MINISTRY OF EDUCATION AND TRAINING
HA NOI UNIVERSITY OF MINING AND GEOLOGY

TRAN NGOC TRUNG

RESEARCH ON IMPROVING THE EFFICIENCY OF NATURAL GAS PROCESSING AT HAI THACH FIELD

Major: Petroleum engineering
Code: 9520604

SUMMARY OF TECHNICAL DOCTORAL THESIS

Ha Noi, 2023

The dissertation is completed at: **Department of Drilling and Production, Faculty of Oil and Gas, Hanoi University of Mining and Geology**.

Scientific Supervisors:

1. Assoc. Prof., Dr. Trieu Hung Truong

2. Dr. Ngo Huu Hai

Reviewer 1: **…………………………**

Reviewer 2 **: …………………………**

Reviewer 3: **…………………………**

The dissertation will be defended before the Academic Review Board at the University level at Hanoi University of Mining and Geology at … : ..... of date ..... month ..... year ....

The thesis is available at the: **National Library of Vietnam or the Library of Hanoi University of Mining and Geology.**

INTRODUCTION

1. The necessity of the study

Natural gas offers substantial environmental advantages compared to other fossil fuels and is a remarkably safe energy source during transportation, storage, and usage. Given the actual production requirements, research aimed at enhancing the efficiency of natural gas processing at condensate gas fields, such as those operated by the Bien Dong Petroleum Operating Company (POC), is of utmost importance and urgency. At the Hai Thach - Moc Tinh (HT-MT) field, maintaining uninterrupted gas processing from the wells to the gas compressor outlet is always the most crucial challenge. Any interruption in the subsystem's operation will have a domino effect on the processing sequence, leading to gas flaring and potentially posing a risk to the platform's safe operation. As a result, it is imperative to utilize an intelligent data analysis tool that can monitor equipment anomalies and predict the input parameters of the gas processing system, allowing for quick, precise, and ongoing adjustments to the system. Consequently, the use and development of artificial intelligence in natural gas exploitation and processing at the Hai Thach field is a cutting-edge trend in contemporary oil and gas production management. This approach helps to minimize unplanned downtime, enhance production and processing efficiency, ensure safe and continuous operation, and yield economic benefits for the natural gas processing systems at the HT field.

2. The purpose of this thesis

- Enhance the productivity and processing efficiency of natural gas at the Hai Thach field by incorporating artificial intelligence to assist in monitoring, analyzing, and predicting processes.
- Establish an innovative and suitable approach to predict the production of gas condensate at the HT field, guaranteeing the recovery and production of each well; provide data sources and stabilize input parameters

for the processing system.

- Create a methodology for detecting anomalies in production wells and gas compressors to ensure pressure and flow stabilization for the gas processing systems.

3. Target and area of this research

- Target of this research: Natural gas production and processing at the HT field.
- Area of this study: To research and develop artificial intelligence techniques to enhance the effectiveness of natural gas production and processing in the HT field, focusing on supporting the monitoring, analysis, and forecasting processes.

4. Research subjects

This research will encompass a review of the literature regarding the research area and the current state of natural gas processing technology in Vietnam, focusing on the Hai Thach-Moc Tinh field. Additionally, an assessment and analysis of the advancements in advanced analytics in the oil and gas industry will be conducted, both globally and in Vietnam.

Research and development of a gas-condensate prediction method using an Improved Adaptive Neuro-Fuzzy Inference System (Improved ANFIS) to improve well recovery and production, providing a reliable data source and stabilizing conditions for the processing system.

Research and development of anomaly detection method for water production and natural gas compressors using an Improved Long Short Term Memory based Autoencoder Network (Improved LSTM-AE), which stabilizes the pressure and ensures smooth flow of flash gas during the gas treatment process.

5. Approaches and Methodology

5.1. Approaches

Overview of the worldwide and Vietnamese advancements in the oil and gas industry concerning scientific accomplishments, cutting-edge technology, and machine learning algorithms. Subsequently, proposing ways to enhance the efficiency of gas production and processing in the HT field through further research and analysis.

5.2. Methodology

- Investigate the theoretical aspects of machine learning methodology and data processing techniques through research.
- Use a quantitative approach: gathering, examining, and analyzing the historical production data at the HT field.
- Conduct a model study: developing, testing, and evaluating the machine learning models.
- Involvement of the domain expert: to guide the research approach and ensure the research outcomes' validity and robustness.

6. The scientific and practical meaning of this thesis

6.1. Scientific meaning:

Applying the Machine Learning method can fully realize the potential of the historical and real-time gas production and processing database at the HT field. This is a topic of significant interest among researchers for its potential applications across various aspects of life and society.

The data processing model in this study has the capability to evaluate, fine-tune and guarantee high reliability, enabling operators to make informed and prompt decisions to mitigate production risks. As a result, this innovative and scientifically robust oil and gas management system guarantees safety and continuous operation while enhancing the efficiency of natural gas production in the HT field.

6.2. Practical meaning

Integrating cutting-edge technology with optimizing existing infrastructure and utilizing local resources is crucial in lowering costs and enhancing production and processing efficiency in oil and gas fields. The thesis demonstrates that the use of machine learning in predicting well production rate and detecting anomalies contributes to ensuring the safety, efficiency, and stability of the natural gas processing system, reducing the unplanned downtime or well closures caused by incidents in an offshore production platform. The practical application of the research findings contributes to the advancement and optimization of production and processing efficiency at gas condensate fields, reducing the self-investment time and cost for oil and gas operating companies and thereby enhancing the investment efficiency of oil and gas exploration and production projects in Vietnam.

7. The new highlights of this thesis

7.1. The thesis presents a novel, suitable, and effective solution to the practical issues encountered during production at the Hai Thach field through the application of Machine learning. This approach leverages existing infrastructure and historical data at the HT gas-condensate field.

7.2. Implementing the Improved AFNIS network has demonstrated the ability to prevent the overfitting phenomenon and improve prediction accuracy. When applied in the HT-MT field, the forecasting of the virtual flow meter supports the well-testing and performance monitoring workflow, increasing the predictability of the allocation plan.

7.3. An Improved LSTM-AE network for detecting anomalies in multivariate time-series features was trained to minimize the Mean Absolute Error (MAE) loss function using the Random Search optimization technique on a set of hyperparameters. The optimal threshold value was then determined to yield the highest F-score among the optimized models.

8. Primary arguments of this thesis

Argument 1:

The effective deployment of the Improved ANFIS network for complex multiple non-linear time-series datasets improves the prediction accuracy of gas-condensate production and stabilizes the gas processing by achieving a Root Mean Square Error (RMSE) of 0.0645 and a coefficient of determination (R2) of 0.9482.

Argument 2:

The Improved LSTM-AE network was successfully applied to categorize anomalies in production wells and natural gas compressors, achieving the best F-score of 0.57143. This outcome reduces risks and ensures the stability, safety, and efficiency improvement of natural gas processing in the HT field.

9. Reference material for this thesis

The formation of the thesis is based on the incorporation of various types of documents and data:

- Thesis-related material includes references, analysis, statistics, and data from books, articles, and research findings.
- The author's research results are presented in articles published in specialized scientific journals in Vietnam and abroad.
- Research data and results of state-level projects where the author is the key project member. Research project ID ĐT.CNKK.QG.0077/21 "Research and development of Artificial Intelligence tools to support the combined analysis of geological, petrophysical, and production data to improve the management and operation efficiency of Hai Thach - Moc Tinh gas-condensate fields, Blocks 05-2 and 05-3, East Sea, Vietnam" under the "National key science and technology program for innovation and modernization of mining and processing technology to 2025".

10. The scope and structure of this thesis

The thesis consists of an introduction, three chapters, a conclusion, a suggestion section, and an index of references. The entire content of this thesis is illustrated in 202 pages of A4 size, including 9 tables and 59 figures.

CHAPTER 1:
A GLOBAL REVIEW OF RESEARCH ON NATURAL GAS PROCESSING, THE SITUATION IN VIETNAM, AND STRATEGIES TO ENHANCE EFFICIENCY

1.1. Overview of natural gas processing on an offshore production platform

The initial phase of the natural gas value chain, offshore production, involves extracting the gas from the reservoir and bringing it to the customers. The processing of natural gas encompasses the separation of gas streams and compression to meet pipeline specifications, the recovery of flash gas, stabilization and transmission of condensate, treatment of produced water, and removal of impurities. Some stages are carried out near or directly at the wellhead and the processing facilities, while the entire deep processing of natural gas is performed at a processing/refining facility, typically located onshore.

In recent years, significant scientific research has been directed toward addressing key issues related to the processing, treatment, and transmission of natural gas to convert it into hydrocarbon products that are easily transportable and marketable. Most research endeavors aim to enhance the efficiency of processes that result in the most significant economic benefits for each natural gas facility while also considering technical, safety, and environmental considerations.

1.2. The position and role of HT-MT gas condensate field

The Bien Dong 1 Project encompasses the offshore development of the Nam Con Son Basin’s gas and condensate fields, which are situated approximately 300 km from the coast of Vietnam. This is a nationally significant project with challenging and unconventional offshore conditions, including deep waters (ranging from 118 to 145 meters) and an area of abnormally high pressure (890 atm) and temperature (exceeding 190 degrees Celsius). The daily production of the Hai Thach - Moc Tinh gas condensate field is estimated to be around 6-7 million standard cubic meters of gas and 8000-9000 barrels of condensate. The achievement of the HT-MT gas condensate project aligns with Vietnam's oil and gas industry development strategy and holds significant value in terms of strategy, economy, society, politics, and national security.

1.3. The production and processing systems in HT-MT

The fluid from the HT and MT gas/condensate wells will be directed to the central production facility located in Hai Thach, referred to as the Production and Quater Platform (PQP-HT). The purpose of this central production facility is to separate and compress the gas stream to meet the specifications of the NCSP gas pipeline, stabilize and export the condensate product to a Floating Storage and Offloading (FSO) vessel in Hai Thach, and remove oil contamination from the produced water to comply with Vietnamese regulations on the discharge of process wastewater into surface water.

1.3.1. Production rate forecasting to optimize operation mode and ensure flow assurance on the in-field pipeline system

Increasing the accuracy of production flow rate forecasting is of utmost importance as it enables real-time monitoring of well performance, enhancing the predictability of allocation plans without the need for a comprehensive well-testing program. Process optimization is continuously solving fluid dynamic and thermodynamic equations and constantly finding and changing equipment and processes' appropriate dynamic behavior and performance. This is in response to changes in operation and fluid conditions, leading to cost savings in both Capital Expenditure (CAPEX) and Operating Expenditure (OPEX) while maintaining safety and control in operations.

1.3.2. Anomaly detection system for wells and equipment to enhance gas processing efficiency at the HT field

The gas production and processing system at the HT field has a proven track record of safe and efficient operation. However, it currently faces numerous difficulties and challenges due to declining flow and pressure. Maintaining a continuous and stable operation of wells and processing equipment is a crucial concern for the HT-MT field, as it is vital to achieving production goals such as exporting gas and delivering condensate to the FSO. To address these challenges, it is imperative to integrate and utilize data intelligently, enabling advanced analytics to detect anomalies in the wells and equipment of the gas processing system, aiming to optimize and enhance processing efficiency.

1.4. Methods to improve the efficiency of natural gas processing

1.4.1. Real-time Optimization

The main functions of a Real-time Optimization (RTO) system include (1) System monitoring and optimization condition determination, utilizing both steady-state and dynamic models to monitor operating performance; (2) Reconciliation and parameter estimation, which involves gathering operating conditions and reconciling the plant-wide model to determine the current state of the plant; (3) Optimization, where the current operating limits (constraints) are taken into account and an optimization problem is solved to find the most profitable operating conditions; (4) Setpoint update, which involves implementing the optimization results by downloading the optimized set points for use by the control system.

1.4.2. Steady-state and dynamic modeling

Integrating and constructing system simulation models using data and intelligent algorithms is crucial for enhancing production and processing efficiency and predicting future asset performance. By combining the principles of thermodynamics, hydraulic simulation, and the operation of each device module, the model can accurately predict complex facility simulations in both steady-state and dynamic scenarios.

1.4.3. Real-time Advanced Process Control

Advanced Process Control models in real-time can be constructed through highly precise dynamic simulation systems and a sequence of calibration procedures designed to align with actual historical data. Typically, these control models undergo extensive analysis to determine complete and specific control conditions and settings to guarantee the quality of the commercial products produced by the gas processing plant.

1.4.4. Intelligent data analysis techniques

The operation of an oil and gas field presents challenges in maintaining stable and continuous operations due to unexpected changes in production wells or flow patterns in multiphase pipelines. Oil and gas operators must monitor, control, and mitigate these fluctuations safely to minimize their impact on processing systems. Many gas processing projects have adopted intelligent data analysis technology and Industry 4.0 solutions to achieve this.

1.4.5. Elaborate on the current methods being utilized and the future trends

RTO or Real-time APC models require a fundamental understanding of physical, chemical equations, and mechanical relationships, as well as the definition of relevant parameters in each model. The systems are equipped with an equation solver/optimizer and an interface that continuously monitors and assesses plant conditions and scenarios, determining and calculating the optimal operating parameters. However, due to their complexity, including the use of numerous uncertain fluid variables, inter-module interactions, and high hardware demands, as well as the need for expensive software and specialized engineering services for installation, commissioning, and maintenance, these methods can be challenging to implement across the entire gas processing plant.

For the natural gas processing system at the HT-MT field to operate efficiently, it is essential to maintain the stable and continuous operation of equipment such as natural gas compressors. Improving the accuracy of forecasting production well flow rates is crucial for monitoring well performance, optimizing well allocation, and planning production effectively. Hence, the need for a tool that can analyze data, detect and predict anomalies in production wells and equipment on gas processing systems and provide prompt and appropriate responses is a pressing practical issue that needs to be addressed in the HT field.

1.5. Conclusion for Chapter 1

Implementing advanced technology while preserving the existing infrastructure is crucial in lowering costs and enhancing the effectiveness of processing systems in gas condensate fields. Intelligent data analytics systems can analyze and extract patterns from data, select relevant information, and uncover underlying relationships among data layers during operation. This enables accurate and reliable monitoring, diagnosis, analysis, and forecasting. The utilization of intelligent data analysis tools for production data, detecting and predicting anomalies in wells and equipment within the gas processing system at the HT-MT field, and providing prompt and effective responses, is a pressing and practical issue that requires resolution to guarantee the efficiency of the gas processing system. Thus, the focus of this thesis will be to explore and implement methods of enhancing the efficiency of natural gas processing through intelligent data analysis techniques that utilize the existing infrastructure and historical data available at the HT gas condensate field.

CHAPTER 2:
DEVELOPING INTELLIGENT DATA ANALYSIS MODELS TO IMPROVE THE EFFICIENCY OF NATURAL GAS PROCESSING IN THE HAI THACH FIELD

2.1. Research to improve the accuracy of flow rate prediction method, ensuring recovery factor and productivity at HT field

2.1.1. Introduction of the method being applied to predict the flow of gas condensate at the HT field

The design process for simulating a condensate gas processing system employs software that models steady-state behavior, which requires inputting process conditions such as pressure, temperature, flow, composition, and physical properties of heavy hydrocarbons for each stream.

The well-testing analysis represents a critical component of reservoir engineering management and analysis by providing the estimates of reserves, deliverability, permeabilities, supporting allocation calculation, understanding of the status of wellbore conditions, and inflow profile along the well. However, implementing a well-testing and performance monitoring program can be costly for oil and gas companies due to equipment and personnel expenses. Moreover, obtaining all necessary data for a reservoir may not be feasible due to production demands or changes in surface process conditions. Conducting tests and monitoring gas well performance can be time-consuming and costly to meet gas extraction requirements or avoid changes in surface processing system conditions. Thus, effectively balancing production and allocation for each well to optimize recovery factors and operations heavily relies on the accuracy of production forecasting.

2.1.2. Intelligent data analysis tool in predicting production flow rate

The use of Machine Learning (ML) for time series data prediction in the oil and gas industry has gained popularity and achieved numerous advancements. These studies have demonstrated high accuracy and illustrated the connection between academic research and practical implementation.

2.1.3. Research and development of a gas-condensate prediction method using an Improved ANFIS

ANFIS model processes the calculating and formulating of giving non-linear inputs to the set of ‘crisp’ values, represented in the form of membership functions and fuzzy rules. Then performs a de-fuzzy procedure to generate ‘crisp’ output out of fuzzy rules for reasoning. The input dataset is separated into groups of input-output data with clustering techniques, then tunable membership functions and rules are generated into the desired ANFIS architecture. Besides, ANFIS learns by tuning all its membership functions and consequent parameters to formulate the inputs to the desired output with minimum error.

2.1.4. Model development

The author aims to construct an Improved ANFIS model for predicting the gas condensate production flow rate utilizing historical and real-time data. The model incorporates two main partitioning algorithms, Subtractive clustering and Fuzzy C-mean (FCM). To verify the accuracy of the ANFIS model, the author also plans to conduct a linear regression analysis between the actual and predicted values to identify any instances of overfitting.

2.1.4.1. Data preprocessing

Using raw data directly in machine learning models can lead to unsatisfactory results as the algorithms cannot automatically extract relevant features. To address this, data preprocessing is commonly used to bring the data to a normal or Gaussian distribution.

2.1.4.2. Data smoothing techniques

The four data smoothing techniques selected for the FCM and Subtractive models in the research are: (1) a moving average with a window length of 5, (2) a 1-D median filter with an order of 3, (3) a local regression ‘*loess*’ with a span of 4, and (4) a robust local regression ‘*rloess*’ with a span of 4. These techniques aim to reduce the computation time while ensuring the required accuracy of the time series prediction model.

2.1.4.3. Partition styles for Fuzzy models

The ANFIS model relies heavily on partitioning techniques. By decreasing the radius for Subtractive clustering or increasing the number of clusters in the ANFIS FCM model, the natural groupings of a large data set can be better identified, the variation within each cluster can be reduced, and a more accurate representation of the system's behavior can be formed.

2.1.4.4. K-fold cross-validation

The performance assessment was done using the k-fold cross-validation method, which has been deemed superior to other methods such as holdout, bootstrap, and leave-one-out cross-validation in determining the generalization error in the selection of models.

2.1.4.5. Evaluation metrics

Five of the most common evaluation metrics are calculated to assess the accuracy and efficiency of the prediction model.

2.1.4.6. Evaluate the effectiveness of k-fold cross-validation

The linear regression technique was utilized to observe the overfitting phenomenon when using the Subtractive clustering ANFIS model with shuffling and without the k-fold cross-validation method. The best results obtained from the Subtractive and FCM clustering were compared with those from improved Particle Swarm Optimization-ANFIS and Genetic Algorithm-ANFIS models from prior research by Elbaz et al. [1, 2].

2.2. Research on anomaly detection model that supports stabilize pressure and ensures smooth flash gas flow for gas processing at HT field

2.2.1. Introduction of the method being applied to categorize and detect anomalies in the HT field

The gas processing system at the PQP-HT platform is contingent upon the stability of the well fluid input and the continuous operation of the equipment. Various methods for categorizing and detecting anomalies are utilized at the HT field for each occurrence to ensure smooth functioning.

2.2.2. Intelligent data analysis method to classify anomalies

To guarantee effectiveness at the HT-MT field, it's crucial to maintain stable and continuous operation of the well and processing equipment. Thus, the author researched to develop an anomaly detection method that leverages historical and real-time data sets, utilizing a deep learning tool to provide more precise and rapid analytical classification, supporting engineers in monitoring, analyzing, forecasting, and adjusting the gas processing system.

2.2.3. Research and development of anomaly detection method using an Improved Long Short Term Memory based Autoencoder Network

2.2.3.1. Long Short Term Memory networks (LSTM)

The LSTM recurrent manner makes it an appropriate method for sequence data and incredibly complex multivariate time series. LSTM network contains a chain of the repeated neural cell containing the inputting, outputting, and forget gates. This design controls how much data is kept, forgotten, and delivered to the output. The modules in the LSTM consist of 4 special interoperable gates that differ from the '*tanh*' single-function architecture commonly found in standard RNNs.

2.2.3.2. Autoencoder

Autoencoder is an unsupervised feed-forward neural network that recreates the input data while extracting its features through different dimensions. By constraining the latent space to have a smaller dimension than the input, the autoencoder is forced to learn the most critical features of the training data. The comparison or deviation of error vectors when feeding with normal and abnormal data can be used to classify the anomaly.

2.2.3.3. Improved LSTM-AE

The Improved LSTM-AE used in this research was constructed as a sequence-to-sequence model containing stacked LSTM layers. This network allows extracting essential features from the multivariate time series more efficiently. Indeed, the LSTM encoder learns a fixed-length vector representation of the input time series, and the LSTM decoder uses this representation to reconstruct the time series using the current hidden state and the value predicted at the previous time step.

2.2.4. Model development for anomaly detection of production well and natural gas compressor in HT field

The optimization process of the LSTM deep learning model includes the fine-tuning process for the following hyperparameters: number of neurons, number of layers, number of epochs, batch size, dropout size, learning rate, activation function, the optimizer for algorithm training, etc. The network, combined with the Random Search optimization technique and the predefined set of hyperparameters, was trained to achieve an optimized loss function of Mean Absolute Error (MAE).

2.3. Conclusion for Chapter 2

Using an improved ANFIS for enhancing the precision of the well production flow rate prediction is an appropriate approach. In the HT-MT field, the implementation of this method supports the well-testing and performance monitoring process by improving the virtual flow meter prediction, thereby raising the predictability of the allocation plan.

Using an LSTM-AE deep learning network facilitates the extraction of vital features from time series data, effectively learning fixed-length vector representation of the input time series. This makes the method well-suited for detecting anomalies in production wells and natural gas compressors, which play a crucial role in maintaining stable pressure and ensuring a consistent flow for the processing and recovery of flash gas.

CHAPTER 3:
APPLICATION OF INTELLIGENT DATA ANALYSIS MODEL TO IMPROVE GAS PROCESSING EFFICIENCY IN HAI THACH FIELD

3.1. Application of gas condensate prediction model at HT field

3.1.1. Input dataset

The authors utilized real-time surface input parameters from the same reservoir, including choke valve opening, tubing and downhole pressure, upstream choke temperature, downstream choke pressure and temperature, test separator pressure and temperature, and the inlet pressure of the gas processing system, to predict the gas and condensate production of each well in the database at the HT field.

3.1.2. Results and discussion when applying in the HT field

Using data smoothing techniques reduced the computation time while preserving the accuracy required for the time series prediction model. The evaluation metrics indicate that local regression "*loess*" provided the best results for both Subtractive and FCM clustering algorithms regarding different data preprocessing techniques (i.e., highest R-value and lowest MSE, mean error, and error standard deviation). Using k-fold cross-validation techniques revealed its ability to prevent overfitting and improve prediction accuracy, particularly for the Subtractive clustering model. In addition, the Improved ANFIS prediction model method is much simpler to set up than the current method, which requires regression and correction steps to estimate the fluid flows from the well entering the MT-HT pipeline.

3.1.3. Applying the research results to optimize the production of wells in the HT-MT field

The Genetic Algorithm (GA) is based on the concept of evolution, where genes in living beings symbolize a distinct feature separate from other genes. The genetic algorithm was designed to identify the optimal "well opening" gene sequence that aligns with the required gas nomination for 13 wells at the HT-MT field. The number of bits used to represent the specific characteristics of the well depends on the range of the well opening (fixed or adjustable choke size). The Improved ANFIS model employs a blend of genetic algorithms to calculate the optimal nomination plan. Various choke openings will be combined to attain maximum condensate production.

3.1.4. Evaluate the results at the HT field

The important results are highlighted as follows:

- The evaluation metrics show that local regression '*loess*' gives both clustering algorithms (Subtractive and FCM) the best results for different data preprocessing techniques (highest value of R and lowest value of MSE, error mean and error standard deviation);
- The k-fold cross-validation techniques demonstrated the capability to avoid the overfitting phenomenon and enhance prediction accuracy.
- The proposed model provided the lowest RMSE of 0.0645 and 0.0733 for training and testing data, respectively. It also has the highest R^2 of 0.9482 and 0.9337 for training and testing data and the highest Variance account of 0.9482 and 0.9334 for training and testing data.
- Compared with the method being applied, the prediction results using the Improved ANFIS give smaller RMSE values of 0.0733 and 0.12, respectively, with a higher coefficient of determination R^2.

3.2. Application of anomaly detection model for gas processing systems at HT field

3.2.1. Dataset for production well and natural gas compressor in HT field

For the anomaly detection model with the centrifuge gas compressor, the input dataset contains the multivariate time series inputs from multiple sensors installed at different compressor locations: pressure, temperature, flow, current, and power. Input data for well anomaly detection research includes time data; choke size opening; wellhead pressure before and after the choke valve; wellhead temperature before and after the choke valve; bottomhole pressure and temperature of all production wells in HT. In addition, data is labeled with anomaly instances for further performance evaluation.

3.2.2. Anomaly detection mechanism

3.2.2.1. Reconstruction error

The reconstruction error (the error between the input and the prediction vectors) is used to evaluate the machine learning model's performance and classify outliers. The reconstruction error during training autoencoder networks with normal instances is much lower than feeding with an anomalous sequence. The comparison or deviation of error vectors when feeding with normal and abnormal data can be used to classify the anomaly.

3.2.2.2. Binary confusion matrix

The binary confusion matrix expresses the relationship between predicted and actual classes or prediction/reconstruction error between the predicted /forecasted and the actual values. Then, the optimal threshold value was selected to achieve the highest F-score among the optimized models.

3.2.2.3. Threshold selection

The Random Search method efficiently optimized the Improved LSTM-AE network to attain the minimum MAE of the reconstruction error vector for various hyperparameter combinations. Subsequently, the optimal threshold value was determined to achieve the maximum F-score among the optimized models. Two key hyperparameters, namely timesteps and activation functions, were evaluated across different trials to identify the model with the highest F-score of 0.57143.

3.2.2.4. Timestep and activation function optimizations

In many instances, the activation functions ReLU and LeakyReLU yield similar results. For instance, with a timestep value of 3, both functions produce equivalent outcomes. When the activation function is set to ReLU and the timestep is 2 or 4, the highest F-score is 0.57143. However, with a timestep of 2, the MAE is 0.07077, which is lower than the MAE of 0.1189 for a timestep of 4. When using Leaky ReLU as the activation function, the best timestep value is still 4, resulting in a maximum F-score of 0.5. With both activation functions, as the timestep increases, the F-score decreases and the MAE of the reconstruction error increases.

3.2.3. Results and discussion when applying in the HT field

Previous studies have shown that any deep-learning model's activation function is crucial. However, it can be seen that, with the same timestep value, ReLU gives better F-score results. The loss distribution (MAE) of all the data is represented, and the threshold value at 0.1189 is selected to get the best F-score, and the anomaly instances have a much higher MAE of the reconstruction error. By comparing the timestep and the activation function, the authors can select the corresponding threshold value to get an F-score of 0.57143, True Positive Rate (TPR) of 0.667, and a False-Positive Rate (FPR) of 0.0003 with all the anomalies and an early finding correctly identified.

3.2.3.1. Anomaly detection results for production well in HT

Applying the DP Deviation monitoring method at the HT field reveals water production starting from the 20th of September, 2019. By utilizing periodic well testing with varying choke size openings and measuring the water salinity at the test separator, it can be observed that the Liquid Gas Ratio (LGR) has a tendency to rise, and there was a spike in salinity on the 22nd of September, 2019. However, one drawback of these methods is high latency, as the well-testing process is not conducted in real-time. When utilizing the Eclipse E300 hydrodynamic model in conjunction with well geophysical data, it becomes unfeasible to pinpoint the exact time of water production when the initial Gas Water Contact (GWC) cannot be determined and has gradually risen until the present time. When applying an anomaly detection method using an Improved LSTM-AE, the anomaly was shown on the 16th of September 2019. The benefit of this method is its ability to perform better with increased data and features. The model can be regularly trained and evaluated based on the time series data.

3.2.3.2. Anomaly detection for natural gas compressor, ensuring pressure stability for gas processing at HT field

Using the LSTM-AE anomaly detection model, anomalies were detected more accurately and earlier when evaluating the mean of error vectors for the features. An adaptable and accurate LSTM-AE network was developed to obtain the highest value of F-score of 0.57143 with all the anomalies and correctly identified early findings. The network consisted of 4 LSTMs with 448, 224, 224, and 448 neurons, respectively, 1 Dropout of 0.15 fraction, 1 Repeat Vector, and 1 TimeDistributed layer. The activation function was ReLU, L2 regularization of 0.025, the learning rate of 0.001, and MAE's loss function to provide the best performance metric.

3.2.4. Evaluation of detection result at HT field

The important results are highlighted as follows:

- The Random Search technique effectively optimized the improved LSTM-AE network to obtain the lowest MAE of reconstruction error vector for different hyperparameter sets such as network architecture, the number of neurons, sliding window size, activation function, regularization, dropout, batch size, and learning rate.

- The optimal threshold value is then selected to correctly classify the most anomaly instances with the highest achieved F-score among the optimized models.

CONCLUSION AND RECOMMENDATION

Conclusion

The gas production and processing system at the HT field have a long history of safe and efficient operation. Despite this, it faces various difficulties and challenges due to declining flow and pressure. Ensuring a continuous and stable operation of wells and processing equipment is a critical issue for the HT-MT field, as it is vital to meeting production and processing targets. Research results have verified the feasibility and cost-effectiveness of utilizing ML methods by leveraging the existing infrastructure and historical production data at the HT gas condensate field.

Implementing the Improved ANFIS algorithm has increased the accuracy of gas condensate gas flow prediction, allowing for the proactive opening and closing of wells based on consumption conditions. This helps to eliminate waste and inefficiency while maintaining high condensate flow and preserving reservoir energy, ensuring optimal recovery and production from each well. Additionally, the algorithm helps stabilize the processing system's input conditions. Data preprocessing and smoothing techniques such as local regression using weighted linear least squares and a 2nd-degree polynomial model '*loess*', random shuffling techniques, and k-fold cross-validation have shown the capacity to prevent overfitting and enhance the accuracy of the prediction model for the HT field when using both Subtractive and FCM partitioning algorithms.

The anomaly detection method has enhanced reliability and safety, resulting in increased equipment uptime of up to 99.99%, and guaranteed pressure and flow stability for the processing and recovery of flash gas. An Improved LSTM-AE network designed for anomaly detection in multivariate time-series features was trained using the Random Search optimization technique and the MAE loss function. The optimal hyperparameters were determined through this process. Afterward, the best threshold value was established to achieve the highest F-score among the optimized models.

Recommendation

Continue to research and develop ML algorithms for applications of reservoir management, predictive maintenance, and characterization analysis to serve the exploration and evaluation in Block 05-2; 05-3.

Continue to promote research and development of applied Machine Learning algorithms in Vietnam’s oil and gas industry, focusing on areas such as exploration, drilling, processing, and transportation.

The research results of this thesis can be applied to other oil and gas fields with similar offshore conditions in Vietnam.

LIST OF PUBLICATION

The Ph.D. candidate has earned recognition as the author or co-author of various scholarly works that have been honored at the state and ministerial levels, including the prestigious Ho Chi Minh Prize. One of these works, an article published in the SPE Journal, has been ranked within the top tier (Q1 group) of the ISI/SCOPUS list. Moreover, the Ph.D. candidate has contributed to a specialized conference of the SPE by publishing an article, namely:

Vietnamese

[1]. Tran Ngoc Trung, Ngo Huu Hai, Trieu Hung Truong, "Real-time monitoring and assessing the fouling resistance and performance of Plate Type heat exchanger in Hai Thach gas condensate platform", *Journal of Mining and Earth Sciences*, Volume 60, Issue 1 (2019), 81-89.

[2]. Tran Ngoc Trung, Ngo Huu Hai, Tran Vu Tung, Ly Van Dao, Trieu Hung Truong, "Enhancing production with natural gas ejector for Hai Thach gas condensate field - CFD model development", *PetroVietnam Journal*, Vol 5 (2020), page 14-24, ISSN 2615-9902.

[3]. Tran Ngoc Trung, Ngo Huu Hai, Tran Vu Tung, Ly Van Dao, Trieu Hung Truong, "Structural investigation of supersonic ejector for boosting low pressure wells in Hai Thach gas condensate field", *PetroVietnam Journal*, Vol 9 (2020), page 4-19, ISSN 2615-9902.

[4]. Tran Ngoc Trung, Tran Vu Tung, Hoang Ky Son, Ngo Huu Hai, Dao Quang Khoa, "Implementation of a Digital transformation platform at Hai Thach-Moc Tinh fields", *PetroVietnam Journal*, Vol 12 (2020), ISSN 2615-9902.

[5]. Nguyen Van Thinh, Trieu Hung Truong, Tran Thanh Hai, Ngo Huu Hai, Tran Ngoc Trung. "Research on the application of Artificial intelligence tools to diagnose the common failure of centrifugal pumps applied to gas

condensate transportation system at HT-MT field", *Journal of Mining and Earth Sciences*, Vol 63, Issue 4 [08-2022], page 81-90, DOI:10.46326/JMES.2022.63(4).08.

[6]. Co-author of the scientific project "Research and implementation of the Digital Transformation and Smart oilfield management program to improve the efficiency of management and production of the Hai Thach - Moc Tinh condensate gas field, Lot 05-2; 05-3, in the East Sea of Vietnam", Second prize, Vietnam Science and Technology Innovation Award (Vifotec).

[7]. Co-author of the scientific project "Nghiên cứu, phát triển và ứng dụng công nghệ để khai thác các mỏ khí - condensate với điều kiện đặc biệt phức tạp, thềm lục địa Việt Nam" đạt giải A, Giải thưởng Khoa học và Công nghệ Dầu Khí Việt Nam 2015-2020 (Lần 2) Quyết định số 5862/QĐ-DKVN, Giải thưởng Hồ Chí Minh và Giải thưởng Nhà nước về Khoa học và Công nghệ Đợt 6.

English

[8]. Tran Ngoc Trung, Trieu Hung Truong, "Increase the accuracy of Condensate metering by applying the new metering correction formula and standardize to follow API standard at HaiThach-MocTinh Condensate Field", *Earth Sciences and Natural Resources for Sustainable Development Vietnam 2018*.

[9]. Tran Vu Tung, Tran Ngoc Trung, Ngo Huu Hai, Nguyen Thanh Tinh, "Digital transformation in oil & gas company: A case study of Bien Dong POC", *Tạp chí Dầu Khí*, Số 10 năm 2020, ISSN 2615-9902.

[10]. Tran Ngoc Trung, Trieu Hung Truong, Tran Vu Tung, Ngo Huu Hai, Dao Quang Khoa, Hoang Ky Son, "Virtual multiphase flow metering using Adaptive Neuro-Fuzzy Inference System (ANFIS) – a case study of Hai Thach-Moc Tinh field, Offshore Vietnam", *SPE Journal*

SPE_J_206741, 202.

[11]. Son Hoang Ky, Tung Tran Vu, Tan Nguyen Ngoc, Tu Truong Anh, Duy Pham Hoang, Tran Ngoc Trung, Vinh Trinh Xuan, Anh Ngo Tuan, "Successful Application of Machine Learning to Improve Dynamic Modeling and History Matching for Complex Gas-Condensate Reservoirs in Hai Thach Field, Nam Con Son Basin, Offshore Vietnam", *SPE Symposium: Artificial Intelligence – Towards a Resilient and Efficient Energy Industry*, October 11-12, 202, DOI: 10.2118/208657-MS.

[12]. Son Hoang Ky, Tung Tran Vu, Tan Nguyen Ngoc, Tu Truong Anh, Duy Pham Hoang, Tran Ngoc Trung, Vinh Trinh Xuan, Anh Ngo Tuan. "Successful Case Study of Machine Learning Application to Streamline and Improve History Matching Process for Complex Gas-Condensate Reservoirs in Hai Thach Field, Offshore Vietnam", *SPE Middle East Oil & Gas Show and Conference*, December 2021, 10.2118/204835-MS.

[13]. Tran Ngoc Trung, Trieu, H. T., Tran, V. T., Ngo, H. H., & Dao, Q. K. (2021), "An overview of the application of machine learning in predictive maintenance", *Petrovietnam Journal*, 10, 47 – 61, https://doi.org/10.47800/PVJ.2021.10-05.

[14]. Hai H. Ngo, Tran Ngoc Trung, Tung V.Tran, Khoa Q. Dao, Trung T. Nguyen, Son K. Hoang, Truong H. Trieu, "Anomaly Detection for Centrifuge Natural Gas Compressor Using LSTM-Based Autoencoder in Hai Thach – Moc Tinh Field, Offshore Vietnam", *Integrated Petroleum Engineering for Energy Transition and Sustainable Development the 6th of October, 2022, Hanoi, Vietnam*, page 197-215.