IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 2, Ser. IV (Mar - Apr 2019), PP 56-64
www.iosrjournals.org

Nonlinear Auto-Regressive with eXogenous (NARX) input model for Liquid-Gas Dehydration and control Systems: Data Driven Modelling Approach.

Akaninyene J. Ntuk[2], Enyenihi H. Johnson[1], Uduak I. Akpan[1]
[1]Department of Elect/Elect Engineering, Akwa Ibom State University, Ikot Akpaden, Nigeria
[2] Department of Elect/Elect Engineering, Exxon Mobil, Nigeria
Corresponding Author: Akaninyene J. Ntuk

Abstract: *The operation and management of dehydration systems for the determination of liquid and gas dew points is becoming increasingly complex. Several methods for prediction of the liquid and gas dew points have been developed by researchers. These methods differ in the available inputs, their classification and the horizon of the prediction. In this paper, a nonlinear auto-Regressive with eXogenous (NARX) input model has been proposed. The data driven modelling approach was adopted in the proposed model with System Identification toolbox in MATLAB software. Implementation and simulation of the model which was achieved by feeding the measured input data into the developed model shows that the model is able to reproduce the dew-point (measured output) of the dehydrator bed, and hence gives a good prediction of it.*

Keywords: *final prediction error, loss function evaluation, model estimation, Modelling Structure, system identification*

Date of Submission: 15-04-2019 Date of acceptance: 30-04-2019

I. Introduction

The operation and management of dehydration systems for the determination of liquid and gas dew points is becoming increasingly complex. Several methods for prediction of the liquid and gas dew points exist in some literatures. They differ in the available inputs, their classification and the horizon of the prediction. Some of these methods are based on linear models such as Linear Regression (LR), Auto-Regressive Moving Average (ARMA) and Auto-Regressive (AR) [1]. However, because of the nonlinear behaviour of the liquid and gas dew points, researchers propose several nonlinear models based on wavelet-based methods, fuzzy models, Adaptive Neural Fuzzy Inference Systems (ANFIS) Random Forests (RF), k-Nearest Neighbours (KNN) and Artificial Neural Networks (ANN) [2-4]. In this paper, the work of the actual data driven modelling will be carried out with System Identification toolbox in MATLAB program. The dehydrator bed model will be developed with the modelling structure of the Non-linear Auto-Regressive with exogenous input (NARX). However, several dehydrator bed models were developed for other model structures such as linear Auto-Regressive with exogenous input (ARX) model, state-space model, Box-Jenkins (BJ) model and Autoregressive Moving Average with exogenous inputs model (ARMAX).

II. Molecular Sieve Dehydration Bed Regeneration Process

Once a molecular sieve dehydrator bed that is on-line completes the dehydration cycle, it immediately switches and commences the process of removing the moisture and other heavier components adsorbed by the molecular sieve during the process of the natural gas stream dehydration [3,8,10].

However, Akpabio and Aimikhe in [1] pointed out that the regeneration process is achieved by flowing a hot gas from the bottom to the top of the molecular sieve dehydrator bed which is in an opposite direction to the direction of flow of the wet natural gas stream during the dehydration process. More so, the regeneration of solid desiccant of which the molecular sieve is a type can be successfully carried out in two ways namely:

- The thermal swing regeneration and
- The pressure swing regeneration

Of the above mentioned ways of carrying out regeneration of the dehydrator beds, the thermal swing regeneration is the easiest and more pronounced approach. In a large scale practice when using thermal swing approach, the temperature at which the hot gas is introduced into the dehydration bed is determined using the properties of the molecular sieve [1]. Furthermore, Barrow and Veldman in [3] clearly state that the temperature for the regeneration process of a molecular sieve should be higher than 450 °F and it is usually about 550 °F. However, [1] reveals that if the regeneration gas temperature is too high, it will make the dehydration adsorption capacity reduce faster and if the regeneration gas temperature is too low compared to the design, it will cause the bed to still have water in the adsorbent at the end of the bed regeneration cycle which can lead to carrying over of water into the cryogenic processing plant. The dehydration bed regeneration temperature is very important to the operations of the molecular sieve dehydration beds as it can increase or decrease the life span of the molecular sieve. The pressure swing regeneration approach as explained in [1] is an alternative to the thermal swing approach. It is applied in situations where:

- Adsorption process is carried out at high pressure
- Regeneration process is done at a low pressure and
- Dehydration beds with shorter cycle times owing to the fact that changes in pressure as a process parameter occurs faster than changes in the temperature.

According to [5-7], to achieve the desired and an accurate dew point of the natural gas stream leaving the dehydrator prior to the cryogenic process, certain process parameters must be maintained within the specified ranges in the natural gas stream to the molecular sieve dehydrator bed. Such parameters include:

- the feed gas temperature (°F)
- the feed gas flow-rate (MMscfd)
- the feed gas pressure (psi) and
- the pressure drop within the dehydrator beds (psi)

These parameters are critical to the operation of the molecular sieve dehydration bed in achieving a good dew point for the natural gas liquid production. However, the pressure drop and the feed gas flow-rate tends to be the two most critical parameters that require strict monitoring during the operations of the dehydrator beds as they have the potential of causing breakdown of the molecular sieve into finer particles, thus affecting the water content of the natural gas stream exiting the dehydrator bed. Table 1 shows a typical operating parameter of a molecular sieve dehydrator beds as pointed out in [7].

Table 1: Typical operating conditions for molecular sieve dehydration beds [1]

Feed Rate	10 - 1500 MMscfd
Superficial Velocity	Approximately 30 - 35 Ft/Min
Pressure Drop	Approximately 5 psi, not exceeding 10 psi
Cycle Time	4 - 24 Hours; 8 or a multiple thereof is common
Temperatures and Pressures	
Adsorption	Temperatures: 50 to 115 °F
	Pressures: 0 to 1500 psig
Regeneration	Temperatures: 400 to 600 °F
	Pressures: adsorption pressure or lower

III. System Modelling Structure

The choice of model structure is crucial in data-driven modelling as the replication of the plant model or process solely depends on it. Considering a simple structure with a static gain **K** mapping a set of input data u(t) to the output data y(t), the output can be expressed as;

$$y(t) = Ku(t) \tag{1}$$

The structure may or may not be able to approximate a complex system such as that of the molecular sieve dehydrator bed with a reasonably level of accuracy. Although there are banks of model structures available in System Identification, the dew point of natural gas leaving the molecular sieve dehydration bed for natural gas liquid production would first be estimated with a Nonlinear Auto-Regressive with exogenous inputs (NARX) model as suggested by the advice command. The structure of this model is written as shown in the generalized equation;

$$y(t) = f(y(t-1), \dots, y(t-na), u(t-nk), \dots, u(t-nk-nb+1)) \tag{2}$$

Where:

f is a function that relies on known number of previous input u and output y,

na is the number of past output terms used to predict the current output,

nb is the number of past input terms used to predict the current output and

nk is the delay from the input to the output, specified as the number of samples

The Diagram of Nonlinear Auto-Regressive with eXogenous (NARX) input model
The block diagram of the NARX model is shown in Fig. 1**.**

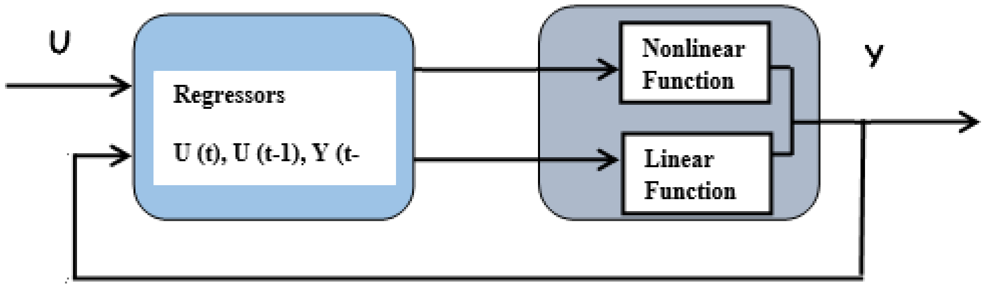

Figure 1: block diagram of nonlinear auto-regressive with exogenous (NARX) input model

The NARX model is made up of standard and custom (if specified) regressors, linear and nonlinear functions. NARX model first, computes regressors from the current and past input values and past output data. Regressors are simply delayed inputs and outputs, and could be written as shown in Fig. 1. Unless it is specified, all regressors are inputs to both the linear and the nonlinear function blocks of the nonlinearity estimator. The nonlinearity estimator block can either use one of the functions or a combination of nonlinear and linear functions for transforming the regressors to the model output. There are several nonlinear estimators such as tree-partition or binary tree, wavelet networks, sigmoid networks and multi-layer neural networks.
In this research, a tree-partition estimator is used for the model estimation and its function is given by the form;

$$y = F(x) \tag{3}$$

Where: F is a piecewise-linear function of x, y is scalar, and x is a 1-by-m vector of the model regressors. The network or binary tree is presented with the equation

$$F(x) = d + xL + (1, x)C_k \tag{4}$$

Where:
L is 1-by-m vector, d is a scalar called output offset, common for all elements of the partition, C_k is a 1-by-(m+1) vector and the subscript k represents each node of the tree.
In further optimization of the NARX model parameters after estimation, the wavelet estimator is use and it structure is as shown in (5)

$$F(x) = (x - r)PL + \sum_{s=1}^{s=ns} a_s\, f\left(b_s\left((x - r)Q - c_s\right)\right) + \sum_{w=1}^{w=nw} a_w\, g\left(b_w\left((x - r)Q - c_w\right)\right) + d \tag{5}$$

Where:
$f$is a scaling function,
g is the wavelet function,
P and Q are m-by-p and m-by-q projection matrices, correspondingly.
Also, r is a 1-by-m vector and represents the mean value of the regressor vector computed from estimation data, while d, a_s, b_s, a_w, and b_w are scalars with the s subscript for scaling parameters, and w subscript for wavelet parameters. L is a p-by-1 vector, c_s and c_w are 1-by-q vectors.
f and g are radial functions given as

$$f(x) = e^{-0.5xx} \tag{6}$$

$$g(x) = (N_r - xx)e^{-0.5xx} \tag{7}$$

Model Evaluation Criteria: Models in this project were evaluated based on the following criteria:
- Fitness (**FIT**),
- Loss Function (**V**),
- Final Prediction Error (**FPE**),

Fitness Evaluation Criteria: Models are evaluated based on **FIT** values calculated as follows:

$$\mathrm{FIT} = \left(1 - \frac{|y-y'|}{|y-y''|}\right) \times 100 \quad (8)$$

Where: variable y is the measured output, y' is the predicted model output, and y'' is the mean of the measured output y. A 100% of fitness depicts that the estimated model perfectly captured the system dynamics, and 0% of fitness indicates a very poor fit.

Loss Function Evaluation Criteria: The Loss Function **V** is well-defined by the following equation:

$$V = \det\left(\frac{1}{N}\sum_1^N \varepsilon(t,\theta_N)\left(\varepsilon(t,\theta_N)\right)^T\right) \quad (9)$$

Where, θ_N represents the estimated parameters and ε is the output error. The model structure with the lowest **V** shows the best model.

Final Prediction Error (FPE) Evaluation Criteria: FPE is written as

$$\mathrm{FPE} = V\left(\frac{\left(1+\frac{d}{N}\right)}{\left(1+\frac{d}{N}\right)}\right) \quad (10)$$

Where V is the loss function, d is the number of estimated parameters, N and is the number of values in the estimation data set. The best model produces the least FPE.
However, the evaluation criteria are critical and are used in the selection of the Non-linear auto-regressive with exogenous input model.

IV. Selection of NARX Model Orders

The next step, after selecting a suitable model structure either from expert knowledge of the plant or intuition, was to configure the model for estimation. This was done by specifying model orders and delays. In research, na denotes number of past output terms used to predict the current output, nb denotes number of past input terms used to predict the current output and n_k is delay from input to the output in terms of the number of samples. In order to select a suitable model order, the loss function evaluation criteria was used and various choices of delays were evaluated. However, the structure with the smallest loss function depicts the best model order. Selecting model order of (na = 1:4, nb = 1:4) and trying out time delay between 1 and 10, the loss function (V) for the various models were calculated. Consequently, several combinations of model orders were generated as a 256000 by 7 matrix using the estimation and validation data sets. Amongst these, the best model order was found to be [3,4,2,3,2,1,9]; where the first digit being 3 in the row denotes na = 3, the next three digits being 4,2 and 3 denotes nb = [4 2 3] and the last three digits being 2, 1 and 9 stands for nk = [2 1 9]. Furthermore, the program coding for the model order selection is as shown in Appendix I, lines 151 - 170.

Model Estimation

In training or estimation of the model for the dew point of the natural gas, half of the obtained and pre-processed data were used in order to capture the entire attribute present in the data and to replicate a suitable representation of the data. Using model order of [3,4,2,3,2,1,9], where na = 3, nb = [4 2 3] and nk = [2 1 9] as a guide, ten model orders were selected by trial and error method. The experiment and estimation of the ten NARX models were carried out using tree partitioning as the nonlinear estimator. Fig. 2 Plot of estimated Nonlinear Auto-Regressive with eXogenous (NARX) input Models with 1-step a-head prediction and the results in a percentage best of fit.

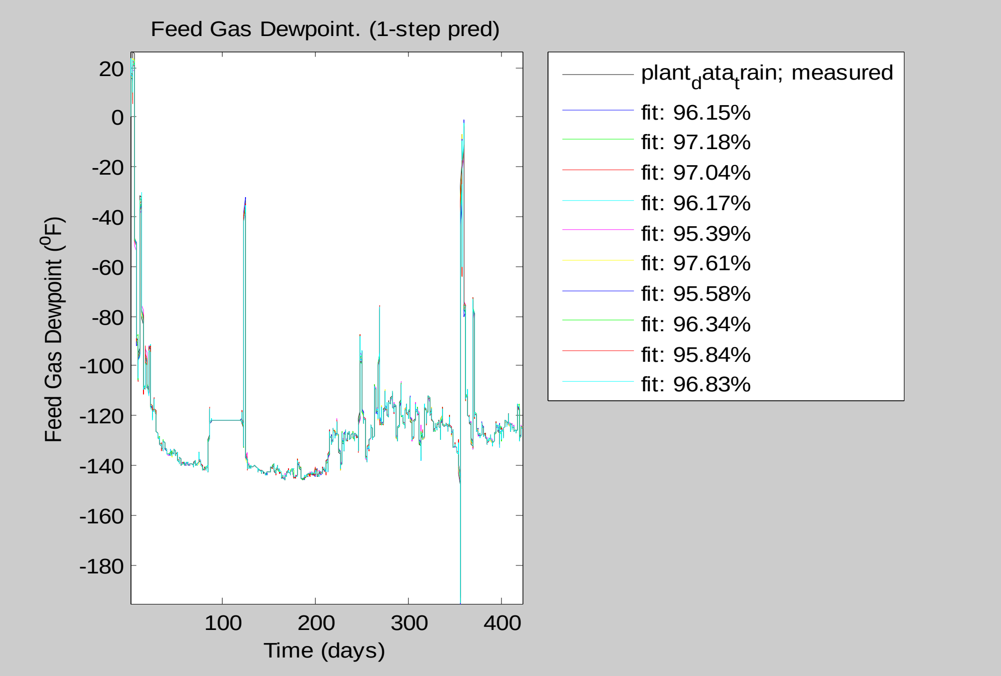

Figure 2: plot of estimated nonlinear auto-regressive with exogenous (narx) input models

Model Validation

It is necessary to examine the models-output plot to see how well the models' outputs match the measured output in the validation data set. The validation of the estimated ten non-linear auto-regressive with extra input models were done using a set of program codes. The plot of percentage of fitness of the ten non-linear auto-regressive with extra input models after validation is as shown in Fig 3.

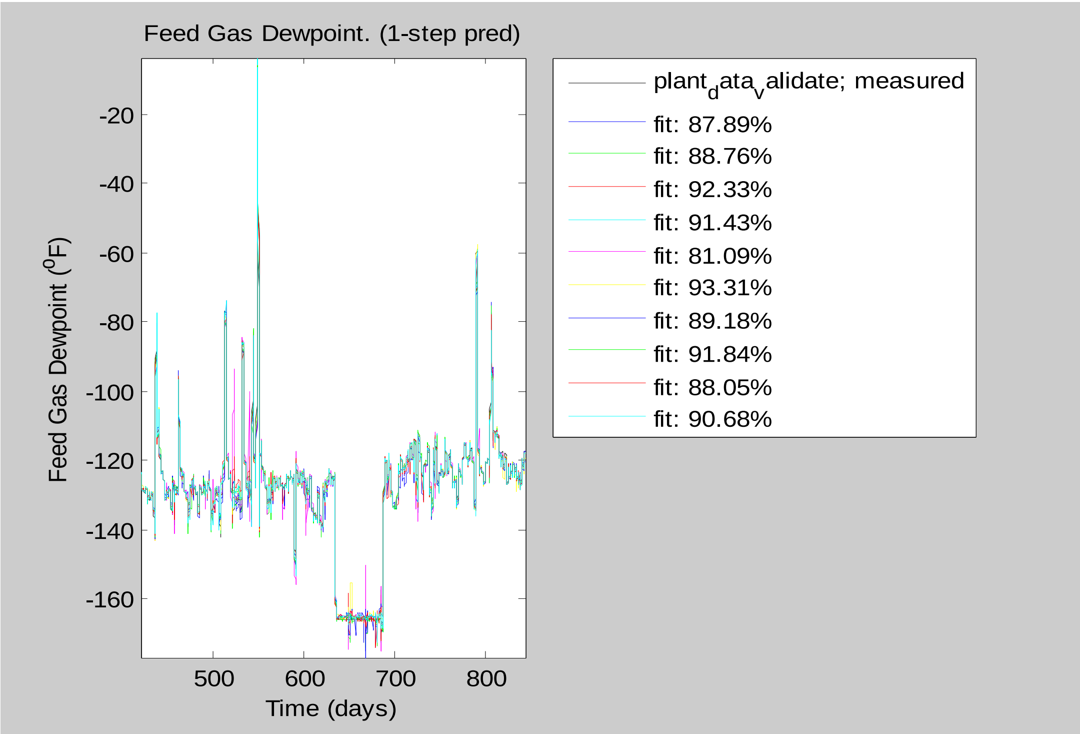

Figure 3: plot of validated nonlinear auto-regressive with exogenous (NARX) input models

Model Parameter Estimation and Optimization

In estimating and optimizing the Nonlinear Auto-Regressive models with exogenous inputs (NARX), the MATLAB inbuilt function known as **pem** was used alongside with the wavelet nonlinear estimator to perform parameter estimation and optimization. This function adopts parameters that produce the best fit, by minimizing errors between the measured and the modeled outputs as well as the loss function during the estimation process. Although, there are several search methods available in System Identification that are used for iterative parameter estimation, which include the Grid Search method, Gradient Search method (Steepest Descent-Newton, Quasi-Newton, Levenberg-Marquardt) and Random Search method (Bremermann Optimizer), this work uses Gradient search method with a Gauss-Newton subspace to estimate and optimize the model parameters based on prediction error criteria. The graphs in Figs. 4 and 5 illustrate the level of fitness of the models to the measured estimation and validation data.

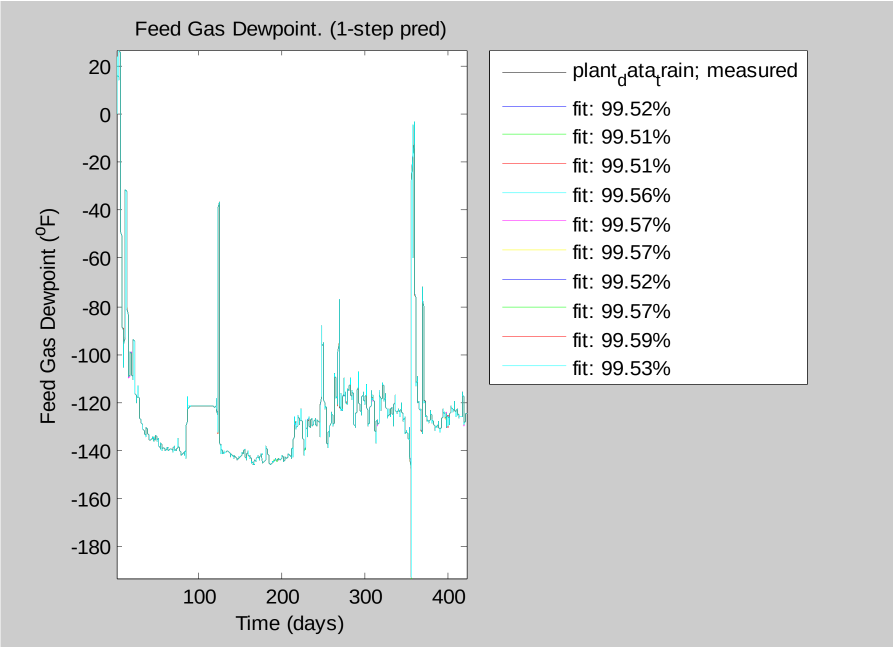

Figure 4: plot of model parameter estimation of NARX input models

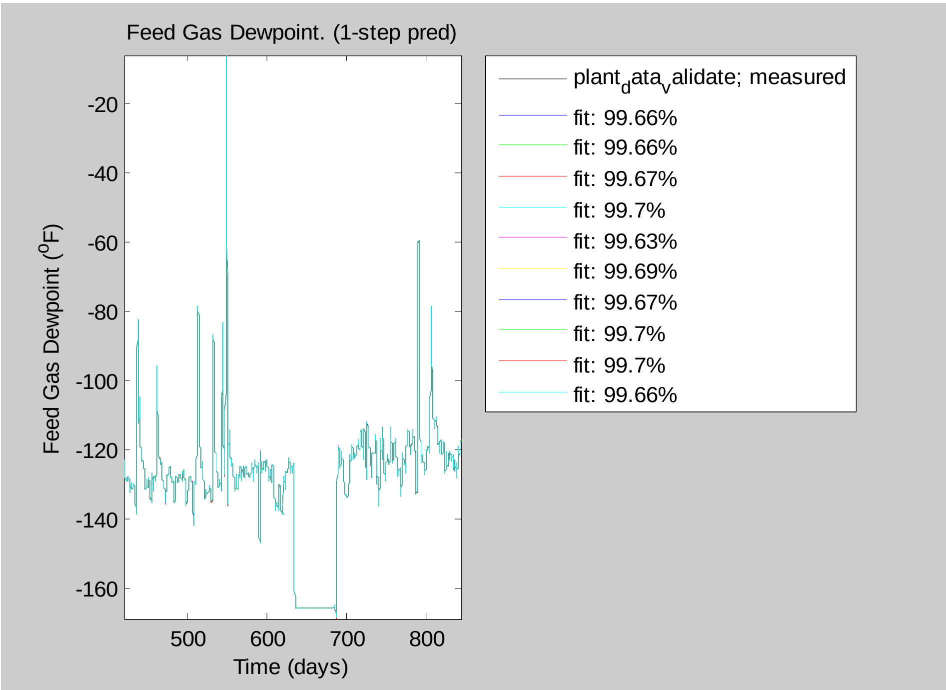

Figure 5: plot of model parameter validation of NARX input models

Simulation and Cross-Validation

The first step in this regard is to compare the NARX models with linear model structures to verify whether the suggestions given by the "advice" command in System Identification toolbox was appropriate. Four different models were estimated for this purpose and they include Auto-Regressive with exogenous input (ARX) model, state-space model, Box-Jenkins (BJ) model and Autoregressive Moving Average with exogenous inputs model (ARMAX). Ten models were estimated for each of these structures. however, the structures of ARX and state-space models developed were based on the on the model orders used in developing the NARX models, while BJ and ARMAX have slight differences in their parameters and thus, were computed with another set of model orders. Furthermore, the BJ has the following parameters: nb, nc, nd, and nf which are orders of the B, C, D, and F polynomials, respectively and nk is the input delay, specified as the number of samples. Model one of the BJ sets would have a model order specified as nb = [1 2 1], nc = 2, nf = [4 2 1] and nk = [0 1 1].

The respective best fit plots for all the linear models after estimation and validation with one-step predicted response are shown in Figs. 6, 7, 8, and 9.

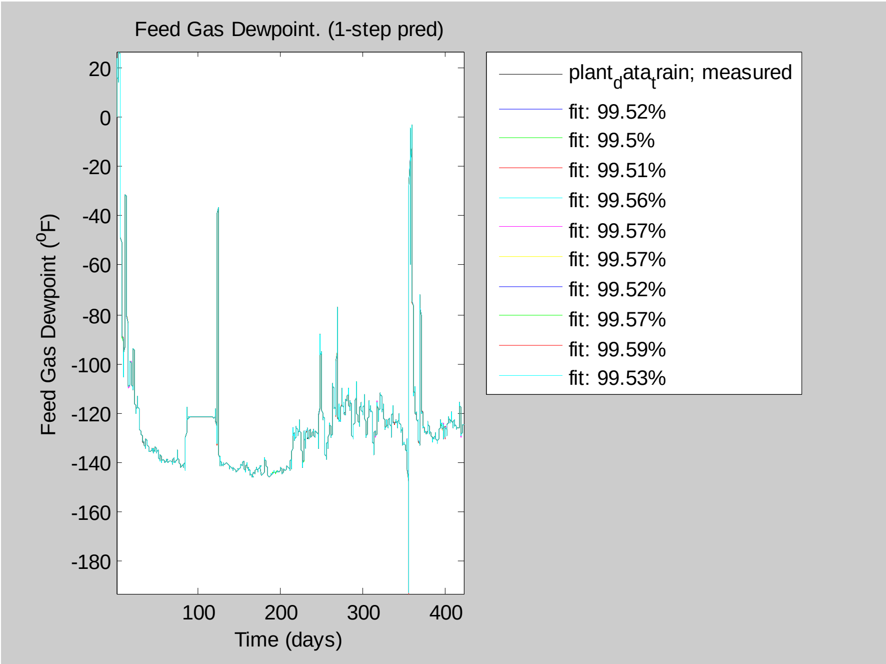

Figure 6: linear ARX models estimation

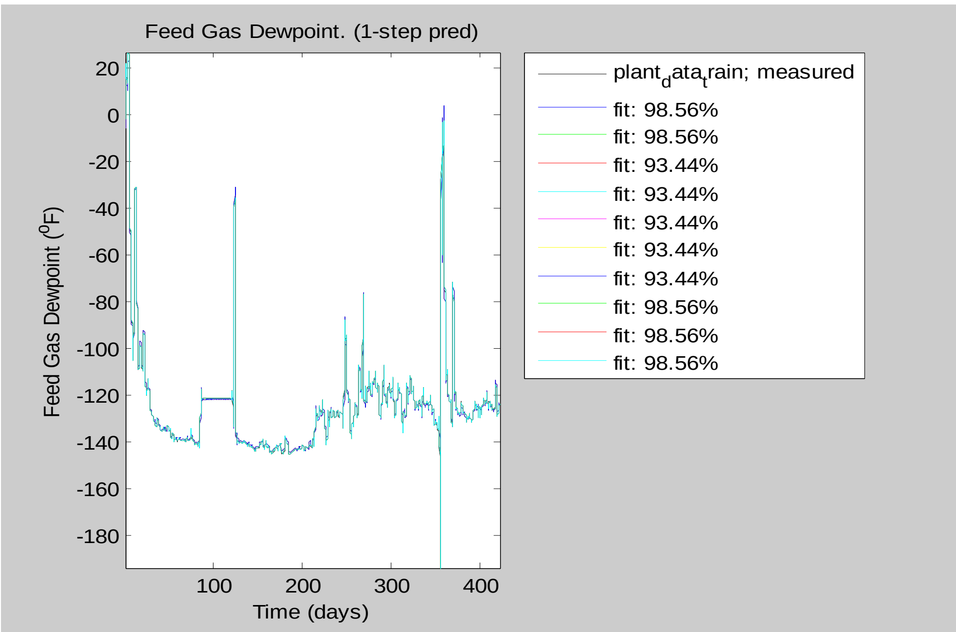

Figure 7: linear state-space models estimation

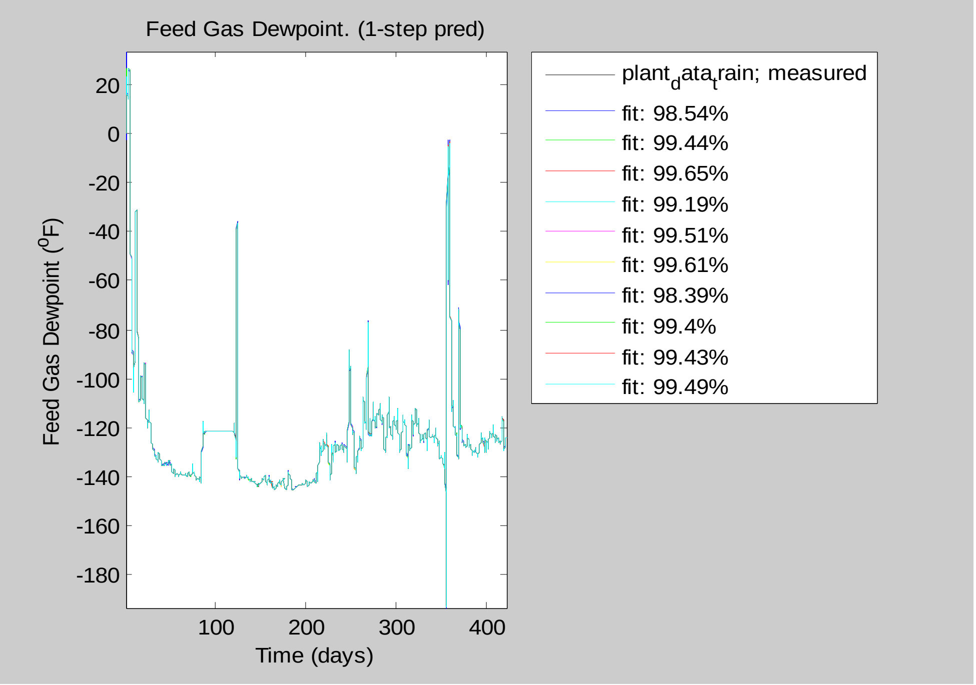

Figure 8: linear box-jenkins (bj) models estimation

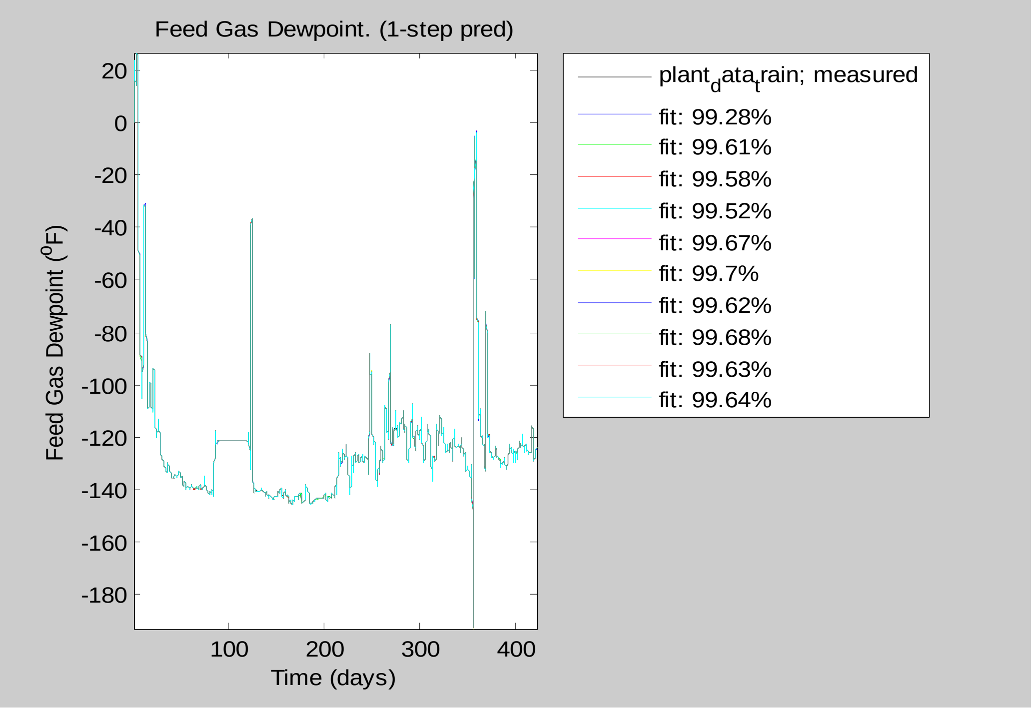

Figure 9: linear ARMAX models estimation

Simulation and cross validation are procedures for testing whether a model is capable of reproducing the measured output when driven by the actual input. This simply means computing the model response using input data and initial conditions. The difference in initial conditions of estimated model and the actual system can affect and cause disparity between the simulated responses and the measured responses. To minimize this variation, the initial state values from the data must be estimated and specified the initial states as input arguments to the simulation algorithm.

V. Conclusion

In this paper, data driven modelling was carried out with System Identification toolbox in MATLAB program. The dehydrator bed model was developed with the modelling structure of the Non-linear Auto-Regressive with exogenous input (NARX), as its suitability was suggested by the advice command. However, several dehydrator bed models were developed for other model structures such as linear Auto-Regressive with exogenous input (ARX) model, state-space model, Box-Jenkins (BJ) model and Autoregressive Moving Average with exogenous inputs model (ARMAX). The first step in the model development process was to select an optimum model order of the NARX model using loss function as the evaluation criteria for both estimation and validation data sets, and the best model order indicated the smallest loss function value and was found to be [3,4,2,3,2,1,9]. Based on this model order, ten other orders were selected by trial-error approach and used for the estimation of the Non-linear Auto-Regressive with exogenous input (NARX) model. Another criterion called best fit was used to visually assess the viability of the estimated dehydrator bed model. Equally, the validation of the trained model was accomplished using the set of program coding with the validation data set and their percentage of fit were plotted. Finally, NARX models was simulated by feeding the measured input data into the developed models to see how well they could reproduce the dew-point (measured output) of the dehydrator bed. Plots of the measured output against the simulated output were done for both estimation and validation.

References

[1]. E. J. Akpabio and V. J. Aimikhe, Dynamics of Solid Bed Dehydration in a Niger Delta Natural Gas Liquids Plant, *International Journal of Engineering and Technology*, 2(12), 2012, 2026-2032.

[2]. A. Elshorbagy, G. Corzo, S. Srinivasulu and D. P Solomatine, Experimental investigation of the predictive capabilities of data driven modelling techniques in hydrology - Part 2: Application, *Hydrology and Earth System Sciences*, 14(10), 2010, 1943-1961.

[3]. S. Guillaume and B. Charnomordic, Fuzzy inference systems: An integrated modelling environment for collaboration between expert knowledge and data using FisPro, *Expert Systems with Applications*, 39(10), 2015, 8744- 8755.

[4]. K. J. Åström and P. Eykhoff, System identification—A survey, *Automatica*, 7(2), 2009, 23-162.

[5]. S. Avila, A. Benito, C. Berro, S. Blanco, S. Otín, and I. Velasco, Dew-Point Curves of Natural Gas. *Measurement and Modelling,Industrial & Engineering Chemistry Research*, 45(14), 2002, 179-5184.

[6]. K. L Kar, C. G. Kar, V. Dohare, and B. Mangaraj, Optimization of the performance of ANN using VHDL. *International Journal of Scientific & Engineering Research*, 4(7), 2013, 2229-5518.

[7]. E. H. Johnson, A. J. Ntuk, and U. I. Akpan, Automated Electrically Operated Dehydration Bed Modelling: Automatic Generation of Expression Approach. *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, 2018, 6(1), 43-55.

[8]. G. J. Bowden, H. R. Maier, and G. C. Dandy, Optimal division of data for neural network models in water resources applications, 38(2): 2002. 1010-104

[9]. M. Y. Chen and D. Linkens, Rule-base self-generation and simplification for data-driven fuzzy models, *Fuzzy Sets and Systems*, 142(2), 2004, 243-265.

[10]. S. Simani, Application of a Data-Driven Fuzzy Control Design to a Wind Turbine Benchmark Model, *Advances in Fuzzy System*, 2012, 1-12

[11]. A. T. Azar, Adaptive neuro-fuzzy systems. *Fuzzy systems*, 2010, 85-110.

[12]. S. L. Barbour, J. Krahn, Numerical Modelling: Prediction or Process? *Geotechnical News,* 22 (4), 2010, 44-52.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Akaninyene J. Ntuk. " Nonlinear Auto-Regressive with eXogenous (NARX) input model for Liquid-Gas Dehydration and control Systems: Data Driven Modelling Approach." IOSR Journal of Computer Engineering (IOSR-JCE) 21.2 (2019): 56-64.