Determination of Antoine Equation Parameters

December 4, 2012
PreFEED Corporation
Yoshio Kumagae

Introduction

Physical property data is extremely important for performing process design and plant data analysis. Physical properties include equilibrium physical properties (pure substance vapor pressure, vapor-liquid equilibrium, specific heat, heat of evaporation, etc.) and transport properties (viscosity, thermal conductivity, diffusion coefficients). Among these, the pure substance vapor pressure is significantly important as it is the basis for estimating other physical property values.

Here we will introduce a parameter determination method for the Antoine equation which is often used for pure substance vapor pressure calculations.

Related materials: "On the B Value of the Antoine Equation" (Tips #0811)

PreFEED

Simple Determination Methods of Antoine Constants

PreFEED

Vapor Pressure Correlation Equation - Antoine Equation

$$3^{rd}\ \text{ed.}: \log_{10} P[\text{mmHg}] = A - \frac{B}{C + t[^\circ\text{C}]}$$

$$5^{th}\ \text{ed.}: \log_{10} P[\text{kPa}] = A_5 - \frac{B_5}{C_5 + t[^\circ\text{C}]}$$

$$A = A_5 + 0.87510, B = B_5, C = C_5$$

$$6^{th}\ \text{ed.}: \ln P[\text{Pa}] = A_6 - \frac{B_6}{C_6 + T[\text{K}]}$$

$$A = 0.43429 \times A_6 - 2.1249,$$
$$B = 0.43429 \times B_6, C = 273.15 + C_6$$

A, *B*, and *C* in the Antoine equation are called Antoine constants (parameters), and their values differ depending on the unit system that is used. In the case of the "Kagaku Kougaku Binran" (*in Japanese* - Chemical Engineering Handbook), the format and units of the equation differ according to the publication year as shown on the left.

For reference, the conversion of constants is also shown (denoted *A*, *B*, *C* for the third edition, A_5, B_5, C_5 for the fifth edition, and A_6, B_6, C_6 for the sixth edition).

In the explanations that follow, the notation of the third edition will be used.

Antoine Equation Constants for Water

Kagaku Kougaku Binran	A	B	C
3rd ed.	7.8097	1572.53	219
5th ed.	7.07406	1657.46	227.02
6th ed.	23.1964	3816.44	-46.13

$$\log_{10} P = A - \frac{B}{t + C}$$

The easiest way to determine the Antoine constants *A*, *B*, *C* from experimental data is to linearly approximate the equation so that a linear plot can be made in a graph.

Linear Approximation 1: Clausius-Clapeyron Equation
By approximating $C = 273.15$, $t+C$ can be replaced with the absolute temperature T.

$$\log_{10} P = A - \frac{B}{T}$$

By plotting $\log_{10}(P)$ against $1/T$, A and B can be determined from the intercept and slope of the straight line.

Linear Approximation 2: Cox Equation
Approximating C=230 gives a good approximation for many substances.

$$\log_{10} P = A - \frac{B}{t + 230}$$

By plotting $\log_{10}(P)$ against $1/(t+230)$, A and B can be determined from the intercept and slope of the straight line.

Although the above equations are simple, they are not sufficiently accurate for distillation calculations and the like. It is necessary to determine the Antoine constants including the C value.

Difficulty of Determining Antoine Constants

Parameter Determination Method by Excel Solver

Raw Data

t, C	p, torr
-16.2	10
-6.3	20
11.4	60
34.03	200
64.51	760

Antoine Equation

$$\log_{10} P = A - \frac{B}{t + C}$$

$$P[\text{mmHg}],\ t[^\circ\text{C}]$$

Error Definition

$$Q = \sum [\log P_{i\,(\text{Measured Value})} - \log P^{cal}_{i\,(\text{Calculated Value})}]^2$$

$$= \sum [\log P_i - (A - \frac{B}{t_i + C})]^2$$

The values of A, B and C that minimize the error are determined with the Excel solver.

Excel Solver Results

	Initial Value	After Optimization	t, C	Measured Value p, torr	Calculated Value pcal, torr
A=	8	8.8444	-16.2	10	10.2
B=	2000	2002.2486	-6.3	20	19.9
C=	300	271.6758	11.4	60	59.0
			34.03	200	197.1
			64.51	760	773.7

	Initial Value	After Optimization	t, C	Measured Value p, torr	Calculated Value pcal, torr
A=	10	7.9198	-16.2	10	10.0
B=	1500	1497.6571	-6.3	20	20.0
C=	200	232.5704	11.4	60	60.4
			34.03	200	200.5
			64.51	760	756.1

	Initial Value	After Optimization	t, C	Measured Value p, torr	Calculated Value pcal, torr
A=	10	1.9909	-16.2	10	98.0
B=	10	0.0017	-6.3	20	19.5
C=	10	6.3024	11.4	60	97.9
			34.03	200	97.9
			64.51	760	97.9

Convergence results are different for each set of initial values of A, B, C. The standard boiling point is the most important data, but it is greatly affected by the initial value. The reason for this is that the C value is within the denominator of the fraction and is a nonlinear parameter.

PreFEED

Improvement of Determination of Antoine Constants
Part 1 Nonlinear Least Squares Method

PreFEED

Nonlinear Least Squares Method using Normal Equation

$$Q = \sum [\log P_i - \log P_i^{cal}]^2 = \sum [\log P_i - (A - \frac{B}{t_i + C})]^2$$

The following condition holds for the values of *A*, *B*, *C* which minimize *Q*.
This is called a normal equation.

$$\frac{\partial Q}{\partial A} = 0, \frac{\partial Q}{\partial B} = 0, \frac{\partial Q}{\partial C} = 0$$

As can be seen from the form of the expression of *Q*, a linear equation is obtained when it is partially differentiated by *A* and *B*.

Solutions for R&D to Design

Derivation of Normal Equation

$$Q=\sum[\log P_i-\log P_i^{cal}]^2=\sum[\log P_i-(A-\frac{B}{t_i+C})]^2=\sum[(\log P_i)^2-2\log P_i(A-\frac{B}{t_i+C})+(A-\frac{B}{t_i+C})^2]$$

$$\frac{\partial Q}{\partial A}=\sum[-2\log P_i+2(A-\frac{B}{t_i+C})]=2[-\sum(\log P_i)+A\sum(1)-B\sum(\frac{1}{t_i+C})]=0$$

$$\therefore -\sum(\log P_i)+An-B\sum(\frac{1}{t_i+C})=0...(1)$$

$$\frac{\partial Q}{\partial B}=\sum[\frac{2\log P_i}{t_i+C}+2(A-\frac{B}{t_i+C})(\frac{-1}{t_i+C})]=2[\sum(\frac{\log P_i}{t_i+C})-A\sum(\frac{1}{t_i+C})+B\sum(\frac{1}{t_i+C})^2]=0$$

$$\therefore \sum(\frac{\log P_i}{t_i+C})-A\sum(\frac{1}{t_i+C})+B\sum(\frac{1}{t_i+C})^2=0...(2)$$

From Eq.(1) and (2)

$$B=\frac{\sum\log P_i\cdot\sum(\frac{1}{t_i+C})-n\sum(\frac{\log P_i}{t_i+C})}{n\sum(\frac{1}{t_i+C})^2-[\sum(\frac{1}{t_i+C})]^2},\quad A=\frac{\sum\log P+B\sum(\frac{1}{t_i+C})}{n}$$

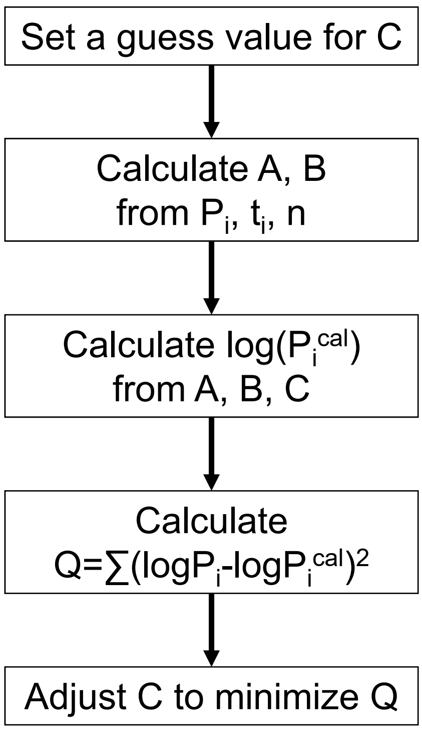
A and *B* can be calculated from the measured values (t_i, P_i) by guessing a value for *C*. In other words, there is no need to independently search for *A*, *B*, *C*, and it is possible to search for *C* only.

Algorithm

$$B=\frac{\sum\log P_i\cdot\sum(\frac{1}{t_i+C})-n\sum(\frac{\log P_i}{t_i+C})}{n\sum(\frac{1}{t_i+C})^2-[\sum(\frac{1}{t_i+C})]^2}$$

$$A=\frac{\sum\log P_i+B\sum(\frac{1}{t_i+C})}{n}$$

With this method, it is not necessary to change three parameters. The problem becomes a single variable search.



Optimization Results

A=	8.11118	Equation for calculating A, B from C
B=	1596.03	
C=	240.644	Variable to change

t	p	logP	1/(t+C)	logP/(t+C)	1/(t+C)^2	log10(pcal) =A-B/(C+t)	square of error (log base)	pcal
-16.2	10	1.000	0.00446	0.00446	1.99E-05	1.0001	2.121E-08	10.003
-6.3	20	1.301	0.00427	0.00555	1.82E-05	1.3006	2.254E-07	19.978
11.4	60	1.778	0.00397	0.00705	1.57E-05	1.7788	4.699E-07	60.095
34.03	200	2.301	0.00364	0.00838	1.33E-05	2.3005	2.311E-07	199.779
64.51	760	2.881	0.00328	0.00944	1.07E-05	2.8809	1.549E-08	760.218
		9.261	0.01961	0.03488	7.78E-05	Q=	9.631E-07	

It is extremely difficult to obtain optimal values simply by applying a solver to *A*, *B*, and *C*, but it is possible to obtain relatively easily the true optimal values by using the normal equation and setting only the search variable *C*.

Improvement of Determination of Antoine Constants Part 2 Multivariate Linear Equation Coefficient Regression Method

There is a multivariate linear equation regression function among the Excel Add-in : Data Analysis - Regression functions.

$$y = a_0 + a_1 x_1 + a_2 x_2 + .. + a_n x_n$$

a_0, a_1, a_2, a_n: Constants
x_1, x_2, x_n, y: Data

It is possible to determine linear constants by using this function.

Transformation into Multivariate Linear Equation

Format 1*

$$\log_{10} P = A - \frac{B}{t+C}$$

$$(t+C)\cdot\log_{10} P = A(t+C) - B$$

$$t\cdot\log_{10} P + C\cdot\log_{10} P = At + AC - B$$

$$t\cdot\log_{10} P = At + AC - B - C\cdot\log_{10} P$$

$$= (At) + (AC - B) - (C\cdot\log_{10} P)$$

$$\log_{10} P = A + (AC - B)\frac{1}{t} + (-C)\frac{\log_{10} P}{t}$$

Format 2

$$\log_{10} P = A - \frac{B}{t+C}$$

$$(t+C)\cdot\log_{10} P = A(t+C) - B$$

$$t\cdot\log_{10} P + C\cdot\log_{10} P = At + AC - B$$

$$C\cdot\log_{10} P = At + AC - B - t\cdot\log_{10} P$$

$$= (At) + (AC - B) - (t\cdot\log_{10} P)$$

$$\log_{10} P = \frac{AC - B}{C} + \frac{A}{C}t + (-\frac{1}{C})(t\cdot\log_{10} P)$$

It is possible to determine the coefficients since both equations are linear expressions with two variables of the following type:

$$y = a_0 + a_1 x_1 + a_2 x_2$$

*) As indicated, for instance, in "Problem Solving in Chemical and Biochemical Engineering with Polymath, Excel and MATLAB" (Cutlip and Shacham, Prentice Hall p37)

Comparison of Multivariate Linear Regression Results with Normal Equation Method

Measured Value		Two Variable Linearization 1			Two Variable Linearization 2		
t, C	p, torr	log10(p)	1/t	log10(p)/t	log10(p)	t, C	t*log10(p)
-16.2	10	1.000	-0.0617284	-0.0617284	1.000	-16.2	-16.2
-6.3	20	1.301	-0.1587302	-0.2065127	1.301	-6.3	-8.196489
11.4	60	1.778	0.0877193	0.1559782	1.778	11.4	20.270924
34.03	200	2.301	0.0293858	0.0676177	2.301	34.03	78.304051
64.51	760	2.881	0.0155015	0.0446569	2.881	64.51	185.84128

		Coefficient
"A"	Intercept	8.12814663
"AC-B"	1/t	357.156565
"-C"	log10(p)/t	-241.50139
A=	8.12814663	
B=	1605.80217	
C=	241.501393	

		Coefficient
(AC-B)/C	Intercept	1.47886856
A/C	t	0.03370968
"-1/C"	t*log10(p)	-0.004157
A=	8.10920993	
B=	1594.99674	
C=	240.560274	

t	Measured Value	Nonlinear Optimization	Linearization 1	Linearization 2
-16.2	10	10.003	10.018	10.003
-6.3	20	19.978	19.989	19.978
11.4	60	60.095	60.066	60.098
34.03	200	199.779	199.585	199.787
64.51	760	760.218	759.666	760.181

By recalculating, it can be seen that all methods converge with sufficient accuracy.

Conclusion

- When attempting to perform a parameter regression without rearranging the Antoine equation, it is difficult to obtain the true convergence values by a simple three variable regression (by using the Excel solver) because the C parameter is nonlinear. In this case, it is considerably easier to obtain the true convergence values by using the normal equation and optimizing only the C parameter.
- By transforming the Antoine equation into a multivariate linear equation, it is possible to obtain the true convergence values by performing a multivariate linear regression (by using the data analysis function of Excel).