25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

# How much data should be accumulated for reliable water pollution source identification? Critical time profile discovery and monitoring process design

**R.Y. Yang** [a], **J.P. Jiang** [a], **T.R. Pang** [b], **Y. Zheng** [a] **and Z.H. Yang** [c]

[a] *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China*
[b] *School of Environment, Harbin Institute of Technology, Harbin, China*
[c] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China*
*Email: 997078439@qq.com*

**Abstract:** Identifying hidden pollutant emissions sources is crucial for surface water quality management. The numerical inversion method, based on a pollutant transport model, is an effective approach for pollution source identification (PSI). Nevertheless, existing research tends to concentrate on the algorithm rather than its real-world application. For instance, many proposed methodologies rely on fixed water quality sensors deployed in key monitoring sections for data collection. In order to obtain reliable PSI results, complete pollutant Breakthrough Curve (BTC) were usually used as inputs to numerical PSI models. In some watersheds, this sampling process may last for more than ten hours, which clearly violates the requirement of timeliness.

Scientific problems arise accordingly. How much data should be accumulated for reliable PSI? How long does it take to obtain this data under different data collection conditions and river hydrodynamic conditions? How can the emergency monitoring process be improved in the context of the environmental Internet of Things? Understanding the above problems will help design the emergency monitoring process and ultimately improve the application performance of numerical source identification methods.

To this end, the study adopts the Adaptive Metropolis-Markov Chain Monte Carlo (AM-MCMC) Bayesian Inference method as the benchmark PSI framework. Subsequently, a one-dimensional instantaneous point source emission pollutant transport model was established based on the Advection-Diffusion-Reaction equation (ADR), and the model was calibrated using dye tracer test data from the United States Geological Survey (USGS) on the Truckee River for hypothetical case data generation. Based on the numerical experiments of hypothetical cases, the source inversion error level and uncertainty under different sampling cumulative time were evaluated. It was found that there is a critical moment of low error level and uncertainty convergence, after which the error slowly decreases and remains in a convergence state. Then, the basic characteristics of critical time were discussed through numerical experiments: increasing the number of monitoring sections will shorten the critical time; To a certain extent, the monitoring frequency and monitoring error level has little effect on the critical time. Furthermore, numerical experiments were carried out on the variation law of critical time under different river hydrodynamic conditions. The above results were proved based on the information entropy theory, and the universal equation of Peclet number (representing hydrodynamic conditions) and relative critical time ($\Lambda$) , namely the critical time profile, was fitted based on the information entropy analysis expression.

Finally, based on the critical time profile, a quantitative design method of emergency monitoring network was proposed, and the practical technical route of emergency PSI was discussed, which provides reference for the layout of monitoring section and the formulation of process in actual source identification operation, particularly for the watershed management in the era of Internet of Things.

***Keywords:*** *Source identification, numerical inversion, critical time, information entropy, monitoring network design*