

온라인저장소, 클라우드기반 JupyterHub와 모델 APIs를 활용한 수자원 모델링의 재현성 개선

Advancing Reproducibility in Hydrological Modeling: Integration of Open Repositories, Cloud-Based JupyterHub, and Model APIs

최영돈*

Young Don Choi

요 지

지속적인 학문의 발전을 위해서는 선행연구에 대한 재현성이 무엇보다도 중요하다고 할 수 있다. 하지만 컴퓨터와 소프트웨어의 급속한 발달로 인한 컴퓨터 환경의 다양화, 분석 소프트웨어의 지속적 최신화로 인해서 최근 구축된 모델도 짧게는 몇 달, 길게는 1~2년후면 다양한 에러로 인하여 재현성이 불가능해지고 있다. 이러한 재현성의 극복을 위해서 온라인을 통한 데이터와 소스코드의 공유의 필요성이 제시되고 있으나, 실제로는 개인마다 컴퓨터 환경, 버전, 소프트웨어 설치에 필요한 라이브러리의 버전 또는 디렉토리 등이 달라 단순히 온라인을 통한 데이터와 소스코드의 공유만으로 재현성을 개선하기는 힘든 것이 현실이다. 따라서 이러한 컴퓨터 모델링 환경의 공유는 과거의 형태와 같이 데이터, 소스코드와 매뉴얼의 공유만으로 불가능하다고 할 수 있다. 따라서 본 연구에서는 수자원 모델링의 재현성 개선을 위해 1) 온라인 저장소, 2) 클라우드기반 JupyterHub 모델링 환경과 3) 모델 APIs 3개의 핵심 구성요소를 제시하고, 최근 미국에서 개발된 SUMMA(Structure for Unifying Multiple Modeling Alternative) 수자원 모델에 적용하여 재현성 달성을 위한 3개의 핵심 구성요소의 필요성과 용이성을 검증하였다.

첫 번째, 데이터와 모델의 온라인 공유는 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙으로 개발된 수자원분야의 대표적인 온라인 저장소인 HydroShare를 활용하여 모델입력자료를 메타데이터와 함께 공유하였다. 두 번째, HydroShare에서 Web App의 형태로 제공되는 클라우드기반 JupyterHub환경인 CUAHSI JupyterHub(CJH)와 일루노이대학에서 제공하는 CyberGIS-Jupyter for water JupyterHub(CJW)환경에 수자원모델링 환경을 컨테이너(Docker) 환경을 통해 구축·공유하였다. 마지막으로, 클라우드에서 수자원모델의 효율적 이용을 위해 Python기반의 SUMMA모델 API인 pySUMMA를 개발·공유하였다. 이와같이 구축된 3개의 핵심 구성요소를 이용하여 2015년 Water Resources Research에 게재된 SUMMA 논문의 9개 Test Cases 중에서 5개를 누구나 쉽게 재현할 수 있음을 증명하였다.

재현성의 중요성에 대한 인식의 증가로 Open과 Transparent Hydrology에 대한 요구가 증대되고 있으며, 이를 위해서 클라우드 기반의 모델링 환경구축 및 제공이 확대되고 있다. 본 연구에서 제시한 HydroShare와 같은 온라인 저장소, CJH와 CJW와 같은 클라우드기반 모델링환경, 모델의 효율적 이용을 위한 모델 APIs는 급속도로 발달하고 있는 컴퓨터 및 소프트웨어 환경에서 핵심구성요소이며, 연구의 재현성 개선을 통해 수자원공학 발전에 기여할 것으로 기대된다.

핵심용어 : Reproducibility, Open Hydrology, JupyterHub, APIs

* 정회원 · K-water 연구관리처 AI연구센터 책임연구원 · E-mail : choiyd1115@kwater.or.kr