.H. Williams and M.L. Goodman, A simple method for the construction of empirical confidence limits for economic forecasts, *J. Amer. Stat. Assoc.*, 66:752 (1971).

9. M.A. Stoto, The accuracy of population projections, *J. Amer. Stat. Assoc.*, 78:13 (1983).

10. N.W. Keilman, "Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations," Amsterdam, Swets and Zeitlinger, Publications of the Netherlands Interdisciplinary Demographic Institute (NIDI) and the Population and Family Study Centre (CBGS), vol. 20. (1990).

11. V. Zarnowitz, *Business Cycles: Theory, History, Indicators, and Forecasting*, The University of Chicago Press, 1992.

12. A.I. Shlyakhter and D.M. Kammen, Sea-level rise or fall?, *Nature* , 357:25 (1992).

13. A.I. Shlyakhter and D.M. Kammen, Estimating the range of uncertainty in future development from trends in physical constants and predictions of global change, CSIA discussion paper 92-06, Kennedy School of Government, Harvard University, July 1992.

14. A.I. Shlyakhter and D.M. Kammen, Uncertainties in modeling low probability/high consequence events: application to population projections and models of sea-level rise, *Proc. ISUMA'93: 2nd Intl. Symp. Uncertainty Modeling and Analysis*, University of Maryland, College Park, Maryland, April 25-28, IEEE Computer Soc. Press, Los Alamitos, California, 246-253 (1993).

15 A.I. Shlyakhter, et al., Estimating uncertainty in physical measurements and observational studies: lessons from trends in nuclear data, *ibid.* p. 310-317.

16. D.M. Kammen, et al., Non-Gaussian uncertainty distributions: historical trends and forecasts of the United States Energy sector, 1983-2010, *ibid.* p. 112-119.

17. A.I. Shlyakhter, et al., Quantifying the credibility of energy projections from trends in past data: the U.S. energy sector," *Energy Policy*, 22:119 (1994).

18. A.I. Shlyakhter, Improved framework for uncertainty analysis: accounting for unsuspected errors, *Risk Analysis*, in press.

19. A.I. Shlyakhter, Uncertainty estimates in scientific models: lessons from trends in physical measurements, population and energy projections, in *Uncertainty Modeling and Analysis: Theory and Applications*, B.M.Ayuub and M.M.Gupta, eds., North Holland, in press.

20. A.P. Bukhvostov, On the statistical meaning of experimental results, Leningrad Nuclear Physics Institute Preprint, LNPI-45 (1973). (in Russian)

21. LBL (1991) Data file of elementary particle masses and lifetimes maintained by Nuclear Data Center, Lawrence Berkeley National Laboratory.

22. M.P. Avotina, I.A. Kondurov, and O.N. Sbitneva, "Tables Of Nuclear Moments and Deformation Parameters of Atomic Nuclei," Leningrad Nuclear Physics Institute Report, 1982.

23. L. Koester, H. Rauch, and E. Seymann, Neutron scattering lengths: a survey of experimental data and methods, *Atomic Data and Nuclear Data Tables*, 49:65 (1991).

24. K.A. Stevenson and E.P. Hardy, Estimate of excess uranium in surface soil surrounding the feed materials production center using a requalified data base, *Health Physics*, 65:283 (1993).

25. F.R. Hampel, et al., *Robust Statistics: The Approach Based On Influence Functions*, John Wiley and Sons, New York, 1986.

26. United Nations, *World Population Prospects*, Population Studies Paper No. 120 (1991).

27. *Annual Energy Outlook with Projections to 2010*, Energy Information Administration, US Department of Energy, Washington, DC 20585, DOE/EIA-0383, 1992.

28. U. Cubasch and R.D. Cess in "Climate Change", The Intergovernmental Panel on Climate Change (IPCC) Scientific Assessment. Report Prepared for IPCC by Working Group 1. Edited by J.T. Houghton, G.J. Genkins and J.J. Ephraums. Cambridge University Press, 1990, p.75-91.

29. International Organization for Standardization (ISO), "Guide to the Expression of Uncertainty in Measurement," Geneva, Switzerland, 1993.

22. M.P. Avotina, I.A. Kondurov, and O.N. Sbitneva, "Tables Of Nuclear Moments and Deformation Parameters of Atomic Nuclei," Leningrad Nuclear Physics Institute Report, 1982.

23. L. Koester, H. Rauch, and E. Seymann, Neutron scattering lengths: a survey of experimental data and methods, *Atomic Data and Nuclear Data Tables*, 49:65 (1991).

24. K.A. Stevenson and E.P. Hardy, Estimate of excess uranium in surface soil surrounding the feed materials production center using a requalified data base, *Health Physics*, 65:283 (1993).

25. F.R. Hampel, et al., *Robust Statistics: The Approach Based On Influence Functions*, John Wiley and Sons, New York, 1986.

26. United Nations, *World Population Prospects*, Population Studies Paper No. 120 (1991).

27. *Annual Energy Outlook with Projections to 2010*, Energy Information Administration, US Department of Energy, Washington, DC 20585, DOE/EIA-0383, 1992.

28. U. Cubasch and R.D. Cess in "Climate Change", The Intergovernmental Panel on Climate Change (IPCC) Scientific Assessment. Report Prepared for IPCC by Working Group 1. Edited by J.T. Houghton, G.J. Genkins and J.J. Ephraums. Cambridge University Press, 1990, p.75-91.

29. International Organization for Standardization (ISO), "Guide to the Expression of Uncertainty in Measurement," Geneva, Switzerland, 1993.

NUREG/CP-0138

Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis

Model Uncertainty: Its Characterization and Quantification

Held at
Governor Calvert House and Conference Center
Annapolis, Maryland
October 20–22, 1993

Edited by
A. Mosleh, Univ. MD, N. Siu, INEL, C. Smidts, Univ. MD, C. Lui, NRC

Sponsored by
U.S. Nuclear Regulatory Commission
EG&G Idaho, Inc.
University of Maryland

Prepared for
U.S. Nuclear Regulatory Commission

NUREG/CP-0138

Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis

Model Uncertainty: Its Characterization and Quantification

Held at
Governor Calvert House and Conference Center
Annapolis, Maryland
October 20–22, 1993

Edited by
A. Mosleh, Univ. MD, N. Siu, INEL, C. Smidts, Univ. MD, C. Lui, NRC

Sponsored by
U.S. Nuclear Regulatory Commission
EG&G Idaho, Inc.
University of Maryland

Prepared for
U.S. Nuclear Regulatory Commission