2nd International Symposium on Computational Mechanics (ISCM II)
12th International Conference on Enhancement and Promotion of Computational Methods in Engineering and Science (EPMESC XII)
November 30 – December 3, 2009
Hong Kong – Macau Paper no. ISCM2-265

Comparison of Different Correlating Methods for the Single-Phase Heat Transfer Data in Laminar and Turbulent Flow Regions

H. K. Tam[a, *], L. M. Tam[a,b], A. J. Ghajar[c] and C. U. Lei[a]

[a] *Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau, Macau, China (*hktam@umac.mo)*
[b] *Institute for the Development and Quality, Macau, China*
[c] *School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, Oklahoma, USA*

Abstract. The experimental heat transfer results, typically expressed as Nusselt number, are usually put in the form of an empirical correlation. Different correlating methods can be used to establish the correlation, such as the traditional least-squares method, artificial neural networks (ANN), and symbolic regression, to name a few. In this study, ANN and symbolic regression will be compared with the traditional least-squares method. Correlations using ANN and symbolic regression were developed first and through contribution analysis we were able to evaluate which correlating method can directly extract more physical meaning. Correlations were developed based on the single-phase heat transfer data [1]. The accuracy of these new correlations is comparable to the traditional least-squares correlations. From the contribution analysis, the importance of each independent variable in the ANN correlations can be examined by the index of contribution [2]. For symbolic correlation, the insignificant variable can be eliminated from the correlation. Therefore, both methods in addition to providing accurate predictions of the heat transfer data also provide correct physical insight as to the contribution of the input variables.

Keywords: Heat Transfer Correlation, Artificial Neural Networks (ANN), Symbolic Regression.
PACS: 44.15.+a

INTRODUCTION

Heat transfer inside horizontal tubes in the laminar, transitional and turbulent flow regimes have been studied experimentally by various researchers in the past. Usually, the experimental results are presented in the form of heat transfer correlations. The form of the correlations is based either on different theoretical models or they are completely empirical based. The coefficients of the correlations are usually determined by the conventional least squares method. Kakac et al. [3] and Tam and Ghajar [4] documented some of the most well accepted correlations in the above-mentioned flow regimes. Ghajar et al. [5] used an unconventional method, the artificial neural network (ANN), and successfully correlated the transitional heat transfer experimental data of Ghajar and Tam [1] for three different inlet configurations. Their ANN heat transfer correlation when compared with the conventional least squares based correlation of Ghajar and Tam [1] showed to be more accurate [5]. However, as briefly discussed in Ghajar et al. [5], the ANN method provides little explanation as to the relative importance of the independent variables that appear in these correlations. Therefore, Tam and his coworkers [2] developed contribution analysis with the index of contribution for indication of the importance of the independent variables. That method was successfully applied to the laminar and turbulent flow data [1] and provided physically realistic results.

Symbolic regression is a method to seek a good mathematical expression, in symbolic form, for correlating the independent variables and the associated dependent variables. Koza [6] proposed to use the genetic programming (GP) for symbolic regression. GP involves the evolutionary computation techniques and the "tree-structured representation" into the computational process. Such as GA [7], the optimized variables (or the best-fit correlation form in symbolic regression) can be obtained through the process of evolution, in which the selection process, the mutation, and the crossover are taken into account, to extremize the fitness function. In recent studies [8, 9], the symbolic regression were used for the heat transfer problems. The accuracy of prediction of the symbolic correlations was shown to be better than the traditional correlations.

The objective of this study was to compare different correlating methods (ANN and symbolic regression) with the traditional method (least-squares), using heat transfer data of [1] in laminar and turbulent regions. Moreover, the contribution analysis was used to evaluate whether the new correlations can predict physically realistic heat transfer results.

EXPERIMENTAL DATASET AND TRADITIONAL CORRELATIONS

The heat transfer experimental data used in this study, along with a detailed description of the experimental apparatus and procedures used, were reported in Ghajar and Tam [1]. A schematic of the overall experimental setup used for heat transfer measurements is shown in Figure 1. The local forced and mixed convective measurements were made in a horizontal stainless steel circular straight tube under uniform wall heat flux boundary condition and three types of inlet configurations (re-entrant, square-edged, and bell-mouth). Thermocouples were placed on the outer surface of the tube wall at close intervals near the entrance and at greater intervals further downstream. Twenty-six axial locations were designated, with four thermocouples placed at each location. The inside wall temperatures and the local heat transfer coefficients were calculated [10]. As reported by Ghajar and Tam [1], the maximum uncertainty for the heat transfer coefficient calculation was 9% and the heat balance error was less than 5%.

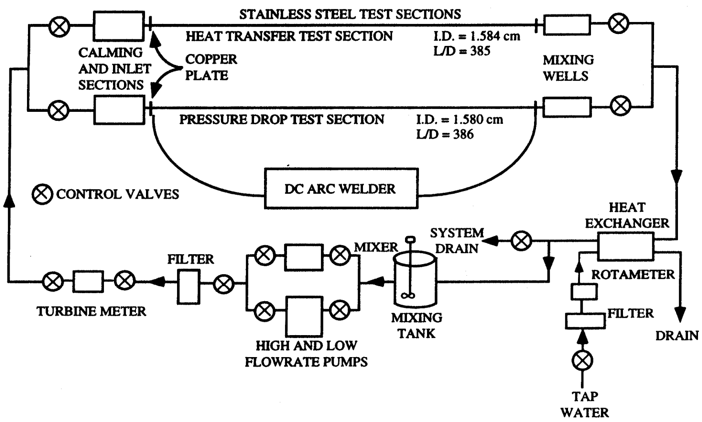

FIGURE 1. Schematic diagram of experimental setup.

The application of heat to the tube wall produces a temperature difference in the fluid. This temperature difference may produce a secondary flow (vortex-like flow) due to free convection. The superposition of forced convection (primary flow) and the free convection (secondary flow) is called the mixed convection. According to [11], the mixed convection only appeared in the laminar heat transfer data. Therefore, it can be concluded that the laminar Nusselt number (dimensionless heat transfer coefficient) is a function of five dimensionless variables, which are Reynolds number (Re), Prantl number (Pr), Grashof number (Gr), length-to-diameter ratio (x/D), and bulk-to-wall viscosity ratio (μ_b/μ_w). The ranges of dimensionless variables considered in this study for laminar flow are $13 \leq Nu \leq 65$, $280 \leq Re \leq 3800$, $40 \leq Pr \leq 160$, $1000 \leq Gr \leq 2.8\times10^4$, $3 \leq x/D \leq 192$, and $1.2 \leq \mu_b/\mu_w \leq 3.8$. Ghajar and Tam [1] collected a total of 546 data points for the forced and mixed convection in the entrance and fully developed flow regions. The experimental data were then correlated in a form similar to the one proposed by Martinelli and Boelter [12].

$$Nu_l = 1.24\left[(RePrD/x) + 0.025(GrPr)^{0.75}\right]^{1/3}(\mu_b/\mu_w)^{0.14} \tag{1}$$

For contribution analysis [2], the dimensionless numbers are grouped into the following form,

$$Nu_l = 1.24\left[(Gz) + 0.025(Ra)^{0.75}\right]^{1/3}(\mu_b/\mu_w)^{0.14} \tag{2}$$

where Gz denotes the Graetz number (for entry length effect) and Ra denotes the Rayleigh number (for buoyancy effect). As shown in Table 1 for the traditional method, the total number of data points for forced convection (212 data points) are predicted by Eq. (1) within -15.4% to +16.9% and those for mixed convection (334 data points) are predicted within -13.17% to +16.5%. Around 86% and 84% of the forced and mixed convection data are predicted with less than ±10% deviation, respectively.

Regarding the turbulent region, the turbulent Nusselt number is a function of the dimensionless variables, Re, Pr, x/D, and μ_b/μ_w. Unlike the laminar region, Gr is not considered here since the buoyancy effect is not important in the turbulent region. The ranges of the dimensionless variables considered in this study for turbulent flow are summarized as follows: $52.3 \le Nu \le 242.4$, $7{,}000 \le Re \le 49{,}000$, $4.0 \le Pr \le 34.0$, $3.2 \le x/D \le 173.1$, and $1.1 \le \mu_b/\mu_w \le 1.7$. Similar to the correlation form proposed by Sieder and Tate [13], Ghajar and Tam [1] developed the following correlation for their turbulent forced convection data in the entrance and fully developed regions for all three inlets:

$$Nu_t = 0.023 Re^{0.8} Pr^{0.385} (x/D)^{-0.0054} (\mu_b/\mu_w)^{0.14} \quad (3)$$

A total of 604 data points were used to develop the above correlation, correlating 100% of experimental data with less than ±11% deviation and 73% of the measured data with less than ±5% deviation. The details for how well Eq. (3), traditional method, predicted the experimental data are given in Table 2.

RESULTS OF ANN METHOD AND SYMBOLIC REGRESSION

To correlate the data in general, the ANN with single hidden layer is employed. Figure 2 shows a typical example of ANN model of this kind. The weight and the bias of the optimal ANN model are usually determined by the back propagation algorithms [14]. In order to determine the contribution of each independent variable to the correlation, the matrix form of the optimal ANN model as shown by Equation (4) has to be examined first.

$$N(p_1,\cdots,p_R) = [w_1^2,\cdots,w_S^2]\begin{bmatrix} f(\sum_{j=1}^{R} w_{1j}^1 p_j + b_1^1) \\ f(\sum_{j=1}^{R} w_{2j}^1 p_j + b_2^1) \\ \vdots \\ f(\sum_{j=1}^{R} w_{Sj}^1 p_j + b_S^1) \end{bmatrix} + b_1^2 \quad (4)$$

where $(p_1,\cdots,p_R)$ are the R inputs, S is the number of hidden neurons, $f(t) = (1+e^{-t})^{-1}$ is the transfer function and w's and b's are the weights and biases of the ANN, respectively.

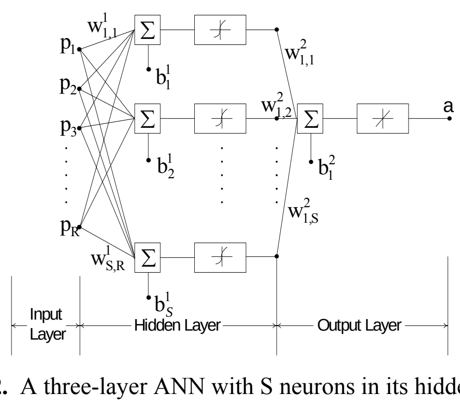

FIGURE 2. A three-layer ANN with S neurons in its hidden layer.

Based on the contribution analysis [2], the contribution of the independent variables p_j to the output of the k^{th} neuron in the hidden layer $f(\sum_{j=1}^{R} w_{kj}^1 p_j + b_k^1)$ is simply $|w_{kj}^1|$ and the relative contribution of the k^{th} neuron output to the ANN model is defined as $Q_k = |w_k^2| / \sum_{i=1}^{S} |w_i^2|$. Therefore, the contribution of the independent variables p_j to the ANN model is $P_j = \sum_{k=1}^{S} Q_k * |w_{kj}^1|$. Finally, in order to compare with the other independent variables, the index of contribution of p_j is defined to be $\text{index}(p_j) = (P_j / \sum_{i=1}^{R} P_i) \times 100\%$. Hence, the most significant independent variable would have the largest index of contribution. On the other hand, the variable with small index appears to be less important.

For contribution analysis, Tam et al. [2] obtained the stable and correct results for turbulent flow data when the gradient method chosen was steepest descent algorithm (SDA). The number of iterations used was 10,000, and the number of neurons used was taken as 6. The form of the ANN correlation used is given by Equation (4). For reliability purposes, ninety percent of the 604 data points were used for training and the remaining data was involved for verification. The initial values of the free parameters (weights and biases) were randomly selected within ±1. For satisfying the log-sigmoid transfer function, the normalized input variables, Re, Pr, x/D, μ_b/μ_w, were arranged into the input vector, **p**. In Equation (4), the dependent output is the Nusselt number for the turbulent heat transfer data. As shown in Table 2 for the ANN method, all the experimental data were predicted within -10.6% to +9.9%. Nearly 100% of all the data (603 data points) were predicted within ±10% deviation. As compared to the results given by Equation (3), good accuracy in predicting the turbulent data was maintained by using the ANN correlation. For the calculation of the index of contribution for each variable, the above-mentioned contribution analysis was employed according to the weight matrices, $\mathbf{w}^1$ and $\mathbf{w}^2$. The index of contribution for Re, Pr, x/D, μ_b/μ_w is 34.5%, 36.7%, 10.1%, and 18.7%, respectively. The results are physically realistic because the Re and Pr are the major variables whatever the flow regime is and the remaining terms are less important due to their purpose as the correction factors in the correlation.

For the analysis of the laminar flow heat transfer data, the 546 data points for the three different inlet configurations were divided into two sets of forced convection data (212 data points) and mixed convection data (334 data points). In the ANN training, the SDA, 50,000 iterations, and 6 neurons were selected [2]. The normalized input variables, Gz, Ra, and μ_b/μ_w, are arranged into the input vector, **p**. As shown in Table 1 for the ANN method, the total number of data points for forced convection is predicted within -12.2% to +17.0% and those for mixed convection are predicted within -11.1% to +8.5%. Over 97% of the data are predicted within ±10% deviation for the two convection modes. As compared to the results given by Equation (1) and presented in Table 1, significant improvement is observed. More data were predicted within ±10%. For forced convection data, the index of contribution for Gz, Ra and μ_b/μ_w was 53.9%, 9.2%, and 36.9%. For mixed convection data, the index of contribution for each variable was 46.8%, 25.7%, and 27.5%. From the above values of the index of contribution, it can be seen that for the forced and mixed convection data, Gz is the most important dimensionless parameter since the thermal entry length is long and most of the data are in the entry length region. On the other hand, Ra is the least important dimensionless parameter in forced convection since near the tube entrance; the secondary flow effect has not yet developed. Compared to the forced convection data, Ra in mixed convection is important since the secondary flow effect has established. For both forced and mixed convection cases, the viscosity ratio is important due to the significant effect of variation of physical properties. Therefore, it can be concluded that the index of contribution calculated completely agrees with the physics of the problem.

TABLE (1). Prediction results for the laminar flow correlations (forced and mixed convection mode) developed based on different correlating methods.

Correlating methods (Total of 546 data points)	Forced Convection (212 data points)				Mixed Convection (334 data points)			
	No. of Data within ±10%	No. of Data within ±5%	Abs. Mean Dev. (%)	Range of Dev. (%)	No. of Data within ±10%	No. of Data within ±5%	Abs. Mean Dev. (%)	Range of Dev. (%)
Traditional	183	113	5.4	-15.40% to +16.9%	282	148	6.0	-13.70% to +16.5%
ANN	207	164	3.6	-12.20% to +17.0%	333	308	2.4	-11.10% to +8.5%
Symbolic regression	200	144	4.3	-13.80% to +17.6%	325	236	3.7	-16.70% to +10.9%

TABLE (2). Prediction results for the turbulent flow correlations developed based on different correlating methods.

Correlating methods (Total of 604 data points)	No. of Data within ±10%	No. of Data within ±5%	Abs. Mean Dev. (%)	Abs. Range of Dev. (%)
Traditional	600	439	4.4	-10.3% to +10.6%
ANN	603	492	3.0	-10.6% to +9.9%
Symbolic regression	599	444	3.4	-10.8% to +10.1%

For symbolic regression, the linear-in-parameters correlation form proposed by [15] will be used in this study:

$$y(p_1,\cdots,p_R) = a_0 + \sum_{i=1}^{M} a_i F_i(p_1,\ldots,p_R), \qquad (5)$$

where $(p_1,\cdots,p_R)$ are the R inputs, $(F_1,\ldots, F_M)$ are the M nonlinear functions, and $a_0,\ldots,a_M$ are the correlation parameters. Using Equation (5), the correlation parameters, $a_1,\ldots, a_M$, can be computed by the traditional gradient method. Prior to calculation of the parameters, the symbolic functions, $F_1(p_1,\cdots,p_R)$, … , $F_M(p_1,\cdots,p_R)$, represented in tree structure (see Figure 3), was found by GP [6]. In this study, the basic arithmetic operators {+, -, *, /} and mathematical functions {$\log_{10}$, sqrt} are used in the tree structure. By the evolutionary process, the optimized correlation can be obtained and it gives out the minimum value of the fitness function, i.e., the mean square error, $MSE = \frac{1}{N}\sum_{k=1}^{N}\left(y_{exp} - y(p_1,\cdots,p_R)\right)$, between the calculated and the experimental output values.

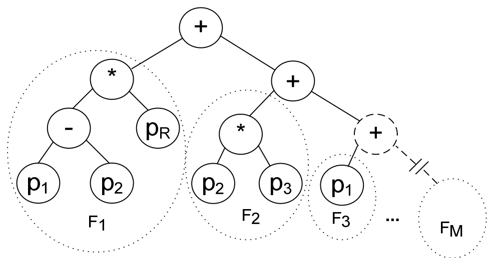

FIGURE 3. Representation of linear combination of the functions $(F_1,\ldots, F_M)$ in tree structure.

In this study, the evolutionary parameters such as the population size, the maximum number of generations, the type of selection, the generation gap, the probability of crossover, the probability of mutation, and tree depth are necessary to be pre-determined. The type or numeric value of those parameters was 80, 50, roulette-wheel type, 0.9, 0.5, 0.5, and 3, respectively. For reliability purposes, ninety percent of the total data points were also used for training and the remaining data was used for verification. For turbulent data, the Re, Pr, x/D, μ_b/μ_w were treated as the input variables. With the minimum MSE, the best correlation, which is Eq. (6), for turbulent data was obtained from 50-times trials,

$$Nu_t = 17.18 + 2.263\times10^{-3}\times Re\times Pr^{0.5} - 1\times10^{-7}Re^2 \qquad (6)$$

In Equation (6), the most important variables, Re and Pr numbers, were kept in the correlation. Other variables like x/D and μ_b/μ_w were taken out from that correlation automatically. As shown in Table 2 for the symbolic regression method, all the experimental data were predicted within -10.8% to +10.1%. The absolute deviation of all the predictions is around 3.4%. Around 99% of all the data (599 data points) were predicted within ±10% deviation. As compared to the traditional and ANN correlations, the predictions given by Equation (6) also have good accuracy.

For the laminar data, the above-mentioned evolutionary parameters were also used for establishment of the symbolic correlations. The input variables were Gz, Ra, μ_b/μ_w. From 50-times trials, the best correlation for laminar forced convection data was obtained,

$$Nu_{l,force} = 10.49 - 0.149[(\mu_b/\mu_w)^2 - \sqrt{(Gz)(\mu_b/\mu_w)}] \qquad (7)$$

and the best correlation for laminar mixed convection data was represented,

$$Nu_{1,mix} = 5.566 + 0.189\left[Gz^{0.5} + (\mu_b/\mu_w)\right] + 0.00235\left[Ra^{0.5} \times (\mu_b/\mu_w)\right] \quad (8)$$

As shown in Table 1, the total number of data points for forced convection was predicted within -13.8% to +17.6% and those for mixed convection were predicted within -16.7% to +10.9%. About 94% of all forced and mixed convection data were predicted within ±10% deviation. As shown in Table 1, more data were predicted accurately by Equations (7) and (8) than the traditional correlation. Through the symbolic regression, the variables, Gz and μ_b/μ_w, were kept in the Equations (7-8). Those variables are important for the heat transfer coefficient. However, the insignificant variable, Ra number, was also eliminated from the forced convection correlation. The reduction of Ra number in forced convection absolutely matched with the physics of the problem as discussed before.

CONCLUSIONS

Based on the heat transfer data in laminar and turbulent regions, the ANN and symbolic regression correlations were compared with the traditional correlations. The accuracy of these new correlations is comparable to the traditional least-squares correlations. From the contribution analysis, the importance of each independent variable in the ANN correlation can be examined by the index of contribution method. For symbolic correlation, the insignificant variable can be eliminated from the correlation. Therefore, both methods in addition to providing accurate results also provide insight as to the contribution of the input variables.

ACKNOWLEDGMENTS

This research is supported by the Fundo para o Desenvolvimento das Ciencias e da Tecnologia under project no. 033/2008/A2.

REFERENCES

1. Ghajar AJ, Tam LM**,** Heat transfer measurements and correlations in the transition region for a circular tube with three different inlet configurations, *Experimental Thermal and Fluid Science*, 1994, **8**, 79-90.
2. Tam LM, Ghajar AJ, Tam HK, Contribution Analysis of Dimensionless Variables for Laminar and Turbulent Flow Convection Heat Transfer in a Horizontal Tube Using Artificial Neural Network, *Heat Transfer Engineering*, 2008, **29**(**9**), 793-804.
3. Kakac S, Shah RK, Aung W, *Handbook of Signle-Phase Convective Heat Transfer*, Wiley, New York, 1987.
4. Tam LM, Ghajar AJ, Transitional Heat Transfer in Plain Horizontal Tubes, *Heat Transfer Engineering*, 2006, **27**(**5**), 23-38.
5. Ghajar AJ, Tam LM, Tam SC, Improved heat transfer correlation in the transition region for a circular tube with three inlet configuration using artificial neural network, *Heat Transfer Engineering*, 2004, **25**(**2)**, 30-40.
6. Koza JR, *Genetic Programming Paradigm, On the Programming of Computers by Means of Natural Selection,* The MIT Press, Cambridge, MA, 1992.
7. Holland J, *Adaptation in Natural and Artificial Systems,* MIT Press, 1975.
8. Cai W, Pacheco-Vega A, Sen M, Yang KT, Heat transfer correlations by symbolic regression, *International Journal of Heat and Mass Transfer*, 2006, **49**, 4352–4359.
9. Zdaniuk GJ, Luck R, Chamra LM, Linear correlation of heat transfer and friction in helically-finned tubes using five simple groups of parameters, *International Journal of Heat and Mass Transfer*, 2008, **51**, 3548-3555.
10. Ghajar AJ**,** Kim J, Calculation of Local Inside-Wall Convective Heat Transfer Parameters from Measurements of the Local Outside-Wall Temperatures along an Electrically Heated Circular Tube, *Heat Transfer Calculations,* edited by Myer Kutz, McGraw-Hill, New York, NY, pp. 23.3-23.27, 2006.
11. Ghajar AJ, Tam LM, Flow Regime Map for a Horizontal Pipe with Uniform Heat Flux and Three Different Inlet Configurations, *Experimental Thermal and Fluid Science*, 1995, **10**(**3**)287-297.
12. Martinelli RC, Boelter LMK, The Analytical Prediction of Superposed Free and Forced Viscous Convection in a Vertical pipe, *Univ. Calif. Publ. Eng.,* 1942, **5**(**2**), pp.23.
13. Sieder EN, Tate GE, Heat Transfer and Pressure Drop in Liquids in Tubes, *Ind. Eng. Chem.*, 1936, **28**, 1429–1435.
14. Rumelhart DE, Hinton GE, Williams RJ, *Learning Internal Representations by Error Propagation, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Eds: D. E. Rumelhart, and J. L. McClelland, Vol. 1*, MIT Press, Cambridge, Mass., pp. 318-362, 1986.
15. Madár J, Abonyi J, Szeifert F, Genetic programming for identification of nonlinear input-output models, *Ind. Eng. Chem. Res.*, 2005, **44**, 3178-3186.