I. Introduction

Natural gas is a key player in the global energy landscape, offering a cleaner alternative to traditional fossil fuels and serving various sectors. Its significance lies in its lower carbon dioxide emissions during combustion compared to other fossil fuels, making it an attractive option for industries seeking to reduce their environmental impact [1]. The widespread use of natural gas underscores its importance in meeting energy demands efficiently and sustainably.

However, the transportation of natural gas from production sites to end consumers presents logistical challenges, especially in the transportation phase. The extensive pipeline networks connecting offshore production platforms to onshore facilities are crucial for ensuring a continuous and stable supply of natural gas to consumers [2]. Maintaining optimal gas pressure throughout these transmission networks is essential to guarantee the safe and efficient delivery of natural gas.

Traditionally, various methods have been used to simulate gas flow throughout pipelines, such as the adaptive method of lines [3], Runge-Kutta discontinuous Galerkin method [4], transfer function models [5], and state space models [6]. These methods have been applied to simulate transient gas flow, state estimation, and control algorithms for natural gas pipelines. However, these traditional methods have limitations in forecasting future trends. For instance, they may not fully capture the complex dynamics and nonlinear relationships affecting gas flow, especially in scenarios with time-varying injections, withdrawals, and compression [7].

Data-driven methodology has gained popularity in the development of data science and machine learning techniques, as it is applicable to numerous areas of research. These techniques offer promising solutions for forecasting, predictive maintenance, and optimization tasks, providing a data-driven approach to complex challenges [8]. Data science, when applied to natural gas production and transmission, has the potential to significantly improve forecasting accuracy and operational efficiency.

In the realm of natural gas demand forecasting, researchers have explored different approaches such as univariate statistical methods and machine learning techniques [9][10]. The use of machine learning in natural gas price forecasting has gained significant interest due to its potential to enhance prediction accuracy [10]. Furthermore, the application of hybrid models combining different forecasting techniques has shown promise in improving forecasting performance [11]. Specifically, in the context of natural gas consumption forecasting, studies have employed methods like the Multi-Layer LSTM model, fuzzy decision trees, and deep learning approaches for more accurate predictions [12][13].

Although advanced forecasting methods that utilise artificial intelligence and deep learning offer certain benefits, classic models such as ARIMA have demonstrated their value in natural gas forecasting. ARIMA models are desirable because of their simplicity and interpretability, particularly in situations that require a compromise between accuracy and complexity. ARIMA models have been quite effective in predicting natural gas production. Research has demonstrated the efficacy of ARIMA models, such as SARIMA, SARIMAX, and ARIMAX, in accurately forecasting production levels and consumption patterns in the natural gas sector [14][15][16].

This study compares four ARIMA-based models to forecast the pressure at a sink node in a gas pipeline network that receives gas flow from four separate sources. The models are evaluated to determine which one performs best in predicting pressure. Pressure measurements from all sources, along with historical pressure data at the sink node, are used as input features for the models.