



Nested NARIMA Model of the Atmospheric Distillation Column

Michał Falkowski¹(✉) and Paweł D. Domański²

¹ Faculty of Automotive and Construction Machinery Engineering,
Warsaw University of Technology, Narbutta 84, 02-524 Warsaw, Poland
mfalko@stud.elka.pw.edu.pl

² Institute of Control and Computation Engineering,
Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
p.domanski@ia.pw.edu.pl

Abstract. Atmospheric distillation column plays a very important role in the crude oil processing. It is multivariate, strongly nonlinear process. Its operation and proper control impacts significantly the overall refinery performance. One may find in literature several papers describing column modeling with full spectrum of approaches starting from the first principle models up to regression *black-boxes*. Presented model composes of subsequent sections associated with technology. Each section is modeled as the nested NARIMA model. This methodology has been previously tested in several other chemical applications. It is extended by decomposition and coordination of subsequent sub-models in the considered case. Such a structure enables to obtain process nonlinear model with clear technological meaning of all considered elements. Further, it may be directly used in the process of control philosophy design. The procedure is illustrated with real data originating from the industrial installation.

Keywords: Atmospheric distillation column · NARIMA ·
Nonlinear process · Black-box

1 Introduction

Presented paper addresses two important issues: nonlinear methodology of nested NARIMA models and evaluation of atmospheric distillation column simulation model. In presented case the need for simulation model has appeared first. The modeling using NARIMA approach was adopted as it fits the best technology requirements and identification constraints. The subject of atmospheric column modeling exists in the research for several years. The first principle models are the most popular candidates [1–3], especially once unit simulation is an ultimate goal. They give an insight into the process and can be implemented in a simple way.

However, they are simplified and their calibration with industrial data might be difficult or even impossible due to the limited information and data. Secondly, it may be hard or not reasonable to use such a model in real time control

or optimization applications. Empirical models deliver an alternative. One may choose between different classes [4], like linear or nonlinear regression, Wiener-Hammerstein, neural networks, fuzzy, hybrid models, etc. Hammerstein [5] approach was selected, although not in a standard form. It has given the inspiration, further developed into the novel methodology.

The next design decision is associated with the structure identification [6]. Appropriate model variables (model inputs and outputs) have to be selected. Output selection is straightforward. They directly originate from the modeling goals. Other limiting variables are also often taken into consideration. Selection of input signals [7] is not simple and frequently depends on data availability and quality, process knowledge, statistical analysis, applied benchmarking indexes, etc. Concluding, model structure identification is rather the art.

Once the structure identification phase is achieved, we may progress towards estimation of model parameters. We focus on Hammerstein structure in our case. Thus, our model consists of two elements: nonlinear static part and linear dynamic one. There are many various algorithms [8]. Parameters of the nonlinear part are determined by any non-linear optimization, while the linear part may be determined by least squares approach. We may find many papers considering various approaches, like blind approach [9], iterative [10], multi-variable methods [11] and aspects such as an identification in closed loop [12].

In course of the work, various approaches were tested. Linear dynamic element could be satisfactory estimated with least squares. Nonlinear part was designed as polynomials using the least squares algorithm as well.

The paper is organized as follows. In the starting Sect. 2 the modeling methodology is presented. Then Sect. 3 presents the process of the crude oil atmospheric column. Those two background sections are followed by presentation of the modeling results (Sect. 4). The paper concludes in Sect. 5 with the discussion of obtained results and presentation of open issue for future research.

2 Modeling Methodology

Modeling methodology for the complex installation consists of four main phases:

1. process decomposition into separable dynamic subprocesses that may be identified independently, ideally, with one common approach,
2. structure identification for selected subprocesses, i.e. proper selection of model inputs and outputs, its type, size, order, etc.
3. estimation of the model parameters and its validation,
4. consolidation and coordination of all sub-models into single model.

2.1 Process Decomposition

Process decomposition derives from the considered modeling problem. The main goal is to select such sub-processes that are clearly separable, being independent on the rest of the process. They are often associated with the process technological design. The main two approaches for the selection are serial or parallel

structures, however hybrid configurations are frequently used as well [4]. In presented work selection is done according to the P&ID (Process & Instrumentation Diagram).

2.2 Model Structure Identification

Model structure identification is quite a tedious task, not well defined and very often customized to the problem. It frequently consists of several steps, while each of them gives fragmentary contribution to the overall methodology [13,14]. In the considered case the process is nonlinear and identification is based on the historical data. The results significantly depend on the data thus variables time series review and proper selection of data files constitutes an important step (time trends, histograms and statistical analysis). Methodology includes combination of various methods, like autocorrelation and cross-correlation backed with the basic modeling of different structures and their performance comparison.

2.3 Model Parameter Estimation

Estimation of the model parameters is strictly connected with the selection of the process model type. In the considered case the Nested ARIMA (NARIMA) model structure was selected (see Fig. 1) with modeled output in form of (1).

$$y_m(k+1) = y_m(k) + \Delta y_m(k+1) \quad (1)$$

It belongs to the class of nonlinear dynamical models called block-oriented ones. It consists of a nonlinear memoryless static element followed by linear dynamics and as such is very similar to the Hammerstein models. While the nonlinear part is interpreted in a similar way to Hammerstein models (with output denoted as $y_m(k)$), the approach to dynamics modeling differs. Dynamic part identifies model error of the nonlinear static element $\Delta y_m(k+1)$. Static nonlinear function and linear dynamic part can be generally of different origin and structure [8]. In our case we use static MISO polynomials and ARIMA regression models.

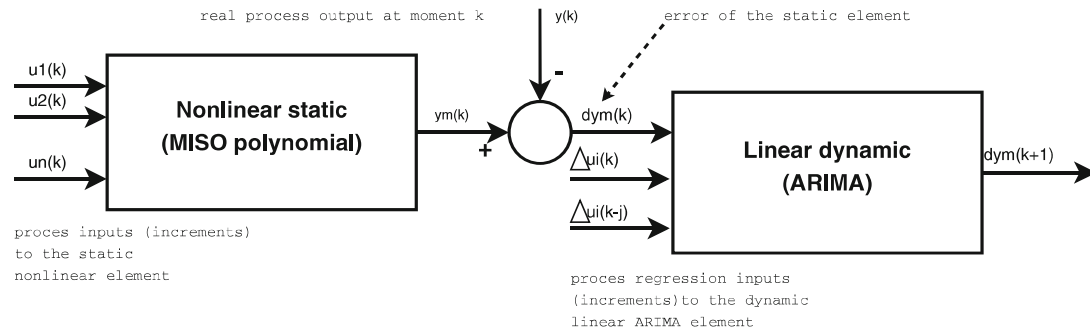


Fig. 1. Nested ARIMA (NARIMA) model scheme

Model selection determines estimation algorithms. Identification of the static MISO polynomials is done using standard least squares. ARIMA regression models are also derived with the least squares method implemented within Matlab System Identification Toolbox. In all cases data are divided into two data sets, one used for identification and the other one for validation purposes.

Identification of the structures is done separately for each of the sub-models. At first the size of the static polynomial is selected. In most cases it is first or the second order. Orders of the dynamic regression models are chosen in an experimental way, selecting those number which result in the best model fitting.

It has to be noted that all the models are multivariate. Selection of the appropriate input signals is done on the basis of the process technology consulted with site experts accompanied by correlation analysis.

3 Atmospheric Column for Crude Oil

The crude oil Atmospheric Distillation unit (column), is one of the most important plants in a refinery. It is crucial to have a proper simulation tools to study thoroughly such a multicomponent distillation process, because of its complexity and the fact that its products are the feeds to other site production units. The crude oil is distilled to produce distillate streams, which are the basic streams for the refinery product slate. These streams are either subject to further treating downstream or become feed stock for other conversions in the refinery.

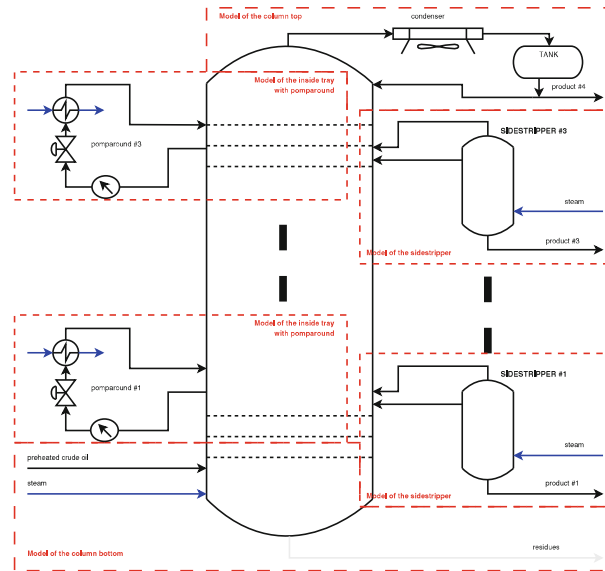


Fig. 2. Crude oil atmospheric column with the decomposition scheme to subprocesses

A schematic flow diagram of an atmospheric crude column is sketched in Fig. 2. The process is considered in literature for many years and thus its discussion may be found in several reports [15–17]. The process is complex, however it

consists of several similar elements. For the sake of decomposition the following sub-process are defined and considered, like [18]:

- model of the column bottom,
- model of the inside tray with a pumparound,
- model of the sidestripper,
- model of the column top.

Although models of the top and bottom of the column appear only once there might be several models of the inside tray and sidestrippers. In our case there are three inside tray models and three sidestrippers. Modeled variables are interesting product flows and temperature of products and at each tray level.

The structure of the selected sub-models, i.e. inputs and outputs selection is the same for each model instance. Model of the column bottom has four inputs: inflow of the superheat steam, inflow and temperature of the crude oil and outflow of the atmospheric residues. Temperature at the top of the column section forms the output of the sub-model.

Inside tray with pumparound model has tray top temperature as the output. The inputs are as follows: tray bottom temperature (it is the tray top temperature of the lower section), pumparound flow and the tray product outflow.

In case of sidestrippers models each one has temperature on the output. There are four inputs: tray bottom product flow and temperature (within the same section), superheated steam and product flow which is further processed. All the data used originates from the full-scale operational anonymous industrial installation.

4 Modeling Process

The modeling described in the paper is performed according to the industry requirements constrained by technology and limits of Instrumentation & Control infrastructure. There are two main issues that should be addressed. Historical data access, collection and preprocessing embedded in plant data historian is the first one. Aspects of identification in the mixed mode with some process loops operating in MANUAL and unfortunately the most of them in AUTO (resulting in close loop identification).

4.1 Data Collection and Preprocessing

Working with the real object is associated with a number of complications and problems. Considerations of economy and security of the plant, which is modeled object does not allow for testing of identification. The only solution appears to work with historical data collected during normal process operation.

Target plant is working under control of the well-tuned system. Operation is additionally frequently altered by human interventions (MANUAL model operation). Process frequently is non-stationary, due to external weather conditions, varying parameters of substrates and volumes of consumption.

It is very important to notice that consultation process with plant personnel is crucial and may significantly help. However, before it will be possible to work with data, they must be properly selected, grouped, filtered. Data may be subject to various errors and modifications, like for instance due to the properties of data compression in SCADA systems, transmission failures, etc. The best practice shows that human data inspection is very important. Autonomous identification often leads toward unreliable results.

4.2 Close Loop Identification

Identification of the dynamics, which is inside of the close loop is a challenge. In industrial reality it is often impossible to open the loop, apply external identification input (step, PRBS, ...) and collect data. Apart from the cost of altering normal production and adding of external disturbance, i.e. identification signals, everybody have concerns about installation safety and associated risks. However, in closed loop case standard identification policies, like for instance least squares fail due to the convergence criteria. Thus formally there are two options:

- Indirect Process Identification: An overall model of the closed-loop is identified. The controller has to be known and than process model is evaluated.
- Direct Process Identification: Process is identified as it is.

In reality it is difficult to practically use knowledge about real controller. It is complex structure several other blocks, like nonlinear functions $f(x)$, feed-forwards, override controls, interlocking, normalizations, interactions with other loops and human interventions (biases, setpoint changes, etc.). Thus we are left to the Direct Process Identification option despite all deficiencies.

4.3 Identification Results

The results of the modeling process of the atmospheric column are presented for the complete column model in the form of trends of the static and NARIMA models. In addition, a residual analysis was performed to determine the quality of these models. The analysis procedure follows the modern approach presented for the chemical engineering plants [19].

Figure 3 presents the results of identification of the static and NARIMA model of the temperature of the upper part of column. Sampling is equal to $T_s = 1$ [min].

Analysis of trends of received global models of atmospheric column for crude oil shows that static model despite following real signal trend achieves worse results than NARIMA model - Fig. 3. Both identified models do not fully accurately reflect real signal however the NARIMA model in all cases follows a real signal better and keeps dynamics of atmospheric column for crude oil. For more accurate comparison series of coefficients determining the quality of the obtained global models of crude distillation unit were placed in Table 1.

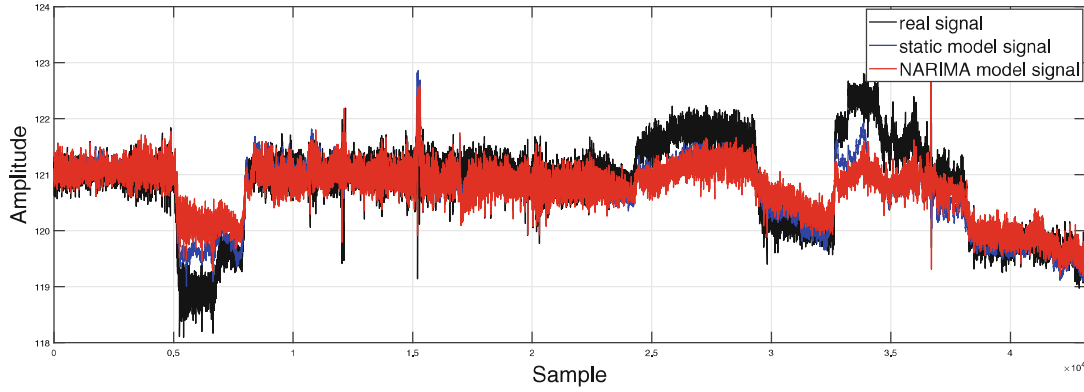


Fig. 3. Comparison of global static and NARIMA atmospheric column for crude oil models

When comparing coefficients listed in Table 1 for both global static and NARIMA models error, it can be noticed that static model has better quality compared to dynamic one. Value of the integral absolute error is in both cases comparable, however the value of mean square error indicates in favor of a static model. Analysing trends and coefficients, it can be concluded that, after all global NARIMA model of atmospheric column for crude oil has a higher quality compared to a global static model.

Table 1. Coefficients comparison of the atmospheric column for crude oil models error

Coefficient	Value	
	Static model	NARIMA model
Mean	−0.261	−0.177
Logistic location	−0.257	−0.172
Bias	0.009	0.098
Standard deviation	0.751	0.861
Scale (Cauchy)	0.407	0.356
Scale (Laplace)	0.568	0.599
Scale (Logistic)	0.663	0.587
IAE	0.601	0.613
MSE	0.633	0.722

Further Figs. 4 and 5 show global error histogram of a NARIMA model. First obtained histogram is characterized by two occurring maxima. This situation indicates that the measurement data set should be divided and some data rejected. Re-analysis of input and output signals trends allowed to reject a certain range of data and to re-determine the error histogram of the NARIMA model. Results

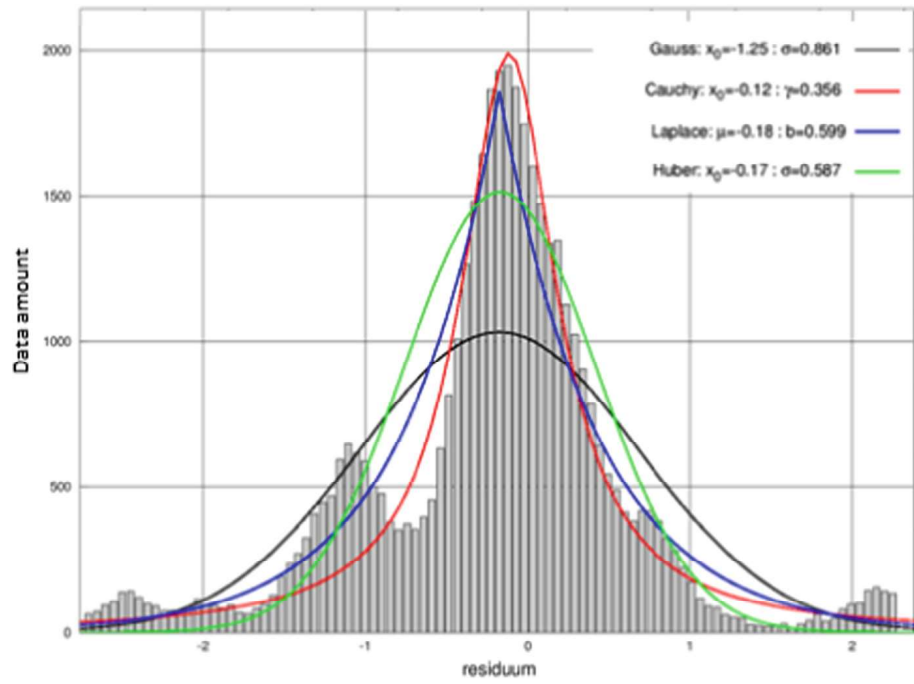


Fig. 4. Error histogram of NARIMA model of the atmospheric column for crude oil with distributions

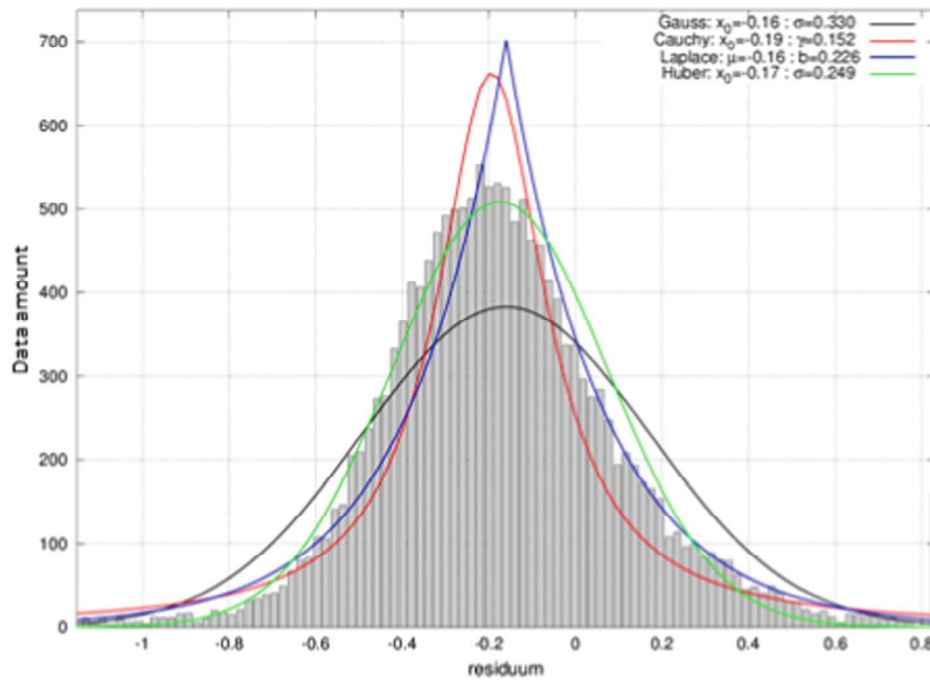


Fig. 5. Error histogram of NARIMA model of the atmospheric column for crude oil with distributions after data rejection

of such an operation are shown in Fig. 5. It is clearly visible that the undesirable maxima from Fig. 4 disappeared and Gauss, Cauchy, Laplace and Huber distributions are better matched [20].

5 Conclusions and Further Research

The paper includes results of the dynamic modeling of the crude oil atmospheric distillation unit based on the data originating from the real industrial site.

First step was to analyze received real data and actual object diagrams, which results in division the atmospheric column for crude oil into numerous local models. For the sake of decomposition the process are defined and considered like model column bottom, insidetray with a pumparound, sitestripper and column top.

Second step was to extract the correct input and output signals. Inputs like steam flow were considered as energy of this steam, referring to such an approach used in literature. Many methods have been used to verify the correctness of chosen input and output signals, e.g. by determining their correlation.

An important step is also selection of delays in increments of input signals. This element is used in implementation of NARIMA model dynamic block by identifying error model obtained from static block.

During identification process of the atmospheric column for crude oil model a number of problems were found. Operator's influence or various unknown external factors had negative impact on the quality of obtained static and NARIMA model. Situation presented in histogram (Fig. 4) indicates that part of the entire measurement data set should be rejected. This is also evidenced by the values of coefficients presented in Table 1.

In conclusion, it can be concluded that final result is satisfactory after a detailed analysis of the trends of both global static and NARIMA models and residual analysis. Quality of the developed models can be described as good with a slight advantage in favor of a NARIMA model.

References

1. Gani, R., Ruiz, C., Cameron, I.: A generalized model for distillation columns - I. *Comput. Chem. Eng.* **10**(3), 181–198 (1986)
2. Cameron, I.T., Ruiz, C.A., Gani, R.: A generalized model for distillation columns - II. *Comput. Chem. Eng.* **10**(3), 199–211 (1986)
3. Chaudhuri, U.R.: *Fundamentals of Petroleum and Petrochemical Engineering*. Chemical Industries. CRC Press, Boca Raton (2010)
4. Nelles, O.: *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, Heidelberg (2001)
5. Hammerstein, A.: Nichtlineare Integralgleichungen nebst Anwendungen. *Acta Math.* **54**(1), 117–176 (1930)
6. Nelles, O.: *Process Dynamics and Control: Modeling for Control and Prediction*. Wiley, Hoboken (2007)

7. Sindelar, R., Babuska, R.: Input selection for nonlinear regression models. *IEEE Trans. Fuzzy Syst.* **12**(5), 688–696 (2004)
8. Isermann, R., Münchhof, M.: Identification of Dynamic Systems: An Introduction with Applications. Springer, Heidelberg (2011)
9. Bai, E.W.: A blind approach to Hammerstein model identification. *IEEE Trans. Signal Process.* **50**(7), 1610–1619 (2002)
10. Vörös, J.: An iterative method for Hammerstein-Wiener systems parameter identification. *J. Electr. Eng.* **55**(11), 328–331 (2004)
11. Wills, A., Schön, T.B., Ljung, L., Ninness, B.: Identification of Hammerstein-Wiener models. *Automatica* **49**(1), 70–81 (2013)
12. Han, Y., de Callafon, R.: Closed-loop identification of Hammerstein systems using iterative instrumental variables. In: *Proceedings of the 18th IFAC World Congress*, Miedzyzdroje, Poland, vol. 18, pp. 13930–13935 (2011)
13. Haber, R., Keviczky, L.: Nonlinear System Identification - Input-Output Modeling Approach. Volume 2: Nonlinear System Structure Identification. Kluwer Academic Publishers, Dordrecht, Boston, London (1999)
14. Sun, N.Z., Sun, A.: Model Calibration and Parameter Estimation: For Environmental and Water Resource Systems. Springer, New York (2015)
15. Watkins, R.N.: Petroleum Refinery Distillation, 2nd edn. Gulf Publishing Company, Houston (1979)
16. Hsie, W.H.L., McAvoy, T.J.: Modeling, simulation and control of crude towers. *Chem Eng Comm.* **98**, 1–29 (1990)
17. Green, D.W., Perry, R.H.: Perry's Chemical Engineers' Handbook, 8th edn. McGraw-Hill Professional, New York (2008)
18. Radulescu, G., Paraschiv, N., Marinoiu, V.: A model for the dynamic simulation of a crude oil unit. *Control. Eng. Appl. Inform.* **2**(1), 43–50 (2000)
19. Domański, P.D., Golonka, S., Jankowski, R., Kalbarczyk, P., Moszowski, B.: Control rehabilitation impact on production efficiency of ammonia synthesis installation. *Ind. Eng. Chem. Res.* **55**, 10366–10376 (2016). <https://doi.org/10.1021/acs.iecr.6b02907>
20. Domański, P.D., Ławryńczuk, M.: Assessment of predictive control performance using fractal measures. *Nonlinear Dyn.* **89**, 773–790 (2017)