

2021년 지질자원 데이터 활용 및 인공지능 경진대회 최종 제안서

딥러닝기반 셰일가스 미래 생산량 예측

김하영¹⁾, 허승원²⁾

목 차

I. 서론 ----- 1

1. 연구개발 개요 ----- 1

 (1) 연구 배경 및 필요성 ----- 1

 (2) 연구 목적 ----- 2

II. 본론 ----- 3

1. 몬트니 셰일플레이 상세분석 ----- 3

 (1) 지질특성 분석 ----- 3

 (2) 부존특성 분석 ----- 7

2. 사용 데이터 ----- 10

 (1) 데이터 출처 ----- 10

3. 분석 방법 ----- 11

 (1) 기초통계분석 ----- 11

 (2) 생산자료 전처리 ----- 13

4. 분석 결과 ----- 17

 (1) 미래 생산량 예측을 위한 핵심인자 추출 ----- 17

5. 활용방안 ----- 23

 (1) 모델 학습 알고리즘 선정 ----- 23

 (2) 학습을 위한 데이터셋 분리 ----- 24

 (3) 케이스 선정 ----- 25

 (4) LSTM/Random forest 예측 결과 비교 ----- 31

 (5) 케이스 결과 비교 및 분석 ----- 31

 (6) 피크 이전 생산이력 제거 결과 분석 ----- 36

 (7) DCA기법과의 예측 결과 비교 ----- 38

 (8) 다른 셰일 플레이 적용 결과 ----- 39

III. 결론 ----- 41

IV. 참고문헌 ----- 42

1) 공주대학교/자연과학대학/지질환경과학과 (nicehy1023@gmail.com)

2) 공주대학교/자연과학대학/지질환경과학과 (hjsun1006@naver.com)

I. 서론

1. 연구개발 개요

(1) 연구 배경 및 필요성

북미지역은 현재 셰일가스 생산이 활발히 진행 중이다. 셰일가스개발에 있어 핵심 기술은 수평시추와 다단계수압파쇄이다. 기존의 전통자원인 근원암에서 형성된 탄화수소화합물이 트랩에 집적되어 수직시추만으로도 효율적인 생산이 가능하다. 하지만 비전통자원인 셰일가스는 비교적 넓은 지역에 분포하므로 생산성이 셰일가스 부존층을 통과하는 표면적을 넓히기 위해 수평시추를 수행한다. 또한, 셰일가스는 투수성이 0.001 md미만으로 매우 낮아 저류층에 자극을 준 후 투수성을 높이는 수압파쇄기법이 이용된다. 특히, 수압파쇄 시 stage 수를 늘려 생산성을 극대화하는 다단계수압파쇄기법이 사용된다.

수평시추와 다단계수압파쇄를 통해 개발이 어려웠던 셰일가스를 생산하기 시작했지만, 그와 동시에 기법들을 시행함에 있어 소비되는 비용 또한 고려해야 할 문제이다. 경제적으로 셰일가스를 생산하기 위해서는 효율적인 생산계획이 필수적이며, 이 과정에 있어 정확한 미래 생산량을 예측하는 것이 중요하다.

대표적인 생산예측 기법에는 저류층 시뮬레이션과 감퇴곡선기법(decline curve analysis, DCA)이 있다. 저류층 시뮬레이션은 효율적인 생산계획을 수립하는 데에 사용되는 방법이지만, 신뢰할 수 있는 정보를 얻기 위해서 필요로 하는 입력자료가 많고, 수압파쇄기법에 따른 균열망모델링에 불확실성이 존재한다(Alolayan et al., 2021). 생산이력만을 입력자료로 사용하는 DCA는 수백 개의 석유 생산정의 생산량 정보를 바탕으로 한 경험식이다. 감퇴지수 b값에 따라 세 가지의 대표적인 식이 있으며(표 1), 입력자료가 적어 적용이 간편하고 준수한 예측 성능으로 인해 현장에서 주로 사용되는 방법이다.

[표 1] Three types of Arp's equation

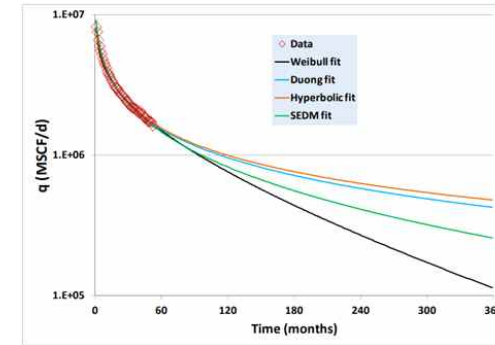
Case	b	equation
Exponential	b=0	$q_t = q_i \exp(-D_i t)$
Hyperbolic	0 < b < 1	$q_t = q_i / (1 + b D_i t)^{1/b}$
Harmonic	b=1	$q_t = q_i / (1 + D_i t)$

셰일 저류층의 경우 투수성이 매우 낮고, 수평시추 및 다단계수압파쇄와 같은 특징적인 생산방식으로 인해 기존의 전통자원과 다른 생산거동이 나타난다. 높은 초기 생산량을 보인 후, 빠르게 생산량이 감소하고 적은 생산량이 오랫동안 유지되는 셰일가스 생산양상 때문에 기존의 DCA기법을 셰일가스에 적용하는 것은 적합하지 않다(Yang and Lim, 2013). DCA기법 적용의 한계점을 극복하기 위해 비전통자원에 적합하게 변형된 여러 기법들이 개발되고 있지만, 그림 1에 나타나듯이, 동일한 과거 셰일가스 생산이력에도 기법별로 예측되는 미래 생산량의 값에 차이가 크게 나타난다(Mishra, 2012).

최근에는 앞서 언급한 저류층 시뮬레이션의 적용 한계와 DCA기법별 불확실성을 개선하기 위해 기계학습을 적용하려는 시도가 이루어지고 있다. 차원축소 알고리즘의 하나인 locality preserving projection을 적용한 후 k-nearest neighbor 알고리즘을 통해 생산량을 예측하는 연구가 진행되었고(Zhang and Hu, 2020), 적합한 클러스터링 변수 및 전이학습(transfer learning)을 통해 향상된

예측 결과가 제시되었다(Ashayeri and Jha, 2021). 전이학습을 통한 생산량 예측 결과가 DCA기법의 예측 결과보다 더 낮은 오차를 보였으며(Alolayan et al., 2021), 일시적으로 생산을 중단하는 첫인(shut-in)기간을 입력자료로 가용하여 DCA기법보다 향상된 예측 성능을 보였다(Lee et al., 2019). 하지만 기존 연구는 생산이력과 관련된 인자가 거의 고려되지 않았고, 소수의 인자만이 예측을 위해 사용되었다.

본 연구에서는 사용되는 기법에 따라 예측되는 값이 다른 DCA기법의 불확실성을 개선하기 위해, 과거 생산이력 데이터를 활용하여 딥러닝을 기반으로 한 셰일가스 미래 생산량을 예측하는 기술을 제안하고자 한다.



[그림 1] Uncertainty of DCA(Mishra, 2012).

(2) 연구 목적

본 연구에서는 동적자료인 과거 생산이력 데이터를 활용하여 유정의 미래 생산량을 예측하는 기술을 제안하고자 하며, 해당 기술을 통해 기존 DCA기법의 적용 한계 및 여러 기법에 따라 예측 결과 편차가 큰 단점을 보완할 수 있을 것으로 예상된다. 또한 DCA기법은 가용되는 입력인자의 종류가 제한적인 이유로 생산정 운영인자가 고려되지 않아 곡선 형태의 결과가 도출되어 생산이 감퇴하는 경향만을 파악할 수 있다. 이러한 DCA기법의 한계를 개선하기 위해 본 연구에서는 생산량 자료뿐만 아니라 생산시간과 같은 생산정 운영인자를 같이 고려하여 미래 생산량을 예측하고자 한다.

또한, 캐나다 몬트리 셰일플레이 생산정 자료를 연구의 목적에 맞게 전처리한 후, 핵심인자 추출로 미래 생산량 예측 결과의 신뢰도를 향상시키고자 한다. 전처리 과정은 모델의 성능을 좌우하는 매우 중요한 과정으로, 전처리 과정이 제대로 이루어지지 않는다면 원하는 수준의 결과를 기대하기 어렵다. 본 연구에서는 몬트리 셰일플레이를 대상으로 지질 및 부존특성 상세분석을 실시한 후, 생산자료 기초통계분석 및 도메인 지식(domain knowledge)을 통한 자료 전처리 과정을 수행하고자 한다. 기계학습 진행 시 핵심인자를 추출하는 과정 또한 모델의 신뢰도를 향상시키는 데 있어 중요한 단계이다. 출력층 인자와의 상관관계가 거의 나타나지 않는 입력층 인자 또는 불필요하게 중복되는 인자를 가용할 경우, 기계학습을 진행하는데 있어 불확실성을 높이는 요소가 될 수 있다. 본 연구에서는 각 인자 간의 상관관계 분석을 통해 미래 생산량 예측을 위한 핵심인자를 추출하고자 하며, 추출된 핵심인자를 바탕으로 상대오차 약 15% 이내의 정밀도 높은 예측 모델을 도출하고자 한다.

II. 본론

1. 몬트니 셰일플레이 상세분석

캐나다의 오일과 가스생산은 크게 동쪽과 서쪽으로 구분된다. 동쪽은 유티카, 호턴 블러프 등의 플레이가 위치하고 서쪽은 혼 리버, 몬트니, 듀버니 등의 플레이가 위치한다. 서쪽에 위치한 듀버니를 대상으로 머신러닝 알고리즘을 이용한 생산량 예측 연구가 진행된 바 있다(Lee et al., 2019). 본 연구 지역인 몬트니 셰일플레이는 듀버니의 서남부에 위치하며, 셰일층의 깊이 변화로 인해 동부에서는 셰일오일, 서부에서는 셰일가스가 주로 생산된다고 알려져 있다(Vishkai et al., 2017). 연구 진행에 앞서, 몬트니 셰일플레이의 지질 및 부존특성 분석을 수행하여 사전 도메인 지식 확보를 위한 기반으로 활용하고자 한다.

(1) 지질특성 분석

① 일반환경

몬트니 셰일층은 그림 2와 같이 캐나다 브리티시컬럼비아주 북동부에서 앨버타주 서부까지 분포하고 있으며 면적은 약 140,000 km²로 한반도 면적의 약 64%에 해당하는 넓은 범위이다. 북서-남동으로 분포하는 몬트니층은 암상을 기준으로 상부, 하부로 구분된다. 상부 몬트니층은 지표면으로부터 실수직거리 2,800~3,500m이며 두께는 평균 200m로 동부에서 서부로 갈수록 점점 두꺼워진다. 생성시기는 중생대 트라이아스기 초기 및 중기이고 당시 현 북미판의 북미지괴 서부 가장자리에 퇴적되었다. 몬트니층은 비활성 대륙주변부에서 퇴적되었으며, 건조한 계절풍의 영향을 받은 것으로 나타난다.

하부 몬트니층은 외해 환경 또는 외대륙붕에서, 상부 몬트니층은 내해 연안환경 또는 내대륙붕에서 각각 퇴적되었다(그림 2). 현재 상부는 동쪽, 하부는 서쪽에 위치한다.



[그림 2] The location of the Montney formation in Canada(Modified from Yoon and Jo, 2019).

② 암상조성분석

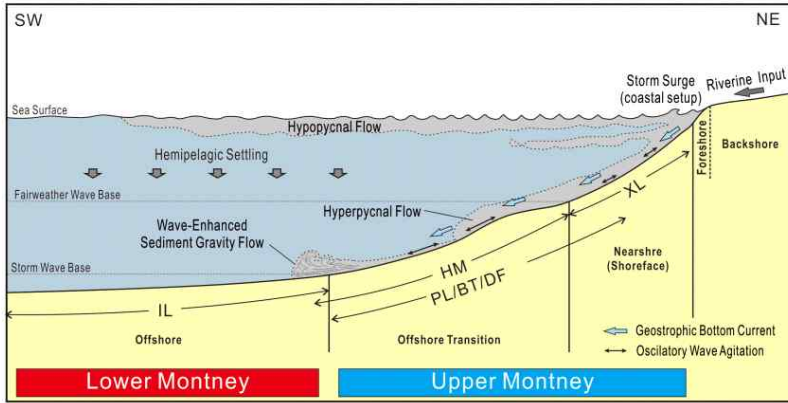
암상분석 결과 동쪽은 연안면 사암(shoreface sandstones), 백운암질 패각암(coquinas), 하부 연안면 실트암(lower shoreface siltstones)으로 구성되며, 서쪽은 storm dominated silt rock, 저탁류 실트암 혹은 사암이 교호하는 것으로 나타남에 따라 동쪽이 상부, 서쪽이 하부인 것으로 판단된다. 하부 몬트니층의 광물 조성은 석영 28~52%(평균 45%), 장석류 6~15%(평균 12%)를 보이며 쇄설성 실트 입자로 나타난다. 또한, 백운석 8~42%(평균 15%), 황철석·능철석·백철석 함량은 3~9%(평균 5%)를 보인다. 또한 층상 규산염 광물은 11~44%(평균 26%)를 보이며 백운석은 교결물질과 이질 기질(Muddy Matrix)의 역할을 하고 층상 규산염 광물은 기질의 역할을 한다. 결과적으로 점토질 백운석 실트암으로 구성되며 드물게 점토질 실트암이나 실트질 이암도 나타난다. 이러한 암석들의 입자는 분급이 양호하고 조립질 혹은 중립질의 실트사이즈를 가진다. 또한 원마도가 높게 나타나며 total organic carbon(TOC)은 0.18~1.32%(평균 0.68%)로 나타난다.

상부 몬트니층의 광물 조성은 석영 21~62%(평균 49%), 장석류 9~19%(평균 14%), 황철석·능철석·백철석 2~11%(평균 3%), 백운석 8~33%(평균 19%), 층상 규산염 광물 4~50%(평균 15%)로 이루어져 있다. 하부 몬트니층과 조성은 비슷하지만, 원마도가 하부에 비해 불량하며 약점토질 백운석 실트암으로 구성되어있다. TOC 또한 0.06~1.28%(평균 0.26%)로 낮게 나타난다.

③ 퇴적환경

몬트니층의 퇴적환경은 그림 3과 같이 대륙붕과 대륙사면 혹은 완만한 램프지형에서 퇴적된 것으로 밝혀졌다(Yoon and Jo, 2019). 외해로 갈수록 일반적으로 퇴적물에 미치는 에너지의 영향이 미약하므로 입자크기가 작은 셰일 등이 퇴적된다. 하부 몬트니층은 외해에서 퇴적되어 연안지역에서 나타나는 파도의 흔적(연흔) 등이 나타나지 않는다. 사층리 또한 나타나지 않고 세립질 퇴적상을 보이며, 대다수 외해에서 퇴적이 일어난 층에서는 점토 및 유기물의 함량이 높게 나타난다.

이와 반대로 상부 몬트니층에서는 내해에서 퇴적된 특징이 강하게 나타난다. 내해에서 발생하는 파도에 의한 침식과 변형 및 생물에 의해 생성되는 생고란 작용이 나타난다. 이런 퇴적상의 변화를 보았을 때, 하부가 퇴적된 후 상부가 퇴적되기까지 해수면의 하강 혹은 대륙지각의 융기가 일어난 것을 알 수 있다. 내해 연안환경에서도 파도와 폭풍 등에 의해 이암이 퇴적될 수 있으며, 이 때 파도와 조류 등의 영향으로 인한 퇴적구조가 나타난다. 때문에 상부 몬트니층은 백운암질 실트암으로 구성되어있다. 전체적인 몬트니층은 여러 차례에 걸친 해수면의 변동으로 인하여 실트암과 셰일이 교호하며 퇴적되었으며 이암부터 사암까지 다양한 종류가 나타난다.



[그림 3] Schematic diagram showing depositional environments and sedimentary processes of the Montney formation(Yoon and Jo, 2019).

④ 퇴적상

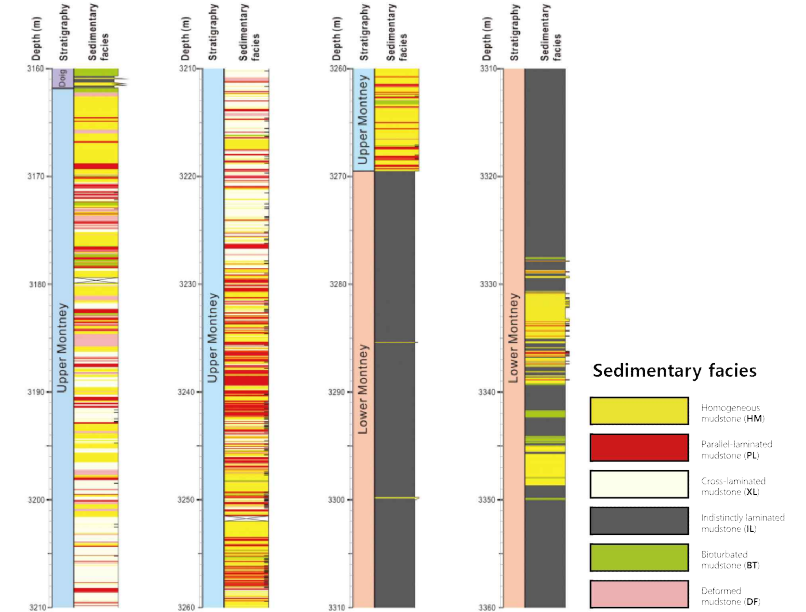
퇴적상이란 퇴적물의 특징 중 하나로 베커니즘과 생성 당시 퇴적환경에 대한 정보를 담고 있다. 몬트니층의 셰일에서는 입자의 조직특성 및 퇴적구조에 따라 6가지의 퇴적상이 나타난다. 주로 석영과 장석으로 이루어진 규산쇄설암이며 실트암 혹은 이암이 엽층리, 사엽층리, 평행엽층리, 무구조, 생교란작용, 가소성 변형 구조 등을 나타낸다(표 2).

[표 2] Sedimentary facies of Lower Montney formation and Upper Montney formation(Modified from Yoon and Jo, 2019)

Sedimentary facies	Sedimentary processes	Depositional environments	Characteristic
균질 이암, Homogeneous mudstone(HM)	세립질 저탁류의 최후기 단계	Offshore	대체로 층리나 엽리가 관찰이 안됨. 희미하게 엽층리 및 점이층리가 관찰됨(상부, 하부)
희미한 엽층리 이암, Indistinctly laminated mudstone(IL)	조립질 퇴적물이 공급되지 않는 에너지가 안정적인 퇴적환경	Offshore or deeper marine setting below storm wave base	이암 혹은 셰일의 조직특성을 보임. 내부적으로 점이층리나 환경사의 사층리 발달(하부)
사엽층리 이암, Cross-laminated mudstone(XL)	파도에 의한 침식-퇴적과정	Nearshore above fair-weather wave base	소규모 사층리(경사져 발달하는 사층리 포함, 상부)
평행 엽층리 이암, Parallel-laminated mudstone(PL)	고농도의 퇴적물 부유이동과정	Nearshore to offshore above storm wave base	담회색 및 암회색, 1 mm 이하의 엽층리 발달. 부분적으로 사엽층리 발달(상부)
생교란 이암, Bioturbated mudstone(BT)	해저 저서생물의 섭식활동 및 주거구조	Oxic nearshore to offshore	1 cm이하 생물 활동 및 서식구조에 의해 원래 조직이나 구조가 교란되어 나타남(상부)
변형 이암, Deformed mudstone(DF)	가소성 변형	Nearshore to offshore above wave base	층경계면 혹은 엽층리가 구겨져 접히거나 말림(상부)

외해 환경에서 퇴적된 하부 몬트니층은 낮은 에너지에서 퇴적이 진행되었다. 그로 인하여 점토가 우세하게 퇴적이 되었다. 하부 몬트니층을 주로 이루고 있는 이암의 전체적인 색은 암회색을 띄며, 석영질과 실트질이 희미하게 엽층을 이루고 있지만 불연속적으로 나타난다. 또한 점이층리와 낮은 경사의 사층리를 보이기도 하며 불규칙하게 나타난다. 이러한 특징을 가지는 이암을 희미한 엽층리 이암(indistinctly laminated mudstone, IL)이라고 한다. 하부 몬트니층의 주된 퇴적상

은 IL로 나타나고, 부분적으로 IL과 비슷한 특징을 가지는 균질 이암(homogeneous mudstone, HM)도 나타난다. 상부 몬트니층의 전반적인 조성은 하부와 유사하지만 구조적인 특징에서 차이가 있다. 상대적으로 내해 환경에서 퇴적되어 다양한 엽층이 나타나며 그에 따라 다양한 퇴적상이 발달하였다. 주로 나타나는 퇴적상은 세 가지로 균질 이암, 평행 엽층리 이암(parallel laminated mudstone, PL), 사엽층리 이암(cross-laminated mudstone, XL)이다(그림 4).



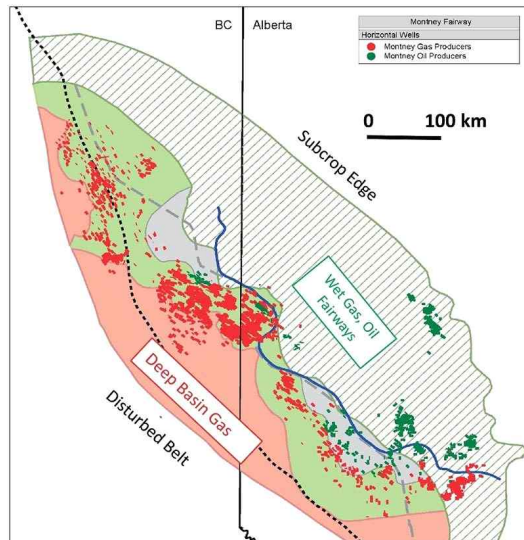
[그림 4] Summarized borehole plots of stratigraphy and sedimentary facies(Modified from Yoon and Jo, 2019).

부분적으로 층리가 나타나며 HM의 두께는 수~30 cm 이상으로 나타난다. 또한 상하부 층과의 경계면이 매우 뚜렷하며 상부에 존재하는 HM에서는 부분적으로 생교란작용이 나타난다. 이러한 HM 퇴적상이 형성되려면 갑작스럽게 퇴적이 일어나야 하며, 그 원인은 해저에 발달하는 세립질 저탁류에 의하여 고농도의 퇴적물이 이동하다 빠르게 가라앉아서 생성된다. 평행 엽층리 이암은 암회색 및 담회색을 띄며 사엽층리가 나타나고 두께는 약 1 mm 이하로 나타난다. 엽층리는 완만한 경사인 언덕 사엽층리(hummocky lamination)가 나타나며 유기물과 규질입자가 서로 교호하며 퇴적되어 형성되었다. 이러한 사엽층리는 해저에 가라앉은 퇴적물을 띄우는 폭풍파와 폭풍에 의해 발생되는 해저 지형류의 영향으로 생성된다. 사엽층리 이암은 PL과 유사하게 규질과 점토가 교호하며 생성되고 담회색인 엽층이 나타난다. 또한 사층리층 내의 퇴적면에 경사져 생성된 사엽층리가 포함된다. 이러한 사층리의 두께는 매우 얇으며 1~2 cm 해류의 흐름 혹은 파도에 의해 형성되는 연흔이 이동을 하며 형성된다.

(2) 부존특성 분석

① 일반현황

몬트니는 1950년대부터 자원탐사의 대상이 되기 시작했다. 몬트니의 사암과 백운암에서 자원을 개발하였지만, 실트암에 있는 비전통자원은 2005년까지 개발이 진행되지 않았다. 기존의 전통자원과는 달리 비전통자원은 치밀한 미세 공극 내에 존재하며, 치밀가스화 세일가스의 유체투과율이 각각 0.1 md, 0.001 md 수준으로 매우 낮은 자원을 개발하기 어렵기 때문에 다른 추가적인 기술이 요구된다. 비전통자원은 전통자원에 비해 넓은 분포를 가져 기존의 수직시추가 아닌 수평시추를 적용하였다. 또한, 유체투과율이 매우 낮은 점을 고려하여 저류층에 인공적인 균열을 발생시킨 후 저류층의 유체투과율을 높여 생산성을 향상시키는 방법인 수압파쇄기술이 사용되었다. 기존의 수평시추, 수압파쇄기술과 다단계수압파쇄기술을 함께 적용함으로써 세일혁명이 시작되었고 몬트니의 실트암과 같은 비전통자원에 대한 개발이 이루어지게 되었다. 몬트니층의 탄화수소자원은 두께와 심도에 따라 다양하게 나타나는데, 심도가 깊은 서쪽으로 갈수록 세일가스가 주로 생산되며, 동쪽으로 갈수록 세일오일이 생산된다(그림 5).



[그림 5] Montney play fairway, straddling the British Columbia/Alberta border(THEWAYAHEAD, 2018).

② 부존량

몬트니의 비전통가스자원은 브리티시컬럼비아주와 앨버타주에 고루 분포되어 있다. 세일가스의 기술적으로 회수 가능한 자원(technically recoverable resource, TRR)의 양은 약 49 Tcf로 많은 편에 속하지 않지만, 총 비전통가스자원의 TRR은 약 449 Tcf로 2012년 캐나다 천연가스 사용량의 약 160배에 해당한다(표 3).

몬트니는 비전통가스자원 뿐만 아니라 비전통오일자원과 NGL자원의 추정되는 TRR 또한 각각 1,125 MBO, 14,521 MBO로 많은 양에 해당한다(표 3). 몬트니의 비전통자원을 주별로 구분한 표

4와 표 5를 보면 비전통가스자원의 경우 브리티시컬럼비아주와 앨버타주에서 추정되는 TRR이 각각 271 Tcf, 178 Tcf로 큰 차이를 보이지는 않는다. 하지만 NGLs와 비전통오일자원은 주별 차이가 크다. NGLs의 경우 브리티시컬럼비아주에서 추정되는 TRR은 12,647 MMb이지만, 앨버타주는 1,874 MMb로 브리티시컬럼비아주에서 더 많은 양의 NGLs가 추정되는 반면, 비전통오일자원은 브리티시컬럼비아주에서 29 MBO, 앨버타주에서 1,096 MBO로 앨버타주에서 더 많은 양이 추정된다.

비전통자원에 대한 브리티시컬럼비아주와 앨버타주의 회수율을 비교해본 결과 전반적으로 앨버타주의 자원 회수율이 더 적은 것으로 나타나는데, 이는 앨버타주에서 추정되는 자원은 대부분 개발의 불확실성이 큰 shallower areas에 존재하기 때문이다(National Energy Board et al., 2013). 브리티시컬럼비아주에 비해 앨버타주의 NGLs 및 비전통오일자원 평가 시 부존량과 TRR의 불확실성이 크고 회수율도 작음을 알 수 있다.

[표 3] Ultimate potential for Montney unconventional petroleum in British Columbia and Alberta(Modified from National Energy Board. et al, 2013)

Hydrocarbon Type	In-Place			TRR			Recovery
	Low	Expected	High	Low	Expected	High	
Natural Gas - billion m ³ (trillion cubic feet)	90,559 (3,197)	121,080 (4,274)	153,103 (5,405)	8,952 (316)	12,719 (449)	18,257 (645)	10.5%
NGLs - million m ³ (million barrels)	13,884 (87,360)	20,173 (126,931)	28,096 (176,783)	1,540 (9,689)	2,308 (14,521)	3,344 (21,040)	11.4%
Oil - million m ³ (million barrels)	12,865 (80,949)	22,484 (141,469)	36,113 (227,221)	72 (452)	179 (1,125)	386 (2,430)	0.8%

[표 4] Ultimate potential for Montney unconventional petroleum in British Columbia(Modified from National Energy Board. et al, 2013)

Hydrocarbon Type	In-Place			TRR			Recovery
	Low	Expected	High	Low	Expected	High	
Natural Gas - billion m ³ (trillion cubic feet)	42,4326 (1,498)	55,664 (1,965)	69,630 (2,458)	5,666 (200)	7,677 (271)	10,311 (364)	13.8%
NGLs - million m ³ (million barrels)	11,974 (75,340)	15,310 (96,332)	19,172 (120,633)	1,418 (8,920)	2,010 (12,647)	2,760 (17,366)	13.1%
Oil - million m ³ (million barrels)	211 (1,328)	439 (2,763)	739 (4,652)	1 (8)	5 (29)	11 (70)	1.0%

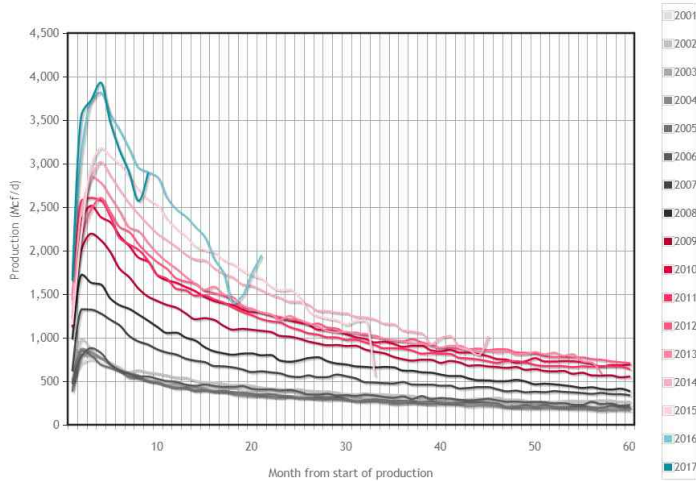
[표 5] Ultimate potential for Montney, including lowermost Doig siltstone, unconventional petroleum in Alberta(Modified from National Energy Board. et al, 2013)

Hydrocarbon Type	In-Place			TRR			Recovery
	Low	Expected	High	Low	Expected	High	
Natural Gas - billion m ³ (trillion cubic feet)	48,124 (1,699)	65,415 (2,309)	83,474 (2,947)	3,286 (116)	5,042 (178)	7,946 (281)	7.7%
NGLs - million m ³ (million barrels)	1,910 (12,020)	4,863 (30,599)	8,924 (56,150)	122 (769)	298 (1,874)	584 (3,674)	6.1%
Oil - million m ³ (million barrels)	12,654 (79,621)	22,045 (138,706)	35,373 (222,569)	71 (444)	174 (1,096)	375 (2,360)	0.8%

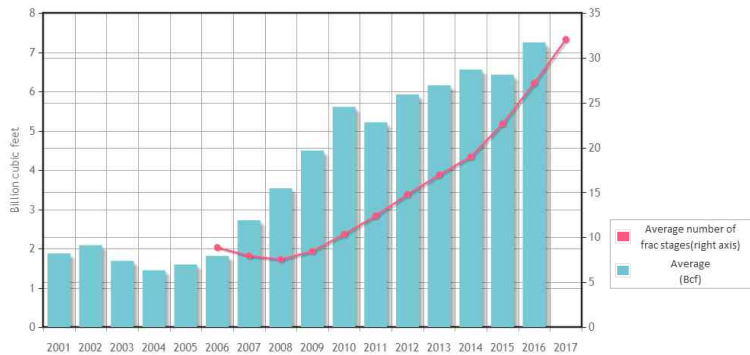
③ 생산특성

현재 몬트니의 탄화수소자원 중 콘덴세이트는 거의 생산이 되지 않는 상태이며, 가스과 오일의 초기 생산량은 각각 3~4 MMcfd, 575 BOPD이다. 세일가스 탐사 및 개발 경험이 축적되면서 sweet spot 탐사기술이 고도화되고 다단계수압파쇄 디자인이 최적화되어 최근에 시추된 유정일수록 이전에 시추된 유정보다 더 나은 생산 퍼포먼스를 보인다(그림 6). 2005년에 시추된 유정들의 경우 평균 0.9 MMcfd의 생산량을 보이지만 2017년에 시추된 유정들은 평균 4 MMcfd로 약 4배 증가하였다.

최근에 시추된 유정일수록 궁극가채매장량(estimated ultimate recovery, EUR) 또한 증가하는 모습을 보인다. 2005년에 시추된 유정은 1.8 Bcf, 2016년에 시추된 유정은 7.2 Bcf로 EUR이 약 4배 증가하였고, 수압파쇄 시의 stage 수 또한 평균 8개에서 30개 이상으로 증가하였다(그림 7).



[그림 6] Montney Formation Gas Well Performance by Year(Canada Energy Regulator, 2018).



[그림 7] Average well EUR and number of fracture stages(Canada Energy Regulator, 2018).

2. 사용 데이터

(1) 데이터 출처

본 연구에 사용된 데이터는 지오빅데이터 오픈플랫폼에 업로드 되어있는 서캐나다 몬트니 세일 가스 월간 생산량 자료이다. 해당 데이터는 몬트니 5,408개 유정의 기본적인 정보를 포함하여, 수압파쇄에 관한 정보 및 각 유정별 생산되는 유체의 종류와 과거 생산이력이 나타나있다. 생산자료 기초통계분석에 앞서 총 5,408개의 유정 데이터에서 연구 목적에 부합하는 유정을 선별하는 과정을 진행하였다.

① 생산되는 유체에 따른 유정 선별

총 5,408개 유정의 266,564개 데이터에서 값이 존재하지 않는 5,407개의 결측치를 제거한 뒤, 5,408개 유정의 261,157개 데이터를 바탕으로 연구를 진행하였다. 서로 다른 유정에서 생산되는 유체의 종류가 다를 시, 압력 등 여러 요인들로 인해 유체의 생산거동에 차이가 발생할 수 있다. 즉, 세일가스만 생산되는 유정과 세일가스과 세일오일이 생산되는 유정을 비교하면, 동일 종류의 유체일지라도 생산 감퇴경향이 각각 다르게 나타날 수 있다. 본 연구에서는 세일가스와 물이 생산되는 유정으로 생산 유체의 종류를 한정하여, 5,408개 유정에서 세일가스와 물이 생산되는 4,830개의 유정 중 오일과 콘덴세이트가 생산되는 3,724개 유정을 제외한 1,106개 유정을 선별하였다.

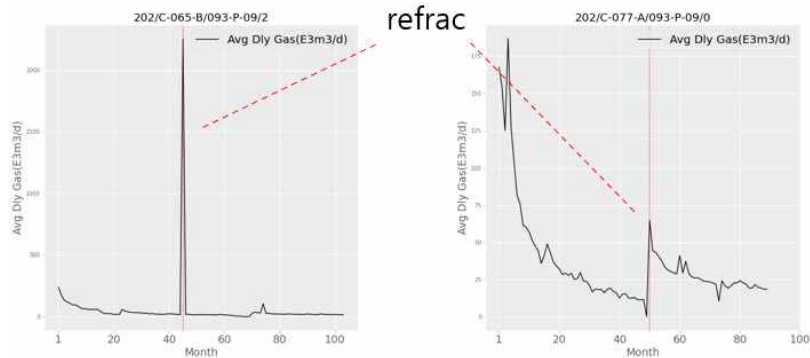
② 분지에 따른 유정 선별

이전 선별 과정을 거친 1,106개 유정의 58,170개 데이터를 대상으로 분지에 따른 유정을 선별하였다. 1,106개의 유정 중 서로 다른 분지에 해당하는 유정의 수는 각각 Alberta Syncline은 25개, Athabasca Oil Sand Zone은 0개, Peace River Embayment는 1,007개, Northern Alberta Sub-basin은 74개로 확인되었다. 세일가스는 각 분지 또는 지역 간 지질학적 특성의 차이로 인해 생산거동이 다르게 나타날 수 있으므로, 본 연구에서는 대상 분지를 가장 많은 1,007개의 유정이 존재하는 Peace River Embayment로 선정하였다.

③ 제수압파쇄 유정 제외

제수압파쇄(refracturing)은 생산 초기 수압파쇄를 진행한 후, 시간이 지남에 따라 감소하는 생산성을 다시 회복시키기 위해 사용하는 기법이다. 제수압파쇄를 수행할 시, 단혔던 균열이 확장되거나 새로운 균열이 만들어짐으로써 유정의 생산성이 급격히 증가하게 된다(그림 8). 그림 8은 1,007개의 유정 중 제수압파쇄를 수행한 이력이 있는 2개 유정의 생산거동을 나타낸 그래프이다. 두 개의 유정 모두 초기 생산을 시작한 후, 유정의 생산성이 감소하지만 제수압파쇄(빨간색 점선)를 한 이후 생산성이 급증한 것을 확인할 수 있다. 이러한 제수압파쇄 후의 특징적인 생산이력은 중요한 데이터로 가용될 수 있지만, 1,007개의 유정 중 제수압파쇄를 한 유정은 총 2개로 학습되는 과정에 있어 불충분한 자료 수라고 판단하여 제외하였다.

위의 유정 선별 과정을 거쳐, 1,005개 유정의 55,031개의 데이터를 본 연구에 활용하였다.



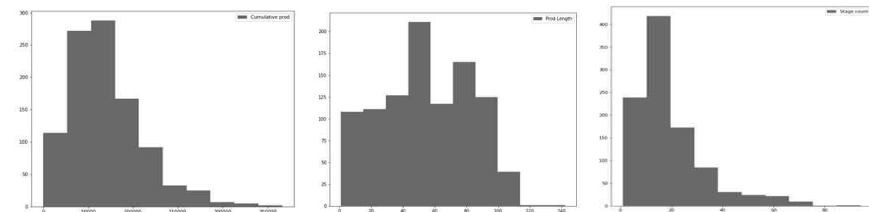
[그림 8] Effects of refracturing.

3. 분석 방법

(1) 기초통계분석

유정 선별 과정을 거친 1,005개 유정을 대상으로 기초적인 통계분석을 수행하였다. 세일가스생산을 시작한 연도는 2006~2018년으로 다양하며 유정별 누적생산량, 생산기간 및 수압파쇄 정보를 히스토그램을 통해 분석하였다. 그림 9a는 유정별 누적생산량으로 34~266,212 E3m³의 범위를 가지며, 평균 누적 생산량은 70,280 E3m³이다. 생산시간 분포를 나타낸 그림 9b의 x축은 생산이력의 길이, y축은 유정 개수로 최소 1개월에서 최대 142개월의 생산이력을 가지며, 평균 생산이력의 길이는 약 55개월이다. 유정 생산이력의 길이는 전체적으로 고른 분포를 보이며, 생산이력의 길이가 100개월 이상인 유정은 적다. 그림 9c는 유정의 수압파쇄 정보 중 stage 개수에 대한 분석 자료로, 평균 stage 수는 20개로 다단계수압파쇄가 이뤄졌지만, 일부 stage의 수가 1개, 즉, 단일 수압파쇄 유정도 존재한다.

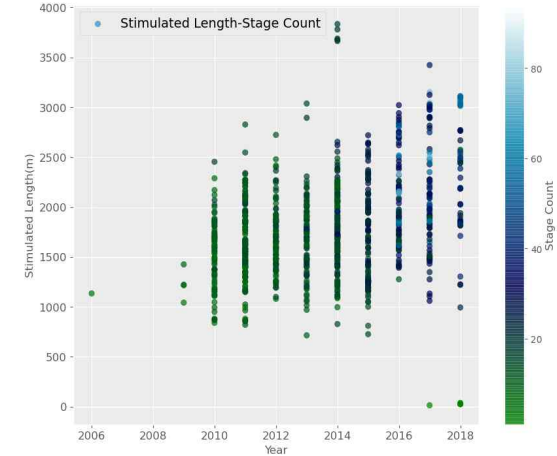
(a) Cumulative production (b) Production length (c) Stage count



	Count	Mean	Std	Min	25%	50%	75%	Max	단위
Cumulative prod	1005	70279.9	40858.4	34	41635.3	64410.8	90917.1	266211.7	E3m ³
Prod length	1005	54.7	27.8	1	32	53	79	142	Month
Stage count	1005	19.8	13.4	1	11	15	25	94	개

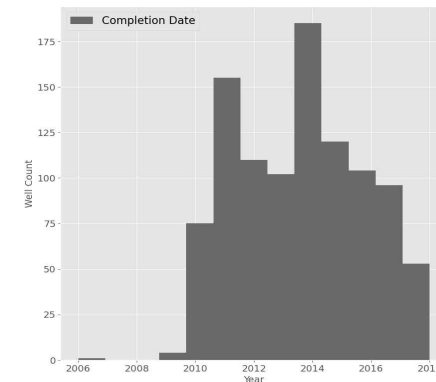
[그림 9] Statistical Analysis of 1,005 wells in Peace River Embayment basin: (a) Cumulative production, (b) Production length and (c) Stage count.

그림 10은 각 유정의 완결시점(completion date)에 따른 자극길이(stimulated length)의 변화 추세를 보여주며 색은 stage 개수를 의미한다. 2015년 이후 완결된 유정들은 stage 개수가 많이 증가하는 경향을 보이며 이에 따라 자극길이도 전체적으로 증가한다.



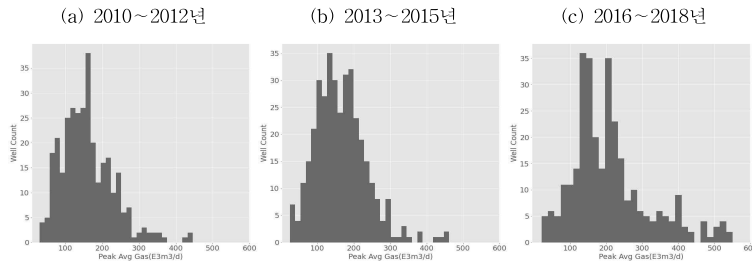
[그림 10] Scatter plot of stimulated length and stage count by year.

20개 내외의 stage 수를 설정하다 최근에는 80개 이상의 stage를 설정하는 경우도 보고되고 있다. 그림 11은 유정마다 완결연도의 분포를 파악한 히스토그램이다. 2006년부터 2010년까지의 유정은 5개로, 유정의 완결연도는 대부분 2010년 이상인 것을 확인하였다. 이때, 2010년 이전의 유정들은 개수가 적고 이후의 유정들과 completion design의 차이가 커 제외하고 분석을 진행하였다.



[그림 11] Histogram of well count of completion date by year.

완결연도에 따른 생산량의 변화를 알아보기 위해 2010~2012년, 2013~2015년, 2016~2018년의 세 구간으로 나눠 분석을 진행하였다(그림 12). x축의 생산량은 유정의 생산능력 중 Avg Dly Gas가 가장 높은 Peak Avg Gas 생산량을 사용하였다. 세 구간의 평균 Peak Avg Gas의 양은 각각 164.4 E3m³/d, 165.1 E3m³/d, 210.1 E3m³/d이다. 그림 12a구간에서 그림 12c구간으로 갈수록 유정의 평균 Peak Avg Gas가 증가하는 경향을 볼 수 있으며, 특히, 그림 12b구간에서 그림 12c구간으로 가는 2016년부터 값이 크게 증가한 것을 확인할 수 있다. 2016년 이후 stage 개수가 급격히 늘어남에 따라 자극깊이가 증가한 것이 원인으로 보인다. 분석 결과 연도에 따른 수압파쇄 디자인 차이와 그에 따른 유정 생산량 차이를 확인하였다.

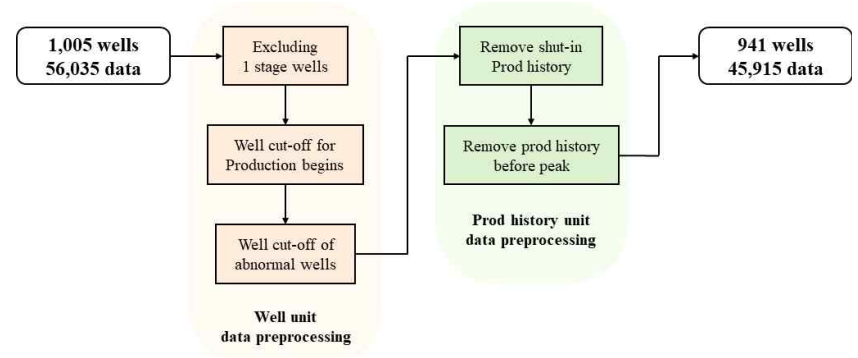


	Count	Mean	Std	Min	25%	50%	75%	Max	단위
2010~2012	324	164.4	78.5	29.9	110.7	152.8	200.3	768.8	E3m ³ /d
2013~2015	367	165.1	71.4	24.4	113.4	160	202.8	461.7	E3m ³ /d
2016~2018	307	210.1	108	20.5	139.3	193.5	247.3	550.4	E3m ³ /d

[그림 12] Analysis of Peak Avg Gas production for the three sections: (a) 2010~2012, (b) 2013~2015 and (c) 2016~2018.

(2) 생산자료 전처리

전처리 과정은 가공되지 않은 원시자료(raw data)를 모델의 목적에 적합하게 데이터를 재구성하는 것이다. 이러한 전처리 과정은 모델의 성능을 결정하는 데에 매우 중요한 과정으로, 전처리 과정이 제대로 이루어지지 않는다면 원하는 수준의 분석 결과를 기대하기 어렵다. 본 연구의 목적에 맞는 모델을 만들기 위해 기초통계분석 및 도메인 지식을 바탕으로 모델 훈련에 방해될 수 있는 자료를 유정 단위, 상세생산능력 단위로 구분하여 전처리해주었고, 이후 전처리된 데이터를 시계열로 분리하는 과정을 수행하였다. 그림 13에 생산자료 전처리 과정의 전체적인 작업 흐름을 도시하였다.

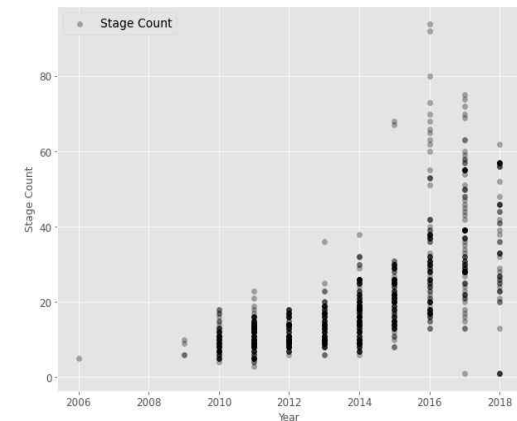


[그림 13] Workflow of data preprocessing.

① 유정 단위 전처리

- 단일스태지 유정 전처리(1,005개→1,000개)

세일가스는 기존의 전통자원과 달리 투수성이 매우 낮은 특성을 가지고 있다. 이러한 이유로 액체를 주입하여 투수성을 높은 인공균열을 생성한 뒤 가스를 생산하는 수압파쇄기법을 사용한다. 그림 14와 같이 stage 개수는 연도가 증가할수록 점점 증가하며, 2016년에는 증가하는 폭이 매우 큰 것을 알 수 있다. 하지만, 2017년과 2018년의 5개 유정은 다단계수압파쇄가 아닌, stage 개수가 1인 단일수압파쇄를 수행하여 해당 유정들을 훈련자료에서 제외하였다.



[그림 14] Scatter plot of stage count by year.

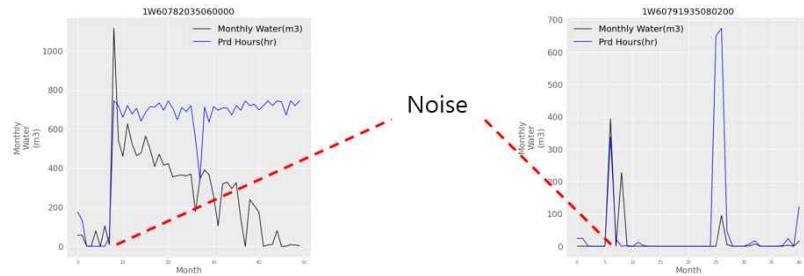
- 완결연도에 따른 전처리(1,000개→995개)

연도에 따른 수압파쇄 디자인의 차이로 인해 유정의 생산경향도 달라진다. 그림 12에서 제외했던 2006~2010년 구간에 속한 유정들은 그림 14에서 보듯이 stage 수가 상대적으로 적어 평균

Peak Avg Gas양이 약 90 E3m³/d정도로 다른 구간에 비해 많이 낮은 수준이다. 모델을 학습시킬 시, 연도 차이에 따른 생산량 차이가 커 딥러닝모델 학습에 혼동을 일으킬 수 있으므로 2006~2010년 구간의 5개 유정을 제외하였다.

- 노이즈 유정 전처리(995개→942개)

정상적인 생산정은 생산시간이 0일 경우 생산량 또한 0이 되어야 한다. 그림 15는 유정의 생산 시간(파란색)과 월간 물 생산량(검은색)을 나타낸 그래프이다. 두 유정 모두 생산시간은 0이지만 약 100m³, 200m³의 물이 생산된 이력을 볼 수 있다. 이러한 유정처럼 생산시간은 0이지만 생산량이 0이 아닌 이력을 가진 유정을 노이즈 유정이라고 판단하여, 총 53개의 유정을 제외해주었다.

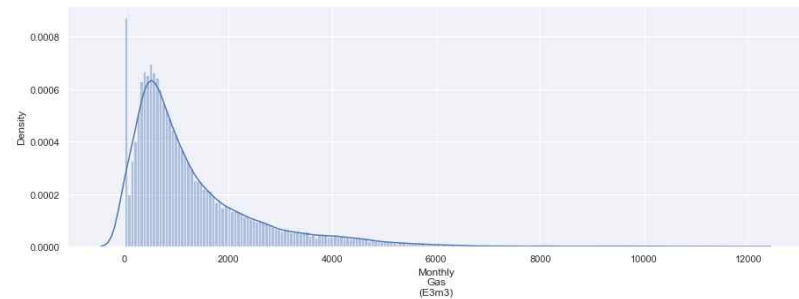


[그림 15] Production hours and monthly water production rate for abnormal wells.

② 상세생산이력 단위 전처리

- 첫인이력 전처리

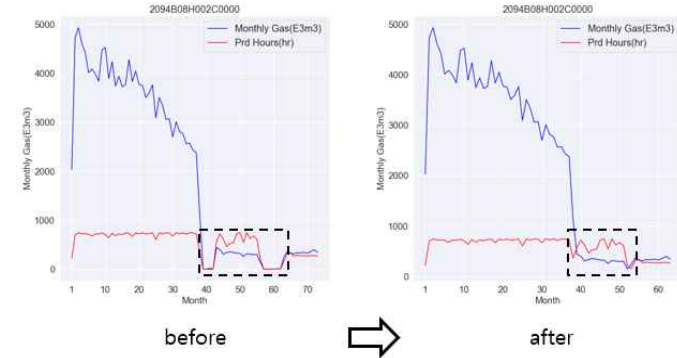
유정의 생산이력 중 여러 사유로 인해 일시적으로 생산을 중지하는 기간을 첫인기간이라고 한다. 이 첫인기간은 생산시간이 0이며, 생산량 또한 0인 구간으로 해당 기간의 데이터를 제외하지 않고 학습에 사용할 경우 주기적으로 생산량을 0으로 예측하는 잘못된 모델이 만들어질 수 있다. 따라서 Monthly Gas와 Monthly Water가 100 E3m³미만으로 매우 적거나(그림 16), 월 생산시간이 24시간미만으로 한 달 동안 하루도 생산하지 않은 경우 해당하는 생산이력을 제외하였다.



	Count	Mean	Std	Min	25%	50%	75%	Max	단위
Monthly Gas	51171	1307.8	1282.1	0	471.7	890.6	1713.7	11995.7	E3m ³

[그림 16] Histogram of monthly gas using 51,171 data from 942 wells.

첫인조건에 부합하는 데이터를 삭제한 후, 그림 17과 같이 생산이력을 촘촘하게 이어(검은색 점선) 생산이력을 수정하였다. 일반적인 경우 첫인기간동안 유정 주변의 압력이 증가하여 생산재개 시, 생산량이 증가하는 경향이 있기 때문에(Lee et al., 2019), 첫인이력을 별도의 주요인자로 사용해주었다.



[그림 17] Removal of shut-in period and concept of dense production.

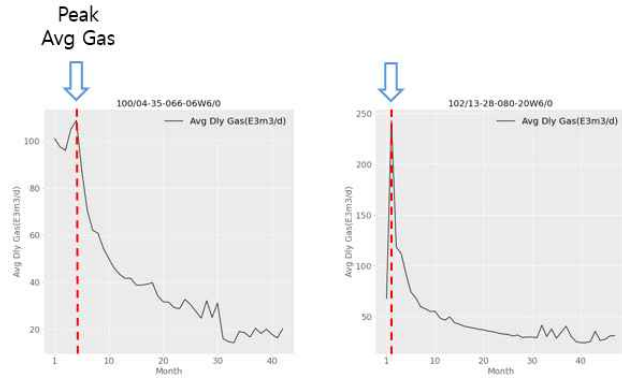
- 피크 이전 생산이력 전처리

Arp의 감퇴곡선기법 수식은 초기생산량인 q_i를 어떻게 설정하느냐에 따라 예측하고자 하는 t시점의 생산량인 q_t가 크게 달라질 수 있다(표 1). 즉, 미래생산량을 예측함에 있어 어느 시점을 생산량이 감퇴하기 시작하는 시점으로 볼 것인가가 매우 중요하다. 셰일가스 생산이력을 보면 첫 번째 개월에 가장 높은 Avg Dly Gas양을 보이는 유정도 있지만, 대부분의 유정은 평균 1~3개월의 안정기를 거친 후 가장 높은 Avg Dly Gas양을 나타낸다.

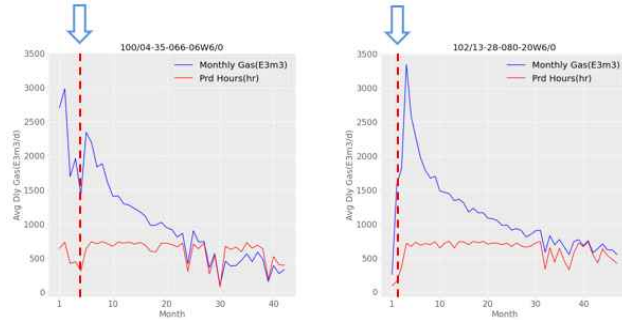
초기 Avg Dly Gas가 증가하는 구간은 유정의 전체적인 감퇴경향을 학습시킴에 있어 불필요한 부분이라고 판단하여, Avg Dly Gas가 가장 높은 시점인 Peak Avg Gas 이전의 생산이력을 삭제하였다(그림 18a). Monthly Gas의 경우 생산시간이 얼마인지에 따라 변동이 크기 때문에(그림 18b), Avg Dly Gas를 기준으로 피크(붉은색 점선)를 선정하였다. 하늘색 화살표는 Avg Dly Gas가 가장 높을 때로, Avg Dly Gas와 Monthly Gas의 피크가 다르다는 것을 나타낸다. 한 개의 유정 당 평균 3개월의 생산이력이 제거되어 총 2,223개의 생산자료가 제거되었다.

최종적으로 기존 1,005개 유정에 대한 56,035개의 데이터에서 위의 전처리 과정을 거친 후, 941개 유정에 대한 45,915개의 데이터가 딥러닝모델 학습을 위해 준비되었다.

(a) Average gas production rate by month



(b) Production hours and monthly gas production by month



[그림 18] Determination of peak production for each well: (a) Average gas production rate by month, and (b) Production hours and monthly gas production by month.

4. 분석 결과

(1) 미래 생산량 예측을 위한 핵심인자 추출

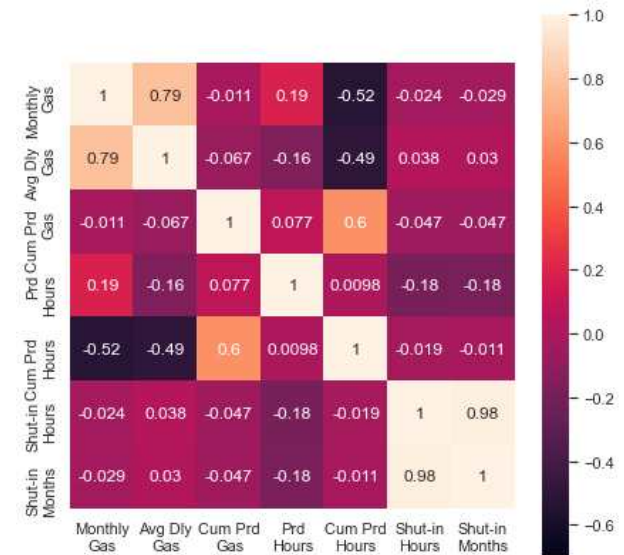
전처리 과정을 거친 후, 미래 생산량 예측을 위한 총 7개의 주요인자를 선정하였다. Monthly Gas, Avg Dly Gas, Prd Hours 이외에 모델 학습 시 과거의 정보를 잘 전달하기 위한 목적으로, 유정마다 Monthly Gas를 누적한 Cum Prd Gas, Prd Hours를 누적한 Cum Prd Hours 인자를 선정하였다. 또한, 전처리 과정에서 제외된 첫인정보는 첫인기간을 시간 단위로 나타낸 Shut-in Hours, 월 단위로 나타낸 Shut-in Months의 인자를 새롭게 설정하였다. 이때, 첫인이 일어나지 않은 경우에는 0으로 입력하고, 첫인이 일어난 경우에만 정보를 입력해주었다. 이러한 첫인정보는 첫인기간 이전이 아닌 첫인 직후의 시계열 자료로 배정하여 첫인기간 이후 생산량에 반영될 수 있도록 하였다. 선정된 7개 인자들을 바탕으로 딥러닝모델을 학습시키기 위한 핵심인자를 추출하였다.

① 인자 간의 상관계수

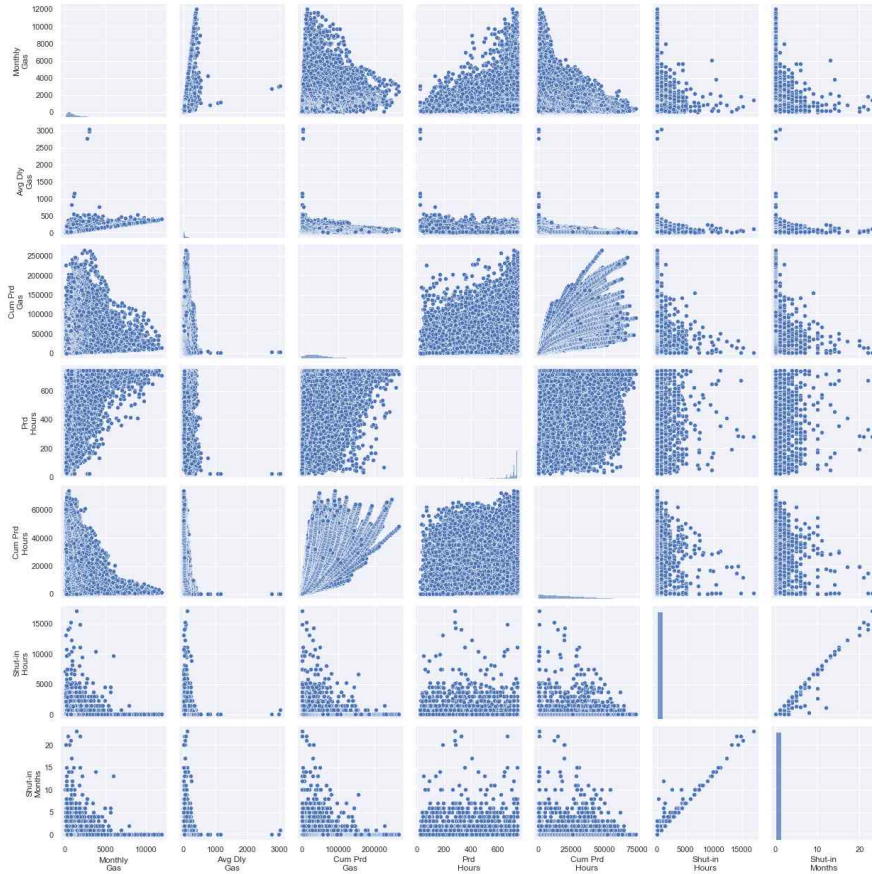
그림 19와 그림 20은 각각 7개 인자 간의 상관계수를 나타낸 heatmap과 산점도를 나타낸 pairplot이다. Monthly Gas와 Avg Dly Gas는 각각 생산량을 나타내는 인자로써 서로 높은 상관관계(0.79)를 보이며, 과거의 정보를 누적하여 나타낸 Cum Prd Gas와 Cum Prd Hours 또한 서로 0.6의 비교적 높은 양의 상관관계를 나타낸다.

Cum Prd Hours는 유정의 생산시간이 증가함에 따라 생산되는 양이 줄어 Monthly Gas와 -0.52의 음의 상관관계를 보이지만, Cum Prd Gas와 Monthly Gas는 상관관계를 보이지 않는다. Cum Prd Hours는 지속적으로 증가하는 반면 Cum Prd Gas는 생산이 진행됨에 따라 수렴하는 경향을 보이기 때문이다. 이러한 이유로, Cum Prd Gas는 강한 음의 상관관계에서 상관계수가 점점 증가해 -0.011의 약한 상관관계가 나타난다.

Prd Hours는 약하지만 Monthly Gas와는 양의 상관관계(0.19)를, Avg Dly Gas와는 음의 상관관계(-0.16)를 보인다. 마지막으로 Shut-in Hours와 Shut-in Months는 생산량과 관련된 Monthly Gas, Avg Dly Gas와의 상관계수가 매우 낮게 나타난다. 이는 대부분의 첫인정보가 0으로 구성되어 있기 때문이다.



[그림 19] Heatmap(correlation map) for the seven features.



[그림 20] Pairplot for the seven features.

위와 같은 상관계수 분석을 바탕으로 딥러닝모델 설계안에 대해 고찰하여 딥러닝 예측모델의 결과 출력층 인자를 Monthly Gas와 Avg Dly Gas 중 Monthly Gas로 설정하였다. Avg Dly Gas는 Monthly Gas와 Prd Hours가 함께 고려된 인자로(그림 21), 선행연구(Lee et al., 2019)에서 출력층 인자로 설정되었지만 Monthly Gas가 Avg Dly Gas보다 입력층 인자와의 상관관계가 높은 것으로 분석되었기 때문이다(그림 19).

$$Avg\ Dly\ Gas\ \left[\frac{E3m3}{d}\right] = \frac{Monthly\ Gas\ prod.\ total\ [E3m3]}{Monthly\ prod.\ time\ [hr]} \times 24\ \left[\frac{hr}{d}\right]$$

[그림 21] Equation of Avg Dly Gas.

② 데이터 시계열 분리

총 7개의 입력층 인자들은 크게 두 가지로 나뉜다. 하나는 예측하고자 하는 생산량과 관련된 인자인 Monthly Gas, Avg Dly Gas 및 Cum Prd Gas이며 다른 하나는 생산정 운영인자와 관련된 Prd Hours, Cum Prd Hours, Shut-in Hours, Shut-in Months이다. 후자의 경우, 자연현상이 아닌 조절 가능한 인자이므로 입력층 인자로 활용 시 시점에 유의해야 한다.

기본적인 예측모델 구조는 출력층의 t시점의 Monthly Gas양을 예측하기 위해 (t-1)시점의 Monthly Gas, Avg Dly Gas, Cum Prd Gas와 t시점의 Prd Hours, Cum Prd Hours, Shut-in Hours, Shut-in Months가 입력된다(그림 22). 즉, 우리가 예측하고자 하는 t시점의 이전 생산량 정보와 t시점의 운영인자를 어떻게 설정할 것인지에 대한 정보가 입력층에 사용된다.

이러한 모델구조에서 t시점의 생산량을 예측하기 위해 과거 이력을 얼마나 이용할지도(생산이력 길이) 중요한 요소이다. 이 연구에서는 과거 생산이력 길이를 3개월, 6개월 또는 12개월로 설정하는 세 가지 경우를 가정하였다.

그림 22는 이전 12개월의 생산이력을 사용하는 경우로 7개의 입력층 인자가 어떠한 시간텀으로 활용되는지 나타낸다. 그림 22에서 가장 마지막 시점 t의 Monthly Gas를 예측하기 위해 (t-12)부터 (t-1)까지의 생산량 정보, (t-11)부터 t까지의 생산시간인자와 첫인인자가 입력층에 사용된다.

Sort UWI	Prod Date	Monthly Gas (E3m ³)	Avg Dly Gas (E3m ³ /d)	Cum Prd Gas (E3m ³)	Prd Hours (hr)	Cum Prd Hours (hr)	Shut-in Hours (hr)	Shut-in Months (month)
1W60791601010000	2012-06-01	1188.1	39.99214586	15494.3	713	6020	0	0
1W60791601010000	2012-07-01	1204.2	39.10798376	16698.5	739	6759	0	0
1W60791601010000	2012-08-01	1067.1	34.65548038	17765.6	739	7498	0	0
1W60791601010000	2012-09-01	305.3	33.92222222	18070.9	216	7714	0	0
1W60791601010000	2012-10-01	920.4	58.28390501	18991.3	379	8093	0	0
1W60791601010000	2012-11-01	827.2	40.93360825	19818.5	485	8578	0	0
1W60791601010000	2012-12-01	1000.4	38.47692308	20818.9	624	9202	0	0
1W60791601010000	2013-01-01	922.8	33.81251908	21741.7	655	9857	0	0
1W60791601010000	2013-02-01	764.6	32.82719141	22506.3	559	10416	0	0
1W60791601010000	2013-03-01	336.3	32.67692308	22842.6	247	10663	0	0
1W60791601010000	2013-05-01	1016.4	57.12786885	23859	427	11090	744	1
1W60791601010000	2013-06-01	920.2	35.67819063	24779.2	619	11709	0	0
1W60791601010000	2013-07-01	837.6	32.84705882	25616.8	612	12321	0	0

Output(t) Input(t-12 ~ t-1) Input(t-11 ~ t)

[그림 22] Example of usage of time-series data for the seven input features in case of the 12-month history case.

③ 시계열에 따른 상관계수의 변화 분석

입력층으로 들어가는 이전 생산이력을 3개월, 6개월, 또는 12개월로 구분하여, 시계열 자료를 전처리한 후 7개 입력인자와 연속시점 t의 Monthly Gas간의 상관관계를 분석하였다(그림 23). 생산이력 길이에 상관없이 과거 Monthly Gas와 Avg Dly Gas는 예측시점의 Monthly Gas와 높은 양의 상관관계를 보였으며, (t-1)로 갈수록 점차 증가한다. 즉, t시점의 Monthly Gas를 예측하기 위해서 최근 시점인 (t-1)의 생산량 정보가 가장 중요한 인자라는 것을 알 수 있다.

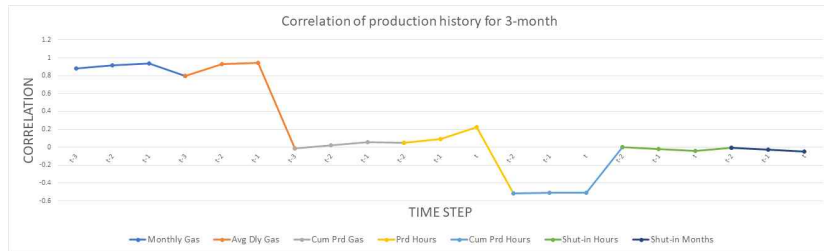
인자 간의 상관계수를 나타낸 그림 19에서는 Monthly Gas와 Cum Prd Gas가 매우 낮은 상관계수를 보였지만, 12개월의 생산이력을 입력했을 때의 상관계수 변화를 나타낸 그림 23에서는 Cum Prd Gas의 상관계수가 점차 증가하여 (t-1)시점에서 약 0.35의 값이 확인되었다. 이는 초기 생산이력에서 Monthly Gas의 양이 감소하고 Cum Prd Gas는 증가하여 낮은 상관계수를 나타내지만, 생산이 지속될수록 Monthly Gas가 감소하는 폭이 작아져 점차 높은 상관계수를 가짐을 알 수 있다(그림 24a).

Prd Hours는 대부분 0.1미만의 낮은 상관관계를 보이지만, 예측하고자 하는 시점 t의 상관계수

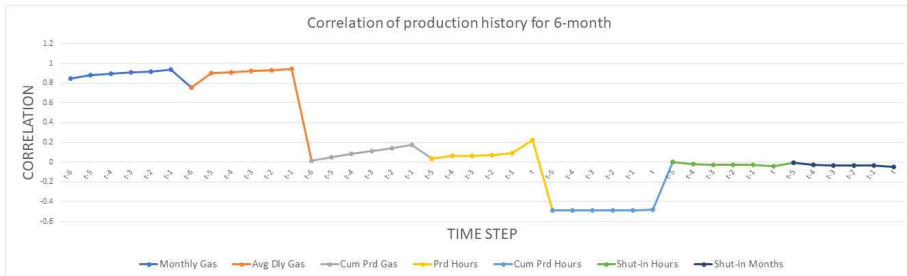
는 비교적 높게 나타나는 것을 볼 수 있다. Prd Hours를 누적한 인자인 Cum Prd Hours는 생산을 계속함에 따라 생산량은 지속적으로 감소하므로, 약 -0.51 ~ -0.44의 음의 상관관계를 보인다. 또한, Cum Prd Hours는 다른 인자와 달리 시계열에 따른 변화가 거의 없는 특징이 있다. 같은 누적인자인 Cum Prd Gas는 시간이 지남에 따라 Monthly Gas가 감소하며, 누적인자가 증가하는 폭이 감소하여 상관계수의 변화가 나타난다. 하지만, Cum Prd Hours의 경우 시간이 지남에 상관없이 Prd Hours가 감소하는 경향이 없고 최댓값인 744시간에 근접하여 유지되므로 누적인자의 증가 폭이 거의 일정하기 때문이다(그림 24b).

Shut-in Hours와 Shut-in Months는 그림 19와 마찬가지로 특정 조건에서 증가하는 생산량을 나타내는 지표이므로 모든 구간에서 가장 낮은 상관관계가 나타난다.

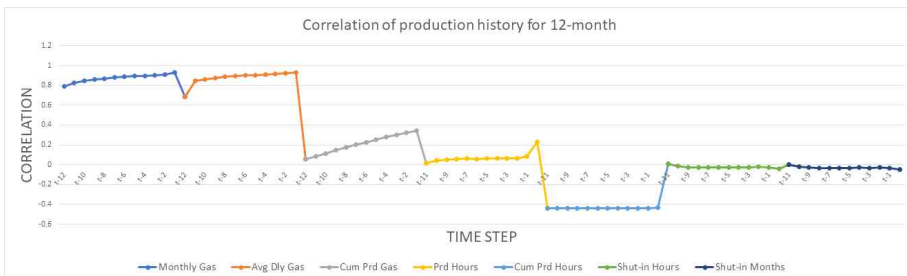
(a) 3-month sequence data



(b) 6-month sequence data

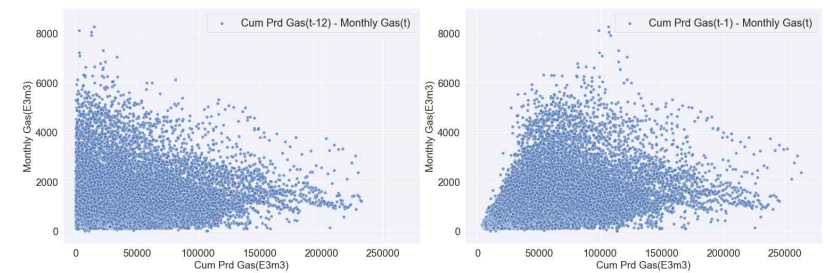


(c) 12-month sequence data

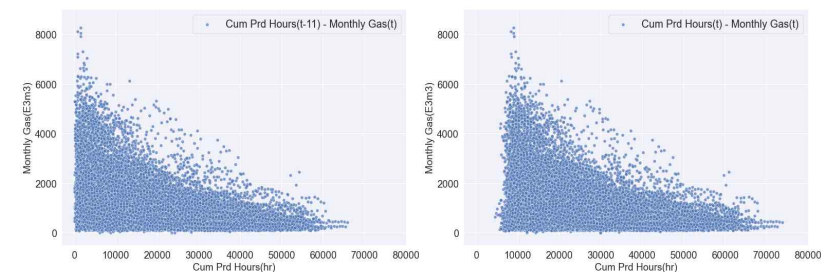


[그림 23] Correlation between future Monthly Gas and the seven factors: (a) 3-month sequence data, (b) 6-month sequence data, and (c) 12-month sequence data.

(a) Cum Prd Gas-Monthly Gas



(b) Cum Prd Hours-Monthly Gas



[그림 24] Changes in correlation of accumulated factors for time: (a) Cum Prd Gas-Monthly Gas, (b) Cum Prd Hours-Monthly Gas.

그림 23의 세 경우를 비교하면 동일한 시점에서의 상관계수가 정확히 일치하지 않는 것을 알 수 있다. 예를 들어 Avg Dly Gas를 비교하였을 때, 그림 23a의 (t-3)~(t-1)과 그림 23c의 (t-3)~(t-1)은 동일한 시점이지만 상관계수 차이가 있다. 그림 23c의 경우, 생산이력이 13개월 미만인 유정의 데이터는 상관계수 평가에 포함되지 않는 반면 그림 23a는 생산이력이 4개월 이상만 되면 해당 유정의 데이터가 상관계수 계산에 포함되기 때문이다.

하지만, Monthly Gas, Avg Dly Gas, Prd Hours, Shut-in Hours, Shut-in Months는 세 경우에 대해 같은 시점에서 나타나는 상관계수의 차이가 크지 않다. 반면, 누적인자인 Cum Prd Gas와 Cum Prd Hours는 생산이력 길이에 따라 동일시점에서도 상관관계 차이가 크고 경향도 달라짐을 확인할 수 있다.

먼저 Cum Prd Gas의 경우, 이전 생산이력을 적게 사용할수록 (t-1)시점의 Cum Prd Gas 상관계수가 0.34에서 0.05로 점차 감소하는 경향을 보인다. 이는 그림 23a와 같이 생산이력이 짧은 유정이 포함된 경우, 해당 유정은 생산량이 감퇴하는 폭이 큰 초기의 생산이력만이 사용되는 데 이 구간의 상관계수가 강한 음의 상관관계를 보여 상관계수 값을 낮추는 영향을 미치기 때문이다. 분석한 결과를 바탕으로 Cum Prd Gas는 이전 생산이력의 길이가 중요한 인자라는 것을 확인할 수 있다. 또한, Cum Prd Gas인자를 사용하기 위해서는 적어도 12개월 이상의 생산이력을 입력해야 유의미한 상관계수가 나타나는 것을 확인할 수 있었다.

또 다른 누적인자인 Cum Prd Hours도 Cum Prd Gas와 마찬가지로 3개월 생산이력의 경우가 강한 음의 상관관계를 보이는 자료를 포함하기 때문에 그림 23a에서의 상관계수(-0.51)가 그림

23c에서의 상관계수(-0.44)에 비해 더 큰 음의 상관관계를 보인다.

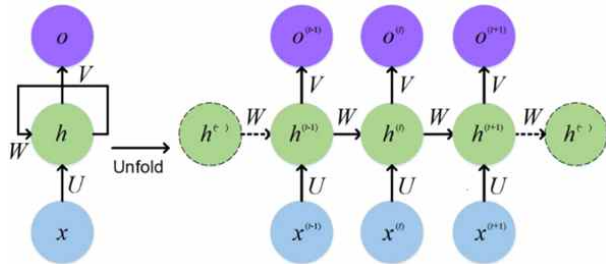
상관계수의 변화를 분석하여 각각의 인자마다 어느 시점의 생산이력이 중요한지 파악할 수 있었고, 생산이력길이에 따른 상관계수의 변화를 확인하였다. 과거 Monthly Gas와 Avg Dly Gas가 t시점의 Monthly Gas 생산량을 예측할 때 가장 중요한 인자이며, t시점의 Prd Hours 또한 중요한 인자로 분석되었다. 누적인자인 Cum Prd Gas와 Cum Prd Hours는 입력된 이전 시계열 자료의 길이에 따른 상관계수의 변화가 크다는 것을 확인할 수 있었다. 첫인자들은 상대적으로 상관계수가 낮았으나, 딥러닝모델 설계 및 최적화 과정에서 추가 분석이 필요할 것으로 보인다.

5. 활용방안

(1) 모델 학습 알고리즘 선정

미래 생산량 예측을 위한 핵심인자 추출 후, 모델 학습을 위한 알고리즘 선정 과정을 수행하였다. 모델 학습 자료는 유정의 생산이력 데이터로, 시간 순서가 중요한 시계열 자료이다. 딥러닝 알고리즘 중, 일반적으로 시계열 자료 처리에 뛰어나다고 알려진 recurrent neural network(RNN)와 long short-term memory(LSTM)를 비교한 후 훈련에 적합한 알고리즘을 선정하였다.

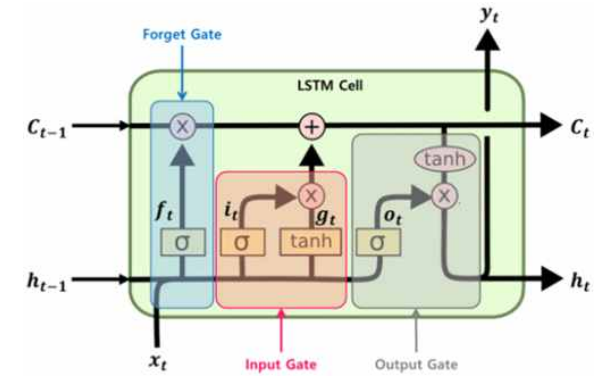
RNN은 딥러닝의 한 종류로, 다른 인공 신경망과 구분되는 특징이 존재한다. 기존 인공 신경망의 경우 입력층에서 출력층 방향으로 즉, 한 방향으로만 계산된 값이 전달되었지만 RNN은 출력층으로 값이 전달될과 동시에 입력층으로도 값이 전달된다는 특징이 있다(그림 25). 이러한 특징으로 인해 RNN은 과거의 자료가 미래에 중요한 영향을 미치게 되는 시계열 자료에 많이 사용된다. 하지만 길이가 긴 시계열 자료의 경우에는 기울기소실(gradient vanishing) 또는 기울기폭주(gradient exploding)가 발생하여 학습이 제대로 이루어지지 않는 문제가 발생한다.



[그림 25] Cell structure of RNN(Ki et al., 2019).

이러한 RNN의 문제점을 보완하기 위해 LSTM이 제시되었다. 출력층으로 전달된 값이 입력층으로도 전달된다는 점은 RNN과 동일하지만 구조에서 LSTM과 RNN의 차이가 존재한다. LSTM은 장기상태(long-term state)인 C_t 가 추가되었으며, 추가된 C_t 로 인해 길이가 긴 시계열 자료의 경우 학습이 잘 이루어지지 않았던 RNN의 문제를 해결할 수 있게 되었다(Ki et al., 2019). 또한, 입력된 장기기억 중 불필요한 정보를 삭제하는 Forget Gate, 장기기억에 입력할 정보를 결정하는 Input Gate, 다음 셀에 전달할 단기기억을 정하는 Output Gate 세 개의 게이트로 구성되어 있다(그림 26).

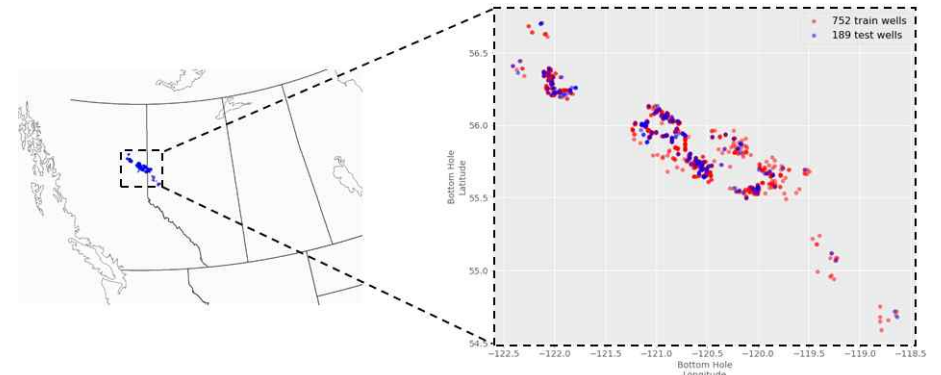
두 알고리즘을 비교한 결과 LSTM이 상대적으로 길이가 긴 시계열 자료에서도 신뢰도 높은 결과를 얻을 수 있음을 확인하였고, 본 연구를 진행함에 있어 모델 학습에 사용할 알고리즘을 LSTM으로 선정하였다.



[그림 26] Cell structure of LSTM(Ki et al., 2019).

(2) 학습을 위한 데이터셋 분리

LSTM을 통한 학습을 진행함에 앞서, 데이터를 학습을 위한 자료와 성능 평가를 위한 자료로 분리하는 과정을 수행하였다. 총 941개의 유정 중 80%인 752개의 유정을 train 자료로, 20%인 189개의 유정을 test 자료로 분리한 후 train 자료의 20%를 모델의 성능을 평가하기 위한 validation 자료로 사용하였다. train과 test 자료의 유정이 고르게 분포하지 않고 특정 지역에 밀집되어 있을 경우 잘못된 모델 학습이 이루어질 수 있으며, 해당 모델의 예측 결과 또한 신뢰도가 낮을 수 있다. 본 연구에서는 그림 27과 같이 train 및 test 자료의 유정 분포를 균일하게 구성하였다.



[그림 27] Location of train/test wells.

분리된 자료의 유정 및 데이터 수는 사용하고자 하는 시계열 길이에 따라 변화할 수 있다. 표 6, 표 7에 나타나듯이 이전 3개월의 생산이력을 사용하였을 때보다, 이전 12개월의 생산이력을 사용하였을 때 사용할 수 있는 유정 및 데이터의 수가 감소한다. 과거 생산이력을 3개월만 사용하

는 경우 유정의 생산이력이 최소 4개월만 되더라도 해당 유정을 자료로 사용할 수 있지만, 과거 12개월의 생산이력을 사용하는 경우에는 최소 13개월 이상의 생산이력이 존재하는 유정만이 자료로 사용될 수 있다.

941개 유정의 45,915개 데이터를 분석한 결과 4개월 이상의 생산이력이 존재하는 유정의 수는 921개, 7개월 이상은 899개, 13개월 이상인 유정은 833개이며, 모델 훈련에 사용할 수 있는 유정의 개수는 각각 736개, 717개, 667개로 더 많은 과거 생산이력을 사용할수록 개수가 줄어드는 것을 확인하였다(표 6). 유정 개수가 감소함에 따라 모델 훈련에 사용할 수 있는 데이터 수 또한 3개월의 경우 34,563개, 12개월의 경우 28,214개로 6,349개의 데이터가 줄어드는 것을 확인하였다(표 7).

[표 6] Variation in the number of available wells

	train(well)	test(well)	total(well)
3-month	736	185	921
6-month	717	182	899
12-month	667	166	833

[표 7] Variation in the number of available data

	train(data)	test(data)	total(data)
3-month	34563	8550	43113
6-month	32371	7995	40366
12-month	28214	6941	35155

데이터를 구성하고 있는 시계열의 길이가 긴 경우, 길이가 짧은 경우에 비해 더 좋은 예측 성능이 기대될 수 있다. 하지만 표 6, 표 7과 같이 시계열 길이에 따른 자료 수의 차이가 큰 경우, 과거 오랜 기간을 사용하는 것보다 짧은 기간으로 분리하여 많은 훈련 자료를 확보하는 것이 좀 더 신뢰도 높은 예측 결과가 도출될 수 있다. 본 연구에서는 시계열 길이와 자료 수의 차이 중 모델 훈련에 있어 더 긍정적인 영향을 미치는 요소를 알아보기 위한 과정을 수행하였다.

(3) 케이스 선정

① 생산시간 관련 인자

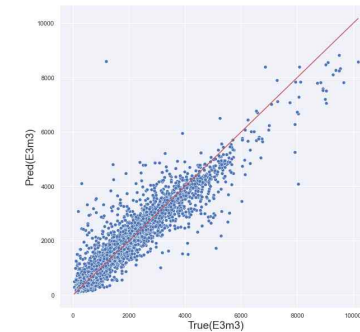
추출된 핵심인자를 바탕으로 케이스 선정 과정을 수행하기에 앞서, 생산시간 관련 인자가 생산량 예측에 미치는 영향을 분석하였다. 분석에 사용한 방법은 입력층 인자로 생산량 관련 인자인 Monthly Gas만을 고려한 경우와 생산시간 관련 인자 중 하나인 Prd Hours를 같이 고려한 경우로 구분하여 Prd Hours 유무에 따른 예측 결과를 비교하였다.

그림 28, 그림 29는 데이터를 3개월로 구분한 후, 입력층 인자로 Monthly Gas만을 사용하였을 때, Monthly Gas와 Prd Hours를 같이 사용하였을 때의 예측 결과이다. 그림 28a와 그림 29a는 한 달 후의 미래 생산량 예측 결과를 나타낸 그래프이다. x축은 실제 값, y축은 예측 값을 나타내며, y=x축을 나타내는 빨간색 직선상에 많은 점이 수렴할수록 더 신뢰도 높은 모델이라고 판단된다. Prd Hours를 고려한 경우, 고려하지 않은 경우에 비해 실제 값과 예측 값이 y=x축 직선상에 잘 수렴하고 있는 것을 확인했으며, 상대오차 값(relative error) 또한 각각 11.3%, 18.6%로 많은 차이가 나타났다(표 8).

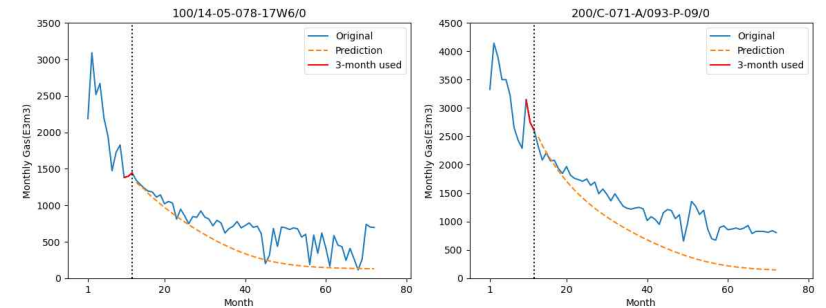
Monthly Gas만을 입력층 인자로 사용한 모델의 낮은 예측 신뢰도는 그림 28b와 그림 29b의 예측 결과를 통해서도 확인할 수 있었다. 그림 28b와 그림 29b는 유정이 12개월까지의 생산이력이 있다고 가정하였을 때(검은색 점선), 마지막 10개월부터 12개월까지 3개월의 생산이력만(빨간색)을 가지고 예측한 결과 값을 다시 입력 값으로 전달하여 번 미래의 생산량을 예측한 결과이다. 주황색 그래프는 모델의 생산량 예측 결과를 나타내며, 파란색 그래프는 예측 결과를 비교하기 위해 도시한 실제 유정의 생산이력이다. Prd Hours를 사용하지 않은 경우(그림 28b), DCA기법의 예측 결과와 같이 곡선 형태의 결과가 도출되어 생산이 감퇴하는 경향만을 확인할 수 있었지만, Prd Hours를 사용한 경우에는(그림 29b) 생산시간에 따라 변동하는 Monthly Gas의 경향을 어느 정도 예측하는 것으로 나타났다.

분석 결과를 통해 Monthly Gas를 예측할 시, Prd Hours를 같이 고려한다면 고려하지 않았을 때보다 더욱 신뢰도 높은 예측 결과가 나타나는 것을 확인하였다. 이는 Monthly Gas가 생산시간이 얼마인지에 따라 변동이 큰 인자인 이유라고 생각된다. 그림 29b 또한 그림 29a에 비해 더 좋은 예측 결과를 확인할 수 있지만, 정확한 생산량을 예측하는 데에는 한계가 있는 것으로 나타났다. 이러한 모델의 성능을 최적화하기 위해 다른 추가적인 입력층 인자를 사용하여 모델을 구성하는 과정을 수행하였다.

(a) Future production forecast in a month

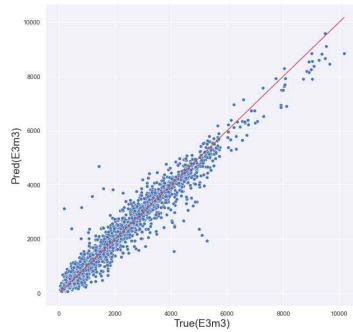


(b) Production forecast for the far future with 3-month production history

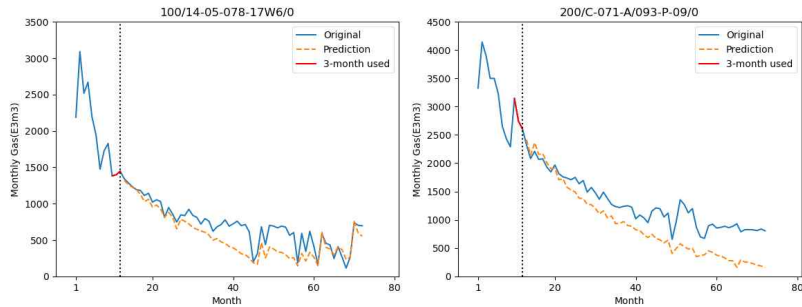


[그림 28] Using Monthly Gas as input feature.

(a) Future production forecast in a month



(b) Production forecast for the far future with 3-month production history



[그림 29] Using Monthly Gas and Prd Hours as input feature.

[표 8] Metric values for Prd Hours

Input features		MSE	R ²	relative error
생산량 인자	생산정 운영인자			
Monthly Gas	Not used	0.0977	0.9087	18.6%
Monthly Gas	Prd Hours	0.0349	0.9673	11.3%

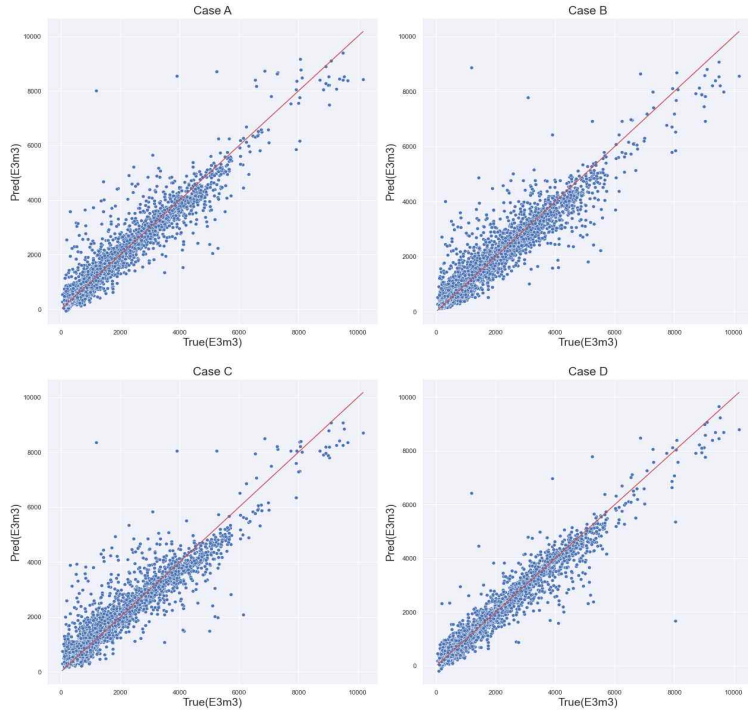
생산시간 관련 인자는 Prd Hours 이외에 Cum Prd Hours가 고려될 수 있다. Cum Prd Hours는 유정마다 Prd Hours를 누적한 값으로 Cum Prd Gas와 같이 누적인자로 활용된다. 이전 핵심인자 추출과정에서 Cum Prd Hours를 하나의 입력층 인자로 선정한 이유는 생산시간을 누적한 인자로써, Prd Hours를 대체할 수 있는 생산정 운영인자라고 판단하였고, 동시에 누적된 값을 통해 과거 시계열 중 어느 시점의 정보인지를 나타내는 인자라고 생각하였기 때문이다. 이와 같은 견해를 바탕으로 Cum Prd Hours가 Prd Hours를 대체할 수 있는 지, 혹은 더 나은 입력층 인자가 될 수 있는 지 알아보기 위해, 표 9와 같이 생산정 운영인자가 Cum Prd Hours 단독으로 사용된 케이스를 선정하여 그에 따른 결과를 분석하였다.

[표 9] Case in which operating factors are used as Cum Prd Hours

Case	Input features	
	생산량 인자	생산정 운영인자
A	Monthly Gas, Avg Dly Gas	Cum Prd Hours
B	Monthly Gas, Cum Prd Gas	Cum Prd Hours
C	Avg Dly Gas	Cum Prd Hours
D	Monthly Gas, Avg Dly Gas, Cum Prd Gas	Cum Prd Hours

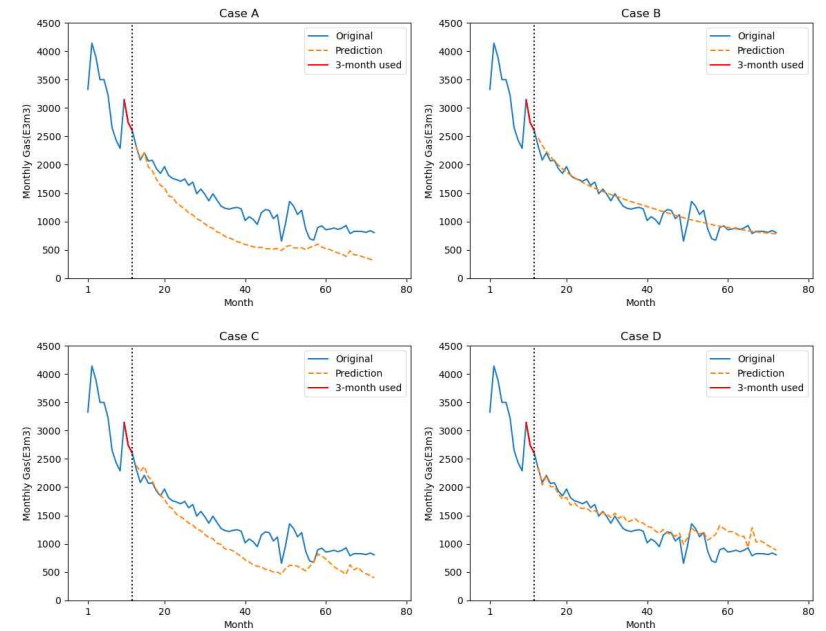
이전 Prd Hours의 유무에 따른 결과 비교 방법과 마찬가지로 데이터를 3개월로 구분한 뒤, 모델의 예측 결과를 분석하였다. Case A~Case D의 한 달 후 미래 생산량 예측 결과를 나타낸 그래프는 그림 30에 도시하였다. 그래프 분석 결과 전체적으로 예측 값과 실제 값의 차이가 크게 나타나는 것을 볼 수 있었으며, 입력층 인자로 Monthly Gas와 Prd Hours를 사용한 경우(그림 29a)보다 오차가 큰 결과 값이 도출된 것을 확인할 수 있다. 먼 미래의 생산량을 예측한 결과 또한 마찬가지로(그림 31), Cum Prd Hours가 Prd Hours를 대체하지 못하여 생산시간에 따라 변동하는 생산이력을 예측하지 못하고 있음을 알 수 있었다. 상대오차를 계산한 결과, 생산시간이 학습되지 않아 곡선 형태의 결과가 도출되었지만, 도출된 예측 결과가 생산 감퇴경향을 잘 대표하는 Case B를 제외한 나머지는 상대오차 값이 크게 나타나는 것을 확인하였다.

생산시간 관련 인자를 분석한 결과, Monthly Gas는 Prd Hours와 밀접한 관련이 있는 것을 확인하였고 누적인자인 Cum Prd Hours는 Prd Hours를 대체하여 사용할 수 없다는 사실을 알 수 있었다. 다음과 같은 분석 결과를 참고자료로 활용하여 케이스 선정을 위한 입력층 인자 선정 과정을 수행하였다.



Case	MSE	R ²	relative error
A	0.0768	0.9282	16.3%
B	0.0958	0.9104	19.8%
C	0.1019	0.9048	22.0%
D	0.0493	0.9538	13.0%

[그림 30] Future production forecast in a month of Case A~Case D.



200/C-071-A/093-P-09/0		
Case	MSE	relative error
A	0.1721	37.8%
B	0.0132	8.4%
C	0.1028	27.8%
D	0.0315	16.1%

[그림 31] Production forecast for the far future of Case A~Case D.

② 최종 케이스 선정

본 연구를 진행함에 있어 추출된 핵심인자가 미래 예측 결과에 미치는 영향을 알아보기 위해 총 9개의 최종 케이스를 선정하였다. 또한, 시계열 길이에 따른 자료 수의 차이가 큰 경우 시계열 길이와 자료 수의 차이 중 모델의 예측 결과에 더 큰 영향을 미치는 요소를 알아보기 위해 과거 생산이력의 길이를 3개월, 6개월, 12개월로 구분하여 총 27개의 케이스를 선정하였다. 각 케이스에 사용된 입력층 인자의 종류는 표 10에 도시하였으며, 케이스 뒤 괄호안의 숫자를 통하여 해당 케이스의 데이터가 몇 개월로 구분되었는지 나타내었다. 즉, Case 1(3)은 입력층 인자로 Monthly Gas와 Prd Hours를 사용하고 데이터를 3개월로 구분한 경우를 의미한다.

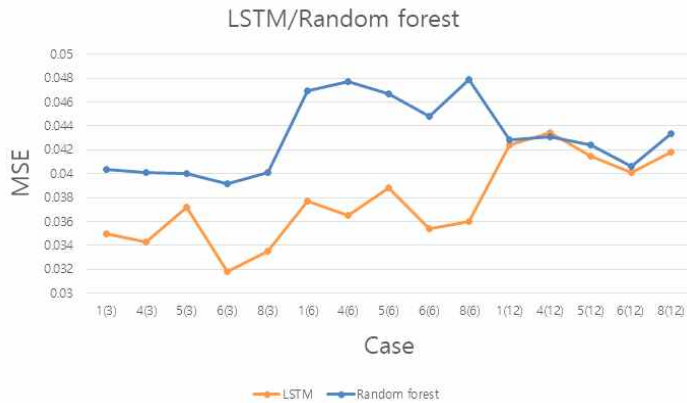
생산량 관련 인자와 생산정 운영인자를 적어도 하나씩은 포함시켜 케이스를 선정하였고, 생산정 운영인자의 경우 앞서 생산시간 관련 인자를 분석한 결과를 참고하여 모든 케이스에 Prd Hours 인자를 포함하여 선정하였다.

[표 10] Input features used for each case

Case	Input features	
	생산량 인자	생산정 운영인자
1	Monthly Gas	Prd Hours
2	Avg Dly Gas	Prd Hours
3	Monthly Gas, Avg Dly Gas	Prd Hours
4	Monthly Gas, Cum Prd Gas	Prd Hours
5	Monthly Gas	Prd Hours, Shut-in Hours, Shut-in Months
6	Monthly Gas, Avg Dly Gas, Cum Prd Gas	Prd Hours, Cum Prd Hours
7	Monthly Gas, Avg Dly Gas, Cum Prd Gas	Prd Hours, Cum Prd Hours, Shut-in Hours, Shut-in Months
8	Monthly Gas, Cum Prd Gas	Prd Hours, Cum Prd Hours
9	Monthly Gas, Avg Dly Gas	Prd Hours, Cum Prd Hours, Shut-in Hours, Shut-in Months

(4) LSTM/Random forest 예측 결과 비교

선정된 케이스들을 바탕으로 한 결과 비교 분석에 앞서, 알고리즘 간 예측 성능의 차이를 확인해보았다. LSTM 알고리즘과 앙상블 기반 머신러닝 알고리즘인 Random forest를 적용하였으며, 두 알고리즘의 예측 결과에서 차이를 확인할 수 있었다(그림 32). 그림 32는 LSTM과 Random forest의 결과를 비교한 그래프로 x축은 알고리즘을 적용한 케이스, y축은 MSE 값을 나타낸다. 생산이력 길이를 12개월로 분리한 케이스에서는 뚜렷한 차이가 나타나지 않았지만, 대체적으로 LSTM을 적용한 학습 결과의 오차가 더 낮게 측정이 되는 것을 확인하였다. 생산이력 길이를 3개월, 6개월로 분리한 케이스는 두 알고리즘의 유의미한 차이가 나타났다. 해당 구간의 모든 케이스에서 LSTM의 오차가 더 낮게 측정이 되었으며, Random forest와 비교하였을 때 큰 폭으로 감소한 것을 확인할 수 있다. 특히, Case 4(6)과 Case 8(6)은 약 23%, 25%의 오차 감소율을 보였다.

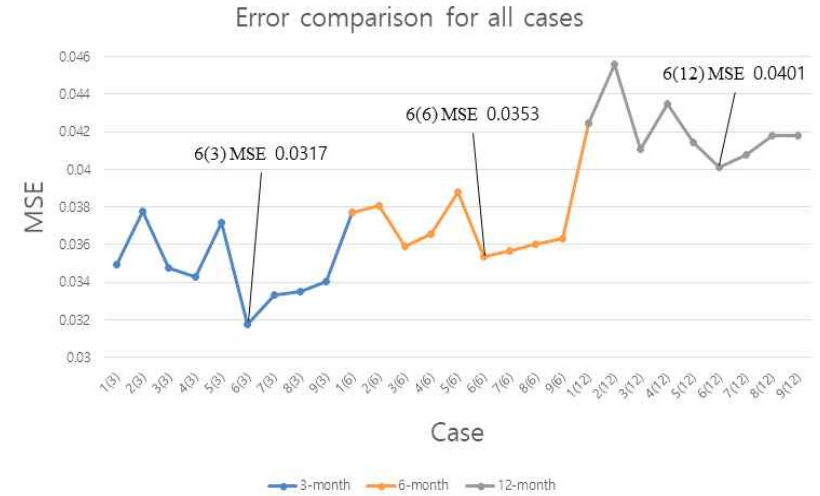


[그림 32] MSE comparison of LSTM and Random forest.

(5) 케이스 결과 비교 및 분석

앞서 선정된 케이스를 통해 모델 학습 과정을 거친 후, 각 인자가 미래 예측 결과에 미치는 영향을 알아보기 위한 비교 분석을 수행하였다. 그림 33에 최종 선정된 27개 케이스의 MSE 값을

도시하였으며, 파란색 그래프는 생산이력을 3개월로 분리하였을 때, 주황색은 6개월, 회색은 12개월로 분리한 경우이다. 표 11은 각 케이스에서 측정된 MSE 및 R²의 값을 확인할 수 있다. 그래프의 경향을 분석한 결과, 사용된 시계열 길이가 길수록 오차 값이 증가하는 것이 확인되었다(그림 33). 이는 12개월 단위로 자료를 구성하였을 때보다, 3개월 단위로 구성하였을 때 사용가능한 훈련 자료 수가 증가하기 때문이다(표 7). 즉, 시계열 길이에 따른 자료 수의 차이가 클 때, 길이가 긴 시계열 단위로 자료를 구성하기보다, 길이가 짧은 시계열 단위로 세분하여 많은 자료를 확보하는 것이 예측 결과의 신뢰도를 높일 수 있다.



[그림 33] Comparison of error trends in cases.

[표 11] Comparing MSE and R² of the entire cases

Case	MSE			R ²		
	3-month	6-month	12-month	3-month	6-month	12-month
1	0.0350	0.0377	0.0424	0.9674	0.9658	0.9609
2	0.0378	0.0381	0.0456	0.9647	0.9654	0.9579
3	0.0348	0.0359	0.0411	0.9675	0.9674	0.9621
4	0.0343	0.0365	0.0435	0.9680	0.9668	0.9599
5	0.0372	0.0388	0.0414	0.9653	0.9648	0.9618
6	0.0318	0.0354	0.0401	0.9703	0.9679	0.9630
7	0.0333	0.0356	0.0408	0.9689	0.9676	0.9624
8	0.0335	0.0360	0.0418	0.9687	0.9673	0.9615
9	0.0340	0.0363	0.0418	0.9682	0.9670	0.9614

시계열 길이에 따른 오차의 경향을 분석한 후, 각 케이스에서 나타나는 오차를 비교 분석하여 사용된 인자가 예측 결과에 미치는 영향을 알아보는 과정을 수행하였다.

① Monthly Gas

Monthly Gas를 입력층 인자로 가용했을 경우와 가용하지 않았을 경우의 예측 결과차이를 분석하였다. Case 1은 생산량 관련 인자로 Monthly Gas만을 사용하였고, Case 2는 Avg Dly Gas만을 사용하였다. 이전 핵심인자 추출을 위해 시계열에 따른 상관계수의 변화를 분석한 결과(그림 23), Avg Dly Gas는 Monthly Gas와 비슷하거나 더 높은 상관계수 값이 나타나는 구간이 확인되어 Case 2의 생산량 인자로 사용되었다. Case 1과 Case 2를 비교한 결과, Case 1(3)의 오차는 약 0.035, Case 2(3)은 0.038로 Monthly Gas를 입력층 인자로 사용하였을 때 더 낮은 오차가 나타났으며, 생산이력 길이를 6개월, 12개월로 설정하였을 때도 마찬가지로 Case 2에 비해 Case 1의 오차가 더 낮은 수준임을 확인할 수 있다. 이를 통해 Monthly Gas는 t시점의 Monthly Gas를 예측하기 위한 핵심 인자로 판단하였다.

② Avg Dly Gas

Case 1과 Case 2를 비교함으로써 Avg Dly Gas만을 생산량 인자로 가용할 경우 Monthly Gas만을 가용한 경우에 비해 예측 신뢰도가 떨어지는 것을 확인하였지만, Avg Dly Gas와 다른 생산량 관련 인자 및 생산정 운영인자를 같이 고려할 경우, 예측 신뢰도가 높아지는 것을 확인하였다. 생산량 인자로 Monthly Gas와 Avg Dly Gas를 같이 고려한 Case 3과 Monthly Gas만을 사용한 Case 1을 비교한 결과 모든 시계열 길이에서 Case 3의 오차가 더 낮게 나타났으며, Case 6과 Case 8의 비교 결과 또한 Avg Dly Gas를 고려한 Case 6의 예측 결과 값이 더 좋은 것으로 확인되었다.

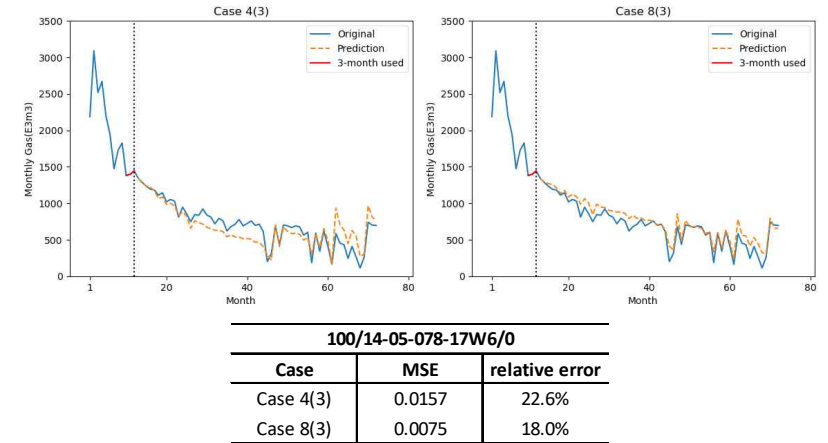
③ Cum Prd Gas

누적인자인 Cum Prd Gas가 미래 예측 결과에 미치는 영향을 분석하기 위해 Case 1과 Case 4를, Case 7과 Case 9의 예측 결과를 확인하였다. Case 1에 사용된 인자에서 Cum Prd Gas가 추가적으로 고려된 Case 4는 Case 1보다 좋은 예측성능이 확인되었다. 7개의 모든 인자가 고려된 Case 7과 Cum Prd Gas만을 고려하지 않은 Case 9의 오차 비교 결과 또한, Cum Prd Hours를 같이 고려하였을 때 예측성능이 향상되는 것을 확인하였다. 이를 통해 누적인자인 Cum Prd Gas는 생산량 예측에 긍정적인 영향을 미친다고 판단하였다.

④ Cum Prd Hours

Cum Prd Hours는 Prd Hours를 누적한 인자로, 생산정 운영인자 중 하나이다. 이전 생산시간 관련 인자 분석과정에서, 선정된 Case A~Case D(표 9)의 결과 분석을 통해 Cum Prd Hours가 Prd Hours를 대체하여 사용될 수 없다는 것을 확인하였다. 본 과정에서는 Case 4와 Case 8을 활용하여, Cum Prd Hours가 Prd Hours를 포함한 다른 인자와 같이 사용된 경우의 예측 결과를 확인하였다. Case 4는 Monthly Gas, Cum Prd Gas와 생산정 운영인자로 Prd Hours만이 사용되었고, Case 8은 Case 4에서 고려된 인자와 Cum Prd Hours를 같이 사용하였다.

두 케이스를 비교한 결과 Cum Prd Hours를 같이 사용한 Case 8이 더 낮은 오차를 가지는 것을 확인했으며, 3개월의 생산이력을 사용하여 미래 생산량을 예측한 경우, Case 8(3)의 예측 결과가 Case 4(3)의 예측에 비해 불확실성이 개선된 모습을 확인할 수 있다(그림 34). 분석 결과, 생산정 운영인자로 Cum Prd Hours를 단독으로 사용할 경우 예측 결과의 신뢰도가 낮게 나타났지만, 다른 인자와 함께 고려하여 사용한다면 더 신뢰도 높은 예측 결과가 나타나는 것을 확인하였다.

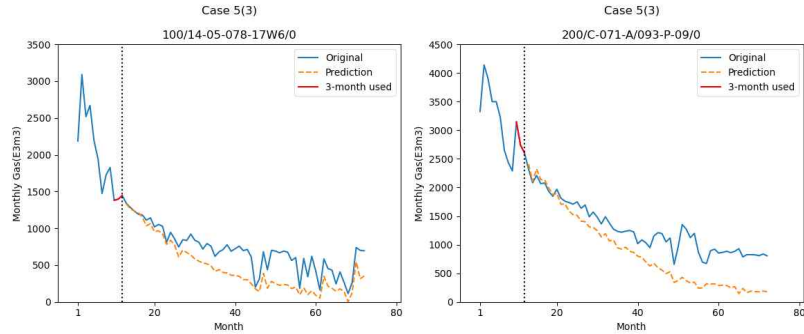


[그림 34] Production forecast for the far future of Case 4(3) and Case 8(3).

⑤ Shut-in Hours, Shut-in Months

이전 상관계수 분석 과정에서 Monthly Gas와 셋인 인자와의 상관계수는 각각 -0.024, -0.029(그림 19)로 생산량 관련 인자와의 상관관계가 거의 관찰되지 않았다. 시계열에 따른 상관계수의 변화를 분석한 결과 또한 다른 인자와 비교하였을 때, 가장 낮은 상관관계가 나타났다(그림 23). 이는 셋인 인자가 특정 조건을 나타내는 인자이기 때문이라고 판단하였고, 실제 입력층 인자로 가용하였을 때 나타나는 예측 결과의 변화를 분석하였다.

분석에 활용한 케이스는 셋인 인자를 사용한 케이스와 사용하지 않은 케이스인 Case 1과 Case 5, Case 6과 Case 7의 오차를 비교하였다. 분석한 결과 두 경우 모두 셋인 인자를 고려하였을 때의 오차가 고려하지 않았을 때보다 오차가 증가하는 것을 확인하였다. 3개월의 생산이력을 이용하여 예측한 결과를(그림 35), Case 1(3)의 결과인 그림 29와 비교하였을 때, 셋인 인자를 사용한 Case 5(3)은 Case 1(3)에 비해 예측 결과의 개선 정도를 확인할 수 없었으며, 오히려 측정된 오차가 증가한 것을 확인할 수 있었다. 이와 같은 분석을 통해, 셋인 인자는 후후 연구를 통해 최적화 과정이 필요할 것으로 생각된다.

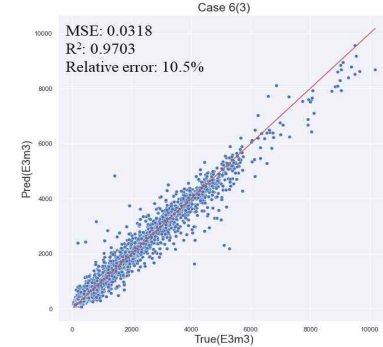


Case	100/14-05-078-17W6/0		200/C-071-A/093-P-09/0	
	MSE	relative error	MSE	relative error
Case 1(3)	0.0321	26.1%	0.1431	36.1%
Case 5(3)	0.0578	40.1%	0.1778	39.8%

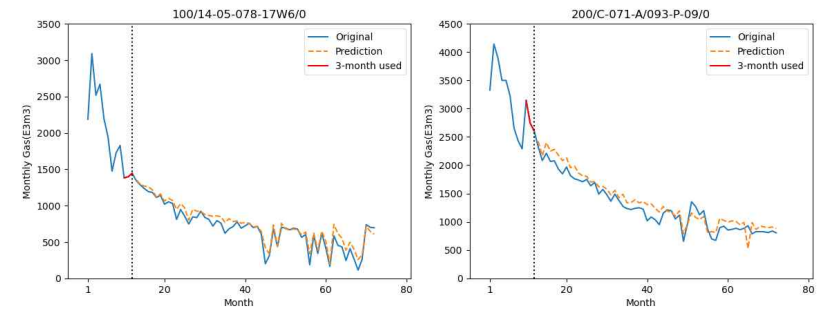
[그림 35] Production forecast for the far future of Case 5(3).

케이스마다 다르게 나타나는 오차를 비교 분석한 결과, 각 인자가 예측성능에 미치는 영향을 확인할 수 있었다. 확인한 결과 Shut-in Hours, Shut-in Months를 제외한 나머지 인자는 입력층 인자로 사용될 경우, 예측성능이 향상되는 것을 확인하였지만, 셋인 관련 인자인 Shut-in Hours와 Shut-in Months는 입력층 인자로 사용될 시, 모델의 예측 불확실성이 높아지는 것을 확인했다. 이와 같은 결과를 바탕으로 셋인 인자를 제외한 Monthly Gas, Avg Dly Gas, Cum Prd Gas, Prd Hours, Cum Prd Hours의 5개 인자가 미래 예측을 위한 주요인자로 판단되었다.

5개의 주요인자가 입력층 인자로 사용된 Case 6은 그림 33에 도시되어 있듯이, 모든 시계열 길이 구간에서 가장 낮은 오차를 확인할 수 있었으며, 자료를 3개월로 분리하였을 때의 MSE 및 상대오차는 각각 0.0318, 10.5%로 나타났다. 이를 통해 Case 6(3)의 모델이 미래 생산량을 예측하는데 있어 가장 최적화된 모델임을 알 수 있다. Case 6(3)의 한 달 후 생산량 예측 결과는 그림 36에서 볼 수 있듯이 실제 값과 예측 값이 잘 수렴하고 있는 것을 확인할 수 있다. 3개월의 생산이력을 통해 먼 미래를 예측해본 결과(그림 37), 60개월 이후의 생산량을 비교적 잘 예측하는 것을 확인할 수 있었으며, 생산시간에 따라 변동하는 생산이력 또한 잘 나타나고 있음을 확인하였다.



[그림 36] Future production forecast in a month of Case 6(3).



Case	100/14-05-078-17W6/0			200/C-071-A/093-P-09/0		
	MSE	R ²	relative error	MSE	R ²	relative error
Case 6(3)	0.0054	0.9172	15.4%	0.0173	0.8851	11.1%

[그림 37] Production forecast for the far future of Case 6(3).

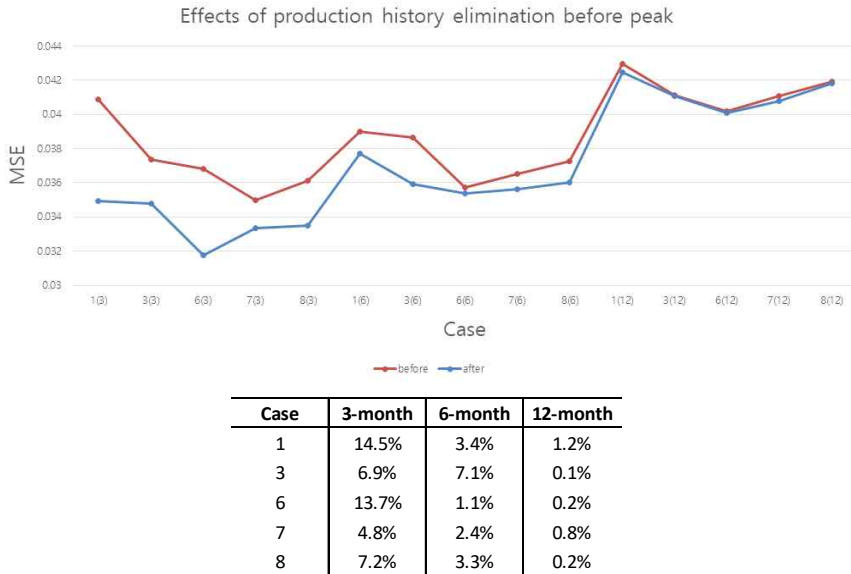
(6) 피크 이전 생산이력 제거 결과 분석

이전 자료 전처리 과정에서 피크 이전의 생산이력 제거를 수행하였다. 초기 생산성이 증가하는 구간의 자료는 생산성이 감퇴하는 경향을 예측하기 위한 모델의 훈련자료로 적합하지 않다고 판단하여 총 2,223개의 생산자료를 제거하였다. 본 과정에서는 해당 전처리 과정을 수행했을 때와 수행하지 않았을 때의 결과를 비교하여, 피크 이전 생산이력을 제거하는 것이 생산량 예측에 있어 적합한 전처리 과정임을 나타내고자 한다.

각 케이스에서 도출된 결과와 피크 이전의 생산이력을 제거하지 않고 모델 학습을 진행하였을 때의 결과를 비교하였다. 그림 38은 제거 전과 제거 후의 오차를 도시한 그래프로, 빨간색 그래프와 파란색 그래프는 각각 제거 전, 제거 후의 결과를 나타낸다. 해당 전처리 과정을 수행하였을 때 전체적으로 오차가 줄어든 모습을 확인하였다, Case 1(3)은 14.5%의 오차가 개선된 것을 확인하였고, 생산량 예측에 가장 최적화된 Case 6(3)은 13.7%의 오차 개선 효과를 보였다.

앞서 시계열 길이와 자료 수 차이의 영향을 분석한 결과, 짧은 생산이력 단위로 세분하여 훈련

자료 수를 증가시키는 것이 예측 결과의 신뢰도를 높이는 것을 확인하였다. 이는 모델의 예측성능에 있어 훈련 자료의 수가 중요한 요소이기 때문이다. 해당 전처리 과정을 수행하면서 2,223개의 피크 이전 생산이력이 제거됨에 따라, 모델 학습에 사용되는 훈련 자료의 수 또한 감소하였다. 그림 38의 그래프를 분석한 결과, 모델 훈련에 사용된 자료 수가 감소하였음에도 불구하고 예측성능이 향상된 것을 확인하였다. 이는 생산량 예측을 위해 불필요한 데이터가 제거되었음을 의미하고, 해당 전처리 과정은 모델이 학습하는데 있어 유의미한 전처리 과정임을 나타낸다.



[그림 38] Effects of production history removal before peak.

피크 이전의 생산이력을 제거한 후, 전체적인 오차 개선이 확인되었지만, 과거 이력을 얼마나 사용하였는지에 따라 오차가 감소하는 경향이 다르게 나타난다. 자료를 12개월 단위로 분리하였을 때보다 3개월 단위로 분리하였을 때, 전처리 과정을 통한 오차 감소율이 크게 나타났으며, 사용된 시계열 길이에 따른 오차 감소율을 그림 39에 도시하였다. 파란색, 빨간색, 연두색 그래프는 각각 이전 3개월, 6개월 및 12개월의 생산이력을 사용한 경우이다.

그래프를 확인한 결과 자료를 3개월로 분리한 경우의 평균 오차 감소율은 9.4%, 12개월은 0.5%로 12개월의 생산이력을 사용하였을 때 예측성능이 거의 개선되지 않는 것을 확인하였다. 이는 사용된 시계열 길이에 따라 자료의 특성이 다르기 때문이다. 16개월의 생산이력 중, 초기 3개월 동안 생산성이 증가하는 유정에서 피크 이전의 생산이력을 제거한 경우를 가정하였을 때, 3개월 단위로 분리한 경우는 초기 생산성이 증가하는 1~3개월의 자료가 하나의 데이터로 학습이 되어 해당 구간의 자료를 제외한다면 학습 결과에 차이가 발생할 수 있다. 하지만, 12개월 단위로 분리한 경우는 1~12개월의 자료가 하나의 데이터로 학습이 진행되기 때문에 초기 생산성이 증가하는 1~3개월의 자료를 제외하더라도 학습 결과에 큰 영향을 미치지 않는다. 이와 같은 이유로 분리

한 시계열 길이에 따라 해당 전처리 과정의 영향에 차이가 발생한다.

위의 분석을 통하여 피크 이전의 생산이력을 제거하는 과정이 신뢰도 높은 예측 결과를 위해 필요한 작업이라고 판단하였고, 해당 전처리 과정은 자료를 12개월로 분리하였을 때보다 3개월 단위로 세분하였을 경우 그 효과가 더 크게 나타나는 것을 확인하였다.



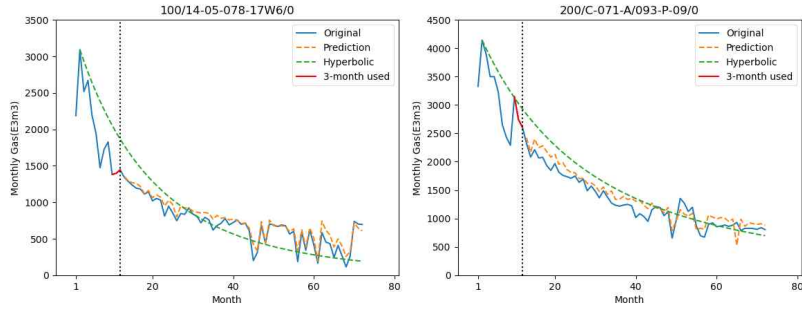
[그림 39] Variation of error reduction rate with time-series.

(7) DCA기법과의 예측 결과 비교

케이스 비교 분석 과정에서 가장 좋은 예측성능이 확인된 Case 6(3)과 DCA기법의 예측 결과를 비교하였다. 미국 에너지 관리청(Energy Information Administration, EIA)에서 자원량을 평가하기 위해 사용하는 hyperbolic식을 활용하였으며(Alolayan et al., 2021), 감퇴지수인 b는 유정의 생산이력에 최적화된 값을 사용하였다. Case 6(3)은 10~12개월의 생산이력만을, DCA기법은 생산량이 가장 높은 시점부터 12개월까지의 자료를 이용하여 생산량을 예측하였다.

Case 6(3)과 DCA기법의 예측 결과는 그림 40과 같이 나타났다. 주황색 그래프는 Case 6(3) 모델의 예측 결과, 연두색 그래프는 DCA기법의 예측 결과이며, 파란색 그래프는 유정의 실제 생산이력을 나타낸다. DCA기법은 곡선 형태의 결과가 도출되어 유정의 생산성이 감퇴하는 경향만을 확인할 수 있었지만, Case 6(4)는 입력층 인자로 생산량뿐만 아니라 생산시간이 같이 고려되어, 변동하는 생산이력이 반영된 예측 결과를 보인다. 그에 따라 오차 값에도 차이가 나타나는 것을 확인하였다. 상대오차를 비교한 결과, Case 6(3)에서는 15.4%, 11.1%의 값이 확인되었지만, DCA기법은 32.3%, 16.9%로 DCA기법을 적용하였을 때보다 해당 모델을 적용할 시, 더 신뢰도 높은 결과가 도출되었다.

이러한 결과를 통해, 해당 모델은 생산시간이 고려되어 DCA기법과 차별성 있는 생산량 예측이 가능하며, 그에 따라 예측 결과의 오차가 개선되는 것을 확인하였다.

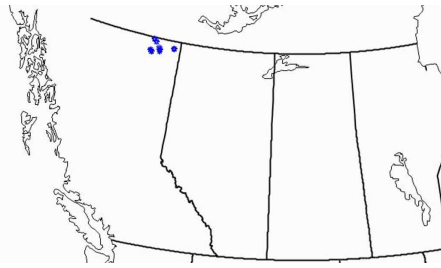


	100/14-05-078-17W6/0		200/C-071-A/093-P-09/0	
	MSE	relative error	MSE	relative error
Case 6(3)	0.0054	15.4%	0.0173	11.1%
Hyperbolic	0.0456	32.3%	0.0597	16.9%

[그림 40] Comparison of the prediction results of Case 6(3) and DCA.

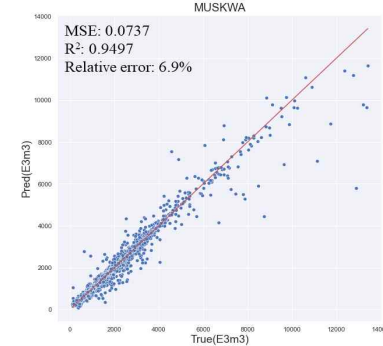
(8) 다른 셰일플레이 적용 결과

본 과정에서는 최적 모델로 선정된 Case 6(3)을 무스카와에 적용하여 생산량을 예측하였다. 무스카와는 브리티시컬럼비아주 북동부에 위치하며(그림 41), 국내 기업이 보유한 셰일 자산이 다수 존재한다.



[그림 41] Wells location of Muskwa in British Columbia, Canada.

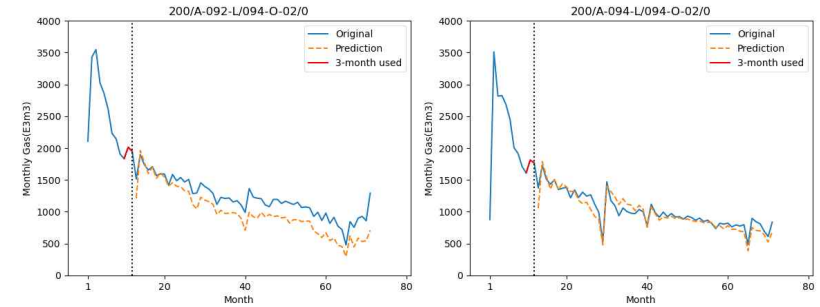
몬트니를 대상으로 수행한 유정 선별 및 자료 전처리 과정을 무스카와에 적용하여 47개 유정의 3,521개 데이터를 확보하였다. 확보한 데이터를 test 자료로 구성한 후, Case 6(3) 모델을 적용하여 무스카와의 미래 생산량을 예측한 결과(그림 42), 특정 예측 값을 제외하면 대부분 오차가 낮은 예측 결과를 나타내고 있다. MSE는 0.0737로 Case 6(3) 모델을 몬트니에 적용하였을 때 (0.0318)보다 높은 오차가 도출되었다. 이는, test 자료로 사용한 몬트니 유정의 최대 생산량은 약 100,000 E3m³이지만 무스카와 유정의 최대 생산량은 약 140,000 E3m³로 생산량 범위의 차이가 큰 이유라고 판단되지만, 두 셰일플레이 간 존재하는 지질학적 특성 차이의 영향일 가능성도 배제할 수 없다. R²과 상대오차는 각각 0.9497, 6.9%로 해당 모델을 통한 예측 결과가 신뢰성 있음을 나타낸다.



[그림 42] Future production forecast in a month of Muskwa using Case 6(3).

10~12개월의 생산이력을 이용하여 무스카와 유정의 미래 생산량을 예측해보았다. 사용된 유정은 실제 국내 기업이 보유한 셰일 자산으로, 예측된 결과는 그림 43과 같이 나타났다. 몬트니 유정에 비해 예측된 결과의 오차가 증가한 경우도 나타났지만, 전반적으로 신뢰할 수 있는 예측 성능이 확인되었다.

몬트니의 생산량 예측에 최적화된 모델을 무스카와에 적용한 결과, 신뢰도 있는 예측 결과가 나타나는 것을 확인했으며, 추후 전이학습(transfer learning)을 통한 최적화 과정을 수행한다면 더욱 신뢰도 높은 예측 모델이 개발될 수 있을 것이라 예상된다.



Case	200/A-092-L/094-O-02/0		200/A-094-L/094-O-02/0	
	MSE	relative error	MSE	relative error
Case 6(3)	0.0466	20.3%	0.0076	7.8%

[그림 43] Production forecast for the far future of Muskwa using Case 6(3).

III. 결론

셰일가스는 전통자원과 생산거동의 차이가 존재하여 기존의 DCA기법을 적용할 경우 예측 결과의 불확실성이 크다. 셰일가스의 생산거동이 반영된 여러 기법이 연구되고 있지만, 사용되는 기법에 따라 예측 결과의 차이가 크며, 연구자의 주관이 들어갈 수 있다는 문제점이 있다. 본 연구에서는 이러한 문제점을 개선하기 위해 과거 생산이력 데이터를 활용한 셰일가스의 미래 생산량을 예측하는 기술을 제안하였고, 연구 과정을 통해 다음과 같은 결과를 얻을 수 있었다.

1. 몬트니 유정의 생산자료 전처리를 수행한 후, 미래 생산량 예측을 위한 핵심인자를 추출하기 위해 각 인자와 출력층 간의 상관관계를 분석하였다. 분석한 결과 생산량 관련 인자는 Monthly Gas, Avg Dly Gas, Cum Prd Gas가, 생산정 운영인자는 Prd Hours, Cum Prd Hours, Shut-in Hours, Shut-in Months가 예측을 위한 핵심인자로 선정되었다.
2. 생산량 예측에 가장 최적화된 모델을 도출하기 위해, 추출된 핵심인자를 바탕으로 9개의 케이스를 구성하였으며, 사용한 과거 이력의 길이에 따른 예측 결과의 차이를 확인하기 위해 3개월, 6개월, 12개월의 생산이력을 사용한 경우로 구분하여 총 27개의 케이스가 선정되었다. 동일 케이스에서 과거 이력 길이에 따른 오차의 경향을 분석한 결과, 짧은 시계열 길이를 사용할수록 오차가 감소하는 것을 확인하였다. 이는, 자료를 짧은 시계열 길이로 세분할수록 모델 훈련에 사용되는 학습자료 수가 증가하여 예측성능에 도움이 된 것으로 판단된다.
3. 케이스에서 도출된 예측 결과를 비교함으로써 각 인자가 예측에 미치는 영향을 확인할 수 있었다. 섯인 인자를 제외한 Monthly Gas, Avg Dly Gas, Cum Prd Gas, Prd Hours, Cum Prd Hours가 주요 인자로 판단되었으며, 해당 주요인자가 사용된 Case 6(3)이 예측을 위한 최적의 모델로 선정되었다. 해당 모델의 MSE는 0.0318, R²은 0.9703이며 상대오차는 10.5%로 신뢰도 높은 예측 결과가 도출되었다.
4. 피크 이전의 생산이력을 제거하는 것이 적합한 전처리 과정임을 확인하였다. 피크 이전의 생산이력을 제거하였을 때, 해당 과정을 수행하지 않았을 때보다 오차가 개선되는 것을 확인하였으며 최적 모델로 선정된 Case 6(3)의 경우 약 13.7%의 오차가 감소하였다. 또한, 해당 전처리 과정은 자료를 3개월로 분리하였을 때 그 효과가 크게 나타났다. 이는, 사용한 시계열 길이에 따른 자료 특성의 차이로 판단된다.
5. Case 6(3) 모델은 DCA기법보다 예측 결과의 오차가 더 낮았으며, 해당 모델은 생산시간이 고려되어, 예측 결과가 실제 유정의 변동하는 생산이력을 잘 나타내는 것이 확인되었다. 또한, 모델을 무스카와 유정에 적용해본 결과, 신뢰성 있는 예측 결과가 나타났다.

본 연구의 최종 모델에 선정된 5개 주요인자는 입력층 인자로 사용될 시, 예측 결과가 향상되는 것을 확인하였지만, 섯인 인자는 예측 결과의 개선이 이루어지지 않았다. 이에 후후 연구에서는 섯인 이후 생산성이 증가하는 경향을 입력층 인자에 잘 반영하여, 출력층과의 상관관계를 높이는 최적화 과정을 수행할 예정이다. 또한, 전이학습을 통해 타 셰일플레어로 확장해 나갈때, 셰일가스자원에 범용적인 모델이 개발될 수 있을 것이라 예상된다.

IV. 참고문헌

- Alolayan, O.S., Raymond, S.J., Montgomery, J.B., Williams, J.R., 2021. Towards Better Shale Gas Production Forecasting Using Transfer Learning. arXiv:2106.11051 [cs].
- Ashayeri, C., Jha, B., 2021. Evaluation of transfer learning in data-driven methods in the assessment of unconventional resources. Journal of Petroleum Science and Engineering. 207. 109178.
- Canada Energy Regulator, 2018. "Market Snapshot: Evolving technology is a key driver of performance in modern gas wells: a look at the Montney Formation, one of North America's biggest gas resources." last modified Jan 29, 2021, accessed Mar 20, 2021.
- Ki, S., Seo, J., Kwon, O., Jang, I., 2019. Prediction of Missing Tubing Head Pressure Using Recurrent Neural Network. Journal of the Korean Society of Mineral and Energy Resources Engineers. 56(5). pp. 416-426.
- Lee, K., Lim, J., Yoon, D., Jung, H., 2019. Prediction of Shale-Gas Production at Duvernay Formation Using Deep-Learning Algorithm, Society of Petroleum Engineers. 24. pp. 2423-2437.
- Mishra, S., 2012. A New Approach to Reserves Estimation in Shale Gas Reservoirs Using Multiple Decline Curve Analysis Models. Presented at SPE Eastern Regional Meeting, 3-5 October, Lexington, Kentucky, USA. SPE 164092.
- National Energy Board, Alberta Energy Regulator. British Columbia Oil & Gas Commission, British Columbia Ministry of Natural Gas Development, 2013. The Ultimate Potential for Unconventional Petroleum from the Montney Formation of British Columbia and Alberta-Energy Briefing Note.
- "The Current State of Canadian Unconventional Oil and Gas Plays." THEWAYHEAD. last modified Jul 26, 2018, accessed Mar 17, 2021, <https://jpt.spe.org/twa/current-state-canadian-unconventional-oil-and-gas-plays>.
- Vishkai, M., Wang, J., Wong, R.C.K., Clarkson, C.R., Gates, I.D., 2017. Modeling geomechanical properties in the montney formation, Alberta, Canada. International Journal of Rock Mechanics and Mining Sciences. 96. pp. 94-105.
- Yang, S., Lim, J., 2013. Shale Gas Production Rate Forecasting and Type Curve Estimation by Decline Curve Analysis of Horn River Basin in Canada. Journal of the Korean Society of Mineral and Energy Resources Engineers. 50(4). pp. 502-511.
- Yoon, S., Joe, Y., 2019. Sedimentary facies and depositional environments of the Montney formation in Alberta, Canada: Sedimentological characterization for a shale gas formation. Journal of the Geological Society of Korea. 55(3). pp. 277-289.
- Zhang, Y., Hu, J., 2021. Application of Locality Preserving Projection-Based Unsupervised Learning in Predicting the Oil Production for Low-Permeability Reservoirs. Society of Petroleum Engineers. 26(3). pp. 1302-1313.