JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 17, No. 1, January–February 1994

Statistics for Spacecraft Pointing and Measurement Error Budgets

Marc Weinberger*
European Space Agency, 2200 AG Noordwijk, The Netherlands

This contribution to the calculation of spacecraft pointing and attitude measurement error budgets provides a rational method, based on Bayesian maximum a posteriori probability and related to maximum entropy principles, for the determination of probability distribution functions of the error components in these budgets. It takes into account the kind and amount of information on these error components that is usually available. It thereby improves on the conventional method, which assumes that all error components are normally distributed. Guidelines are provided for the choice of the confidence levels, i.e., the probability with which actual pointing errors must lie below the specified value. An example is given.

I. Introduction

BARRING some rare missions like Starlette and Lageos, which were equipped with multiple laser reflectors, all spacecraft have attitude pointing and/or measurement requirements and therefore require performance verification by means of analytical error budgets. This is an important task during all design and manufacturing phases and even beyond. At some time during a spacecraft project life, very detailed mathematical models, possibly coupled with hardware, are used to verify performance by simulation and test. But this topic is not the subject of the present paper. In the rest of the paper the expression "pointing" will be shorthand for "pointing and/or attitude measurement."

Global pointing errors are combinations (see Sec. II) of error components associated with sensors, actuators, structures, mathematical models, control laws, and electronics. Completely known systematic components can be calculated once for all and will not be further considered. The remaining components, which are the majority, are best treated as random variables, characterized by probability distribution functions (PDFs). Except where the distinction is needed, the acronym PDF will henceforth indiscriminately apply to densities or cumulative distributions.

Specifications usually prescribe maximum allowable errors together with an associated probability or confidence level of not exceeding this error. Section III explains why unrealistically demanding confidence levels must be avoided.

Section IV discusses conventional approaches, where very often all errors are assumed to have normal distributions, because these are easy to work with and linear combinations are also normal. Reasons for rejecting this blanket assumption are given there.

Section V defines PDFs for periodic errors and limit cycles, and so also serves as an argument against general use of the normal distribution.

These considerations bring us to the aim of this paper, namely, to provide a rational methodology for the determination of these required PDFs, which will permit a realistic estimation of performance. In Sec. VI, Bayesian maximum a posteriori probability (MAP)[1,2] is used. MAP provides a framework to combine prior knowledge with later, experimental data. MAP is related to the maximum entropy principle,[3-5] which in an information-theoretical sense produces the least prejudiced PDFs that take into account the available information.

Section VII provides an example, and Sec. VIII gives conclusions.

This paper is situated with respect to prior and related work as follows. Maximum entropy (MAXENT),[3,4,6] MAP, and maximum likelihood[1] (a special case of MAP) have all been used for the estimation of frequency-of-occurrence laws or PDFs and for the problem of image reconstitution, which shares with PDFs the requirement for nonnegative estimators (frequencies or intensities). Reference 1 presents a very readable account of MAP and related methods. Gull and Daniell[6] use the MAXENT principle in image reconstitution. The present paper uses Bayesian MAP and differs from Ref. 1 in using different models for the measurements and for the prior knowledge that are more appropriate for the spacecraft pointing case. The detailed differences between Refs. 1, 3, and 6 and the present paper are explained in Sec. VI, when the necessary formulas are available. Reference 7 uses a different approach: errors are classified in various categories (such as biases, long term errors, harmonic or periodic errors, and noise), and types of PDFs are recommended per category. We use elements of this approach in the selection of prior PDFs (see Sec. VI).

II. Error Combination

Accuracy requirements are usually stated as follows: the pointing error E must be below a maximum permissible value $E_{\max}$ with a specified probability or confidence level P_C. This can be expressed as follows ("prob" stands for "probability"):

$$\text{prob}\left(|E| \leq E_{\max}\right) = P_C \tag{1}$$

The choice for P_C will be discussed in Sec. III; here we present the combination of several errors into global single axis errors and into cone errors. Reference 7 has much of the material in this section, which is given to make the present paper more self-contained.

A. Single Axis Errors

Many missions require the computation of rotation errors around single axes, e.g., three orthogonal body axes such as roll, pitch, and yaw.

The total rotation error E around one axis is usually the linear combination[7]

$$E = \Sigma_k a_k E_k \tag{2}$$

Received March 29, 1992; presented as Paper 92-4547 at the AIAA Guidance, Navigation, and Control Conference, Hilton Head, SC, Aug. 10-12, 1992; revision received Jan. 29, 1993; accepted for publication Feb. 16, 1993.

*Staff Member, Automation and Informatics Department, ESTEC, P.O. Box 299.

of individual error components. There are situations, especially with high performance missions, where higher order terms must be retained. Such terms may contain powers, products, trigonometric functions, etc., of components, and the standard formulas[8,9] for functions of one- or multidimensional variables can be used to compute the PDF of the nonlinear entities. Some of the terms in Eq. (2) can result from such a calculation. If E_1, E_2, etc., are independent, as assumed here, the density $f(E)$ of E is the result of successive scaled convolutions of the densities $f_k(E_k)$ of all E_k and is found by repeated application of the formula for two components:

$$f(E) = |a_1 a_2|^{-1} \int f_1(E_1'/a_1) f_2[(E - E_1')/a_2]\, dE_1' \qquad (3)$$

The hypothesis of independence of the components E_k is usually appropriate because entirely different and unrelated pieces of hardware and software are responsible for them. But orbit-related periodic thermal effects may well cause connections between several components; these should first be combined deterministically with the proper phasing, insofar as this is known, and the resulting error is then combined statistically with the others.

If the components E_K have a nonzero mean, the mean of E is simply the weighted sum of these means. In the sequel only zero-mean error components are considered as it is simple to add on nonzero means afterward.

B. Cone Errors

Many mission requirements bear on the total cone angle error in the pointing or measurement of an axis. If E_X and E_Y are the rotation angle errors for single axis rotations around x and y, the semicone pointing error E of the z axis for small E_X and E_Y is given by[7,10]

$$E = (E_x^2 + E_y^2)^{1/2} \qquad (4)$$

The PDF for E can be obtained by two-dimensional integration of the PDFs of E_X and E_Y over circular areas satisfying Eq. (1). This and more complicated expressions are discussed in Ref. 7. Other discussions of pointing error criteria can be found in Refs. 10 and 11.

III. Choice of the Confidence Level

Confidence levels were defined at the start of Sec. II, Eq. (1). Conventional budgets, assuming normal PDFs for everything, often use 3σ confidence levels, corresponding to a probability of 0.9973. This very high probability is then often carried over to the other PDFs. For the tails of any PDF to be known with some accuracy at this level requires a large volume of experimental data, much larger than is normally available from tests.

The Kolmogorov-Smirnov statistics[8] can be used to find the number of experimental trials needed to determine a cumulative PDF $F(E)$ to an accuracy δF. Since we only need the tails of the PDF, we can use less conservative statistics.[12] But what really is to be determined is the value E_C of E where $F(E_C) = P_C$, and this calls for inversion of the function $F(E)$ at the tail, where the density $f(E) = (dF/dE)$ is near zero and uncertain, and hence where dE/dF is very large. An asymptotic analysis for large N indicates that the required number of trials N is proportional to $(1 - P_C)$ and inversely proportional to the square of the permitted error δF on $F(E_C)$. If we allow a 10% error on E_C, the result is that the ratio $N(P_C = 0.9973)/N(P_C = 0.95) = 6.2$.

This result constitutes a more than sixfold amount of testing required for the higher confidence level compared with $P_C = 0.95$.

It is therefore strongly recommended to choose confidence levels P_C not in excess of something like 0.95 to avoid extreme sensitivity to the modeling of the tails of the PDF. Reference 7 gives the same recommendation, which it bases on the graphical comparison of the normal PDF and several beta PDFs. This shows that for P_C above 0.95 these curves strongly diverge, whereas their differences are much smaller at lower values of P_C.

A perhaps paradoxical but justifiable exception can be made when the worst case must be considered and $P_C = 1$ is to be used: a finite upper bound is usually known and therefore we do not meet the difficulty of large sample size here. It may be justified to have separate requirements depending on error category and, e.g., to use $P_C = 1$ for some categories and a lower P_C for others because of different influences on the total performance. As an example, biases are very important if absolute pointing or measurement is demanded, and $P_C = 1$ may be justified here. (This then usually has the practical and beneficial effect of motivating the users to use calibration as a means for reducing bias whenever possible.)

IV. Conventional Methods

Conventional pointing error compilations assume that all error components obey a normal distribution. The limit theorems of probability are often invoked as the justification, even though they are not necessarily applicable (e.g., because the error sources are not the sum of many effects or because a few nonnormal components dominate). Section V gives two counterexamples. Other errors are better modeled by uniform distributions, and all errors have finite bounds, unlike the infinite tails of the normal PDF.

If all of the contributions E_k were normal, so would E be normal, and this fact is one reason why budget compilers like to assume that everything is normal. Since numerical integrations discussed in Sec. II can now be performed on simple personal computers, there is no reason to stick to normal PDFs just for computational convenience.

In these conventional budget compilations, one calculates only variances (var) from Eq. (2)

$$\text{var}(E) = \Sigma_k a_k^2 \,\text{var}(E_k) \qquad (5)$$

and uses the standard confidence level 0.9973 for 3σ, or 0.955 for 2σ, or 0.68 for 1σ (σ is the standard deviation). The 3σ value is often considered to be the equivalent of the worst case in a practical sense. The example in Sec. VII shows that the conventional method can be very unrealistic. There are also ad hoc variations, some of which are discussed in Ref. 7, devised in attempts to get more realistic error budgets. But as one does not know when such an ad hoc rule is good, there is a need for a rational method. This method must also be practical: it must neither demand excessively complicated calculations nor require more information than is usually available to the budget compilers.

V. Special Error Distributions

To give some immediate counterexamples to the conventional, normal-only distributions, periodic errors and limit cycles are discussed here.

A. Periodic Errors

Assume first an error contribution of the form

$$E = A \sin(wt + \text{phase}) \qquad (6)$$

For fixed amplitude A and frequency w this is a deterministic signal if we exclude random phase shifts. Nevertheless, if our performance specifications are interpreted as a percentage of time that errors have to be below a specified bound, it is acceptable to use a density based on the temporal distribution

$$f(E) = (1/\pi)(A^2 - E^2)^{-1/2}; \qquad |E| \le A \qquad (7)$$

This is a bimodal PDF with maxima at the extreme points, very unlike the normal distribution, which concentrates probability near the mean.

A more accurate picture of reality is that the signal E is periodic with well-known fundamental frequency w, e.g., orbit frequency, but harmonics can be present and the amplitude A is only approximately known. This more realistic case is subsumed by the method of Sec. VI.

B. Limit Cycles

For idealized limit cycles one can derive PDFs based on temporal distribution as was done for periodic errors. The ideal bang-bang limit cycle in the absence of disturbing torques can lead to a uniform pointing error PDF within the deadband. The ideal one-sided limit cycle with constant disturbing torque, which produces parabolic motion between control pulses, leads to a bimodal PDF with maxima at the endpoints.

VI. Probability Distribution Function Estimation via Bayesian MAP

A. Basics

Take an error component x (x will be used here instead of the earlier symbol E). We want to have a rationally justifiable model for its probability density function $f(x)$ that incorporates any a priori knowledge we may have, produces consistent answers, and is minimally prejudiced in that it avoids drawing conclusions that are not justified by the available data.

For a simple presentation of the method, following Ref. 1, assume that x is a discrete or, if necessary, a discretized variable with m possible outcomes x_I, $i=1$ to m, and we want to determine the discrete PDF $f_I = f(x_I)$. Take N measurements z_J, related to the true values c_J, by

$$z_J = c_J + e_J \tag{8}$$

The notation c_J is used to indicate the true value of the discrete or discretized variable x, such that numerically c_J is equal to one of the values in the set $\{x\}$. Braces $\{x\}$ denote sets (x_1, x_2, ...). The error e_J is c-independent data noise, with inessential and easily removable restriction to zero mean, and with variance σ_E^2. The noise e_J is large enough to be able to produce wrong data points z; i.e., z_J can be misidentified with a wrong value x_I in $\{x\}$.

Count the number n_I of occurrences of the outcome x_I. Then

$$N = \sum_{i=1}^{m} n_I \tag{9}$$

The relative frequency of experimental occurrence $f_I = n_I/N$, if it were available, could be taken as the posterior (after the measurement) estimate of the true f_I. Assume that knowledge acquired before the test can be interpreted as prior probabilities q_I of the outcomes x_I. The conditional probability of finding a particular set $\{f\}$, $\{e\}$, given $\{q\}$, is then given by Ref. 1

$$P\big(\{f\}, \{e\}/\{q\}\big) = P_0\big(\{f\}/\{q\}\big) \cdot P_N\big(\{e\}\big) \tag{10}$$

P_0 is the probability when there is no measurement noise and is given by

$$P_0\big(\{f\}/\{q\}\big) = N! \prod_{k=1}^{m} q_k^{f_k N}/(f_k N)! \tag{11}$$

P_N is the probability for the corresponding noise vector $\{e\}$

$$P_N\big(\{e\}\big) = \prod_{i=1}^{N} \text{prob}(e = e_i) \tag{12}$$

The independence of the variables $\{c\}$ and $\{e\}$ in Eq. (8) permits the split into factors in Eq. (10) and makes P_N independent of $\{f\}$ and $\{q\}$.

Often the prior knowledge is that the error can lie between a minimum and maximum and that intermediate values are possible. Then, different approaches such as group theory[13] and the method of least informative priors[14] agree that it is reasonable to assume equal q_I, i.e., a flat prior distribution. Structural misalignment errors usually belong to this category. If we do not even know that intermediate values can exist, these theories[13,14] indicate that it is better to take a bimodal distribution for q_I with maxima at the extreme values. For periodic errors a nonuniform distribution of q_I may also be appropriate: if its amplitude changes slowly around the extrema, a bimodal distribution is suitable (see Sec. V). As an example of the latter category, $\{q\}$ could be based on the law equation (7) where A is given the most probable or mean amplitude.

Most often budget compilers do not have the record of the set $\{z\}$, i.e., of $\{n\}$ or $\{f\}$ for the individual error components. The measurements have already been lumped together in quantities that are functions of the set $\{f\}$ and that are corrupted by noise. These quantities usually belong to either or both of two broad categories much used in statistics. The first category involves the knowledge of moments, especially means and variances. The second category deals with order parameters, e.g., medians, quantiles, or more general ranges. For some errors we do not have real measurements, but rather estimations based on mathematical models. This condition usually applies to structure-related errors. We will now discuss examples of the two categories: moments and order parameters.

B. Knowledge of Mean and Standard Deviations

Assume that the original and now unavailable measurement data $\{z\}$ have been lumped into an experimental value s^2 of the variance σ^2 of x, as

$$s^2 = \left(\sum_{i=1}^{N} z_i^2\right) \Big/ N \tag{13}$$

The conditional mean $\overline{s^2}$ and variance V of s^2, given $\{f\}$ and the noise statistics, are provided in Appendix A. It is an intractable problem to obtain the exact conditional PDF of s^2 given $\{f\}$; therefore an approximation by a normal PDF, requiring knowledge of mean and variance only, is used here. Another way of looking at this approximation is to consider it as the fundamental term either in a Hermite-normal expansion of the true PDF[9] or in an Edgeworth-like N-dependent expansion[9] applied to s^2.

The conditional probability of finding the value s^2, given $\{f\}$, $\{e\}$, and $\{q\}$, can then be represented by

$$P\big(s^2/\{f\}, \{e\}, \{q\}\big) = (2\pi V)^{-1/2} \exp\left[-(s^2 - \overline{s^2})^2/(2V)\right] \tag{14}$$

The arguments $\{q\}$ in Eq. (14) can be omitted because s^2 does not depend explicitly on them in this approximation. They have been kept for clarity, to preserve the standard sequence of arguments in the Bayes's formula equation (16).

It is worth noting that taking a PDF like the gamma distribution for s^2 to avoid any possibility of negative values leads effectively to the same ultimate results but requires more algebra. The modeling of the conditional distribution of s^2 is further discussed in Appendix A, together with practical methods for estimating V.

We now determine $\{f\}$ by requiring that the estimate for $\{f\}$ maximizes the posterior probability or, equivalently, its logarithm (log means natural logarithms) $\log P(\{f\}, \{e\}/\{q\}, s^2)$ subject to the normalization condition

$$\sum_{i=1}^{m} f_i = 1 \tag{15}$$

This is the Bayesian MAP approach.[2]

The posterior log probability can be obtained by Bayes's theorem

$$\log P(\{f\}, \{e\}/\{q\}, s^2) = \log\left[P(s^2/\{f\}, \{e\}, \{q\}) \cdot P(\{f\}, \{e\}/\{q\})/P(s^2/\{q\})\right] \quad (16)$$

If we had assumed $P(\{f\}, \{e\}/\{q\})$, the probability of $\{f\}$ and $\{e\}$ before s^2 is determined to be uniform, the MAP would be specialized to the maximum likelihood principle,[2] which says that the estimate for $\{f\}$ must maximize $P(s^2/\{f\}, \{e\}, \{q\})$.

Note that $P(s^2/\{q\})$ is independent of $\{f\}$ and thus does not affect these maximizations. The same remark applies to $P_N(\{e\})$: the noise parameters $\{e\}$ are effectively decoupled by our hypothesis of additive noise and do not play an explicit role in the maximizations.

Appendix B describes the solution method. For continuous errors x with prior $q(x)$, the sought-for density $f(x)$ is

$$f(x) = \begin{cases} cq(x)\exp(\lambda x^2) & |x| \le A \\ 0 & \text{elsewhere} \end{cases} \quad (17)$$

The coefficient λ, determined in Appendix B, can be positive, zero, or negative. The factor c serves to normalize $f(x)$.

It is interesting to note what happens for the two extreme cases of very high noise, V infinite, and very low noise, V zero. For high noise the unreliable datum s^2 is discarded and we find the maximum entropy solution where the posterior probability $\{f\}$ is the same as the prior $\{q\}$. For low noise the solution $\{f\}$ is constrained to produce the accurate datum s^2 exactly and, subject to this constraint, does its best to maximize entropy.

The special case with uniform priors (and continuous distributions), $q(x) = \text{const}$, leads to the truncated normal density[4] (λ can be >0, $=0$, <0), as illustrated in Fig. 1.

$$f(x)\,dx = \begin{cases} (c/A)\exp\left[\lambda(x/A)^2\right] dx & |x| \le A \\ 0 & \text{elsewhere} \end{cases} \quad (18)$$

with the necessary formulas for λ, c, and variance of x given in Appendix C. For high noise, i.e., increasing V, $f(x)$ becomes more uniform, as expected from maximum entropy considerations.

At this point we are ready to discuss in more detail the differences between this paper and Refs. 1, 3, 4, and 6. Tribus[4] takes the MAXENT principle as his starting point and assumes noise-free measurements for integral quantities like s^2. His results correspond to the low noise limit for uniform priors here. Gull and Daniell[6] treat noisy image reconstruction (our $\{f\}$ corresponds to their pixel intensity) and also start from the MAXENT principle. Their method then entails the constrained maximization of $P(\{f\}/\{q\})$ for uniform priors. The effects of noise are approximately accounted for by the constraint that the sum of weighted squares of differences between true values and measurements be equal to its expected value. This can be interpreted as one among several ad hoc generalizations of the noise-free case to a noisy case. Frieden[1] mostly treats noise-free data, but he also considers the case where the measurements have additive noise, as in our Eq. (8), and where these measurements remain available to the analyst (i.e., the condensation of the measurements with additive noise into something like the experimental variance s^2 does not occur in Ref. 1). The MAP equations equivalent to Eq. (B2) then have other noise-dependent terms and maximization is performed over not only $\{f\}$ but also the best noise-vector estimate. Jaynes[3] gives what he calls a "handwaving" argument to justify an approach that is equivalent to Ref. 1. The contribution of our paper is to allow noise and nonuniform priors, to incorporate measurement models for the data that are representative of spacecraft pointing budget conditions, to employ MAP [although we are forced to use approximate PDFs like Eq. (14) and Eq. (D1) for the clumped data], and to set up an approach in which other types of data structures can be treated beside the variances and quantiles dealt with here. An additional goal is to alert spacecraft budget compilers to some of the relevant literature in applied statistics.

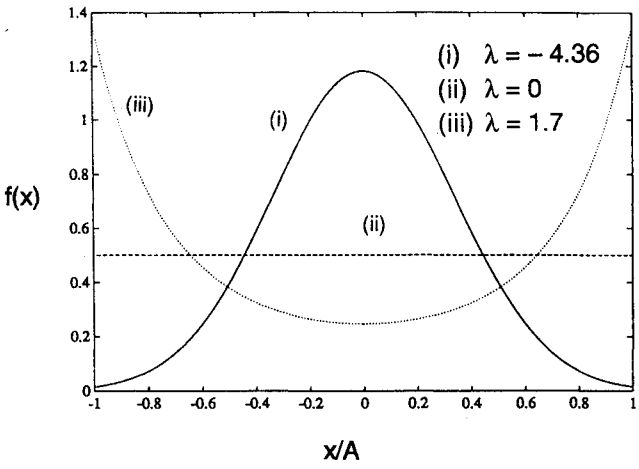

Fig. 1 Symmetric truncated Gaussian distributions.

A. Knowledge of Quantiles (Order Parameters)

Sometimes we cannot estimate the variance, but we have experimental evidence that a certain fraction p_1 of the errors lies in an interval I_1, another fraction p_2 in another interval I_2 etc., with m being the total number of intervals or bins. A three-interval example can arise as follows: the error is uniform between expected limits $-A$ and $+A$ with, say, 90% probability. To introduce a measure of conservatism, we suppose that the remaining 10% are uniformly distributed between A and, say, $1.5A$, and $-A$ and $-1.5A$.

The discrete frequencies $\{f\}$ to be determined are the best posterior estimates for the true $\{p\}$. Following the method explained in Sec. VI, we assume that p_I is normally distributed with mean f_I and variance V_I, and that all p_I are independent, except of course for the normalization requirement Eq. (12). The details of finding the solution $\{f\}$ are in Appendix D.

Low measurement uncertainty $V_I \to 0$ gives the expected constrained maximum entropy solution $f_I = p_I$ (posterior probabilities are equal to measured data), whereas the high measurement noise $V_I \to \infty$ gives the maximum entropy solution $f_I = q_I$ (posterior probabilities are the same as the prior ones).

B. Choice of Prior Distribution

Any prior knowledge should be incorporated in the choice of the prior PDF $\{q\}$. Models that we have found useful are (also see Ref. 7) uniform priors for structural errors and for many biases; bimodal (upward concave) priors like Eq. (7) or the truncated normal Eq. (18) with positive λ for harmonic errors; and the truncated normal Eq. (18) with relative standard deviation $\sigma/A = 1/3$ (a pseudonormal with 3σ identified with worst value) for noise and for field-of-view distortion effects (sometimes called "local biases" in star sensors).

Prior PDFs with asymmetry about the mean are possible, but usually there is insufficient information to justify their choice. The most likely exceptions are one-sided limit cycles (see Sec. V) when the disturbing torques have predominantly one polarity.

VII. Example

The example is taken from the pitch axis error budget of the Earth Resources Satellite ERS-1 of the European Space Agency. ERS-1, launched in July 1991, is in a sun-synchronous, i.e., near-polar Earth orbit. Its yaw axis is Earth-oriented. The pitch axis is on the average perpendicular to the orbit plane; one payload, a side-looking radar, requires yaw

Table 1 Error sources

Type of error source	Amplitude, deg	Prior λ	Posterior λ
Bias $b1$	0.0580	0	0
Bias $b2$	0.0200	0	0
Bias $b3$	0.0100	0	0
Harmonic $h1$	0.0387	0	1.43
Harmonic $h2$	0.0344	0	1.43
Random noise n	0.0230	−4.36	−4.36

steering, i.e., there is a commanded orbit-periodic yaw motion such that the angle between pitch axis and orbit normal has a sinusoidal behavior with maximum excursion at the equatorial passages and zero excursion at the extreme latitudes. This yaw steering causes orbit-periodic body rates along yaw and the third axis roll.

The pitch angular measurement is obtained as the output of complementary filters[15] operating on the outputs of an Earth sensor and of gyros. The pitch error θ can be represented by

$$\theta = G(s)\theta_E + \left[1 - G(s)\right]\dot{\theta}_G/s \tag{19}$$

where θ_E is the Earth sensor error, $\dot{\theta}_G$ the gyro pitch rate error, s the Laplace transform variable, and $G(s)$ the low-pass transfer function, with $G(0)=1$, which attenuates the mostly higher frequency Earth sensor noise. The complementary filter $(1-G)/s$ suppresses constant gyro bias and also high frequency gyro noise. The six largest error contributions have been retained; they are listed in Table 1 and will be explained next.

Earth sensor biases pass unchanged through the filter G and produce the errors $b1$ = long term alignment shifts in the Earth sensor structure in orbit; $b2$ = Earth sensor alignment shift relative to the payload reference, due to launch vibrations; and $b3$ = very low frequency, thermally caused drift in the Earth sensor. The Earth sensor further contributes a twice-orbital frequency, infrared radiation modeling error with amplitude e_{IR}, whose amplitude $h2$ at the filter output is given by $|G(i2w_o)|e_{IR}$, where i is the imaginary unit and w_o is the orbital angular frequency.

The gyros are skewed relative to the spacecraft reference axes; the pitch rate error $\dot{\theta}_G$ in Eq. (19) is obtained by mathematical projection of three real gyro rate outputs onto the pitch axis. It therefore contains the orbit-periodic error $h1$ because the yaw steering-induced periodic rates, multiplied by small errors in the knowledge of the constant gyro scale factors, are not perfectly suppressed by the filter $(1-G)/s$. Finally, there is gyro random noise, modeled as normal, which is not perfectly suppressed by the complementary filter and that contributes the normally distributed noise entry n in Table 1.

All six errors are modeled by PDFs of the form defined by Eq. (18), and the amplitudes are given in the second column of Table 1.

In accordance with the recommendations at the end of Sec. VI, the prior values for λ in the third column of Table 1 have been chosen to correspond with uniform PDFs for the biases and with a pseudonormal PDF for the noise. We assume now that the PDFs for biases and noise are well known (their individual V is zero) and therefore their posterior (in the fourth column of Table 1) and prior PDFs are identical.

For the harmonics we also take a uniform prior (higher harmonics may give a flatter distribution than a pure sine), but we assume that the ratio of experimental variance to amplitude squared, s^2/A^2, = ½. For these harmonics we assume that the 95% confidence level uncertainty ds^2 on the experimental variance s^2 is ±10%. (We use method B of Appendix A.) The method of Appendix B [for discretization into 101 values, $m=101$ in Eq. (B4)] then gives a posterior value $\lambda=1.43$, shown in the fourth column of Table 1, which is indeed nearer the uniform PDF than the PDF with $\lambda=1.7$ that we would have found if there were no uncertainty ($V=0$). Repeated convolution, Eq. (3) gives the errors displayed in Table 2, second column, for several popular confidence levels quoted in the first column.

Table 2 Total errors (deg)

Confidence level P_c, %	Exact result	Total is normal	All sources are normal
50	0.0363	0.0345	0.0189
68	0.0529	0.0508	0.0279
95	0.0991	0.1002	0.0550
99.73	0.1347	0.1533	0.0841
100	0.1841	∞	∞

The third column of Table 2 gives the results for the errors if we assume that the total error is normally distributed with the exact standard deviation (0.0511). Except for P_C very near unity, there is a reasonable agreement between columns 2 and 3 of Table 2, because the six contributions have roughly comparable sizes and their sum is not too far from normal. But in other cases wider differences can be found.

The fourth column of Table 2 results from the conventional method (all individual sources are normal with a standard deviation equal to one-third of the amplitude A in Table 1). It is obvious that this crude approach leads to very optimistic error budgets for this example, except for confidence levels extremely close to 100% where they are just the opposite—very pessimistic.

VIII. Conclusion

Conventional pointing and measurement error budgets, which model all components by a normal distribution, can produce very wrong results, since many components often are more appropriately modeled by different probability laws. To remedy this situation, we have presented a rational approach for the determination of distribution functions. Two specific cases, where variances and quantiles are measured, were treated in detail. The techniques of this Bayesian MAP approach can be adapted to cases where other data types are measured. The new approach is a worthwhile improvement over past and present practices.

Appendix A: Distribution of the Measured Variance

The conditional mean s^2 and conditional variance V of s^2 used in Eq. (14) are given by

$$\overline{s^2} = \text{expectation}\left(\overline{s^2}/\{f\}, \{e\}, \{q\}\right) = \sigma^2 + \sigma_E^2 \tag{A1}$$

$$V = \text{var}\left(s^2/\{f\}, \{e\}, \{q\}\right) = N^{-1}(m_{4E} - \sigma_E^4 + 4\sigma^2\sigma_E^2) \tag{A2}$$

with

$$\sigma^2 = \sum_{k=1}^{m} f_x x_k^2 \tag{A3}$$

The fourth moment m_{4E} and the variance σ_E^2 of the measurement noise e in Eqs. (A1) and (A2) are known, because we assume that the statistics for the error e_J in Eq. (8) are known. If e is normal

$$m_{4E} = 3\sigma_E^4 \tag{A4}$$

The normal PDF equation (14) permits negative values for s^2, which have no physical meaning. However, for moderate to large values of N [even if $(NV)^{0.5}$ is not small compared to s^2], $V^{0.5}$ will be much smaller than s^2 and negative values of s^2 are very improbable. This defect can be avoided altogether by PDFs like the gamma, but the results obtained are very similar to those of the normal PDF anyway (they are identical to the level of approximation used here).

Later, for the maximization of posterior probability, we will need the partial derivatives

$$\partial \log\left[P(s^2/\{f\}, \{e\}, \{q\}\right]/\partial f_k = x_k^2 \left| \frac{s^2-\overline{s^2}}{V} + \frac{2\sigma_E^2(s^2-\overline{s^2})^2}{(NV)V} - \frac{2\sigma_E^2}{(NV)} \right| \quad \text{(A5)}$$

For moderate and large N, the second and third terms in Eq. (A5) are much smaller than the first one and will be neglected. This procedure is the same as neglecting changes in V due to changes in $\{f\}$ in Eq. (A2) for NV.

Appendix B shows that the maximization process requires knowledge of the combination NV, which does not depend on N [see Eq. (A2)], but not of N itself.

To be able to use the proposed method, we need (approximate) values for NV and $\overline{s^2}$, i.e., for σ_E and m_{4E}, as per Eqs. (A1) and (A2). We will next describe two methods.

Method A: The value to be taken for σ_E is related to the uncertainty of the measurements equation (8) that produce s^2 via Eq. (13). The following simple approximation is proposed: generally the budget compiler has at least a rough idea of how accurate the original measurements equation (8) were. Normally the signal-to-noise ratio σ/σ_E is well above unity; otherwise the then large and probably inaccurately known bias σ_E^2 in the value of $\overline{s^2}$ [Eq. (A1)] would make the procedure dubious. We can then estimate σ_E from the preceding signal-to-noise ratio or its inverse, the fractional uncertainty, of the measurements equation (8). As an example, assume normal statistics for e and approximate σ by the experimental value s. Then choosing a value r for the ratio σ_E/s means that we believe that on the average the measurements $\{y\}$ did have fractional errors $|e_i|/s$ less than r with a probability of 68%. With this assumption of normality for e, the fourth moment is then given by Eq. (A4).

Method B: An even simpler way of looking at this is to consider s^2 as a super measurement from a normal distribution with one sample, $N=1$, in Eq. (A1). If we believe that s^2 has a probable (50% level) error $\mathrm{d}s^2$ (different probability levels can be taken), we get the estimate

$$V^{1/2} = 1.47\ \mathrm{d}s^2 \quad \text{(A6)}$$

If we had chosen 95% instead of 50%, the coefficient 1.47 would be replaced by 0.5. Usually it is most perspicuous to express $\mathrm{d}s^2$ as a fraction or percentage of s^2. In this simple approach we neglect the bias σ_E^2, presumed small compared with σ^2, in Eq. (A1), and we do not need this equation for NV since this quantity is now a directly available experimental datum.

Appendix B: MAP Solution when Variance is Measured

If we use Stirling's approximation formula[1] for factorials and adjoin the normalization equation (15) by means of the Lagrangian multiplier $N\lambda_o$ (the factor N is inessential but simplifies later results), the conditions for MAP [see Eqs. (10), (14), and (16) for MAP and Eqs. (A1) and (A2)]

$$\frac{\partial}{\partial f_k}\left[\log P\left(s^2/\{f\}, \{e\}, \{q\}\right) + \log P_0\left(\{f\}/\{q\}\right) - N\lambda_o\left(\sum_{i=1}^{m} f_i - 1\right)\right] = 0; \qquad k = 1 \text{ to } m \quad \text{(B1)}$$

yield

$$f_k = q_k \exp(-1-\lambda_o)\exp(\lambda x_k^2); \qquad k = 1 \text{ to } m \quad \text{(B2)}$$

where

$$\lambda = (s^2-\overline{s^2})/(NV) \quad \text{(B3)}$$

and where $\overline{s^2}$ and V depend on all of the unknowns $\{f\}$ according to Eqs. (A1) and (A2). To obtain these equations, a further simplifying approximation has been made in the calculation of the partial derivatives as mentioned in Appendix A.

The numerical solution of the nonlinear equations (B2) is easily accomplished as follows: If we somehow knew the quantity λ in Eq. (B3), the PDF $\{f\}$ would be given by

$$f_k(\lambda) = \left[q_k \exp(\lambda x_k^2)\right] \Big/ \sum_{k=1}^{m} q_k \exp(\lambda x_k^2) \quad \text{(B4)}$$

Now form

$$\sigma^2(\lambda) = \sum_{k=1}^{m} f_k(\lambda) x_k^2 \quad \text{(B5)}$$

Then consistency requires

$$\rho(\lambda) = \lambda \quad \text{(B6)}$$

where

$$\rho(\lambda) = \frac{s^2 - \sigma^2(\lambda) - \sigma_E^2}{m_{4E} - \sigma_E^4 + 4\sigma_E^2\sigma^2(\lambda)} \quad \text{(B7)}$$

This is a nonlinear equation in one variable. A straightforward application of the Schwarz inequality shows that $\mathrm{d}\sigma^2/\mathrm{d}\lambda \geq 0$, and it is clear that $\mathrm{d}\rho/\mathrm{d}\sigma^2 < 0$. Hence $\rho(\lambda)$ decreases when λ increases. The quantities $\sigma^2(\lambda)$ and $\rho(\lambda)$ remain bounded for all values of λ; therefore the expression $\rho(\lambda) - \lambda$ is negative when λ is very large and positive when λ is very negative $(-\infty)$. So there will be a single root λ for Eq. (B6), which can be bracketed and accurately determined by methods like the van Wijngaarden-Dekker-Brent algorithm.[16] The PDF $\{f\}$ now is given by Eq. (B4) with λ as just determined. The parameter λ can be negative, zero, or positive.

When method B of Appendix A is employed to describe experimental uncertainty, σ_E^2 is neglected in the numerator of the right-hand side of Eq. (B7), and the denominator is replaced by (NV), which does not depend on λ.

On return to continuous variables x, the PDF takes the form of Eq. (17).

Appendix C: Truncated Normal Distributions

The symmetric zero-mean truncated normal distribution is given by Eq. (18).

The normalizing coefficient c is

$$c = \begin{cases} \left(|\lambda|/\pi\right)^{1/2}/\mathrm{erf}\left(|\lambda|^{1/2}\right) & \lambda < 0 \\ 0.5 & \lambda = 0 \\ 0.5\lambda^{1/2}\exp(-\lambda)/D(\lambda^{1/2}) & \lambda > 0 \end{cases} \quad \text{(C1)}$$

where D is the Dawson integral

$$D(x) = \exp(-x^2)\int_0^x \exp(t^2)\,\mathrm{d}t \quad \text{(C2)}$$

The relations between variance σ^2 and parameter λ are

$$(\sigma/A)^2 = \begin{cases} 0.5|\lambda|^{-1}\left[1 - 2\left(|\lambda|/\pi\right)^{1/2}\exp(\lambda)/\mathrm{erf}\left(|\lambda|^{1/2}\right)\right] & \lambda < 0 \\ 1/3 & \lambda = 0 \\ 0.5\left[\lambda^{-1/2}/D(\lambda^{1/2}) - \lambda^{-1}\right] & \lambda > 0 \end{cases} \quad \text{(C3)}$$

Appendix D: MAP Solution when Quantiles are Measured

We use the simple model

$$P\left(\{p\}/\{f\}\right) = C\exp\left[-0.5\sum_{k=1}^{m}(p_k - f_k)^2/V_k\right]\delta\left(1 - \sum_{k=1}^{m} p_k\right) \quad \text{(D1)}$$

where $\delta(\,)$ is the Dirac function.

C is a normalizing constant, independent of $\{f\}$. Again we could use another PDF like the multivariable gamma distribution to avoid negative p_k, but the end results would be similar, as in the case of variances.

The conditions for Bayesian maximum posterior probability are

$$f_k = q_k \exp(-1-\lambda_o)\exp\left[(p_k - f_k)/V_k\right]; \quad k = 1 \text{ to } m \quad \text{(D2)}$$

subject to Eq. (12).

The following procedure describes the solution method.

Set $B = \exp(-1-\lambda_o)$, to be determined. B is necessarily positive. The nonlinear system of equations (D2) for $\{f\}$ can be solved numerically as follows.

If B were known, each equation (D2) could be solved easily for f by the same algorithms as in Appendix B, because the scalar function

$$h(f) = f - qB \exp\left[(p-f)/V\right] \quad \text{(D3)}$$

varies monotonically with f and its single, positive root f^* is bracketed by $f = 0$ and f very large positive (∞). This constitutes the inner loop of the procedure for any fixed positive value of B.

It is easy to see that f^* increases from zero when B increases from zero. Therefore, the function $\Sigma_i f_i^* - 1$ increases from -1 to a large positive value when B increases from zero to a large positive value, and its zero, which gives the desired solution B, and hence $\{f\}$, is bracketed by $B = 0$ and B large positive, and can be found by the same algorithm in the outer loop of the procedure.

References

[1]Frieden, B. R., "Unified Theory for Estimating Frequency-of-Occurrence Laws and Optical Objects," *Journal of the Optical Society of America,* Vol. 73, July 1983, pp. 927–938.

[2]Kendall, M. G., and Stuart, A., *The Advanced Theory of Statistics,* Vol. I, 3rd ed., Griffin, London, 1969, Chap. 8.

[3]Jaynes, E. T., "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE,* Vol. 70, Sept. 1982, pp. 939–952.

[4]Tribus, M., *Rational Descriptions, Decisions and Designs,* 1st ed., Pergamon, New York, 1969, Chap. 5.

[5]Shore, J. E., and Johnson, R. W., "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory,* Vol. IT-26, Jan. 1980, pp. 26–37.

[6]Gull, S. F., and Daniell, G. J., "Image Reconstruction from Incomplete and Noisy Data," *Nature,* Vol. 272, April 20, 1978, pp. 686–690.

[7]Virdee, L., "Handbook of Pointing Errors and Their Statistical Treatment," European Space Agency, Contractor Rept. CR(P)-1363, Noordwijk, The Netherlands, June 1981.

[8]Fisz, M., *Probability Theory and Mathematical Statistics,* 3rd ed., Wiley, New York, 1963, Chaps. 2 and 10.

[9]Cramer, H., *Mathematical Methods in Statistics,* Princeton Univ. Press, Princeton, NJ, 1946, Chaps. 14 and 17.

[10]Coffey, T. C., "An Error Criterion for the Pointing of Axially Symmetric Spacecraft Payloads," *Journal of Guidance, Control, and Dynamics,* Vol. 6, No. 6, 1983, pp. 484–490.

[11]Lipton, A. H., "Alignment of Inertial Systems on a Moving Base," NASA TN D-4110, Sept. 1967.

[12]Alexander, C. H., "Simultaneous Confidence Bounds for the Tail of an Inverse Distribution Function," *Annals of Statistics,* Vol. 8, No. 6, 1980, pp. 1391–1394.

[13]Jaynes, E. T., "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics,* Vol. SSC-4, Sept. 1968, pp. 227–240.

[14]Spall, J. C., and Hill, S. D., "Least-Informative Bayesian Prior Distributions for Finite Samples Based on Information Theory," *IEEE Transactions on Automatic Control,* Vol. 35, May 1990, pp. 580–583.

[15]Higgins, W. T., "A Comparison of Complementary and Kalman Filtering," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-11, No. 3, May 1975, pp. 321–325.

[16]Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., *Numerical Recipes: The Art of Scientific Computing,* Cambridge Univ. Press, Cambridge, England, UK, 1986, pp. 251–254.