Using Simulation to Increase Yields in Chemical Engineering

William C. Conley
Business Administration, University of Wisconsin
Green Bay, Wisconsin 54311-7001, USA

ABSTRACT

Trying to increase the yields or profit or efficiency (less pollution) of chemical processes is a central goal of the chemical engineer in theory and practice. Certainly sound training in chemistry, business and pollution control help the engineer to set up optimal chemical processes. However, the ever changing demands of customers and business conditions, plus the multivariate complexity of the chemical business can make optimization challenging.

Mathematical tools such as statistics and linear programming have certainly been useful to chemical engineers in their pursuit of optimal efficiency. However, some processes can be modeled linearly and some can not. Therefore, presented here will be an industrial chemical process with potentially five variables affecting the yield. Data from over one hundred runs of the process has been collected, but it is not known initially whether the yield relationship is linear or nonlinear. Therefore, the CTSP multivariate correlation coefficient will be calculated for the data to see if a relationship exists among the variables.

Then once it is proven that there is a statistically significant relationship, an appropriate linear or nonlinear equation can be fitted to the data, and it can be optimized for use in the chemical plant.

Keywords: CTSP Test, Nonlinear Chemical Yield Optimization

1. INTRODUCTION

The data presented here consists of the percentage increases in temperature, pressure, length of reaction time, catalyst one, and catalyst two from the standard settings and amounts currently being used. These five variable (X_1, X_2, X_3, X_4, X_5) values are noted and $Y=X_6$ (the percentage increase in yield from the old standard average) also is recorded.

Then the general purpose multi stage Monte Carlo optimization (MSMCO) technique is used to calculate the CTSP statistic and test it for correlation for the process. If the results show a promising correlation, then the subsequently fitted yield equation from the data will be optimized using MSMCO. The advantage of doing the CTSP statistical test first is that if it fails to show a correlation, then the process perhaps needs rethinking or rearranging. However, if it shows a good relationship (linear or nonlinear among the variables) then it can be pursued. That way much time is not lost on multivariate modeling of relationships that do not exist.

Briefly, MSMCO finds a shortest route connecting the *n* six dimensional data points in a closed loop tour through six dimensional space. This value is the numerator of CTSP (correlation of the traveling salesman problem shortest route). Then a random set of *n* points in the same range of the six dimensional space that the data came from, is also worked on by MSMCO to find a shortest route. This value is the denominator of CTSP. Therefore, CTSP=numerator/denominator. The further this value is from one (less than one, or greater than one if you use its reciprocal), the more doubt that is cast on the hypothesis of no correlation between the variables represented by the data under consideration.

Several more runs of the MSMCO shortest route program on sets of random data will produce enough numerators and denominators of CTSP's under the hypothesis of no correlation. That way the CTSP's sampling distribution can be estimated, for the subsequent crucial test of hypothesis.

2. A LINEAR EXAMPLE

Consider the following short illustration of a CTSP test. A sample of n=119 six dimensional points (X_1, X_2, X_3, X_4, X_5, X_6) is tested for the hypothesis of no correlation. Its shortest route calculated with

MSMCO was 4603.048 (using the standard six dimensional metric obtained by generalizing the Pythagorean theorem to six dimensions).

Then six random sets of n=119 points in the same ranges (0-100) were generated with a random number generator, and their MSMCO generated shortest routes were 5343.728, 5367.962, 5416.098, 5428.896, 5529.960 and 5597.184. The 6x5=30 CTSP quotients from these random data shortest routes are all in the range of .9547 to 1.0474. Our CTSP for the real data is CTSP=4603.048/5422.4997=.8487 (using the median of the six random set's distances as the denominator). This .8487 is surely statistically significant as the majority of the CTSP's sampling distribution, under the null hypothesis of no correlation, is between .9547 and 1.0474. Also, probably 99+% of it is greater than .9, so .8487 shows that the data here is highly correlated. This turned out to be a linear relationship.

However, the six variable chemical yield problem that will be featured here is nonlinear. It is nice to have a correlation coefficient (CTSP) for our computer age, that can pick up both linear and nonlinear relationships with equal ease. All calculations for this presentation were done on a 1999 inexpensive desktop PC. Multi stage Monte Carlo optimization (MSMCO) is an affordable and practical simulation based solution approach for engineers dealing with the complex problems of our 21st century.

3. A NONLINEAR STUDY

The chemical engineers at our plant believe that increasing the temperature (X_1), pressure (X_2), length of reaction time (X_3) and catalyst one and two (X_4 and X_5) will lead to a significant increase in the yield of our chemical and hence cut costs and increase profit for our company. They have tried n=119 sample runs of the chemical yield process and have recorded the following data. All six variable values are in percentage increases (from 0 to 100%) in the standard values for the five input variables and the sixth variable (the dependent yield variable increase also).

	X_1	X_2	X_3	X_4	X_5	X_6
1	74	19	84	14	61	73
2	36	64	58	91	21	85
3	35	82	66	79	91	93
4	88	9	99	93	85	73
5	1	5	20	77	68	54
6	8	57	57	37	25	76
7	36	79	42	68	28	88
8	21	29	92	78	54	75
9	85	17	73	73	97	85
10	39	70	73	12	50	80
11	27	78	8	81	98	92
12	66	70	68	38	38	79
13	10	55	29	98	11	73
14	56	99	93	83	79	74
15	83	88	95	95	23	53
16	38	9	30	94	83	81
17	87	100	55	20	48	50
18	96	44	29	87	53	82
19	80	56	42	56	64	90
20	67	15	7	14	19	65
21	41	79	100	17	4	55
22	2	61	30	56	62	86
23	92	100	93	72	17	36
24	11	82	75	24	100	80
25	65	29	9	4	44	73
26	8	14	12	47	68	64
27	99	55	70	26	56	69
28	67	67	11	90	21	79
29	25	80	43	63	100	95
30	21	43	71	36	100	84
31	65	20	62	84	47	87
32	74	16	59	100	80	87
33	92	8	62	84	52	80
34	100	11	11	67	91	80
35	66	15	21	97	70	88
36	27	16	59	100	91	78
37	57	28	74	86	62	90
38	51	72	81	43	53	86
39	53	100	32	5	53	67
40	6	97	100	50	21	64
41	64	45	9	71	30	86
42	14	17	74	81	29	62
43	87	45	11	83	83	87
44	58	82	77	38	63	82
45	30	23	74	30	35	71
46	58	45	66	94	71	94
47	100	59	84	80	37	66
48	21	18	22	69	4	60
49	67	83	3	3	93	68
50	27	13	45	44	67	77
51	25	51	75	95	0	69
52	43	26	86	65	37	80
53	2	28	39	91	84	73
54	87	98	14	45	40	55
55	86	57	91	32	5	56

56	84	97	32	53	30	59
57	12	67	80	91	31	79
58	45	40	41	93	85	96
59	81	2	13	58	24	71
60	15	48	51	78	10	75
61	39	28	31	34	29	79
62	0	26	50	63	37	63
63	97	37	83	58	26	69
64	69	59	48	67	75	95
65	1	89	100	34	100	71
66	83	32	78	71	33	80
67	2	86	74	88	75	85
68	97	99	36	59	31	47
69	54	85	19	87	28	81
70	54	76	70	48	78	90
71	54	79	92	44	70	82
72	21	41	57	74	38	85
73	64	32	35	23	35	82
74	100	87	25	100	83	59
75	81	26	59	11	54	76
76	33	40	46	24	43	83
77	77	70	43	46	100	85
78	9	49	4	100	60	81
79	84	38	98	34	61	75
80	100	11	69	100	2	57
81	3	35	79	53	22	62
82	64	91	71	28	54	73
83	54	82	9	50	2	68
84	80	10	82	78	46	78
85	88	35	46	65	77	90
86	17	92	80	72	25	77
87	45	11	82	76	37	74
88	25	31	61	87	77	87
89	72	23	13	80	43	86
90	92	93	45	67	95	64
91	97	48	43	38	45	76
92	9	62	84	44	37	77
93	79	55	43	77	70	92
94	50	10	32	100	5	66
95	100	80	78	62	95	62
96	28	100	31	18	37	72
97	89	75	74	79	17	62
98	96	39	100	47	100	71
99	40	42	97	97	27	75
100	0	100	93	47	70	75
101	97	69	60	56	5	55
102	51	59	100	76	68	86
103	52	46	15	98	25	85
104	5	100	51	2	83	72
105	6	56	64	90	70	86
106	43	34	30	80	99	94
107	49	34	80	65	91	91
108	63	58	93	32	92	82
109	97	36	84	71	88	79
110	99	50	37	84	64	81
111	100	9	89	27	96	67
112	70	70	100	0	3	43
113	78	88	38	67	57	77
114	3	83	52	88	61	88
115	46	4	36	71	94	82
116	91	24	59	36	100	82
117	88	99	13	23	21	42
118	36	35	97	55	90	83
119	85	100	71	12	59	49

The MSMCO algorithm, adjusted for a TSP shortest route through six dimensions with n=119 is used on these 119 data points collected by the chemical engineers. Its calculated distance is 4,406.754. Therefore, we can use the median of the six random TSP shortest routes (that we used in the previous linear example) as the denominator here so CTSP=4,406.754/5,422.497=.8127 which is clearly statistically significant when compared with the bounds of .9547 and 1.0474 on our 30 CTSP quotients (in the linear problem) that were used to estimate the sampling distribution of CTSP under the null hypothesis of no correlation between the six variables.

Therefore, the variables are clearly correlated. Some additional experimenting finds a satisfactory nonlinear curve fit of

$$\begin{aligned} Y = X_6 &= 7.128 - .00648X_1^2 + .94608X_1 - \\ &.00648X_2^2 + 1.01736X_2 - .00648X_1X_2 - \\ &.00324X_3^2 + .25272X_3 - .00324X_4^2 + \\ &.46656X_4 - .00324X_5^2 + .52488X_5 \end{aligned} \qquad \text{Eq. (1)}$$

Then MSMCO can be used to optimize this function in the ranges of 0 to 100 for each input variable. The optimal solution is X_1=45, X_2=56, X_3=39, X_4=72 and X_5=81. Therefore, increasing the temperature by 45%, the pressure by 56%, the length of reaction time by 39% and catalysts one and two by 72% and 81%, respectively, should come close to optimizing the yield of our company's important chemical for each run of the process. Also, of course, as conditions change (new equipment is brought on line, etc.), the model can be updated to suit. Additionally, costs and a demand equation could be added so that the yield equation could be expanded into a profit equation to be maintained and updated and maximized as business conditions change.

4. GENERAL MSMCO

The multi stage Monte Carlo optimization (MSMCO) technique is a general purpose optimization technique for engineering and business. It is not limited to traveling salesman problems (TSP) [1] and the CTSP correlation coefficient statistic calculations. It was used in [2] to optimize pollution control equipment and in medical research in [3] as examples of its versatility.

Multi stage Monte Carlo optimization (MSMCO) makes repeated searches in an ever narrowing and moving subset of the feasible solution space of the optimization problem in question. Its goal is to cross the sampling distribution to the optimal solution region and then further close in on and find the optimal solution (or a useful approximation if the exact optimal can not be located).

The first few stages of multi stage do not produce very good answers. However, the MSMCO computer program "learns" from these early "mistakes" where the optimal solution region is. Then MSMCO redirects the rest of the simulation to this area of the feasible solution space and refines the answer so it is accurate enough to be useful to the engineer or scientist doing the study. It is a computer intensive optimization approach for our 21st century computer age.

5. CONCLUSION

Presented here was a multivariate data analysis using the CTSP correlation statistic and its estimated sampling distribution on a multivariate chemical yield industrial process. The CTSP calculation and subsequent statistical test of hypothesis showed a strong correlation between the variables in the study. Then a yield equation was fitted to the data and optimized to give the manager of the process and the engineers a good idea of how to improve the yield of the chemical by adjusting the values of the input control variables.

The CTSP statistic may be useful in screening other multivariate data sets to see if any relationships exist between the variables under consideration. It can pick up linear or nonlinear relationships with equal ease. The multi stage Monte Carlo optimization (MSMCO) simulation technique that was featured here can also be used in general multivariate optimization. Additional examples of MSMCO's versatility are in [4], [5] and [6].

The world will face many problems in the 21st century. Not all of these can be quantified. However, those that can be expressed mathematically (from physics and engineering principles and/or data analysis) can be optimized exactly (or approximately) with MSMCO. This gives the engineer, scientist, and decision maker one more tool when faced with difficult decisions in our complex multivariate world.

6. REFERENCES

[1] W. Conley, "Multi Stage Monte Carlo Optimization Applied to a Six Hundred Point Traveling Salesman Problem", **International Journal of Systems Science**, Vol. 24, No. 4, 1993, pp. 609-626.

[2] W. Conley, "Ecological Optimization of Pollution Control Equipment", **Proceedings of the 13th European Simulation Society Conference**, October 2001, pp. 543-548.

[3] W. Conley and H. Wichowski, "Simulation and Model Optimization of Three Bio-Medical Systems", **Journal of the Japan Society for Simulation Technology**, Vol. 7, No. 4, 1988, pp. 47-57.

[4] D. Szarkowicz, "Investigating the Brachistochrone with a Multi Stage Monte Carlo Method", **International Journal of Systems Science**, Vol. 26, No. 2, 1995, pp. 223-243.

[5] J. Wong, "A Note on Optimization in Integers", **International Journal of Mathematical Education in Science and Technology**, Vol. 27, No. 6, 1996, pp. 865-874.

[6] W. Conley, "Multi State Monte Carlo Optimization Applied to Multivariate Decision Making in Finance", **Computers in Education Journal**, Vol. 5, No. 2, 1995, pp. 68-75.