ASAP*prime*® Software Technical Notes
Error Bar Calculations

This technical note describes the process of calculating error bars used in ASAP*prime*® (v6 and later). These error bars are the basis for the probability calculations.
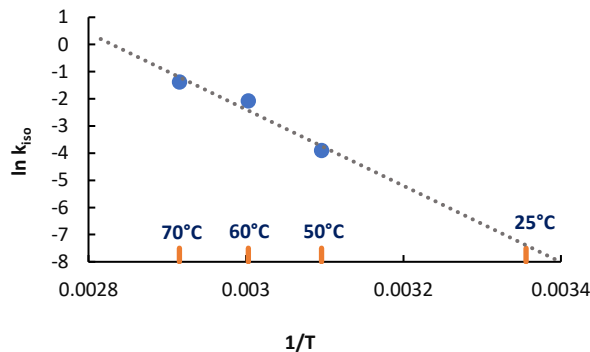
Error of the Fit

When data are fit to a mathematical function, a linear or non-linear least squares regression analysis can be performed to minimize the distance between the best fit of the function and each of the values (the residual). In this process, the regression line (for linear functions) is the line that minimizes the sum of squared residuals. There are different ways to capture the error bars for that fit. In this Technical Note, the process used in ASAP*prime*® for taking the isoconversion times at each condition to determine the error bars for fitting the Arrhenius or modified Arrhenius equations is described. These error bars are then used to determine the corresponding probabilities based on that distribution.

For simplicity, the situation of temperature only will be considered here, i.e., no relative humidity term, since the expansion to add the relative humidity term should be straightforward to understand. Also, for simplicity we will discuss the growth of degradant but recognize that loss of potency will be handled in an analogous manner. This means that we are fitting a linear function of ln $[(deg_{spec}-deg_{init})/t_{iso}]$ (ln k_{iso}) vs. 1/T, where deg_{init} is the amount of degradant at time zero, deg_{spec} is the specification limit, t_{iso} is the isoconversion time (time to fail) and T is the stress temperature in Kelvin. Fitting the line provides two parameters: ln A (the intercept) and E_a/R (the slope). For simplicity, we can consider the case of just three conditions and with the deg_{init} = 0 and deg_{spec} = 0.5% (Table 1 and Figure 1).

Table 1. Example data set for error bar calculations.

Temp. (°C)	1/T	t_{iso} (days)	ln k_{iso}	ln k_{iso} [best fit]	Residual Squared
50	0.003096	25	-3.912	-3.739	0.030
60	0.003003	4	-2.079	-2.433	0.125
70	0.002915	2	-1.386	-1.203	0.034
Average	**0.003005**		**-2.459**		0.189 (sum)

Figure 1. Plot of example data set.

The fitted line represents the lowest sum of the square residuals for each of the three points. The mean square error (MSE) will equal the square root of the sum of the square residuals divided by the number of degrees of freedom. The square residuals are shown in Table 1. With three points and two parameters, there is only one degree of freedom. This means that the MSE equals 0.593/1 = 0.593.

To calculate the error bars for the fit, we need to calculate the standard error (SE) in the ln k_{iso}. This is determined using the following equation:

$$ SE = \sqrt{\frac{1}{n-2}\left\{\sum (lnk_{iso}-\overline{lnk_{iso}})^2 - \frac{[\sum(\frac{1}{T}-\overline{\frac{1}{T}})(lnk_{iso}-\overline{lnk_{iso}})]^2}{\sum(\frac{1}{T}-\overline{\frac{1}{T}})^2}\right\}} $$

Here n is the number of conditions and the bars over a value indicate the average of that value. For the example in Table 1, SE = 0.437. This leads to the error bars for the isoconversion rates as shown in Figure 2.

Figure 2. Error bars representing the error of the fit calculation for the example data from Table 1.
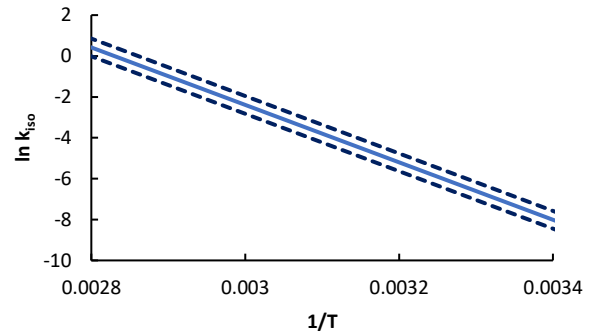
From the error bars in the logarithm of the isoconversion rates, the degradation curve, with error bars, can be calculated at 25°C, as shown in Figure 3. Even though the fit to the data is relatively good (R^2 = 0.94), there is still a wide range of possible degradant levels due to the small number of points used and the translation of logarithmic to normal rates.

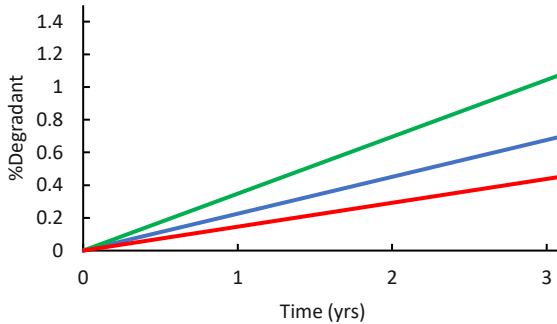FreeThink Technologies, Inc. +1 860 237 5800 **www.freethinktech.com**

ASAP*prime*® Software Technical Notes
Error Bar Calculations

Figure 3. Distribution of potential degradant levels as a function of time at 25°C based on the error of the fit for the example data set. The green and red lines represent plus/minus one standard deviation rates vs. the blue line (mean prediction).

Figure 4. Plot of example data set used to illustrate propagated calculation of error bars.

As can be seen, the errors of the points are much larger than any error from the fit itself. We can generate a random distribution of isoconversion times for each temperature with the indicated average and standard deviation. We can use a total of 50 simulations as illustrated in Table 3 (ASAP*prime*® defaults to 2500 simulations).

Error Propagation from Isoconversion Times

Another situation can also occur, especially with only a small number of conditions. If the points lie directly on a line, there will be no error bar at all from the error of the fit. However, since each isoconversion time has an error bar, i.e., each point represents a part of a distribution, another type of error calculation becomes important to reflect the error bars of the isoconversion times themselves.
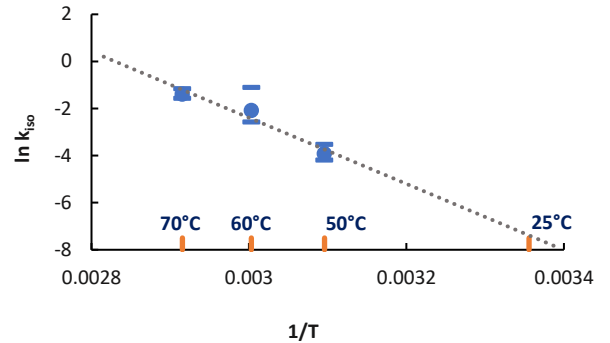
The error bars for isoconversion times reflect the error bars from the data (i.e., amount of degradation at a single time point and condition) coupled with the amount of extrapolation each requires to hit the specification limit. We can envision three scenarios: (1) The error bars for the isoconversion times are much larger than the error bars for the fit; (2) the error bars for the fit are much larger than the isoconversion time errors; and (3) both errors are comparable.

In calculating the error bars for degradation at lower temperatures propagating from the isoconversion time error bars at accelerated conditions, the fact that each isoconversion error bar is likely to be different makes a closed form solution problematic. Because of this, a Monte-Carlo simulation is carried out in ASAP*prime*® to determine this error bar. If we assume three conditions with isoconversion times corresponding to a perfect Arrhenius fit but now incorporate varying error bars at each condition (Table 2 and Figure 4), we can illustrate how this works.

Table 2. Example data set used to illustrate propagated calculation of error bars.

Temp. (°C)	t_{iso} (days)	SD
50	25	8
60	4	2.5
70	2	0.4

Table 3. Monte-Carlo simulation for example data from Table 2.

#	50°C	60°C	70°C
1	20.8	3.4	1.8
2	32.9	0.3	2.6
3	22.7	6.4	2.5
4	7.5	8.0	1.9
5	19.6	7.1	2.1
6	32.2	4.6	2.1
7	24.4	6.1	2.1
8	35.0	3.6	2.5
9	18.9	1.2	1.4
10	18.2	4.6	1.9
11	27.5	7.1	2.1
12	21.1	6.1	1.7
13	23.8	3.8	1.9
14	22.8	7.5	2.4
15	13.8	1.7	1.6
16	22.9	0.7	2.1
17	29.1	8.5	2.3
18	19.8	6.1	2.2
19	15.9	5.6	2.3
20	23.4	4.3	2.2
21	32.9	5.7	1.8
22	24.1	3.9	2.1
23	18.3	8.2	2.4
24	26.6	3.1	2.1
25	32.4	8.3	2.5
26	18.2	4.5	2.3
27	16.7	2.7	2.5
28	47.7	5.7	1.3
29	31.0	1.9	2.1
30	24.1	5.0	2.4

FreeThink Technologies, Inc.   +1 860 237 5800   **www.freethinktech.com**
Copyright 2019

2

|			
31	24.5	0.9	1.8
32	19.0	5.3	1.7
33	16.3	2.1	2.2
34	16.5	0.7	2.4
35	46.0	1.4	1.9
36	43.1	3.8	2.5
37	16.5	5.3	2.0
38	30.2	4.3	2.0
39	25.9	8.6	1.8
40	29.2	6.3	2.0
41	34.2	5.1	1.9
42	34.9	1.8	1.8
43	20.9	0.8	1.6
44	24.5	7.1	2.7
45	26.0	1.2	1.8
46	33.0	3.7	1.7
47	21.8	5.2	2.6
48	34.7	4.2	2.2
49	34.1	3.0	1.6
50	23.6	1.0	1.6
Avg.	25.6	4.4	2.1
SD	8.1	2.4	0.3

The 50 simulations have averages and standard deviations that are close to the targets from Table 2 and are expected to get even closer with greater numbers of simulations. For each of the 50 simulations, the isoconversion times can be fit to the Arrhenius equation, and the resulting parameters for ln A and E_a provide a distribution of values which can be approximated as a normal distribution. The results are shown in Table 4.

Table 4. Results of fifty-iteration Monte-Carlo simulation for data from Table 3.

|	Monte-Carlo Mean	Monte-Carlo Standard Deviation	Regression Line Mean
ln A	39.1	6.3	40.5
E_a (kcal/mol)	27.4	4.2	27.9

As can be seen, the regression means and Monte-Carlo means are not quite the same because of the limited number of simulations used in this example. In ASAP*prime*®, the regression mean is used to center the distribution with the standard deviations calculated from the Monte-Carlo simulation. The confidence bands can be seen in Figure 5 and can be compared to those in Figure 2 based on the error of the fit.

Figure 5. Confidence bands for example data set based on propagating the error bars for the isoconversion rates using a Monte-Carlo calculation.
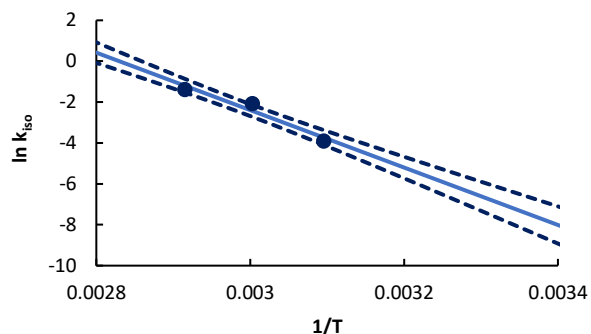
Using these confidence bands, the standard deviation for the isoconversion rates at 25°C can be used to generate the projected degradation error bars shown in Figure 6.

Figure 6. *Distribution of potential degradant levels as a function of time at 25°C based on the Monte-Carlo propagated error bars for the data set. The green and red lines represent plus/minus one standard deviation rates from the blue line (mean prediction).*
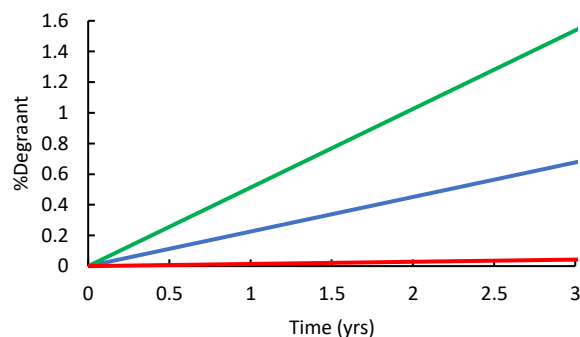
Combining Errors

In a sense, one would expect that if a fitting model (in this case an Arrhenius fit) is correct, the error bars for that model should be encompassed naturally within the error bars for the points used. In other words, if the fit to the model with its error bars does not encompass the points with their error bars, one would have to consider that the model is not correct for that data set or that the error bars used are too narrow. This was the assumption implicit with ASAP*prime*® versions earlier than v6. However, a bad fit to the model could easily result in a tight error bar if the isoconversion errors are small, giving many ASAP*prime*® users confident values that were not accurate. For example, if one condition showed a discontinuity due to melting, the data for each condition could be tight, yet the overall model fit would be poor. Removing the point would result in a more accurate model but not necessarily a smaller error bar. To remedy this situation, we combine both the error of the fit and the isoconversion error propagation

into a new, effective error bar that takes both types of error bars into account. This is done by using the following calculation:

$$SD_{total} = \sqrt{SD_{fit}^2 + SD_{propagated}^2}$$

In the present example, the combination of error bars produces the behavior shown in Figure 7. At 25°C, this corresponds to the degradant growth curve shown in Figure 8. As can be seen, the Monte-Carlo propagated error is dominant in this case; however, the additional error due to the goodness of fit is noticeable in the error bars. For example, at three years, the one standard deviation higher than the mean predicted value is 1.05% using just the error of the fit, 1.56% using just the propagated error, and 1.71% using the square root of the sum of the square errors.

Figure 7 Confidence bands for example data set based on the square root of the sum of the square errors from the error of the fit and the propagated error bars from the isoconversion rates.
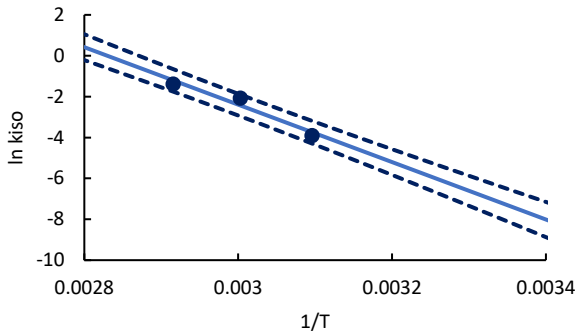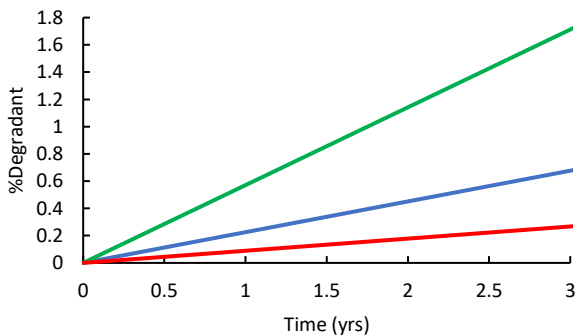

Figure 8 Distribution of potential degradant levels as a function of time at 25°C based on the square root of the sum of the square errors from the error of the fit and the Monte-Carlo propagated error bars for the data set. The green and red lines represent plus/minus one standard deviation rates from the blue line (mean prediction).

Summary

Error bar calculations need to reflect the uncertainty in modeling predictions. For ASAP*prime*®, the two major modeling errors are the error of the fit from the isoconversion rates (i.e., how well the data fit the model) and the error bar propagation from the isoconversion errors. The latter term reflects a combination of the experimental errors and the degree of extrapolation used to estimate the isoconversion rate at each condition. The software combines the two error sources using the square root of the sum of the squared errors. The overall error bars are then used to determine the probability of passing. The combination results in a more conservative estimate of shelf-life compared to using only the propagated error that was used in earlier versions of ASAP*prime*®.