CyberWater:
Making Water Science
Open Science

Professor Xu Liang
Professor Yao Liang

Scientia

CYBERWATER: MAKING WATER SCIENCE OPEN SCIENCE

Water is possibly the most important resource on our planet, and being able to model and understand the systems that transport it through and around our planet is essential. As water science straddles many diverse scientific disciplines, gathering and representing heterogeneous data in a standardised way for analysis and modelling can be challenging. A joint project among eight institutions, led by **Dr Xu Liang** (University of Pittsburgh) and **Dr Yao Liang** (Indiana University–Purdue University Indianapolis), is creating a generalised, open-source platform to allow complex and diverse data to automatically and directly flow into diverse water models.

Collaborative Computer Models

Many of the natural processes studied by researchers are incredibly complex. Scientists often have to integrate different models to obtain a broader and deeper view of a system under investigation. This is further complicated by the fact that often, various data providers offer their data products in different formats, structures and organisations (e.g., ASCII, Binary, Geotiff, GRIB/GRIB2, HDF5, HGT, NetCDF and XML), as well as data access protocols (e.g., FTP, HTTP/HTTPS and OPeNDAP). On top of this, different models are usually developed by individual and independent research groups in their own domains rather than being standardised across an area of scientific study. This makes it very difficult for scientists wanting to make use of these data with their own models or their data with other models, as they're all using different techniques, data sets, and computing languages.

Water science suffers from this abundance of incompatible data and models. It is a very broad area of study, with researchers investigating groundwater levels, river flow, water quality, and making predictions on the frequency of floods and droughts affected by climate changes, amongst many other topics. With so many interrelated scientific disciplines under one umbrella, it is not surprising that different groups approach designing computer models differently. When research groups attempt to combine models written by different groups, they have to write some kind of 'glue' code to integrate the models together. This process is difficult and time-consuming, with each pair of models having to have separate glue code written for them. This is bad enough when using two models but quickly becomes a huge task when researchers are trying to get multiple models to work in sync with each other.

Dr Xu Liang from the University of Pittsburgh and Dr Yao Liang from Indiana University–Purdue University Indianapolis have been working with their team to build a platform, CyberWater, to make diverse data and computational models in water science easier to use. The CyberWater project aims to create an open-source framework for researchers and students to easily integrate and couple complex models and to have diverse and heterogeneous data from data providers directly and automatically flow to these models in a user-friendly and workflow-controlled environment.

Going with the Workflow

The CyberWater framework is developed to significantly simplify complex data and model integration with little to no coding. It supports high-performance computing (HPC) on demand, the reproducibility and reusability of the
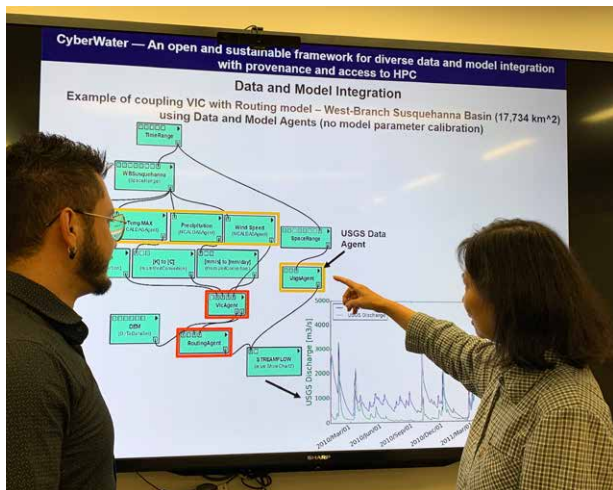
Dr Xu Liang working with a student. Credit: Xu Liang.

model system, and allows for the easy integration of other systems. It also provides tools, such as a generic model agent toolkit, so that the user can construct model agents to integrate their own models into the CyberWater system without coding. In addition, CyberWater provides tools through its static parameter agent toolkits to automatically generate model parameter input files with various file formats to facilitate the time-consuming and error-prone data and model parameter files pre-processing process for diverse models. With the CyberWater framework, diverse data products and complex models are integrated via data and model agents, respectively.

At the heart of the CyberWater framework is the development of meta-scientific modelling (MSM), an open-data, open-model platform that provides a sophisticated workflow-controlled modelling environment for heterogeneous data and model integration without the need for central administration. This means that models and data shared with the platform are available for other researchers to use. Users can also easily and freely add any new data sources/products and models to the CyberWater system by themselves.

This open architecture allows researchers to bypass the issue of having multiple different models requiring multiple unique pairwise glue codes to work together. Once a model has been entered into the MSM system, it can be coupled with any other models that already exist in the system. In the same way, any data integrated into the CyberWater system can be accessed by any model on the platform without the need for further data preparation or processing.

As a further advantage, the MSM was developed to handle the input and output of the models it hosts and does not require model source code to be integrated into the system. The benefits of this are two-fold – by not needing the actual code of each programme to integrate it, MSM simplifies the integration work itself. It also allows the integration of models where the source code is not available, such as models based on commercial software.

The CyberWater project embodies several functionalities to make it easier for researchers to use the platform. A generic model agent toolkit helps researchers construct the model agents to integrate their own models into the system without writing any computer code; the static parameter toolkit helps users prepare complicated model input parameter files. As such, this set of toolkits is designed to make the platform as user-friendly as possible.

To make the platform as welcoming as possible, the team has also created several system integration engines to support popular system platforms, including GIS, Matlab and Docker. Additionally, when a researcher has a computationally expensive model they want to run, the team has created a mechanism to offload that processing with on-demand high-performance computation. This will compute the complex models efficiently and effectively without slowing down the user's local computer.

Open Water Science

The CyberWater project tackles a hugely broad scope in water science. Excitingly, CyberWater2 is now under development, meaning that models concerning the atmosphere, land surface, rivers and coastal regions can all be coupled together. Other capabilities (such as supporting intelligent site recommendations for HPC/Cloud access on demand) are being added to the existing CyberWater system through the ongoing CyberWater2 project.

Using the CyberWater/CyberWater2 platform, researchers from across different areas of science can collaborate on solving some of the big challenges that society currently faces, including coastal flooding, landslide risk assessment, forest fire, severe drought predictions, and depleting groundwater levels that are associated with climate change.

Perhaps most importantly, this project is designed to remove as many barriers as possible for its users, making it much easier to conduct large-scale collaborations across heterogeneous computing platforms, disciplines and organisations to more effectively solve large and complex problems than previously possible. By being open-source and providing multiple toolkits for the science communities, the CyberWater/CyberWater2 project lowers the financial and technical cost for researchers and students wanting to participate in this area of science.

The work of Dr Xu Liang, Dr Yao Liang, and their colleagues on building CyberWater/CyberWater2 has resulted in a platform that provides a very powerful cyber-infrastructure framework – and perhaps shows the way the current is flowing for a more open and collaborative scientific future. The beta version of CyberWater will be released to the public soon.

Meet the Researchers

Professor Xu Liang
Department of Civil and Environmental Engineering
University of Pittsburgh
Pittsburgh, PA
USA

Professor Yao Liang
Department of Computer and Information Science
Indiana University–Purdue University Indianapolis (IUPUI)
Indianapolis, IN, USA

Dr Xu Liang is a Professor of Hydrology. Her core research spans three main areas: land surface modelling and eco-hydrology, hydroinformatics using advanced machine learning methodologies, and cyber system development. She actively engages in interdisciplinary collaboration work with atmospheric scientists, plant biologists, and computer scientists. Dr Liang has been instrumental in the initial and subsequent development of the VIC land surface model and the VIC+ model. She has been an elected Fellow of the American Meteorological Society since 2016. She is also the recipient of the Chancellor's Distinguished Research Award (senior category) of the University of Pittsburgh in 2016, the recipient of the 2014 Carnegie Science Environmental Award, and the recipient of the Hellman Foundation Junior Faculty Research Award of the University of California at Berkeley in 2000. She held the William Kepler Whiteford Professorship from 2014 to 2019. Prior to joining the University of Pittsburgh, she was a member of the faculty at the University of California, Berkeley. Dr Liang received her PhD in hydrology from the University of Washington (Seattle) and completed postdoctoral work at Princeton University.

Dr Yao Liang is a Professor in the Department of Computer and Information Science. His research interests include wireless sensor networks, the Internet of Things, cyberinfrastructure, open data and model integration, machine learning, neural networks, data engineering, and distributed systems. His research projects have been funded by the National Science Foundation, the National Aeronautics and Space Administration, and the Department of Transportation. In 2019, he received the Glenn W. Irwin, Jr., M.D., Research Scholar Award at IUPUI. He served as the General Co-Chair of the International Conferences on Big Data Engineering from 2019 to 2022. Dr Liang has given invited talks and lectures at various universities across the USA, Europe and China. He is a senior member of the Institute of Electrical and Electronics Engineers and a member of the Association for Computing Machinery. He also gained extensive industrial R&D expertise as a Technical Staff Member at Alcatel. Dr Liang received his BS degree in Computer Engineering and MS degree in Computer Science from Xi'an Jiaotong University and his PhD in Computer Science from Clemson University.

CONTACT

E: xuliang@pitt.edu

CONTACT

E: yaoliang@iupui.edu

KEY COLLABORATORS

A. Castronova, D. McCay and R. Hooper, Consortium of Universities for the Advancement of Hydrologic Science
I. Demir and W. Krajewski, University of Iowa
L. Ruby Leung, Pacific Northwest National Laboratory, Department of Energy
J. Lin, University of Pittsburgh
L. Lin, Ball State University
S. Pamidighantam and R. Quick, Indiana University
F. Song, Indiana University–Purdue University Indianapolis
Y. Zhang, Northeastern University

FURTHER READING

R Chen, D Luna, F Li, et al., CyberWater: An Open Framework for Data and Model Integration in Water Science and Engineering. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022. DOI: https://doi.org/10.1145/3511808.3557186

R Chen, D Luna, Y Cao, et al., Open data and model integration through generic model agent toolkit in CyberWater framework, Environmental Modelling and Software, 2022, 152, 105384. DOI: https://doi.org/10.1016/j.envsoft.2022.105384

D Salas, X Liang, M Navarro, et al., An open-data open-model framework for hydrological models' integration, evaluation and application, Environmental Modelling and Software, 2020, 126, 104622. DOI: https://doi.org/10.1016/j.envsoft.2020.104622

FUNDING