



ALLDATA 2025

The Eleventh International Conference on Big Data, Small Data, Linked Data and
Open Data

ISBN: 978-1-68558-282-1

May 18 - 22, 2025

Nice, France

ALLDATA 2025 Editors

Alexey Cheptsov, TIK – Technische Informations und Kommunikationsdienste,
Stuttgart, Germany

ALLDATA 2025

Forward

The Eleventh International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2025), held between May 18-22, 2025 in Nice, France, continued a series of events bridging the concepts and the communities devoted to each of data categories for a better understanding of data semantics and their use, by taking advantage from the development of Semantic Web, Deep Web, Internet, non-SQL and SQL structures, progresses in data processing, and the new tendency for acceptance of open environments.

The volume and the complexity of available information overwhelm human and computing resources. Several approaches, technologies and tools are dealing with different types of data when searching, mining, learning and managing existing and increasingly growing information. From understanding Small data, the academia and industry recently embraced Big data, Linked data, and Open data. Each of these concepts carries specific foundations, algorithms and techniques, and is suitable and successful for different kinds of application. While approaching each concept from a silo point of view allows a better understanding (and potential optimization), no application or service can be developed without considering all data types mentioned above.

We welcomed academic, research and industry contributions. The conference had the following tracks:

- Big Data
- Open Data
- Linked Data
- Challenges in processing Big Data and applications

We take here the opportunity to warmly thank all the members of the ALLDATA 2025 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to ALLDATA 2025. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also thank the members of the ALLDATA 2025 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that ALLDATA 2025 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of all data. We also hope that Nice provided a pleasant environment during the conference and everyone saved some time to enjoy the historic charm of the city.

ALLDATA 2025 Chairs

ALLDATA Steering Committee

Anastasija Nikiforova, European Open Science Cloud Task Force "FAIR metrics and data quality" & University of Tartu, Estonia

Gerold Hoelzl, University of Passau, Germany

Vikas Thammanna Gowda, Champlain College, USA

ALLDATA Publicity Chairs

José Miguel Jiménez, Universitat Politècnica de Valencia, Spain

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain

Ali Ahmad, Universitat Politècnica de València, Spain

Laura Garcia, Universidad Politécnica de Cartagena, Spain

Sandra Viciano Tudela, Universitat Politècnica de Valencia, Spain

ALLDATA 2025

Committee

ALLDATA Steering Committee

Anastasija Nikiforova, European Open Science Cloud Task Force "FAIR metrics and data quality" & University of Tartu, Estonia
Gerold Hoelzl, University of Passau, Germany
Vikas Thammanna Gowda, Champlain College, USA

ALLDATA 2025 Publicity Chairs

José Miguel Jiménez, Universitat Politècnica de Valencia, Spain
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
Vikas Thammanna Gowda, Champlain College, USA
Laura García, Universidad Politécnica de Cartagena, Spain
Sandra Viciano Tudela, Universitat Politècnica de Valencia, Spain

ALLDATA 2025 Technical Program Committee

Alessandra Alaniz Macedo, University of São Paulo (USP), Brazil
Hugo Alatrasta-Salas, Universidad del Pacífico, Peru
Farah Alshanik, Clemson University, USA
Qiushi Bai, Microsoft (Azure Data R&D), USA
Syed Raza Bashir, Toronto Metropolitan University, Canada
Gábor Bella, University of Trento, Italy
Sandjai Bhulai, Vrije Universiteit Amsterdam, Netherlands
Ayan Biswas, Los Alamos National Laboratory, USA
Jean-Yves Blaise, CNRS (French National Centre for Scientific Research) | UMR CNRS/MC 3495 MAP, France
Doulkifli Boukraa, LaRIA Lab | University of Jijel, Algeria
Ozgu Can, Ege University, Turkey
Ramon Alberto Carrasco Gonzalez, Universidad Complutense de Madrid, Spain
Richard Chbeir, Université de Pau et des Pays de l'Adour (UPPA), France
Rachid Chelouah, Ecole Internationale des Sciences du Traitement de l'Information (*EISTI*), Cergy, France
Haihua Chen, University of North Texas, USA
Wei-Kuo Chiang, National Chung Cheng University, Taiwan
Esma Nur Cinicioglu, Istanbul University - School of Business, Turkey
Stefano Cirillo, University of Salerno, Italy
António Correia, University of Jyväskylä, Finland
Cinzia Daraio, Sapienza University of Rome, Italy
Subhasis Dasgupta, University of California San Diego, USA
Bidur Devkota, Asian Institute of Technology (AIT), Thailand
Christian Dirschl, Wolters Kluwer Deutschland GmbH, Germany

Gokila Dorai, School of Computer & Cyber Sciences - Augusta University, USA
Ricardo Eito Brun, Universidad Carlos III de Madrid, Spain
Mounim A. El Yacoubi, Telecom SudParis / Institut Mines Telecom / Institut Polytechnique de Paris, France
Mahmoud Elbattah, Université de Picardie Jules Verne, France
Denise Beatriz Ferrari, Instituto Tecnológico de Aeronáutica, São José dos Campos - SP, Brazil
Hacene Belhadef, University of Constantine 2, Abdelhamid Mehri, Algeria
Matteo Francia, University of Bologna, Italy
Munehiro Fukuda, University of Washington Bothell, USA
Chiara Gallese Nobile, Eindhoven University of Technology, Netherlands / Carlo Cattaneo University - LIUC, Italy
Thirusubramanian Ganesan, Cognizant Technology Solutions, Texas, USA
Fausto Pedro Garcia Marquez, University of Castilla-La Mancha, Spain
Raji Ghawi, Technical University of Munich, Germany
William F. Godoy, Oak Ridge National Laboratory, USA
Vikas Thammanna Gowda, Champlain College, USA
Piotr Grochowalski, University of Rzeszów, Poland
Jerzy Grzymala-Busse, University of Kansas, USA
Venkat N. Gudivada, East Carolina University, USA
Boujemaa Guerhazi, Toronto Metropolitan University, Canada
Yifan Guo, Case Western Reserve University, USA
Samrat Gupta, Indian Institute of Management Ahmedabad, India
Leila Hamdad, Ecole Nationale Supérieure en Informatique (ESI), Algeria
Qiwei Han, Nova School of Business & Economics, Portugal
Arvin Hekmati, University of Southern California, USA
Gerold Hoelzl, University of Passau, Germany
Tsan-sheng Hsu, Academia Sinica, Taiwan
Xin Huang, University of Maryland at Baltimore County, USA
Sayem Mohammad Imtiaz, Iowa State University, USA
Athraa Juhi, Al-Nahrain University, Iraq
Hanmin Jung, Korea Institute of Science and Technology Information, South Korea
Sokratis K. Katsikas, Norwegian University of Science and Technology, Norway
Eleni Kaldoudi, Democritus University of Thrace, Greece
Verena Kantere, National Technical University of Athens, Greece
Ashutosh Karna, UPC, Barcelona / HP Inc., Spain
Yasuko Kawahata, Rikkyo University, Japan
Majida Kazmi, NED University of Engineering and Technology, Pakistan
Rasib Khan, Northern Kentucky University, USA
Olivera Kotevska, Oak Ridge National Laboratory, USA
Boris Kovalerchuk, Central Washington University, USA
Shao Wei Lam, SingHealth, Singapore
Dominique Laurent, University of Cergy-Pontoise, France
Saïd Mahmoudi, University of Mons, Belgium
Sebastian Maneth, University of Bremen, Germany
Venugopal Mani, Walmart Global Tech, India
Felice Antonio Merra, Politecnico di Bari, Italy
Óscar Mortágua Pereira, University of Aveiro, Portugal
Fabrice Mourlin, Université Paris-Est Créteil Val de Marne, France

Hidemoto Nakada, National Institute of Advanced Industrial Science and Technology (AIST), Japan
Rodica Neamtu, Worcester Polytechnic Institute, USA
Dong Quan Ngoc Nguyen, University of Notre Dame, USA
Marko Niemelä, University of Jyväskylä, Finland
Anastasija Nikiforova, European Open Science Cloud Task Force "FAIR metrics and data quality" & University of Tartu, Estonia
Nikolay Nikolov, SINTEF Digital, Norway
Shin-ichi Ohnishi, Hokkai-Gakuen University, Japan
Nagham Osman, University College London (UCL), UK
Jisha Jose Panackal, Sacred Heart College, Kerala, India
Edivaldo Pastori Valentini, Federal Institute of Education, Science and Technology of São Paulo, IFSP - Catanduva, São Paulo, Brazil
João Pereira, Eindhoven University of Technology, Netherlands
Van Vung Pham, Sam Houston State University, USA
Elaheh Pourabbas, National Research Council (CNR), Italy
Livia Predoiu, University of Oxford, UK
Christian Prehofer, DENSO AUTOMOTIVE Deutschland GmbH / TU München, Germany
Stephane Puechmorel, ENAC, France
Ela Pustulka, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
Sophia Rahaman, Manipal Academy of Higher Education, UAE
Franck Ravat, IRIT - Université Toulouse Capitole, France
Ivan Rodero, Rutgers University, USA
Amine Roukh, University of Mons, Belgium
Peter Ruppel, CODE University of Applied Sciences, Berlin, Germany
David Sánchez, Universitat Rovira i Virgili, Spain
Jason Sawin, University of St. Thomas, St. Paul Minnesota, USA
Daniel Schneider, Federal University of Rio de Janeiro (UFRJ), Brazil
Monica M. L. Sebillio, University of Salerno, Italy
Florence Sèdes, IRIT - University Toulouse 3 Paul Sabatier, France
Ivan Miguel Serrano Pires, Polytechnic Institute of Viseu, Portugal
M. Omair Shafiq, Carleton University, Canada
Rahul Sharma, Ajay Kumar Garg Engineering College, Ghaziabad, India / Tallinn University of Technology, Estonia
Babak Maleki Shoja, Duke Energy, USA
Suzanne Shontz, University of Kansas, USA
Vyacheslav Sidelnik, St.Petersburg State University, Russia
Fernando Silva, Polytechnic University of Leiria, Portugal
Andrzej Skowron, Systems Research Institute - Polish Academy of Sciences / Digital Science and Technology Centre of UKSW, Poland
Hongyang Sun, University of Kansas, USA
Yingcheng Sun, Columbia University, USA
K. Suresh, Government Engineering College, India
Nasseh Tabrizi, East Carolina University, USA
George Tambouratzis, Institute for Language and Speech Processing, Athena, Greece
Farhan Tanvir, Oklahoma State University, USA
David Tormey, Institute of Technology Sligo, Ireland
Jorge Valverde-Rebaza, Visibilia, Brazil
Costas Vassilakis, University of the Peloponnese, Greece

Sirje Virkus, Tallinn University, Estonia
Marco Viviani, University of Milano-Bicocca, Italy
Haoxin Wang, Toyota Motor North America R&D InfoTech Labs, USA
Jieling Wu, Iwate University, Japan
Zijun Yao, University of Kansas, USA
M. Amin Yazdi, RWTH Aachen University, Germany
Feng Yu, Youngstown State University, USA
Xiong (Bill) Yu, Case Western Reserve University, USA
Jack (Yunpeng) Zhang, University of Houston, USA
Wenbin Zhang, Carnegie Mellon University, USA
Qiang Zhu, University of Michigan - Dearborn, USA
Souad Taleb Zouggar, Oran 2 University, Algeria

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Physics-Informed Neural Network Surrogate Models for River Stage Prediction <i>Maximilian Zoch, Edward Holmberg, Pujan Pokhrel, Ken Pathak, Steven Sloan, Kendall Niles, Jay Ratcliff, Maik Flanagan, Elias Ioup, Christian Guetl, and Mahdi Abdelguerfi</i>	1
---	---

Physics-Informed Neural Network Surrogate Models for River Stage Prediction

Maximilian Zoch

CoDiS-Lab ISDS
Graz Technical University of Technology
Graz, Austria
maximilian.zoch@tugraz.at

Edward Holmberg

Gulf States Center for Environmental Informatics
University of New Orleans
Louisiana, United States
eholmber@uno.edu

Pujan Pokhrel

Gulf States Center for Environmental Informatics
University of New Orleans
Louisiana, United States
ppokhre1@uno.edu

Ken Pathak

US Army Corps of Engineers
Vicksburg District
Mississippi, United States
ken.pathak@usace.army.mil

Steven Sloan

US Army Corps of Engineers
Vicksburg District
Mississippi, United States
steven.sloan@usace.army.mil

Kendall Niles

US Army Corps of Engineers
Vicksburg District
Mississippi, United States
kendall.niles@usace.army.mil

Jay Ratcliff

Gulf States Center for Environmental Informatics
University of New Orleans
Louisiana, United States
jratclif@uno.edu

Maik Flanagan

US Army Corps of Engineers
New Orleans District
Louisiana, United States
maik.c.flanagan@usace.army.mil

Elias Ioup

Center for Geospatial Sciences
Naval Research Laboratory
Mississippi, United States
elias.z.ioup.civ@us.navy.mil

Christian Guetl

CoDiS-Lab ISDS
Graz Technical University of Technology
Graz, Austria
c.guetl@tugraz.at

Mahdi Abdelguerfi

Gulf States Center for Environmental Informatics
University of New Orleans
Louisiana, United States
gulfsceidirector@uno.edu

Abstract—This work investigates the feasibility of using Physics-Informed Neural Networks (PINNs) as surrogate models for river stage prediction, aiming to reduce computational cost while maintaining predictive accuracy. We demonstrate PINNs successfully approximate Hydrologic Engineering Center’s River Analysis System (HEC-RAS) solutions, achieving strong predictive accuracy, despite some variation among river segments. By integrating the governing Saint-Venant equations into the learning process, the proposed PINN-based surrogate model enforces physical consistency and significantly improves computational efficiency compared to HEC-RAS. We evaluate the model’s performance in terms of accuracy and computational speed, demonstrating that it closely approximates HEC-RAS predictions while enabling real-time inference. These results highlight the potential of PINNs as effective surrogate models for single-river hydrodynamics, offering a promising alternative for computationally efficient river stage forecasting. Future work will explore techniques to enhance PINN training stability and robustness across a more generalized multi-river model.

Keywords—Physics-Informed Neural Networks; Surrogate Modeling; River Stage Prediction; HEC-RAS.

I. INTRODUCTION

Rivers and waterways play a critical role in sustaining agricultural, industrial, and urban infrastructure. Understanding river stage dynamics is essential for a wide range of applications, including crop irrigation, drinking water supply, drainage planning, and flood risk assessment [1, 2]. Accurate river stage prediction enables informed decision-making in these domains, with economic, environmental, and societal benefits. During extreme weather events such as hurricanes or heavy rainfall, the ability to make real-time predictions of river behavior is particularly crucial for flood forecasting and emergency response [4, 5].

Traditional hydrodynamic models such as the Hydrologic Engineering Center’s River Analysis System (HEC-RAS) provide high-fidelity simulations of river stage by solving the Saint-Venant equations [1, 2]. While these models are widely

used for flood risk analysis, they are computationally expensive, requiring extensive parameter calibration and fine spatial and temporal resolution. As a result, simulating future water levels can take hours or days, making real-time forecasting infeasible in rapidly evolving flood scenarios.[3].

To address this challenge, this work investigates the development of a *Physics-Informed surrogate model* for river stage prediction. Unlike purely data-driven approaches, *Physics-Informed neural networks (PINNs)* integrate the governing *Saint-Venant equations* into the learning process, enforcing physics-based constraints while improving model generalization beyond the training domain [6, 7]. This approach enables more *computationally efficient* predictions while maintaining *physical consistency*.

This study focuses on developing and evaluating a *single-river PINN-based surrogate model* that approximates HEC-RAS river stage predictions. The primary objectives are:

- **Assessing the accuracy and computational efficiency** of a Physics-Informed surrogate model trained on a single river.
- **Comparing PINN predictions to HEC-RAS outputs** to determine the feasibility of using surrogate modeling for river hydrodynamics.
- **Identifying challenges and limitations** in training PINNs for river stage prediction, establishing a foundation for future work in extending these models to multiple river systems.

By demonstrating the feasibility of Physics-Informed surrogate models for single-river applications, this work aims to provide a stepping stone for future research into broader hydrodynamic modeling frameworks.

The remainder of this paper is organized as follows. Section II describes relevant background and related work. Section III explains the methodology and model architecture. Section IV covers the experimental setup, while Section V presents the

benchmarking of HEC-RAS and the PINN surrogate. Section VI reports our results and discusses key findings. Finally, Section VII concludes the paper and suggests future research directions.

II. BACKGROUND AND RELATED WORK

A. HEC-RAS: A Computational Numerical River Model

HEC-RAS is developed by the U.S. Army Corps of Engineers (USACE), is an industry-standard numerical model for simulating open-channel flow [2]. HEC-RAS is widely used in flood forecasting, infrastructure planning, and water resource management [10].

At its core, HEC-RAS numerically solves the *Saint-Venant equations*, a system of *shallow water PDEs* that govern mass and momentum conservation in river channels [1]. The *1D Saint-Venant equations* are given by:

$$\frac{\partial A}{\partial t} + \frac{\partial(Au)}{\partial x} = 0, \quad (\text{Continuity}) \quad (1)$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + g \frac{\partial h}{\partial x} + S_f - S_0 = 0, \quad (\text{Momentum}) \quad (2)$$

where:

- $A(x, t)$ is the cross-sectional flow area,
- $u(x, t)$ is velocity,
- $h(x, t)$ is water surface elevation,
- S_f is the friction slope,
- S_0 is the bed slope.

In this formulation, x denotes the river mile along the flow direction, and t represents the day of the year.

1) HEC-RAS Inputs:

- **Geometric Data:** Cross-sectional profiles of riverbanks and channel bottoms, which are often stored in geospatial databases and used for hydrodynamic simulations [20, 22].
- **Boundary Conditions:** Flow rates, water levels, and upstream/downstream conditions.
- **Hydraulic Parameters:** Manning's roughness coefficients, channel slopes, and obstructions [11].

2) HEC-RAS Outputs:

- **Stage Predictions:** Water surface elevations over time.
- **Flow Predictions:** Velocity distribution across river stations.

While HEC-RAS provides high-fidelity results, it is *computationally expensive*, requiring iterative solvers and extensive parameter calibration [3]. This computational burden makes real-time forecasting impractical in flood response scenarios.

B. Surrogate Models in Computational Science

To mitigate computational costs, *surrogate models* approximate numerical solvers by learning the relationship between *input parameters* (e.g., river geometry, boundary conditions) and *output predictions* (e.g., water surface elevation) without directly solving PDEs. These models are trained on diverse input-output pairs from numerical simulations, enabling them to predict approximate solutions at significantly reduced computational cost [13].

Surrogate models have been successfully applied in *fluid dynamics*, *aerodynamics*, and *weather prediction*, demonstrating their ability to reduce the complexity of PDE-based simulations [14]. However, purely data-driven surrogate models, such as artificial neural networks (ANNs) and Gaussian processes, *lack physical consistency*, leading to poor generalization when applied to dynamic, unseen river conditions [15].

C. Physics-Informed Neural Networks (PINNs)

PINNs provide an alternative approach by *embedding governing physics equations into their training process*. Unlike traditional surrogate models that rely solely on input-output mappings, PINNs enforce *physical laws* (e.g., conservation of mass and momentum) as constraints in their loss function [6].

By minimizing residuals from PDEs during training, PINNs generate solutions that remain *consistent with known physics*, even in *data-scarce environments*. PINNs have demonstrated effectiveness in *computational fluid dynamics (CFD)*, hydrodynamic simulators, and geophysics [8, 9], but *their application in river stage prediction remains limited* [7].

D. Fourier Feature Encoding in Surrogate Models

Machine learning applications in fluid dynamics often employ *Fourier feature encoding* to improve the model's ability to learn fine-scale variations in spatial and temporal data [24]. Standard neural networks exhibit a spectral bias toward learning low-frequency functions [23], which can lead to poor generalization in high-variability physical systems.

Fourier feature encoding mitigates this bias by transforming input coordinates into a high-dimensional space:

$$\gamma(x) = [\cos(2\pi Bx), \sin(2\pi Bx)]^T \quad (3)$$

where B is a matrix of random Fourier base frequencies. This transformation enables neural networks to capture fine-grained variations in river stage predictions.

E. Research Gap and Motivation for Our Approach

Despite advancements in surrogate modeling and Physics-Informed learning, *PINNs have not been widely applied as surrogate models for river stage prediction*. Previous work on PINNs has primarily focused on *idealized fluid simulations* rather than real-world hydrodynamic systems governed by HEC-RAS data.

This work addresses these gaps by:

- Developing a *PINN-based surrogate model* for river stage prediction that enforces the *Saint-Venant equations* as physical constraints.

- Investigating the effectiveness of Fourier feature encoding in improving the model's ability to capture fine-scale variations in river flow.
- Evaluating whether *our Physics-Informed surrogate model* can achieve accuracy comparable to HEC-RAS while significantly reducing computational cost.

The following sections describe our proposed methodology in detail.

III. METHODOLOGY

A. Problem Formulation

The primary objective of this study is to develop a computationally efficient surrogate model for river *stage prediction*, denoted as $h(x, t)$, using a PINN. The model is trained on simulated river data from the HEC-RAS, a numerical solver developed by the USACE. HEC-RAS provides high-fidelity water surface elevation predictions by solving the *Saint-Venant equations*, but its computational complexity makes real-time forecasting infeasible. Our approach seeks to approximate the HEC-RAS stage predictions while significantly reducing inference time.

Given a river cross-section and boundary conditions, the proposed surrogate model minimizes:

$$\mathcal{F}(h, u, A) = 0, \quad (\text{Physics-Constrained Objective}) \quad (4)$$

where:

- $h(x, t)$ is the *water surface elevation*,
- $u(x, t)$ is the *flow velocity*,
- $A(x, t)$ is the *cross-sectional area*.

The function $\mathcal{F}$ represents the **Saint-Venant equations**, ensuring that predictions adhere to known physical constraints.

B. Surrogate Model Architecture

The surrogate model consists of a deep neural network that approximates HEC-RAS river stage predictions while incorporating Physics-Informed regularization. The architecture follows a supervised learning approach with additional physics-based constraints to ensure compliance with governing hydrodynamic equations. As shown in Figure 1, the overall pipeline takes spatial and temporal inputs and outputs physically consistent predictions of stage and velocity.

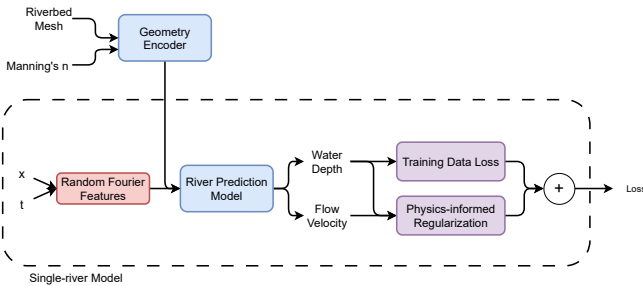


Figure 1. System overview of the single-river PINN surrogate model.

1) *Fourier Feature Encoding*: Neural networks typically exhibit a bias toward learning low-frequency functions [23]. To mitigate this and improve fine-scale resolution, **Fourier feature encoding** is applied to the input coordinates:

$$\gamma(x) = [\cos(2\pi Bx), \sin(2\pi Bx)]^T \quad (5)$$

where B is a matrix of random Fourier base frequencies [24]. This transformation enhances the model's ability to capture complex variations in river stage across time and space. The standard deviation σ of the Fourier base matrix B is optimized through grid search over values $\{0.5, 1.0, 2.0, 4.0, 8.0\}$ to balance spectral bias and overfitting. A value of $\sigma = 4.0$ was selected based on validation loss and qualitative error reduction in river stage predictions.

2) *Neural Network Architecture*: The surrogate model adopts an implicit neural representation, mapping spatial and temporal inputs (x, t) to predicted river stage $h(x, t)$ and flow velocity $u(x, t)$. The architecture is structured as follows:

- **Input Layer**: Encodes river mile x and time t using Fourier features.
- **Hidden Layers**: 6 fully connected residual blocks with 512 hidden dimensions per layer.
- **Output Layer**: Predicts $h(x, t)$ (water depth) and $u(x, t)$ (flow velocity).

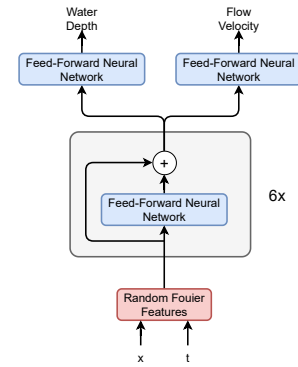


Figure 2. Structure of the single-river surrogate model.

Residual connections are used to improve training stability and convergence [25]. Figure 2 depicts the architecture of the single-river model.

C. Loss Function

The surrogate model is trained using a hybrid loss function combining supervised learning and Physics-Informed regularization:

$$\mathcal{L} = \mathcal{L}_{\text{HEC-RAS}} + \lambda \mathcal{L}_{\text{physics}}, \quad (6)$$

where:

- $\mathcal{L}_{\text{HEC-RAS}}$ is the *supervised loss*, measuring deviation from HEC-RAS-generated stage predictions.

- $\mathcal{L}_{\text{physics}}$ enforces compliance with the *Saint-Venant equations*, ensuring physically valid water surface elevations.

The Physics-Informed term is derived from the continuity and momentum equations:

$$\mathcal{L}_{\text{physics}} = \left\| \frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} \right\|^2 + \left\| \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + g \frac{\partial h}{\partial x} \right\|^2. \quad (7)$$

These residual terms are computed using automatic differentiation in PyTorch.

IV. EXPERIMENTAL SETUP

A. Dataset Description

The dataset used in this study represents segments of the Mississippi River. Each segment is treated as a single-river model. In total, 63 river segments were analyzed, containing 3,240 strategically placed river stations used for high-fidelity hydrodynamic simulation. These stations were selected by expert modelers to capture critical variations in river stage and flow while minimizing redundancy. On average, stations are spaced approximately 0.74 miles apart, and each contains time-series data generated for 1D unsteady flow analysis.

TABLE I. DATASET OVERVIEW FOR BENCHMARKING STUDY

Parameter	Value
River System	Mississippi River
River Segments	63
Total River Stations	3,240
Average Station Spacing	0.74 miles

Station Features: Each river station records key hydrodynamic and geometric attributes, as summarized in Table II.

TABLE II. KEY ATTRIBUTES RECORDED AT EACH RIVER STATION

Feature	Description
Water Surface Elevation	Height of water column at a station.
Water Discharge	Volume of water flowing per unit time.
Geometric Information	Cross-sectional profiles, bed elevation, and channel width.

B. Baseline Model for Comparison

To evaluate the effectiveness of the surrogate model, we compare it against HEC-RAS, which serves as the ground truth for single-river model validation. HEC-RAS is the industry-standard numerical solver for simulating 1D open-channel flow using the Saint-Venant equations and provides high-fidelity predictions of water surface elevation.

This comparison focuses on two key aspects:

- **Predictive Accuracy:** How closely the surrogate model approximates HEC-RAS stage predictions.
- **Computational Efficiency:** The reduction in inference time when using the PINN-based model.

C. Training Details

The training follows a hybrid loss framework combining supervised learning with Physics-Informed constraints.

TABLE III. TRAINING CONFIGURATION FOR SURROGATE MODEL

Parameter	Description
Network Architecture	6 hidden layers, 512 neurons/layer, ReLU activation
Optimizer	Adam with learning rate 10^{-3} and exponential decay
Batch Size	1024 samples per iteration
Loss Weighting	Optimized via grid search
Training Duration	100,000 iterations

V. BENCHMARKING

To assess computational efficiency, we compared the execution time of HEC-RAS and our PINN-based surrogate model on the same machine (Intel Xeon E5-1620, 32GB RAM, NVIDIA GTX 970). The surrogate model was implemented using PyTorch and ran on GPU; HEC-RAS was executed as a CPU-based process using version 5.0.1. Table IV shows the results for a full 1D unsteady flow simulation across all river segments.

TABLE IV. TOTAL EXECUTION TIME COMPARISON (SECONDS)

Model	Total Time (sec)
HEC-RAS	8317
PINN Surrogate	82.9

The PINN model achieved a 100× speedup while maintaining predictive accuracy. These results underscore the practicality of Physics-Informed surrogates for near real-time river forecasting. While HEC-RAS is limited by CPU-bound solvers, the PINN model leverages GPU acceleration, enabling faster inference and potential scalability for broader deployment.

VI. RESULTS AND DISCUSSION

A. Evaluation Metrics

To evaluate the performance of the single-river PINN model, we consider three key metrics: predictive accuracy, physical consistency, and computational efficiency. Each metric is described below.

a) Mean Relative Absolute Error (MRAE): This metric measures the average deviation of predicted river stage values from the ground truth, normalized by the magnitude of the ground truth values. It provides a scale-invariant assessment of accuracy across rivers with varying stage magnitudes.

$$\text{MRAE} = \frac{\frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|}{\frac{1}{N} \sum_{i=1}^N y_i} \quad (8)$$

where $\hat{y}_i$ is the predicted stage at time i , y_i is the corresponding HEC-RAS value, and N is the total number of predictions.

b) Physics Residual Loss: We assess adherence to physical laws by evaluating residuals of the Saint-Venant equations (Equation 7). This ensures physically realistic predictions, even with limited data.

c) Inference Time: We also measure total runtime for full-river inference on GPU hardware. This metric quantifies the practical efficiency of the surrogate model for real-time or near real-time forecasting.

B. Single-River Model Evaluation

Single-river models are evaluated independently to assess how accurately the PINN surrogate approximates HEC-RAS predictions without retraining.

The model architecture and training configuration were tuned using the Arkansas River segment. Therefore, the results on the remaining 62 rivers serve as a strong evaluation of the model's generalization ability under fixed hyperparameters.

Across all river stations, the model achieves a **low mean relative absolute error (MRAE)**, indicating strong predictive accuracy. Some rivers, such as the Tensas River, exhibit particularly low errors, while others, such as the Arkansas River, show slightly increased deviations.

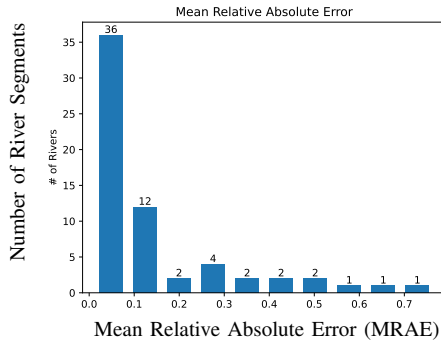


Figure 3. Histogram of relative error scores across river stations for the single-river model.

Figure 3 depicts a histogram of the relative errors across different river stations. The majority of stations maintain low error rates, demonstrating the model's ability to accurately capture river dynamics. The previously defined MRAE metric is used to evaluate stage prediction accuracy across river segments.

In addition to the histogram in Figure 3, Figure 4 provides a more detailed view of per-river model performance. Each boxplot summarizes the distribution of absolute prediction errors over time for an individual river segment. This breakdown highlights variability in model accuracy across river systems. Rivers such as the Tensas River and Yazoo River segments exhibit consistently low errors, while others like the Arkansas River and Forked Deer segments show higher or more variable errors.

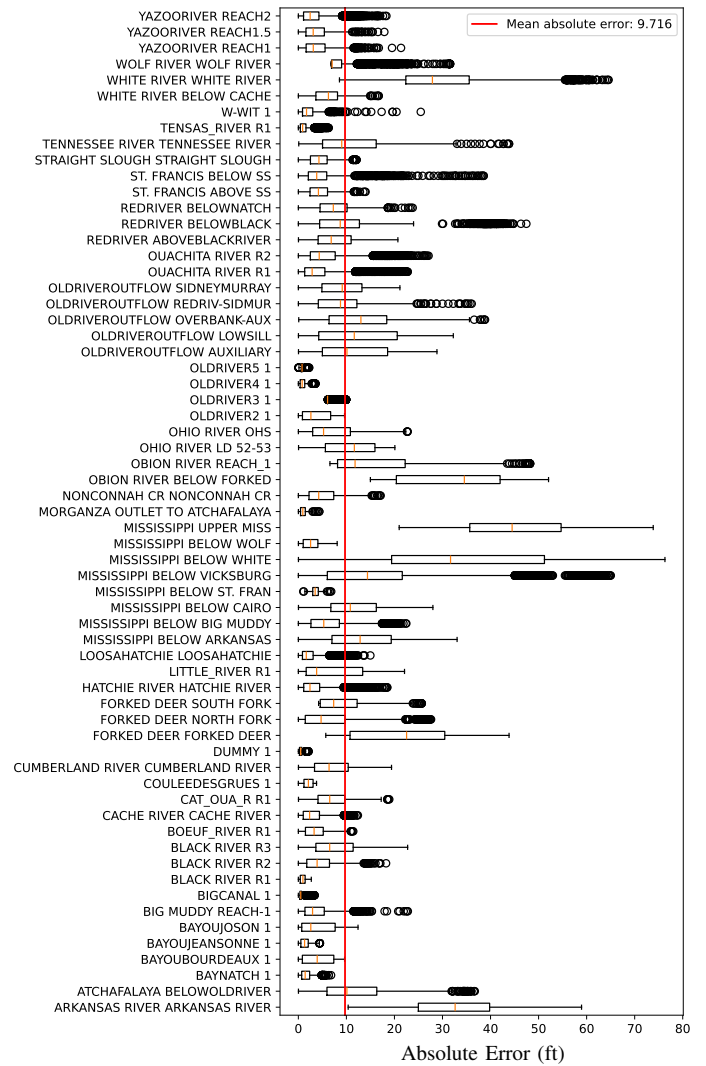


Figure 4. Distribution of absolute prediction errors for each river segment. Each boxplot summarizes model performance over time for a single river, with the red vertical line denoting the overall mean absolute error (9.716 ft).

C. Global Evaluation Across All Rivers

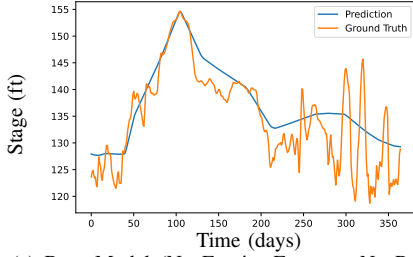
We compute a global normalized error based on the full dataset stage range (-1.11 to 380.56 ft; total span 381.67 ft). The model's mean absolute error of 9.716 ft corresponds to a normalized error of only 2.55% , indicating predictions remain within physically plausible bounds (see Table V).

TABLE V. GLOBAL NORMALIZED ERROR SUMMARY

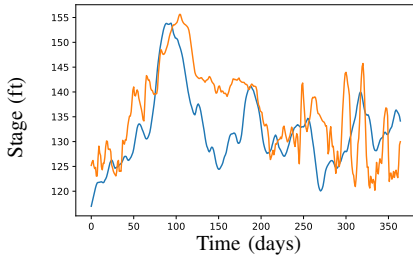
Metric	Value
Global Minimum Stage	-1.11 ft
Global Maximum Stage	380.56 ft
Mean Absolute Error (MAE)	9.716 ft
Global Normalized Error	2.55%

D. Ablation Study

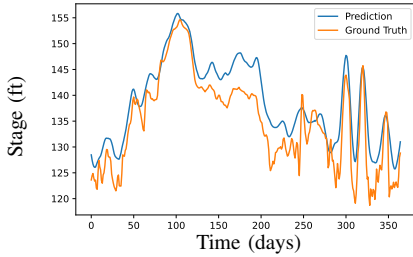
To evaluate the contribution of key model components, we conducted an ablation study analyzing the effects of Fourier feature encoding and Physics-Informed regularization. The results are presented in Figure 5, which compares the predicted water surface elevation (blue) against the HEC-RAS ground truth (orange) across a representative river segment over a full annual cycle (x-axis: time in days, y-axis: water surface elevation in feet).



(a) Base Model (No Fourier Features, No Regularization)



(b) With Random Fourier Features



(c) With Physics-Informed Regularization

Figure 5. Ablation study results demonstrating the impact of Fourier features and Physics-Informed regularization.

In Figure 5(a), the base model without Fourier features or regularization fails to capture the fine-scale variability present in the ground truth data. The predictions are overly smooth, and rapid changes in stage are poorly approximated, highlighting the model's spectral bias toward low-frequency components.

Figure 5(b) shows the effect of adding random Fourier feature encoding. The model's predictions more closely follow high-frequency fluctuations in the river stage, significantly improving local resolution. However, the inclusion of these features introduces overfitting in some regions, particularly when the model lacks regularization.

Finally, Figure 5(c) demonstrates the combined effect of Fourier features and Physics-Informed regularization. This version of the model achieves the best alignment with the ground truth, maintaining both high-frequency resolution and global physical consistency. These results confirm that Physics-Informed regularization is essential to prevent overfitting and enforce hydrodynamic realism when using expressive encodings such as Fourier features.

E. Findings and Discussion

The evaluation of the single-river PINN models reveals several key findings:

First, the PINN-based surrogate models closely approximate HEC-RAS stage predictions, with strong accuracy observed across most river stations. As shown in Figure 3, the distribution of Mean Relative Absolute Error (MRAE) is heavily skewed toward lower values. Specifically, 36 out of 61 river segments (59%) achieve an MRAE below 0.1, and nearly 80% fall below 0.2. This indicates that the majority of rivers are predicted with high fidelity using the surrogate model.

TABLE VI. SUMMARY STATISTICS OF MRAE ACROSS RIVER SEGMENTS

Metric	Mean MRAE	Median MRAE	Mode MRAE
Value	~0.15	< 0.1	0.0–0.1

Using multiple summary statistics helps capture this skewed distribution: while the mean MRAE is around 0.15, the median and mode indicate that most models perform substantially better. This supports the claim that the PINN surrogate is highly accurate for most rivers, even if a few outlier cases raise the overall average.

Second, the use of Fourier feature encoding improves the model's capacity to resolve fine-grained temporal dynamics in water surface elevation. However, without Physics-Informed regularization, this expressiveness introduces overfitting, particularly in regions with high variability. The ablation study confirms that combining Fourier encoding with Physics-Informed loss yields the best trade-off between local resolution and global physical consistency.

Finally, the proposed PINN framework provides significant computational advantages. As demonstrated in Section V, the surrogate model executes predictions approximately 100× faster than HEC-RAS, reducing runtime from over two hours to under 90 seconds. Importantly, this speedup is achieved without sacrificing accuracy: the model maintains a mean MRAE of approximately 0.15, which is acceptable for many practical forecasting applications. These results highlight the promise of PINNs for enabling fast, physically grounded river stage prediction in real-world decision-support systems.

VII. CONCLUSION

This study demonstrates that single-river PINNs effectively serve as computationally efficient surrogate models for river stage prediction, offering substantial speed improvements over HEC-RAS without sacrificing accuracy. Future work should

generalize the approach across multiple rivers and enhance accuracy via ensemble methods.

A. Toward a Generalized Multi-River Model

Extending PINNs to multiple rivers without retraining remains a key challenge. Two promising strategies include:

- **Geometry Encoding:** Robust cross-sectional or graph-based encodings to capture diverse river geometries [21].
- **Ensemble Robustness:** Weighted fusion of independently trained PINNs, with weights proportional to validation accuracy, to enhance predictive robustness.

B. Future Directions

Additional research should focus on:

- **Scaling to 2D/3D Hydrodynamics:** Extending PINNs to multi-dimensional flow modeling.
- **Real-World Validation:** Refining accuracy through comparisons with observed river data.
- **Adaptive Loss Weighting:** Balancing Physics-Informed and data-driven constraints to improve stability.

Advancing these areas will help PINNs evolve into scalable, physically consistent, and computationally efficient tools for real-time hydrodynamic forecasting.

REFERENCES

- [1] H. Chanson, *Hydraulics of Open Channel Flow*, Butterworth-Heinemann, 2004.
- [2] USACE Hydrologic Engineering Center, *HEC-RAS Hydraulic Reference Manual*, U.S. Army Corps of Engineers, 2024.
- [3] A. Malcherek, *Fliessgewässer: Hydraulik, Hydrologie, Morphologie und Wasserbau*, Springer Vieweg, 2019.
- [4] Intergovernmental Panel on Climate Change (IPCC), *Summary for Policymakers*, Contribution to the Sixth Assessment Report, 2022.
- [5] J. Henley, "Death toll reaches 16 as 'dramatic' flooding in central Europe continues," *The Guardian*, 2024.
- [6] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [7] D. Feng, Z. Tan, and Q. He, "Physics-informed neural networks of the Saint-Venant equations for downscaling a large-scale river model," *Water Resources Research*, vol. 59, no. 5, 2023.
- [8] X.-D. Bai, Y. Wang, and W. Zhang, "Applying Physics-Informed neural networks for flow data assimilation," *Journal of Hydrodynamics*, vol. 32, no. 5, pp. 943–954, 2020.
- [9] J. Donnelly, A. Daneshkhah, and S. Abolfathi, "Physics-informed neural networks as surrogate models of hydrodynamic simulators," *Science of The Total Environment*, vol. 904, p. 166945, 2024.
- [10] USACE Hydrologic Engineering Center, *HEC-RAS Hydraulics User's Manual*, U.S. Army Corps of Engineers, 2024.
- [11] J. Flanagan, R. David, and M. Young, "Application of Manning's roughness coefficients in river modeling," *Journal of Hydraulic Engineering*, vol. 133, no. 9, pp. 1021–1033, 2007.
- [12] USACE Hydrologic Engineering Center, *HEC-RAS Floodplain Management and Mapping*, U.S. Army Corps of Engineers, 2024.
- [13] P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Review*, vol. 57, no. 4, pp. 483–531, 2015.
- [14] H. Bi, J. Lin, and X. Zhao, "Data-driven surrogate modeling in computational fluid dynamics: A review," *Computers and Fluids*, vol. 258, p. 104764, 2023.
- [15] Z. Takkiri-Borujeni, J. T. Smith, and M. D. White, "Challenges of using surrogate models for environmental simulations: A case study in hydrology," *Environmental Modelling and Software*, vol. 134, p. 104882, 2020.
- [16] A. Sanchez-Gonzalez, V. Heess, and J. T. Springenberg, "Learning physics with graph networks," *Nature Communications*, vol. 11, p. 3223, 2020.
- [17] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, 2020.
- [18] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660.
- [19] Z. Wu, L. Hu, and X. Guo, "Graph-based deep learning for modeling spatial-temporal dynamics in hydrological systems," *Water Resources Research*, vol. 57, no. 4, 2021.
- [20] M. Flanagan, A. Grenotton, J. Ratcliff, K. B. Shaw, J. Sample, and M. Abdelguerfi, "Hydraulic splines: A hybrid approach to modeling river channel geometries," *Computing in Science and Engineering*, vol. 9, no. 5, pp. 4–15, 2007, IEEE Computer Society. doi: 10.1109/MCSE.2007.98.
- [21] M. Abdelguerfi, *3D Synthetic Environment Reconstruction*, vol. 611, Springer Science and Business Media, 2012. doi: 10.1007/978-1-4614-1803-2.
- [22] M. Chung, R. Wilson, R. Ladner, T. Lovitt, M. Cobb, M. Abdelguerfi, and K. Shaw, "The geospatial information distribution system (GIDS)," in *Succeeding with Object Databases*, John Wiley and Sons, 2001, pp. 357–378.
- [23] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the Spectral Bias of Neural Networks," *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, pp. 5301–5310, 2019. Available: <http://proceedings.mlr.press/v97/rahaman19a.html>
- [24] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in *European Conference on Computer Vision (ECCV)*, Springer, 2020, pp. 405–421. doi: 10.1007/978-3-030-58452-8_24.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 770–778. doi: 10.1109/CVPR.2016.90.