7. Uncertainty Analysis

Background

The ability to calculate the average concentration (and temperature) time profile with uncertainty (represented by either standard deviation bars or confidence bands) is, most likely, the most important and last procedure one should do to validate a model. Turanyi et al.[28] has shown that NOT performing such a procedure and using only one-run time profiles is quite naïve as it is the chemical kinetics/thermodynamics and **NOT** transport/convection processes that are controlling the concentration and temperature profiles. A good analogy would be a hurricane forecaster that only shows one single possible path a hurricane can take with no uncertainty bands or an average path! In addition, the uncertainty analysis in Kintecus also gives one a "smeared-out" sensitivity analysis on ALL the parameters/constants in a model as well as possible maximum and minimum time profiles concentrations (and temperature) can take on during the course of all the simulations.

Uncertainty analysis/confidence band predictions and the related can now be calculated in Kintecus V3.7. Kintecus V3.7 now incorporate uncertainty analysis with an extremely versatile, easy and quick way specifying uncertainty in ANY of the constants (rate constants, TROE factors, residence times, temperature, Cp, Cv, H, S, G, A, concentrations, etc. etc.) in several distributions (uniform, gaussian, etc.) with any number of repeated runs (which can range from several hundred to several thousand depending on the size of the model) in a very easy, straightforward manner with the standard deviations and/or confidence band plots in Excel (though Excel is NOT required). Also, the maximum and minimum concentrations that a species (or temperature) can take on during all the runs can also be displayed and plotted. There are two sample Kintecus-Excel spreadsheets that you may wish to examine: "Enzyme_Uncertainty_Analysis.xls" and "Combustion_workbook_OH_CONF.xls".

Implementation

Uncertainty analysis in Kintecus is turned on by specifying the "-CONF" switch. This switch has numerous options:

-CONF{:a:b:c:d:e:f:g}

All options a, b, c, d, e, f and g are all optional, but when setting options b to g all preceding options must be specified. All options have a default setting which can be specified with the letter "D" or "d" (most Kintecus switches have this feature). Here is a breakdown of all the options (defaults are listed in brackets, [] or by entering the letter "D" for an option indicates to use default value.):

a) Option “a” states to Kintecus the total number of samples (simulations) to run to gather final averages, standard deviations, confidence predictions and maximum/minimum values. **[100]** For the beginning of each simulation, the parameters the user has specified (see other options below and **Selective Parameter Uncertainty Analysis)** are changed by using either a Gaussian or Uniformly generated random number and implementing the parameter as the average (for Gaussian) or central number (for Uniform). The range of the generated random number is given as **one-sigma** for Gaussian generated (which means there is about 1/100 change that a three-sigma generated number can “popup”) or exactly one-sigma in the plus and one-sigma in the minus direction centered on the parameter-mean for the uniform distribution. The type of distribution can be selected for whole sets or separate parameters as described below.

b) Option “b” sets of types parameters to randomize. It has a range of 0 to 31. The types include:

Type or Set Number ID	Randomization Includes:
1	Rate constants, k or A, m and Ea
2	Initial concentrations, Residence Times and External Concentrations (if specified)
4	Initial Temperature, External Temperature, External Heat Source
8	Any and all values enter in M[] or S[] type reactions, such as third-body enhancements, TROE, SRI, Lindeman parameters, etc.
16	All thermodynamic parameters from the thermodynamic database (Cp, H, U, S)
31	ALL sets from 1 to 16.

Table 14.

In addition, the sets or types described above include default relative standard devations that are used in the course of the uncertainty analysis:

Type or Set Number ID	Randomization Includes:	DEFAULT Relative Standard Deviation Implement in Uncertainty Analysis
1	Rate constants, k or A, m and Ea	0.05
2	Initial concentrations, Residence Times and External Concentrations (if specified)	0.02
4	Initial Temperature, External Temperature, External Heat Source	0.02
8	Any and all values enter in M[] or S[] type reactions, such as third-body enhancements, TROE, SRI, Lindeman parameters, etc.	0.02
16	All thermodynamic parameters from the thermodynamic database (Cp, H, U, S)	0.02

Table 15.

To define some sets and exclude others, only add up the type numbers for the sets you wish to randomize. Example, to only include rate constants and initial concentrations, enter a "3" for option "b" (1+2=3), or just to randomize thermodynamic parameters and initial temperature just enter "12" (8+4) for option "b". **[31]. You can selective specify items in any worksheet to include in the uncertainty analysis by using the "number(%std.dev/1)?" operator. Please see below under "Selective Parameter Uncertainty Analysis".**

c) Option "c" multiplies the standard deviations calculated for all species concentrations at all times (and temperature if –THERM is specified) by this value. The default is 1 or one sigma. **[1].** The equation Kintecus uses for standard deviation is:

$$std.dev. = \sqrt{\frac{\sum_{j=1}^{N}(x_i - \bar{x})^2}{n-1}} \bullet (option\,c)$$

where n is the total number of samples or runs (from option a above), x_i is the value of a concentration at some time for sample run j, x-bar is the average value for a concentration at some time t for all runs.

d) Specifying one of the below confidence intervals for option "d" specifies to Kintecus to calculate confidence bands instead of standard deviations. A confidence interval of 0 specifies Kintecus to calculate standard deviations instead. There are four confidence intervals a user can specify: 68%, 95%, 99% and 99.9%. The default is 0 **[0]** which is do not calculated confidence bands. The equation Kintecus utilizes to calculate confidence bands is:

$$Conf.Band. = \left(\frac{std.dev.}{\sqrt{n}}\right) \bullet t - value \bullet (option\ c)$$

Std.dev., option "c" and **n** are described above for option "c", the t-value at the percent confidence interval the user has specified is calculated as:

Confidence Interval	t-value
99.9%	3.28
99%	2.57
95%	1.96
68%	1.00
0%	Confidence Bands NOT CALCULATED (default), but standard deviations are calculated.

Table 16.

e) Option "e" overrides the default relative percent standard deviations used in the uncertainty analysis as described in Option "a" and shown in Table 15 above. A user can specify a single number here to override ALL DEFAULT "Relative Standard Deviations Implement in Uncertainty Analysis" as shown in Table 15 above. The user can also specify a list followed by a Relative Standard Deviations all delimited by a comma: (set or type keyword), (relative standard deviation). So to specify the range of random numbers for the initial concentrations as 15% of the initial mean and the starting initial temperature with a random range of 9% on the initial value, we would use the "-CONF" switch as **–CONF:D:D:D:D:CONC,0.15,TEMP,0.09** (note the commas are used to delimit the keyword from its value).

Keyword(s)	Type of Parameters Included
K or A	Rate constants or Arrhenius Parameter
M or MT	Exponent, m, in T^m for expanded Arrhenius expressions
EA or E	Energy of Activation
SPEC	Special Reaction parameters (the values in S[]) and/or Third-body enhancement factors or pressure fall-off reactions as specified in M[].
CONC	Initial Species Concentrations
RESTIME or RESTIM	Residence Time
BOUNDC or BOUND	External Concentration for Species
TEMP or TEMPERATURE	Initial Temperature
EXTERNTEMP or EXTERNTEMPERATURE	External Temperature
HSOURCE or HEATSOURCE	External heat source/sink
CP or H	All thermodynamic parameters from all thermodynamic databases.

Table 17.

f) Option "f" is very similar as option "e" with the exception option "f" handles how the parameters will be randomly distributed: Gaussian or Uniform. This option overrides the default random number distribution used in the uncertainty analysis as described in Option "a". Option "f" has only two options: 1 for a Gaussian distribution or 2 or a uniform distribution for random numbers. The default is **[1],** Gaussian distribution, for all parameters. Just as in option "e", a list of keywords followed by a distribution type all delimited by a comma: (set or type keyword), (relative standard deviation). Again, the keywords and their respective representation is given in Table 17. So to specify the random distribution type for the initial concentrations as Gaussian and the starting initial temperature using a uniform random distribution about the initial value, we would use the "-CONF" switch as
–CONF:D:D:D:D:CONC,0.15,TEMP,0.09**:CONC,1,TEMP,2** (note the commas are used to delimit the keyword from its value).
A note on using Gaussian (normal) distributions in that the value you give represents **one sigma**, there is about 1/100 chance probability that a value that is **three times greater** than the relative standard deviation you give will be selected. You should stay away from using relative standard deviations of 0.30 (30%) or greater. Using relative standard deviations of 0.30 (30%) or greater can result in negative values for initial rate constants, initial/external concentrations, residence times. In uniform distributions, the relative standard deviation you give will NEVER BE EXCEEDED in the plus or minus direction. All numbers between the parameter +/- the relative standard deviation you give have an equal chance of being selected.

g) Option “g” states to Kintecus whether to remove all the simulation concentration files (that look like CONCnnnnn.txt where n is a numeric digit) after the program is done or keep them. There are only two options for this: KEEP or DELETE. The default is to **[KEEP].**

There are two sample Kintecus-Excel spreadsheets that you may wish to examine: “Enzyme_Uncertainty_Analysis.xls” and “Combustion_workbook_OH_CONF.xls”.

Here are some sample "-CONF" switches and a short description on what they do:

Sample "-CONF" switch	What does it do?
-CONF:10000:1	Run 10,000 simulations varying ONLY the rates constants or (if present) the three Arrhenius parameters (A, m and Ea) with a relative standard deviation of 5% using Gaussian distributed random numbers. All the error bars for the average concentration plot are all calculated at one-sigma.
-CONF:D:2	Run 100 simulations varying ONLY the initial concentrations and (if present) residence times and external concentrations using standard deviation of 2% using Gaussian distributed random numbers. All the error bars for the final average concentration plot are all calculated at one-sigma.
-CONF:1000:1:1:99%:k,0.01:k,1	Run 1,000 simulations varying ONLY the rates constants or (if present) the three Arrhenius parameters (A, m and Ea) using a relative standard deviation of 1% using Gaussian distributed random numbers. Calculate confidence bands at 99% confidence level for all the for the final average concentrations.
-CONF:D:4:D:D:temp,.1:temp,2	Run 1,000 simulations varying ONLY the initial temperature or (if present) the external heat source or external bath temperature using a relative standard deviation of 10% using uniform distributed random numbers. All the error bars for the final average concentration plot are all calculated at one-sigma.
-CONF:1000:0	Run 1,000 simulations varying ONLY those constant/parameters that have been "flagged" by the user's "**(#/#)?**" operator. See below under the "Selective Parameter Uncertainty Analysis". All the error bars for the final average concentration plot are all calculated at one-sigma.
-CONF:D:D:D:D: temp,0.05,k,0.02,m,0.08,ea,0.08:temp,1,k,1, m,2,ea,1	Run 100 simulations varying ONLY the rate constants, k, m, Ea and the initial temperature. Rate constants can change 2%, Ea, 8% and temperature 5% all temperature using a relative standard deviation, but the exponent m will vary AT MOST 8% using uniform random numbers. Keep in mind, there is 1 in 100 change that percent changes can triple using a Gaussian (Normal) distribution!. The percent change listed for m (0.08=8%) is the MAXIMUM upper and lower boundary change allowed since a uniform distribution is assumed for this parameter! **Very important to remember that!**

Table 18. Various sample switch values for the "-CONF" switch to perform various uncertainty analyses. The "-CONF" switch can be overrided for single constants/parameters using the "(#/#)?" operator (see below under Selective Parameter Uncertainty Analysis).

Selective Parameter Uncertainty Analysis

There are some parameters/constants whose values are known quite precisely (such as rate constants, initial concentrations, enhanced third-body factors, TROE, SRI factors, etc.) and then there are some parameter values that are not known quite precisely. To specify an average percent standard deviation for such values would be wrong, so Kintecus provides a way to override all the options given in the above section for such parameters. This is accomplished through the "**(#/#)?**" operator that will follow any number that may appear in any spreadsheet (model, species, parameter). The two numbers located in the parenthesis dictate the relative standard deviation and the second number dictate the type of distribution to generate the random number (1=Gaussian, 2=Uniform) about the mean, the mean would be the actual preceding parameter/constant. For example, for the reaction of OH + HO2➔H2O + O2, if we assume the rate constant was 8.0 x 10^{12} +/- 2 x 10^{12} (assuming 2 x 10^{12} represents **one-sigma!)** then the relative standard deviation is 20% or 0.2. This reaction can be written in the model spreadsheet with an uncertainty override and assuming a Gaussian distribution:

8.0E+12**(0.2/1)?** , OH + HO2➔H2O + O2

Again, the **(#/#)?** uncertainty override can practically appear after any number: initial concentrations, residence times, TROE factor, enhanced third-body factors, temperature, etc.

What About Zero Means?

If you haven't realized it, but if your parameter is zero then it is NOT INCLUDED in the uncertainty analysis. For some parameters this makes sense such as initial concentrations, residence times, external concentrations, but for a few (like the exponent **m** for $T^{\mathbf{m}}$ in the expanded Arrhenius equation) in which one would like to include some variation in it during an uncertainty run. One can accomplish this by using the selective parameter uncertainty operator, "**(#/#)?**", and by specifying a negative value instead of a relative standard deviation for the first value. The negative value is actually the ABSOLUTE RANGE for the spread of the random number. For example:

8.0E+12, 0(-0.75/1)?, 4.33, OH + HO2➔H2O + O2

The m=0 shown above will be randomly spread about (using a Gaussian distribution) –0.75 to +0.75 for 1 sigma. Again, there is a 1/100 chance that a number equal to 2.25 or greater (three sigma x range) might be selected since a Gaussian distribution has been selected. If the –0.75 were a +0.75, then there would be no change during ALL uncertainty runs for m=0 because 0 x 0.75 x (random number) is always equal to zero.

The Output

Kintecus will output several files that contain the average, standard deviation/confidence bands, maximum, minimum concentration (and temperature) time profiles. These files are always named **CONCAVG.TXT, CONCSTD.TXT, CONCMAX.TXT** and **CONCMIN.TXT** and should be located in the same directory as Kintecus.exe. In addition, if the "g" suboption of the –CONF switch is either not given or set to "KEEP" ALL uncertainty runs will also be kept in the same directory as Kintecus.exe and have the prefix **CONCnnnnn.TXT**, where n is a digit, 0-9.

The updated Kintecus-Excel spreadsheets will automatically load and plot all four files into the spreadsheet. If you do not see such loaded plots and worksheets then be sure you are using the latest Kintecus-Excel spreadsheets! This is very important!

The Kintecus-Excel files (suffixed with .xls) contain separate VBA code from the main Kintecus program that permit execution of Kintecus using the Excel interface. These files routinely are changed with Kintecus releases, so if you have created a Kintecus model using an Excel-Kintecus worksheet from Kintecus V3.5 or earlier, then **you must update** your Excel-Kintecus worksheet. You can do this by loading a blank Kintecus worksheet (Kintecus Blank Model in this distribution) and copy all your worksheets from your old to the new ones. This is easy to do as it requires you to right-click on the worksheet tab, select "Move or Copy", then under "To book" select the name of the loaded, NEW Kintecus worksheet, click OK.

The VBA Macros in the new Kintecus Excel worksheets are more reliable and have extra graphing capabilities. Such graphing capabilities as the plotting the Uncertainty Analysis with confidence limit plotting and will also recognize the "-o:y:y:y:y" and automatically plot all related files.

Some Examples!

The first example implementing an uncertainty analysis run is with the small kinetic system of an enzyme binding, reacting, and then releasing a product. The following page shows the output from a single run followed by a plot of the average concentrations with one-sigma errors computed using the uncertainty analysis feature in Kintecus V3.7. Figure 14 below shows a single run using the nominal values. Figure 15 shows an average run from the 100 uncertainty runs. Note the great concentration variations of over 1,000% (!!!) for [S] and [ES] about the time of 900 seconds. If [S] or [ES] were dangerous, poisonous intermediates, figure 15 demonstrates that exact control of some initial conditions are required. All is not lost as the final, equilibrium concentrations of all species are all very close.

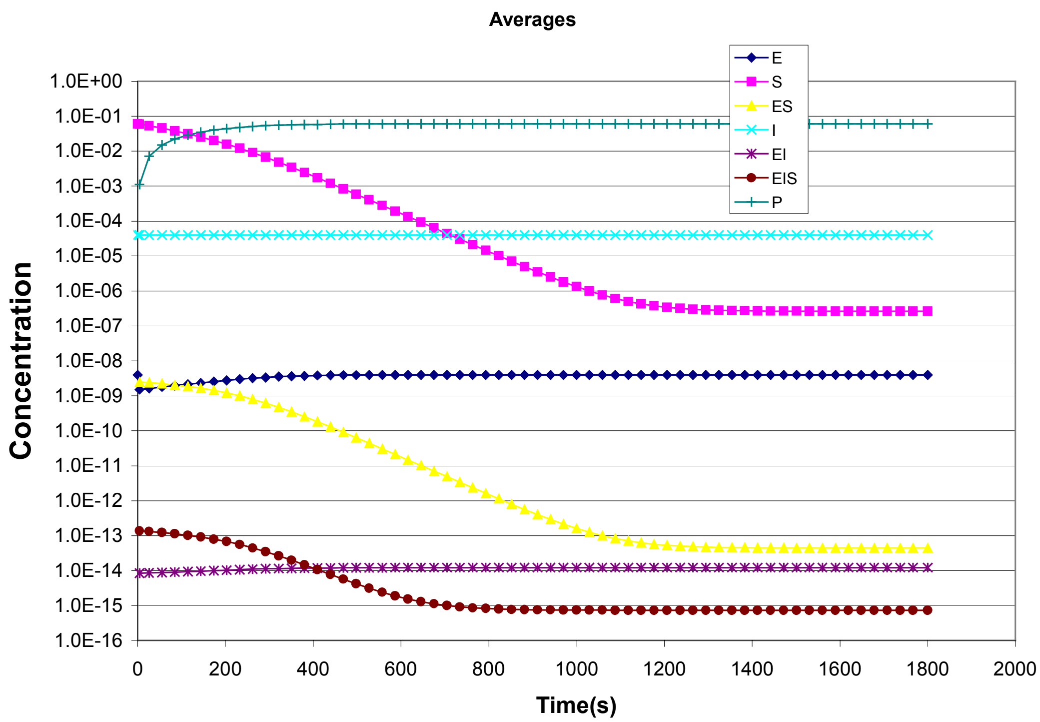

Figure 14. Plot of an enzyme model utlizing one single run. The y-axis is logarithmic.

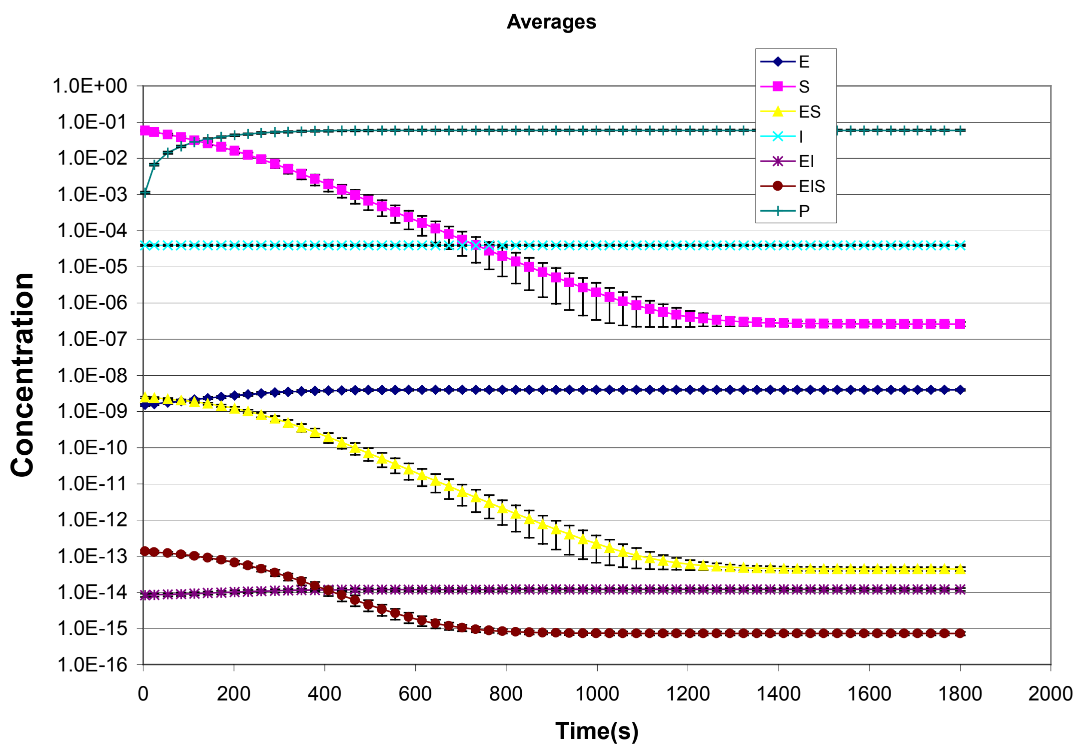

Figure 15. An average plot with one-sigma deviations of an enzyme model utilizing 100 single runs in a Kintecus uncertainty analysis run. The y-axis is logarithmic. Note the great concentration variations of over 1,000% (!!!) for [S] and [ES] about the time of 900s.

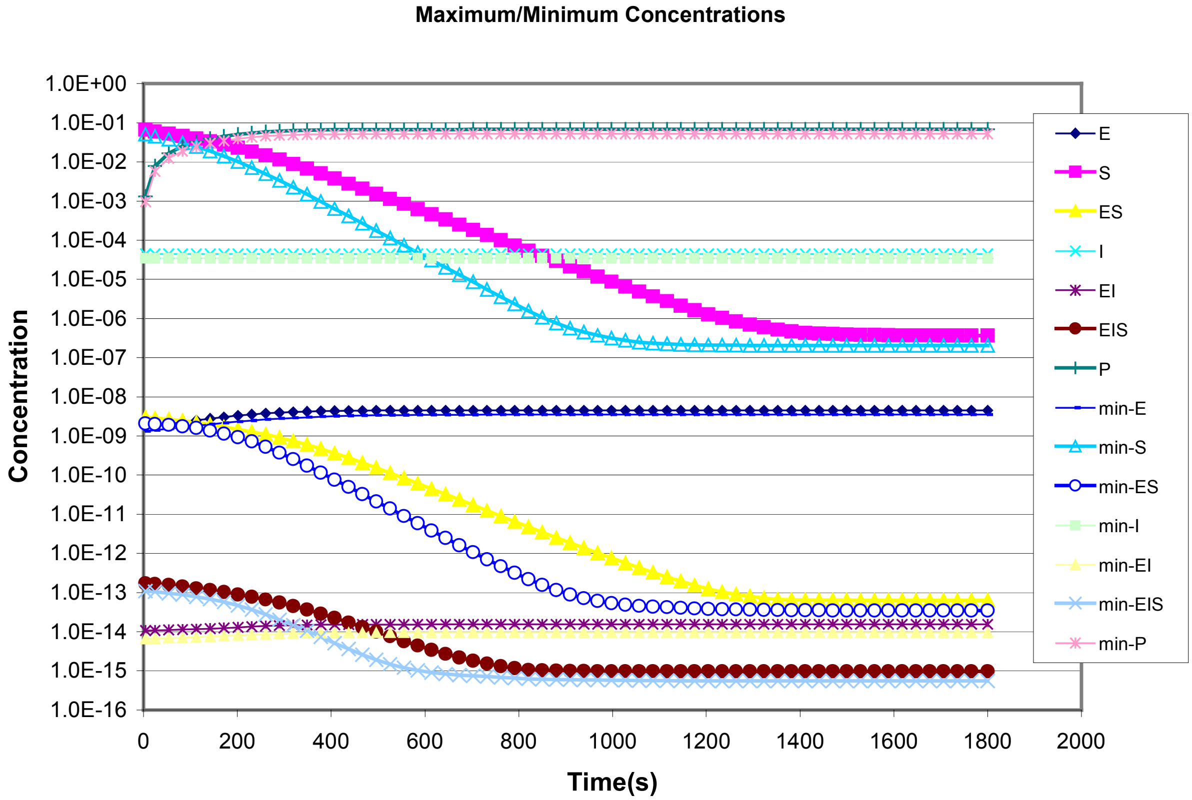

Figure 16. Maximum/minimum of an enzyme model utilizing 100 single runs in a Kintecus uncertainty analysis run. The y-axis is logarithmic. Only the species S and ES show great concentration ranges under the uncertainty analysis, but, the final equilibrium concentrations show little changes.

The second example applying uncertainty analysis is with the H2-O2 combustion model. Figures 17 and 18 below show a single run, but figures 19 and 20 below show an average model with error bars (one-sigma) generated with the –CONF switch in Kintecus V3.7. As one can clearly see that the average does differ significantly from the single run, with the concentration profile of O_2 containing the greatest scatter. Figure 22 shows a zoom plot of the HO2 and H2O2 average concentrations with error bars of one-sigma for the uncertainty run of the H2-O2 combustion under adiabatic conditions, constant pressure. These average concentrations are close to the concentrations in figure 21 that were generated from single-run using nominal initial conditions, but there is a great intermediate variation in the concentration time profiles. Interestingly, the concentration of H_2O_2 around 1.9x10(-4) seconds shows very little scatter indicating that at this time point, $[H_2O_2]$ is quite invariant to the accuracy of rate constants, thermodynamics, third-body reaction enhancements and even initial concentrations!

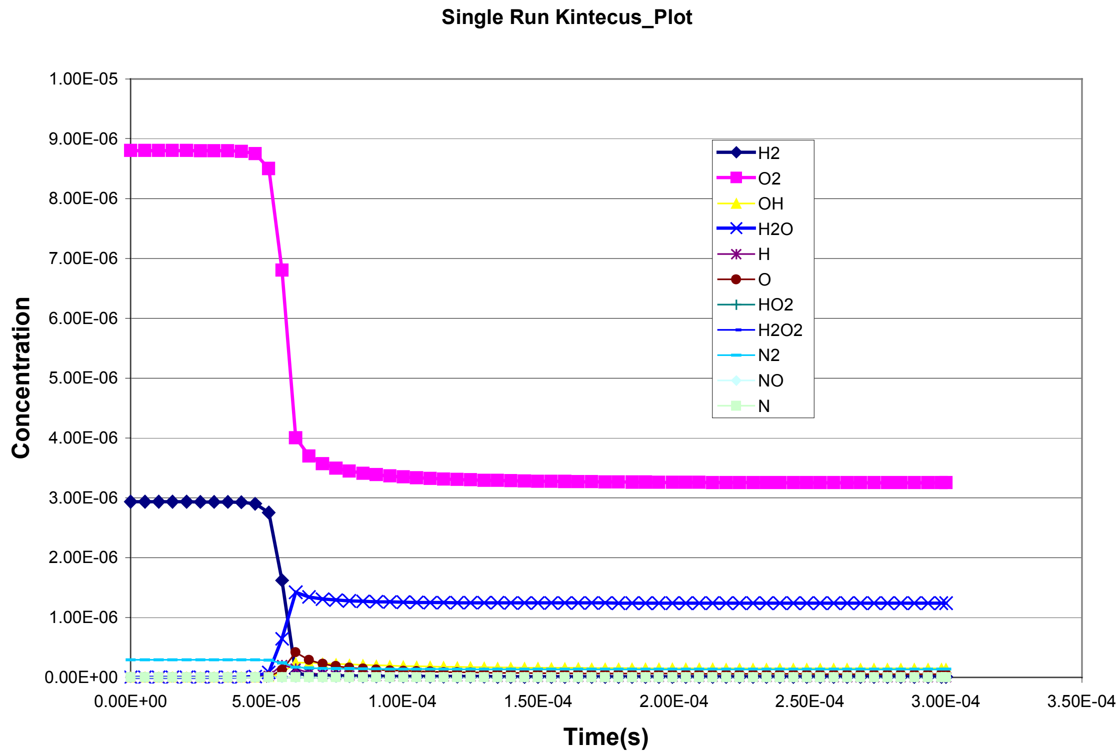

Figure 17. Concentration profiles from a single run of the combustion of H2 and O2 at constant pressure.

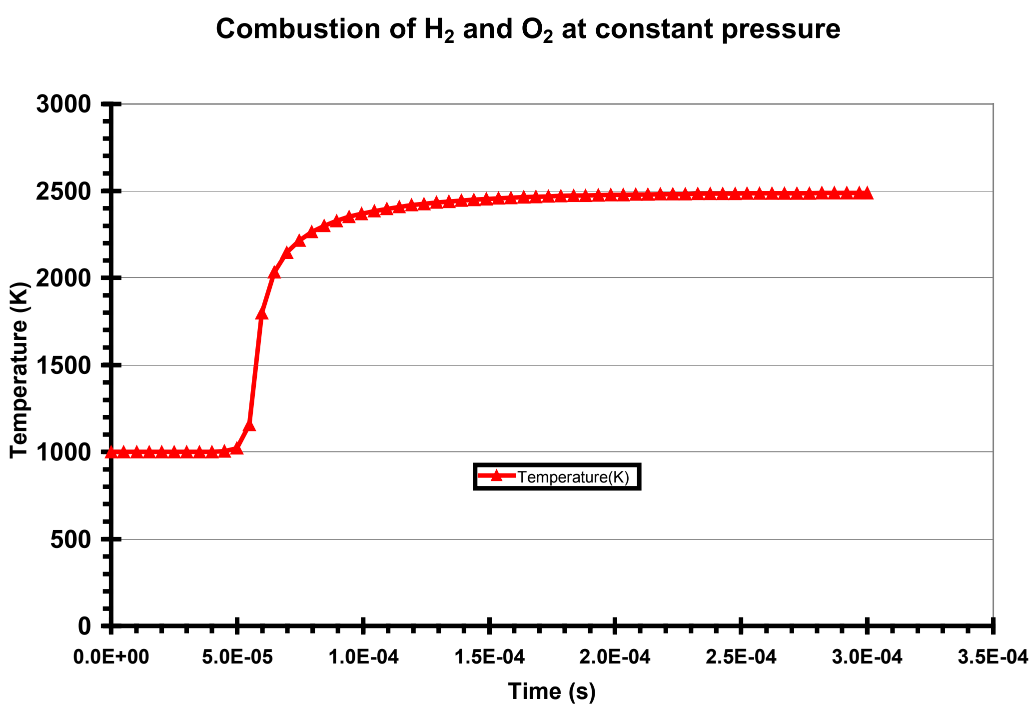

Figure 18. Temperature profile from a single run from a single run of the combustion of H2 and O2 at constant pressure.

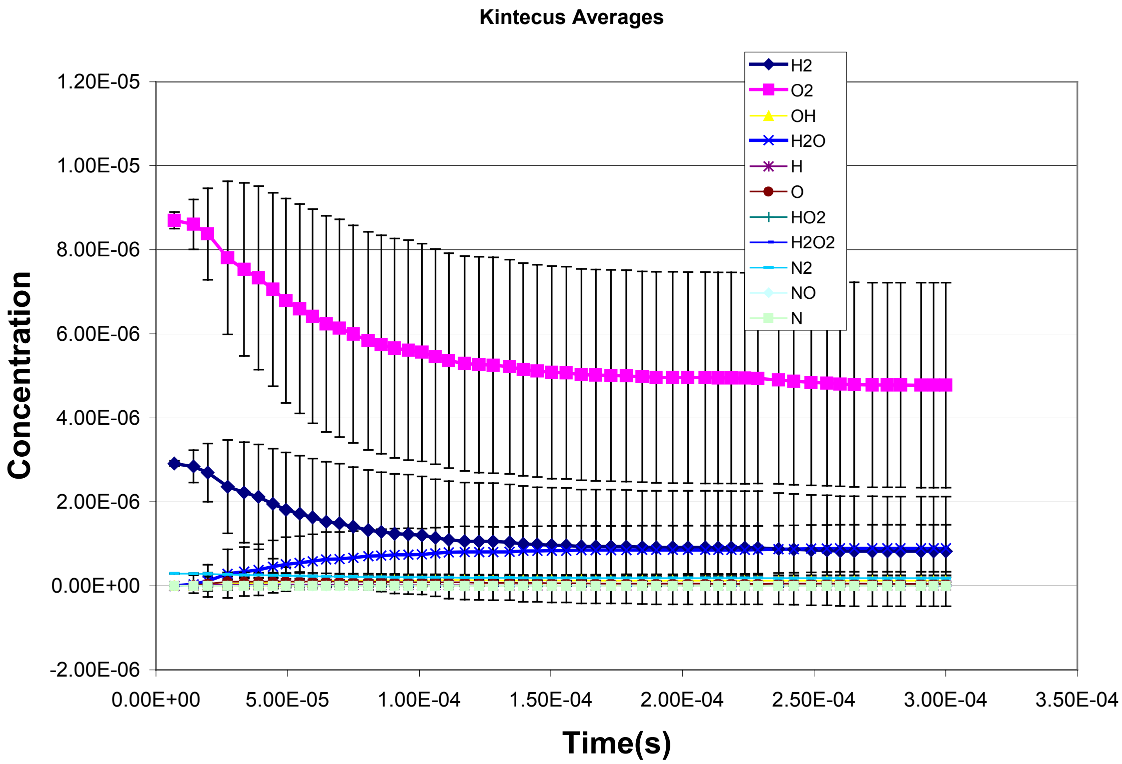

Figure 19. Average concentration profiles from the H2-O2 combustion run under constant pressure with error bars from 100 sample runs with "-CONF" switch. These average results are very much different from the single run shown in figure 17 above!

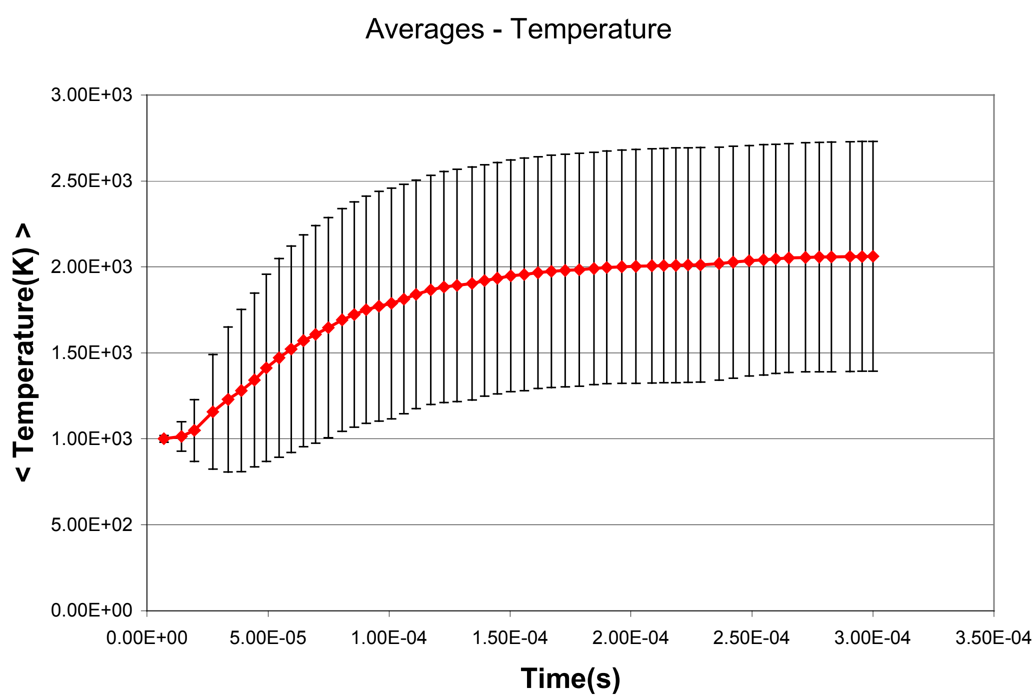

Figure 20. Average temperature profiles from the H2-O2 combustion run under constant pressure with error bars from 100 sample runs with "-CONF" switch. Again, these average results are vastly different from the single run shown in figure 18 above!

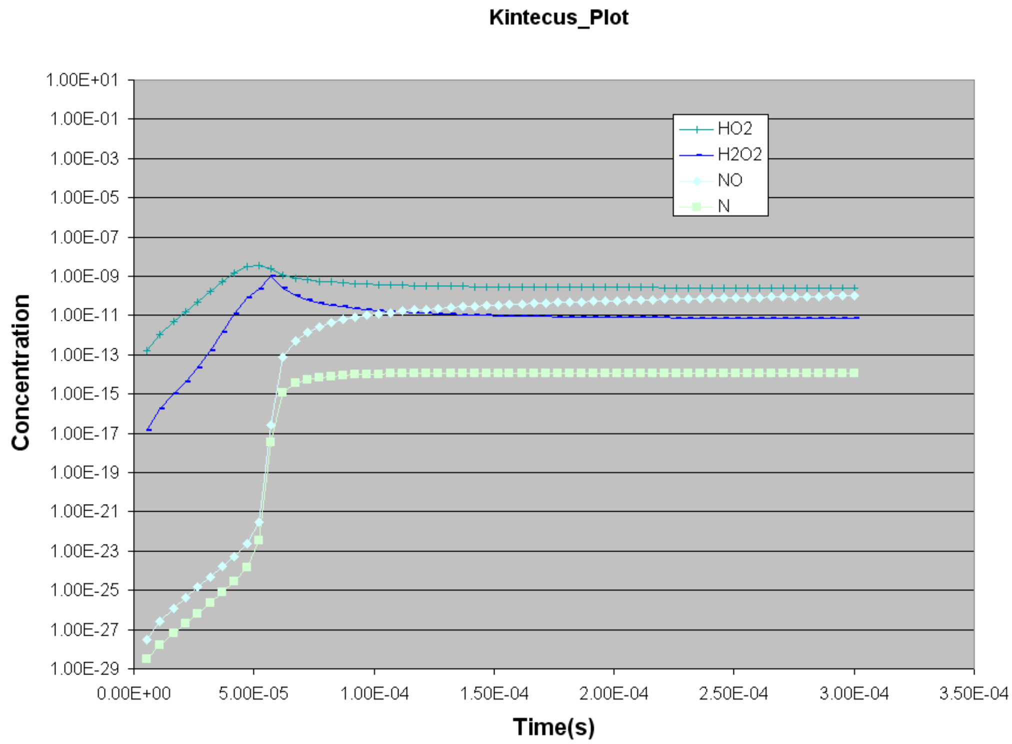

Figure 21. A single run showing the time profile for HO_2, H_2O_2, NO and N.

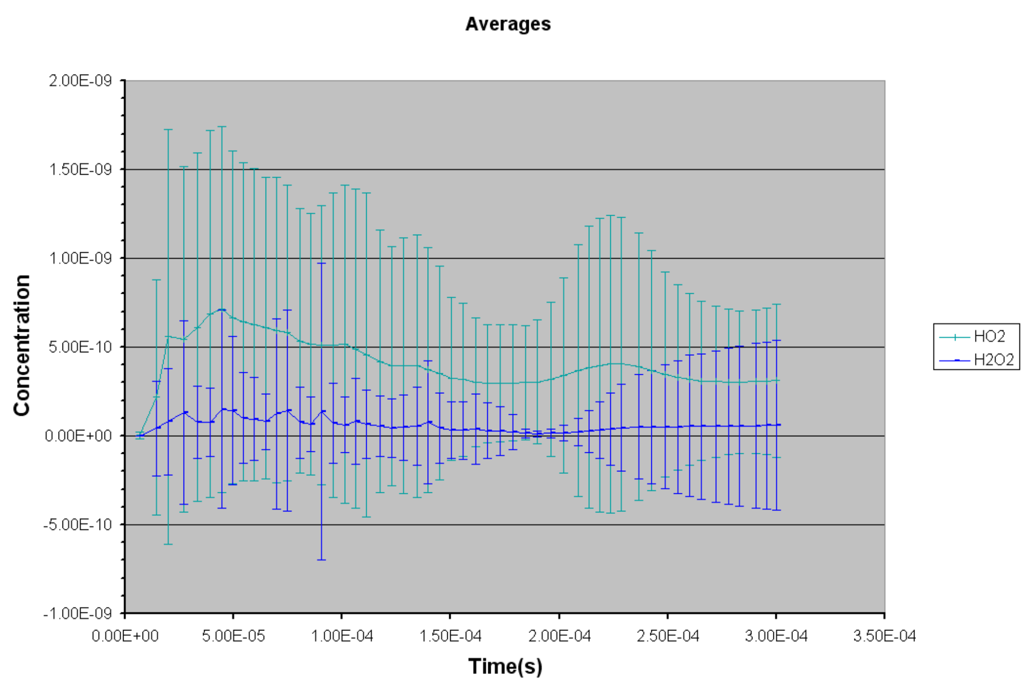

Figure 22. Plot of the HO2 and H2O2 average concentrations with error bars of one-sigma for the uncertainty run of the H2-O2 combustion under adiabatic conditions, constant pressure. These average concentrations are close to the concentrations in figure 21 that were generated from single-run using nominal initial conditions, but there is great intermediate variations in the concentration time profiles.