JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Laboratory Experiment

Multivariable Model Fitting as Applied to Air, a Physical Chemistry Experiment

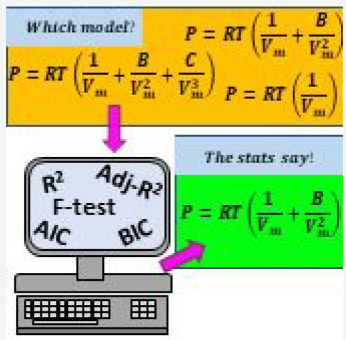
Jonathan P. Antle, Jerry T. Godbout, and Scott Simpson*

Cite This: *J. Chem. Educ.* 2022, 99, 2107−2111

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Students in an upper-level physical chemistry course utilized an open-sourced statistical software package to construct models fitted to experimental pressure−volume data. In the first part of the experiment, students familiarize themselves with model fitting. In the second part of the experiment, students determine which truncated version of the virial equation of state is statistically valid to properly model air. This experiment demonstrates proper multivariable linear regression model construction, along with the nonideality of air.



KEYWORDS: *Upper-Division Undergraduate, Physical Chemistry, Problem Solving/Decision Making, Gases, Mathematics/Symbolic Mathematics, Theoretical Chemistry, Thermodynamics*

INTRODUCTION

The virial equation of state is often investigated in most undergraduate physical chemistry courses and is introduced early on in most physical chemistry textbooks.[1−5] Several experiments and articles investigating the behavior of real gases have been published in this *Journal* and in textbooks.[6−14] Our approach to this makes the experimental setup simpler for the instructor and more cost-effective for institutions lacking resources, implements more intensive model fitting/statistical comparisons of models, and utilizes the open-sourced R statistical software package.

R is an open access language and environment that provides a wide variety of modeling and statistical techniques.[15] Statistics and model fitting is a required skill set expected of chemists in the future workforce.[16] With increasing computational power, model construction is becoming more widespread, accessible, and is more utilized by those in academia and industry. The divisions of cheminformatics, QSRR, and machine learning (ML) applied to chemistry, etc., are fields that require incoming chemists to learn modern tools to construct predictive models and adequate statistics to determine if one has arrived at a meaningful result.[17−19] However, most undergraduate chemistry programs do not cover multivariate model construction or adequate statistical knowledge leaving a gap in the skill set required for employment or research.[16,19−25] To prepare undergraduates for the future of chemistry and model construction, we present the following experiment in which students utilize statistical arguments to determine a sufficient model for the behavior of a mixture of real gases, air.

However, a brief introduction to the virial equation of state is beneficial to the reader. To these ends, just as many others have done in past derivations,[12−14] we start at the equation of state for a real gas mixture:

$$PV = n_{\text{mixture}}RTZ_{\text{mixture}} \tag{1}$$

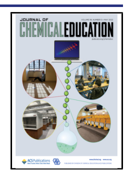
where P is the total pressure, V is the volume of the gaseous mixture, R is the ideal gas constant, n_{mixture} is the total number of gaseous particles in moles, and Z_{mixture} is the compressibility of the mixture. Equation 1 can be rewritten in terms of molar volume (V_m):

$$PV_m = RTZ_{\text{mixture}} \tag{2}$$

where $V_m = V/n_{\text{mixture}}$. Depending on the application or preference of the investigator, Z_{mixture} has two forms of the virial expansion: One is in terms of molar volume:

$$Z_{\text{mixture}} = \frac{PV_m}{RT} = 1 + \frac{B_{\text{mixture}}}{V_m} + \frac{C_{\text{mixture}}}{V_m^2} + ... \tag{3}$$

Received: December 17, 2021
Revised: April 1, 2022
Published: April 20, 2022



https://doi.org/10.1021/acs.jchemed.1c01241
J. Chem. Educ. 2022, 99, 2107−2111

Table 1. Assessment Questions and Preassessment/Postassessment Responses

<table>
<tr><th rowspan="2">Assessment Question</th><th colspan="2">% Correct Student Responses (N=12)</th></tr>
<tr><th>Pre-assessment</th><th>Post-assessment</th></tr>
<tr><td>1. Write a general formula for the Virial Equation of State.</td><td>0.0 %</td><td>100.0 %</td></tr>
<tr><td>2. Given the following data, which model should be utilized?
<table>
<tr><th>Model</th><th>AIC</th><th>BIC</th></tr>
<tr><td>A</td><td>22.5</td><td>38.7</td></tr>
<tr><td>B</td><td>50.4</td><td>40.9</td></tr>
</table></td><td>41.7 %</td><td>91.7 %</td></tr>
<tr><td>3. Write an example of a multivariable linear regression.</td><td>25.0 %</td><td>66.7 %</td></tr>
<tr><td>4. Write the command required to make a single variable regression titled “Regression” in R using the dependent variable Y and independent variable X.</td><td>0.0 %</td><td>16.7 %</td></tr>
<tr><td>5. Write the command required to make a single variable regression titled “Regression” in R using the dependent variable Y and independent variables X_1 and X_2.</td><td>0.0 %</td><td>16.7 %</td></tr>
<tr><td>6. The results of a F-test show that the p-value of the F-statistic is 0.23 for two models. Assuming a α-value of 0.05. are the two models significantly different from one another?</td><td>25.0 %</td><td>91.7 %</td></tr>
<tr><td>7. Which model would be considered a better fit given the following information?
<table>
<tr><th>Model</th><th>R^2</th><th>Adjusted R^2</th></tr>
<tr><td>A</td><td>0.978</td><td>0.654</td></tr>
<tr><td>B</td><td>0.921</td><td>0.889</td></tr>
</table></td><td>58.3%</td><td>100.0 %</td></tr>
</table>

where B_{mixture} is the second virial coefficient for the mixture and C_{mixture} is the third virial coefficient.

The other is in terms of pressure:

$$Z_{\text{mixture}} = \frac{PV_{\text{m}}}{RT} = 1 + B'_{\text{mixture}}P + B'_{\text{mixture}}P^2 + \ldots \tag{4}$$

where B'_{mixture} is the second virial coefficient for the mixture and C'_{mixture} is the third virial coefficient. The values of these coefficients change not only on the basis of composition, but also on the temperature of the mixture.

For a mixture of gases (such as air), the second virial coefficient B is an average of all possible pairs of molecules that comprise the mixture, B_{mixture}, as described by eq 5.

$$B_{\text{mixture}} = \sum_i \sum_j \chi_i \chi_j B_{ij} \tag{5}$$

Here, χ_i is the mole fraction of species i, and B_{ij} are the mixed coefficients for the pairwise interactions between species i and species j. As an example, a binary mixture is described by 2 pure-component second virial coefficients (B_{11} and B_{22}), and 1 so-called cross-term second virial coefficient (B_{12}).[14]

However, one does not need to split the second virial coefficient into its individual components to uncover meaningful results or to model the behavior of a real gas. This could be expanded upon in the future by instructors if interested as this is outside the scope of this work.

In this experiment, physical chemistry students utilized generic forms of eqs 2, 3, and 4 to determine the second and the third virial coefficients for a mixture of gases, ambient air, over the range 0−2.1 atm. The number of terms included from eq 3, and some simple manipulation, results in 3 different models investigated by students:

$$\text{model 1: } P = RT\left(\frac{1}{V_{\text{m}}}\right) \tag{6}$$

$$\text{model 2: } P = RT\left(\frac{1}{V_{\text{m}}} + \frac{B}{V_{\text{m}}^2}\right) \tag{7}$$

$$\text{model 3: } P = RT\left(\frac{1}{V_{\text{m}}} + \frac{B}{V_{\text{m}}^2} + \frac{C}{V_{\text{m}}^3}\right) \tag{8}$$

These 3 models are investigated by students for the best performance and are compared with varying statistics, such as R^2, adjust R^2, residual standard error, F-test, Akaike information criterion (AIC), and Bayesian information criterion (BIC).[26,27]

■ EXPERIMENTAL OVERVIEW

The goals of this experiment were to

(1) Have students rationalize which equation of state best models the pressure−volume relationship of a real gas utilizing statistical arguments
(2) Familiarize students with the open-sourced statistical package R
(3) Familiarize students with multivariable regressions
(4) Familiarize students with statistical information for model selection
(5) Determine values for the second virial coefficient of ambient air

This experiment was implemented in a first semester physical chemistry course for chemistry and biochemistry majors at St. Bonaventure University. This first semester physical chemistry course usually covers thermodynamics and

https://doi.org/10.1021/acs.jchemed.1c01241
J. Chem. Educ. 2022, 99, 2107−2111

kinetics. Currently, physical chemistry is taught at St. Bonaventure University every other year, meaning junior and senior chemistry/biochemistry students take the course. Analytical chemistry and instrumental analysis is also currently offered every other year at St. Bonaventure University. Therefore, approximately half the class had taken the analytical sequence and the corresponding statistics covered in those courses.

The experiment was performed in one 4 h session. Students worked in groups of 2−3 people. Additionally, a postlab presentation was required of each student group that included the group's analysis, interpretation, and results of the experiment. Each presentation was ≈15 min in length. Two weeks after the experiment was conducted, a written lab report was required for each student.

The experiment is composed of 2 parts. In part A, students utilize the Galton Height Data set to become familiar with RStudio, writing and executing scripts in RStudio, and interpreting the results produced by R. RStudio is an open-sourced integrated development environment for R.[15] This section of the experiment is completely conducted on the computer. The Notes for the Instructor (Supporting Information) contain more information on the interpretation of the input/output of R.

In part B, students utilize a pressor sensor from Vernier to collect pressure−volume data of air at the temperature of the lab. It is assumed that the temperature of the room is kept constant and is monitored periodically to confirm this. Students then utilize the knowledge and programming skills developed in part A to answer a simple question, which of the 3 previously mentioned models should be used to describe the real gas behavior of air.

Materials Required

The gas pressure sensor kit from Vernier was utilized in this experiment. This kit includes an apparatus that is easy to set up and troubleshoot and is relatively inexpensive.[28] This kit also includes a 20 mL syringe required for this experiment, that easily screws onto the pressure sensor.

The temperature of the room was monitored with a Vernier temperature probe. The relative humidity was monitored with a relative humidity sensor from Vernier.[29]

A computer with Microsoft Excel, the statistical software package R, and the R Studio user interface is also required for model construction and statistical analysis.

■ HAZARDS

Safety glasses should be worn when taking pressure−volume data. There are no chemical hazards involved with this experiment.

■ RESULTS AND DISCUSSION

Student learning of content was assessed by a pre/postassessment, a 15 minute postexperiment oral presentation given by student groups explaining the theory and procedure of the laboratory exercise, and a summative laboratory report prepared by the students. Table 1 contains the questions that were asked, along with the correct number of student responses on both the preassessment and postassessment. Each assessment question showed an increase in the number of correct responses. However, there was still poor performance in questions 5 and 6 after the experiment was conducted. These questions asked for the specific commands utilized in R Studio; most students could not recall the exact syntax. However, in the summative report assessment and in the postexperiment presentation, it was clear that students could utilize these commands to produce multivariable regression models.

Before students were permitted to conduct the experiment, each student conducted a preassessment at the start of the laboratory period, while the post assessment was conducted after the postexperiment presentation.

It should be noted that it is not only the volume indicated on the side of syringe that the gas occupies. The sensor also has an internal volume of 0.8 mL which needs to be accounted for when calculating the volume the gas occupies.[28] Additionally, the pressure sensor's specifications state that the usual pressure range for this device is 0−210 kPa. Any data outside of this range should be discarded, according to the sensor's manufacturer.

Table 2 contains values obtained by the instructor under different temperatures and relative humidity. It is clear to see that both humidity (composition) and temperature play a role in the value of B.

Table 2. Experimentally Determined Second Virial Coefficient (B) of Air along with Percent Relative Humidity (% Humidity)

% Humidity	Laboratory Temperature (K)	B (cm^3/mol)
60.57%	301.2	0.93 ± 0.13
62.92%	298.0	1.15 ± 0.18
62.08%	299.1	1.28 ± 0.11
60.92%	296.5	0.75 ± 0.09
56.31%	295.4	1.28 ± 0.06
50.68%	297.2	0.61 ± 0.10

The majority of students determined that the virial equation truncated to the second term (model 2, eq 7) was the most statistically valid model. Generally, F-tests comparing models 2 and 3 showed that there was not a significant difference between the models. Only 2 of the 12 students found that models 1 and 2 were not significantly different. Students elected to utilize the less complicated model (model 2) over model 3. They rationalized their selection based upon lower AIC and BIC values for model 2 over the others, residual standard error, and results of F-tests. R^2 and adjusted-R^2 values were often reported to be close to 1 for all models, and the students determined that theses parameters were inadequate to select one model over another. The weakness of the reliance on R^2 for model selection has been the target of several articles in this *Journal*.[30,31] The average reported value of B obtained by the students was 3 ± 2 cm^3/mol at an average 296.1 ± 0.3 K and an average relative humidity of 47 ± 1%.

As pointed out by one of the reviewers, the 95% confidence interval can be utilized to determine the uncertainty in the virial coefficients. For example, the data set provided in the SI results in a second virial coefficient of 2.32 cm^3/mol with a standard error of ±0.32 cm^3/mol. This indicates that B is well-defined as the magnitude of the standard error is small in comparison to the value of B. However, when fitted to the three-termed virial equation, the value of the third virial coefficient (C) was determined to be −0.028 cm^6/mol with a standard error of ±0.026 cm^6/mol^2. This standard error is essentially the same magnitude of C itself, indicating that the fitted value of this coefficient is highly uncertain.

https://doi.org/10.1021/acs.jchemed.1c01241
J. Chem. Educ. 2022, 99, 2107−2111

SUMMARY

We have developed an experiment measuring pressure−volume data of air to different expansions of the virial equation of state utilizing the open-source software package R for an upper-level undergraduate physical chemistry laboratory course. Students first learn how to write and execute scripts in R and then apply these skills to determine which model best describes their pressure−volume data. Our preassessment/postassessment shows growth in student understanding in topics associated with the experiment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c01241.

Sample pressure−volume data (XLSX)
Galton height data (XLSX)
Grading rubric (PDF, DOCX)
Laboratory handout (PDF, DOCX)
Notes for instructors (PDF, DOCX)
Preassessment/postassessment (PDF, DOCX)
Sample scripts (R) and Galton height script (R) (ZIP)

AUTHOR INFORMATION

Corresponding Author

Scott Simpson − *Department of Chemistry, St. Bonaventure University, St. Bonaventure, New York 14778, United States;* orcid.org/0000-0001-8093-1814; Email: ssimpson@sbu.edu

Authors

Jonathan P. Antle − *Department of Chemistry, St. Bonaventure University, St. Bonaventure, New York 14778, United States; Department of Chemistry, University at Buffalo, the State University of New York (SUNY), Buffalo, New York 14260, United States*

Jerry T. Godbout − *Science & Wellness Division, University of New Mexico—Valencia County Campus, Los Lunas, New Mexico 87031, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.1c01241

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge National Science Foundation (awards 1904825 and 2142874) for support of this research and the Donors of the American Chemical Society Petroleum Research Fund (PRF-58954-UNI5). The authors acknowledge the support of the Leo E. Keenan Faculty Endowment. Without the donation of Les Quick III, this research would have not been possible.

REFERENCES

(1) Raff, L. M. *Principles of Physical Chemistry*; Prentice Hall, 2001; pp 25−50.
(2) Atkins, P.; de Paula, J. *Physical Chemistry: Thermodynamics, Structure, and Change*, 10th ed.; Freeman, 2014; pp 45−53.
(3) Ball, D. W. *Physical Chemistry*; Thomson, 2003; pp 10−21.
(4) Engle, T.; Reid, P. *Thermodynamics, Statistical Thermodynamics, & Kinetics*, 2nd ed.; Prentice Hall, 2010; pp 155−169.
(5) Halpern, A. M.; McBane, G. C. *Experimental Physical Chemistry: A Laboratory Textbook*, 3rd ed.; W. H. Freeman: New York, 2006.
(6) Varberg, T. D.; Bendelsmith, A. L.; Kuwata, K. T. Measurement of the Compressibility Factor of Gases: A Physical Chemistry Laboratory Experiment. *J. Chem. Educ.* **2011**, *88*, 1166−1169.
(7) Baskett, W. P.; Mathews, G. P. An Experiment of Measure Gas Imperfection. *J. Chem. Educ.* **1985**, *62*, 353−355.
(8) Dannhauser, W. Determination of compressibility. An experiment for physical chemistry. *J. Chem. Educ.* **1972**, *49*, 563−564.
(9) Halpern, A. M.; McBane, G. C. *Experimental Physical Chemistry: A Laboratory Textbook*, 3rd ed.; W. H. Freeman: New York, 2006.
(10) Rybolt, T. R. A virial treatment of the Joule and Joule-Thomson coefficients. *J. Chem. Educ.* **1981**, *58*, 620−624.
(11) Wisniak, J. Interpretation of the Second Virial Coefficient. *J. Chem. Educ.* **1999**, *76*, 671−675.
(12) Privat, R.; Jaubert, J.-N.; Moine, E. Improving Students' Understanding of the Connections between the Concepts of Real-Gas Mixtures, Gas Ideal-Solutions, and Perfect-Gas Mixtures. *J. Chem. Educ.* **2016**, *93*, 2040−2045.
(13) Hayez, B. Approximate Equation to Calculate Partial Pressure in a Mixture of Real Gases. *J. Chem. Educ.* **2018**, *95*, 1982−1988.
(14) Sattar, S. Thermodynamics of Mixing Real Gases. *J. Chem. Educ.* **2000**, *77*, 1361−1365.
(15) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. URL https://www.R-project.org/.
(16) Schlotter, N. E. A Statistic Curriculum for the Undergraduate Chemistry Major. *J. Chem. Educ.* **2013**, *90*, 51−55.
(17) Lafuente, D.; Cohen, B.; Fiorini, G.; García, A. A.; Bringas, M.; Morzan, E.; Onna, D. A Gentle Introduction to Machine Learning for Chemists: An Undergradaute Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling. *J. Chem. Educ.* **2021**, *98*, 2892−2898.
(18) Thrall, E. S.; Lee, S. E.; Schrier, J.; Zhao, Y. Machine Learning for Funcational Group Identification in Vibrational Spectroscopy: A Pedagogical Lab for Undergraduate Chemistry Students. *J. Chem. Educ.* **2021**, *98*, 3269.
(19) Oberg, T. Introducing Chemometrics To Graduate Students. *J. Chem. Educ.* **2006**, *83*, 1178−1181.
(20) Antonelli, T. M.; Olivieri, A. Developing and Implementing an R Shiny Application to Introduce Multivariate Calibration to Advanced Undergraduate Students. *J. Chem. Educ.* **2020**, *97*, 1176−1180.
(21) Williamson, J. C.; Silverstein, T. P. Sweetness of Aspartame: Introducing Key Statistical Concepts in an Upper Division Bioanalytical Laboratory Project. *J. Chem. Educ.* **2021**, *98*, 1233−1241.
(22) Davidian, M.; Kutal, C. Collaboration to Meet the Statistical Needs in the Chemistry Curriculum. *J. Chem. Educ.* **2014**, *91*, 12.
(23) McCartney, M. Stats for Scientists. *Science* **2013**, *339*, 629.
(24) Davidian, M.; Louis, T. A. Why Statistics? *Science* **2012**, *336*, 12.
(25) Holme, T. A. Can Today's Chemistry Curriculum Actually Produce Tomorrow's Adaptable Chemist? *J. Chem. Educ.* **2019**, *96*, 611−612.
(26) Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19* (6), 716−723.
(27) Stone, M. Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)* **1979**, *41* (2), 276−278.
(28) Gas Pressure Sensor. https://www.vernier.com/product/gas-pressure-sensor/ (accessed Oct 2021).
(29) Relative Humidity Sensor. https://www.vernier.com/product/relative-humidity-sensor/ (accessed Oct 2021).
(30) Zielinski, T. J.; Allendoerfer, R. D. Least squares fitting of nonlinear data in the undergraduate laboratory. *J. Chem. Educ.* **1997**, *74*, 1001.

(31) Davis, W. H.; Pryor, W. A. Measures of goodness of fit in linear free energy relationships. *J. Chem. Educ.* **1976**, *53*, 285.

https://doi.org/10.1021/acs.jchemed.1c01241
J. Chem. Educ. 2022, 99, 2107−2111