



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년01월22일

(11) 등록번호 10-2756953

(24) 등록일자 2025년01월15일

(51) 국제특허분류(Int. Cl.)

G06Q 50/26 (2024.01) G01F 23/00 (2022.01)

G01W 1/14 (2006.01) G06N 20/00 (2019.01)

G06N 3/08 (2023.01) G06Q 10/04 (2023.01)

G06Q 50/10 (2012.01) G08B 21/10 (2014.01)

H04N 7/18 (2023.01)

(52) CPC특허분류

G06Q 50/26 (2024.01)

G01F 23/00 (2022.01)

(21) 출원번호 10-2024-0003909

(22) 출원일자 2024년01월10일

심사청구일자 2024년01월10일

(30) 우선권주장

1020230164141 2023년11월23일 대한민국(KR)

(56) 선행기술조사문헌

KR101138098 B1*

(뒷면에 계속)

전체 청구항 수 : 총 1 항

심사관 : 유선중

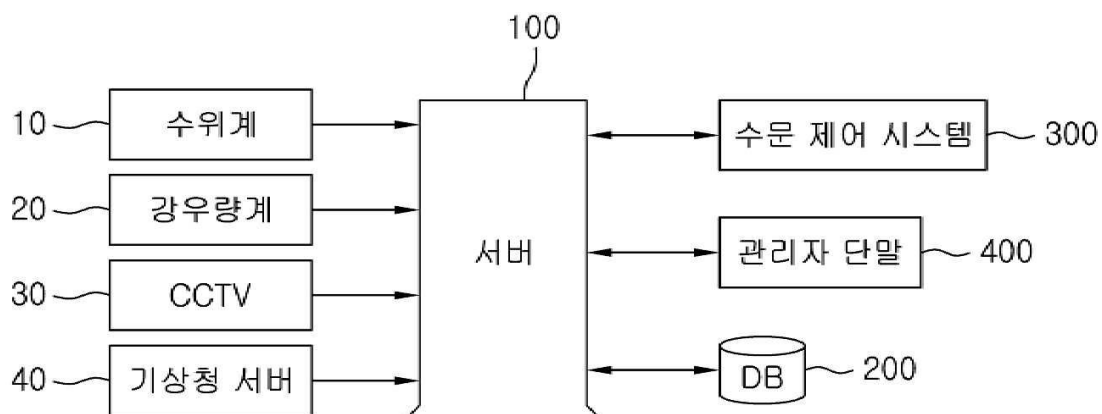
(54) 발명의 명칭 인공지능을 이용한 저수지 관리 시스템

(57) 요약

본 발명은 저수지 관리 시스템에 관한 것으로서, 관리 대상 저수지 주변의 다수의 지점에 설치된 수위계로부터 수위 데이터를 수신하고, 상기 관리 대상 저수지 주변의 다수의 지점에 설치된 강우량계로부터 강우량 데이터를 수신하고, 기상청 서버로부터 기상 정보를 제공받고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행하고, 이를 통해 상기 저수지의 수위를 예측하고, 수위 예측 결과를 포함하는 저수지 관리 정보를 제공하는 서버, 상기 수위 데이터, 강우량 데이터 및 기상 정보를 저장하기 위한 데이터베이스 및 상기 서버로부터 저수지 관리 정보를 제공받는 관리자 단말을 포함한다.

본 발명에 의하면, 인공지능을 활용한 저수지 수위 사전 예측 기술과 이를 이용한 스마트 방류 기술을 제안함으로써, 극한 호우 발생시 효율적으로 저수지에 저류된 물을 사전에 방류할 수 있다는 효과가 있다.

대 표 도 - 도1



(52) CPC특허분류

G01W 1/14 (2013.01)
G06N 20/00 (2021.08)
G06N 3/08 (2023.01)
G06Q 10/04 (2023.01)
G06Q 50/10 (2015.01)
G08B 21/10 (2013.01)
H04N 7/18 (2023.01)
G01W 2203/00 (2013.01)
Y02A 90/10 (2020.08)

(56) 선행기술조사문헌

KR1020220149105 A*
KR1020230072876 A
KR1020230061012 A
KR1020230012815 A
KR1020230052440 A
KR1020140068436 A
*는 심사관에 의하여 인용된 문헌

명세서

청구범위

청구항 1

관리 대상 저수지 주변의 다수의 지점에 설치된 수위계로부터 수위 데이터를 수신하고, 상기 관리 대상 저수지 주변의 다수의 지점에 설치된 강우량계로부터 강우량 데이터를 수신하고, 기상청 서버로부터 기상 정보를 제공받고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행하고, 이를 통해 상기 저수지의 수위를 예측하고, 수위 예측 결과를 포함하는 저수지 관리 정보를 제공하는 서버;

상기 수위 데이터, 강우량 데이터 및 기상 정보를 저장하기 위한 데이터베이스; 및

상기 서버로부터 저수지 관리 정보를 제공받는 관리자 단말을 포함하고,

상기 서버는 상기 저수지의 상류에 위치하여 물이 유입되는 상류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 상기 저수지의 하류에 위치하여 물이 방류되는 하류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행하고,

상기 저수지에 설치된 댐의 수문을 제어하는 수문 제어 시스템을 더 포함하고,

상기 서버는 예측된 저수지 수위의 변화량을 이용하여 사전에 방류해야 할 방류량을 산출하고, 상기 수문 제어 시스템을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 하고,

상기 서버는 유입량에 따른 방류량 시나리오를 생성하고, 방류량 시나리오에 따른 하류 하천의 피해를 예측하고, 하류 하천의 피해를 최소화하는 방류량 시나리오를 선택하고, 선택한 방류량 시나리오에 따라 상기 수문 제어 시스템을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 하고,

상기 서버는 자동 머신러닝(Automated Machine Learning, AutoML) 기법을 수행하여 저수지 수위를 예측하되, TPOT(Tree-based Pipeline Optimization Tool)에서 추천한 알고리즘을 이용하여 저수지 수위를 예측하고,

상기 서버는 상기 관리자 단말에 저수지 관리 정보 화면을 제공함에 있어서, 지도 화면에서 한 지점을 클릭하면 해당 지점의 예측 홍수위, 실측 및 예측 홍수위 관련 상세 화면이 표시되도록 하고, 현재 저수율과 유입 홍수 수문곡선을 고려한 사전 방류량을 결정할 수 있는 화면을 제공하고, 화면 좌측에 실시간 저수지 현황 정보, 일 평균 저수율 정보가 표시되고, 화면 중앙 상단에 저수지를 촬영하고 있는 CCTV 화면이 표시되고, 화면 중앙 하단에 상,하류 하천 수위 현황, 실시간 시간당 강우량, 날씨 정보가 표시되도록 하며,

상기 서버는 TPOT(Tree-based Pipeline Optimization Tool)에서 추천한 알고리즘을 이용하여 저수지 수위를 예측함에 있어서, TPOT에서 최적화된 유전 알고리즘을 사용하여 최적의 파이프라인을 생성하기 위해, 스택킹 기법을 사용하여 복수의 모델을 앙상블하고, 각 모델의 예측 결과를 사용하여 새로운 모델을 학습하고, 이를 통해 상기 수위 데이터, 상기 강우량 데이터 및 상기 기상정보에 따라 시간당 수위 예측 또는 실시간 수위 예측과 방류 시나리오 생성에 최적화된 예측모델을 생성하여 저수지 수위를 예측하는 것을 특징으로 하는 저수지 관리 시스템.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

발명의 설명

기술 분야

[0001] 본 특허출원은 한국농어촌공사의 오봉저수지 운영관리 연구용역(2022.02.24~2022.12.21 / 2023.02.09~2023.12.31)의 결과물로 출원된 것이다.

[0002] 본 발명은 인공지능을 이용한 저수지 관리 시스템에 관한 것으로서, 더욱 상세하게는 저수지 수위 예측, 사전 방류 기술, 하류 하천의 피해 예측을 포함하는 저수지 관리 시스템에 관한 것이다.

배경 기술

[0004] 그동안 큰 규모의 농업용 저수지를 운영하는데 있어서, 유지관리 경험을 바탕으로 저수지 수위를 관리하였으나, 최근 기후변화, 기후위기 시대에는 급변하는 강우 패턴으로 인해 저수지 수위 유지 관리의 어려움이 커지고 있다.

[0005] 최근 200년이상 빈도 강우 발생으로 인해 저수지 붕괴 위험 지속적으로 증가되고 있으며, 저수지 실측 수위 자료를 학습자료로 활용하여 사전 저수지 변동성 예측 기술 개발에 대한 요구가 증가하고 있고, 이에 따라 저수지 수위 사전 예측을 통한 사전 방류 기술 개발에 대한 요구가 증가하고 있다.

[0006] 또한, 저수지 하류부 침수피해 방지(방류)를 고려한 홍수기 적정 저수량·저수위 관리가 필요하다. 저수지 하류부 침수피해는 인적, 물적, 사회적으로 심각한 문제를 야기하고, 집중호우에 따른 수리시설의 안정성 확보 및 유지관리 필요성이 증대되고 있다.

[0007] 자연재해 중 위험성이 큰 홍수 통제뿐만 아니라 효과적인 수자원 관리 차원에서도 저수지의 수위를 정확히 예측한다는 것은 중요하다고 볼 수 있으며, 기존의 통계적 방법과 다양한 머신러닝(Machine Learning) 알고리즘들이 저수지 수위 예측에 활용되어 왔다.

[0008] 저수지의 수위 변화를 정확히 예측하고자 지도 머신러닝 알고리즘을 사용한 사례들이 다수 있으며, 그 중 인공신경망(artificial neural network) 기반의 모델들이 하천 유량과 수위, 지하수위 등과 같은 비선형 수문변수들의 모델링에 지속적으로 활용되어 오고 있다.

[0009] 특히 저수지 수위 예측 관련한 선행 기술에서는 홍수 기간의 수위 예측을 위해 뉴로 퍼지 방법(neuro-fuzzy approach)을 적용한 바가 있으며, 일단위 저수위 예측을 위해서 SVM(Support Vector Machine)을 적용하여 타 알고리즘과의 비교를 진행한 바가 있다. 또한 강수량과 수위를 입력변수로 사용하여 저수위를 예측한 것으로서는 베이저안 선형회귀(Bayesian Linear Regression)가 우수했고 강수량, 수위, 방류량이 입력변수인 경우는 부스팅 결정트리 회귀(Boosted Decision Tree Regression)가 결정 포레스트 회귀(Decision Forest Regression)와 신경망 회귀(Neural Network Regression) 보다 우수한 결과를 보여준 바가 있다.

[0010] 국내 충주댐에 대하여서도 다양한 머신러닝 알고리즘을 사용하여 저수지의 수위를 예측하는 방법을 제시한 바가 있다. 예측 모델을 개발하기 위해 저수지의 과거 수위 데이터와 기상 데이터를 활용하였으며, ARIMA(Autoregressive Integrated Moving Average), SVM(Support Vector Machines), ANN(Artificial Neural Networks) 등의 머신러닝 모델을 비교 평가하였고 실험 결과, SVM과 ANN 모델이 저수지 수위 예측에 효과적인 성능을 보였으며, 특히 ANN 모델의 경우 다른 모델보다 더 높은 예측 정확도를 나타내었다. 또한, 저수지 수위 예측에 기상 데이터가 중요한 영향을 미침을 확인하였다.

[0011] 이처럼 수위 예측에 있어서 다양한 알고리즘들의 비교 연구는 많으나 앙상블(ensemble)이나 하이브리드(hybrid) 방식의 알고리즘 적용 연구는 아직 미미한 상황이다. 기존에 순환 신경망(Recurrent Neural Network, RNN)의 일종으로, 데이터의 단기 및 장기적인 종속성을 모델링한 양방향 장기 및 단기 기억(Bidirectional Long Short-Term Memory, Bi-LSTM)과 시계열 데이터에서 다양한 모드나 성분을 분리하는 신호 분해 방법인 변동 모드 분해(Variational Mode Decomposition, VMD)라는 두 가지 기법을 결합하여 유입량 예측 모델의 정확도를 향상시킨 바가 있다. 하지만 일일 저수지 유입량 예측으로 시간위의 저수지 정보 예측 관련 연구는 그 빈도가 낮은 편이다.

[0012] 이러한 머신러닝 방법들은 수위 시계열 데이터의 복잡한 동적 특성과 시간적 의존성을 포착하는 데 어려움을 겪는 경우가 많다. 이러한 이유로 정확하고 신뢰성 있는 수문학 시뮬레이션으로 효율적인 홍수 예측이 이루어져 왔으며 SWAT(Soil and Water Assessment Tool)과 HEC-RAS(Hydrologic Engineering Center's River Analysis System)을 연계하여, 지상 강우-유출 과정과 강 하천 유체 역학을 동적으로 연결하는 자동화시스템을 개발하여, 강 수계에서 홍수기 분석시 R2 0.8 이상의 성능과 홍수 크기와 타이밍을 정확히 예측한 바가 있다. 이처럼 저수지나 강에서 역학을 이용한 방식의 수문학 물리 모델의 적용으로 홍수기 동안의 정확한 피크와 기간을 모의하는 바는 많지만 아직 머신러닝에서는 실시간으로 변화는 홍수기의 저수위 예측에 어려움이 수반되는 실정이다.

선행기술문헌

특허문헌

[0013] (특허문헌 0001) 대한민국 등록특허 10-2572746

발명의 내용

해결하려는 과제

[0014] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 저수지 수위 학습 자료와 인공지능 기술을 이용하여 저수지 수위를 예측하고, 저수지 수위 변동 예측에 따른 사전에 방류를 결정하고, 방류에 따른 하류 하천의 피해를 예측하는, 인공지능을 이용한 저수지 관리 시스템을 제공하는데 그 목적이 있다.

[0015] 본 발명의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0017] 이와 같은 목적을 달성하기 위한 본 발명은 저수지 관리 시스템에 관한 것으로서, 관리 대상 저수지 주변의 다수의 지점에 설치된 수위계로부터 수위 데이터를 수신하고, 상기 관리 대상 저수지 주변의 다수의 지점에 설치된 강우량계로부터 강우량 데이터를 수신하고, 기상청 서버로부터 기상 정보를 제공받고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행하고, 이를 통해 상기 저수지의 수위를 예측하고, 수위 예측 결과를 포함하는 저수지 관리 정보를 제공하는 서버, 상기 수위 데이터, 강우량 데이터 및 기상 정보를 저장하기 위한 데이터베이스 및 상기 서버로부터 저수지 관리 정보를 제공받는 관리자 단말을 포함한다.

[0018] 상기 서버는 상기 저수지의 상류에 위치하여 물이 유입되는 상류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 상기 저수지의 하류에 위치하여 물이 방류되는 하류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행할 수 있다.

[0019] 본 발명의 일 실시예에서 상기 저수지에 설치된 댐의 수문을 제어하는 수문 제어 시스템을 더 포함하고, 상기 서버는 예측된 저수지 수위의 변화량을 이용하여 사전에 방류해야 할 방류량을 산출하고, 상기 수문 제어 시스템을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 할 수 있다.

[0020] 상기 서버는 유입량에 따른 방류량 시나리오를 생성하고, 방류량 시나리오에 따른 하류 하천의 피해를 예측하고, 하류 하천의 피해를 최소화하는 방류량 시나리오를 선택하고, 선택한 방류량 시나리오에 따라 상기 수문 제어 시스템을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 할 수 있다.

[0021] 상기 서버는 자동 머신러닝(Automated Machine Learning, AutoML) 기법을 수행하여 저수지 수위를 예측할 수 있다.

발명의 효과

[0023] 본 발명에 의하면, 인공지능을 활용한 저수지 수위 사전 예측 기술과 이를 이용한 스마트 방류 기술을 제안함으로써, 극한 호우 발생시 효율적으로 저수지에 저류된 물을 사전에 방류할 수 있다는 효과가 있다.

- [0024] 본 발명에 의하면, 저수지의 안전성 확보 및 하류 지역 호우로 인한 침수를 예방할 수 있는 효과가 있다.
- [0025] 본 발명에 의하면, 방류 수문을 가진 저수지나, 비상 수문을 갖춘 저수지에 효율적으로 활용될 수 있을 것으로 보이고, 또한 비상 수문이 없는 작은 규모 저수지에도 활용될 수 있을 것으로 기대된다.
- [0026] 또한, 본 발명에 의하면, 저수지 수위 사전 예측 정보를 이용하여 저수지 하류 지역 주민과 농민들에게 경고방송과 문자 전송을 통해 안전을 확보할 수 있는 효과가 있다. 또한 댐, 저수지 그리고 다양한 수리 시설의 안전한 유지관리가 가능할 것으로 기대된다.

도면의 간단한 설명

- [0028] 도 1은 본 발명의 일 실시예에 따른 인공지능을 이용한 저수지 관리 시스템의 구성을 개략적으로 도시한 것이다.
- 도 2는 본 발명의 실험 대상 지역을 도시한 것이다.
- 도 3은 본 발명에서 TPOT를 통해 추천된 저수위 예측 알고리즘들 정보를 정리한 도표이다.
- 도 4는 본 발명에서 머신러닝(machine learning) 모델을 위한 입력 데이터의 통계 기술 정보값을 정리한 도표이다.
- 도 5는 본 발명의 실험에서 특성 중요도 분석 결과를 도시한 그래프이다.
- 도 6은 본 발명의 실험에서 교차검증을 위해 학습과 테스트 자료 분할 방식을 실험데이터셋으로 직접 분할 진행된 결과 그래프를 도시한 것이다.
- 도 7은 본 발명의 실험에서 1시간 후 저수위 예측모델의 총 4개 방식의 5차 교차검증 후 각 성능평가 지표들의 평균값을비교 그래프로 도시한 것이다.
- 도 8은 본 발명의 실험에서 12시간 후 저수위 예측모델의 총 4개 방식의 5차 교차검증 결과를 도시한 것이다.
- 도 9는 본 발명의 실험에서 1~12 시간 후 저수위 예측 중 대표적으로 1, 3, 6, 12시간 후 저수위 예측 결과의 성능평가를 도시한 것이다.
- 도 10은 본 발명의 실험에서 관측값과 모델들의 예측값을 시계열 변화로 도시한 그래프이다.
- 도 11은 본 발명에서 빅데이터 구축 및 분석 과정을 개념적으로 예시한 것이다.
- 도 12는 본 발명의 실험에서 하천 수위의 시계열 변화를 도시한 것이다.
- 도 13은 본 발명의 실험에서 하천의 유입량 시계열 변화를 도시한 그래프이다.
- 도 14는 본 발명의 저수지 관리 시스템에서 실시간 유입량, 저수위, 저수량, 저수율 예측 프로세스를 개념적으로 예시한 것이다.
- 도 15는 본 발명의 실험에서 실시간 저수위 및 저수량 예측을 위한 인공지능 알고리즘의 성능을 나타낸 그래프이다.
- 도 16은 본 발명의 실험에서 저수량과 강수량을 도시한 그래프이다.
- 도 17은 본 발명의 실험에서 TPOT 추천 알고리즘을 정리한 도표이다.
- 도 18은 본 발명의 실험에서 저수량과 강수량을 도시한 그래프이다.
- 도 19는 저수지에 설치된 수문을 예시한 것이다.
- 도 20은 시간에 따른 저수지 유입량을 예시한 그래프이다.
- 도 21은 본 발명의 실험에서 시간에 따른 저수율, 강수량 및 게이트 방류량을 도시한 그래프이다.
- 도 22는 예측 유입량, 예측 저수율, 예측 방류량, 예측 개도 높이의 다양한 시나리오를 예시한 것이다.
- 도 24는 본 발명의 실험에서 오봉 저수지의 방류량에 따른 저수위를 도시한 그래프이다.
- 도 25는 본 발명의 실험에서 저수지 하류부 홍수 위험도를 반영한 시나리오 기반 저수지 운영 방안을 예시한 것이다.

도 26 및 도 27은 본 발명의 실험에서 오봉 저수지 하류부 유입 지류를 고려한 홍수 추적 및 분석을 설명하기 위한 도면이다.

도 28은 본 발명의 실험에서 FLDWAV 모델의 신뢰성을 검토하기 위해 예시한 도면이다.

도 29는 본 발명의 실험에서 방류 시나리오에 따른 각 지점별 홍수위 결과를 도시한 것이다.

도 30은 본 발명의 실험에서 방류 시나리오에 따른 각 지점별 홍수위 결과를 정리한 도표이다.

도 31은 본 발명의 실험에서 시나리오에 따른 홍수 범람 예상도이다.

도 32는 본 발명의 실험에서 각 시나리오에 따른 방류량을 도시한 그래프이다.

도 33은 본 발명의 실험에서 시나리오에 따른 방류 결과를 보여주는 것이다.

도 34는 본 발명의 실험에서 방류 시나리오에 따른 하류 범람지점 예측 과정을 도시한 것이다.

도 35는 방류 시나리오에 따른 하류 범람지점 개수를 나타낸 도표이다.

도 36은 본 발명의 실험에서 방류 시나리오에 따른 하류 범람지점 예측 과정을 예시한 것이다.

도 37 내지 39는 본 발명에서 관리자 단말에서 디스플레이되는 화면예를 도시한 것이다.

도 40은 본 발명에서 TPOT 추천 알고리즘의 성능을 비교한 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0029] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시 예를 가질 수 있는 바, 특정 실시 예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0030] 본 출원에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0031] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 갖는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0032] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0033] 도 1은 본 발명의 일 실시예에 따른 인공지능을 이용한 저수지 관리 시스템의 구성을 개략적으로 도시한 것이다.

[0034] 도 1을 참조하면, 본 발명의 일 실시예에 따른 저수지 관리 시스템은 서버(100), 데이터베이스(200), 수문 제어 시스템(300) 및 관리자 단말(400)을 포함한다.

[0035] 서버(100)는 관리 대상 저수지 주변의 다수의 지점에 설치된 수위계(10)로부터 수위 데이터를 수신하고, 관리 대상 저수지 주변의 다수의 지점에 설치된 강우량계(20)로부터 강우량 데이터를 수신하고, 관리 대상 저수지 주변의 다수의 지점에 설치된 CCTV(30)로부터 영상 데이터를 수신하고, 기상청 서버(40)로부터 기상 정보를 제공받는다. 이때, 서버(100)는 AWS(Automatic Weather System) 지점으로부터 오픈 API(Application Programming Interface)를 통해 단기 예보 강수량 정보를 포함하는 기상 정보를 제공받을 수 있다.

[0036] 서버(100)는 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행하고, 이를 통해 저수지의 수위를 예측하고, 수위 예측 결과를 포함하는 저수지 관리 정보를 제공한다.

- [0037] 데이터베이스(200)는 수위 데이터, 강우량 데이터, 영상 데이터 및 기상 정보를 저장하는 역할을 한다.
- [0038] 관리자 단말(400)은 저수지를 관리하는 관리자가 사용하는 단말로서, 서버(100)로부터 저수지 관리 정보를 제공 받는다. 본 발명의 일 실시예에서 관리자 단말(400)은 데스크탑 컴퓨터, 랩탑 컴퓨터, 태블릿 PC, 스마트폰, 이동통신 단말기 등 다양한 형태로 구현될 수 있다.
- [0039] 수문 제어 시스템(300)은 저수지에 설치된 댐의 수문 개폐 동작 등 수문을 제어하는 역할을 한다.
- [0040] 서버(100)는 저수지의 상류에 위치하여 물이 유입되는 상류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 저수지의 하류에 위치하여 물이 방류되는 하류 하천의 주변의 다수의 지점에 설치된 수위계와 강우량계로부터 수위 데이터와 강우량 데이터를 수신하고, 수신한 수위 데이터, 강우량 데이터 및 기상 정보를 입력으로 하여 인공지능 기반 머신러닝을 수행할 수 있다.
- [0041] 서버(100)는 예측된 저수지 수위의 변화량을 이용하여 사전에 방류해야 할 방류량을 산출하고, 수문 제어 시스템(300)을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 할 수 있다.
- [0042] 서버(100)는 유입량에 따른 방류량 시나리오를 생성하고, 방류량 시나리오에 따른 하류 하천의 피해를 예측하고, 하류 하천의 피해를 최소화하는 방류량 시나리오를 선택하고, 선택한 방류량 시나리오에 따라 상기 수문 제어 시스템을 통해 수문의 개폐를 제어하는 방식으로 산출한 방류량이 방류되도록 할 수 있다.
- [0043] 서버(100)는 자동 머신러닝(Automated Machine Learning, AutoML) 기법을 수행하여 저수지 수위를 예측할 수 있다. 특히, 서버(100)는 TPOT(Tree-based Pipeline Optimization Tool)에서 추천한 알고리즘을 이용하여 저수지 수위를 예측할 수 있다.
- [0045] 본 발명에서 제안하는 저수지 관리 시스템을 실제로 구축하고 그 결과를 실험하였으며, 그 내용을 기술하면 다음과 같다.
- [0046] 도 2는 본 발명의 실험 대상 지역을 도시한 것이다.
- [0047] 도 2에서 보는 바와 같이, 본 발명의 실험에서는 강릉시 성산면에 위치한 오봉저수지를 대상으로 실시간 저수위 측정을 위한 실험을 진행하였다. 본 발명의 실험에서 오봉저수지를 포함한 저수지로 유입되는 하천인 왕산천, 도마천 그리고 방류되는 하류지역인 강릉남대천과 오봉저수지 인근 어홀천을 포함한다.
- [0048] 본 발명의 실험에서 예측 모형의 입력자료가 되는 측정값을 생성하는 측정지점들은 A~E로 표기된 지점들로 A지점인 도마1과 C지점인 왕산1에는 수위계와 강우량계를 설치하여 자료를 제공받고, B지점인 도마2, D지점인 왕산2에서는 강우량 정보를 제공받는다. 어홀천에 위치한 E지점과 강릉남대천에 위치한 F지점에서는 수위 정보를 수신받고 이 측정자료들은 모두 실시간으로 입력자료로 사용한다. 이처럼 본 발명의 실험에서 계측을 위해, 다수 현장에 강우량계, 수위계, CCTV 등의 다양한 계측기가 설치되어 있으며, 정기적으로 계측기를 유지 관리하였다.
- [0049] 그리고 기상청의 초단기 실황 강수량 제공은 왕산면 AWS(Automatic Weather System) 지점 자료를 오픈 API(Application Programming Interface)로 초단기 1~4시간 후 예보 강수량과 1~12시간 후 단기예보 강수량 정보를 함께 가져온다.
- [0050] 본 발명의 실험에서 홍수량 예측에 적합한 미래 기후변화 시나리오 도출 및 극한 강우사상 분석을 수행하고, 기후변화 시나리오에 따른 일단위 강우자료의 시간 단위 강우자료로의 변환 및 데이터베이스를 구축하였다.
- [0051] 그리고, 유입 홍수량에 따른 저수지 운영률(Operation rule)을 반영한 저수지 유출량을 예측하고, 다양한 강우 시나리오 기반 저수지 유입량을 반영한 데이터베이스를 구축하여 머신러닝의 학습 데이터로 활용하였다. 그리고, 저수지 방류를 고려한 시계열 저수위 변동 예측하여 시나리오 기반 모의결과를 포함하는 데이터베이스를 구축하였다.
- [0052] 본 발명의 실험에서 머신러닝 기법을 적용하여 저수위와 같은 시계열 자료를 학습하고, 시간에 따른 목표값(target)을 예측하려면 지도학습 회귀모델을 사용해야 한다.
- [0053] 지도학습 회귀모델은 크게 선형과 비선형 모델로 분류할 수가 있다. 선형모델로는 최소제곱법(Ordinary Least Squares, OLS)과 정규화(regularized) 방법을 적용한 라쏘(Least Absolute Shrinkage and Selection Operator, LASSO), 능형 회귀분석(ridge regression analysis), 엘라스틱 넷(Elastic Net)을 많이 적용한다.
- [0054] 비선형모델의 경우는 신경망(NN, Neural Network)이 대표적이며, 신경망 기반의 RNN(Recurrent Neural Network), LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit) 및 CNN-LSTM(Convolutional Neural

Network-LSTM) 등의 딥러닝 모델들이 있으며, 그 외 랜덤포레스트(Random Forest)와 엑스트림 그라디언트 부스트 등의 부스트 트리와 같은 트리 기반 모델들을 주로 사용한다. 이 모델들 외에도 배깅(Bagging)기법과 같은 앙상블 방식과 완전 하위집합 회귀(CRS, complete subset regression)나 선형 및 비선형 모델의 아이디어를 결합한 새로운 머신러닝 예측 방법을 생성하는 하이브리드 방법 등이 있다.

[0055] 이처럼 많은 회귀 알고리즘들 중 저수위 예측에서 우수한 모델을 보다 빠르게 최적화시켜서 추천받기 위해 자동 머신러닝(Automated Machine Learning, AutoML) 기법을 활용할 수 있다. 자동머신러닝은 머신러닝의 모델 선택, 하이퍼파라미터 튜닝, 특성 공학 등의 과정을 자동화하여 효율적인 예측, 분류, 판단 작업을 수행하는 인공지능 분야이다.

[0056] 자동머신러닝은 데이터 전처리, 알고리즘 선택, 하이퍼파라미터 최적화, 모델 평가 등의 여러 핵심 구성요소로 구성되며, 머신러닝 파이프라인을 자동화하여 데이터 과학자와 머신러닝 전문가의 수작업을 줄이고 효율성을 향상시킨다. 자동머신러닝 종류로는 Keras 모듈을 기반으로 하며 자동 딥러닝(Deep Learning) NAS(Neural Architecture Search)에 중점을 둔 파이썬 라이브러리인 Auto-Keras와 PyTorch 프레임워크와 다중 충실도 최적화를 사용하여 최상의 아키텍처의 네트워크 유형, 계층 수, 활성화 기능 등의 매개변수를 검색하는 방식을 사용하는 Auto-PyTorch 등이 있다. 그 외 사이킷런(Scikit-Learn) 프레임워크 위에 구축된 자동머신러닝 라이브러리로 대표적인 Auto-Sklearn이 있으며, 본 발명에 사용되는 TPOT(Tree-based Pipeline Optimization Tool)도 회귀모형에서 많이 사용되고 있다. 그 외 AutoGluon, H2O AutoML 등이 있다.

[0057] TPOT은 최적의 파이프라인을 생성하기 위해 유전 알고리즘을 사용한다. 유전 알고리즘은 진화의 개념을 사용하여 최상의 솔루션을 찾는 기술이다. TPOT은 적합한 특성, 전처리, 알고리즘, 하이퍼파라미터를 찾기 위해 최적화된 유전 알고리즘을 사용하여 최적의 파이프라인을 생성하기 위해 여러 개의 모델을 앙상블한다. 앙상블을 위해 스택킹 기법을 사용하는데 이는 여러 모델을 학습하여 각 모델의 예측 결과를 사용하여 새로운 모델을 학습한다. 즉 예측 결과를 기반으로 새로운 모델을 학습하므로 앙상블 모델은 개별 모델보다 더 나은 예측 성능을 제공할 수 있다. 또한 다양한 모델을 탐색하여 최상의 모델을 찾는데 TPOT은 선형 모델, 결정 트리, 랜덤 포레스트, 그라디언트 부스팅 등 다양한 모델을 사용한다. 그리고 알고리즘을 결합하여 더 나은 성능을 제공하는데 예를 들어, 랜덤 포레스트와 그라디언트 부스팅을 결합하여 랜덤 포레스트의 안정성과 그라디언트 부스팅의 예측 성능을 결합할 수가 있게 되는 것이다. 따라서, TPOT은 다양한 모델과 알고리즘을 결합하여 최적의 파이프라인을 생성하고 스택킹 기법을 적용하여 예측 성능을 최대화한다. 본 발명의 실험에서는 TPOT에서 추천해주는 알고리즘들을 4개 방식의 시계열 자료 5차 교차검증 방식을 통해 모델을 선정하고자 한다.

[0058] 도 3은 본 발명에서 TPOT를 통해 추천된 저수위 예측 알고리즘들 정보를 정리한 도표이다.

[0059] 도 3을 참조하면, 본 발명에서 12개 알고리즘들 중 유사한 알고리즘은 생략하고 총 7개 종류의 파이프라인 모델들만 선정하여 이중에서 1~12시간 후 저수위 예측에 있어서 모두 우수한 결과를 보여줄 대표모델을 선정하고자 한다. 먼저, 7시간과 12시간 후 예측에 추천된 엑스트라 트리, 스택킹한 능형(Ridge) 회귀, 단순 능형회귀의 앙상블모델(ET_S(R)_R), 1, 6, 11시간 후 예측에는 라쏘라즈(LL)가, 그 외 최소-최대 스케일러 전처리를 적용한 능형회귀(m_R), 원핫인코더 전처리 적용한 능형회귀 (o_LL), 스택킹 라쏘라즈와 라쏘라즈 앙상블 모델(S(LL)_LL), 리버스트 스케일러 전처리 후 라쏘라즈(r_LL), 스탠다드 스케일러 전처리 후 능형회귀(s_R)이며 차후 시계열 모델 교차검증 방식을 통해 가장 우수한 모델을 선정하고자 한다. 여기서 한가지 모델로 통일하려고 하는 바는 실시간 예측 시스템에서 시간별 예측값의 일관성을 주기 위함이다. 즉 각 시간별로 알고리즘을 달리하면 시계열 증가 혹은 감소 추세에서 벗어나는 결과를 나타내는 현상을 보이기도 하기에 대표 알고리즘으로 통일하고자 한다.

[0060] 오봉저수지 저수위 예측 모형 개발을 위한 학습과 교차 검증에는 2021년 8월 2일 15시부터 2023년 6월 22일 3시의 시단위 자료들을 사용하였다. 그리고 이 기간동안 측정된 강우량과 하천 수위, 기상청 조단기와 단기예보 강수량, 그리고 농촌용수종합정보시스템 (<https://rawris-am.ekr.or.kr/wrms>)의 실시간 계측정보를 활용하였다.

[0061] 홍수기 예측시 저수위 예측에 있어서 정확도를 향상시키기 위해 학습기간 동안의 부족한 홍수 발생 저수 정보를 보강하고자 고수위 발생 시나리오에 대한 정보를 추가 확보하였다. 100년 빈도 중 10시간 동안 강수량 307.7 mm, 12시간 동안 340.9 mm, 24시간 동안 523.0 mm 발생 시나리오와 200년 빈도 중 10시간 동안 강수량 338.6 mm, 12시간 동안 377.5 mm, 24시간 동안 580.3 mm 시나리오, 그리고 500년 빈도의 10시간 동안 강수량 379.3 mm, 12시간 동안 423.5 mm, 24시간 동안 654.5 mm 발생 시나리오로 총 9개 고수위 발생 시나리오 기간의 입력자료를 생성하였다. 또한 누적 강수량이 359 mm로 강릉지역 대부분의 도로가 침수되었던 태풍 “미탁” 기간(2019년 10월 2일 9 ~ 10월 3일 15시) 동안의 오봉저수지 고수위 저수량 정보도 포함하였다. 이때 고수위 정보를 생

성하기 위해 FLDWAV 모델링 결과와 하천 수위 머신러닝 비선형 회귀모형을 사용하였다.

- [0062] 기상자료개방포털 (<https://data.kma.go.kr>)에서 오픈 API로 제공받고 있는 자료들로는 동네예보 구역(5km 격자)에 대한 대표 AWS 관측값과 예보값으로 왕산면 지점(128.840219444444° E, 37.6698555555555° N, x좌표 92, y좌표 130)의 초단기 실황(http://apis.data.go.kr/1360000/VilageFcstInfoService_2.0/getUltraSrtNcst)과 초단기 예보(http://apis.data.go.kr/1360000/VilageFcstInfoService_2.0/getUltraSrtFcst), 단기 예보(http://apis.data.go.kr/1360000/VilageFcstInfoService_2.0/getVilageFcst)를 받고 있다. 초단기 실황과 예보는 예보시점부터 6시간 이내에 대하여 행하는 예보로 매시 30분에 발표되며, 단기예보는 예보기간과 구역을 시공간적으로 세분화하여 발표하는 자료로 3시간 간격으로 발표되고 있으며 최대 72시간 이내 예보 정보를 얻을 수 있으며, 본 발명에서는 매 3시간마다 제공받아서 6시간과 12시간 누적강수량을 입력자료로 구성하고 있다.
- [0063] 도 4는 본 발명에서 머신러닝(machine learning) 모델을 위한 입력 데이터의 통계 기술 정보값을 정리한 도표이다.
- [0064] 도 4를 참조하면, 강수량 예보정보와 고수위 저수 정보까지 보유한 총 12,977개 샘플로 구성된 23개 입력변수들의 평균, 표준편차, 최솟값, 1분위수, 중앙값, 3분위수, 최댓값, 왜도 (skew), 첨도 (kurtosis)와 같은 통계기술정보값이 정리되어 있다.
- [0065] 입력변수들을 살펴보면 baseDate는 날짜정보로 1~365일의 연중일(day of year)을 사용하여 계절적 정보를 학습할 수 있도록 하였고, baseTime은 시간적 변화를 인식할 수 있게 하였다. 여기서, w_t_add_1~4는 왕산면 초단기예보 1~4시간 후 강수량 예보정보이며, w1~d2_r은 왕산과 도마에 설치된 강우량계 측정값이다. w~k_l은 왕산천, 도마천, 어흘천, 강릉남대천에서 측정되고 있는 수위계 실시간 측정값이며, o_wa과 o_wl은 웹크롤링 (web crawling) 방식으로 가져온 농촌용수종합정보시스템의 저수량과 저수위 실시간 정보이며, o_wa_b1~3은 농촌용수종합정보시스템 오봉저수지의 1~3시간 전 측정 저수량이고, w_t_add_6~12는 단기예보를 통한 6시간과 12시간 예보누적강수량 정보이다. 통계분석 결과에서 저수위와 저수량의 분포는 왼쪽으로 치우쳐진 비대칭 분포를 가지며 첨도값이 음수로 중앙 부분이 완만하고 꼬리가 얇은 분포를 나타냄을 알 수 있으며, 저수위 자료의 중앙값은 118.13 m이며, 최대값은 119.55 m를 나타내었다.
- [0066] 입력자료들 가운데 머신러닝시에 중요한 역할을 하는 특성을 살펴보기 위해 트리기반 모델들 중 우수한 성능을 보이는 엑스트라 트리회귀(ExtraTree Regressor)을 이용해서 특성중요도를 살펴보았다. 엑스트라 트리 알고리즘은 앙상블 학습의 일종으로, 여러 개의 의사결정 트리를 구성하고 특성 중요도를 계산하는 데 활용된다. 알고리즘의 'feature_importances_' 속성을 통해 각 특성의 상대적 중요도를 추출하였다. 'feature_importances_' 속성은 엑스트라 트리 알고리즘에서 제공되며, 각 특성이 예측 모델에 기여하는 정도를 나타내는 값으로 해석된다.
- [0067] 본 발명에서는 자동화 머신러닝 TPOT을 통해 우수 알고리즘을 다시 선정하는 과정을 거치기 때문에 엑스트라 트리 모형에서의 특성중요도가 높다고해서 다른 알고리즘에서의 중요도도 크다고 볼 수는 없지만 어느 정도의 객관성을 입증하고자 입력변수별 중요도를 살펴보았다. 또한 본 발명의 최종 결과에서 엑스트라 트리 알고리즘을 포함한 파이프라인 모델이 우수한 결과를 보여준 바가 있기에 트리기반 알고리즘들 중 엑스트라 트리 모형으로 특성 중요도를 분석하였다. 총 23개 중 상위 10개만을 나타내도록 그래프를 제시하였다.
- [0068] 도 5는 본 발명의 실험에서 특성 중요도 분석 결과를 도시한 그래프이다.
- [0069] 도 5 (a)와 (b)는 모든 자료에 대한 중요도 평가이며, (c)와 (d)는 왕산면 AWS 강수량 측정값 기준으로 강수량이 0.5 mm 이상 발생한 시간대들의 자료들만 필터링한 자료들을 기반으로 중요도를 분석한 결과이다.
- [0070] 모든 자료에서 중요도가 높은 특성은 1시간 후 예측에서는 o_wa_b1인 1시간 전 저수량 정보이며, 12시간 후 예측에서는 현재 저수량 정보가 가장 높은 중요도로 모두 25% 넘는 수치를 나타내었다. 저수량과 저수위 정보 외에는 연중일을 나타내는 baseDate 중요도가 높았으며 그 외 기상청의 초단기와 단기예보 정보(w_t_add_1~12)의 영향력은 낮은 편이었다. 반면 강수가 발생한 자료의 경우에는 저수량과 저수위의 실시간 측정값과 이전 저수량 정보뿐만 아니라 w_t_add_12인 단기예보를 통한 12시간 예보누적강수량이 저수위 예측에서의 특성중요도가 높음을 확인할 수 있다.
- [0072] 독립적이고 동일하게 분산된 것으로 가정되는 정규분포 데이터와 달리 시계열 데이터는 이러한 가정을 따르지 않는 고유한 특성을 가진다. 실제로 각 샘플은 이전 샘플에 의존하므로 샘플 순서를 변경하면 데이터 해석이 달라지기 때문이다. 이에 본 발명에서는 시계열 자료 특성에 맞는 교차검증기법을 적용하였고 K-Fold 교차검증도

같이 진행했으며 이에 총 4개 방식의 교차검증이 진행되었다.

[0073] 도 6은 본 발명의 실험에서 교차검증을 위해 학습과 테스트 자료 분할 방식을 실험데이터셋으로 직접 분할 진행 된 결과 그래프를 도시한 것이다.

[0074] 도 6에서 (a) KFOLD (shuffle=T)는 K-Fold 교차검증(섞음)으로 데이터를 무작위 섞은 후 K개의 폴드(fold)로 나누어서 모델을 K번 반복 학습하고 평가하는 기법이며 데이터과학자들이 모델 검증에 주로 사용하는 방식이다. (b) KFOLD (shuffle=F)는 K-Fold 교차검증 (섞지 않음)으로 (a)와 동일하지만 데이터를 섞지않고 분할하는 것이고, (c) BlockingTimeSeriesSplit는 블로킹 시계열 분할 교차검증으로 시계열 데이터에 적합한 방식으로, 시계열 데이터에서는 시간의 흐름에 따라 데이터에 상관관계가 있을 수 있기에 이를 고려하여 시계열 데이터를 블록(block)으로 나누어서 모델을 학습하고 평가하는 기법이다. (d) TimeSeriesSplit는 시계열 분할 교차검증으로 시계열 데이터를 연속된 순서대로 분할하여 모델을 학습하고 평가하는 기법으로 시간의 흐름에 따라 데이터의 패턴이 변화하는 시계열 데이터에서 모델의 성능을 정확히 평가하기 위해 사용된다. K-Fold와 유사하게 작동하여 테스트 세트를 생성하지만 차이점은 데이터가 처음부터 섞이지 않고 다음 반복의 훈련 세트가 이전 반복의 상위 세트라는 것이다. 즉, 반복 횟수에 따라 훈련 세트가 계속 커진다.

[0075] 모든 회귀 모델 성능평가를 위해서 사용된 방식으로는 NSE, R^2 , RMSE, MAE를 사용하였으며, NSE는 수리모델의 예측과 관측값 사이의 일치도를 평가하는 지표로 $-\infty$ 에서 1까지의 범위를 가지며, 1에 가까울수록 모델이 관측값과 잘 일치함을 나타낸다. R^2 는 회귀 분석에서 사용되는 예측 모델의 설명력을 나타내는 지표로 0에서 1 사이의 범위를 가지며, 1에 가까울수록 모델이 종속 변수의 변동을 잘 설명한다는 것을 의미한다. RMSE는 예측값과 관측값 차이를 제곱하여 평균한 뒤 제곱근을 취한 값으로, 예측 모델의 예측 오차의 크기를 나타내며, MAE는 예측 모델의 예측값과 관측값 간의 절대적인 차이를 측정하는 지표로 RMSE와 MAE는 작을수록 모델의 예측 정확도가 높다고 판단할 수 있다. 수식은 아래와 같고 여기서 O_t 는 t번째의 실제 값이고, $\bar{O}_t$ 는 실제 값의 평균이고, M_t 는 t번째의 예측값이며, n은 자료의 총 개수이다.

[0076]
$$NSE = 1 - \frac{\sum (O_t - M_t)^2}{\sum (O_t - \bar{O}_t)^2} \quad (1)$$

[0077]
$$RMSE = \sqrt{\frac{\sum (O_t - M_t)^2}{n}} \quad (2)$$

[0078]
$$MAE = \frac{1}{n} \sum |M_t - O_t| \quad (3)$$

[0079]
$$R^2 = \frac{[\sum (O_t - \bar{O}_t)(M_t - \bar{M}_t)]^2}{\sum (O_t - \bar{O}_t)^2 \sum (M_t - \bar{M}_t)^2} \quad (4)$$

[0081] 도 7은 본 발명의 실험에서 1시간 후 저수위 예측모델의 총 4개 방식의 5차 교차검증 후 각 성능평가 지표들의 평균값을비교 그래프로 도시한 것이다.

[0082] 도 7에서 K-Fold 5차 교차검증(섞음), K-Fold 5차 교차검증(섞지 않음), 블로킹 시계열 분할 5차 교차검증, 시계열 분할 5차 교차검증 결과 모두 NSE, R^2 이 1.00의 값을 나타내었고, RMSE, MAE도 0.08 m 이하이며 모든 검증방식에서 7개의 모델들의 성능 차이도 크게 찾아볼 수가 없이 대부분 우수한 결과를 나타내었다.

[0083] 도 8은 본 발명의 실험에서 12시간 후 저수위 예측모델의 총 4개 방식의 5차 교차검증 결과를 도시한 것이다.

[0084] 도 8에서 교차검증 방식에 따른 성능평가 결과의 민감성을 살펴볼 수 있으며, 특히 시계열 자료 교차검증 방식으로 효과적인 블로킹 시계열 분할 5차 교차검증과 시계열 분할 5차 교차검증에서는 K-Fold 5차 교차검증 방식들 보다 성능평가 점수들이 낮았으며 7개 모델들의 차별성을 보여주었다.

[0085] 성능평가 지표 중 NSE를 기준으로 블로킹 시계열 분할 5차와 시계열 분할 5차 교차검증 평균값은 ET_S(R)_R인 엑스트라 트리회귀와 스택킹 방식이 적용된 능형회귀, 그리고 단순 능형회귀 알고리즘들의 앙상블모델이 NSE, R^2 이 0.93 ~ 0.95로 가장 우수하였고, 블로킹 시계열 분할 5차 교차검증결과에서는 m_R로 표기된 최소-최대 스케일러 전처리를 적용한 능형회귀 파이프라인모델이 0.79로 가장 성능이 떨어졌고, 시계열 분할 5차 교차검증결과

에서는 스탠다드 스케일러 전처리 후 능형회귀 알고리즘을 사용한 s_R 이 0.81로 가장 낮은 NSE를 나타내었다. 시계열 데이터는 불균형이나 트렌드의 변화가 자주 발생하기에 K-Fold 교차 검증방식으로는 모델의 일반화 능력을 정확하게 평가하지 못할 수 있다. 이에 블로킹 시계열 분할과 시계열 분할 5차 교차검증 결과를 중심으로 살펴보면 1~12시간 후 저수위 예측 모델로 $ET_S(R)_R$ 앙상블모델을 사용함이 좋을 것으로 판단된다.

[0087] 본 발명의 실험에서 학습과 5차 교차검증에 사용된 총 12,977개 샘플로 학습한 후 머신러닝 모델을 저장하고 생성된 모델들로 2023년 6월 22일 4시부터 2023년 8월 12일 23시까지의 테스트 기간 동안의 오봉저수지 저수위를 실시간 예측한 결과를 분석하였다. 태풍 카눈(Khanun)도 테스트 기간에 포함되며, 8월 9일 22시에 왕산면 AWS 측정 최대 강수량이 20.5 mm로 기록되었다.

[0088] 시계열 모델 교차검증 방식을 통한 가장 우수한 모형으로 선정한 모형을 포함하여 교차검증에 적용된 7개 알고리즘 모두를 적용하여 테스트기간 동안의 저수위를 예측하고 그 모형들의 성능평가와 시계열 그래프도 함께 제시하였다.

[0089] 도 9는 본 발명의 실험에서 1~12 시간 후 저수위 예측 중 대표적으로 1, 3, 6, 12시간 후 저수위 예측 결과의 성능평가를 도시한 것이다.

[0090] 도 9를 참조하면, 본 발명의 실험에서 1~12 시간 후 저수위 예측 중 대표적으로 1, 3, 6, 12시간 후 저수위 예측 결과의 성능평가는 대부분 결과가 좋았지만 6시간 후 예측에서는 엑스트라 트리와 스테킹 능형회귀와 능형회귀 앙상블모델 ($ET_S(R)_R$)의 예측 성능이 NSE, R^2 이 0.99, RMSE, MAE는 0.26 m 이하로 타 알고리즘들에 비해 더 높은 예측정확도를 보여주었다. 하지만 12시간 후의 경우는 앞서 시계열 분할 교차검증에서 최고 높은 성능을 보인 S(LL)_LL의 결과가 좋은 것을 확인할 수 있다.

[0091] 도 10은 본 발명의 실험에서 관측값과 모델들의 예측값을 시계열 변화로 도시한 그래프이다.

[0092] 도 10을 참조하면, 1시간 후 예측인 (a)에서는 모델들간의 차이성이 나타나지 않으며 모두 유사한 시계열변화 패턴을 나타냈지만, 6시간 후, 12시간 후 저수위 예측인 (b, c)에서 모델 예측값들은 관측값에 비해 변동성이 크고 불안정한 패턴 변화에서 이상치들이 나타나고 있다. 특히 태풍 발생 기간 (2023. 8. 8 ~ 2023. 8. 10)의 추세 일치성은 다소 떨어짐을 살펴볼 수 있다. 이 기간동안 입력자료인 도마천과 강릉남대천의 수위 결측값이 발생하여 품질관리 진행 후의 값들이 입력자료로 사용되었지만 평상시 예측 보다는 다소 낮은 예측 경향을 살펴볼 수 있다.

[0093] 6시간 후 예측에서는 앞서 교차검증에서 추천된 엑스트라 트리와 스테킹 능형회귀와 능형회귀 앙상블모델 ($ET_S(R)_R$)의 예측값이 관측값의 추세 변화와 일치하며 우수한 결과를 보여주었다. 반면 12시간 후 예측시에는 시계열 분할 5차 교차검증에서 추천한 스테킹 라쏘라즈와 라쏘라즈 앙상블 모델 (S(LL)_LL)이 NSE 0.99, MAE 0.16 m로 가장 좋은 성능을 보여주었으며, 시계열 그래프에서도 태풍기간 동안 가장 유사한 패턴으로 실측 저수위 추세를 따르고 있음을 확인할 수 있다.

[0094] 1~12시간 후 저수위 예측에 있어서 각 시간대별로 다른 알고리즘들로 모델을 생성할 경우 예측되는 값들의 추세가 매끄럽지 못하고 증가와 감소를 반복하는 패턴이 발생할 수가 있다. 이에 본 발명에서는 최대한 전 시간대 모두를 잘 예측할 수 있는 1개의 알고리즘을 정하기 위해 총 4개 기법의 교차검증 방식을 적용하였으며, 태풍 기간과 같은 홍수기에서는 12시간 후 예측에 있어서는 시계열 분할 5차 교차검증 방식의 타당함을 살펴볼 수가 있었다. 즉 일반적으로 머신러닝시 진행되는 샘플들을 섞어서 분할하는 K-Fold 교차검증 방식으로는 우수한 알고리즘을 선별하는 데에는 어려움이 있음을 확인하였고 블로킹 분할이나 시계열 분할 방식 등 시계열 데이터에서 유용하게 활용할 수 있는 교차 검증 방식이 필요함을 확인할 수 있다.

[0095] 자연재해 중 홍수의 경우는 폭우로 유발되는 2차 재해이기에 장마와 태풍 기간 동안의 기상예보에 따른 폭우 정보를 기반으로 홍수 발생을 사전에 예방할 수가 있다. 즉 기상예보를 통한 강수 정보를 제공받아서 저수지의 저수위 정보를 실시간으로 정확히 예측한다면 하류 지역의 침수피해를 예방함으로써 인해 홍수 조절이 가능할 수 있을 것이다. 이에 본 발명에서는 실시간 제공되는 기상청의 강수 기상예보 정보를 바탕으로 학습된 머신러닝 모델을 통해 오봉저수지의 저수위 예측의 정확도를 상승시키는 방안을 살펴보았다.

[0096] 자동화된 머신러닝 도구인 TPOT은 진화적인 알고리즘과 머신러닝 파이프라인의 자동 생성 기능을 결합하여 최적의 머신러닝 모델을 찾아주는 역할을 한다. TPOT은 사용자가 복잡한 머신러닝 파이프라인을 수동으로 설계하고 최적화하는 번거로움을 줄여주며 자동화된 접근 방식을 통해 더 효율적이고 정확한 결과를 얻을 수 있으며, 머신러닝 모델 개발과 발명을 빠르고 효율적으로 진행할 수 있게 해주는 장점이 있다. 이에 본 발명에서는 TPOT이

추천해주는 알고리즘들을 사용하여 1~12시간 후 오봉저수지 저수위를 예측할 수 있는 모델들을 제안하였다.

- [0097] 각 시간 후 저수위를 예측하는 모델들은 시계열 분할 교차검증 기법을 포함한 총 4개 방식의 교차검증 기법들을 사용하여 5번의 학습과 검증을 거쳤으며, 데이터의 시간적 특성을 고려하여 평가한 것으로서, 한층 더 시계열 자료 특성을 잘 평가할 수 있는 모델 선정 과정을 거쳤다고 판단된다. 이렇게 선정된 모델은 실제 테스트 기간 동안에 12시간 후 저수위 예측에 있어서 NSE 0.99 이상의 성능평가를 보여주었고 시계열 변화 패턴이 일치함을 알 수 있었다. 이처럼 TPOT이 추천한 모델이지만 시계열 교차검증 방식들을 통해 선정된 모델을 사용할 시에 정밀한 시계열 변화를 보다 더 일치하게 예측할 수 있는 우수한 모델을 개발할 수 있을 것으로 기대된다.
- [0098] 도 11은 본 발명에서 빅데이터 구축 및 분석 과정을 개념적으로 예시한 것이다.
- [0099] 도 11에서 보는 바와 같이, 본 발명에서 실시간 수위 및 강수량 예측, 기상청 기상 실시간 예보 자료 등 빅데이터 시스템을 보완하고 고도화하며, 기상청의 초단기실황, 초단기예보, 단기예보, 농촌용수종합정보시스템(RAWRIS)의 예측정보 등을 활용한 4지점 하천 수위, 오봉저수지 저수량, 저수위, 저수율을 예측한다.
- [0100] 도 12는 본 발명의 실험에서 하천 수위의 시계열 변화를 도시한 것이다.
- [0101] 도 12를 참조하면, 왕산천, 도마천, 어흘천, 강릉남대천의 4개 측정지점에서 1, 2, 3, 4, 6, 12 시간 후의 하천 수위를 예측한 것이 도시되어 있고, 상기 4개 측정지점에서의 2022.7.1~11.30 하천 수위 시계열 변화가 도시되어 있다. 여기서, 왕산천의 경우 1~12시간 이후 예측 정확도가 NSE, R^2 이 0.81~0.93으로 나타났다.
- [0102] 도 13은 본 발명의 실험에서 하천의 유입량 시계열 변화를 도시한 그래프이다.
- [0103] 도 13을 참조하면, 엑스트라 트리를 적용하여 1~12시간 후 왕산천, 도마천의 하천 수위 예측한 것으로서, 수위-유량 곡선 활용하여 산출한 왕산천과 도마천의 1~12시간 후 유입량 예측 정확도는 NSE, R^2 0.71~0.84로 나타났다.
- [0104] 도 14는 본 발명의 저수지 관리 시스템에서 실시간 유입량, 저수위, 저수량, 저수율 예측 프로세스를 개념적으로 예시한 것이다.
- [0105] 도 14에서 저수지 상류 강수량 자료를 이용하여 시단위 유입량, 저수위, 저수량, 일단위 저수율을 예측할 수 있고, 1, 3, 6, 12 시간 후의 하천 수위, 오봉저수지의 저수위, 저수량 예측을 위한 머신러닝 모델을 구축할 수 있다. 그리고, 예측 시단위 저수율을 유효저수량 대비 현재 예측 저수량의 백분율로 산정할 수 있다. 예를 들어, 머신러닝 모델을 이용하여 오봉저수지에 대해 실시간으로 저수위, 저수량을 예측할 수 있다.
- [0106] 도 15는 본 발명의 실험에서 실시간 저수위 및 저수량 예측을 위한 인공지능 알고리즘의 성능을 나타낸 그래프이다.
- [0107] 도 15를 참조하면, 결정트리, K-최근접 이웃, 서포트 벡터 머신, 랜덤 포레스트, 그래디언트 부스팅, 엑스트림 그래디언트 부스팅, Auto-Sklearn, TPOT 알고리즘의 성능이 표시되어 있고, 이 중에서 NSE와 R^2 1.00인 TPOT의 성능이 가장 우수한 것을 확인할 수 있다. 따라서, 본 발명에서 TPOT를 실시간 저수지 정보 예측 우수 알고리즘으로 선정하고, 정기적으로 저수위 및 저수량의 1, 3, 6, 12시간 후 예측과 내일과 모레 저수율 예측을 수행할 수 있다.
- [0108] 도 16은 본 발명의 실험에서 저수량과 강수량을 도시한 그래프이다.
- [0109] 도 16에서 22.4.1~22.9.8 기간의 저수량과 강수량이 도시되어 있다. 본 발명에서 머신러닝 모형 학습자료의 우수성을 향상시키기 위하여, 측정값 품질관리(Quality Control, QC)를 수행할 수 있다. 예를 들어, 전시간 대비 수위 200%(왕산, 도마, 어흘, 강릉남대)인 경우, 저수위 및 저수량 5% 초과인 경우, 수위 빈도분포 검사 후 극단 이상값 초과시 측정값 품질관리를 적용할 수 있다. 그리고, 자료 스케일링(Scaling), 특성 선택(Feature Selection)에 있어서, TPOT 추천 방식을 적용할 수 있다.
- [0110] 도 17은 본 발명의 실험에서 TPOT 추천 알고리즘을 정리한 도표이다.
- [0111] 도 17에서 예측 항목별 파이프라인 추천 알고리즘이 표시되어 있다. 본 발명에서 시단위 저수위 및 저수량, 일단위 저수율 예측 결과에 있어서, 저수위, 저수량, 저수율 예측 모형 정보에 최근 누적학습자료로 최신 최적화 방안 적용하여 진행할 수 있다. 본 발명에서 AutoML TPOT 추천 알고리즘 및 하이퍼파라미터를 사용할 수 있으며, 추천 알고리즘의 예는 ExtraTreesRegressor, LassoLarsCV, RidgeCV, GradientBoostingRegressor,

LinearSVR 등이다.

- [0112] 도 18은 본 발명의 실험에서 저수량과 강수량을 도시한 그래프이다.
- [0113] 도 18에서 2022년 7월과 8월의 저수량과 강수량을 분석한 결과 그래프이다. 예를 들어, 제공되는 화면의 대시보드에서 '저수량(위) 예측검증' 명령버튼이 클릭되면, 최근 6개월 동안의 실제 저수량과 저수위 측정값과 예측값의 검증결과와 표시된 시계열 변화 그래프가 출력될 수 있다.
- [0114] 도 18에서 1~12시간 후 예측 검증 NSE, R^2 0.997 이상의 우수한 결과를 보이고, 7~8월의 강수일 자료 분석 결과 NSE 0.983 이상의 결과를 보이는 것을 확인할 수 있다.
- [0115] 본 발명에서 12시간 사전 예측을 통한 방류량을 결정하는 기술을 제안한다.
- [0116] 도 19는 저수지에 설치된 수문을 예시한 것이다.
- [0117] 본 발명의 실험에서 오봉저수지에 4개 수문(폭 9.2m * 높이 11m)이 설치되어 있고, 최대방류량은 632 cms이고, 저수지 수문 개도 속도 30cm/min이다. 이처럼 수문을 개폐하는데 시간이 소요되므로, 정확한 유입량(홍수량)을 예측하여 방류량을 사전에 결정할 필요가 있다.
- [0118] 도 20은 시간에 따른 저수지 유입량을 예시한 그래프이다. 본 발명에서 유입량을 측정하기 위해 단위도법, 모델링 기법을 활용할 수 있다.
- [0119] 도 21은 본 발명의 실험에서 시간에 따른 저수율, 강수량 및 게이트 방류량을 도시한 그래프이다.
- [0120] 도 21에서 왕산 AWS 기준 2021년 강수량은 1,096mm 이고, 이는 전년 대비 67% 수준이다. 그리고, 8~10월 강수량이 493.5mm로서 전체 강수량 대비 45%를 차지하며 집중 호우가 발생하는 경향을 나타내고 있다. 이처럼, 극한 호우시 저수율에 따라 저수지 운영에 어려움이 발생할 수 있다. 따라서 12시간 전에 유입량을 예측하여 사전에 수문을 제어하여 방류량을 제어하는 것이 필요하다. 참고로 저수지 운영 관리 수위는 태풍시 80%, 장마기 85%, 동절기(평시) 100%이다.
- [0121] 도 22는 예측 유입량, 예측 저수율, 예측 방류량, 예측 개도 높이의 다양한 시나리오를 예시한 것이다.
- [0122] 도 22에서 보는 바와 같이, 본 발명에서 다양한 시나리오를 분석하여 강수량 및 유입량에 따라 수문의 개도 높이를 결정하여 방류량을 조절할 수 있다. 도 22에서 현재 저수율과 개도 높이별 방류량이 표시되어 있고, 개도 속도를 고려한 유입량 대비 방류량을 예측할 수 있다. 도 23은 방류 시작 6시간 전에 관리자 단말로 알림 메시지를 전송하는 것을 예시한 것이다.
- [0123] 도 24는 본 발명의 실험에서 오봉 저수지의 방류량에 따른 저수위를 도시한 그래프이다.
- [0124] 도 24에서 오봉 저수지의 각 게이트의 오픈 시 방류량에 따른 저수위가 도시되어 있다. 본 발명에서 시뮬레이션을 통한 총 방류량 데이터베이스를 구축하고, 총 방류량을 시계열로 분산하여 순차적으로 진행하는 방식으로 최적 방류량 결정 시나리오를 도출한다.
- [0125] 도 25는 본 발명의 실험에서 저수지 하류부 홍수 위험도를 반영한 시나리오 기반 저수지 운영 방안을 예시한 것이다.
- [0126] 도 25를 참조하면, 오봉댐을 기준으로 상류부에서 HEC-HMS를 이용하여 저수지 유입량을 예측하고, FLDWAV를 이용하여 저수지 방류량을 산정하고, 실측 자료를 이용하여 저수지 방류량을 산정한다.
- [0127] 그리고, 오봉댐을 기준으로 하류부에서 FLDWAV를 이용하여 하도추적을 수행한다. 예를 들어, 본류(강릉남대천)과 지류(어흥천)를 고려한 병렬 하도추적 모델링을 수행한다.
- [0128] 도 26 및 도 27은 본 발명의 실험에서 오봉 저수지 하류부 유입 지류를 고려한 홍수추적 및 분석을 설명하기 위한 도면이다.
- [0129] 도 26에서 홍수위 데이터베이스를 구축하기 위한 지점을 선정하되, 유량변화가 크게 일어나는 지류의 합류점, 주거시설, 상업시설 밀집지역 인근 하천단면으로 선정한다(예를 들어, 강릉 남대천 12개). 예를 들어, 강릉남대천 하천기본계획을 바탕으로 주요지점을 1차 선정하고, 수치해석결과 실제 월류발생 위험구역을 추가하여 선정할 수 있다.
- [0130] 도 27을 참조하면, 주요 유입지류를 고려한 하도망 구축 및 병렬 홍수추적 모의를 위한 지형단면을 구축하고,

하도추적을 위한 상하류 경계조건을 구축할 수 있다. 도 27의 예시에서 강릉 남대천 하천 네트워크 구성이 도시되어 있고, 지류단면이 표시되어 있다. 상류 경계조건으로 오봉저수지 유입량, 어흘천 유입량이 도시되어 있고 하류 경계조건이 도시되어 있다.

- [0131] 도 28은 본 발명의 실험에서 FLDWAV 모델의 신뢰성을 검토하기 위해 예시한 도면이다.
- [0132] 도 28은 2019 태풍 '미탁' 당시 회산교 홍수위와 비교한 것으로서, 2019년 홍수피해 상황조사보고서(환경부)에 강릉 남대천 회산교 지점의 홍수위와 비교한 것이다. 여기서 강릉 남대천 회산교 최고 관측 수위는 2.23m이고, 최고 계산 수위는 2.31m으로서, 오차율은 3.46%이다.
- [0133] 도 29는 본 발명의 실험에서 방류 시나리오에 따른 각 지점별 홍수위 결과를 도시한 것이다. 도 29에서 500년 빈도(380nm-19hr) 규모 총 방류 시나리오이고, 게이트는 완전 개방(full opening)된 상태이다.
- [0134] 도 30은 본 발명의 실험에서 방류 시나리오에 따른 각 지점별 홍수위 결과를 정리한 도표이다. 도 30에서 500년 빈도(380nm-19hr) 규모 총 방류 시나리오이고, 게이트는 완전 개방(full opening)된 상태이다.
- [0135] 도 30에서 최대 방류량은 1,800 CMS 규모이고, 39개 하천 단면 구간에서 범람이 발생하였다. 이렇게 500년 빈도 규모로 지점별 시계열 수위, 월류수심, 유량 등의 데이터를 포함하는 데이터베이스를 구축할 수 있다.
- [0136] 도 31은 본 발명의 실험에서 시나리오에 따른 홍수 범람 예상도이다.
- [0137] 도 31은 500년 빈도 시나리오를 적용하고, 하도 버퍼링 기법을 적용한 홍수범람 결과로서, 월류 24개 지점을 확인할 수 있다.
- [0138] 도 32는 본 발명의 실험에서 각 시나리오에 따른 방류량을 도시한 그래프이다. 도 32는 극한 강우 사상의 조건이므로 수문은 4련 모두 개방을 원칙으로 설정된 것이다.
- [0139] 도 32는 500년 빈도 유입량에 따른 방류 시나리오 착오범을 통한 적정범위 산출 후 최종 8개 방류 시나리오가 예시되어 있다.
- [0140] 도 33은 본 발명의 실험에서 시나리오에 따른 방류 결과를 보여주는 것이다.
- [0141] 도 33에서 S1은 댐이 없는 경우로서 유입이 그대로 방출되는 경우이고, S2은 기존 저수지 운영 시나리오에 따른 결과이고, S3은 본 발명의 최적 시나리오에 따라 12시간 사전 예측에 따른 사전 방류 결과이다. 본 발명에서 8개 시나리오 중 최적의 시나리오를 도출하여 저수지 안정성, 저수율을 확보하고 하류부 홍수피해를 고려하였고, 수문 4련에서 6.5m를 개도하는 방식으로 동시 개방하였다.
- [0142] 도 33에서 보는 바와 같이, S1의 경우 39 지점, S2의 경우 24 지점, S3의 경우 12 지점이 월류된 것을 확인할 수 있다.
- [0143] 도 34는 본 발명의 실험에서 방류 시나리오에 따른 하류 범람지점 예측 과정을 도시한 것이다.
- [0144] 도 34의 방류 시나리오에서 기준방류량은 100년 빈도이고, 3번은 200년 빈도이고, 6번은 500년 빈도이고, 12번은 PMF 방류시나리오별 하도 추적과 홍수 범람 지점을 예측한 것이다.
- [0145] 도 34를 참조하면, 하도 추적 결과를 반영한 기준 방류량을 산정하고(시행착법), 점두 방류량을 포함하는 방류 시나리오를 생성하고, 방류량 시나리오를 산정하고, 방류 시나리오별 홍수 범람 예상도를 작성하고, 방류 시나리오별 하도 추적을 수행한다.
- [0146] 도 35는 방류 시나리오에 따른 하류 범람지점 개수를 나타낸 도표이다.
- [0147] 도 35를 참조하면, 하도 안정성을 고려한 기준 방류량(Scenario 2 : 1,161.5 cms)을 통해 21개 방류량 시나리오를 생성한다.
- [0148] 도 36은 본 발명의 실험에서 방류 시나리오에 따른 하류 범람지점 예측 과정을 예시한 것이다.
- [0149] 도 36의 예시에서 평시 지도 화면과 PMF 시나리오에 따른 지도 화면이 예시되어 있다. PMF 시나리오에 따른 지도 화면에서 홍수위에 따른 마커 색상이 변경되고, 홍수위에 해당하면 고수부지 홍수 예보와 홍수 알림 문자 메시지를 단말에 전송한다.
- [0150] 도 37 내지 39는 본 발명에서 관리자 단말에서 디스플레이되는 화면예를 도시한 것이다.
- [0151] 도 37에서 지도에서 한 지점을 클릭하면 해당 지점의 예측 홍수위, 실측 및 예측 홍수위 관련 상세 화면이 표시

된다.

[0152] 도 38에서 현재 저수율과 유입 홍수 수문곡선을 고려한 사전 방류량을 결정할 수 있다.

[0153] 도 39에서 오봉 저수지 운영 관리 시스템의 관리자 단말 화면예를 예시한 것으로서, 화면 좌측에 실시간 저수지 현황 정보, 일평균 저수율 정보가 표시되고, 화면 중앙 상단에 오봉 저수지를 촬영하고 있는 CCTV 화면이 표시되고, 화면 중앙 하단에 상,하류 하천 수위 현황, 실시간 시간당 강우량, 날씨 정보가 표시되어 있다.

[0154] 본 발명의 일 실시예에서 강우(시간당 1mm 이상)와 비강우(시간당 1mm 미만)로 데이터의 샘플들을 분할하여, 각 시간대별 2개의 모델들을 생성한 후, 예측 시점에 강우 발생 유무에 따른 구분된 모델을 적용하는 하이브리드 방식의 자료별 분리형 알고리즘을 선정한 실시간 시간위 저수위 예측 모델을 제안한다. 이렇게 함으로써, 강수 발생일 6, 12 시간 후 저수위 예측 정확도가 낮아지는 것을 개선할 수 있다.

[0155] 도 40은 본 발명에서 TPOT 추천 알고리즘의 성능을 비교한 그래프이다.

[0156] 도 40은 강우시와 비강우시로 자료를 분리한 자료별 분리형 알고리즘으로서, 비강우시는 SelectPercentile+ElasticNetCV (sp_E)로서, 12시간 후 예측 5차 교차검증 평균이 NSE, R^2 0.99, MAE 0.55으로 나타났다. 강우시는 SelectFwe+LassoLarsCV (sf_LL)로서 12시간 후 예측 5차 교차검증 평균이 NSE 0.96, R^2 0.97, MAE 0.55으로 나타났다.

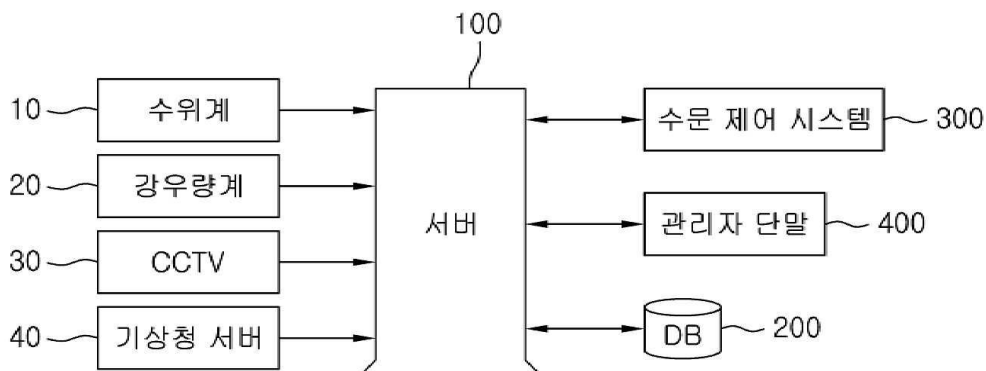
[0157] 이상 본 발명을 몇 가지 바람직한 실시 예를 사용하여 설명하였으나, 이들 실시 예는 예시적인 것이며 한정적인 것이 아니다. 본 발명이 속하는 기술분야에서 통상의 지식을 지닌 자라면 본 발명의 사상과 첨부된 특허청구 범위에 제시된 권리범위에서 벗어나지 않으면서 다양한 변화와 수정을 가할 수 있음을 이해할 것이다.

부호의 설명

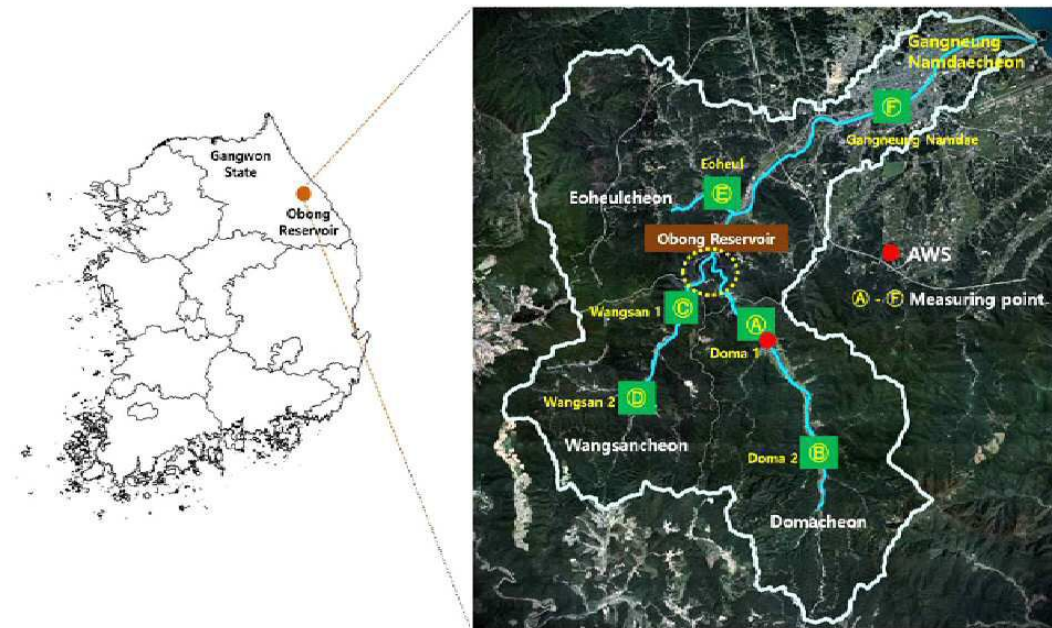
[0159] 100 서버
200 데이터베이스
300 수문 제어 시스템 400 관리자 단말

도면

도면1



도면2



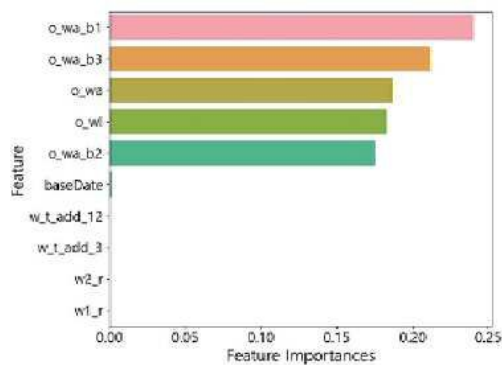
도면3

Prediction target	Notation	Selected model
Reservoir water level after 1 hour	LL	LassoLarsCV(normalize=True)
Reservoir water level after 2 hours	m_R	make_pipeline(MinMaxScaler(),RidgeCV())
Reservoir water level after 3 hours	LL	LassoLarsCV(normalize=True)
Reservoir water level after 4 hours	s_R	make_pipeline(StandardScaler(),RidgeCV())
Reservoir water level after 5 hours	o_LL	make_pipeline(OneHotEncoder(minimum_fraction=0.1, sparse=False, threshold=10), LassoLarsCV(normalize=True))
Reservoir water level after 6 hours	LL	LassoLarsCV(normalize=True)
Reservoir water level after 7 hours	ET_S(R)_R	make_pipeline(SelectFromModel(estimator=ExtraTreesRegressor(max_features=0.8500000000000001, n_estimators=100), threshold=0.1),StackingEstimator(estimator=RidgeCV(), RidgeCV()))
Reservoir water level after 8 hours	S(LL)_LL	make_pipeline(VarianceThreshold(threshold=0.2),MaxAbsScaler(),StackingEstimator(estimator=LassoLarsCV(normalize=True)), LassoLarsCV(normalize=False))
Reservoir water level after 9 hours	r_LL	make_pipeline(RobustScaler(), LassoLarsCV(normalize=False))
Reservoir water level after 10 hours	ET_S(R)_R	make_pipeline(SelectFromModel(estimator=ExtraTreesRegressor(max_features=0.8500000000000001, n_estimators=100), threshold=0.1),RidgeCV())
Reservoir water level after 11 hours	LL	LassoLarsCV(normalize=True)
Reservoir water level after 12 hours	z_LL	make_pipeline(ZeroCount(), LassoLarsCV(normalize=True))

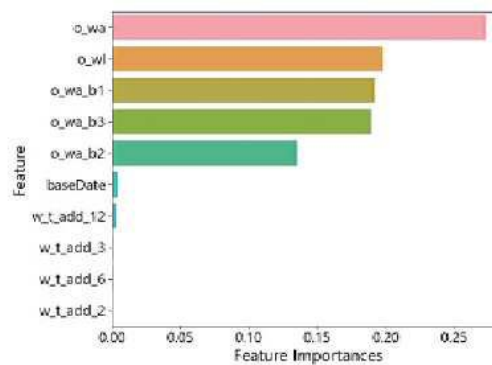
도면4

Feature name	Mean	Std	Min	25%	50%	75%	Max	Skew	Kurtosis
baseDate	190.54	100.02	1	112	189	273	365	-0.08	-1.11
baseTime	1161.1	689.19	0	600	1200	1800	2400	-0.03	-1.19
w_obs	0.22	1.71	0	0	0	0	50.56	16.06	335.48
w_t_add_1	0.26	1.89	0	0	0	0	55.13	16.44	342.29
w_t_add_2	0.27	2.15	0	0	0	0	59	16.25	324.76
w_t_add_3	0.29	2.44	0	0	0	0	88.51	17.47	384.01
w_t_add_4	0.3	2.53	0	0	0	0	63.26	15.94	295.7
w1_r	0.19	1.61	0	0	0	0	50.56	17.81	408.92
w2_r	0.16	1.55	0	0	0	0	50.56	19.42	471.07
d1_r	0.22	1.64	0	0	0	0	50.56	17.18	384.87
d2_r	0.28	1.7	0	0	0	0	50.56	15.82	332.66
w_l	98.87	17.7	58.04	88.22	95.85	104.26	330.63	3.48	23.77
d_l	28.35	12.49	0.19	21.88	26.51	31.8	176.68	3.05	19.27
a_l	49.5	20.87	3.73	39.35	45.54	56.49	332.18	2.95	20.33
k_l	70.59	37.6	0.02	42.75	84.86	98.12	238.57	-0.64	-0.53
o_wa	10,450.0	2,643.3	4,283.4	8,306.6	11,367.3	12,572.4	16,086.5	-0.67	-0.81
o_wl	116.71	3.66	106.63	114.39	118.13	119.55	123.36	-0.92	-0.26
o_wa_b1	10,448.4	2,641.9	4,283.4	8,299.4	11,367.3	12,570.9	15,783.3	-0.67	-0.82
o_wa_b2	10,446.9	2,640.5	4,283.4	8,293.7	11,367.3	12,568.1	15,606.9	-0.67	-0.82
o_wa_b3	10,445.4	2,639.3	4,283.4	8,293.7	11,367.3	12,568.1	15,352	-0.67	-0.82
w_t_add_6	1.88	13.74	0	0	0	0	316.9	15.12	265.83
w_t_add_12	3.83	26.37	0	0	0	0	511.31	13.25	193.7

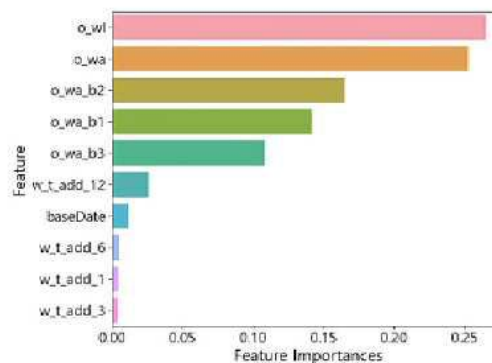
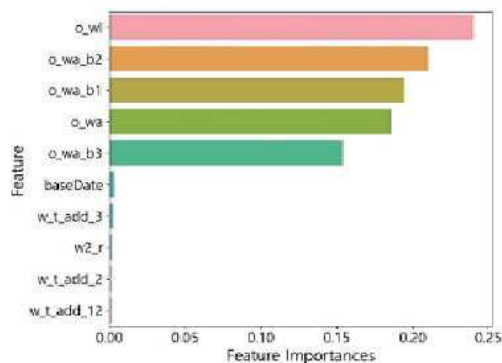
도면5



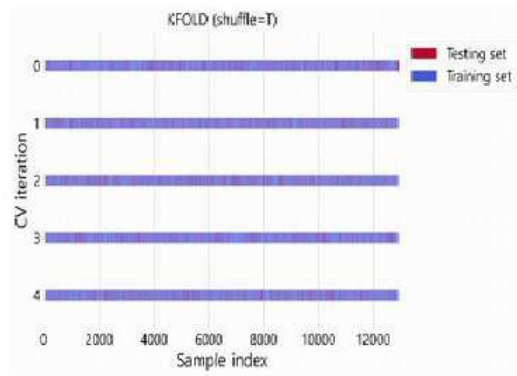
(a) After 1 hour



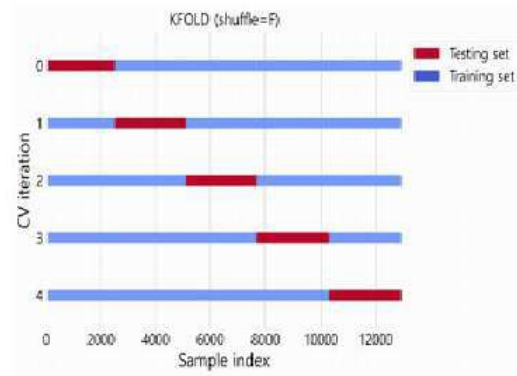
(b) After 12 hours



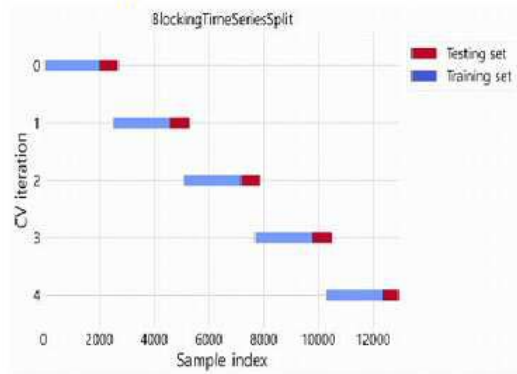
도면6



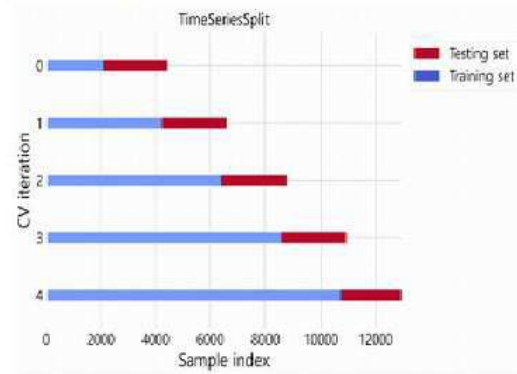
(a) KFOLD (shuffle=T)



(b) KFOLD (shuffle=F)

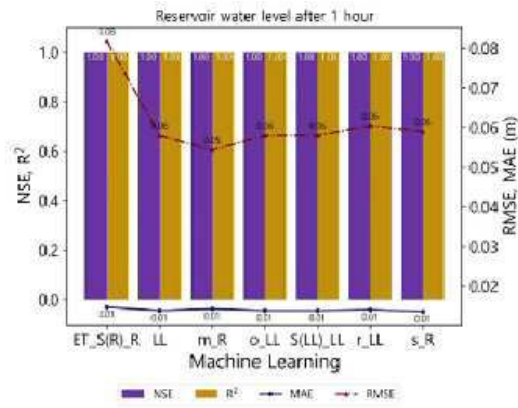


(c) BlockingTimeSeriesSplit

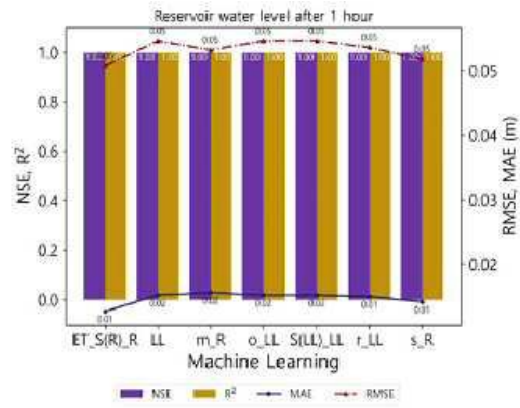


(d) TimeSeriesSplit

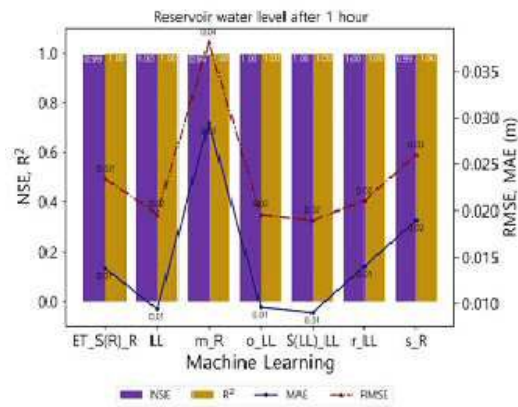
도면7



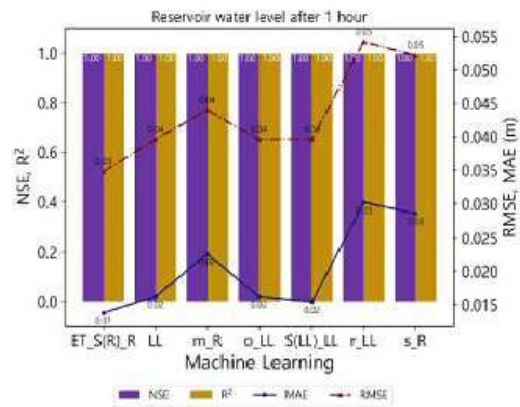
(a) KFOLD (shuffle=T)



(b) KFOLD (shuffle=F)

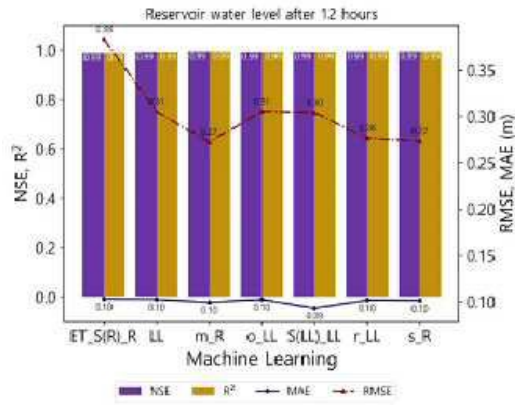


(c) BlockingTimeSeriesSplit

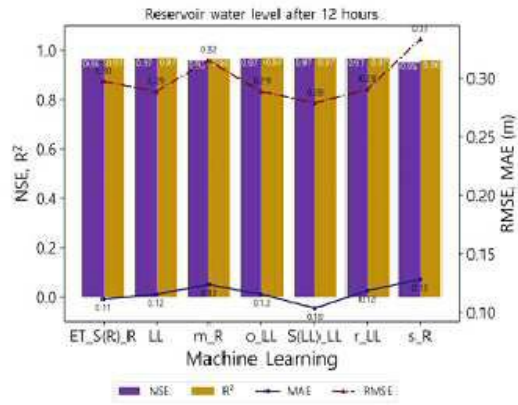


(d) TimeSeriesSplit

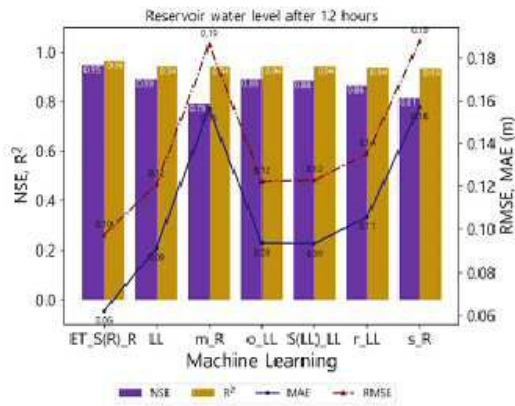
도면8



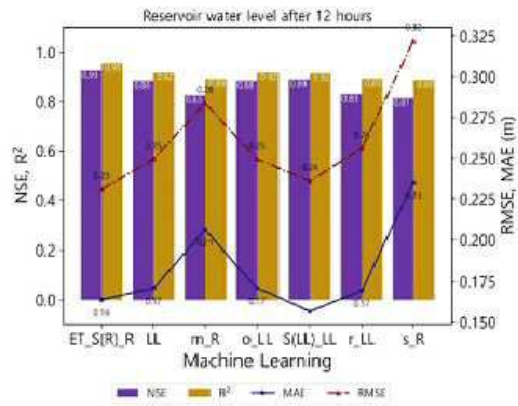
(a) KFOLD (shuffle=T)



(b) KFOLD (shuffle=F)

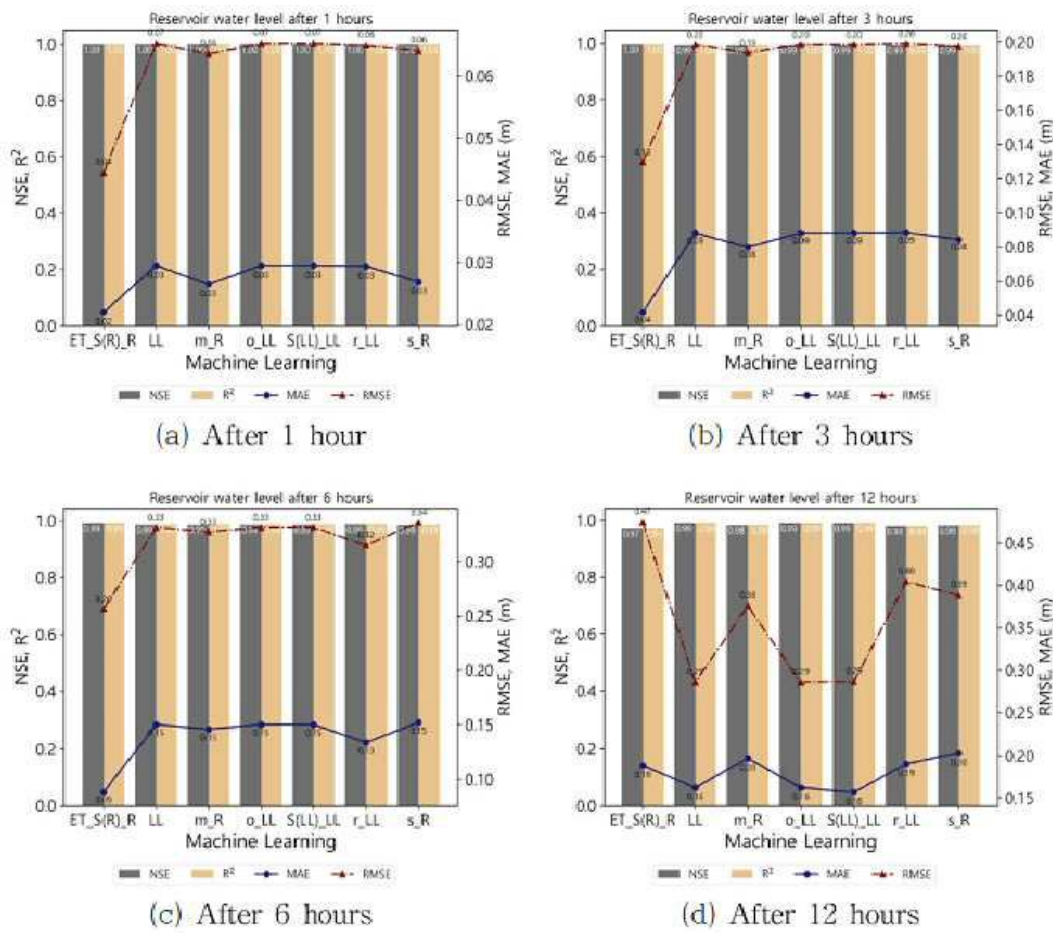


(c) BlockingTimeSeriesSplit

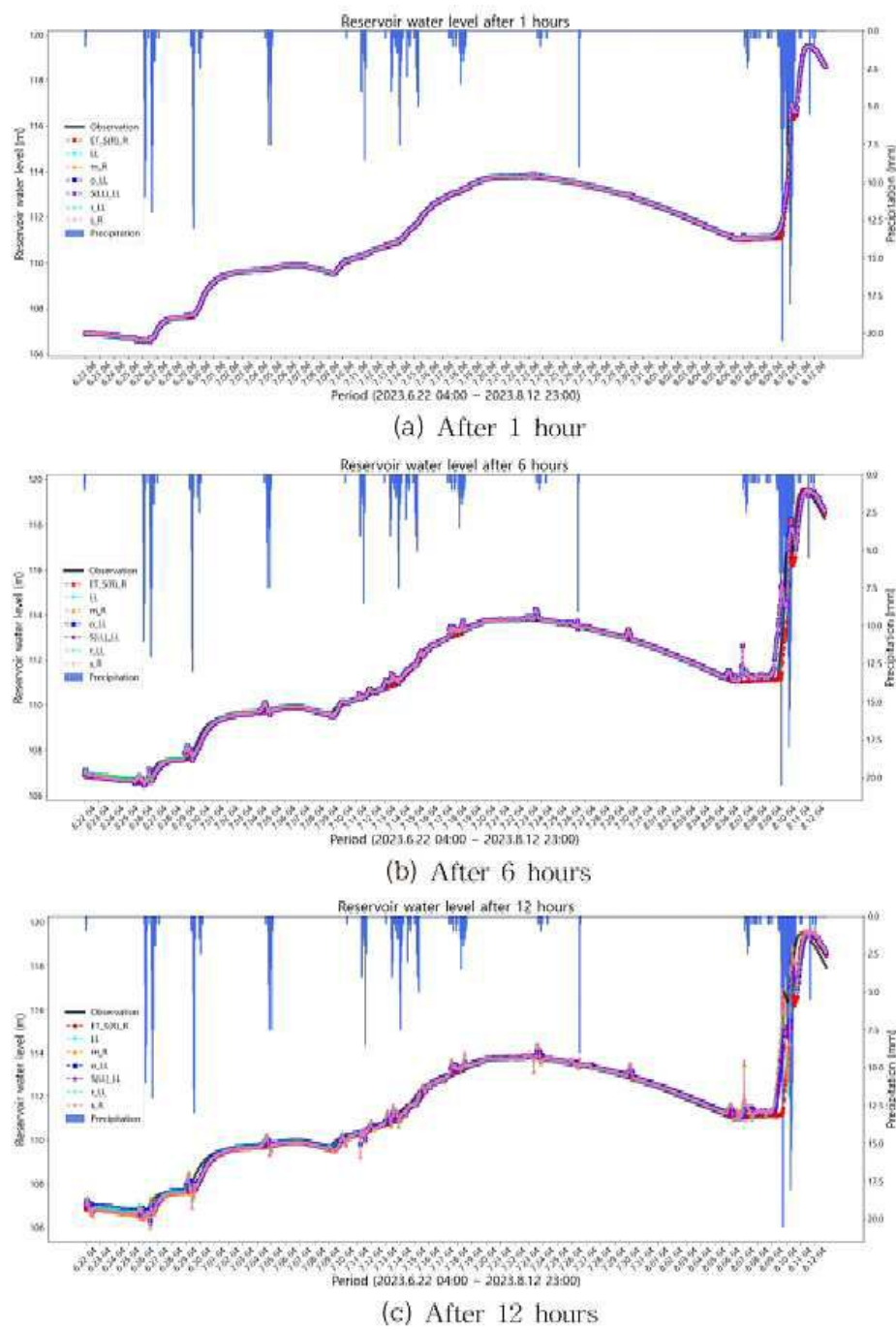


(d) TimeSeriesSplit

도면9



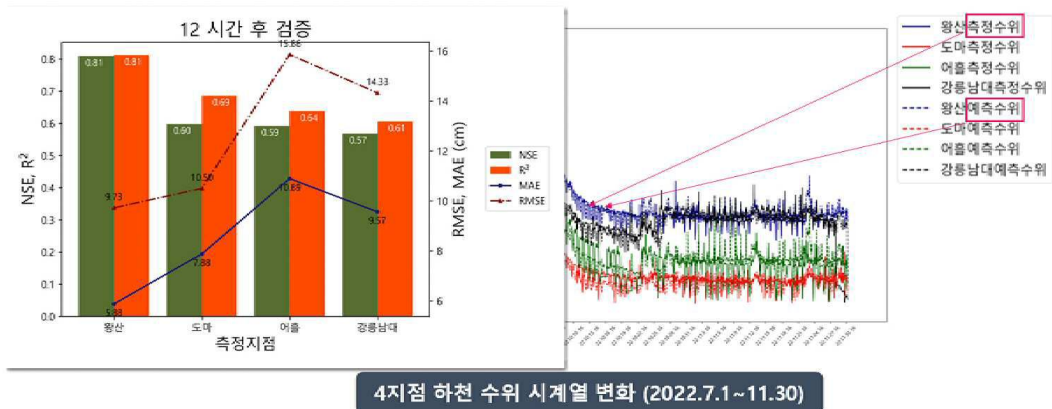
도면10



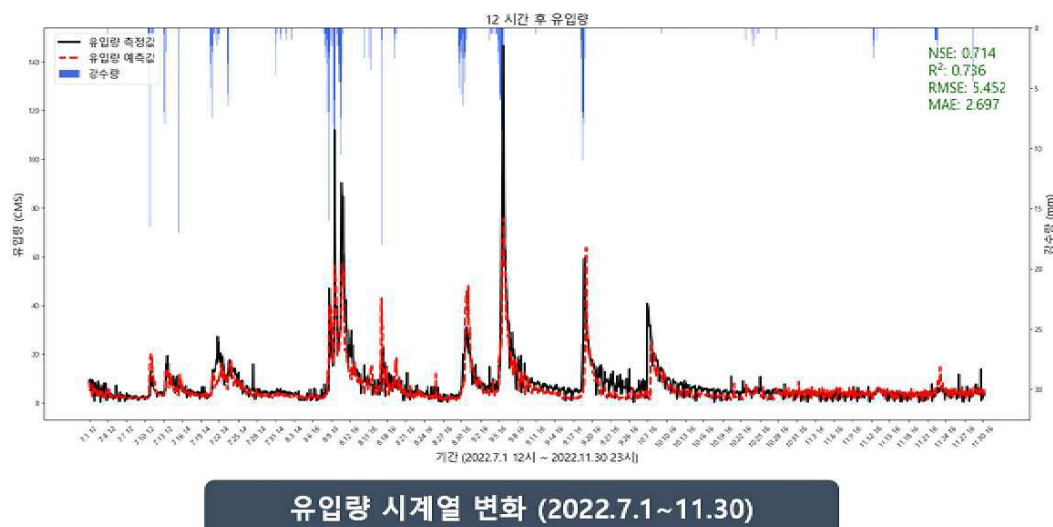
도면11



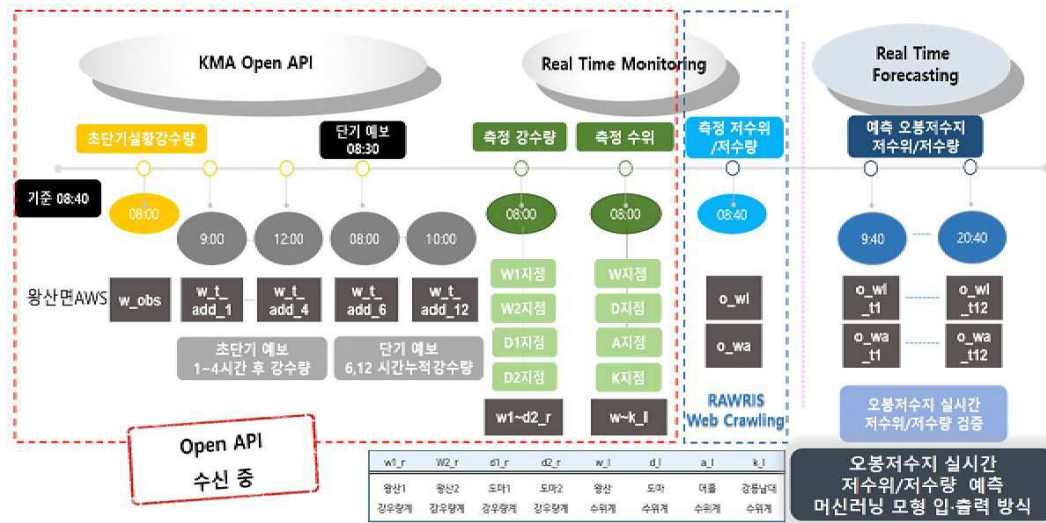
도면12



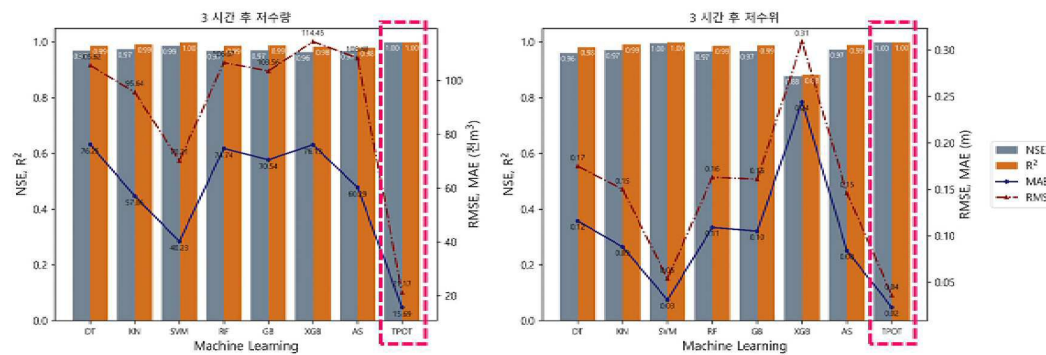
도면13



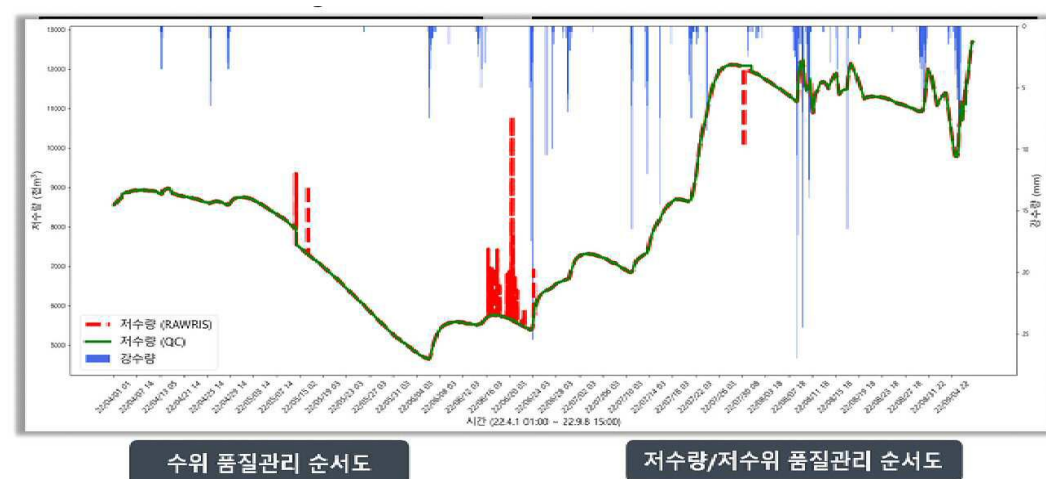
도면14



도면15



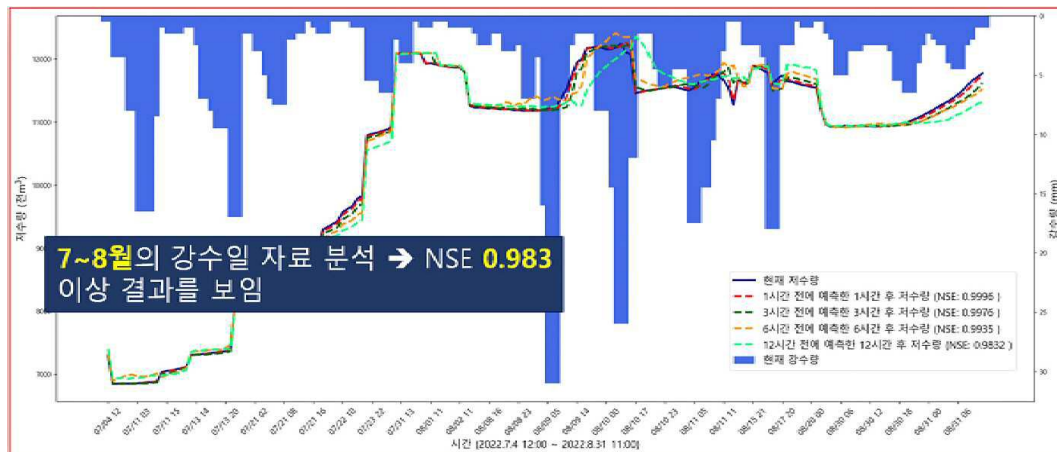
도면16



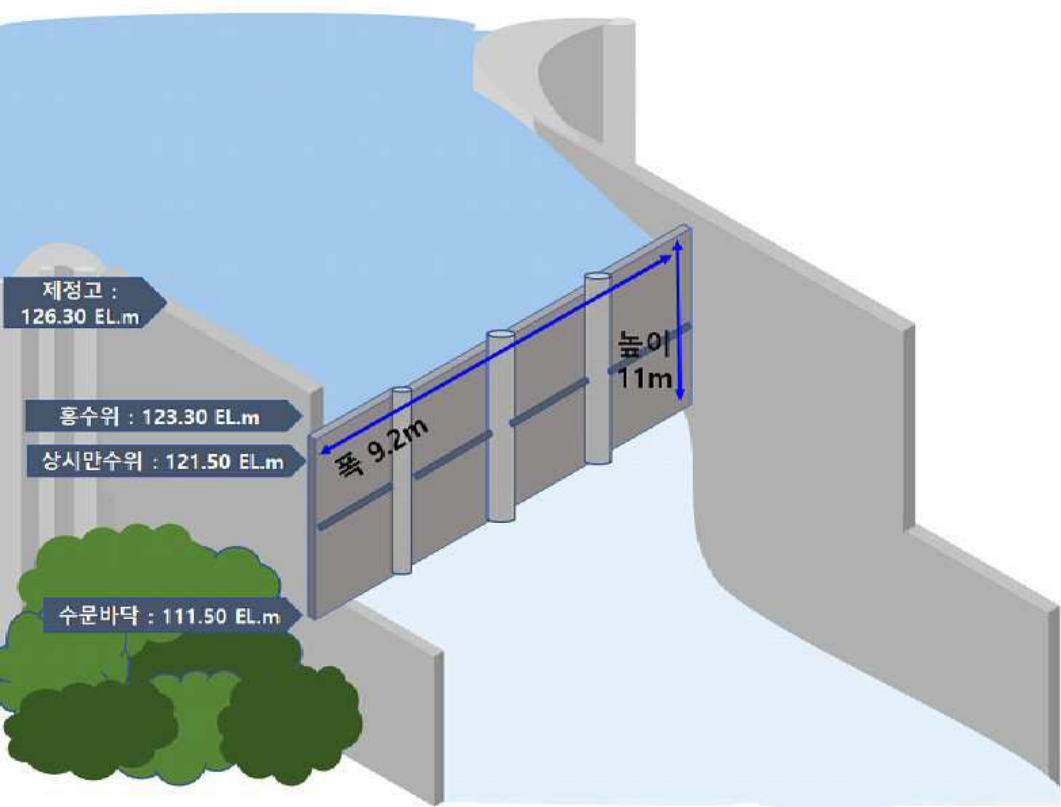
도면17

예측 항목	파이프라인 추천 알고리즘
1시간 후 저수량	<code>make_pipeline(StackingEstimator(estimator=ExtraTreesRegressor(bootstrap=True, max_features=0.1, min_samples_leaf=9, min_samples_split=14, n_estimators=100)),LassoLarsCV(normalize=False))</code>
3시간 후 저수량	<code>LassoLarsCV(normalize=False)</code>
6시간 후 저수량	<code>make_pipeline(FeatureAgglomeration(affinity='euclidean', linkage='ward'),RidgeCV())</code>
12시간 후 저수량	<code>LinearSVR(C=1.0, dual=True, epsilon=0.01, loss='squared_epsilon_insensitive', tol=0.1)</code>
1시간 후 저수위	<code>make_pipeline(StackingEstimator(estimator=GradientBoostingRegressor(alpha=0.85, learning_rate=0.01, loss='lad', max_depth=2, max_features=1.0, min_samples_leaf=6, min_samples_split=19, n_estimators=100, subsample=0.3)),LassoLarsCV(normalize=True))</code>
3시간 후 저수위	<code>make_pipeline(StackingEstimator(estimator=GradientBoostingRegressor(alpha=0.85, learning_rate=0.01, loss='huber', max_depth=2, max_features=1.0, min_samples_leaf=6, min_samples_split=19, n_estimators=100, subsample=0.3)),LassoLarsCV(normalize=True))</code>
6시간 후 저수위	<code>make_pipeline(SelectPercentile(score_func=f_regression, percentile=4),RidgeCV())</code>
12시간 후 저수위	<code>make_pipeline(SelectPercentile(score_func=f_regression, percentile=54),StackingEstimator(estimator=ElasticNetCV(l1_ratio=0.35000000000000003, tol=0.01),RidgeCV())</code>
내일 저수율	<code>make_pipeline(SelectFwe(score_func=f_regression, alpha=0.046),RidgeCV())</code>
모레 저수율	<code>LinearSVR(C=0.1, dual=True, epsilon=0.1, loss='epsilon_insensitive', tol=0.1)</code>

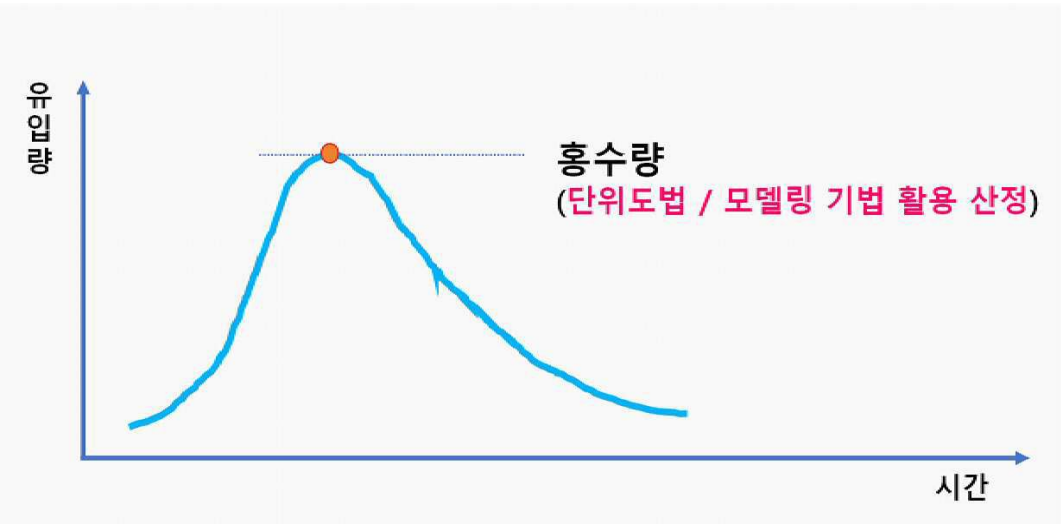
도면18



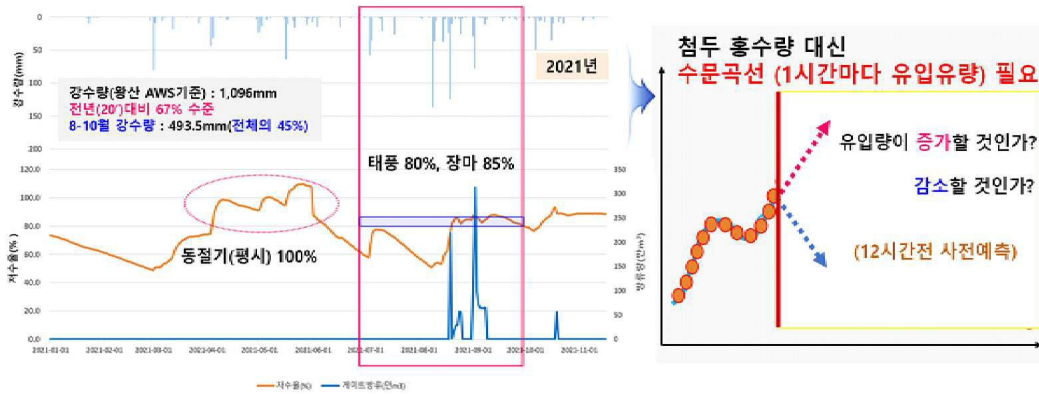
도면19



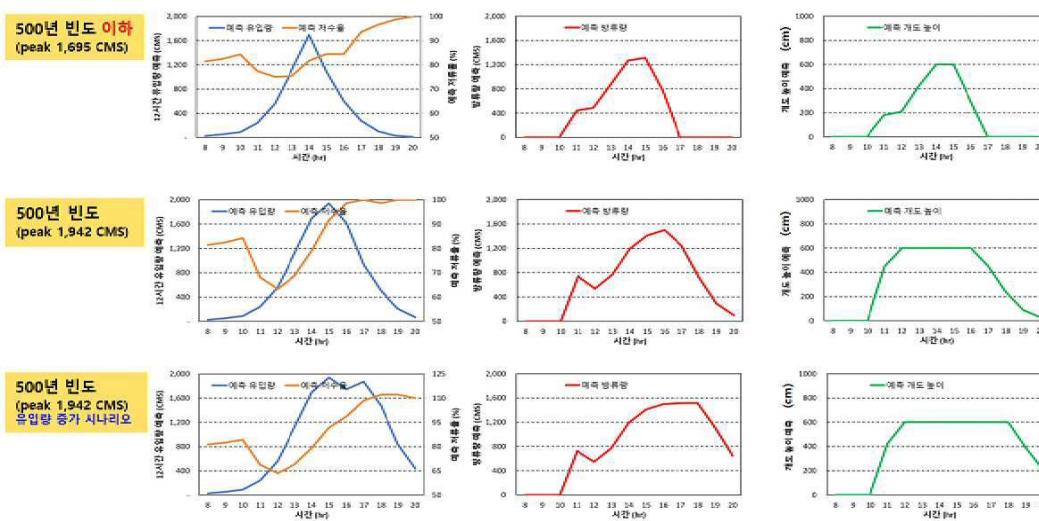
도면20



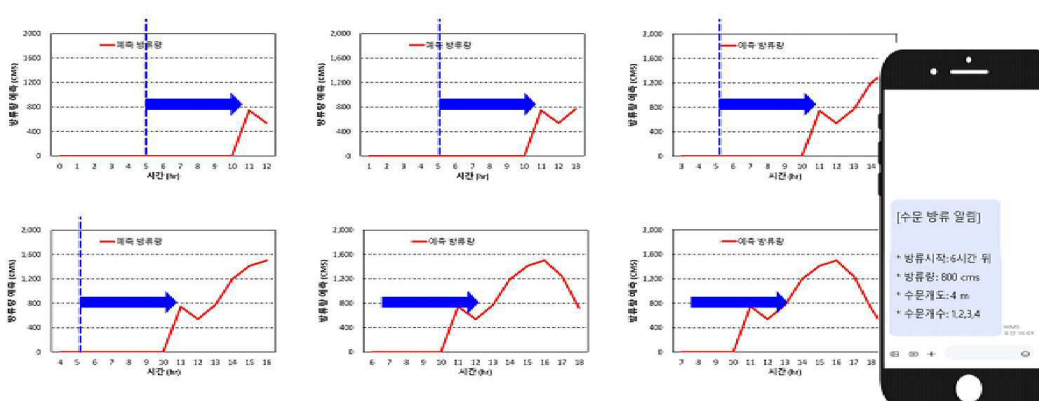
도면21



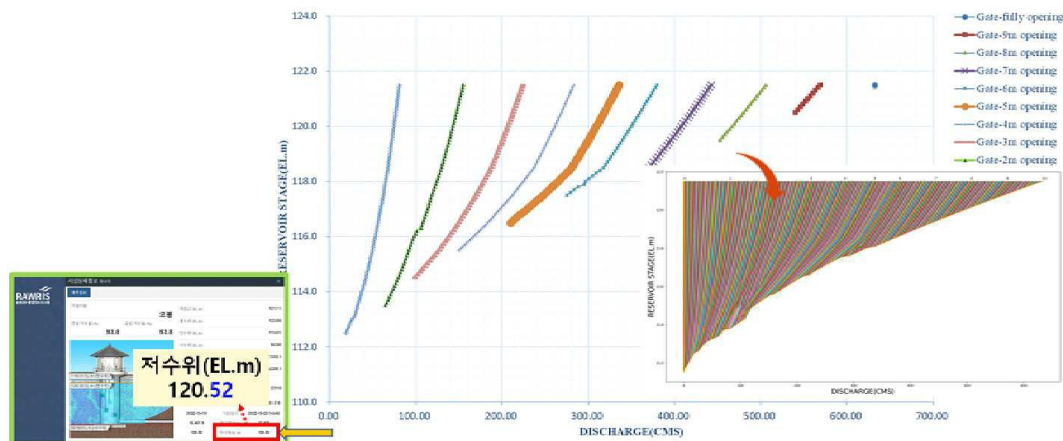
도면22



도면23



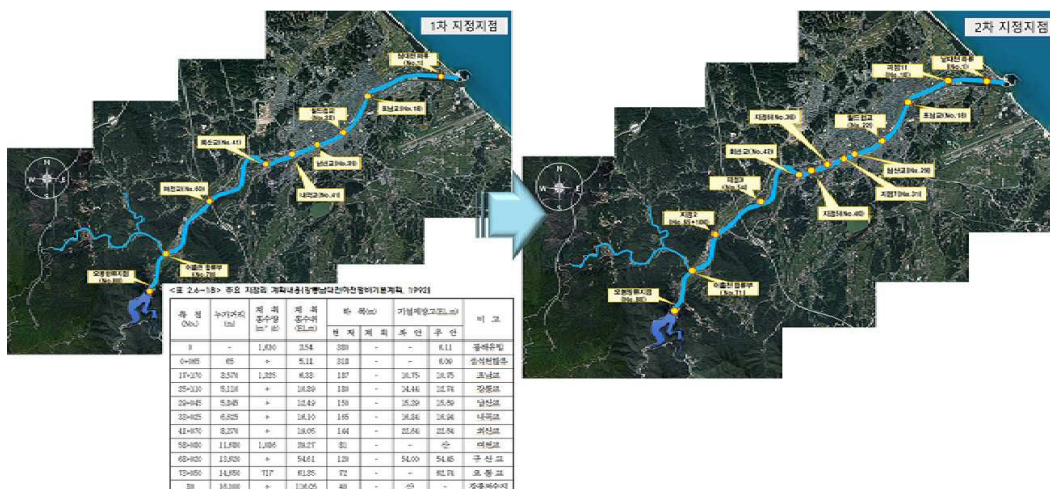
도면24



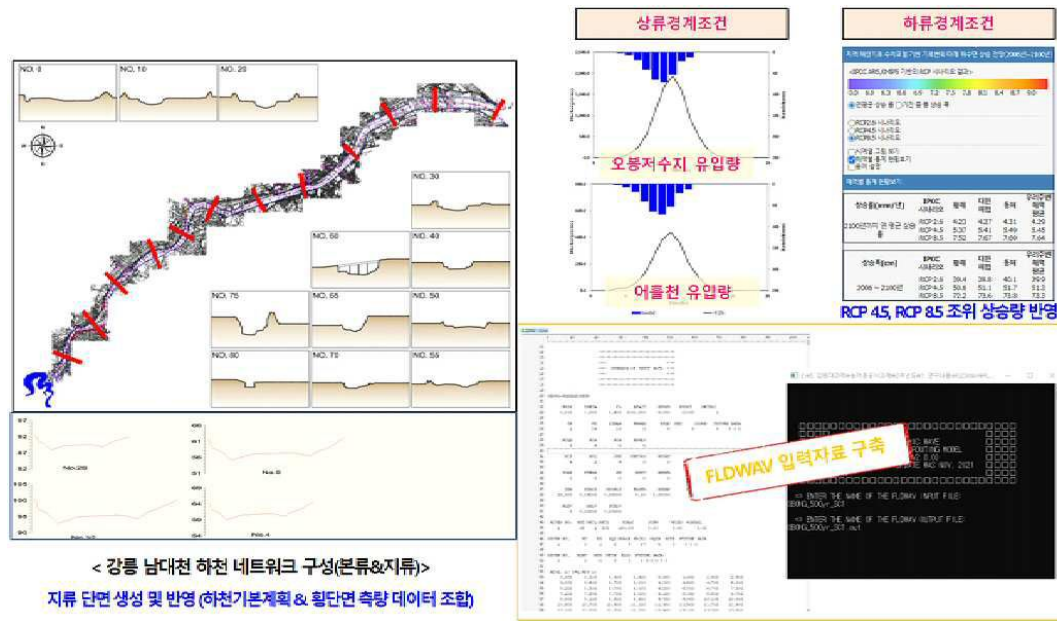
도면 25



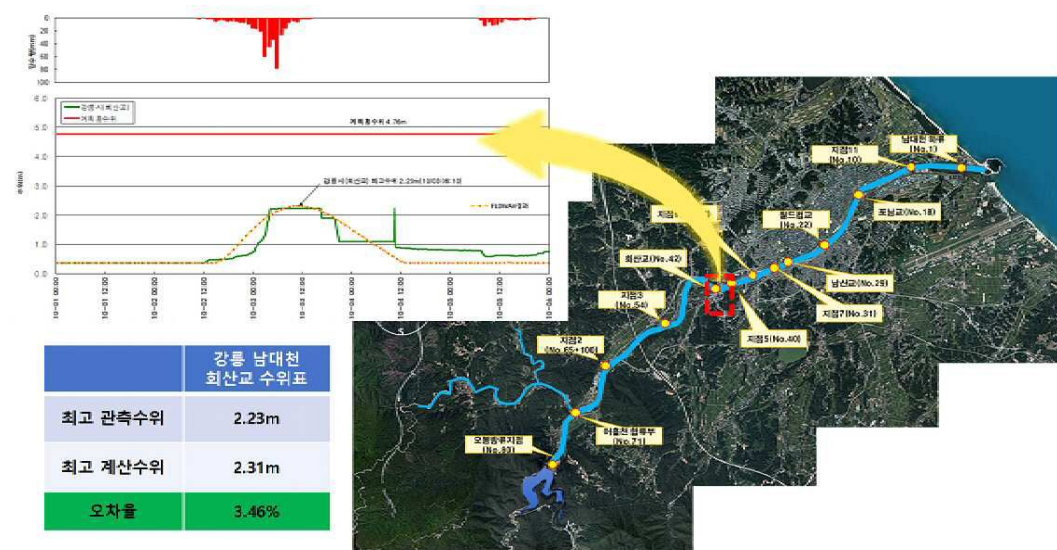
도면 26



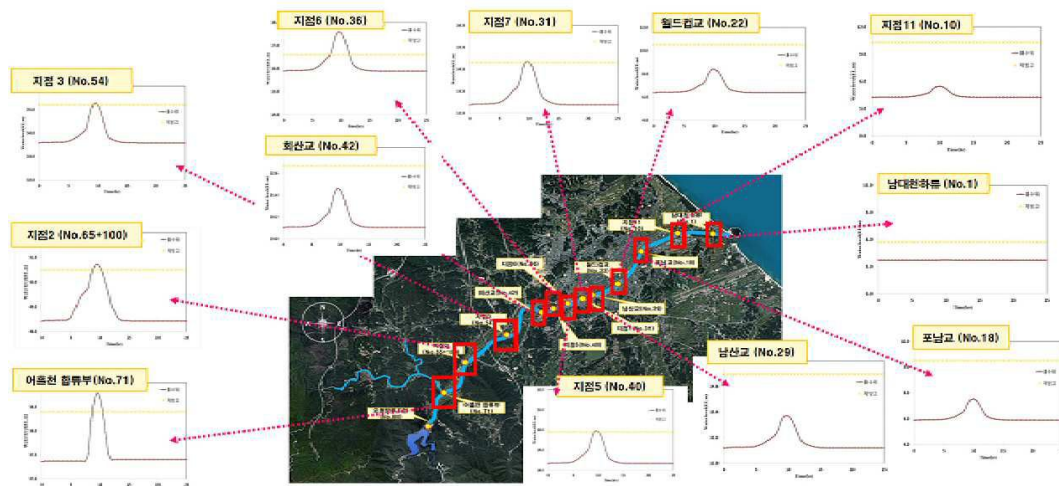
도면27



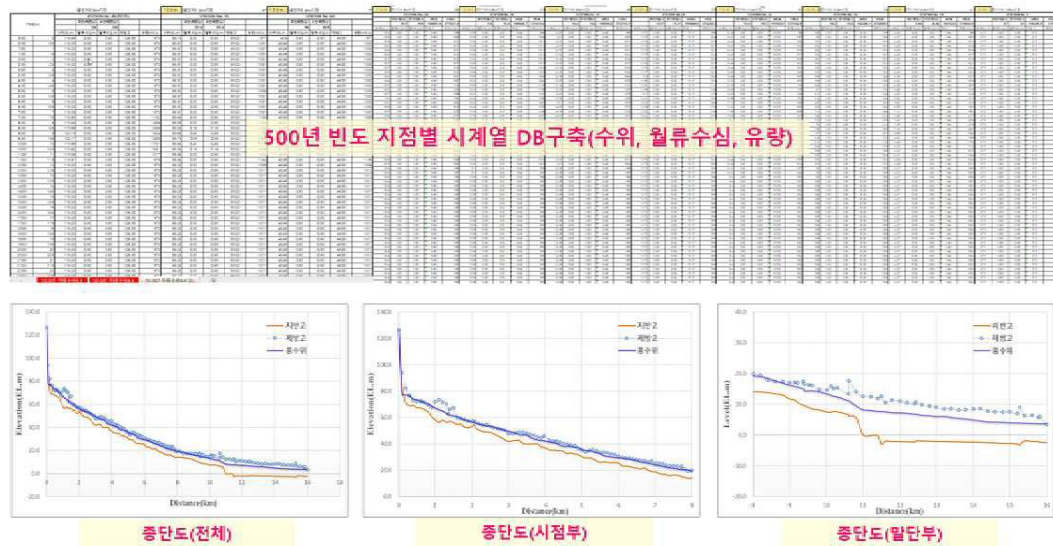
도면28



도면29



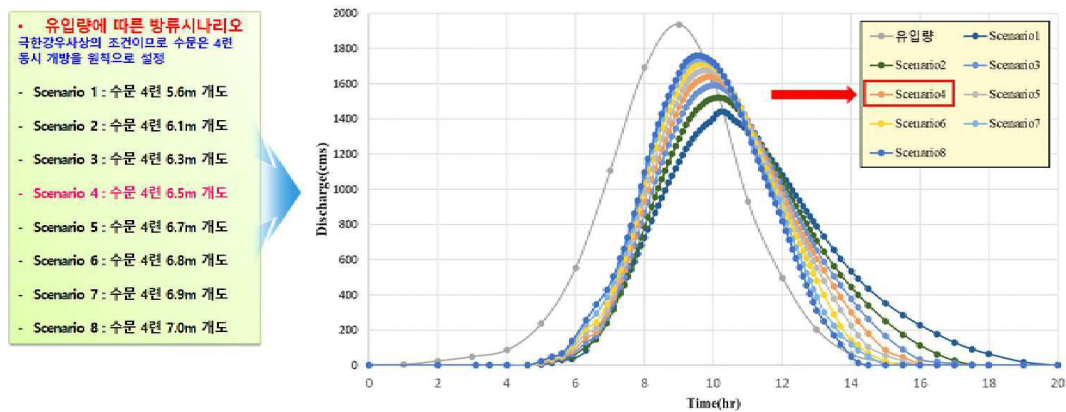
도면30



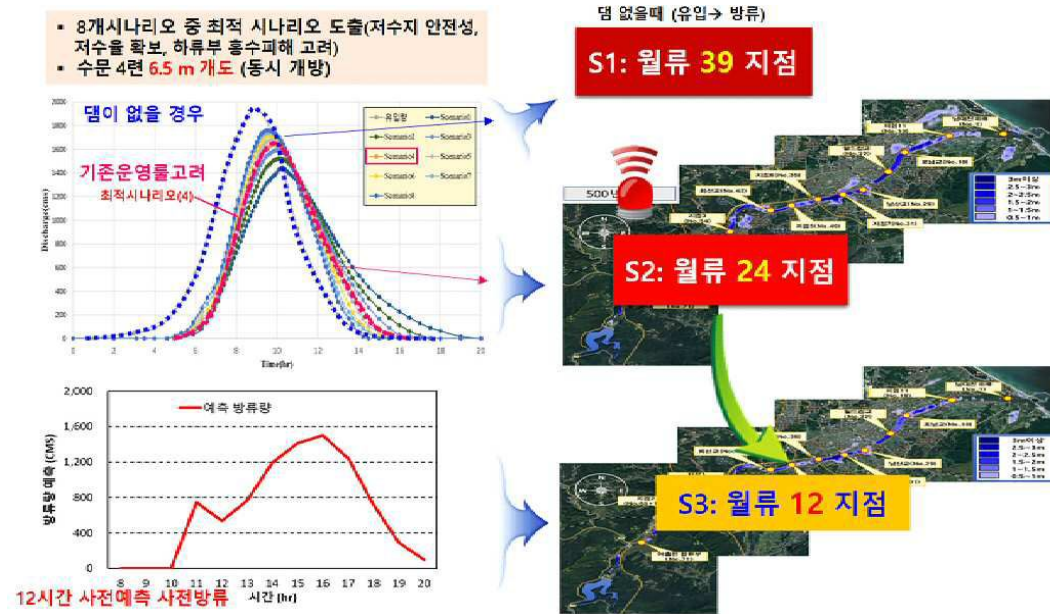
도면31



도면32



도면33



도면34



도면35

										점수별 점수 감소					예) 500년 빈도 우발량 발생									
범류구분	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12			
모형+대물평균점수(점수)	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4			

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8	7	6	5	4	3	2	1	기초점점수	1	2	3	4	5	6	7	8	9	10	11	12					
모형	30%	50%	60%	70%	80%	85%	90%	95%	100%	105%	110%	115%	120%	130%	140%	150%	160%	180%	200%	220%	231.2%					
대물	348.5	580.8	696.3	813.1	912.2	987.3	1,045.4	1,103.4	1,161.5	1,219.6	1,277.7	1,335.7	1,393.8	1,510.0	1,626.1	1,742.3	1,858.4	2,090.7	2,323.0	2,555.3	2,885.4					

범류

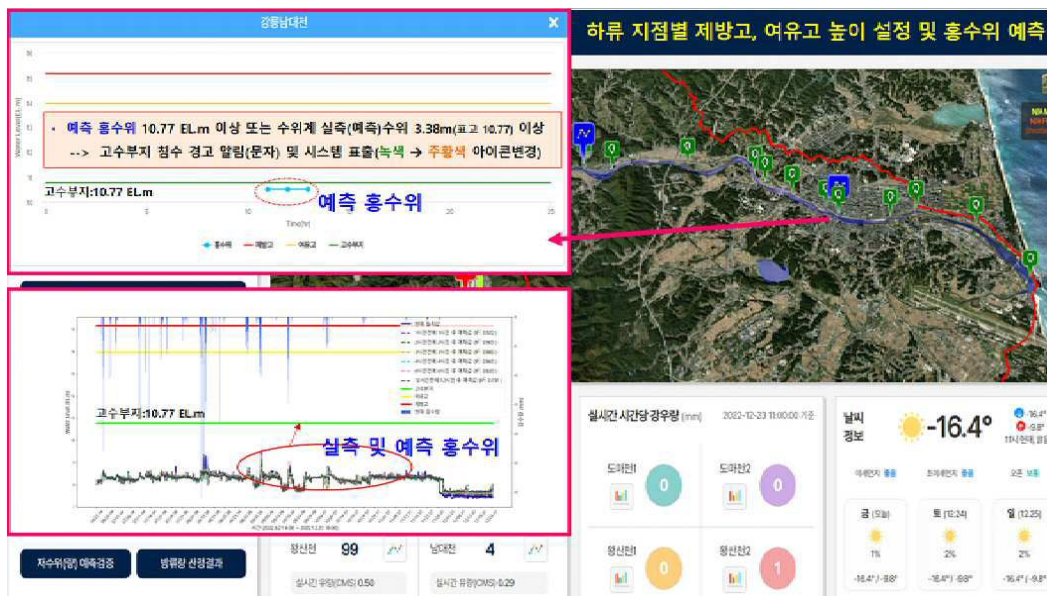
시나리오

										100년 빈도					200년 빈도					300년 빈도					PMF	
범류	8</																									

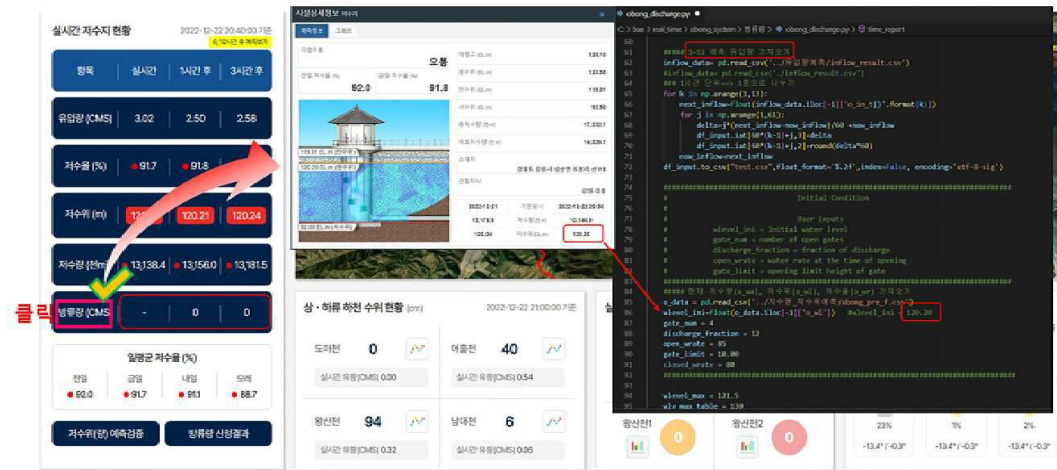
도면36



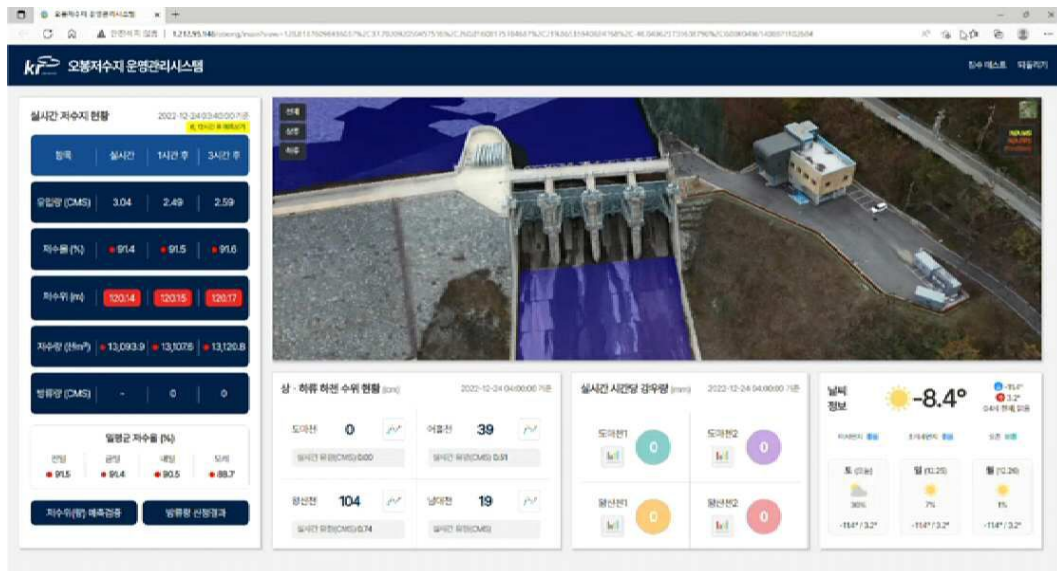
도면37



도면38



도면39



도면40

