



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

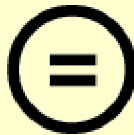
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

머신러닝을 활용한 셰일층 석유 생산 주요인자 기반 생산성 분석 연구

Productivity Analysis Based on Key Factors of Oil Production
in Shale Formation using Machine Learning



지도교수 임종세

2020년 2월

한국해양대학교 대학원

해양에너지자원공학과

신 효 진

본 논문을 신호진의 공학박사 학위논문으로 인준함.

위원장 신 성 렬 (인)

위 원 임 종 세 (인)

위 원 정 우 근 (인)

위 원 김 광 염 (인)

위 원 장 일 식 (인)

2019년 12월

한국해양대학교 대학원

목 차

List of Tables	iii
List of Figures	iv
Abstract	ix
제 1 장 서 론	1
제 2 장 세일 저류층에서의 생산성 예측	10
2.1 세일 저류층 평가방법	10
2.2 세일 저류층에서의 생산 예측 기법	13
2.3 확률론적 기법을 적용한 생산량 예측	21
제 3 장 데이터 기반 분석을 이용한 생산성 예측	24
3.1 머신러닝 기법	24
3.1.1 자료 특징 파악	27
3.1.2 비지도학습	30
3.1.3 지도학습	34
3.1.4 인공신경망 기법	40
3.2 석유 생산 관련 인자	43
3.3 연구 사례 분석	48
제 4 장 기존 생산정의 주요인자를 활용한 생산특성 분석	55
4.1 연구대상 지역	55
4.2 세일층의 석유 생산 주요인자 파악	58

4.2.1 자료 전처리 과정	58
4.2.2 생산량 자료 분석	61
4.2.3 생산 관련 인자 간 관계 파악	66
4.2.4 차원 감소 및 주요인자 파악	70
4.3 주요인자를 활용한 생산특성 분석	79
4.3.1 그룹 분류를 위한 군집분석	81
4.3.2 그룹별 생산특성 및 특징 분석	85
4.4 분류된 그룹을 활용한 누적 생산량 추정	89
 제 5 장 새로운 유정의 생산량 예측	93
5.1 그룹별 특징 및 확률론적 입력값 추출	94
5.2 새로운 유정에 대한 그룹 분류	98
5.3 새로운 유정의 누적 생산량 예측	102
5.3.1 누적 생산량 예측을 위한 인공신경망 모델 개발	102
5.3.2 생산량 예측 모델의 검증	104
5.3.3 확률론적 값을 이용한 생산량 예측	105
 제 6 장 결 론	111
 참고문헌	114
 맺음말	123

List of Tables

Table 1	Tools used to estimate reserves	12
Table 2	Flow rate decline analysis	13
Table 3	Comparison of the original and modified RTA methods	15
Table 4	Arps equation	17
Table 5	Research on productivity prediction of existing wells in shale reservoir ..	51
Table 6	Research on productivity prediction of new wells in shale reservoir	54
Table 7	List of input attributes in this study	66
Table 8	Results of PCA in Case 1	71
Table 9	Results of PCA in Case 2	73
Table 10	Results of PCA in Case 3	75
Table 11	Results of PCA in Case 4	76
Table 12	Summary of fuzzy c-means clustering	82
Table 13	Mean value of key factors for each group	87
Table 14	Mean value of production on the effect of key factors by group	87
Table 15	AAPE of estimated production results by group	92
Table 16	Probability value of DCA parameters by group	97
Table 17	Case study for input attributes	99

List of Figures

Fig. 1 Primary energy consumption by fuel	1
Fig. 2 Worldwide map of unconventional resources	2
Fig. 3 U.S. oil and gas production from 2018 to 2050	2
Fig. 4 Average production profiles for shale wells	3
Fig. 5 2-D schematic of multi stage hydraulically fractured horizontal well in shale reservoir	4
Fig. 6 U.S. tight oil and shale gas production and well counts	5
Fig. 7 Global hydraulic fracturing market size and forecast	5
Fig. 8 U.S. shale average breakeven oil price	6
Fig. 9 Data-driven analytics workflow	6
Fig. 10 Workflow of this study	9
Fig. 11 Flow chart and generalized division of resource and reserve categories ...	11
Fig. 12 Flow regime in multi-fractured horizontal wells	14
Fig. 13 Variation of production behavior depending on Arps parameters	18
Fig. 14 Changing decline exponent in Arps hyperbolic	19
Fig. 15 Effect of increasing data in Pinedale horizontal wells production forecast ..	20
Fig. 16 Monte Carlo simulation	22
Fig. 17 Example of machine learning process	24
Fig. 18 Machine learning algorithm	26
Fig. 19 Example of applying PCA	29
Fig. 20 Membership function for clustering of a one dimensional dataset	31
Fig. 21 Example of incorrect cluster analysis	33
Fig. 22 Data classification using discriminant analysis	35
Fig. 23 Illustrative example of distance	37

Fig. 24	Example of separating class using SVM	38
Fig. 25	Typical neural network representation of neuron	40
Fig. 26	The sigmoid function defined	41
Fig. 27	List of native, design and dynamic parameters in shale reservoir	44
Fig. 28	Impact of lateral length and stage count	45
Fig. 29	Effect of reservoir and fracture conditions	46
Fig. 30	Cross plot of 30 days cumulative production in Marcellus shale	47
Fig. 31	Monthly count of DUCs	52
Fig. 32	Monthly U.S. oil and gas production	56
Fig. 33	Well permitted and completed in the Eagle Ford Shale	57
Fig. 34	Data pre-processing for machine learning	59
Fig. 35	Example of histogram to remove data	60
Fig. 36	Production history in study area	62
Fig. 37	Change of decline curve factors according to production period	63
Fig. 38	Example of production behavior when applying Arps hyperbolic	64
Fig. 39	Histogram of DCA parameters after pre-processing	65
Fig. 40	Highly correlated input attributes	67
Fig. 41	Scatter plot of input attributes	68
Fig. 42	Correlation coefficient between input attributes	69
Fig. 43	Scree plot of 16 attributes in Case 1	72
Fig. 44	Scree plot of 12 attributes in Case 2	74
Fig. 45	Scree plot of 8 attributes in Case 3	75
Fig. 46	Scree plot of 5 attributes in Case 4	77
Fig. 47	Visualization of PCs in 3D plot in Case 4	78
Fig. 48	Variable distribution of the five clusters based on 6months production	79
Fig. 49	Variable distribution of the four clusters based on quartile	80
Fig. 50	Convergence of objective functions in fuzzy c-means clustering	82
Fig. 51	Cluster analysis results	83

Fig. 52 Location map by cluster	84
Fig. 53 Normalized Production of each group	82
Fig. 54 Box plot of key factors and production by group	88
Fig. 55 Estimated production results of the total wells using Arps hyperbolic with 12 months data	90
Fig. 56 Relationship between NP 6 and NP 48 by group	91
Fig. 57 Relationship between NP 12 and NP 48 by group	92
Fig. 58 Box plot of key factors by group	95
Fig. 59 Cumulative probability of DCA parameters by group	96
Fig. 60 Accuracy for testing wells by methods	99
Fig. 61 Accuracy for new wells by methods	101
Fig. 62 Classification results of group by methods	101
Fig. 63 Structure of ANN model in this study	103
Fig. 64 Comparison of measured and predicted production for testing wells	104
Fig. 65 Comparison of measured and predicted production for new wells	106
Fig. 66 Comparison of measured data and probabilistic prediction for new wells ..	107
Fig. 67 Probability distribution of DCA parameters with 48 months data	109
Fig. 68 Estimation results of total wells	110

머신러닝을 활용한 셰일층 석유 생산 주요인자 기반 생산성 분석 연구

신 호 진

한국해양대학교 대학원
해양에너지자원공학과

초 록

대규모 연속체로 부존되어 있는 셰일 저류층은 불균질(heterogeneous)하고 낮은 유체투과도(permeability)를 가지므로 개발을 위해 수평시추(horizontal drilling), 수압파쇄(hydraulic fracturing) 기술을 적용한다. 이로 인하여 유동 메커니즘(mechanism)의 파악에 어려움이 있어 기존의 생산 예측 기법 적용에 한계가 존재하며, 생산 변동성이 큰 셰일층에서는 단일정을 이용한 생산량 예측보다 다수 생산정을 기반으로 생산거동 및 궁극가채량(estimated ultimate recovery)에 대한 추정이 이루어진다. 이와 더불어 석유 E&P 산업에 존재하는 다양하고 많은 양의 자료에 대한 효율적인 분석을 위해 데이터 기반 분석이 수행되고 있으며, 이를 활용하여 생산 운영에 필요한 자료를 도출할 수 있다. 생산 관련 인자에는 셰일층의 특성(native parameters), 수압파쇄 설계인자(design parameters), 동적인자(dynamic parameters)가 있으며, 생산성과 일정한 경향이 나타나지 않는 인자들의 복합적인 영향에 대한 분석이 필요하다. 따라서 다양한 변수와 생산량과의 관계를 파악하고 이로 인해 발생하는 생산특성을 파악하여야 한다. 이 연구에서는 머신러닝을 적용하여 셰일층의 석유 생산 주요인자(key factors)를 파악하고 이에 따른 생산특성을 기존

생산정에 대하여 분석하였으며, 그 결과를 활용하여 새로운 유정에서의 생산량을 예측하고자 하였다. 이를 위해 기존의 다수 오일 생산정의 제한된 현장 자료에서 유의미한 속성 값을 추출하고자 극단값 제거, 단위길이 당 인자 값 도출 등의 전처리 과정을 수행하였다. 취득한 다양한 인자에 대하여 차원 감소와 주요인자를 파악하고자 상관관계를 분석하였으며, 입력 속성에 따른 주성분분석을 수행함으로써 연구대상 지역에서 생산정 깊이와 수압파쇄 설계인자가 주요인자임을 파악하였다. 이를 바탕으로 하나의 개체가 여러 그룹에 속할 가능성으로 표현되는 퍼지군집분석을 적용하여 3개의 그룹으로 분류하였으며, 수압파쇄 설계 인자의 영향에 대한 생산특성을 분석하였다. 그 결과 생산성이 높은 그룹은 수압파쇄 간격이 좁으며 파쇄 시 단위길이 당 주입 프로판트 및 유체의 양이 생산정 중 많은 편에 속해 이와 같은 현장 조건에서 최적의 생산이 가능할 것으로 판단된다. 기존 생산정 분석으로 도출한 그룹을 바탕으로 분류 알고리즘을 통해 새로운 유정에 대한 그룹 예측을 수행하였으며, 제한된 주요인자를 활용하여 생산특성이 다른 그룹 분류가 가능함을 확인하였다. 또한, 새로운 유정의 생산량 예측 시 신뢰성을 향상시키고자 통계적 분석으로부터 각 그룹의 특징을 파악하였으며, 입력 인자에 대한 불확실성 고려를 위해 확률론적 입력값을 도출하였다. 이를 통해 생산정 깊이, 수압파쇄 설계인자, 감퇴곡선인자를 입력인자로 구축한 인공신경망 모델을 활용하여 생산 초기와 4년의 시점에서의 누적 생산량을 예측하였다. 이 연구에서 제시한 기존 및 새로운 셰일 수평정에 대한 생산량 예측은 머신러닝 기반의 시행착오를 통한 분석으로 새로운 현장 자료에 대해서도 적용가능성이 높으며, 광범위한 변수 속에서 자료 운용이 제한적인 경우에도 활용할 수 있다. 또한, 셰일층 개발 시 추가적인 자료 취득으로 인한 시간 및 비용 손실 없이 분석을 수행할 수 있으므로 보유하고 있는 자료를 활용하여 생산 운영에 필요한 자료를 도출할 수 있을 것으로 사료된다.

주요어: 셰일층; 주요인자; 생산특성; 생산성 예측; 머신러닝

Productivity Analysis Based on Key Factors of Oil Production in Shale Formation using Machine Learning

Hyo-Jin Shin

Department of Ocean Energy & Resources Engineering
Graduate School of Korea Maritime & Ocean University

Abstract

Shale reservoirs, which exist in large-scale continuum, are heterogeneous and have very low permeability, thus applying techniques such as horizontal drilling and hydraulic fracturing for development is needed. As a result, it is difficult to understand the flow mechanism, which limits the application of conventional techniques. In the shale formation with high production variability, the production behavior and estimated ultimate recovery is predicted based on multi-wells rather than a single well. As data-driven analytics is being conducted to analyze various and large amounts of data efficiently in the petroleum E&P industry, it can be used to obtain data for production operations. Production-related factors include native parameters, design parameters and dynamic parameters. An analysis of the complex influence is needed for factors that do not exhibit uniform trends with productivity. Therefore, it is necessary to understand the relationship between the various factors and productivity, and the production characteristics of results should be identified. In this study, machine learning was used to define key factors of the shale formation, to analyze the production characteristics according to the key factors and predict production in new wells based on the results of

existing wells. For this purpose, the pre-processing was performed to extract data with high priority from the limited field data of existing oil wells, and to remove outlier and the normalized value per unit length. In order to identify the dimensional reduction and the key factors of the acquired data, correlations were analyzed, and principal component analysis depending on the input attributes was performed to determine the true vertical depth of production well and the fracturing design factor were the key factors in the study area. Based on this, fuzzy cluster analysis, which represents the possibility of one individual to be involved to several clusters, was applied to classify the data into three groups. As an analysis result of production characteristics on the influence of design factors, the group with high productivity had a narrow spacing, and a large amount of injection proppant and fluid per lateral length when fracturing. It is believed that optimal production will be possible under these field conditions of the group. Using the results of group analyzing the existing wells, the classification algorithm of machine learning was performed to predict group of the new wells. Limited key factors were used to identify groups with different production characteristics. In addition, the characteristics of each group were identified from the statistical analysis to improve the reliability of new well productivity prediction, and probabilistic inputs were derived to consider the uncertainty of the input factors. Using these values, the artificial neural network was constructed with the true vertical depth, fracturing design and decline curve factor as input attributes, and the cumulative production was predicted at the beginning of production and four years. The productivity prediction for the existing and new shale horizontal wells suggested in this study are applicable to new data based on machine learning through trial and error, and could be used even with the limited data among the inclusive data. As shale formation development can be performed without the time and cost loss for additional data acquisition, it could be possible to obtain the data for production operations from database.

KEY WORDS: Shale Formation; Key Factors; Production Characteristics; Productivity Prediction; Machine Learning

제 1 장 서 론

세계 에너지 수요는 지속적으로 증가하는 추세이며, 전체 1차에너지 소비량 중 유·가스의 비중이 가장 높다(Fig. 1). 이 중 막대한 부존량과 생산기술의 발달로 비전통자원(unconventional resources)의 개발 비중이 커지고 있으며, Fig. 2와 같이 전 세계적으로 분포하여 북미를 중심으로 개발이 진행되고 있다.

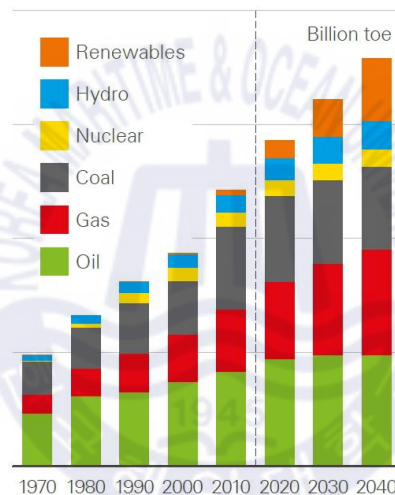


Fig. 1 Primary energy consumption by fuel (BP, 2019)

이러한 비전통자원은 대규모 연속체로 부존이 확인되어 상대적으로 탐사 위험도는 작으나, 회수 기술 및 생산성에 따라 경제성 여부가 결정되고 시장가격과 회수 기술 개발에 따라 상업적 가치가 변동될 수 있다(Sung et al., 2009). 비전통자원의 종류는 회수 기술 난이도 및 개발 비용에 따라 분류되며, 초중질유(extra heavy oil), 비튜멘(bitumen), 오일샌드(oil sand), 치밀오일(tight oil), 치밀가스(tight gas), 셰일오일(shale oil), 셰일가스(shale gas) 등이 포함된다. 그 중 치밀·셰일 가스 및 오일의 생산이 증가하고 있으며, 앞으로도 많은 양을 차지할 전망이다(Fig. 3).

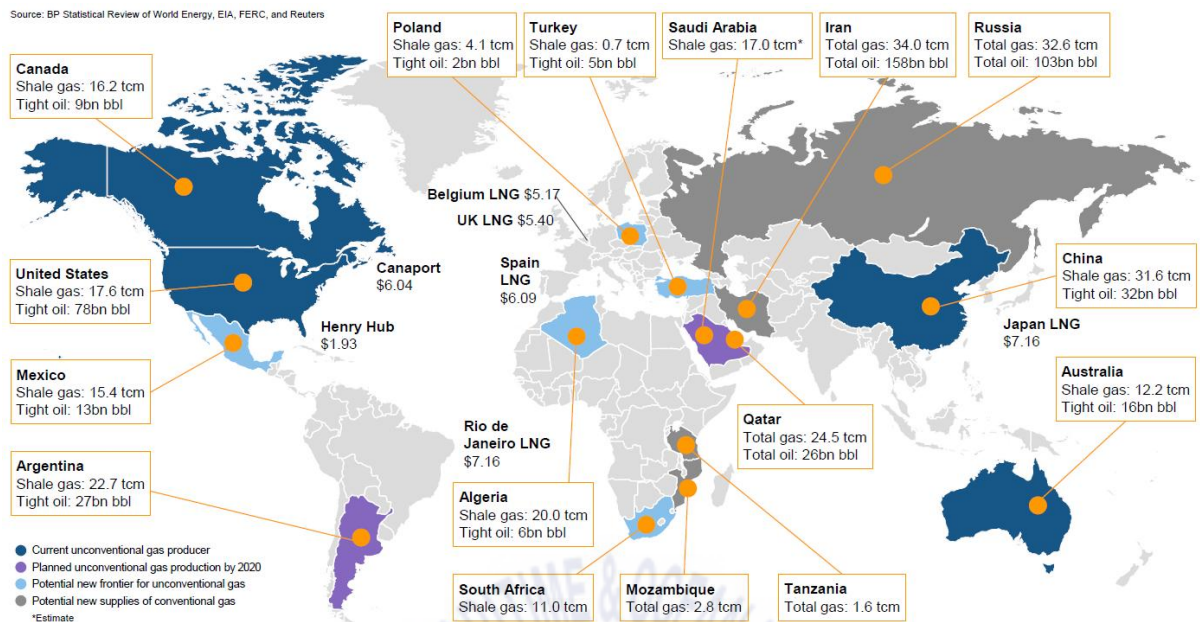


Fig. 2 Worldwide map of unconventional resources (WEC, 2016)

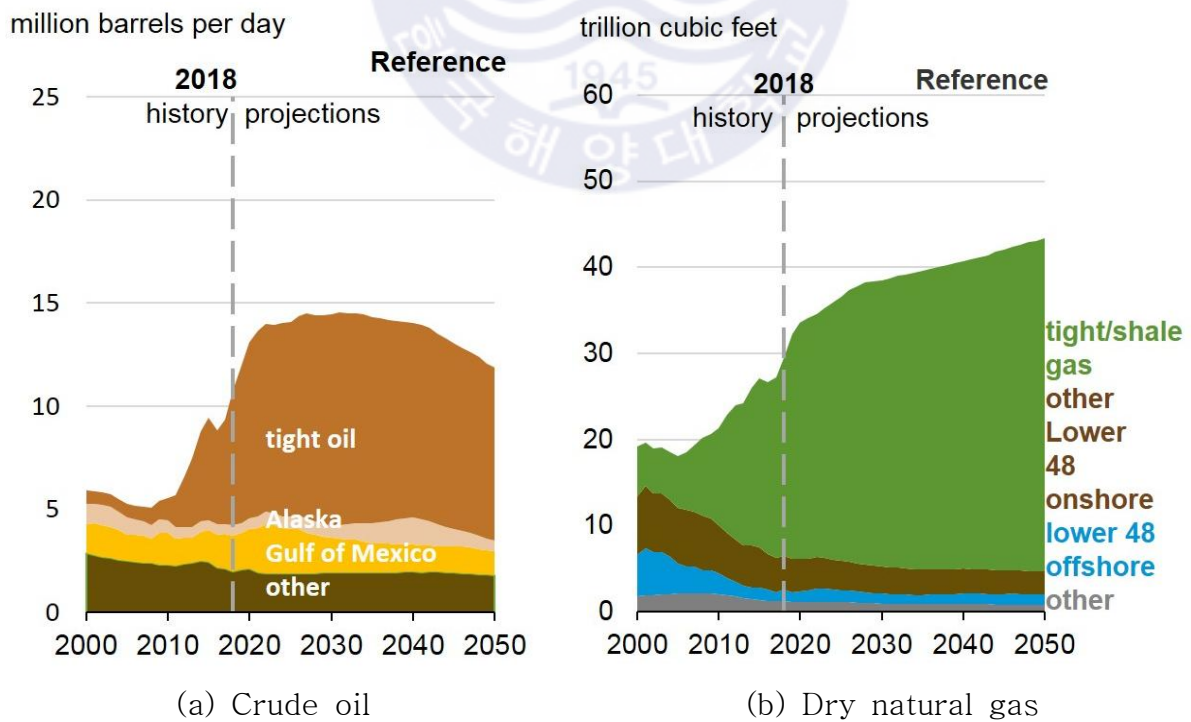


Fig. 3 U.S. oil and gas production from 2018 to 2050 (EIA, 2019a)

비전통자원 중 셰일층 개발을 위해서는 지질학적인 특성과 생산경향에 대한 이해가 필요하다. 특정 지질구조에 집중적으로 축적된 일반적인 석유자원(conventional resources)과 달리 셰일층은 넓은 지역에 걸쳐 연속적인 형태로 부존되어 있으며, 동일 분지라도 수직 및 수평에 따라 특성 변화가 심하다. 셰일 저류층은 근원암과 저류암이 동일하고 일반적으로 생산 시작 후 공극에 존재하는 자유가스(free gas)가 먼저 생산되며, 흡착가스(adsorbed gas)는 케로젠(kerogen)으로부터 탈착이 되기까지 일정 압력 강하가 필요하므로 후반에 생산이 이루어진다. 공극률은 일반적으로 6~12%, 유체투과도는 nano-darcy 정도로 매우 낮아 생산을 위해 수압파쇄(hydraulic fracturing)를 반드시 적용하여야 한다(Zendehboudi and Bahadori, 2017). 이로 인해 초기 생산과 감퇴율이 커서 Fig. 4와 같이 1~5년 사이에서 주된 생산 감퇴가 일어나며, 이후로는 낮은 생산량으로 지속되게 된다.

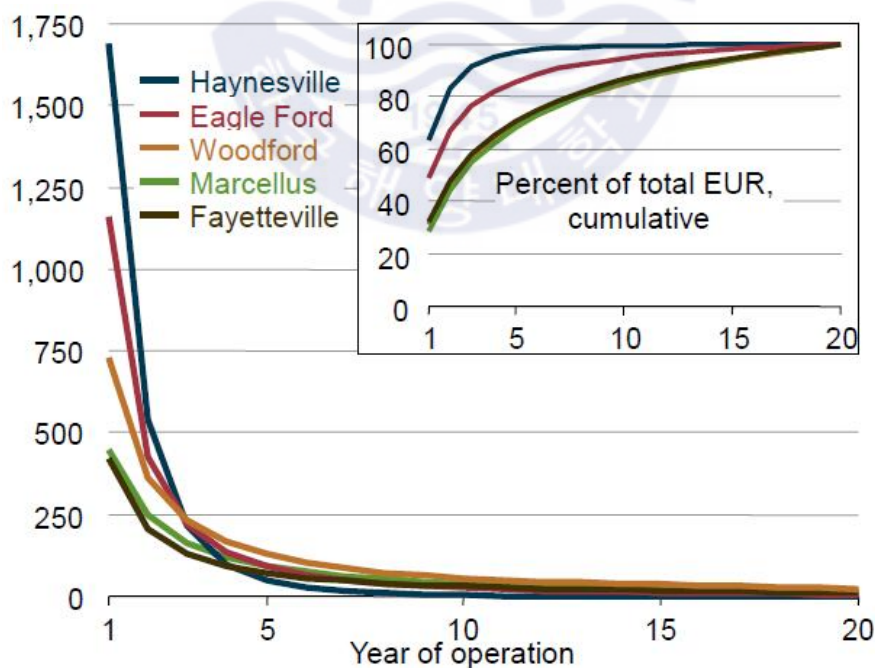


Fig. 4 Average production profiles for shale wells (EIA, 2013)

셰일층 개발을 위한 기술 중 수평정 시추(horizontal well drilling)는 Fig. 5와 같이 지표에서 수직방향으로 시추 후 저류층에 교차하도록 수평하게 유정을 시추 및 완결하는 기술이다. 대부분의 석유 저류층은 광범위한 수직 균열망을 형성하고 있으므로 저류층에 수평하게 시추하여 더 많은 저류암을 시추공 표면에 노출시켜 생산 효율을 향상시키게 된다. 이러한 수평정에 대해 일정 간격, 주입압 등을 설정하여 수압파쇄를 적용하게 되며, 이는 유체투과도가 낮은 저류층의 생산성을 높일 수 있다. 고압의 파쇄유체(fracture fluid)를 대상 저류층에 주입하여 인공균열을 발생시킴으로써 유동이 이루어지게 되며, 균열의 간극(aperture)은 저류층 상부 지층압에 의해 다시 작아질 수 있어 파쇄유체와 함께 균열지지체인 프로판트(proppant)를 주입한다. 프로판트는 직경이 커질수록 전도도(conductivity) 및 강도(strength)가 높아지는 경향을 보이며, 파쇄유체의 경우 지층과의 민감도, pH, 온도에 따른 유체의 거동 및 안정성을 고려하여 적합한 유체를 선정해야 한다(Shin et al., 2015).

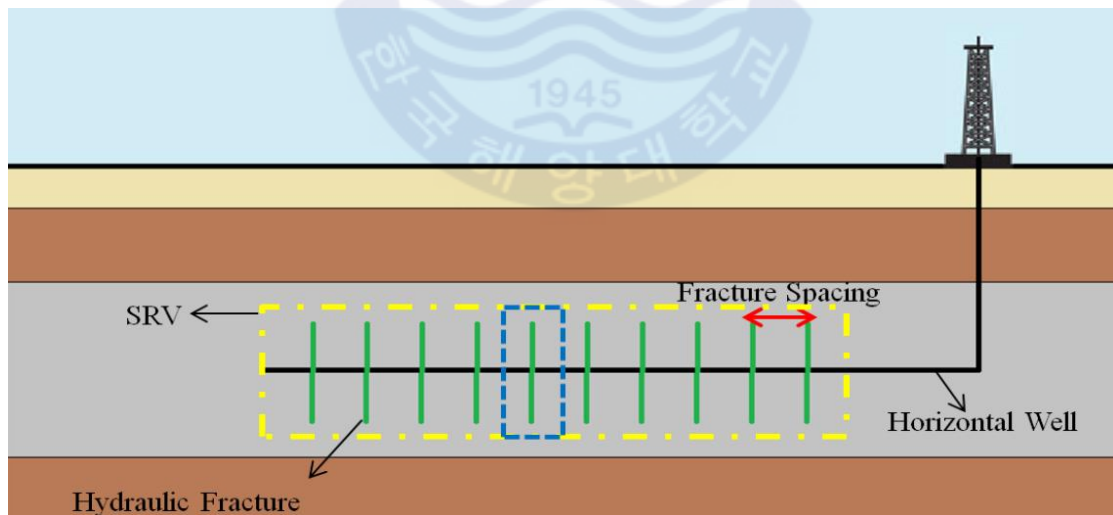


Fig. 5 2-D schematic of multi stage hydraulically fractured horizontal well in shale reservoir (Zanganeh et al., 2014)

이렇듯 연속적으로 분포된 셰일층에서 대규모의 유체 생산을 위하여 수평정 시추와 수압파쇄 등의 생산기술이 발달하였으며, 수평정의 개수와 효율적인 생산을 위한 수압파쇄 시장 역시 계속해서 규모가 커질 전망이다(Fig. 6; Fig. 7). 이와 같은 기술 향상에 따라 개발 비용이 낮아지고 상업적 개발 성공률이 높아지고 있으며, 셰일층에서 유·가스 생산 시 일정 기간 수익과 비용이 같아지는 가격(break-even price)이 점차 감소하는 추세이다(Fig. 8).

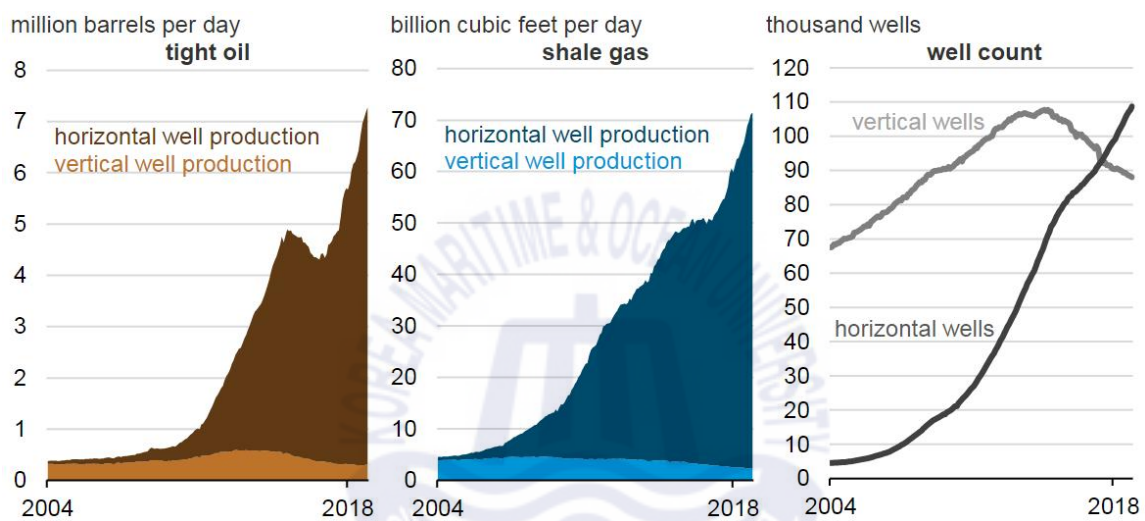


Fig. 6 U.S. tight oil and shale gas production and well counts (EIA, 2019b)



Fig. 7 Global hydraulic fracturing market size and forecast (CMR, 2017)



Fig. 8 U.S. shale average breakeven oil price (Mandel, 2017)

최근 들어 석유 E&P 산업에서는 데이터 기반 분석(data-driven analytics)을 이용한 차별화를 통해 개발 및 생산을 진행하고 있으며, 이는 중요한 경쟁 기술이 되고 있다. 비전통자원에 대한 생산, 특히 셰일과 관련하여 기업들은 지난 몇 년 동안의 수집을 통해 다양하고 많은 양의 자료를 보유하고 있으며, 이를 활용하여 생산 예측 뿐 아니라 탐사 위험도, 생산 시나리오 개발, 시추·완결 및 생산 운영 과정 등에 대해 Fig. 9와 같은 일련의 과정으로부터 최적화를 수행할 수 있다.



Fig. 9 Data-driven analytics workflow (Dykes, 2019)

일반적으로 셰일 저류층에서의 생산 예측은 저류층 시뮬레이션(reservoir simulation), 생산천이분석(Rate Transient Analysis; RTA), 생산감퇴곡선분석(Decline Curve Analysis; DCA) 등과 같은 기법을 적용하고 있다. 그러나 셰일층 생산 시 유동 메커니즘(mechanism)이 불분명하며, 기존 분석기법은 입력 변수 설정, 생산 경향 예측 등을 수행함에 있어 주관적인 판단과 많은 가정으로 인해 불확실성과 한계점이 존재하게 된다. 따라서 자료에 대한 분석 혹은 모델링 수행 시 기존의 분석 기법에 대한 한계점 극복과 결과 분석에 필요한 통찰력을 제공할 수 있도록 인공지능(artificial intelligence) 및 머신러닝(machine learning)을 활용한 연구가 수행되고 있다.

셰일층의 생산성 예측에 있어 데이터 기반 분석의 적용성을 확인하기 위한 다양한 연구가 진행되고 있으며, 단일 생산정에 대하여 시간에 따른 생산량 예측을 위해 DCA와 인공신경망(artificial neural network; ANN) 예측 결과를 비교·분석하여 ANN 적용 결과의 우수성을 확인하였다(Cao et al., 2016; Suhag et al., 2017). 또한, 순환신경망(recurrent neural network)을 이용하여 생산량을 예측함으로써 실제 생산량과의 비교를 수행한 바 있다(Sun et al., 2018; Klie and Florez, 2019). 그러나 넓은 지역에서 불균질하게 분포되어 있는 셰일층의 경우 다수의 생산정에 대한 분석이 필요하며, 이를 위해 셰일 저류층 및 암석역학적 물성, 수압파쇄 설계인자와 생산량과의 상관관계를 파악하고 패턴인식을 통해 영향 정도를 평가한 연구가 수행되었다(Mohaghegh, 2017). 또한, 셰일층에서 취득한 다양한 자료를 기반으로 각 인자와 생산성의 영향을 파악하기 위한 연구가 수행되었으며(Gaurav, 2017), 주요인자 파악을 바탕으로 생산량 예측을 위해 머신러닝 기법을 적용하여 감퇴곡선인자 혹은 이로부터 도출된 궁극가채량을 추정하였다(Alabboodi and Mohaghegh, 2016; Li and Han, 2017; He, 2017; Vyas et al., 2017; Bowie, 2018). 이외에도 수압파쇄 인자 혹은 수압파쇄 시 역학적 물성 영향 평가를 위해 머신러닝을 적용한 바 있다(BuKhamseen and Ertekin, 2017; Tandon, 2019).

이와 같이 셰일층에 대하여 기존 생산정에 대한 생산 경향 및 누적 생산량 예측과 새로운 유정에 대한 생산량 예측 연구가 수행되고 있다. 궁극적으로 새로운 지역의 경우, 어느 지역에 시추할 것인지, 어떠한 수압파쇄 설계를 적용할 것인지, 그리고 이로부터 얼마만큼의 생산량이 나올 것인지를 파악하는 것이 필요하다. 일반적으로 새로운 지역에 대한 개발 시 인근 생산정(neighbor wells) 정보를 바탕으로 개발계획을 수립하며, 이와 관련된 생산성 예측 연구는 시작 단계이다. 새로운 유정에서도 기존 생산정과 유사한 입력인자를 통해 머신러닝 기법을 적용하여 생산거동 및 6개월의 누적 생산량을 예측한 바 있다(Mohaghegh et al., 2017). 또한, 새로운 지역에 대한 제한된 정보로 인해 인근 생산정의 자료를 활용하여 예측 모델을 개발하였으나(Amr et al., 2018), 저류층 특성을 고려하였음에도 불구하고 불균질한 셰일층의 영향으로 결과에 대한 불확실성이 존재한다.

따라서 이 연구에서는 기존 및 새로운 셰일 수평정의 생산량 예측을 위하여 우선적으로 기존의 다수 생산정에 대해 유의미한 입력인자를 도출하고, 서로 상관되어 있는 반응 변수로부터 주요인자를 파악하고자 한다. 이로부터 생산량과 주요인자와의 관계를 파악하고자 하며, 생산 예측의 신뢰성 향상을 위해 생산특성이 다른 그룹을 분류하고 각 그룹의 추세선을 바탕으로 기존 생산정의 누적 생산량을 추정하고자 한다. 또한, 새로운 다수 유정에 대하여 주요인자의 제한된 입력속성만으로 그룹을 분류할 수 있는 모델을 개발하고, 분류된 그룹의 확률론적 입력값을 활용하여 불확실성을 고려한 누적 생산량 예측을 수행하고자 한다(Fig. 10).

2장에서는 일반적인 석유자원과 다른 특성을 가지는 셰일 저류층에 대하여 석유 생산성 예측 방법에 대한 기존의 분석 기법과 개선한 방법에 대해 서술하였다.

3장에서는 데이터 기반 분석 방법과 셰일층의 생산 관련 인자의 종류 및 특징을 요약 설명하였으며, 이를 활용하여 연구를 수행한 기존 연구결과들에 대해 서술하였다.

4장에서는 기존 생산정에 대한 데이터 기반 분석을 통해 연구대상 지역의 셰일층 석유 생산 주요인자를 파악하고 생산특성이 다른 그룹의 분류 방법과 결과를 설명하였으며, 분류된 그룹으로부터 추정한 생산량 결과를 서술하였다.

5장에서는 새로운 유정에 대한 생산량 예측을 위해 기존 생산정 분석 결과를 바탕으로 그룹을 분류하고 확률론적 입력 값을 도출하는 방법 및 결과를 서술하였으며, 단위길이 당 누적 생산량 예측 결과에 대해 설명하였다.

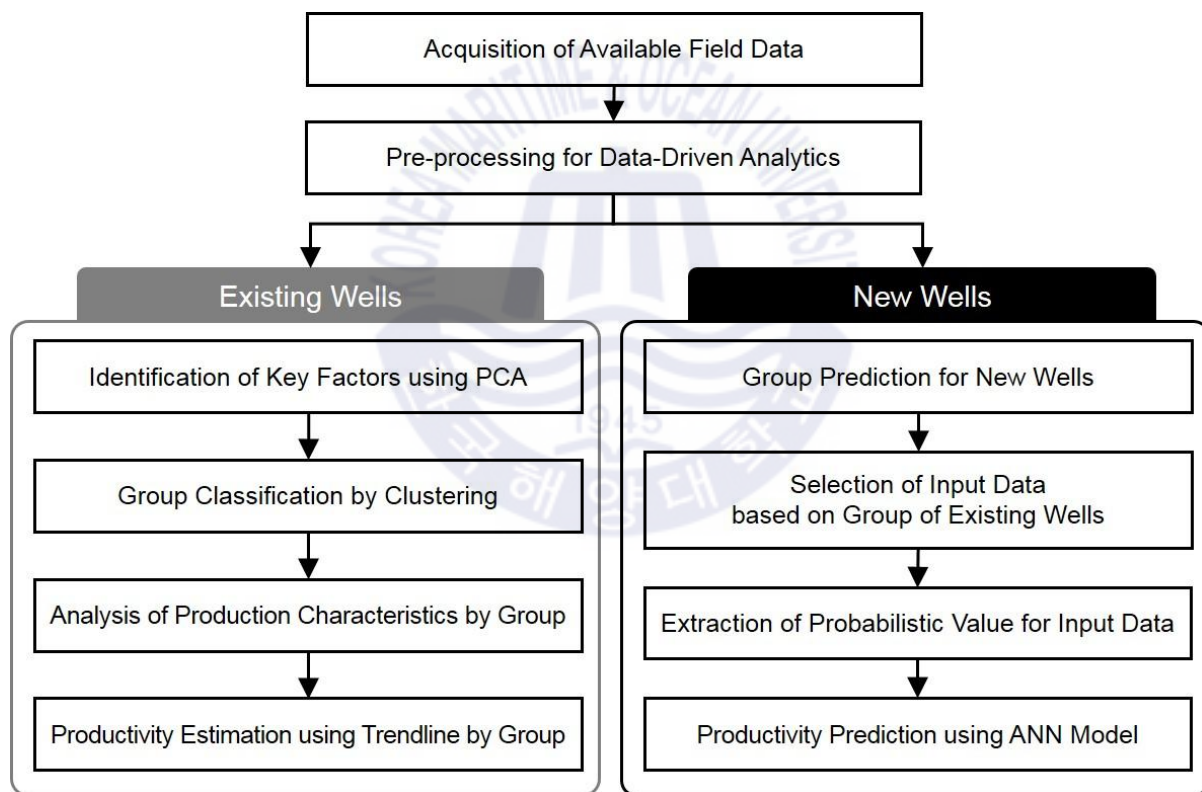


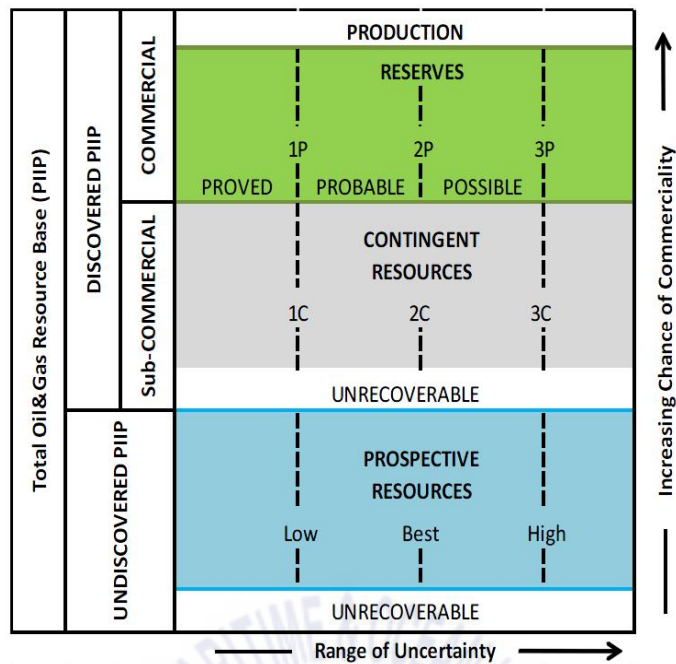
Fig. 10 Workflow of this study

제 2 장 셰일 저류층에서의 생산성 예측

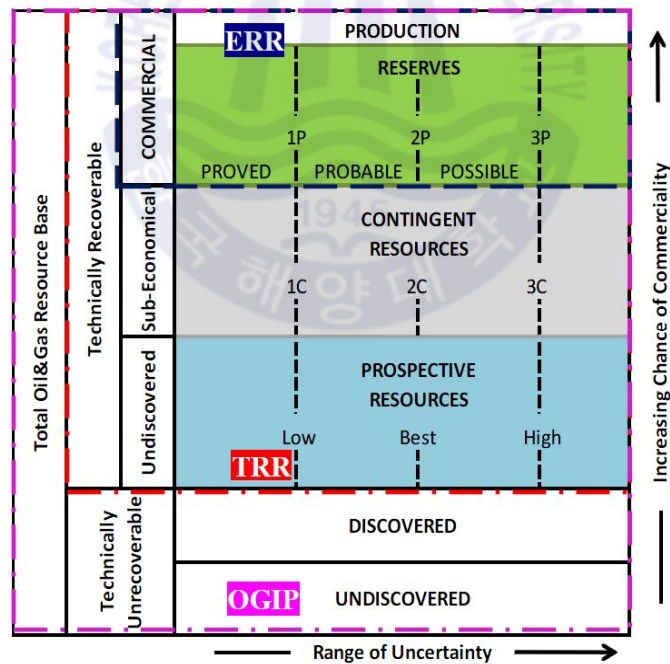
2.1 셰일 저류층 평가방법

비전통자원량은 일반적으로 시추를 통해 부존을 확인한 상태에서 시장 판매가격 및 회수기술의 난이도에 따라 평가를 수행하며, 일반적인 석유자원의 분류체계를 일부 변형하여 Fig. 11과 같이 ERR(economically recoverable resources), TRR(technically recoverable resources)라는 개념을 이용하여 구분한다. 이렇듯 비전통자원량을 매장량으로 평가하기 위해서는 파일럿 시험에 근거하여 상업성 판단과 회수 가능성이 입증되어야 한다(Sung et al., 2009).

비전통자원량은 일반적으로 궁극가채량과 혼용하여 사용하며, 탐사 성공률, 부존량, 생산거동 변수, 개발비용, 광구권 매입전략 등을 종합적으로 분석하여 결정한다(MKE, 2009). 자원량 평가 방식에는 유추법(analogy), 용적법(volumetric), 물질평형법(material balance method), 저류층 시뮬레이션, DCA가 있으며, 일반적인 석유자원과 비전통자원의 차이를 Table 1에 나타내었다. 비전통자원은 파일럿 시험, DCA, 궁극가채량 추정법을 통해 평가를 수행하며, 전체 저류층의 부존량 산출이 중요한 일반적인 저류층과 달리 셰일층과 같은 경우 생산정 당 회수량이 상업성에 큰 영향을 미침을 알 수 있다.



(a) Resource classification of PRMS



(b) EIA definitions mapped to PRMS

Fig. 11 Flow chart and generalized division of resource and reserve categories (Dong et al., 2013)

Table 1 Tools used to estimate reserves (Rezaee, 2015)

Method	Conventional reservoir	Unconventional reservoir
Analogy	• Can be applied in both conventional and unconventional assets	
Volumetric method	• Accurate in blanket reservoir	• Used only when no wells have been drilled
Material balance	• Accurate in depletion drive reservoir	• Should never be used • Average pressure cannot be measured accurately
Decline curve analysis	• Hyperbolic (small b) or exponential decline usually accurate	Must use hyperbolic decline: • CBM: $b=0-0.5$ • Shale and tight gas: b may be larger than 1 • Use best fit “ b ” until predetermined minimum decline rate reached, then impose exponential decline • Set “ b ” to proper “terminal value”
Reservoir simulation	• Used to simulate field	• Used to simulate individual wells

2.2 셰일 저류층에서의 생산 예측 기법

일반적으로 셰일 저류층에서의 생산 예측 시 저류층 시뮬레이션, RTA, DCA 등과 같은 기법을 적용하고 있다. 저류층 시뮬레이션은 셰일층의 유동특성과 자연 및 수압파쇄 균열을 고려한 모델 구축에 어려움이 있으며(Tan *et al.*, 2018), 생산 자료 분석 기법의 경우 낮은 유체투과도, 수평시추와 수압파쇄, 긴 천이유동(transient flow) 기간 등의 이유로 셰일 저류층 적용에 제한이 있다. RTA와 DCA의 차이는 Table 2와 같다. 이와 같은 방법은 입력 변수 설정, 생산 경향 예측 등을 수행함에 있어 주관적인 판단과 많은 가정으로 인해 불확실성과 한계점이 존재하게 된다.

Table 2 Flow rate decline analysis (Lacayo, 2013)

Decline Curve Analysis	Rate Transient Analysis
• Production rate only • Using historical trends to predict future • Empirical (curve fitting) • Based on analogy • Deliverables: Production forecast Recoverable reserves under current conditions	• Production rate and flowing pressures • Based on physics, not empirical • Reservoir signal extraction and characterization • Deliverables: OOIP/OGIP and skin Drainage area and shape Production optimization screening Infill potential

OOIP : Original Oil in Place

OGIP : Original Gas in Place

생산 자료로부터 저류층 및 균열의 특성, 부존량을 도출할 수 있는 RTA는 적용에 앞서 유동 영역(flow regime)을 구분하여야 한다. 유동 영역은 크게 이중선형(Bilinear)유동, 선형(Linear)유동, 경계영향유동(Boundary-Dominated Flow; BDF)으로 나뉘지며, Fig. 12와 같이 나타낼 수 있다. 이중선형 및 선형유동에는 균열 전도도 및 길이가 주로 영향을 미치며, 균열 끝단으로 흐르는 유동을 무시할 수 있고 각각의 균열은 서로 다른 균열에 발생하는 유동에 영향을 미치지 않는다. 생산이 진행될수록 점차 수평정에 수직방향으로 유동이 발생하며, 균열의 영향이 저류층에 영향을 미치게 된다. 그리고 저류층 경계에 도달하여 준정상상태(pseudo-steady state)의 유동 형태가 나타나는 BDF로 흐르게 된다.

세일층에 대한 분석 시 기존의 RTA에 균열 표면적(fracture surface area)에 대한 정보를 추가하면 정확도(accuracy)를 향상시킬 수 있으며, Wattenbarger et al.(1998)에 의해 제안된 방법이 일반적으로 사용된다. 이는 생산 감퇴로 인한 균열 표면적을 직선(straight line)을 이용하여 추정할 수 있다. 최근에는 Table 3과 같이 직선 분석 시 응력(stress)에 의존적인 암체(matrix)의 유체투과도를 고려하여 수정된 분석방법이 제안되었다(Soni et al., 2017). 그러나 RTA를 적용하기 위해서는 압력자료 등 분석에 필요한 입력 변수 취득에 있어 DCA에 비해 상대적으로 어려움이 따른다.

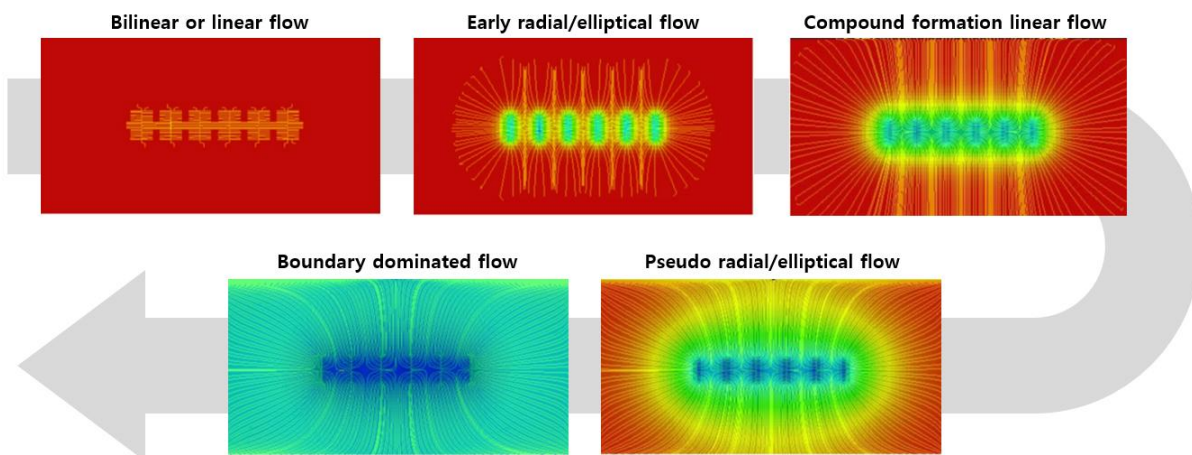


Fig. 12 Flow regime in multi-fractured horizontal wells (Lacayo, 2013)

Table 3 Comparison of the original and modified RTA methods

	Wattenbarger et al.(1998)	Modified by Soni et al.(2017)
$\frac{1}{q_D}$	$\frac{kh [m(p_i) - m(p_{wf})]}{1424 q_g T}$	$\frac{x_f h [m_k(p_i) - m_k(p_{wf})]}{2236 q_g L T}$
t_D	$\frac{0.00633 k}{(\phi \mu c_t)_i x_f^2} t$	$\frac{0.00633 k}{\phi_i L^2} t_{ap}$
$\frac{m(p)}{m_k(p)}$	$2 \int_{p_0}^p \frac{p}{z \mu} dp$	$2 \int_{p_b}^p \frac{k_g p}{z \mu} dp$
$\frac{t}{t_{ap}}$	$time$	$\int_0^t \frac{k_g}{\mu c_t} dt$
$\frac{m_{cp}}{m_{k_{cp}}}$	$\frac{315.4 T}{x_f h \sqrt{(\phi \mu c_t)_i}} \frac{1}{\Delta m(p) \sqrt{k}}$	$\frac{315.4 T}{x_f h \sqrt{\phi_i}} \frac{1}{\Delta m_k(p)}$
A_{Frac}	$f_{cp} \cdot \frac{1261.2 T}{\sqrt{(\phi \mu c_t)_i}} \cdot \frac{1}{\sqrt{k} m_{cp} \Delta m(p)} \cdot n$	$f_{cp} \cdot \frac{1261.2 T}{m_{cp} \sqrt{\phi_i}} \cdot \frac{1}{\Delta m_k(p)} \cdot n$

- A_{Frac} : total fracture surface area [acre] or [ft²]
 c_t : total compressibility of the shale formation [1/psia]
 f_{cp} : slope correction factor
 h : net formation thickness [ft]
 k : constant shale matrix permeability [md]
 k_g : dynamic shale matrix permeability [md]
 L : lateral length [ft]
 $m(p)$: real gas pseudo-pressure at pressure p [psia²/cp]
 $m_k(p)$: real gas pseudo-pressure with dynamic permeability [md psia²/cp]
 m_{cp} : slope of $1/q_g$ vs. $\sqrt{t}$ plot [D^{1/2}/Mscf]
 p : reservoir pressure [psia]
 p_{wf} : bottom-hole flowing pressure [psia]
 q_D : dimensionless flow rate
 q_g : gas flow rate [Mscf/d]
 T : absolute temperature [R]
 t_{ap} : real gas pseudo-time [md psia D/cp]
 t_D : dimensionless time
 x_f : fracture half-length [ft]
 z : gas deviation factor
 μ : gas viscosity [cp]
 ϕ : porosity [fraction]

DCA는 생산량(production rate) 자료만을 이용하여 생산 감퇴 거동을 파악하고 생산성을 예측하는 방법이며, 유정 주변의 물리적 변화가 심하지 않을 경우 점진적으로 변화하는 생산량은 저류층의 상태를 나타낸다. 이러한 양상으로 외삽(extrapolation)한 값은 잔존하는 석유자원의 양과 생산 수명 예측에 효과적으로 이용할 수 있다.

유·가스전 생산량 예측 시 일반적으로 활용하는 Arps(1945) 방정식은 Table 4와 같으며, 감퇴지수(decline exponent)에 따라 지수함수(exponential), 쌍곡선함수(hyperbolic), 조화함수(harmonic)의 세 가지 방법으로 구분할 수 있다. 지수함수적으로 감퇴하는 특성은 통상 균열 저류층 혹은 매우 불균질한 저류층에서 나타나며, 매장량을 과소평가할 수 있어 체적법이나 유추법을 이용한 검토가 필요하다. 조화함수의 감퇴경향은 수공법(water flood)을 적용하였거나 대수층 배압 생산메커니즘(water drive)으로 생산되는 저류층에서 발견되나 일반적인 특성은 아니다(KIGAM, 2002). 쌍곡선함수는 중력배유(gravity drainage)에 의해 생산되는 저류층에서 특히 많이 발견되었으며, 대부분의 일반적인 석유자원은 감퇴지수가 0~0.7 범위를 가지나 1보다 큰 값을 도출할 수 있는 셰일 저류층에서는 쌍곡선함수를 활용한다. 이를 보완하고자 Modified Hyperbolic(Seshadri and Matter, 2010), LGM(Clark et al., 2011), Modified SEPD(Yu et al., 2013), Duong(Joshi and Lee, 2013) 등의 다양한 기법이 제안되었다.

그러나 현장에서는 셰일 저류층의 생산거동을 예측하기 위해 Arps 방정식을 대부분 사용하며, 초기생산량(initial production rate)과 초기감퇴율(initial decline rate)이 크고 천이유동 구간이 길어지므로 감퇴지수 값이 1을 초과하게 되어 부존량을 과대 산정하거나 실제 유량의 거동을 반영하지 못하는 한계가 발생한다. 이러한 감퇴곡선인자 중 초기감퇴율은 값이 커질수록 초기생산의 감퇴가 급격히 일어나며, 감퇴지수 값은 커질수록 후반부로 생산이 진행될수록 생산거동이 점차 완만해 진다(Fig. 13).

Table 4 Arps equation

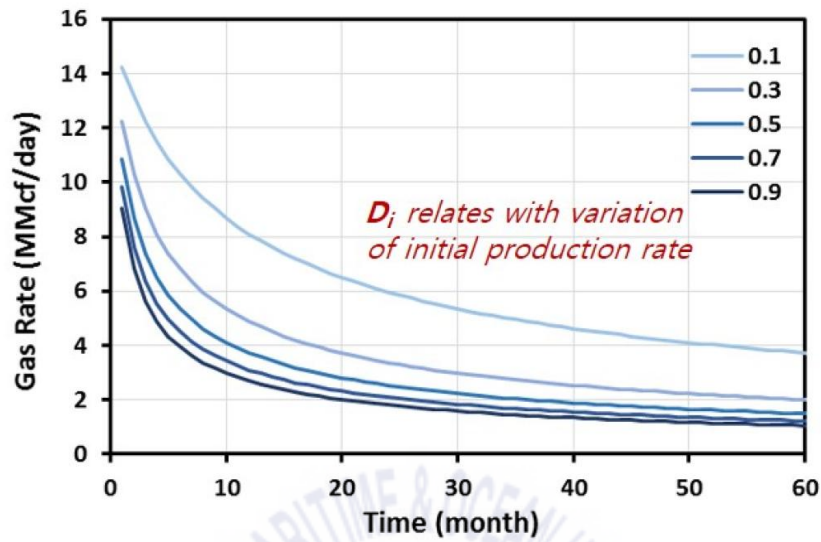
Case	b-value	Rate relation	Cumulative relation
Exponential	b=0	$q_t = q_i \exp(-D_i t)$	$N_p = q_i D_i [1 - \exp(-D_i t)]$
Hyperbolic	0 < b < 1	$q_t = \frac{q_i}{(1 + b D_i t)^{1/b}}$	$N_p = \frac{q_i}{(1-b)D_i} [1 - (1 + b D_i t)^{1-1/b}]$
Harmonic	b=1	$q_t = \frac{q_i}{(1 + b D_i t)}$	$N_p = \frac{q_i}{D_i} \ln(1 + D_i t)$
Beyond Hyperbolic	b ≥ 1	same as hyperbolic, used in shale reservoirs	

b : decline exponent

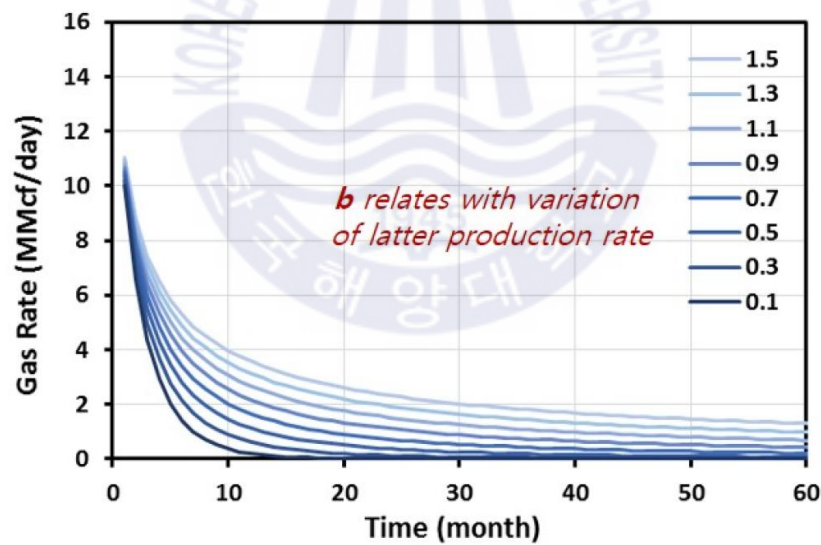
D_i : initial decline rate [%/month]

q_i : initial production rate

q_t : production rate at time t



(a) Initial decline rate



(b) Decline exponent

Fig. 13 Variation of production behavior depending on Arps parameters
(Kang et al., 2017)

세일층은 수압파쇄를 적용함으로써 발생한 균열을 통해 생산이 되므로 일반적으로 생산 시작 초반의 몇 일 동안은 감퇴지수가 4의 값을 가지며, 몇 주 혹은 몇 달 동안 2의 값으로 유지되다 0이 된다(Rezaee, 2015). 이로 인해 세일 저류층에서 감퇴지수 값은 분석에 이용한 생산량 자료의 기간에 따라 상이한 결과가 추정된다(Fig. 14). 또한, Fig. 15와 같이 DCA 기법별로 차이는 발생하지만 생산 초기의 충분하지 않은 생산량 자료만을 사용할 경우, 생산량 예측 결과에 대한 불확실성이 존재한다(Meyet Me Ndong et al., 2013).

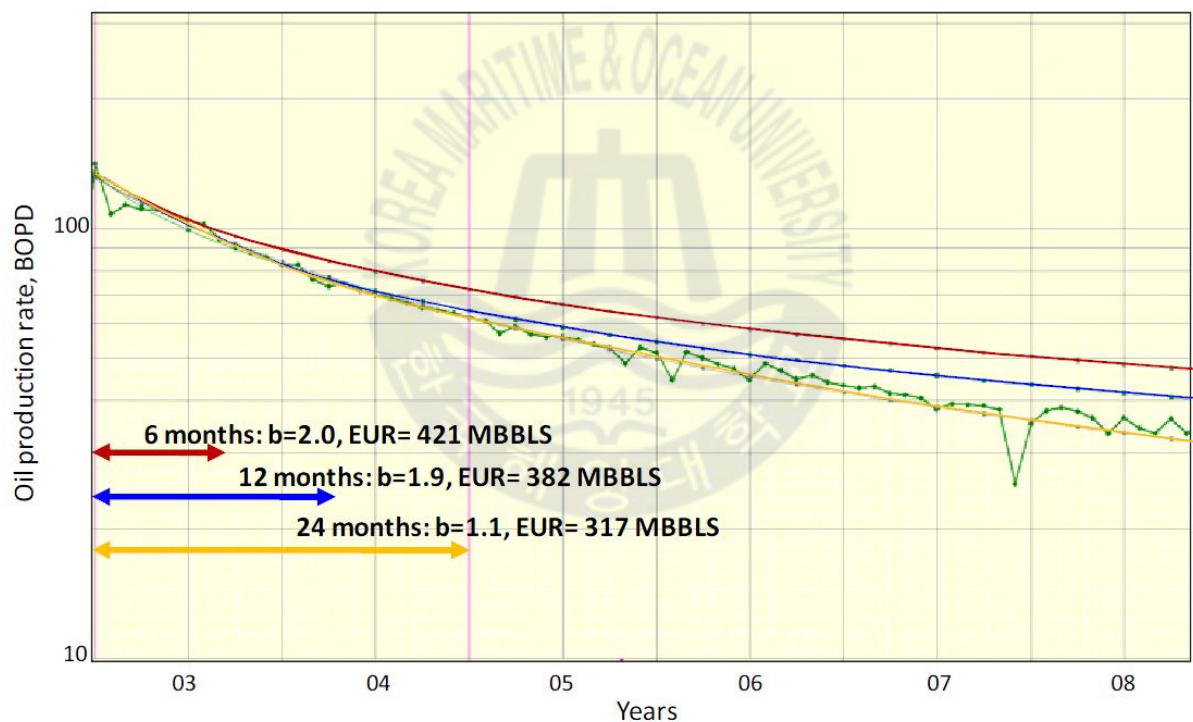


Fig. 14 Changing decline exponent in Arps hyperbolic (Kurtoglu et al., 2011)

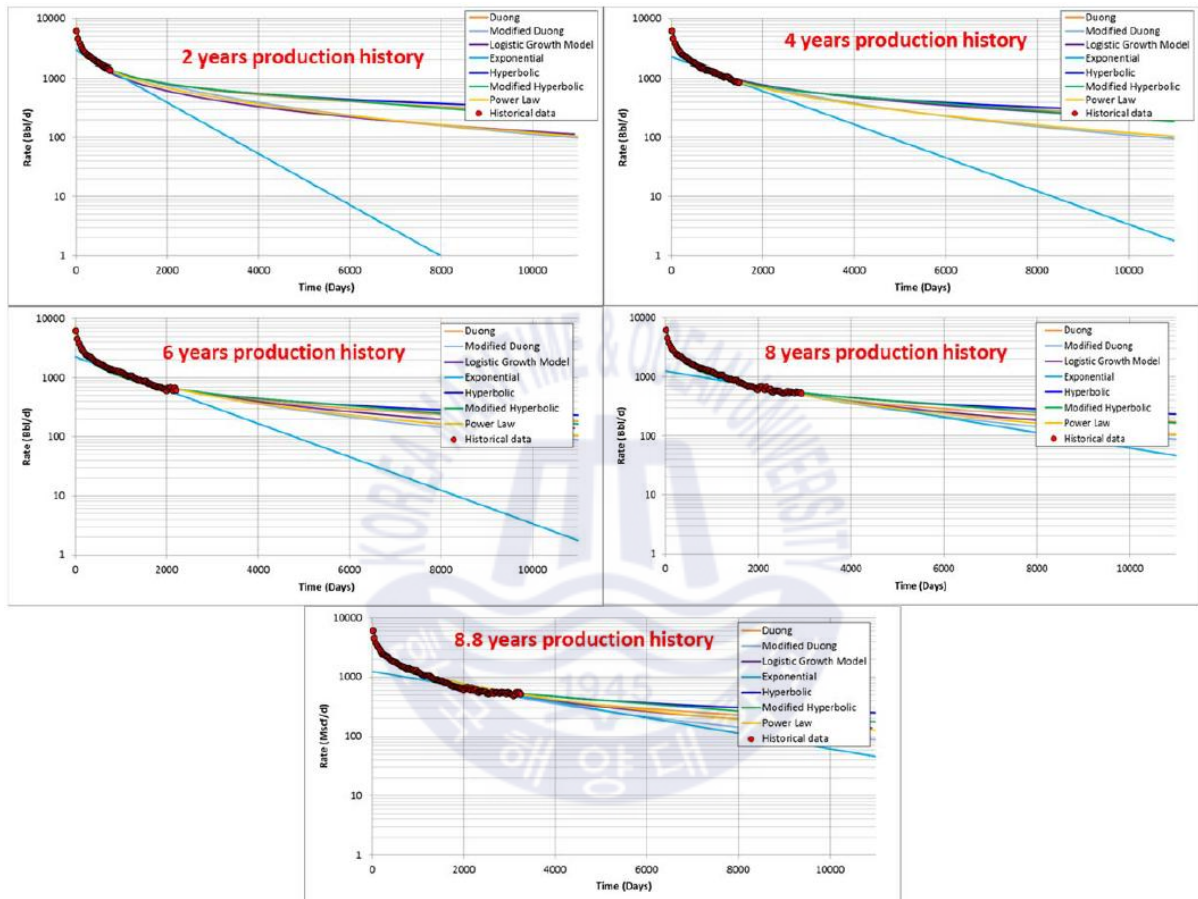


Fig. 15 Effect of increasing data in Pinedale horizontal wells production forecast (Meyet et al., 2013)

2.3 확률론적 기법을 적용한 생산량 예측

세일층에서는 천이유동 구간이 길어 BDF 형태의 유동흐름이 생산하는 동안 나타나지 않을 수 있으며, 장기간의 생산자료가 없거나 생산 운영압력(operating pressure)이 일정하지 않을 수 있다(Rezaee, 2015). 또한, 용적법을 이용한 자원량 추정 시 필요한 물성 측정에 어려움이 있으므로 자원량에 대한 불확실성을 고려하고자 몬테카를로 시뮬레이션(monte carlo simulation)을 적용하고 있다(Ma and Holditch, 2016). 이는 자원량 산정에 필요한 저류층 입력변수들의 가능한 모든 범위를 확률분포로 나타내고, 확률적 의미를 가질 만큼 난수를 발생시켜 각 변수 값을 추출함으로써 자원량의 예상 분포범위와 확률을 계산하는 방법이다(Fanchi, 2012).

해당 확률론적 기법 적용을 위해서는 자료의 특성에 따른 확률분포(probability distribution) 형태를 우선적으로 선정해야 하며, 가용할 수 있는 자료 개수에 따라 많은 경우 정규분포(normal distribution) 또는 로그정규분포(log-normal distribution)를 사용하고 균등분포(uniform distribution) 또는 삼각분포(triangular distribution)는 자료가 적은 경우에 주로 적용한다(Budzaya, 2013). 확률분포 형태의 최적화를 위해서 많은 시간과 모의시험이 소요되나, 몬테카를로 시뮬레이션은 난수를 생성하여 보다 유효한 결과도출이 가능하다(Chung, 2013). Fig. 16과 같이 확률분포 형태를 결정한 후에는 이를 누적확률분포(cumulative probability distribution; CDF)로 나타내며, 난수를 발생시킨 y축에서 해당하는 x축의 값을 추출하여 얻어진 값을 통해 확률론적 결과를 도출할 수 있다.

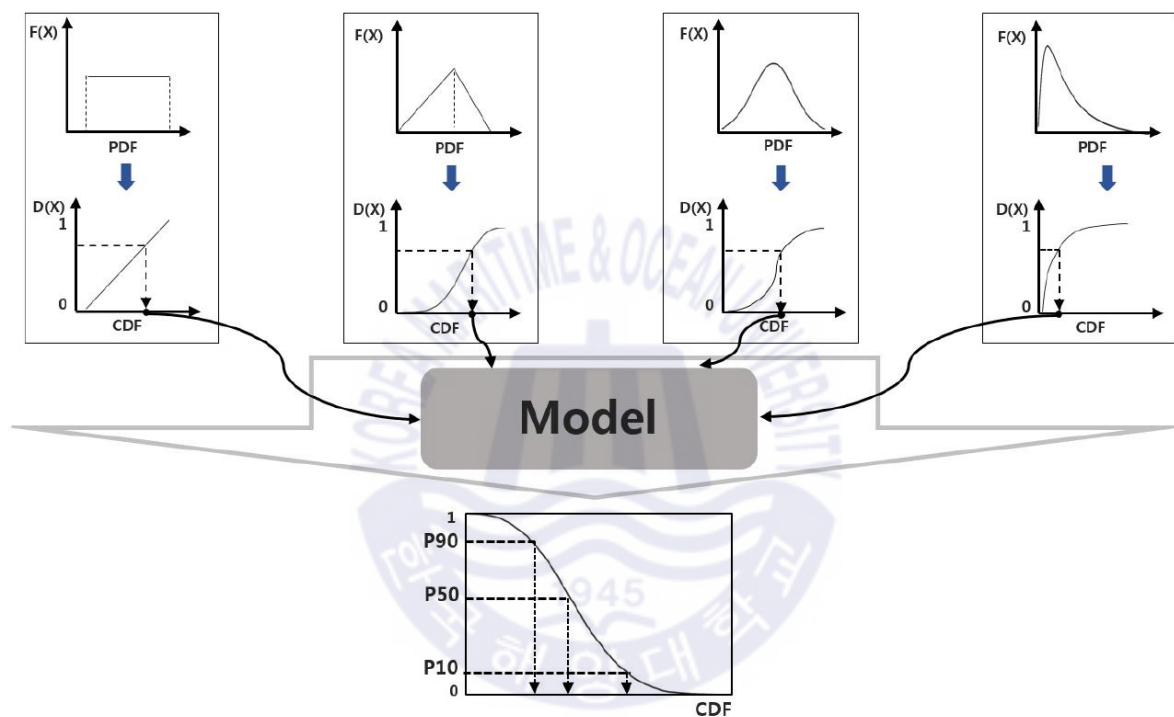


Fig. 16 Monte Carlo simulation (Kim et al., 2014)

단일 예측값을 산출하는 DCA를 이용할 경우 생산 초기의 감퇴율이 크고 후반에는 변화율이 작은 세일층에서의 생산 변동성을 고려하기 어렵다(Shin et al., 2014). 따라서 DCA 적용 시 확률론적 기법을 활용하여 세일가스의 궁극가채량 추정을 위한 연구가 수행되었으나(Kim et al., 2014), 적은 수의 자료, 왜곡된 자료 등에 대해 계급구간의 간격이나 시작점 선택에 따라 확률밀도함수(probability density function; PDF)의 모양이 달라지는 히스토그램(histogram)을 적용하기에는 한계가 있다. 이에 부정확한 가정을 전제로 하지 않고 통계량을 분석하는 비매개변수적(non-parametric) 방법인 핵밀도함수(kernel density function)를 적용하여 몬테카를로 시뮬레이션 적용 시 감퇴곡선인자에 대한 적정 분포형을 파악하였으며, 자료의 특성을 반영한 확률분포를 활용하여 궁극가채량을 추정할 바 있다(Shin et al., 2017).



제 3 장 데이터 기반 분석을 이용한 생산성 예측

3.1 머신러닝 기법

머신러닝은 자료로부터 특징을 도출함으로써 일종의 알고리즘 형태로 더 나은 의사결정과 예측을 수행할 수 있다. 또한, 대용량의 자료와 많은 변수가 관련되어 있으나, 기존 공식이나 방정식이 없는 복잡한 작업 등에 사용 가능하다. 즉, 직접 수많은 규칙을 미리 정하는 것을 대신하여 자료를 통해 스스로 학습하도록 하는 방법으로 Fig. 17과 같은 일련의 과정으로부터 결과를 도출한다.

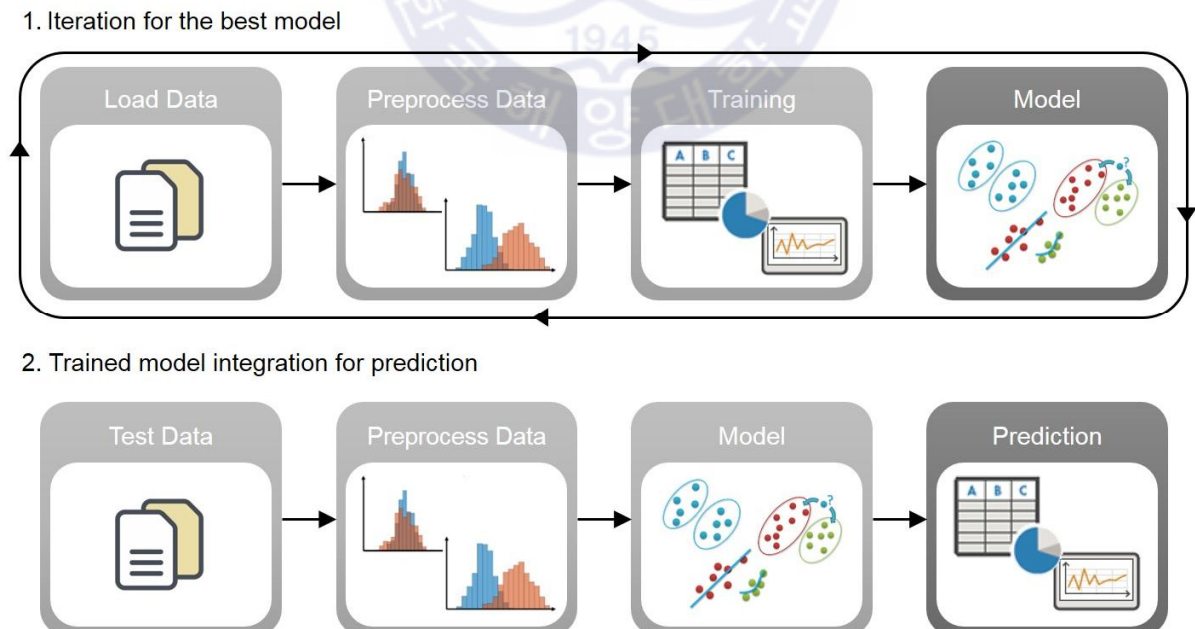


Fig. 17 Example of machine learning process

머신러닝은 학습방법에 따라 Fig. 18과 같이 구분된다. 지도학습(supervised learning)은 특정 입력에 대해 정답 혹은 레이블(label)이 있는 출력 자료 집합(data set)이 주어지는 경우로 프로그램은 이 정보로부터 입력과 출력의 관계를 유추하게 된다. 이러한 방법에는 주어진 자료의 유형을 나누는 분류(classification)와 주어진 자료에 대해 연속적인 값을 예측하는 회귀(regression)로 구분된다.

비지도학습(unsupervised learning)은 정답이 없는 자료 집합을 이용하여 정보 내 숨어있는 어떤 관계를 찾아내는 방법이다. 그 중 대표적으로 군집화(clustering)는 자료 사이의 관계를 찾아내 군집화하고 변수가 가질 수 있는 값들의 확률분포를 찾아내는 모델이며, 이상탐지(anomaly detection)를 통해 비정상적인 상태의 자료를 찾아내는 방법이 있다. 이외에도 보상을 최대화시키도록 결과를 선택하는 것을 학습하는 방법으로 강화학습(reinforcement learning)이 있으며, 보상은 즉각적이거나 지연된 것일 수 있다(Kim, 2017a).

이러한 방법 중 특정 머신러닝 알고리즘으로 모든 문제에 적용할 수는 없으며, 분석하고자 하는 자료에 적합한 알고리즘을 파악하기 위해서는 많은 시행착오가 필요하다. 또한, 훈련 속도, 새로운 자료에 대한 예측 정확도, 사용 편의성과 같이 튜닝(tuning)을 통해 적합한 알고리즘을 선택하여야 한다.

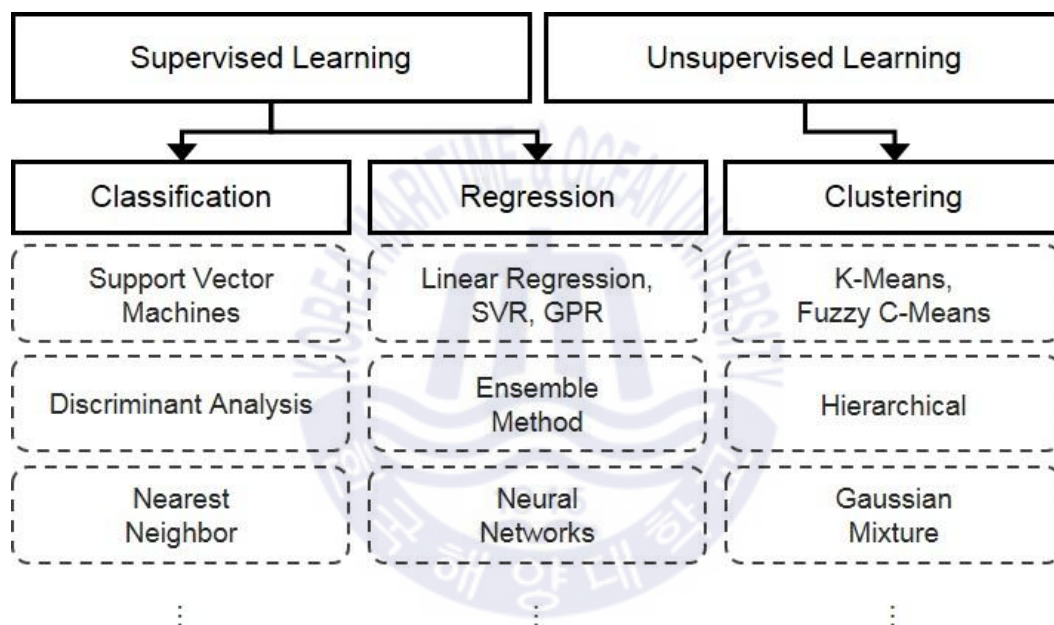


Fig. 18 Machine learning algorithm

3.1.1 자료 특징 파악

머신러닝은 자료의 양과 질에 의존적이므로 가장 이상적인 입력 자료는 부족하지도 과하지도 않은 정확한 정보만 포함할 수 있는 것이 유리하다. 그러나 관찰을 통해서 정확한 정보만을 얻기에는 한계가 존재하며, 해결하고자 하는 문제에 대한 정확한 배경지식이 있다면 가능하지만 대부분의 상황에서 그렇지 않다. 따라서 머신러닝 알고리즘을 적용하기에 앞서 속성 혹은 특징에 대한 파악과 관련이 있는 특징을 선별하는 과정이 필요하다.

일반적으로 이상값 제거, 누락된 자료의 대체, 정규화(normalization) 등을 수행 후, 충분한 자료를 취득하여 관측하고자 하는 대상에 대한 특징이 유효한지 아닌지 확인하는 특징 추출(feature extraction), 선택(selection) 등의 과정을 거친다. 이러한 방법은 기존의 입력을 토대로 새로운 입력 자료를 생성하며, 보통 학습과정 전에 실행되어 머신러닝 구조에서 핵심적인 전처리 과정 중 하나이다. 이를 통해 정확한 결과를 생성할 확률이 높은 자료에 중점을 두어 머신러닝 알고리즘을 적용하게 된다.

특징 추출은 학습 단계의 효율성을 위해 여러 종류의 속성으로부터 자료에 존재하는 중복성을 제거하는 과정을 수행하게 된다. 예를 들어 현장에서 측정되는 온도, 압력, 유량 등이 특징이 될 수 있으며, 이러한 특징의 선택은 전문가의 지식이 요구된다(Kumar and Bhatia, 2014). 또한, 자료의 필수 패턴을 파악할 수 있는 특징을 선택해야 한다. 고차원의 복잡한 자료에 대해서는 일반적으로 중복 혹은 관련이 없는 인자가 포함이 되어있으며, 이를 이용할 경우 낮은 일반화(generalization)로 인해 과적합(overfitting)을 초래할 수 있다. 따라서 특징 선택을 통해 관련 없는 자료를 제거하는 것이 중요하며, 클러스터링, 분류 등의 방법을 통해 가장 효과적인 하위집합(subset)을 선택한다.

이렇듯 고차원의 자료 집합은 많은 특징 혹은 인자가 서로 상관관계가 있으며, 이러한 종류의 자료에 대해 특징 파악과정 없이 분석하게 될 경우 과적합, 저장공간(storage) 낭비 등과 같은 문제가 발생할 수 있다. 이를 해결하기 위한 방법 중 머신 러닝 및 패턴인식기술에 주성분분석(principal component analysis; PCA)이 널리 사용되고 있으며, 입력 인자의 복잡성과 독립성을 감소시켜 특징을 추출 및 선택할 수 있다. PCA는 선형적으로 상관관계가 없는 벡터에 대하여 관계가 있는 변수를 생성하기 위한 통계적 절차이며(Li and Han, 2017), 특이값분해(singular-value decomposition)를 기반으로 여러 가지 반응 변수들에 대해 얻어진 다변량 자료 분석함으로써 손실을 최소화하여 차원축소가 가능하다(Fig. 19). 또한, 가장 중요한 독립변수를 발견하고 전체 자료에 대해 기여하는 정도에 따라 순서를 정하며, 기본적으로 자료의 분산(variance)을 가장 많이 보존하는 하위공간(subspace)을 찾아내는 기법이다(Jolliffe, 2002). 이는 서로 직교화 (orthogonalization)하는 새 기저로 자료를 변환하기 때문에 변수 간 상관관계가 높은 자료를 위한 분석에도 적용 가능하다.

PCA는 먼저 중앙으로 집중되어 있는 자료에 대해 식 (1)과 같이 기존 자료의 공분산 행렬을 계산하고, 새로운 축을 찾기 위해 식 (2)와 같은 고유치 분해(eigenvalue decomposition)를 수행한다. 그 후 고유값의 크기 순서대로 고유벡터(eigenvectors)를 나열하고 정렬된 고유벡터 중 일부를 선택하며, 해당 고유벡터와 자료의 내적(inner product)을 구하게 된다.

$$cov(A) = \frac{1}{n-1} AA^T \quad (1)$$

$$AA^T = U\Lambda U^T \quad (2)$$

A : data matrix

여기서, Λ 는 대각성분이 공분산 행렬의 고유값이고 나머지 요소는 0인 행렬이며, 대각성분은 행렬 A 의 각 변수에 해당하는 분산을 의미한다. U 는 열벡터가 공분산 행렬 AA^T 의 고유벡터로 이루어진 행렬이다. 이러한 분석 기법은 히스토리 매칭(history matching), 탄성과 해석(seismic interpretation) 등과 같은 여러 분야에 적용되고 있다(Chen et al., 2016). 또한, 해석의 효율성을 위해 새로운 차원에 대한 PCA 결과를 군집분석의 입력 변수로 활용할 수 있다(Lim et al., 1997; Yoo et al., 2009).

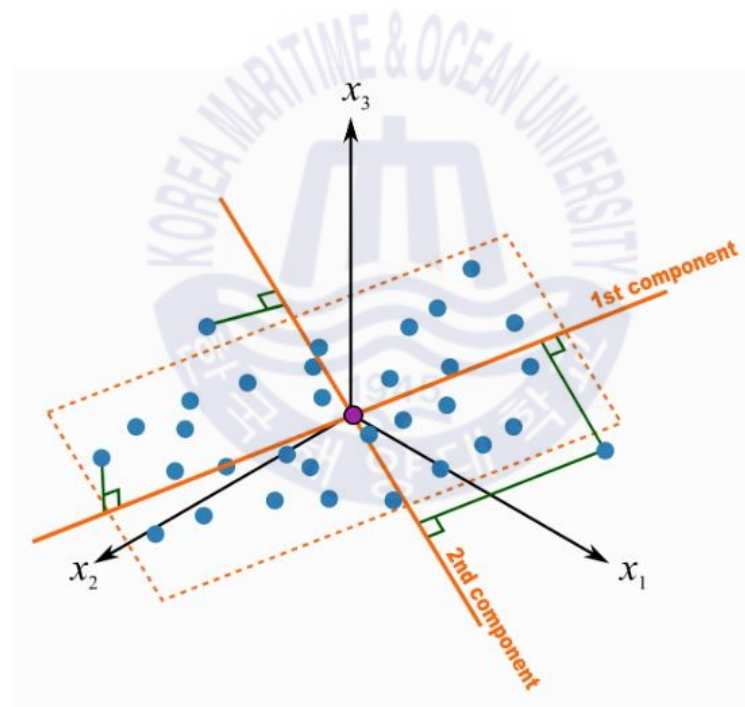
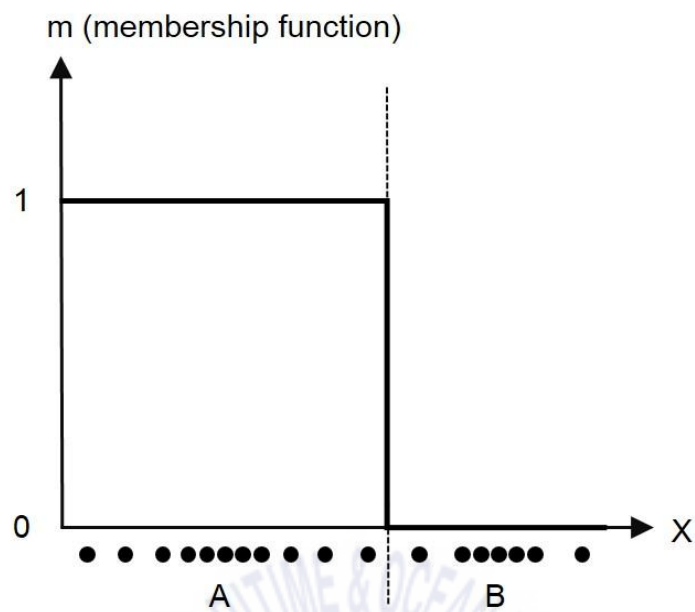


Fig. 19 Example of applying PCA (Dunn, 2019)

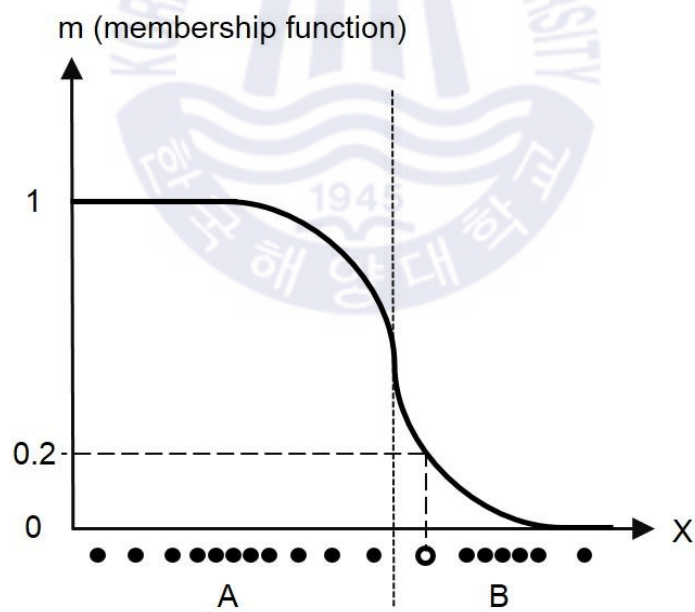
3.1.2 비지도학습

비지도학습은 자료에 포함된 정보가 무엇인지 불확실하거나 자료 탐색의 목표가 구체적이지 않은 경우 유용하게 활용할 수 있다. 대부분의 비지도학습은 군집분석(cluster analysis)의 형태로 여러 속성을 가지는 다수의 객체에 대하여 유사성 또는 속성의 측정값을 기반으로 그룹을 분할한다. 이는 한 군집 안에 부분 군집의 유무에 따라 계층적 군집(hierarchical clustering)과 비계층적 군집(non-hierarchical or partitional clustering)으로 나눌 수 있다. 계층적 군집은 한 군집 안에 부분 군집이 있는 반면, 비계층적 군집은 군집 간 부분집합이나 중복 없이 상호 배타적으로 존재한다(Tan et al., 2005).

계층적 군집분석은 유사한 자료끼리 군집으로 묶어가는 응집형 방법(agglomerative or bottom-up method)과 모든 자료를 하나의 군집에 속한다고 가정하여 세부 군집으로 나누는 분리형 방법(divisive or top-down method)이 있다. 비계층적 군집분석에는 k-중심(centroid) 군집, 퍼지(fuzzy) 군집 등의 프로토타입 기반(prototype-based), 분포 기반(distribution-based), 밀도 기반(density-based), 그래프 기반(graph-based)의 알고리즘이 있다(Tan et al., 2005). 또한, 이러한 방법은 Fig. 20과 같이 각 측정점(data point)이 하나의 범주에 속하는 하드 클러스터링(hard clustering)과 각 측정점이 두 개 이상의 범주에 속할 수 있는 소프트 클러스터링(soft clustering)으로 구분가능하다(Zaefferer, 2012).



(a) k-means (hard clustering)



(b) fuzzy c-means (soft clustering)

Fig. 20 Membership function for clustering of a one dimensional dataset

일반적인 하드 클러스터링 알고리즘에는 사전에 군집수 ‘k’를 정한 후 각 군집의 대푯값을 정하고 거리에 따라 각 개체를 k개 중 하나의 군집에 배정하는 방법이 있으며, 대푯값이 평균(mean), 중간점(median) 등에 따라 방법이 나뉜다. 소프트 클러스터링에는 하나의 개체가 여러 군집에 속할 가능성(possibility), 확률(probability)로 표현하는 퍼지 c-평균, 가우시안 혼합 모델 등이 있다. 퍼지 클러스터링은 k-평균과 유사하지만 하나의 포인트가 둘 이상의 클러스터에 속할 수 있도록 퍼지니스를 포함하며, 가우시안 모델은 하나의 포인트가 군집에 속할 확률을 제시하므로 군집의 크기와 속성의 상관관계 구조체가 다양할 경우 용이하게 활용할 수 있다(MathWorks, 2017).

이러한 군집분석 방법 중 k-평균과 퍼지 c-평균에 대하여 식 (3)과 식 (4)~(7)에 나타내었다. k-평균은 목적함수가 최소가 되는 즉 중심값과의 직선 거리를 측정하여 최소가 되는 점을 찾게 되며, 각 개체는 하나의 고유한 군집에 할당되므로 잘못된 값이나 한쪽으로 치우친 결과를 도출할 수 있다(Fig. 21). 퍼지 c-평균 분석은 하나의 개체가 여러 군집에 속할 가능성으로 표현하며, 목적함수의 지역적 최소(local minimum) 혹은 안장점(saddle point)에 수렴하는 값을 찾기 위하여 반복적인 membership function의 업데이트를 통해 식 (7)의 조건에 맞는 해를 찾는 방법이다(Politecnico di Milano, 2003).

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (3)$$

- c_j : center of cluster j
- J : objective function
- k : number of cluster
- n : number of data
- $x_i^{(j)}$: i th of measured data of cluster j

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (4)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (5)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \times x_i}{\sum_{i=1}^N u_{ij}^m} \quad (6)$$

$$\max_{ij} \{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \} < \varepsilon, \quad 0 < \varepsilon < 1 \quad (7)$$

c_j : center of cluster j
 C : number of cluster
 J_m : objective function
 k : iteration steps
 m : any real number greater than 1
 N : number of data
 u_{ij} : degree of membership of x_i in the cluster j
 x_i : i th of measured data
 ε : termination criterion

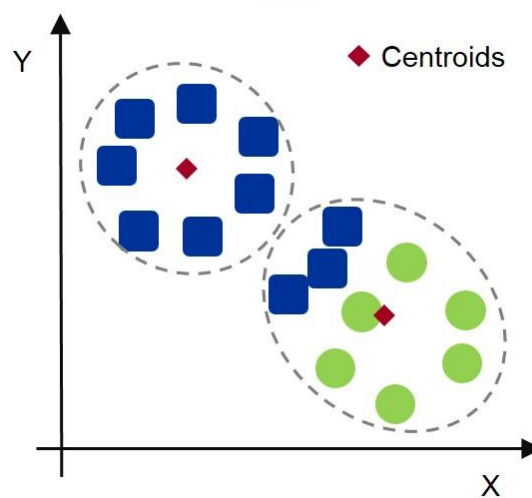


Fig. 21 Example of incorrect cluster analysis

3.1.3 지도학습

지도학습은 알려진 입력과 출력 자료 집합으로 모델을 훈련하고 새로운 입력 자료에 대해 합리적인 예측을 수행할 수 있으며, 모델 개발을 위해 분류와 회귀 알고리즘을 적용한다. 일반적인 분류 알고리즘으로 의사결정 트리(decision tree), 판별분석(discriminant analysis), 로지스틱 회귀(logistic regression), 나이브 베이즈(naïve bayes), 서포트 벡터 머신(support vector machine; SVM), 최근접이웃(k-nearest neighbor; k-NN) 등의 방법이 있으며, 회귀 알고리즘에는 선형(linear) 및 비선형(non-linear) 회귀, SVM 회귀, 가우시안 프로세스 회귀(gaussian process regression; GPR) 등이 있다. 다양한 알고리즘 중 적합한 방법을 결정하기 위해서는 충분한 자료 분석과 자료를 획득하는 분야에 대한 사전 지식이 필요하다.

의사결정 트리는 시작하는 루트(root)에서 아래쪽 리프(leaf) 노드까지 트리의 의사결정을 따르는 방식으로 응답을 예측하며, 훈련을 통해 분기수와 가중치(weights)가 결정된다. 로지스틱 회귀는 한 군집에만 속하는 이진 응답의 확률을 예측할 수 있는 모델을 피팅하며, 단순성으로 인해 일반적으로 이진 분류 문제의 시작점으로 활용한다. GPR 모델은 연속 응답 변수 값 예측이 가능한 비모수적(non-parametric) 모델이며, 공간 분석 분야에서 불확실성이 있을 경우 보간을 위해 널리 사용된다(MathWorks, 2017).

판별분석은 가우시안 분포를 기반으로 특징의 일차 결합을 찾아 자료를 분류하며, 새로운 자료 분류에 적용하여 클래스 레이블을 생성한다. 자료를 범주로 구분하는 직선을 찾거나 다양한 곡선에 맞출 수 있으며, Fig. 22와 같은 경우에는 선형 및 2차선(quadratic line)의 결과가 동일하다고 할 수 있다. 또한, 판별분석은 변수가 정규 분포를 따르고 그룹 내 공분산 행렬이 같다는 가정을 가지며, 일반적으로 분류기 개발 시 첫 번째로 활용하는 방법이다(Harwood and Wipat, 2012).

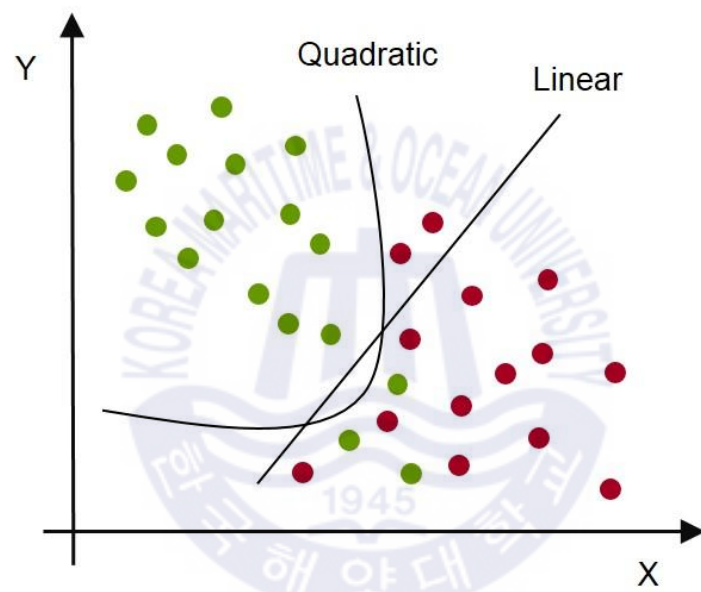


Fig. 22 Data classification using discriminant analysis

k-NN은 자료 집합 내에서 최근방의 군집을 기반으로 객체를 범주화하며, 서로 가까이 있는 객체가 비슷하다는 가정 하에 식 (8), (9)와 같은 유클리드(euclidean), 코사인(cosine) 등의 거리가 사용된다(Fig. 23). 자료 샘플의 레이블을 결정하는 가장 간단한 비모수적 절차이며, 투표(voting) 메커니즘을 기반으로 가장 가까운 k개의 인접 레이블이 결정된다(Cunningham and Delany, 2007).

$$d(x_i, x_j) = \|x_i - x_j\| = \sqrt{\sum_{l=1}^N (x_{i,l} - x_{j,l})^2} \quad (8)$$

$$d(x_i, x_j) = \cos(\theta_x^{ij}) = 1 - \frac{1}{2} \left\| \frac{x_i}{\|x_i\|} - \frac{x_j}{\|x_j\|} \right\|^2 \quad (9)$$

d : distance

θ_x^{ij} : the angle between x_i and x_j

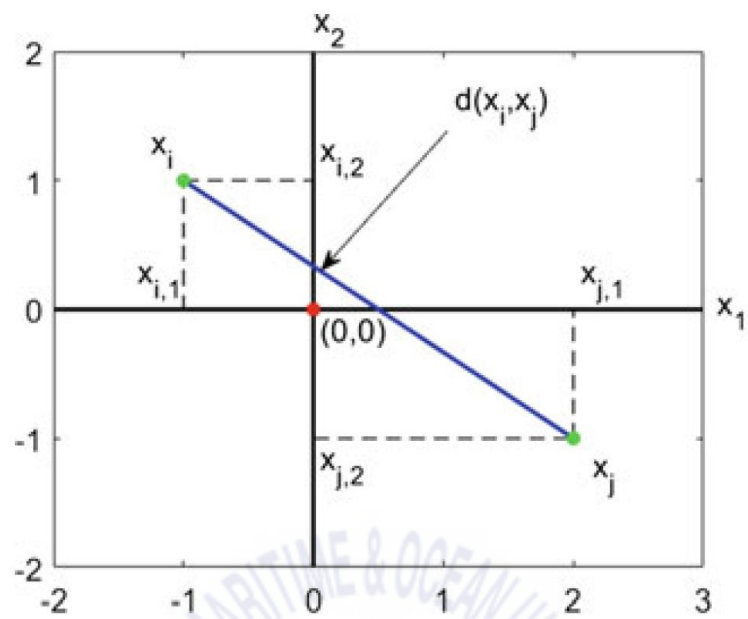
나이브 베이즈 분류기(classifier)는 기존 확률 및 통계의 개념을 바탕으로 하여 널리 연구에 사용되고 있는 분류 알고리즘 중 하나이다(Theodoridi and Koutroumbas, 2008). 이는 식 (10)과 같은 조건부 확률(conditional probability)을 기반으로 군집에 있는 서로 다른 특징들이 통계적으로 독립적이라는 가정 하에 분석을 수행한다. 새로운 자료가 특정 군집에 속할 가장 높은 확률을 기반으로 군집을 분류하며, 적은 수의 훈련 자료로 분류가 가능하여 계산 효율이 높은 장점을 가지고 있다.

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (10)$$

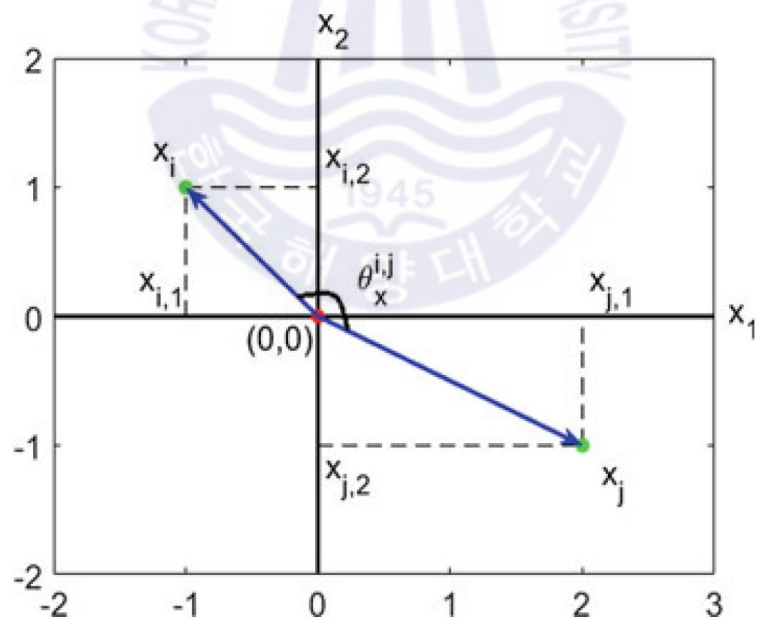
P : probability

X : dependent event

Y : dependent event



(a) Euclidean



(b) Cosine

Fig. 23 Illustrative example of distance (Angelov and Gu, 2019)

SVM은 한 군집에 모든 측정점을 다른 군집의 측정점과 구분하는 초평면(hyperplane)을 찾아서 자료를 분류하며, 가장 가까운 각 면에서 거리가 최대인 초평면 주위의 구역인 최대 마진(maximum margin)에 따라 결과의 정확도가 달라진다(Fig. 24). 비선형적인 경우에는 커널 변환을 사용하여 초평면을 찾을 수 있는 상위 차원으로 변환하여 적용하며, 일반적으로 사용하는 커널인 polynomial과 gaussian radial basis function에 대해 식 (11)과 (12)에 나타내었다(Angelov and Gu, 2019).

SVM 회귀 알고리즘은 SVM 분류 알고리즘과 같이 작동하지만 연속 응답을 예측할 수 있도록 수정되며, 자료를 구분하는 초평면을 탐색하는 대신 오류에 대한 민감성을 최소화하고자 가능한 적은 파라미터 값을 사용하여 측정된 자료에서 작은 값만큼 벗어나는 모델을 찾는 방법이다.

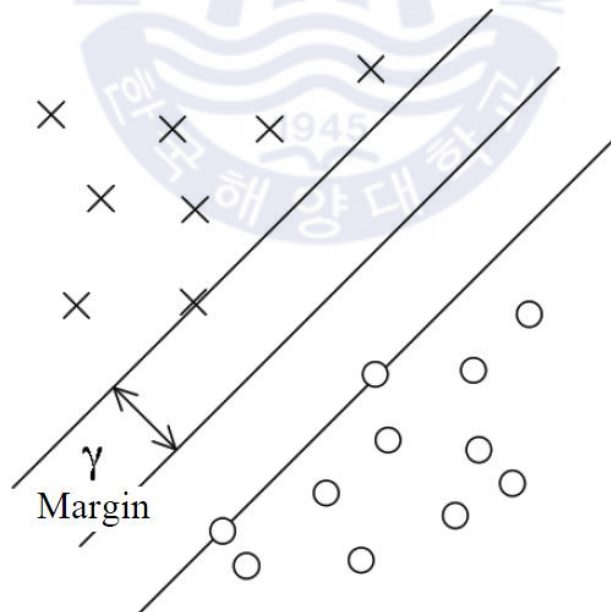


Fig. 24 Example of separating class using SVM (Crowley, 2016)

$$K(x_i, x_j) = (x_i^T x_j + 1)^n, \quad i, j = 1, 2, \dots, K \quad (11)$$

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}, \quad i, j = 1, 2, \dots, K \quad (12)$$

K : kernel

x_i^T : symmetric matrix of x_i

σ : standard deviation

회귀 알고리즘 중 선형 회귀는 연속 응답 변수를 하나 이상의 예측 변수에 대한 일차 함수로 표현하며, 모델이 간단하지만 입출력 사이의 관계에 대해 적절하고 해석 가능한 설명을 제공할 수 있다(Hastie et al., 2009). 식 (13)을 통해 출력을 예측할 수 있으며, 보편적으로 식 (14)와 같은 최소 제곱(least square)을 구하여 활용한다.

$$y = a_0 + \sum_{j=1}^N a_j x_j = \bar{x}^T a, \quad \bar{x} = [1, x_1, x_2, \dots, x_N]^T, \quad a = [a_0, a_1, a_2, \dots, a_N]^T \quad (13)$$

$$a = (\bar{x}^T \bar{x})^{-1} \bar{x}^T y \quad (14)$$

N : number of data

x : matrix of input vector

y : output

3.1.4 인공신경망 기법

ANN은 신경계와 생물학적 뉴런(neuron)의 동작원리를 모방하여 신경세포들 간의 수학적 연결 관계로 나타낸 것이며, 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)로 구성된다(Fig. 25). 식 (15)와 같이 연결강도를 변화시키면서 학습을 수행하게 되며, 활성화 함수(activation function)의 종류에 따라 뉴런의 출력이 달라질 수 있다.

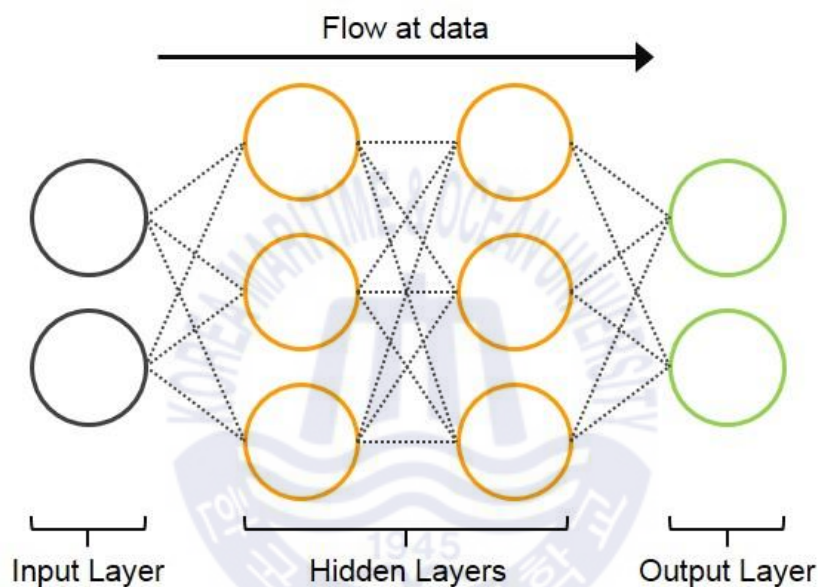


Fig. 25 Typical neural network representation of neuron

$$a^i = \sigma(W^i a^{i-1} + b^i) \quad (15)$$

- a^i : the neurons in layer i
- a^{i-1} : the neurons in layer $i-1$
- b^i : the bias vector
- W^i : the weight matrix for each layer i
- σ : the activation function

활성화 함수는 대표적으로 binary, linear, rectified linear, sigmoid, gaussian, hyperbolic tangent function이 있으며, binary function은 단극성 또는 양극성의 이진 함수로 입력 가중합이 임계치보다 작은 경우에 뉴런의 출력이 0이고 임계치보다 크거나 같을 때 뉴런의 출력은 1을 가진다. 시그모이드 함수는 Fig. 26과 같은 형태로 단극성 또는 양극성 비선형 연속 함수이며, 어떠한 형태의 입력값도 0과 1사이의 값으로 표현 가능하다.

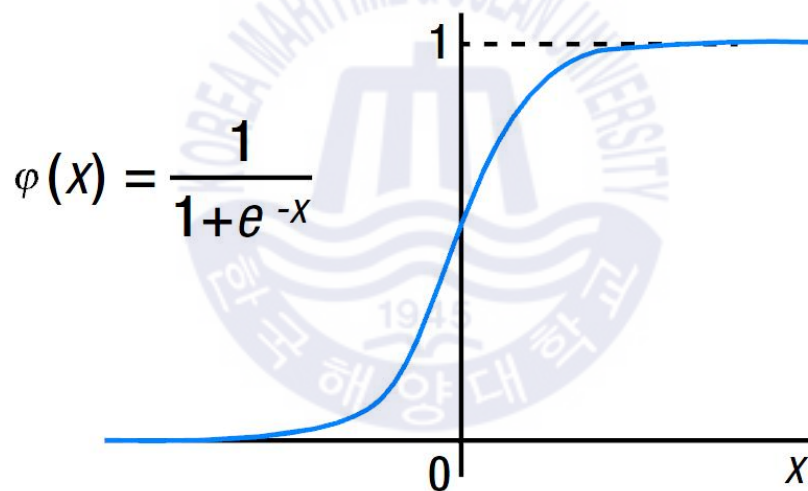


Fig. 26 The sigmoid function defined (Kim, 2017b)

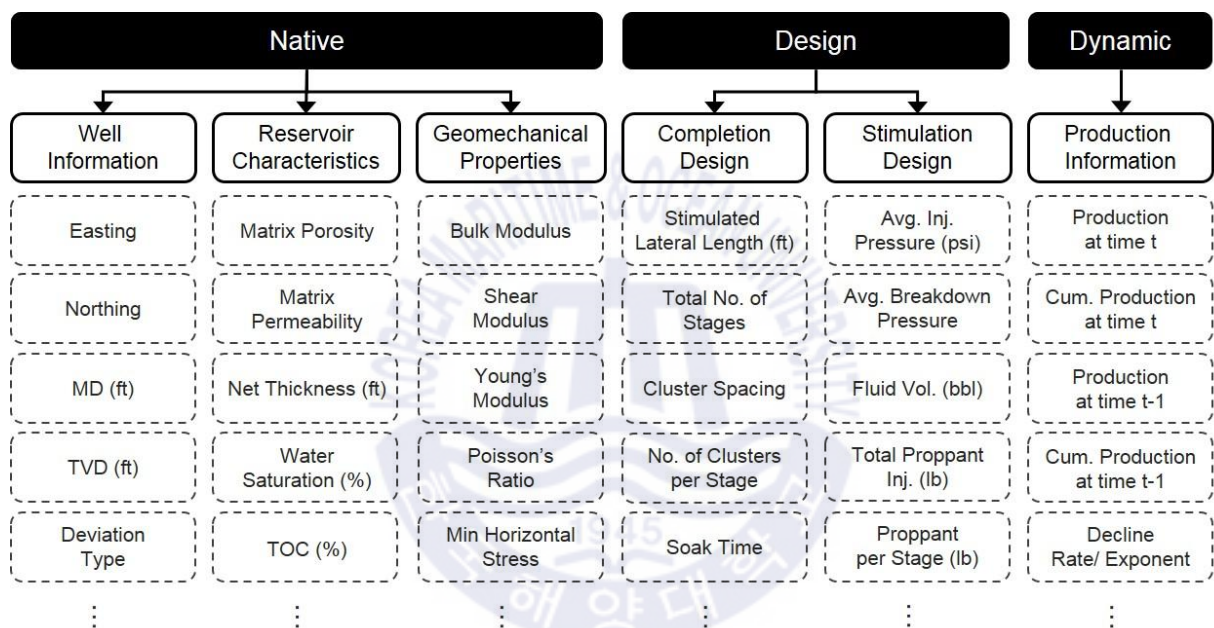
자료에 대하여 ANN은 지도학습이나 비지도학습을 적용할 수 있다. 전자의 경우 출력값을 주어진 입력에 따른 목표치의 오차범위 이내가 되도록 연결강도를 변경하며, 대표적인 모델로 Perceptron, Hopfield, 오류역전파(backpropagation) 등이 있다. 후자의 경우 비슷한 입력패턴들이 같은 출력으로 학습되도록 연결강도를 조정하며, Kohonen의 경쟁학습(competitive learning), Grossberg의 ART(adaptive resonance theory) 모델 등이 있다.

이러한 ANN은 뉴런간의 연결강도를 변화시키는 학습과정을 통해 입력 자료가 완전하지 않아도 패턴인식 및 비선형적으로 나타나는 특징을 도출할 수 있으며, 많은 양의 자료를 빠른 시간 내에 병렬로 처리할 수 있다(Ertekin and Silpngarmlers, 2005). 또한, 학습을 통해 연결강도를 조정함으로써 사전지식이 없는 상태에서 입력 및 출력 자료의 관계를 파악할 수 있으며, 새로운 정보를 추가하거나 변경할 수 있는 적응 특성을 가진다. 따라서 비선형적인 시스템의 모델링, 모델을 지속적으로 업데이트 하려는 경우, 입력 자료에 예기치 않은 변동이 있을 수 있는 경우 등에 활용 가능하다.

3.2 석유 생산 관련 인자

세일층에서의 생산은 저류층 특성에 의한 영향이거나 수압파쇄 설계의 영향을 받는다. 이러한 수압파쇄 성공 정도에 따라 생산성과 관련 있는 저류층자극부피(simulated reservoir volume; SRV)가 균질하게 생성되지 못하거나 미균열 구간이 발생할 수 있다. 이러한 균열의 판단을 위해 micro-seismic을 활용하여 세일 저류층을 모니터링 함으로써 자극된 영역과 그렇지 못한 영역을 구별하거나(Grechka et al., 2018), micro-seismic 자료, 생산 검층(production logging), 이미지 검층(image logging), tracer test를 사용하여 균열 특성화를 수행할 수 있다(Herz et al., 2017). 그러나 이러한 방법은 자료 취득 비용이 비싸거나 복잡한 균열 네트워크(fracture network)에서 평면(planar)의 균열 하나를 개별적으로 식별함에 있어 적용이 제한적이다(Baek et al., 2019).

따라서 Fig. 27과 같이 세일층의 생산과 관련한 현장 자료를 요약하였으며, 다양한 인자들이 존재한다. 먼저 ‘Native Parameters’에는 시추공의 위치, 총길이를 의미하는 생산정 깊이(measured depth; MD), 생산정 깊이(true vertical depth; TVD), 방향 등과 저류층 특성(matrix porosity, matrix permeability, net thickness, water saturation, total organic carbon etc), 역학적 물성(bulk modulus, shear modulus, poisson's ratio, minimum horizontal stress etc)이 포함된다. ‘Design Parameters’에는 자극된 수평 길이(stimulated lateral length), 수압파쇄 단계 수(stages), 균열 간격(spacing) 등의 유정완결(well completion) 요소와 프로판트 및 유체의 총양, 주입 압력 등의 유정자극(well stimulation) 요소와 같이 조절 가능한 인자가 포함된다. 그리고 시간에 따른 생산량 변화에 대한 요소를 포함한 ‘Dynamic Parameters’에는 생산량, 누적 생산량, 감퇴율 등 생산량과 관련된 정보가 포함된다.



Source: Modified Mohaghegh(2017) & He(2017)

Fig. 27 List of native, design and dynamic parameters in shale reservoir

조절 가능한 변수인 수압파쇄 설계인자 중 Lateral length와 Stage 수는 일반적으로 증가할수록 생산량 또한 높아지는 현상을 반영하여 세일층을 개발하여 왔다. Marcellus와 Eagle Ford의 경우 2008년에 10회에서 2012년 16회로 Stage 수를 증가시켰고, Bakken은 2008년 14회에서 2012년 27회로 횃수를 늘렸다(Ma and Holditch, 2016). 그러나 수압파쇄 설계가 이상적으로 이루어졌다는 가정 하에 Fig. 28과 같이 Stage 수에 따라 유정 완결 비용도 증가하므로 적절한 값을 찾기 위한 연구가 수행되어야 한다.

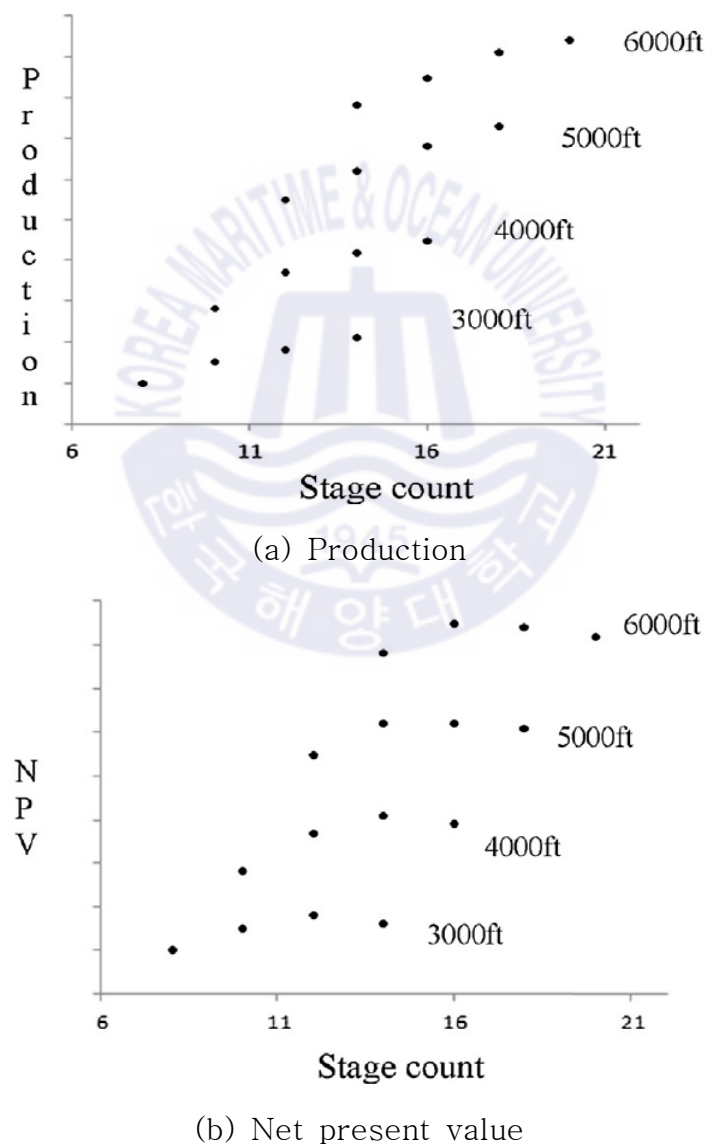
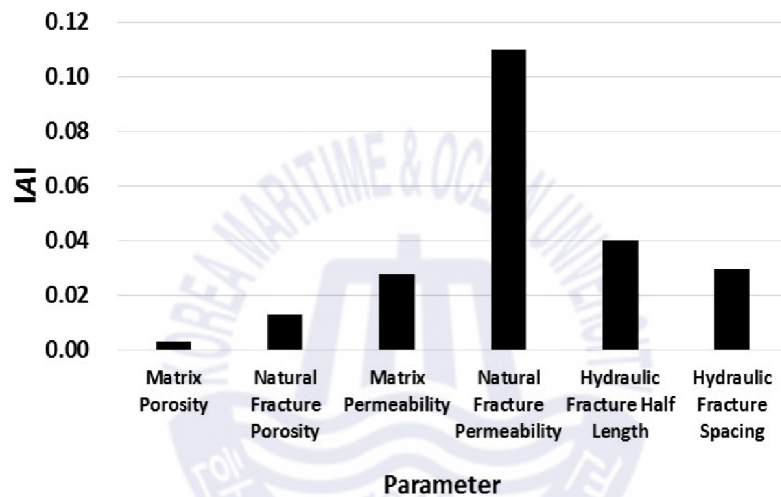
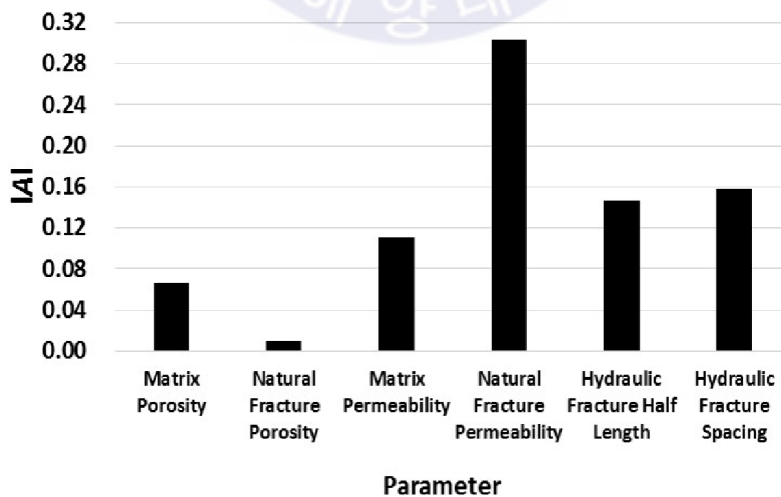


Fig. 28 Impact of lateral length and stage count (Ma and Holditch, 2016)

이외에도 저류층 물성과 수압파쇄 요소가 셰일층의 생산거동에 미치는 영향을 파악하고자 현장 저류층 조건을 반영한 시뮬레이션을 수행하였으며(Kang et al., 2017), 결과에 대한 분석을 통해 각 요소가 감퇴곡선인자에 미치는 영향정도를 나타내는 $|A|$ 를 도출하여 자연균열 유체투과도, 수압파쇄 균열길이 및 간격의 영향이 큰 것을 파악하였다(Fig. 29). 따라서 복합적인 영향으로 변화하는 생산성에 대한 분석을 위해서는 자료에 대한 이해와 주요 정보를 추출하는 것이 필요하다.



(a) Initial decline rate



(b) Decline exponent

Fig. 29 Effect of reservoir and fracture conditions (Kang et al., 2017)

그러나 이러한 인자들은 각 인자 간의 관계나 Fig. 30과 같이 생산량과의 관계가 불분명하며, 다수의 인자의 복합적인 영향으로 인해 생산량이 변화하게 되므로 인자들의 관계 파악 및 분석을 위해 취득 가능한 현장 자료를 수집하여 본 연구에서 활용하고자 한다.

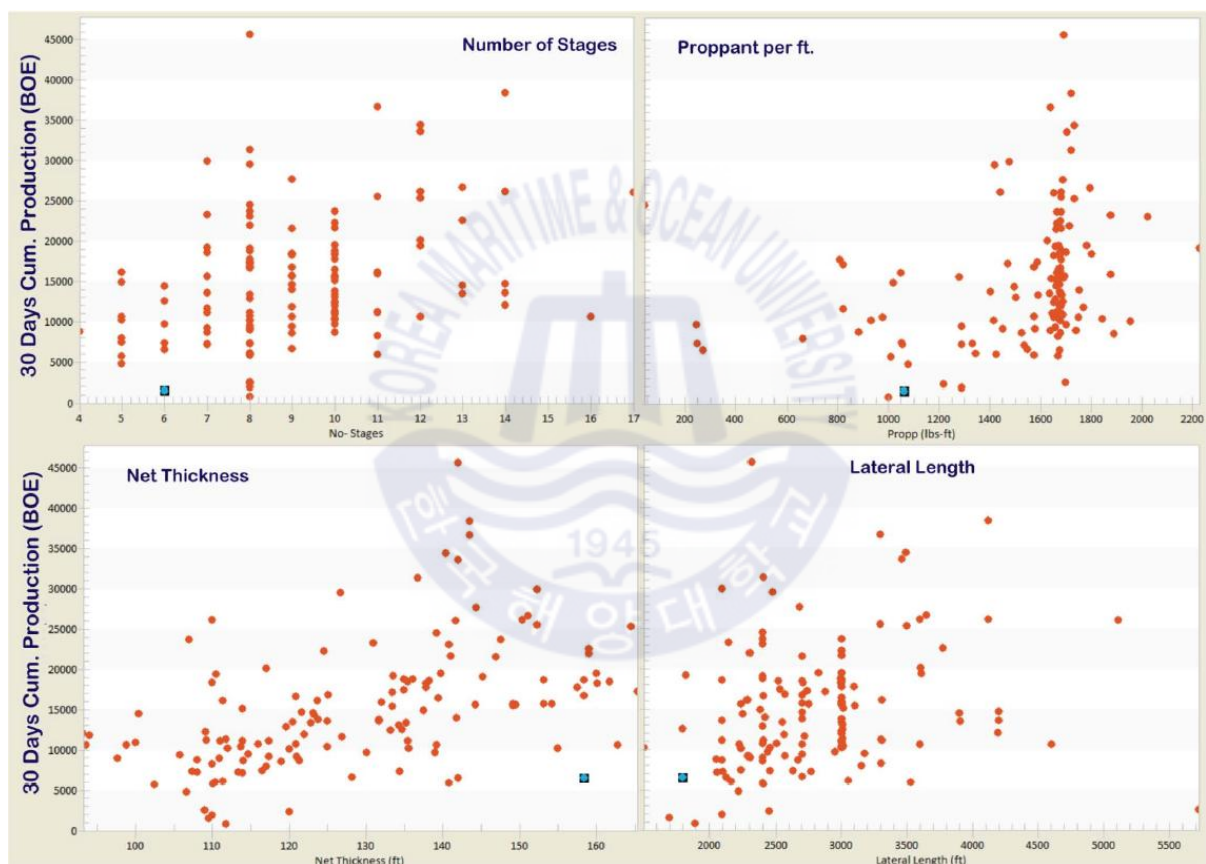


Fig. 30 Cross plot of 30 days cumulative production in Marcellus shale (Mohaghegh et al., 2017)

3.3 연구 사례 분석

세일층의 생산 거동이나 생산량에 영향을 미치는 인자들의 경우 상관관계가 불명확하므로 데이터 기반 분석을 통해 인자들 간의 영향을 파악하고 이를 활용한 생산 예측 연구가 진행되고 있다. 일반적인 석유자원과 달리 세일 저류층의 특성에 대하여 규명이 되지 않아 지난 2~3년 사이에 자료의 특징을 반영할 수 있는 머신러닝을 활용하여 많은 연구들이 수행되었다.

이러한 연구의 궁극적인 목표는 기존의 생산정에 대하여 미래 생산거동이 어떻게 바뀔 것인지와 새로운 위치에서 생산을 하게 될 경우 어떠한 계획을 가지고 얼마만큼의 생산량이 나올 것인지를 예측하는 것이다. 해당 절에서는 세일층의 자료의 특징을 반영한 머신러닝을 통해 생산 예측을 수행한 연구에 대하여 분석하였으며, 이와 관련한 주요 내용을 Table 5와 Table 6에 정리하였다.

세일층에 대하여 기존의 생산정 자료를 바탕으로 플레이 별 연구가 진행되고 있으며, 미래의 생산 거동과 생산량 추정을 위해 여러 가지 인자 중 주요한 인자에 대한 파악을 우선적으로 수행하였다. Li and Han(2017)은 저류층 시뮬레이션을 기반으로 세일층 특성과 수압파쇄 설계인자의 10가지 인자에 대해 PCA를 수행하였으며, 각 주성분에 대한 계수를 바탕으로 7가지의 주요한 인자를 ANN 모델의 입력 인자로 활용하였다. 출력층에는 DCA 중 하나인 Logistic Growth Model(LGM)의 3개의 변수를 위치시켰으며, 훈련 70%, 검증 15%, 테스트 15%로 ANN 모델을 학습시킨 결과 0.013 Mscf/day의 MSE(mean squared error)와 0.92의 상관계수(correlation coefficient)를 가짐으로써 실제 현장에 대한 적용이 가능할 것으로 판단하였다.

He(2017)은 저류층 특성, 수압파쇄 설계인자, 생산량 인자에 대해 Intelligent Solution Inc에서 개발한 IMPROVE™을 활용하여 Key Performance Indicator(KPI) 파악하고 이를 바탕으로 ANN 모델을 개발하였다. 20개의 입력층과 50년의 EUR 추정을 위해 훈련 80%, 검증 10%, 테스트 10%의 자료를

오류역전과 알고리즘을 이용하여 학습한 결과 대부분 일치하는 결과를 보였으며, 이를 통해 생산량에 대하여 세일층 특성과 유정완결 인자에 대한 영향 분석을 수행하였다.

Vyas et al.(2017)은 생산량 자료를 이용하여 4가지의 DCA를 수행하였으며, 도출한 감퇴곡선인자와 EUR을 응답변수로 세일층의 생산과 관련된 9가지 인자를 입력하여 3가지의 머신러닝 알고리즘을 적용하였다. 다양한 DCA 별 결과에 대해 Bayesian Model Averaging(BMA)와 Generalized Likelihood Uncertainty Estimation(GLUE)로 평균화한 결과 SEPD의 SVM 결과가 가장 정확도가 높았으며, 입력인자에 대한 Relative Influence(RI)를 구한 결과 초기생산량, 총 프로판트 양, 생산정 깊이 순으로 나타났다.

Bowie(2018)는 Maturity, Pad wells, Hydraulic fracture size와 관련된 인자들에 대한 상관계수로부터 Variance Inflation Factor(VIF)를 계산하여 다중공선성(multicollinearity)을 정량화하였으며, 단순 및 다중 선형 회귀, ANN을 적용한 결과 ANN 모델의 예측 정확도가 결정계수(coefficient of determination) 0.93으로 가장 높았다.

Luo et al.(2018)은 세일층의 특성, 수압파쇄 설계인자에 대하여 상관계수를 구하여 서로 관련 있는 변수를 파악하였으며, 특징 선택을 통해 도출한 8가지 입력 인자를 다수의 은닉층과 뉴런으로 구성된 인공신경망인 딥러닝(deep learning) 모델로 학습하였다. 그러나 많은 생산정 자료와 은닉층 및 뉴런수가 많은 모델임에도 불구하고 결정계수가 0.61로 나타났다. 또한, 개발된 예측 모델을 활용하여 지질학적 인자와 유정완결 인자가 단위길이 당 생산량에 미치는 영향을 파악하였으며, 낮은 공극률에서 세일층의 두께(thickness)가 두껍고 Stage 수가 많으며 심도가 깊어질수록 생산량이 증가함을 확인하였다.

선행 연구 결과 ANN을 활용한 경우 예측 성능이 높은 편으로 나타났으며, 수치적인 분석 시 머신러닝을 활용한 예측이 가능함을 확인하였다. 이 때, 분석에 활용한 생산정에 대하여 주요한 영향을 미치는 특징의 선택, 추출

과정이 수행되어야 하며, 생산정 수에 관계없이 동일 분지라도 생산 변동성이 크게 나타나는 셰일층에 대한 고려가 필요하다. 이 연구에서는 기존 생산정에 대하여 그룹을 분류하여 다양한 생산특성이 나타나는 셰일층의 현장 조건을 고려하고자 하였으며, 그룹을 활용하여 신뢰성 있는 생산량 추정을 수행하고자 한다.

기존 분석기법 적용에 있어 한계가 존재하고 생산 시 유동 메커니즘이 불확실한 셰일층은 일반적인 석유자원과 달리 수압파쇄 설계인자가 생산성에 미치는 영향을 파악하여야 한다. 또한, 생산을 위한 유정 단위의 개발 계획과 더불어 새로운 유정에 대한 생산량 예측이 필요하다. 최근 들어 유·가스 가격에 대한 영향으로 북미지역에서는 시추는 하였으나 유정완결 작업을 하지 않은 유정(drilled but uncompleted wells; DUCs)이 점차 증가하는 추세를 보이고 있으며(Fig. 31), 이 경우에는 유정의 위치나 깊이 정보는 존재한다.

이처럼 셰일층의 정보가 없는 새로운 유정에 대한 개발 시 일반적으로 인근 생산정의 정보를 바탕으로 계획을 수립하고 있다. 이와 관련한 개발 계획이나 생산성 예측 관련 연구는 시작 단계이며, 새로운 유정에서도 기존 생산정과 유사한 입력인자를 활용하여 생산거동 및 특정 시점의 누적 생산량을 예측한 바 있다.

Table 5 Research on productivity prediction of existing wells in shale reservoir

	Li & Han (2017)	He (2017)	Vyas et al. (2017)	Bowie (2018)	Luo et al. (2018)
Basin	Reservoir simulation	Marcellus	Eagle Ford	Duvernay	Bakken
Fluid	Oil & Gas	Gas	Oil	Oil	Oil & Gas
Well Num.	100	100	120	262	2061
Method	ANN [7 50 3]	ANN [20 20 1]	RF SVM MARS	Linear Regression Multiple LR ANN	Deep Learning [8 4x100 1]
	7	20	9	21	8
Input	• Permeability • Porosity • Fracture width • Fracture half length • Fracture conductivity • Formation pressure • Formation temperature	• Total proppant • Lateral length • Stages • MD • Fracture clusters • Average pressure • Easting Northing • TOC • Azimuth - Etc	• Latitude • Longitude • Initial prod. • Total proppant • Total fluid • Stages • Lateral length • TVD of heel • TVD heel-toe difference	• Azimuth • TVD • Total proppant • Total fluid • Lateral length • Avg. pump rate • Spacing • Stages • Acid volume - Etc	• Total thickness • Norm. proppant • Depth • Porosity • Stages • Norm. fluid spacing • Water saturation
Output	LGM (K, a, n)	50 years EUR	DCA parameters & EUR (Arps, SEPD. Duong, Weibull)	Well performance (individual well/ type curve)	1 year production by stage
Results	R 0.92	The points of the actual and predicted data are almost identical	RMSE (for EUR) 45,32,44,30(RF) 44,31,43,29(SVM) - (MARS)	R ² 0.41(LR) 0.76(Multi LR) 0.93(ANN)	R ² 0.75(Training) 0.61(Testing)

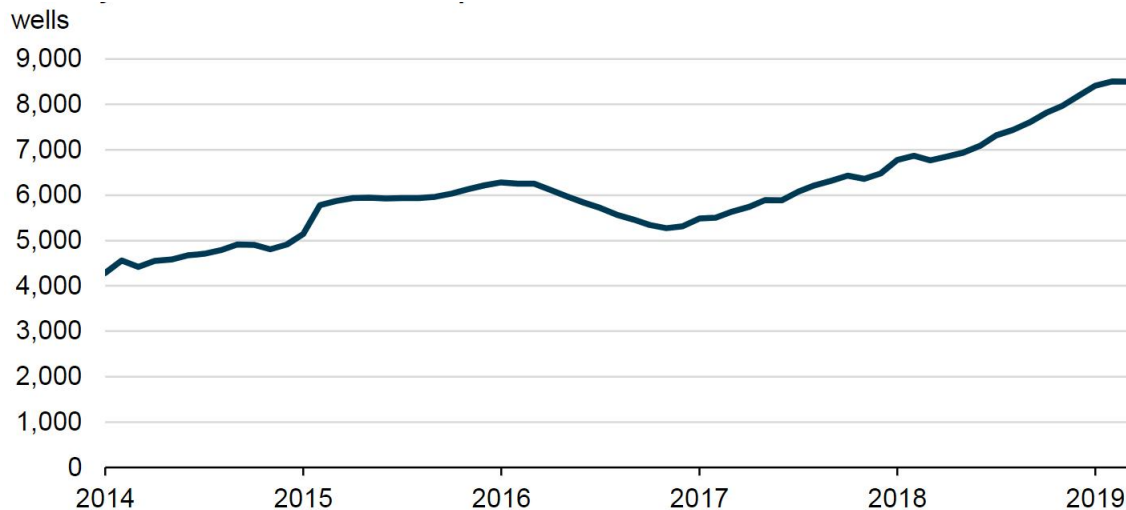


Fig. 31 Monthly count of DUCs (EIA, 2019c)

Cao et al.(2016)은 하나의 유정에 대한 생산거동 예측을 위해 ANN 모델을 적용하였으며, 기존 생산정의 경우 Tubing Head Pressure(THP)와 생산량 자료를 이용하였다. 새로운 유정에 대해서는 타겟의 위치, 인근 생산정 정보로부터 파악한 지질학적 지도, THP, 생산량 자료를 입력하여 결과를 도출하였으며, 4년 이상의 생산량 자료에 Arps, SEPD, Duong, Power law의 DCA를 적용한 결과와 비교분석하였다. 그 결과 DCA를 활용한 경우 실제 생산량과 생산경향은 유사한편이었으나, THP를 이용하여 분석한 ANN 모델의 정확도가 높았다.

Mohaghegh et al.(2017)은 데이터 기반의 셰일층에 적용할 수 있는 Shale Analytics의 개념을 도입하였으며, IMPROVE™을 활용하여 Well Quality Analysis(WQA)를 수행하고 Native 및 Design 인자에 대하여 생산량과의 관계를 KPI로 파악하였다. 이를 ANN 모델의 입력 인자로 6개월의 누적 생산량을 추정하였으며, 기존 생산정에 대해 96%의 정확도를 보였다 새로운 유정에 대한 생산 예측을 위해 기존의 생산정을 Blind well로 가정하였으며, 그 결과 77%의 정확도와 평균 12.9%의 오차율을 보였다. 또한, 생산량에 대한 수압파쇄 설계인자의 영향 파악을 위해 몬테카를로 시뮬레이션과 예측 모델을 활용하여

Look-Back Analysis를 수행하였으며, 단일 생산정에 대하여 최적화된 수압파쇄 설계인자를 제시할 수 있는 방법을 제안하였다.

Amr et al.(2018)은 생산 지역(producing location; PLs)과 미생산 지역(non producing location; NPLs)에 대한 생산량 예측을 위해 많은 양의 자료와 여러 지역의 자료를 활용하였다. 이 때, 업계의 기준인 Arps 방정식의 초기감퇴율과 EUR을 추정하였으며, 인근 생산정에 대한 정보를 활용하여 P10, P50, P90의 Type Curve를 도출하였다. 생산량 예측을 위해 세일층 특성, 수압파쇄 설계인자에 대해 거리가 2500 ft 이하 등 조건에 만족하는 인근 생산정 자료를 활용하였으며, 초기감퇴율과 EUR을 추정할 수 있는 머신 러닝 모델을 개발하였다. 7개 알고리즘을 적용하였으며, 그 중 Extreme gradient boosting tree(xgbTree)의 EUR 예측 정확도는 PLs 93.27%, NPLs 76.74%로 나타났다. NPLs에 대한 정확도 향상을 위해 여러 지역의 자료를 추가하여 분석하였으나, 가장 높은 경우는 67.43%로 불균질한 세일층의 특성에 의해 저류층 특성과 인근 생산정 자료를 활용함에도 불구하고 생산하지 않은 지역에 대한 결과의 정확도가 낮은 편이다.

따라서 이 연구에서는 수직, 수평에 따른 특성 변화가 심한 세일층의 경우 인근 생산정 정보 사용에 한계가 존재하므로 생산특성이 다른 그룹의 확률론적 입력값을 활용하여 제한된 자료만으로 새로운 유정에서 불확실성을 고려한 생산량 예측을 수행하고자 한다.

Table 6 Research on productivity prediction of new wells in shale reservoir

	Mohaghegh et al.(2017)	Amr et al.(2018)
Basin	Marcellus	DJ, Williston, Anadarko, Powder River
Fluid	Oil & Gas	Oil
Well Num.	128	713 (DJ)
Method	ANN [9 15 1]	Extreme gradient boosting tree (xgbTree) (Best performing of 7 algorithms)
	9	31
Input	• TVD (ft) • Net thickness (ft) • Porosity (%) • TOC (%) • Lateral length (ft) • Total No. of stage • No. of clusters per stage • Fluid per lateral (bbl/ft) • Proppant per lateral (lbs/ft)	• Surface/Bottom latitude & longitude • County • Lateral length, Proppant per lateral • 3, 6, 12 Cum. Oil • Bulk volume, Thickness, S_w, Porosity (net reservoir) • Distance, Angle (neighbors) • Avg. 3, 6, 12 Cum. Oil (neighbors) • Avg. Bulk volume, Thickness, S_w, Porosity (neighbors) - Etc
Output	6 Months Cum.	Arps (q_i , D_i , EUR) (one for each parameter)
Results	R^2 0.96(Training) 0.77(Blind)	Accuracy(%) (for best case) 97.31, 87.82, 93.27(PLs) 81.89, 80.97, 76.74(NPLs)

제 4 장 기존 생산정의 주요인자를 활용한 생산특성 분석

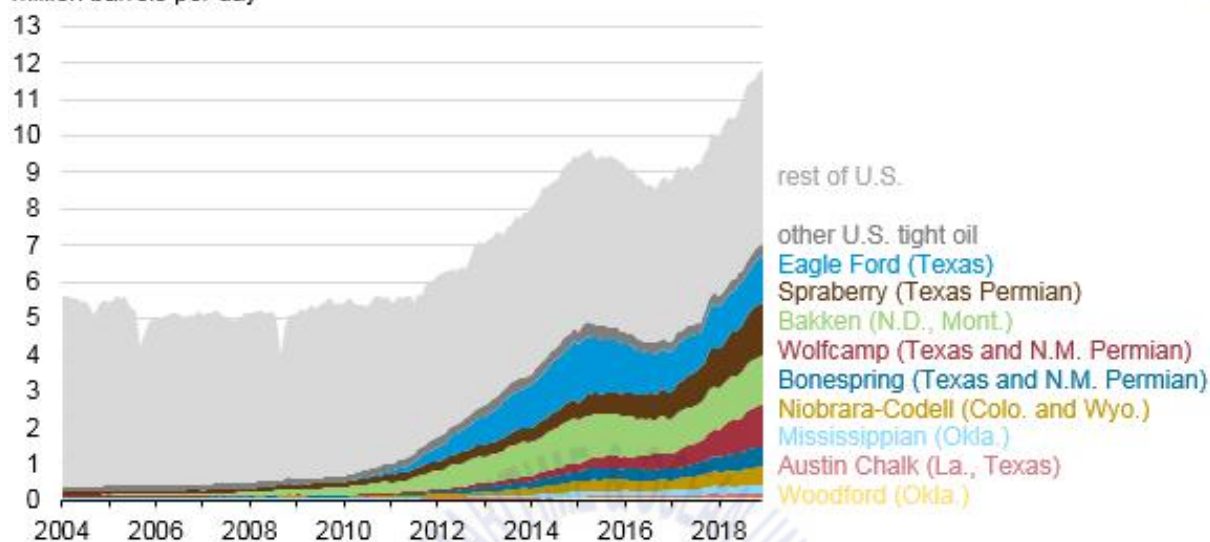
4.1 연구대상 지역

Fig. 32와 같이 미국 내 활발한 생산이 이루어지고 있는 셰일 플레이 중 하나인 ‘Eagle Ford Shale’은 100개 이상의 리그가 운영되고 있다. 해당 지역은 20.81 Tcf의 천연가스와 3.351 Bbbl의 오일이 부존되어 있을 것으로 추정되며, dry gas, wet gas, NGLs, gas condensate, crude oil과 같이 다양한 상의 유체가 생산된다. 근원암은 Austin Chalk이며, 북서쪽으로 갈수록 얕아지고 셰일 함량이 증가한다. 또한, 셰일층은 50 miles의 너비와 평균 250 ft의 두께를 가지며, 4,000~14,000 ft 사이의 심도에 위치한다(James, 2013).

연구대상 지역의 현장 자료는 DrillingInfo에서 AAPG geological province 중 Texas & Louisiana Gulf Coast Basin, 운영권자(operator), 생산 기간, 유정 완결 기간을 동일하게 설정하였으며, 약 60개월 생산된 250개의 자료를 취득하였다. 이 연구에서는 Fig. 33에서 오일 생산이 주를 이루는 곳이 대상 지역으로 분석 시 오일 생산량 자료를 활용하였으며, 데이터 기반 분석을 위해 MathWorks사의 MATLAB R2019b를 사용하였다.

Monthly U.S. crude oil production (2004-2018)

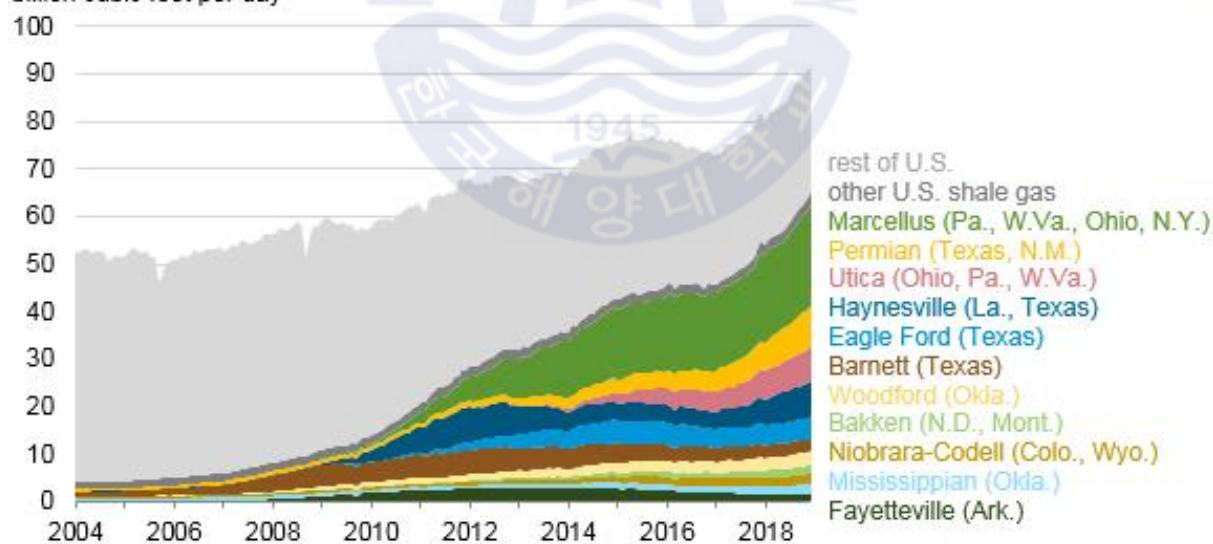
million barrels per day



(a) Oil

Monthly U.S. dry natural gas production (2004-2018)

billion cubic feet per day



(b) Gas

Fig. 32 Monthly U.S. oil and gas production (EIA, 2019d)

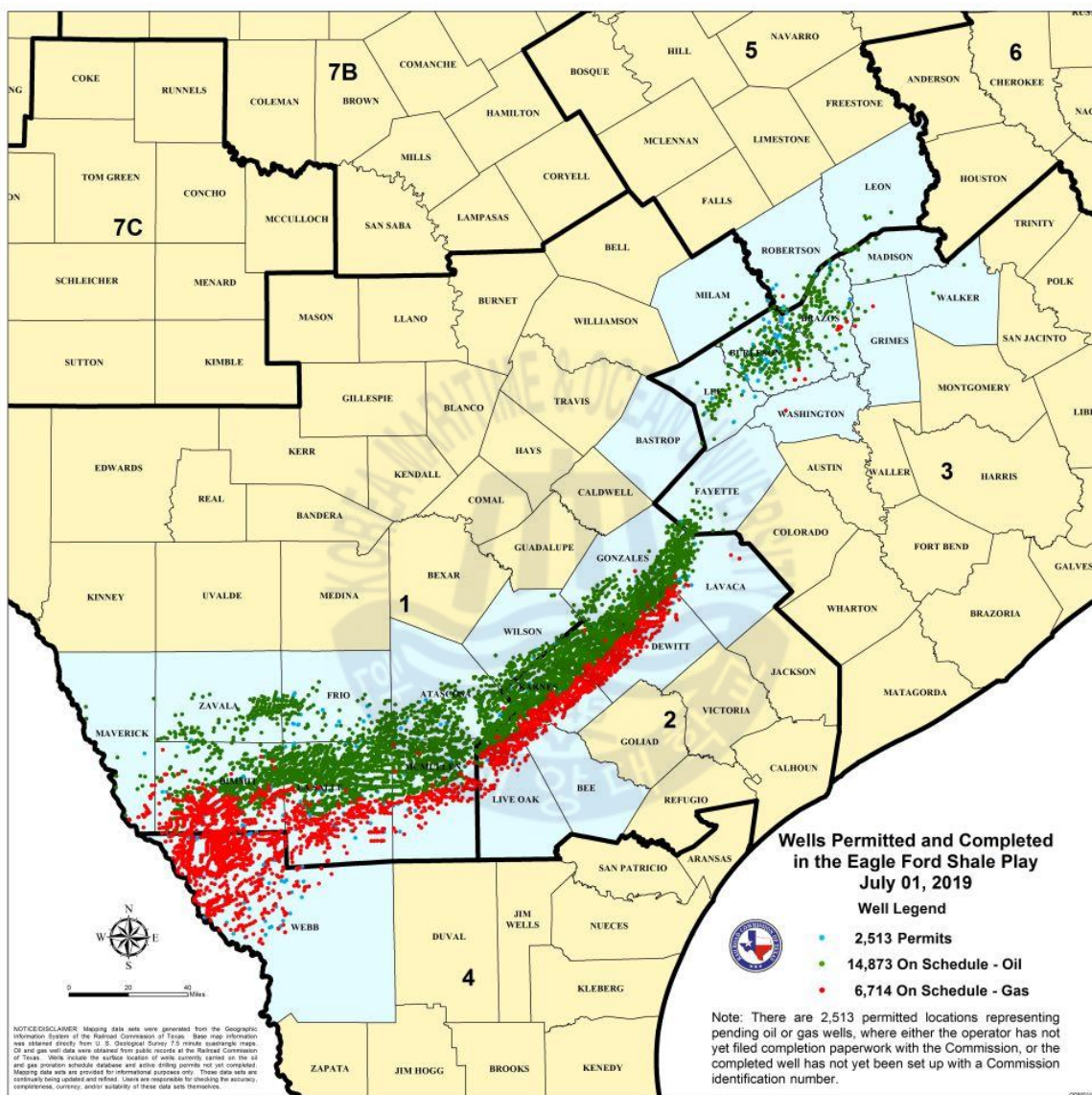


Fig. 33 Well permitted and completed in the Eagle Ford Shale (Texas RRC, 2019)

4.2 셰일층의 석유 생산 주요인자 파악

지질학적 물성 혹은 수압파쇄 설계에 따라 유정들 간 생산성 차이가 발생할 수 있으며(Perrier et al., 2018), Native, Design, Dynamic 속성 중 저류층 물성, 지질학적 정보 등의 취득에 한계가 존재하므로 제한된 자료 중 서로 관련 있는 변수를 파악하여 셰일층의 석유 생산 주요인자(key factors)를 도출하고자 한다.

4.2.1 자료 전처리 과정

현장 자료에 대하여 다수의 생산정 간의 비교가 가능하도록 Lateral length 혹은 Stage 수로 나눔으로써 정규화 하였으며(Fig. 34), Normalized Volume(NV)과 Normalized Production(NP)을 도출하여 Lateral length 길이나 Stage 수에 대하여 동일하게 분포되어 있다는 가정 하에 단위길이 당 인자 값을 구하였다(Luo et al., 2018). Landing direction의 경우 Surface와 Bottom hole에서의 생산정 위치를 통해 방위각을 도출하였으며, 이는 수평정의 방향으로 수압파쇄 균열 생성 시 참고할 수 있다. 또한, 각 속성(attribute)의 Fig. 35와 같은 분포를 통해 매우 작거나 큰 극단값(outlier), 값이 없는 생산정 등 생산 관련 인자에 전처리 과정을 수행하고 분석을 진행하였다.

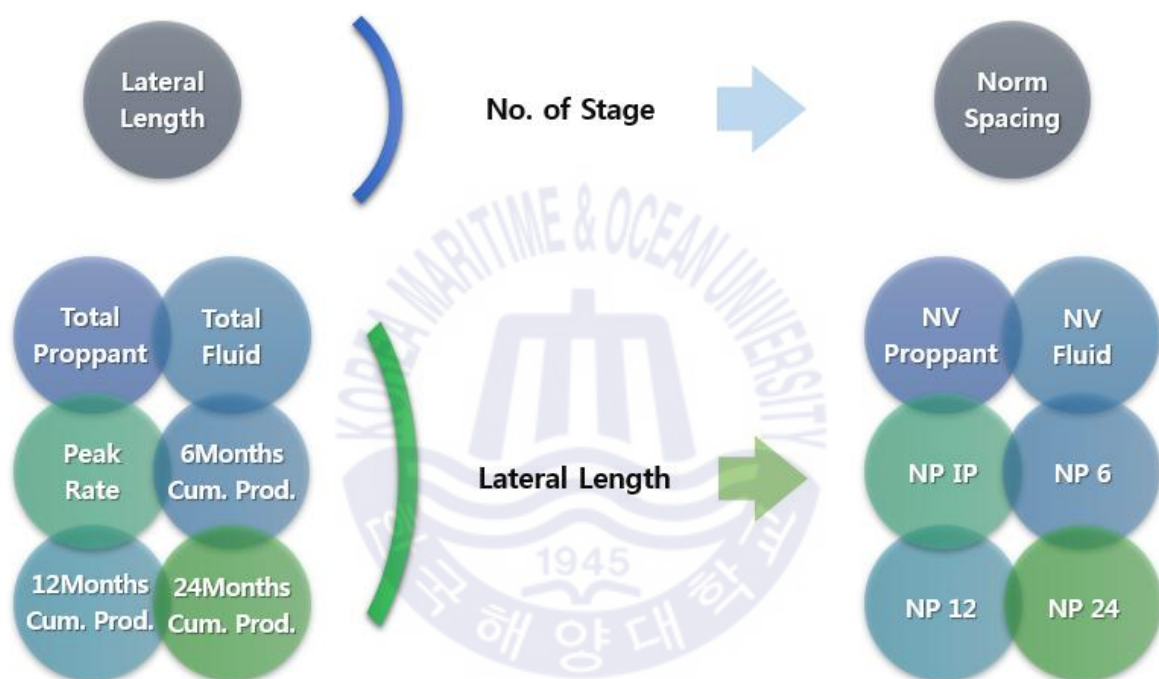
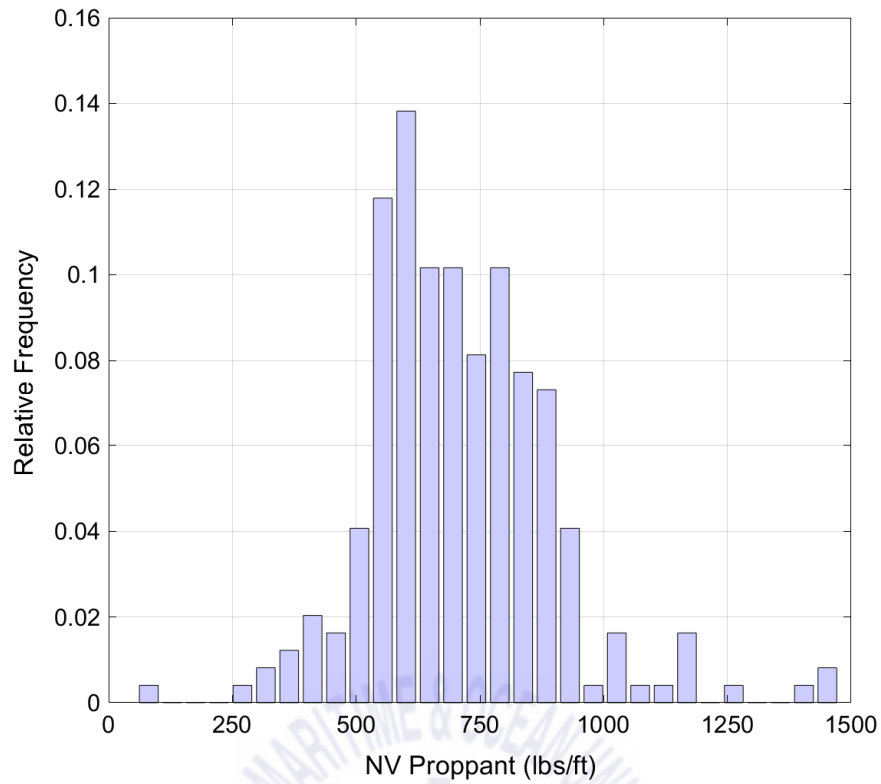
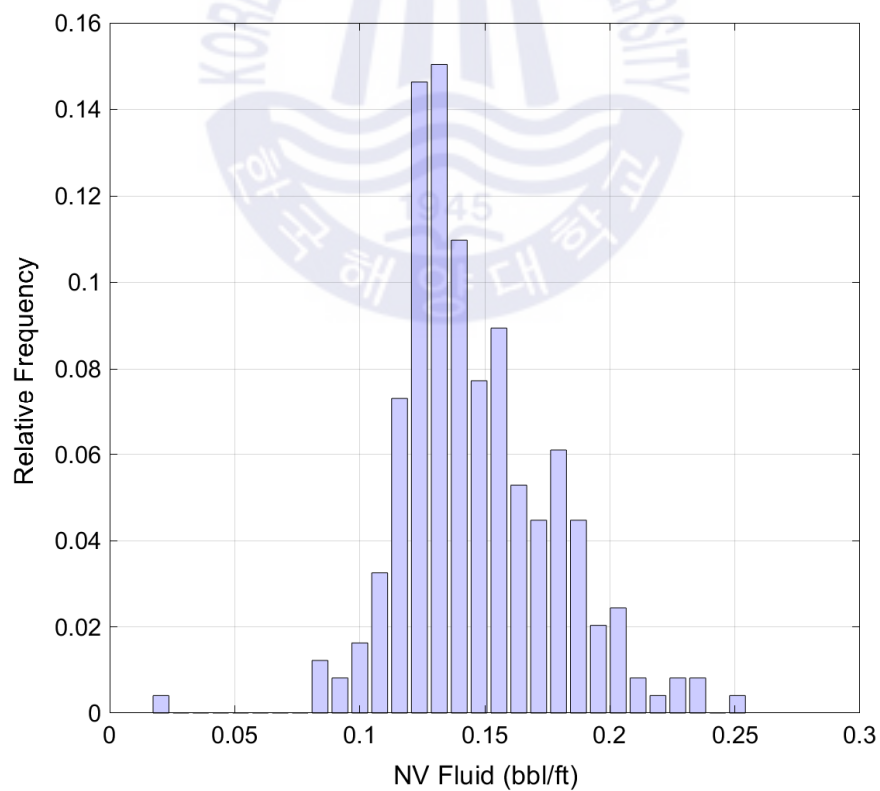


Fig. 34 Data pre-processing for machine learning



(a) NV Proppant



(b) NV Fluid

Fig. 35 Example of histogram to remove data

4.2.2 생산량 자료 분석

선행연구에서는 감퇴곡선인자 추정 시 12개월 이상의 자료를 사용하였으나(Bowie, 2018; Amr et al., 2018), DCA 적용 시 장기간의 생산량 자료를 사용한 경우 안정적인 생산 거동을 보이게 된다(Shin et al., 2015). 따라서 이 연구에서는 Dynamic 속성을 도출하기 위하여 각 생산정마다 동일한 기간인 48개월의 생산 이력자료를 활용하여 Arps 방정식의 쌍곡선함수를 적용하였다(Fig. 36). 또한, 생산기간에 따른 DCA를 수행하였으며, 전체 생산정에 대하여 초기감퇴율과 감퇴지수의 누적확률분포를 도출하였다(Fig. 37). 짧은 생산 기간의 자료를 적용한 경우 비정상적인 결과값은 제외하였으며, 후반부의 생산거동 관련 있는 감퇴지수에 대하여 추정된 결과의 불확실성이 큰 것을 파악할 수 있다.

감퇴곡선인자 중 초기감퇴율과 감퇴지수를 추정하기 위해 비선형 회귀분석을 활용하였다. 비선형 회귀분석은 변량 X로부터 다른 변량 Y를 추정하는 경우에 X에 대한 Y의 평균적인 경향이나 불균형의 정도가 2차 곡선 등 고차의 곡선으로 특정 짓는 관계에 있는 변량군을 분석하기 위한 방법이다. 비선형회귀분석을 적용한 각 생산정 별 DCA 결과를 확인하였으며, 결정계수 값이 약 0.9보다 작거나 생산경향이 맞지 않는 유정을 제외시켰다(Fig. 38). 추정한 감퇴지수의 값이 사용할 수 없는 결과인 경우는 생산 감퇴 경향이 초기에서 급격하게 일어나는 생산정이 많았으며, 생산량의 변동성이 크거나 감퇴경향이 큰 경우 비정상적인 결과를 도출하는 것으로 사료된다.

현장 자료에 대하여 이러한 전처리 과정을 수행한 결과 최종적으로 220개 생산정의 48개월 생산량 자료를 활용하여 DCA를 적용하였으며, 감퇴곡선인자는 Fig. 39와 같은 분포를 보이는 것을 확인할 수 있었다.

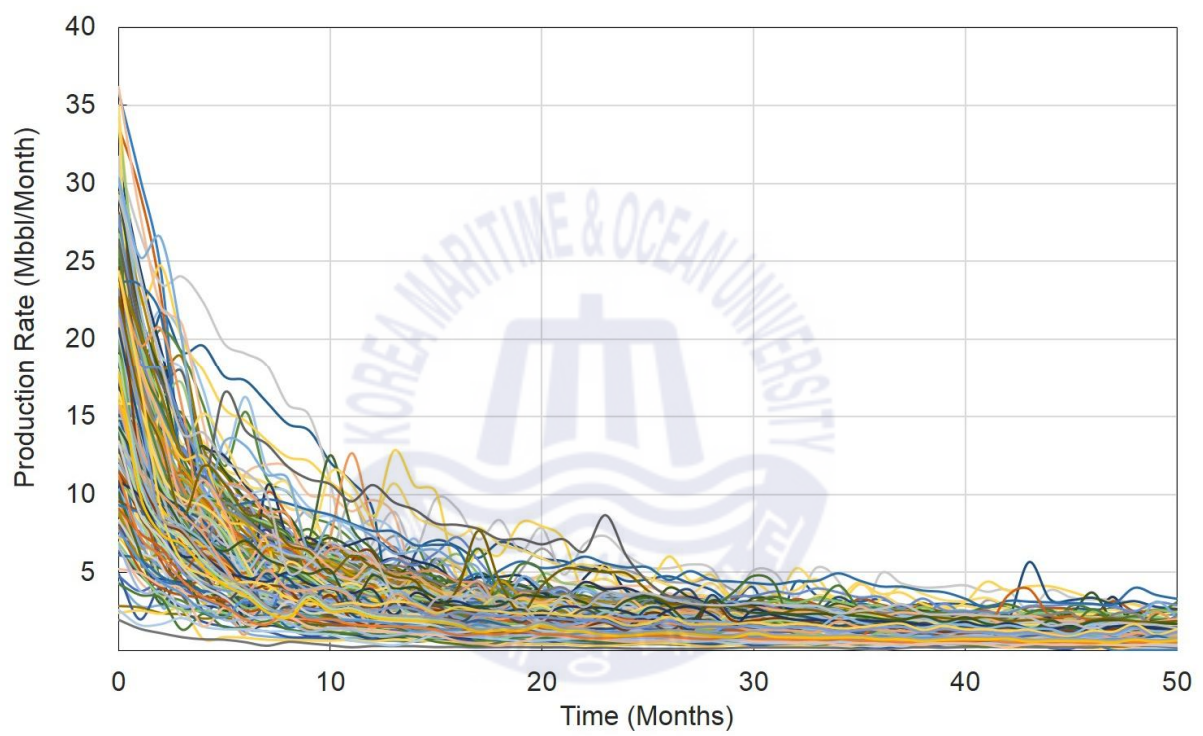
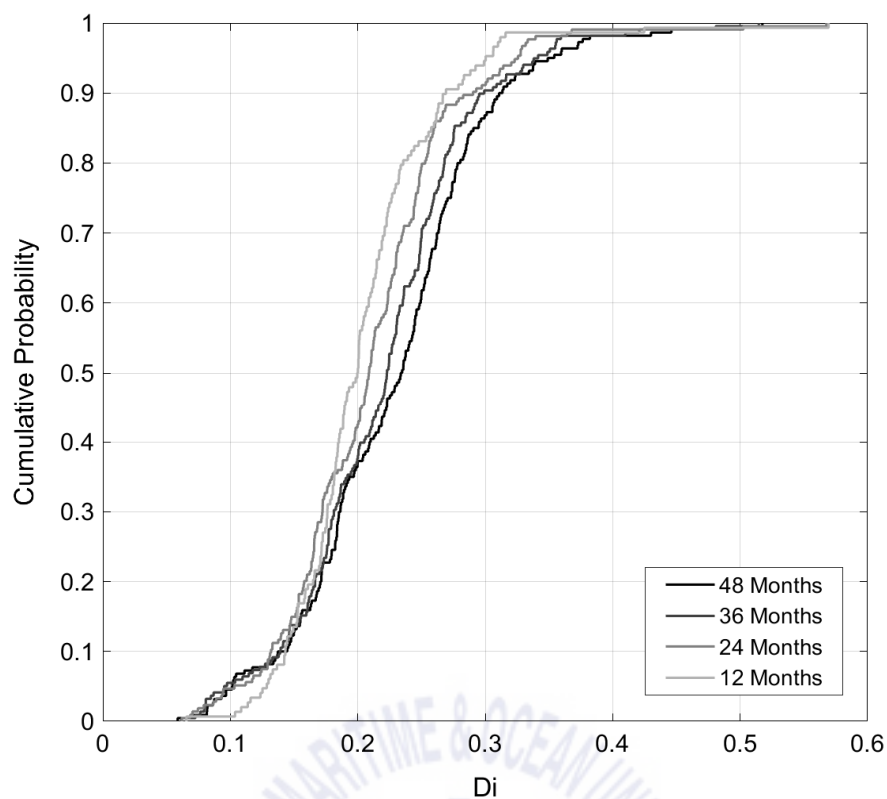
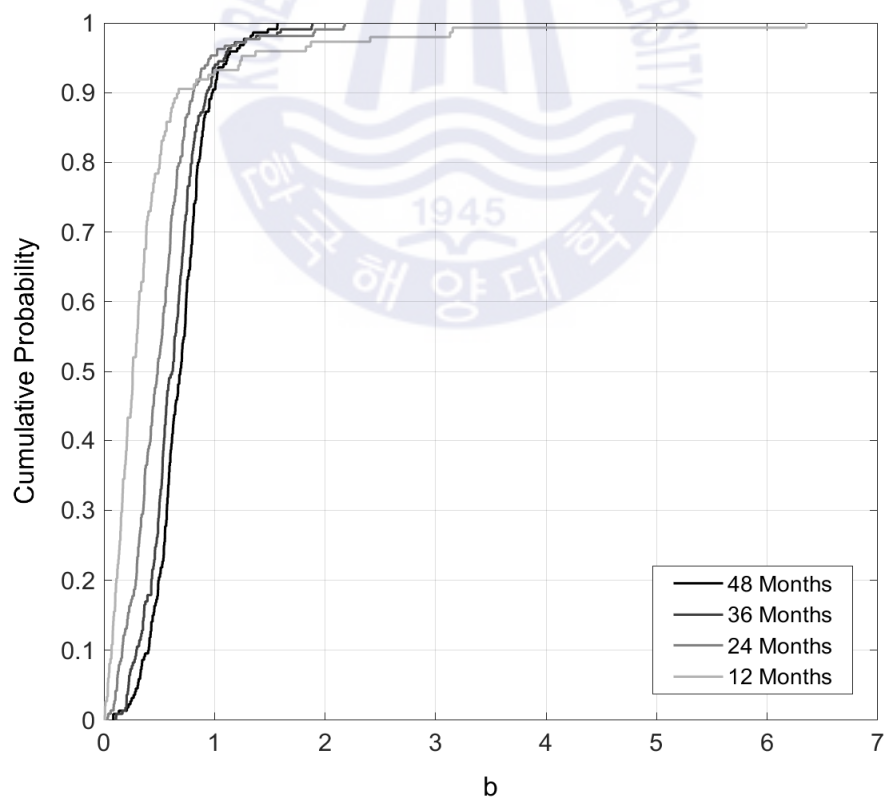


Fig. 36 Production history in study area

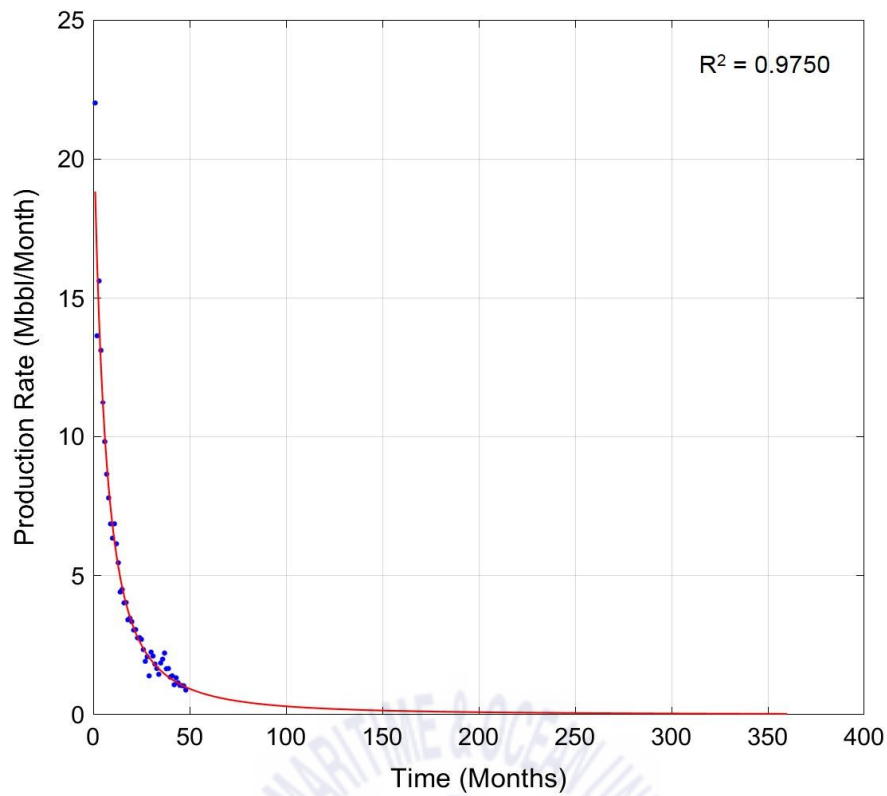


(a) Initial decline rate

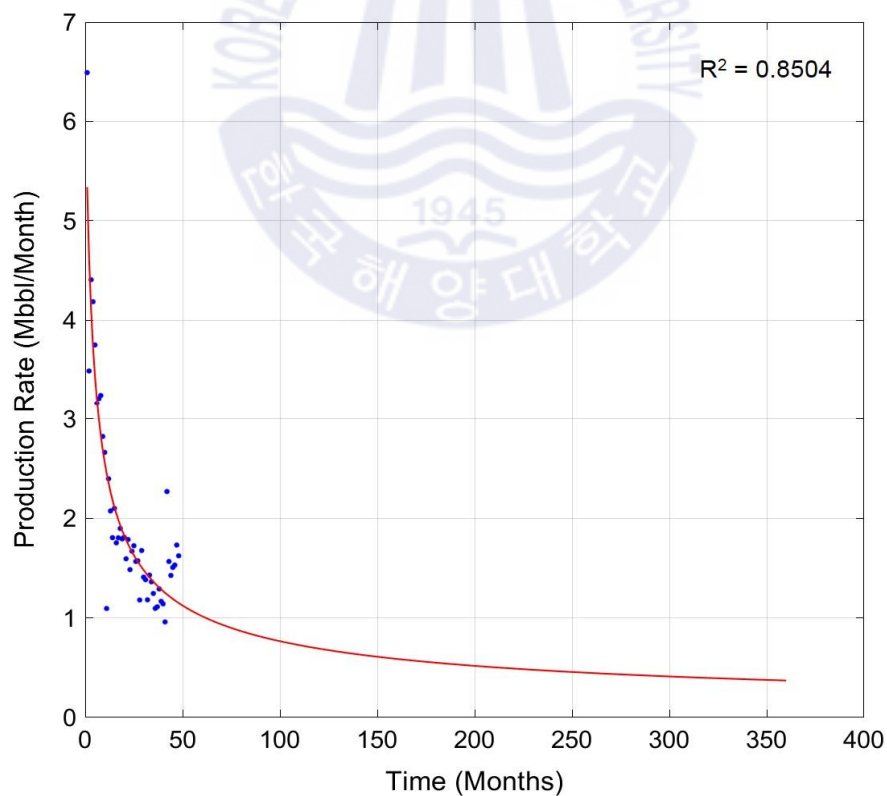


(b) Decline exponent

Fig. 37 Change of decline curve factors according to production period



(a) Matched case



(b) Mismatched case

Fig. 38 Example of production behavior when applying Arps hyperbolic

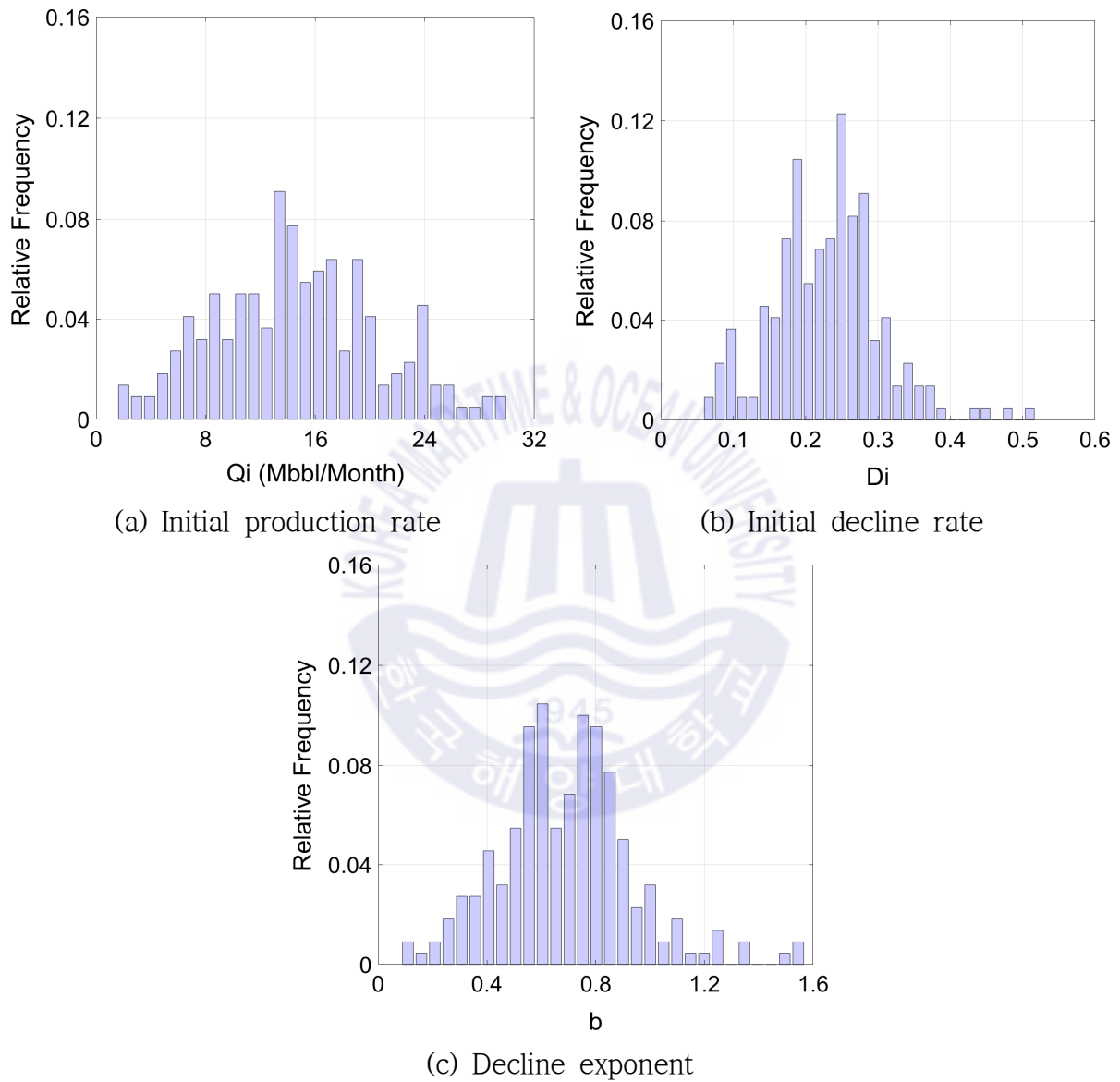


Fig. 39 Histogram of DCA parameters after pre-processing

4.2.3 생산 관련 인자 간 관계 파악

앞서 수행한 전처리 과정 및 DCA 적용 결과를 바탕으로 본 연구에서 분석에 활용하고자 하는 총 22개의 입력 자료를 Table 7에 항목별로 정리하였다. 속성에 대한 관계를 파악하고자 상관관계를 분석하였으며, Fig. 40과 같이 생산정 위치와 관련된 4가지 속성과 생산량과 관련된 5가지 속성의 경우 유사한 인자로 상관관계가 높은 것을 알 수 있다. 이를 통해 상관관계가 없는 인자는 제외하고 경향을 나타내는 인자에 대해서 산점도(scatter plot)를 도시하였으며(Fig. 41), Fig. 42에 상관계수를 나타내었다.

그 결과 Lateral length와 MD는 0.8의 상관관계를 나타냈으며, Stage와 Lateral length, Spacing과 NV Fluid는 0.7 이상의 관계를 보였다. 또한, MD와 Stage, NV Proppant와 Spacing, NV Fluid와 NV Proppant는 0.6 이상의 값을 보였으며, Stage와 Spacing은 0.5 이상의 상관관계를 나타내었다. 그러나 그 외인자들의 경우 상사성(similarity)이 비선형적인 관계를 갖는 것을 파악하였다.

Table 7 List of input attributes in this study

Native	Design	Dynamic
• Surface latitude • Surface longitude • Bottom hole latitude • Bottom hole longitude • Measured depth (MD) • True vertical depth (TVD) • Landing direction • Choke size • County Number • Well Elevation	• Lateral length • HF Stage • Norm Spacing • NV Proppant • NV Fluid	• NP IP • NP 6 Months • NP 12 Months • NP 24 Months • NP Q_i (Arps) • D_i (Arps) • b (Arps)

Surface Latitude	0.9802	0.9982	0.9805
0.9802	Surface Longitude	0.9814	0.9995
0.9982	0.9814	Bottomhole Latitude	0.9807
0.9805	0.9995	0.9807	Bottomhole Longitude

(a) Location

NP IP (bbl/ft)	0.9332	0.8876	0.8462	0.9873
0.9332	NP 6 (bbl/ft)	0.9835	0.9505	0.9766
0.8876	0.9835	NP 12 (bbl/ft)	0.9870	0.9429
0.8462	0.9505	0.9870	NP 24 (bbl/ft)	0.9025
0.9873	0.9766	0.9429	0.9025	Arps NP Qi (bbl/ft)

(b) Production

Fig. 40 Highly correlated input attributes

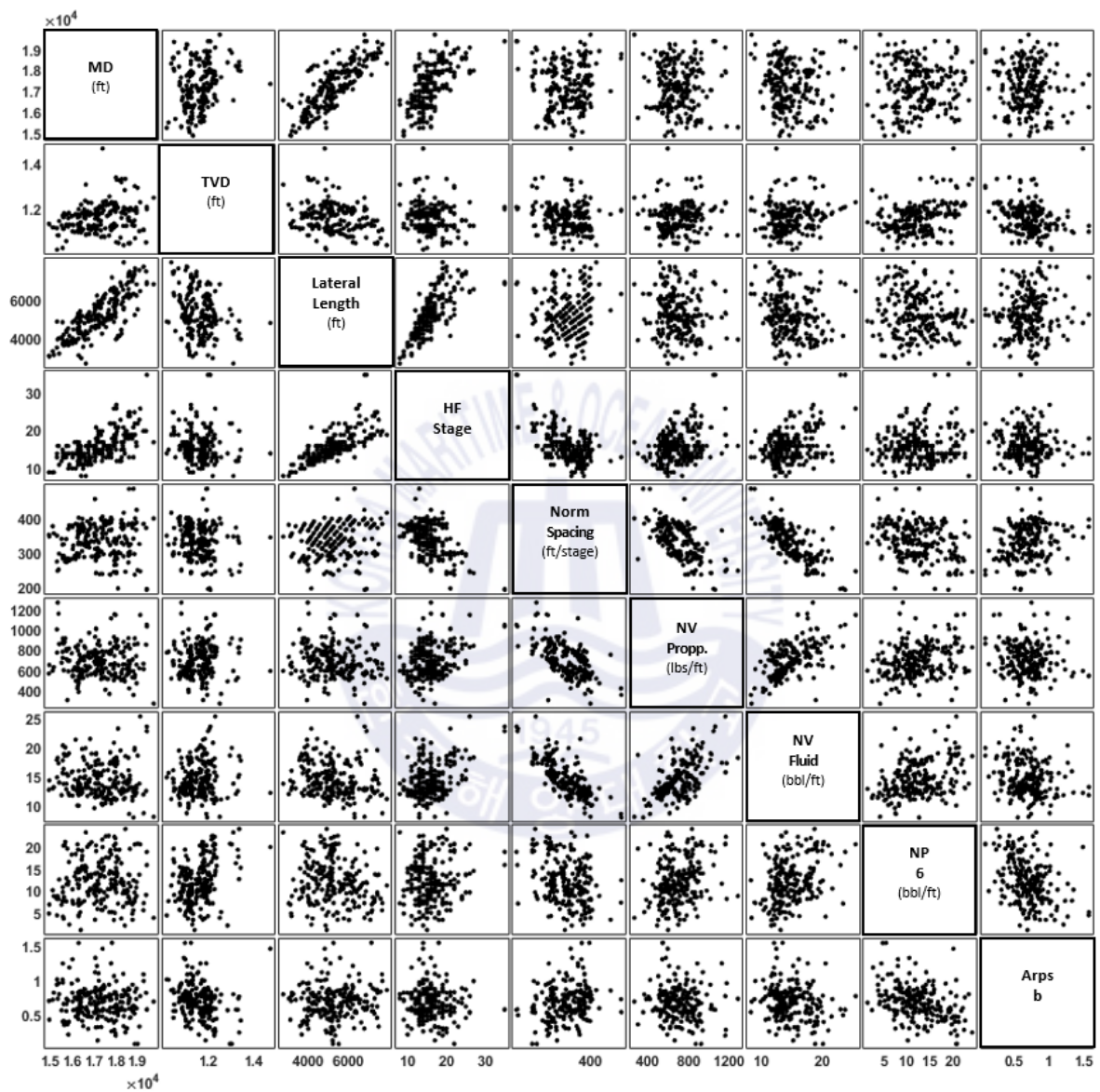


Fig. 41 Scatter plot of input attributes

MD (ft)	0.2381	0.8007	0.6506	0.0551	-0.0323	-0.0514	0.0400	-0.0944
0.2381	TVD (ft)	-0.2840	-0.0953	-0.1668	0.0931	0.1697	0.3748	-0.2502
0.8007	-0.2840	Lateral Lenth (ft)	0.7316	0.1449	-0.1204	-0.1630	-0.1589	0.0884
0.6506	-0.0953	0.7316	HF Stage	-0.5386	0.3155	0.3996	0.1151	-0.0893
0.0551	-0.1668	0.1449	-0.5386	Norm Spacing (ft/stage)	-0.6122	-0.7707	-0.3351	0.2305
-0.0323	0.0931	-0.1204	0.3155	-0.6122	NV Proppant (lbs/ft)	0.6095	0.2894	-0.1197
-0.0514	0.1697	-0.1630	0.3996	-0.7707	0.6095	NV Fluid (bbl/ft)	0.3422	-0.1929
0.0400	0.3748	-0.1589	0.1151	-0.3351	0.2894	0.3422	NP 6 (bbl/ft)	-0.3781
-0.0944	-0.2502	0.0884	-0.0893	0.2305	-0.1197	-0.1929	-0.3781	Arps b

Fig. 42 Correlation coefficient between input attributes

4.2.4 차원 감소 및 주요인자 파악

생산성이 높은 경우는 셰일층의 고유의 특성에 의한 영향이거나 수압파쇄 설계의 성공 정도에 따라 상이하다. 이러한 다차원적인 변수들을 축소, 요약하여 단순화하고 서로 상관되어 있는 반응 변수를 파악하고자 PCA를 수행하였으며, 이러한 결과로부터 셰일층에서의 석유 생산 주요인자를 파악하고자 한다.

이 연구에서는 입력 속성들에 대한 분산의 역을 가중치로 사용하여 PCA를 수행하였으며, 속성들에 대하여 계수 행렬, 점수, 분산 등을 도출할 수 있다. 이 중 계수 행렬은 가중치의 사용으로 정규 직교가 아니므로 새로운 계수 행렬을 계산하여 정규 직교성을 확인하는 과정이 필요하다. 이러한 과정으로부터 구한 주성분 변수에 대하여 선형 계수 혹은 부하(loadings)의 행렬을 도출할 수 있으며, 이 값이 크다는 것은 주성분 변수가 계산될 때 그 변수의 영향이 크다는 것을 의미한다. 부하는 주성분을 계산할 때 입력 속성 각각에 대한 가중치로 주성분 계수로 이해될 수 있다. 즉, 주성분 계수 값이 음인 경우 다른 변수들과 음의 상관관계가 있음을 의미하며, 값이 클수록 변수와의 유사성이 존재함을 일컫는다. 따라서 주성분과 상관관계가 높은 즉 주성분 계수 값이 큰 입력 속성들을 이 연구에서는 주요인자라고 말하고자 한다.

이 연구에서는 PCA 적용 시 주요인자 파악을 위해 입력 속성에 대하여 Case study를 수행하였다. 총 22개의 인자 중에서 생산정 위치, 5가지의 생산량 값과 같이 상관관계가 0.9이상으로 유의한 경우 중복된 상태로 볼 수 있으며, 이에 대한 고려를 위해 위치의 경우 Surface latitude와 Surface longitude, 생산량은 NP 6 외의 속성을 제거하고 16개 속성에 대한 Case 1의 분석을 수행하였다. 그 결과 Table 8과 같으며, 3개의 주성분을 기준으로 보았을 때 전체 속성의 약 51%를 나타내었다(Fig. 43). 이를 통해 County Num, Well elevation, Landing direction, Choke size의 주성분 계수 값이 다른 속성에 비해 적은 것을 파악하였다.

Table 8 Results of PCA in Case 1

Attributes	PC 1	PC 2	PC 3
County Num.	0.0724	0.0509	0.2144
Surface Latitude	0.2887	-0.3061	0.3049
Surface Longitude	0.3213	-0.3079	0.3361
Well Elevation (ft)	-0.0669	0.1389	0.1378
Landing Direction (degree)	0.0053	0.0434	-0.0523
Measured Depth (ft)	-0.0023	0.4124	0.4456
True Vertical Depth (ft)	0.2920	-0.1146	0.2904
Lateral Length (ft)	-0.1610	0.4578	0.2973
HF Stage	0.1510	0.5206	0.0575
Norm Spacing (ft/stage)	-0.4144	-0.1905	0.2951
NV Proppant (lbs/ft)	0.3269	0.1816	-0.3031
NV Fluid (bbl/ft)	0.4102	0.1588	-0.2872
NP 6 (bbl/ft)	0.3758	-0.0168	0.0808
Choke Size	0.0411	0.1513	-0.0759
D _i (Arps)	-0.1248	-0.0685	-0.2429
b (Arps)	-0.2534	-0.0405	-0.1382
Eigenvalue	3.2448	2.8752	1.9909

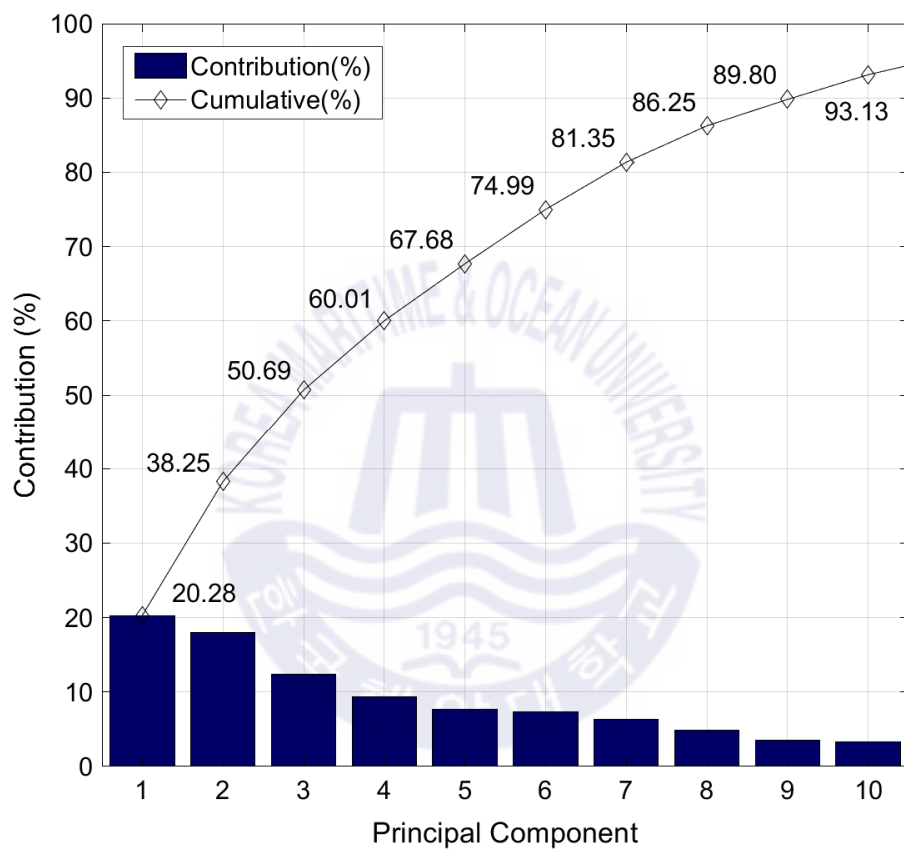


Fig. 43 Scree plot of 16 attributes in Case 1

Case 1에서 주성분과의 상관관계가 적은 4개의 속성을 제외하고 12개 속성에 대한 Case 2의 결과를 Table 9에 나타내었으며, 3개의 주성분이 전체 속성의 약 76%를 표현하였다(Fig. 44). Case 3에서는 주성분 계수가 상대적으로 작은값을 가지는 감퇴곡선인자 2가지와 모든 주성분에 대해 유사한 범위의 계수 값을 나타내는 생산정 위치를 제외시킨 결과를 도출하였으며(Table 10), Fig. 45와 같이 8개 속성의 약 82%를 나타냄으로써 Case 1과 2에 비해 주요 속성에 대한 영향을 고려함을 파악하였다.

Table 9 Results of PCA in Case 2

Attributes	PC 1	PC 2	PC 3
Surface Latitude	0.2882	-0.3031	0.3510
Surface Longitude	0.3203	-0.3087	0.3715
Measured Depth (ft)	-0.0042	0.4117	0.4525
True Vertical Depth (ft)	0.2914	-0.1363	0.2392
Lateral Length (ft)	-0.1627	0.4710	0.3386
HF Stage	0.1510	0.5397	0.0983
Norm Spacing (ft/stage)	-0.4170	-0.2035	0.2820
NV Proppant (lbs/ft)	0.3314	0.1831	-0.3318
NV Fluid (bbl/ft)	0.4126	0.1679	-0.2847
NP 6 (bbl/ft)	0.3790	-0.0307	0.0900
D _i (Arps)	-0.1290	-0.0788	-0.2408
b (Arps)	-0.2568	-0.0367	-0.1253
Eigenvalue	3.2167	2.7887	1.9383

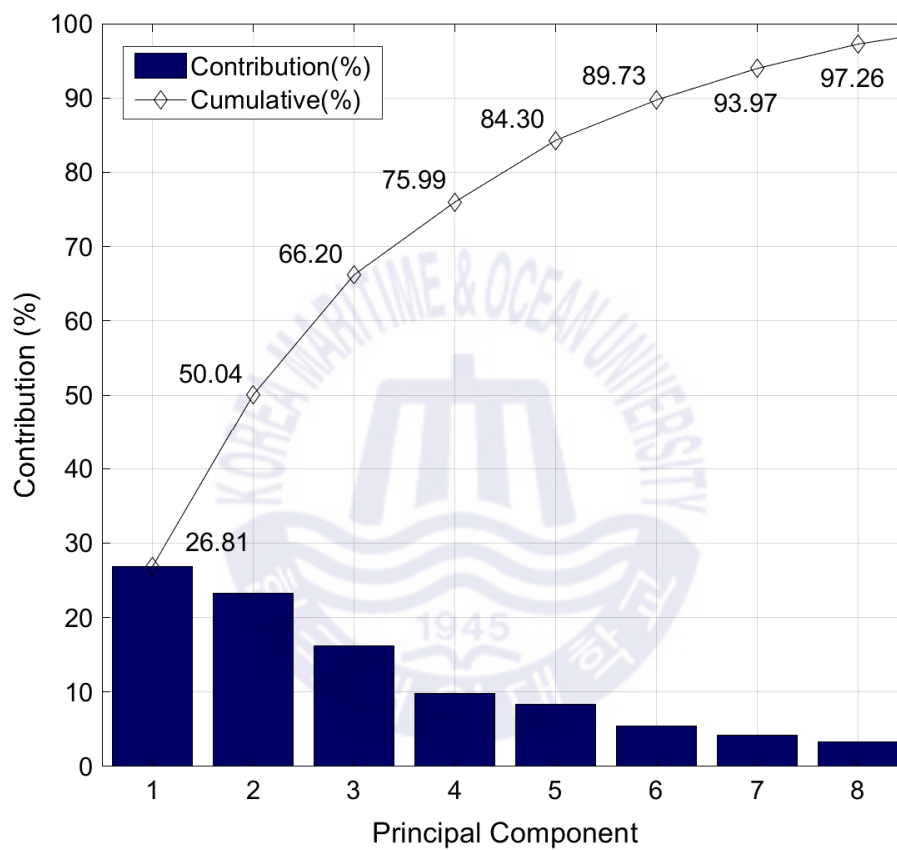


Fig. 44 Scree plot of 12 attributes in Case 2

Table 10 Results of PCA in Case 3

Attributes	PC 1	PC 2	PC 3
Measured Depth (ft)	0.1593	0.5382	0.3643
True Vertical Depth (ft)	0.1574	-0.1589	0.7494
Lateral Length (ft)	0.0680	0.6279	-0.0454
HF Stage	0.4099	0.4297	-0.1305
Norm Spacing (ft/stage)	-0.5103	0.1452	0.1781
NV Proppant (lbs/ft)	0.4347	-0.1551	-0.1944
NV Fluid (bbl/ft)	0.4893	-0.1775	-0.1530
NP 6 (bbl/ft)	0.2974	-0.1719	0.4400
Eigenvalue	2.9033	2.4117	1.2711

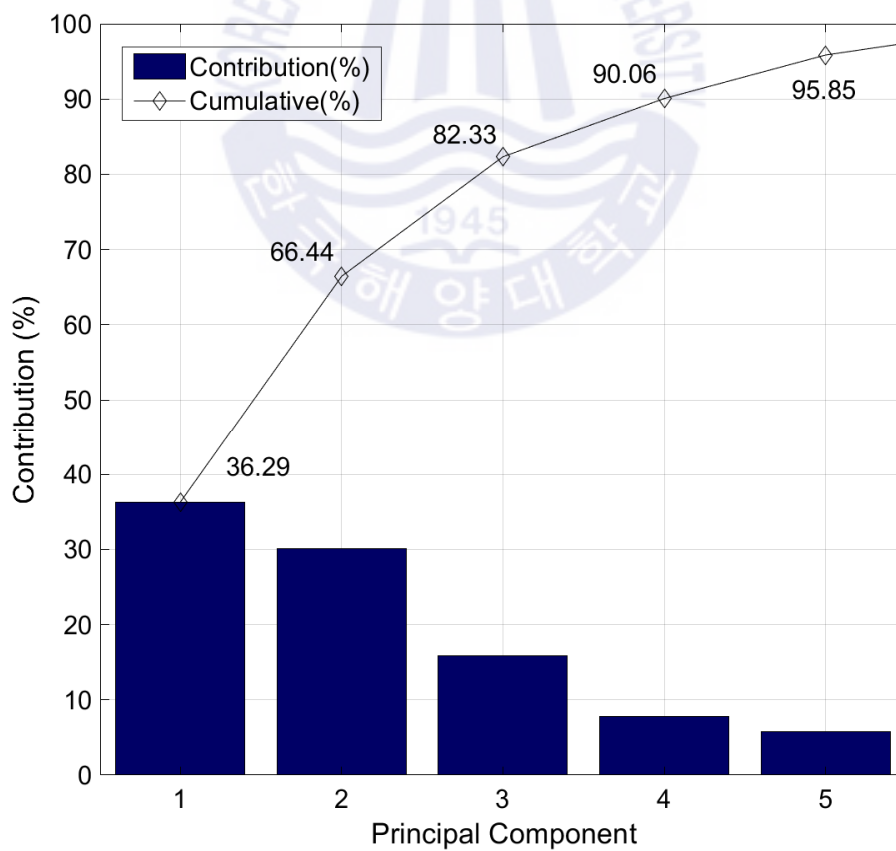


Fig. 45 Scree plot of 8 attributes in Case 3

영향을 미치는 주요 속성에 대한 분석을 바탕으로 정규화에 사용한 Lateral length 제외하고 Case 4에서는 생산정 깊이와 수압파쇄 설계인자 간의 영향을 파악하고자 PCA를 수행하였으며, Table 11과 같이 5개의 제한된 자료만을 사용하여 3개의 주성분이 속성의 약 88%로 전체 속성 중 많은 부분을 표현함을 파악하였다(Fig. 46). 그 결과 식 (16) ~ (18)과 같이 주성분 1은 수압파쇄 설계인자 중 Spacing, NV Fluid, NV Proppant, 주성분 2는 TVD, 주성분 3은 Stage 수와 주된 상관관계가 있었으며, 주성분과 각 속성간의 영향정도를 Fig. 47에 나타내었다.

이를 통해 연구대상 지역에서의 취득 가능한 자료 중 석유 생산 주요인자는 생산정 깊이, 수압파쇄 설계인자임을 파악하였다. 또한, 연구 대상 지역 혹은 취득한 자료에 따라서 PCA를 적용 결과에 따른 주요인자가 상이할 수 있으며, 파악한 주요인자를 바탕으로 그룹 분류에 활용 가능할 것이다.

Table 11 Results of PCA in Case 4

Attributes	PC 1	PC 2	PC 3
True Vertical Depth (ft)	0.1146	0.8905	0.3957
HF Stage	0.3921	-0.4446	0.7112
Norm Spacing (ft/stage)	-0.5601	-0.0111	-0.0583
NV Proppant (lbs/ft)	0.4803	0.0396	-0.5525
NV Fluid (bbl/ft)	0.5374	0.0876	-0.1702
Eigenvalue	2.6757	1.0705	0.6504

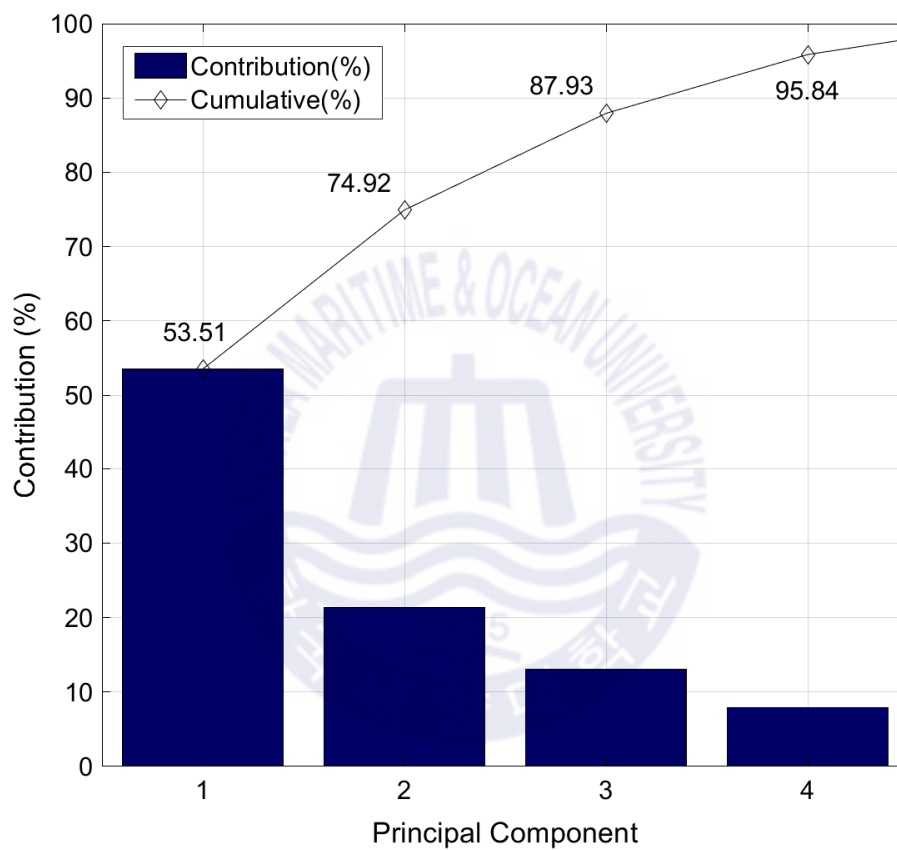


Fig. 46 Scree plot of 5 attributes in Case 4

$$PC1 = 0.1146 TVD + 0.3921 Stage - 0.5601 Spacing + 0.4803 NVP + 0.5374 NVF \quad (16)$$

$$PC2 = 0.8905 TVD - 0.4446 Stage - 0.0111 Spacing + 0.0396 NVP + 0.0876 NVF \quad (17)$$

$$PC3 = 0.3957 TVD + 0.7112 Stage - 0.0583 Spacing - 0.5525 NVP - 0.1702 NVF \quad (18)$$

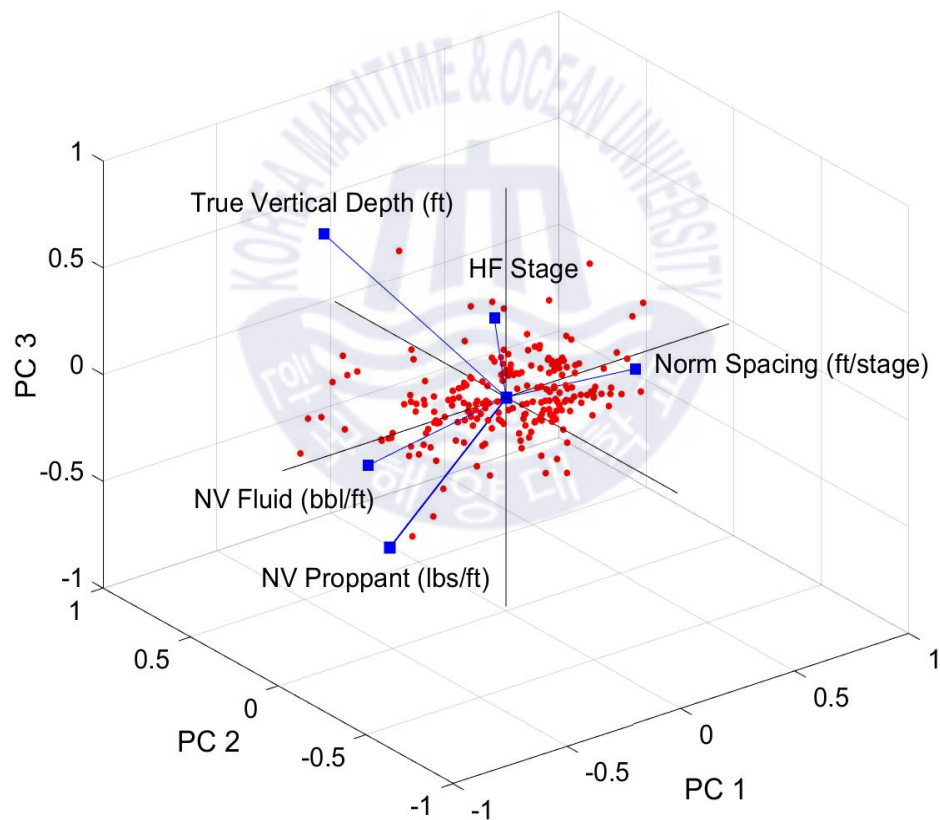


Fig. 47 Visualization of PCs in 3D plot in Case 4

4.3 주요인자를 활용한 생산특성 분석

동일 분지라도 불균질한 셰일층의 특성과 수압파쇄 적용에 따른 생산경향이나 누적 생산량이 상이하다. 이로 인해 데이터 기반 분석을 수행할 때, 특징이 서로 다른 전체 자료를 이용할 경우 예측의 신뢰성이 낮아질 수 있다. 또한, 선행연구에서 유정마다 다른 생산성에 대한 고려를 위해 생산량을 기준으로 그룹을 나누었으며, 셰일층 특성 혹은 수압파쇄 설계인자 등이 생산량 그룹에 미치는 영향을 분석하였으나 도출되는 특징이 그룹별로 유사하게 나타났다(Fig. 48; Fig. 49). 따라서 본 연구에서는 해석의 효율성을 위해 새로운 차원에 대한 PCA 결과를 이용하여 그룹을 분류하고자 하며, 이로부터 도출된 그룹별 생산특성을 분석하고자 한다.

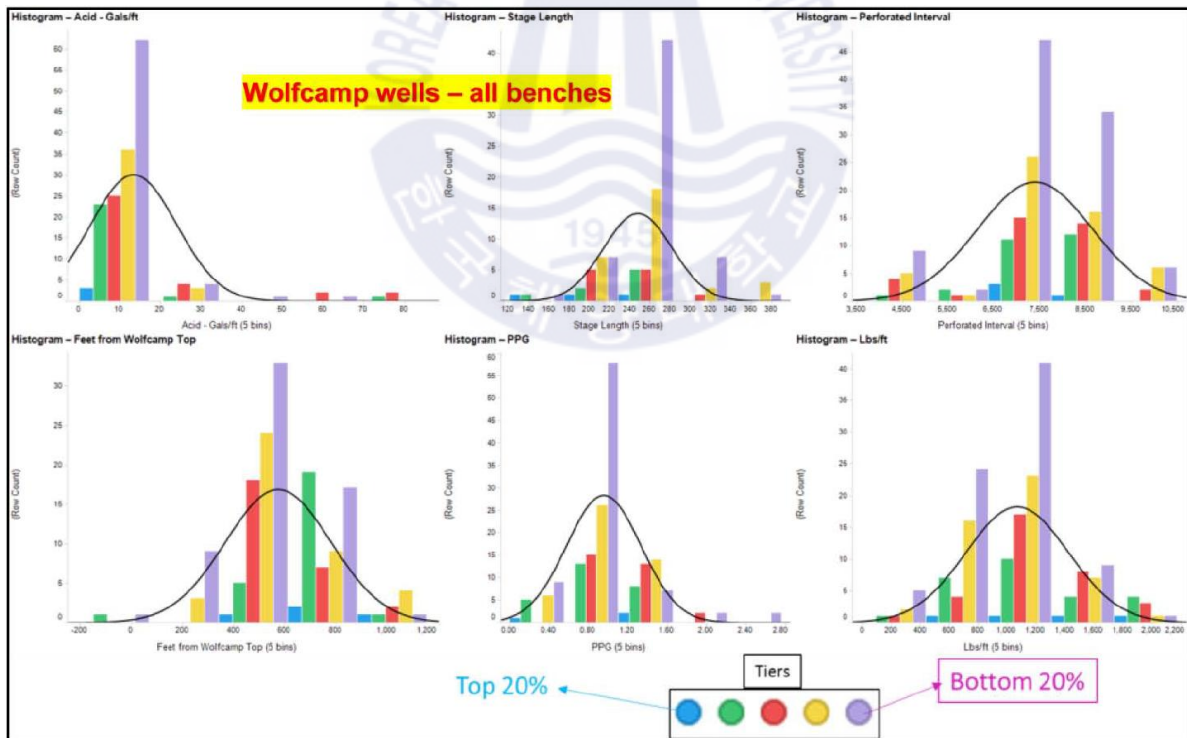


Fig. 48 Variable distribution of the five clusters based on 6months production (Gaurav, 2017)

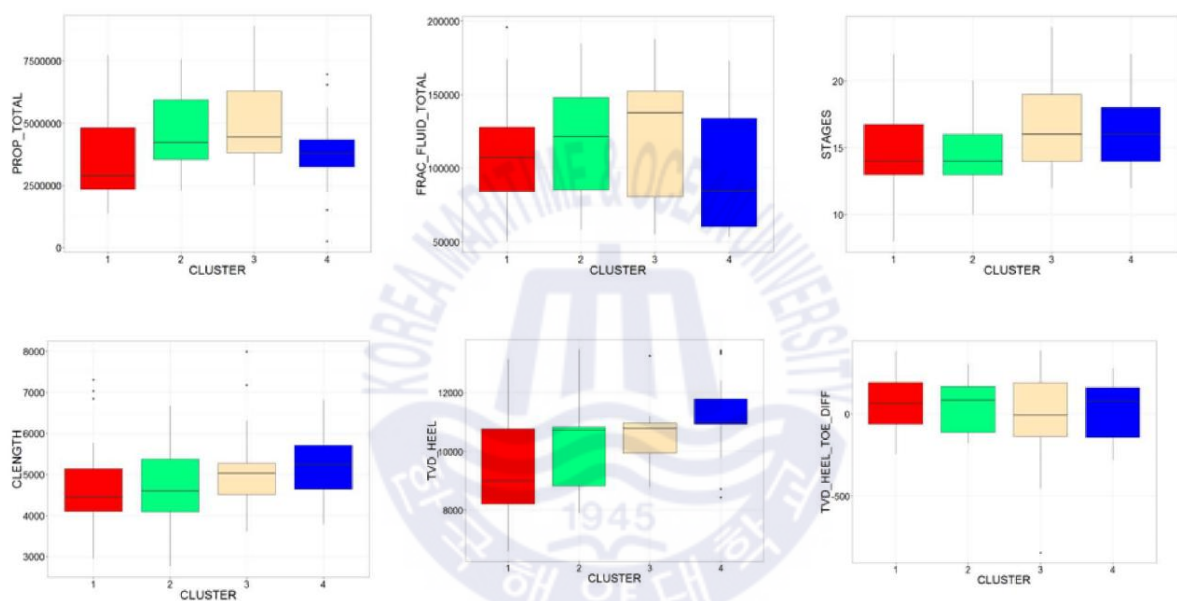
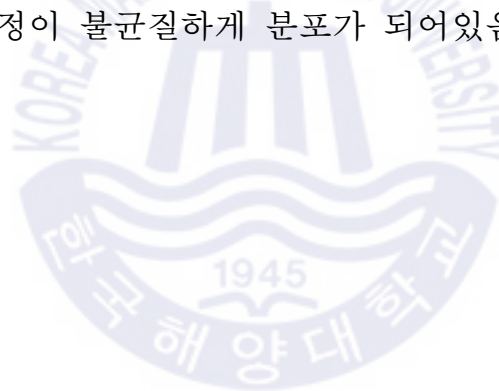


Fig. 49 Variable distribution of the four clusters based on quartile (Vyas et al., 2017)

4.3.1 그룹 분류를 위한 군집분석

PCA 결과를 바탕으로 그룹을 분류하고자 하나의 개체가 여러 군집에 속할 가능성으로 표현되는 퍼지군집분석을 수행하였으며, 대표적인 퍼지군집분석 기법 중 하나인 퍼지 c-평균 기법을 사용하였다.

이 연구에서는 PCA 분석 시 주요인자로 구성된 Case 4의 3개의 주성분을 활용하여 총 10회의 퍼지군집분석을 수행하였으며, 10회 모두 Fig. 50과 같이 약 12회에서 목적함수가 일정한 값에 수렴하는 것을 확인하였다. 그룹 속성 개수에 대하여 Table 12에 나타내었으며, 3개의 그룹으로 분류하였다(Fig. 51). 이를 통해 각 그룹은 수압파쇄 설계인자 중 Spacing, NV Fluid, NV Proppant와 주된 상관관계가 있는 주성분 1에 의한 분류가 뚜렷한 것을 파악할 수 있다. 또한, 그룹이 분류된 생산정의 지도를 도시하였으며(Fig. 52), 세일층의 특성에 의해 같은 그룹의 생산정이 불균질하게 분포가 되어있음을 파악할 수 있다.



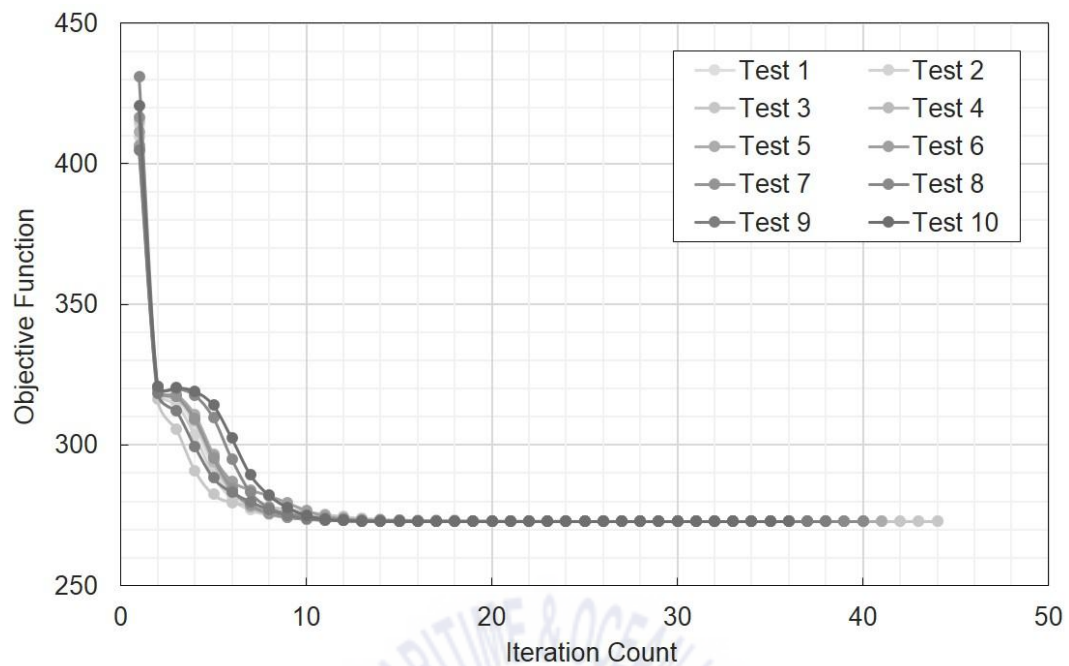


Fig. 50 Convergence of objective functions in fuzzy c-means clustering

Table 12 Summary of fuzzy c-means clustering

	Cluster 1	Cluster 2	Cluster 3
Test 1	74	50	96
Test 2	96	74	50
Test 3	74	50	96
Test 4	50	96	74
Test 5	74	50	96
Test 6	96	50	74
Test 7	50	96	74
Test 8	96	74	50
Test 9	50	96	74
Test 10	50	96	74

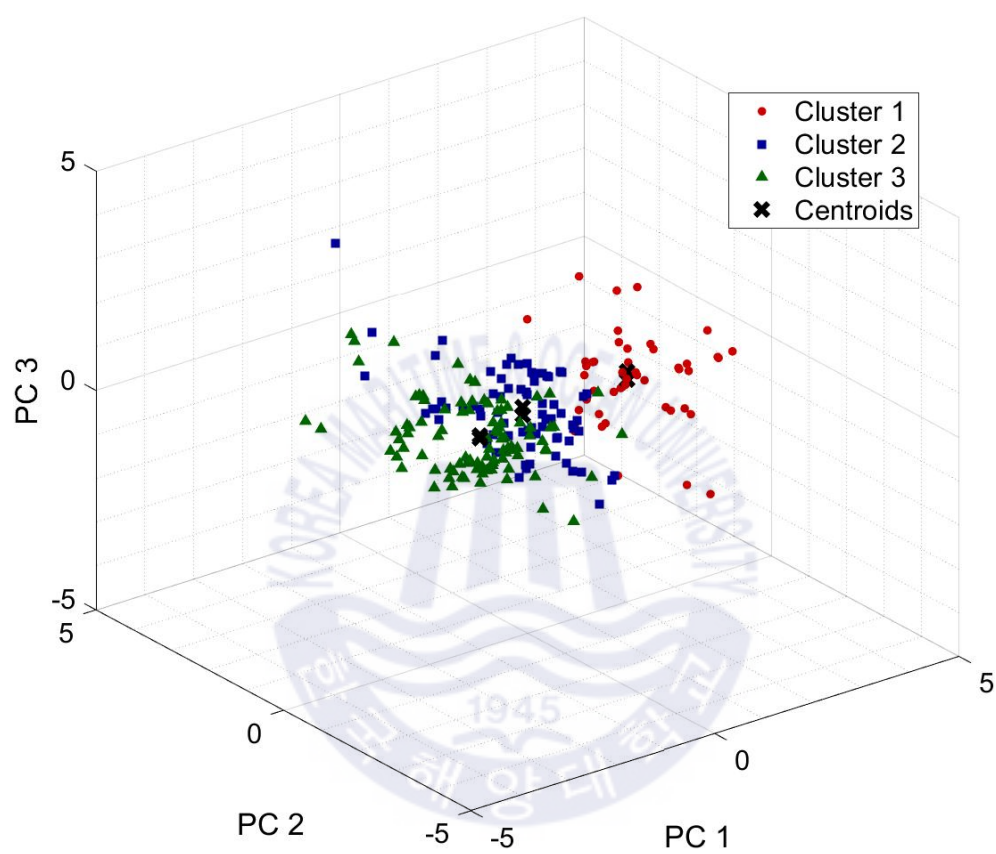


Fig. 51 Cluster analysis results

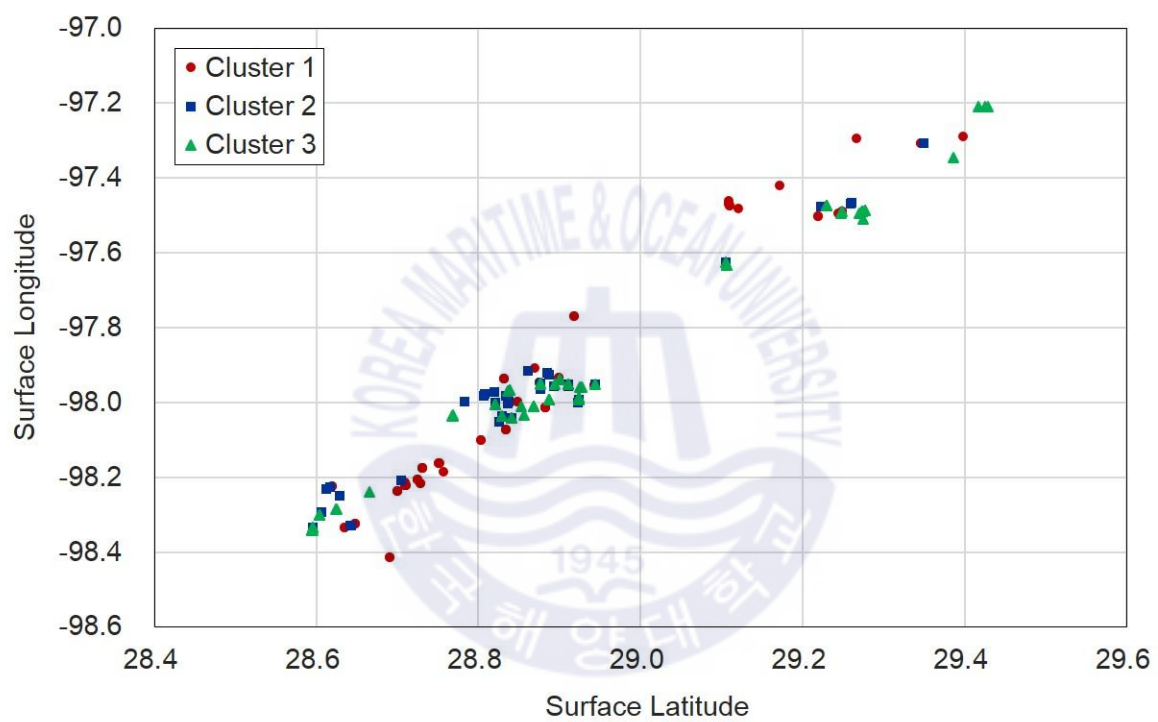


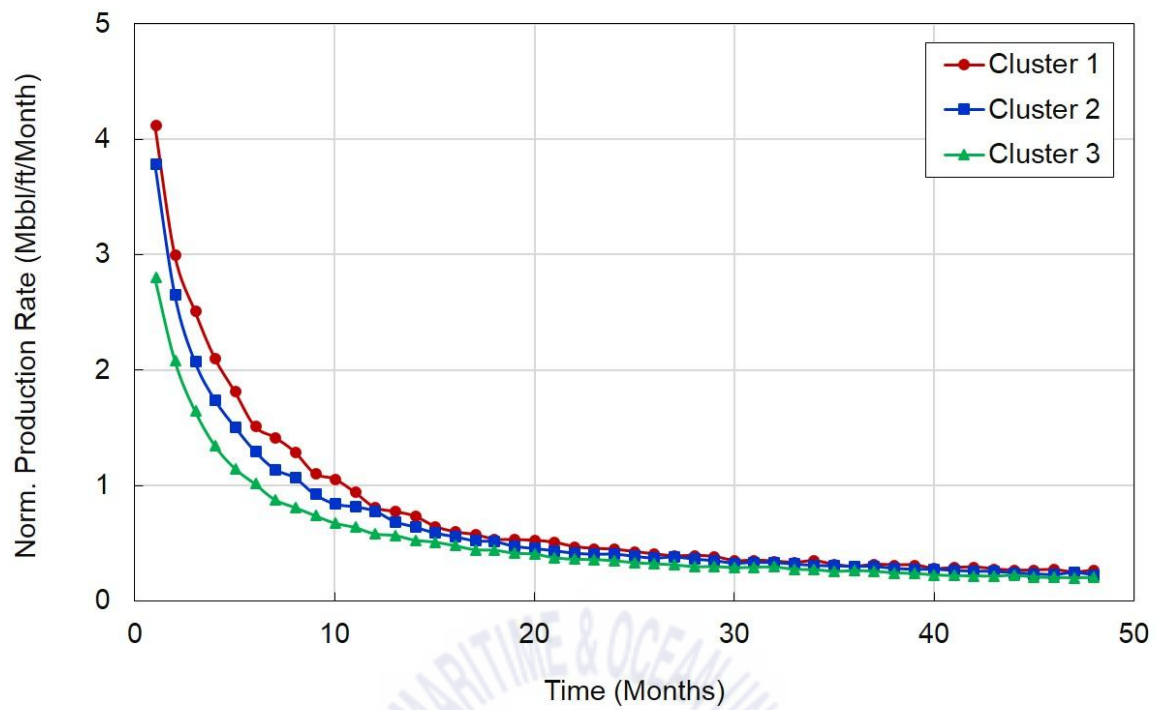
Fig. 52 Location map by cluster

4.3.2 그룹별 생산특성 및 특징 분석

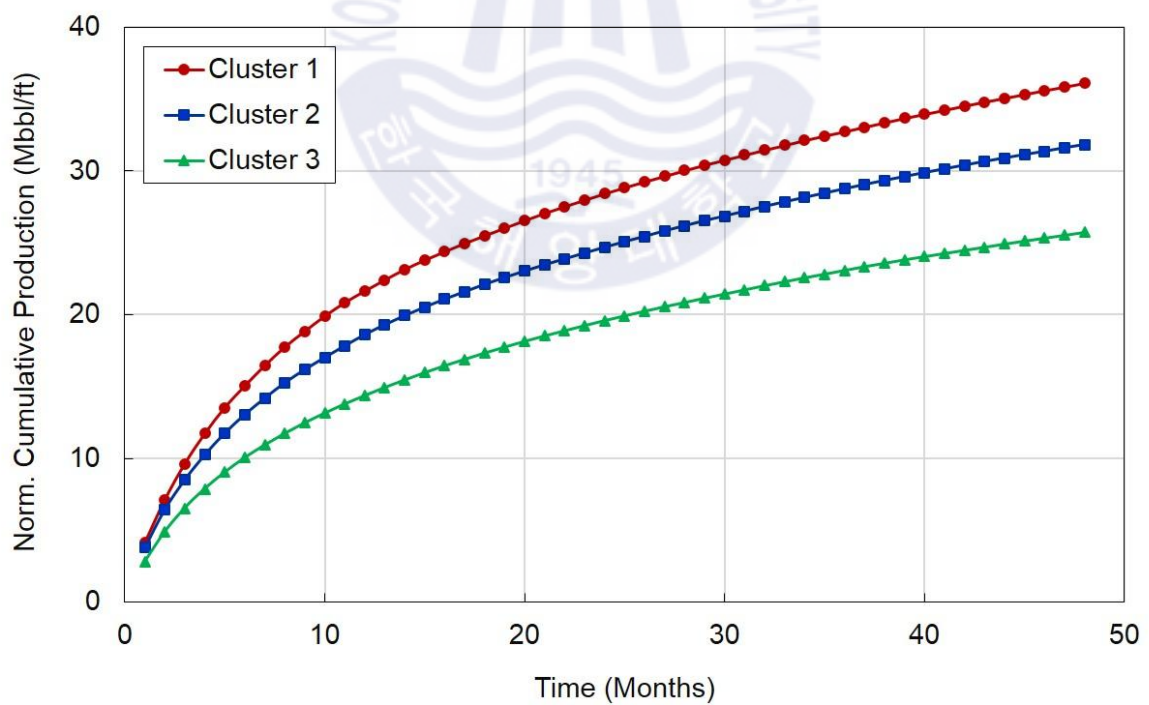
생산정 간의 비교를 위해 정규화한 단위길이 당 생산량을 바탕으로 그룹별 평균 생산량과 누적 생산량을 파악하고자 하였으며(Fig. 53), Cluster 1, 2, 3 순으로 초기생산량 및 누적 생산량이 높게 나타났다. 또한, 약 15개월까지의 감퇴 경향이 그룹별로 차이가 났으며, 생산 후반부에서는 유사한 것을 파악할 수 있다.

그룹에 대한 특징 도출을 위해 각 그룹별 주요인자의 통계적 분석을 수행하여 주요인자의 평균값을 도출하였으며(Table 13), 이로부터 도출된 단위길이 당 누적 생산량을 Table 14에 나타내었다. 또한, 그룹에 대하여 주요인자 항목별로 평균, 최대값, 일사분위, 중간값, 삼사분위, 최소값의 특징을 파악할 수 있는 박스플롯을 도출하였다(Fig. 54). 이로부터 Spacing이 좁고, 수압파쇄 시 단위길이 당 주입 프로판트 및 유체 양이 많은 경우 생산량이 높았다. 또한, Cluster 1과 2는 TVD가 유사하지만 수압파쇄 설계인자의 영향으로 생산성이 좋은 유정이 포함되어 있으며, Cluster 2와 3은 Stage 수가 유사하나 수압파쇄 설계인자에 의해 단위길이 당 누적 생산량이 큰 편으로 나타났다.

생산량으로 그룹을 분류한 선행연구 결과와 달리 세일층의 주요인자를 이용하여 그룹을 나눈 결과 수압파쇄 설계인자의 범위가 서로 다른 특징을 보였으며, 분석 지역 대하여 생산성이 높은 현장 조건을 파악할 수 있었다. 이는 추후 새로운 유정에 대하여 생산특성이 다른 그룹 분류 시 특징 값을 활용할 수 있을 것으로 판단된다.



(a) Average production rate



(b) Average cumulative production

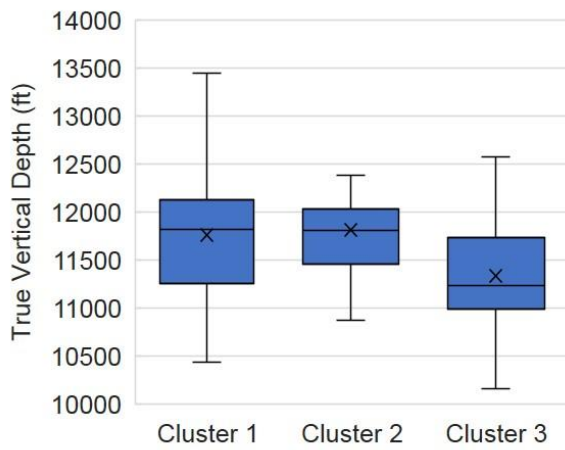
Fig. 53 Normalized Production of each group

Table 13 Mean value of key factors for each group

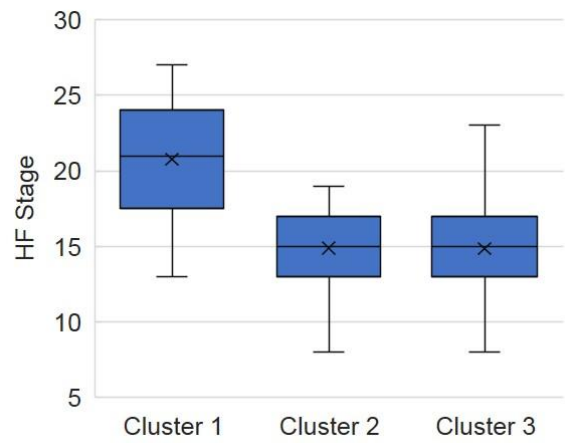
Attributes	Cluster 1	Cluster 2	Cluster 3
True Vertical Depth (ft)	11760.16	11814.96	11335.57
HF Stage	20.76	14.88	14.84
Norm Spacing (ft/stage)	270.95	327.45	373.22
NV Proppant (lbs/ft)	877.65	739.28	592.64
NV Fluid (bbl/ft)	18.49	14.98	12.57

Table 14 Mean value of production on the effect of key factors by group

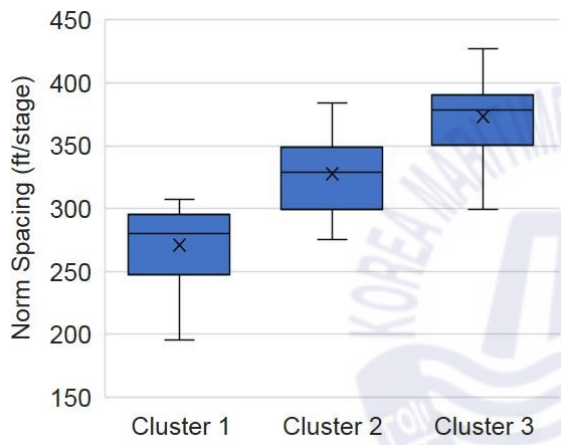
Attributes	Cluster 1	Cluster 2	Cluster 3
NP IP (bbl/ft)	4.18	3.85	2.83
NP 6 (bbl/ft)	15.19	13.20	10.15
NP 12 (bbl/ft)	21.77	18.80	14.48
NP 24 (bbl/ft)	28.44	24.94	19.71



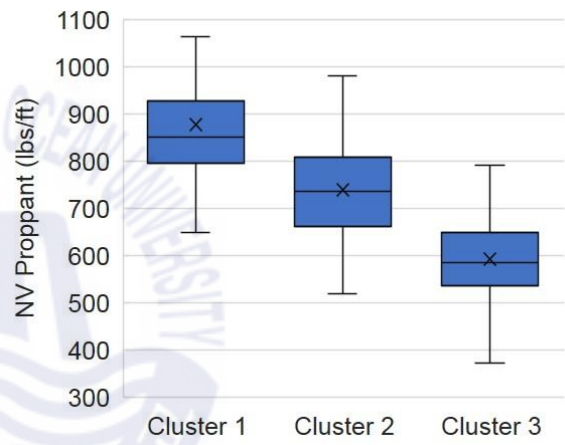
(a) True Vertical Depth



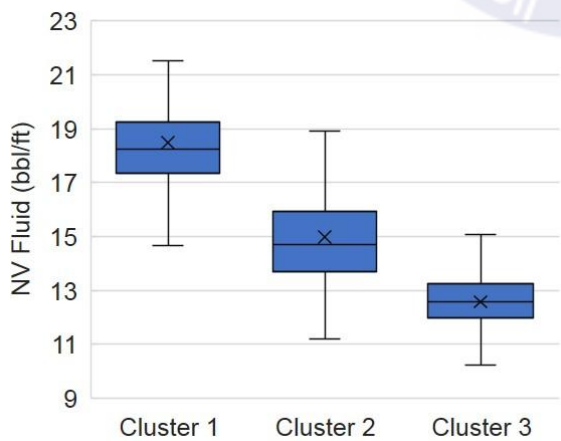
(b) HF Stage



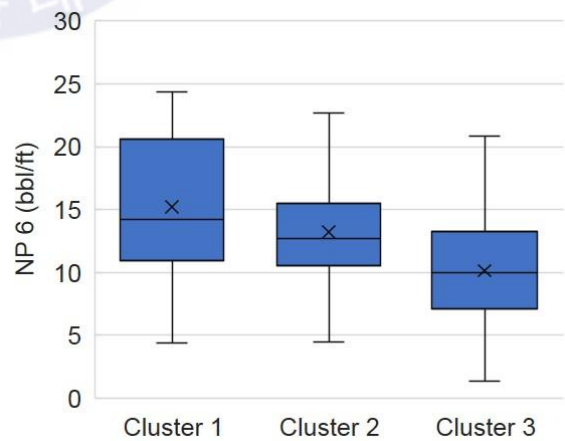
(c) Norm Spacing



(d) NV Proppant



(e) NV Fluid



(f) NP 6

Fig. 54 Box plot of key factors and production by group

4.4 분류된 그룹을 활용한 누적 생산량 추정

세일층의 초기생산량 자료를 보유하고 있는 경우 일반적으로 DCA를 통해 생산량 예측을 수행하며, 대략적인 EUR 혹은 누적 생산량을 도출한다. Arps의 쌍곡선함수를 12개월의 생산량 자료에 적용하였으며, 48개월 생산량 추정 결과에 대한 분포를 Fig. 55에 나타내었다. 이상값을 제외하고 식 (19)로 도출할 수 있는 AAPE(average absolute percent error)를 도출하였으며, 그 결과 전체 생산정에 대하여 약 26%로 짧은 생산 기간에 대한 생산 변동성으로 인해 비교적 큰 오차율을 보였다. 6개월의 생산량 자료를 DCA에 적용할 경우 비정상적인 추정값을 도출하는 생산정이 많아 비교대상에서 제외하였다.

따라서 이 연구에서는 생산특성이 서로 다른 그룹에서 초기의 단위길이 당 누적 생산량을 활용하여 추세선을 도출함으로써 48개월의 누적 생산량을 추정하고자 하였다. Fig. 56과 Fig. 57에 NP 6과 NP 12에 대하여 NP 48과의 추세선 식과 결정계수를 나타내었으며, 각 그룹의 추세선을 활용하여 추정된 값의 AAPE는 Table 15와 같다. 이를 통해 기존에 생산중인 유정에 대하여 6개월 혹은 12개월의 누적 생산량 값이 있을 경우, 입력 자료에 대한 특성 파악 과정을 통해 도출한 추세선을 활용하여 48개월의 누적 생산량 추정이 가능함을 확인하였다.

$$AAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Measured_i - Estimated_i}{Measured_i} \right| \quad (19)$$

n : No. of data set

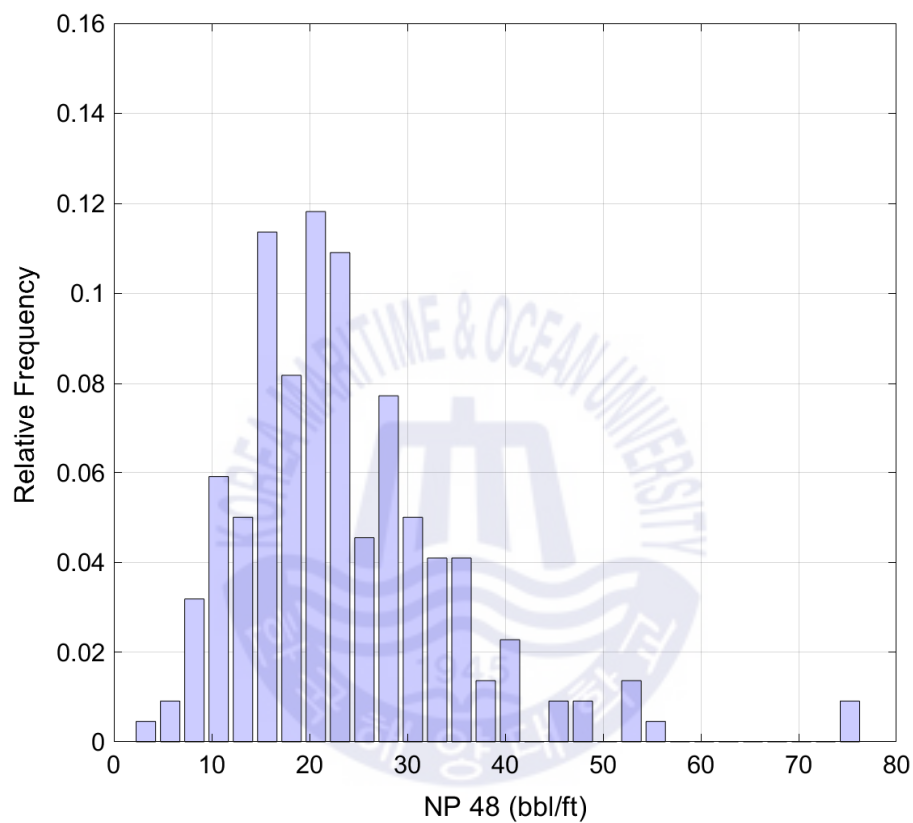
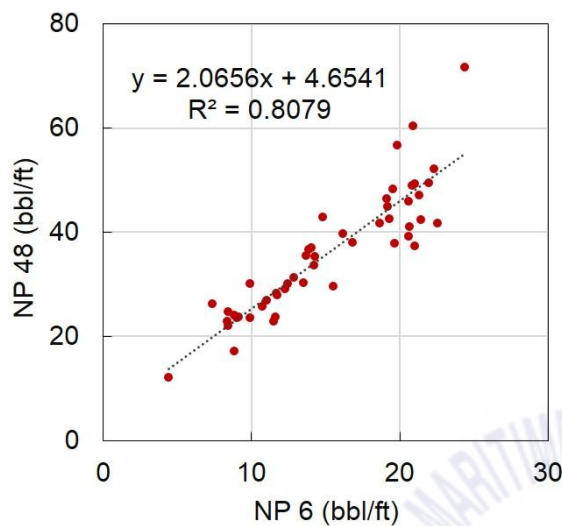
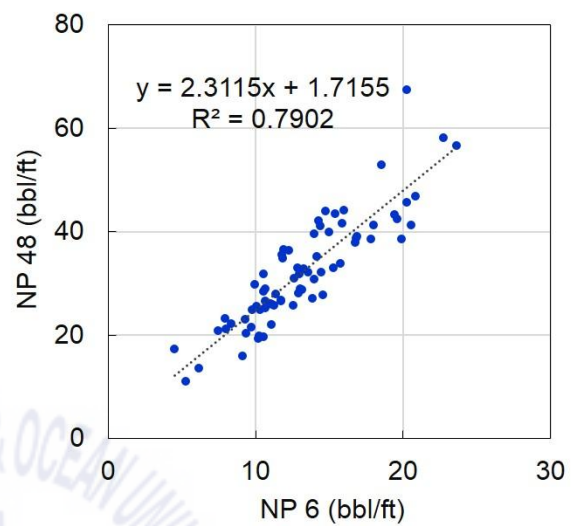


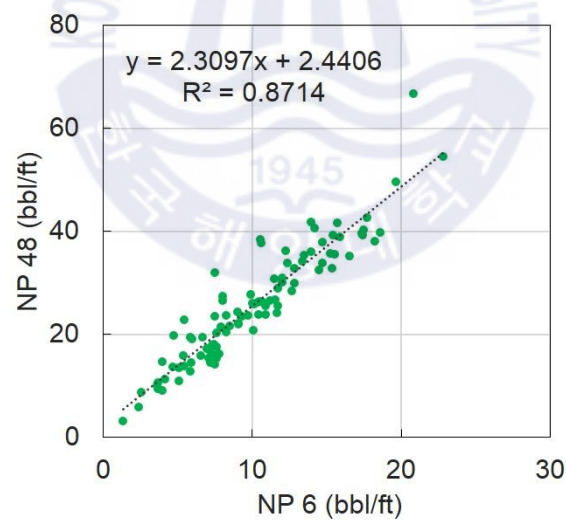
Fig. 55 Estimated production results of the total wells using Arps hyperbolic with 12 months data



(a) Cluster 1

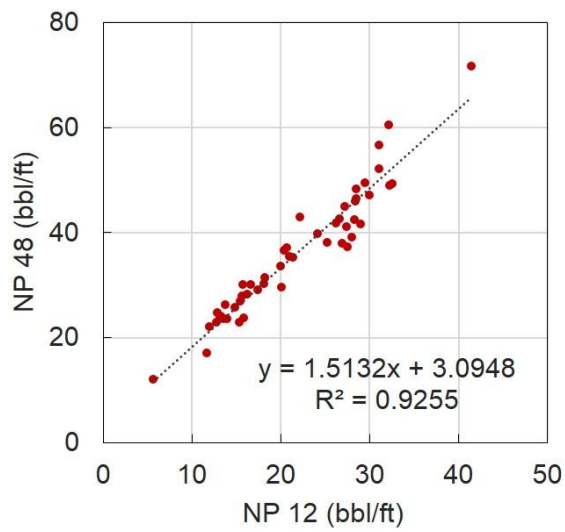


(b) Cluster 2

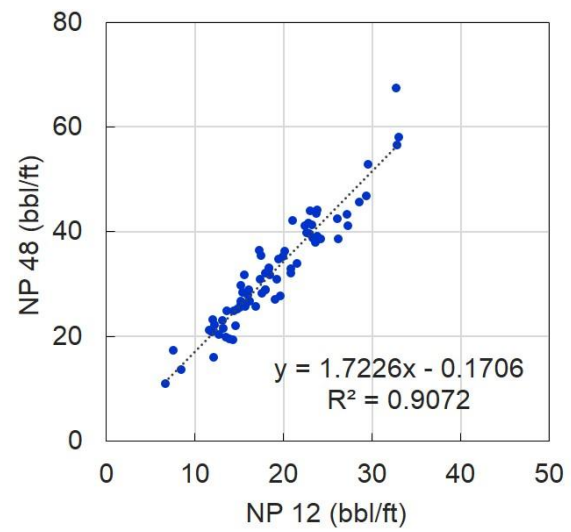


(c) Cluster 3

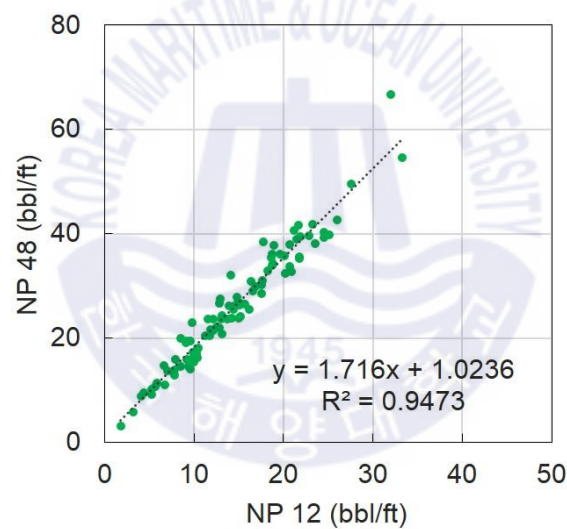
Fig. 56 Relationship between NP 6 and NP 48 by group



(a) Cluster 1



(b) Cluster 2



(c) Cluster 3

Fig. 57 Relationship between NP 12 and NP 48 by group

Table 15 AAPE of estimated production results by group

	Cluster 1	Cluster 2	Cluster 3
NP 6 vs. NP 48	9.72 %	12.02 %	12.60 %
NP 12 vs. NP 48	6.44 %	7.94 %	7.87 %

제 5 장 새로운 유정의 생산량 예측

생산이 진행되고 있는 기존의 유정에서의 생산거동 예측 연구는 활발히 이루어지고 있으며, 일반적인 석유자원보다 생산성 예측에 어려움이 따르는 셰일층에 대해 최근 들어 머신러닝 기법의 적용으로 개선된 결과를 보이고 있다. 그러나 정보가 없는 새로운 유정에서의 생산량 예측 연구는 상대적으로 적으며, 기존 생산정에 대한 분석보다 많은 가정으로 인해 어려움이 따른다. 또한, 새로운 유정에 대한 정보를 취득하기 위해서는 많은 시간과 비용이 소모되며, 이를 통해 취득한 정보에도 불확실성이 존재한다. 따라서 해당 장에서는 기존 생산정으로부터 보유하고 있는 데이터베이스를 활용하여 제한된 입력인자만으로 다수의 새로운 유정에서의 생산량 예측을 위한 방법을 적용하고자 하며, 생산특성이 다른 그룹을 분류하고 예측값에 대한 불확실성 고려를 위해 단일값이 아닌 확률론적 값을 도출하고자 한다.

5.1 그룹별 특징 및 확률론적 입력값 추출

새로운 유정에서의 신뢰성 있는 생산량 예측을 위해 분류된 그룹의 특징을 활용하고자 하며, 이를 위해 전체 22개 속성 중 주요인자와 생산량과 관계있는 감퇴곡선인자에 대하여 통계적 특징을 박스플롯으로 도시하였다(Fig. 58). 이를 통해 각 그룹별 0~1 사이로 정규화한 그룹별 각 주요인자의 평균, 최대값, 일사분위, 중간값, 삼사분위, 최소값의 특징이 다른 것을 알 수 있다. 그 중 수압파쇄 설계인자 중 Spacing, 수압파쇄 시 단위길이 당 주입 프로판트 및 유체 양의 그룹별 특징이 뚜렷하게 구분이 되며, 그룹 분류 시 주요한 속성으로 활용 가능할 것으로 사료된다. 또한, Cluster 1의 수압파쇄 설계인자의 특징 값이 다른 그룹과 차이가 나며, 이로 인해 Cluster 1은 생산량이 높은 생산정을 포함하고 있음을 파악할 수 있다. 분석에 사용된 생산정이 최적의 수압파쇄 설계를 적용하였다고 가정하였을 때, 그룹을 분류함으로써 새로운 유정에 대한 수압파쇄 설계인자 정보의 추정이 가능할 것이다.

새로운 유정의 입력값 선정 시 불균질한 세일층의 특성상 인근 생산정 정보만으로 생산량 예측 시 불확실성이 존재한다. 이를 고려하기 위한 확률론적 값(P-value)을 Fig. 59와 같은 누적확률을 통해 각 그룹별 모델 입력인자에 대한 P10, P50, P90을 도출하였다(Table 16). 이러한 결과로부터 주요인자의 영향으로 분류된 생산 특성이 다른 그룹의 특징을 반영하여 새로운 유정의 생산량 예측 시 활용 가능할 것으로 사료된다.

- ① True Vertical Depth (ft) ② HF Stage ③ Norm Spacing (ft/stage)
- ④ NV Proppant (lbs/stage) ⑤ NV Fluid (bbl/stage) ⑥ Qi (Arps) (bbl/ft)
- ⑦ Di (Arps) ⑧ b (Arps)

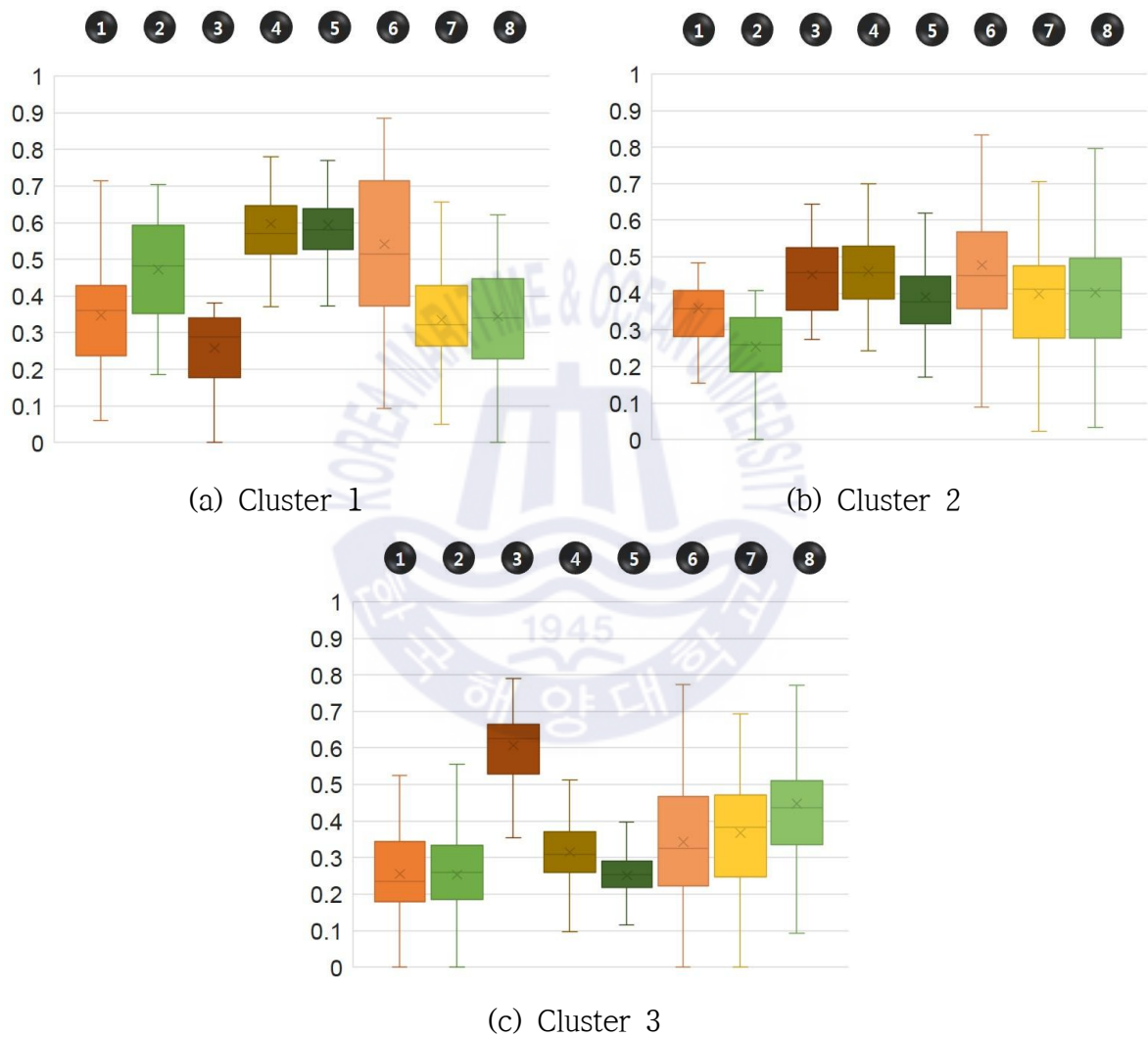


Fig. 58 Box plot of key factors by group

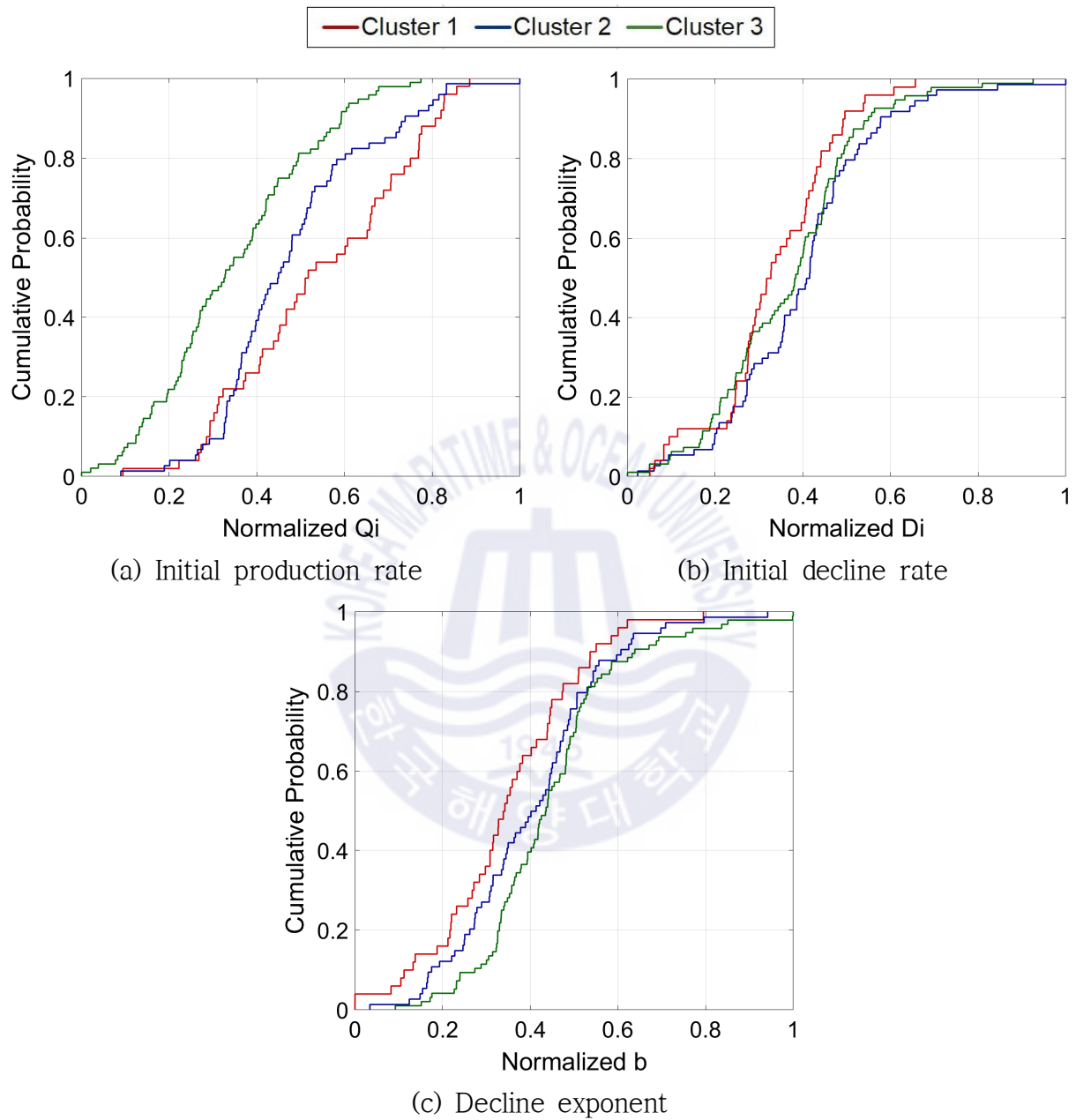


Fig. 59 Cumulative probability of DCA parameters by group

Table 16 Probability value of DCA parameters by group

	Cluster 1			Cluster 2			Cluster 3		
	P10	P50	P90	P10	P50	P90	P10	P50	P90
Initial production rate (bbl/ft/Month)	1.95	3.25	4.97	1.99	2.89	4.57	1.03	2.18	3.73
Initial decline rate	0.10	0.20	0.28	0.15	0.25	0.32	0.14	0.23	0.31
Decline exponent	0.25	0.59	0.88	0.33	0.68	0.99	0.49	0.73	1.02

5.2 새로운 유정에 대한 그룹 분류

이 연구에서는 세일층의 주요인자를 파악하고 생산특성이 다른 그룹을 분류하는 일련의 과정을 거치지 않고 주요인자 값으로 그룹을 분류하고자 머신러닝을 적용하여 모델을 개발하였다. 이를 위해 지도 학습 중 분류 알고리즘을 적용하여 학습 80%, 테스트 20%로 구성하고 학습 시 5겹 교차검증(k-fold cross validation)을 수행하였다.

생산정의 위치 정보를 고려하여 입력 변수에 따른 Case study를 수행하였으며(Table 17), 입력 변수는 PCA 결과 상위에 위치하는 주요인자와 생산량에 대한 속성을 활용하고자 하였다. 6~7개의 경우 생산량 속성이 제외되었으며, 4개는 TVD를 고려하지 않은 상태에서 분석하였다.

각 속성 개수의 Case에 대해 가장 높은 정확도를 보이는 결과를 도출하였으며(Fig. 60), 속성의 영향으로 분류 알고리즘 별 결과가 상이하게 나타났다. 분류 알고리즘 중 판별 분석, 나이브 베이즈, SVM의 정확도가 80% 내외로 나타났으며, 알고리즘 별로 포함하고 있는 속성에 따른 결과가 다른 것을 파악할 수 있다. 판별분석의 경우 생산량만을 포함하고 있지 않은 6~7개에서 테스트 자료에 대한 정확도가 높게 나타났으며, SVM은 주요인자 중 TVD를 제외한 4개의 속성을 고려하였을 때 높은 예측 정확도를 보였다. 나이브 베이즈는 속성에 따른 변화가 크지 않았으나, 그 중에서도 주요인자를 포함하고 있는 3~5개의 경우에 예측 정확도가 높았다. 이를 통해 속성의 영향에 따라 정확도가 높은 분류 알고리즘이 다른 것을 파악할 수 있었으며, 모든 정보가 존재하는 기존 생산정의 그룹 분류 결과는 입력 변수에 대한 영향이 비교적 적은 것으로 사료된다.

Table 17 Case study for input attributes

8 attributes	...	5 attributes	...	3 attributes
• TVD • Lateral length • HF Stage • Norm Spacing • NV Proppant • NV Fluid • NP IP • NP 6		• TVD • HF Stage • Norm Spacing • NV Proppant • NV Fluid		• Norm Spacing • NV Proppant • NV Fluid

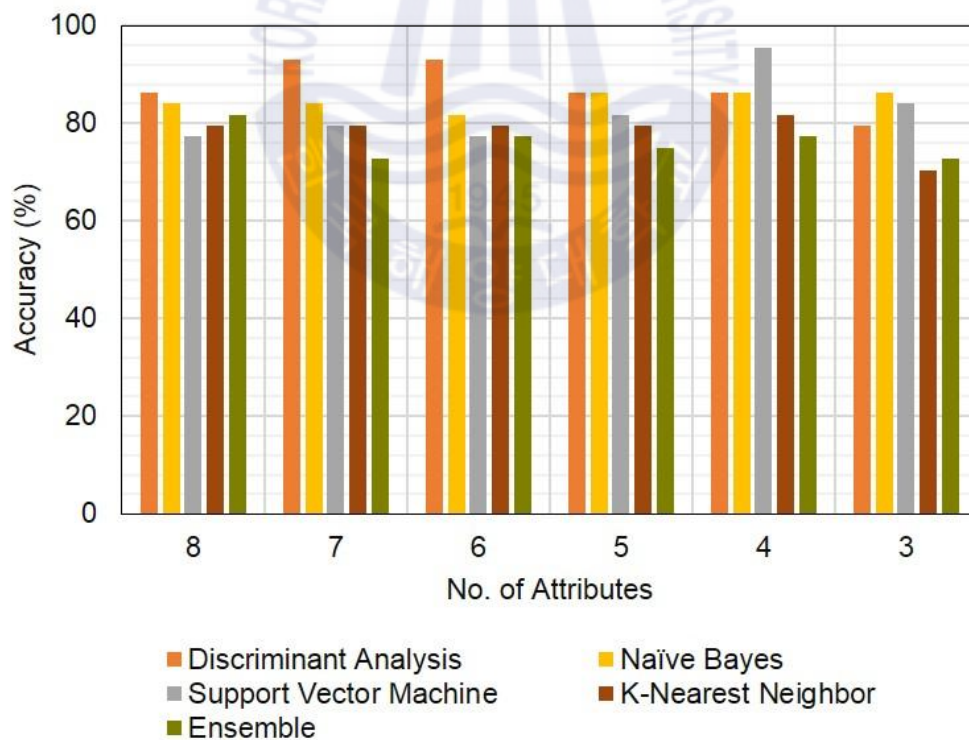


Fig. 60 Accuracy for testing wells by methods

기존 생산정을 통해 검증된 분류 모델을 활용하여 44개의 Blind well 혹은 새로운 유정을 대상으로 그룹을 분류하였다. 생산정 위치, TVD 이외의 새로운 유정의 입력값이 존재하지 않는 속성에 대한 고려를 위해 Amr et al.(2018)의 인근 생산정 후보 조건을 참고하여 유정 사이의 거리가 2500 ft 이하에 해당되는 인근 생산정의 각 인자 평균값을 활용하였다.

새로운 유정에 대해서도 입력 변수에 따른 Case study를 수행하였으며, 그 결과 Fig. 61과 같이 속성의 개수에 따라 알고리즘 별 정확도가 유사하게 나타났다. 그러나 기존 생산정으로 테스트한 경우와 달리 주요인자 중 그룹별 분류가 뚜렷하게 나타난 Spacing, NV Proppant 및 Fluid의 3개 속성을 이용한 예측 결과의 정확도가 가장 높게 나타났다. 이는 새로운 유정에 대한 입력 자료를 인근 생산정의 평균 정보를 활용함에 따른 불확실성으로 인해 도출된 결과로 사료된다.

또한, 3개 속성 대한 그룹별 오차와 정확도 결과를 Fig. 62에 도시하였으며, 알고리즘 중 나이브 베이즈의 정확도가 약 70%로 도출되었다. 기존 생산정을 활용한 분류 모델의 검증에서도 나이브 베이즈의 경우 속성에 따른 정확도가 높은 범주에 속하였으며, 이러한 제한된 입력 인자를 이용한 그룹 분류 시에 확률로 그룹을 분류하는 알고리즘이 적합한 것으로 판단된다.

새로운 유정에 대하여 생산특성이 다른 그룹 분류를 위해 입력 변수에 따른 Case study를 수행하였으며, 이를 통해 포함하고 있는 속성과 자료의 특징에 따라 정확도가 높은 혹은 적합한 알고리즘이 다른 것을 파악할 수 있었다. 따라서 데이터 기반 분석을 수행함으로써 분석에 활용하는 입력 변수나 자료 특성의 영향으로 적합한 알고리즘이 바뀔 수 있을 것이다. 또한, 주요인자를 포함한 제한된 속성을 이용하여 그룹 예측이 가능함을 확인하였으며, 새로운 유정의 그룹을 분류함으로써 생산량 예측에 있어 신뢰성 향상이 가능할 것으로 사료된다.

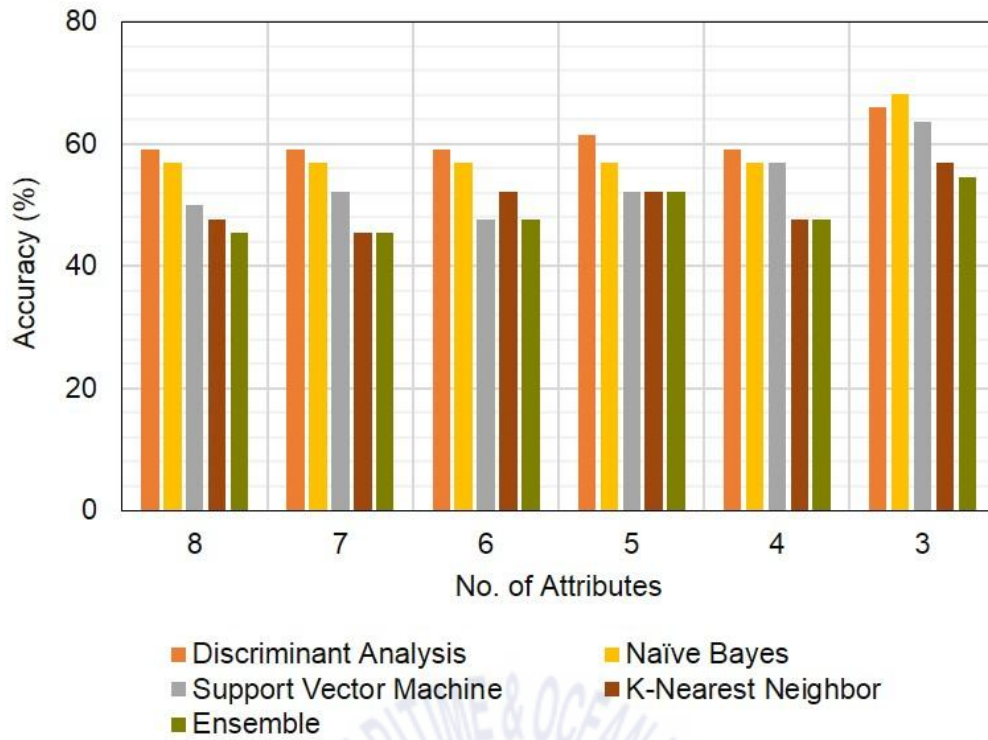


Fig. 61 Accuracy for new wells by methods

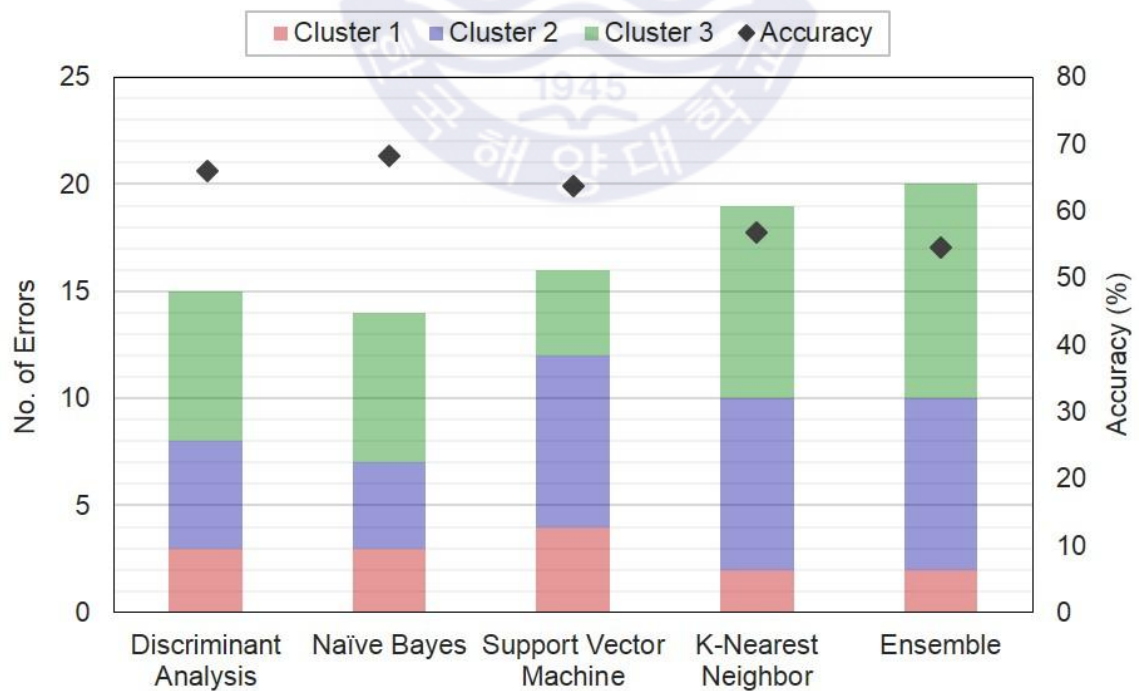


Fig. 62 Classification results of group by methods

5.3 새로운 유정의 누적 생산량 예측

새로운 유정에서는 주어진 정보가 없는 경우를 고려하여 생산량 예측을 수행하여야 한다. 이를 위해 인근 생산정의 정보를 활용하지만, 셰일층의 불균질한 특성으로 이에 대한 불확실성이 존재한다. 따라서 이 연구에서는 주요인자를 활용하여 분류한 그룹의 특징을 통해 생산량을 예측을 수행하고자 한다. 이 때, 셰일층의 특성상 초기 Peak Rate 이후의 생산이 급격히 감소하는 경향을 보이므로 이에 대한 고려를 위해 선행연구에서 정확도가 높은 결과를 보인 ANN을 활용하여 6, 12개월의 생산량 관련 인자(NP 6, 12)를 추정하고자 한다. 또한, 많은 생산이 이루어진 시점인 4년에서의 단위길이 당 누적 생산량(NP 48)에 대한 예측을 수행하였다.

5.3.1 누적 생산량 예측을 위한 인공신경망 모델 개발

Fig. 63과 같은 형태의 ANN 모델의 입력층은 기존 생산정 분석 결과와 제한적인 정보를 가진 새로운 유정에 대하여 고려하고자 하였으며, 유정의 위치 및 깊이, Spacing, Arps의 감퇴곡선인자를 활용하였다. ANN 모델 개발 시 은닉층의 뉴런수, Validation check, Epoch를 고려하여 최적화하였으며, 수치적 안정성을 위해 0~1 사이의 정규화된 입력값을 적용하였다.

이로부터 오차 신호를 역전파하여 연결 강도 갱신이 가능한 scaled conjugate gradient(scg) 학습 함수와 각 12개의 은닉층을 가지는 모델로 결정하였으며, 그룹별로 훈련자료 80%는 학습 80%, 검증(validation) 20%로 구성하고 그 외 20%를 테스트 자료로 활용하였다. 또한, 입력층과 첫번째 은닉층 사이, 첫 번째 은닉층과 두번째 은닉층 사이는 양극성 시그모이드 함수를 적용하였으며, 두 번째 은닉층과 출력층 사이에는 선형함수를 사용하였다.

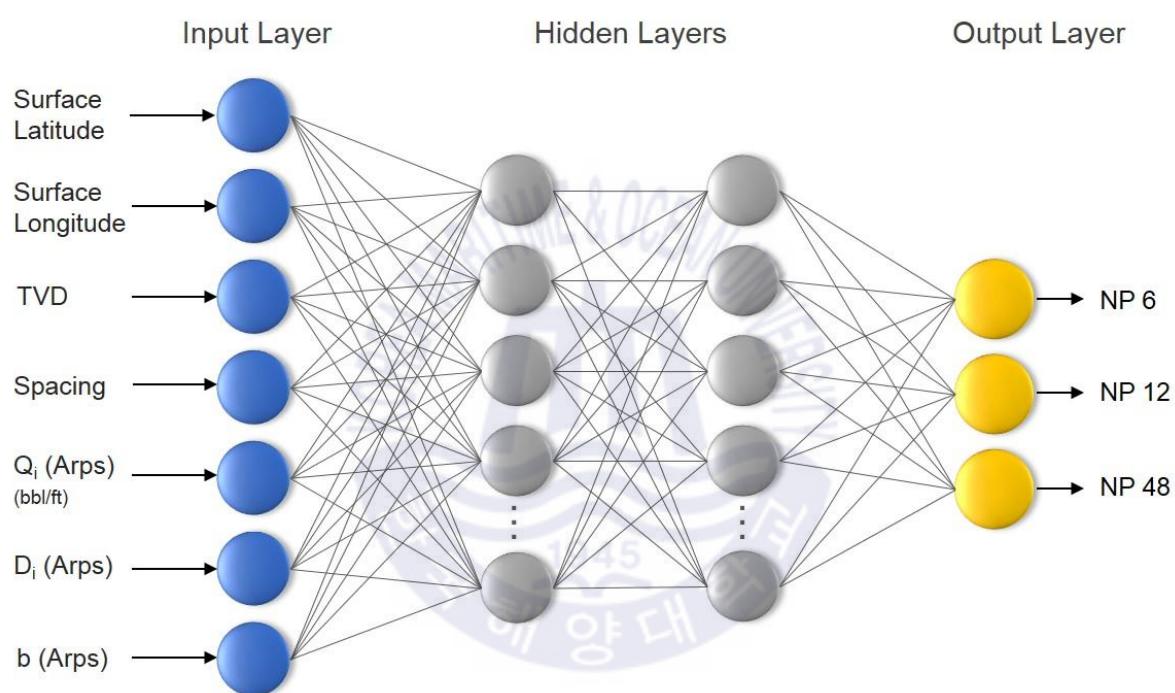


Fig. 63 Structure of ANN model in this study

5.3.2 생산량 예측 모델의 검증

ANN 모델의 검증을 위해 기존 생산정에 대한 입력인자로부터 추정한 결과와 실제 생산량 결과를 비교하였으며(Fig. 64), 44개의 생산정에 대해 NP 6, 12, 48은 각각 2.85, 2.69, 3.24%의 AAPE와 약 98%의 정확도를 보였다. 이를 통해 다수의 생산정에서 단위길이 당 누적 생산량 추정이 가능하며, 이는 짧은 기간 동안 생산이 이루어진 유정의 경우에도 활용할 수 있을 것이다.

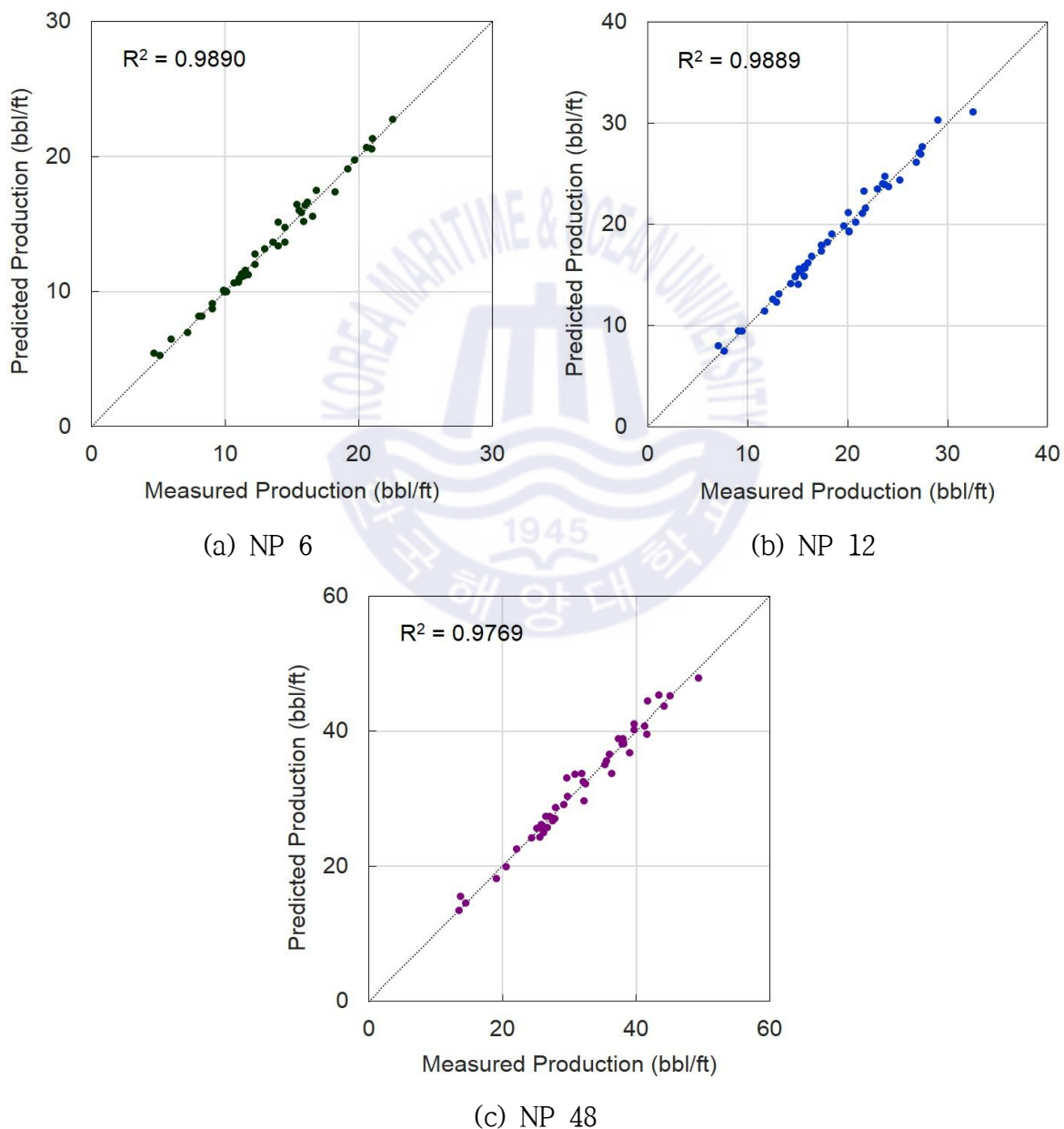


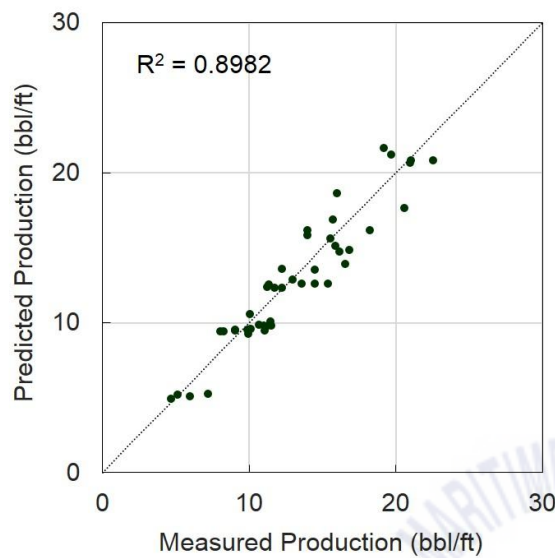
Fig. 64 Comparison of measured and predicted production for testing wells

5.3.3 확률론적 값을 이용한 생산량 예측

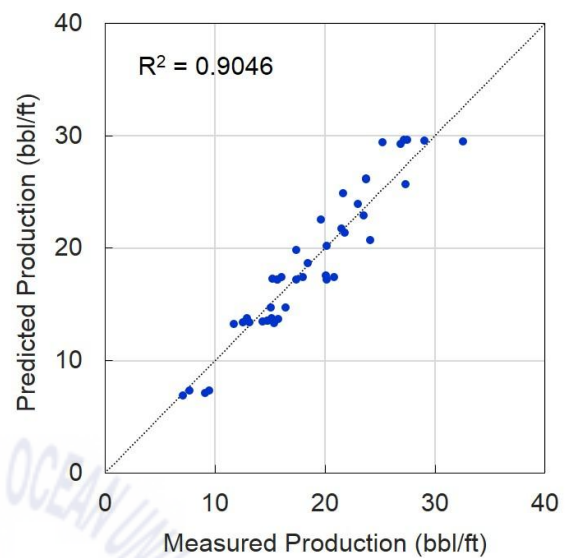
ANN 모델 검증 시 위치, 깊이, 감퇴곡선인자는 생산정에 따라 변화되는 값을 입력하였으나, 새로운 유정에 대해서는 입력인자에 대한 불확실성을 고려하고자 Spacing과 감퇴곡선인자에 대해 3가지의 P-value를 활용하여 생산량을 예측하였다. 44개의 유정 대하여 각 그룹별 P10, P50, P90의 세가지 결과 중 최소의 오차를 가지는 경우 NP 6, 12, 48은 각각 9.52, 8.85, 11.01%의 AAPE를 나타내었다. 또한, 검증된 ANN 모델을 이용하여 실제 생산량과 새로운 유정에 대한 입력인자로부터 예측한 결과를 비교하였으며(Fig. 65), 약 80% 이상의 정확도를 보였다.

이러한 새로운 유정별 확률론적 생산량 예측 결과를 Fig. 66에 도시하였으며, 비교적 안정적인 생산거동을 나타내는 48개월에서의 누적 생산량 범위는 실제값이 P10과 P90 사이에 포함되는 것을 알 수 있다. 세일층에서 생산 변동성이 큰 6, 12개월 시점에서는 생산량이 크거나 작은 경우에 대해서 일부 포함하지 못하였으나, 확률론적 생산량 범위에 근접하게 나타났다.

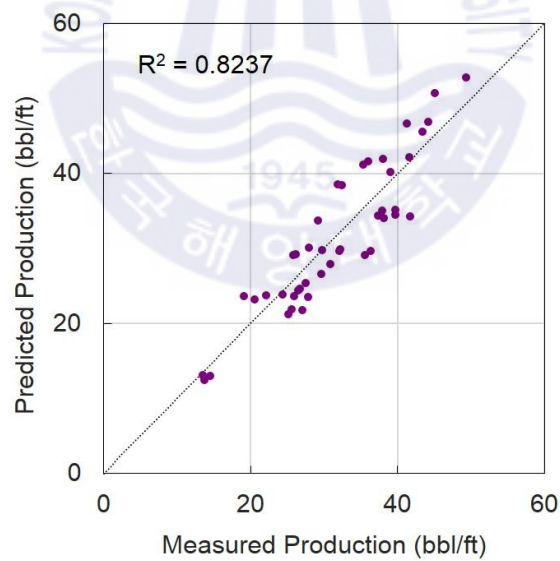
이러한 결과를 통해 6, 12개월에서의 생산량 값으로부터 초기의 생산 감퇴 정도를 간접적으로 확인 가능하며, 낮은 생산량으로 지속되는 후반부 거동 이전의 생산에 대하여 48개월의 누적 생산량을 통해 파악할 수 있을 것으로 사료된다. 따라서 확률론적 결과를 도출함으로써 결정론적인 결과에 대한 불확실성을 고려하였으며, 특정 시점의 누적 생산량, 그룹별 수압파쇄 설계인자와 감퇴곡선인자를 활용하여 단위길이 당 누적 생산량 예측이 가능함을 확인하였다.



(a) NP 6

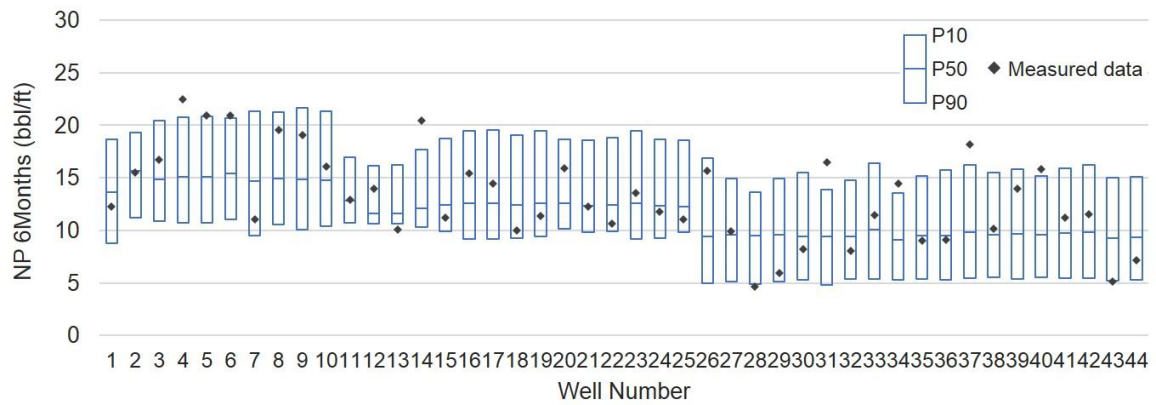


(b) NP 12

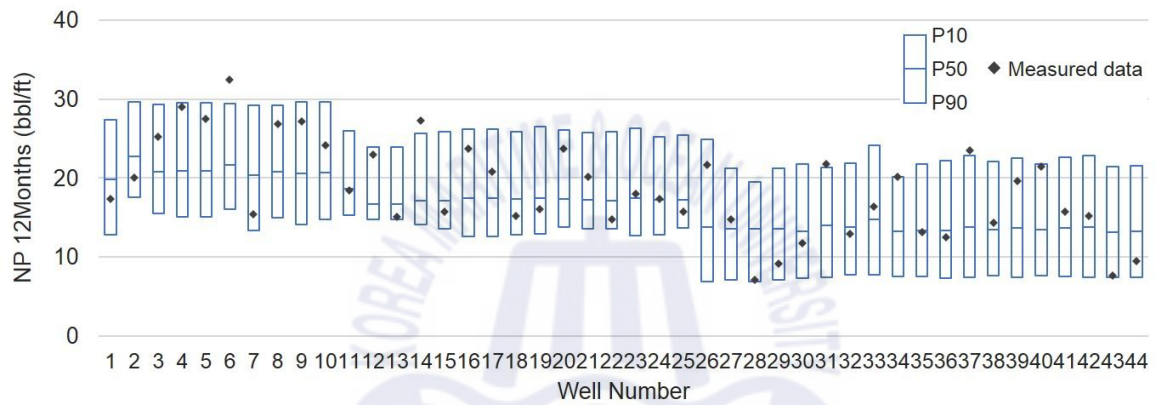


(c) NP 48

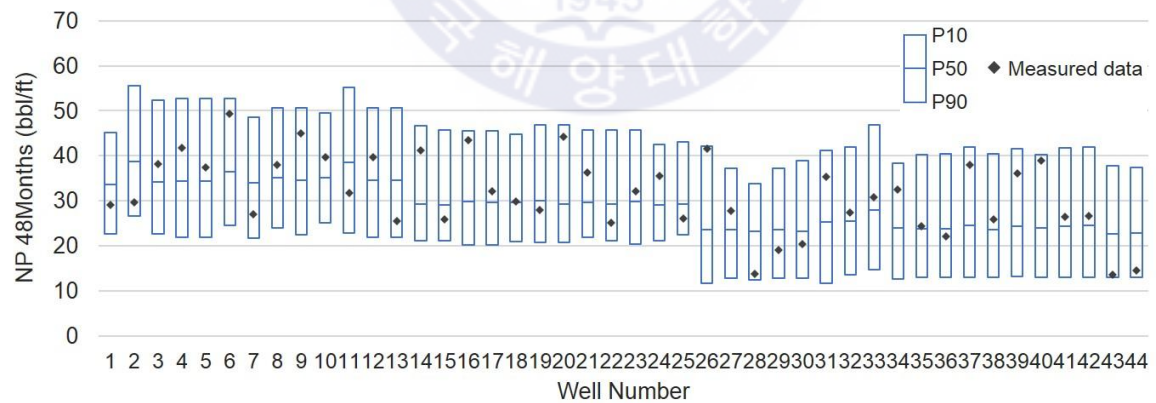
Fig. 65 Comparison of measured and predicted production for new wells



(a) NP 6



(b) NP 12



(c) NP 48

Fig. 66 Comparison of measured data and probabilistic prediction for new wells

새로운 유정에 대한 생산량 예측 결과의 신뢰성 비교를 위해 2장에서 언급한 확률론적 기법을 적용한 생산량 예측 방법으로 결과를 도출하고자 하였다. 이를 위해 일반적으로 활용하는 Arps의 쌍곡선함수와 몬테카를로 시뮬레이션을 이용하여 생산량을 추정하였으며, 이는 생산량의 변동성이 큰 셰일층에 대해 적용 가능한 방법이다. 이 때, Shin et al.(2017)에서 제안한 분석 절차를 활용하였으며, 전체 유정의 감퇴곡선인자에 대한 확률분포 형태 선정 시 자료의 특성을 반영할 수 있는 핵밀도 함수를 적용하여 결과를 도출하였다(Fig. 67). 이러한 확률분포를 바탕으로 몬테카를로 시뮬레이션을 적용하여 Fig. 68과 같이 기존 생산정에 대한 48개월에서의 누적 생산량을 추정하였다.

그 결과 전체 유정에 대하여 P10 228.80 Mbbl, P50 117.34 Mbbl, P90 48.69 Mbbl로 추정하였으며, 단위길이 당 생산량을 도출하고자 새로운 유정으로 가정한 44개의 Blind well의 Lateral length로 각각 나누었다. 이러한 단위길이 당 누적 생산량에 대해 연구에서와 동일하게 P-value 중 최소의 오차를 가지는 값을 통해 AAPE가 약 17%로 해당 연구의 결과보다 비교적 높은 오차율을 보임을 파악하였다.

이와 같이 연구대상 지역에서의 기존 다수 생산정에 대한 확률론적 결과는 새로운 유정의 생산량 예측에 있어 표준곡선(type curve)과 같이 큰 범위의 생산량으로 파악할 수 있다. 이와 더불어 이 연구에서 예측한 생산량 범위는 이보다 좁은 범위의 확률론적 생산량 결과로써 유정 단위의 결과를 제안할 수 있을 것으로 사료된다.

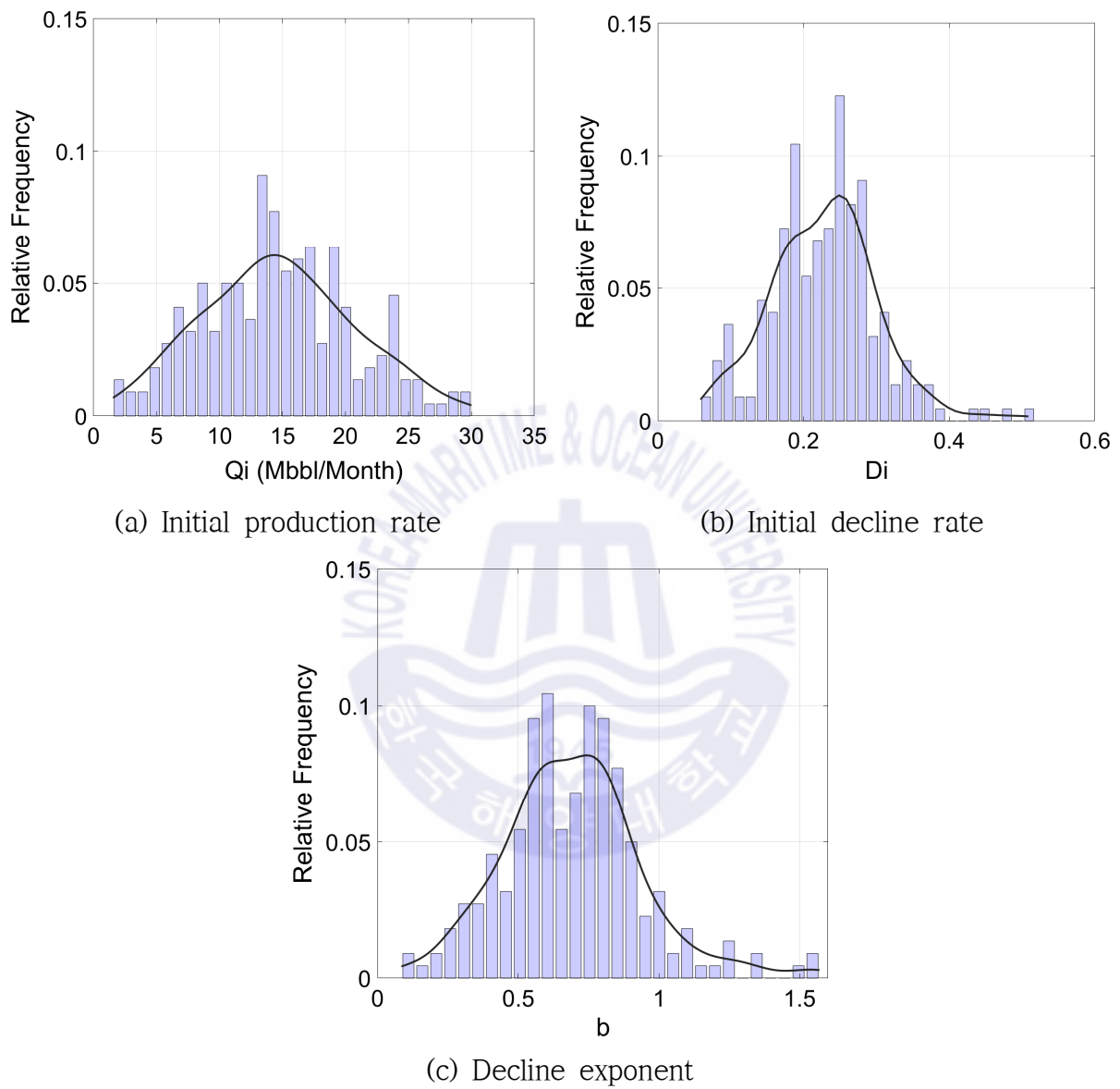
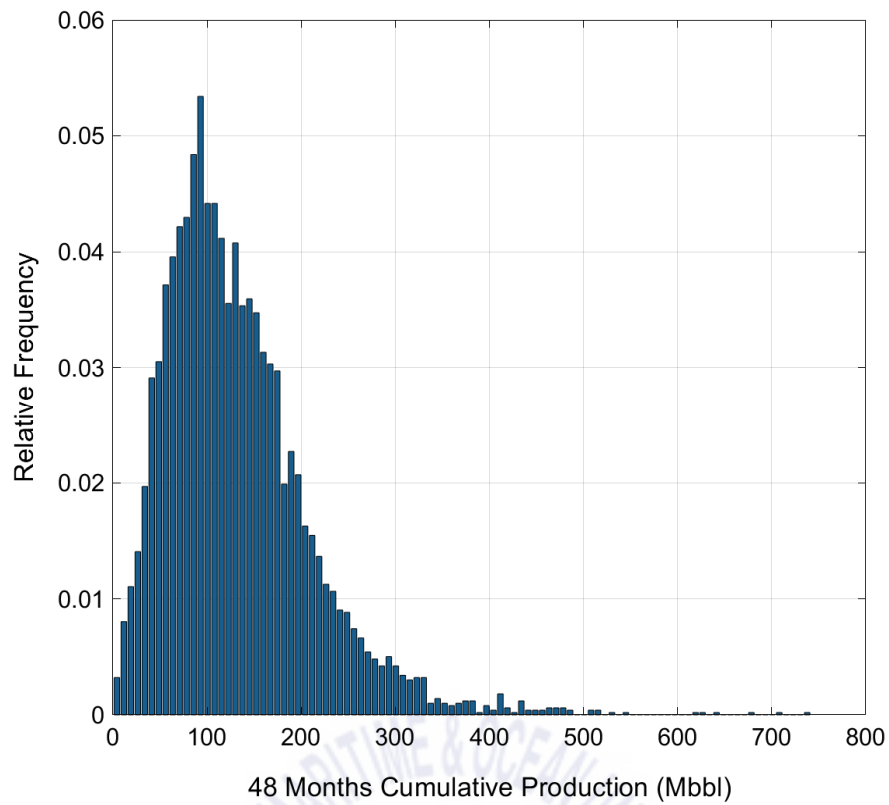
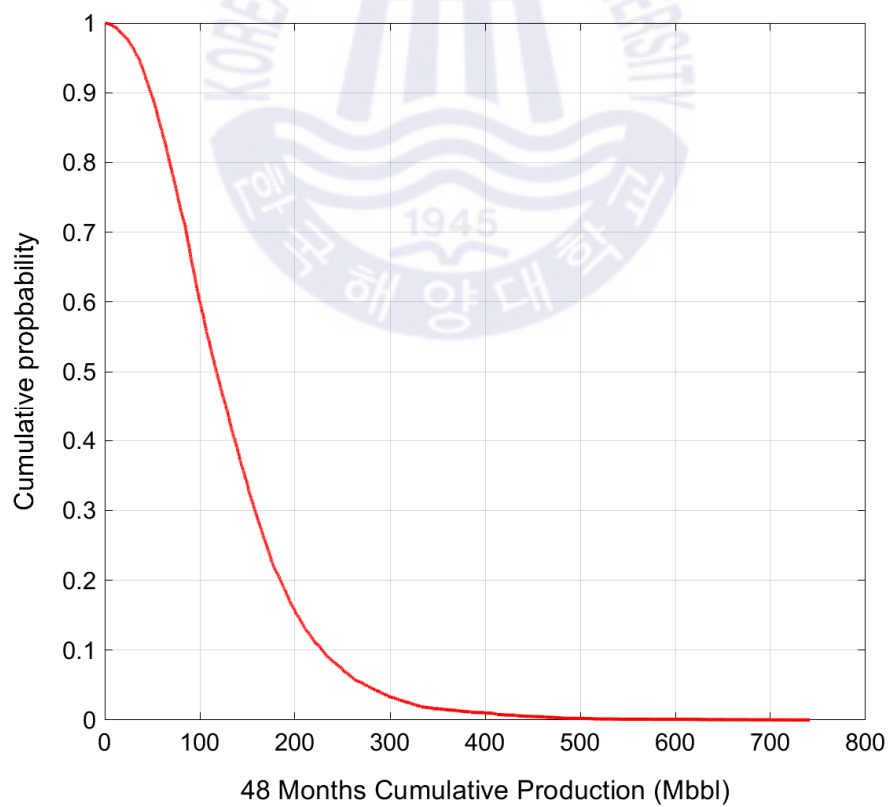


Fig. 67 Probability distribution of DCA parameters with 48 months data



(a) Histogram



(b) Cumulative probability distribution

Fig. 68 Estimation results of total wells

제 6 장 결 론

이 연구에서는 미국 Eagle Ford Shale의 약 60개월 생산된 220개의 생산정을 대상으로 셰일층에서의 석유 생산 관련 주요인자를 파악하였다. 이를 바탕으로 그룹을 분류함으로써 수압파쇄 설계인자의 영향에 대한 생산특성을 분석하였으며, 분류된 그룹의 생산량 관계를 통해 기존 생산정에서의 누적 생산량을 추정하였다. 또한, 기존 생산정에 대한 생산량 예측뿐 아니라 새로운 유정의 생산량 예측을 수행하고자 생산특성 파악을 위한 일련의 과정에 대하여 주요인자만으로 그룹을 분류하였으며, 그룹별 특징과 확률론적 입력값을 활용하여 입력인자에 대한 불확실성을 고려한 확률론적 예측값을 도출하였다.

1. 현장 자료에 대하여 다수의 생산정 간 비교가 가능하도록 Lateral length 혹은 Stage 수로 정규화 하였으며, 각 입력 인자의 분포를 통해 극단값을 제외 등 전처리 과정을 수행하였다. 또한, 인자들의 상관관계를 분석하였으며, 수압 파쇄 설계인자 사이의 관계, 생산량과 생산정 깊이와의 관계가 비교적 높은 것을 파악하였다.
2. 다차원적인 변수들에 대한 차원 감소 및 주요인자 파악을 위하여 영향이 적은 입력 속성을 제외시킴으로써 생산정 깊이와 수압파쇄 설계인자에 대해 PCA를 수행하였으며, 5개 인자에 대해 3개의 주성분이 속성의 약 88%를 표현하였다. 주성분 1은 Spacing, NV Fluid, 주성분 2는 TVD, 주성분 3은 Stage, NV Proppant와 주된 상관관계가 있음을 파악하였다.

3. PCA 결과를 바탕으로 퍼지군집분석을 수행하여 3개의 그룹으로 분류하였으며, 생산정 간 비교를 위해 도출한 단위길이 당 누적 생산량이 Cluster 1, 2, 3 순으로 크게 나타났다. 그룹별 주요인자 평균값을 통해 Cluster 1은 타 그룹에 비해 Spacing이 좁고, 수압파쇄 시 단위길이 당 주입 프로판트 및 유체 양이 많음을 확인하였다. 이에 따라 분석 지역에서 Cluster 1과 같은 현장 조건에서 최적의 생산이 가능한 것으로 사료된다. 생산량과 수압파쇄 설계인자와의 관계를 파악하고자 박스플롯을 도시하여 그룹별로 비교·분석하였다. 그 결과 각 그룹별 경향과 포함 범위가 상이하였으며, 주요인자 중에서도 Spacing 간격, 단위길이 당 수압파쇄 주입 프로판트 및 유체 양이 그룹별로 뚜렷하게 구분되었다.

4. 분류된 각 그룹에 대하여 NP 6과 NP 12에 대하여 NP 48과의 추세선을 도출하였으며, 전자의 경우 그룹별로 약 10% 내외의 오차율을 보였고 후자는 10% 미만의 오차율을 보였다. 이를 통해 기존 생산정에 대하여 입력인자의 생산특성 파악 과정을 통해 추세선을 이용하여 단위길이 당 누적 생산량 추정이 가능할 것으로 사료된다.

5. 데이터 기반 분석의 일련의 과정을 거치지 않고 세일층의 주요인자를 활용하여 생산특성이 다른 그룹을 구분하고자 지도학습 중 분류 알고리즘을 적용한 모델을 개발하였다. 이를 위해 위치 정보를 고려하여 속성에 따른 Case study를 수행하였으며, 포함하고 있는 속성과 자료의 특징에 따라 정확도가 높은 혹은 적합한 알고리즘이 다른 것을 파악할 수 있었다. 또한, 그룹별로 뚜렷하게 구분된 주요인자를 활용한 경우 새로운 유정에 대한 그룹 분류 시 정확도가 높게 나타남을 파악하였다.

6. 새로운 유정의 신뢰성 있는 생산량 예측을 위하여 분류된 그룹에 대한 통계 분석을 수행하였으며, 예측 값에 대한 불확실성을 고려하고자 모델 입력 인자의 P-value 값을 도출하였다. 그 결과 수압파쇄 설계인자의 특징값이 그룹별로

구분되었으며, 이로 인해 Cluster 1에 생산량이 높은 생산정을 포함하고 있음을 파악하였다. 또한, 최적의 수압과쇄 설계를 적용한 생산정을 분석에 사용하였다고 가정하였을 때, 그룹을 분류함으로써 새로운 유정에 대한 수압과쇄 설계인자 정보의 추정이 가능할 것이다.

7. 생산 변동성이 큰 셰일 저류층의 특성 상 초기 생산 이후에 생산 경향이 급격히 감소할 수 있으므로 이에 대한 고려를 위해 6, 12개월의 단위길이 당 누적 생산량과 4년의 시점에서의 생산량 예측 모델을 개발하였다. 생산정 위치와 깊이, 수압과쇄 설계인자, 감퇴곡선인자로 개발한 ANN 모델은 학습 80%, 테스트 20%로 구성하였으며, 기존 생산정 자료로 검증한 결과 약 98%의 정확도와 2~3%대의 오차율을 나타냈다. 이를 통해 짧은 기간 동안 생산이 이루어진 경우의 유정에 대해서도 적용 가능할 것으로 사료된다.

8. 새로운 유정에 대한 생산량 예측 시 입력 인자에 대한 불확실성이 존재하므로 그룹별 확률론적 값을 이용하여 결과를 도출하였다. 검증한 모델에 대하여 Spacing과 감퇴곡선인자의 P10, P50, P90 값을 이용하여 단위길이 당 누적 생산량을 예측하였으며, 그 결과 약 80% 이상의 정확도와 10% 내외의 오차율을 보였다. 또한, 예측한 확률론적 값의 범위에 실제 생산량 값이 포함되는 것을 파악하였으며, 이를 통해 새로운 유정에서의 불확실성을 고려한 생산량 예측이 가능함을 확인하였다.

이 연구에서 제시한 기존 및 새로운 셰일 수평정에 대한 생산성 분석은 데이터 기반의 시행착오를 통한 분석으로 새로운 데이터에 대해서도 적용 가능성이 높으며, 이러한 방법론은 다른 지역의 셰일층에 대한 분석 시에 활용할 수 있을 것이다. 또한, 셰일층 개발 시 예측의 신뢰성 향상을 위해 추가적인 데이터 취득으로 인한 시간 및 비용 손실 없이 분석을 수행할 수 있으므로 보유하고 있는 데이터를 활용하여 생산 운영에 필요한 자료를 도출할 수 있을 것으로 사료된다.

참고문헌

- Alabboodi, M.J. and Mohaghegh, S.D., 2016. Conditioning the Estimating Ultimate Recovery of Shale Wells to Reservoir and Completion Parameters. SPE paper 184064 presented at *SPE Eastern Regional Meeting*, Canton, Ohio, USA, September 13-15.
- Amr, S., El Ashhab, H., El-Saban, M., Schietinger, P., Caile, C., Kaheel, A. and Rodriguez, L., 2018. A Large-Scale Study for a Multi-Basin Machine Learning Model Predicting Horizontal Well Production. SPE paper 191538 presented at *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, September 24-26.
- Angelov, P.P. and Gu, X., 2019. *Empirical Approach to Machine Learning*. Springer.
- Arps, J.J., 1945. Analysis of Decline Curves. *Trans. AIME*, 160, pp.228-247.
- Baek, S., Akkutlu, I.Y., Lu, B., Ding, S. and Xia, W., 2019. Shale Gas Well Production Optimization using Modified RTA Method – Prediction of the Life of a Well. URTEC paper 185 presented at *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Denver, Colorado, USA, July 22-24.
- Bowie, B., 2018. Machine Learning Applied to Optimize Duvernay Well Performance. SPE paper 189823 presented at *SPE Canada Unconventional Resources Conference*, Calgary, Alberta, Canada, March 13-14.
- BP, 2019. *BP Energy Outlook 2019 Edition*. London: BP.
- Budzaya, U., 2013. *A Study on Estimation of CBM Resources using Probabilistic Approach from Coal Basin Information*. MS Thesis. Gwangju: Chonnam

National University.

- BuKhamseen, N. Y. and Ertekin, T., 2017. Validating Hydraulic Fracturing Properties in Reservoir Simulation Using Artificial Neural Networks. SPE paper 188093 presented at *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, April 24-27.
- Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W. and Jeyachandra, B., 2016. Data Driven Production Forecasting Using Machine Learning. SPE paper 180984 presented at *SPE Argentina Exploration and Production of Unconventional Resources Symposium*, Buenos Aires, Argentina, June 1-3.
- Chen, C., Gao, G., Gelderblom, P. and Jimenez, E., 2016. Integration of Cumulative-Distribution-Function Mapping with Principal-Components Analysis for the History Matching of Channelized Reservoirs. *SPE Reservoir Evaluation & Engineering*, 19(2), pp.278-293.
- Chung, Y., 2013. Analysis and Evaluation of Net Present Value by means of Monte Carlo Simulation. *J. of the Korean Society of Hazard Mitigation*, 13(6), pp.15-20.
- Clark, A.J., Lake, L.W. and Patzek, T.W., 2011. Production Forecasting with Logistic Growth Models. SPE paper 144790 presented at *SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA, October 30 - November 2.
- Crowley, J.L., 2016. Intelligent Systems: Reasoning and Recognition [Online] Available at: <http://www-prima.imag.fr/Prima/Homepages/jlc/Courses/2015/ENSI2.SIRR/ENSI2.SIRR.S5.pdf> [Accessed 29 October 2019].
- Crystal Market Research(CMR), 2017. *Hydraulic Fracturing Market By Technology, Material and Application- Global Industry Analysis and Forecast to 2022*. Pune: CMR.
- Cunningham, P. and Delany, S.J., 2007. K-nearest neighbour classifiers. *Multiple Classifier Systems*. 34(8), pp.1-17.

- Dong, Z., Holditch, S. and McVay, D., 2013. Resource Evaluation for Shale Gas Reservoirs. *SPE Economics & Management*, 5(1), pp.5-16.
- Dunn, K., 2019. Process Improvement using Data [Online] Available at: <https://learnche.org/pid/PID.pdf?490-f09d> [Accessed 10 January 2019].
- Dykes, 2019. What Do We Mean by Data-Driven? [Online] Available at: <https://www.oreilly.com/library/view/creating-a-data-driven/9781491916902/ch01.html> [Accessed 28 May 2019].
- Ertekin, T. and Silpngarmlers, N., 2005. Optimization of formation analysis and evaluation protocols using neuro-simulation, *J. of Petroleum Science and Engineering*, 49, pp.97-109.
- Fanchi, R., 2012. Forecasting Shale Gas Recovery using Monte Carlo Analysis of Decline Curves [Online] Available at: http://www.fanchi.com/wp-content/uploads/2012/11/DCA_Fanchi_Website.pdf [Accessed 18 May 2013].
- Grechka, V., Li, Z., Howell, B., Garcia, H. and Woollorton, T., 2018. Microseismic imaging of unconventional reservoirs. SEG paper 2995627 presented at *2018 SEG International Exposition and Annual Meeting*, Anaheim, California, US A, October 14-19.
- Gaurav, A., 2017. Horizontal Shale Well EUR Determination Integrating Geology, Machine Learning, Pattern recognition and MultiVariate Statistics Focused on the Permian Basin. SPE Paper 187494 presented at *SPE Liquids-Rich Basins Conference – North America*, Midland, Texas, USA, September 13-14.
- Harwood, C. and Wipat, A., 2012. *Methods in Microbiology 39*. Academic Press.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The elements of statistical learning: Data mining, inference, and prediction*. 2nd Ed. Springer.
- He, Q., 2017. Smart Determination of Estimated Ultimate Recovery in Shale Gas Reservoir. SPE Paper 187514 presented at *SPE Eastern Regional*

Meeting, Lexington, Kentucky, USA, October 4-6.

Hetz, G., Kim, H., Datta-Gupta, A., King, M.J., Przybysz-Jarnut, J.K., Lopez, J. L. and Vasco, D., 2017. History Matching of Frequent Seismic Surveys Using Seismic Onset Times at the Peace River Field, Canada. SPE Paper 187310 presented at *SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, October 9-11.

James, S., 2013. *Shale gas production processes*. Gulf Professional Publishing.

Jolliffe, I.T., 2002. *Principal Component Analysis*. 2nd Ed. Springer.

Joshi, K. and Lee, W.J., 2013. Comparison of various deterministic forecasting techniques in shale gas reservoirs. SPE Paper 163870 presented at *SPE Hydraulic Fracturing Technology Conference*, The Woodlands, Texas, USA, February 4-6.

Kang, P.S., Shin, H.J. and Lim, J.S., 2017. Effect of Shale Reservoir Property and Condition of Hydraulic Fracture on Decline Curve Analysis Factor. *J. of the Korean Society of Mineral and Energy Resources Engineers*, 54(3), pp.223-232.

Kim, J.S., Shin, H.J. and Lim, J.S., 2014. Probabilistic Decline Curve Analysis for Forecasting Estimated Ultimate Recovery in Shale Gas Play. *J. of the Korean Society of Mineral and Energy Resources Engineers*, 51(6), pp.808-819.

Kim, K.I., 2017a. Deep Learning [Online] Available at: http://www.keei.re.kr/web_keei/d_results.nsf [Accessed 10 August 2019].

Kim, P., 2017b. *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. Apress.

Klie, H. and Florez, H., 2019. Data-Driven Discovery of Unconventional Shale Reservoir Dynamics. SPE Paper 193904 presented at *SPE Reservoir Simulation Conference*, Galveston, Texas, USA, April 10-11.

- Korea Institute of Geoscience and Mineral Resources(KIGAM), 2002. *Development of Integrated Analysis Modules for Evaluation of Oil and Gas Reservoirs and Exploitation*. Daejeon:KIGAM.
- Kumar, G. and Bhatia, P.K., 2014. A Detailed Review of Feature Extraction in Image Processing Systems, *IEEE Fourth International Conference on Advanced Computing & Communication Technologies*, Rohtak, Haryana, India, February 8-9.
- Kurtoglu, B., Cox, S.A. and Kazemi, H., 2011. Evaluation of Long-Term Performance of Oil Wells in Elm Coulee Field. SPE paper 149273 presented at *Canadian Unconventional Resources Conference*, Calgary, Alberta, Canada, November 15-17.
- Lacayo, J., 2013. *Pressure normalization of Production rates Improves Forecasting results*. MS Thesis. College Station: Texas A&M University.
- Li, Y. and Han, Y., 2017. Decline Curve Analysis for Production Forecasting Based on Machine Learning. SPE paper 189205 presented at *SPE Symposium: Production Enhancement and Cost Optimisation*, Kuala Lumpur, Malaysia, November 7-8.
- Lim, J.S., Kang, J.M. and Kim, J., 1997. Multivariate Statistical Analysis for Automatic Electrofacies Determination from Well Log Measurements. SPE paper 38028 presented at *SPE Asia Pacific Oil and Gas Conference*, Kuala Lumpur, Malaysia, April 14-16.
- Luo, G., Tian, Y., Bychina, M. and Ehlig-Economides, C., 2018. Production Optimization Using Machine Learning in Bakken Shale. URTEC paper 2902505 presented at *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, Texas, USA, July 23-25.
- Ma, Y. and Holditch, S., 2016. *Unconventional Oil and Gas Resources Handbook*. Gulf Professional Publishing.
- Mandel, M., 2017. AVERAGE U.S. SHALE BREAK-EVEN OIL PRICE HAS

- DROPPED 42 PERCENT IN FOUR YEARS [Online] Available at: <https://www.energyindepth.org/average-u-s-shale-breakeven-oil-price-has-dropped-42-percent-in-four-years/> [Accessed 26 September 2013].
- MathWorks, 2017. Machine Learning [Online] Available at: <https://kr.mathworks.com/campaigns/offers/machine-learning-with-matlab.html> [Accessed 20 October 2018].
- Meyet Me Ndong, M.P., Dutta, R. and Burns, C., 2013. Comparison of Decline Curve Analysis Methods with Analytical Models in Unconventional Plays. SPE paper 166365 presented at *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, September 30 – October 2.
- Mohaghegh, S.D., 2017. *Shale Analytics*. Springer.
- Mohaghegh, S.D., Gaskari, R. and Maysami, M., 2017. Shale Analytics: Making Production and Operational Decisions Based on Facts: A Case Study in Marcellus Shale. SPE paper 149273 presented at *SPE Hydraulic Fracturing Technology Conference and Exhibition*, The Woodlands, Texas, USA, January 24–26.
- Ministry of Knowledge Economy(MKE), 2009. Petroleum Reserves & Resources Definitions [Online] Available at: http://eiec.kdi.re.kr/skin_2016/common/epicdownload.jsp?num=104875&filenum=1 [Accessed 23 October 2019].
- Perrier, S., Delpeint, A., Shawuti, Z. and Shrestha, A., 2018. Machine-Learning Based Analytics Applied to Stimulation Performance in the Utica Shale: Case Study and Lessons Learned. URTEC paper 2902017 presented at *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, Texas, USA, July 23–25.
- Politecnico di Milano, 2003. A Tutorial on Clustering Algorithms [Online] Available at: http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/ [Accessed 29 October 2019].
- Rezaee, R., 2015. *Fundamentals of gas shale reservoirs*. John Wiley & Sons.

- Seshadri, J. and Matter, L., 2010. Comparison of power law and modified hyperbolic decline methods. SPE paper 137320 presented at *Canadian Unconventional Resources and International Petroleum Conference*, Calgary, Alberta, Canada, October 19-21.
- Shin, H.J., Lim, J.S. and Shin, S.H., 2014. Estimated ultimate recovery prediction using oil and gas production decline curve analysis and cash flow analysis for resource play. *Geosystem Engineering*, 17(1), pp.78-87.
- Shin, H.D., Lee, K.S., Lee, D.S., Jang, I.S. and Park, C., 2015. *Unconventional resources development*. CIR.
- Shin, H.J., Kim, J.S. and Lim, J.S., 2015. Application of various deterministic decline curve analyses in resource plays. *Geosystem Engineering*, 18(1), pp.61-72.
- Shin, H.J., Hwang, J.Y., Lim, J.S., 2017. Probabilistic Prediction of Estimated Ultimate Recovery in Shale Reservoir using Kernel Density Function. *J. of the Korean Institute of Gas*, 21(3), pp.61-69.9
- Soni, L.P., Akkutlu, I.Y. and Maggard, J.B., 2017. Shale Gas Well Total Fracture Surface Area Calculation Re-Visited for a Dynamic Formation Permeability. SPE paper 185809 presented at *SPE Europec featured at 79th EAGE Conference and Exhibition*, Paris, France, June 12-15.
- Suhag, A., Ranjith, R. and Aminzadeh, F., 2017. Comparison of Shale Oil Production Forecasting using Empirical Methods and Artificial Neural Networks. SPE paper 187112 presented at *SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, October 9-11.
- Sun, J., Ma, X. and Kazi, M., 2018. Comparison of Decline Curve Analysis DCA with Recursive Neural Networks RNN for Production Forecast of Multiple Wells. SPE paper 190104 presented at *SPE Western Regional Meeting*, Garden Grove, California, USA, April 22-26.
- Sung, W.M., Kim, S.J., Lee, K.S. and Lim, J.S., 2009. Korean Petroleum

- Resources Classification System. *J. of the Korean Society for Geosystem Engineering*, 46(4), pp.498–508.
- Tan, P., Steinbach, M., Karpatne, A. and Kumar, V., 2005. Introduction to Data Mining. Pearson.
- Tan, L., Zuo, L. and Wang, B., 2018. Methods of Decline Curve Analysis for Shale Gas Reservoirs. *Energies*, 11(3), pp.1–18.
- Tandon, S., 2019. Integrating Machine Learning in Identifying Sweet Spots in Unconventional Formations. SPE paper 195344 presented at *SPE Western Regional Meeting*, San Jose, California, USA, April 23–26.
- Texas RRC, Eagle Ford Shale Information, 2019. [Online] Available at: <https://www.rrc.state.tx.us/oil-gas/major-oil-and-gas-formations/eagle-ford-shale-information/> [Accessed 30 July 2019].
- Theodoridi, S. and Koutroumbas, K., 2008. *Pattern Recognition*. 4th Ed. Academic Press.
- U.S. Energy Information Administration(EIA), 2013. *Annual Energy Outlook 2013*. Washington D.C.: EIA.
- U.S. Energy Information Administration(EIA), 2019a. *Annual Energy Outlook 2019*. Washington D.C.: EIA.
- U.S. Energy Information Administration(EIA), 2019b. [Online] Available at: <https://www.eia.gov/todayinenergy/detail.php?id=39752> [Accessed 24 August 2019].
- U.S. Energy Information Administration(EIA), 2019c. [Online] Available at: <https://www.eia.gov/todayinenergy/detail.php?id=39332> [Accessed 6 October 2019].
- U.S. Energy Information Administration(EIA), 2019d. [Online] Available at: <https://www.eia.gov/todayinenergy/detail.php?id=38372> [Accessed 13 October 2019].
- Vyas, A., Datta-Gupta, A. and Mishra, S., 2017. Modeling Early Time Rate Decline in Unconventional Reservoirs Using Machine Learning Techniques. SPE paper 188231 presented at *Abu Dhabi International Petroleum Exhibition & Co*

nference, Abu Dhabi, UAE, November 13-16.

Wattenbarger, R.A., El-Banbi, A.H., Villegas, M.E. and Maggard, J.B. 1998. Production Analysis of Linear Flow Into Fractured Tight Gas Wells. SPE paper 39931 presented at *SPE Rocky Mountain Regional/Low-Permeability Reservoirs Symposium*, Denver, Colorado, USA, April 5-8.

World Energy Council(WEC), 2016. Unconventional gas, a global phenomenon [Online] Available at: <https://www.worldenergy.org/publications/entry/unconventional-gas-a-global-phenomenon> [Accessed 15 October 2019].

Yoo, H.J., Lim, J.S. and Kim, S.J., 2009. Electrofacies Determination from Well Logs using Fuzzy Clustering Analysis. *J. of the Korean Society for Geosystem Engineering*, 46(4), pp.424-430.

Yu, S., Lee, W.J., Miocevic, D.J., Li, D. and Harris, S., 2013. Estimating Proved Reserves in Tight/Shale Wells Using the Modified SEPD Method. SPE paper 166198 presented at *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, September 30 – October 2.

Zaefferer, M., 2012. *Optimization and Empirical Analysis of an Event Detection Software for Water Quality Monitoring*. MS Thesis. Cologne: Technical University of Cologne.

Zanganeh, B., Ahmadi, M., Hanks, C. and Awoleke, O., 2014. Proper Inclusion of Hydraulic Fracture and Unpropped Zone Conductivity and Fracturing Fluid Flowback in Single Shale Oil Well Simulation. SPE paper 169511 presented at *SPE Western North American and Rocky Mountain Joint Meeting*, Denver, Colorado, USA, April 17-18.

Zendehboudi, S. and Bahadori, A., 2017. *Shale Oil and Gas Handbook*. Gulf Professional Publishing.