UNCERTAINTY IN ORBITAL DEBRIS MEASUREMENTS AND MODELS

M. J. Matney

NASA Johnson Space Center, 2101 NASA Parkway, Mail Code KX, Houston, TX, 77058, USA,
Email: mark.matney-1@nasa.gov

ABSTRACT

The orbital debris environment is described in the form of mathematical models. However, all models and the measurements they are based upon are subject to uncertainty. Measuring instruments never really measure exactly what it is we want to know – they always measure some auxiliary quantity. Knowledge of the limitations of the instruments (measurement uncertainty) is only half the problem. The actual transformation of the data into models adds more uncertainty to the final model result.

In this paper, I provide examples of the basic types of uncertainties that we encounter in making orbital debris measurements. I outline the distinctions between Bayesian and frequentist interpretations of statistics and how this influences the types of conclusions one should or should not draw from uncertainty assessments. I also outline how uncertainties tie into such calculations as probabilistic risk assessments (PRAs), and the benefits and pitfalls of such analyses.

1. INTRODUCTION

The study of orbital debris consists of two primary activities: measurements and modeling. Both activities have limitations due to uncertainties. These activities are intended to give users of the space environment an accurate assessment of the risks to their assets or how much risk their missions pose to other space assets or to people on the ground, and how to plan their missions so as to reduce that risk. They also should help nations and other launching entities know what kinds of actions they can take today to mitigate future growth of orbital debris hazards.

Measurement uncertainties arise for a variety of reasons. All instruments have an intrinsic measurement uncertainty. In addition, an observer never really measures what he wants to measure. For example, a radar operator may want to measure particle size, but in fact measures voltages in the receiver that are interpreted as radar cross section. Measurements near the noise limit of an instrument (such as occurs when measuring small debris) are notorious for introducing uncertainties in measurements. In practice, an observer cannot observe an infinite region for an infinite amount of time, so there will be "sampling error" due to this finite coverage.

Modeling can also introduce uncertainties. Often these are model choices to make the model more tractable on a computer, such as binning or numerical integration. Sometimes simplifying assumptions are made, such as treating objects as if they were in perfect Kepler orbits. Other uncertainties are conceptual, such as projecting unknown launch rates or solar activity into the future.

In addition to these uncertainties, adapting the measurement data for modeling purposes can introduce uncertainty. This is because the models usually need orbit and other debris characterization information that is missing, incomplete, or inferred from measurements.

Over the years, a variety of mathematical tools have been developed to help quantify and work with uncertainties. Many of these are applicable to debris studies.

It is the purpose of this paper to highlight the importance of doing proper uncertainty analysis, especially for the products provided to the spacefaring community. While such analysis is not always easy, accurately understanding the uncertainties in our models and measurements is needed in order to make meaningful decisions regarding the space environment.

2. "TRADITIONAL" METHODS

A significant fraction of statistical work is done under the assumption of normal (Gaussian) distributions. Many practical problems do lend themselves well to normal distributions, and over the years, a number of analytic tools have been developed to work with such distributions. Even the language of statistics (e.g., standard deviation) is often based on analyses using normal distributions.

However, few systems are truly normal, and so the application of Gaussian distributions can introduce unwanted errors. The analysis of many types of problems using normal distributions may actually result in misleading results. These include Poisson problems with sparse data and non-normal noise analysis.

An orbital debris application where normal distributions are used is in collisions avoidance. This is a classic case where the perturbations are small and well-behaved. The uncertainty in the future position of an object in orbit can be defined as a 3-dimensional Gaussian func-

Proceedings of the Fourth European Conference on Space Debris, Darmstadt, Germany, 18-20 April 2005
(ESA SP-587, August 2005)

tion described by a 3x3 covariance matrix. The probability of collision for the conjunction of two objects turns out to be a straightforward mathematical problem.

However, this example reveals one of the recurring problems of statistical analysis. Often the choice of statistical models introduces unstated assumptions that may bias the answer. In the case of collision avoidance, the tacit assumption has been made that the positional uncertainties of the two conjoining objects are independent of one another. This would be true if the terms in the covariance matrix were only due to measurement uncertainty. However, part of the downrange uncertainty comes from an uncertainty in the projected drag, which is correlated to unknown future variations in the solar activity. If the solar activity and atmospheric density goes up or down, then there may be a correlation between downrange positions of both objects. A preliminary study (Matney, 2003) showed that there was, indeed, some correlation in the downrange position uncertainty that could be attributed to atmospheric drag. This correlation means that the problem is not one of two 3x3 covariance matrices, but a single 6x6 covariance matrix that identifies all correlations (and anti-correlations if they exist).

Even in the case of collision avoidance, the normal approximations break down after a time, and the positional uncertainty distributions no longer lend themselves to convenient mathematics.

3. THE QUESTION

The first thing one must ask oneself when doing a statistical analysis is "what is it that I seek"? This sounds like a trivial question, but many statistical tools may answer a different question than the one you wish to answer.

In the measurement context, we are typically faced with a set of data from which we would like to extract some set of parameters and their associated uncertainties. For instance, a radar may observe some rate of objects of a particular type for some length of time, and we would like to know how many such objects we would see if we were to observe an "infinite" amount of time.

4. FREQUENTIST STATISTICS

The "standard" tool is supplied by a branch of the statistical arts known as "frequentist statistics". Frequentist statistics tells you that of a given a set of possible parameters, which ones have the highest confidence of producing the observed data. The answer is usually given as "confidence intervals" – lower, upper, or central, depending on the question being asked. A very good explanation of such confidence intervals can be found in Feldman and Cousins, 1998.

Frequentist confidence intervals have a very nice property known as "coverage". Consider a case where you know the actual set of parameters that determine the probability of various measurements (e.g., you know what average rate for a Poisson process). Now you randomly draw data from that distribution. For each data set drawn, pretend you don't know the correct answer and compute the appropriate confidence limits (e.g., 90% upper confidence limits). If you were to repeat that procedure many times – randomly drawing data and computing the confidence intervals – you would find that the confidence intervals would "bracket" or "cover" the correct answer a fraction of the time equivalent to the percentage confidence limit chosen. Consequently, if you computed 90% confidence limits for all your data sets, you would know that approximately 9 out of every 10 estimates would "cover" the correct answer.

Computing confidence limits for multidimensional problems can lead to some problems. As can be seen in Fig. 1, if you wish the confidence limits on one parameter out of many, you end up "over-covering" the correct answer. One solution is a procedure that computes the "profile likelihood" (Rolke and López, 2001). This is the maximum likelihood for each value of the parameter of interest held fixed and allowing the other parameters to vary. The limits of this curve within the confidence bounds approximate the one-dimensional confidence bounds of the parameter of interest.

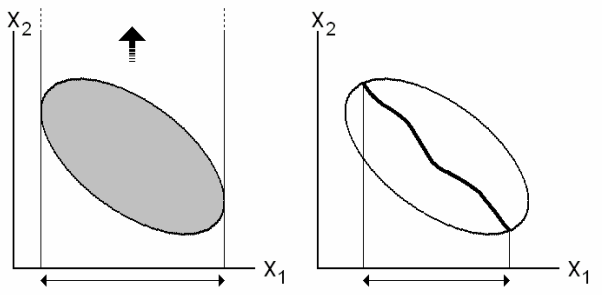

Figure 1. Consider a two-dimensional confidence region in parameters x_1 and x_2 where we wish to construct the confidence region for parameter x_1 only. Simply projecting the two-dimensional confidence region onto the x_1 axis results in over-coverage – the one-dimensional confidence region covers more possible solutions than desired. Instead, compute the curve that defines maximum likelihood for fixed x_1 lying within in the two-dimensional confidence region. This "profile likelihood" gives approximate coverage for the one-dimensional confidence limits of x_1.

There are some computer-intensive tools that are very useful for computing systems that do not follow the textbook rules. They are resampling methods, where the actual data is resampled multiple times to compute the desired information (Efron, 1982). These procedures work because the data you have sampled represents a single subset of the "true" data set, which you cannot access. However, a sub-sample from your data imitates multiple actual samples from the "true" data set.

One method is known as the jackknife. Suppose you have a set of N data points (x_1,x_2,…,x_N). In the jackknife, you compute the parameter of interest (e.g., the median) N times, excluding one data point in turn each step. You end up with N realizations of the parameter which are used to estimate the uncertainty on the parameter θ:

$$\hat{\theta}_{(j)} = \hat{\theta}(x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_N)$$
$$\hat{\theta}_{(\bullet)} = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_{(j)} \qquad (1)$$
$$\hat{\sigma}_{Jackknife} = \sqrt{\frac{N-1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_{(j)} - \hat{\theta}_{(\bullet)}\right)^2}$$

Another method is known as the bootstrap. In this method, you randomly resample N data points from the original set of N points with replacement. This means that for a resample you may choose any one data point one, two, or three times, or not at all. For each resample, the parameters of interest are computed. The distribution of these parameters over many bootstrap samples corresponds to the approximate uncertainties in the parameters, and can be used to construct confidence limits.

There is another type of bootstrap method known as the parametric bootstrap. In this method, you assume that you have information on the parametric form of the probabilities. The parameters are estimated from the original data (e.g., the mean and the standard deviation of a normal distribution), and the sampling is done from that fitted distribution.

Bootstrap methods in particular tend to be very computer intensive. In many cases, the number of bootstrap resamples may need to exceed 1000 in order to achieve the desired results. There are some parametric bootstrap techniques using a pivot – a point with one or more parameters held fixed – that give excellent results in multidimensional problems (Scholz, 1994). However, they require bootstrap resampling from the bootstrap resample itself, and so are very computer-intensive.

Despite all the advances with frequentist techniques in recent years, the results are not always what we would like. While the coverage property gives us confidence we have "covered" the answer, frequentist techniques do not tell us the relative probability of the various possible solutions within the confidence limits. This is due, in part, to the fact that there is no unique set of confidence limits for a particular problem. For example, if one wishes to construct 90% confidence limits, one can construct upper limits, lower limits, central limits, or anything in between. In order to assess the relative probability of different solutions, we need to turn to another method.

5. BAYESIAN STATISTICS

The question we are usually asking is "given a set of data x, what is the probability that the actual parameter of interest is y?" This can be written

$$p(y \mid x). \qquad (2)$$

What we usually have, however, is a model where if we know y, then we can compute the probability of sampling a data set x, given by

$$p(x \mid y). \qquad (3)$$

The relationship between these two was originally proposed by Reverend Thomas Bayes in the 18th century:

$$p(y \mid x)\ p(x) = p(x \mid y)\ p(y) \qquad (4)$$

$$p(y \mid x) \propto p(x \mid y)\ p(y). \qquad (5)$$

This requires knowledge of the function *p(y)*, known as the "prior". The prior describes the relative probability of the various parameters before data has been added to the mix. For instance, an observer might have some vague idea of the range of possible values of a parameter before it is actually measured.

There is substantial literature on computing priors, especially for what are termed "non-informative priors" (Kass and Wasserman, 1994). These are priors where no subjective information is brought in before the measurements are made, other than the form of *p(x|y)*. In general, non-informative priors are not simple constructs, but are based on information and entropy arguments. Use of naïve priors such as arbitrarily setting *p(y)* to a constant may lead to erroneous results.

Bayesian analysis has the potential for providing the kinds of answers we seek. However, there are some drawbacks. Bayesian solutions do not always have simple frequentist properties, so it may be difficult to

determine if a particular non-informative prior is giving the correct results. Computation of multidimensional non-informative priors is often difficult and computer-intensive (Lafferty and Wasserman, 2001). Also, if the observer brings subjective knowledge to a situation, how is that knowledge to be rigorously quantified in the prior?

Because Bayesian analysis can provide the right kinds of answers, we return to the original question, "What is it we seek?" Consider the notional chart in Fig. 2, showing the cumulative distribution of some orbital debris population as a function of size. Historically, the question has always been "What is the cumulative flux or population and the associated uncertainties at a particular size x?" If the "error bars" are constructed correctly, then we simply look up the answer on the chart. However, this assumes that we are asking for the value at x_1 independently of the value at x_2. What if we ask for the flux and uncertainties at x_1, x_2, and many other points simultaneously? The values of the flux and their uncertainties at these points are often correlated/anti-correlated with one another, so that the confidence limits on a variety of fluxes will generally look different than those for a single parameter. Consequently, one must be very careful how the problem is framed, or you may not get the answer you desire.

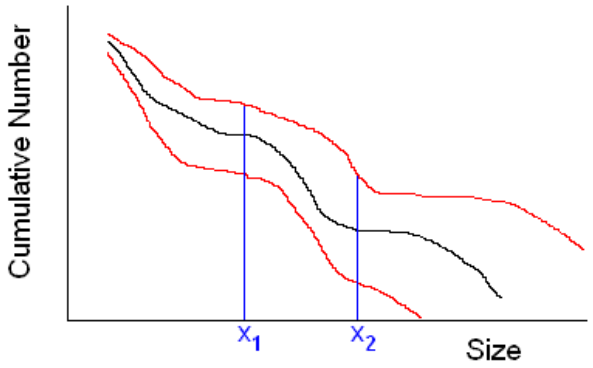

Figure 2. Consider the cumulative distribution of some parameter (like flux) as a function of particle size x. One can ask what the uncertainty of the flux estimate at a point x_1 independent of x_2 and compute the answer. However, if one is looking for the uncertainties of the flux at both x_1 and x_2, the resulting confidence limits or "error bars" will not be the same as the point-by-point answer. This is due to correlations or anti-correlations between the parameters. Such situations may arise in estimating uncertainties in Bumper-type models that compute fluxes at different sizes simultaneously.

This situation arises in detailed risk analyses of spacecraft hazard. NASA's Bumper code models a spacecraft as a series of surfaces. Each surface is susceptible to perforation by particles of a given size at a given angle and impact speed. The orbital debris environment is provided by the NASA ORDEM2000 Engineering model. Because Bumper is integrating the flux on a surface from different debris sizes, directions, and speeds simultaneously, any attempts to assign uncertainty to the final risk numbers will need to accurately assess the correlations and anti-correlations of the debris fluxes across the size, direction, and speed spectrum.

6. PROBABILISTIC RISK ASSESSMENTS

Probabilistic risk assessments (PRAs) represent an attempt to incorporate statistics and their associated uncertainties to determine failure modes in engineering systems. For spacecraft, one of the potential failure modes is orbital debris/micrometeoroid collisions, so there has been a desire to tailor engineering and risk assessment models to incorporate the necessary structures to be used in PRAs.

PRAs model the different failure modes by using the associated probability of a given event occurring. These probabilities are usually based on a combination of experiments and engineering judgment, and themselves have an intrinsic uncertainty. For instance, a particular component may have a failure mode that behaves in a Poisson manner with some average failure time T. This failure time in turn has some uncertainty associated with it σ_T – itself described by some model distribution (normal, log-normal, etc)

PRAs typically model failures by running Monte Carlo simulations and computing random failure modes. These computations can be quite complex, with certain failure modes triggering others. The only limit is the complexity of the model selected.

The uncertainties in the failure modes are handled by randomly choosing the parameters according to their distribution given by the user (in the example above, randomly choosing T from the user-supplied model each Monte Carlo run). The final results, in principle, offer an improved estimate of the total probability of the risk to the system.

However, there are some drawbacks. PRA analyses are only as good as the data supplied to them. In general, the parameters are computed based on some data, and the uncertainties on these parameters are typically computed using Frequentist methods. However, the manner in which the parameters are sampled in a PRA – selection based on a probability model – is manifestly Bayesian. Therefore in order to do an accurate PRA it may be important that a proper Bayesian analysis is performed on the uncertainties of the failure mode in question.

7. CONCEPTUAL UNCERTAINTIES

A different type of uncertainty occurs when we use models to extrapolate into areas where we have insufficient information. One example is projecting future launch rates and types. How do we bound that problem?

Typically, orbital debris models use Monte Carlo simulations to sample the range of future activities based on a set of assumptions. It would be prudent to construct our models to accurately reflect the range of possible future outcomes. For example, it is widely recognized that future solar activity has a strong effect on the future evolution of debris. However, only the gross behavior of the 11-year cycle of solar activity can be predicted. It is not known if any particular future cycle will be high or low.

Some researchers have computed approximate high and low confidence limits on the next solar cycle based on past cycles. But the how do we use these? One method is to simply randomly sample from the distribution over the complete cycle (Fig. 3). This does not reflect history, however, where higher solar cycles tend to stay high through the complete cycle and low solar cycles stay low throughout the complete cycle. Another option is to just take the average solar cycle and project it out into the future. However, this does not reflect solar activity in the past, and probably will not reflect the future behavior.

The solution I propose takes advantage of the Monte Carlo nature of the calculations and draws on the inspiration from PRA analyses. For each Monte Carlo run, have a different future solar activity with randomly alternating high and low solar activity profiles. In this method, each entire solar cycle is chosen at random, presumably from a catalog of possible solar cycles based on history. That way, the solar cycles themselves will behave in a reasonable manner, and the future solar profile for each Monte Carlo run will be different. The Monte Carlo runs of the model will now better "map out" the future possibilities. A similar method can be used for launch traffic and other future assumptions. The idea behind this approach is to properly formulate the question, "What is the reasonable range of possible futures we could expect to see?"

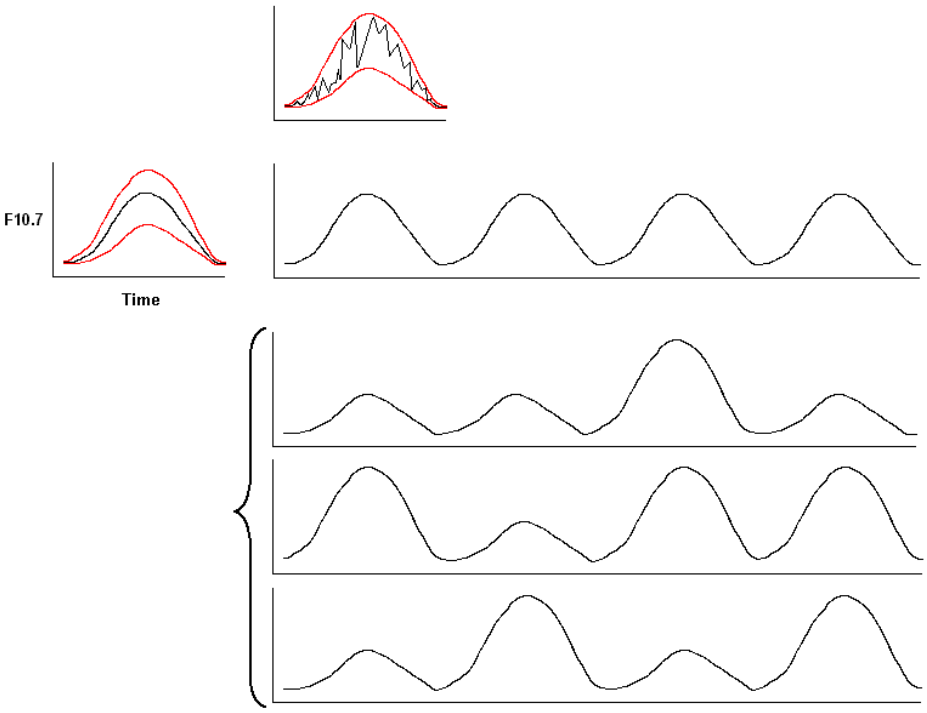

Figure 3. It is possible to estimate the upper and lower confidence limits on a solar cycle by examining historical behavior of the Sun (figure on left). But how are these uncertainties implemented in models that seek to predict future behavior of the orbital debris environment? One way (top) is to have the F10.7 cm flux of the Sun vary randomly each point in time. However, this does not reflect the way the Solar activity has behaved in the past, where the entire cycle tends to stay low or high – indicating correlations of solar activity within an entire cycle. Historically, models have used the average solar activity (center) to model the future solar activity. However, the solar activity has never behaved like that in past, nor is it likely to do so in the future. Another option (bottom) is for each Monte Carlo realization of the future projection model to have a different set of high and low solar activities that reflect the range of possible futures. A method like this is more likely to sample the range of possible futures.

8. CONCLUSIONS

In this paper, I have endeavored to highlight the importance of doing proper statistical analyses on orbital debris models and measurements. I have also outlined some old and new tools that are available. The following are a set of heuristic rules that I have found useful when doing statistical analyses.

Be sure that you are answering the question you wish to answer. Too often, statistical tools are applied that are easy to use, but answer the wrong questions. Simple statistical tools make a set of tacit assumptions about your problem that may not be appropriate, so be sure you are using the right tool for the job.

Be sure you are formulating and implementing your question correctly. Be careful about your assumptions on the inter-relation of the parameters and data. Are the parameters correlated?

Frequentist statistics may be easy to calculate, but they may not be answering the questions you really want answered. Bayesian statistics may give you the answers you really want, but the road to get to that answer may be difficult.

Be careful how you implement conceptual uncertainties in models. Be sure they model the world of solutions you wish to explore.

We in the orbital debris science community should make it a priority to better understand and report the uncertainties in our products. The road is long and difficult, but the accurate presentation of our models allows users to make meaningful decisions. The task is worth the effort.

9. REFERENCES

Efron B., *The Jackknife, the Bootstrap and Other Resampling Plans,* Society for Industrial and Applied Mathematics, 1982.

Feldman G. J. and Cousins R. D., A Unified Approach to the Classical Statistical Analysis of Small Signals, *Phys.Rev.* D57, 3873-3889, 1998.

Kass R. and Wasserman L., Formal rules for selecting prior distributions: A review and annotated bibliography, *Journal of the American Statistical Association*, Vol. 91, 1343-1370, 1996.

Lafferty J. and Wasserman L., Iterative Markov chain Monte Carlo computation of reference priors and minimax risk**,** *Uncertainty in Artificial Intelligence (UAI '01),* 2001

Matney M., Anz-Meador P. A., and Foster J. L., Covariance Correlations in Collision Avoidance Probability Calculations, *Advances in Space Research*, Vol. 34, Issue 5, 1109-1114, 2003.

Rolke W. A. and López, A. M., Confidence Intervals and Upper Bounds for Small Signals in the Presence of Background Noise, *Nuclear Instruments and Methods in Physics Research A*, Vol. 458, 745-758, 2001.

Scholz F. W., On Exactness of the Parametric Double Bootstrap, *Statistica Sinica*, Vol. 4, 477-492, 1994.