

# Literature Values for $\rho$ , $a_{\text{H}_2\text{O}}$ & $H_0$

Barry Linkletter

This is an update to the interpolation exploration that I have previously presented. Here I will use more recent data sets for  $a_{\text{H}_2\text{O}}$  and  $H_0$  vs. concentration of  $\text{H}_2\text{SO}_4$  (expressed in molality, molarity and % mass). I will use polynomial fits and create parameterized models to convert between the data sets. This will be more portable than the interpolator functions used previously.<sup>1</sup>

## Introduction

In a previous exploration document, I presented methods for interpolating tables of data for values of  $H_0$ ,  $a_{\text{H}_2\text{O}}$  and density of mixtures of water and sulphuric acid.<sup>2</sup> Using data tables from the literature, a smoothing interpolation function was applied to create an interpolation function that would return a value when given a concentration of a  $\text{H}_2\text{SO}_4/\text{H}_2\text{O}$  mixture. These interpolator functions, provided from the `scipy.interpolate` library, tracked very closely to the experimental data but tended to deviate at the extremes of the data sets. These deviations were small but became a problem when the interpolated values were also tiny. I ran into an issue where the interpolation for  $a_{\text{H}_2\text{O}}$  in very concentrated sulphuric acid mixtures was returning negative values. If I want to use data at greater than 99 %  $\text{H}_2\text{SO}_4$ , I will need a better system for extracting data from these data tables.

Also, the interpolator functions generated by the `scipy.interpolate` library are a black box, and cannot be expressed in printed form. Having the code available via an interactive *Python* notebook does allow the reader to access exactly what was done,<sup>3</sup> but I find it more satisfying when the methodology can be expressed in writing.

The data that I had used to build the interpolator functions for  $a_{\text{H}_2\text{O}}$  was from a data set reported by Giaque in 1960.<sup>4</sup> I also had access to a data set produced by Rard in 1976.<sup>5</sup> I have since come to realize that these data sets are based on polynomial fits of previous literature data. Over the years other groups have measured  $a_{\text{H}_2\text{O}}$  using a variety of methods and this data has been used in similar models based on polynomial fits.<sup>6</sup>

I will be using polynomial fits to create parameterized models that predict  $a_{\text{H}_2\text{O}}$  in units of apparent mole fraction and apparent molar concentration from concentration data reported in units of %mass. I will compare my results with those presented by Cox, who seems to have performed a similar exercise.<sup>7</sup>

This document was produced using the  $\text{\LaTeX}$  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*.

<sup>1</sup> WARNING: After 19 pages of work, I concluded that interpolation works best in most cases for these data sets. Skip to the end and read the rest if you want to learn more about polynomial modelling in *Python*.

<sup>2</sup> See “Interpolating Literature Data Sets”, available via our Moodle site.

<sup>3</sup> The *Python* notebook for the previous interpolation exploration can be accessed via Google Colab at <https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class.30/Yates-interpolators.ipynb>. If you have internet access, you can investigate the code and reproduce my results easily. But, when the end of civilization comes to pass, those who seek to follow in my footsteps will only have these printed words found on singed paper in the rubble. That is why I want to build these parameterized models.

<sup>4</sup> “The Thermodynamic Properties of Aqueous Sulfuric Acid Solutions and Hydrates from 15 to 300 K.” W.F. Giaque, E. W. Hornung, J. E. Kunzler, T. R. Rubin., *J. Am. Chem. Soc.*, 1960, 82, 62–70. <https://doi.org/10.1021/ja01486a014>

<sup>5</sup> “A review of the osmotic coefficients of aqueous sulfuric acid at 25 °C” J.A. Rard, A. Habenschuss, F.H. Spedding, *J. Chem. Engin. Data*, 1976, 21, 374–379. <https://doi.org/10.1021/jc60070a002>

<sup>6</sup> “Activity and osmotic coefficients of aqueous sulfuric acid at 298.15 K.” Bert R. Staples, *J. Phys. Chem. Ref. Data*, 1981, 10, 779–798. <https://doi.org/10.1063/1.555648>

<sup>7</sup> “Excess acidities.” Robin A. Cox, *Adv. Phys. Org. Chem.*, 2000, 35, 1–66. [https://doi.org/10.1016/S0065-3160\(00\)35011-0](https://doi.org/10.1016/S0065-3160(00)35011-0)

## The Plan

In my own exploration<sup>8</sup> of the results reported by Yates and McClelland for ester hydrolysis<sup>9</sup> in concentrated sulphuric acid, I chose to express the value for  $a_{\text{H}_2\text{O}}$  in terms of effective molar concentration rather than mole fraction. The tables of Giaque<sup>4</sup>, Rard<sup>5</sup> and Staples<sup>6</sup> all report values for  $a_{\text{H}_2\text{O}}$  in terms effective mole fraction of water versus the concentration of  $\text{H}_2\text{SO}_4$  expressed as molality. However, the data tables that I have for  $H_0$  use %mass of  $\text{H}_2\text{SO}_4$ . I need a model that will give me a value for  $a_{\text{H}_2\text{O}}$  when given % $\text{H}_2\text{SO}_4$ . With this tool, I can have values for both  $a_{\text{H}_2\text{O}}$  and  $H_0$  at any given value of % $\text{H}_2\text{SO}_4$ .

The CRC book<sup>10</sup> provides this very information in data tables for the density of sulphuric acid mixtures. I hope to demonstrate that I can reach the same results. Below is my plan for building the parameterized polynomial models....

1. I will convert the molal concentrations reported in the data sets of Giaque, Rard and Staples to % $\text{H}_2\text{SO}_4$  and molar concentrations. I will compare my results to those in the data tables of the CRC Handbook to increase my personal confidence that I did this correctly.
2. I will also establish my models for converting molal to % $\text{H}_2\text{SO}_4$  and compare them to literature models. I will create a data set for converting molar concentration ( $M$ ) to % $\text{H}_2\text{SO}_4$  ( $w$ ) and vice-versa. Molality ( $m$ ) can be converted to  $w$  by math. To convert  $M$  we will need the measured density included in the calculation.
3. I will build a model for  $a_{\text{H}_2\text{O}}$  using the data sets of Giaque, Rard and Staples and fit the data to a polynomial<sup>11</sup> using % $\text{H}_2\text{SO}_4$ . I will compare the data. If there is no major disagreement, I will combine the data sets and make a model. I will also compare my model to literature models for  $a_{\text{H}_2\text{O}}$ .<sup>12</sup>
4. I will combine the acidity function data sets of Gillespie & Peel, Johnson, and Jorgenson and fit this to a polynomial model. I will produce a parameterized model that represents the data table and predicts the value of  $H_0$  from % $\text{H}_2\text{SO}_4$ . I will not use the data set of Tickle, which differs significantly from the above, or the data set of Paul & Long, which is very different at high acidities. I think that I will plot all five sets of data to show where they agree and diverge.
5. I will convert the data from Cox<sup>7</sup> that is given in terms of molarity to %mass. I will compare this to his second data set that is uses %mass. This would be an attempt to confirm that our conversion method is working as well as the undocumented method that was used by Cox.<sup>13</sup>

<sup>8</sup> See "Extreme Ester Hydrolysis", available via our Moodle site.

<sup>9</sup> "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>

<sup>10</sup> CRC Handbook of Chemistry and Physics, 106<sup>th</sup> 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](#).

<sup>11</sup> All three of these data tables are, in fact, the output of parameterized models that are based on extensive data obtained by the authors or from literature. I am reverse-engineering models for which I have the outputs, but not the parameters.

<sup>12</sup> Zhang et al. have recently presented a complex parameterized model for predicting the value of  $a_{\text{H}_2\text{O}}$  in terms of mole fraction from the calculated mole fraction of the mixture. See "An accurate water activity model for sulfuric acid solutions and its implementation on moisture sorption isotherm determination." L. Zhang, P.M. Grace, D.-W. Sun, *Drying Technology*, **2022**, *40*, 2540–2549. <https://doi.org/10.1080/07373937.2020.1869037>.

<sup>13</sup> I am "training" a model using a data set and then validating that model against an independent data set. If all goes well, I can go forward with my model to predict a property from an input. This is how AI models work. They are all just models that predict and output when given an input. The whole world is currently fascinated by words produced from a very complex curve fit.

## 1. Molality to Molarity and Percent

Molal? Why not molar or percent weight? Molarity ( $M$  or  $\text{mol L}^{-1}$ ) is a favourite of bench chemists. It related directly to rate constants and can be used as a measure of concentration in gas-phase chemistry, as long as the volume does not change. Volume changes in liquids are considered to be insignificant by most milquetoast room-temperature sea-level chemists.<sup>14</sup> Molarity also takes a bit of fiddling. One must add so many moles of solute to a solution, mix until dissolved, and then carefully dilute to the final volume.

Molality ( $m$  or  $\text{mol kg}^{-1}$ ) is a measure of concentration that is independent of volume. It is a standard concentration in chemical engineering data sets because it will not change with temperature or pressure. It is also very easy to set molal concentration. One just throws a set amount of solute into a kilogram of water. Done. As a result, many data tables from the literature use units of molality ( $m$ ) or percent weight ( $w$ ),<sup>15</sup> both of which are interconvertible by simple math.<sup>16</sup>

$$w = \frac{m \cdot 98.079}{1000} 100 \quad (1)$$

$$m = \frac{\frac{w}{100} 98.079}{\frac{w}{100} 98.079 + (1 - \frac{w}{100}) 18.015} \quad (2)$$

Converting from molality to molarity is trivial, but one needs to know the density of the mixture.

$$M = \frac{m \cdot \rho}{m \cdot 98.079 + 1000} \quad (3)$$

Densities must be measured or found in data tables. A table that reports for densities of mixtures of sulphuric acid and water according to molal, molar and percent concentrations is available in the CRC Handbook.<sup>10</sup> I can use this table to check my math for interconverting the three concentration values and you can check my math by examining the *Python* notebook that I have made available.<sup>17</sup>

I have used the molal concentrations ( $m$ ) of  $\text{H}_2\text{SO}_4$  to calculate values for percent mass ( $w$ ) and molarity ( $M$ ) using the density data table from the CRC handbook. I plotted the differences between my calculations and the reported values in that CRC table, and the results are identical within the precision of the numbers available. See Figure 1 on the next page for these plots. My math agrees with the math of the editors of the CRC Handbook. More importantly, this exercise shows that there are no typographical errors in the four columns of data in that table.<sup>18</sup> We have completed the first goal in our plan: we can convert between units of  $m$ ,  $M$ , and  $w$  within a data table for density.

<sup>14</sup> We salute the rocket scientists and chemical engineers that operate far beyond the borders of standard temperature and pressure. No one really needs all ten of their fingers.

<sup>15</sup> In this document the following abbreviations will be used:  $M$  is molarity (moles in a total volume of one litre);  $m$  is molality (moles added to a kilogram of solvent);  $w$  is percent mass (mass solute/total mass);  $\rho$  is density ( $\text{g/cm}^3$  or  $\text{kg/m}^3$ ).

<sup>16</sup> The following constants will be used: mole weight of  $\text{H}_2\text{O}$  is  $18.015 \text{ g mol}^{-1}$ ; mole weight of  $\text{H}_2\text{SO}_4$  is  $98.079 \text{ g mol}^{-1}$ .

<sup>17</sup> The *Python* notebook for all following calculations and plots can accessed via Google Colab at [https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class\\_30\\_Yates\\_New/ParameterizedModels.ipynb](https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class_30_Yates_New/ParameterizedModels.ipynb).

<sup>18</sup> I typed these values or used sketchy OCR data from images of the written pages and the corrected the many errors. Transcribing data is a common source of error. Always validate your databases.

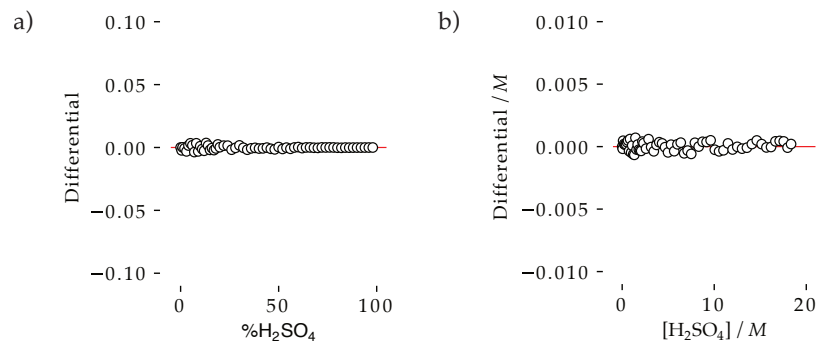


Figure 1: Plots of differences between converted values for concentration and the corresponding reported values in the CRC data set. All differences are within the precision of the reported values:

a) Percent weight ( $w$ ) from molality ( $m$ );  
b) molarity ( $M$ ) from molality ( $m$ ).

←

## 2. A Model for Density vs. $w$

The values of concentration in terms of molality or  $\%\text{H}_2\text{SO}_4$  ( $w$ ) are directly related to each other. However, conversion of these two units to molarity requires knowing the density of the given mixture, which is a measured value. We used the densities listed in the data tables for the exercise in section 1 above. But if we want a value for density at any value of  $w$  or  $m$ , then we will need a model that predicts density vs.  $w$ . First we will need a reliable data set. Let us compare some data sets other than the one presented in the CRC handbook.

### Literature Density Data Sets

As stated above, the density data for sulfuric acid–water mixtures at  $20^\circ\text{C}$  was obtained from the CRC Handbook. A second table available in the online-only edition of the CRC handbook also presents density data at a range of temperatures between  $0$  and  $100^\circ\text{C}$ .<sup>10</sup> This data appears in many works throughout the last century, such as Perry.<sup>19</sup> All of these data sets are clearly results of a parameterized model. These models are slight adjustments to a data set presented in 1905 by Bien & Domke.<sup>20</sup> I did my best to translate the German text. It appears that the authors assembled sets of density data reported by many previous works dating back as far as 1861 and also measured their own data. They then fit this data to an undocumented parameterized model and produced a final table of densities at every integer percent weight between 1 and 100 % at temperatures between  $0$  and  $60^\circ\text{C}$ . Their data was standardized to a density of pure water of  $1.0000\text{ g cm}^{-3}$  at  $15^\circ\text{C}$ . The modern tables adjust for our current definition of the density of water, but otherwise use the data as presented in 1905. There are many reports in the literature of density data at various concentrations and temperatures. When I have examined them, I find that they usually agree to the fifth decimal point with the Domke data set.

That is a long way to say that the data set from the CRC handbook is the gold standard for today's chemist. Even though it is likely to be a polynomial fit of a data table from 1905 that was itself a polynomial fit of data<sup>21</sup> from

<sup>19</sup> "Perry's Chemical Engineers' Handbook, 7<sup>th</sup> ed." McGraw-Hill, New York, 1997, D.W. Green, J.O. Maloney, Eds., Available via interlibrary loan. See table 2-101 on page 2-107.

<sup>20</sup> "Über Dichte und Ausdehnung der Schwefelsäure in wässriger Lösung, ein Beitrag zu ihrem physikalisch-chemischen Verhalten." J. Domke, W. Bein, *Z. Anorg. Chem.*, 1905, 43, 125–181. <https://doi.org/10.1002/zaac.19050430114>. Available via interlibrary loan.

<sup>21</sup> Or perhaps the data was carefully plotted on a large sheet of fine graph paper and then a French curve was used to create a best line through the data. That method is the essentially the same as modern bspline parameterized models made using a fancy computer.

the 19<sup>th</sup> century. Densities of liquids are easy to measure and it is no surprise that these values have been precisely known for hundreds of years.

### Our Data

First I will gather the data. I have access to the following data sets...

- 100 data pairs of  $\rho$  at integer values of  $w$  (0 to 100 %) for temperatures in the range of 0 and 100 °C. This data is from Perry<sup>19</sup> and is based on an undocumented parameterized model from Domke.<sup>20</sup> This model is also the basis for the CRC Handbook data tables.<sup>10</sup>
- 30 data pairs of measured values of  $\rho$  and  $w$  (0 to 99.6 %) for the temperatures of 0, 25, 50 and 75 °C.<sup>22</sup>
- 27 data pairs of measured values of  $\rho$  and  $w$  (0.1 to 40 %) for temperatures in the range of 0 and 100 °C.<sup>25</sup>
- 92 data pairs of values of  $\rho$  and  $w$  (1.7 to 100 %) for the temperature of 20 °C. This data set is very detailed in the range of 98 to 100 %  $\text{H}_2\text{SO}_4$ .<sup>23</sup> This data has no identified source. It appears to be a result of a parameterized model but I do not know that for sure.

I will use the data table from Perry, that presents densities at every integer value of % $\text{H}_2\text{SO}_4$ . I am aware that this data is the result of an undocumented parameterized model and that I am “making a model of a model”.

### Literature Parameterized Models

There are sophisticated models available for predicting the density of mixtures of sulphuric acid and water. Clegg & Wexler have reported a model based on two bspline fits of 16 parameters each combined with a polynomial fit with 12 parameters.<sup>24</sup> That's a model with 44 parameters. I'm not doing that. Oca et al. surveyed several models and data sets and produced experimental data for  $\text{H}_2\text{SO}_4$  mixtures from 0.1 to 40 %.<sup>25</sup> The model they ultimately proposed has six parameters. That seems doable.

References in Oca led me to a much simpler parameterized model by Hyvärinen et al.<sup>26</sup> that uses a basic 4<sup>th</sup>-degree polynomial. That's all you need. I will attempt to make a simple polynomial model using *Python* tools and see how it compares to the literature data and the literature models. With reassurance that the method is working within acceptable precision, I will then proceed to make models for  $a_{\text{H}_2\text{O}}$  and  $H_0$ .

<sup>22</sup> “The Viscosities of Mixtures of Sulfuric Acid and Water.” F. H. Rhodes, C. B. Barbour, *Ind. Eng. Chem.*, **1923**, 15, 850–852. <https://doi.org/10.1021/ie50164a033>.

<sup>23</sup> Page 222 in “Tables of physical and chemical constants and some mathematical functions, 15<sup>th</sup> ed.” G.W.C. Kaye, T.H. Laby, *Longman Inc.*, New York, 1986. Available in UPEI library reference section QC61.K3 1986 REF

<sup>24</sup> “Densities and Apparent Molar Volumes of Atmospherically Important Electrolyte Solutions.” S. L. Clegg, A. S. Wexler, *J. Phys. Chem. A*, **2011**, 115, 3393–3460. <https://doi.org/10.1021/jp108992a>. This paper has an extensive survey of available data sources for density of  $\text{H}_2\text{SO}_4$ .

<sup>25</sup> “Review and Analysis of Thermophysical Properties of a Sulfuric Acid–Water Electrolyte.” L. Oca, J.M. Campillo-Robles, M.M. Bou-Ali, *J. Chem. Eng. Data*, **2018**, 63, 3572–3583. <https://doi.org/10.1021/acs.jced.8b00466>

<sup>26</sup> “Surface Tensions and Densities of Sulfuric Acid + Dimethylamine + Water Solutions.” A.-P. Hyvärinen et al., *J. Chem. Eng. Data*, **2004**, 49, 917–922. <https://doi.org/10.1021/je034225b>.

### *The Path to Our Model*

The *Python* code for separating the dataset from Perry into a training and validation set and for fitting the model is available in the notebook described previously.<sup>27</sup> I randomly removed 25 % of the data from the data set and used the remaining 75 % in the polynomial curve fit. I will produce the following plots...

<sup>27</sup> See reference 17 on page 3

- **COMPARING DATA SETS:** A plot of three data sets for  $\rho$  vs  $w$  at 25 °C – Perry (model), Rhodes & Oca (measured). Does the literature agree? I will also plot the data at 20 °C for Perry (model), Rhodes (measured) & Kaye (model?). See Figures 2 and 3 on the following page.
- **4<sup>TH</sup>-DEGREE POLYNOMIAL FIT:** I will use the data from Perry to produce a 4<sup>th</sup>-degree polynomial parameterized model. I will plot the model against the data set and plot the differentials. Hyvärinen reported a model of 4<sup>th</sup>-degree.<sup>26</sup> I will compare that model to my own.
- **6<sup>TH</sup>-DEGREE POLYNOMIAL FIT:** I will repeat the above exercise using a 6<sup>th</sup>-degree polynomial parameterized model and compare my results to the 6<sup>th</sup>-degree model reported by Oca.<sup>25</sup> How did I do?

The final result will be a parameterized model for the density of a sulphuric acid solution at 25 °C. The equation can be simply stated and used by anyone. If this exercise is successful then I will be more confident in making models for  $a_{\text{H}_2\text{O}}$  and  $H_0$ .

### *Comparing Data Sets*

Figures 2 and 3 on the following page are plots of density data using data sets that are available for 25 °C and 20 °C. I observe that the literature data is in good agreement across the century of time between these data sets.

When I plotted the data from Rhodes,<sup>22</sup> I observed **five typographical errors** in the data set. The data at the concentration of 51.2 %H<sub>2</sub>SO<sub>4</sub> was off by 0.1 g cm<sup>-3</sup> in all four temperature series. That seems like too much of a coincidence. I corrected the data and made a note in the data tables presented on my textitGitHub repository. Others can judge my decisions easily since I make my data and methods available as open source materials.

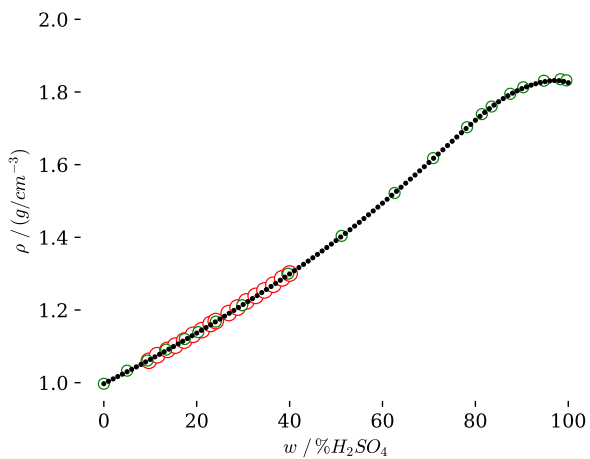


Figure 2: Plots of  $\rho$  vs  $w$  for data sets of Perry (small closed circles), Rhodes (black open circles) and Oca (red open circles) at  $25^\circ\text{C}$

←

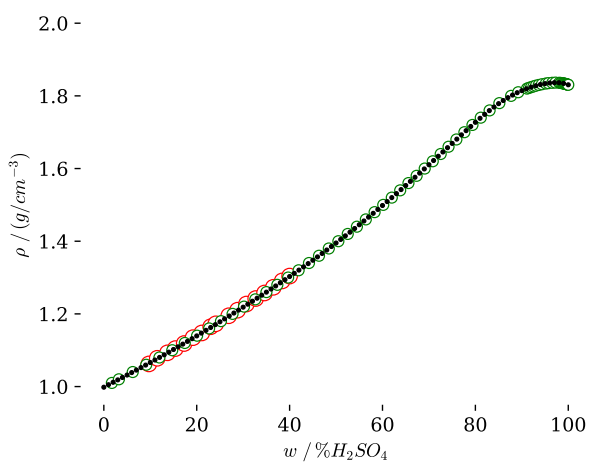


Figure 3: Plots of  $\rho$  vs  $w$  for data sets of Perry (small closed circles), Kaye (black open circles) and Oca (red circles) at  $20^\circ\text{C}$

←

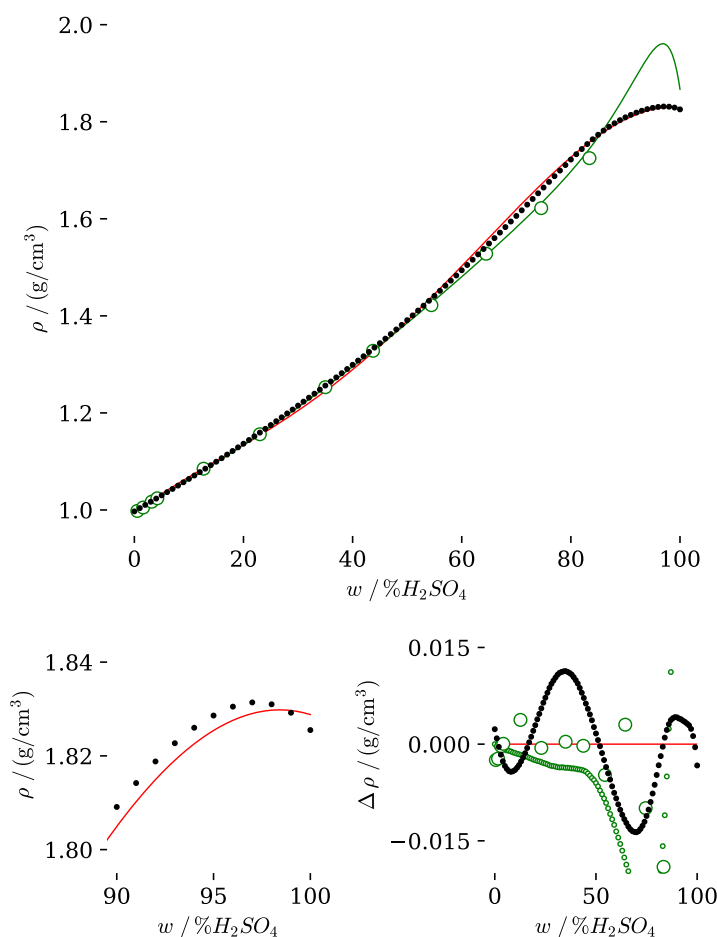


### 4<sup>th</sup>-Degree Polynomial Fit

A 4<sup>th</sup>-degree polynomial model fits adequately, but we can do better. The data set was weighted at the ends to try to force a better fit<sup>28</sup> in the high concentration region (see lower left plot).

The model of Hyvärinen deviated badly. The authors had used their own measured density data. I extracted the data and observed that the model fits its own data very well, within its bounds. The lesson is that polynomial models should never be used for extrapolation. The Hyvärinen model matches its data between 0.5 and 84.5 %H<sub>2</sub>SO<sub>4</sub>. It will go off the rails outside of that range, as it should.

The experimental data of Hyvärinen deviated from that of Perry (and the other data sets) above 60 %H<sub>2</sub>SO<sub>4</sub>. The model outperformed my own polynomial fit until that point, and then it followed the deviant data and diverged.



<sup>28</sup> All of the details are presented in the code that generated the plot in Figure 4 and it is available in the *Python* notebook. See reference 17 on page 3.

HYVÄRINEN MODEL: The equation used in the polynomial fit is...

$$\rho = a_4 \chi^4 + a_3 \chi^3 + a_2 \chi^2 + a_1 \chi + a_0$$

$\chi$  is mole fraction of H<sub>2</sub>SO<sub>4</sub> and it returns the density,  $\rho$ , in kg m<sup>-3</sup>

$$a_0 = 997.05 \quad a_1 = 3420$$

$$a_2 = -7230 \quad a_3 = 9090$$

$$a_4 = -4410$$

Figure 4: Plots a 4<sup>th</sup>-degree polynomial model.

TOP: The top plot presents the data set from Perry as small closed circles. The polynomial fit is traced by the red line. The parameterized model of Hyvärinen is traced in green and the experimental data used to make that model is shown in larger open green circles.

LOWER LEFT: This plot is a closeup of the highest concentration region.

LOWER RIGHT: This is the differential plot. We can see that our model deviates significantly (small closed black circles). The Hyvärinen model is worse (small closed green circles). However, when the differentials are calculated between the Hyvärinen model and the data that was used to build it, we see good agreement (large green open circles).

POLYNOMIAL MODEL: The equation used in my own polynomial fit is...

$$\rho = a_4 w^4 + a_3 w^3 + a_2 w^2 + a_1 w + a_0$$

$w$  is %H<sub>2</sub>SO<sub>4</sub> and it returns the density,  $\rho$ , in g cm<sup>-3</sup>

$$a_0 = 9.94729 \times 10^{-1} \quad a_1 = 8.29495 \times 10^{-3}$$

$$a_2 = -1.30170 \times 10^{-4} \quad a_3 = 3.58003 \times 10^{-6}$$

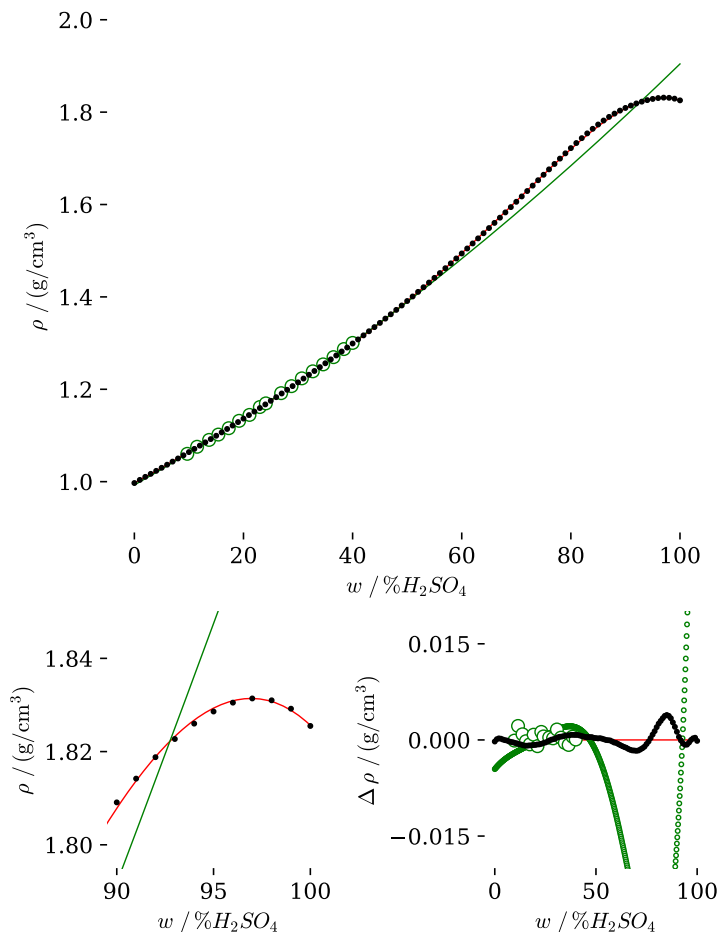
$$a_4 = -2.27374 \times 10^{-8}$$

Next, we will try a six-parameter fit. Lets see how that improves the model.



### 6<sup>th</sup> Degree Polynomial Fit

A 6<sup>th</sup>-degree polynomial model fits very well. the differentials are much smaller, but we will never eliminate the artifacts inherent in polynomial fits.<sup>29</sup> The data set was weighted at the ends.



The Oca model is really a second degree polynomial w.r.t each of mass fraction ( $\bar{w}$ ) and temperature ( $T$ ). We chose a fixed temperature of 25 °C. It is a three-parameter model and will not fit the entire range of the data from Perry. It performs adequately in the range of 0 to 40 %H<sub>2</sub>SO<sub>4</sub>, but fails quickly if you venture outside the range of data used to create it.

The 6<sup>th</sup>-degree polynomial model is a seven-parameter fit and does an excellent job of staying close to the training data.

We have completed the second goal of this exercise: we can make polynomial models that relate concentration to density for sulphuric acid mixtures and have compared these models to literature models. I conclude that it is best to fit my own bespoke models rather than rely on literature polynomial models. As we have seen, there are more sophisticated models available in

<sup>29</sup> More parameters will make the deviations smaller and smaller. Try a 12-parameter fit using the *Python* notebooks that I have provided.<sup>17</sup>

Figure 5: Plots a 6<sup>th</sup>-degree polynomial model.

TOP: The top plot presents the data set from Perry as small closed circles. The polynomial fit is traced by the red line (it fits the data so closely that it is hidden by the points). The parameterized model of Oca is traced in green. The experimental data used to make the model of Oca is shown in larger open green circles.

LOWER LEFT: This plot is a closeup of the highest concentration region.

LOWER RIGHT: This is the differential plot. We can see that our model is much closer to the data (small closed black circles). The differentials between the Oca model and the Perry data set (small closed green circles) is also a great fit within the bounds of its data set. Again we see why we should never extrapolate with polynomial models. The differentials between the Oca model and the data that was used to build it show excellent agreement (large green open circles).

OCA MODEL: The equation used in the polynomial fit is...

$$a_{0,0} \bar{w}^0 T^0 + a_{0,1} \bar{w}^0 T^1 + a_{0,2} \bar{w}^0 T^2 + a_{1,0} \bar{w}^1 T^0 + a_{1,1} \bar{w}^1 T^1 + a_{2,0} \bar{w}^2 T^0$$

$\bar{w}$  is mass fraction ( $w/100$ ) of H<sub>2</sub>SO<sub>4</sub> and it returns the density,  $\rho$ , in kg m<sup>-3</sup>.  
 $T = 298.15$  K

$$\begin{aligned} a_{0,0} &= 1122 & a_{0,1} &= -0.5076 \\ a_{0,2} &= 2.484 \times 10^{-4} & a_{1,0} &= 976.4 \\ a_{1,1} &= -1.015 & a_{2,0} &= 237.8 \end{aligned}$$

POLYNOMIAL MODEL: The equation used in my own polynomial fit is...

$$\rho = a_6 w^6 + a_5 w^5 + a_4 w^4 + a_3 w^3 + a_2 w^2 + a_1 w + a_0$$

$w$  is %H<sub>2</sub>SO<sub>4</sub> and it returns the density,  $\rho$ , in g cm<sup>-3</sup>

$$\begin{aligned} a_0 &= 9.97298 \times 10^{-1} & a_1 &= 6.34762 \times 10^{-3} \\ a_2 &= 4.29036 \times 10^{-5} & a_3 &= -6.35832 \times 10^{-7} \\ a_4 &= 3.16815 \times 10^{-9} & a_5 &= 1.74392 \times 10^{-10} \\ a_6 &= -1.66035 \times 10^{-12} \end{aligned}$$

the literature that predict density across a wide range of temperatures and concentrations. I did not evaluate those — I'm not going to transcribe 25 or 75 parameters. I'll role my own models using data sets that align with my needs.

### A Spline Fit

After completing this exercise (see below) I learned a lot about fitting literature data tables. These tables were created using a parameterized model and are usually closely spaced in the series. the tables for  $\rho$  are a good example of this. I can spend a lot of mental energy on these polynomial fits and their validation, or I can accept the data as it is and perform a simple interpolation. If the data is closely spaced along both the  $x$  and  $y$ -axes then this will be the best option. Figure 6 demonstrates a line created using a spline interpolation. More on this method will be presented later in this document. I have travelled back from the last page to add this information here.

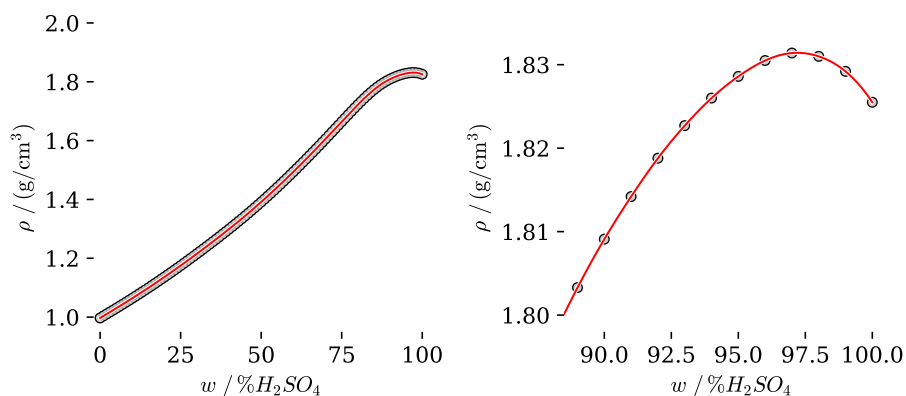


Figure 6: Plots  $H_0$  vs  $m$  ( $\text{H}_2\text{SO}_4$ ) for the data of Gillespie and Johnson combined. Where duplicate values existed, the data from Gillespie was used. The spline interpolation is presented on the plot. This combined data set is precise and closely spaced so an interpolation is likely the best way to represent it.

### 3. A Model for $a_{\text{H}_2\text{O}}$ vs $\text{H}_2\text{SO}_4$

We will now move on to our third goal, constructing a model to predict  $a_{\text{H}_2\text{O}}$  at any given concentration of a sulphuric acid-water mixture. I have data sets for this relationship from publications by Giaque<sup>4</sup>, Rard<sup>5</sup> and Staples.<sup>6</sup> First let us plot these data sets together to see if they agree.

Each data set reports values of  $a_{\text{H}_2\text{O}}$  (as fractional vapour pressure of water,  $p_x/p_0$ ) against concentration of  $\text{H}_2\text{SO}_4$  (as molal values). I will convert concentration from  $m$  to  $w$  and plot  $a_{\text{H}_2\text{O}}$  vs.  $w$  to compare the data sets as shown in Figure 7 on the following page.

We can see that the Giaque data set (black) is sparse at lower concentrations and contains data from 9% to 100%  $\text{H}_2\text{SO}_4$ . The Rard data set (green) tracks perfectly with Giaque and contains more data points with a span

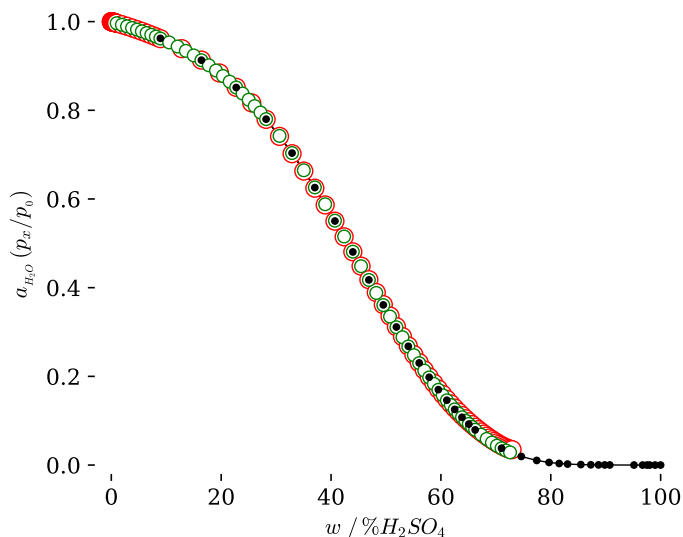


Figure 7: Plots  $a_{\text{H}_2\text{O}}$  vs  $w$  for the data sets of Giaque (small closed black circles), Rard (medium open green circles), and Staples (large open red circles).

←

from 1 % to 72 % $\text{H}_2\text{SO}_4$ . I suspect it was derived from the same parameterized model or used the same data sets to generate the model. The Staples data had more data for low concentrations and spanned from 0.1 % to 73 % $\text{H}_2\text{SO}_4$ . It tracked almost perfectly with the both Giaque and Rard at concentrations below 60 % $\text{H}_2\text{SO}_4$ , but diverged slightly at higher concentrations.

I will combine the data sets from Rard and Giaques.<sup>30</sup> This will give me a data set for  $a_{\text{H}_2\text{O}}$  vs  $w$  that will cover from 1 % to 100 % $\text{H}_2\text{SO}_4$ . The results of Yates et al.<sup>9</sup> were in the range of 10 % to 99 % $\text{H}_2\text{SO}_4$ , so that is all that I need.

### Fitting the Model

I combined the data of Rard<sup>5</sup> and Giaque<sup>4</sup> by concatenating and sorting the data sets. I weighted the last ten points. See Figure 8 on the next page for the results of this extreme polynomial fit.

I was unable to obtain a model that provided an acceptable level of accuracy for predicting values of  $a_{\text{H}_2\text{O}}$  in the high concentration region. As you can see, in desperation I used a 21-parameter fit. Still not good.

The range in the values is from 1 down to  $10^{-9}$ . This is just too large of a change in magnitude. The fit is actually excellent with a median error of  $8 \times 10^{-5}$ . But when data values are smaller than the error, we can have huge relative errors. The error in the final point is 325 %

I tried lots of ways to improve the fit. Setting the weights to values of  $1/a_{\text{H}_2\text{O}}$  puts increasing weight toward the smaller values. this resulted in a fit that followed the data with 2 % but the *Python* numerical engine warned of errors. We needed such a high degree of a polynomial we were getting past the limits of double-floating point precision.

<sup>30</sup> I might remove duplicate points, or I might not. The *Python* notebooks will have the recipe for my data sauce. Not only will others be able to examine my exact method, I will also be able to see what I have done. One month from now I will have no clue what was done to create the model. I could write very detailed notes, or I could unambiguously describe process in the *Python* code. See reference 17 on page 3.

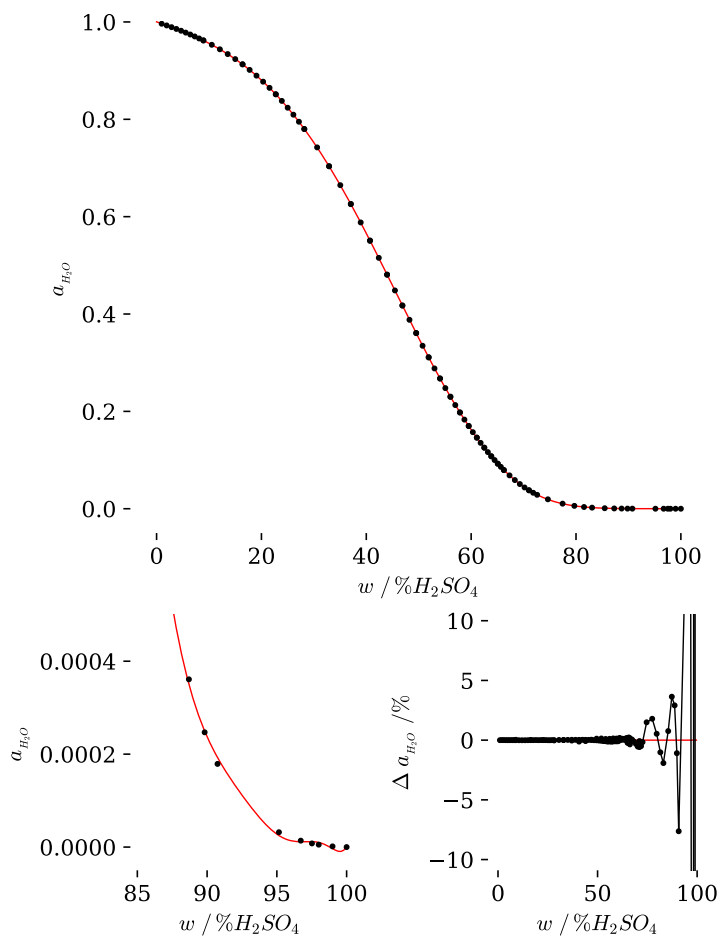


Figure 8: Plots a 20<sup>th</sup>-degree polynomial model for  $a_{H_2O}$ .

TOP: The top plot presents the data set from the combined data sets of Rard and Giauque for  $a_{H_2O}$  vs  $w$ . The red line is calculated from the polynomial fit parameters.

LOWER LEFT: This plot is a closeup of the highest concentration region.

LOWER RIGHT: This is the differential plot (% difference). The model fits very well until we enter the region with very small values of  $a_{H_2O}$ . The average error between predicted and measured values was about  $10^{-5}$ . This is great for most of the plot where  $a_{H_2O} > 10^{-2}$ , it is a large error when the values of  $a_{H_2O}$  fall below  $10^{-4}$  and become enormous for the final value where  $a_{H_2O} \approx 10^{-9}$

←

### A Piecewise Model

I tried splitting the data up into three sections and performing a piecewise fit. Each section still required high degrees of parameterization to get an adequate fit. Table 1 describes regions and models that were used.

$x$ Range % $\text{H}_2\text{SO}_4$	$y$ range	# of Points	Degree
$0 \leq w < 60 \%$	$1$ to $10^{-1}$	62	12
$60 \leq w < 85$	$10^{-1}$ to $10^{-3}$	28	8
$w \geq 85 \%$	$10^{-3}$ to $10^{-9}$	11	6

The piecewise polynomeial fit is displayed in Figure 9. It fits very well with most relative deviations between predicted and given data points being less that 0.05 %. The largest relative error was at the final data point. The magnitude of that point was  $10^{-9}$  and we hit it within a relative error of about 1 %.

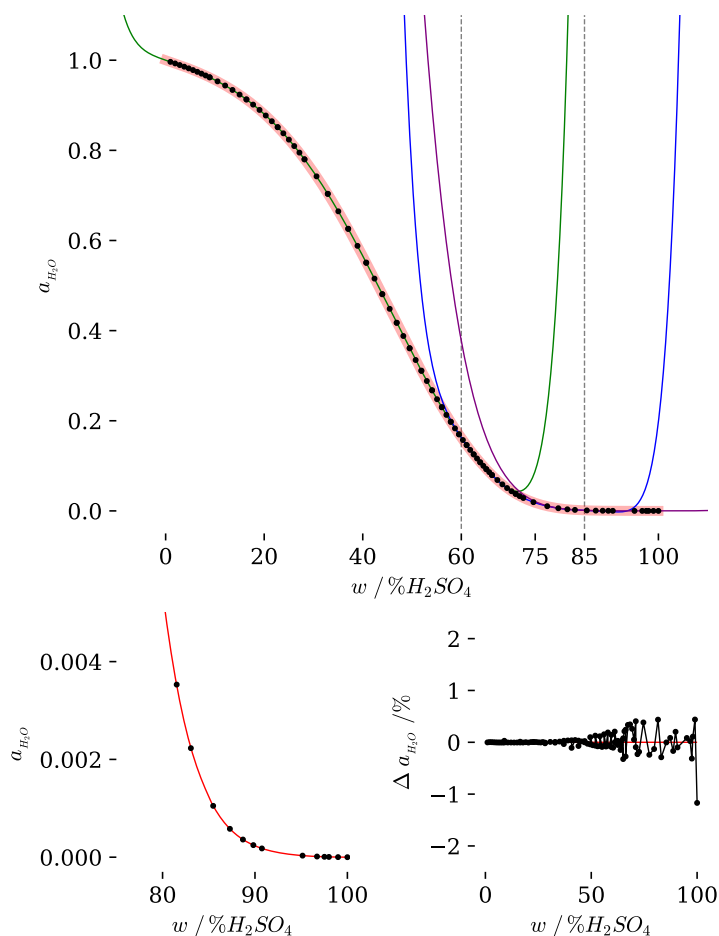


Table 1: Pieces for the piecewise model. Small piece require fewer terms for an accurate fit.

Figure 9: Plots a piecewise polynomial model for  $a_{\text{H}_2\text{O}}$ .

TOP: The top plot presents the data set from the combined data sets of Rard and Giauque for  $a_{\text{H}_2\text{O}}$  vs  $w$ . The thick transparent red line is calculated from the piecewise polynomial fit parameters. The piecewise polynomials that were used are also plotted. You can see why we never extrapolate with polynomials. Each was precise only within its assigned range:  $0 \leq x < 60$  (green),  $60 \leq x < 85$  (blue), and  $x \geq 85$  (purple).

LOWER LEFT: This plot is a closeup of the highest concentration region. We observe an excellent fit using the 6<sup>th</sup>-degree polynomial fit in the region where  $x \geq 85$ .

LOWER RIGHT: This is the differential plot (% difference). The model fits very well until we enter the region with very small values of  $a_{\text{H}_2\text{O}}$ . The relative error only becomes visible toward the end of the plot. the average error for the final piece in the range  $x \geq 85$  was 0.1 %. The relative error for the final point was -1.1 %.

In a span of predicted values  $10^0$  to  $10^{-9}$ , we stayed very close to the given values.

### Describing the Piecewise Function

The model consists of three polynomials of degrees 12, 8 and 6 used in the regions of  $w < 60\%$ ,  $60\% \leq w < 85\%$ , and  $w \geq 85\%$ .

$$f(x) = \begin{cases} w < 60\% & \sum_{n=0}^{12} a_n x^n \\ 60\% \leq w < 85\% & \sum_{n=0}^8 b_n x^n \\ w \geq 85\% & \sum_{n=0}^6 c_n x^n \end{cases} \quad (4)$$

The values for the coefficients are given in Table 2. Take note of the extreme degree of precision expressed. These high degree of the polynomials and the large range of magnitude in the values require every digit of precision that a 64-bit float can provide.

Coefficient	Value
$a_0$	$9.999\,602\,819\,100\,265 \times 10^{-1}$
$a_1$	$-3.758\,393\,717\,937\,19 \times 10^{-3}$
$a_2$	$7.228\,976\,593\,850\,476 \times 10^{-5}$
$a_3$	$-3.263\,212\,544\,746\,975\,6 \times 10^{-5}$
$a_4$	$3.926\,131\,352\,850\,760\,5 \times 10^{-6}$
$a_5$	$-3.265\,857\,346\,023\,434\,4 \times 10^{-7}$
$a_6$	$1.753\,610\,829\,845\,118\,6 \times 10^{-8}$
$a_7$	$-6.266\,897\,770\,870\,015 \times 10^{-10}$
$a_8$	$1.503\,635\,173\,094\,747\,8 \times 10^{-11}$
$a_9$	$-2.386\,415\,901\,562\,021\,4 \times 10^{-13}$
$a_{10}$	$2.400\,749\,367\,443\,683 \times 10^{-15}$
$a_{11}$	$-1.386\,588\,628\,647\,250\,2 \times 10^{-17}$
$a_{12}$	$3.502\,829\,658\,557\,864\,3 \times 10^{-20}$
$b_0$	$5.278\,125\,388\,514\,924 \times 10^3$
$b_1$	$-5.699\,778\,844\,885\,892 \times 10^2$
$b_2$	$2.683\,568\,358\,232\,527 \times 10^1$
$b_3$	$-7.193\,893\,888\,613\,687 \times 10^{-1}$
$b_4$	$1.200\,962\,559\,880\,584 \times 10^{-2}$
$b_5$	$-1.278\,611\,615\,731\,464 \times 10^{-4}$
$b_6$	$8.478\,783\,316\,886\,211 \times 10^{-7}$
$b_7$	$-3.202\,141\,249\,445\,121\,7 \times 10^{-9}$
$b_8$	$5.273\,726\,317\,476\,769 \times 10^{-12}$
$c_0$	$1.116\,478\,506\,456\,988\,5 \times 10^2$
$c_1$	$-6.869\,569\,447\,713\,705\,5$
$c_2$	$1.762\,746\,136\,695\,284 \times 10^{-1}$
$c_3$	$-2.414\,407\,034\,715\,221 \times 10^{-3}$
$c_4$	$1.861\,608\,264\,919\,891\,6 \times 10^{-5}$
$c_5$	$-7.660\,823\,213\,573\,518 \times 10^{-8}$
$c_6$	$1.314\,440\,485\,846\,466\,8 \times 10^{-10}$

Table 2: Pieces for the piecewise model. Small piece require fewer terms for an accurate fit.

←

The high precision in these numbers is needed. Try rounding them off to eight digits and watch the fit wander far away from the given values.

Good luck typing all that in. I do not believe that I achieved my goal of a text-based, portable expression of the polynomial model. It is text, but it's not very easy to carry.

### Using $\log a_{\text{H}_2\text{O}}$ Values for the Model

The huge range in values for  $a_{\text{H}_2\text{O}}$  was too much for a simple polynomial fit, and is on the edge of possible for the piecewise model that I used above (there are better ways to do that. I am not an expert).

The values of  $\log a_{\text{H}_2\text{O}}$  would have a dynamic range of less than one order of magnitude. Can I create a polynomial plot that predicts  $\log a_{\text{H}_2\text{O}}$  vs  $\% \text{H}_2\text{SO}_4$ ?

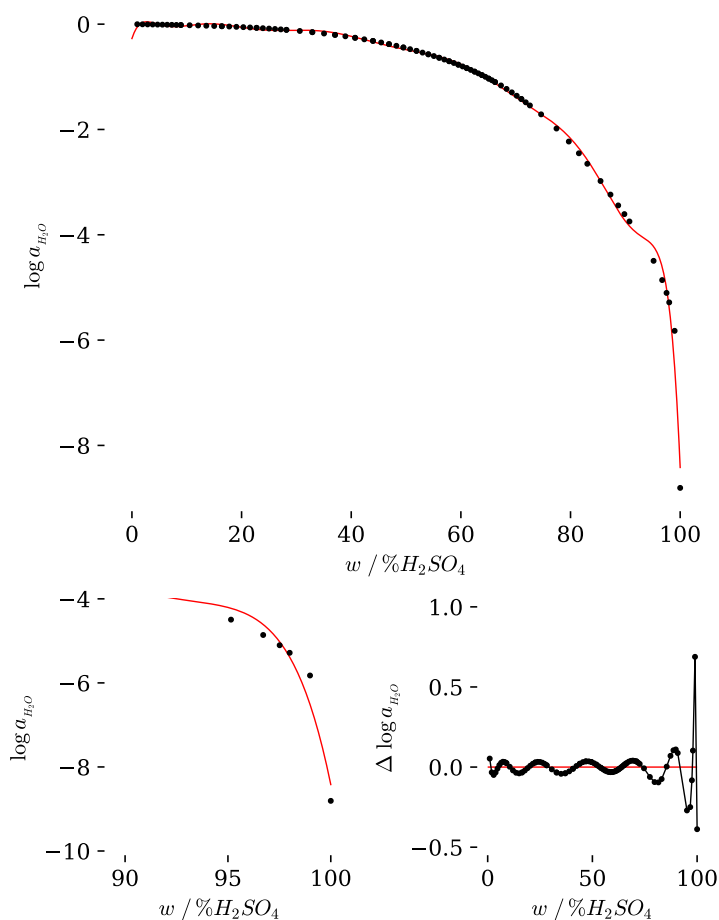


Figure 10: Plots a polynomial model for  $\log a_{\text{H}_2\text{O}}$ .

TOP: The top plot presents the data set from the combined data sets of Rard and Giauque for  $\log a_{\text{H}_2\text{O}}$  vs  $w$ . The red line is a 12<sup>th</sup>-degree polynomial fit. It was not a satisfying result.

LOWER LEFT: This plot is a closeup of the highest concentration region. We observe a poor fit using the 12<sup>th</sup>-degree polynomial fit in the region where  $x \geq 85$ . Remember that, on a log scale, a difference of 1 is equivalent to an error of 10-fold (1000 %).

LOWER RIGHT: This is the differential plot. The model fits poorly across the whole range. A difference of 0.01 in  $\log a_{\text{H}_2\text{O}}$  is the same as a 25 % error in the predicted value for  $a_{\text{H}_2\text{O}}$ .

←

Clearly, my amateur efforts have failed. For whatever reason, the shape of the  $\log a_{\text{H}_2\text{O}}$  vs  $\% \text{H}_2\text{SO}_4$  curve defies a good polynomial fit. The large change between the last and second-last values again result in a poor fit. We do not have balanced data. The extremely steep slope at the end means that small imperfections in the model will yield large differences in the prediction. this high-concentration region was the most important in my analysis.



### Using Python Interpolation Functions

I am using a polynomial fit to make a model for data that was likely derived from an undocumented polynomial model (at least, I haven't found the description yet). Perhaps I should just interpolate between the points.

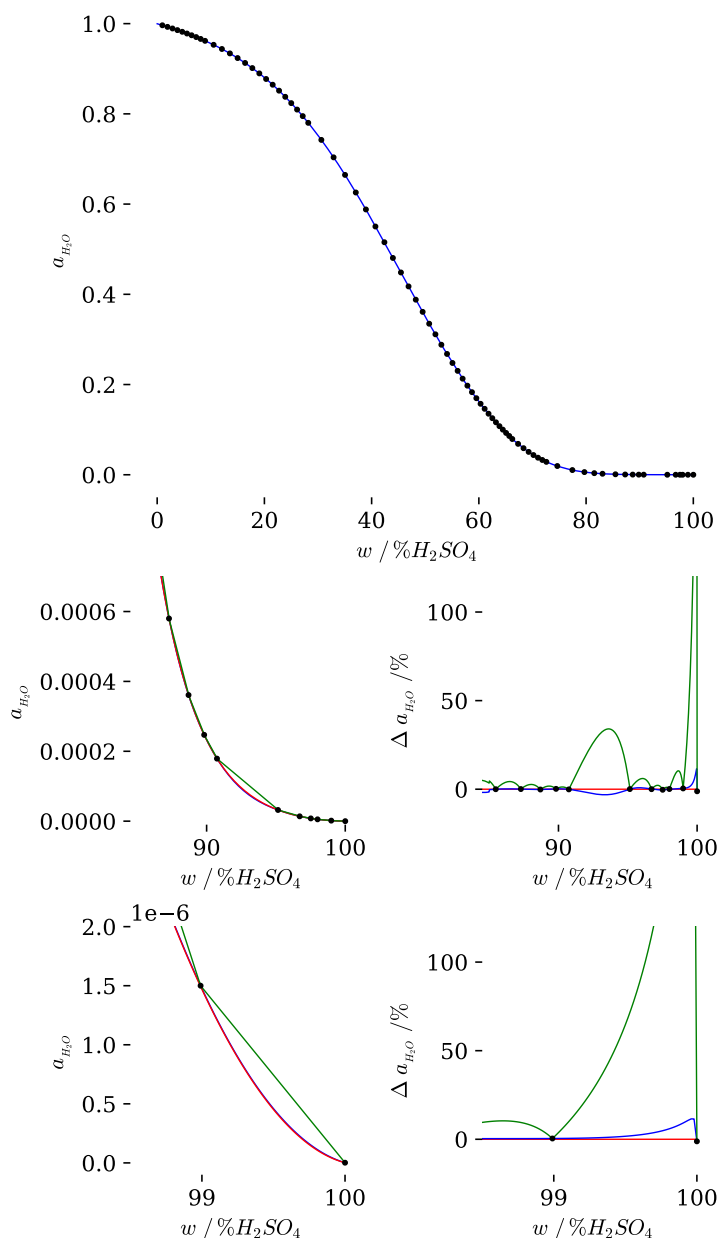


Figure 11: Plots of *Python* interpolation functions compared to the polynomial model for  $\log a_{\text{H}_2\text{O}}$ .

TOP: The top plot presents the data set from the combined data sets of Rard and Giauque for  $a_{\text{H}_2\text{O}}$  vs  $w$ . The blue line is an interpolation function based on Bsplines. It is essentially a piecewise polynomial fit with many, many pieces.

MIDDLE AND LOWER LEFT: These plots are magnifications of the high concentration regions where the greatest difference in magnitude for values of  $a_{\text{H}_2\text{O}}$  occur. Observe that a linear interpolation (green line, connect-the-dots) results in large relative errors when there is a significant span between points. My piecewise plot (red) and the interpolation function (blue) are nearly identical.

MIDDLE AND LOWER RIGHT: These are the corresponding relative differential plots. It is interesting to observe that the Bspline interpolation tracks very closely to the piecewise polynomial fit (3 pieces) that I had hacked together.

←

NOTE: The relative deviation between my own piecewise polynomial model and the spline interpolation is insignificant below 60 % $\text{H}_2\text{SO}_4$ . At higher concentrations, as the value of  $a_{\text{H}_2\text{O}}$  becomes very small, the relative deviations increase. The maximum deviations occur only at the highest concentrations of  $\text{H}_2\text{SO}_4$  and are less than  $\pm 5\%$ . The linear deviations for the linear interpolation are much greater, especially when there are larger spans between values and become unacceptable above 90 % $\text{H}_2\text{SO}_4$ .

The model would simply be the data set and the python command to return values from the interpolation function. That actually seems simpler and more portable than the text-based description of my polynomial model in Table 2 on page 14.

#### 4. A Model for $H_0$ vs $\% \text{H}_2\text{SO}_4$

Taking a lesson from all of the above wandering through various fits and interpolations I will now build a model to predict values of  $H_0$  at any given value for  $w$  ( $\% \text{H}_2\text{SO}_4$ ).

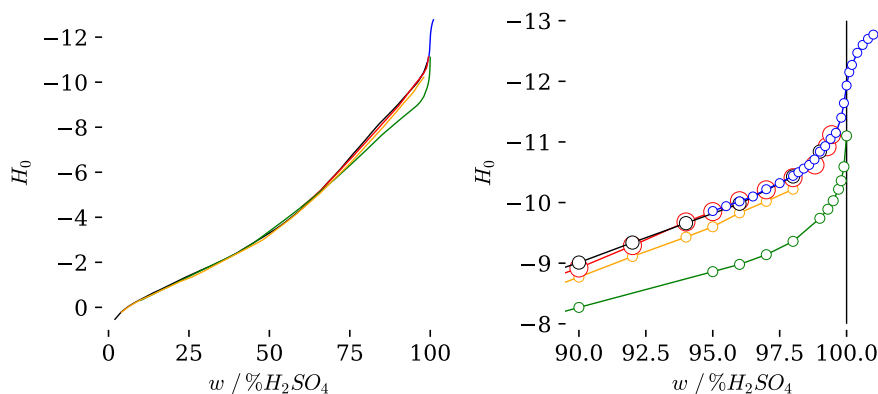


Figure 12: Plots  $H_0$  vs  $m$  ( $\% \text{H}_2\text{SO}_4$ ) for various data sets available in the literature. Data is from: Gillespie, 1971,<sup>31</sup> blue; Johnson, 1969,<sup>32</sup> black; Jorgenson, 1963,<sup>33</sup> red; Tickle, 1970,<sup>34</sup> orange; Paul & Long, 1953,<sup>35</sup> green

#### Assembling the Data

The data reported by Gillespie et al.,<sup>31</sup> Johnson et al.,<sup>32</sup> and Jorgenson et al.,<sup>33</sup> agreed closely. The values reported by Tickle et al.,<sup>34</sup> tracked closely with the above but were significantly different at higher concentrations. The values of Paul & Long<sup>35</sup> diverged greatly above 70  $\% \text{H}_2\text{SO}_4$ .

All five data sets described above are clearly outputs of parameterized models fit to experimental data. I am sure the original data could be found in the literature, but I am not going to spend the time. The data sets of Johnson and of Jorgenson are nearly identical across their entire mutual range of 60 to 99  $\% \text{H}_2\text{SO}_4$ . Johnson covers a broad range from 2 to 99  $\% \text{H}_2\text{SO}_4$ . The data reported by Gillespie is at extreme concentrations of 95 to 101  $\% \text{H}_2\text{SO}_4$ .<sup>36</sup> I will use these data sets to build my own model for  $H_0$  vs  $m$  from 2 to 101  $\% \text{H}_2\text{SO}_4$ .

One approach is to combine the three data sets to train a polynomial model. I can use all the data and obtain the best fit. Or I could use a spline interpolation using selected data points that create a smooth set of precise points assembled in the order of increasing  $\% \text{H}_2\text{SO}_4$ .

<sup>31</sup> "Hammett acidity function for some super acid systems." R.J. Gillespie, T.E. Peel, and E.A. Robinson, *J. Am. Chem. Soc.*, 1971, 93, 5083–5087. <https://doi.org/10.1021/ja00749a021>

<sup>32</sup> "Temperature variation of the  $H_0$  acidity function in aqueous sulfuric acid solution." C.D. Johnson, A.R. Katritzky, S.A. Shapiro *J. Am. Chem. Soc.*, 1969, 91, 6654–6662. <https://doi.org/10.1021/ja01052a021>

<sup>33</sup> "A Critical Re-evaluation of the Hammett Acidity Function at Moderate and High Acid Concentrations of Sulfuric Acid. New  $H_0$  Values Based Solely on a Set of Primary Aniline Indicators." M.J. Jorgenson, D.R. Hartter, *J. Am. Chem. Soc.*, 1963, 85, 878–883. <https://doi.org/10.1021/ja00890a009>

<sup>34</sup> P. Tickle, A.G. Briggs, J.M. Wilson, *J. Chem. Soc. B*, 1970, 65–70. <https://doi.org/10.1039/j29700000065>

<sup>35</sup> " $H_0$  And Related Indicator Acidity Functions." M.A. Paul, F.A. Long, *Chem. Rev.*, 1957, 57, 1–45. <https://doi.org/10.1021/cr50013a001>

<sup>36</sup> Your little-league coach was right. you can give more than 100 %. At least that is true for sulphuric acid. Pure  $\text{H}_2\text{SO}_4$  is, in fact, only 98  $\% \text{H}_2\text{SO}_4$ . This is due to the equilibrium dissociation of  $\text{H}_2\text{SO}_4$  to give  $\text{H}_2\text{O}$  and  $\text{SO}_3$  (observe that 100 % has a non-zero value for  $a_{\text{H}_2\text{O}}$ ). What is called "100  $\% \text{H}_2\text{SO}_4$  is pure  $\text{H}_2\text{SO}_4$  to which excess  $\text{SO}_3$  has been added. This is the famous "fuming sulphuric acid."

### A Polynomial Model

Again, I created a 12<sup>th</sup>-degree polynomial model and fit it to the data. Figure 13. It diverged at the highest values of  $m$ . The differential plot shows how the fit followed between the Johnson and the Jorgenson data sets, and then began diverging as it moved into the Gillespie data at very high values of  $w$ . There are probably better ways to do this such as splitting it into pieces like before but I feel that interpolation will be the best option here.

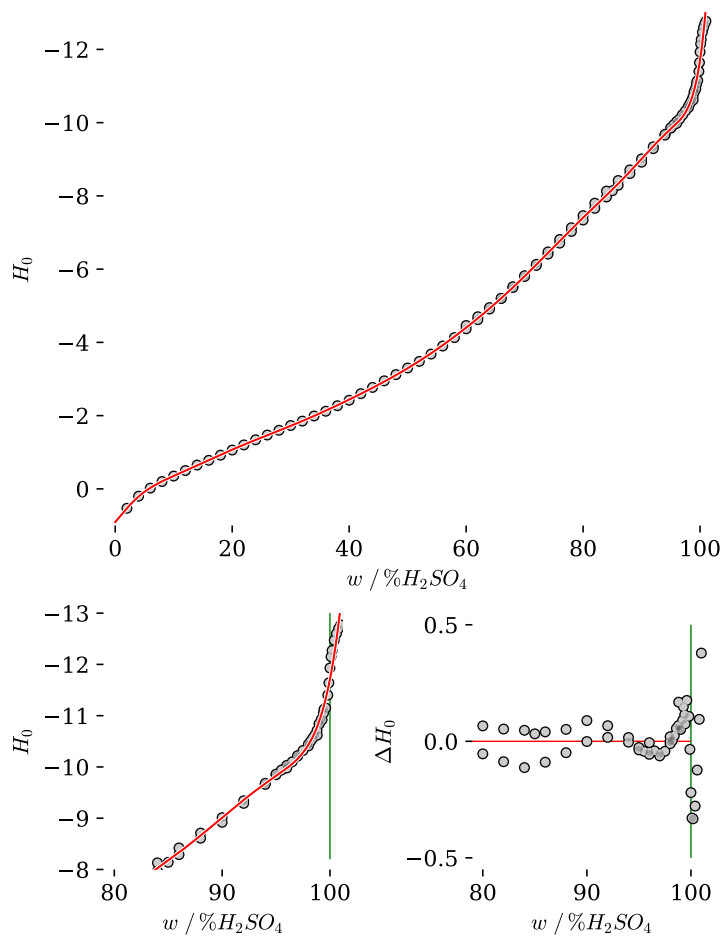


Figure 13: Plots of the polynomial model for  $H_0$  vs  $\%H_2SO_4$ .

TOP: The top plot presents the data set from the combined data sets of Johnson, Jorgenson and Gillespie. The red line is a 12<sup>th</sup>-degree polynomial fit. It was satisfactory at all but the most extreme concentrations.

LOWER LEFT: This plot is a closeup of the highest concentration region. We observe a divergence the region where  $w > 95\%H_2SO_4$ . Remember that, on a log scale, a difference of 0.3 is equivalent to an error of two-fold (100%).

LOWER RIGHT: This is the differential plot. The model fits adequately across the whole range with errors become excessive only above 100  $\%H_2SO_4$ .

←

### An interpolated Model

To use an interpolator it is required to have a set of single values on an increasing x-axis series. I could average the values where there are duplicates at a given value of  $m$ . I decided to use the Johnson and Gillespie data sets to cover the range from 2 to 101  $\%H_2SO_4$ . The data is precise and closely spaced and so the interpolation is a good representation of the data. Figure 14 on the following page shows the result.

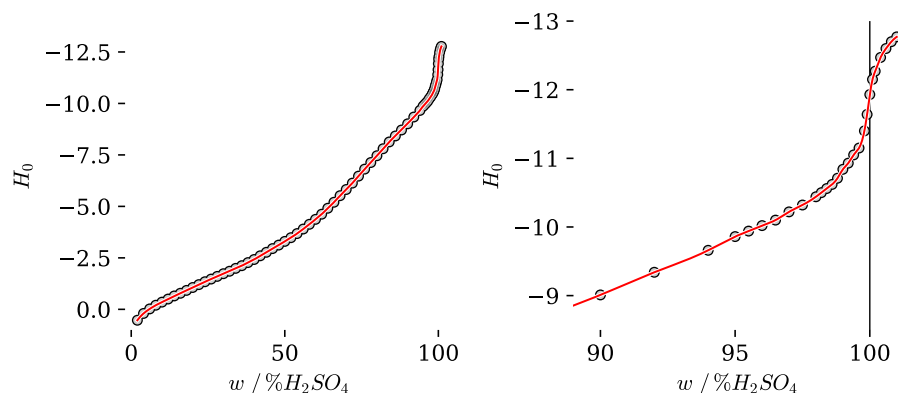


Figure 14: Plots  $H_0$  vs  $m$  (% $\text{H}_2\text{SO}_4$ ) for the data of Gillespie and Johnson combined. Where duplicate values existed, the data from Gillespie was used. The spline interpolation is presented on the plots. The right-hand plot is a detail region. This combined data set is precise and closely spaced, so an interpolation is likely the best way to represent it.

### Where Are We Now?

What a long strange journey it has been. I began this exercise seeking a better way to interpolate the literature data sets for  $\rho$ ,  $a_{\text{H}_2\text{O}}$  &  $H_0$  of sulfuric acid mixtures. I was unhappy with the results of interpolation that I was using previously and wanted to try polynomial models for this interpolation. In the end I am all the way back to recommending a simple interpolation of the points in the data table as the solution for most cases. In the case of  $a_{\text{H}_2\text{O}}$  values, where there is an enormous difference in magnitude between the final data points at high concentrations, the polynomial fit might be best. However, it was an complicated piecewise model requiring extreme precision in the parameters. The simple interpolation was good enough.

The reason that my interpolation for  $a_{\text{H}_2\text{O}}$  was taking an excursion into negative values and causing errors in my calculations was because I had previously been using a smoothing interpolation function (useful for experimental data that contains random error). This did a little too much smoothing. I used a more traditional Bspline interpolator function, `SciPy.interpolate.make_interp_spline`, and it produced a better result. When someone says they interpolated their data, the exact method matters. You can see exactly what I have done by examining my *Python* code for yourself.<sup>37</sup>

<sup>37</sup> Using *Python* notebooks will enable you to “show your work” for any problem that you are solving. No need for long winded documents like this one. See reference 17 on page 3

### 5. Comparing Our $a_{\text{H}_2\text{O}}$ Against Cox

The analysis that I have been using for the results of Yates and McClelland<sup>9</sup> are based on a method called the “excess acidity method.” To use this method, we will need good values for  $H_0$  and  $a_{\text{H}_2\text{O}}$ . Robin Cox, a longtime collaborator of Yates, published a review of the excess acidity method<sup>7</sup> in the year 2000 and, in that paper, produced tables of excess acidity values (related to  $H_0$  values) and  $a_{\text{H}_2\text{O}}$  values. I want to compare my values with those in the table by Cox.

### Converting Units

The tables of  $a_{\text{H}_2\text{O}}$  from Giaque<sup>4</sup> and others are presented as fractional vapour pressures ( $p_x/p + 0$ ), but the values reported by Cox are in units of apparent molar concentration. I will need to obtain an activity coefficients from the value of  $a_{\text{H}_2\text{O}}$  and the formal mole fraction of  $\text{H}_2\text{O}$  and then multiply that against the formal molar concentration of  $\text{H}_2\text{O}$ . Will I get the same values?

I extracted activity coefficients for water in  $\text{H}_2\text{SO}_4$  ( $\gamma_w$ ) by using the formal mole fraction of water in the mixture ( $x_w$ ) and the apparent mole fraction (fractional vapour pressure,  $p_x/p_0$  or  $a_{\text{H}_2\text{O}}$ ).

$$\gamma_w = \frac{a_{\text{H}_2\text{O}}}{x_w}$$

I then calculated the formal molar concentration of water in the mixture using the interpolated density tables to convert  $w$  (% $\text{H}_2\text{SO}_4$ ) to  $M$  ( $\text{mol L}^{-1}$ ). The product of the activity coefficient and molar concentration will give the activity of water in units of molarity.

$$a_{\text{H}_2\text{O}}^M = \gamma_w [\text{H}_2\text{O}]$$

I converted the data set that I had assembled from Giaque<sup>4</sup> and Rard<sup>5</sup> (see a plot of that data set in Figure 11 on page 16) so that the values of  $a_{\text{H}_2\text{O}}$  were in units of molarity. I designated this set of activity values as  $a_{\text{H}_2\text{O}}^M$ .

### Comparing Cox and Giaque

A plot of  $a_{\text{H}_2\text{O}}^M$  vs  $w$  for both the values that I had just converted and the values reported by Cox<sup>7</sup> is presented in Figure 15.

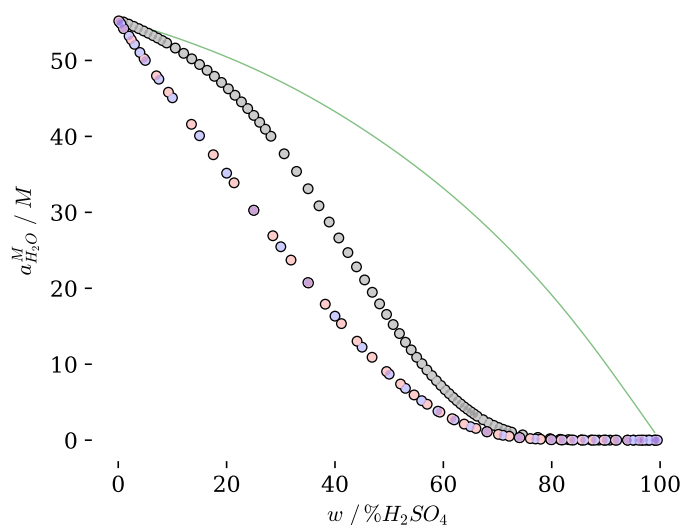


Figure 15: Plots of  $a_{\text{H}_2\text{O}}^M$  ( $\text{mol L}^{-1}$ ) vs  $w$  (% $\text{H}_2\text{SO}_4$ ) for the data of Giaque and Rard combined (black) along with the data set presented by Cox (red, Table 3 in the paper; blue, Table 4). The green line is the calculated formal concentration of  $\text{H}_2\text{O}$  in the mixture.

←

We can clearly see that the values for  $a_{\text{H}_2\text{O}}^M$  that were used by Cox are different than the ones derived from Giaque and Rard. The values above 70 %H<sub>2</sub>SO<sub>4</sub> are not able to be separated from each other. Plotting using a log scale will be more useful. The plot of  $\log a_{\text{H}_2\text{O}}^M$  vs  $w$  is presented in Figure 16. I observe that the difference is less than an order of magnitude throughout most of the range, with the deviation being about 0.3 to 0.4 log units (the values for  $a_{\text{H}_2\text{O}}^M$  used by Cox are about one half of those from Giaque and Rard). This is not really a big difference in the world of physical organic chemistry, but I still need to explain it.

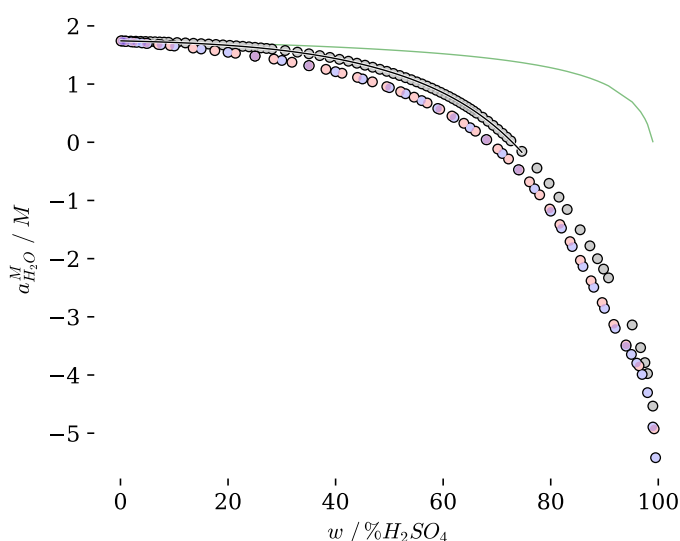


Figure 16: Plots of  $\log a_{\text{H}_2\text{O}}^M$  (mol L<sup>-1</sup>) vs  $w$  (%H<sub>2</sub>SO<sub>4</sub>) for the data of Giaque and Rard combined (black) along with the data set presented by Cox (red, Table 3 in the paper; blue, Table 4). The green line is the calculated formal concentration of H<sub>2</sub>O in the mixture. The black line is data calculated from osmotic activities reported by Zeleznick.

### Investigating the Data Sources

The review by Cox<sup>7</sup> was written in the year 2000. The data that we will be considering that was presented by Yates and McClellan<sup>9</sup> is from 1967 and the values they used for  $a_{\text{H}_2\text{O}}$  were from Giaque<sup>4</sup> in 1960. Cox had access to more recent data for values of  $a_{\text{H}_2\text{O}}$ . He used a data set presented by Frank Zeleznick<sup>38</sup> in 1991. The data presented in Table 8 of that paper reported the osmotic coefficient vs molal concentration for sulphuric acid/water mixtures up to 30 mol kg<sup>-1</sup> (75 %H<sub>2</sub>SO<sub>4</sub>). Osmotic coefficients are a common way to determine values for  $a_{\text{H}_2\text{O}}$  and I converted these values to units of activity as partial pressure using the standard equation<sup>39</sup> displayed below as eq. 5.

$$\phi = - \left( \frac{1000}{M_W v m} \right) \ln a_{\text{H}_2\text{O}} \quad (5)$$

Where the osmotic coefficient,  $\phi$  is defined in terms of  $a_{\text{H}_2\text{O}}$  ( $p_x/p_0$ ).  $M_W$  is the molar mass of water, 18.015 g mol<sup>-1</sup>;  $v$  is the number of ions that sulfuric

<sup>38</sup> "Thermodynamic Properties of the Aqueous Sulfuric Acid System to 350 K." F.J. Zeleznick, *J. Phys. Chem. Ref. Data* 1991 20, 1157–1200. <https://doi.org/10.1063/1.555899>. Obtained via inter-library loan. You can also access the paper via <https://www.nist.gov/system/files/documents/srd/jpcrd426.pdf>. I eventually found a better copy at <https://srd.nist.gov/jpcrdreprint/1.555899.pdf>. Both of those are US government sources, so I presume that they are freely available.

<sup>39</sup> "Thermodynamic Properties of Aqueous Sulfuric Acid." H. Sippola, P. Taskinen, *J. Chem. Eng. Data*, 2014, 59, 2389–2407. <https://doi.org/10.1021/jc4011147>

acid can produce on complete dissociation ( $\nu = 3$  in this case); and  $m$  is the concentration of  $\text{H}_2\text{SO}_4$  in molal units.

I plotted values of  $a_{\text{H}_2\text{O}}^M$  calculated from the osmotic coefficient data of Zeleznick. It is presented as the black line in Figure 16 on the previous page and I observed that it tracked perfectly with the data of Giauque and Rard and not with the values reported by Cox.

### *Digging Deeper*

So where did Cox get his data for  $a_{\text{H}_2\text{O}}^M$ ? The Zeleznick paper reports the osmotic coefficients only to 75 %  $\text{H}_2\text{SO}_4$  and the data table of Cox goes to 99 %. Cox did not report  $a_{\text{H}_2\text{O}}^M$  calculated from the osmotic coefficient data table, but used the parameterized curve fit model described in the paper. There is no model for osmotic coefficients in the paper so I had to read it carefully and not just look at the pictures and the data tables.

The model that is proposed by Zeleznick is completely inscrutable to me. The photocopy of the paper that I obtained via interlibrary loan had about half an inch cut off along the text on the sides of each page. Some terms in the equations are missing. Even if I wanted to get into the weeds with this model, it has 50 parameters and uses a complex algorithm with three summation loops. If someone can get me a good copy of page 1185 and help me parse through equations 11, 12 and 13 of the Zeleznick paper<sup>38</sup>, please contact me.<sup>40</sup> I tried following the trail of references to the papers where Zeleznick first presented these equations<sup>41,42</sup> but these were intense works of mathematics and completely defeated me.

### *Using the Pictures*

Zeleznick used his model to calculate the relative chemical potential for water in sulphuric acid,  $\mu_1 - \mu_1^*$  where  $\mu_1$  is the chemical potential for a water-sulphuric acid mixture and  $\mu_1^*$  is the chemical potential for the standard state of water. From this value, the osmotic coefficient,  $\phi$ , was calculated according to Eq. 6 and then the value of  $a_{\text{H}_2\text{O}}$  can be calculated using Eq. 5 on the previous page.

$$\frac{\mu_1 - \mu_1^*}{RT} = 3 \frac{x_1}{x_2} \phi \quad (6)$$

Zeleznick presents values of osmotic coefficients,  $\phi$ , that were calculated using his model in Table 8 of his paper. But Cox must have used his model and calculated values of  $\mu_1 - \mu_1^*$  and then converted these to values of  $a_{\text{H}_2\text{O}}^M$ . I have tried using the equations for a few data points and do not get anything near what is reported by Zeleznick. I'm doing it wrong and don't know how to do it right. I see in Figure 25 of the paper that Zeleznick presents a graph of  $-\mu_1 - \mu_1^*/RT$  values for sulphuric acid and water. I have reproduced this graph in Figure 17 on the following page. Zeleznick presents data calculated to

<sup>40</sup> I eventually found a better copy of the paper and could follow the equation now, but I think I still need help.

<sup>41</sup> "A class of nonideal solutions. 1: Definition and properties." F.J. Zeleznick, NASA TP-1929, 1983. <https://ntrs.nasa.gov/citations/19830014918>

<sup>42</sup> "A class of nonideal solutions. 2: Application to experimental data" F.J. Zeleznick, L.F. Donovan, NASA TP-1930, 1983. <https://ntrs.nasa.gov/citations/19830014917>



99.5 %. The data in his Table 8 gives me data only to a mole fraction of 0.35 (75 %).

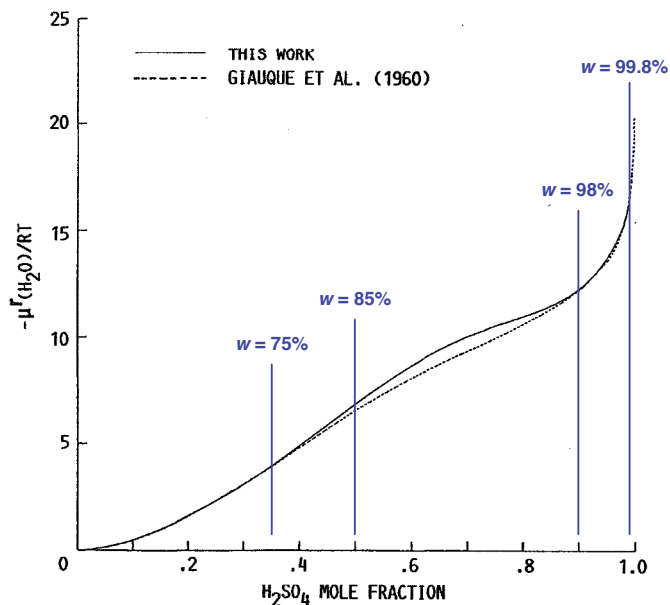


Figure 17: Plots of  $-\mu_1 - \mu_1^0/RT$  vs mole fraction  $\text{H}_2\text{SO}_4$  for values calculated from the models of Zeleznick and Giauque. I have added indications for values of concentration in units of  $w$ , % $\text{H}_2\text{SO}_4$ , for comparison sake. I observe that the values are nearly identical to those of Giauque until we exceed a mole fraction of 0.5 (85 % $\text{H}_2\text{SO}_4$ ).

I will use a plot digitizer<sup>43</sup> to extract a set of points from the line drawn by Zeleznick for his own calculated results. I can compare my extracted results to those derived from his table of osmotic coefficients. The data extracted is values for  $-\mu_1 - \mu_1^0/RT$ . Using Eq. 6 as reported by Zeleznick did not give correct values for  $\phi$ . My physical chemistry is weak, but I do know that there is no need to convert chemical potential to osmotic coefficients and then to activity. Activity can be calculated directly from the chemical potential.<sup>44</sup> I will use the formula for calculating activity from the difference between the chemical potential in the standard state and that observed in the given system,<sup>45</sup> as shown in Eq. 7.

$$a_x = e^{\frac{\mu_x - \mu_x^0}{RT}} \quad (7)$$

I now have a set of data for chemical potentials of water as calculated from the models of Giauque and of Zeleznick according to Figure 25B in Zeleznick's paper.<sup>38</sup> I converted these chemical potential values to activity and then to molar activity for water and plotted the data in Figure 18 on the next page. As you can see, the data tracked very well with the points that I had derived from Giauque's data tables previously. The potentials calculated from Zeleznick's model and presented in Figure 25B of the paper agreed almost perfectly with the Giauque data except in the region above 85 % $\text{H}_2\text{SO}_4$ .

Zeleznick's discussion about the difference between his model and the model of Giauque expresses some frustration with a lack of information

<sup>43</sup> I used WebPlotDigitizer by Ankit Rohatgi. <https://automeris.io/v4/>

<sup>44</sup> Recall that chemical potential of a reactant is actually a measure of the activity of a component as expressed by the activity coefficient and concentration,  $a_x = \gamma_x \cdot c_x$

<sup>45</sup> 'activity' in *IUPAC Compendium of Chemical Terminology*, 5<sup>th</sup> ed. International Union of Pure and Applied Chemistry, 2025. Online version 5.0.0, 2025. <https://doi.org/10.1351/goldbook.A00115>

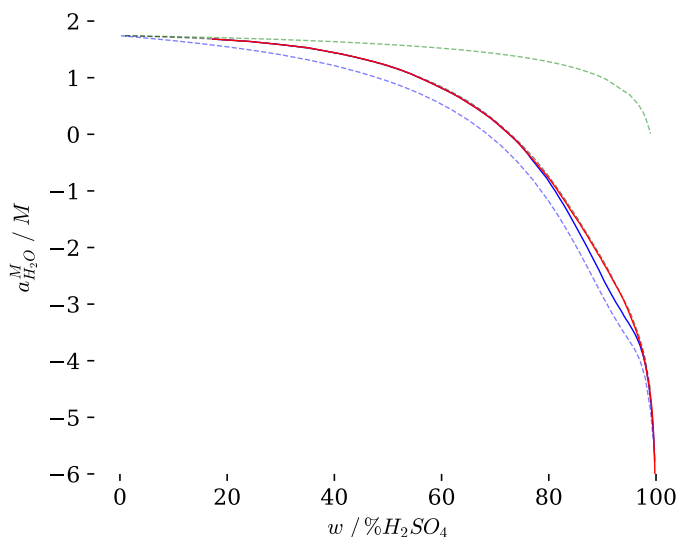


Figure 18: Plots of  $\log a_{\text{H}_2\text{O}}^M$  ( $\text{mol L}^{-1}$ ) vs  $m$  ( $\%\text{H}_2\text{SO}_4$ ) derived from Figure 25B in Zeleznick.<sup>38</sup>

←

Red solid line: the digitized data from Giaugue's model as presented in Zeleznick's figure.

Blue solid line: the digitized data from Zeleznick's model.

Green dashed line: The green line is the calculated formal concentration of  $\text{H}_2\text{O}$  in the mixture.

Black dashed line: The data of Giaugue and Rard combined together (this line is mostly obscured by the red line).

Blue dashed line: the data set presented by Cox in Table 3 of his paper<sup>7</sup>.

about the exact methodology and math in that model. Zeleznick proposes some possible reasons for the difference but cannot be sure as the math was not documented in detail. If only *Python* notebooks had been available in 1960, alas.

I still cannot explain the difference between my calculated molar activity values for water and those of Cox. It is clear that he used the model of Zeleznick (observe how the Cox data set is nearly parallel with the Zeleznick model in the region above 85 %) but there is a significant difference across the whole data set. I cannot explain this difference. I have no information about the math that was used by Cox but I suspect that the difference lies there. Cox states that he calculated the molar concentration of water using a standard equation found in Robinson and Stokes.<sup>46</sup> This equation is the same one that I used. You can check my math, but we cannot check the math in the paper of Cox.<sup>47</sup>

I am now more confident in my own calculations of values of  $a_{\text{H}_2\text{O}}^M$  in molar units for use in my analysis of literature data for ester hydrolysis in sulphuric acid mixtures. I will use the data set that I extracted from Figure 25B of the Zeleznick paper and combine that with the data set calculated from the osmotic potential data in table 8 of the same paper. This will give me a set of values for  $a_{\text{H}_2\text{O}}$  from near zero to 99.5  $\%\text{H}_2\text{SO}_4$ . I am confident of the accuracy of this scanned data because of the very close agreement between the scanned data for Giaugue and the data reported in the original paper of Giaugue.

The new data for  $a_{\text{H}_2\text{O}}^M$  from 1991 will differ significantly from that used by Yates and others in the 1960's and 70's. Will it make a difference in the interpretation of the data? Probably not. The region of interest in most of the data is below 85  $\%\text{H}_2\text{SO}_4$ . But let us proceed and see what we get.

<sup>46</sup> "Electrolyte Solutions – The Measurement and Interpretation of Conductance, Chemical Potential and Diffusion in Solutions of Simple Electrolytes. 2<sup>nd</sup> ed." R. A. Robinson and R. H. Stokes, 1970, *Dover Publications, Mineola, NY*, reprint edition, 2002. Page 31.

<sup>47</sup> The *Python* notebook for all following calculations and plots involving the data of Cox can be accessed via Google Colab at [https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class\\_30\\_Yates\\_New/cox.ipynb](https://colab.research.google.com/github/blinkletter/4410PythonNotebooks/blob/main/Class_30_Yates_New/cox.ipynb).

## Conclusion

This was a fun exercise in data archeology. I have accomplished all my goals and now express the following conclusions.

1. I am now confident in my mathematical method for converting between molal, molar and %mass expressions of concentrations. I have written *Python* functions for all of these and documented the methods.<sup>17,47</sup>
2. To accomplish the first goal I had to create models that predicted density of  $\text{H}_2\text{SO}_4$  mixtures. I developed polynomial models that gave a good approximation. In the end, the best approach was a direct interpolation of the data set using a spline interpolation function available in the *SciPy* library of *Python*.
3. I investigated literature data sets for values of  $a_{\text{H}_2\text{O}}$  and decided to combine the data sets of Rard<sup>5</sup> and Giauque.<sup>4</sup> I then investigated several approaches to making a parameterized polynomial model, including a piecewise model. Again, a simple spline interpolation applied to the combined data set was deemed sufficient.
4. I compared several literature data sets for  $H_0$  values. I investigated polynomial curve fits to combinations of these data sets. In the end I chose to use a model based on a simple spline interpolation applied to the combined data set of Johnson<sup>32</sup> and Gillespie.<sup>31</sup>
5. I derived a method to convert the partial pressure version of  $a_{\text{H}_2\text{O}}$  to a molar concentration version,  $a_{\text{H}_2\text{O}}^M$ . I plotted the values of  $a_{\text{H}_2\text{O}}^M$  vs % $\text{H}_2\text{SO}_4$  for the data of Giauque & Rard and also for the data of Cox<sup>7</sup>. They were significantly different over the entire concentration range. I obtained the data used by Cox and showed that it gave the same results as my own calculations except where Zeleznick<sup>38</sup> and Giauque disagreed above 85 % $\text{H}_2\text{SO}_4$ . I cannot explain the differences between the data presented by Cox for  $a_{\text{H}_2\text{O}}^M$  and my own calculations based on the data of Zeleznick. I decided to construct a data set for  $a_{\text{H}_2\text{O}}$  values based on tabulated and digitized data found in the paper of Zeleznick. I will use that data set to construct interpolation models.

I will someday revisit the math of Zeleznick's model and attempt to produce values for  $a_{\text{H}_2\text{O}}$  derived directly from the math rather than picking points from Zeleznick's own presentation of the data in a plot.