

# 2021 빅콘테스트 대회 설명

데이터 분석분야 퓨처스 리그

2021. 7.



*01*

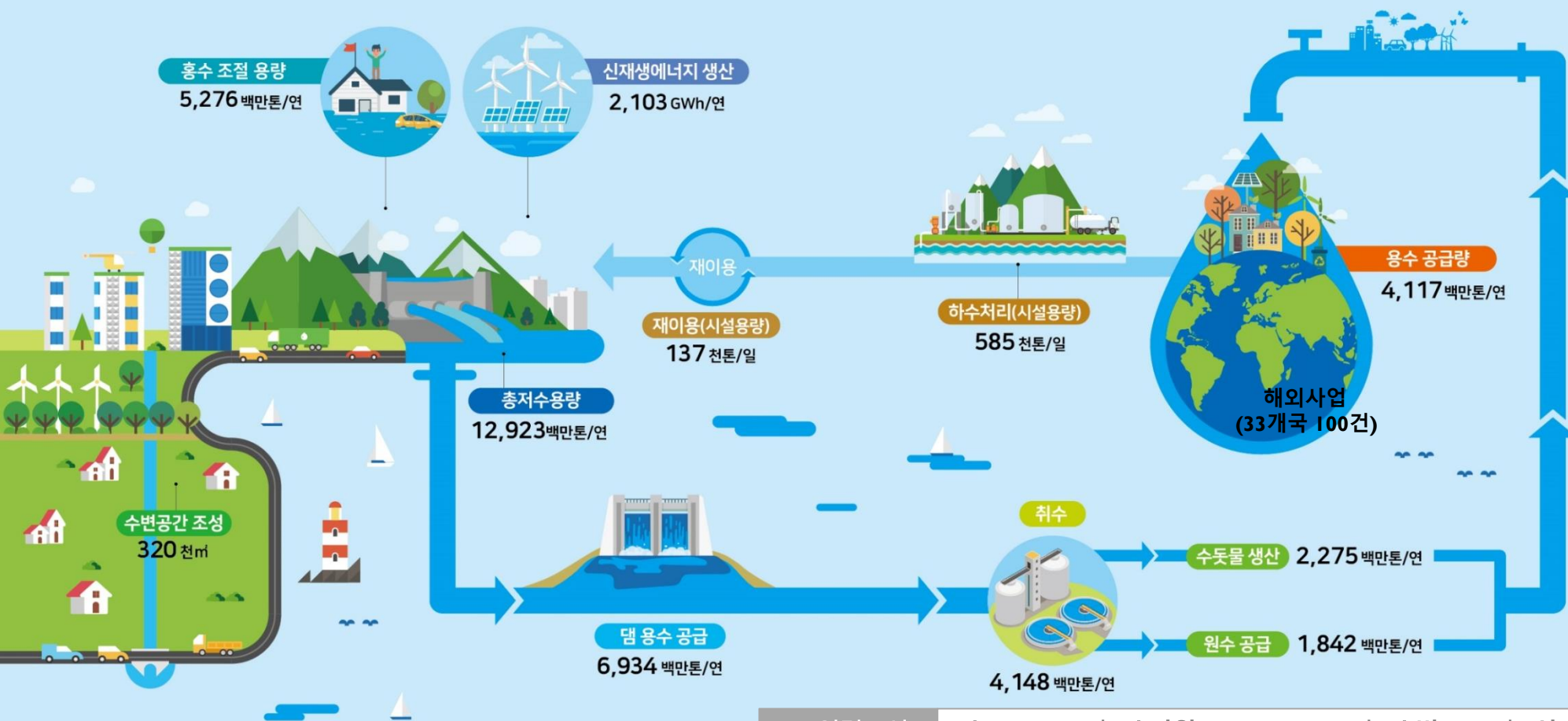
---

## 한국수자원공사 소개

---

## 물순환 전과정에 대한 다양한 업무 영역

물(홍수조절 95%, 재이용 69%, 용수공급 60%, 수돗물 생산 31%), 신재생에너지(8.6%), 스마트시티, 해외사업



※ 인력구성

수도 2,711명, 수자원(물관리) 1,358명, 수변 316명, 신재생 81명, R&D 292명, 해외 97명

## 기존사업 공익성 강화와 함께 혁신적 물관리 위한 도약(퀀텀 점프) 추진



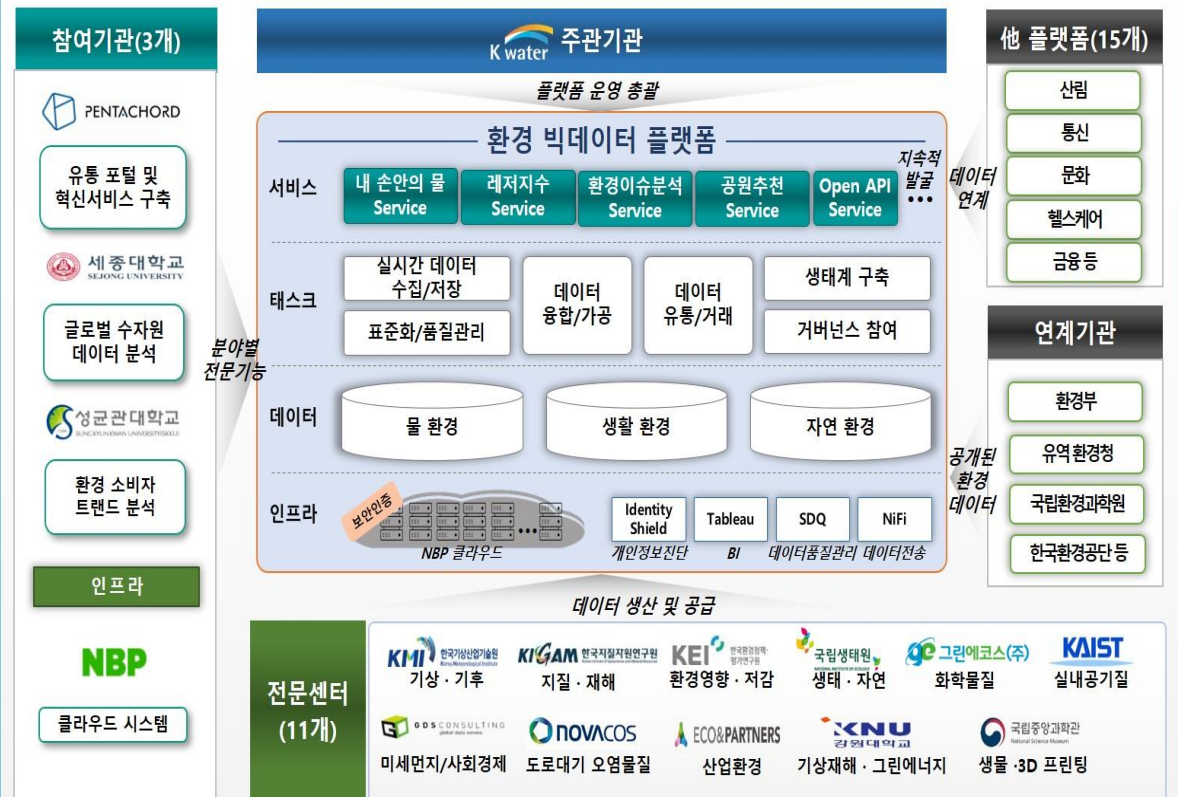


K-water는 환경부 산하 물 전문 공기업으로서  
대국민 환경 빅데이터 점점, 활용 확대 공공성 실현 중!

## 사업개요

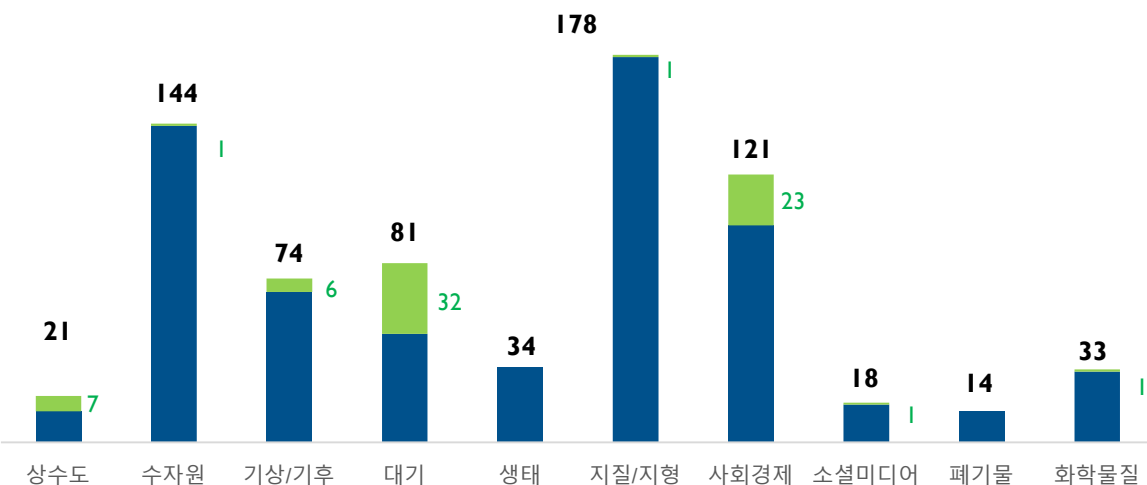
- ❖ 과제명 환경 비즈니스 빅데이터 플랫폼 및 센터 구축사업
- ❖ 수행기관 K-water(주관기관), 플랫폼 참여기관(3개), 센터 참여기관(11개)
- ❖ 사업기간 2021. 1. 1 ~ 2021. 12. 31 (\* 총 사업기간 : 2019~2021)
- ❖ 수행내용
  - ▶ 전문기업 협업을 통한 안정적 · 효율적 플랫폼 구축 및 운영
  - ▶ 환경매체간 · 타 분야간 융합데이터 생산 · 제공으로 新 부가가치 창출
  - ▶ 공공 · 민간 데이터 활용 기반 혁신서비스 발굴
  - ▶ 데이터 유통 · 거래 가격산정 모델 개발, 정책 수립 및 시스템 구축
  - ▶ 민간 신규 비즈니스 창출 지원 및 데이터 혁신 생태계 조성
  - ▶ 환경 빅데이터 분석 전문인력 양성 및 범국가적 거버넌스 구축
- ❖ 최종목표 환경 현안문제 개선의 새로운 돌파구 마련 및 환경 분야 데이터 유통 거래 활성화로 환경 산업 육성 및 일자리 창출에 기여

## 수행체제도

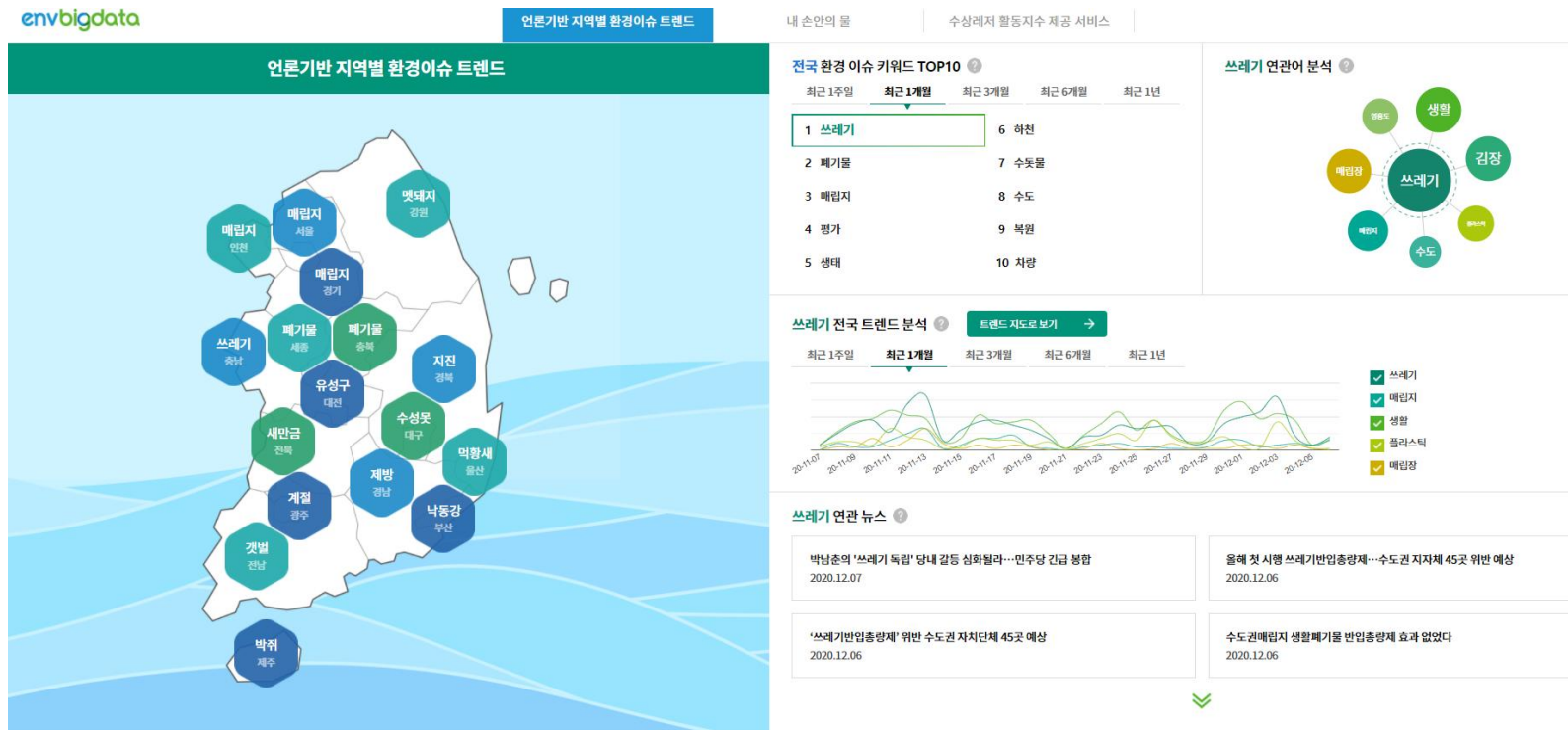


**물, 생활, 자연환경 분야 601 데이터종, 718 데이터상품, 8천여 데이터셋 제공**

구분	데이터 명	판매수
1위	실내 공기질 데이터	2248건
2위	사후환경조사서	775건
3위	화학물질 기본정보	539건
4위	하수처리시설 방류 수질 현황	479건
5위	생태 전국 자연환경 조사(식생)	423건
6위	가뭄 예 경보 정보(기상 가뭄)	384건
7위	하수처리시설 슬러지 반출 정보	332건
8위	미세먼지 관련 질병 지역별 통계	308건
9위	지하수 관측소별 가뭄 분석정보	276건
10위	지자체 / 아파트 배출내역(시간별) 목록	265건



## HOT 언론보도, 키워드 기반 지역별 환경이슈 트렌드 분석 서비스 제공



- 전국 지도 기반 실시간 환경 이슈 키워드 Top10 및 연관 검색어 제공
- 환경 이슈 연관어 분석 및 트렌드 제공



02

---

## 홍수분석 모형 소개

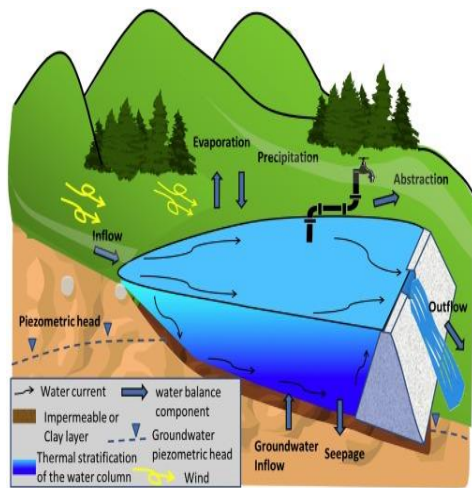
---



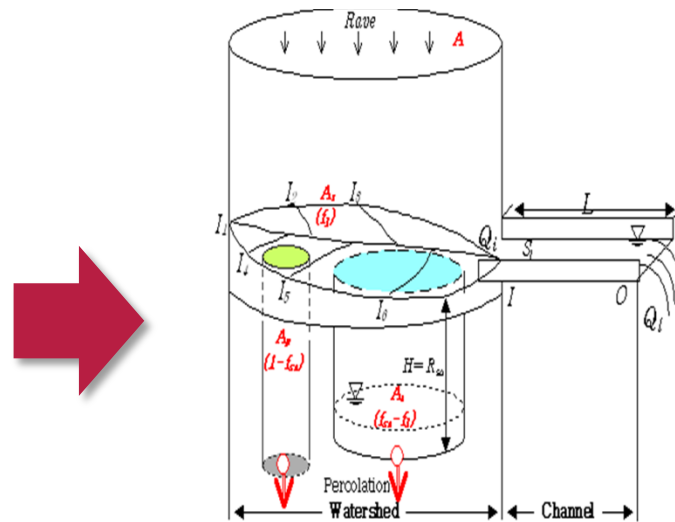
# 물리 모델을 활용한 홍수 예측

K-water는 댐 홍수 예측시 물리적 수치모형인 COSFIM을 개발 활용 中

\* 저류함수 이론에 기반한 물리적(집중형) 수치 모형

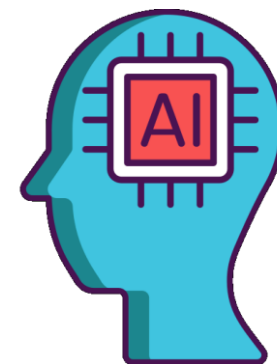


자연 상태



## 물리 모형의 한계

- ✓ 몇 가지 변수로 자연 상태 홍수량을 정확히 예측하는데 한계
- ✓ 정확도 향상을 위해서 물리적 요소들을 추가 고려 필요하나 수식화 되지 않은 요소는 반영이 곤란
- ✓ 미계측 유역 분석 시 가용정보 부족으로 예측 정확도 한계



AI 활용  
도입  
고려

※ (참고) 수자원 분석모형 비교

	물리기반(수치해석) 모형	자료기반(머신러닝) 모형
장점	<ul style="list-style-type: none"> <li>물리적인 과정을 고려한 상세 설계 가능</li> <li>경계조건하에서 흐름 현상에 대한 명확한 특성 파악</li> <li>흐름영역에서 상세한 흐름 특성 규명 가능</li> </ul>	<ul style="list-style-type: none"> <li>관측된 결과자료와 입력자료 연계 과정 우수</li> <li>결측자료 보완 가능</li> <li>복잡한 모의결과 재현 가능</li> <li>물리기반 모형에서의 오차 최소화</li> </ul>
단점	<ul style="list-style-type: none"> <li>모델링에 의한 정확한 지형 자료 필요</li> <li>대상시스템의 물리현상이 규명되지 않거나 수식화 되지 않으면 모델링 불가</li> <li>관측치와 모의치의 오차 처리 어려움</li> </ul>	<ul style="list-style-type: none"> <li>물리적 현상과 연계 과정에서 어려움 발생</li> <li>충분한 학습 및 검증자료 확보 필요</li> <li>입출력 관계의 규명을 위해서 대상시스템의 물리적 이해 필요</li> </ul>



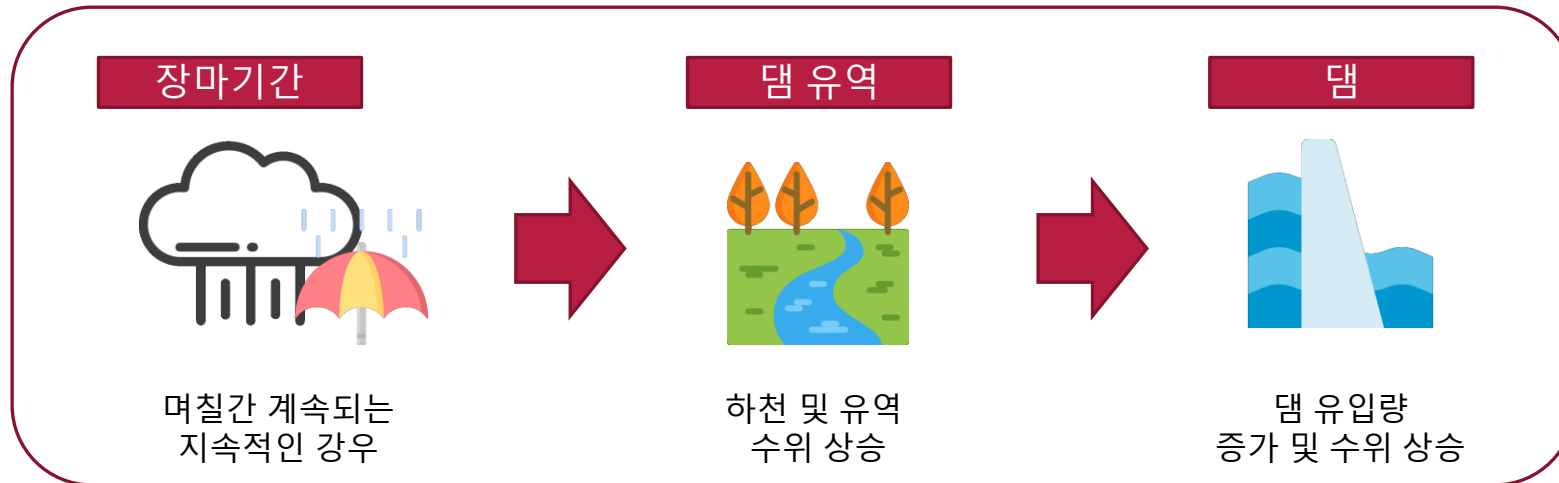
03

---

문제 설명

---

장마 및 태풍기간 동안 댐 주변 지역 강수량, 수위데이터 분석을 통해  
댐에 유입되는 수량(水量)을 예측하여 홍수기 댐 운영 효율화

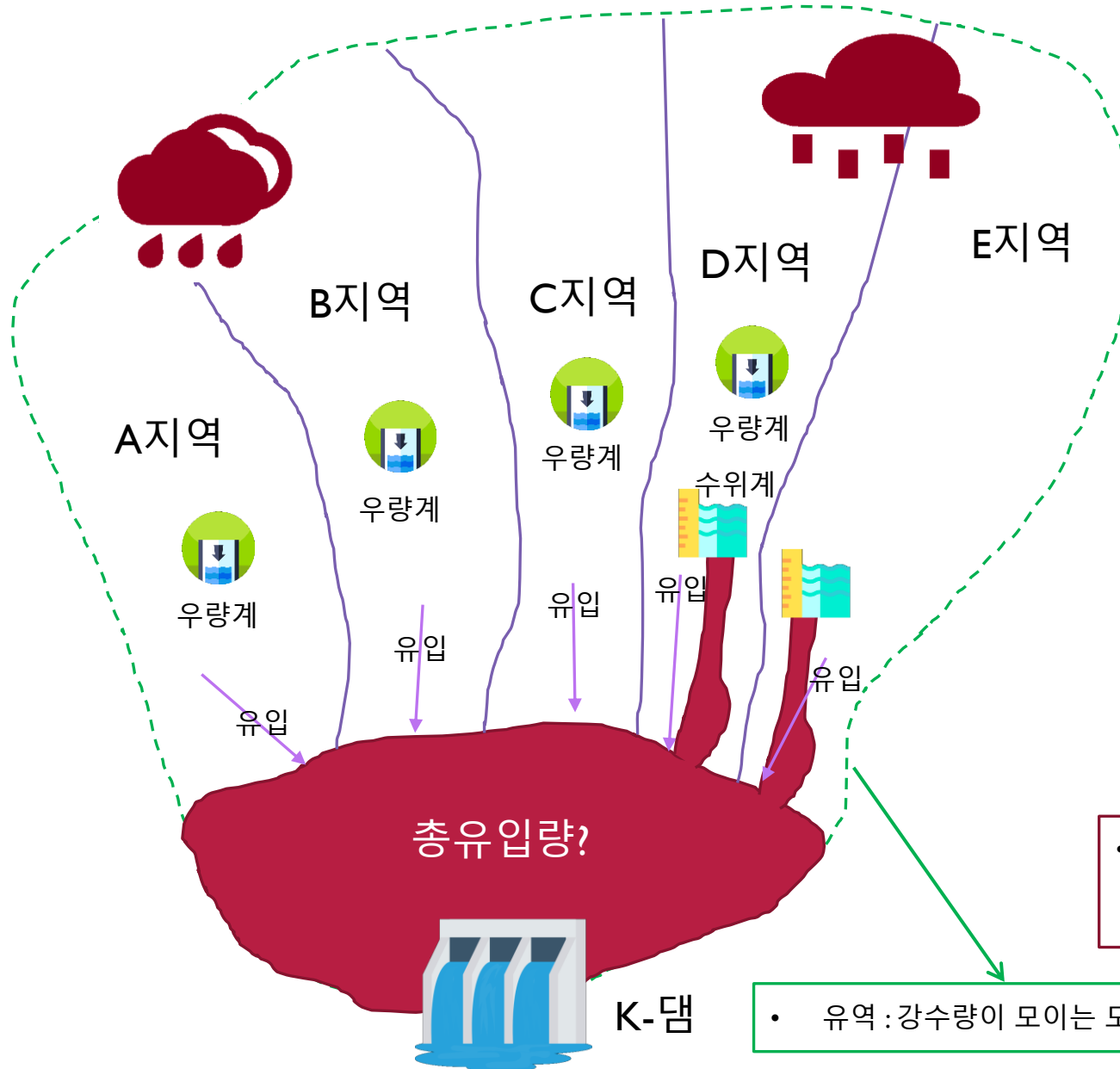


성과활용

기존 물리모형과 결합하여 홍수예측 모형 최적화로 댐 운영 활용  
물 관리 분야 데이터 분석기술 활용, 검증으로 향후 통합물관리 도입 검토



## 문 제



K-댐의 수위는 상류 및 주변지역 5곳으로부터 유입되는 유량의 영향을 받으며 각 지역은 표와 같이 계측기기가 설치되어 있다.

지역	설치 계측기기	기기설명
A, B, C, D	우량계	비가 온 양을 계측
D, E	수위계	물의 높이를 계측

최근 10년간 발생했던 **홍수사상(총 25개)**을 대상으로 유입량에 영향을 미치는 관측소의 강우량 및 수위 데이터를 학습하여 **홍수사상 26의 댐 유입수량(水量)을 예측하시오.**

- 홍수사상 : 강우가 상대적으로 커 토양/지반이 포화된 후 댐으로 많은 양의 유량이 흘러 들어온 기간  
※ 강우가 적을 때는 땅에 침투되어 댐 유입량에 변화 없음

- 유역 : 강수량이 모이는 모든 지역

# 문 제

홍수사상 번호	연	월	일	시간	유입 량	데이터집단 1						데이터집단 2							
						유역평 균강수	강우 (A지역)	강우 (B지역)	강우 (C지역)	강우 (D지역)	수위 (E지역)	수위 (D지역)	유역평 균강수	강우 (A지역)	강우 (B지역)	강우 (C지역)	강우 (D지역)	수위 (E지역)	수위 (D지역)
26	2018	7	1	6		14.3	32.0	0.0	0.0	0.0	1.9	120.5	11.0	32.0	0.0	0.0	0.0	1.9	120.5
26	2018	7	1	7		11.0	20.0	1.0	0.0	0.0	1.9	120.5	7.9	20.0	1.0	0.0	0.0	1.9	120.5
26	2018	7	1	8		7.9	11.0	5.0	0.0	0.0	1.9	120.5	7.9	12.0	5.0	0.0	0.0	1.9	120.5
26	2018	7	1	9		7.9	3.0	11.0	0.0	0.0	1.9	120.5	13.3	13.0	11.0	1.0	0.0	1.9	120.5
26	2018	7	1	10		13.3	4.0	25.0	1.0	8.0	1.9	120.5	20.1	18.0	25.0	14.0	8.0	1.9	120.5
26	2018	7	1	11		20.1	13.0	48.0	14.0	24.0	2.0	120.5	27.0	20.0	48.0	19.0	24.0	2.0	120.5
26	2018	7	1	12		27.0	18.0	58.0	19.0	33.0	2.0	120.5	34.5	22.0	58.0	22.0	33.0	2.0	120.5

예측해야 하는 26번 홍수사상의 종속변수인 유입량

# 데이터 설명(독립 및 종속변수)

❖ 데이터집단 : 종속변수(유입량)와 상관성이 높은 데이터 집단(총 6개 데이터 집단 (1번~6번)으로 구성)\*

\* 성격 : 강우·수위관측소~댐 구간 거리, 시간 등을 달리 설정하여 얻은 독립변수 집단

홍수사상 번호	연	월	일	시간	유입량	데이터집단 1							데이터집단 2						
						유역평 균강수	강우 (A지역)	강우 (B지역)	강우 (C지역)	강우 (D지역)	수위 (E지역)	수위 (D지역)	유역평 균강수	강우 (A지역)	강우 (B지역)	강우 (C지역)	강우 (D지역)	수위 (E지역)	수위 (D지역)
1	2006	7	10	8	189.1	6.4	7.0	7.0	7.0	8.0	2.5	122.6	6.3	7.0	7.0	7.0	8.0	2.5	122.5
1	2006	7	10	9	217.0	6.3	7.0	8.0	7.0	8.0	2.5	122.6	6.4	7.0	8.0	7.0	8.0	2.5	122.6
1	2006	7	10	10	251.4	6.4	7.0	9.0	7.0	8.0	2.5	122.6	7.3	7.0	9.0	7.0	8.0	2.5	122.6
1	2006	7	10	11	302.8	7.3	7.0	10.0	7.0	8.0	2.5	122.6	8.2	7.0	10.0	8.0	8.0	2.5	122.6

❖ 홍수사상번호 : 홍수 고유번호

❖ 유입량 : K-댐에 흘러 들어오는 유량(종속변수)

- ❖ 유역평균강수 : 전체 유역 평균 누적강수량
- ❖ 강우 (A지역) : A관측소 누적강수량
- ❖ 강우 (B지역) : B관측소 누적 강수량
- ❖ 강우 (C지역) : C관측소 누적 강수량
- ❖ 강우 (D지역) : D관측소 누적 강수량
- ❖ 수위(E지역) : E관측소 수위
- ❖ 수위(D지역) : D관측소 수위

04

---

평가 방법

---



2018년 26번 홍수 사상에 대한 예측 결과 160개 데이터에 대하여 RMSE  
(Root Mean Square Error) 평가

$$\text{RMSE} = \sqrt{\sum_{i=1}^{160} (\text{예측값}_i - \text{관측값(실제값)}_i)^2}$$



감사합니다!