Interpolating Literature Data Sets

Barry Linkletter

Many data sets in the literature list values such as H_0 or activity coefficients for mixtures of acids (and other solvent conditions). However they may not have a data point exactly where you need it to be. To extract the information that you want you will need to interpolate.

This document was produced using the LaTeX typesetting language with the Tufte-handout document class. Chemical diagrams were created in *ChemDoodle* and calculations and plotting were performed using *Python* tools in a *Jupyter Notebook*. Diagrams and plots were further edited in *Affinity Designer*

Hunting for Data

I have a system where rate constants of a reaction, k_{obs}, in various sulphuric acid mixtures have been reported. I am constructing a plot that will relate the activity of water, a_{H_2O}, the acidity function to the mixture, H_0, to this rate constant. I need values for a_{H_2O} and H_0 for the mixtures used in the experiment. Fortunately, the chemical literature has many examples of such tables, and a few have become the standards in the field. Table 1 below lists some data sources that might be useful for our investigation.

Data	H_2SO_4 Mixture	Source
a_{H_2O}	1 to 75 %	Rard, 1976
a_{H_2O}	9 to 99 %	Giauque, 1960
a_{H_2O}	20 to 42 %	Yang, 2016
H_0	1 to 60 %	Johnson, 1969
H_0	4 to 98 %	Tickle, 1970
H_0	60 to 99 %	Jorgenson, 1963
H_0	98 to 100 %	Gillespie, 1971
D_0	8 to 99.6 %	Sierra, 1970
H_0 & D_0	$10^{-3.8}$ to 1 mol L^{-1}	Högfeldt, 1960
H_R	6 to 90 %	Cook, 1975
H_R	0.5 to 98 %	Deno, 1955
density	0.5 to 100 %	CRC Handbook

Table 1: Chemical data for properties of mixtures of sulphuric acid. ⟵

There are many options for a particular property in a given acid mixture. Hopefully, they all agree. Spoiler alert: they sometimes do not. However, subsequent publications will reveal the data sets most accepted by the field. So, do not grab just any data set; read widely and get a sense of which is most trusted. Over time, even well-respected groups have their data cast aside as more modern methods reveal better measurements.

References are listed below…

J.A. Rard et al., *J. Chem. Engin. Data*, **1976**, *21*, 374-379. https://doi.org/10.1021/je60070a002

W.F. Giauque et al., *J. Am. Chem. Soc.*, **1960**, *82*, 62-70. https://doi.org/10.1021/ja01486a014

H. Yang al., *J. Solution Chem.*, **2016**, *45*, 1580–1587. https://doi.org/10.1007/s10953-016-0516-4

C.D. Johnson et al., *J. Am. Chem. Soc.*, **1969**, *91*, 6654-6662. https://doi.org/110.1021/ja01052a021

P. Tickle et al., *J. Chem. Soc. B*, **1970**, 65-70 https://doi.org/10.1039/J29700000065

M.J. Jorgenson, D.R. Hartter, *J. Am. Chem. Soc.*, **1963**, *85*, 878-883. https://doi.org/10.1021/ja00890a009

R.J. Gillespie et al., *J. Am. Chem. Soc.*, **1971**, *93*, 5083-5087. https://doi.org/10.1021/ja00749a021

J. Sierra et al,, *J. Chem. Soc. B*, **1970**, 1570-1573. https://doi.org/10.1039/J29700001570

E. Högfeldt, J. Bigeleisen, *J. Am. Chem. Soc.*, **1960**, *82*, 15-20. https://doi.org/10.1021/ja01486a005

M.J. Cook et al., *J. Am. Chem. Soc.*, **1975**, *97*, 760-764. https://doi.org/10.1021/ja00837a012

N.C. Deno et al., *J. Am. Chem. Soc.*, **1955**, *57*, 3044-3051. https://doi.org/10.1021/ja01616a036

CRC Handbook of Chemistry and Physics, 106th ed. **2025**, J.R. Rumble, ed., *CRC Press, Taylor & Francis, Boca Raton, FL.* https://library.upei.ca/chem or get the PDF here or here.

The Land of In-Between

When analyzing literature results, we sometimes find it informative to replot published data against more modern data sets and see if we reach the same conclusions as the authors (a 100 % success rate so far). In the analysis that we are currently engaged in, I will need to obtain data for the activity of acid (h_0), the activity of water (a_{H_2O}) and the molar concentration of water (will require the density of the mixture) in given mixtures of sulphuric acid and water.

We are investigating the observed rate constant for methyl acetate hydrolysis in concentrated sulphuric acid mixtures.[1] I will use the data set from The data from the paper Giauqueet (1960) for a_{H_2O}, the data from Tickle (1970) for H_0 (We can calculate h_0 from H_0), and the density data from the CRC Handbook (from which we can calculate $[H_2O]$.)[2]

Table 2 on the following page presents data from all my sources. You can see that the data points for the various values seldom line up. How will I obtain a literature value for H_0, a_{H_2O} and density at the experimental concentrations used in the rate study?[1] The answer is to interpolate values between the data points of the literature data sets.

I could use an interpolation function that provides values on straight lines between each data point. This linear interpolation method will deliver excellent results with most data sets. A slightly better method is to use a Blpline interpolation to give me values on a curving spline between data points. This is useful if the data has a smooth curve and significant spacing between points. When using a spline, one must ensure that we are not getting a wiggle between data points as it attempts to run the spline through each point. I chose to use a smoothing Bspline function provided by the *SciPy.Interpolate* library in *Python*.[3] This function will create a smooth interpolation on a spline, but it is not mathematically required to pass through each data point. It is suitable for moderately noisy data. Our data sets are produced by authors who reported the values based on polynomial curve fits through their own experimental data and are noise-free. We do not have access to the underlying data and only have the data points reported in each publication. So we will interpolate.

The Tools for Interpolation

I have written short blobs of *Python* code that will read in a data file and create an interpolation function based on that data. Once the interpolation function is created it will return a y-value when given an x-value. The result should give a very good estimate for values between points of our data set. Since I am using a smoothing interpolation function, the results at each data point where we have a known x,y pair may differ slightly. This will be due the precision of the numbers (most values have only three significant figures)

[1] "Mechanisms of ester hydrolysis in aqueous sulfuric acids." K. Yates, R.A. McClelland, *J. Am. Chem. Soc.*, **1967**, *89*, 2686-2692. https://doi.org/10.1021/ja00987a033

[2] The references are given in table 1

[3] See the interactive *Python* notebooks on Google Colab.

$\%H_2SO_4$	$k_{obs}/10^{-2}\,min^{-1}$	H_0	$a_{H_2O}/10^{-2}$	$\rho/g\,cm^{-3}$
8.9			96.22	1.0947
14.1	1.50	−0.59		
20		−0.99		1.1398
20.7	2.61			
22.7			85.14	
22		−1.13		1.1554
28.3	4.22		78.00	
30		−1.63		1.2191
32.9			70.35	
34.8	6.41			
36		−2.09		1.2685
37.0			62.57	
40.4	8.14			
40.7			55.03	
44		−2.69	48.06	1.3386
45.4	10.4			
50		−3.21		1.3952
51.9			31.13	
50.2	11.4			
54		−3.65	26.78	1.4351
55.2	13.3			
58		−4.10	19.80	1.4770
60.4	13.8			
64		−4.85	10.77	
65.2	11.9			
66		−5.10	7.929	
70		−5.68		1.6105
70.4	7.25			
71.0			3.799	
74		−6.24		
74.1	3.83			
79.7			0.590	
80.0	0.931	−7.17		1.7272
81.5			0.353	

Table 2: Pseudo-first order rate constants for hydrolysis of methyl acetate at 25 °C.[1] Selected data for H_0, a_{H_2O} and density are included. Observe that we do not have exact matches between the reaction conditions reported and the values for our literature data.
←

For example: Observe that we have a rate reported at 45.4 % H_2SO_4. We have data for H_0 for 44 % and 50 % mixtures, but not at 45.4 %. Data for water activity is not available for 50 % but is at 51.9 %. How shall we find values in the spaces between data points in literature tables? The answer is interpolation.
←

and also slight drift at the ends of the data sets where the spline-creating math will have less data to work with.

In figures 1, 2 and 3 on the next page present the data sets chosen for this project with the Bspline interpolation plotted as well.[4] With each plot there will also be presented the calculated residuals at the known data points. You can see that the spline interpolation tracks with the data within a few tenths of a percent in most cases. You can change how closely the spline tracks the data points by editing the *Python* code.

[4] The *Python* notebook for the plots in Figures 1, 2 and 3 on the next page above can accessed via Google Colab at https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class_30/Yates-interpolators.ipynb

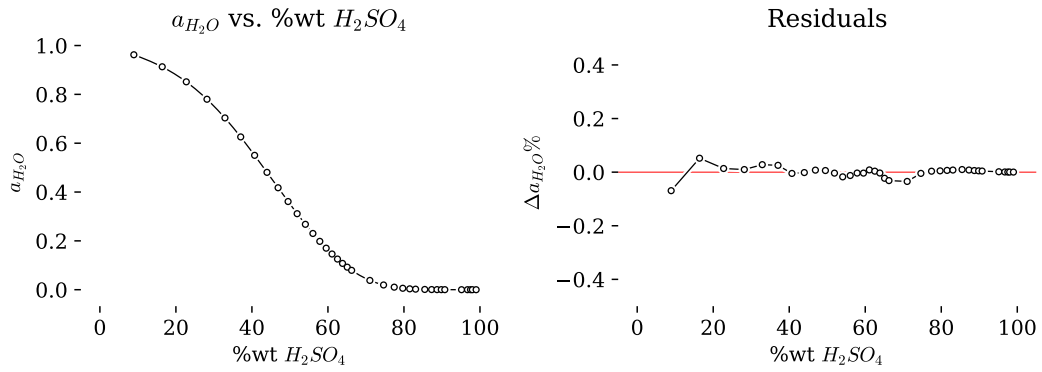

Figure 1: Data and interpolation of the a_{H_2O} data from Giauqueet, 1960. ↑

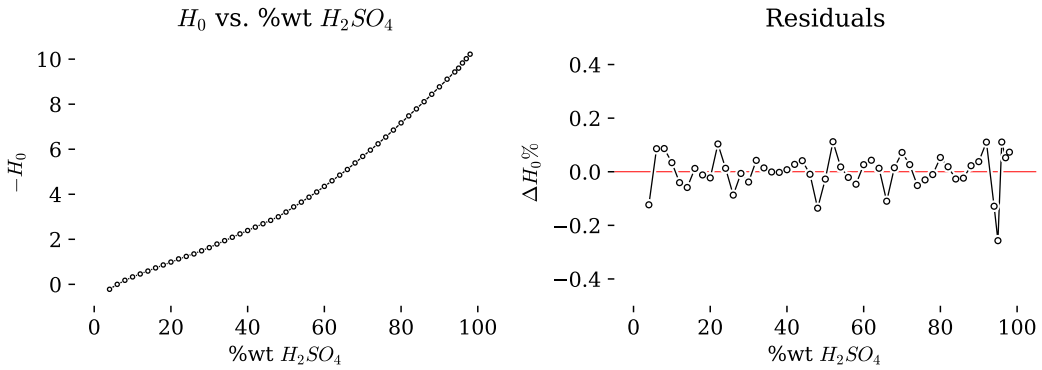

Figure 2: Data and interpolation of the H_0 data from Tickle, 1970. ↑

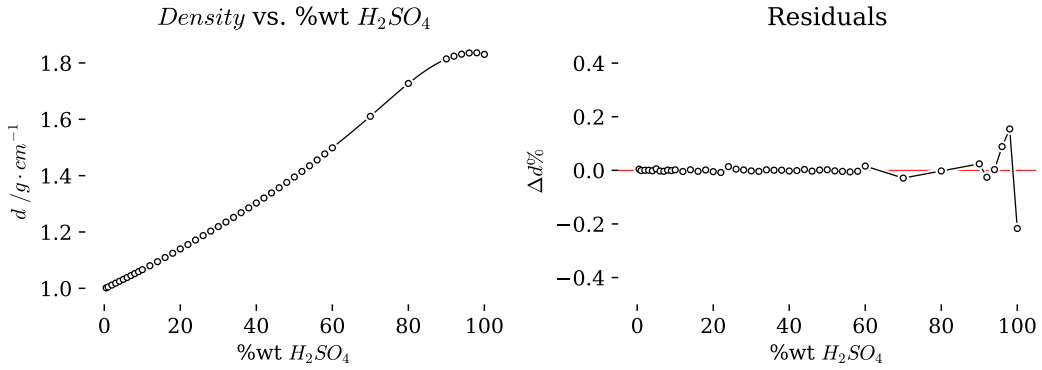

Figure 3: Data and interpolation of the density data from the CRC handbook, 2023. ↑

Applying the Interpolations

By applying the three interpolation functions at the values for $\%H_2SO_4$ in the data table for rates of hydrolysis[1] we can estimate values for H_0, a_{H_2O} and density. The results of this are presented in table 3.

$\%H_2SO_4$	$k_{obs}/10^{-2}\,min^{-1}$	H_0	$a_{H_2O}/10^{-2}$	$\rho/g\,cm^{-3}$
14.1	1.50	−0.59	93.0	1.095
20.7	2.61	−1.04	87.4	1.145
28.3	4.22	−1.51	78.0	1.205
34.8	6.41	−2.00	66.9	1.258
40.4	8.14	−2.42	55.7	1.306
45.4	10.4	−2.79	45.0	1.351
50.2	11.4	−3.23	34.7	1.397
55.2	13.3	−3.78	24.6	1.448
60.4	13.8	−4.40	15.6	1.503
65.2	11.9	−5.00	9.09	1.556
70.4	7.25	−5.74	4.23	1.615
74.1	3.83	−6.26	2.157	1.659
80.0	0.931	−7.17	0.538	1.727

Table 3: Pseudo-first order rate constants for hydrolysis of methyl acetate at 25 °C.[1] Interpolated data for H_0, a_{H_2O} and density are included. Only five data points matched up in the tables against the concentrations used by the authors. All other values for H_0, a_{H_2O} and density were estimated by interpolation. ←

We now have all the data necessary to construct the plot of the model for the $A_{Ac}2$ acid-catalyzed ester hydrolysis in concentrated acidic conditions.