

Multivariable Contrived Experiments

Contrived experiments are run to reveal the physical laws that control the observed phenomena. Sometimes we may know, or think we know, the functional form of the law; in other cases we don't, but we assume that there is indeed some law that controls the events. We then approximate the form of this law over a small range with an empirical relationship. The empiricism comprises an equation or equations that state how the observed phenomena will vary depending upon physically measurable and controllable variables and unknown parameters. A very complete law (theory), such as the ideal gas law, may give the values of parameters (e.g., the gas "constant"), and our experiment could then have the purpose of testing the theory. With the gas absorption experiment, the observation is the mass-transfer coefficients, $K_g a$, while the independent variables are, at least, caustic concentration, flow rates, CO_2 concentration, and temperature. There could be other variables, but we attempt to hold these as constant as possible.

Think square

Statisticians have long known that it pays to run your contrived experiment such that the variables that you wish to study and that are under your control (actually, or in principle) are changed in some orthogonal pattern. (The more-intuitive, hold-everything-constant-except-one-variable approach is far more tedious variation of the orthogonal pattern, and may not even be possible.) The pattern of variation is called the experimental design. The design also forces (if followed) the technician to avoid confounding variables, that is, changing them both at once in the same direction. Thus, if you wish to study the effects of temperature and concentration on a reaction, if both were raised and lowered together, you would not know which one was really influencing the result. Efficient designs thus have a combination of low-high, high-low etc.

Save time by design

In your design, imagine the variables plotted in space, and locate experimental conditions at the extreme corners of this space, as far as is physically possible. This gives maximum leverage for determining the influence of each variable. If one has two variables to study (e.g., gas flow rate and liquid flow rate) and can make only 4 runs, then Design A will give more precise results than Design B.

Design	Liquid Flow Rate, L/min	Gas Flow Rate, L/min
A	1	10
	5	50
	1	50
	5	10
B	3	10
	3	50
	1	30
	5	30

Intuitively, one would think that it would be necessary to hold all variables exactly constant and vary only one at a time, but this is not the case as long as the variables are included in the proposed model. Also, in real life, variables can be hard to control exactly. The caustic concentration in the gas absorption experiment is an example. Some students feel they must hold a variable constant, and end up assuming that it is constant, rather than measuring it accurately. For example, setting gas flow in the gas absorption column to exactly 60 L/min really can't be done. Don't try!! However, do record the meter reading accurately when your sample is taken. This may require several readings if the meter is "noisy." When the dependent variable is observed and recorded, one run is complete.

Random is better than orderly

While one can plan to avoid confounding independent variables, it is easy to confound known variables with ones that you don't anticipate. For example, it would be a very risky plan to vary flow rate monotonically from low to high as the experiment progresses. Why? You may not know why, but one possibility is that the temperature of the run may simultaneously increase, or the analysis technique drifts. Thus one executes the planned runs in a *random* order, so that these unknown effects appear as random error (along with all the other random errors, such as misreading a gauge). This technique is called blocking.

Enter the model and the computer

The problem with the approach recommended above is that one has to do something with data that is gathered under all combinations of conditions. It looks like a mess. Fortunately, tools are available on every spreadsheet to untangle the mess. The most useful to everyday engineers is multivariable modeling. The model refers to the law or empiricism mentioned earlier. It is usually (hopefully?) an algebraic equation featuring the dependent variable (it's great if this can be all alone on the left-hand side of the equation, but such is not necessary), the independent variables, the unknown parameters, and perhaps a few constants (constants, in this context, are real constants like $\pi = 3.14159\dots$). If you can't find an equation to use, you will have to make one up. Use some sense: make it physically realistic, or at least defined over the entire possible range of the independent variables.

If the variable range is small, then one might assume that the real law governing the phenomena can be expanded in a Taylor series. For example, we might argue that the law (known only to the divine, and perhaps a few professors) for mass transfer in a packed bed could be expanded as:

$$K_g a = \mathbf{b}_0 + \mathbf{b}_1 V + \mathbf{b}_2 L \quad (1)$$

where $K_g a$ is the mass transfer coefficient, V is the gas flow rate, and L is the liquid flow rate. The \mathbf{b} 's are the Taylor coefficients, and they become our unknown parameters. A slightly more sophisticated model would be

$$K_g a = a_0 L^{a_1} V^{a_2} C^{a_3} \quad (2)$$

which incorporates the concentration of NaOH (C) as a variable. It also introduces the possibility of curvature with no more parameters.

Empiricism vs. Theory

So what good is this? It certainly does not give us the real theory (although the result might help the theorist). However, it does tell us if V and L are influencing the mass transfer, and it may give us a design equation for the column to predict the performance of a plant. It may permit us to compare our results with similar experiments reported in the literature. But best of all, it allows us to do something with the messy data from an experiment and end up with a positive result.

Let's try out Eq. 2 on some real data. These data are appended. Note that the group getting this data was trying the OVAAT approach. Plotting $K_g a$ vs V at constant liquid flow rate L is attempted in Figure 1, but the results are not very informative because the other variables are having a huge effect or there is too much error. It looks like the experiment is a failure.

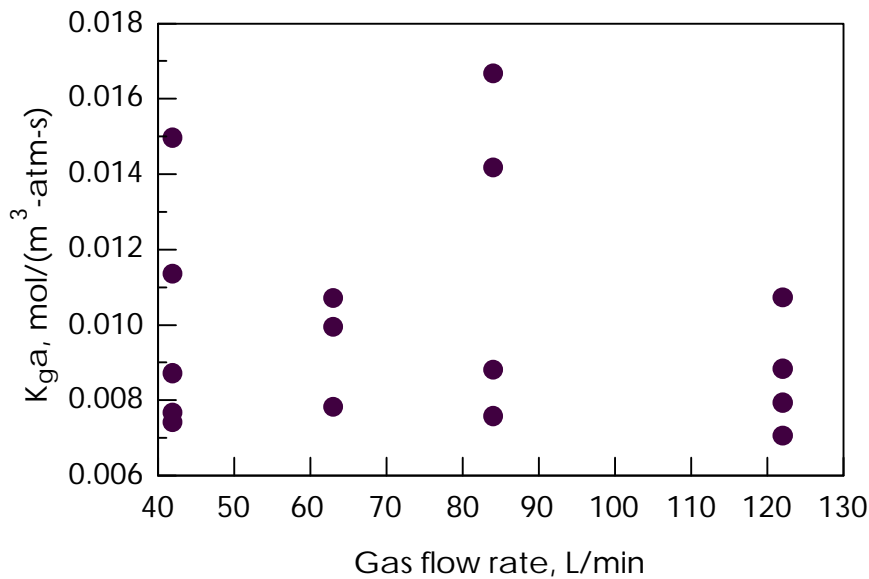


Figure 1. Mass-transfer coefficients for absorption of CO₂ into NaOH solution in packed column.

However, application of Eq. 1 gives the following results:

Model Equation(s):

$$KGA=A0 *V^{A1}*L^{A2} *C^{A3}$$

Parameter: A0= 0.00292220426

95 % Confidence Interval

LOW: -0.0018247393
HIGH: 0.0076688245

Parameter: A1= 0.34756350

95 % Confidence Interval

LOW: 0.016897327
HIGH: 0.67822967

Parameter: A2= 0.40309582

95 % Confidence Interval

LOW: 0.16198528

HIGH: 0.64420636

Parameter: A3= 0.28802167

95 % Confidence Interval

LOW: 0.11818315

HIGH: 0.45786019

The coefficients appear to all be significant, as their 95% confidence intervals are positive (although large). The reason that we are able to achieve this result is that there is an adequate quantity of data which has "overcome" the scatter in the data. To get a picture of the scatter, we plot the mass-transfer coefficients that are predicted by the model against the observed coefficients (or the other way). The result, shown in Figure 2, can be helpful in looking for regions where the model is not working well.

One must be careful about the logic of the problem. By fitting the model to the 16 data points, we are assuming that the points are independent, that is, there is no more relationship between any two points than there is for any two others. On examining the data set, it is clear that the data were taken at fixed L and V as the NaOH was depleted during a single run. Thus it could be argued that there are only 5 independent observations. In making this decision, it is important to know the sources of error and their magnitude. The titration analysis is a source of considerable error and is run independently, to a large extent, on each sample drawn from the column.

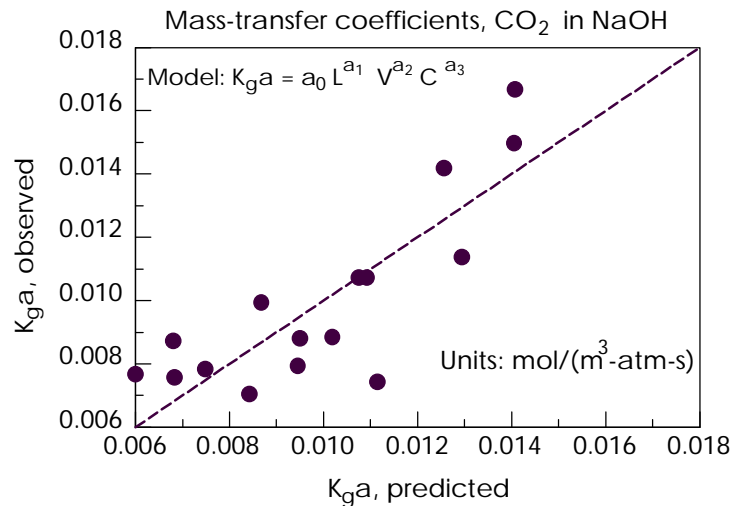


Figure 2. Application of multivariable analysis of mass-transfer coefficients with liquid flow rate (L), gas flow rate (V) and NaOH concentration (C) as factors. (See text for values of coefficients a_i)

APPENDIX. Data for above examples:

L, L/min	V, L/min	[NaOH]	Kga, mol/m ³ -atm-s
2	42	0.07844	0.008715
2	42	0.05088	0.007676
4	63	0.09434	0.010718
4	63	0.04240	0.009945
4	63	0.02544	0.007824
8	42	0.13992	0.014966
8	42	0.10494	0.011360
8	42	0.06254	0.007416
6	84	0.09116	0.016666
6	84	0.06148	0.014182
6	84	0.02332	0.008802
6	84	0.00742	0.007577
2	122	0.106	0.010727
2	122	0.08798	0.008837
2	122	0.06784	0.007932
2	122	0.04558	0.00705