MATERIALS AND METHODS

The chapter presents the specification of hardware and software used to complete the research work. Also, it describes development of the versatile process model (named VP Model) in MS Excel utilizing the concept of ANN.

3.1 Materials

Specifications of hardware and software utilized for the research are listed below:

<u>Laptop</u>

Manufacturer: Lenovo

Model: Lenovo Win7 PC

Processor: Intel(R) Core(TM) i5-2430M CPU @ 2.40 GHz

Installed Memory (RAM): 4.00 GB

System Type: 64-bit Operating System

<u>Operating system</u>

Windows7 Home Premium Operating system

<u>Microsoft Excel</u>

Version: 14.0.7116.5000 (32-bit), Part of Microsoft Office Professional 2010

3.2 Methods

3.2.1. Research Design

The trigger for this research was to develop process model for DDP that may be used at oilfield plant to simulate various operating scenarios in order to support decisions for optimizing plant's operation. However, the research was not limited to simply develop a model for a specific process scenario of dehydration and desalting process, rather it explored to attain a versatile process model that may serve as a modelling tool as well as a decision support tool, keeping in view ANN's strength as a universal approximator and MS Excel's versatility and outreach.

Therefore, first by applying the techniques of ANN model development, an MLP type artificial neural network was developed in MS Excel to attain a general tool what is termed here as Versatile Process (VP) Model. Basically, this neural network correlated 5 independent process variables to one dependent process variable.

Thereafter, it was populated, trained and tested for process modelling applications. Two applications, pertaining to modelling, simulation and optimization of DDP are mentioned in this thesis.

3.2.2. Development of the Model

Piuleac, Rodrigo and others [43] presented a modelling methodology for modelling of electrolysis process in 10 iterative steps by using neural networks, as a promising alternative to phenomenological models. These have been listed as: (i) model purpose, (ii) modelling context: scope and resources, (iii) system conceptualization, data specification, prior knowledge, (iv) selection of model features and family, (v) selection of model structure and parameter values, (vi) choice of estimation performance criteria and technique, (vii) identification of model structure and parameters, (viii) conditional verification and diagnostic testing, (ix) quantification of uncertainty, and (x) model evaluation or testing.

Further, Basheer and Hajmeer [10] suggested that the development of a successful ANN project constitutes a cycle of six phases, namely: (i) the problem definition and formulation,

(ii) system design, (iii) system realization, (iv) system verification, (v) system implementation, and (vi) system maintenance.

Objective of these steps is to find an optimal solution for a particular problem, and deployment of an ANN model through a complete project cycle. This study has utilized some of the relevant steps and phases according to the purpose and scope of this work, which have been further clarified in the thesis at appropriate places.

In general, there exist a large number of alternative ANN physical architectures and learning methods, and therefore design of optimal neural networks is problematic. A heuristics that utilize domain knowledge to produce an artificial neural network to attain optimal output performance is found in [52]. The above mentioned heuristic ANN design approach involves the following steps:

i. Knowledge based selection of input values
ii. Selection of a learning method, and
iii. Design of the hidden layers (numbers and nodes per layer)

However, it is not intended to limit this research work, to achieve an optimal performance ANN to model DDP for predicting a specific dependent variable, e.g. salt removal efficiency, based on certain independent variables e.g. temperature, chemical injection, fresh water injection, mixing time, and settling time for a plant of particular make at a particular location. Instead, this work has a broader perspective. In this work, the idea is to explore building a versatile process model that may serve as modelling tool or decision support tool to handle various process issues: pertaining to DDP optimal operation, in particular, and which model, owing to its inherent versatility may remain useful for many other purposes, in general, e.g. as a building block for emerging decision support systems.

To further clarify, previous studies have already suggested some specific ANN process model for a very specific problem pertaining to DDP for a specific set of data [3, 15, 25, 41]. The purpose of this study is to extend the previous work towards actual deployment of such models at plant. A step towards attaining it was to develop a basic versatile model in MS Excel which would serve the purpose in many ways. In this way, the purpose of this study does not require, at this stage, to focus on finding optimal ANN to be used for a specific reason, for a specific input / output combination for a given data set, but is focused towards attaining a versatile model which can serve as a modelling framework, which may be used as a tool during day-to-day decision making. Every stage of any ANN development project requires trial and error [44] and MS Excel offers several options to perform such trial and error, which may be used to find optimal solutions at some advanced stage.

In short, the present work attempted to utilize the method and findings used in previous studies including data and ANN configuration from the literature. Particularly, one of the previous studies, which has been reported in [41, 42] was used as a baseline to attain a reasonable and beneficial extension of the previous work. There are several reasons, including availability of enough data needed to complete and evaluate the current work, and the ANN configuration selection, that made it reasonable to consider it as a baseline case, which are further indicated at appropriate places in this thesis to maximize conciseness of this thesis, as far as practicable.

With the above clarity regarding the perspective with which ANN development was carried out in the present work, some of the core issues that were handled in the development of a versatile process model by developing an ANN process model in MS Excel are indicated below.

ANN type selection:

There are many types of ANNs depending upon several factors, e.g. the topology of the network, direction of dataflow, type of network learning etc. A taxonomy is provided by Huang and Zhang in [24].

The most general form would be fully connected networks in which every node is connected to every node in both directions. Evidently, for a network with n nodes there would be n^2 weights. Despite its generality, such model is seldom used because of large number of parameters [34].

Another type of networks are layered networks in which nodes are partitioned into subsets, called layers, with no connections that lead to from layer j to layer k if j > k. Intra-layer connection may exists. However, if intra-layer connection does not exist, it can be termed as acyclic network, a special case of layered network [34].

Further, networks that are not acyclic are recurrent networks. Clearly, computational processes in acyclic networks would be simpler than those have inter-layer connections, or recurrence [34].

Moreover, such acyclic networks in which a connection is allowed from a node in k^{th} layer only to nodes in $k+1^{th}$ layer are referred to as feedforward networks; though sometimes the term feedforward is used to refer to layered or acyclic networks [34].

Also, besides topology and connections, ANN models may also be differentiated on the basis of the mode of training, i.e. supervised and unsupervised. These two are very broad classifications and there are numerous alternatives [52].

Thus, there are numerous types of ANNs; some of the famous neural networks (NNs) are [35]:

i. Hopfield Network
ii. Kohonen Network

iii. RBF (Radial Basis Function)

iv. MLP (Multilayer Perceptron)

According to a study [35], network utilization was: MLP: 81.2%, Hopfield: 5.4%, Kohonen: 8.3%, and others: 5.1% of times.

MLP is a feedforward neural network in which each neuron output is connected to every neuron in subsequent layers, with no connections between neurons in the same layer. For the subject development, MLP is the preferred type ANN to reap those benefits, owing to which, this type is being widely used [11, 19]. Some of the previous researchers used MLP for building model for DDP also [3, 25, 41]. In the aforementioned "baseline" study also, MLP was used [42]. According to Vafaei and Eslamloueyan [50], MLP networks are the most commonly used one for the function approximation. This type of network has the potential of approximating most types of nonlinear functions, irrespective of how much they are complex. Thus, for the purpose of this work, too, MLP type neural network was selected for further process modelling.

Selection of topology for the MLP:

Number of inputs / outputs:

First, the number of inputs (independent variables) and outputs (dependent variables) to be correlated through the MLP should be decided in order to select the topology. *A priori* knowledge regarding the process would help in deciding the outputs and inputs to be mapped through the ANN. Too many input and output parameters with sparce training data sets may affect the learning process and generalization feature of the ANN. Therefore, it is desired to optimize the number of input and output parameters. Couple of techniques to carry-out such exercise is listed in [44].

From the literature it is known that while solving a process problem the entire process need not be modelled through a single ANN having multiple inputs and multiple outputs (termed as MIMO configuration). Rather, for each output different independent ANN may be used. As reported in [4], it was done for developing an optimizer, using ANN, for optimal operation of a separation plant which involved several process equipment and several independent and dependent variables. Such single output ANN may enhance the performance of ANN [4].

Therefore, in order to attain a versatile model that may be used as a modelling tool or a building block for advanced decision support system, single output ANN could be a preferred choice, as single output ANN can be used to predict more than one dependent variables of interest by using one ANN for each dependent variable of interest.

Some of the previous researchers also used single output ANN for building model for DDP [3, 15, 25, 41]. In fact, in the studies reported in [15, 25] only one output was of interest whereas in the study reported in [41], two outputs were of interest but separate ANN with single output was used for determining each dependent process variables of interest. In the aforementioned "baseline" study also, single output MLP was used [42]. Thus in this study also single output ANN was selected for further modelling.

On the other hand, in previous studies [15, 25, 41], number of inputs for modelling DDP were different. 7, 3 and 5 inputs were considered for ANNs development as reported in [25], [41], and [15], respectively. Further, in [6], author has presented an MLP model of 3 input, 3 nodes at hidden layer and 3 output besides bigger networks, for modelling of HDS, which is a key process at refinery, to predict removal of sulfur from naphtha. According to the literature [6], prediction of the smaller 3 x 3 x 3 model was good.

In view of the above, to attain development of a versatile process model, 5 inputs, which is also the number of inputs that was considered in the baseline study [42], was decided, expecting that this number was neither too small nor too big, and would be able to handle several modelling cases. However, as the work is being done in MS Excel, it can be customized for different needs, within certain limitations inherent to MS Excel.

Such MLP, selected for the present work, is termed as Multiple Input Single Output (MISO)

Number of hidden layer:

There may be one or more hidden layer of neurons. Various researchers have established the universal approximation property of feed forward neural networks with various activation functions. In most of the studies, one hidden layer is considered sufficient for approximating any continuous or measurable functions [20]. Therefore, one hidden layer is the preferred choice in view of the purpose of the network development. Amongst the previous studies [3, 15, 25, 41] related to DDP modelling through ANN, one study [25] considered two hidden layers of neurons; whereas, even for modelling complex HDS process, only one hidden layer was considered [6]. All these ANN process models, wherein there was only one hidden layer, reportedly performed well. For the purpose of the current work, it was thought appropriate to select single hidden layer, which is also the number of hidden layer that was considered in the baseline study [42].

Number of neurons (i.e. nodes) in the hidden layer:

Number of neurons in a hidden layer may affect the outcome of the network training. In case of fewer neurons in the hidden layer, the network may not be able to differentiate between complex patterns, leading to only a linear estimate of the actual trend. However, in case of too many neurons in the hidden layer, the network may follow the noise in the data leading to poor generalization for untrained data [10]. Also, with increasing number of neurons in the hidden layer, training may become time consuming [10]. In the literature, some authors provide rule of thumb relating the number of hidden nodes to the number of input and output variables and the number of training patterns [44]. However, the most popular approach,