RESULTS AND DISCUSSIONS

The chapter presents results achieved through the completed research work. Outputs of the process models that were developed (as explained in the previous chapters) for two different modelling problems pertaining to DDP: (i) one problem involved calculating SRE and WRE, and (ii) another problem involved calculating wash water flowrate, are presented. Also important conclusions derived from such observations and significance of the completed work are highlighted.

5.1 Background

After applying material and methods as discussed in Chapter-3, VP Model was attained, whose features were further described in Chapter-4. Two process modelling problems pertaining to DDP were also described as applications of the versatile process model in the previous chapter. Results are further discussed and concluded in the following sections:

5.2 Modelling Problem-1: Modelling of DDP to predict SRE and WRE

As was explained in section 4.2.1, the developed VP Model i.e. the modelling framework / tool was used to model DDP to predict SRE and WRE based on temperature (^{o}C), chemical injection (ppm), fresh water injection %, mixing time (min.) and settling Time (min.). The modelling primarily involved populating VP Model with Training Data and using Solver to attain 71 model parameters (weights and biases), separately for SRE and WRE, based on the training data. Then this solved model was used to predict SRE and WRE for the Testing Data. In the following sub-section results are discussed.

5.2.1. Results

To verify the performance of the developed model, output (predicted efficiencies) of thus developed model in MS Excel, was compared with

i. known experimental value, and
ii. known value computed from ANN model implemented through MATLAB

In literature, coefficient of correlations (R^2) is found as widely used criterion to evaluate performance of ANN model. In this thesis also R^2 is used. Observations pertaining to SRE and WRE are presented separately, as these were calculated through separate models using VP Model tool.

Observations and discussions pertaining to SRE:

1. **VP Model (this study) Output vs. Experimental value (lit.) [for Training Data]**

   The graphical comparison of SRE [%] calculated from VP Model (the ANN Process model developed in MS Excel in the current research) versus corresponding experimental value of SRE found in the literature (lit.) for each Training Data set consisting of independent variables as temperature (^oC), chemical injection (ppm), fresh water injection %, mixing time (min.) and settling time (min.) is presented as follows:

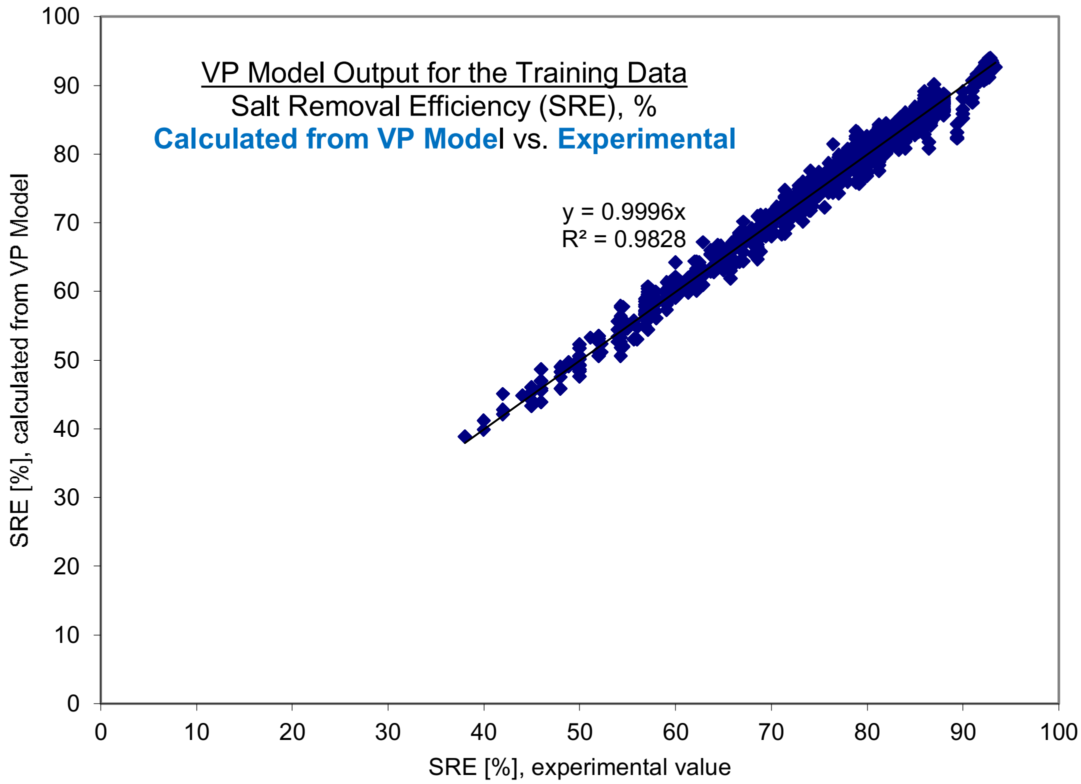

Fig. 5.1: SRE: VP Model Output vs. Experimental data (for Training Data)

Observed data distribution and R^2 value (0.9828) indicate model performance as encouraging.

Also, it may be noted that MSE (mean squared error) determined for the value predicted by model as compared to the known experimental value is 0.00074, which is also considered indicative of the model performance.

However, as Training Data represents that data which was actually used to solve the model to determine model's parameters (71 numbers), there remains a possibility that model simply "fitted" the data and did not "learn" appropriately to "generalize". Generalization is the main feature which is desired in any ANN process model. To cross-check, the "learning of the model to be able to generalize", further analysis is needed, which is described below.

2. **VP Model (this study) Output vs. Experimental value (lit.) [for Testing Data]**

   To check model's ability to generalize, SRE was determined for another set of unseen data (termed as Testing Data) which was not used for solving model's parameters. Accordingly, SRE [%] calculated from VP Model (the ANN model developed in MS Excel in the current research) vs. experimental value found in the literature for each Testing Data is plotted as follows:

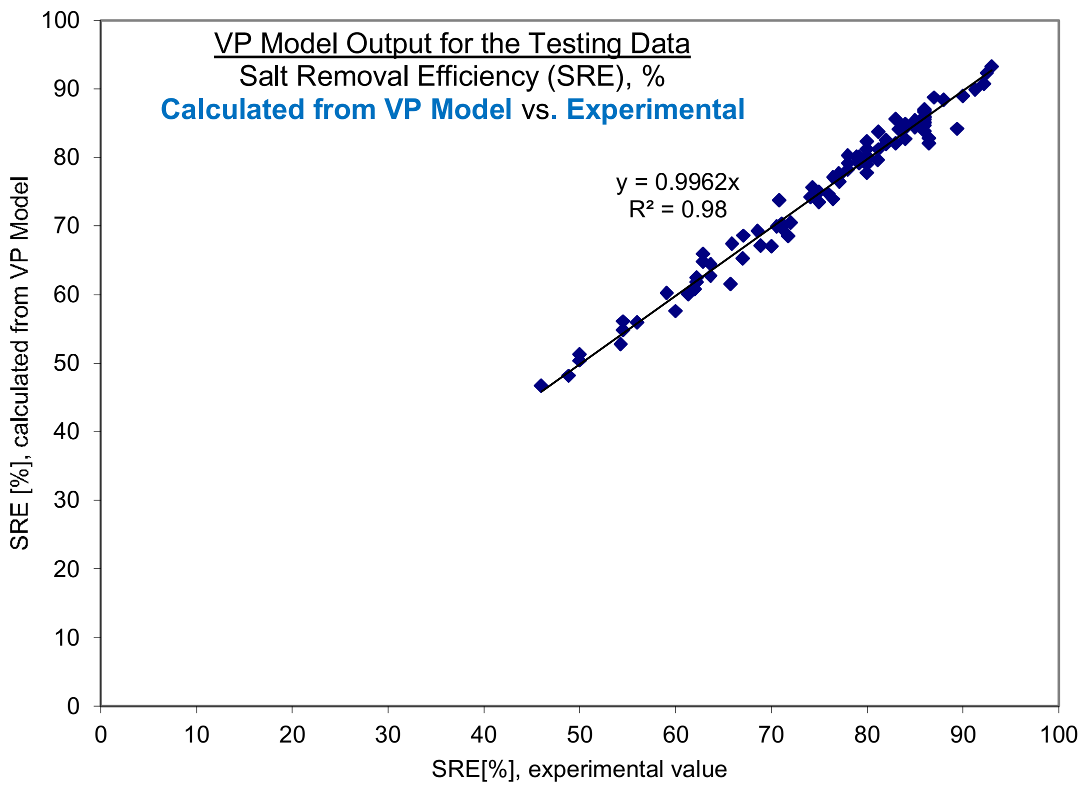

Fig. 5.2: SRE: VP Model Output vs. Experimental data (for Testing Data)

   Observed data distribution and R^2 value (0.98), even for previously unseen data for the model, indicate model performance as more encouraging.

However, it would be interesting to know how the model developed in MS Excel in this research performs as compared to the model developed through reputed modelling tool in previous study (reported in literature), particularly because developing ANN model in Excel in order to combine the power and versatility of ANN and MS Excel is one of the main novelties of this study.

3. VP Model/ANN in Excel (GRG2, this study) Output vs. ANN in MATLAB (LM, lit.) Output [for Training data]

In order to compare the performance of the ANN process model developed in MS Excel in this research with model developed through reputed modelling tool in previous study, SRE calculated from VP Model (the ANN model developed in MS Excel in this research) vs. SRE computed from ANN model implemented in MATLAB (reported in the literature) for each Training Data is plotted as follows:

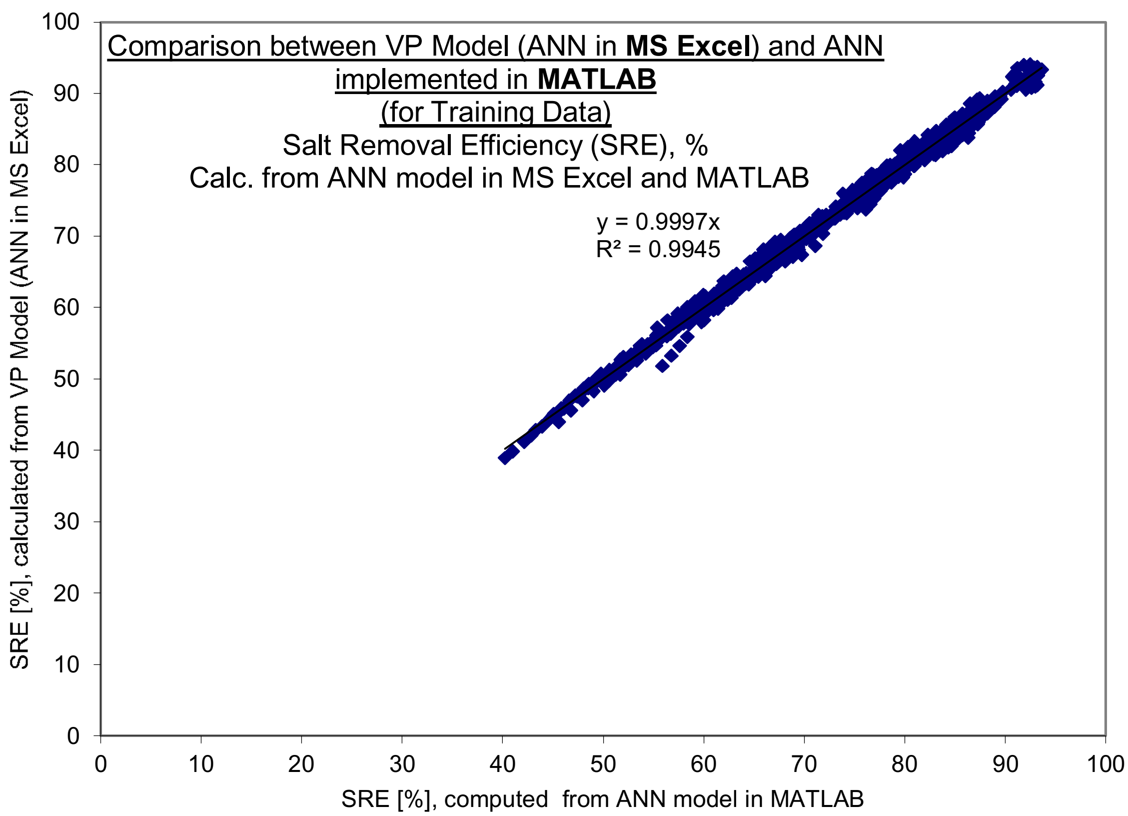

Fig. 5.3: SRE: VP Model (Excel) Output vs. MATLAB Output (for Training Data)

Observed data distribution and R^2 value (0.9945) indicate model performance as encouraging. It may be noted that several previous studies pertaining to developing ANN process model utilised MATLAB tools. The output of ANN in MATLAB corresponded to Training achieved through Levenberg-Marquardt (LM) algorithm [42], which is considered as a prominent algorithm [30]. However, the current study utilised Solver in MS Excel using the GRG Nonlinear Solving Method for nonlinear optimization that uses the Generalized Reduced Gradient (GRG2) code.

It is known that LM outperforms the Solver incorporated in MS Excel in terms of accuracy [30], and there is way to implement LM algorithm in MS Excel [30], but current study focused on attaining ANN modelling framework in MS Excel workbook, without embedding any macros / programming, in order to achieve various benefits from such modelling, besides modelling particular cases of DDP presented in this thesis. Accuracy of the result obtained by LM may be limited by the noise present in data [30].

Despite not implementing sophisticated algorithm like LM, performance of ANN model achieved through standard Solver in MS Excel appears encouraging from the above plot. From the literature, it is found that other researchers have also tried to utilised the underlying GRG2 algorithm for ANN training [59]. The results achieved in the present work in MS Excel for modelling of DDP further added variety regarding application of GRG2.

Further, as explained previously, output generated for Testing Data is likely to be a better indicative of model's performance. Therefore, the performance of the ANN model developed in this study for Testing Data was further compared with the performance of the ANN model developed through reputed software like MATLAB for the same set of Testing Data.

4. VP Model/ANN in Excel (GRG2, this study) Output vs. ANN in MATLAB (LM, lit.) Output [for Testing Data]

The graphical comparison of SRE [%] calculated from VP Model (ANN process model developed in MS Excel) vs. SRE computed from ANN model implemented in MATLAB for the Testing Data sets is as follows:

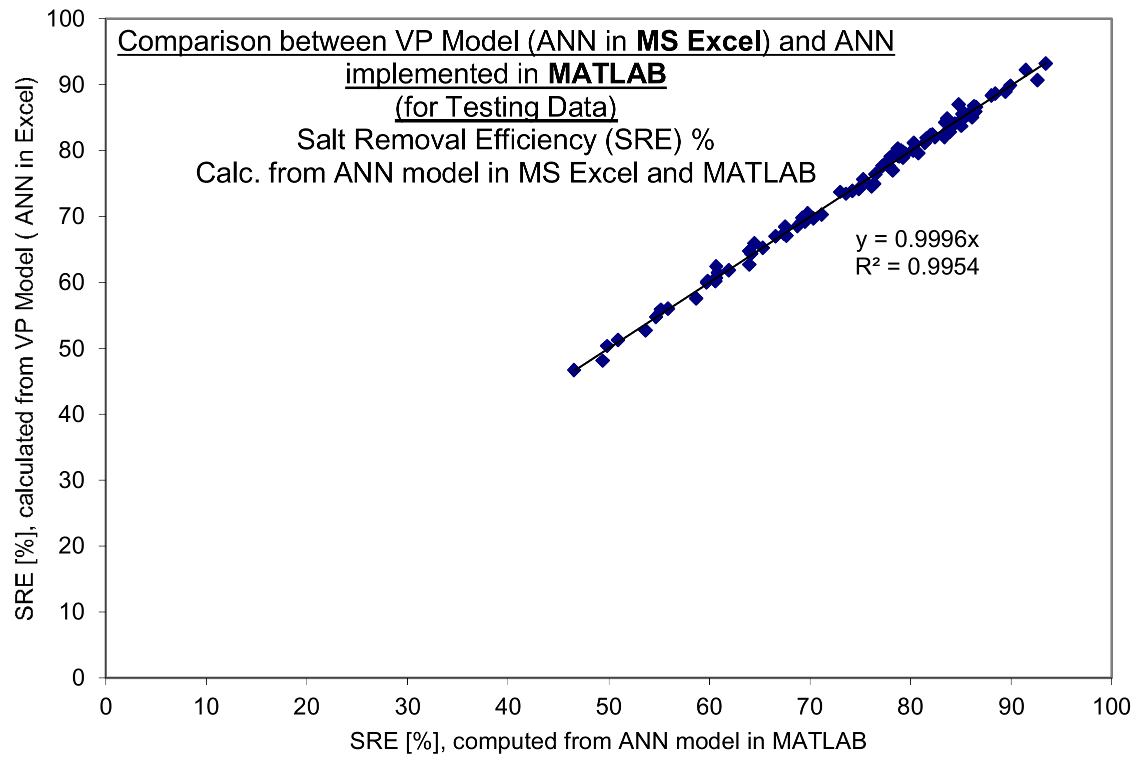

Fig. 5.4: SRE: VP Model (Excel) Output vs. MATLAB Output (for Testing Data)

Observed data distribution and R^2 value (0.9954) indicate model performance as encouraging.

Thus the performance of the Versatile Process (VP) Model (developed as ANN in MS Excel in this research) appears comparable with the performance of the ANN model implemented through LM method in MATLAB that enjoys popularity amongst the academician for scientific and engineering research.

The ANN modelling tool / basic framework developed in this research as MS Excel spreadsheet, might be considered like a basic version, which does not include features those are present in MATLAB. However, it is important to remember that the research, in general, was meant, not to replace any reputed software, whether it is process simulation software or ANN modelling tool or any other decision support tools in doing for what they are considered great, but is meant to supplement them with the combined power and versatility of ANN and MS Excel for wider usage.

Observations and discussions pertaining to WRE:

1. **VP Model (this study) Output vs. Experimental value (lit.) [for Training Data]**

   The graphical comparison of WRE [%] calculated from VP Model (the ANN model developed in MS Excel in the current research) vs. experimental value found in the literature for each Training Data is presented as follows:

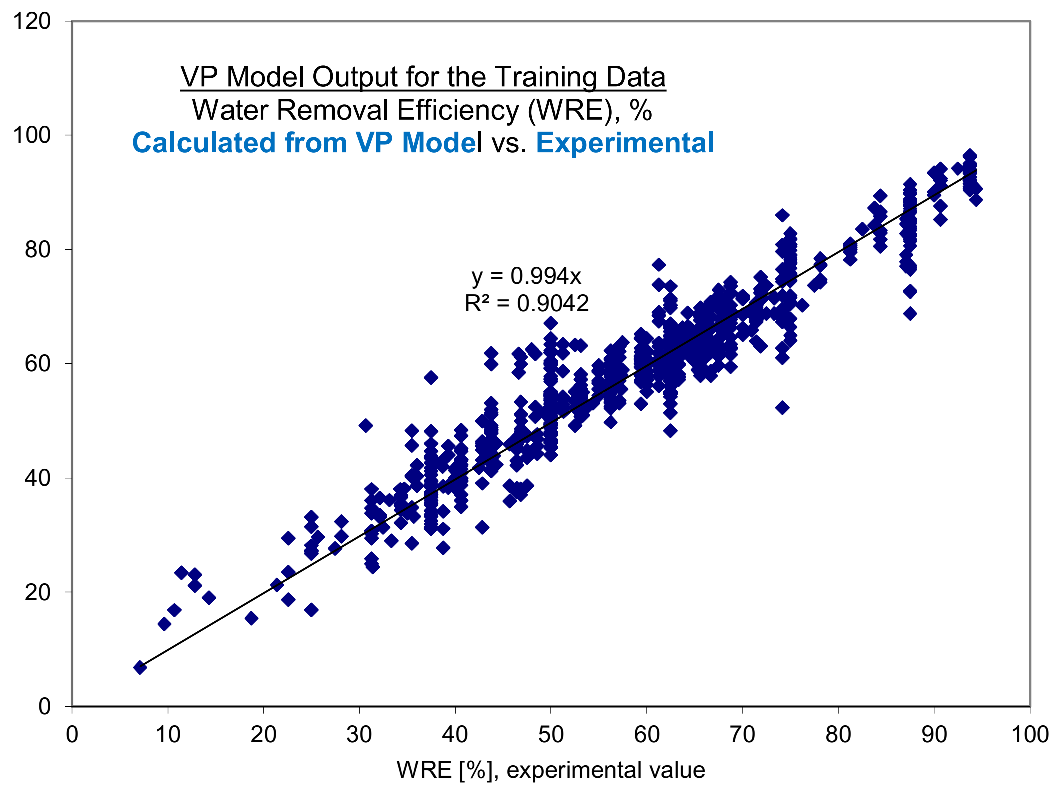

Fig. 5.5: WRE: VP Model Output vs. Experimental data (for Training Data)

Observed data distribution and R^2 value (0.9042) indicated model performance as encouraging. Also, it may be noted that MSE (mean squared error) determined for the value predicted by model as compared to the known experimental value is 0.00296, which is also considered indicative of the model's performance.

However, as Training Data represents that data which was used to solve the model to determine model's parameters, there remains a possibility that model simply "fitted" the data and did not "learn" appropriately to "generalize". To cross-check, the "learning of the model to be able to generalize", further analysis is needed, which is further discussed.

2. **VP Model (this study) Output vs. Experimental value (lit.) [for Testing data]**

   To check model's ability to generalize, WRE [%] calculated from VP Model (the ANN model developed in MS Excel in the current research) vs. experimental value found in the literature for each Testing Data is plotted as follows:

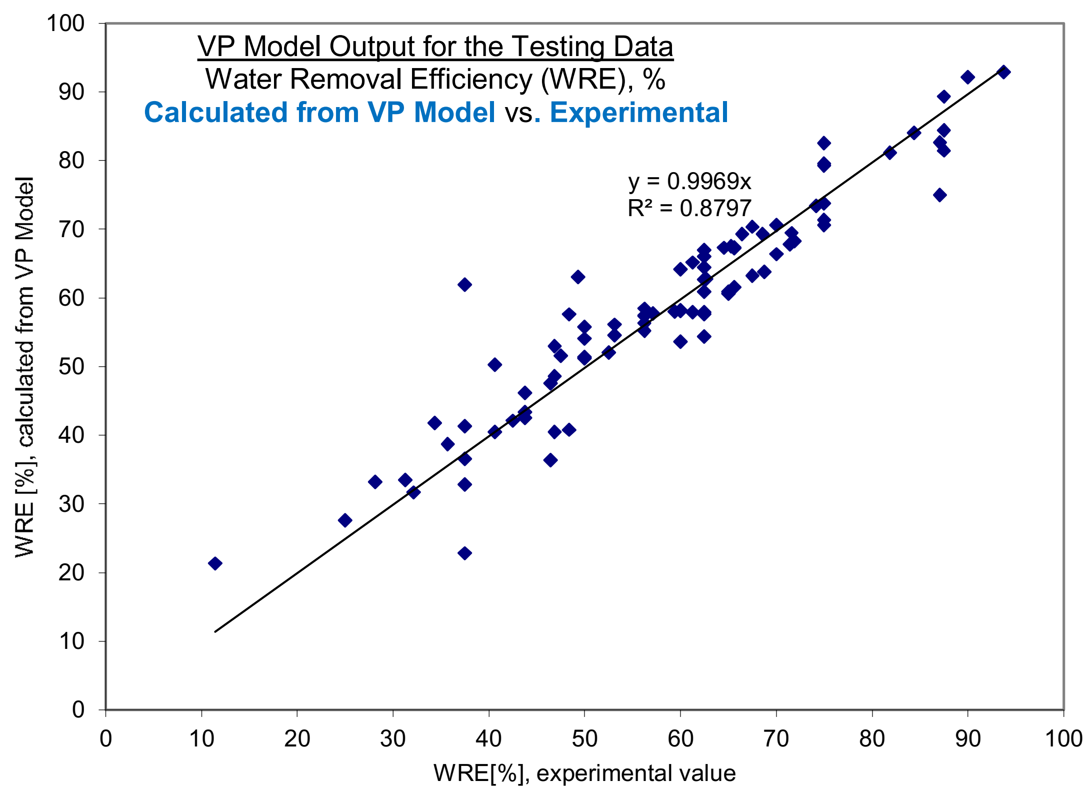

Fig. 5.6: WRE: VP Model Output vs. Experimental data (for Testing Data)

Observed data distribution and R^2 value (0.8797), even for previously unseen data for the model, indicates model performance as more encouraging.

However, it would be interesting to know how the model developed in MS Excel in this research performs as compared to the model developed through reputed modelling tool in previous study (reported in literature).

3. **VP Model/ANN in Excel (GRG2, this study) Output vs. ANN in MATLAB (LM, lit.) Output [for Training Data]**

   WRE calculated from VP Model (the ANN model developed in MS Excel in this research) vs. WRE computed from ANN model implemented in MATLAB (reported in the literature) for each Training Data is plotted as follows:

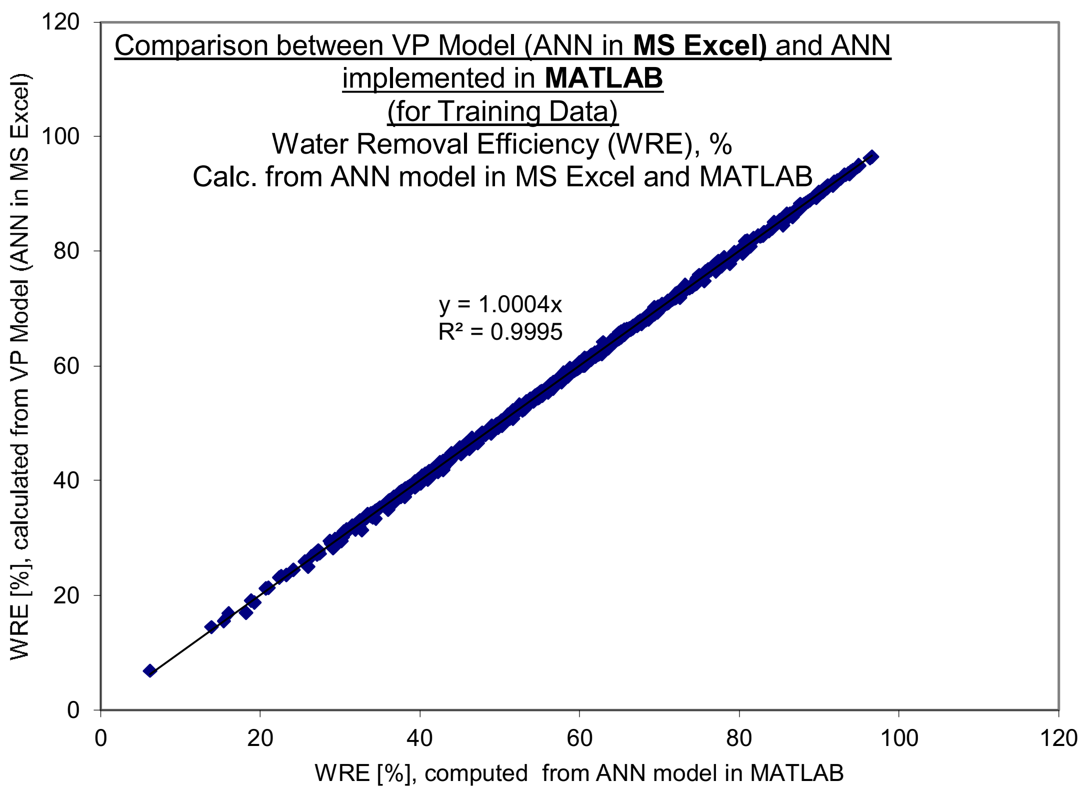

Fig. 5.7: WRE: VP Model (Excel) Output vs. MATLAB Output (for Training Data)

Observed data distribution and R^2 value (0.9995) indicate model performance as encouraging. As explained for the case of SRE, despite not implementing prominent algorithm like LM, performance of ANN model achieved through standard Solver in MS Excel appears encouraging from the above plot.

Further, as explained previously, output generated for Testing Data is likely to be a better indicative of model's performance. Therefore, the performance of VP Model (ANN developed in Excel) for Testing Data was further compared with the performance of the ANN model developed through MATLAB for the same set of Testing Data.

4. **VP Model/ANN in Excel (GRG2, this study) Output vs. ANN in MATLAB (LM, lit.) Output [for Testing data]**

   The graphical comparison of WRE [%] calculated from VP Model (the ANN model developed in MS Excel in this research) vs WRE computed from ANN model implemented in MATLAB for the Testing Data sets is as follows:

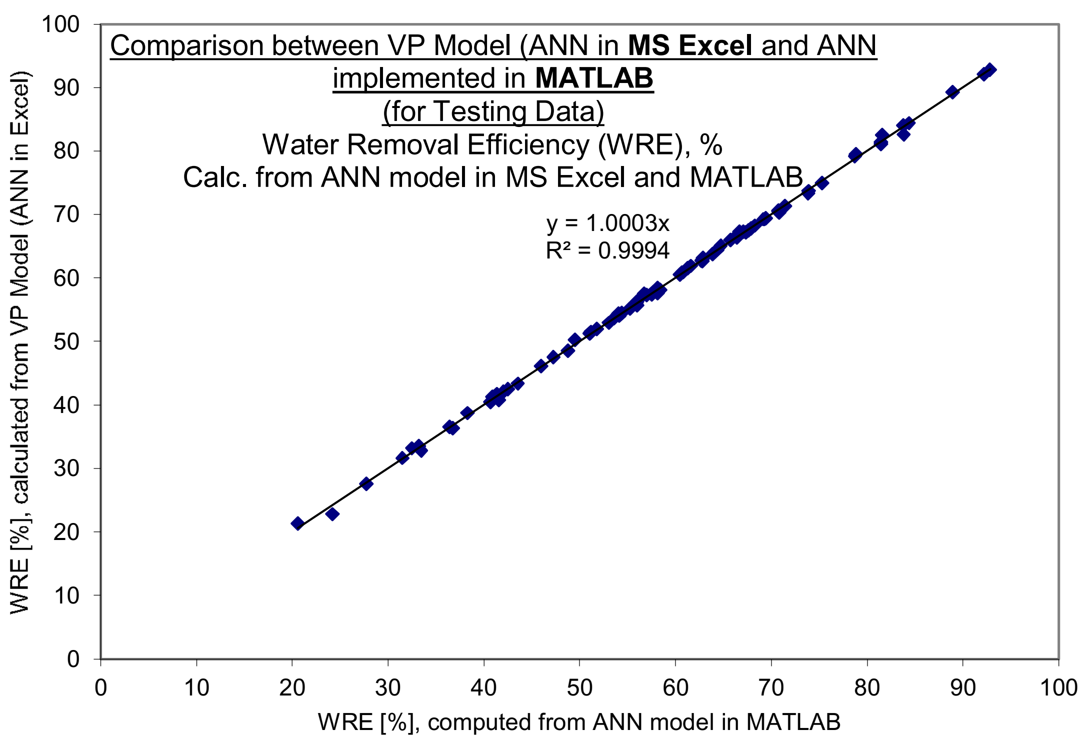

Fig. 5.8: WRE: VP Model (Excel) Output vs. MATLAB Output (for Testing Data)

Observed data distribution and R^2 value (0.9994) indicates model performance as encouraging.

Thus the performance of the Versatile Process (VP) Model (ANN process model developed in MS Excel in this research) appears comparable with the performance of the ANN model implemented through LM method in MATLAB, which enjoys popularity amongst the academician for scientific and engineering research.

5.2.2. Conclusions

On the basis of above results following conclusions are noted:

1. As the ANN process model's prediction is in agreement to the experimental values, it is inferred that ANN model has the potential to be used for modelling performance of a process (like DDP) which depends upon several parameters in a complicated manner. This reaffirms the previous finding.
2. It is demonstrated that ANN process model can also be successfully implemented in widely available MS Excel, in contrast to several previous ANN process modelling attempts, which utilized special codes / package like MATLAB. Though Levenberg-Marquardt (LM) algorithm available in MATLAB to train ANN model is considered better but the output of the ANN model trained in MS Excel using Solver is found in agreement with the output of ANN trained through LM to a satisfactory level in above cases. Such observation strengthens the notion that even though LM can offer better accuracy but for many practical problems it makes little difference, because the accuracy of the result obtained may be limited by experimental noise in the data.
3. MATLAB used in previous study, requires some license that is generally not available in PCs, and usage also requires skills in MATLAB, whereas current work implemented

ANN in MS Excel which is commonly available and widely used, and thus the current work extended the previous work by enhancing its applicability and outreach.

4. As development of ANN did not involve embedding any phenomenological information in it, and because even blank cells in MS Excel can be correlated to each other through powerful mathematical formula, development of ANN in MS Excel i.e. combining the benefits and versatility of ANN and MS Excel resulted into Versatile Process (VP) Model, that is, a model-based decision support tool in Excel that may serve variety of purposes. The current work was not limited to solve a specific modelling problem of DDP but to attain a versatile modelling tool / framework, which solved the specific problems, too, while attempting to explore modelling, simulation and optimization for DDP.
5. By virtue of being an ANN, the model may be populated, trained and tested for making simulation and optimization of DDPs of different makes present at different locations, as these are data driven model, not incorporating any physical / geometrical aspects of the plants, nature of fluid, barometric pressure etc. in the model.

Thus, this work, through novel means (by combining benefits and versatility of ANN and MS Excel), enhanced applicability of previous works in many ways.

5.3 Modelling Problem-2: Modelling of DDP to predict Wash Water flowrate

As was explained in section 4.2.2, the developed versatile process modelling framework / tool was used to model DDP for predicting wash water flowrate for given production flow rate (barrel per day), inlet temperature (^{o}F), inlet and outlet salt content (ppm) and chemical dosing rate (ppm). Performance of the new model is presented below. The modelling primarily involved populating VP Model with Training Data, and using Solver to attain 71 parameters (weights and biases) associated with the model based on the training data. Then this solved model was used to predict Wash Water flowrate [bbl/day] for the Testing Data. In the following sub-section results are discussed.

5.3.1. Results

To verify the performance of the model, output (predicted wash water flow rate) of thus developed model in MS Excel, was compared with known plant data. As discussed previously, coefficient of correlations (R^2) is widely used to review performance of ANN model.

1. **VP Model (this study) Output vs. Plant measured value (lit.) [for Training data]**

   The graphical comparison of wash water flowrate calculated from VP Model (the ANN model developed in MS Excel in this research) vs. plant measured value found in literature for each Training Data is presented as follows:

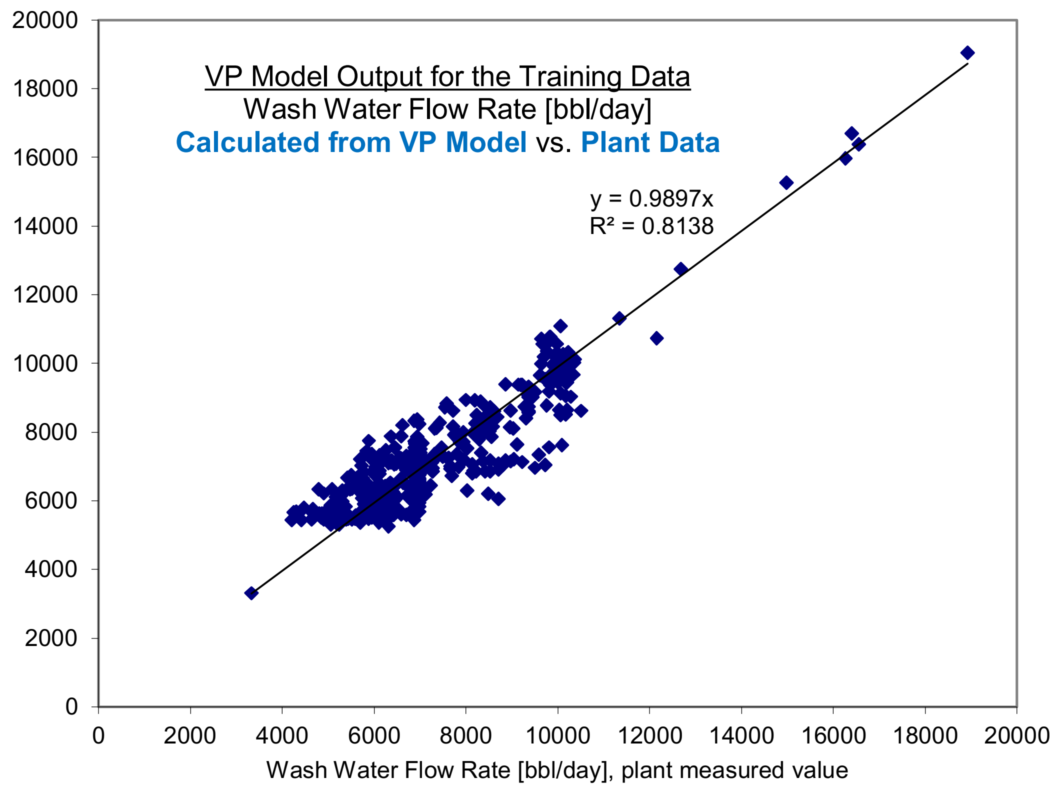

Fig. 5.9: Wash Water Flowrate: VP Model Output vs. Plant data (for Training Data)

Observed data distribution and R^2 value (0.8138) indicate model performance as encouraging. Also, it may be noted that MSE (mean squared error) determined for the value predicted by model as compared to the known experimental value is 0.002347, which is also considered indicative of the model's performance.

However, as Training Data represents that data which was used to solve the model to determine model's parameters (71 numbers), there remains a possibility that model simply "fitted" the data and did not "learn" appropriately to "generalize". Generalization is the main feature which is desired in any ANN process model. To cross-check, the "learning of the model to be able to generalize", further analysis is needed.

2. **VP Model (this study) Output vs. Plant measured value (lit.) [for Testing Data]**

   Wash water flow rate [bbl/day] calculated from VP Model (the ANN model developed in MS Excel in the current research) vs. plant measured value found in the literature for each Testing Data is plotted as follows:

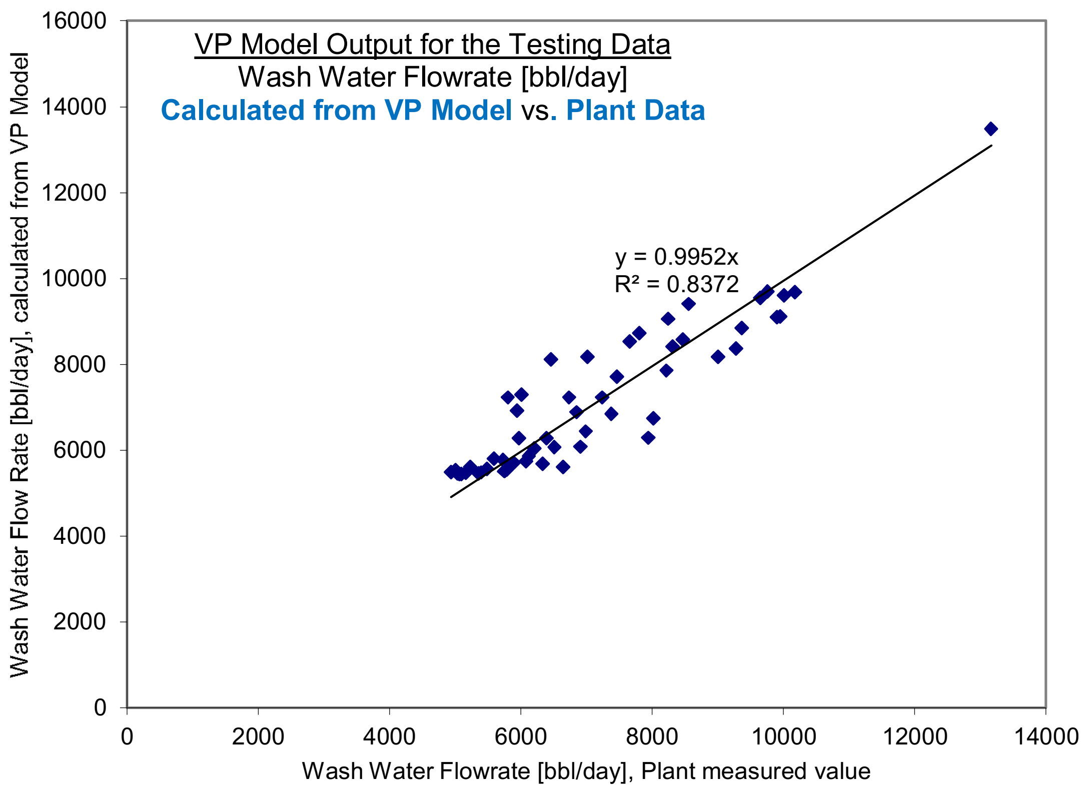

Fig. 5.10: Wash Water Flowrate: VP Model Output vs. Plant data (for Testing Data)

Observed data distribution and R^2 value (0.837), even for previously unseen data for the model, indicate model performance as encouraging.

To further analyze the performance of the model, wash water flowrate of Training Data sets obtained from plant was arranged in increasing order and it was compared with corresponding VP Model output to observe deviations.

3. **VP Model Output (this study) and Plant data (lit.) on the same vertical axis [for Training data]**

Wash water flowrate of Training Data sets obtained from plant was arranged in increasing order and it was compared with corresponding wash water flowrate predicted by VP Model on the same axis. The plot is as follows:

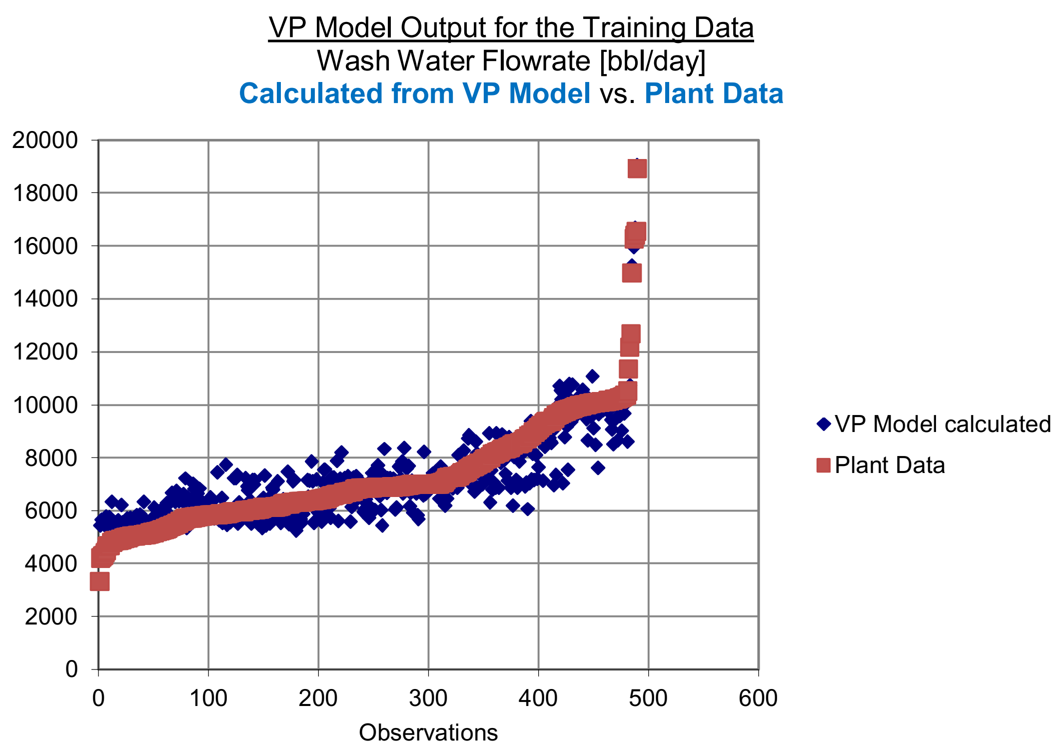

Fig. 5.11: WW Flow: VP Model Output vs. Plant data on same axis (for Training Data)

4. VP Model Output (this study) and Plant data (lit.) on the same vertical axis [for Testing data]

Wash water flowrate of Testing Data sets obtained from plant was arranged in increasing order and it was compared with corresponding wash water flowrate predicted by VP Model on the same axis. The graph is as follows:

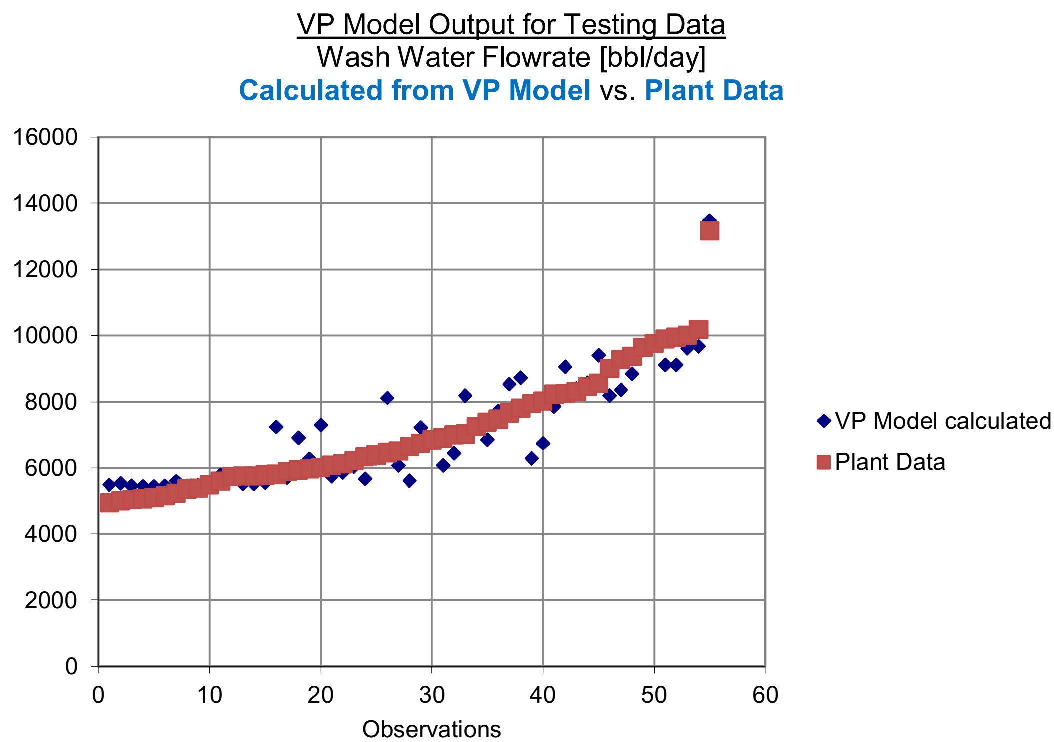

Fig. 5.12: WW Flow: VP Model Output vs. Plant data on same axis (for Testing Data)

Above plots shown in fig. 5.11 and 5.12 give an idea about the deviations of the wash water flowrate predicted by VP Model (ANN in Excel) for a given set of operating condition with respect to the actual plant measured wash water flowrate for the same set of operating conditions. Though there are deviations but still the performance of the model appears encouraging.

5.3.2. Conclusions

Above modelling problem-2 is included in the thesis to demonstrate the versatility of the achieved model as a modelling tool / framework by virtue of its being an ANN in MS Excel. The developed model which solved the specific case of predicting SRE and WRE as problem-1, was populated, trained and tested, as a modelling tool, to predict the wash water flowrate based on data for another set of dependent variables for problem-2.

Results presented in previous sub-section showed models' performance as encouraging. Overall prediction by model is seen somewhat comparable to the real plant data. It may be noted that to achieve best result, data need to be appropriately collected / generated and adequately pre-processed. However, in this research the focus was not to achieve the best numerical result for a particular modelling case but the focus was to attain a versatile process model, that is, a model-based decision support tool in Excel that may serve variety of purposes by virtue of its being ANN in MS Excel. Above result demonstrates attainment of such versatile process model, though achieved through the modelling related to DDP.

Thus, through the result of this problem-2, it is emphasized that the developed model may be populated, trained, and continually updated for various operating parameters of interest for *different* DDPs *irrespective of their different design / construction*, to be used as a decision support tool for optimizing plant operation.

5.4 Overall Significance of the Work

As is seen from the previous sections, that in this research, process models were developed for DDP, which is a crucial process at oilfield. These may be used at oilfield to ***simulate*** *various operating scenarios in order to support decisions for* ***optimizing plant's operation.*** Such exercise can be carried out in variety of ways through the models as these are basically ANN process models implemented in MS Excel. Therefore, these models inherently offer all those numerous benefits as well as great versatility which ANN and Excel do offer, particularly for solving optimization problems. As ANN is considered as universal approximator, and any work in MS Excel, if done with appropriate vision, can serve as a software tool without the need to repeat all the work again to solve similar problems, the current work was done to avail both features of ANN and MS Excel to extend the previous works to enhance its applicability towards actual deployment at plant. Thus, while designing the development of above models, instead of limiting the work to develop a model only to meet a very specific requirement, methods were incorporated to attain a versatile process model (named VP Model), that is, a model that can serve as a modelling tool or framework to handle variety of problems, even beyond the realm of DDP.

Thus developed VP Model may be populated, trained, and continually updated for various operating parameters of interest for different DDPs irrespective of their different design / construction, to be used as a decision support tool for optimizing plant operation.

Further, in view of its wider significance, it may be noted that the research was meant not to develop any software to do something to replace any other commercial software which are already much developed to do what they are reputed for, but it was intended to highlight the power and versatility of ANN and Excel and combine them together as a widely available basic tool to handle variety of real-life decision making problems, and complementing the existing tools to derive maximum benefits from all these reputed software tools. To explain further, even for those items which can be modelled well in Aspen Hysys, Unisim or Pro II like reputed process simulation software, these advanced software tools can be used to generate large number of data sets of good quality to train VP Model (ANN process model developed in MS Excel in this research). Such appropriately trained, tested and validated model may be incorporated in real-time optimization decision support system to avoid spending precious computational power and time in performing loops of iterative calculations involved in phenomenological process models, because ANN once trained, tested, validated and deployed, need only algebraic equations to compute output, and MS Excel is also considered quite good in returning instant result for practical purposes.

Also the importance of accessing even rigorous process model built in Hysys through MS Excel by plant personnel has already been recognized by the developers of such rigorous tool. MS Excel removes many types of barriers (which is not the topic of the current discussion) those otherwise might inhibit actual usage of the model in optimization of plant operation. At practical level, having a powerful and sophisticated process model at plant is one thing and availing benefits of it is another thing. Moreover, even utilizing such proprietary model-based decision support tool would require, each time, extra-ordinary computation power (derived through license) and some time (in accessing license even if the user is lucky that at the time of accessing the model network license is not being used by other user in the company). Such things may cause reluctance for using it regularly for model-based decisions; whereas one can keep open MS Excel files, and instantly check impact of some variation in operating parameters in VP Model (an ANN process model built in MS Excel), well trained and tested through the data generated by state-of-the art softwares like Hysys and Pro II.

Similarly, MATLAB or any other well developed tools may be utilized by the 'experts' available in the company to identify the optimal topology of the ANN taking into consideration variety of issues, because well-developed commercialized tools are supposed to have appropriate features for carrying out such actions, and then based on the outcome of such work, VP Model (an ANN model in MS Excel) may be used for implementing the particular application, to capture the implicit relationship between input / output to ensure its wider applications (without needing any special license other than what is generally available in all PCs).

Thus, as explained above the purpose of VP Model (an ANN model in Excel) developed in this research should not be seen in limited sense that such model is to be applied only in those cases where direct modelling provisions are not present in the state of the art softwares, rather such basic tool offers certain benefits from the perspective of usage.

Having mentioned the above, overall significance of the current work is summarized below:

1. The developed VP Model may be populated, trained, and continually updated for various operating parameters of interest for different DDPs irrespective of their different design / construction, to be used as decision support tool for optimizing plant operation.

   For deploying such model at real plant appropriate measures may be taken to generate sufficient number of data sets envisaging entire operating range expected during its service life, while conducting Performance Acceptance Test for any newly installed DDP and plant historian / lab test data can be used for old facilities.

2. The model may be coupled with widely used proprietary process simulation softwares like Aspen Hysys (which does not contain direct modelling provision for DDP) that allow input /output interface with Excel.

3. Though versatility of the model was demonstrated in reference to dehydration and desalting process, but it may be utilized as a versatile process model (VP Model) to correlate any 5 independent process variables to any dependent variable for a given data set. Thus, it may serve as a fundamental tool / framework for further process studies and development even beyond the realm of DDP.

4. Such model can be utilised to save expenditure on computation power (e.g. usage license), time (taken by proprietary softwares to access license each time it is opened, to converge each time any parameters are varied) and prevent wastage of resources (in view of reluctance to use it), by using these proprietary softwares to simulate data needed to train such model and to ascertain optimum design of the ANN by experts before actual deployment for wider usage by the end users.

5. The model may be customized for different applications by end users without having any specialized computer program coding skills or advanced mathematical skills except usual working knowledge of excel
6. The model may be utilized to demonstrate the concept and application of Artificial Neural Network, in general, and MLP in particular, for education, as a short course. Transparent implementation in Excel allows user to develop clarity about the underlying concepts unlike other software packages developed using some software code**.**
7. The model does not require special expensive software license other than what is generally available in any computer in industries and academic / research institutions.
8. The model may help to promote ANN, as an 'operations research tool" like other statistical tools for researchers / scholars.

In short, the model may be easily adapted for variety of chemical engineering applications: for industrial as well as academic needs, especially, pertaining to process modeling, simulation and optimization as well as process control. However, it is important to consider the current work as basic and further enhancement would be needed for actual deployment in the industry. Needless to mention, it shall suffer from all the limitations which ANN as well as MS Excel spreadsheets are prone to. Irrespective of any limitations, its ease in deployment for decision support would definitely help rather than taking decisions without any such model but completely arbitrarily. Further, it can be used in the development of various advanced decision support systems emerging in the Oil & Gas industry as a building block or can enhance their use as separate block by coupling with them; being in Excel it is likely to be coupled with any emerging tool very conveniently.