Congrès annuel de la Société canadienne de génie civil

Annual Conference of the Canadian Society for Civil Engineering

Montréal, Québec, Canada
5-8 juin 2002 / *June 5-8, 2002*

DESIGN OF EXPERIMENTS IN CIVIL ENGINEERING: ARE WE STILL IN THE 1920'S?

L. M. Lye[A]
A Faculty of Engineering and Applied Science, Memorial University, St. John's, Newfoundland, Canada

ABSTRACT: In this paper, a very useful class of statistically designed experiments, the two-level factorial design will be introduced and the pitfalls of the commonly used one-factor-at-a-time (OFAT) method of experimentation will be discussed. Applications of statistically designed experiments or DOE in civil engineering in particular will be emphasized. Statistically designed experiments have been around for the last eighty years and have been widely used with great success in many industries for the past thirty years. However, they have yet to make any inroads in civil engineering except for certain areas of environmental engineering. In fact, many experienced researchers in civil engineering have never heard of it and have continued to use the inefficient trial-and-error or OFAT methods. These include many authors of papers in journals and presenters of papers at CSCE conferences over the years. It is hope that this paper can serve as a tutorial and will provide a stimulus for experimenters in civil engineering to abandon the trial-and-error and OFAT methods of experimentation and begin to learn and use the well established statistically based design of experiments.

1. INTRODUCTION

Like many scientists and other engineers, civil engineers carry out a fair amount of experimentation in laboratories and in design offices in areas of structural engineering, hydraulic engineering, geotechnical engineering, environmental engineering, etc. We carry out experiments to 1) evaluate and compare basic design configurations, 2) evaluate material alternatives, 3) select design parameters so that the design will work well under a wide variety of field conditions (robust design), and 4) determine the key design parameters that impact performance. The experiments may be computer simulations or actual laboratory experiments. The question is - How do we design the experiment so that we can achieve the above objectives efficiently and with a certain level of confidence?

A prevalent, but potentially disastrous, type of experimentation that is commonly used in civil and other disciplines of engineering is the "one-factor-at-a-time" or OFAT experiment. OFAT experiments are regarded as better than the trial-and-error approach of experimentation, and have been perceived by some as "the scientific method". It will be shown that OFAT experiments are inefficient, and that they do not provide any information on how the factors interacts with one another which may be far more important that the effect of individual factors. OFAT experimentation has been outdated since the 1920s with the invention of modern statistically based methods of experimentation based on the factorial designs. Factorial designs are far more efficient, provide information on factor interactions, and conclusions drawn are generally applicable to a wide range of conditions because several factors have been changed simultaneously. By using factorial designs, we can learn about the process we are investigating, screen important variables, build a mathematical model for prediction, and if required optimize our design. Statistically designed experiments or commonly referred to as DOE have been around for the last eighty years and have been widely used with great success in many industries for the past thirty years. However, they have yet to make any inroads in civil engineering except for certain areas of environmental engineering. In fact, many experienced researchers in civil engineering especially those outside of environmental engineering have never even heard of it and have continued to use the inefficient OFAT method. These include many authors of papers in journals and presenters of papers at CSCE conferences over the years.

In this paper, the pitfalls of using the OFAT method of experimentation will be highlighted and a very useful class of statistically designed experiments, the two-level factorial design will be introduced. Applications in civil engineering will be discussed. It is hope that this paper can serve as a tutorial and will provide a stimulus for experimenters in civil engineering to abandon the trial-and-error and OFAT methods of experimentation and begin to learn and use DOE.

2. EXPERIMENTATION STRATEGIES

An experiment is basically a test or a series of tests in which purposeful changes are made to the input variables or factors of a system so that we may observe and identify the reasons for changes that may be observed in the output response (Montgomery, 2001). In the field of civil engineering, many empirical equations in use today ranging from hydraulics to soil mechanics are the result of experiments carried out by researchers in the field, laboratory, or in computers using sophisticated computer models. As engineering systems and designs become more and more complex, to figure out the system response(s) as the many input factors are varied may be out of the question using analytical methods. One may have no choice but to resort to the experimental method.

Consider the design of an engineering system that has 3 input factors (A, B, and C) and 2 output responses (Y_1 and Y_2). It is desired to determine:

- the relative contribution of A, B, and C to the responses Y_1 and Y_2;
- which factors have a synergistic or antagonistic effect on the responses;
- an equation that can be used to predict Y_1 and Y_2 given values of the input factors; and
- what combination of the factors would maximize Y_1 but minimize Y_2?

As with most engineering problems, we are faced with limited time and budget. Hence we would like to gain as much information as possible and do so as efficiently as possible. How would one proceed to conduct such an experiment?

2.1 Best-Guess (with engineering judgment) approach

In engineering, one often-used approach is the best-guess (with engineering judgment) approach. It works reasonably well at times, because the engineer often has a great deal of technical or theoretical knowledge of the system they are studying, as well as considerable practical experience. However, there are at least two disadvantages of the best-guess approach according to Montgomery (2001). First, suppose the initial best-guess does not produce the desired results. Now the engineer has to take another guess at the correct combination of factor levels. This could continue for a long time, without guarantee of success. Second, suppose the initial best-guess produces an acceptable result. Now the engineer is tempted to stop testing, although there is no guarantee that the best solution has been found. In addition, it is often difficult to explain to others how the “optimal” combination of factors are arrived at, to quantify how much each factor contributes to the responses, or to obtain a prediction equation.

2.2 One-factor-at-a-time (OFAT) approach

Another strategy of experimentation that is prevalent in practice is the one-factor-at-a-time or OFAT approach. The OFAT method is considered as the standard, systematic, and accepted method of scientific experimentation. This method consists of selecting a starting point, or baseline set of levels, for each factor, then successively varying each factor over its range with the other factors held constant at the baseline level. After all tests are performed, a series of graphs are usually constructed showing how the response variables are affected by varying each factor with all other factors held constant. If we consider only a low and a high value for each of the factors, then Figure 1 would result.

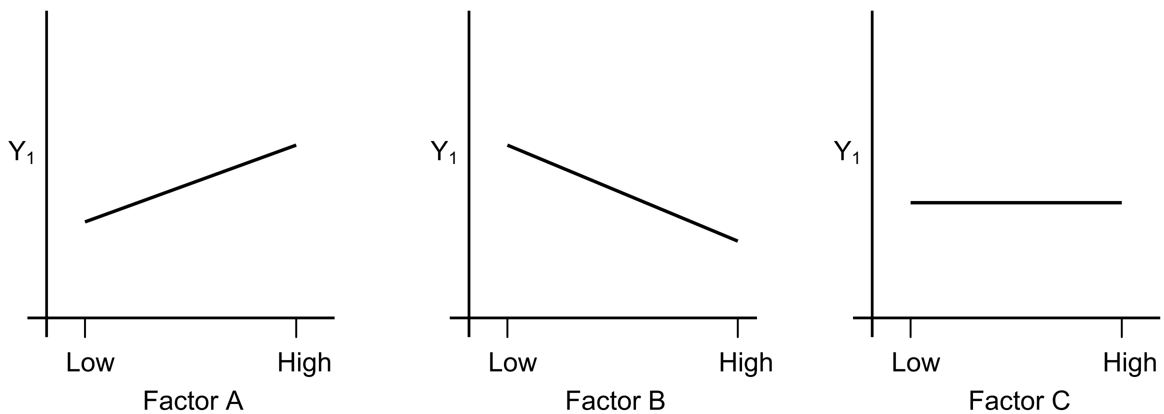

Figure 1: Results of the OFAT strategy of the experiment with 3 factors

The interpretation of the Figure 1 is straightforward. Since the objective is to maximize the response Y_1, we would want to keep factor A at the high level, and factor B at the low level. Factor C seems to have no effect of Y_1. In addition, we may also conclude that Y_1 may increase further when higher levels of factor A are used. An effects graph would then be constructed for response Y_2 and similarly interpreted.

The major disadvantage of the OFAT strategy is that it fails to consider any possible of interaction between the factors. An interaction is the failure of the one factor to produce the same effect on the response at different levels of another factor. Large 2-factor interactions among the variables measure the change in the effect of one variable as another is changed and so are "second-order" effects. Large two-factor interactions point to the likely presence of curvature effects as well, for these are also second-order effects measuring the change in the effect of one variable as a change in level is made in that same variable.

Interactions between factors are very common in engineering experiments as in everyday life, and if they occur, the OFAT method will produce poor and sometimes disastrous results. Many people including experienced researchers do not recognize this, or if they did, do not know how to quantify it. Consequently, the OFAT method is still frequently used in practice. An example of an interaction effect from our daily lives is drinking alcohol when taking certain prescription or non-prescription drug. When the drug and alcohol are taken together it may be fatal, yet when taken separately, may be quite harmless or at least has no serious consequences. One can see that the only way that the interaction effect can be discovered is to take the two substances together!

2.3 Factorial design approach

The OFAT way of experimentation became outdated in the early 1920s when Ronald A. Fisher discovered much more efficient methods of experimentation based on factorial designs. These were further developed in England and the United States to include fractional factorial designs, orthogonal arrays, and respond surface methodology. Instead of varying one factor at a time, the factorial approach says that one should:

- vary all factors simultaneously to deal with interactions and maximize efficiency (experimental precision); and
- do the experiments in random order to maximize accuracy and reduce the chance of wrong conclusions.

Design and analysis of experiments using the factorial concept or simply called design of experiments (or DOE) is an extremely important concept. In this paper, a very useful class of factorial design, the two-level full factorial design, will be presented. An intuitive and non-statistical approach to analyze the results is presented. It must be stressed that the statistics used are simple and straightforward and have been covered in the first course in probability and statistics by all engineering students in Canada. In fact the method of analysis presented can be learned by most high school students.

Basically, DOE is a methodology for systematically applying statistics to experimentation. DOE lets experimenters develop a mathematical model that predicts how input variables *interact* to create output variables or responses in a process or system. DOE can be used for a wide range of experiments for various purposes including nearly all fields of engineering and science and even in marketing studies. The use of statistics is important in DOE but not absolutely necessary. In general, by using DOE, we can:

- learn about the process we are investigating;
- screen important factors;
- determine whether factors interact;
- build a mathematical model for prediction; and
- optimize the response(s) if required.

It has been recognized for many years that the factorial-based DOE is the correct and the most efficient method of doing multi-factored experiments. In many organizations however, not only is it true that insufficient attention is given to experimentation, but often experimentation that is done is done extremely inefficiently. Sometimes not even OFAT is used but just “pick and try”. The extraordinary truth is, that about 80 years after Fisher invented modern experimental design, it is still not widely taught in schools of engineering and of science in our universities. In science and engineering as in all other endeavors, it is difficult to change accepted practice, even when other methods are shown to be better. It is hoped that this paper will encourage civil engineers to learn and use the factorial-based DOE and abandon the trial and error and OFAT approaches. Surely an engineer who does not know how to run an efficient experiment is not a very good engineer. Fisher once said about poorly designed experiments: “nothing much can be gained from statistical analysis; about all you can do is to carry out a postmortem and decide what such an experiment died of”.

3. TWO-LEVEL FULL FACTORIAL DESIGN

In this section, the two-level full factorial design or the 2^k design will be introduced. The two levels are normally referred to as a low and a high level. The levels may be qualitative (e.g. Brand A vs Brand B) or quantitative (e.g. 20 vs 50). The two-level, two-factor or in short 2^2 factorial design will first be discussed to illustrate the concept, terminology, and factor effects calculations. Simple graphical approaches will be used as much as possible for interpretation of the results. This will be followed by another example with more factors.

3.1 Example of a 2^2 factorial design – degradation of hydrocarbons in a landfarm

The 2^2 design is the simplest of the two-level factorial design. This design was recently used by one of the author’s student (Hejazi, 2002) to study the degradation of petroleum hydrocarbons in a land-farm when oily sludge from refinery tank bottoms mixed with sand, are subjected to various treatments. Tilling the sludge-sand mixture, addition of water to the mixture, and addition of nutrients to the mixture. Although there were three factors to be studied, from literature reviews and past experiences, it is known that tilling the sludge-sand mixture greatly enhances the degradation, as measured by the decrease in the oil and grease concentration, over time. Hence, because of the limitation of time and budget, only the effects of the other two factors in the presence of tilling was studied. That is, the effects of tilling alone, tilling with and without addition of water, and tilling with and without the addition of nutrients. The question to be answered was - does the addition of water and/or nutrients in addition to tilling enhances the degradation of oil and grease? The experiments were carried out in the vicinity of an actual land-farm in Saudi Arabia. The oily sludge was from a refinery belonging to Saudi ARAMCO, the largest oil company in the world. Details of the test cells and other results are discussed in Hejazi (2002). The test factors for the hydrocarbon degradation experiment and the response of interest are shown in Table 1.

Table 1: Test factors and response for the hydrocarbon degradation experiment

Factor	Name	Low Level (-1)	High Level (+1)
A	Water	No Water	Water
B	Nutrients	No Nutrients	Nutrients

Response Y: Decrease in oil and grease (O & G) concentration after one year (%)

For any 2^k experiment, all combinations of the k factors must be considered. With two factors, there will be four treatment combinations. The treatment combinations can be conveniently laid out in the form of a sign table, as shown in Table 2.

Table 2: Treatment combinations

Run	Combination	A	B	Description of combinations
1	(1)	-1	-1	no water, no nutrients
2	a	+1	-1	water, no nutrients
3	b	-1	+1	no water, nutrients
4	ab	+1	+1	water and nutrients

In Table 2, under the heading of “Combination”, is a shorthand way of writing down the various combinations. The “(1)” means that all factors are at the low level, “a” means that only factor A is at the high level, all other factors at the low level, “b” means that only factor B is the high level and all other factors at the low level, and “ab” means that both factors are at the high level. The “+1” and “-1” under the headings A and B also indicate the level of factor A and B for each combination. The “+1” means a high level and the “-1” means a low level. Using the “+1” and “-1” to indicate the combinations is the preferred method in most software for design of experiments.

It is understood that as described earlier, these experiments were done in the presence of tilling. If tilling were to be considered as a factor, then we would need to conduct 2^3 experiments or 8 test cells which would add considerably to the construction and analytical costs.

The next step is to carry out the experiment and obtain the response for each treatment combination. For the above experiment, the response of interest is the percentage decrease in the oil and grease concentration in the sludge-sand mixture. The percentage decreases were measured in each of the four test cells and the results are shown in Table 3. Table 3 will also be used to calculate the effect of each factor and the interaction of the factors. For a 2^k design, $2^k - 1$ effects can be estimated. That is, if there are say k=3 factors, we can estimate from the 2^3 experiments 7 effects: the main effects A, B, and C, the two factor interactions AB, AC, and BC, and one three factor interaction ABC. Since we have only two factors in our example, we can estimate from the four experiments 3 effects: main effects A and B, and the interaction effect AB.

Table 3: Responses and estimation of effects

Combination	A	B	AB	Y (%)
(1)	-1	-1	+1	76
a	+1	-1	-1	71
b	-1	+1	-1	40
ab	+1	+1	+1	75
$\overline{Y}_+$	73.0	57.5	75.5	
$\overline{Y}_-$	58.0	73.5	55.5	
Effect (Δ)	15.0	-16.0	20.0	

The AB column is for estimating the interaction between A and B. The "+1" and "-1" signs are obtained by multiplying the signs in columns A and B. Notice that there are equal numbers of positive and negative signs meaning that the design is orthogonal which is a very desirable property. To calculate the effect of factor A, we take the average of the responses with the "+1" and subtract the average of the "-1" responses under column A. For effect A,

Average of (+) responses = (71 + 75)/2 = 73
Average of (-) responses = (76 + 40)/2 = 58
Effect of A or Δ_A = 73 – 58 = 15

Δ_A measures the average change in Y as A changes from a low level to a high level. In this case, an average increase of 15% in oil and grease concentration reduction with the addition of water. The effects of B and the interaction of A and B or AB effect are similarly calculated. From the results, one can see that A has a positive effect of 15, B has a negative effect of –16 (meaning a decrease in oil and grease concentration reduction) with the addition of nutrients. The largest effect is the interaction effect AB of 20 which measures the change in the effect of one factor as another is changed. In the presence of a large interaction effect, we must be very careful in interpreting the main effects of A and B separately since it is their joint effect that is important. The best way to interpret the interaction effect is via the interaction plot shown in Figure 2.

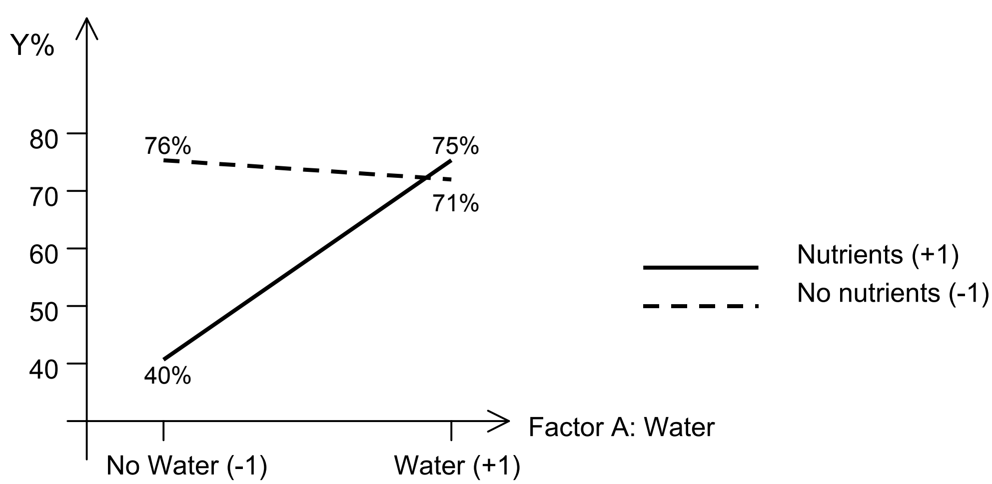

Figure 2: Interaction plot of factors A and B

From Figure 2, the interaction between the two factors A and B can be clearly seen. One can see that when there is no water and no nutrients, we get the best response of 76%. When water is added, the presence or absence of nutrients gave similar responses (75% vs 71%). However, without water, there is a large drop in the response with the addition of nutrients. Since all the treatment combinations were conducted with tilling, this means that the best treatment is tilling alone without addition of either water or nutrients. This is an important conclusion especially in Saudi Arabia where water and nutrients (fertilizers) are expensive and in short supply. Thus tremendous savings can be achieved if only tilling is used to enhance the degradation of the sludge-sand mixture.

Another advantage of factorial designs is that all responses are used for the estimation of each effect. Suppose that the OFAT method had been used instead. To estimate with the same precision, the effect of changing each of two factors would have required not 4, but 6 test runs. But even with 6 runs the OFAT design would not have allowed us to determine interactions between the factors. But it was one such interaction that resulted in the considerable savings. Thus, not only would the OFAT method have required 1.5 times as many runs, it would have failed to find the savings unless the two factors were changed together. With more factors, the efficiency of the factorial design is even more dramatic. For example, with 5 factors, instead of 32 runs, the OFAT method would require 96 runs for the same precision yet it can estimate only 5 instead of 31 effects.

3.2 Example of a 2^4 factorial design – a replacement model for Penman's equation

This next example illustrates the use of factorial designs for developing simple models in the form regression equations as a replacement for complex equations or time consuming computer models. Developing a simple replacement model for Penman's equation for estimating evaporation from open water will be used as an example. The Penman equation is considered as the most physically-based and accurate method of estimating evaporation based on measured climatic variables. It is based on a weighted combination of the aerodynamic method and the energy balance method. For a given latitude and month of the year, the solar radiation received at the top of the atmosphere is known. Then given the average temperature, wind speed, relative humidity, and the amount of sunshine hours, the evaporation from an open body of water can be estimated using Penman's equation. However, to obtain the final estimate of the evaporation, several tables must be consulted and a series of equations must be applied and then combined. The procedure for Penman's method is given in most hydrology texts and hence will not be repeated here.

Estimating evaporation losses from an open water body such as a reservoir or lake used for water supply is very important especially in tropical areas such as Indonesia. Hence in this example, the climatic conditions in Indonesia will be used. The evaporation E_o in mm/day for a monthly average incoming solar radiation under various combinations of the four climatic factors is the response variable. This can later be adjusted for any month in the year. The test factors are shown in Table 5.

Table 5: Test factors for Penman's equation

Factor	Name	Low Level (-1)	High Level (+1)
A	Mean Temperature (°C)	20	35
B	Actual sunshine hours to maximum possible (%)	10	90
C	Relative humidity (%)	20	90
D	Wind speed (m/s)	0.2	5.0

Since this is a 2^4 factorial experiment, 16 estimates of evaporation E_o from 16 combinations of the test factors are required. The factor combinations and the resulting responses are shown in Table 6. The responses are for an average monthly incoming solar radiation at the equator of 35.74 MJ/m^2/day.

Table 6: Run combinations and estimated evaporation from Penman's equation

Run	Combination	A	B	C	D	E_o (mm/day)
1	(1)	-1	-1	-1	-1	2.82
2	a	+1	-1	-1	-1	3.52
3	b	-1	+1	-1	-1	4.72
4	ab	+1	+1	-1	-1	5.88
5	c	-1	-1	+1	-1	2.45
6	ac	+1	-1	+1	-1	3.40
7	bc	-1	+1	+1	-1	5.37
8	abc	+1	+1	+1	-1	7.97
9	d	-1	-1	-1	+1	6.82
10	ad	+1	-1	-1	+1	8.79
11	bd	-1	+1	-1	+1	8.72
12	abd	+1	+1	-1	+1	11.15
13	cd	-1	-1	+1	+1	2.95
14	acd	+1	-1	+1	+1	4.06
15	bcd	-1	+1	+1	+1	5.87
16	abcd	+1	+1	+1	+1	8.63

With 16 runs, we can estimate 15 effects. Four main effects (A, B, C, and D), six two-factor interactions (AB, AC, AD, BC, BD, and CD), four three-factor interactions (ABC, ABD, ACD, and BCD), and one four factor interaction (ABCD). However, not all estimated effects are important. Usually, some two-factor and most three-factor interactions and above can be ignored. This is known as the "sparsity of effects" principle. Due to limited space, details of effects and other calculations will not be shown here. Suffice to say that a more detailed analysis indicated a square-root transform of the responses gave a better model. More details can be found in Susilo and Lye (2002). The 15 estimated effects are shown in Table 7 arranged in descending order by absolute effect size, and plotted in as a Pareto chart shown in Figure 3. Using the Pareto chart, one can determine visually the few important effects. The square-root of the responses were used for effects calculations.

Table 7: Estimated effects (arranged in order of absolute magnitude)

Effect	Effect size (Δ)
B	0.640
D	0.532
CD	-0.397
A	0.348
C	-0.297
BC	0.200
AB	0.067
BD	-0.054
AC	0.051
ABC	0.045
AD	0.040
ACD	-0.033
BCD	0.028
ABD	-0.004
ABCD	0.002

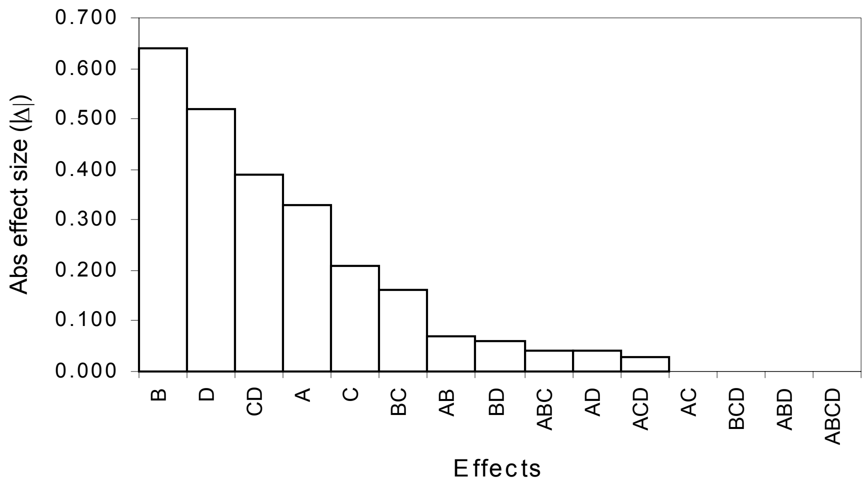

Figure 3: Pareto chart of effects

From Table 6 and Figure 3, it is clear that the most important effects in order of absolute magnitudes are: B (sunshine ratio), D (wind speed), CD (interaction of humidity and wind speed), A (mean temperature), C (humidity), and BC (interaction of sunshine ratio and humidity). Using the important effects, one can then develop a regression equation as a replacement for Penman's equation. The regression equation is of the form:

[1] $$\sqrt{Y} = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_D D + \beta_{BC} BC + \beta_{CD} CD$$

Where: $\sqrt{Y}$ = square-root of the evaporation, β_0 = grand mean of the square-root transformed data, and β_i = regression coefficients which is equal to half the effect size, and the value of the factors are in coded units (-1 and +1). The final equation in coded factors after substituting the estimated coefficients and grand mean is:

[2] $$\sqrt{Y} = 2.35 + 0.17A + 0.32B - 0.15C + 0.27D + 0.10BC - 0.20\ CD$$

The regression equation gave a R^2 of 0.985, and a predicted R^2 of 0.953, and the residuals were normally distributed and homoscedastic. The excellent fit means that the regression equation can be used in place of the complicated Penman's equation without much loss of accuracy. In addition, the contribution of each factor can be clearly seen and easily interpreted. For example, the negative coefficient for C (relative humidity) means that evaporation decreases with humidity. However, there is also a large interaction effect between humidity (C) and wind speed (D). This interaction can be explained using the interaction plot shown in Figure 4.

From Figure 4, it can be seen that at high relative humidity the evaporation is only marginally higher at high wind speeds than at low wind speeds. However, at low relative humidity, there is a dramatic increase in evaporation at high wind speeds compared to low wind speeds. This of course makes physical sense and the interaction plot easily explained the phenomenon. To estimate evaporation for the various months in the year, the estimated evaporation using [2] is multiplied by a factor for each respective month shown in Table 8.

Table 8: Monthly factors (for latitude 0° only)

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dep
0.989	1.129	1.029	1.027	0.941	0.941	0.928	1.040	1.045	1.015	1.023	0.972

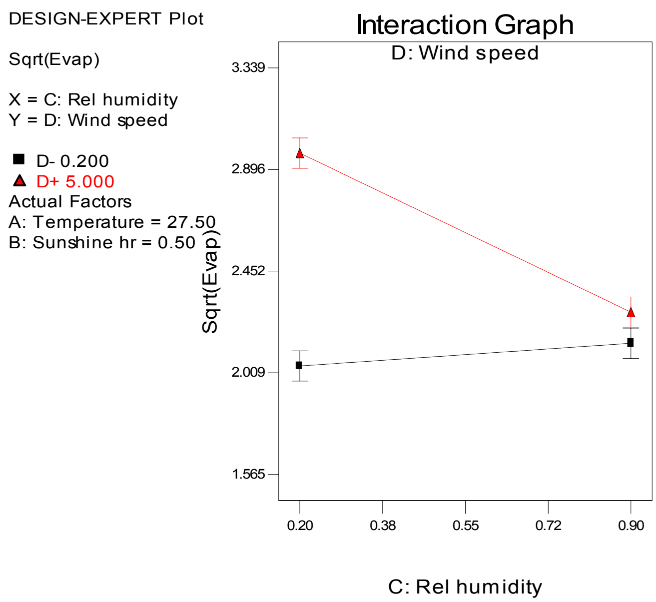

Figure 4: Interaction plot of Factors C (relative humidity) and D (Wind speed)

In this example, the important effects were selected using a simple graphical procedure called a Pareto chart. A more objective approach is to use a statistical approach based on the analysis of variance (ANOVA) to test for the significance of the effects and model chosen. Regression diagnostics are also performed such as checking for normality of residuals, constant variance of the residuals, checking for outliers, and whether there is a need for transformation of the responses to fulfill the assumptions of regression. These tests are normally carried using computer software. However, there is a need to apply common sense to choose effects that are of practical significance rather than effects that are statistically significant. Another technique for choosing significant effects is to use a half-normal plot. Details of these other methods can be found in Montgomery (2001).

4. CONCLUSIONS

A very brief introduction to DOE was given with applications to two different types of experiments. It can be seen that the method is quite straightforward. With widely available easy to use software, the analysis is even easier. The potential for applications of DOE in civil engineering is numerous. As the number of factor increases, a full factorial design would require too many test runs. For these cases, a fractional factorial design can be used. For these designs only a fraction of the test runs are required. For certain engineering problems, we can combine dimensional analysis with factorial designs. We can first use dimensional analysis to reduce the number of variables and then use a factorial design with the dimensionless groups as factors. These and other applications are discussed in Lye and Sharp (2002).

5. REFERENCES:

Hejazi, R. (2002): *Oily sludge degradation study in arid conditions*: *landfarming and bioreactors*. PhD thesis, Memorial University of Newfoundland, St. John's, Canada.

Montgomery, D. G. (2001): *Design and analysis of experiments*, 5th Edition, John Wiley and Sons, Inc.

Lye, L. M. and Sharp, J. J. (2002): *Guide to design of experiments for engineers*, book in progress.

Susilo, G. and Lye, L. M. (2002): *A simple replacement model for Penman's equation*, submitted to Hydrological Sciences Journal.