# Objective:

The GRACE program generates an optimal correlation between a dependent variable (say, y) and multiple independent variables (say, x1, x2, x3 .....up to x30). This is accomplished through non-parametric transformations of the dependent and independent variables. Non-parametric implies that no functional form is assumed between the dependent and independent variables and the transformations are derived solely based on the data set. The final correlation is given by plotting the transformed dependent variable against the sum of the transformed independent variables. The correlation thus obtained can be shown to be optimal (Breiman and Friedman, 1985; Xue et al, 1996).

# Installation:

You need Windows. Unzip the GRACE.ZIP file by clicking it. The following files will be created: GRACE.EXE (this is the program), perm.dat and pvt.dat (these are sample data files) and GRACE\_2013.XLSM (how to use this file will be discussed later.) You may wish to put this stuff into a separate folder.

# Steps in running GRACE:

(1) Create a data file arranging your data in columns. The first line should contain the names of the columns. If possible, use simple names for columns without spaces and other fancy characters inside (underscore is allowed.) Use space and/or tabulator to separate columns. Sample data files are included. Any data line can be temporarily left out from calculations by putting an asterisk at the beginning.

(2) Start Grace\_2013.xlsm. Click Run Grace.exe Button to start GRACE.

(3) Input your data using the input menu. You can either select the variable itself or for positive data, you can select the natural logarithm of the variable.

(4) Execute calculations using the RUN menu item.

(5) The program generates and plots optimal transformations for the dependent and independent variables. (Several options are available for these transformations. The default option for the dependent variable is 'monotonic' transformation and for the independent variables is simply 'orderable' transformation. You can select the appropriate transformations using the option menu and repeating the RUN command. The monotonic transformation is more restrictive but is necessary if you are interested in back transformations.)

(6) The program generates a plot of transformed dependent variable and sum of transformed independent variables. This is the optimal correlation.

(7) The program generates a plot of observed vs. predicted values of the dependent variable based on the correlation developed. Mean absolute deviation and standard errors are computed. (If you selected the ln of a variable as dependent variable, you will see the prediction of the original variable, i.e. the exp of the ln, as well.)

(8) Finally, the GRACE program generates an EXCEL file (Gracetr.xls) that summarizes the results for use in generating functional forms as discussed below.

# Deriving functional forms:

The non-parametric approach adopted by GRACE program generates a transformed value corresponding to each data point for the dependent and independent variables. However, it does not give you a functional form for these transformations. In order to generate a functional form for the final correlation, you must fit these transformations using appropriate functions. In our experience, simple polynomials are generally good enough to fit these transformations. This is accomplished using the macro in Grace\_2013.xlsm that is provided in the disk.

Clicking the Post Processing Button to start the macro. Specify the directory and point to the output file Gracetr.xls and then click open. The macro will generate the plots and polynomial fits to the transformations. The default order of the polynomials is 2. You may change this by putting the appropriate polynomial orders (3, 4,....) into subsequent empty cells (starting with cell C1) on the first line of the spreadsheet Gracetr.xls. The number on cell C1 will correspond to the order of polynomial for the first variable, the number on cell D1 will correspond to the second variable and so on. Then you can save the spreadsheet Gracetr.xls and click the Post Processing Button to rerun the macro. If you like what you see, save it under a unique name as an EXCEL (.xls) file.

# Variable Selection using Stepwise Regression:

GRACE with stepwise regression can be used to select “best” subset of independent variables and build the “optimal” correlation between selected independent variables and dependent variable. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are often used as selection criteria in stepwise regression for model selection. The AIC and BIC are different ways to penalize the number of independent variables in the model. In our EXCEL file, we calculate both AIC and BIC along with RSS (Residual Sum of Squares). These values can be located on the right of Gracetr.xls spreadsheet.

Stepwise regression includes backward elimination and forward selection. Backward elimination means starting from all the independent variables and eliminating one variable at a time until a stop criteria is met. Forward selection is the reverse of backward elimination. It means starting from zero independent variables and adding one variable at a time until stop criteria. You can choose backward elimination or forward selection or both.

The detailed steps are as follows:

Backward elimination (eliminate variables one by one):

1. In GRACE program, choose all your independent variables (either in natural logarithm or not). And then run EXCEL to get the AIC value. You can find AIC and BIC by moving Gracetr.xls spreadsheet to the right. We call the model with all the independent variables as starting model.
2. Eliminate the first independent variable from the starting model, run GRACE and get AIC value in EXCEL file; Repeat this for all the other independent variables. Choose the model with the minimum AIC value and compare its AIC with AIC of the starting model.
3. If the starting model has larger AIC value, we update the starting model and go back to step 2. If the starting model has smaller AIC value (meaning no smaller AIC is found), we stop backward elimination and choose the variables in the starting model.

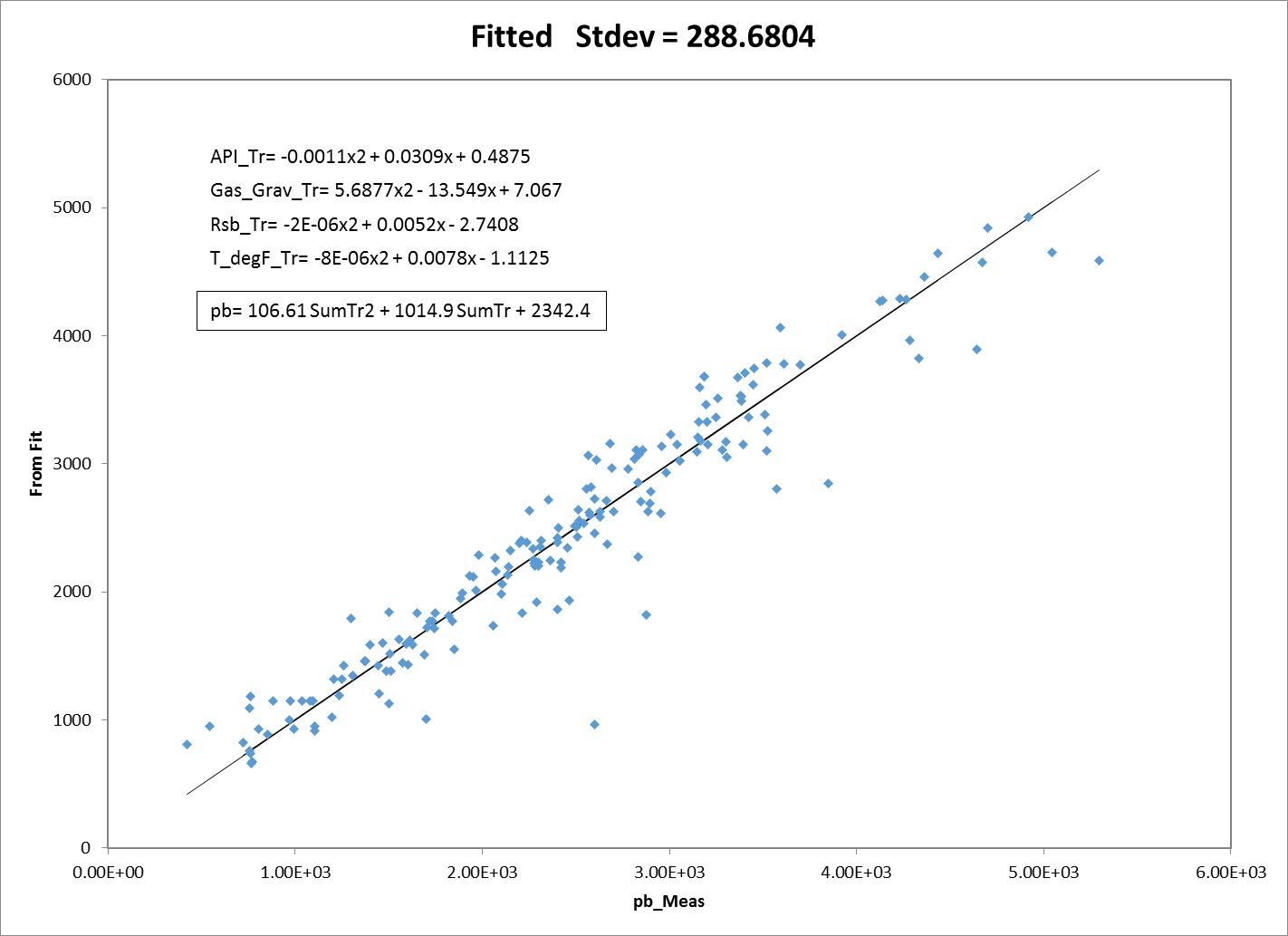
Forward selection (add variables one by one):

1. Choose zero independent variables as the starting model and assume the AIC value is infinity.
2. Add the first non-selected independent variable to the starting model, run GRACE and get AIC value in EXCEL file; Repeat this for all the other non-selected independent variables. Choose the model with the minimum AIC value and compare it with the starting model.
3. If the starting model has larger AIC value, we update the starting model and go back to step 2. If the starting model has smaller AIC value (meaning no smaller AIC is found), we stop forward selection and choose the variables in the starting model.

You can choose either AIC or BIC as your criterion.

# Sample run and functional form:

Run the PVT data set (PVT.dat). Select Pb (the bubble point pressure) as the dependent variable. Select API, Gas gravity, Temperature and Rsb (solution gas ratio) as independent variables. The functional form will be in the ‘Fit’ tab in the Gracetr.xls file as shown in the figure below.



Now given a new set of API, Gas gravity, Temperature and Rsb, first use the equations in the figure above to obtain the corresponding transformations. Sum all the four transformed variables to obtain the variable ‘SumTr’ and use it in the equation to obtain the estimate of Pb.

# Feedback:

Let us know how you like it by sending email to the following address:

[datta-gupta@tamu.edu](mailto:datta-gupta@tamu.edu)

# References:

Breiman, L. and Friedman, J. H., Estimating Optimal Transformations for Multiple Regression and Correlation, Journal of the American Statistical Association (September, 1985) 580.

Xue, G., Datta-Gupta, A., Valkó, P. and Blasingame, T. A., Optimal Transformations for Multiple Regression: Application to Permeability Estimation from Well Logs, SPE/DOE 35412, presented at the Improved Oil Recovery Symposium, Tulsa, OK, April 21, 1996.