LBL-24811
UC-350

Presented at the Annual Illuminating Engineering Society Conference, Minneapolis, MN, October 3–6, 1988, and to be published in the Proceedings

Estimation of Linear Interpolation Error

R. Clear and S. Berman

February 1988

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

Presented at the Annual Illuminating Engineering Society Conference, Minneapolis, MN. February 1988.

L-159
LBL-24811
UC-350

Estimation of Linear Interpolation Error

Robert Clear and Sam Berman

Lighting Systems Research Group
Energy & Environment Division
Lawrence Berkeley Laboratory
Berkeley, California 94720

February 1988

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Building Technologies, Building Equipment Division of the U.S. Department of Energy under Contract No. DE-ACO3-76SFOOO98.

Estimation of Linear Interpolation Error

Robert Clear and Sam Berman

Lighting Systems Research Group
Energy & Environment Division
Lawrence Berkeley Laboratory
Berkeley, California 94720

Abstract

Linear interpolation is used to estimate missing values from candlepower, reflectance and other data tables. We have developed an error estimating program, and have validated it against known functions. Sample error estimates are presented. The procedures should be useful to both the developers and users of data tables.

Introduction

In this paper we briefly discuss three methods of estimating the error in linear interpolation, and then describe a computer implementation and validation of one of the methods. We present a summary of sample analyses of a selection of reflectance, and candlepower data tables suitable for roadway calculations, as an immediate application of our procedure. Average relative error estimates for the sample data sets ranged from 6 to 60%.

The roadway application was selected because the practice being currently developed for computerized roadway calculations uses linear interpolation of reflectance and candlepower data to generate any value that is not in a table.[1] However, the applications are more general. For instance, interpolation is specified in a number of the IES LM reports on Measurements, Testing and Calculations.[2,3,4,5,6] Linear interpolation is the simplest interpolation scheme and may therefore be prefered if it is adequately accurate. The use of linear interpolation is also sometimes implicit in other operations. Zonal constant methods of calculating total lumens are essentially equivalent to assuming bilinear interpolation between integration points.[5,6,7] In short, linear interpolation, either explicitly or implicitly, pervades much of a lighting engineers' work.

The purpose of our program is to provide information to both those who use data tables, and those who generate them. At present guidelines for the generation and use of tables are very loose. The IES approved method for photometric testing of fluorescent lamps notes that the "spacing should permit plotting of well determined candlepower distribution curves", but does not further define what this means.[2,7] The approved method for testing of floodlights specifies the number of points to be measured, but does not provide a guideline as to what accuracy can be expected from use of the resultant tables.[5]

The general error estimation methods described in the theory section can be used for higher order interpolation, and may be of more widespread interest than the remainder of the paper. The section following covers the procedure that we implemented in more detail. The validation section displays the results of comparisons between our estimated error and the actual error when the underlying function is a high order polynomial. The results section presents our error estimates for a selection of reflectance and candlepower tables. The potential significance of the error estimates is covered briefly in the conclusion.

Theory

The procedure we used to estimate errors was derived from the remainder theory of interpolation.[8,9] We assume that the interpolated function is bounded, continuous, and twice differentiable. The first two conditions are necessary both to insure that the error is bounded and for the function to be physically realizable. The last condition can be dropped if the errors are estimated instead by self interpolation or comparison against a fit. These latter methods have their own constraints and problems and we did not develop a procedure based on them. We describe them and their problems briefly at the end of this section.

We only describe the rationale behind the remainder estimate, the reader should refer to the references for proofs. Let f(x) be a function whose values are known and tabulated at the points $x_0, x_1, x_2, ..., x_n$, where $x_0 < x_1 < x_2 < ... x_n$. If $P_1(x)$ is the linear (interpolating) function that matches f(x) at x_0, and x_1, then for any given z in the interval between x_0, and x_1, there exists a point ξ in the interval, such that f(z) can be expanded as follows:

$$f(z) = P_1(z) + f''(\xi) * (z-x_0) * (z-x_1) * (1/2) = P_1(z) + E(z),$$

where E(z) is the remainder, or error term, and the symbol "*" stands for multiplication.

The remainder estimate is a consequence of Rolle's theorem, which essentially is a statement that you cannot get from $f(x_0)$ to $f(x_1)$ on a continuous differentiable curve without finding a point somewhere on the interval where the slope is equal to the average slope.[8]

If $f''(x)$ is known then the maximum value of $f''(x)$ over the interval x_0 to x_1 gives a rigorous maximum error bound. In the situations we are interested in, the best we can do is use the numerical second derivative as an estimate of $f''(x)$. Let $x_1 - x_0 = h$, and $x_2 - x_1 = h*k$, then the numerical second derivative, $\Delta^2 f(x_0)/h^2$ is:

$$\Delta^2 f(x_0)/h^2 = 2 * [f(x_2) - (1+k) * f(x_1) + k * f(x_0)]/[h^2 * k * (1 + k)].$$

The mean value theorem guarantees that $\Delta^2 f(x_0)/h^2$ is equal to $f''(\Xi)$ for some Ξ in the interval x_0 to x_2, but it does not guarantee that $f''(\Xi)$ is the maximum value of $f''(x)$ in the interval x_0 to x_1, or even that it equals a value of $f''(x)$ that would ever be used in the remainder.[9]

The point x where the remainder estimate is largest is $x = (x_0 + x_1)/2$. With the substitutions described we get an estimate for the maximum error of linear interpolation over the interval as:

$$E_{max} \approx |\Delta^2 f(x_0)/8|.$$

Replacing $f''(x)$ with $\Delta^2 f(x_0)/h^2$ is equivalent to estimating the error from the difference between linear and quadratic interpolation. We are <u>not</u> suggesting that linear interpolation should be replaced by quadratic or higher order interpolation. Higher order interpolation will have errors proportional to higher order derivatives, and as figure 1 shows it sometimes gives poorer results than linear interpolation. The point is that the error in linear interpolation is approximately equal to the difference between linear and quadratic interpolation, regardless of the relative accuracy of the two types of interpolation.

It is important to recognize that E_{max} is just an estimate, and not a true upper bound. If the functional values chosen are the zeros of a periodic function then our error estimate will be zero, even though the real error will be arbitrarily large. There is no procedure that can get a reasonable estimate from such poorly selected data, and the user must exercise some common sense when using this or any other estimation procedure.

The reader should note that $\Delta^2 f$ is defined for a continuous function even if f" is not. Although the formal derivation of the error estimate no longer holds in this case, we found that the error estimates were well behaved. We nevertheless looked at, and rejected, two other procedures for error estimation: self interpolation, and least squares.

An intuitive, and very simple procedure for estimating error is to self interpolate at an existing point in a data table and compare the estimated value against the listed value. For example if there are three data points,

$f(x_0)$, $f(x_1)$, and $f(x_2)$, then we would use linear interpolation of $f(x_0)$ and $f(x_2)$ to get an estimate $P_1(x_1)$ for f at x_1. The difference $|f(x_1) - P_1(x_1)|$ is an indication of the absolute error of linear interpolation over these points.

Manipulation of the above error estimate shows that it is actually equal to $k * \Delta^2 f(x_0)/2$. Thus for equally spaced data (k = 1) this error estimate is four times larger than the remainder estimate. For $k \neq 1$ the self interpolated estimate is arbitrarily different from the remainder estimate. If a relative error is calculated at the interpolation points (x_1 and $(x_0 + x_1)/2$ respectively) the estimate is arbitrarily different even for a fixed k. Since self interpolation is less accurate than the remainder estimate, we saw no reason to further consider it.

Least squares fitting techniques offer what appears to be another attractive alternative to remainder estimates. In practice, however, we found that either the fits are so good that there is no real reason to use interpolation of a data table, or the fits are not very good and thus neither are the error estimates. In addition, making sure the fits are good can be an extremely difficult task, and often requires considerably more knowledge than a bare bones table of values.

In the standard least-squares program the user must specify a functional form, and the program estimates coefficients. If the fit is to a polynomial or orthogonal expansion the user must provide a stopping criteria for when further terms are no longer significant. If the expansion or functional form chosen is not appropriate to the problem it can be very difficult to decide when the fit is "good" enough. In addition, although the procedure guarantees that the fit has minimum variance over the data or "net" points, it is a minimum only with respect to the given basis set, and for some basis sets there is no guarantee about the intervals between the net points.[9]

The limit where the number of coefficients equals the number of data points is a special case called orthogonal or Fourier expansion. The same problems apply as with normal least-squares, along with the added one that there is no variance calculation in this limit. A strictly practical problem is that the use of non-equal intervals between points makes it impossible to use the standard Fourier series techniques, as the normal basis sets are not orthogonal over a net of points with uneven spacing. The reader should also be aware that the standard trignometric Fourier expansion converges more slowly to functions that are not periodic.

Least squares techniques are in a sense a "best" way of dealing with data, but it did not appear to be easy to develop a easy, general, and robust procedure that could be used to analyze arbitrary data sets without having to know more than is given by the data itself.

The Errterp program

The remainder estimate described above provided the basis for the Pascal program "Errterp" that we wrote to make error estimates. However, in the actual program we made a number of modifications to improve the estimates, and to adapt them to the problem of error estimation in two dimensions.

The first modification is to note that the ordering of points is arbitrary, and therefore if there are points from x_0 to x_3 it is equally valid to estimate the error for the interval x_1 to x_2 from $\Delta^2 f(x_1)$ as it is from $\Delta^2 f(x_0)$. The program takes the maximum of the two estimates where both exist.

A second modification arose from our awareness that $\Delta^2 f(x_0)/h^2$ is ideally an estimate of f" at the point $x = [\, x_0 + x_1 + x_2 \,]/3$. If the spacing is uneven x may not lie in the interpolation interval, x_1 to x_2. When $\Delta^2 f(x_0)$ and $\Delta^2 f(x_1)$ both exist, and one or both is an estimate for a point outside the interpolation region, Errterp interpolates between the two values to get a better estimate for the interpolation region.

Most of the problems for illuminating engineers involve bilinear interpolation over functions of two variables. This can be formally converted into two-one dimensional problems.[8] The error estimate is therefore a sum of the partial second derivatives in x and y. In the Errterp program we separately summed positive and negative partials and kept the maximum of the two sums. This estimates the maximum error even if the errors on the x and y boundaries have opposite sign and tend to cancel inside the x-y interpolation region. It should be noted that these above changes make the error estimate no longer equivalent to simply taking the difference between linear and quadratic interpolation.

At the option of the user, Errterp can make additional assumptions about the function that were not used in the derivation of the standard remainder estimate. For instance, candlepower and reflectance values are always positive. Telling Errterp that the function is positive forces it to limit negative errors to the size of the interpolated value. A second option tells Errterp to check for possible asymptotic behavior at the boundaries of the table. This would be appropriate if candlepower data far from the beam maximum, or reflectance data well away from the specular region, approaches a constant diffuse limit. If this option is selected, Errterp assumes that there is an asymptote when three points coming off of a boundary lie on a monotonic curve, and the magnitude of the slope is smaller between points one and two than between two and three, where point one is closest to the boundary. The slope of a quadratic function passing through these three points changes sign between points one and two and thus violates the asymptotic edge assumption if $| \Delta^2 f(x_0)/8 | > | \Delta f(x_0)/4 |$. Errterp therefore takes the smaller of these two estimates.

The last option tells Errterp to look for local smoothness (convexity or concavity over an interval). Possible examples are reflection from a near Lambertian surface, or a candlepower distribution from a diffusing fixture. If this option is selected Errterp assumes that the function is locally smooth over the interval x_j to x_{j+3}, when both $\Delta^2 f(x_j)$ and $\Delta^2 f(x_{j+1})$ have the same sign. The quadratic function passing through points x_j to x_{j+2} is not consistent with the assumption of local smoothness if $\Delta^2 f(x_j)/8$ is greater than the difference between the interpolated point in the interval x_{j+1} and x_{j+2} and linear extrapolation from points x_{j+2} and x_{j+3}. This is shown in figure 2. The same test applies to the quadratic function passing through points x_{j+1} to x_{j+3}. Errterp takes the maximum estimate consistent with the local smoothness assumption.

The Errterp program was actually designed to estimate the overall error. The user is prompted to enter a relative precision (zero implies no relative error), and an absolute precision (number of significant figures), for the input data. A zero entry in response to the latter prompt forces Errterp to count the number of digits for each data point and assume that each digit is significant. The final reported error figure is the sum of all the errors.

The Errterp program was designed to read candlepower table data in the IES standard file transfer format.[10] It can also read data in a second less specialized format. The second format can handle tables where the final rows have fewer columns of data than the first rows. This permits it to analyze the roadway reflectance data tables.

For most purposes relative errors are of more interest than absolute errors. However, with candlepower data, the user is usually more interested in the accuracy of the data near the candlepower maximum than elsewhere. Errterp calculates relative errors with the interpolated values as the denominators and outputs a summary table of the average error for all values, and then for only those interpolated values which are within 10%, 20% or 50% of the maximum interpolated value. The summary table also includes the magnitude and location of the maximum relative error in each of these classes. Finally, Errterp can also output the individual relative error values.

The 10/10/87 version of the Errterp program was released to roadway committee members. It runs on IBM PCs and compatibles with at least 256K of memory, and one disk drive. There are two run files, one which uses the 8087 math accelerator chip, and the other which cannot. An updated version will be available from the authors.

Validation and Calibaration

As noted earlier Errterp computes an estimate of the linear interpolation error. After trying to make sure that the program computed what it was supposed to compute, there was still the question of how close the remainder estimate is to the real interpolation error. To get a handle on this question, we had a program generate data tables from both one and two dimensional polynomials from quadratic up to tenth order. The coefficients were chosen randomly within sets of constrained ranges. These ranges were adjusted variously so that we could look at a set of runs where the function was always positive over the domain, or it was sure to have at least one maximum or minimum over the domain, or it was likely to have multiple maxima and minima. A second, slightly shifted domain with uneven intervals was also used with each generated polynomial to get a wider range of results.

Since the data tables were generated from known polynomials it was possible and relatively easy to calculate some actual error values. We calculated two sets of values for each table. The first set of values was simply the set of relative errors at the midpoint of each interval. This point was chosen as being the same point for which Errterp calculates the error, and the point which is likeliest to have the largest absolute error. In addition to knowing maximum errors it is also useful to know average errors. Therefore, a second set of values was derived from the root-mean-square (rms) average error over the interval divided by the interpolated value at the center point of the interval. A fixed value for the denominator was used in the average calculation to make it possible to analytically compute the rms error, and to make the values more comparable to those calculated by Errterp. If the average of the ratios was calculated this would give undue weight to portions of the interval where the interpolated value is small, and in fact would return an infinite average error if the interpolated value is ever zero when the real functional value is not.

The rms error calculation is also useful as a check on how well the midpoint calculation represents the maximum error. The midpoint error calculation was lower than the rms error calculation (albeit occasionally substantially) in only a few percent of the cases . Most of the time it is an estimate of the upper range of the error in an interval.

In all, a sample of 48 different runs was analyzed. Four of these were runs with one dimensional linear interpolation and seven tested the various smoothing options described earlier. These eleven runs were mostly tests for programming errors, and in fact a small bug in the asymptote routine was found in the 10/10/87 version of the program.

The thirty-seven remaining runs were tests of bilinear interpolation with no smoothing options. Twenty-two runs were tests of strictly non-negative data, and therefore directly correspond to most illuminating engineering problems. The other fifteen runs covered both negative and positive values. Table I summarizes the results. Since Errterp computes a summary average error estimate, the table shows the ratios of the average of the calculated errors to the average estimated error in addition to showing the average of the ratios. The second type of average weights those points with large calculated to estimated ratios more strongly than other points. The first type of average strongly weights points which individually have high estimated or calculated errors.

Table 1 shows that on average the rms error is about one-half the Errterp error estimate. For non-negative data this value is fairly stable and is not sensitive to when the average was taken. The variability in the ratios for the mixed data shows that in these cases the actual errors can be small while the estimated errors are still large.

Errterp appears to do a reasonable job of estimating the upper range of the error, but as can be seen it does underestimate occasionally. For non-negative data the fraction of points where the estimated error was less than the calculated error was small, being about 1.5% for the midpoint calculation and about 0.5% for the rms calculation.

As might be expected, Errterp did its best job of estimation for low order polynomials, and its worst for high order polynomials. With high order polynomials, Errterp was more likely to both underestimate one error, and then substantially overestimate another error. The above comments, albeit less strongly, also apply to uneven, versus even, spacing.

The comparisons in Table 1 are indicative of the type of behavior to be expected from Errterp. The reader should bear in mind that the comparison was limited to polynomials, and is not necessarily a random sample of the types of functions that are likely to be found in engineering practice. We feel, however, that it provides a reasonable first guide to interpretation of the Errterp error estimates.

Results

We ran Errterp on a number of roadway fixture candlepower tables, roadway reflectance tables, and the sample files from the IESNA guide, LM-63, to get a handle on the size of errors to be expected with photometric data tables.[10] In the future we hope that manufacturers will use Errterp to provide an estimated error with the photometric file. Most of the data was taken from a disk supplied to IESNA roadway committee members for testing purposes by Merle Keck, the chairman of its Standard Practice subcommittee. We analyzed a selection of twenty-two candlepower tables, representing roadway fixtures from five manufacturers, plus 10 different roadway reflectance tables.

The four roadway reflectance "R" tables are in common use, and we present the results for them separately in table 2.[11] The remaining data is mostly generic data that is only useful as guide. The results for this data is summarized in figures 3a and 3b. For the analyses in table 2 and in figures 3a and 3b, no smoothing options were chosen; the absolute error was calculated from the number of significant figures in the data itself, and the relative error was set to 0.5% of the listed values. For several data sets input values of "0" had to be changed to "0.0" to get the correct absolute error.

Errterp groups the error estimates by the relative size of the interpolated value, s. The groupings in the tables presented here are not the same for the maximum and average. The average values were recalculated from the Errterp output to show the error within the classes.

In general, table 2 and figures 3a and 3b, show that there is a commendably small error in regions of the data tables where the values are large. At the same time the program shows that individual errors may be large, even near the candlepower or reflectance maximums. If Table 1 is any guide, rms errors are about 1/2, and maximum errors are about 2/3 the Errterp estimates in this region of the tables. The considerably larger Errterp estimates for the smaller values in the table is mostly due to round-off error. For example, a set of entries such as "1", "2", and "5" are specified only to within 10 to 50% precision. The numerical second derivative through these numbers is only accurate to ±80%. In short, the relative accuracy for interpolation with this degree of round-off is very poor. The very large error estimates for the smaller values in many tables indicate that calculations over these regions should be viewed with some caution.

Errterp was rerun for a number of the data tables listed above with both the asymptotic edge, and local smoothing options on. To keep things consistent we calculated the ratios between the smoothed error estimates and the basic error estimates. Table 3 summarizes these comparisons. The effect of assuming a smoother function was usually slight.

Conclusion

The validation runs indicate that Errterp does a reasonably consistent job of estimating the nominal error bound level. Errterp overestimated the general rms error level by a factor of 1.5 to 3.5 for non-negative validation data set. Tests of a sample of real data sets indicate that they maintain an estimated general rms level of accuracy of 1 to 15% as for data entries that are within 10% of the maximum size of entries in a table. Individual points may have substantially higher error potentials, and the general error level is also much higher for points which are relatively small.

The Errterp program is a tool for estimating the inherent accuracy of many types of calculations with data tables. It can be used by table developers to assure that enough measurements are made to provide a nominal accuracy level. As endusers, we are using it to provide an estimate for the accuracy of our computer calculations, and by inference a guide to when they no longer make sense. In addition, the nominal error estimates provide a target accuracy for the development of fitting equations.

Acknowledgement

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Building Technologies, Building Equipment Division of the U.S. Department of Energy under Contract No. DE-ACO3-76SFOOO98.

References

1) Ian Lewin. 1988. Personal communication with chairman of the roadway subcommitee on Measurements and Calculations.

2) LM-10. 1975. IES Approved Method for Photometric Testing of Outdoor Fluorescent Luminaires.

3) LM-27. 1967. IES Approved Method for the Photometric Testing of Filament-Type Luminaires for General Lighting Service.

4) LM-34. 1970. General Procedure for Calculating Maintained Illumination.

5) LM-35. 1971. IES Approved Method for Photometric Testing of Floodlights Using Incandescent Filament or Discharge Lamps.

6) LM-36. 1971. IES Practical Guide to Photometry.

7) LM-41. 1972. IES Approved Method for Photometric Testing of Indoor Fluorescent Luminaires.

8) Germund Dahlquist, and Åke Björck. Translated by Ned Anderson. 1974. Numerical Methods. New Jersey, Prentice-Hall.

9) Philip J. Davis. 1975. Interpolation & Approximation. New York, Dover Publications.

10) LM-63. 1986. IES Recommended Standard File Format for Electronic Transfer of Photometric Data.

11) John E. Kaufman and Jack Christensen, eds. 1987. Roadway Lighting, section 14: IES Lighting Handbook, Application Volume. New York, Illuminating Engineering Society of North America.

Table 1

Summary of comparison measures between calculated and estimated interpolation errors.

	Non-negative data			Mixed data		
Measure	Avg. (Median)	Max.	Min.	Avg. (Median)	Max.	Min.
Averages:						
<mid/est>	0.70	0.98	0.40	0.64	0.89	0.40
<mid>/<est>	0.69	0.98	0.37	0.54	0.88	0.14
<rms/est>	0.52	0.71	0.30	0.48	0.64	0.31
<rms>/<est>	0.51	0.72	0.28	0.40	0.64	0.12
Maximums:						
mid/est	1.04(.98)	1.8	0.91	1.6(.98)	3.7	0.89
rms/est	0.75(.71)	1.3	0.64	1.2(.71)	2.7	0.64
Minimums:						
mid/est	0.51*	0.98	0.11*	0.40*	0.72	0.03*
rms/est	0.35	0.71	0.09	0.17	0.51	0.01
Ratio Max/MIn:						
(rms/est)	3.2(2.5)	9.7	1.0	24.(5.3)	91.	1.40

* Only those midpoint values that were higher than the corresponding rms values were considered in the tabulation.

mid = midpoint error calculation.
rms = root-mean-square error calculation
est = Errterp error estimate

The <> stands for an average. In the bulk of the table, values in () are medians.

Table 2

Summary of Errterp percentage error estimates for the R-table roadway reflectance data grouped by the relative size of the interpolated values (s).

R-Table	Average Errors (%) Relative size of interpolated values (s)			
	s<10%	10%<s<20%	20%<s<50%	50%<s
R1	9.5	3.8	3.0	1.9
R2	17.9	5.2	4.3	2.2
R3	16.4	6.3	4.4	2.0
R4	20.4	10.3	4.9	2.6

R-Table	Maximum Errors (%) Relative size of interpolated values (s)			
	All s	10%<s	20%<s	50%<s
R1	45	8	7	7
R2	99	14	13	5
R3	99	24	18	5
R4	100	48	16	9

s = midpoint interpolated value / maximum midpoint interpolated value in table.

Table 3

Ratio of Errterp error estimates with smoothing to Errterp error estimates without smoothing, grouped by the relative size of the interpolated values (s).

VALUE	Ratio of Average Errors (%) Relative size of interpolated values (s)			
	s<10%	10%<s<20%	20%<s<50%	50%<s
(14 data sets)				
Maximum	100	100	100	100
Minimum	62	92	84	84
Average	80	97	95	93
Median	77	97	97	94

VALUE	Ratio of Maximum Errors (%) Relative size of interpolated values (s)			
	All s	10%<s	20%<s	50%<s
Maximum	100	100	100	100
Minimum	52	78	78	78
Average	78	96	96	94
Median	81	100	100	100

s = midpoint interpolated value / maximum midpoint interpolated value in table.

Figure 1. The linear interpolation error estimate

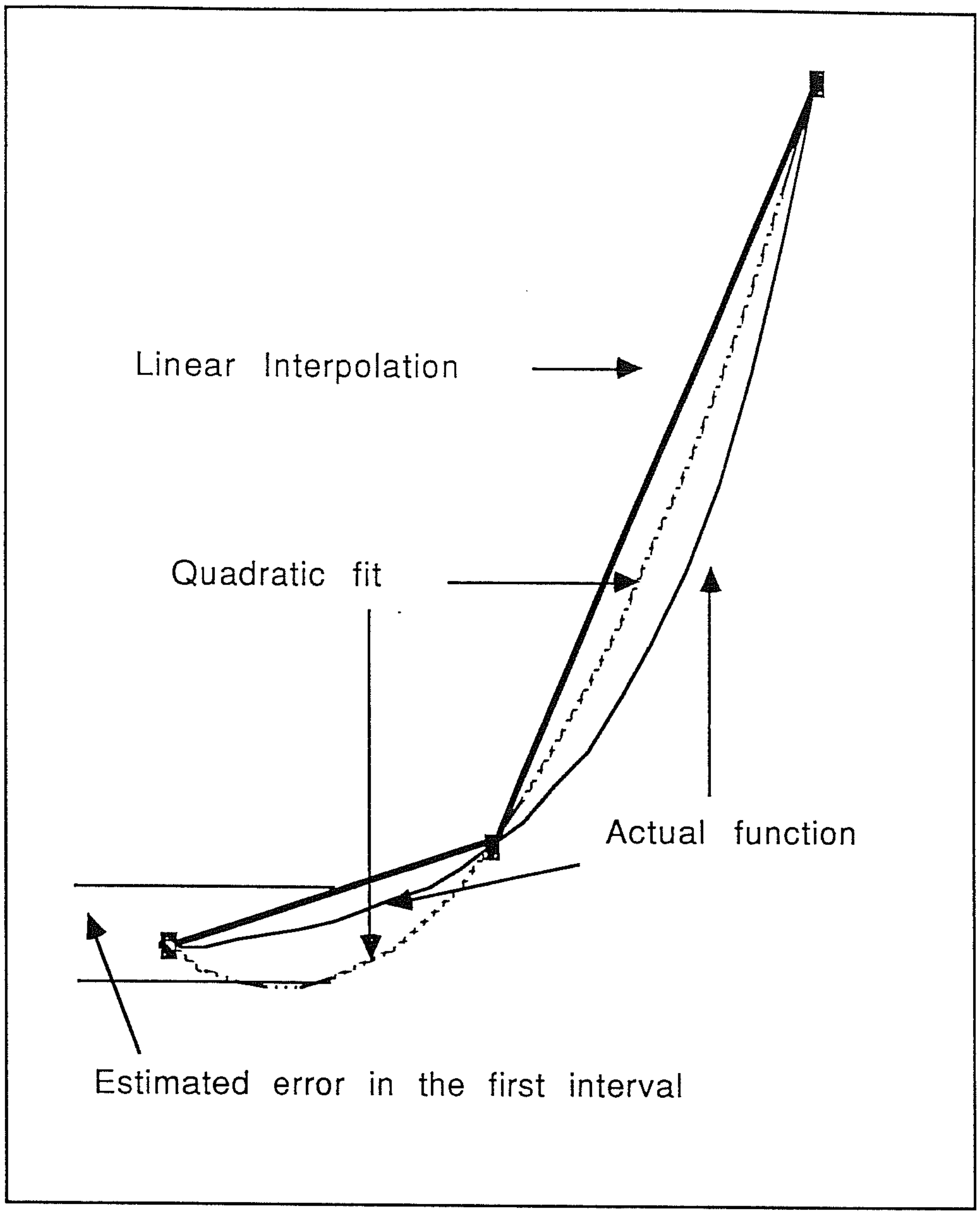


Figure 2: Linear Extrapolation as an Alternate Error Bound to Quadratic Interpolation for Locally Smooth Function

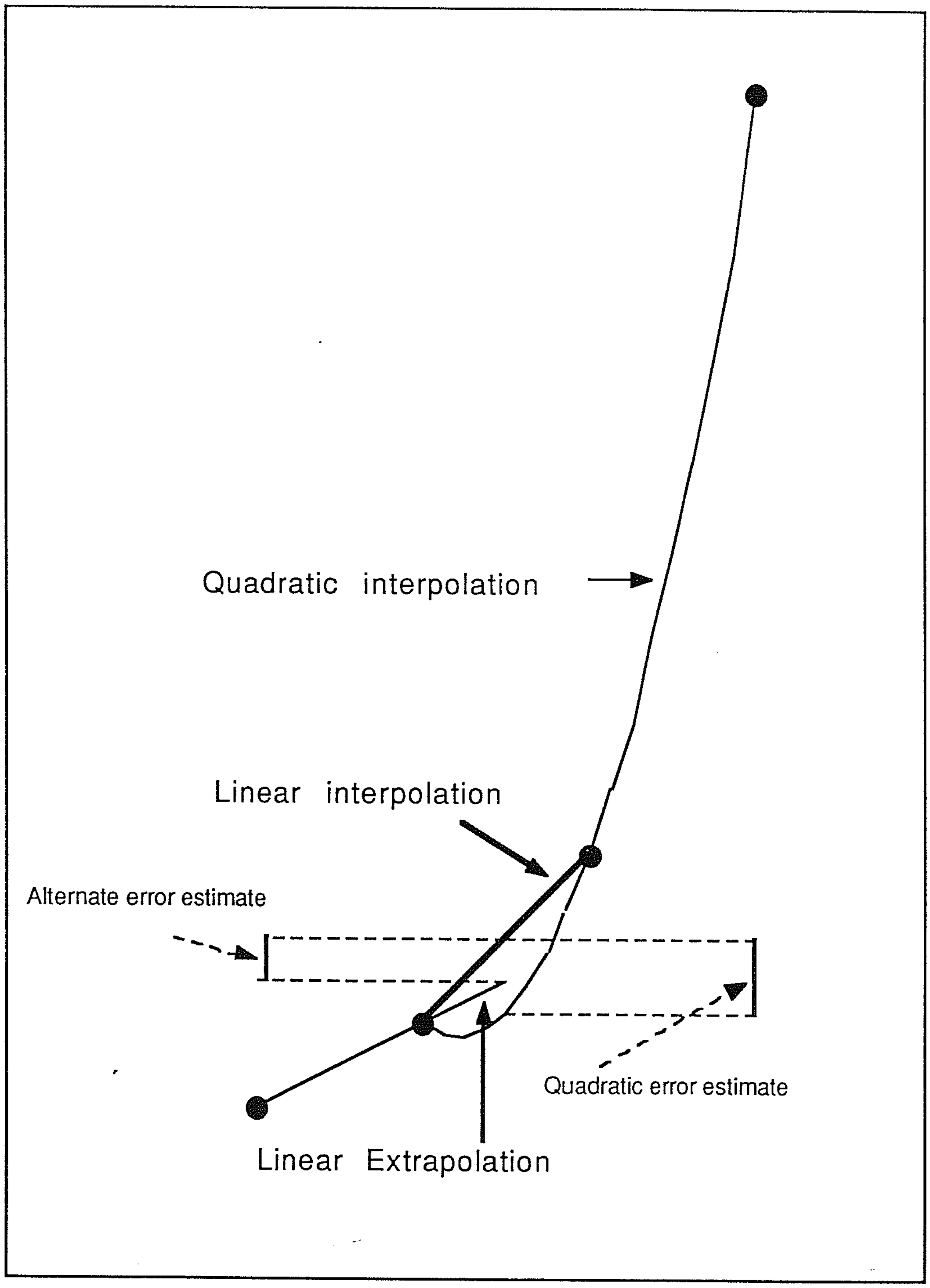
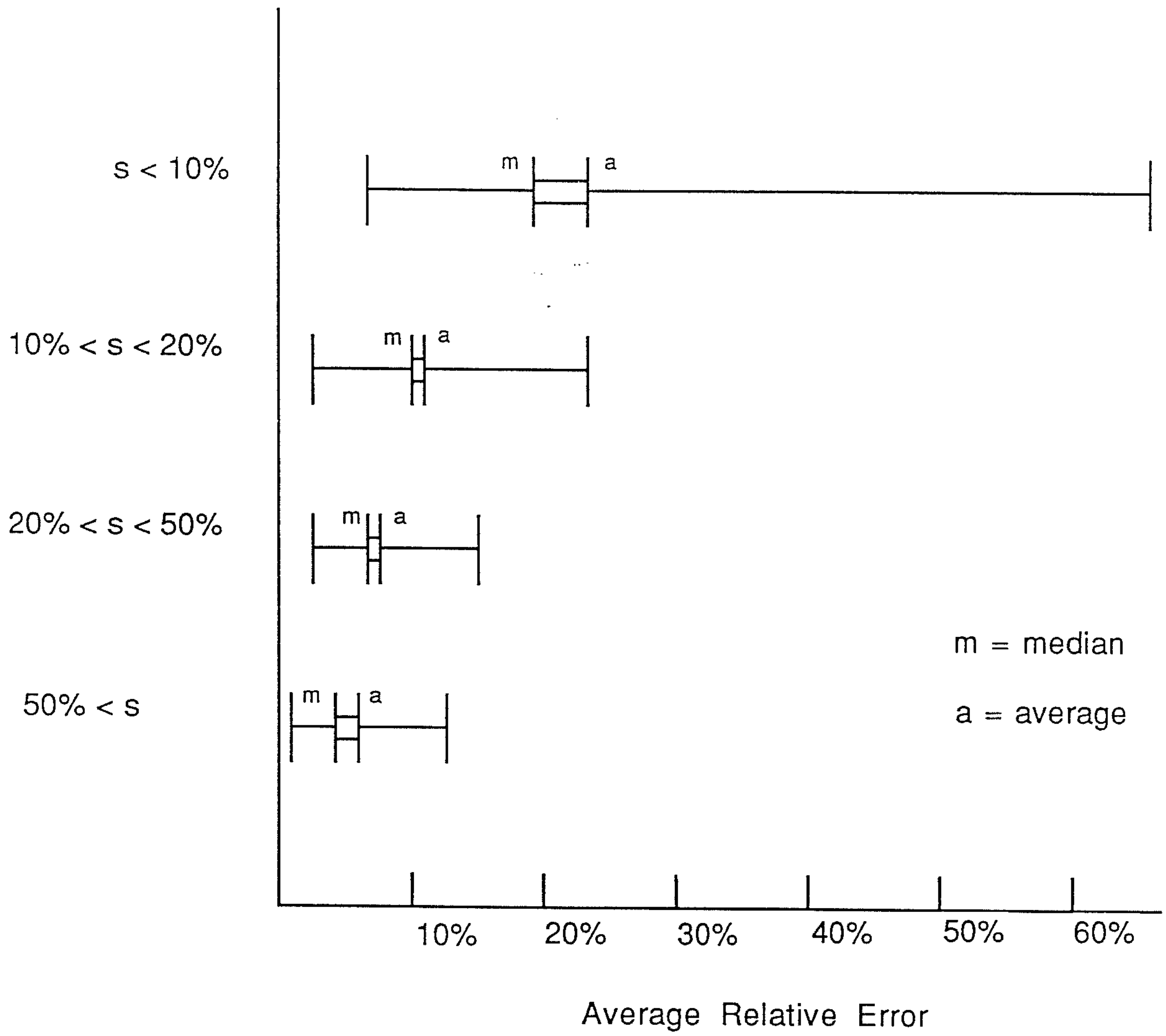


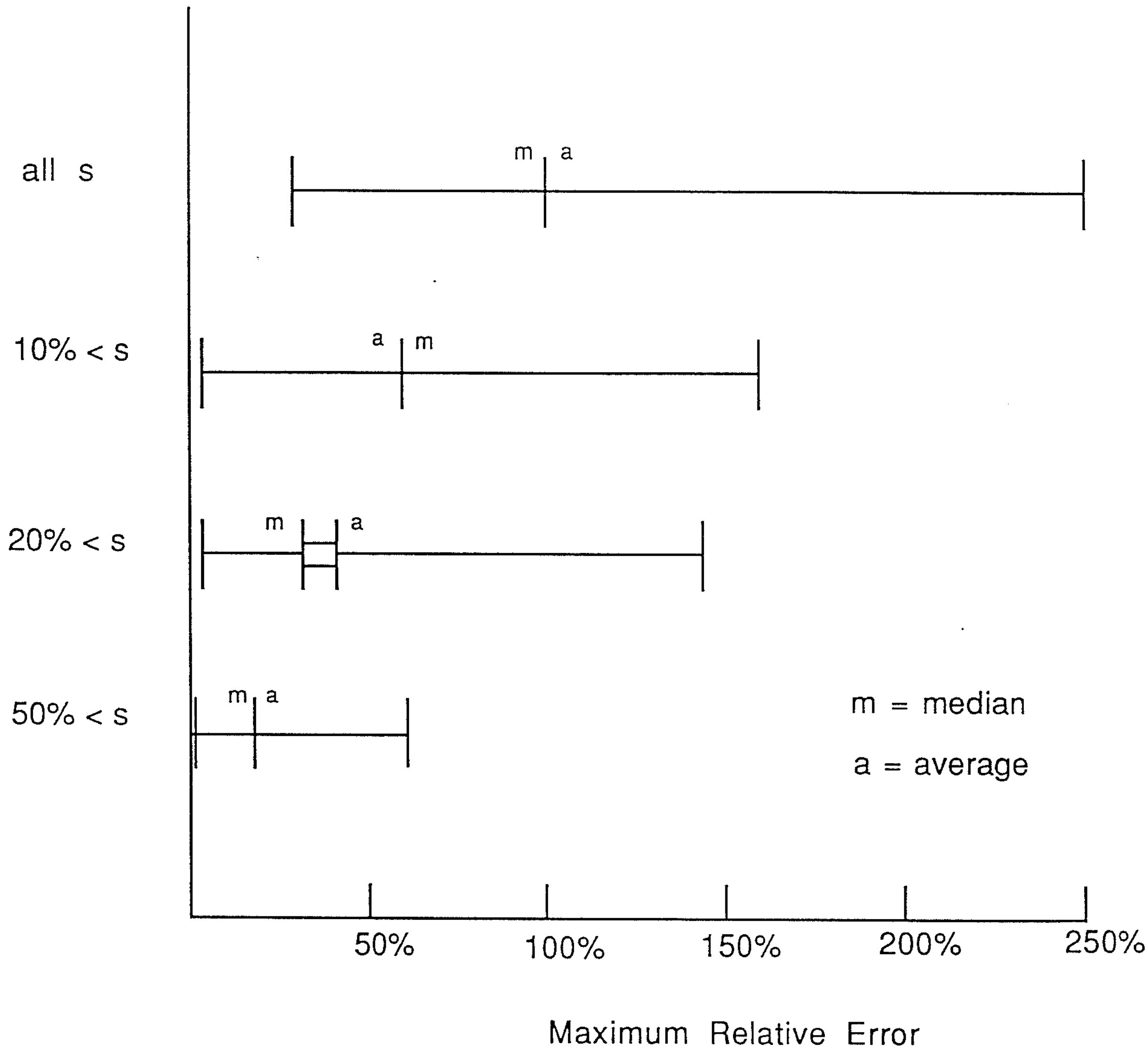
Figure 3a. Summary of average percentage error estimates grouped by the relative size of the interpolated values (s).



s = midpoint interpolated value / maximum midpoint interpolated value in table.

The left and right ends of the bars represent the minimum and maximum values, respectively. The hatch marks marked "m" and "a" are the median and average.

Figure 3b: Summary of maximum percentage error estimates grouped by the relative size of the interpolated values (s).



s = midpoint interpolated value/ maximum midpoint interpolated value in table.

The left and right ends of the bars represent the minimum and maximum values, respectively. The hatch marks marked "m" and "a" are the median and average.