

데이터 사이언스 기술의 지하수 분야 응용 사례 분석 및 발전 방향

정진아¹ · 이재민² · 이수비¹ · 양우종³ · 한원식^{3*}

¹경북대학교 지질학과, ²한국지질자원연구원 지하수환경연구센터, ³연세대학교 지구시스템학과

Applications of Data Science Technologies in the Field of Groundwater Science and Future Trends

Jina Jeong¹ · Jae Min Lee² · Subi Lee¹ · Woojong Yang³ · Weon Shik Han^{3*}

¹Department of Geology, Kyungpook National University, Daegu 41566, Korea

²Groundwater Environmental Research Center, Korea Institute of Geoscience and Mineral Resources, Daejeon 34132, Korea

³Department of Earth System Sciences, Yonsei University, Seoul 03722, Korea

ABSTRACT

Rapid development of geophysical exploration and hydrogeologic monitoring techniques has yielded remarkable increase of datasets related to groundwater systems. Increased number of datasets contribute to understanding of general aquifer characteristics such as groundwater yield and flow, but understanding of complex heterogenous aquifers system is still a challenging task. Recently, applications of data science technique have become popular in the fields of geophysical explorations and monitoring, and such attempts are also extended in the groundwater field. This work reviewed current status and advancement in utilization of data science in groundwater field. The application of data science techniques facilitates effective and realistic analyses of aquifer system, and allows accurate prediction of aquifer system change in response to extreme climate events. Due to such benefits, data science techniques have become an effective tool to establish more sustainable groundwater management systems. It is expected that the techniques will further strengthen the theoretical framework in groundwater management to cope with upcoming challenges and limitations.

Keywords : Data science, Groundwater data analysis, Data acquisition, Data quality, Intelligent groundwater quantity and quality management system

1. 서 론

1.1. 데이터 사이언스(Data science)

데이터 과학은, 다양한 과학적 방법 및 과정, 알고리즘, 시스템 등을 정형 또는 비정형 자료에 접목하여 통찰력 있는 지식과 정보를 추출할 뿐만 아니라 다양한 응용분야에도 적용할 수 있어 다학제적 융합 연구 분야이다. 예를 들어, 데이터 과학은 수학(mathematics), 통계학(statistics), 데이터 분석(data analysis), 정보학(informatics), 컴퓨터 과학(computer science), 그 외 응용분야 전문 지식(domain

knowledge) 등을 통합하는 개념으로 응용분야와 관련된 실제 현상을 이해하고 분석하기 위한 목적으로 적용되고 있다. 즉, 데이터를 분석, 이해 및 해석하기 위해 이용되는 과학적 방법론과 함께 자료수집 또는 분석에 적용할 수 있도록 데이터를 처리하는 과정, 데이터 과학 기술을 적용하여 해결해야 할 문제를 발굴하는 것 등이 포함된 광범위한 개념이라고 할 수 있다.

이와 같이 데이터 과학은 현실 세계로부터 획득할 수 있는 데이터를 어떻게 전산 처리하고 응용할 것인가와 관련된 광범위한 문제를 다루고 있기 때문에, 어떤 이들은 데이터 과학이 인간-컴퓨터 상호작용(human-computer interaction, HCI) 분야 중 하나라고 이야기한다. 최근 데이터 과학이 대용량 데이터에 적용하는 개념으로 이용되고 있으나, 데이터 과학은 데이터의 용량과 무관하게 예전부터 적용되어 오던 개념이라 할 수 있다("Big data needs data science but data science doesn't need big

주저자: 정진아, 조교수

공저자: 이재민, 선임연구원; 이수비, 석사과정; 양우종, 석사과정

*교신저자: 한원식, 교수

Email: hanw@yonsei.ac.kr

Received : 2022. 10. 20 Reviewed : 2022. 11. 03

Accepted : 2022. 11. 17 Discussion until : 2023. 02. 30

Table 1. Acronyms

Abbreviation	Description
AIST	Advanced information systems technology
ANFIS	Adaptive network based fuzzy inference system
ANN	Artificial neural network
ARX	Auto regressive exogenous
BCT	Boosted classification tree
BN	Batch normalization
BR	Bayesian regularization
BRT	Boosted regression trees
CART	Classification and regression tree
CBR	Case based reasoning
CF	Cascade forward
CNN	Convolutional neural network
CPS	Cyber physical system
CVAE	Conditional variational autoencoder
DA	Denoising autoencoder
DNN	Deep neural networks
DO	Dissolved oxygen
DWT	Discrete wavelet transform
DT	Decision tree
EC	Electrical conductivity
EGB	Extreme gradient boosting
FR	Frequency ratio
GA	Genetic algorithm
GB	Gradient boosting
GBR	Generalized boosted regression
GP	Genetic programming
GRNN	Generalized regression neural network
GRU	Gated recurrent unit
HCI	Human computer interaction
IoT	Internet of things
KBC	Knowledge base construction
KDD	Knowledge discovery in databases
KDP	Knowledge discovery process
KNN	K-nearest neighbors
LM	Lavenberg Marquardt
LR	Linear regression
LSTM	Long short term memory
MLP	Multi-layer perceptron
MLR	Multivariate linear regression
MODFLOW	Modular tree-dimensional finite-difference ground-water flow model
NARX	Nonlinear autoregressive exogenous model
NWIS	National water information system
PCA	Principal component analysis
RBF	Radial basis function network
RF	Random forests
RNN	Recurrent neural network

Table 1. continued

Abbreviation	Description
RT	Regression tree
SGI	Standardized groundwater level index
SPI	Standardized precipitation index
SWAT	Soil water assessment tool
SVM	Support vector machine
Wi-Fi	Wireless fidelity
XGB	Extreme gradient boosting

data”- Carla Gentry).

그러나 최근 들어 발달된 과학기술로 야기된 데이터 홍수(data deluge 또는 information explosion)로 인해 데이터의 종류와 양이 폭증하고, 이와 더불어 전처리가 필요한 데이터의 양 또한 증가하였다. 따라서, 보다 복잡한 문제의 해결책을 마련해야하는 등 고차원 기술적응이 필요함에 따라 다학제적 융합을 필요로 하는 방향으로 데이터 과학 분야가 발전하고 있다.

1.1.1. 데이터 마이닝(Data Mining)

데이터 마이닝은 KDP(knowledge discovery process) 또는 KDD(knowledge discovery in databases)로도 알려져 있는데, 대용량 데이터로부터 패턴을 발견 및 추출하고, 응용을 위한 정보로 변환하는 과정을 의미한다. 즉, 데이터 베이스 구축, 저장 및 관리, 데이터 시각화, 데이터 전처리 및 변환, 특징 발견, 현상 해석 및 예측을 위한 알고리즘 개발 등의 연구분야 뿐아니라, 추출한 정보를 이용하여 컴퓨터 기반 의사 결정을 지원해 줄 수 있는 일련의 세부 연구분야들도 모두 통합하는 개념이다.

구체적으로, 데이터들의 연계성 분석을 통해 데이터 군집의 특징 및 패턴을 분석하는 군집화 분석(cluster analysis), 데이터 내 특이한 패턴을 찾아내는 이상치 분석(anomaly detection), 데이터 간 관계를 설명하는 연관규칙 분석(association rule mining), 데이터 간 상관성을 기반으로 다양한 종류의 예측 모델(classification or regression model)을 개발하는 예측 분석(predictive analytics)이 포함되어 있다. 1700-1800년대 베이즈 이론(Bayes' theorem) 및 회귀분석(regression analysis)을 이용한 데이터 마이닝 기술을 시작으로(이는 ‘hands-on’ data analysis라 칭한다) 처리해야할 데이터의 양이 대량화됨에 따라 데이터 처리를 자동화시킬 수 있는 다양한 기술들이 발전하였다. 이러한 과정에서 인공신경망(artificial neural network, ANN), 군집화 분석, 유전 알고리즘(genetic algorithm), 결정 트리(decision tree), 지지벡터머신(support

vector machine) 등이 1900년대 들어 개발 및 적용되고 있다.

1.1.2. 기계학습(Machine Learning)

기계학습은 대용량 데이터의 패턴을 학습하고 처리 및 예측결과를 도출하기 위한 알고리즘을 개발하는 것과 관계가 깊다. 개발된 알고리즘은 독립적으로 작동되며, 데이터로부터 패턴을 학습하고 실시간으로 정보를 제공할 수 있는 예측 모델을 개발한다. 인간은 예측 모델을 이용하여 데이터를 이해하고 개발된 모델을 관심분야에 응용할 수 있다. 최근, 데이터 마이닝, 기계학습, 인공지능(artificial intelligence, AI)이 유사한 의미로 빈번히 사용되고 있으나, 엄격히 말하였을 때 기계학습은 데이터 마이닝을 이행하기 위한 하나의 방법이며, 특히 예측 모델 개발 또는 데이터 분석을 수행하기 위한 수학적 알고리즘을 개발하는 데 초점이 맞추어져 있다.

기계학습의 종류에는 비지도 학습(unsupervised learning)과 지도 학습(supervised learning), 강화학습(reinforcement learning)이 있다. 비지도 학습은 라벨이 없는(unlabelled) 데이터를 대상으로 패턴 분석을 실시하여 통계적 특징과 군집 분석(clustering analysis)을 수행하는 데 주로 이용되는 기술이다. 반면, 지도 학습은 라벨이 주어진 데이터를 대상으로 데이터의 패턴을 학습하는 방법으로 회귀분석(regression), 분류(classification) 문제에 주로 이용된다. 강화학습은 데이터의 학습 과정 중 학습되는 모델의 예측 결과에 반복적으로 보상과 페널티를 부여함으로써 시행착오를 통해 보상이 최대화되는 방향으로 학습을 수행하는 알고리즘이다.

1.1.3. 인공지능(Artificial intelligence)

인공지능은 사람처럼 행동하는 기계를 제작하는 것을 목표로 하는 과학의 한 분야로 정의할 수 있다. 데이터 마이닝이 특정 문제에 대한 솔루션에 도달하기 위하여 대용량 데이터에 이행되는 일련의 과학적 과정이라면, 인공지능은 데이터 마이닝 기술을 이용하여 만든 과정과 결과물(예, 예측 모델)을 체계적으로 자동화 하고 반복 수행할 수 있도록 프로그래밍 된 시스템이라고 할 수 있다. 그리고 기계학습은, 인공지능이 예측을 수행하는 등의 지능적으로 행동할 수 있도록 만들어주는 방법이다. 예를 들어, 알파고(AlphaGo, Silver et al., 2016)는 바둑 기보의 대용량 데이터에 데이터 마이닝 기술과 강화학습이라는 특정 기계학습을 적용하여 바둑 기보의 패턴을 학습하여, 최적의 수를 둘 수 있는 예측 모델을 개발 및 시스템화함

으로써 사람처럼 바둑을 둘 수 있도록 개발된 인공지능의 한 종류이다. 이처럼 현재까지는 한가지 업무를 수행할 수 있는 인공지능의 개발이 일반적이었으나(예를 들어, AlphaGo, Self-driving cars), 인공지능의 궁극적인 목표는 인간과 같이 다양한 종류의 업무를 독립적으로 수행하면서, 생각하고 발전할 수 있는 인공지능을 개발하는 것이다. 이처럼 인간과 같이 다양한 업무를 할 수 있는 인공지능을 일반지능(general AI) 이라고 하며, 사람의 능력을 초월할 수 있는 지능을 초지능(artificial super intelligence) 이라고 한다.

1.1.4. 가상물리시스템(Cyber Physical System)

4차 산업혁명(4th industry revolution)은 정보통신 기술의 융합으로 이루어지는 차세대 산업 혁명으로, 물리적(physical), 디지털적(digital) 세계를 통합시켜 다양한 분야에 적용가능한 기술로 대표된다. 이러한 기술을 통해 모든 산업 분야에 걸쳐 자동화와 객체 간 상호 정보교환 기능을 부여하고자 하는 것이 4차 산업혁명 시대의 기술 추세라 말할 수 있다. 4차 산업혁명 기술을 구성하고 있는 대표적 기술 중 하나로 지능형 시스템(intelligent system)이라고도 정의되는 가상물리시스템(CPS)이다. 이 시스템은 컴퓨터 기반 알고리즘에 의해 실제 현상들이 제어되거나 모니터링 되는 컴퓨터 시스템을 의미한다.

CPS는 단일 학문 분야에 국한되지 않은 연구 분야이며 빅데이터(Big Data) 분석, 사물인터넷(Internet of Things, IoT), 인공지능, 데이터 마이닝, 기계학습, 가상 또는 증강 현실(Virtual or Artificial Reality) 등 다양한 정보통신 관련 기술들을 복합적으로 적용하여 관심 대상이 되는 객체를 지능적으로 분석, 예측 및 제어할 수 있도록 고안한다는 개념에 기초한다. 해당 개념은 스마트 그리드, 자율주행자동차, 스마트 공장, 스마트 농장 등에 적용되고 있으며 지질학 분야에서는 디지털 오일필드(digital oil field)가 대표적인 사례이다. 지질학적 부지를 지능적으로 운영하는 데도 활용성이 높은 CPS는, 다수의 현장 모니터링 센서로부터 현장(physical world) 정보를 습득하고, 예측 모델 및 연산 프로그램(cyber world)을 통해 습득된 현장자료를 가공, 분석, 예측한 후 그 결과로 생성된 제어 정보를 이용하여 원하는 방향으로 현장 부지를 제어하기 위해 적용할 수 있다. 지질학적 부지를 지능적으로 관리하기 위한 시스템으로 CPS 개념을 적용하면, 현장부지에서 획득되는 다양한 종류의 방대한 모니터링 자료를 효율적으로 재생산, 처리 및 예측하는 것을 자동화 할 수 있을 것으로 기대된다.

1.1.5. 디지털 트윈(Digital Twin)

디지털 트윈은 실물 객체를 동일하게 표현 및 모사하는 컴퓨터 기반의 디지털 객체를 의미하며, 실물 객체와 통신 연결을 통해 상태 정보를 공유하고 상호 작용한다. 이를 통해 실물 객체의 변화 특성을 모사 및 예측하고, 그 결과를 통해 실물 객체의 상태를 평가하고 제어하기 위한 정보를 도출하고 실물 객체로 전달하여 실물 객체의 운영을 최적화하는 데 적용된다. 모델링 기술을 의미하는 디지털 트윈에는, 물리적 법칙에 기반한 모델(process-based model)뿐만 아니라 데이터 기반 모델(data-driven model)도 적극 이용될 수 있다. 예를 들어, 지하수 분야에서 특정 지역의 지하수위를 예측하는 모델을 디지털 트윈이라 이야기할 수 있으며, 실제 현장에서 실시간으로 모니터링하고 있는 지하수위 데이터와 디지털 트윈으로 예측한 지하수위와 비교 분석함으로써 대수층의 상태를 감시하도록 응용할 수 있다. 이때, 지하수위 예측에는 기계학습을 이용하여 모델을 제작하고 예측하거나 물리적 법칙에 기반한 모델을 활용할 수 있다. 다만, 실물 객체와 디지털 트윈 사이의 실시간 연결성과 디지털 트윈 구축 기술의 성숙도에 따라 디지털 트윈 기술은 다음과 같이 3단계로 나눌 수 있다:

- 1) 레벨 1(functional digital twin): 가장 단순한 단계로 실물 객체와의 연결성 없이 디지털 객체로서만 존재
- 2) 레벨 2(connected digital twin): 실물 객체와 연결되며 실시간 모니터링 데이터를 포함하고, 단순한 분석과 시뮬레이션이 가능한 디지털 객체
- 3) 레벨 3(intelligent digital twin): 지능형 디지털 트윈으로 디지털 트윈을 구축하기 위하여 대용량의 데이터와 고급분석 기술을 적용하며 전문적인 분석(의사결정, 미래 예측)과 시뮬레이션이 가능할 뿐만 아니라 이를 실물 객체에 전달하고 제어하는 것도 가능하도록 고안된 디지털 객체

2. 지하수 분야 과학적 연구 패러다임의 변화

2.1. 지하수 분야 자료 이용의 어려움(5Vs)

최근 모니터링 및 탐사 기술의 급격한 발전으로 가용 지하수자료나 지질자료의 양이 증가하고 있다. 그러나 복잡한 지질학적 현상으로부터 발생한 지하수 및 지질자료를 다양한 분야에 실용적으로 활용하는 데에는 다음과 같은 어려움이 존재한다(5Vs):

- 1) **가용 자료의 양(Volume)**: 거시 규모의 탐사로 인하여 가용한 자료의 양이 방대함

- 2) **가용 자료 종류의 다양성(Variety)**: 다양한 탐사 기술을 적용함에 따라 다양한 종류의 자료가 존재함
- 3) **지속적인 모니터링 및 지속되는 탐사로 인한 가용 자료의 축적(Velocity)**: 실시간 모니터링 기술의 접목으로 지속해서 자료가 추가 및 분석되어야 하며, 지금도 전세계적으로 다양한 탐사 및 연구 자료들이 쏟아지고 있음
- 4) **양질의 자료 획득의 어려움(Veracity)**: 모니터링 및 탐사 기기 오류, 지하 공간의 지질학적 복잡성 등 여러 요인으로부터 기인한 잡음으로 양질의 자료 획득에는 어려움이 있음(자료의 질에 대한 검증이 반드시 필요함)
- 5) **가치 있는 정보 획득(Value)**: 대용량이며 다양한 종류의 데이터들의 특징을 해석하고, 이로부터 지하수 및 지질분야의 특정 연구 목적에 적합한 가치 있는 정보를 획득하기 위한 전략이 필요함

2.2. 자료 활용 효율성을 높이기 위한 데이터 사이언스 기술의 필요성

지질학은 인류 삶의 터전인 지하 공간에 대한 지적 호기심에서 시작한 관찰에 기반한 학문으로, 관찰한 자연 현상을 다양한 추론 및 분석을 통해 지하 공간의 변화과정을 이해하고 예측하는 데 목적이 있다. 이와 유사한 목적을 가진 지하수학은 그 대상이 지하 대수층 내 지하수라는 차이점이 있다. 지하수에 대한 이해와 예측을 위해 다양한 실험뿐만 아니라 이론적 모델을 기반한 수학적 모델링 연구도 수행되고 있다. 그러나 이론적 모델은 많은 가정이 필요하고 비교적 제한적 자료를 기반으로 해석되기 때문에, 거시적 규모에서 발생하는 복잡한 실제 지질 현상들을 정확하게 이해하고 예측하는 데 어려움이 있다.

기존 연구방법과 달리, 빅데이터 분석 기술은 지하수 변화를 분석하기 위해 현장에서 획득된 방대하고 다양한 자료에 효율적으로 적용할 수 있다. 특히, 획득된 자료가 복잡한 지질학적 현상들을 모두 내재한 결과물임에 따라 유효한 분석이 적용되었을 경우, 보다 현실적인 해석과 예측이 가능하다는 장점이 있다. 또한, 객관적이면서도 신속하게 분석함으로써 보다 많은 양질의 과학적 결과를 제공할 수 있으며, 이를 바탕으로 보다 편리한 지하수자원 활용, 지속 가능한 지하수 자원 개발을 통한 인류 삶의 질 향상과 같은 현실적 문제 해결에도 도움을 줄 수 있다. 이러한 측면에서 데이터 사이언스 기술들은 앞으로 기존에 이뤄져 오던 실험 및 이론 기반의 연구들과 더불어 지하수 분야의 발전에 지속적으로 새로운 활력을 불어넣을 것으로 예상된다.

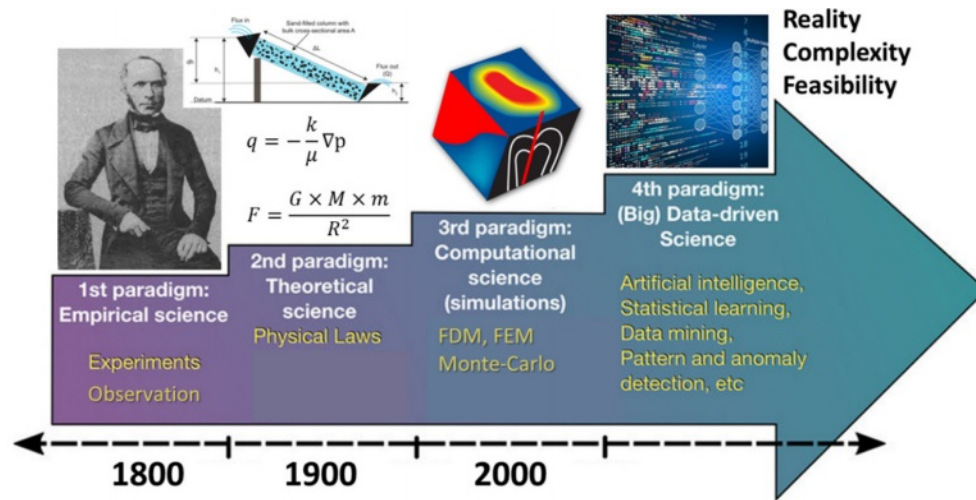


Fig. 1. Evolution of research paradigm in the field of groundwater science (adapted from Schleder et al. (2019)).

3. 지하수 분야 데이터 사이언스 기술 적용 사례

본 논문에서는 지하수 분야 데이터 사이언스 기술 적용 사례를 데이터 취득, 데이터 품질관리, 데이터 분석, 및 지하수 관리시스템 개발을 위한 분석 결과의 현장 적용기술로 나누어 분석하였다. 분석을 위해 검토된 연구논문의 주요 내용 분석(Table 2)에 기초한 구체적인 분석 결과는 다음과 같다.

3.1. 데이터 취득

시계열적 기상 요인(예, 강수, 증발산량)에 의해 변화하는 지하수를 보다 정밀하게 파악하고 분석하기 위해서는 높은 시간 해상도에 기초한 모니터링이 필요하다. 또한, 지하수는 광범위한 국토 내 어디든 분포하면서 지질매체와 유기적 관계를 갖고 있기 때문에, 공간적 해상도 또한 높은 모니터링이 필수적이다. 이러한 이유로 국내 지하수 업무를 담당하는 유관 기관들은 전국의 지하수량과 수질을 모니터링할 수 있는 관측망 시스템을 구축하여 운영하고 있다(한국수자원공사 - 국가지하수정보센터, 한국농어촌공사 - 농어촌지하수관측망시스템/농어촌지하수관리시스템).

최근 해외에서는 공간뿐만 아니라 시간 해상도도 높은 지하수 데이터를 확보하기 위해 IoT 기술을 적극적으로 이용하는 사례가 증가하고 있는 추세이며 이를 통해, 보다 양질의 정확한 데이터를 취득하고 지하수 모니터링 업무를 보다 효율적으로 수행하기 위해 노력하고 있다. 예를 들어, 캐나다 Nova Scotia에서는 지하수 관측정보로부터

터 실시간으로 지하수 수위를 자동 모니터링할 수 있는 네트워크를 구축하였다. 관정 내 실시간 계량기는 초음파 센서를 사용하여 수위를 측정하고, IoT 기술을 사용하여 와이파이가(Wi-Fi) 또는 셀룰러 연결(cellular connection)을 통해 데이터를 인터넷으로 전송한다. 전송된 수위 데이터는 실시간으로 온라인 상에 시계열 그래프로 즉시 표기되고, 이용자는 언제든지 자료를 다운로드할 수 있도록 개발되었다. 캐나다 Nova Scotia의 예시는 데이터 수집을 보다 효율적으로 자동화함으로써 기존 지하수 모니터링 방식의 비용과 관련된 문제를 줄일 수 있음을 입증한다(Drage and Kennedy, 2020).

또 다른 사례로, Wadekar et al.(2016)은 휴대용 기기로부터 데이터를 모니터링 할 수 있는 시스템을 개발하였다. 물탱크에 설치된 센서를 통해 실시간 수위 정보가 클라우드에 업데이트되며, 이 정보를 활용하고자 하는 사용자는 어디에서나 스마트폰으로 수위를 모니터링할 수 있도록 고안되었다. 또한 이를, 홍수가 발생하기 쉬운 장소에 해당 장치를 설치하여 일정 수위 이상으로 수위가 올라갈 경우, 휴대폰 내 어플리케이션 알림을 통해 일반인을 포함한 관련 사람들에게 경고를 보내는 데 이용할 수 있도록 제안하고 있다. 이 외, Senozetnik et al.(2018)은 지하수, 날씨 등의 시계열 데이터를 모니터링하는 데 사용할 수 있는 Middleware IoT 기반 프레임워크를 제안한 바 있다. 이는 검색기, 수집기, API 관리 서비스 및 감시의 네 가지 주요 구성 요소로 이루어져 있으며, 추가로 메타데이터(meta data, 데이터에 대한 정보)와 함께 최신의 시공간 데이터의 정보를 제공한다. 제안된 시스템은 슬로베니아의 Skiathos 지역의 일별 물 소비량과 펌프 작업

Table 2. Summary of research applying data science technologies in the field of groundwater science

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Drage and Kennedy, 2020	자료 취득 효율화	- 초음파 센서를 사용하여 지하수위 측정 - 실시간 모니터링 장비를 사용하여 데이터 수집 자동화	계량기의 실시간 수위	Low-cost real time groundwater level monitoring network
Wadekar et al., 2016	자료 취득 효율화	- 물탱크에 설치된 센서를 통해 실시간 수위 정보가 클라우드에 지속적인 수위로 업로드됨 - 사용하는 어디에서나 스마트폰으로 수위를 모니터링할 수 있음	물탱크에 설치된 센서 수위	Water management using IoT
Senozetnik et al., 2018	자료 취득 효율화	- 지하수, 날씨와 같은 시간 데이터를 모니터링하는 데 사용할 수 있는 Middleware IoT 기반 프레임워크를 제안 - 데이터를 정기적으로 폴링(polling)하여 최신 데이터 검색을 보장	일별 물 소비량, 펌프 작동 시간 에 대한 데이터 및 날씨 데이터 수집	IoT middleware for water management
Aderemi et al., 2022	자료 품질 향상	- 기존 지하수자원 관리 모델과 IoT 기반 모니터링 시스템에 대한 리뷰를 제시 - 점위치 데이터만 제공할 수 있고, 잡음이 생기면 기존 데이터 수집의 문제를 IoT 및 기계 학습 기술을 사용하여 해결할 수 있음	기존 지하수자원 관리 모델, IoT 기반 모니터링 시스템에 대한 리뷰를 리뷰 논문	IoT 및 Data-driven process-based management models(ANN, GP, SVM 등)
Jeong et al., 2020a	자료 품질 향상	- 한국의 지역별 관측된 지하수위 데이터 내 이상치를 판별하여 자료의 질을 향상시킬 수 있는 기법을 개발 - 이상치 판별을 위해 순환 신경망모델을 이용한 강수량 기반 지하수위 예측을 수행	강진·성전(7년) 및 포항-기복(10년) 지하수위	LSTM
Daliakopoulos et al., 2005	분석 기술 개발	- 그리스 지역의 Messara Valley의 지하수위 저하를 예측하기 위하여 ANN 기반의 모델을 개발 - 기온, 강수량, 관측정의 깊이, 하천 유출량을 입력자료로 이용하여 18개월 이후의 지하수위를 예측	Messara Valley의 Pompia 마을에서 획득한 기온, 강수량, 관측정의 깊이, 하천 유출량 데이터(20년)	ANN
Kim et al., 2019a	분석 기술 개발	- 국내 지하수 관측지점의 함양량을 산정하는데 신경망 네트워크 기법을 적용한 모델을 개발 - 개발한 모델을 충청남도 홍성군 서부면 양곡리를 대상으로 지하수 함양량을 산정하는 데 적용	207개 지하수 관측 지점의 지형 표고, 경사, 인근 하천, 지질, 토양 등과 같은 각 관측소의 지점 특성 데이터	MLP
Lee et al., 2018	분석 기술 개발	- 인공신경망을 이용하여 표준지하수지수(SGI)와 표준강수지수(SPI)의 상관관계를 학습 - 위에 기초한 데이터 기반 지하수 가뭄 전망체계를 구축하여 우리나라 전국 시·군에 적용	우리나라 167개 시·군의 지하수 관측자료 (SGI: 21년/ SPI: 43년)	NARX
Lee et al., 2021	분석 기술 개발	- 자료기반 분석 기법을 활용하여 제주 서쪽에 위치한 대정·한경 지역의 지하수 변동특성을 파악하고, 군침형 지하수 이용 관정의 영향력을 파악한 군침형 이용 관정의 영향 범위에 포함되는 지하수위 관측공을 중심으로, 배정 지하수위 유지를 위한 강수량 대비 지하수 이용 제한량 평가 도구를 제안	제주도 대정·한경지역 지하수 이용량, 지하수위 변동 및 강수량 시계열 자료(7년)	Stacked LSTM
Jeong and Park., 2019	분석 기술 개발	- 실시간 지하수 모니터링 및 관리 도구를 개발하기 위하여 지하수위 예측에 적합한 전처리 방법인 DTDS를 제안 - DTDS를 실제 지하수위와 기후 데이터에 적용하고, 여러 심층 신경망 기반의 모델을 사용하여 모델의 지하수위 예측 성능을 비교 및 검증	지하수위(2005-2014년)와 기후 데이터(일평균 기온, 누적 강수량, 평균 상대 습도, 평균 대기압 및 누적 일조 시간)(10년)	NARX-DNN LSTM GRU ARX

Table 2. continued

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Chen et al., 2020	분석 기술 개발	- 중국 북서부 헤이허강 유역의 지하수위를 예측하기 위해 수치 모델 (MODFLOW)과 세 개의 기계학습 알고리즘을 이용 및 결과를 비교함	중국 서부 지역의 디지털 표고 모델 (DEM), 토지 이용 데이터, 지하수 펌핑 수율, 지하수위, 하천 유량	ANN RBF SVM
Hussein et al., 2020	분석 기술 개발	- 아프리카 대륙 남부의 월별 지하수 위성영상(이미지) 자료를 이용 하여 전체 지하수 이미지(map)를 예측하기 위해 회귀 기반의 다양한 기계 학습 기술이 사용됨	월별 지하수 위성영상(이미지)자료 (2002-2019년)	EGB MLR RF MLP SVM
Kim et al., 2019b	분석 기술 개발	- 우리나라의 행정구역(읍·면) 별로 가정용 및 일반·용 관정의 지하수 이용량에 대한 보다 정확한 통계량을 산정하고, 행정구역 별 지하수 이용량 산정 방법을 제시하기 위하여 회귀나무(Regression tree) 방법 이용함	우리나라 32개 시군의 지하수 이용량·실 측 데이터(2006-2012년, 7693개 관정), 충청남도 지역 내 362개 관정의 지하수 이용량 실측 자료(2016-2018년)	RT
Park and Jeong., 2021	분석 기술 개발	- SWAT-MODFLOW를 이용하여 김천 지역의 사공간적 지하수 함양 분포를 제시하고 해당 지역의 위성영상(Landsat-8)과의 상관성 분석을 수행 - 위성영상으로 지하수 함양변화를 추정하기 위해 분류 및 회귀 트리 (CART) 알고리즘을 적용	평균 지하수함양량 분포(2008-2016년), 위성영상(Landsat-8 OLI/TIRS) 자료 (2015-2016년)	CART
Kenda et al., 2018	분석 기술 개발	- 데이터 수집을 위해 middleware 사물 인터넷(IoT)을 사용 - 슬로베니아의 Ljubljana Polje 대수층의 지하수 수위를 예측하기 위해 다양한 기계학습 방법이 적용됨	펌프 센서 데이터 및 기상 데이터 (2010-2019년)	LR DT RF, GB
Zhou et al., 2017	분석 기술 개발	- 중국 Mengcheng 현의 지하수위 데이터를 이용하여 지하수위 예측을 위한 데이터 중심 예측 모델을 제안하고 적용	10개의 관측공에서 측정된 지하수위 데이터(37년)	DWT SVM
Lee et al., 2019	분석 기술 개발	- 원격탐사 데이터에 Frequency Ratio(FR) 모델을 적용하여 지하수와 밀접한 관련성이 있는 데이터를 선별 - 데이터 마이닝 모델을 이용하여 대한민국 고양시 지하수 potential 지도를 생성하고 분석	원격탐사 데이터(topographic, land cover, forest maps), 지하수 양수시험자료(현장관측 8700개) 중 상위 10%에 해당하는 876개의 양수 능력 자료	FR BCT
Ahmed et al., 2021	분석 기술 개발	- 지하수 용도를 분류하기 위한 Recommender System (RS)를 구축 하기 위한 기계학습 기법을 제안 - 제안된 시스템을 통해 음용 및 농업용 지하수 이용지역을 분류	샘플 날짜, 용존칼슘, 용존마그네슘, 용존칼륨, pH, 총 용존고형 (46,400개)	Recommender System(SVM, KNN, Bagging, Decision Stump, CNN)
Najafabadi et al., 2022	분석 기술 개발	- 이란 Sirjan에 위치한 Gohar Zamin 철광석 광산 주변 관측정에 서 획득한 공간 자료 및 시계열 자료를 이용하여 지하수위 변동을 예측 - 지하수위 예측을 위해 최적화 알고리즘을 사용하여 MLP 및 CF의 성능 및 예측 능력을 향상시킴	공간 자료(관측정의 위도와 경도, 수리 전도도, 유효공극률, 전기저항률, 관측 정의 표면 수위, 기반암 높이와 깊이) 및 시계열 자료(배수량, 증발량 및 강우량) 관측정의 1542개 데이터(2019-2020년)	예측 모델: MLP, CF, RBF, GRNN 최적화 방법: Bayesian regularization, Levenberg-Marquardt, Resilient backpropagation, Scaled conjugate gradient

Table 2. continued

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Zounemat-Kermani et al., 2021	분석 기술 개발	<ul style="list-style-type: none"> -수문학의 여러 응용 분야(지표수문학, 수리지질학, 홍수기름 모델링 및 예측)에서 사용되는 다양한 유형의 Ensemble machine learning 방법의 적용을 제시하는 논문을 탐색 -단일 기계학습 방법만을 적용한 것보다 EML 방법을 적용하는 것이 과 적합 문제를 완화할 뿐만 아니라, 성능을 개선하기 위한 적절한 접근 방식으로 긴주됨을 확인할 수 있음 	리뷰논문	Model Averaging Stacking Bagging Boosting Dagging
Sahoo et al., 2017	분석 기술 개발	<ul style="list-style-type: none"> -지하수위 변화를 예측하기 위하여 입력 데이터 전처리 방법과 hybrid artificial neural network 모델을 제안 -해당 기법을 미국의 농업 생산에 이용되는 두 개의 대수층 시스템(High Plains 대수층과 Mississippi River Valley 충적 대수층)에 적용한 결과, 두 지역의 지하수 수위 모델링 시 관계 수요가 가장 중요한 매개변수인 것으로 나타났다. 	High Plains 대수층과 Mississippi River Valley 충적 대수층의 33년치 지하수위 데이터(1980-2012년)	Singular Spectrum analysis Mutual Information GA lag analysis + ANN
Jeong et al., 2020b	분석 기술 개발	<ul style="list-style-type: none"> -지하수위 변동 패턴을 기반으로 대수층 수리 특성 값을 추론할 수 있는 데이터 기반 예측 모델을 제안 -특히, Denoising autoencoder(DA)를 이용해 지하수위 변동 패턴의 저차원 특징을 추출하여 수리진도도와 함양관련 수리 특성 값을 예측하는 회귀모델의 입력 변수로 이용 -제안된 방법은 현장 테스트(예: 양수 시험등) 없이 대수층 수리 특성 값을 효과적으로 유추하는데 활용될 수 있음 	우리나라 59개 관측소의 최소 4년 및 최대 11년 지하수위	DA
Jeong et al., 2021	분석 기술 개발	<ul style="list-style-type: none"> -Jeong et al.(2020b) 연구를 통해 추출된 지하수위 변동 패턴의 저차원 특징이 동일한 수리학적 특성의 대수층에서 획득되더라도 불구하고 해마다 다른 강수 패턴에 의하여 특징의 일관성이 저해되는 문제점을 개선함 -Conditional variational autoencoder(CVAE)를 이용해 지하수위 변동 패턴을 생성 -CVAE 적용한 이후, 관측공별 연 단위 저차원 특징들 간의 유사성이 증가함을 확인함 	전국 기개 관측공에서 획득된 10년치 지하수위 데이터	CVAE
Jeong et al., 2022	분석 기술 개발	<ul style="list-style-type: none"> -지하수 관정에서 관측된 수질 측정 자료를 바탕으로 유해물질과 인자 들 간의 상관성을 평가하고 유해물질 위험 현장 예측 모델을 개발 -지하수 배 자연 방사성 물질 우라늄과 라돈의 음용기준 초과 위험 가능성을 평가할 수 있는 RF 기반 예측 모델을 개발 	국내 전역의 지하수 관정에서 관측된 수 질 및 관정 정보(온도, pH, Eh, DO, EC, 라돈 및 우라늄 농도, 관정 깊이)(2007-2019년), 암상 종류	PCA Sensitivity analysis RF
Majumdat et al., 2020	분석 기술 개발	<ul style="list-style-type: none"> -지하수 양수량을 이용하여 지하수위 하강 예측 -spatial data가 spatiotemporal data보다 예측 성능 향상에 더욱 기여함 	증발산량, 강수량, 토지 피복도, 지하수 양수량 (2002-2018년)	RF
Malakar et al., 2021	분석 기술 개발	<ul style="list-style-type: none"> -지하수위, GWSA, 월 평균 강수량, 기온, 증발산량을 이용하여 지하수 위 하강을 예측 -input data의 양에 따라 예측성능이 변할 수 있으며, 더 많은 input data 가 있을 때, 일반적으로 예측성능이 다양한 스케일에서 증가 -지하수 취수정의 분포, 취수량에 따라 예측성능이 영향을 받을 수 있음 	지하수위, Groundwater storage anomalies, 월 평균 강수량, 기온, 증발산량 (2003-2014년)	SVM

Table 2. continued

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Kim and Lee, 2021	분석 기술 개발	- 지하수위를 예측하고, 지하수위 변동에 미치는 영향인자를 분석 - 연도별 강수량이 입력으로 이용될 때 예측성능이 증가	지하수위, 지하수위의 이동평균, 하천 수 위, 월 평균 강수량, 기온 (2001-2008, 2006-2013년)	ANN
Stackelberg et al., 2021	분석 기술 개발	- 관측정에서 얻어진 pH 지표를 활용하여 Glacial aquifer 내의 pH를 3차원 으로 예측 - 향후 오염원을 선제적으로 식별하여 지하수 모니터링 지점을 선정하는데 유용하게 사용될 수 있음	30년동안 총 18,385개의 관측정에서 얻 어진 지하수의 pH (1988-2018년)	BRT
Zhang et al., 2018	분석 기술 개발	- 중국 허타오 관개지역을 대상으로, 14년간 관측된 지하수위 자료를 활용 하여 지하수위 예측 모델을 제작	증발산량, 강수량, 기온, 227개의 관측정 에서 얻어진 지하수위 자료 (2000-2013년)	LSTM Dropout
Bowes et al., 2019	분석 기술 개발	- 홍수에 취약한 해안 도시를 대상으로 지하수위 예측 모델을 제작 - 모든 시기의 데이터를 활용하는 것 보다 특정 이벤트가 발생하였을 때의 데이터를 활용하여 예측성능을 향상시킬 수 있음	지하수위 자료, 강수량, 해양 수위자료 (2010-2018년)	LSTM RNN
Vu et al., 2021	관리 기술 개발	- 프랑스 노르망디를 연구지역으로 하여, 20 - 50년간 기록된 지하수위 자료를 프랑스로 기계학습 기법을 통해 해당 시기 사이에 존재하는 미기록 자 료에 대한 회귀를 진행 - 장기적인 지하수위 변동 예측을 위한 모델 제작	총 31개의 관측정에서 얻어진 지하수위 자료 (1970-2020년, 불연속 자료 존재)	LSTM
Neyens et al., 2018	관리 기술 개발	- IoT 지인 Environmental data Management Interface(EMI) 응용 프로그램을 사용하여 다양한 데이터를 동일한 플랫폼에서 단순하고 신속하게 표시 및 분석할 수 있는 시스템을 개발 - 데이터 분석을 위한 여러가지 도구들이 탑재되어 있으며, 센서 유형에 따 라 분석할 수 있도록 고안됨	수위, 지하수 품질, 시추공 정보 데이터	Environmental Monitoring Interface(EMI)
Chang et al., 2017	관리 기술 개발	- 실시간으로 수집되는 지하수질 자료가 오염에 대한 특정 기준치를 벗어날 경우, 지정한 관리자에게 메시지를 발송하여 지하수 상태를 인지할 수 있 는 시스템 및 매뉴얼 개발 - 안정적인 데이터 생산, 급격한 수질변동의 관측 여부, 장기관측과 데이터 통신에 적합한 최적의 센서로 활용하기 위한 테스트를 실시하였으며, 사용 화 할 수 있는 수준의 성능을 제시하였음	온도, 수소이온농도, 전기전도도 (30분 간격으로 5일 동안 데이터 수집)	Multi sensor based device for groundwater monitoring
Nova Scotia Department of Energy and Mines, 2017	관리 기술 개발	- 향후 가뭄 관리에 대비해 실시간으로 천부 대수층의 지하수위를 모니터링 할 수 있는 네트워드를 개발 - 맞춤형 수위계는 초음파 센서를 사용하여 지하수 깊이를 측정 - IoT 장치를 사용하여 실시간으로 데이터를 인터넷으로 전송 및 일반인이 열람할 수 있음	총 11개의 관측정의 실시간 지하수위	Real time groundwater level monitoring network
Ahn., 2021	관리 기술 개발	- 캐나다, 호주, 네덜란드, 일본, 한국, 브라질 등의 지질자원 연구기관과 협 력하는 OneGeology Strategic Steering Committee (OSSC)이 구축되어, 실 제 환경에서 모니터링되는 센서 네트워크와 분석(예측 모델, 시각화 등의 결함으로 실시간 4차원 지리공간 데이터 플랫폼을 개발하고자 추진중 - 한국지질자원연구원은 활화산을 주제로 백두산의 디지털 트윈 구축을 주 관하고 있음	리부 논문	Real time monitoring technology

Table 2. continued

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Nevo et al., 2021	관리 기술 개발	- 2021년 Google Research 팀은 인공지능 예측 기법 기반의 홍수 조기 경보 시스템을 구축하였으며, 현재까지 인도 및 방글라데시 전역에 설치 및 적용 중에 있음 - 실제 환경에서 보다 정확하고 확장 가능한 침수 모델을 만들기 위해 물리학 기반 모델링과 기계 학습을 결합한 형태학적 범람 모델이라고 하는 범람 모델링에 대한 새로운 접근 방식을 고안 및 적용 중에 있음	실시간 하천 수위 데이터	Operational flood warning system
Goodall et al., 2008	관리 기술 개발	- 미국의 NWIS(National Water Information System)는 USGS에서 운영 중인 물 관련 자료의 저장 및 검색을 위한 통합 시스템으로 분산 DB 구조를 토대로 측정지점, 시계열적 예측 자료, 유량, 지하수, 수질, 수자원 이용 등에 관한 데이터를 실시간으로 수집/저장 관리하고 있으며, 누구나 손쉽게 웹상에서 다양한 파일형식의 데이터를 제공받을 수 있음 - 지하수정보시스템, 수질정보시스템, 데이터처리시스템, 수자원이용정보시스템의 네 가지 하위시스템으로 구성되어 있으며, 약 150만 지점에서 수집한 수량, pH, 염도, 탁도, 수온, DO, 암모니아, 질산염, 염화물, 염류소, 홍조류, 남조류 등의 측정 정보를 제공	NWIS	
Jeong et al., 2022	자료품질 향상 분석기술 개발	- 국내 전역의 지하수 모니터링 관정 수질측정 자료 기반의 유해물질 위험 현장 예측 모델 개발 - In-situ 현장수질항목을 이용해 지하수 내 자연 방사성 물질(우라늄과 라돈)의 위험성 평가	5,272개 관정에서의 자료: 암상, 심도, pH, EC, DO, Eh, T, HCO ₃ , ²²² Rn, ²³⁸ U (2007-2019년)	PCA 기반 상관분석과 민감도 분석 RF
Amini et al., 2010	분석기술 개발	- 지질기원 불소 농도를 전지구적 규모에서 예측 - 단일 기계학습 기반 예측 모델 뿐 아니라 각 기법을 조합한 예측 수행을 통해 최적의 예측능 도출	> 60,000개 전지구 규모(> 25개국) 불소 농도, 토양, 지질, 고도, 기상, 수리지질 자료	CART KBC MLR ANFIS LR 각 기법 조합: hybrid method
Podgorski et al., 2018	분석기술 개발 관리기술 개발	- 인도 전지역을 대상으로 획득한 자료를 기반으로 지하수 내 불소 농도 예측 모델 개발 - 먹은물 수질기준 초과 위험지역 탐색, 정화방안 제시 - 불소 농도 예측 분포지도 및 예측 모델 주요인자 도출	약 12,600개 지점 지질, 기후, 토양 자료 (2010, 2013년)	RF
Podgorski et al., 2022	관리기술 개발 분석기술 개발	- 지질기원의 물질인 철과 망간의 공간분포 예측 - 예측결과와 확률분포 지도는 고농도의 망간과 철 농도에 대한 정보를 제공해 주고 주로 인구 밀집지 지사: 지하수 관리/활용을 위한 대비에 도움 음 기대	> 6,000개 Mn, Fe 농도, 수리지질 및 수 리화학 자료, 57개 이상의 환경변수(기 상, 지질, 토양, 지형 자료 등)	RF generalized boosted regression
Frederick et al., 2016	자료품질 향상 분석기술 개발	- 미국 3개 광역지역(서부/중부/동부) 개인관정 대상 고농도의 비소 농도 예측 및 지역별 차이와 주요인자 비교 - 고농도 비소와 관련된 공변량간의 관계 분석: 국가규모에서 건조도와 비소 농도는 높은 상관성 지역별로 건조도, pH, 용존이온 등 모델의 주요인자 차이를 보임	40,215개 지점 비소 농도 자료와 30여개 항목의 수화학 인자 및 수문기상 자료	CART

Table 2. continued

참고문헌	목적	내용 요약	이용한 데이터 종류 및 데이터 양	이용 기술
Ahn et al., 2012	분석기술 개발	- 국내 원주 우산공단에서 지하수 오염취약성 평가에 예측 모델 활용 - 기존의 DRASTIC 방법에 예측 모델을 결합: 오염취약성에 영향을 미치는 주요 수리지질인자를 도출할 수 있었고 예측능이 향상됨	TCE농도 및 수리지질항목 자료(2003, 2004, 2008년)	Drastic 기반 취약성평가 Genetic algorithms (ANN, DT, MLR, CBR)과 ridge regression 결합 모델
Nolan et al., 2014	분석기술 개발	- 미국 캘리포니아 Central Valley 공공용수 공급 관정(천부 및 심부관정) 에서의 질산염 농도 예측 - 지표의 질소비료량 및 침투능을 결정하는 수리지질 요인들이 주요인 자로 나타남	1,255개의 천부(317개), 심부(937개) 지 하수 관정 정보, GIS 기반의 수리지질 자 료 및 질소비료 정보(2000-2010년)	Logistic regression과 RF 기반의 classification (RFC) 및 regression (RFR) 방법
Nolan et al., 2015	분석기술 개발	- 미국 캘리포니아 Central Valley 공공용수 공급 관정(천부관정)에서의 질산염 농도 예측능 평가 - 각 모델의 학습과정에서의 overfit 문제를 통해 최적의 예측능을 도출함	437개 지점 관정에서의 질산염 농도 및 41개 항목의 수리지질 인자	BRT ANN BN
Ransom et al., 2022	분석기술 개발 관리기술 개발	- 오염관리 및 대비를 위해 고농도의 질산염 농도 분포지점 예측이 필요 하며, 이에 3차원의 분포 예측 모델 개발(최초의 국가규모 3차원 XGB 모델 적용 사례) - 미국 전역에서 약 1%에 해당하는 지점만이 먹는물 수질기준을 초과하 는 것으로 예측되었지만, 이들 지역 내 거주인구의 음용의존도 높아 관 리 필요	12,082개 관정에서의 질산염 농도 및 지 점정보(기후, 토양, 관정 깊이, 토지피복) 자료 (1988-2018년)	XGB

시간에 대한 데이터 및 날씨 데이터를 수집하는 데 이용되고 있다.

Aderemi et al.(2022)은 점위치 데이터만을 제공할 수 있고, 잡음이 생기기 쉬운 기존 원격 모니터링의 문제를 IoT 및 기계 학습 기술을 사용하여 해결하였다. IoT 기반 취득 데이터는 인터넷 또는 클라우드 컴퓨팅을 통해 원격 리 데이터 센터에서 전산처리되기 때문에, 사용자 보안, 짧은 지연시간, 확장성의 문제가 발생하고 있었다. 따라서, Aderemi et al.(2022)은 이를 극복하기 위하여 계산 비효율성과 확장성에 관한 IoT 기술을 지원하고 지하수 수위 관리 모델에 대한 대체 모델을 개발하여 제안하고 있다.

3.2. 데이터 품질 관리

실제 현장에서 획득되는 데이터는 다양한 이유로 이상치를 포함하고 있다. 대수층에 중요한 영향을 미치는 외부요인(예, 기상, 인간의 개입 등)으로 인해 대수층이 유기적으로 변화함에 따라 이상치가 발생할 경우, 이 이상치는 이러한 현상들을 해석하는 데 중요한 역할을 할 수 있다. 이상치는 단순한 오류로 인해 발생할 수도 있는데, 예를 들어, 실시간 자동 모니터링 기기의 일시적 또는 영구적 결함으로 발생할 수 있으며, 수동 모니터링 데이터의 경우 데이터 취득자의 단순 기재 실수, 자료 취득 절차 미이행과 같은 행위로 인해 발생할 수 있다. 이러한 이유로 발생하는 잡음의 데이터는 데이터 기반의 분석과 모델링 결과에 부정적인 영향을 미칠 수 있기 때문에 모니터링 과정에서 발생할 수 있는 오류를 최소화하기 위한 노력이 필요하다. 최근, 모니터링 데이터 품질을 향상시키기 위한 노력의 일환으로 모니터링 기기의 정확도를 향상시키기 위한 하드웨어적 연구와 품질 향상 알고리즘 개발 등 소프트웨어적 연구가 동시에 진행되고 있는 추세다(Jeong et al., 2020a; Kim et al., 2022).

관련 연구로 Malakar et al.(2021)는 남아시아의 Indus-Ganges-Brahmaputra-Meghna 유역 내 지하수위 이상치를 예측하기 위해 지지벡터머신(Support Vector Machine, SVM)을 이용한 바 있다. 입력 자료로 지하수위, 지하수 저장 이상(groundwater storage anomalies), 월 평균 강수량, 기온 및 증발산량이 이용되었다. 분석 결과, 다양한 공간적 규모에서 예측 가능한 모델을 제작하기 위해서는 더 많은 입력자료가 필요하다는 결론을 도출하였으며, 지하수 취수가 주변보다 비정상적으로 크거나 관측정의 깊이와 공간별 분포 차이가 예측 성능을 저하시킬 수 있음을 지적한 바 있다.

그리고 Jeong et al.(2020a)은 국내 지역 별 관측된 지

하수 수위 시계열 데이터 내 이상치를 판별하여 자료의 질을 향상시킬 수 있는 기법을 개발 및 제안한 바 있다. 시계열 데이터 내 이상치 판별을 위해 순환신경망 모델을 이용한 강수량 기반 지하수위 예측을 수행하고, 예측된 지하수위를 기준으로 이상치를 판별하였다. 특히, 순환신경망 모델 학습에 이용되는 학습용 자료 내 이상치 또한 효과적으로 판별하고 이상치로 교란되지 않은 지하수위를 예측하기 위하여 다양한 형태의 비용함수(cost function)를 제시하였다. 마지막으로, Kim et al.(2022)는 과거 기상 변화에 따른 지하수위 변동 패턴 및 데이터에 내재된 불확실성을 고려하여 지하수위의 정상 범위를 시계열적으로 예측하기 위해 순환신경망 기반 기법을 제안한 바 있다. 해당 기법을 통해 과거 데이터와의 패턴을 비교하여 취득되는 시계열 지하수위 자료 내 모니터링 기기 오류에 의한 데이터 잡음을 정량적으로 감지한다.

3.3. 분석 기술

3.3.1. 수량 분석 기술

1990년대 및 2000년대 초반에 들어 지하수 관련 자료를 대상으로 기계학습 기법 적용 연구가 점차 증가하기 시작하였다(Rizzo and Dougherty, 1994; Kundzewicz, 1995; Lebron et al., 1999; Coulibaly et al., 2000, Coulibaly et al., 2001a, Coulibaly et al., 2001b). 특히, 기상변화에 따른 가용 지하수량을 산정하고 적정 지하수 이용량을 제안하기 위해 심층학습(deep learning)과 같은 기계학습 기법을 이용하여 지하수위 변동을 예측하는 연구 사례가 가장 많았다. 자료활용 측면으로 볼 때, 가뭄 예측, 함양량 추정, 지하수위 저하 요인을 분석하기 위해 다양한 시계열 및 공간 자료들이 이용되었다. 시계열 자료로는 기상자료(강수량, 증발산량 등), 지하수위, 지하수 이용량 등이 이용되었으며, 공간자료로 토지피복, 관정위치 고도, 수리전도도, 유효공극률, 지질 등의 자료가 주로 이용되고 있다. 다양한 시계열 및 공간 자료를 해석하기 위해 다양한 목적의 예측 모델이 개발되었다. 예측 모델을 개발하기 위해서는 인공신경망(artificial neural network, ANN), SVM, 순환신경망(recurrent neural network, RNN), 결정 트리 등의 단일 예측 모델과 랜덤 포레스트(random forest, RF)와 같은 앙상블 예측 기법 등의 기계학습 기법이 주로 이용되었다.

3.3.1.1. 국외사례

Coulibaly et al.(2001b)은 다양한 종류의 ANN 기반의 기계학습 기법을 이용하여 월 단위 지하수위 변동 패턴을

예측하고 이들의 성능을 비교한 바 있다. 강수량 시계열 데이터를 이용하여 지하수위 변화를 예측하였으며, 개발된 예측 모델을 지하수 모니터링 네트워크를 개발하는 데 적용하고자 하였다. Daliakopoulos et al.(2005)은 또한 그리스 지역의 Messara Valley의 지하수위 저하를 예측하기 위하여 ANN 기반의 모델을 개발하였으며, 기온, 강수량, 관측정의 깊이, 하천 유출량을 입력자료로 이용하여 18개월 이후의 지하수위를 예측한 바 있다. Sahoo et al. (2017)는 지하수 수위 변화를 예측하기 위하여 입력 데이터 전처리 방법과 하이브리드(hybrid) ANN(HANN) 모델을 개발하여 미국 농업 생산에 이용되는 두 개의 대수층 시스템(i.e., High Plains 대수층과 Mississippi River Valley 충적 대수층)에 적용한 바 있으며, 지하수위 예측에 이용된 다양한 입력 데이터(온도, 강수량, 하천유량, 관개 수요) 중 관개 수요가 가장 중요한 매개변수임을 도출하였다.

Zhou et al.(2017)은 중국 Mengcheng 현에 있는 10개 관정 내 지하수위 데이터를 이용하여 이산 파형 변환(Discrete wavelet transform, DWT) 전처리 방법과 SVM을 결합한 예측 모델을 제안하였으며, 지하수위 예측을 위해 강수량, 월별 평균 지하수위 및 연간 평균 지하수위 데이터를 이용한 바 있다. 중국의 또다른 지역인 Hetao 관개 지역에서는 14년 동안 관측된 과거 지하수위 자료, 증발산량, 강수량, 기온을 입력 자료로 사용하여 장단기메모리(long short-term memory, LSTM) 기반 지하수위 예측 모델을 제작한 바 있다(Zhang et al., 2018). 해당 연구는 데이터를 얻기 어려운 지역에서의 지하수위 예측에 해당 모델이 효과적인 대안이 될 수 있을 것이라 제안하였다. 또한, 중국 북서부 Heihe 강 유역의 지하수위를 예측하기 위해 ANN, RBF(radial basis function network) 그리고 SVM 기반 모델이 개발된 바 있으며, 이들의 활용 우수성은 MODFLOW를 이용한 수치 모사 결과와 비교되어 검증되었다(Chen et al., 2020).

슬로베니아에서는 Ljubljana Polje 대수층 지하수 수위를 예측하기 위해 선형회귀, 결정트리, RF, GBM(gradient boosting machine)을 적용한 바 있다(Kenda et al., 2018). 그리고 Bowes et al.(2019)는 기후변화에 따른 해수면 상승으로 인한 미국 Virginia 지역 해안 도시의 지하수위 상승을 조기 예측하고 해결하기 위해, 2010-2018년 동안 관측된 지하수위, 강수량, 해양 수위자료를 활용하여 LSTM 기반 지하수위 예측 모델을 개발하였다. 폭풍우 시기의 지하수위를 보다 정확하게 예측하는 모델을 개발하기 위하여 특정 기상 이벤트 발생 시의 자료를 선별하고 모델 학습에 적용하였다.

미국 Kansas 지역에서는 2002 - 2018년 기간 동안 관측된 지하수 이용량을 RF 기법으로 분석하여 지하수위 하강을 예측하였다. 지하수위 하강을 예측하기 위해 증발산량, 강수량, 토지 피복도, 지하수위 양수 데이터를 입력 인자로 이용하였다. 각 인자에 대한 중요도를 분석한 결과, 토지 피복도가 증발산량 및 강수량에 비해 예측 모델 성능 향상에 상대적으로 높은 기여를 하는 것으로 평가되었다(Majumdar et al., 2020). Zounemat-Kermani et al. (2021)은 지표수문학, 수리지질학 및 홍수기름 예측에 활용된 앙상블 기계학습(ensemble machine learning, EML) 기반의 연구들을 분석 후 단일 예측 보다 앙상블 기법을 적용하는 것이 예측의 성능을 개선하기 위해 적합함을 확인하였다.

이란의 Sirjan에서는 Gohar Zamin 철광석 광산 주변에서 획득한 공간 데이터(관측정의 위도와 경도, 수리전도도, 유효공극률, 전기저항률, 관측정의 표면 수위, 기반암 높이와 깊이) 및 시계열 데이터(배수량, 증발량, 및 강수량)를 이용하여 지하수위 변동을 예측하였다(Najafabadi-pour et al., 2022). 예측을 위해 ANN 기반의 다양한 모델 및 최적화 알고리즘을 적용하였으며, 제안된 모델들의 성능 비교 검증을 통해 RBF, ANN-LM(Lavenberg-Marquardt) 그리고 CF-BR(cascade forward using Bayesian regularization)을 통합한 모델을 최종적으로 제안하였다.

Vu et al.(2021)는 프랑스 Normandy 지역 내 지하수위 자료를 설명하는 LSTM 기반 지하수위 예측 모델을 과거 약 20 - 50년동안 관측된 지하수위 데이터를 활용하여 개발하였다. 해당 모델은 결측된 지하수위 데이터를 복원하는데에 사용될 뿐만 아니라, 미래의 지하수위를 예측하기 위해 개발되었다. 이 외에도 Ahmed et al.(2021)은 지하수의 용도를 분류하기 위한 Recommender System(RS)를 구축하기 위하여 기계학습 기법을 제안한 바 있다. 기계학습 기법으로는 SVM, KNN, Bagging, Decision Stump, CNN이 이용되었으며, 학습을 위해 샘플 날짜, 용존 칼슘, 용존 염화물, 용존 마그네슘, 용존 칼륨, 용존 나트륨, pH 등 11가지 종류의 데이터가 이용되었으며, 제안된 시스템을 통해 음용 및 농업용 지하수 이용지역을 분류하였다.

위성 영상을 이용한 해외 연구 사례로는 Hussein et al.(2020)이 있다. 아프리카 대륙 남부의 2002년 3월 - 2019년 5월 동안의 월별 지하수위 위성 영상자료를 이용하여 전체 지하수위 분포를 예측하기 위해 회귀 기반의 다양한 기계 학습 기술(extreme gradient boosting, multivariate linear regression, RF, ANN 및 지지벡터회귀

(support vector regression, SVR)가 사용되었다. 특히 SVR을 예측자로 사용하고, 제공된 재조정된 것이 더 나은 전체 예측 결과에 기여하는 것으로 나타났다.

3.3.1.2. 국내사례

국내 데이터를 대상으로 한 지하수위 예측 연구로 Yoon et al.(2011)을 예로 들 수 있는데, 이 연구에서는 ANN과 SVM을 이용하여 국내 해안지역 지하수위 변동을 예측하는 모델을 개발하였으며, 예측을 위해 과거 지하수위, 강수량, 조위 입력 자료가 이용되었다. 두 모델로 예측한 후, SVM 모델의 예측성능이 더 뛰어난 것을 확인했으며, 기계학습 기법을 이용한 예측 모델을 개발할 시 발생할 수 있는 불확실성 요소에 대하여도 언급하였다. 이 외 Jeong and Park(2019)는 효율적인 지하수위 저하를 감지하기 위한 목적으로 ARX(auto-regressive exogenous), 비선형(non-linear) ARX(NARX), LSTM 및 gated recurrent unit 기반의 지하수위 예측 기법과 데이터 전처리 방법을 제시하였다. 지하수위 변동 예측을 위해 포항-기북 지역, 진도-의신 지역의 자료(2005 - 2014년)를 이용하였으며, 일평균 기온, 강수량, 상대습도, 대기압, 및 일조시간을 입력자료로 이용하고 전처리 방법으로는 추세 및 계절성 변화 제거 방안을 제안하였다. Kim and Lee(2021)에서는 지하수위를 예측하고, 지하수위 변동에 중요한 영향 인자를 규명 및 평가하기 위해 ANN을 이용하였다. 지하수위 예측을 위한 ANN기반 모델을 개발하기 위하여 2001 - 2013년 동안 관측된 지하수위, 하천수위, 강수량, 기온 자료를 입력인자로 고려하였으며, 지하수위의 이동평균 역시 고려되었다. 해당 모델은 국내의 경기도, 서울, 강원도 일부 지역(철산, 마곡, 장위, 광명, 양평, 영월)의 자료를 검증하기 위해 이용하였다.

지하수위 예측에 기계학습을 적용하는 연구 외에 대수층의 함양율, 수리전도도 등의 수리지질학적 특성 값을 효율적으로 예측하기 위한 연구도 진행되었다. Kim et al. (2019a)은 국내 지하수위 관측지점의 함양량을 산정하기 위해 ANN 기반 모델을 개발한 바 있다. 기존 지하수위 데이터에 지하수위 변동법(water table fluctuation)을 적용하여 도출한 함양율 결과를 종속 변수로 하고, 207개 지하수 관측 지점의 지형 표고, 경사, 인근 하천, 지질, 토양 등과 같은 각 관측소의 지점 특성 15개 항목을 입력 변수로 이용하였다. 개발 모델을 충청남도 홍성군 서부면 양곡리를 대상으로 지하수 함양율을 산정하는 데 적용하여 성능을 검증한 바 있다.

이 외 Jeong et al. (2020b)은 지하수위 변동법과 같은

프로세스 기반의 지하수위 모델링 과정 없이 지하수위 변동 패턴을 기반으로 대수층의 수리 특성 값을 추론할 수 있는 딥러닝 기반 방법론을 개발하였으며, 예측 모델의 성능 향상을 위해 denoising autoencoder를 이용한 지하수위 변동 패턴의 저차원 특징 추출 과정을 제안하였다. 국내 전역의 실제 지하수위 데이터를 이용하였으며, 개발된 모델을 통해 수리전도도 및 함양량 관련 수치 값을 예측하였다. 후속 연구로서 Jeong et al.(2021)은 매년 변화하는 강수 패턴 때문에 선행 개발된 모델의 예측 성능이 저하되는 단점을 극복하기 위한 방안으로 conditional variational autoencoder 적용을 제안하고 이의 활용성을 검증한 바 있다.

행정적 활용성을 높이기 위한 기계학습 기반 연구 사례도 꾸준히 증가하고 있다. 예로, Lee et al.(2018)는 전국 월평균 지하수위(1995년 12월 - 2016년 12월)와 월강수량 합계(1974년 1월 - 2016년 12월)를 이용한 지역별 표준지하수지수(standardized groundwater level index, SGI)와 표준강수지수들(standardized precipitation indexes, SPIs)의 상관관계를 인공신경망 종류 중 하나인 NARX를 이용하여 학습시킨 후, 기상청의 기상전망 자료로부터 산정한 지역별 SPI 전망 값을 학습된 인공신경망에 입력하여 우리나라 전국 시군의 SGI를 전망한 바 있다. 또한, Kim et al.(2019b)는 행정구역(읍·면) 별로 가정용 및 일반용 관정의 지하수 이용량의 정확한 통계량을 산정하고, 행정구역 별 지하수 이용량 산정 방법을 제시하기 위하여 회귀나무(regression tree)방법을 이용한 바 있으며, 지하수 이용량 산정을 위해 현장 조사 항목(양수능력, 관정 심도, 읍면동의 구분, 급수인구, 펌프마력) 자료를 입력자료로 이용하였다. Lee et al.(2021)은 누적(stacked) LSTM 기법을 이용한 강수량 및 지하수 이용량 기반 지하수위 예측 모델을 개발 및 활용하여 제주 대정-한경 유역 내 지하수 거동특성을 파악한 바 있다. 또한, 개발된 예측 모델을 이용하여 각 지하수 관측점의 배경 지하수위 유지를 위한 강수량 대비 적정 지하수 이용량 산정 기법을 제안하였다.

국내 인공위성 데이터를 활용한 연구 사례로 Park and Jeong(2021)을 예로 들 수 있다. 해당 연구에서는 SWAT-MODFLOW를 이용하여 김천지역의 시공간적 지하수 함양분포를 제시하고, 함양량 분포 결과와 위성영상(Landsat-8)과의 상관성 분석을 수행하고 분류 및 회귀 트리(classification and regression tree, CART) 알고리즘을 적용하여 영상 기반 함양량 추정 모델을 개발한 바 있다. 또한, Lee et al.(2019)는 지하수 양수능력과 관련 요인 간의 관

계를 분석하여 지하수 잠재력을 추정하기 위해 지하수와 밀접한 관련성이 있는 13개의 데이터(수치 표고모델, 경사면, 토지 피복, 토양배수, 토성, 토양 깊이, 목재 유형, 목재 밀도, 지질, 수문지질, 미고결 대수층, 파쇄 암반 대수층, 누적 강수량)를 frequency ratio 및 boosted classification tree 데이터 마이닝 모델에 적용하였으며, 이를 바탕으로 대한민국 고양시 지하수 잠재 함양 지도를 생성하고 분석한 바 있다.

3.3.2. 수질 분석

앞 절에서 살펴본 것과 같이 지하수를 지속가능한 수자원으로 활용하기 위해서는 지하수량 변화에 대한 모니터링을 통해 안정적인 변동 범위를 평가하고 기후변화와 인위적인 영향(예, 지하수 개발과 사용)에 따른 변동을 예측하여 대비하는 것이 필요할 뿐만 아니라, 지하수 수질의 시·공간적인 분포특성을 정확히 평가하고 예측하는 기술이 필요하다. 지하수 수질 성분은 주로 물-암석 반응에 의해 결정되며, 지질암상, 반응시간, 지하수 유동 경로의 영향에 의한 수질특성을 평가할 수 있다(Edmunds et al., 2003). 또한, 인간활동 기원 오염물질의 유입으로 인한 지하수 수질변화와 오염특성을 평가할 수 있다(Coyte et al., 2019). 지하수 수질오염 특성은 오염원의 종류와 발생 지점, 그리고 지하수 유동 및 수리지질학적 특성에 의해 영향을 받기 때문에 토지피복/토지이용 정보와 오염원에 대한 정보 등을 고려하여 평가할 수 있다(Stigter et al., 2006; Kulabako et al., 2007). 지하수 수질유형은 지하수 관점에서 채취한 지하수 시료의 용존성분 분석을 수행하여 알 수 있으며, 각 성분의 오염농도 기준치 초과여부를 판단하여 수질오염 정도를 평가할 수 있다. 그러나, 수질 분석 과정은 많은 시간과 비용이 소요되고 각 지점에서의 수질특성만을 나타낸다. 이러한 분석 한계를 극복하기 위하여 지하수 수질의 공간적 분포 특성과 오염물질 유동에 관한 예측 모델링 연구들이 필요하며, 최근에는 데이터 기반 기계학습을 적용하여 보다 효율적으로 높은 예측능을 획득할 수 있는 예측 모델들이 개발되고 있다.

지하수 수질 분야에서 기계학습 기법을 적용한 연구들은 크게 지질기원(자연적) 물질 예측연구와 인위적 오염물질 예측연구로 분류된다. 자연발생 물질은 주로 비소(As), 불소(F), 철(Fe), 망간(Mn) 등에 대한 연구가 수행되었고, 인위적 오염물질 연구는 대부분 질산염(NO_3) 농도 예측에 집중되어 있으며, 수질 지수 또는 오염취약성 지표와 결합한 형태의 예측 연구들이 수행되기도 하였다.

3.3.2.1. 자연발생 오염물질 예측 연구

자연발생 물질 중 불소(F)는 화강암 경계대와 밀접한 연관성이 있는 것으로 알려져 있다. Amini et al.(2010)은 전지구에 분포하는 60,000 이상의 지점에서 획득된 불소 농도와 암상특성, 토양, 지질, 고도, 기상, 그리고 수리지질특성 자료를 입력자료로 활용해 CART, KBC(knowledge based clustering), MLR(multiple linear regression), ANFIS(adaptive neuro-fuzzy inference system), LR(logistic regression)과 같은 개별 기계학습 기법과 이들을 결합시킨 형태의 하이브리드 기법을 적용함으로써 불소 농도 분포를 보다 정확히 예측하고 주요 영향 인자를 평가하였다. 또한, 이와 유사하게 Podgorski et al. (2018)은 인도지역 전체에서 약 12,600개 지점의 불소 농도 자료와 지질, 기후, 토양 등의 자료를 바탕으로 RF 기법 기반의 지하수 내 불소 농도 예측 모델을 개발하였다. 예측결과 도출된 불소 분포도를 통해 먹는물 수질 기준을 초과하는 위험지역을 찾아 정화방안을 제시하고 자연적 불소 농도 증가를 유발하는 핵심 인자를 도출하였다. 불소이외에 자연발생 물질로서 많은 연구자들이 철(Fe), 망간(Mn), 비소(As)의 생성기작과 분포 특성에 대해 연구하고 있다.

지하수 내 고농도 망간과 철이 함유되어 있으면 맛이 변질되거나 침전물의 영향으로 지하수 이용이 어렵고 건강 위해성의 문제를 유발시킬 수 있다. 따라서, Podgorski et al.(2022)은 RF 기법과 GBR(generalized boosted regression)을 적용해 동남아시아와 방글라데시 지역 6,000개 이상의 지하수로부터 측정된 철과 망간 농도 자료와 환경인자들(기후, 지질, 토양, 지형)을 활용해 예측 모델을 개발하고 공간분포 특성을 평가하였다. 예측된 확률분포 지도로부터 고농도의 망간과 철이 분포하는 지역은 주로 인구가 밀집된 지역을 지시하고 있어, 지하수 관리 및 활용에 도움을 줄 수 있을 것으로 기대된다. Frederick et al. (2016)은 고농도의 비소와 관련된 공변량간의 관계가 서로 다른 기후 및 수리지질특성을 가지는 미국의 서부/중부/동부 지역에서 어떠한 차이가 나타나는지 CART 기법을 활용해 평가하고자 하였다. 이를 통해 지역별 수문기상과 지하수 수리화학적 특성에 따른 비소 농도 예측 및 고농도의 비소가 나타나는 지역의 주요 인자를 도출할 수 있었다. 특히, 국가 규모에서 건조도(aridity)는 고농도 비소 분포와 밀접한 상관성을 보이는 것을 밝혔으며, 지역별로는 예측 모델에 영향을 미치는 주요인자(건조도와 pH, 용존이온 등)가 다르게 나타났다.

비록 자연발생 기원의 특정 원소는 아니지만, pH는 지

하수 수질 특성을 나타내는 주요 인자이다. Stackelberg et al.(2021)은 빙하 대수층 내 pH 변화를 3차원으로 예측하기 위해, 총 18,386개의 관정으로부터 측정된 pH 자료를 활용하여 gradient boosting의 한 종류인 BRT (boosted regression trees) 모델을 활용하였다. BRT모델의 입력자료로써 일반적인 수리화학적 특성과 함께 지하수 흐름 특성(예, 지하수의 나이, 함양 이후 통과한 경로의 길이 등)을 함께 활용하였다. 비록, 개발된 BRT 모델이 3차원 pH 분포를 정확히 예측하지 못했지만, 향후 오염이 발생시 오염 취약지역을 미리 식별케 하여 지하수 모니터링 지역 선정 시 우선순위를 정하는 데에 기여할 수 있을 것으로 기대된다.

또한, Jeong et al.(2022)은 2007년부터 2020년까지 10년 이상의 기간동안 대한민국 전국에 분포한 지하수 관측공으로부터 얻어진 수질 측정 자료를 분석하여 RF 기반의 유해물질 위험 현장 예측 모델을 개발하였다. 본 연구는 현장에서 간단히 측정할 수 있는 원위치(In-situ) 지하수 측정항목(수온, pH, EC, Eh, DO 등)을 입력변수로 활용하였다는 차별성과 함께, 수리지질특성(암상과 심도 등)을 고려한데 의의가 있다. 특히, PCA (principal component analysis) 기반의 상관성 분석 및 민감도 분석 (sensitivity analysis)을 통해 유해물질과 인자들 간의 상관성을 고려해 자료를 선별함으로써 예측능을 높일 수 있었고, 지하수 내 자연 방사성 물질 우라늄과 라돈의 음용 기준 초과 위험 가능성을 평가하였다. 이러한 원위치 기반의 실시간 현장 자료를 활용하여 개발된 예측 모델이기 때문에 실제 현장 적용 가능성이 높다고 판단된다.

3.3.2.2. 인위적 오염물질 예측 연구

인위적 오염물질의 유입으로 지하수 오염이 발생할 수 있는 지역과 오염 농도를 높은 정확도로 예측할 수 있다면 보다 효율적인 지하수 관리가 가능할 것이다. 특히, 여러 인위적인 오염물질 중 지하수내 질산염의 분포와 농도를 예측하는 연구들이 많이 수행되었다. Nolan et al. (2014)는 미국 캘리포니아 Central Valley내 공공용수 공급 관정으로부터 수리지질인자들과 LR 및 RFC와 RFR (RF 기반의 classification과 regression) 기계학습을 활용하여 질산염 농도 예측 모델을 개발하고 예측능을 실측치와 비교하였다. 또한, 예측 모델은 농경지에 공급된 질소 화학비료의 대수층으로 침투능을 결정하는 수리지질특성들을 주요인자로 판별하였다. 기존 연구를 확장하여 Nolan et al.(2015)은 Central Valley내 천부 지하수만을 대상으로 BRT, ANN, BN(Bayesian networks)의 예측능

을 평가하였고, 추가 연구에서는 모델 학습과정에서의 과적합(overfitting) 문제를 제어함으로써 최적의 예측능을 도출하고자 하였다. Ransom et al.(2022)은 미국 전지역에서 총 12,082개의 관정으로부터 얻어진 질산염 농도를 이용하여 XGB(extreme gradient boosting) 기반의 질산염 농도 예측 모델을 개발하였으며, 해당 연구는 국가 규모에서 3차원의 지하수 품질 관리에 기여할 수 있을 것으로 평가하였다.

국내에서도 지하수의 수질 특성 예측 및 오염 예측 연구에 기계학습을 적용한 사례들이 보고되고 있는데, Ahn et al.(2012)은 원주 우산공단에서 측정한 TCE 농도 및 수리지질자료를 활용해 오염취약성을 평가하고자 기존 DRASTIC 방법에 기계학습을 접목시켰다. 기존 DRASTIC 방법에서는 현장측정 값과의 오염취약성 사이에 상관성이 명확하지 않았지만, ANN, DT(decision tree), MLR, CBR (case-based reasoning) 등의 유전알고리즘들(genetic algorithms)과 리지 회귀(ridge regression) 결합 모델을 적용함으로써 지하수 오염취약성에 영향을 미치는 주요 수리지질인자 4개를 도출하였고, 이를 통해 오염취약성 예측능을 향상시킬 수 있었다.

3.4. 분석결과와 현장 적용: 지능형 수자원 관리 시스템

데이터 마이닝 기술(특히, 기계학습)을 적용한 지하수 분야 연구 사례 및 예측 모델들은 대부분 지하수량 및 수질을 보다 효율적으로 관리 및 감시하고 미래 예측을 통해 행정적인 의사결정을 보조하기위해 개발되는 것으로 보인다. 따라서 본 논문에서는 데이터 사이언스 기술들이 어떻게 지하수량 및 수질 관리 시스템 개발에 활용되고 있는지에 대한 사례를 추가 분석하였다.

데이터 사이언스 기술을 지하수 분야에 적용하기 위해서는 다양한 종류와 다량의 데이터가 활용이 되어야한다. 따라서 데이터를 효율적으로 취합하고 관리하는 것이 관련 융합분야 기술 발전을 위해 선행되어야 한다. 이에 따라, 데이터 수집과 관리를 위한 시스템 개발이 최근 각 기관별로 수행되어지고 있다. 국내의 경우, 한국지질자원 연구원의 지오빅데이터 오픈플랫폼(data.kigam.re.kr), 한국수자원공사의 국가지하수정보센터(www.gims.go.kr), 한국농어촌공사의 농어촌지하수관리시스템(www.groundwater.or.kr)을 예로 들 수 있다. 국외의 경우, USGS의 National Water Information System(waterdata.usgs.gov/nwis)을 예로 들 수 있다. 이러한 시스템 개발과 더불어 개발된 시스템의 범용적 활용성을 확장시키기 위하여 추가적인 기술 개발 또한 동시에 진행되고 있다.

예를 들어, Neyens et al.(2018)은 IoT 지원 Environmental data Management Interface를 사용하여 다양한 형태의 데이터를 동일한 플랫폼에서 단순하고 신속하게 표시 및 분석할 수 있도록 시스템을 개발하였다. 이 시스템에는 데이터 분석을 위한 여러가지 도구(Temporal Chronic: 데이터 변경 사항을 그래프 형식으로 보는 데 사용; Envelop: 이전 연도에 대한 수위의 변화를 보는 데 사용; Interpolation: 1년 동안 전체 대수층에 대한 수리 전도도 변화 지도를 표시하는 데 사용; Log: 다중 지점 데이터를 보는 데 사용; Saltwater intrusion interface: 해수 침투 인터페이스를 사용하여 사용자가 정의한 위치에서의 침투양상을 시간에 따라 추적하는 데 사용)들이 탑재되어 있으며, 센서 유형에 따라 분석할 수 있도록 고안되어 있다. 또한, Chang et al.(2017)은 실시간으로 수집되는 지하수 농도가 오염에 대한 특정 기준치를 벗어날 경우, 지정된 관리자에게 메시지를 발송하여 지하수 수질 상태를 인지할 수 있는 시스템과 대응매뉴얼을 개발한 바 있다. 그리고 Nova Scotia 지질조사국은 가뭄 관리에 대비해 실시간으로 천부 대수층의 지하수위를 모니터링할 수 있는 네트워크를 개발하였으며, 여기 이용되는 맞춤형 수위계는 초음파 센서를 사용하여 지하수 깊이를 측정하고 IoT 장치를 사용하여 실시간으로 데이터를 인터넷으로 전송하여 일반인이 열람할 수 있도록 개발되었다(Nova Scotia Department of Energy and Mines 2017). 미국의 NWIS(National Water Information System)는 미국 지질조사국에서 운영 중인 물 관련 자료의 저장 및 검색을 위한 통합 시스템으로 분산 DB 구조를 토대로 측정지점, 시계열적 계측 자료, 유량, 지하수, 수질, 수자원 이용 등과 같은 데이터를 실시간으로 수집/저장/관리하고 있으며, 누구나 손쉽게 웹상에서 다양한 파일형식의 데이터를 제공받을 수 있도록 개발되고 있다. 지하수정보시스템, 수질정보시스템, 데이터처리시스템, 수자원이용정보시스템의 네 가지 하위시스템으로 구성되어 있으며, 약 150만 지점에서 수집한 수량, pH, 염도, 탁도, 수온, DO, 암모니아, 질산염, 염화물, 엽록소, 홍조류, 남조류 등의 측정 정보를 제공하고 있다(Goodall et al., 2008).

지하수 분야는 아니지만 지하수 분야에서 지하수량 및 수질을 관리하기 위하여 참조할 수 있는 지질 분야 지능형 시스템에 대한 연구 사례를 추가로 분석하였다. 먼저, 영국지질조사소는 지구시스템에 대한 디지털 트윈으로 OneGeology 4.0을 개발하고 있다. 이는 캐나다, 호주, 네덜란드, 일본, 한국, 브라질 등의 지질자원 관련 연구기관과 협력(OneGeology Strategic Steering Committee, OSSC)

하는 프로그램이며, 실제 환경에서 모니터링되는 센서 네트워크와 분석/예측 모델, 시각화 등의 결합으로 실시간 4차원 지리공간 데이터 플랫폼을 개발하고자 추진중에 있다. 이 중, 한국지질자원연구원은 활화산을 주제로 백두산의 디지털 트윈 구축을 주관하고 있다(Ahn, 2021).

NASA 첨단정보 시스템 기술(Advanced Information System Technology, AIST) 프로젝트 중, QUAKE-A(Quantifying Uncertainty and Kinematics of Earthquake Systems Analytic)에서는 합성개구레이더(synthetic aperture radar, SAR) 간섭법(interferometry), 지형 및 측지 영상 데이터를 융합하여 캘리포니아의 활성 판 가장자리에 대한 균일한 지각 변형 참조 모델을 생성하기 위한 프레임워크를 개발 중에 있다(Advanced Information Systems Technology(AIST) Program, 2019). 이외에도 지진 분야의 또 다른 연구로 Lawrence Berkeley National Laboratory의 연구사례를 들 수 있다. 이는 미국 지질조사국과 심층학습 기법 기반의 미세지진 감지 모델을 공동개발하고 캘리포니아 주, 오리건 주, 및 워싱턴 주 시민을 대상으로 휴대전화 애플리케이션(MyShake)을 개발 및 제공하여 시민들의 지진 대응 효율을 증가시키고 있다(Kong, 2021). 이처럼 지질분야에서는 지진과 관련하여 다양한 형태의 효율적 관리 및 정보 시스템이 존재하고 있으며, 이와 같은 형태의 시스템의 개념은 지하수 관리 분야에도 적극 활용될 수 있을 것으로 판단된다.

4. 앞으로의 방향

자료 취득부터 예측 기술 개발, 통합관리시스템 개발 및 활용까지 다양한 측면에서 데이터 사이언스 기술이 지하수 분야에 응용되고 있다. 그리고 데이터 사이언스 기술의 활용성은 기술적 발전과 동시에 융합적 지식이 누적됨에 따라 더욱 증폭될 것으로 예상된다. 국내외 다양한 연구사례와 최신 분야 동향을 종합적으로 살펴보았을 때, 지하수 분야에 있어 데이터 사이언스 융합 기술의 발전이 궁극적으로 추구하는 것은, 결국 지하수자원을 효율적으로 관리하기 위한 통찰력을 얻고, 더욱더 지능적으로 미래의 수자원을 관리하기 위한 보다 실용적인 기술을 개발하는 것으로 판단된다. 그러나 여전히 데이터 사이언스 기술 융합 연구에는 여러 한계점과 제약이 존재하기 때문에 이를 극복하고 지하수자원 관리 지능화의 궁극적 목적을 달성하기 위해서는 연구자들의 깊은 고찰과 노력이 지속되어야 할 것으로 보인다.

먼저, 지하수분야에 데이터 사이언스 기술을 보다 적극

적이고 실질적으로 응용하기 위해서는 다양한 자료와 정보를 통합 관리하는 플랫폼 개발이 반드시 선행되어야 하고 충분히 성숙되어야 할 것으로 보인다. 최근 ‘데이터 개방’, ‘공공데이터 활용’과 같은 용어들은 산업계 뿐만 아니라 학계 등 다양한 커뮤니티에서 흔히 언급되고 있다. 이는 데이터의 중요성에 대하여 많은 사람들이 지각하고 있음을 의미하며, 이에 따라 데이터 통합 관리 중요성에 대한 인식 또한 증가하고 있다. 이러한 추세에 따라 관련된 정부 기관 주도로 지하수자원 데이터 통합관리 플랫폼 개발이 적극적으로 이루어지고 있는 추세이다(예를 들어, K-Water 공공데이터 개방 포털). 개발된 데이터 플랫폼 기반으로 지하수 분야를 적극적으로 발전시키기 위해서는 데이터 플랫폼의 실질적 효용성에 대한 주기적인 점검과 보완이 필요할 것이다. 또한, 관측 데이터를 통합하는 것 뿐만 아니라, 모든 연구자들이 연구데이터와 결과물을 효율적으로 공유하고 토론할 수 있는 시스템의 개발이 필요하며(예를 들어, GitHub와 같은 공유 플랫폼), 각 개발된 기술 분야에 대하여 연구 결과를 서로 비교 검증할 수 있도록 표준자료를 확보하는 것 또한 고려해 볼 필요가 있다. 예를 들어, 컴퓨터 과학분야에서는 기계학습 기술의 성능검증을 위해 대표 자료로 MNIST, Iris 자료 등이 활용되고 있다. 그리고 플랫폼 구축과 더불어 연구자들은 원 자료 뿐만 아니라 연구정보를 기꺼이 공유할 수 있는 자세를 가져야 할 것이다.

현재까지 확보된 다양한 종류의 지하수 관련 자료가 질적 및 양적으로 충분하지에 대한 대대적인 검토가 필요하다. 자료의 신뢰성을 확보하지 못할 경우, 분석된 결과의 신뢰성도 확보되지 않는 것은 당연한 문제이다. 또한, 지금부터 추가로 획득되는 데이터의 품질을 향상시키기 위한 방법론적 기술 개발과 탐사 기기분야의 하드웨어적 기술 개발이 병행 발전되어야 할 것이다. 탐사 기술의 발전으로 방대한 지하수 자료가 축적되고 있지만, 데이터 기반의 연구를 수행하기에는 여전히 자료의 양이 불충분한 경우가 많다. 따라서, 불충분한 학습 자료로 파생되는 문제를 돌파할 수 있는 소프트웨어적 기술 마련이 필요할 것이다. 자료증폭(data augmentation), 전이학습(transfer learning) 등의 기술들이 이러한 문제를 해결하는데 도움이 될 것으로 판단된다.

최근에는 실제 현장 설치된 물리적 센서로부터 자료를 취득하는 것 외에도 가상센서(virtual sensor)를 이용한 자료 취득 연구의 수행 또한 이루어지고 있다(Liu et al., 2009; Cristaldi et al., 2020; Paepae et al., 2021). 가상센서는 취득된 실제 데이터를 기계학습 알고리즘과 결

합하여 이차적인 성질의 정보를 예측하는 소프트웨어적 센서 기술을 의미한다. 예를 들어, 기상자료인 온도, 습도를 측정함으로써 불쾌지수라는 자료를 만들어낼 수 있으며, 이는 기상자료와는 별개의 자료이나 생활의 편의를 제공하는 데 활용될 수 있다. 가상센서 기술을 지하수 수질 관리에도 활용하여 양질의 지하수 수원 찾기(예를 들어, 좋은 물 분포지도-한국지질자원연구원, 한국농어촌공사-기능성 지하수/농업용수)에 적용함으로써 특정 지역의 지하수자원을 명품화하여 지역경제를 활성화할 수 있을 것으로 판단된다. 실제로 청주시에서는 이미 수질로 명성이 높은 조정지하수를 음용수 뿐 아니라 관광자원으로 활용하여 조정치유마을 관광사업을 추진, 지역사회 발전을 도모하고 있고, 순창군에서는 수량이 풍부하고 수질이 우수한 강천음용수를 이용하여 힐링스파 관광산업을 도모하고 있다. 해외의 경우에도 에비앙 샘물, 루르드 샘물 수원지 주변은 지하수 관광휴양산업에 의한 지역경제 활성화가 잘 이루어진 좋은 사례이다. 이처럼 다양한 데이터 사이언스 기술을 가상센서에 적용함으로써 국내에 잘 알려지지 않은 좋은 물을 신규로 발굴해 낼 수 있을 것이며, 수원지의 지역경제 활성화에 기여를 할 수 있을 것으로 기대된다.

예측 모델 개발 분야에 있어서, 기계학습 기법의 예측 결과를 대수층의 수리지질학적 특성과 연계하여 개발 모델을 해석하는 부분에 여전히 분명한 제약이 존재한다. 이를 극복하기 위하여 설명가능한 인공지능(Explainable AI) 기술이 활용되고 있으며, 이의 활용이 증가함에 따라 기계학습을 통해 개발된 예측 모델에 대한 이해도가 향상될 수 있을 것으로 기대된다. 서술된 새로운 기술을 통해 예측 결과를 해석하기 위한 연구는 다양한 분야에서 현재 활발히 진행되고 있기 때문에, 이러한 기술이 지하수분야에도 적용된다면 지하수위 변동과 대수층의 상태를 예측하고 해석하는데 큰 도움이 될 것으로 판단된다. 또한, 자료기반 예측 기술들이 기존 프로세스 기반 모델링에 병행되어 기존 연구의 효율성을 높이고 적용 범위를 넓히는 방향으로 활용될 수 있도록 기술개발이 이루어져야 할 것이다.

다양한 연구결과와 예측 모델을 기반으로 지하수자원을 효율적으로 관리할 수 있는 시스템 개발은 여전히 미흡한 상황이다. 데이터 분석 분야에 있어 기술을 선도하고 있는 미국 또한 수자원 관리를 위한 자료 취합용 정보 시스템 개발 단계에 머물러 있다. 그러나 인공지능 분석 기술이 쏟아져 나오고 있는 현 상황에서 이런 기술력들을 조합하여 보다 효율적으로 지하수자원을 관리하기 위한 시스템 개발은 곧 실현될 것으로 예상된다. 특히, 2022년

미국, 유럽, 중국 등 세계 각국에서는 기상관측 이래 최악의 가뭄과 국지적 폭우가 발생하는 등 극단적인 이상 기상 현상으로 수자원 불균형에 따른 문제들이 나타나고 있다. 이는 기후 변화 문제가 해결되지 않는 이상 더욱 심화될 것으로 보인다. 지표수는 지하수 대비 기상학적 가뭄에 취약하다. 따라서 지표수 외 다른 수자원(지하수, 해수담수화 등)을 확보하는 수자원 다변화가 필요하며, 다변화된 수자원 활용 시설(지표댐, 지하댐, 인공함양 시설 등)들을 연계 개발 및 운영해야 될 것이다. 특히, 해당 시스템은 자료의 취득, 자료 품질 개선, 모니터링 기기 유지 보수, 예측 및 해석, 의사결정, 현장제어 등 모든 과정을 자동화할 수 있도록 개발되어야 하는데, 이에 데이터 사이언스 기술이 적극 활용될 수 있으며, 이를 통해 수자원 관련 시설들을 종합적으로 고려하여 개발 및 운영할 수 있을 것으로 기대된다. 이러한 운영 시스템은 각 다중 시설 별 기상대비 최적 수량을 효율적으로 관리하게 함으로써 수자원 보급 및 관리에 소모되는 에너지를 효율화 함으로써 탄소중립에도 기여할 수 있을 것이다. 앞으로의 미래 기후를 생각하였을 때, 이러한 시스템의 필요성은 지속적으로 증가하고 이는 반드시 실현되어야 할 문제일 것이다.

관련분야에 대한 실질적인 연구가 진행됨에 따라 앞서 기술된 내용 외에도 끊임없는 새로운 한계와 제약들이 생겨날 것이며, 이는 지하수학 및 데이터 사이언스 분야의 발전을 통해 극복되고 새로운 기술 발전의 또 다른 원동력이 됨으로써 지하수 분야의 끊임없는 발전을 이끌어 갈 것으로 기대한다.

사 사

이 성과는 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2020089 80000), 환경부 “지중환경오염·위해관리기술개발사업”(과제번호 2018002440003) 및 한국지질자원연구원의 2022년 기본사업(GP2020-012)의 지원을 받아 수행된 연구입니다.

References

- Aderemi, B., Olwal, T., Ndambuki, J., and Rwanga, S., 2022, A review of groundwater management models with a focus on IoT-Based systems, *Sustainability*, **14**(1), 148.
- Abd El-Aziz, A.A., Alsalem K.O., and Mahmood, M.A., 2021, An intelligent groundwater management recommender system, *Indian J. Sci. Technol.*, **14**(37), 2871-2879.
- Ahn, E., 2021, Analysis of digital twin technology trends related to geoscience and mineral resources after the Korean new deal policy in 2020, *Econ. Environ. Geol.*, **54**(6), 659-670.
- Ahn, J.J., Kim, Y.M., Yoo, K., Park, J., and Oh, K.J., 2012, Using GA-Ridge regression to select hydro-geological parameters influencing groundwater pollution vulnerability, *Environ. Monit. Assess.*, **184**, 6637-6645.
- Amini, M., Abbaspour, K., and Johnson, C., 2010, A comparison of different rule-based statistical models for modeling geogenic groundwater contamination, *Environ. Model. Softw.*, **25**(12), 1650-1657.
- Bowes, B.D., Sadler, J.M., Morsy, M.M., Behl, M., and Goodall, J.L., 2019, Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks, *Water*, **11**(5), 1098.
- Chang, H., Moon, B., Yoon, S., and Jin, T., 2017, Development and performance evaluation of multiple sensor for groundwater quality monitoring and remote control system using IoT, *J. Korea Inst. Inf. Commun. Eng.*, **21**(10), 1957-1963.
- Chen, C., He, W., Zhou, H., Xue, Y., and Zhu, M., 2020, A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China, *Scientific Reports*, **10**, 3904.
- Coulilaly, P., Anctil, F., and Bobée, B., 2000, Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, *J. Hydrol.*, **203**(3-4), 244-257.
- Coulilaly, P., Anctil, F., and Bobee, B., 2001a, Multivariate reservoir inflow forecasting using temporal neural networks, *J. Hydrol. Eng.*, **6**(5).
- Coulilaly, P., Anctil, F., Aravena, R., and Bobee, B., 2001b, Artificial neural network modeling of water table depth fluctuations, *Water Resour. Res.*, **37**(4), 885-896.
- Coyte, R.M., Singh, A., Furst, K.E., Mitch, W.A., and Vengosh, A., 2019, Co-occurrence of geogenic and anthropogenic contaminants in groundwater from Rajasthan, India. *Science of the Total Environment*, **688**, 1216-1227.
- Cristaldi, L., Ferrero, A., Macchi, M., Mehrafshan, A., and Arpaia, P., 2020, Virtual Sensors: a Tool to Improve Reliability, 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, 10.1109/MetroInd4.0IoT48571.2020.9138173.
- Daliakopoulos, I., Coulilaly, P., and Tsanis, I., 2005, Groundwater level forecasting using artificial neural networks, *J. Hydrol.*, **309**(1-4), 229-240.
- Drage, J. and Kennedy, G., 2020, Building a low-cost, internet-of-things, real-time groundwater level monitoring Network, *Ground Water Monit. Remediat.*, **40**(4), 67-73.

- Edmunds, W.M., Shand, P., Hart, P., and Ward, R.S., 2003, The natural (baseline) quality of groundwater: a UK pilot study, *Science of the Total Environment*, **310**(1-3), 25-35.
- Frederick, L., VanDerslice, J., Taddie, M., Malecki, K., Gregg, J., Faust, N., and Johnson, W., 2016, Contrasting regional and national mechanisms for predicting elevated arsenic in private wells across the United States using classification and regression trees, *Water Research*, **91**, 295-304.
- Goodall, J., Horsburgh, J., Whiteaker, T., Maidment, D., and Zaslavsky, I., 2008, A first approach to web services for the National Water Information System, *Environ. Model. Softw.*, **23**(4), 404-411.
- Hussein, E., Thron, C., Ghaziasgar, M., Bagula, A., and Vaccari, M., 2020, Groundwater prediction using machine-learning tools, *Algorithms*, **13**(11), 300.
- Jeong, J. and Park, E., 2019, Comparative applications of data-driven models representing water table fluctuations, *J. Hydrol.*, **572**, 261-273.
- Jeong, J., Park, E., Chen, H., Kim, K., Han, W., and Suk, H., 2020a, Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data, *J. Hydrol.*, **582**, 124512.
- Jeong, J., Jeong, J., Park, E., Lee, B., Song, S., Han, W., and Chung, S., 2020b, Development of an efficient data-driven method to estimate the hydraulic properties of aquifers from groundwater level fluctuation pattern features, *J. Hydrol.*, **590**, 125453.
- Jeong, J., Jeong, J., Lee, B., and Song, S., 2021, The applicability of conditional generative model generating groundwater level fluctuation corresponding to precipitation pattern, *Econ. Environ. Geol.*, **54**(1), 77-89.
- Jeong, J., Choung, S., Jeong, D., Kim, M., Kim, H., and Kim, J., 2022, Development of data-driven models for estimating the probability of high-concentration occurrence of naturally occurring radioactive materials in groundwater, *J. Hydrol.*, **605**, 127346.
- Kenda, K., Cerin, M., Bogataj, M., Senozetnik, M., Klemen, K., Pergar, P., Laspidou, C., and Mladenec, D., 2018, Groundwater modeling with machine learning techniques: ljubljana polje aquifer, *Proceedings*, **2**(11), 697.
- Kim, G., Hwang, C., Shin, H., and Choi, M., 2019a, Applicability of groundwater recharge rate estimation method based on artificial neural networks in unmeasured areas, *J. Geol. Soc. Korea*, **55**(6), 693-701.
- Kim, G., Kim, J., and Shin, H., 2019b, Estimation of groundwater usage for the living (domestic and business) purpose wells by using a regression tree method, *J. Geol. Soc. Korea*, **55**(6), 683-691.
- Kim, I. and Lee, J., 2021, Performance analysis of ANN prediction for groundwater level considering regional-specific influence components, *Groundwater*, **60**(3), 344-361.
- Kim, Y., Jeong, J., Park, H., Kwon, M., Cho, C., and Jeong, J., 2022, Development of a data-driven ensemble regressor and its applicability for identifying contextual and collective outliers in groundwater level time-series data, *J. Hydrol.*, **612**, 128127.
- Kong, Q., 2021, Deep Learning Based Approach to Integrate MyShake's Trigger Data with ShakeAlert for Faster and Robust EEW Alerts, Lawrence livermore national laboratory, LLNL-TR-830065.
- Kulabako, N.R., Nalubega, M., and Thunvik, R., 2007, Study of the impact of land use and hydrogeological settings on the shallow groundwater quality in a peri-urban area of Kampala, Uganda, *Science of the Total Environment*, **381**(1-3), 180-199.
- Kundzewicz, Z. (Ed.), 1995, New Uncertainty Concepts in Hydrology and Water Resources (International Hydrology Series). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511564482.
- Lebron, L., Schaap, M.G., and Suarez, D.L., 1999, Saturated hydraulic conductivity prediction from microscopic pore geometry measurements and neural networks analysis, *Water Resour. Res.*, **35**(10), 3149-3158.
- Lee, J., Kang, S., Kim, T., and Chun, G., 2018, Development of groundwater level monitoring and forecasting technique for drought analysis (II) – Groundwater drought forecasting using SPI, SGI and ANN, *J. Korea Water Resour. Assoc.*, **51**(11), 1021-1029.
- Lee, S., Hyun, Y., and Lee, M., 2019, Groundwater potential mapping using data mining models of big data analysis in Goyang-si, South Korea, *Sustainability*, **11**(6), 1678.
- Lee, S., Jeong, J., Kim, M., Park, W., Kim, Y., Park, J., Park, H., Park, G., and Jeong, J., 2021, Data-driven analysis for developing the effective groundwater management system in Daejeong-Hangyeong watershed in Jeju Island, *Econ. Environ. Geol.*, **54**(3), 373-387.
- Liu, L., Kuo, S.M., and Zhou, M., 2009, Virtual sensing techniques and their applications, 2009 International Conference on Networking, Sensing and Control, 31, doi:10.1109/ICNSC.2009.4919241
- Majumdar, S., Smith, R., Butler Jr, J.J., and Lakshmi, V., 2020, Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning, *Water Resour. Res.*, **56**(11), 11.
- Malakar, P., Mukherjee, A., Bhanja, S.N., Ray, R. K., Sarkar, S., and Zahid, A., 2021, Machine-learning-based regional-scale groundwater level prediction using GRACE, *Hydrogeol.*, **29**(3), 1027-1042.
- Najafabadipour, A., Kamali, G., and Nezamabadi-pour, H.,

- 2022, Application of artificial intelligence techniques for the determination of groundwater level using spatio-temporal parameters, *ACS Omega*, **7**(12), 10751-10764. <https://doi.org/10.1021/acsomega.2c00536>.
- National Water Information System: Mapper, <https://toolkit.climate.gov/tool/national-water-information-system-mapper>, 2022-11-17.
- NASA, 2019, Advanced Information Systems Technology (AIST) Program, <https://esto.nasa.gov/project-selections-for-aist-18/>.
- Nevo, S., Morin, E., Rosenthal, A.G., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yulea, L., Royz, M., Giladi, N., Levi, N.P., Reich, O., Gilon, O., Maor, R., Timnat, S., Shechter, T., Anisimov, V., Gigi, Y., Levin, Y., Moshe, Z., Ben-Haim, Z., Hassidim, A., and Matias, Y., 2021, Flood forecasting with machine learning models in an operational framework, *Hydrol. Earth Syst. Sci.*, **26**, 4013-4032. <https://doi.org/10.5194/hess-2021-554>.
- Neyens, D., Baisset, M., and Lovigghi, H., 2018, Monitoring the groundwater quality/quantity from your desktop – application to salt water intrusion monitoring EMI: Environmental data Management Interface, *E3S Web Conf.*, **54**, 00021.
- Nolan, B., Gronberg, J., Faunt, C., Eberts, S., and Belitz, K., 2014, Modeling nitrate at domestic and public-supply well depths in the central Valley, California, *Environ. Technol.*, **48**(10), 5643-5651.
- Nolan, B., Fienen, M., and Lorenz, D., 2015, A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA, *J. Hydrol.*, **531**, 902-911.
- Nova Scotia Real-Time Shallow Aquifer Monitoring Network, <https://novascotia.ca/natr/meb/water-resources/aquifer-network.asp>, 2022-11-21.
- Paepae, T., Bokoro, P.N., and Kyamakya, K., 2021, From fully physical to virtual sensing for water quality assessment: a comprehensive review of the relevant state-of-the-art, *Sensors*, **21**(21), 6971, <https://doi.org/10.3390/s21216971>.
- Park, S. and Jeong, G., 2021, Variation of seasonal groundwater recharge analyzed using Landsat-8 OLI data and a CART algorithm, *Eng. Geol.*, **31**(3), 395-432.
- Podgorski, J., Labhasetwar, P., Saha, D., and Berg, M., 2018, Prediction modeling and mapping of groundwater fluoride contamination throughout India, *Environ. Sci. Technol.*, **52**(17), 9889-9898.
- Podgorski, J., Araya, D., and Berg, M., 2022, Geogenic manganese and iron in groundwater of Southeast Asia and Bangladesh – Machine learning spatial prediction modeling and comparison with arsenic, *Sci. Total Environ.*, **833**, 155131.
- Ransom, K.M., Nolan, B.T., Stackelberg, P.E., Belitz, K., and Fram, M.S., 2022, Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States, *Sci. Total Environ.*, **807**, 151065.
- Rizzo, D.M. and Dougherty, D.E., 1994, Characterization of aquifer properties using artificial neural networks: Neural kriging, *Water Resour. Res.*, **30**(2), 483-497.
- Sahoo, S., Russo, T., Elliott, J., and Foster, I., 2017, Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., *Water Resour. Res.*, **53**(5), 3878-3895.
- Schleder, G., Padilha, A., Acosta, C., Costa, M., and Fazzio, A., 2019, From DFT to machine learning: recent approaches to materials science-a review, *J. Phys. Mater.*, **2**(11), 032001.
- Senozetnik, M., Herga, Z., Subic, T., Bradesko, L., Kenda, K., Klemen, K., Pergar, P., and Mladenec, D., 2018, IoT middleware for water management, *Proceedings*, **2**, 696.
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D., 2016, Mastering the game of Go with deep neural networks and tree search, *Nature*, **529**(28), 484-489.
- Stackelberg, P.E., Belitz, K., Brown, C.J., Erickson, M.L., Elliott, S.M., Kauffman, L.J., Ransom, K.M., and Reddy, J.E., 2021, Machine learning predictions of pH in the glacial aquifer system, northern USA, *Groundwater*, **59**(3), 352-368.
- Stigter, T.Y., Ribeiro, L., and Carvalho Dill, A.M.M., 2006, Evaluation of an intrinsic and a specific vulnerability assessment method in comparison with groundwater salinization and nitrate contamination levels in two agricultural regions in the south of Portugal, *Hydrogeology Journal*, **14**, 79-99.
- UK Research and Innovation (UKRI), 2019 Digital twins for the next generation of geoscience prediction and understanding: OneGeology 4.0., <http://www.onegeology.org/docs/newsEvents/digital-Twin-leaflet.pdf>.
- Vu, M.T., Jardani, A., Massei, N., and Fournier, M., 2021, Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network, *J. Hydrol.*, **597**, 125776.
- Wadekar, S., Vakare, V., Prajapati, R., Yadav, S., and Yadav, V., 2016, Smart water management using IOT, *IEEE*, doi: 10.1109/WECON.2016.7993425
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., and Lee, K.-K., 2011, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, *J. Hydrol.*, **396**(1-2), 128-138.
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J., 2018,

Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas, *J. Hydrol.*, **561**, 918-929.

Zhou, T., Wang, F., and Yang, Z., 2017, Comparative analysis of ANN and SVM models combined with wavelet preprocess for

groundwater depth prediction, *Water*, **9**(10), 781.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., and Hinkelmann, R., 2021, Ensemble machine learning paradigms in hydrology: A review, *J. Hydrol.*, **598**, 126266.