

연구논문

셰일가스 생산량 예측 모델의 성능개선을 위한 장단기 기억 신경망 기반 생산감퇴 특성 추출 및 활용 연구

이대모¹ · 신호진² · 임중세^{2*}

¹중소조선연구원 경영혁신본부 전략기획실, ²한국해양대학교 에너지자원공학과

Extraction and Application of Decline Characteristics Based on Long Short-Term Memory Neural Networks to Improve the Performance of Shale Gas Production Prediction Model

Dae-Mo Lee¹, Hyo-Jin Shin² and Jong-Se Lim^{2*}

¹Management Innovation Division/Strategic Planning Office, Research Institute of Medium & Small Shipbuilding, Busan, Korea

²Department of Energy & Resources Engineering, Korea Maritime & Ocean University, Busan, Korea

*Corresponding Author. Jong-Se Lim, jslim@kmou.ac.kr, Department of Energy & Resources Engineering, Korea Maritime & Ocean University, Busan, Korea

Received

10 October 2023

Final version Received

16 November 2023

Accepted

27 December 2023

Abstract

Various studies have used decline curve analysis (DCA) to predict shale gas production, with recent proposals focusing on artificial intelligence-based machine learning techniques. Future production is generally predicted using production history as input data, but few studies have examined how to improve predictive performance by processing specific data from given data. This study used Long Short-Term Memory (LSTM) neural networks to predict the production of shale gas, and predictive performance was improved by extracting and using decline characteristics to reflect the decline trend of shale gas. Future predictions of shale gas production based on Long Short-Term Memory neural networks can be improved by processing data on decline characteristics from production history and using them as input data.

Key words : shale gas, long short-term memory neural networks, decline characteristics, predicting production

요약

셰일가스 생산량을 예측하기 위해 생산감퇴곡선분석법을 이용한 다양한 연구가 진행되었으며, 최근에는 인공지능 기반의 머신러닝 기법이 제안되고 있다. 일반적으로 생산이력을 입력자료로 활용하여 미래의 생산량을 예측하는 방식이 적용되고 있지만, 주어진 자료에서 특정한 자료를 가공하여 예측성능을 개선하는 연구는 부족한 실정이다. 이 연구에서는 장단기 기억 신경망을 이용하여 셰일가스의 생산량을 예측하였으며, 생산이력으로부터 셰일가스의 감퇴경향을 반영할 수 있도록 생산감퇴특성을 추출한 후 활용하여 예측성능을 개선하였다. 향후 장단기 기억 신경망 기반의 셰일가스 생산량 예측 시 생산이력으로부터 생산감퇴특성에 관한 자료를 가공하고 입력자료로 이용함으로써 개선된 결과를 도출할 수 있을 것이다.

주요어 : 셰일가스, 장단기 기억 신경망, 생산감퇴특성, 생산량 예측

서 론

일반적으로 유·가스 생산량 예측 시에 활용되는 생산감퇴곡선분석법(Decline Curve Analysis, DCA)은 과거의 생산이력(production history)을 바탕으로 비교적 쉽게 생산량 예측이 가능하다는 이점이 있다(Li *et al.*, 2020). DCA 기법 중 널리 활용되고 있는 Arps(1945) 방정식은 경계영향유동(Boundary Dominated Flow, BDF)을 가정하여 장기간의 천이유동(transient flow)이 발생하는 셰일 저류층에 적용할 경우 부정확한 생산량 예측결과가 도출될 수 있다(Kanfar and Wattenbarger, 2012; Paryani *et al.*, 2016; Tan *et al.*, 2018). 이는 셰일가스(shale gas)가 부존된 저류층이 일반적인 유·가스 저류층과 달리 유체투과도가 낮아 수압파쇄(hydraulic fracturing)를 통해 생산이 이루어지며, 생산 초기에는 가스의 생산감퇴경향이 크게 나타나지만 생산이 지속됨에 따라 생산량의 감퇴율이 매우 작아지고 장기간의 천이유동이 발생하기 때문이다(Kim *et al.*, 2014; Luo and Su, 2022). 이와 같은 유동 특성으로 인하여 셰일 저류층에 대한 생산량 예측 시 Arps 방정식을 활용하기에 어려움이 존재한다. 이에 따라 셰일 저류층의 생산특성을 고려하기 위하여 Power Law Exponential Decline(PLE), Stretched Exponential Production Decline(SEPD), Duong method 등의 DCA 기법이 제안되었다(Ilk *et al.*, 2008; Valko and Lee, 2010; Duong, 2011). 그러나 저류층 특성과 수압파쇄 설계, 자연 균열 등과 같은 유동특성의 상이성으로 인해 경우에 따라 생산량을 과대, 과소 예측하는 결과가 나타났다(Han, 2018).

이러한 한계를 보완하기 위하여 2000년대 이후로 컴퓨터 성능이 발전됨에 따라 인공지능 기반의 머신러닝(machine learning) 기법 적용을 통한 새로운 접근법이 대두되었으며, 다량의 연산을 바탕으로 복잡한 수식을 풀거나 입·출력인자 간 관계를 모사할 수 있게 되었다. 이와 같은 방법을 바탕으로 셰일 저류층에서의 생산량 예측을 위하여 인공신경망(Artificial Neural Networks, ANN), 서포트 벡터 머신(Support Vector Machine, SVM), 랜덤 포레스트(Random Forest, RF) 등을 기반으로 셰일가스의 궁극가채량(Estimated Ultimate Recovery, EUR) 또는 특정 시점의 누적생산량(cumulative production)을 예측하는 연구가 진행되다 있다(Luo *et al.*, 2019; Wang *et al.*, 2019; Oh *et al.*, 2021; Shin *et al.*, 2021). 최근에는 시간에 따른 생산량의 추이를 예측하기 위해 시계열 자료 처리에 유용한 순환신경망(Recurrent Neural Networks, RNN) 기반의 딥러닝(deep learning)이 활용되고 있다. 종전에는 단순한 구조의 RNN을 바탕으로 생산량 예측이 수행되었으나, 시계열 자료가 길어짐에 따라 기울기 소실(gradient vanishing) 문

제가 발생하여 장기적인 예측에 한계가 드러났다. 이에 장기간 기억을 보존함으로써 예측의 정확도를 개선할 수 있는 장단기 기억(Long Short-Term Memory, LSTM)을 기반으로 생산량 예측모델을 개발하기 위한 연구가 활발히 진행되고 있다(Lee *et al.*, 2019; Song *et al.*, 2020; Li *et al.*, 2020; Kocoglu *et al.*, 2021; Lee *et al.*, 2022).

Lee *et al.*(2019)은 생산중단(shut-in) 기간의 생산이력을 제외하여 생산량 자료를 가공하고 생산중단 기간의 길이에 대한 시계열 자료를 만들어 활용하는 방식을 제안하였으며, 생산중단 기간을 추가 입력자료로 이용함으로써 예측 오차를 줄일 수 있음을 파악하였다. Song *et al.*(2020)은 LSTM에 6일간 생산량과 초크 사이즈(choke size)를 입력하여 예측치를 도출하는 방식을 적용하였으며, 예측한 생산량을 그다음 시점의 예측에서 입력자료로 이용함으로써 장기간의 생산량을 예측하였다. 또한, ANN, 단순 RNN, Arps Hyperbolic과 LSTM의 예측결과를 비교하여 LSTM이 신뢰성 있는 예측결과를 도출할 수 있음을 확인하였다. Li *et al.*(2020)은 LSTM 등의 RNN을 이용한 장기간의 석유 생산량 예측 시 생산이력 외에도 생산중단 기간, 초크 사이즈, 튜빙 압력(tubing pressure) 등의 생산운영조건을 추가로 이용하여 예측모델의 성능을 높일 수 있음을 파악하였다. Kocoglu *et al.*(2021)은 LSTM, Bi-directional LSTM(Bi-LSTM), Gated Recurrent Unit(GRU)와 같은 다양한 RNN 기반의 모델을 이용하여 생산 초기 24개월간의 셰일 가스 생산량을 바탕으로 이후 25~64개월에 관한 장기간의 생산량을 예측할 수 있음을 확인하였다. Lee *et al.*(2022)은 6, 12개월의 셰일가스 생산이력만이 주어졌을 경우 Arps Hyperbolic 적용 시 감퇴지수(decline exponent)가 음수로 나타나 비정상적인 예측치가 도출되는 경우가 존재함을 확인하였으며, 대안으로 LSTM 기반 예측모델을 활용하여 생산량 예측이 가능함을 파악하였다. 또한, 생산이력을 이용하는 LSTM 모델과 생산관련인자(유정원결 조건, 수압파쇄 인자 등)를 입력자료로 하는 ANN 모델을 결합하여 개별적인 생산정의 특성을 반영함으로써 셰일가스 생산량 예측 모델의 오차를 개선할 수 있음을 확인하였다. 그러나 앞서 인용한 논문에서 추가 입력자료로 활용한 생산관련인자는 시간에 따라 변화하는 생산량의 감퇴 정도 또는 감퇴 경향을 고려하기에 제한적이다.

이렇듯 LSTM을 활용한 기존의 연구에서는 예측모델 구축 시, 주로 생산이력만을 입력자료로 이용하여 생산량을 예측하거나 생산운영조건을 시계열 자료 형태로 모델에 반영하여 예측성능을 개선하고자 하였다. 그러나 Lee *et al.*(2019)과 같이 주어진 생산이력을 가공하여 특정한 자료를 만들고 이를 입력자료로 활용하여 예측모델의 성능을 개선하고자 하는 연구는 부족한 상황이다. 따라서 주어진 자료

가 제한적이라 하더라도 특정한 정보를 끌어내 입력자료로 활용하는 접근법에 관한 연구가 필요하다고 볼 수 있다. 이 연구에서는 Lee *et al.* (2022)의 연구방법을 활용하여 생산이력만을 이용하는 LSTM 기반의 예측모델을 구축하고 생산감퇴경향이 급격하게 나타나는 셰일가스층의 생산감퇴 특성을 생산이력으로부터 추출 및 입력자료로 반영하여 감퇴특성을 고려할 수 있는 LSTM 모델을 구축하고자 한다. 더 나아가, 궁극적으로는 생산감퇴특성을 이용하여 LSTM 기반의 예측모델의 성능을 개선하고자 한다.

신경망 및 최적화 기법

장단기기의 신경망(LSTM)

시간에 따라 변화하는 생산량 및 생산거동을 예측하기 위해서는 딥러닝의 일종인 RNN 기반의 모델이 범용적으로 활용되고 있다(Lee *et al.*, 2019; Song *et al.*, 2020). RNN은 이전 시점의 정보를 보존하고 다음 시점으로 관련 정보를 전달함으로써 순환적인 흐름을 바탕으로 시계열 자료 처리 및 예측에 이점을 가지고 있다(Graves *et al.*, 2013; Alom *et al.*, 2019; Alakeely and Horne, 2020; Ji *et al.*, 2021; Ng *et al.*, 2022). 그러나 단순 RNN은 시계열의 길이가 길어질 때 기울기 소실 문제가 나타나며, 초기의 입력정보가 후기의 예측결과에 제대로 반영되지 않아 학습성능이 저하되는 문제가 발생한다(Alom *et al.*, 2019; Ki *et al.*, 2019; Kocoglu *et al.*, 2021). 이를 보완하고자 Hochreiter and Schmidhuber(1997)에 의해 LSTM이 제안되었으며, 이는 cell state, hidden state를 통해 이전의 가중치를 기억하거나 제거 여부를 결정하여 RNN에서 발생하는 기울기 소실 문제를 해결하였다(Jeon *et al.*, 2019). 또한, LSTM 모델을 이용할 경우에는 다수의 시계열 자료를 입력자료로 활용할 수 있어 DCA와 달리 생산중단 기간의 발생 여부, 초크 사이즈 변화 등의 생산운영조건을 반영할 수 있다. LSTM 모델은 목적에 따라 입·출력자료를 다양한 형태로 구성할 수 있고 유·가스정 생산량 예측에서는 일정 길이의 시계열 자료를 입력하여 다음 시점에 해당하는 하나의 값을 예측하는 many-to-one 방식이 적용되고 있다(Ki *et al.*, 2019; Li *et al.*, 2020).

베이지안 최적화

기존에는 일정한 범위 및 간격 내에서의 탐색을 통해 해를 도출하는 격자 탐색(grid search), 일정한 규칙 없이 임의의 값을 무작위로 선정하여 모델의 성능을 평가하여 최고의 성능을 보이는 hyperparameter 조합을 최적해로 찾아내는 무작위 탐색(random search)이 적용되어 왔다. 하지만, 이와 같은 방법들은 이전 시점 또는 과거의 탐색을 통해 파악

할 수 있는 사전지식을 반영할 수 없기에 최적해를 찾는 과정이 다소 체계적이지 못하다는 한계가 있다(Choi *et al.*, 2020). 최근에는 사전지식을 바탕으로 다음 탐색 지점을 선정하여 효과적으로 최적해를 도출할 수 있는 베이지안 최적화가 활용되고 있으며, 신경망의 구조가 깊어지거나 복잡해짐에 따라 최적화가 필요한 hyperparameter의 수가 증가하여 최적해를 파악하기 어려움이 있을 때 사용될 수 있는 것으로 알려져 있다(Kocoglu *et al.*, 2021).

현장 자료 전처리 수행

미국 내 천연가스의 주요 생산지인 Marcellus Shale은 2022년 9월을 기준으로 미국의 전체 셰일가스 생산량 중 약 20%를 기록하고 있다(EIA, 2022). 해당 셰일층은 Pennsylvania, West Virginia, Ohio, New York 주에 약 140,000 km²의 면적에 걸쳐 분포하고 있으며, 평균적으로 심도는 약 2 km에 달하고 해당 지층에는 500 Tcf(trillion cubic feet) 이상의 원시부존량이 존재하는 것으로 알려져 있다(Engelder and Lash, 2008; Gihm *et al.*, 2011). 이처럼 많은 양의 셰일가스를 지속적으로 생산하고 있으며, 다수의 생산정이 위치하고 있는 Marcellus Shale을 연구대상 지역으로 선정하였다.

Enverus의 Drillinginfo를 통해 Marcellus Shale 중 Pennsylvania 주에 위치한 셰일가스 수평정의 자료를 획득하였다(Fig. 1). 이 연구에서는 최대 생산량(initial peak rate)을 시작점으로 하여 지속적인 감퇴경향이 나타나는 생산이력 자료를 활용하기 위하여 최대 생산량의 발생 이전에 나타난 생산이력을 제거하였다. 또한, 1~48개월의 생산이력을 기준으로 매월의 생산량에 관한 통계치를 바탕으로 전반적인 생산정의 생산규모를 파악하였다. 이를 통해 대부분의 생산정과 달리 과대 또는 과소 생산(1 MBOE 이하)되어 이

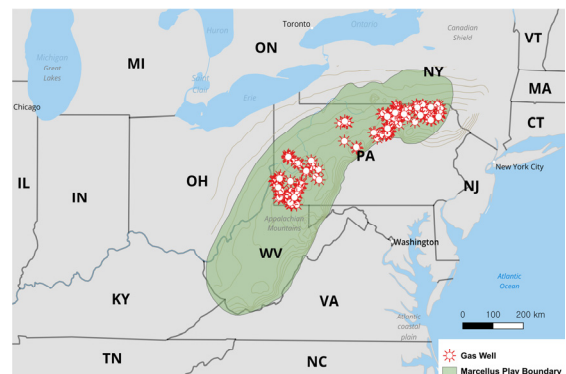


Fig. 1. Spatial distribution of acquired data for horizontal shale gas wells in Pennsylvania.

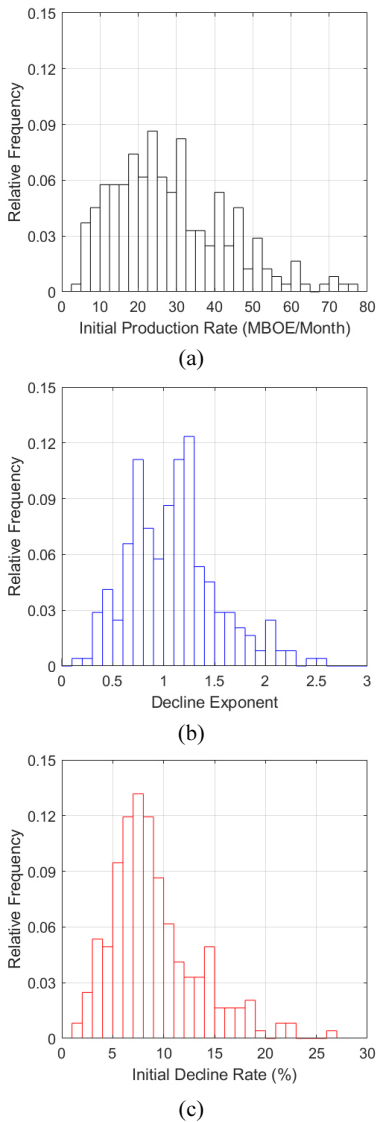


Fig. 2. Histograms of DCA parameters after pre-processing: (a) Initial production rate; (b) Decline exponent; (c) Initial decline rate.

상치(outlier)로 분류될 수 있는 생산량을 기록한 생산정을 제외하여 생산이력 및 생산정 자료에 대한 전처리를 수행하였다. 이와 같은 과정을 통해 취득한 자료에 대해 전처리를 수행하여 48개월간의 생산량 자료를 가진 239개의 생산정 자료를 확보하였다. 이때, 1~48개월의 생산이력을 기준으로 각 생산정 별 Arps Hyperbolic 적용 결과를 통해 초기 생산량(initial production rate), 감퇴지수, 초기 감퇴율(initial decline rate)의 분포를 확인하였으며, 일반적인 석유자원과 달리 셰일가스의 특성상 감퇴지수가 1을 초과하는 생산정이 다수 존재하였다(Fig. 2).

이 연구에서는 생산 변동성이 큰 셰일 저류층의 특성을 고려하기 위하여 다양한 생산거동을 하나의 모델에 학습시켜 여러 생산정에 활용할 수 있는 LSTM 모델을 구축하고자 하였다. 모델 설계 시에는 MathWorks社의 MATLAB R2022a를 사용하였으며, 앞서 전처리를 수행한 전체 자료를 무작위로 배열한 뒤 이를 학습(195개 생산정), 검증(24개 생산정), 테스트(20개 생산정) 자료로 분할하여 활용하였다.

생산이력을 이용한 LSTM 기반 셰일가스 생산량 예측

LSTM 모델 설계

이 연구에서는 Lee *et al.*(2022)과 같이 many-to-one 방식을 반복하여 장기간의 생산량을 예측할 수 있는 LSTM 기반 예측모델을 구축하고자 하였다(Fig. 3). 학습 시 모델의 입·출력자료로는 실제 월간 생산량을 활용하였고, 생산 초기로부터 일정 기간의 생산량을 입력자료로 하여 그다음 단일 시점의 생산량을 출력할 수 있도록 학습하였다. 테스트 시 첫 번째 예측에서는 실제 생산량만을 이용하고 이후에는 앞서 도출한 예측치를 실제 생산량으로 가정하여 다음 시점의 생산량 예측 시 입력자료로 활용하였으며, 이와 같은 방식으로 반복하여 48개월까지의 생산량을 예측하였다.

학습률, 학습 반복 횟수 등과 같은 학습에 관련된 매개변

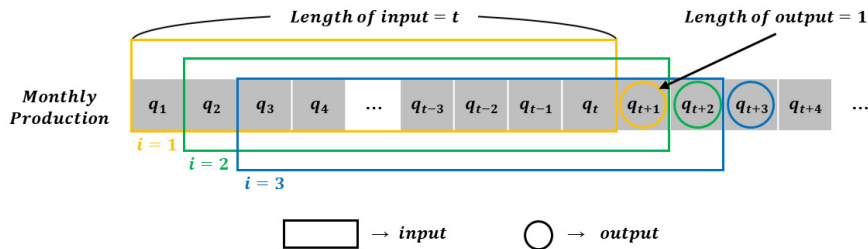


Fig. 3. Illustration of the many-to-one method for the prediction of shale gas production (Lee *et al.*, 2022).

수와 LSTM 층의 수와 같은 모델의 구조에 연관된 매개변수를 hyperparameter라 하며, 예측모델 구축 시 모델의 학습과 구조 관련 hyperparameter에 따라 예측성능이 상이하므로 예측성능을 극대화할 수 있는 최적해를 찾아야 한다. 다만, 다수의 hyperparameter를 동시에 최적화할 경우에는 많은 시간이 요구될 뿐만 아니라 최적해를 찾기에 어려움이 존재한다.

이 연구에서는 Table 1에 제시된 기준을 바탕으로 일부 hyperparameter를 고정하여 Lee *et al.*(2022)의 연구와 같이 초기 예측모델을 구축하였다. 학습 최적화 기법으로는 운동량 개념을 통해 과거의 정보를 일부 반영하면서 새로 계산된 경사 방향을 조합하여 최종적인 갱신 크기를 계산하는 Adaptive Moment Estimation(Adam)을 이용하였으며, 총 100번의 epoch가 수행되는 과정에서 20회의 epoch가 진행될 때마다 0.1의 learning rate drop factor를 적용하여 기존 학습률의 10%로 학습률을 재설정함으로써 예측모델의 학습을 진행하였다.

모델 구조에 관한 hyperparameter 중 LSTM 층의 수는 그 값이 증가할수록 모델의 구조가 복잡해지며, 전체적인 연산의 수가 증가하여 학습에 오랜 시간이 소요될 수 있어 적절한 수를 선정해야 한다. 그리고 학습 관련 hyperparameter 중 Dropout과 L2 Regularization은 예측모델의 복잡도를 낮추어 과적합(overfitting)의 발생을 방지하고 예측성능을 향상할 수 있다(Li *et al.*, 2022). 따라서 Table 2에 나타난 총 3개의 hyperparameter에 대한 최적화를 통해 과

도한 LSTM 층의 활용을 방지하고 적절한 Dropout과 L2 Regularization 값을 이용함으로써 과적합을 피하여 LSTM 모델을 구축하고자 하였다.

이 연구에서는 생산량 예측에 적합한 최적의 모델 구조와 학습조건을 효율적으로 파악하고자 베이지안 최적화를 활용하였다. 이때, 최적화를 위한 척도로 RMSE(Root Mean Square Error)를 사용하였으며, 이는 식 (1)과 같이 나타낼 수 있다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

여기서, \hat{y}_i 는 LSTM 모델을 통해 도출한 i 번째의 예측값, 예측한 생산량에 해당하며, y_i 는 실제 생산량, n 은 예측의 총 횟수를 나타낸다.

학습 자료를 바탕으로 100회 반복 학습을 진행함과 동시에 검증 자료에 대하여 RMSE를 최소화하는 조건을 기준으로 최적의 hyperparameter를 도출하였으며, 이를 활용하여 최적화된 예측모델을 구축하였다. LSTM 모델의 구조는 Lee *et al.*(2022)이 제안한 것과 같이 시계열 입력층(sequence input layer)과 LSTM 층, Dropout 층, 완전 연결층(fully connected layer), 회귀를 통한 출력층으로 구성하였다.

생산이력을 이용한 생산량 예측 및 분석

이 연구에서는 생산 시작부터 12개월의 생산량을 입력자료로 활용하는 경우를 가정하여 예측모델을 구축하고자 하였다. Lee *et al.*(2022)이 수행한 연구와 같이 학습 시의 입·출력자료의 구성은 활용 가능한 생산이력의 길이를 고려하여 12로 설정하였으며, 이와 같은 자료를 바탕으로 최적화된 각 모델의 hyperparameter는 Table 3과 같다. 테스트 시에는 1~12개월의 생산이력을 입력자료로 구성하였으며, 예측한 생산량을 실측치로 가정하여 다음 예측에서 입력자료로 활용하고 many-to-one 방식을 반복함으로써 48개월까지의 생산량을 예측하였다.

최적화된 12개월의 LSTM 모델을 활용하여 테스트 생산

Table 1. Fixed hyperparameter of the LSTM model (Lee *et al.*, 2022)

Type	Hyperparameter	Value
Training	Optimization method	Adam
	Max epoch	100
	Mini batch size	19
	Shuffle	Every-epoch
	Initial learning rate	0.01
	Learning rate drop period	20
	Learning rate drop factor	0.10

Table 2. Range of hyperparameter for optimization

Type	Hyperparameter	Min.	Max.
Architecture	Depth of LSTM layer with Dropout	1	5
	Dropout rate	0	1
Training	L2 Regularization value	10^{-7}	1

Table 3. Optimized hyperparameter of LSTM models (Lee *et al.*, 2022)

Type	Hyperparameter	Model
Architecture	Depth of LSTM layer with Dropout	1
	Dropout rate	0.0033
Training	L2 Regularization value	0.0053

정에 대한 생산량을 예측하였다. 생산정 별 예측결과와 경향을 분석하기 위하여 시계열 자료인 생산이력의 주요 특징으로 볼 수 있는 초기 생산량을 이용하였다. 일반적으로 모델의 예측성능 분석 시에는 정량적인 평가를 위하여 RMSE와 MAPE(Mean Absolute Percentage Error) 등의 예측성능 지표가 활용되고 있다(Botchkarev, 2019). 여러 자료에 대한 평균적인 오차를 의미하는 RMSE는 자료의 크기에 의존적인 경향을 보이나, 예측하고자 하는 값과 동

일한 단위를 가지고 있어 오차를 직관적으로 이해하기에 용이하다(De Gooijer and Hyndman, 2006; Ochella and Shafiee, 2021). 식 (2)와 같이 도출되는 MAPE는 다수의 자료에 관한 평균 오차율을 의미하며, RMSE와 달리 자료의 크기에 의존적이지 않으므로 다른 자료군 또는 모델과의 오차율을 비교할 경우에 활용하고 있다(Ochella and Shafiee, 2021). 따라서 이 연구에서는 생산량 예측결과에 대한 RMSE와 MAPE를 도출하여 LSTM 모델의 예측성능을 파악하고자 하였다.

$$MAPE(\%) = \frac{100}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (2)$$

각 생산정의 초기 생산량에 따른 RMSE를 비교하였을 때, 전반적으로 초기 생산량이 커짐에 따라 RMSE가 증가하는 경향이 나타났다(Fig. 4(a)). 그리고 MAPE를 기준으로 약 15 MBOE 미만의 초기 생산량을 가지는 생산정 중 일부에서 상대적으로 높은 MAPE를 보였다(Fig. 4(b)). 이를 통해 생산정의 생산량이 작아 RMSE의 값이 상대적으로 작게 나타나더라도 MAPE 오차가 크게 발생할 수 있음을 파악하였다. 이처럼 생산량 자료만을 입력자료로 활용할 경우 생산량의 크기에 따라 예측성능에 편차가 존재하므로 추가적인 인자를 활용함으로써 근본적으로 예측모델의 오차 감소 및 성능 개선할 수 있는 방법이 필요하다.

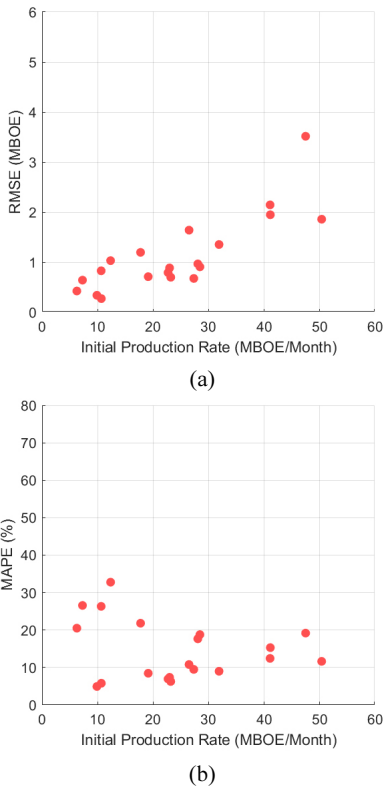


Fig. 4. Predictive performance of the test wells by initial production rate: (a) RMSE; (b) MAPE.

생산감퇴특성을 고려한 LSTM 기반 셰일가스 생산량 예측

생산감퇴특성 추출

복잡한 셰일층의 생산특성을 고려하기 위하여 생산이력 외의 생산감퇴특성에 관한 인자를 입력자료로 활용하고자 하였다. LSTM 모델을 바탕으로 단일 생산정에 관한 생산량 예측 시, Fig. 5와 같이 입·출력자료가 구성된다. 이를 통해 볼 수 있듯이 생산이력의 주요 정보인 초기 생산량의 경

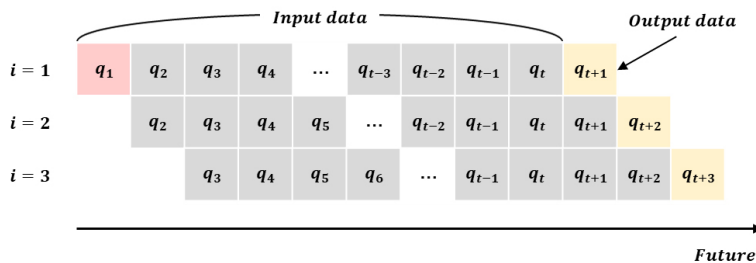


Fig. 5. An illustration of the iteration of LSTM model.

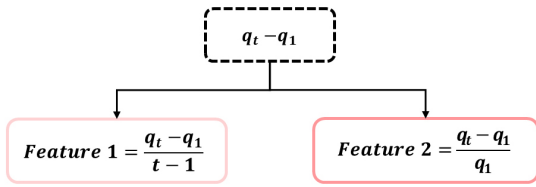


Fig. 6. Features extracted from the production history.

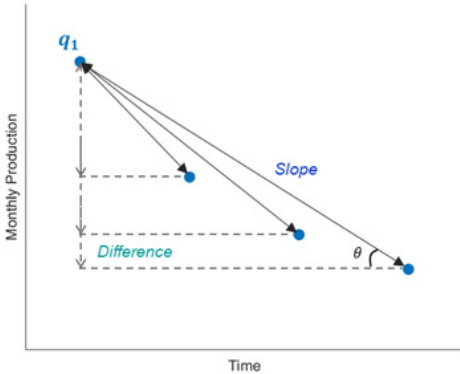


Fig. 7. A conceptual diagram of decline characteristics.

우, 첫 번째 예측에서만 입력자료로 사용된다는 한계가 존재한다. 또한, Lee *et al.*(2022)의 연구에서는 고려하지 못하였던 시간에 따라 변화하는 감소한 생산량, 생산감퇴경향 등을 반영하고자 하였다.

이에 초기 생산량 정보를 지속적으로 예측모델에 반영할 수 있도록 생산이력으로부터 초기 생산량과 생산시점 별 생산량의 차이에 관한 정보를 추출하여 시계열 형태로 예측모델의 입력자료로 활용하고자 하였다. Figs. 6, 7과 같이 생산이력으로부터 각 생산시점의 시간 간격을 고려하기 위하여 초기 생산량 기준 시간에 따라 변화하는 기울기(slope)에 대한 특성(Feature 1), 초기 생산량의 크기를 고려하여 시점별 생산량의 차이(difference)에 관한 정보(Feature 2)를 추가 입력자료의 후보로 선정하였다. 해당 특성은 시점마다 존재하는 정보로써 생산이력과 같은 시계열 자료로 간주할 수 있다.

LSTM 모델 설계

생산이력과 생산감퇴특성을 활용하는 LSTM 예측모델(LSTM_{PH+PD}) 설계 시에는 12개월의 생산이력만을 이용할 수 있는 상황을 가정하였다. 이때, 1~12개월의 실제 생산이력으로부터 생산감퇴특성을 추출하였으며, 생산감퇴특성이 생산이력과 동일한 길이로 예측모델에 입력되도록 설계하였다(Fig. 8). 이와 같은 구조를 바탕으로 생산감퇴특성

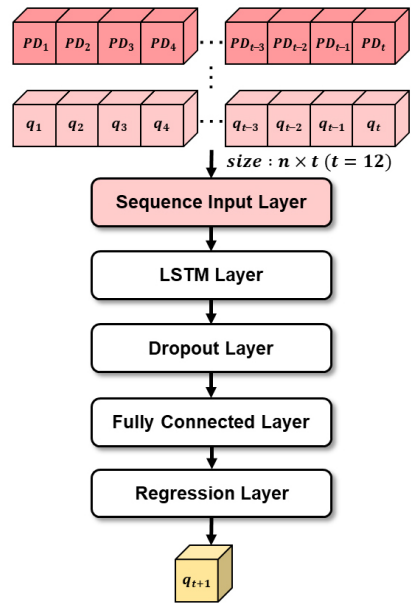


Fig. 8. Structure of the LSTM_{PH+PD}.

Table 4. Additional input data and optimized hyperparameter of LSTM_{PH+PD}

Additional input data	Feature 1 & 2
Depth of LSTM layer with Dropout	1
Dropout rate	0.0183
L2 Regularization value	0.0023

정보인 Feature 1과 2의 조합을 달리하여 개별적인 LSTM 모델을 구축하였다. 그 결과, Feature 1 또는 Feature 2에 관한 단일의 생산감퇴특성을 반영할 경우, 생산이력만을 이용한 LSTM_{PH}에 비하여 RMSE와 MAPE에 관한 예측성능이 다소 개선됨을 파악하였다. 더 나아가, 두 가지의 생산감퇴특성 정보를 모두 활용하였을 때 최상의 예측성능을 보임을 확인하였고 해당 모델의 추가 입력자료 및 hyperparameter의 조건은 Table 4와 같다.

생산감퇴특성을 반영한 생산량 예측

생산이력 및 생산감퇴특성을 활용한 LSTM 모델과 생산이력만을 이용한 모델 간 생산량 예측결과를 비교하였다(Figs. 9, 10). LSTM_{PH+PD}를 이용할 경우 LSTM_{PH}에서 과대 예측이 발생한 Well 3, 10과 더불어 과소 예측된 Well 126, 190에 대해서도 결과가 개선되어 예측한 생산거동이 실측치와 유사하였다. 테스트 생산정의 전반에 걸쳐 LSTM_{PH+PD}의 예측결과가 실제 감퇴거동과의 간극을 줄이

는 경향을 보였으나, 생산운영조건의 변화가 나타난 것으로 보이는 생산정에 대해서는 생산감퇴특성의 활용 여부와 관계없이 예측결과가 실제 생산거동과 차이가 발생하였다.

생산감퇴특성을 추가로 이용한 경우 전체 테스트 생산정 중 14개(70%)의 생산정에서 RMSE를 줄였다. 또한, MAPE를 기준으로 15개(75%)의 생산정에 대한 예측성능을 개선하였으며, 오차율의 증가 폭은 최대 약 2%에 불과한 것에 비하여 감소 폭은 최대 약 19%의 개선 효과가 나타

났다. 따라서 생산이력과 더불어 시간에 따라 변화하는 생산감퇴특성을 추가적으로 이용함으로써 LSTM_{PH}에서 과대·과소 예측이 나타난 생산정의 예측결과를 동시에 개선하였으며, 생산정에 따라 정도는 상이하나 전반적인 예측 성능을 높일 수 있음을 파악하였다. LSTM_{PH+PD}에서는 LSTM_{PH}가 과대 예측한 경우(Well 3, 10)뿐만 아니라 과소 예측한 경우(Well 126, 190) 모두에서 비교적 큰 폭으로 예측결과를 개선하였다. 그 외 Well 97, 132와 같이 생산이 지

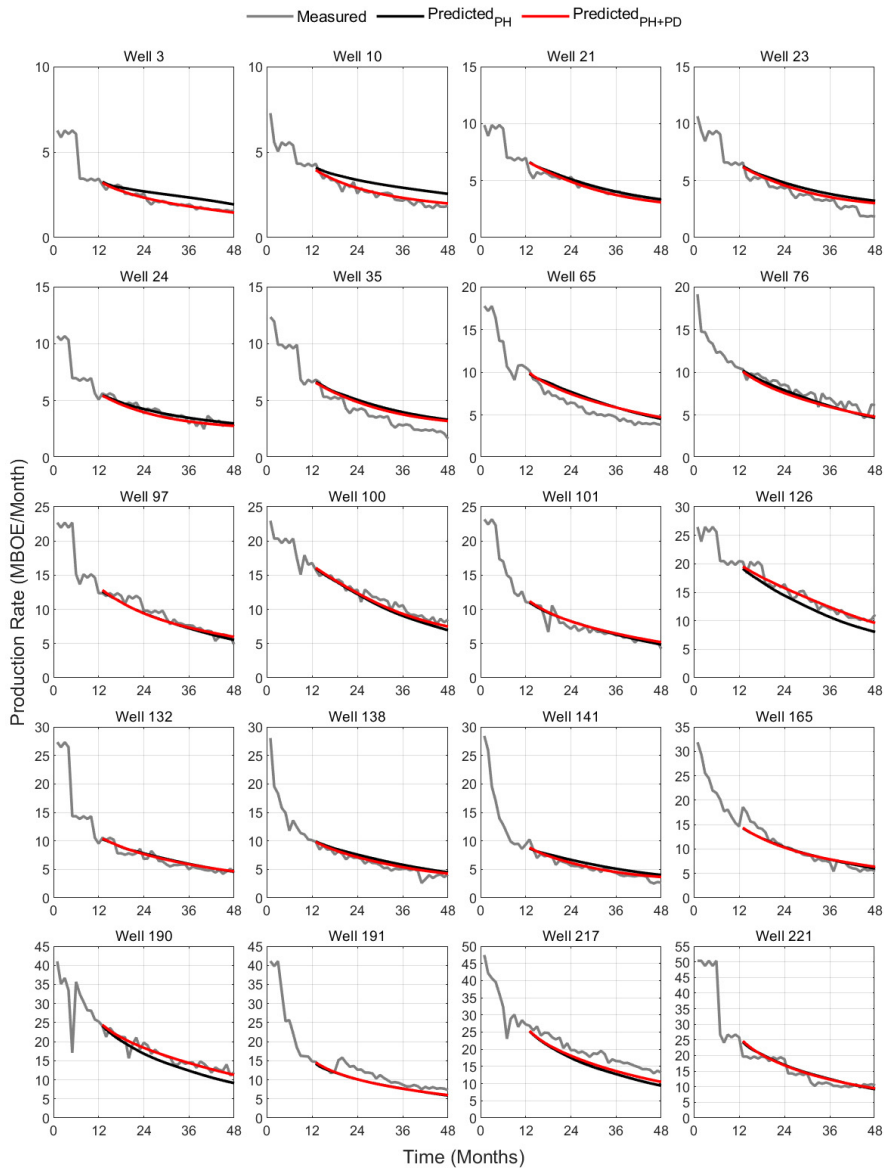


Fig. 9. Prediction results of LSTM_{PH} and LSTM_{PH+PD}.

속됨에 따라 생산량의 감소 폭이 작아져 생산거동의 기울기가 완만해지는 생산정에서는 예측모델에 관계없이 약 10%의 MAPE 이하에서 일정한 예측결과가 나타났다.

LSTM 모델은 입력자료의 마지막 시점에 해당하는 값에 주요한 영향을 받으며, 예측이 반복됨에 따라 생산량의 감소 폭이 지속적으로 작아져 완만한 곡선형의 생산거동을 예측한다. 이로 인하여 생산운영조건 등의 변화로 일정한 생산거동을 벗어나서 생산량이 진동하거나 불규칙한 형태를 보일 경우에는 해당 생산량을 정확히 예측하기에는 어려움이 있다. 이러한 관점에서 볼 때, 20% 수준의 MAPE를 보이는 Well 23, 35의 경우에는 입력자료를 평활(smoothing)하여 계단형 거동을 매끄럽게 만든다면 보았던 결과를 얻을 수 있을 것으로 보인다. 또한, Well 65는 10개월을 기준으로 생산거동이 바뀌어 생산 후반부에서 실측치와 결과 간에 비교적 큰 차이가 발생하였으나, 생산거동이 변화한 부분을 기점으로 재예측한다면 오차를 줄일 수 있을 것

으로 판단된다. Well 165, 191에서는 13~48개월의 예측구간 내 급증한 생산량을 예측하기에 어려움이 존재하였다. 이러한 결과는 LSTM 기반의 예측 시 입력자료로 이용한 생산이력의 감퇴경향이 유지된다는 가정이 동반되기 때문이며, 불확실성을 고려한 확률론적 예측 방법을 도입한다면 신뢰성 있는 결과를 도출할 수 있을 것으로 사료된다.

연구결과의 개선 정도 파악을 위하여 Lee *et al.*(2022)이 제시하였던 생산관련인자(production-related factors)를 활용한 LSTM 모델(LSTM_{PH+PF})의 결과와 비교하였다. Lee *et al.*(2022)은 생산관련인자를 활용할 수 있는 상황을 가정하였으며, 생산이력 외에 초기 생산량, 실수직심도(True Vertical Depth), 수압파쇄 횟수, 총 천공간격을 포함하여 총 4개의 생산관련인자를 추가 입력자료로 활용한 LSTM_{PH+PF}을 구축하였다. 해당 논문에서 제시한 LSTM_{PH+PF}과 LSTM_{PH}, LSTM_{PH+PD} 모델 간 예측성능을 정량적으로 파악하고자 각 생산정의 생산량 크기에 관계없이 오차의 개선정도를

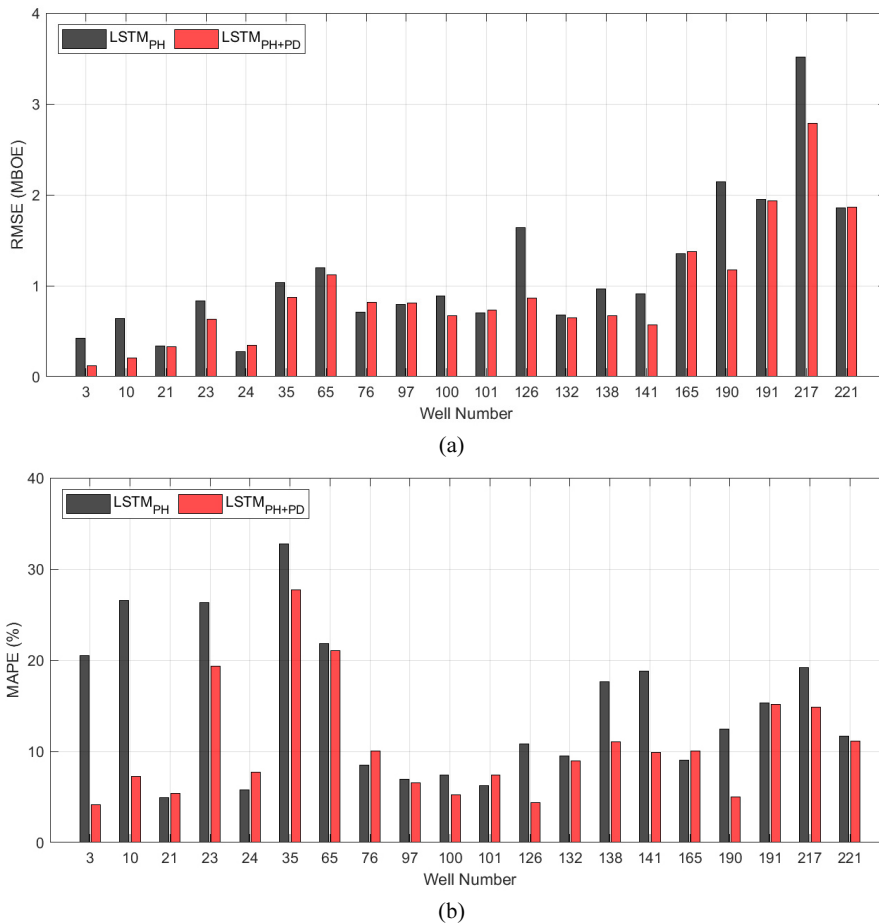


Fig. 10. Comparison of the prediction performance of LSTM_{PH} and LSTM_{PH+PD}: (a) RMSE; (b) MAPE.

Table 5. Comparison of prediction performance of LSTM models

	Type of model		
	LSTM _{PH}	LSTM _{PH+PF} (Lee <i>et al.</i> , 2022)	LSTM _{PH+PD}
Average	14.60	11.51	10.61
Minimum	4.94	6.57	4.17
Median	12.04	11.20	9.40
Maximum	32.78	19.35	27.73

파악하기 용이한 MAPE를 이용하였다. 개별 생산정에 관한 각 모델의 MAPE 통계치를 바탕으로 생산감퇴특성을 추가 입력자료로 활용함으로써 예측성능 개선이 가능함을 파악하였다(Table 5). 생산감퇴특성을 이용한 경우 생산이력만을 활용한 모델, 생산관련인자를 추가로 활용한 모델보다 MAPE의 평균값, 최솟값과 중간값이 상대적으로 작게 나타났다. 이와 같은 결과를 통해 가용한 생산관련인자가 없는 상황에서도 생산이력으로부터 추출한 생산감퇴특성을 이용함으로써 LSTM 모델의 전반적인 예측성능을 향상할 수 있음을 확인하였다.

결론

현장에서 유·가스 생산량 예측 시 일반적으로 활용되고 있는 Arps 방정식은 장기간의 천이유동이 나타나는 세일층에 적용할 경우, 생산량을 과소 혹은 과대 예측할 수 있어 결과에 불확실성이 존재한다. 이 연구에서는 생산 변동성이 큰 세일가스의 생산량을 예측하기 위해 RNN의 일종인 LSTM을 이용하여 예측모델을 구축하였으며, LSTM 모델이 여러 생산정을 바탕으로 다양한 생산감퇴를 학습함으로써 세일 저류층의 감퇴거동을 모사할 수 있음을 확인하였다. 그러나 LSTM의 특성상 초기 생산량에 관한 정보를 지속적으로 활용할 수 없다는 한계가 존재하며, 시간에 따라 변화하는 초기 생산량 대비 생산량의 감소분 및 감퇴경향을 반영하지 못하였다. 따라서 각 생산정의 생산이력으로부터 생산감퇴특성을 추출하여 입력자료로 활용하는 방식을 제안하였으며, Lee *et al.*(2022)이 제시한 연구방법을 활용하여 구축한 예측모델과 비교하였을 때에도 본 논문에서 제시한 생산감퇴특성을 이용한 예측모델이 상대적으로 개선된 예측결과를 도출할 수 있음을 파악하였다. 이와 같은 결과는 생산관련인자를 활용할 수 없는 제한적인 상황일 경우에도 생산이력으로부터 감퇴특성을 추출 및 가공함으로써 LSTM 모델을 활용한 예측이 가능하며, 생산감퇴특성이 예측모델의 학습 및 내부적인 연산 과정에서 개별 생산정의 특성을 반영하여 오차를 보정한 것으로 사료된다.

향후 이 연구에서 제안한 방식을 통해 생산성에 관련된 다양한 자료를 활용하여 LSTM 기반의 모델을 구축한다면 신뢰성 있는 생산량 예측이 가능할 것으로 사료되며, 다량의 자료가 축적되고 있는 현 상황에 있어 해당 연구결과를 유·가스 개발 산업을 위한 디지털 오일필드 기술으로써 효율적인 자산관리 및 개발 계획 수립에 활용할 수 있을 것이다.

사사

이 논문은 2021년도 산업통상자원부의 재원으로 해외자원개발협회의 지원(NO. 2021060002, 디지털 오일필드 전문인력 양성)과 한국에너지기술평가원의 지원을 받아 수행된 연구임(NO. 20216110100050, 소형발전 연계 중소규모 가스전 개발/생산 현장 운영기술 개발).

References

Alakeely, A. and Horne, R.N., 2020. Simulating the behavior of reservoirs with convolutional and recurrent neural networks, *SPE Reservoir Evaluation and Engineering*, 23(3), p.992-1005.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S., and Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures, *Electronics*, 8(3), p.1-66.

Arps, J.J., 1945. Analysis of decline curves, *Transactions of the AIME*, 160(1), p.228-247.

Botchkarev, A., 2019. A new typology design of performance metrics to measure errors in machine learning regression algorithms, *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, p.45-76.

Choi, Y., Yoon, D., Choi, J., and Byun, J., 2020. Hyperparameter search for facies classification with bayesian optimization, *Geophysics and Geophysical Exploration*, 23(3), p.157-167.

De Gooijer, J.G. and Hyndman, R.J., 2006. 25 Years of time

- series forecasting, *International Journal of Forecasting*, 22(3), p.443-473.
- Duong, A.N., 2011. Rate-decline analysis for fracture-dominated shale reservoirs, *SPE Reservoir Evaluation Engineering*, 22(3), p. 377-387.
- Engelder, T. and Lash, G.G., 2008. *Marcellus Shale play's Vast Resource Potential Creating Stir in Appalachia*, American Oil and Gas Reporter, 51(May), Kansas City, Missouri, USA, p.76-87.
- Gihm, Y.S., Hwang, I.G., Kim, H.T., Lee, H.S., and Lee, D.S., 2011. Geological characteristics and development strategy of the marcellus shale, *Journal of Korean Society for Geosystem Engineering*, 48(3), p.371-382.
- Graves, A., Mohamed, A., and Hinton, G., 2013. Speech recognition with deep recurrent neural networks, *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Vancouver, BC, Canada, p.6645-6649.
- Han, D.K., 2018. *Production Forecasting for Shale Gas Well in Transient Flow using Machine Learning Method*, Ph.D. Thesis, Dong-A University, Korea, 100p.
- Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory, *Neural Computation*, 9(8), p.1735-1780.
- Ilk, D., Rushing, J.A., Perego, A.D., and Blasingame, T.A., 2008. Exponential vs. Hyperbolic Decline in Tight Gas Sands-Understanding the Origin and Implications for Reserve Estimates Using Arp's Decline Curves, *SPE Annual Technical Conference and Exhibition*, SPE, Denver, Colorado, p.1-23.
- Jeon, B.K., Lee, K.H., and Kim, E.J., 2019. Development of a prediction model of solar irradiances using LSTM for use in building predictive control, *Journal of the Korean Solar Energy Society*, 39(5), p.41-52.
- Ji, M.S., Kwon, S.Y., Park, G.Y., Min, B.H., and Huy, N.X., 2021. Prediction of water saturation from well log data using deep learning algorithms, *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 58(3), p.215-226.
- Kanfar, M. and Wattenbarger, R., 2012. Comparison of Empirical Decline Curve Methods for Shale Wells, *Paper presented at the SPE Canadian Unconventional Resources Conference*, SPE, Calgary, Alberta, Canada, p.1-12.
- Ki, S.I., Seo, J.G., Kwon, O.K., and Jang, I.S., 2019. Prediction of missing tubing head pressure using recurrent neural network, *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 56(5), p.416-426.
- Kim, J.S., Shin, H.J., and Lim, J.S., 2014. Probabilistic decline curve analysis for forecasting estimated ultimate recovery in shale gas play, *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 51(6), p.808-819.
- Kocoglu, Y., Gorell, S., and McElroy, P., 2021. Application of Bayesian Optimized Deep Bi-LSTM Neural Networks for Production Forecasting of Gas Wells in Unconventional Shale Gas Reservoirs, *Unconventional Resources Technology Conference*, URTEc, Houston, Texas, USA, p.1-21.
- Lee, D.M., Shin, H.J., and Lim, J.S., 2022. Application of long short-term memory neural networks in shale gas production prediction using production-related factors, *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 59(6), p.673-683.
- Lee, K.B., Lim, J.T., Yoon, D.U., and Jung, H.S., 2019. Prediction of shale-gas production at duvernay formation using deep-learning algorithm, *SPE Journal*, 24(6), p.2423-2437.
- Li, X., Ma, X., Xiao, F., Wang, F., and Zhang, S., 2020. Application of gated recurrent unit (GRU) neural network for smart batch production prediction, *Energies*, 13(22), p.1-22.
- Li, X., Xiao, K., Li, X., Yu, C., Fan, D., and Sun, Z., 2022. A well rate prediction method based on LSTM algorithm considering manual operations, *Journal of Petroleum Science and Engineering*, 210, p.1-9.
- Luo, G. and Tian, Y., Bychina, M., and Ehlig-Economides, C., 2019. Production-strategy insights using machine learning: Application for bakken shale, *SPE Reservoir Evaluation and Engineering*, 22(3), p.800-816.
- Luo, S. and Su, H., 2022. Study on the production decline characteristics of shale oil: case study of jimusar field, *Frontiers in Energy Research*, 10, p.1-13.
- Ng, C.S.W., Jahanbani, G.A., and Nait, A.M., 2022. Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm, *Journal of Petroleum Science and Engineering*, 208, p.1-13.
- Ochella, S. and Shafiee, M., 2021. Performance Metrics for Artificial Intelligence (AI) Algorithms Adopted in Prognostics and Health Management (PHM) of Mechanical Systems, *Journal of Physics: Conference Series*, 2020 International Symposium on Automation, Information and Computing, Beijing, China, p.1-10.
- Oh, H., Ki, S., Park, C., and Jang, I., 2021. Analysis of uncertainty trend for estimated ultimate recovery prediction of shale gas with various production periods based on machine learning, *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 58(5), p.475-490.
- Paryani, M., Ahmadi, M., Awoleke, O., and Hanks, C., 2016. Using Improved Decline Curve Models for Production Forecasts in Unconventional Reservoirs, *SPE Eastern Regional Meeting*, SPE, Canton, Ohio, p.1-15.
- Shin, H.J., Lim, J.S., and Jang, I.S., 2021. Probabilistic prediction of multi-wells production based on production characteristics analysis using key factors in shale formations, *Energies*, 14, p.1-30.

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., and Cheng, Z., 2020. Time-series well performance prediction based on long short-term memory (LSTM) neural network model, *Journal of Petroleum Science and Engineering*, 186, p.1-11.

Tan, L., Zuo, L., and Wang, B., 2018. Methods of decline curve analysis for shale gas reservoirs, *Energies*, 11(3), p.1-18.

U.S. Energy Information Administration (EIA), 2022.06.07., <https://www.eia.gov/todayinenergy/detail.php?id=52198>.

Valko, P. and Lee, W.J., 2010. A Better Way to Forecast Production From Unconventional Gas Wells, *SPE Annual Technical Conference and Exhibition*, SPE, Florence, Italy, p.1-16.

Wang, S., Chen, Z., and Chen, S., 2019. Applicability of deep neural networks on production forecasting in Bakken shale reservoirs, *Journal of Petroleum Science and Engineering*, 179, p.112-125.



이 대 모

2023년 한국해양대학교 대학원 해양에너지자원공학과 공학석사

현재 중소조선연구원 경영혁신본부 전략기획실 연구원
(E-mail; dmlee@rims.re.kr)



신 효 진

2020년 한국해양대학교 대학원 해양에너지자원공학과 공학박사

현재 한국해양대학교 에너지자원공학과 박사후연구원
(E-mail; hjshin@kmou.ac.kr)



임 종 세

1998년 서울대학교 대학원 자원공학과 공학박사

현재 한국해양대학교 에너지자원공학과 교수
(E-mail; jslim@kmou.ac.kr)
