Int. J. Heat Mass Transfer. Vol. 33, No. 7, pp. 1566–1567, 1990
Printed in Great Britain

0017-9310/90 $3.00 + 0.00

A statistical procedure for model building in dimensional analysis

CHANG-CHUNG LI and YU-CHENG LEE
Department of Industrial Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China

(Received 21 March 1989 and in final form 26 September 1989)

1. INTRODUCTION

DIMENSIONAL analysis has been used for correlating experimental data of complicated systems in physical and engineering problems. It consists of obtaining a set of dimensionless products through manipulation of the dimensional matrix, and then after taking the logarithm applying the linear regression method to build a model. There have been many research papers on the first part about how to obtain or transform physical quantities into dimensionless products (see Barr [1] for an excellent review). As for the second part, Li and Lee [2] pointed out that the data type should be examined before applying linear regression.

Let $n+1$ be the number of original variables and r be the rank of the corresponding dimensional matrix; then there exists a set of dimensionless products $\pi_0, \pi_1, \pi_2, \ldots, \pi_{n-r}$. A traditional model in dimensional analysis is

$$\pi_0 = K\pi_1^{\beta_1}\pi_2^{\beta_2}\cdots\pi_{n-r}^{\beta_{n-r}}.$$

Unno and Akehata [3] proposed the model

$$\pi_0 = K\pi_1^{\beta_1}\pi_2^{\beta_2}\cdots\pi_{n-r}^{\beta_{n-r}}\left(1+\sum_i \alpha_i\pi_i + \sum_{i,j}\alpha_{ij}\pi_i\pi_j + \sum_{i,j,k}\alpha_{ijk}\pi_i\pi_j\pi_k + \cdots\right)$$

and developed an algorithm based on the group method of data handling (GMDH) by Ivakhnenko [4, 5]. Dovi [6] suggested using the nonlinear regression method for models as

$$\pi_0 = \sum_{j=1}^{m} \beta_{0j} \prod_{i=1}^{n-r} \pi_i^{\beta_{ij}} \tag{1}$$

which is clearly a generalization of Unno's model. Dovi [6] demonstrated his method using data of Singer and Wilhelm [7] concerning heat transfer in packed beds by assuming m in (1) to be 3. In this paper the consideration that m is a known fixed integer is relaxed. We propose to build a model step by step. Terms of dimensionless products are entered into the model sequentially if at each step the value of a test statistic based on the likelihood ratio test is significant at a prespecified level.

2. ANALYSIS

Stepwise-type regression is a well-known technique for model building in statistics. It can be classified into three broad categories: (i) forward selection, (ii) backward elimination and (iii) stepwise regression, which is a popular combination of procedures (i) and (ii). Here only the forward selection method is considered because we do not have a list of candidate regressors as considered in usual regression analyses.

Regressors are now terms of dimensionless products. The forward selection procedure begins with the assumption that there are no regressors in the model. An effort is made to find a best fit regression equation by inserting regressors into the model one at a time. This procedure continues if the calculated value of a test statistic is significant. Otherwise the procedure terminates and a fitted model is obtained.

Assume that at the mth step we have the equation

$$\hat{\pi}_0 = \sum_{j=1}^{m} \hat{\beta}_{0j} \prod_{i=1}^{n-r} \pi_i^{\hat{\beta}_{ij}}$$

where $\hat{\beta}_{ij}$ are the estimates of β_{ij} in (1) and can be obtained by the nonlinear regression method (NLR).

Now, if we are adding another term to the model, then the fitted equation will be

$$\hat{\pi}_0 = \sum_{j=1}^{m+1} \hat{\alpha}_{0j} \prod_{i=1}^{n-r} \pi_i^{\hat{\alpha}_{ij}}.$$

The contribution of the $(m+1)$th term can be tested using a test statistic based on the likelihood ratio theory in statistics. That is, let n_0 be the data points and

$$SSE(m) = \sum_{k=1}^{n_0}\left(\pi_{0k} - \sum_{j=1}^{m}\hat{\beta}_{0j}\prod_{i=1}^{n-r}\pi_{ik}^{\hat{\beta}_{ij}}\right)^2$$

and

$$SSE(m+1) = \sum_{k=1}^{n_0}\left(\pi_{0k} - \sum_{j=1}^{m+1}\hat{\alpha}_{0j}\prod_{i=1}^{n-r}\pi_{ik}^{\hat{\alpha}_{ij}}\right)^2$$

be the computed residual sum of squares in the mth and $(m+1)$th steps; then the natural logarithm of the likelihood ratio

$$LR = n_0 \cdot \ln\,(SSE(m)/SSE(m+1)) \tag{2}$$

is proved to be a Chi-square distributed random variable asymptotically with $(m+1)\cdot(n-r+1)$ degrees of freedom. All we have to do is look into the Chi-square table in any statistics textbook or statistical handbook (e.g. CRC Table) and compare the value of LR in (2) with the tabulated value for $(m+1)\cdot(n-r+1)$ degrees of freedom. If the computed value is bigger than the tabulated value, then the $(m+1)$th term should be retained, otherwise the $(m+1)$th term should not enter the model and the procedure terminates. This procedure will be illustrated with examples in the next section.

To compute $\hat{\beta}_{ij}$, we use a nonlinear regression technique which needs a set of initial values. Good initial values, which are essential for the algorithm to converge, can be obtained by physical insight. In practice, a few different initial values are needed.

3. EXAMPLE

Singer and Wilhelm [7] discussed the transfer of heat in beds of packed solids with heat flowing through the container wall into a fluid flowing through the bed. They obtained a set of data corresponding to the modified Reynolds number Re' and modified Peclet number Pe' from their laboratory for two sizes ($D_p/D_t = 0.0486$ and 0.0236) of Socony-Vacuum Co. bead catalyst as the heat-exchanger packing. We demonstrate our procedure by building a model for the case $D_p/D_t = 0.0486$. Graphic results are shown in Fig. 1. Similar results for $D_p/D_t = 0.0236$ are given in Fig. 2. All

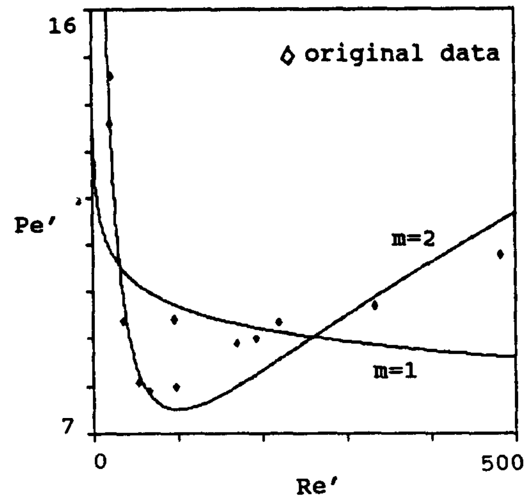

FIG. 1.

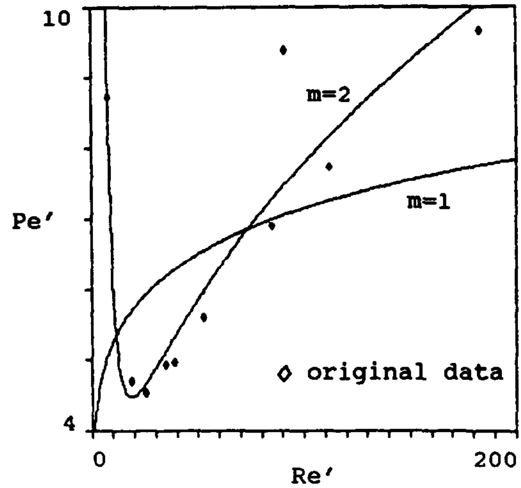

FIG. 2.

analyses were done in CDC CYBER 840 at the National Tsing Hua University using package BMDP3R. Our procedure in detail is as follows.

Step 0. $m = 0$, the total sum of squares for Pe' is 50.4.

Step 1. $m = 1$, the tentative model is $Pe' = 13.7 \cdot Re'^{-0.075}$ with $SSE(1) = 40.3$; but SSE is too large so we continue.

Step 2. $m = 2$, the model becomes $Pe' = 0.5 \cdot Re'^{0.5} + 250 \cdot Re'^{-1.0}$ with $SSE(2) = 8.65$, then the log likelihood ratio LR is $12 \cdot \ln(40.3/8.65) = 18.46 > 9.5$, the tabulated value for Chi-square distribution with four degrees of freedom at the 0.05 level, so we adopt this model and continue our procedure to see if the third term is necessary.

Step 3. $m = 3$, then $Pe' = 0.5 \cdot Re'^{0.53} + 250 \cdot Re'^{-1.0} - 0.0003 Re'^{1.5}$ with $SSE(3) = 5.44$; so LR is $12 \cdot \ln(8.65/5.44) = 5.6 < 12.59$, the tabulated value for Chi-square distribution with six degrees of freedom at the 0.05 level; therefore we decide that the model in step 2 is the 'best' model in a statistical sense.

From Fig. 1, we observe that the residual for obs. 6 is 1.91 which accounts for $1.91^2/8.6 = 42\%$ of $SSE(2)$. Also, the predicted value deviates from the observed value by about 20%. Therefore, this point should be double checked before data analysis and/or model building. A similar comment can be made on obs. 8 in Fig. 2.

4. CONCLUSIONS

From Unno's model to Dovi's model it is clear that traditional models containing only one term of dimensionless products are not enough for model building. Neither should the number of terms be fixed. This number should be unknown and can be determined statistically. However, we should point out that the Chi-square distribution based on the likelihood ratio is for large sample sizes. Therefore for a small sample, as in our examples, the test result should be regarded as a reference. If a powerful graphics software package like STATGRAPHICS is available, then a better judgement can be reached by statistical tests and graphics displays together.

Acknowledgement—This research was supported by the National Science Council of the Republic of China under the contract NSC 78-0415-E007-09.

REFERENCES

1. D. I. H. Barr, Consolidation of basics of dimensional analysis, *Int. J. Engng Mech., ASCE* **110**, 1357–1375 (1984).
2. C.-C. Li and Y.-C. Lee, Computational aspects of dimensional analysis, *Int. Commun. Heat Mass Transfer* **16**, 315–321 (1989).
3. H. Unno and T. Akehata, Simultaneous estimation of functional relationship among non-dimensional variable groups and its coefficients in analyzing experimental data, *J. Chem. Engng Japan* **13**, 173–177 (1980).
4. A. G. Ivakhnenko, The group method of data handling in optimal control with accumulation of information, *Sov. Aut. Control* **14**(5), 57–60 (1969).
5. A. G. Ivakhnenko and Y. V. Koppa, Regularization of decision functions in the group method of data handling, *Sov. Aut. Control* **15**(2), 28–37 (1970).
6. V. G. Dovi, Efficient correlation techniques in dimensional analysis, *Int. Commun. Heat Mass Transfer* **11**, 545–551 (1984).
7. E. Singer and R. H. Wilhelm, Heat transfer in packed beds, *Chem. Engng Prog.* **46**(7), 343–357 (1950).