# Support Vector Regression (SVR) Model to Predict the Boiling/Non-boiling Length of the Heated Tube in a Vertical Tube Thermosiphon Reboiler

**Sadaf Zaidi**

*Aligarh Muslim University, Aligarh*
*Department of Chemical Engineering, Aligarh Muslim University, Aligarh-202002 (U.P)*

*Abstract:* **Conventionally, the thermosiphon reboilers are being widely employed in petroleum, chemical and petrochemical industries as vaporizers, evaporators and reboilers. The hydrodynamics and heat transfer in a thermosiphon reboiler interact with each other making the process very complex. Prediction of the rates of heat transfer and thermally induced flow are the primary requirements for the design of thermosiphon reboilers. Some empirical models have been developed and published in literature by a number of workers for the prediction of the length of the boiling/non-boiling section of the heated tube. However, they suffer from critical infirmities. Thus, with the above observation in perspective and given the recent developments in the application of artificial intelligence techniques, the state-of-the-art technique called support vector machine (SVM) or specifically the SVM variant called Support Vector Regression (SVR) was employed for the first time to develop a model for predicting the length of the boiling/non-boiling section of the heated tube in a thermosiphon reboiler, for different single component liquids with wide variation in thermo physical properties and operating parameters. The choice of this technique was prompted by its many attractive features. Moreover, the extensive survey of the literature revealed that very little work had been reported on the application of SVR to chemical engineering problems in general and to heat transfer in particular. No work was found on the application of SVR to model a closed loop vertical tube thermosiphon reboiler.The SVR-based model developed in this study, based on the superior Structural Risk Minimization principle, has been found to be far more superior than any of the published models based on the Empirical Risk Minimization principle. The study has opened new vistas to apply SVR to the problems of chemical engineering in general and heat transfer and boiling heat transfer in particular.**

## 1. INTRODUCTION

A vertical tube thermosiphon reboiler represents effectively a pump-less system, in which natural, gravity-aided circulation takes place. Thermosiphon reboilers are being used in petroleum, chemical and petrochemical industries as vaporizers, evaporators and reboilers. Apart from this, the device is well adapted for heat transfer in situations prevailing in power plants, cooling systems for nuclear fuel rods in nuclear power reactors, refrigeration systems, pipe stills, and in the electronics industry. It also has relevant for efficiently transferring solar energy with little or no forced flow. The vertical tube thermosiphon reboiler has the advantages of excellent heat transfer rates, cheap manufacturing cost, easy cleaning, little maintenance, simplicity of construction, compactness and low operational costs. The frictional losses in the inlet and outlet piping and the cost of vapor line are minimized as the reboiler can be installed close to the column. The prediction of the rate of liquid circulation (thermally induced flow) and heat transfer is the primary requirement for the design and efficient operation of thermosiphon reboilers. Thermosiphon systems can be used in a wide range of operating temperatures and pressures. In almost all the applications, a subcooled liquid entering the tube gets heated by single phase convection and moves upwards. Depending upon the wall temperature conditions, subcooled boiling may get initiated at the surface. When the liquid attains saturation value, saturated boiling begins with the generation of net vapour, which increases resulting in various flow patterns ranging from bubbly to mist flow. The point at which the two phase begins is known as the incipient point of boiling (IPB) which corresponds to the condition of minimum degree of wall superheat required for the formation and detachment of the vapour bubble from the heated surface. Heat transfer coefficients at the onset of boiling are very high because of the nature of nucleate boiling and the increase in the velocity due to transition from single phase to a two-phase mixture. The IPB effectively divides the tube into two distinct regions; the non-boiling single phase and the two phase with entirely different modes of heat transfer. The prediction of the IPB is thus of paramount importance in the design of various process equipments operating on the thermosiphon reboiler principle. Engineering applications, in which the prediction of incipient boiling is necessary, are numerous. For instance, in some process applications dealing with thin film evaporation, nucleate boiling may not be a desirable phenomenon. However, in boiling water nuclear reactors, it is necessary to be able to predict the location of incipient boiling in the cooling channels, so that the downstream void distribution and the two phase pressure gradient can be determined. Various authors have proposed definitions of boiling incipience.

A number of studies have been conducted to understand the phenomenon of boiling incipience. Models have been proposed based on the semi-empirical approach for predicting the superheat for forced convection boiling in tubes. Important studies on boiling incipience include those by Murphy and Bergles [1], Yin and Messih [2], Hsu and Graham [3], Bergles and Rohsenow [4]. Yin and Abdel Messih [5] determined the liquid superheat during incipient boiling in a uniformly heated forced convection channel, and developed an analytical equation using Freon-11 as the test medium. Agarwal [6], Ali [7], Kamil [8] and Zaidi [9] gave empirical correlations to determine the length of heated tube required for the onset of fully developed boiling ($Z_{OB}$) in a vertical tube thermosiphon reboiler. Experimentally measured values of $Z_{OB}$ were compared with those predicted by the above equations.

Shamsuzzoha, et al. [10] have carried out a theoretical analysis to develop an analytical equation for the incipient point of boiling, including the effect of submergence. The minimum degree of wall superheat required for the onset of fully developed boiling of liquids was related to their physical properties through the theoretically derived equation. The model was tested using the experimental data available in literature. Shamsuzzoha and Alam [11] have carried out an analysis to predict the boiling incipience in a natural circulation flow reboiler tube taking into consideration the effect of turbulent eddies and submergence. An equation has been proposed to estimate the wall superheat for different types of liquids.

Thus, with the above observations in mind and given the recent developments in the application of artificial intelligence techniques, it was decided to explore the possibility of using one such technique for developing a unified correlation for predicting circulation rate in a thermosiphon reboiler.

Data driven modeling have been finding increasing relevance and acceptability in process industries. Of these, the classical methods such as least-squares methods, the maximum likelihood methods and traditional ANN are based on empirical risk minimization (ERM) principle whereas the support vector machine (SVM) method is based on the structural risk minimization (SRM) principle. This enables the SVM to achieve an optimum network structure by striking a right balance between the complexity of the approximation of the given data and the complexity of the approximating function. SVM is a supervised learning theory from the field of machine learning applicable to both nonlinear classification called support vector classification (SVC) and regression or SVR. The crux of the SVM design is solving a quadratic programming (QP) problem with linear constraints, which depends on the training vectors and the selection of few kernel parameters. The solution of a QP problem provides us the necessary information for choosing the most important vectors known as support vectors (SV) among all the data, and these support vectors will play an important role of defining the discriminant hyperplane or predicting function.

The SVM method has been instrumental in heralding a new era in supervised learning paradigm. It has been successfully applied to many fields; some of which are pattern recognition, phase diagram assessment, molecular and materials design, trace element analysis, cancer diagnosis [12], image analysis, drug design, time series analysis, quality control of food, protein structure/ function and genomics [13]. SVM methods have been applied to regression problems with great success [14, 15]. The applicability of SVR-based models in the field of chemical engineering has been well demonstrated [16,17]. Very little work has been reported in literature on the application of SVR to heat transfer in general and boiling heat transfer in particular. Such studies include those by Gandhi et al. [18]and Zaidi [19]. Besides, to the best of the author’s knowledge, no published literature is there on the application of SVR for modeling of the important performance parameters of a thermosiphon reboiler, like the length of heated tube required for the onset of fully developed boiling ($Z_{OB}$) in a vertical tube thermosiphon reboiler.

In the present study, it is for the first time that SVR-based modeling has been used for predicting the length of heated tube required for the onset of fully developed boiling ($Z_{OB}$) in a vertical tube thermosiphon reboiler.

The experimental data from literature was first preprocessed. Regression diagnostic tools were used in order to detect outliers. Thereafter, the data was analyzed for the contributions of the influencing variables to the regression and the insignificant variable was dropped. Using this data, a unified SVR-based model was developed. Further, the estimation performance of this model was comprehensively compared and evaluated with the conventional models. In the present work, the experimental data of Ali [7], Kamil [8], Nihaluddin [20] and Zaidi [9] for the boiling of acetone, benzene, ethanol, ethyl acetate, ethylene glycol, propan-2-ol, toluene and distilled water was utilized for the development and validation of the SVR-based models. The regression models used from literature for comparison were those of Ali [7], Nihaluddin [20], Kamil [8], and Zaidi [9].

## 2. EXPERIMENTATION AND DATA REDUCTION

### *2.1 Experimentation*

The experimental facility employed for the generation of a part of the data (including the author’s own work) that was used in the present study, was a single vertical tube thermo siphon reboiler as installed in the Heat Transfer Research Laboratory of the Department of Chemical Engineering, Aligarh Muslim University, Aligarh, India. Agarwal [6] however used a slightly earlier version of the same setup for the generation of the remaining data. The description of the later setup, operating procedure and data reduction has been necessitated by the fact that to appreciate the models, it was necessary to know these aspects in some detail. The experimental facility consisted of a natural circulation reboiler loop with a condenser and cooling system, power supply system and required instrumentation as shown in the schematic diagram in Fig. 1. The main unit was a U shaped circulation loop made up of two long vertical tubes connected together with the bottom by a short horizontal stainless tube, while the upper ends were connected to a vapor liquid separator and the condenser. One of the vertical tubes was electrically heated and served as the test section. The liquid entered the tube at its bottom end, got heated and rose upwards with subsequent boiling. The vapor liquid mixture entered the separator from where the vapors went to the condenser for total condensation. The condensate and the liquid from the separator were directed towards the top of the other tube serving as down flow cold leg. The entire liquid from the cold leg ultimately entered the test section through a view port. The vapor liquid separator was a cylindrical vessel with a tangential entry of the two phase mixture in the middle. The vapors were condensed by means of two water-cooled condensers used in series. The primary condenser was a spiral coil fitted just below the top cover of the condenser vessel. The condensation took place at the outer surface of the coil and condensed liquid drained down the bottom of the condenser vessel through a vertical tube fitted with a liquid level indicator.

A thermocouple was also inserted in this tube to measure the condensate temperature. The incondensables, if any, from the primary condenser entered the helical coil of the secondary condenser. The exit of the condenser was connected to a glass tube with its free end dipped into a bottle containing the test liquid so as to provide effective.

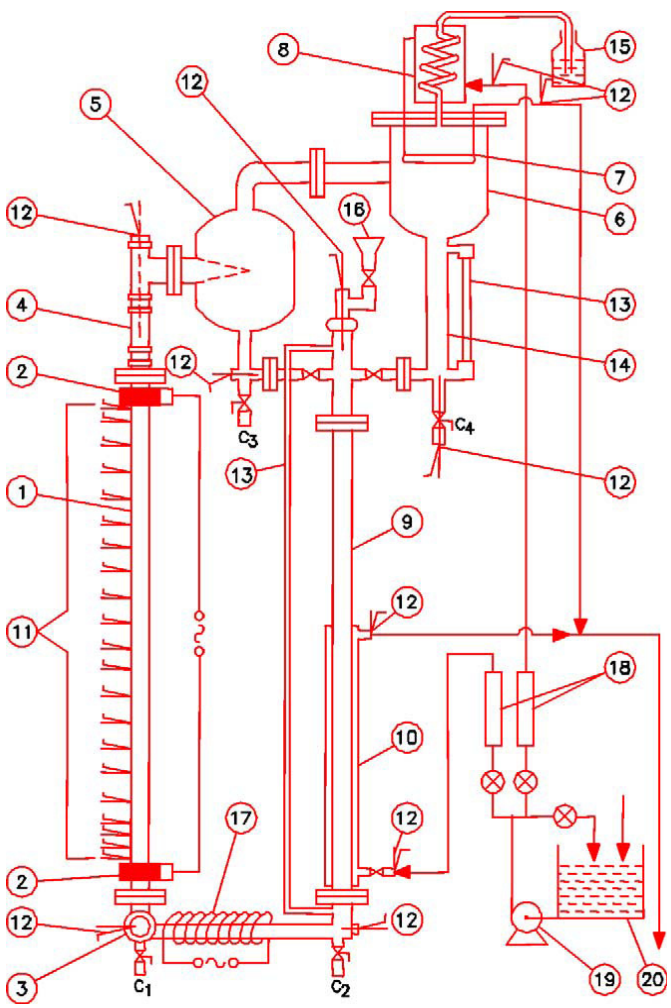
Visual observation of the removal of traces of dissolved air from the test liquids during initial boil off. A centrifugal pump and storage tank arrangement connected to fresh water supply was used for circulating water in the condensers. To measure the total rise in temperature of the cooling water, thermocouple probes of copper-constantan were located at the inlet of the secondary condenser and the outlet of the primary condenser. In order to control the inlet liquid temperature to the test section, the liquid down flow pipe was jacketted from the lower end up to a height of 1000 mm, using a pipe of 80 mm I.D. in which cooling water was passed as and when needed. The inlet and outlet temperatures of the water in the jacket were measured by means of thermocouple probes fitted therein. The temperature of the test liquid exiting from the down flow pipe and entering the horizontal pipe was measured by another thermocouple probe inserted at the bottom of the down flow pipe. The level of the test liquid in the down flow pipe (submergence) was indicated by a glass tube level indicator. This level acts as the driving force for the circulation of liquid through the loop.

Prior to the start of experimentation the setup was hydraulically tested for leaks. It was flushed with distilled water for through cleaning and finally filled with it up to the top of the test section. The connections to the power supply thermocouple and various measuring instruments were made and checking their calibration ensures the satisfactory performance of these.

Power was supplied to the test section and circulation system. Simultaneously, cooling water supply was activated thereby ensuring adequate amount of cooling water to the condensers. The system was kept running for several hours followed by aging in order to ensure stable tube wall nucleating characteristics. This step was essential for the reproducibility of data. Extreme care was taken that once the tube wall got stabilized; it must remain fully submerged with the liquid as the dry surface was very liable to entrap a thin film of air. This air on heating leaves the surface as tiny bubbles and joins the liquid, thereby setting up micro convection near the surface, resulting in additional extraneous turbulence causing error. During startup for conducting a series of runs, the test liquid was boiled off for about 6-8 hours to remove the last traces of dissolved air that was indicated by the cessation of air bubbles in the bubbler. After this, the desired heat flux was impressed upon the test section by proper adjustment and cooling water flow rate was maintained. Adding or draining the necessary amount of test liquid maintained the liquid level in the down flow pipe. When steady state conditions were established readings of thermocouples, various electrical instruments and rotameters were recorded. The liquid level in the down flow pipe was observed and noted from the glass tube level indicator. While keeping the submergence unchanged, readings were taken for different heat fluxes in increasing order. The horizontal pipe connecting the lower end of the down flow pipe/cold with that of the test section was heated by an electric heater made of 22 SWG nichrome wire wound over a length of 500 mm. It was energized by means of an autotransformer and the power supply was measured by a calibrated wattmeter. This arrangement was used in regulating and maintaining the temperature of the liquid and also in measuring the rate at which it entered the test section (Nihaluddin, 1993).

The test section was a stainless steel tube of 25.56 mm I.D., 28.85 mm O.D and 1900 mm long, tapped between two thick copper clamps designed to provide electrical contact to the tube with negligible contact resistance. The test section was electrically isolated from the rest of the setup by means of specially designed flanges fitted at the lower end and the upper glass tube section leading into the vapor liquid separator. This section gave a visual display of the boiling liquid emerging out of the test section. A view port at the junction of the horizontal tube and the test section enabled a visual observation of the test liquid to ensure complete absence of any air or vapor bubbles at the entry to the test section. A thermocouple probe at this location was inserted to measure the temperature of the liquid entering the test section. In order to monitor the heat transfer surface temperatures along the tube length, twenty one copper constantan thermocouples were spot welded on the outer surface of the tube at intervals of 50 mm up to a length of 200 mm from the bottom end and of 100 mm over the remaining length.

***Data Reduction***




***1. Test section; 2. Copper clamps; 3. View port for inlet liquid; 4. Glass tube section; 5. Vapor-liquid separator; 6. Primary condenser; 7. Spiral coil; 8. Secondary condenser; 9. Liquid down-flow pipe; 10. Cooling jacket; 11. Wall thermocouple; 12. Liquid thermocouple probes; 13. Liquid level indicator; 14. Condenser down-flow pipe; 15. Bubbler; 16. Feeding funnel; 17. Auxiliary heater; 18. Rotameters; 19. Centrifugal pump; 20. Cold water tank. $C_1$ to $C_4$. Drain cock valves***

**Figure 1. Schematic diagram of the experimental set up**

In some experiments, readings from only nineteen such thermocouples have been reported as the other three were probably not working. The energy to the test section was supplied through an automatic voltage stabilizer, autotransformer and low voltage high current transformer. The electrical energy input that got converted to heat in the wall of the test section was quantified by measuring the impressed voltage and current.

The operating parameters taken for investigation with each single component liquid were heat flux and submergence. All the data was generated at atmosphere pressure as prevailing in Aligarh, India. The thermal equilibrium model as suggested by Saha and Zuber [21] , formed the basis of determination of circulation rates and liquid bulk temperature distribution along the heated tube length by making a heat balance on the test section. According to it, as the sub cooled liquid enters the test section of the reboiler loop and because of the uniform heat flux at the heating surface, the liquid bulk temperature starts to increase almost linearly and continues up to the saturation value if all the heat added to the system goes to raise the temperature of the liquid only.

According to it, as the sub cooled liquid enters the test section of the reboiler loop and because of the uniform heat flux at the heating surface, the liquid bulk temperature starts to increase almost linearly and continues up to the saturation value if all the heat added to the system goes to raise the temperature of the liquid only. Thereafter, the liquid bulk temperature would remain constant at the saturation value and all the heat added would go as latent heat to generate vapor. For determining the circulation rate it was necessary to know the effective length of the non-boiling or sensible heating region over which the liquid temperature varied linearly. The lengths of the effective boiling and non-boiling zones over the entire heated tube were determined from the quantity of net vapor generation as obtained from the amount of vapor condensed in the condenser. A heat balance around the condenser gave:

$$M_V = \frac{\left[FC_{LC}(T_{C2} - T_{C1})\right]}{\left[\lambda + C_{L2}(T_{L2} - T_V)\right]} \tag{1}$$

Thus,

$$Z_B = \frac{M_V \lambda}{\pi q d} \tag{2}$$

$$Z_{NB} = L - Z_B \tag{3}$$

The rate of liquid circulation caused by buoyancy-induced flow was evaluated by making a heat balance over the non-boiling section.

$$Q = \pi d Z_{NB} q = m C_L \left( T_{L2} - T_{L1} \right) \quad (4)$$

The liquid temperature distribution along the tube length in the non-boiling zone was calculated assuming a linear relationship as mentioned below.

$$T_L = T_{L1} + \frac{(T_{L2} - T_{L1}) Z}{Z_{NB}} \quad (5)$$

where, $Z \leq Z_{NB}$

To obtain the correct inside wall temperature, the temperature drop between the thermocouple bead and the inside surface was estimated using the equation of conductive heat transfer with internal heat generation for a cylinder as:

$$\Delta T_W = \frac{q}{2\pi l k_W (D^2 - d^2)} \left[ D^2 \ln\left(\frac{D}{d}\right) - \left(\frac{D^2 - d^2}{2}\right) \right] \quad (6)$$

The local heat transfer coefficient at the thermocouple locations was calculated as:

$$h = \frac{q}{(T_W - T_L - \Delta T_W)} \quad (7)$$

The length mean heat transfer coefficient for the entire tube length as well as for the boiling and non boiling sections was calculated as:

$$\bar{h} = \frac{q}{\Delta T_{avg}} \quad (8)$$

where, $\Delta T_{avg}$ with proper subscript represents the value for the relevant zone of interest.

## 3. A BRIEF THEORY OF SVR-BASED MODELING

The detailed theory of SVM can be referred to in several excellent works, for example Vapnik [22] and Smola and Schölkopf [23]. Therefore, in this study only an abridged mention is there of the fundamentals of support vector regression (SVR). Support vector machines (SVMs) are a set of non-parametric machine-learning techniques, whose algorithm aims at a constructive learning procedure based on the statistical learning theory [22]. The crux of the SVM design is solving a quadratic programming (QP) problem with linear constraints, which depends on the training vectors and the selection of few kernel parameters. The solution of a QP problem provides us the necessary information for choosing the most important vectors known as support vectors (SV) among all the data, and these support vectors play an important role of defining the discriminant hyperplane or predicting function. In SVM the basic aim is to map the original data into a feature space F with higher dimensionality via a non linear mapping function $\phi$, which is usually unknown, and then carry out linear regression in the feature space. The SVR approximates a function by minimizing the regularized risk function given as:

$$R(C) = C \frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(y, f(x, w)) + \frac{1}{2} \|w\|^2 \quad (9)$$

Where,

$$L_\varepsilon(y, f(\mathbf{x}, w)) = \begin{cases} 0 & \text{if } |y - f(\mathbf{x}, w)| \leq \varepsilon \\ |y - f(\mathbf{x}, w)| - \varepsilon & \text{otherwise} \end{cases} \quad (10)$$

and $\varepsilon$ is a prescribed parameter.

In Eq. (9), the first term is called the empirical risk or error, which is measured by $\varepsilon$-insensitive loss function given by Eq. (10) and indicates that it does not penalize errors below $\varepsilon$. The parameter $\varepsilon$ is the tube size and it is equivalent to the approximation accuracy placed on the data points. It is a user defined value. The second term in Eq. (9), is used as a measurement of function flatness. C is a regularized constant determining the trade-off between the training error and the model flatness. C is also a user defined parameter. When introducing slack variables, the SVR formulation can be expressed mathematically in the form of a convex optimization problem as:

Minimize $\frac{1}{2}\left\|w^2\right\| + C\sum_{i=1}^{N} \xi_i + \xi_i^{(*)}$

Subject to

$$(y_i - w \bullet \phi(x) - b) \leq \varepsilon + \xi_i,$$

$$(w \bullet \phi(x) + b - y_i) \leq \varepsilon + \xi_i^{(*)}, \;\; \xi_i, \xi_i^* \geq 0 \quad (11)$$

The above-mentioned convex optimization problem (Eq.(11)) can be solved by transforming it into its dual form by introducing Lagrange multipliers and exploiting optimality constraints. The final decision function takes the following form:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N_{sv}} (\alpha_i - \alpha_i^*)(\phi(x_i) \bullet \phi(x_j)) + b \quad (12)$$

where, $\alpha$ and $\alpha^*$ are the introduced Lagrange multipliers. Only the non-zero coefficients, $(\alpha_i - \alpha_i^*)$, and the corresponding input vectors, $\mathbf{x}_i$, are called support vectors (SVs). These SVs are the most informative data points that compress the information content of the training set, thereby representing the entire SVR function. By introducing the radial basis function as the kernel function, all the necessary computations related to $\phi$ can be performed implicitly in the input space instead of in the feature space. Thus the basic SVR decision function modeling the data takes the following form:

$$f(x,\alpha_i,\alpha_i^*) = \sum_{i=1}^{N_{sv}} (\alpha_i - \alpha_i^*)(K(x_i,x_j) + b \qquad (13)$$

where, $K(x_i, x_j)$ is the kernel function.

The bias parameter, *b*, is computed by applying Karush–Kuhn–Tucker (KKT) conditions, which state that at the optimal solution the product between dual variables and constraints has to vanish.

A fast and efficient algorithm known as Sequential minimal optimization (SMO) has been adopted. The advantage of SMO lies in the fact that solving for the two Lagrange multipliers can be done analytically [24].The computation was carried out on a computer with the following specifications: Lenovo Intel[R], Pentium [R] M processor 1.73 GHz, 795 MHz with 1.24 GB RAM. In order to maintain a similarity in approach so as to facilitate comparison between the results from SVR model and those obtained by literature correlations the same randomly divided data sets were used for training (240 runs) and testing (60 runs). Out of the different kernels, the RBF kernel has been used in this study because of its good general performance and the few parameters to be adjusted. The best values of *C*, epsilon and the kernel parameter *γ*, were obtained by using the grid search methodology with standard *k*-fold cross-validation procedure on training data set that minimizes the average absolute relative error (AARE) and improves the correlation coefficient (R) values in the direction of unity.

For the statistical analyses of SVR-based model and the models from literature, the following model evaluation parameters were used.

The average absolute relative error (AARE) should be a minimum. It is given as:

$$AARE = \frac{1}{N}\sum_{i=1}^{N}\left\|\left[\frac{y_{ipred} - y_{i\exp}}{y_{i\exp}}\right]\right\| \qquad (14)$$

2. Correlation coefficient (R) should approach unity for a good fit. It is given as:

$$R = \frac{\sum_{i=1}^{N}(y_{i\exp} - y_{imean})\times(y_{ipred} - y_{ipredmean})}{\sqrt{\sum_{i=1}^{N}(y_{i\exp} - y_{i\exp mean})^2} \times \sqrt{\sum_{i=1}^{N}(y_{ipred} - y_{ipredmean})^2}} \qquad (15)$$

The root mean square (RMSE) used to evaluate the model should be minimum. Its mathematical expression is given as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}((y_{i\exp} - y_{ipred}) / y_{ipred})^2}{N}} \qquad (16)$$

The standard measure of how widely values are dispersed from the average value (the mean) is evaluated by the standard deviation (σ). It should be a minimum.

$$\sigma = \sqrt{\sum_{i=1}^{N}\frac{1}{N-1}(\left|(y_{ipred} - y_{i\exp} / y_{i\exp})\right| - AARE)^2} \qquad (17)$$

The internal predictive capability of the SVR-based model was evaluated by leave-one-out cross validation($Q^2_{LOO}$) on the training set, which was calculated by the following equation(Gramatica, 2007):

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{N_{training}}(y_{i\exp} - y_{ipred})^2}{\sum_{i=1}^{N_{training}}(y_{i\exp} - y_{\exp mean})^2} \qquad (18)$$

The external predictive capability of the SVR-based model was evaluated by leave-one-out cross validation($Q^2_{ext}$) on the test set, which was calculated by the following equation:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{N_{test}}(y_{i\exp} - y_{ipred})^2}{\sum_{i=1}^{N_{test}}(y_{i\exp} - y_{tr\,mean})^2} \qquad (19)$$

The mean relative error (MRE) for valuating the model should be as low as possible.

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\left\|\left[\frac{y_{i\exp} - y_{ipred}}{y_{ipred}}\right]\right\| \qquad (20)$$

## 4. RESULTS AND DISCUSSION

***Data preprocessing results***

The length of the heated tube ($Z_{OB}$ or $Z_{NB}$) from the inlet, required for the boiling to be effective has been found to depend upon the wall heat flux (q), submergence (S), inlet liquid sub cooling ($\Delta T_{sub}$ or $t_{sub}$) and ratio of kinematic viscosities of the liquid and the vapor ($\nu_L/\nu_V$) [7,8,9]. The data from the sources listed earlier was preprocessed to detect outliers. The data thus obtained consisted of 300 experimental runs involving the boiling of acetone, benzene, ethanol, ethyl acetate, ethylene glycol, propan-2-ol, toluene and distilled water in a thermosiphon reboiler.

All the correlations for predicting the length of the heated tube from the inlet, required for the boiling to be effective, have correlated the experimentally obtained values of $Z_{OB}$ or $Z_{NB}$ with the influencing variables in terms of dimensionless groups. Thus, $Z_{OB}$ or $Z_{NB}$ has been converted to $Z_{OB}$/Lx100 and q and $\Delta T_{sub}$ have been converted to $Pe_B$ and $K_{sub}$ respectively; S and $\nu_L/\nu_V$ being already dimensionless. The ranges of these dimensionless groups as used in the present study for developing the SVR model are given in Table 1.

In the present study, the RBF kernel was used. For getting a high generalization performance, it was necessary to optimize the model in order to have the right setting of C, $\varepsilon$ and $\gamma$.

**Table 1: Ranges of dimensionless groups used in the present study**

| Group | Peb | Ksub | S | $\nu_L/\nu_V$ | $Z_{OB}$/Lx100 |
|---|---|---|---|---|---|
| Range | 39.3-586.6 | 1.3-1155 | 30-100 | 0.00018-0.23 | 22-82 |

For achieving this, grid search methodology with 10-fold cross-validation, was used on the training data by first varying these parameters coarsely in the ranges: C: [$2^3$, $2^{14}$ ], $\gamma$ [$2^{-12}$, $2^{-1}$ ] and $\varepsilon$ [$2^{-9}$, $2^2$ ] and then making a fine search. The optimal values of the model parameters for the SVR-based model are listed in Table 2.

**Table 2. Optimal parameters for SVR-based model for length of the non-boiling zone in a vertical tube thermosiphon reboiler.**

| Model | C | $\gamma = 1/2\sigma^2$ | ε | Kernel | Loss function | SVs | training points |
|---|---|---|---|---|---|---|---|
| $Z_{OB/L}$x100 | 100 | 0.1 | 0.32 | RBF | ε - insensitive | 122 | 240 |

After optimization of the SVR parameters, the model output was used to construct the training course curve and the test data course curve as shown in Figure 2 and Figure 3, respectively. It can be seen that the SVR-based model is adequately trained and it predicts the unseen test data well.

Table 3 compares the prediction of the SVR-based model in terms of statistical evaluation parameters using the training data set and the test data set. In terms of the model evaluation parameters, an AARE of 8.98 % and an R of 0.8333 on the training data and corresponding values of 10.08 % AARE and an R of 0.8302 on the test data were obtained. This may be considered as a fairly good prediction and generalization ability, considering the diversity of sources from which the data was obtained. It can be concluded that the model has a good accuracy and generalization ability.

Plots between the actual values of $Z_{OB}$/Lx100 and those predicted by the SVR model using the training data set and the test data set are shown in Figure 4. The model predicts the experimental data very well. The results demonstrate that the SVR-based model is a robust model. Table 4 illustrates the distribution of predicted data points of $Z_{OB}$/Lx100 by SVR-based model in terms of absolute deviation for training data.

**Table 3. Model evaluation parameters for SVR-based model using the training data and the test data.**

| SVR Model evaluation parameter | Train data | Test data |
|---|---|---|
| AARE (%) | 8.98 | 10.08 |
| R | 0.8333 | 0.8302 |
| RMSE | 0.1381 | 0.1185 |
| SD | 0.0918 | 0.0833 |
| $Q^2_{LOO}$ (Training data), $Q^2_{ext}$(Test data) | 0.6923- | -0.6778 |
| MRE | 0.0908 | 0.0946 |

**Table 4. Percentage distribution of predicted data points of $Z_{NB}$ by SVR-based model in terms of absolute deviation (AD) for training data.**

| Absolute deviation (AD) (%) | % of SVR model predicted values | Cumulative score |
|---|---|---|
| AD< 10 | 77.91 | 77.91 |
| 10<AD<15 | 7.5 | 85.41 |
| 15<AD<20 | 3 | 88.41 |
| 20<AD<25 | 4.1 | 92.51 |
| AD>25 | 7.49 | 100 |
| Total | 100 | |

It is observed that the SVR-based model predicts nearly 88.41 per cent data points within an absolute deviation of less than 20 % and a total of 92.51 per cent data points within an absolute deviation of less than 25 %. Only 7.49 per cent data points have an AD of more than 25 per cent.

Table 5 depicts the distribution of predicted data points of $Z_{OB}$/Lx100 by SVR-based model in terms of absolute deviation for the test data set. It can be seen that the SVR-based model predicts nearly 88.33 per cent data points within an absolute deviation of less than 20 % and a total of 93.33 cent data points within an absolute deviation of less than 25 %. Only 6.67 per cent data points have an AD of more than 25 per cent. This shows that the SVR-based model has a good ability to predict the unseen test data set.

**Table 5. Percentage distribution of predicted data points of $Z_{NB}$ by SVR-based model in terms of absolute deviation (AD) for test data.**

| Absolute deviation (AD) (%) | % of SVR model predicted values | Cumulative score |
|---|---|---|
| AD< 10 | 58.33 | 58.33 |
| 10<AD<15 | 16.67 | 75 |
| 15<AD<20 | 13.33 | 88.33 |
| 20<AD<25 | 5 | 93.33 |
| AD>25 | 6.67 | 100 |
| Total | 100 | |

## 5. COMPARISON OF SVR-BASED MODEL WITH CORRELATIONS AVAILABLE IN LITERATURE

A comparison of the unified SVR-based model for predicting the length of non-boiling section of the heated tube was made against three models from the literature using the test data set. Table 6 depicts the AARE values for each model. It is observed that the SVR-based data-driven model gives an unprecedented lowest AARE value of 10.08 per cent among all the models. The nearest AARE value is 57.19 % exhibited by Kamil [8]. It can therefore be concluded that the prediction performance of the SVR-based model is the best.

**Table 6. Performance of different models on test data set to predict the length of the non-boiling zone in a vertical tube thermosiphon reboiler.**

| Author(s) | AARE (%) |
|---|---|
| Ali [7] | 78.71 |
| Kamil [8] | 57.19 |
| Zaidi [9] | 77.71 |
| SVR-based model (present work) | 10.08 |

Further, Figure 5 illustrates the prediction performance of the unified SVR-based model vis-à-vis the other three models from literature. It is observed that the SVR-based model gives the most superior performance.

Thus, based on the statistical evaluation parameters, the training and test course curves, the AD values and the comparison with other correlations in literature, the results show that the maiden SVR- based model based on the universal Statistical Learning Theory has very high prediction ability and is very accurate.

## 6. CONCLUSIONS

In the present study, a model has been developed to predict the length of the boiling/non-boiling section of the heated tube in a thermosiphon reboiler. The SVR-based model is superior to the other models considered in this study as it displays a high degree of accuracy and generalization ability. It shows that the prospects of the application of SVR in chemical engineering in general and heat transfer in particular are very bright.

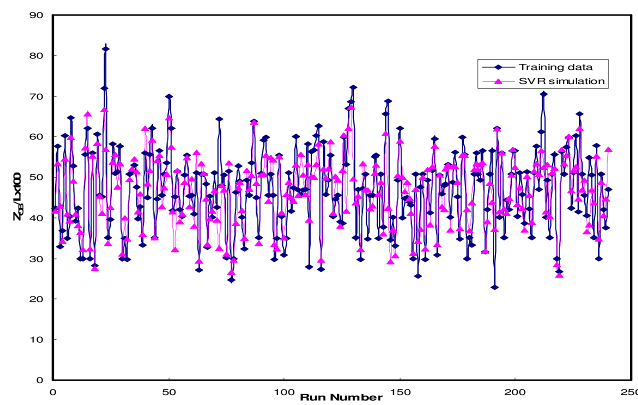



**Figure 2. Training course curve for the length of the non-boiling zone.**

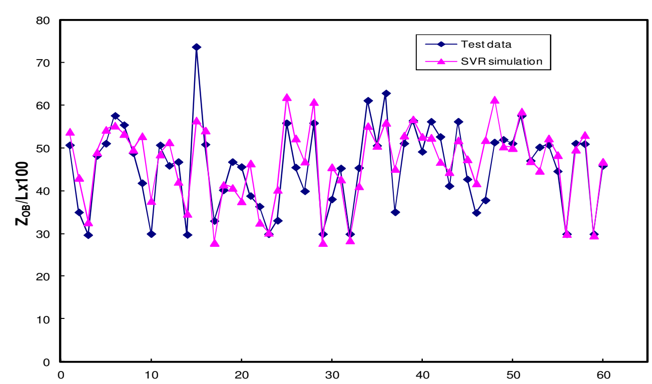



**Figure 3. Test course curve for the length of the non-boiling zone.**

*Notation*

| | | |
|---|---|---|
| A | cross-sectional area of heated tube | |
| AARE | average absolute relative error | % |
| b | bias term | b |
| $C_L$ | heat capacity | J/kg °C |
| C | cost function | C |
| d | inside diameter of the tube | m |
| F | high-dimensional feature space | |
| f(**x**) | regression function | |
| k | thermal conductivity | W/m °C |
| $K(\mathbf{x}_i,\mathbf{x}_j)$ | kernel function | |
| L | Lagrangian function (dual form) | |
| m | circulation rate | kg/s |
| $N_{sv}$ | number of support vectors | |
| Q | heat input | W |
| q | heat flux | $W/m^2$ |
| S | submergence | % |
| $T_{L1}$ | inlet liquid temperature to the tube | °C |
| $T_S$ | liquid saturation temperature in the tube | °C |
| $T_{sub}$ | degree of subcooling | °C |
| **w** | weight vector | |
| $\mathbf{x}_i$ | ith input vector | |
| $x^o$ | vapour fraction | |
| $y_i$ | target output corresponding to the ith vector | |
| $Z_{NB}$ | length of the non-boiling zone | m |

*Greek Letters*

| | | |
|---|---|---|
| λ | latent heat of vaporization | J/kg |
| ρ | density | $kg/m^3$ |
| μ | dynamic viscosity | N s / $m^2$ |
| σ | width of RBF kernel | |
| $\sigma_L$ | surface tension of liquid | N/m |
| α and $\alpha^*$ | Lagrange multipliers | |
| γ | regularization parameter | |
| ε | loss function | |
| $\varphi(\mathbf{x}_i)$ | mapping function to high dimensional feature space for input vector **x** | |

*Subscripts*

| | |
|---|---|
| exp | experimental |
| pred | predicted |
| L | liquid |
| NB | non- boiling |
| S | saturation |
| sub | subcooling |
| tr | training |
| V | vapour |

*Dimensionless Numbers*

| | |
|---|---|
| $K_{sub}$ | subcooling number |
| $Pe_B$ | Peclet number |
| Re | Reynolds number |
| $X_{tt}$ | Lockhart- Martinelli parameter |

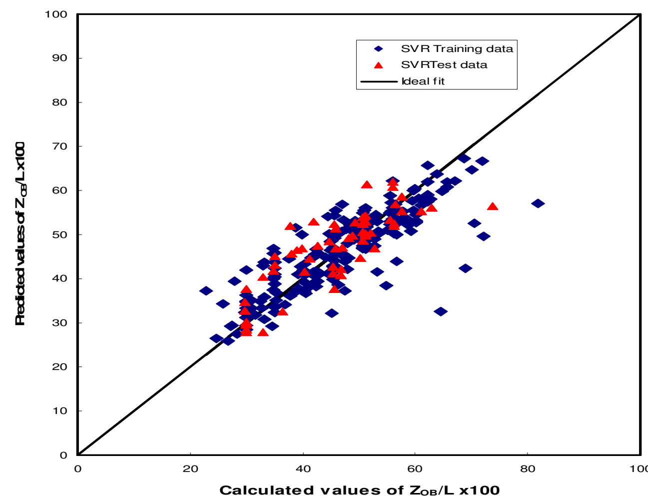



**Figure 4. SVR simulation of the length of the non-boiling zone in a vertical tube.**

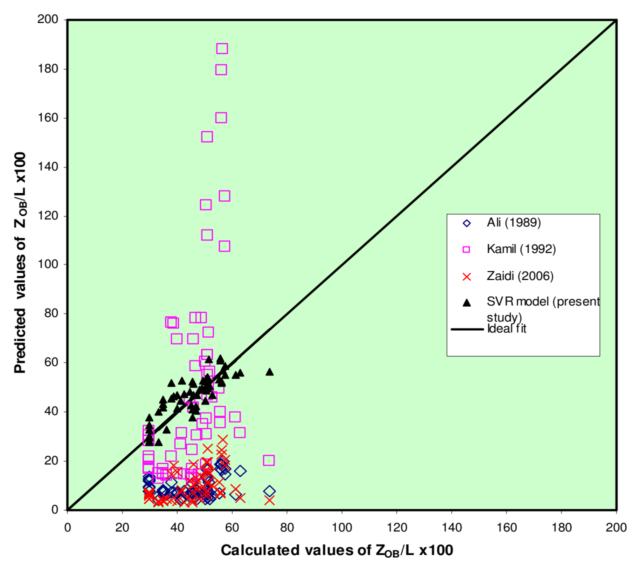



**Figure 5. Comparison of SVR simulation of the length of the non-boiling zone with models in literature using test data.**

## 7. ACKNOWLEDGEMENTS

The author wishes to gratefully acknowledge all the authors whose works have been referred to in the paper, Also, the author is grateful to the Aligarh Muslim University for having provided all the facilities for this work.

## REFERENCES

[1] Murphy R. W., Bergles, A. E. Subcooled flow boiling of fluro-carbons- hysteresis and dissolved gas effects on heat transfer,

Proceedings of the Heat Transfer and Fluid Mechanics Institute, Stanford Univ. Press, (1972) 400-416.

[2] Yin, S. T., Abdelmessih, A. H. Prediction of incipient flow boiling from a uniformly heated surface, AIChE Symp. Ser.164, 73, (1977) 236-243.

[3] Hsu, Y.Y., Graham, R.W. Transport Processes in Boiling and Two- Phase Systems, Hemisphere Publishing Corporation, Washington (1976).

[4] Bergles, A.E., Rohsenow, W. M. The determination of forced convection surface boiling heat transfer, Trans. ASME. Ser. C, Vol. 86, (1964) 365-372.

[5] Yin, S. T., Abdelmessih, A. H. Measurements of liquid superheat, hysteresis effects and incipient boiling oscillations of Freon-11 in forced convection vertical flow, University of Toronto, Mech. Eng. Tech. Pub. Ser., TP-7401 (1974).

[6] Agarwal, C.P. Heat Transfer Studies in a Vertical Tube of Closed-Loop Thermo siphon, PhD Thesis in Chemical Engineering, University of Roorkee, India (1980).

[7] H. Ali, Studies on Thermosiphon Reboiler, PhD Thesis in Chemical Engineering, Aligarh Muslim University, Aligarh (1989).

[8] M. Kamil, Studies on a vertical tube thermosiphon reboiler, Ph.D. Thesis, Aligarh Muslim University, Aligarh (1992).

[9] Zaidi, S. Studies on Vertical Thermo siphon Reboiler, M. Tech. Chemical Engineering (Petroleum Processing) Thesis, Aligarh Muslim University, Aligarh (2006).

[10] Shamsuzzoha, M., Kamil, M., Alam, S. S. Conditions of Onset of Boiling in a Vertical Thermo siphon Reboiler, Journal of Thermophysics and Heat Transfer, 18 (4) (2004) 448-456.

[11] Shamsuzzoha, M., Alam, S. S. Effect of Bubble Turbulence and Submergence on Boiling Incipience in Vertical Thermosiphon Reboiler, Heat transfer engineering, Vol 29(3) (2008) 306-320.

[12] Chen, N, Lu, W., Yang, J., Li, G., 2004. Support Vector Machine in Chemistry. World Scientific, Singapore.

[13] Li, H., Liang, Y., Xu, Q., 2009. Support vector machines and its applications in chemistry. Chemometrics and Intelligent Laboratory Systems 95, 188-198.

[14] Al-Anazi A.F., Gates I.D., 2010. Support vector regression for porosity prediction in heterogeneous reservoir: A comparative study. Computers & Geosciences 36 1494–1503.

[15] Kavaklioglu, K., 2011. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Applied Energy 88, 368–375.

[16] Zhao, B., 2009. Modeling pressure drop coefficient for cyclone separators: a support vector machine approach. Chemical Engineering Science **64**, 4131–4136.

[17] Azamathulla, H. M., Wu, F. C., 2011. Support vector machine approach for longitudinal dispersion coefficients in natural streams. Applied Soft Computing 11, 2902-2905.

[18] Gandhi, A.B., Joshi, J.B., 2010. Estimation of heat transfer coefficient in bubble column reactors using support vector regression. Chemical Engineering Journal 160, 302-310.

[19] Zaidi,S. 2012. Development of support vector regression (SVR)-based model for the prediction of circulation rate in a vertical tube thermosiphon reboiler. Chemical Engineering Science 69, 514-521.

[20] Nihaluddin, 1993. Heat transfer to boiling liquids in a vertical thermosiphon reboiler. M. Sc. Chemical Engineering (Petroleum Processing) Thesis, Aligarh Muslim University, Aligarh, India.

[21] Saha, P., Zuber, N., 1978. An analytical study of the thermally induced two-phase flow instabilities including the effect of thermal non-equilibrium. International Journal of Heat and Mass Transfer 21, 415- 426.

[22] V. N. Vapnik, 1995. The Nature of Statistical Learning Theory, Springer, New York.

[23] A. J. Smola and B. Schölkopf, A tutorial on support vector regression. Statistics and Computing, Vol 14, (2004 ) 199-222.

[24] Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in Kernel Methods-Support Vector Learning* (pp. 185-208). Cambridge, MA: The MIT Press.