



Making Waves: Towards data-centric water engineering

Guangtao Fu^{a,*}, Dragan Savic^{a,b}, David Butler^a

^a Centre for Water Systems, University of Exeter, Exeter EX4 4QF, United Kingdom

^b KWR Water Research Institute, Nieuwegein 3430 BB, the Netherlands

ARTICLE INFO

Keywords:

Artificial intelligence
Data-centric
Model-centric
Scientific paradigm
Water engineering

ABSTRACT

Artificial intelligence (AI) is expected to transform many scientific disciplines, with the potential to significantly accelerate scientific discovery. This perspective calls for the development of data-centric water engineering to tackle water challenges in a changing world. Building on the historical evolution of water engineering from empirical and theoretical paradigms to the current computational paradigm, we argue that a fourth paradigm, i.e., data-centric water engineering, is emerging driven by recent AI advances. Here we define a new framework for data-centric water engineering in which data are transformed into knowledge and insight through a data pipeline powered by AI technologies. It is proposed that data-centric water engineering embraces three principles – data-first, integration and decision making. We envision that the development of data-centric water engineering needs an interdisciplinary research community, a shift in mindset and culture in the academia and water industry, and an ethical and risk framework to guide the development and application of AI. We hope this paper could inspire research and development that will accelerate the paradigm shift towards data-centric water engineering in the water sector and fundamentally transform the planning and management of water infrastructure.

1. The era of big data and artificial intelligence

Our world is currently undergoing the most profound change in human history driven by big data and artificial intelligence (AI). Digital technologies have advanced more rapidly than any innovation in our history, and have moved our society into the era of big data with the proliferation of sensing devices and data facilities. AI technologies, in particular machine learning, have been developed and deployed to process large amounts of data for problem-solving and decision-making tasks, significantly improving efficiency and productivity in many sectors (Russell et al., 2015). The impact of AI on industry and day-to-day life is just beginning, with the potential to bring unprecedented benefits to humanity. The development of general-purpose AI systems employing what we already have more effectively and at greater scale was estimated to be worth a net present value of \$13.5 quadrillion and produce a tenfold increase in global gross domestic product, which last time required over 190 years from 1820 to 2010 in history (Russell et al., 2015).

AI could potentially transform how scientific research is conducted in many disciplines and its impact is just emerging. Jim Gray, a pioneering computer scientist, described a new paradigm of scientific research - data-intensive science - as the fourth paradigm of science

following empirical, theoretical and computational paradigms (Hey et al., 2009). A wide range of fundamental sciences, including mathematics, medical science, physics, chemistry and geoscience, are already being affected by AI (Xu et al., 2021). For example, the success of AlphaFold in predicting 3D protein structures has expanded the size of the protein structure database by 200 times, with the potential to accelerate scientific discovery in biology (Hassabis, 2022).

Water engineering is similarly poised to be significantly affected by AI technologies. Water engineering is a sub-field of engineering that addresses complex issues of sustainable water management in a changing world, generally covering water resources, water treatment, water distribution, stormwater, and wastewater systems. AI applications have been found in various problems including anomaly detection, prediction, asset condition assessment, operation, planning and maintenance (Fu et al., 2022; Li et al., 2023) and the development of pathways towards sustainable and resilient water systems (Fu et al., 2023). The water sector has come a long way in terms of digital transformation, though water utilities are at varying stages of adopting digital technologies (Boyle et al., 2022; Daniel et al., 2023).

This perspective aims to propose a new data-centric paradigm for water engineering to tackle water management problems in a changing world. This paradigm is empowered by AI technologies focusing on how

* Corresponding author.

E-mail address: g.fu@exeter.ac.uk (G. Fu).

<https://doi.org/10.1016/j.watres.2024.121585>

Received 8 December 2023; Received in revised form 14 February 2024; Accepted 7 April 2024

Available online 8 April 2024

0043-1354/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

they effectively extract knowledge and insight from the data pipeline to improve, enhance and develop water management. We will first review the historical evolution of water engineering, then define the data-centric framework and three underlying key principles, and finally discuss ways forward to enable the paradigm shift in water research and practice. This paper is positioned as a call for action to move towards data-centric water engineering in the water research and practice communities.

2. Paradigms of water engineering

Historically, water engineering as a research discipline and as a practising profession has gone through empirical, theoretical, and computational paradigms. In recent years, the interest in data-driven research and development has grown substantially in water engineering (Eggimann et al., 2017; Fu et al., 2022; Makropoulos and Savic, 2019). This calls for the recognition of a new paradigm – data-centric water engineering – as a pillar of water research and development. Before introducing the new paradigm, the previous paradigms are explained briefly to understand the historical evolution of water engineering (Fig. 1).

2.1. Empirical water engineering

Water engineering can be traced back to ancient civilisations when many hydraulic structures were built for water supply and drainage. Mesopotamia built lengthy branching canals in 2950 – 2400 BCE and, a sophisticated network of small tanks connected by canals was built for irrigation in Sri Lanka about 2500 years ago (Vairavamorthy, 2022). The Romans are well known for building elaborate aqueducts to transport water long distances for urban water supply and artificial drains (e.g. cloaca maxima) to take used water away to prevent flooding. In China, a large-scale hydraulic complex consisting of dams, levees and ditches was built to support the ancient city of Liangzhu about 5100 years ago (Liu et al., 2017). Dujiangyan was another Chinese engineering feat constructed in 256 BCE to solve irrigation and flooding problems with three key structures - a fish mouth levee, flying sand weir and bottle-neck channel - working together (Zhang et al., 2013).

At this stage, water engineering was based on rules of thumb developed through observation and understanding of the local environment and natural phenomena rather than scientific principles. With limited knowledge of flow dynamics, however, it was not uncommon for projects to fail or result in unintended consequences, with examples such as the use of lead in water pipes resulting in public health hazard for the Romans. On the contrary, successful projects such as Dujiangyan resulted from a detailed understanding of geography and water forces. Advanced knowledge in hydraulic engineering was normally regarded as a key driver for social, political and economic developments in

ancient civilisations.

2.2. Theoretical water engineering

The science of hydraulics did not advance much until Leonardo da Vinci summarised the state-of-the-art in a book circa 1500 (Walski, 2006). Classic water engineering began to develop in the 17th century with the advances in hydraulic experiments and theories. Hydrostatics took shape with the discovery of Pascal's law in 1653. With the advances in mathematics and physics, the 18th and 19th centuries saw the arrival of hydrodynamics, with some fundamental theories gradually developed, for example, Bernoulli's theory, Chezy and Manning formulae, Prandtl and von Karman's laws and Navier-Stokes equations (Chadwick et al., 2013). Experimentation was a key tool to determine key coefficients, validate theories and understand flow dynamics. Significant advances in instrumentation were also made, such as the Piezometer and Pitot tube, enabling pressure and flow velocity measurement respectively. In another line of development, the activated sludge process was discovered by laboratory experiments in the 1910s and was quickly adopted to solve the water body pollution problem that caused public health crises in many European cities (Jenkins and Wanner, 2014).

Hydraulic theories significantly enhanced the understanding of fundamental hydraulic behaviour compared to the first paradigm and thus enabled rapid development of large and complex water infrastructure for growing cities. For example, in the mid-18th century, water in London was delivered through a network of approximately 50 km of wood and cast-iron pipes, pumped from rivers by water wheels and later steam engines (Walski, 2006; Sedlak, 2014). At this stage, however, water engineers had a much higher capacity in constructing systems than in conducting analysis to understand hydraulic behaviours as they relied on a combination of simplifications, rules-of-thumb, and conservatism (Walski, 2006).

2.3. Computational water engineering

A new era of water engineering began when digital computers were first applied to solve hydraulic problems in the early 1950s. The early models could only solve steady-state hydraulic problems and required punch-card input on large mainframe computers, but water utilities began to use such models (e.g., the McIlroy Network Analyzer) for flow simulation (Walski, 2006). With the popularity of personal computers from the 1970s and their rapidly increasing computational power, physics-based models were developed to represent more detailed processes and higher temporal-spatial resolutions. For example, flow dynamics are simulated with increasing dimensions ranging from 1D and 2D to 3D, runoff processes are represented from lumped to grid cells at sub-metre levels and water quality is increasingly included in water systems modelling for problems such as water distribution system design, green infrastructure planning and effluent quality consenting. As a result, hydroinformatics was established as a research field to focus on computer simulation/optimisation modelling for supporting informed decision-making (Makropoulos and Savic, 2019).

Compared to the previous stage, this stage significantly improved the capacity of understanding detailed processes, predicting their behaviour and evaluating engineering solutions under various conditions before they were implemented in the real world. Computers were, in effect, used as a virtual laboratory for hydraulic experiments. However, a model-centric approach was generally followed with efforts focused on improving modelling accuracy by increasing model realism, complexity or choosing appropriate models. This is true even for data-driven models (Liu et al., 2023). The use of computer simulation models and optimisation methods to support decision making has been widely accepted in water systems planning and management.

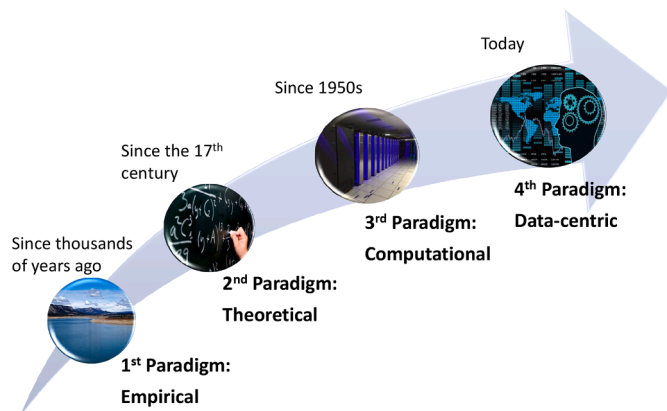


Fig. 1. Four paradigms of water engineering.

2.4. The need for data-centric water engineering

As discussed above, data has been playing a central role in the history of water engineering: from activities related to observing and measuring water processes, through extracting physical laws from experiments to building computer models which could be used for design and operational purposes. In an age of big data and AI, however, what is fundamentally different from the past is our ability to gather, manage and analyse data in resolutions, scales and volumes that have grown beyond our imagination. Data-centric water engineering is emerging as the 4th paradigm of water engineering and represents a radical change in the use of data for water system planning, management and operation.

In particular, the advance of physics-based models has been gradual in recent years and met with many challenges such as complex interactions, large numbers of model parameters, intensive computing resources, and limited human resources and skills (Fu et al., 2022; Nearing et al., 2021). AI could be just the right technology to help develop the next-generation models of water systems, as demonstrated by an increasing number of AI-based or hybrid models of water systems, e.g., flood models (Zhang et al., 2023) and water distribution models (Li et al., 2024). The huge potential of AI has started to be realised as one of the most useful technologies for advancing scientific research and practical adoption in the real world.

Building on the previous paradigms, this paradigm will significantly improve the speed and efficiency of the data-to-knowledge transformation process for water system planning and management. It will also lead to higher levels of automation in decision making compared to previous paradigms.

3. Defining data-centric water engineering

Data-centric engineering is emerging as a new research field sitting at the interface of data science and engineering (Butler et al., 2019). It is the systematic fusion of fundamental physical and chemical laws with data-enabled and empirically derived laws (Girolami, 2021). Data-centric engineering has gained significant recognition in recent years due to the increasing availability of data in many industries and AI advances, which enable the leverage of large amounts of data to understand complex system behaviours and make informed decisions. Water engineering, as a branch of the engineering discipline, charts a similar development pattern of data-centric engineering. The framework and key principles for data-centric water engineering are discussed below.

3.1. New framework

At its most fundamental level, a water system can be thought of as an information processing system, albeit an extraordinarily complex and dynamic one, as it handles a range of diverse data related to flow, water quality, assets, socioeconomic systems, and interdependent environment and engineering systems. Data-centric water engineering is used here as a term to describe the paradigm for designing and managing water systems that focus on the handling and management of data as the primary concern. Data-centric water engineering can be viewed as an interdisciplinary academic discipline of integrating water research with cutting-edge AI technologies to provide meaningful insight, actionable knowledge and high-performing interventions for sustainable and resilient management of water systems in the face of social and environmental changes.

Fig. 2 shows a new framework of data-centric water engineering in which the data pipeline plays a central role in water infrastructure planning and management and stakeholder engagement. In the pipeline, data and information are collected from water infrastructure, its environment and interdependent systems, and various stakeholders. They are fed into the computational core of the framework – an integrated data-driven (i.e., machine learning) and physics-based modelling engine – for new knowledge and insight extraction. Knowledge and insight are applied to water infrastructure systems and fed back to stakeholders. This process is iterated to meet the needs of the society and environment that are fed into the data pipeline through stakeholder engagement.

The integrated modelling approach in the data pipeline, empowered by AI technologies, is thought to be fundamentally different from traditional modelling approaches. ChatGPT has demonstrated the ability to automate modelling and coding tasks, with its application in hydrology and earth sciences (Foroumandi et al., 2023) and water resources management (Egbemhenge et al., 2023). It can be envisioned that the modelling process could be automated so that the transformation from data to knowledge and insight through the data pipeline could be highly efficient concerning human resources and cost requirements, enabling effective decisions for water system management.

3.2. Fundamental principles

We propose the following key principles of data-centric water engineering, which make it fundamentally distinct from the previous paradigms.

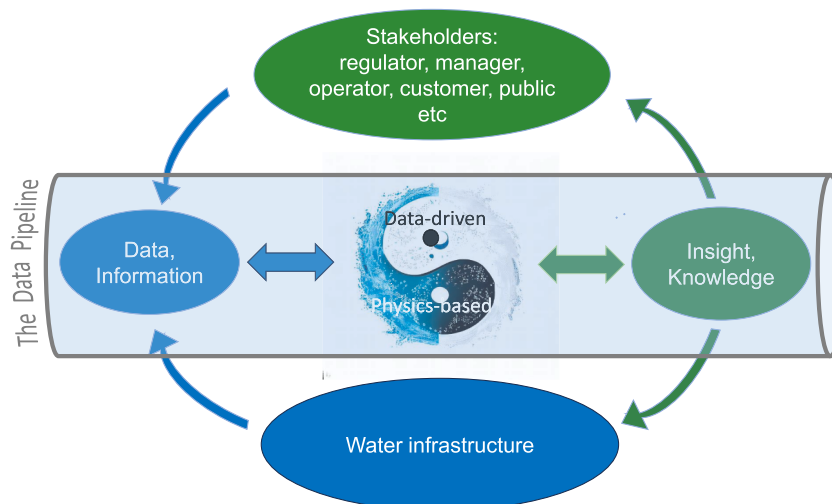


Fig. 2. Proposed framework of data-centric water engineering.

3.2.1. Data-first

Data should be treated as a first-class citizen in the planning, design, operation and management of water systems. This means that the data pipeline should be maintained and updated at every stage along with the infrastructure. This approach is in contrast to traditional systems approaches in the 3rd computational paradigm, which tend to focus on the functionality of the system and the processes it will support, rather than the data itself. The data-first principle requires the consideration of data-related activities: capture, transmission, storage, curation, and analytics, and the design and management of data infrastructure should become an integrated part of water infrastructure management.

Water systems are now recognised as cyber-physical systems (CPS) that integrate sensors, controllers, data management and computational capabilities to control and monitor physical processes. The cyber system should be planned and managed together with the physical system in an integrated way in the whole life cycle. This involves many problems such as optimal sensor placement, data quality assurance, sensor anomaly detection, security, proactive maintenance and interdependency with other systems (e.g., communication), and thus affects investment decisions. In addition to data collected from water systems, data from other sources such as weather and geo-spatial data from satellites, radars, unmanned aerial vehicles and social media should be included in the data pipeline, and they can be of varying scales and shapes, such as text and videos.

3.2.2. Integration

Data-centric water engineering should be regarded as an integrating framework that not only unifies the previous empirical, theoretical and computational paradigms but also provides a new approach for them to interact and improve. This is primarily exhibited by how the data and information are processed and learned to form new knowledge and insights for water system planning and management.

AI-human integration is fundamental in knowledge and insight generation. In the previous paradigms, the generation of knowledge and insight predominantly relies on humans: physical theories and laws are developed and tested based on observations and experiments to understand water systems, and then used to build physics-based models for predictions of environmental changes and human interventions. However, many processes and properties of water systems are still understood less or are highly heterogeneous in space and time, in particular with changing environment and increasing scales. This is reflected in assumptions and empirical relationships in physics-based models. Physics-based models often have limitations in high computational demands and require high human resources and skills in transferability across systems and scales. The advances in machine learning provide a new (often automated) way to understand the behaviour of water systems directly from growing amounts of data collected, though it still has challenges in improving explainability and extrapolation. Thus integration of physics-based and machine learning models can effectively improve our ability to develop a digital representation of natural and engineered water systems that can be used for planning and management decisions.

The integration of physics-based and machine learning models can take different forms. Various terminologies and approaches have been proposed, such as physics-informed or theory-guided machine learning, hybrid modelling, and differentiable modelling (Shen et al., 2023). However, a key challenge is to identify synergies between physical and data-driven models and suitable ways to leverage the AI power for accurate system modelling. Most importantly, AI advances could significantly improve the efficiency of model development and reduce the human efforts required. One example is the development of foundation models such as the Geospatial AI Foundation Model (Jakubik et al., 2023), which could be used to re-train deep learning models for a specific region or be integrated with hydrodynamic models for flood simulations.

3.2.3. Decision making

AI enables automation of decision making inherently. In data-centric water engineering, a move from decision support to higher levels of automation in decision making is inevitable.

In the 3rd paradigm, computer simulation and optimisation models have been playing an increasingly significant role in water management involving hydraulics, hydrology and environmental engineering. However, these models are generally regarded as a tool to support informed decision making by improving understanding of water systems, estimating potential impacts of interventions or exploring decision space, though their scope is not limited to water systems but also embraces social-economic issues that are linked to water systems. This is referred to as the 'human in the loop' approach in which human does the decision making and AI provides only decision support (Ross and Taylor, 2021).

Compared to the human-in-the-loop approach, there are higher levels of automation in decision-making approaches driven by AI (Ross and Taylor, 2021): 1) human in the loop for exceptions where most decisions are automated but human intervention is required for exceptional conditions which could be extreme events or AI systems of high uncertainty; 2) human on the loop where AI is assisted by humans in automated decision making but humans review decision outcomes and adjust parameters for future decisions; 3) human out of the loop where AI makes every decision but humans intervene only by setting new constraints and objectives, for example, in response to evolving needs of stakeholders.

Automated decision making is already happening in the water industry. For example, control systems of varying automation levels are implemented throughout the urban water and wastewater system mainly for the control of process units in water and wastewater treatment plants. With data-centric water engineering, it is envisioned that automated system control becomes more prolific in water systems, in particular for real-time control problems such as pump scheduling, stormwater control, and green infrastructure control. Further, decisions related to planning, design and maintenance tasks can also be automated, though human intervention may be required at different levels and through different means. One example is predictive maintenance where investment plans are developed and implemented from monitor systems using machine learning. All these examples of higher levels of automation mean that different training of personnel involved with water systems is needed, thus may require even more training to embrace AI and automation of the water cycle (Savic, 2022).

4. Ways forward

An interdisciplinary research community is needed to develop a common vision and enable a paradigm shift to data-centric water engineering which is, by its nature, interdisciplinary (Ley et al., 2020). For water engineering, combined knowledge from many disciplines such as water engineering, data and computer sciences, and social sciences is needed to tackle increasingly complex water challenges. The interdisciplinary approach has already been widely recognised in the context of hydroinformatics (Makropoulos and Savic, 2019). In the era of AI, however, developing interdisciplinary collaborations between data scientists and water researchers becomes more important for solving highly complex water problems with increasing amounts of data. Further, collaborative partnerships with big IT companies could play a key role in leveraging AI to tackle water challenges as they are increasingly driving AI advances as demonstrated by recent successes such as AlphaGo, ChatGPT and the Geospatial AI Foundation Model (Jakubik et al., 2023).

A shift in mindset and culture in academia and the water industry is required. The focus on data and the development of the data pipeline requires significant changes to the way that organisations work and water infrastructure is managed, including the adoption of new tools and processes. A key challenge is to improve the level of data literacy within organisations, which means that engineers and researchers must be trained to reach a deeper understanding of data and how it can be

used to drive value for improved water management (Wagener et al., 2021). Developing sufficient sensing systems and data infrastructure could be challenging due to significant investments required, in particular for the Global South.

An ethics and risk management framework is needed to address important concerns related to the use of AI systems and related data. This is a fundamental prerequisite for the successful implementation of data-centric water engineering. Risks could arise in many respects including data privacy, security, bias and discrimination, inequity and social injustice, design errors and misuse. Failures or design errors in AI systems could generate direct impacts on water system operations, cascading water system failures and even impacts across other interdependent systems and wider society (Richards et al., 2023). For example, unintended consequences on global climate change mitigation may occur when an AI system optimises urban wastewater treatment processes but does not consider both direct and indirect carbon emissions (Sweetapple et al., 2014). However, progress is being made to manage AI risks, such as, published guidelines for secure AI system design, development, deployment, operation and maintenance published (UK National Cyber Security Centre, 2023), and the EU AI Act - the world's first comprehensive AI law to regulate the use of AI. Notwithstanding the above, specific guidelines for robust data management and curation protocols as well as responsible, trustworthy AI systems in the water sector are still needed to achieve the widespread adoption of data-centric water engineering.

Data-centric water engineering is emerging as a new paradigm for water research and development. The data pipeline through which knowledge and insight are extracted from data could be a fundamental feature of this new paradigm and it could be powered by AI advances for significant improvements in efficiency and productivity. Though challenges may arise from dimensions of cyber-physical infrastructure, institutional governance, social-economic systems and technological development in wider society (Eggmann et al., 2017; Fu et al., 2022, 2023), we envision that the new paradigm will transform the way water systems are planned and managed to allow for more effective knowledge and insight extraction from data at scale and thus the creation of more sustainable and resilient water systems.

5. Conclusions

- Data-centric water engineering is emerging as a new research paradigm for water research, development and practice, following the historical evolution of empirical, experimental and computational paradigms.
- The fundamental feature of data-centric water engineering is the data pipeline empowered by AI for knowledge and insight extraction from data.
- Data-centric water engineering should embrace three key principles – data-first, integrated modelling and automated decision making.
- The development of data-centric water engineering calls for an interdisciplinary research community, a shift in mindset and culture in academia and the water industry, and an ethical framework to guide the development and application of AI.

CRedit authorship contribution statement

Guangtao Fu: Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review & editing. **Dragan Savic:** Resources, Writing – review & editing. **David Butler:** Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the Royal Society under the first author's industry fellowship scheme (Ref: IF160108). Dragan Savic has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [951424]).

References

- Boyle, C., Ryan, G., Bhandari, Pratik, Law, K.M.Y., Gong, J., Creighton, D., 2022. Digital Transformation in Water Organizations. *J. Water Resour. Plann. Manage.* 148 (7), 03122001 [https://doi.org/10.1061/\(ASCE\)WWR.1943](https://doi.org/10.1061/(ASCE)WWR.1943).
- Butler, L.J., Lau, D.H., Gregory, A., Girolami, M., Elshafie, M.Z.E.B., 2019. Introducing data-centric engineering to instrumented infrastructure. In: *International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making*, 2019, pp. 343–349. <https://doi.org/10.1680/icsic.64669.343>.
- Chadwick, A., Morfett, J., Borthwick, M., 2013. *Hydraulics in Civil and Environmental Engineering*, 5th Edition. CRC Press, London.
- Daniel, I., Ajami, N.K., Castelletti, A., Savic, D., Stewart, R.A., Cominola, A., 2023. A survey of water utilities' digital transformation: drivers, impacts, and enabling technologies. *Npj Clean Water* 6 (1), 1–9. <https://doi.org/10.1038/s41545-023-00265-7>.
- Egbemhenge, A.U., Ojeyemi, T., Iwuozor, K.O., Emenike, E.C., Ogunsanya, T.I., Anidiobi, S.U., Adeniyi, A.G., 2023. Revolutionizing water treatment, conservation, and management: harnessing the power of AI-driven ChatGPT solutions. *Environ. Challenges* 13 (October), 100782. <https://doi.org/10.1016/j.envc.2023.100782>.
- Eggmann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy De Vitry, M., Beutler, P., Maurer, M., 2017. The potential of knowing more: a review of data-driven urban water management. *Environ. Sci. Technol.* 51 (5), 2538–2553. <https://doi.org/10.1021/acs.est.6b04267>.
- Foroumandi, E., Moradkhani, H., Sanchez-vila, X., Singha, K., 2023. ChatGPT in hydrology and earth sciences : opportunities, prospects, and concerns. *Water Resour. Res.* 59, e2023WR036288 <https://doi.org/10.1029/2023WR036288>.
- Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler, D., 2022. The role of deep learning in urban water management: a critical review. *Water Res.* 223, 118973 <https://doi.org/10.1016/j.watres.2022.118973>.
- Fu, G., Sun, S., Hoang, L., Yuan, Z., Butler, D., 2023. Artificial intelligence underpins urban water infrastructure of the future: a holistic perspective. *Cambridge Prisms: Water* 1 (e14), 1–11. <https://doi.org/10.1017/wat.2023.15>.
- Girolami, M., 2021. What is data-centric engineering? DCEng Summit - YouTube. <https://www.youtube.com/watch?v=uFyHd4S20bY>.
- Hassabis, D., 2022. AlphaFold reveals the structure of the protein universe. <https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe> (accessed 28/12/2022).
- Hey, T., Tansley, S., Tolle, K., 2009. *The Fourth paradigm: Data-Intensive Scientific Discovery*. Microsoft Corporation, Redmond Washington.
- Jakubik, J., Roy, S., Phillips, C.E., Fraccaro, P., Godwin, D., 2023. Foundation models for generalist Geospatial artificial intelligence. *ArXiv*. <https://doi.org/10.1038/s41586-023-05881-4>.
- Jenkins, D., Wanner, J., 2014. *Activated Sludge –100 Years and Counting*. IWA Publishing, London.
- Ley, C., Tibolt, M., Fromme, D., 2020. Data-Centric Engineering in modern science from the perspective of a statistician, an engineer, and a software developer. *Data-Centric Engineering* 1, 1–10. <https://doi.org/10.1017/dce.2020.2>.
- Li, Z., Liu, H., Zhang, C., Fu, G., 2023. Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. *Environ. Sci. Ecotechnol.* 14, 100231 <https://doi.org/10.1016/j.ese.2022.100231>.
- Li, Z., Liu, H., Zhang, C., Fu, G., 2024. Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. *Water Res.* 250, 121018.
- Liu, B., Wang, N., Chen, M., Wu, X., Mo, D., Liu, J., Xu, S., Zhuang, Y., 2017. Earliest hydraulic enterprise in China, 5,100 years ago. In: *Proceedings of the National Academy of Sciences of the United States of America*, 114, pp. 13637–13642. <https://doi.org/10.1073/pnas.1710516114>.
- Liu, G., Savid, D., Fu, G., 2023. Short-term water demand forecasting using data-centric machine learning approaches. *J. Hydroinf.* 25 (3), 895–911. <https://doi.org/10.2166/hydro.2023.163>.
- Makropoulos, C., Savic, D.A., 2019. *Urban hydroinformatics: past, present and future*. *Water (Switzerland)* 11 (10), 1959.
- Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V., 2021. What role does hydrological science play in the age of machine learning? *Water Resour. Res.* 57 (3) <https://doi.org/10.1029/2020WR028091>.

- Richards, C.E., Tzachor, A., Avin, S., Fenner, R., 2023. Rewards, risks and responsible deployment of artificial intelligence in water systems. *Nat Water* 1, 422–432. <https://doi.org/10.1038/s44221-023-00069-6>.
- Ross, M., Taylor, J., 2021. Managing AI Decision-Making Tools. *Harv. Bus. Rev.* <https://hbr.org/2021/11/managing-ai-decision-making-tools> (accessed 02/12/2023).
- Russell, S., Dewey, D., Tegmark, M., 2015. Research priorities for robust and beneficial artificial intelligence. *AI Magazine* 36 (4), 105–114. <https://doi.org/10.1609/aimag.v36i4.2577>.
- Savić, D., 2022. Digital water developments and lessons learned from automation in the car and aircraft industries. *Engineering* 9, 35–41.
- Shen, C., Appling, A.P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fencia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C.J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Lawson, K., 2023. Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Rev. Earth and Environ.* 4 (8), 552–567. <https://doi.org/10.1038/s43017-023-00450-9>.
- Sweetapple, C., Fu, G., Butler, D., 2014. Multi-objective optimisation of wastewater treatment plant control to reduce greenhouse gas emissions. *Water Res.* 55, 52–62.
- UK National Cyber Security Centre, 2023. Guidelines For Secure AI System Development. <https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development> (accessed 13/2/2024).
- Vairavamoorthy, K., 2022. Tapping Into tradition: Water Insights from Down the centuries, *The Source* (thesourcemagazine.org). IWA (accessed 28/12/2022).
- Wagener, T., Savic, D., Butler, D., Ahmadian, R., Arnot, T., Dawes, J., Djordjevic, S., Falconer, R., Farmani, R., Ford, D., Hofman, J., 2021. Hydroinformatics education—the water informatics in science and engineering (WISE) centre for doctoral training. *Hydrol. Earth Syst. Sci.* 25 (5), 2721–2738.
- Walski, T.M., 2006. A history of water distribution. *Journal of the American Water Works Association* 98 (3), 110–121. <https://doi.org/10.1002/j.1551-8833.2006.tb07611.x>.
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., ..., Zhang, J., 2021. Artificial intelligence: a powerful paradigm for scientific research. *The Innovation* 2 (4), 100179. <https://doi.org/10.1016/j.xinn.2021.100179>.
- Zhang, L., Qin, H., Mao, J., Cao, X., Fu, G., 2023. High temporal resolution urban flood prediction using attention-based LSTM models. *J Hydrol (Amst)* 620, 129499. <https://doi.org/10.1016/j.jhydrol.2023.129499>.
- Zhang, S., Yi, Y., Liu, Y., Wang, X., 2013. Hydraulic Principles of the 2,268-Year-Old Dujiangyan Project in China. *J. Hydraul. Eng.* 139 (5), 538–546. [https://doi.org/10.1061/\(asce\)hy.1943-7900.0000675](https://doi.org/10.1061/(asce)hy.1943-7900.0000675).