Leadership in Sustainable Infrastructure
Leadership en Infrastructures Durables

Vancouver, Canada
May 31 – June 3, 2017/ *Mai 31 – Juin 3, 2017*

A SIMPLE DOE-BASED REPLACEMENT MODEL FOR PENMAN'S EVAPORATION EQUATION

Lye, Leonard M.[1,2]
[1] Faculty of Engineering and Applied Science, Memorial University of Newfoundland, Canada.
[2] llye@mun.ca

Abstract: Penman's equation for the estimation of evaporation from measured climatic variables is the most well-known and considered to be the most accurate method for open water evaporation. It is based on a weighted combination of evaporation estimated using energy balance and aerodynamic methods. However, usage of Penman's equation is not straightforward. It requires applying a series of equations and reference to at least one table. In addition, it is difficult to determine from the series of equations, the relative contribution of the each climatic variable to the estimated evaporation and how the variables interact with one another. In this paper, a simple replacement model for Penman's equation obtained by applying statistical design of experiment (DOE) methodology will be demonstrated. The replacement model is based on a simple two-level factorial design. The replacement model is a simple one-line regression equation and the effects of each variable and their interactions are clearly seen and explained. Moreover, the replacement model gives practically identical results to those obtained by the full Penman's equation.
Keywords: Penman's equation, factorial design, replacement models, design of experiment, regression analysis.

1. INTRODUCTION

Estimating the loss of water from evaporation in a lake or open water body is important for water resources management and planning. The rate of evaporation can be estimated from evaporation pans or from measured climatic variables. Many methods for estimating evaporation from open water based on climatic variables are available and one of the best-known methods is the one developed by (Penman 1948). This method is based on a weighted combination of the aerodynamic and energy balance methods (Chow et al 1988). This method is widely used around the world because it is known to give accurate results, and the data required, namely incoming solar radiation, mean temperature, relative humidity, sunshine hours, and wind speeds are usually available.

However, the estimation of the evaporation using Penman's method is not straightforward. It requires the user to make several assumptions, applying a series of about 10 equations, and reference to at least one table. While nomograms have been developed to simplify the calculations, it is not convenient to use in computer-based hydrological models and is subject to reading error. An example of such a nomogram is given in (Wilson 1990). In addition to the many steps required, it is also difficult to determine from the series of equations, the relative contribution of each climatic variable to the estimated evaporation and how the variables interact with one another. While this is not of primary interest when estimating evaporation in practice, it is of importance from a teaching point of view and in understanding the role played by each of the climatic variable and their possible interaction.

In this paper, a simple replacement (or surrogate) model for the Penman equation is presented. The replacement model is in the form of a linear regression equation developed using design of experiment or DOE methodology. The replacement model is much simpler to use and the interpretation of the coefficients and terms in the model is straightforward. The contribution of each climatic variable (or factor) to the rate of evaporation and the interactions among the variables can be clearly interpreted. In the following sections, Penman's method for estimating the rate of evaporation is first discussed. DOE methodology and how it can be used to develop a replacement model for Penman's equation then follows. Verification of the replacement model for a variety of conditions and how it compares with the full Penman's equation is then presented. Conclusions and references are in the last two sections.

2. PENMAN'S COMBINATION METHOD

Penman (1948) combined the evaporation computed by the aerodynamic method and evaporation computed by the energy balance method to obtain a weighted estimate of the rate of evaporation. The equation is given by:

[1] $E_o=\Delta/(\Delta+\gamma)\ E_r + \gamma/(\Delta+\gamma)\ E_a$

Where: E_o is the evaporation from open water (mm/day), E_r is the evaporation estimated by the energy balance method (mm/day), E_a is the evaporation estimated by the aerodynamic method (mm/day), Δ is the slope of the vapor pressure curve at a particular temperature, and γ is the psychrometric constant. The step and series of equations required to apply Penman's method are given in Appendix A.

Data required to use Penman's method are: (1) incoming solar radiation arriving at the top of the atmosphere, R_A in $MJ/m^2/day$. This is known and tabulated for a given latitude and month of the year (e.g. Wilson 1990; Gupta 2001); (2) the mean air temperature, T in °C; (3) the ratio of actual bright sunshine hours and maximum possible sunshine hours, n/N; (4) the relative humidity, h; and (5) the wind speed at 2 m height, U_2 in m/s. In addition to the series of equations required, several constants in the equations must be assumed as these vary with different regions of the world.

3. DESIGN OF EXPERIMENT (DOE) METHODOLOGY

In engineering, one often-used strategy for experimentation is the one-factor-at-a-time or OFAT approach. The OFAT method was once considered the standard, systematic, and accepted method of scientific experimentation. This method has been shown to be inefficient and in fact can be disastrous (Lye 2002; Czitrom 1999; Montgomery 2012). More efficient methods of conducting experiments based on factorial designs have since been discovered by R.A. Fisher and others. This class of experimental designs includes the general factorial, two-level factorial, fractional factorial, and response surface designs among others. These statistically based experimental design methods are now simply called design of experiment methods or DOE methods. Authors such as (Islam and Lye; 2009; Lye and Hawkins 2012; Wu et al 2012; and Lye et al 2015) provide a variety of examples on the use of DOE methodologies in civil engineering.

DOE is a basically 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. By using DOE, the hypothesis about cause-effect relationships in the system can be tested, quantified, and modelled with the fewest experimental runs. DOE can be used for a wide range of multi-factored experiments in nearly all fields of engineering, science, and business. In general, by using DOE, one can: (i) learn about the process being investigated, (ii) screen for important factors, (iii) determine whether factors interact, (iv) build a mathematical model for prediction, and (v) optimize the response(s), if required (Montgomery 2012).

DOE methods are also very useful as a strategy for building mechanistic models, and they have the additional advantage that no complicated calculations are needed to analyze the data produced from the designed experiment. It has been recognized for many years that the factorial-based DOE is the correct and the most efficient method of conducting multi-factored experiments; they allow a large number of factors to be investigated in few experimental runs. The efficiency stems from using settings of the independent factors that are completely uncorrelated with each other. That is, the experimental designs are orthogonal. The consequence of the orthogonal design is that the main effect of each experiment factor, and also the interactions between factors, can be estimated independent of the other effects (Montgomery 2012). DOE methodology can be applied to physical experiments as well as numerical experiments such as the development of simple replacement models for complex numerical models.

Many industries have recognised this fact and DOE methods are widely used in many industries. Yet it is surprising that almost a century after modern experimental designs were invented they are still not widely taught in engineering programs in our universities (Tiao et al 2000 ; Lye 2016).

3.1 Two-Level (2^k) Factorial Design

In this section, a very useful class of experimental designs, the two-level, k factors, factorial design or 2^k design, which is used herein, is described. 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 estimation of the effect of each factor and their interactions, development of the predictive model, and interpretation of the results are rather simple and straightforward with the two-level factorial design. These can be found in (Montgomery 2012), among many other texts.

3.2 Application of the 2^k factorial design to Penman's Equation

Five variables or factors must be known to use Penman's equation: the incoming solar radiation, R_A, received at the top of the atmosphere; the mean temperature, T; the ratio of actual bright sunshine hours to the maximum possible, n/N; the relative humidity, h; and the wind speed at 2 m height, U_2.

The climatic factors and their operating ranges used in this paper are shown in Table 1. Since there are five factors, one can use a full 2^5 factorial experiment with 32 run combinations or a resolution V, 2^{5-1}, fractional factorial design with 16 runs. However, with a resolution V design, three factor interaction effects are lost due to aliasing with two-factor interaction effects. Since there may be a possibility of three factor interactions with Penman's equation, a full five factor 2-level factorial design with 32 runs will be used. The factor combinations and the estimated evaporation rates (the responses) are shown in Table 2.

Table 1: Operating ranges of the climatic factors

Factor	Name	Low Level	High Level
A	Mean Temperature, T (°C)	20.0	35.0
B	Actual sunshine hours to maximum possible (n/N)	0.1	0.9
C	Relative humidity (h)	0.2	0.9
D	Wind speed at 2 m height, U_2 (m/s)	0.2	5.0
E	Incoming solar radiation, R_A ($MJ/m^2/day$)	20.0	45.0

3.3 Estimation of Factor Effects

From the 32 combinations, 31 effects can be estimated. These are the five main effects, 10 two-factor interaction effects, 10 three factor interaction effects, five four factor interaction effects, and one five factor interaction effect. However, based on the 'sparsity of effects' principle, usually only main and only some two-factor are important. Some two-factor interactions and most three factor interactions and above are negligible and can be ignored (Montgomery 2012). The 31 effects can be estimated by means of a sign table, Yate's algorithm, or by regression analysis. Details are given in (Montgomery 2012). In this paper, Design-Expert Version 10 software (Statease 2016) will be used to estimate the effects, and develop the replacement model. The 31 estimated effects are shown in Table 3 arranged in descending order by absolute effect size.

From Table 3, the most important effects in order of absolute magnitudes are: E (incoming solar radiation), D (wind speed), B (sunshine ratio), CD (interaction of humidity and wind speed), A (mean temperature), BE (interaction of sunshine ratio and solar radiation), C (humidity), BC (interaction of sunshine ratio and humidity), AB (interaction of mean temperature and sunshine ratio), AD (interaction of mean temperature and wind speed), and AE (interaction of temperature and solar radiation). All other effects are negligible and many are zero. Another approach is to use a statistical approach based on the analysis of variance or ANOVA to test for the significance of the effects. Graphical approaches such as a Pareto plot shown in Figure 1 or a normal probability plot of the effects can also be used. However, there is a need to apply common sense and hydrological knowledge to choose effects that are of practical significance.

Table 2: Factor combinations and estimated evaporation from Penman's equation

Factors: Run	A:T (°C)	B: n/N	C: h	$D:U_2$ (m/s)	$E:R_A$ ($MJ/m^2/day$)	E_o (mm/day)
1	20	0.1	0.2	0.2	20	1.228
2	35	0.1	0.2	0.2	20	1.542
3	20	0.9	0.2	0.2	20	1.506
4	35	0.9	0.2	0.2	20	1.975
5	20	0.1	0.9	0.2	20	1.302
6	35	0.1	0.9	0.2	20	1.838
7	20	0.9	0.9	0.2	20	2.365
8	35	0.9	0.9	0.2	20	4.082
9	20	0.1	0.2	5.0	20	4.902
10	35	0.1	0.2	5.0	20	6.550
11	20	0.9	0.2	5.0	20	5.180
12	35	0.9	0.2	5.0	20	6.983
13	20	0.1	0.9	5.0	20	1.761
14	35	0.1	0.9	5.0	20	2.464
15	20	0.9	0.9	5.0	20	2.824
16	35	0.9	0.9	5.0	20	4.708
17	20	0.1	0.2	0.2	45	3.161
18	35	0.1	0.2	0.2	45	3.894
19	20	0.9	0.2	0.2	45	6.015
20	35	0.9	0.2	0.2	45	7.463
21	20	0.1	0.9	0.2	45	3.234
22	35	0.1	0.9	0.2	45	4.190
23	20	0.9	0.9	0.2	45	6.873
24	35	0.9	0.9	0.2	45	9.569
25	20	0.1	0.2	5.0	45	6.835
26	35	0.1	0.2	5.0	45	8.902
27	20	0.9	0.2	5.0	45	9.689
28	35	0.9	0.2	5.0	45	12.471
29	20	0.1	0.9	5.0	45	3.693
30	35	0.1	0.9	5.0	45	4.816
31	20	0.9	0.9	5.0	45	7.333
32	35	0.9	0.9	5.0	45	10.195

Table 3: Estimated absolute effect sizes arranged in descending order

Term	Name	Effect Size
E	R_A, Incoming solar radiation	3.57
D	U_2, Wind speed	2.44
B	n/N, Sunshine ratio	2.43
CD	Interaction of relative humidity and wind speed	1.90 (-)
A	T, mean temperature	1.48
BE	Interaction of sunshine ratio and incoming solar radiation	1.43
C	H, Relative humidity	1.07 (-)
BC	Interaction of sunshine ratio and relative humidity	0.65
AB	Interaction of mean temperature and sunshine ratio	0.47
AD	Interaction of mean temperature and wind speed	0.38
AE	Interaction of mean temperature and incoming solar radiation	0.35

Other effects are negligible or zero.	<0.20

**Note: (-) means that the effect is negative.*

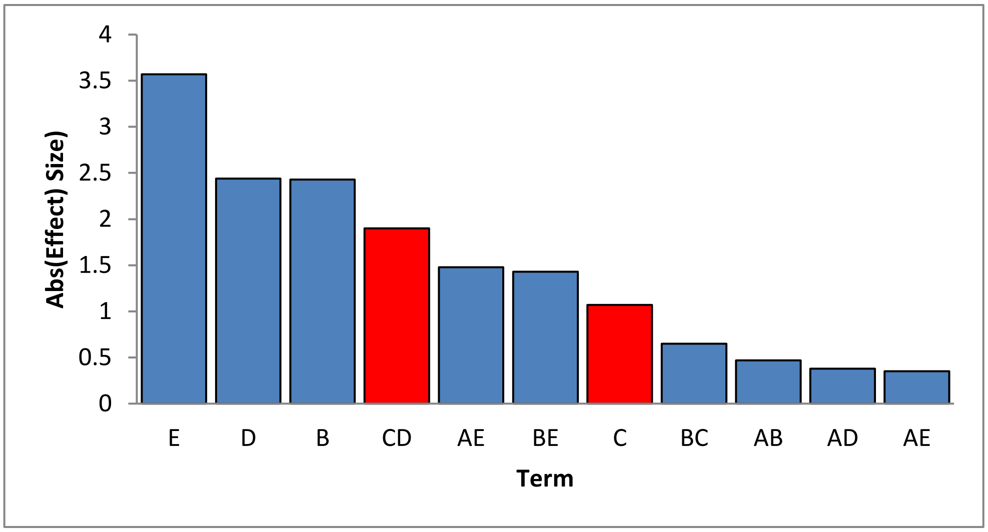

Figure 1: Pareto plot of absolute effect size. (Red = negative effect)

Using the selected effects, one can then develop a regression model as a replacement for Penman's equation. The regression model is in the form:

[2] $$\mathrm{E}_m = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_D D + \beta_E E + \beta_{AB} AB + \beta_{AD} AD + \beta_{AE} AE + \beta_{BC} BC + \beta_{BE} BE + \beta_{CD} CD$$

Where: E_m is the rate of evaporation (mm/day), β_0 is the grand mean of the responses, and β_i are regression coefficients which is equal to half the effect sizes, and the values of the factors are in coded units (-1 and +1). The final equation in coded factor scale after substituting the estimated coefficients and grand mean is:

[3] $$E_m = 4.99 + 0.74\,A + 1.22B - 0.53C + 1.22D + 1.79E + 0.24AB + 0.19\,AD + 0.17AE + 0.32BC + 0.71BE - 0.95CD$$

The regression equation gave an adjusted R^2 of approximately 0.992, and a prediction R^2 of 0.987. The high predicted R^2 means that the regression equation can be used in place of the Penman's equation without much loss of accuracy. In addition, the contribution of each factor and how the factors interact are quantified and easily interpreted. The model term with the largest coefficient has the greatest effect on the estimated evaporation. Negative coefficients indicate a negative effect. That is, as the factor increases, the estimated evaporation decreases, and vice versa. From Table 3, there is a large interaction effect between humidity (C) and wind speed (D). This interaction effect can be explained using an interaction plot shown in Figure 2. From Figure 2, 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, which makes perfect physical sense. If the air is saturated, wind or no wind would not increase evaporation.

The next largest interaction effect is BE, the interaction between sunshine ratio (n/N) and the incoming solar radiation (R_A). The interaction plot is shown in Figure 3. It can be seen that higher amounts of incoming solar radiation produces higher rates of evaporation in general but the effect is more pronounced as the sunshine ratio increases. There is little effect of the sunshine ratio at the lower level of incoming radiation which again makes physical sense. There are also other interactions involving the mean temperature and other terms. But they are less pronounced.

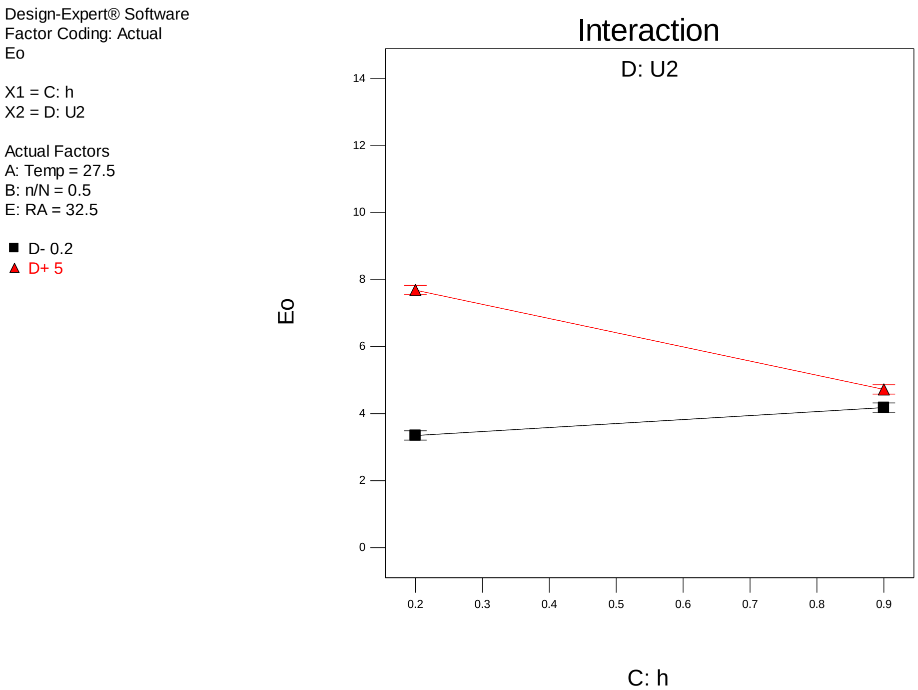

Figure 2: Interaction of relative humidity (h) and Wind speed (U_2).

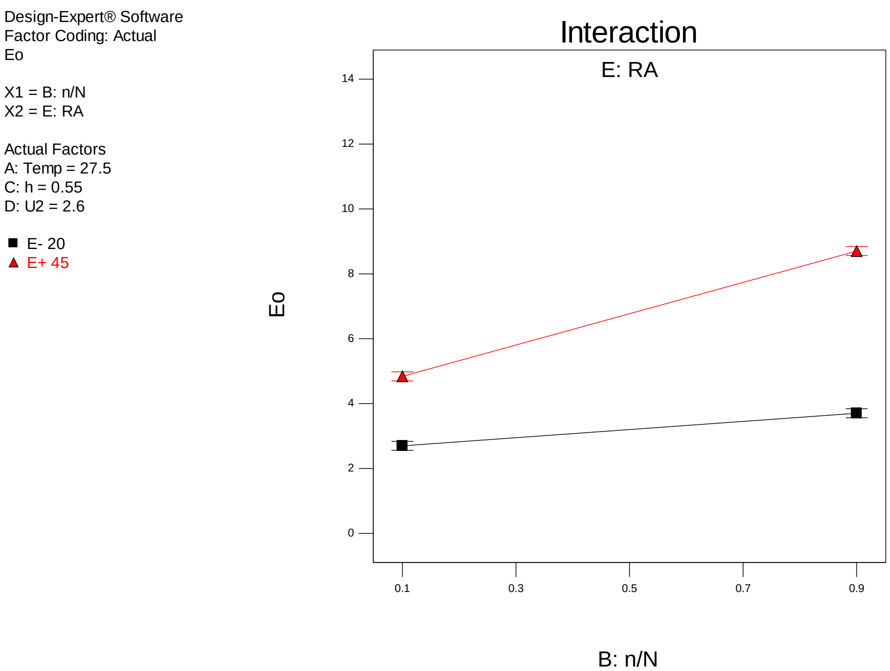

Figure 3: Interaction of sunshine ratio (n/N) and Incoming solar radiation (R_A).

It is important that interpretation of the relative magnitudes and signs of the coefficients are done in coded factor (-1, +1) scale. The signs and coefficients in actual factor scales are not directly comparable because of the different units used. However, in application of the replacement model, the model given by Equation [3] can be converted to actual factor scale giving:

$$[4] \quad E_m = 0.458 - 0.028\,(T) - 5.048\,\left({}^{n}/_{N}\right) + 0.258\,(h) + 0.844\,(U_2) + 0.021(R_A) + 0.079\,(T)\left({}^{n}/_{N}\right) + 0.010(T)(U_2) + 0.002(T)(R_A) + 2.318\left({}^{n}/_{N}\right)(h) + 0.143\left({}^{n}/_{N}\right)(R_A) - 1.130(h)(U_2)$$

Where all terms have been previously defined. Although equations [3] and [4] give the same estimates, only the coefficients in coded scale are proportional to the observed effect.

4. VERIFICATION OF THE REPLACEMENT MODEL

The ultimate test of the replacement model is to compare the model, Equation [4], with that obtained by the full Penman's equation, Equation [1], for a variety of conditions that were not used in the development of the replacement model. The two methods were compared for 100 randomly generated sets of the five climatic factors within the ranges used for each factor. The result of the comparison is shown in Figure 4. From Figure 4, the intercept and slope of the line are practically 0 and 1.0, respectively indicating that the replacement equation is unbiased. The maximum absolute percentage error is 4.24, and the maximum absolute error is about 0.20 mm/day. This shows that the replacement model provides accurate estimates of the rate of evaporation when compared with Penman's equation. Any error will probably be smaller than the probable margin of error in estimating the (n/N) factor, which is normally estimated from the amount of cloud cover as:

$$[5] \quad \frac{n}{N} = (1 - \%C)$$

Where: n/N is defined earlier and %C is the percent of cloud cover during the day (Wilson 1990).

As part of the replacement model development, a second-order model with quadratic terms was also fitted. However, it was found that there is no significant improvement in performance. This means that Penman's equation can be essentially modeled using just a linear model.

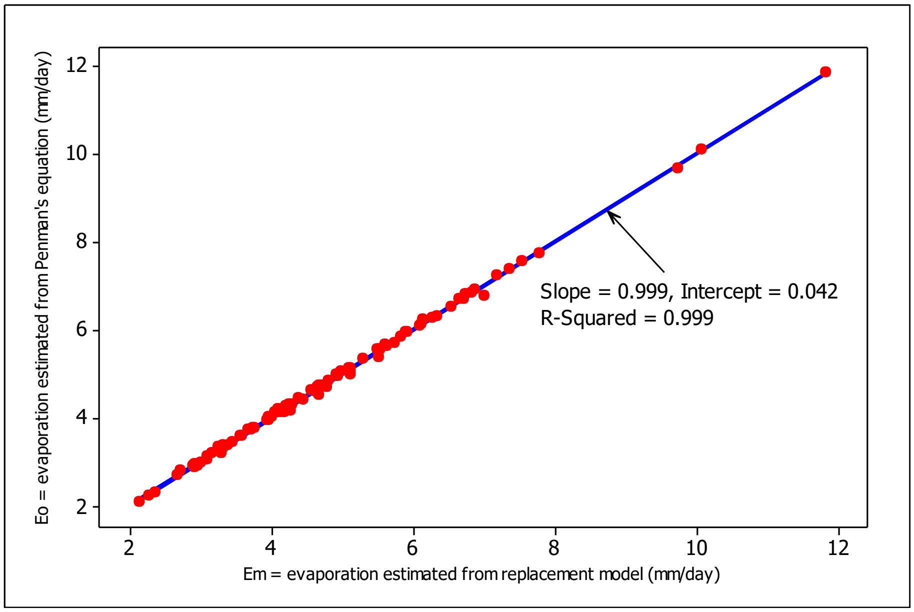

Figure 4: Comparison of estimated evaporation: Penman's equation (E_o) vs. Replacement equation (E_m)

5. CONCLUSION

In this paper, a simple replacement model for Penman's equation was developed using a five factor two-level factorial design. The replacement model is simple to use and the interpretation of the coefficients and the terms in the model is straightforward. The contribution of each climatic factor to the rate of evaporation and the interactions among the factors can be clearly seen and explained. Moreover, the replacement model gave practically identical estimates (to the nearest 0.2 mm/day) as those given by the full Penman's equation. The replacement model is valid for the ranges of the climatic factors that were used to develop the model and if the constants used in the Penman's equation are acceptable. The same procedure can be repeated for other latitudes except that slightly different replacement models would result due to the change in some of the constants used and change in the climatic factor ranges.

6. REFERENCES

Chow, V. T, Maidment, D. R. and Mays, L. W. 1988: *Applied Hydrology*, McGraw-Hill, New York.

Czitrom, V. 1999: One-Factor-at-a-Time versus Designed Experiments. *The American Statistician*, 53:2, pp. 126-131.

Gupta, R. S. 2001: *Hydrology and Hydraulic Systems, 2nd Edition*, Waveland Press, Inc., Illinois, USA.

Islam, I., and Lye, L. M. 2009: Combined Use of Dimensional Analysis and Modern Statistical Design of Experiments in Hydrodynamics Experiments. *Ocean Engineering*, 36(3), 237-247, March.

Lye, L. M., Hussein, A., and Dey, R. 2015: Improved Sensitivity Analysis in Civil Engineering Design Using Design of Experiments, Proceedings of the Annual Conference of the *Canadian Society for Civil Engineering*, Regina, Saskatchewan. 10 pages. Paper in CD-ROM Proceedings.

Lye, L. M., and Hawkins, D. P. 2012: Application of DOE Methodologies to the FEA of a Payload Suspended on Drill Pipe from a Small-Submersible Drill Rig. Proceedings of the *3rd International Structural Specialty Conference*, Edmonton, June. 10 pages. Paper in CD-ROM Proceedings.

Lye, L. M. 2016: Graduate Attribute 3.1.3: Investigations – How civil engineers can possess this attribute. Proceedings of the *Canadian Society for Civil Engineering Annual Conference*, London, Ontario. 10 pages. Paper in CD-ROM Proceedings.

Montgomery, D.C. 2012: *Design and Analysis of Experiments,* 8th Edition, John Wiley and Sons, Inc. N. Y., USA.

Penman, H. L. (1948): Natural Evaporation from Open Water, Bare Soil and Grass, *Proceedings Royal Society (London)*, A, Vol. 193.

Statease, 2016: Design-Expert Version 10.0: Software for Design of Experiments, *Statease Inc*., Minneapolis. USA.

Tiao, G. C., Bisgaard, S., Hill, W. J., Pena, D., and Stigler, S. M. (Editors), 2000: *Box on Quality and Discovery: With Design, Control and Robustness*, John Wiley and Sons, USA.

Wilson, E. M. 1990: *Engineering Hydrology*, 4th Edition, Macmillan. 348 pp, London, England.

Wu, H., Lye, L. M., and Chen, B. (2012): A DOE-Aided Sensitivity Analysis and Parameterization for Hydrological Modelling, *Canadian Journal of Civil Engineering,* Vol. 39, 1-13.

APPENDIX A

Calculation of Open Water Evaporation Using Penman's Equation (Metric Units)

Given data:

1. For a given latitude and month, incoming solar radiation, $\mathbf{R_A}$, is known and tabulated in Gupta (2001) or Wilson (1990). Wilson's table is in gm calories/cm^2/day which has to be converted to MJ/m^2/day by multiplying by 0.0419.
2. Mean daily temperature, **T** °C.
3. Ratio of actual sunshine duration and maximum possible hours, **n/N** (ranges from 0 – 1) or from cloud cover, n/N = 1 - %C. Where %C is the percentage of cloud cover.
4. Relative humidity, **h** (ranges from 0 – 1).
5. Wind speed at 2 m height, $\mathbf{U_2}$ m/s.

Calculation steps:

1. Saturation vapor pressure,
2. Actual vapor pressure of air, $e_a = e_s h$
3. Slope of vapor pressure curve at T °C,
4. Psychrometric constant, where P_{atm} = atmospheric pressure = 101.3 kPa.
5. Latent heat of vaporization,
6. Net incoming radiation absorbed, $R_i = R_A(1-\alpha)\left(a + b\ {}^{n}/_{N}\right)$ MJ/m^2/day

Where: constants a and b are assumed to be 0.25 and 0.54, respectively, α is the albedo assumed to be 0.08 for open water.

7. The long wave radiation emitted, $R_e = \left(c + d\ {}^{n}/_{N}\right)\left(0.34 - 0.14\sqrt{e_a}\right)\sigma\,(T + 273.2)^4$ MJ/m^2/day

Where: c and d are constants assumed to be 0.2 and 0.8, respectively, and σ is the Stefan-Boltzmann constant which is equal to 4.903 x 10^{-9} MJ/m^{2}K^4 day.

The evaporation computed from the energy balance is then given by:

8. $E_r = \frac{R_n}{\lambda \rho_w}$

Where: ρ_w is the density of water (kg/m^3), $R_n = R_i$-R_e, the net radiation available for evaporation in MJ/m^2/day, If λ is expressed in MJ/kg, E_r will be in m/day which when multiplied by 1000 can be converted to mm/day.

Finally, for estimating evaporation using the aerodynamic method, many versions are available. This paper will use the general version given in (Chow et al 1988). It is given by:

9. $E_a = B(e_s - e_a)$

Where: B is the vapor transfer coefficient given by:

10. $B = \frac{0.622\, k^2 \rho_a\, U_2}{P_{atm} \rho_w \left[ln\left({}^{z_2}/_{z_0}\right)\right]^2}$ x 86.4

Where: k is the von Karman constant = 0.4, U_2 is the wind speed (m/s) at 2 m height, ρ_a is the air density (kg/m^3) = 1.18 kg/m^3, P_{atm} the atmospheric pressure (kPa) = 101.3 kPa, ρ_w the water density (kg/m^3) = 997 kg/m^3, z_2 equals 2 m, and z_0 the roughness height at the water surface = 0.0003 m. Once E_r and E_a are obtained; the evaporation from open water using Penman's combination equation can be estimated:

11. $E_o = \frac{\Delta}{\Delta+\gamma} E_r + \frac{\gamma}{\Delta+\gamma} E_a$ mm/day. Where all terms are as previously defined.