CASTLE: A Cascaded Spatio-Temporal Approach for Long-lead Streamflow Forecasting

Yong Zhuang[1], David L. Small[4], Patrick D Flynn[3], Wahid Palash[5], Shafiqul Islam[2], Ping Chen[3], Wei Ding[3]
[1]*School of Computing*, *Grand Valley State University*, Allendale, MI, USA
[2]*Department of Civil and Environmental Engineering*, *Tufts University*, Boston, MA, USA
[3]*Department of Computer Science*, *University of Massachusetts Boston*, Boston, MA, USA
[4]*Verisk Extreme Event Solutions*, *Lafayette City Center*, Boston, MA, USA
[5]*Integrated Sustainability*, 1600-400 3rd Avenue Southwest, Calgary, AB T2P 4H2, Canada
yong.zhuang@gvsu.edu
{patrick.flynn003, ping.chen, wei.ding }@umb.edu
Shafiqul.Islam@tufts.edu
wahid.palash@integratedsustainability.ca
bigdavesmall@gmail.com

***Abstract*—Effective early warning systems for extreme flood events in large river basins necessitate reliable long-lead streamflow forecasts. However, the inherent uncertainty within each phase of the weather system—rainfall prediction, runoff generation, and streamflow prediction—amplifies with each stage, rendering accurate long-lead streamflow estimations challenging. In response to this, our study introduces a novel deep-learning-based model, the Cascaded Spatio-Temporal Learning Deep Network (CASTLE). CASTLE synergistically integrates observed upstream precipitation, recent streamflow data, and short-term precipitation forecasts derived from a selection of quantitative climate models to produce an accurate streamflow estimate.**

Specifically, we employ deep residual architectures on both observed and forecasted precipitation data to model the cascading spatio-temporal processes, which begin with upstream rainfall, move to rainfall-runoff, and finally conclude with downstream discharge. Our aim is to identify hidden space-time patterns that can be used to forecast future downstream flow over extended periods. We assess CASTLE's efficacy by forecasting the downstream discharge of the Ganges River over a long lead time. Results show that our approach outperforms the current state-of-the-art streamflow forecasting models.

***Index Terms*—Spatio-temporal climate data, Climate data mining, Long-lead streamflow prediction**

I. Introduction

Flooding is the most lethal and expensive natural disaster in the developing world. According to the World Resources Institute (WRI), approximately 80% of the population exposed to river flood risk worldwide resides in just 15 countries, all of which are in stages of development or underdevelopment. Among these, Bangladesh stands out, suffering significant human and material losses due to flooding. As 80% of its terrain lies within a low-lying flood plain, and with an average of 844,000 million cubic meters of upstream water flooding the country during the monsoon season (May through September)[1], the nation is particularly susceptible to extreme flood events. Establishing a reliable flood warning system with a long lead time could dramatically alleviate the impact of these disasters. The potential benefits of such a system are considerable, potentially equating to an annual fiscal advantage of around $100 billion.[2].

Fig. 1: Ganges Basin and Hardinge Bridge

Nevertheless, generating accurate flood forecasts, particularly with a lead time of 5-10 days crucial for effective flood preparedness, presents notable challenges. This is particularly true for flood-prone regions like Bangladesh. The development of short (3-5 days) to mid-range (7-10 days) flood forecasts that are both reliable and accurate is hindered by a range of significant obstacles.

A. Challenges in Long Lead Time Flood Forecasting

1) Low Skill Precipitation Forecasting: Despite recent advancements in the resolution and accuracy of weather forecasting models, which have spurred discussions on extending streamflow forecasts to longer lead times using model-based quantitative precipitation forecasts[3], improvements in precipitation prediction haven't kept pace with advancements in operational numerical weather prediction models for large-scale circulation patterns[4]. Precipitation forecast models are often found to underestimate precipitation totals even over short lead times (1 - 3 days) [5]. Their predictive capabilities over longer lead times (5–10 days) are primarily confined to the occurrence, not the magnitude, of precipitation[6]. ensemble forecasts of precipitation is a advanced model, which can extend magnitude of precipitation predictions beyond 48 hours, a forecast lead time of 5–10 days remains limited[7].

2) High Uncertainty: Due to the chaotic nature of the weather system and the poor quality of quantitative forecasts of precipitation, uncertainty cascades from rainfall forecasts through runoff generation to flood wave forecasts and is amplified at each stage. This amplification is due to the non-linearity of the governing dynamics of each subsequent model, resulting in non-quantifiable uncertainty within hydrological models[8], which represents a significant challenge in producing reliable flood forecasts with longer lead times.[9].

3) Missing Geological Attributes: Reliable rainfall-runoff-flow process modeling is able to capture the relationships between upstream rainfall and the changes of the downstream flow. However, because the soil and geological properties of catchments are always unknown at the scale needed to model the relevant dynamics, hydrological models are limited in their ability to capture local characteristics of the rainfall-runoff process. Without observations of the small spatial scale heterogeneity of the local soils, vegetation and geology, estimation of parameters during calibration of distributed hydrological models is an ill-posed mathematical problem[8]. Similar parameter estimation concerns are also applicable to hydraulic models for which the grid scales are larger relative to the scales of momentum dissipation, turbulence and secondary currents induced by local heterogeneities in channel and flood plain geometry [10]. This makes rainfall-runoff-flow process modeling extremely hard.

4) Limited Data Availability: The data from upstream basin areas is the key to accurately predict the downstream flow. However, the limited availability of data in the upstream basin places a fundamental limitation on accurate long-lead flood forecasting in downstream areas. Take Bangladesh as an example, more than 90% of the upstream river basin is located outside Bangladesh and it generates about 80% of the flood season (June through September) flow in Bangladesh [11]. Overcoming this limitation imposed by upstream data availability is a big problem to solve.

B. Overcoming Challenges with Deep Learning and Physics-Guided Approaches

Deep learning has made significant strides in diverse fields such as natural language processing, video analysis, and time series forecasting. Its success underscores its ability to understand complex relationships that are non-linear in nature. However, when it comes to forecasting over a longer duration in hydrological studies, the inherent uncertainties across extended periods pose unique challenges. Because domain-specific knowledge is essential to navigate these uncertainties, leaving the fusion of deep learning with hydrological expertise a largely unexplored territory.

In response to these challenges, our study presents a forecasting framework rooted in deep learning. This framework merges forecasts from hydrological models with the prowess of deep learning. By doing so, it seamlessly combines ensemble numerical weather forecasts, observational data, and the capabilities of deep learning to deliver reliable streamflow forecasts for downstream areas over extended durations, such as 5 to 10 days. Consequently, the primary contributions of our research are:

- **Cascade Modeling of Spatio-Temporal Events:** We put forth the concept of a cascading Spatio-Temporal Events Chain. To model the sequential connections present in these events, we employ sophisticated deep learning methods.
- **CASTLE:** We propose CASTLE, a cascaded spatio-temporal deep learning network. This innovative architecture fuses observational data with ensemble numerical weather predictions, enabling an extension of the forecasting lead time.
- **Empirical Validation in a Real-world Scenario:** We apply CASTLE to a practical use case, studying historical streamflow data gathered from the Hardinge Bridge station on the Ganges (See Fig. 1). Our empirical analyses demonstrate that CASTLE significantly outperforms existing state-of-the-art methods in long-lead streamflow forecasting.

The structure of this paper is organized as follows: Section II delves into a review of relevant literature. In Section III, we introduce the concept of the Spatio-Temporal events chain, with a comprehensive discussion on its modeling presented in Section IV. Our empirical studies conducted on a real-world precipitation data set, including a comparison with leading long-lead precipitation prediction models, are detailed in Section V. We encapsulate our findings and conclusions in Section VI.

II. Related Work

Streamflow forecasting has paramount importance in the region of Bangladesh, which, as the most downstream riparian country of the GBM river system, faces annually recurring flooding events. Approximately one-fifth of Bangladesh's area is inundated by flood water every year, and up to two-thirds is submerged during extreme events [12]. A notable challenge for effective flood forecasting in Bangladesh is the limited data availability from upstream basin areas in India, which restricts the country's capacity to produce and disseminate skilled flood forecasts of 5-10 days lead time. Over 90% of the GBM drainage areas lie outside of Bangladesh, contributing around 80% of the flood season flow inside Bangladesh [13].

A. Traditional Models

Short-term streamflow forecasting is traditionally approached through time series forecasting models. Among these, the auto-regressive moving average (ARMA) and the auto-regressive integrated moving average (ARIMA) [14] are prominent. However, their efficacy hinges on the series being stationary, implying that inherent seasonality or trends must be isolated beforehand. To cater to seasonal data, variants such as the seasonal ARIMA (SARIMA) and SARIMA with exogenous regressors (SARIMAX) have been developed. While adept at handling monthly seasonal data, their performance dwindles for more extended timescales and larger datasets.

B. Hydrological Models

In response to the data limitation issue, the Flood Forecasting and Warning Centre (FFWC) of Bangladesh has been utilizing a numerical one-dimensional hydrodynamic model since 1992 [15]. The model includes 38 upstream boundary condition points, three of which are of utmost importance. However, the FFWC model's success is limited by its high computational time and difficulty in disseminating ensemble forecasts for operational purposes. [7] investigates the potential of a simplified flood forecasting model framework, focusing on pinpointing the crucial hydrological variables and processes inherent to a basin. The study demonstrates that certain simplified models, which capitalize on vital components such as flow persistence, aggregated upstream rainfall, and basin travel time, can produce flood forecasts that rival the accuracy of more complex methodologies. Such an approach underscores the viability of these elements for enhancing the accuracy and reliability of long-lead streamflow forecasting.

C. Data-based Modeling

Data-based models have been proposed as potential alternatives to the hydrological models. These models follow an inductive approach, allowing the data to suggest an appropriate model structure [16]. Young and Beven [17] have shown that a data-based mechanistic modeling approach can provide sufficient and reasonable explanations of the system behavior. This approach could potentially decompose complex nonlinear natural processes into several serial, parallel, or feedback connections of simple processes. However, it is important to account for the associated variable and parameter uncertainties and to use an adaptive mechanism to train the model with multiple time series dataset for a robust model equation [18].

D. Machine Learning Models

Machine learning models, with a significant emphasis on deep learning, have recently risen to prominence in hydrological flood forecasting [19]. Artificial Neural Networks (ANNs) are favored for their adeptness at modeling the non-linearities inherent in chaotic climate systems [20]. In the realm of sequence and temporal processes, Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs) have distinguished themselves [21], [22]. These architectures excel in capturing the temporal dependencies and nuances present in sequential data. On the other hand, Convolutional Neural Networks (CNNs) shine in the analysis of spatial and spatio-temporal patterns, adeptly detecting relationships between features across varied scales and complexities [23]. Interestingly, analogous forecasting methodologies have found utility in the financial sector, particularly in stock price prediction—a domain sharing attributes like non-linearity and chaos with streamflow forecasting [24].

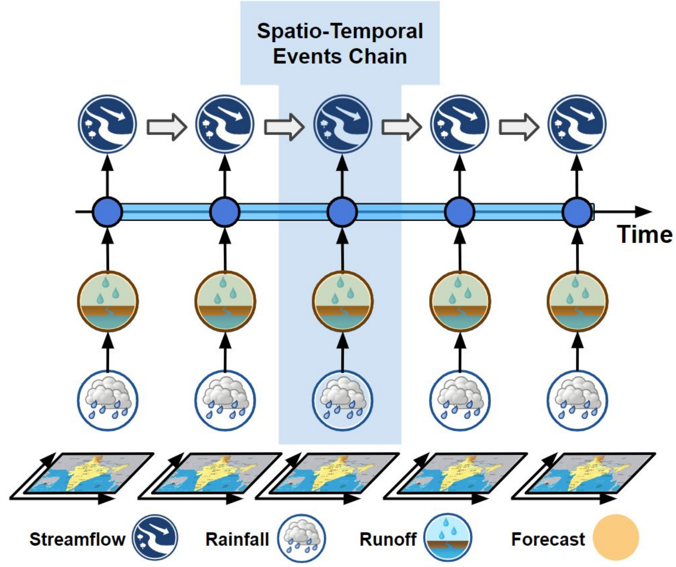

Fig. 2: Spatio-Temporal events chain for Long-Lead Streamflow Predictive Modeling: Streamflow is depicted as a progression through a three-step Spatio-Temporal Events Chain (STEC): $rainfall \triangleright runoff \triangleright flow\text{-}increment$

E. Novelty of Our Approach

In this work, we synthesize multiple approaches to present a novel paradigm: the Cascading Spatio-Temporal Events Modeling. Through the deployment of an advanced deep network architecture named CASTLE, we aim to surmount the constraints observed in current models, thereby enhancing both the reliability and accuracy of long-lead streamflow forecasts.

III. Spatio-Temporal Data Modeling and Problem Formulation

Modeling spatio-temporal (ST) data requires a unique approach due to its inherent properties, such as spatial and temporal correlations, nonlinearity, continuity, and partial order. In this section, we introduce the concept of a Spatio-Temporal Events Chain (STEC) to model these unique data characteristics and provide a formulation for long-lead streamflow forecasting.

A. Modelling Spatio-Temporal Events

ST data can be represented in terms of events and processes. While processes represent continuous phenomena that evolve over time, like urbanization or global climate change, events are more instantaneous, presenting punctuated occurrences like storms or festivals. These events, especially when influenced by preceding ones, dictate the trajectory of the encompassing process. To capture this interconnected progression of events, we put forth the concept of the Spatio-Temporal Events Chain.

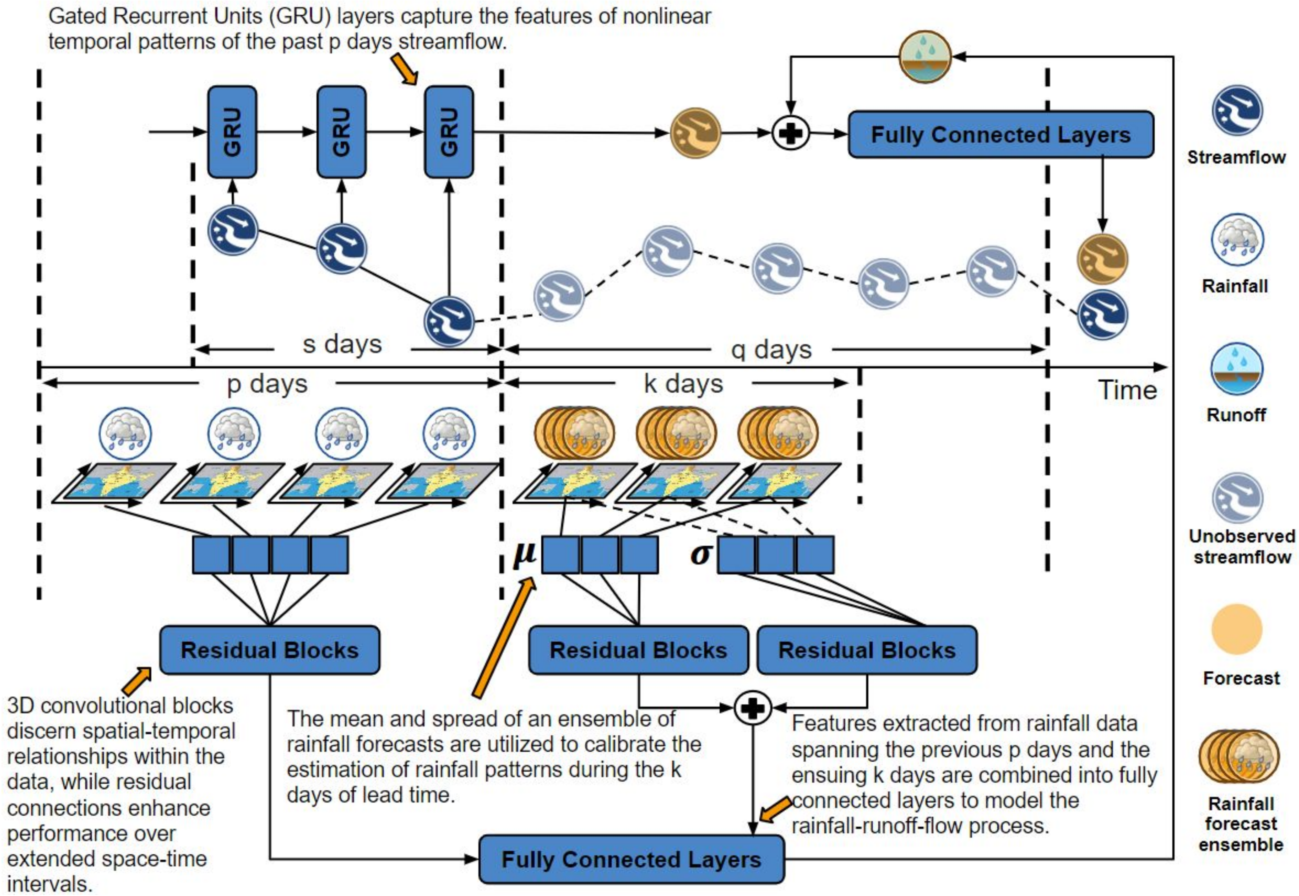

Fig. 3: The CASTEL Network Architecture for Enhanced Long-lead Streamflow Forecasting. It consists of Gated Recurrent Units (GRUs) for capturing nonlinear temporal streamflow patterns, 3D convolutional residual blocks for spatial-temporal relationship discernment, and fully connected layers for synthesizing features across time scales. The network is calibrated using ensemble rainfall forecasts to predict streamflow with extended lead times.

Definition 1 (*Spatio-Temporal Events Chain*): A Spatio-Temporal Events Chain (STEC) is an ordered series of spatially-related events that occur sequentially over time.

In an STEC, events are intricately ordered, both spatially and temporally. This sequencing forms the backbone of the model. Underpinning this structure is the "follow relationship", a pivotal concept that delineates the progression and interdependencies within the chain of events.

Definition 2 (*Follow Relationship*): A follow relationship, denoted as '$\triangleright$', indicates a preceding event e_i is followed by a subsequent event e_j.

Thus, a $STEC$ can be formed by cascaded events, which follow each other in spatio-temporal context.

$$STEC = (e_1 \triangleright e_2 \triangleright ... \triangleright e_i), \tag{1}$$

where e_i represents the i_{th} event.

B. Formulation of Streamflow Forecasting

Streamflow emerges from a sequence of interconnected events, as shown in Figure 2. Rainfall that occurs in certain upstream areas generates runoff. This runoff then enters river channels, leading to an increase in flow, termed as "flow-increment". The sequence of rainfall, runoff, and flow-increment transpires in distinct stages and can experience time lags. This sequence can be aptly described by a three-step Spatio-Temporal Events Chain: $rainfall \triangleright runoff \triangleright flow\text{-}increment$.

Let $\mathcal{Q}t$ denote the streamflow at the forecast location at time t. The historically observed streamflow over the fixed time period s, where $s > 0$, is represented as $\mathcal{Q}_{t-s}^{t}$. Given a lead time of q, where $q > 0$, the streamflow forecasting problem can be formulated as:

$$\hat{\mathcal{Q}}_{t+q} = \mathcal{F}(\mathcal{Q}_{t-s}^{t}, flow\text{-}increment), \tag{2}$$

Here, $\mathcal{F}$ represents the estimation model. The term $flow\text{-}increment$ results from the progression defined by the 3-step STEC: $rainfall \triangleright runoff \triangleright flow\text{-}increment$. Let's use f to represent the follow relationship $\triangleright$, R to represent rainfall in the upstream basin, and Z to represent the runoff generated by R. If we consider p time steps prior, we obtain:

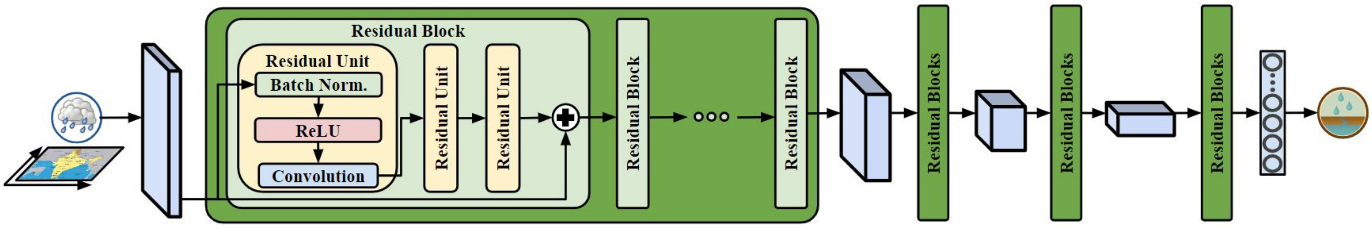

Fig. 4: Residual architecture for spatio-temporal learning.

$$
\begin{aligned}
Z_{t-p}^{t} &= f_z(R_{t-p}^{t}) \\
flow\text{-}increment &= f_q(Z_{t-p}^{t}) = f_q(f_z(R_{t-p}^{t}))
\end{aligned} \tag{3}
$$

Where f_z and f_q are the transformation functions for rainfall-to-runoff and runoff-to-flow-increment, respectively.

Substituting into equation 2, we get:

$$\hat{\mathcal{Q}}_{t+q} = \mathcal{F}(\mathcal{Q}_{t-s}^{t}, f_q(f_z(R_{t-p}^{t}))) \tag{4}$$

To optimize forecasting over a long lead time q, we incorporate forecasted rainfall ($\hat{R}$) of the upstream basin areas up to time k into the scheme. Rewriting equation 2 with this inclusion:

$$\hat{\mathcal{Q}}_{t+q} = \mathcal{F}(\mathcal{Q}_{t-s}^{t}, f_q(f_z(R_{t-p}^{t})), f_q(f_z(\hat{R}_{t}^{t+k}))) \tag{5}$$

The model formulated above underpins our proposed CASTLE model, which will be elaborated on in the subsequent sections.

IV. CASTLE: A Deep Learning Approach to Streamflow Forecasting

To model the intricate spatio-temporal dependencies in equation 5, we propose a Cascaded Spatio-Temporal Learning (CASTLE) model by leveraging deep learning to holistically integrate information from observed and forecasted rainfall in the upstream area and the historical stream flow at the prediction site. This provides an enriched feature set to estimate $\hat{\mathcal{Q}}_{t+q}$. Figure 3 provides an architectural overview.

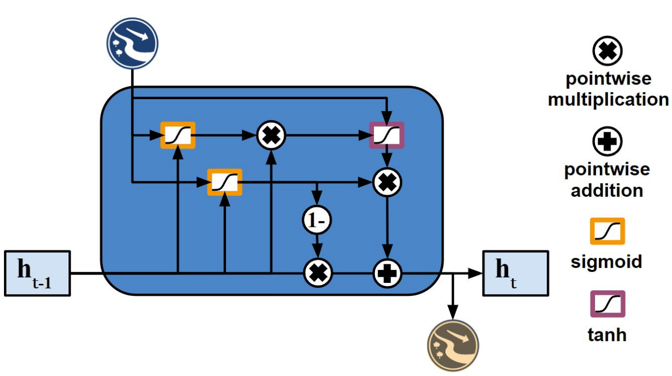

Fig. 5: GRU cell used for temporal sequence learning

A. Temporal Sequence Learning

With time being an inherent component in hydrological processes, there's an imperative need to capture historical dependencies in streamflow data. While traditional RNNs serve this purpose, we specifically employ the GRU cell for this task (See Fig. 5). GRU, being computationally efficient and having fewer parameters, aptly captures temporal dependencies in the streamflow data $\mathcal{Q}_{t-s}^{t}$, preparing the stage for the subsequent prediction of $\hat{\mathcal{Q}}_{t+q}$.

B. Spatial-temporal Learning

For the $rainfall \triangleright runoff \triangleright flow\text{-}increment$ STEC, since the water systems, by their nature, interact over spatial domains, and these interactions evolve over time. Addressing this dual challenge requires an architecture that can encapsulate both spatial and temporal dependencies. we employ deep residual networks with 3D Kernels for two key reasons:

- **Depth for Spatial Coverage:** Hydrological events across a basin can have cascading impacts. A deeper architecture ensures that spatial interactions, even those from distant locations, are effectively captured. Residual networks, with their skip connections, make training such deep architectures feasible.
- **3D Kernels for Temporal Depth:** The temporal axis added by 3D kernels in the residual units ensures that not just spatial, but also temporal relationships are maintained. This recognizes the temporal lag between causative events like rainfall and resultant effects like streamflow increment.

After an extensive evaluation of various residual architecture configurations, we determined that a structure consisting of 18 residual blocks equipped with 3x3x3 filters delivers the most optimal performance for our specific application. This particular setup strikes a balance between model complexity and computational efficiency, making it well-suited for capturing the nuanced spatial-temporal dynamics inherent in hydrological processes. The 3x3x3 filter size was found to be particularly effective in encapsulating local spatial features while maintaining a manageable number of parameters, thus reducing the risk of overfitting. Additionally, the depth provided by 18 blocks allows the network to learn a hierarchy of features, from basic patterns at the lower levels to more complex interactions at the higher levels. This depth is crucial for accurately modeling the multifaceted relationships between rainfall, runoff, and flow increment. The combination of these

TABLE I: Stream flow forecasting performance for the Ganges River during June-September over the period 2010-2016.

Metrics	QQ			QQ+ObsR			QQ+ObsR+ForeR$_P$			CASTLE$_{2D}$			CASTLE$_{3D}$		
	5-d	7-d	10-d	5-d	7-d	10-d	5-d	7-d	10-d	5-d	7-d	10-d	5-d	7-d	10-d
MAE (m^3/s)	3165	4540	6196	2694	3529	4566	2687	3480	4341	2435	3028	3580	**2411**	**3010**	**3565**
RMSE (m^3/s)	4380	5973	7875	3623	4595	5871	3594	4494	5552	3567	4389	5367	**3497**	**4316**	**5291**
R^2	0.88	0.78	0.63	0.93	0.89	0.83	0.93	0.9	0.86	**0.94**	**0.91**	**0.87**	**0.94**	**0.91**	**0.87**

factors makes the 18-block architecture with 3x3x3 filters not only robust in terms of predictive capability but also practical for implementation, offering a refined balance of depth and breadth in feature representation. Therefore, in this study, we will utilize this configuration, leveraging its proven effectiveness in capturing the intricate spatial-temporal dynamics of hydrological processes. The details are shown in Fig. 4.

When modeling the interactions of rainfall-runoff and optimizing streamflow forecasting over an extended lead time q, it's crucial to incorporate both observed and forecasted data. However, it's common knowledge that systematic biases often plague rainfall forecasts. These biases can originate from various factors, such as model inaccuracies, insufficient resolution, sub-optimal parameterizations, or even less-than-ideal methods employed for generating the initial conditions. To address this, for the forecasted segment, we utilize ensemble predictions from hydrological models. While these enrich our data source, they simultaneously introduce computational complexities. To maintain model efficiency, we've streamlined the architecture to work on ensemble means and spread, rather than individual ensemble predictions. Accordingly, we've constructed three specialized 3D residual networks: two are dedicated to processing ensemble statistics, while one is tailored for observed rainfall data. These networks process different data sources, and their outputs are further integrated into subsequent layers to ensure a holistic and accurate representation of the hydrological processes.

C. Fusion

We employ fully connected layers to model estimation function $\mathcal{F}$ in Eq.5 and integrate the temporal and spatial output streams from the temporal sequence model and the spatial-temporal models, blending them, ensuring a cohesive representation of hydrological processes, leading to the final estimate of $\hat{\mathcal{Q}}_{t+q}$.

V. Experiments

A. Datasets used in the study

In this study, we applied the CASTLE model to three distinct data sources:

- Historical daily discharge (streamflow) records from Hardinge Bridge (Ganges).
- Observed precipitation data from 1985 to 2016 with a one-degree resolution, produced by the PERSIANN system (Remotely Sensed Information using Artificial Neural Networks) [25].
- Ensemble precipitation forecasts from 1985 to 2016, sourced from the Global Forecast System (GFS) of the Environmental Prediction[1].

B. Experiment Settings

The model ingests lagged observations of precipitation and streamflow from day -14 to day 0 (p=15). It also considers ensemble weather forecast data (both mean and spread) from day 1 to day 5 (k=5) and historical streamflow from day -9 to day 0 (s=10). The other configurations are as follows:

- **Rainy Season:** To account for the reduced flow and variability during the dry months, we only consider data from the monsoon season, namely June to September.
- **Training Set:** 90% of the data from 1985 to 2009.
- **Validation Set:** The remaining 10% of the data from the same period.
- **Testing Set:** Data from 2010 to 2016.

All experiments were conducted on a dedicated Linux GPU server equipped with 48 CPU cores, 1TB RAM, and eight Nvidia 1080ti GPUs, each with approximately 11GB memory. Each model iteration was trained using a single GPU on Tensorflow 2.0. The training capped at 1000 epochs, but early stopping was applied if no improvement was detected after 15 epochs. Should there be no improvement over five consecutive epochs, we reduced the learning rate by 10%.

C. Comparison with ReqSim models

1) Overview of ReqSim models: To evaluate how well CASTLE performs, we compared it with the Requisitely Simple (ReqSim) streamflow forecasting models, which include three models:

- **Flow Persistence Model (QQ):** This model primarily bases its predictions on past streamflow data or water levels at the location we're forecasting for. It provides a basic way to understand flow patterns.
- **Flow Persistence with Observed Rainfall Model (QQ+ObsR):** Building on the QQ model, this version not only considers past streamflow data or water levels but also includes recent rainfall data from areas upstream. It adjusts these predictions considering the typical time it takes for water to travel downstream.

[1]https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

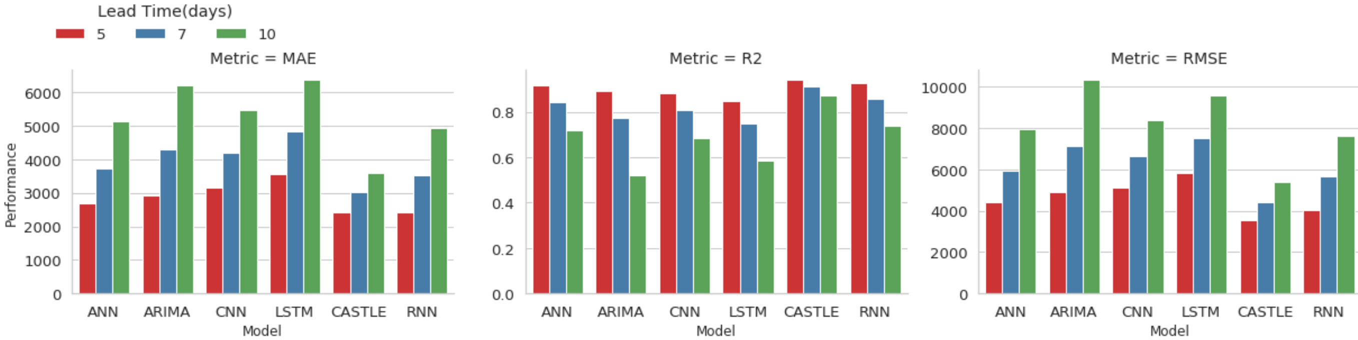

Fig. 6: Comparative model performance for 5, 7, and 10-day lead times, measured by MAE, R², and RMSE.

- **Flow Persistence with Observed and Forecasted Rainfall Model (QQ+ObsR+ForeR):** The most advanced model in the ReqSim group, it combines the features of the previous two models and also factors in predicted rainfall data from upstream areas, offering a more comprehensive forecast.

2) Results: To really test the importance of using 3D to capture time-related patterns, we also looked at how the 2D version of CASTLE performed. We used common evaluation metrics: Mean Absolute Errors (MAE), Root Mean Square Errors (RMSE), and R-squared values (R^2). The details are in Table I.

From the table, we can see:

a) Performance Enhancement with Rainfall Data Integration: Transitioning from the base QQ model, which primarily relies on flow persistence, to the more data-enriched QQ+ObsR and further to QQ+ObsR+ForeR underscores a marked performance boost. The addition of observed and forecasted rainfall data not only improves prediction accuracy but also lends credence to the viability of the rainfall-runoff-flow increment Spatio-Temporal Events Chain. This progression is evident across all lead times and metrics, particularly in the reduction of MAE and RMSE values, indicating tighter error margins and more reliable forecasts.

b) Dominance of CASTLE Models: Both CASTLE$_{2D}$ and CASTLE$_{3D}$ variants tower above the ReqSim models across the board. Regardless of the lead time (5, 7, or 10 days), CASTLE models consistently reported lower errors and higher R^2 values. These findings suggest that the CASTLE architecture, in both its dimensions, holds significant promise in streamflow forecasting.

c) Superiority of 3D Temporal Learning: The most crucial takeaway emerges when comparing CASTLE$_{2D}$ with CASTLE$_{3D}$. The consistently superior metrics exhibited by the 3D version advocate for its enhanced ability in capturing intricate time-related patterns. For instance, the 10-day lead time predictions show MAE and RMSE values of 3580 (m^3/s) and 5367 (m^3/s) for the 2D version, whereas the 3D version tightens these values to 3565 (m^3/s) and 5291 (m^3/s), respectively. Furthermore, the R^2 values for both models stand at an impressive 0.87, indicating a strong fit to the actual data.

In summary, our experiments have affirmed the inherent advantage of integrating 3D kernels in temporal sequence learning, especially when it comes to streamflow forecasting. The meticulous design of CASTLE, complemented by its adept handling of time patterns, marks a significant stride in predictive hydrology.

D. Benchmarking

1) Models: In this experiment, we compared the 3D version of CASTLE against five streamflow predictive models, including Artificial Neural Networks (ANN), Auto-regressive Integrated Moving Average (ARIMA), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory Networks (LSTM). The comparative landscape was set against three varied lead time benchmarks: 5, 7, and 10 days.

2) Results: The results, visually encapsulated in Fig. 6, offer compelling evidence in favor of CASTLE$_{3D}$'s superior predictive capabilities over the other tested models. Across all three lead times of 5, 7, and 10 days, CASTLE$_{3D}$ exhibited the lowest Mean Absolute Error (MAE) and the highest R-squared (R^2) values, confirming its more accurate and consistent performance.

a) Mean Absolute Error (MAE): At a 5-day lead time, CASTLE$_{3D}$ achieved an MAE of 2435, significantly better than the closest competing model, RNN, which had an MAE of 2417.6. The gap in MAE widened even more at longer lead times. At a 7-day lead time, CASTLE$_{3D}$'s MAE of 3028 was noticeably lower than ANN's 3721.4, and at a 10-day lead time, it further reduced to 3580, considerably lower than the next best, which was ANN at 5140.8. These MAE results validate CASTLE$_{3D}$'s robustness in short to medium-term streamflow forecasting.

b) R-squared (R^2): R^2 values further substantiated the model's proficiency. At a 5-day lead time, CASTLE$_{3D}$'s R^2 value was 0.94, outpacing the closest competitor, ANN, with an R^2 of 0.914. The trend continued at 7 and 10-day lead times, where CASTLE$_{3D}$ posted R^2 values of 0.91 and 0.871, respectively, remaining above all competing models. The high R^2 values indicate that CASTLE$_{3D}$ could explain more of the variance in the observed data, reaffirming its predictive quality.

c) Root Mean Square Error (RMSE): Although RMSE is generally not the primary measure for comparison, CASTLE$_{3D}$ also led in this aspect, with RMSE values of 3567, 4389, and 5367 for 5, 7, and 10-day lead times, respectively. These numbers were lower than those of all the other models, revealing CASTLE$_{3D}$'s capability to minimize both the frequency and the magnitude of errors.

In summary, the results conclusively show that CASTLE$_{3D}$ delivers a powerful, accurate, and reliable streamflow forecasting model that significantly outperforms other models across multiple metrics and lead times.

E. Discussion

Our study clearly highlights the advantages of CASTLE, especially its 3D version, in tackling streamflow prediction challenges. Compared to traditional benchmarks, CASTLE's consistent performance shows its strong capability and flexibility. A key finding was how 3D Kernels excel in capturing detailed temporal patterns—a feature deserving more attention in future research. While our current results are promising, testing CASTLE in various hydrological situations and on larger datasets would further validate its broad applicability.

VI. Conclusion

In this study, we aimed to assess the performance of CASTLE, with a special focus on its 3D variant, within the domain of streamflow forecasting. Our detailed experiments confirmed that CASTLE$_{3D}$ outperforms traditional models. The introduction and testing of 3D Kernels hint at a new direction for more accurate forecasting methods. Incorporating both observed and forecasted rainfall data has shown to improve streamflow predictions. Looking ahead, there's ample opportunity to further refine and expand upon CASTLE's potential.

VII. Acknowledgements

This material is based upon work partially supported by the National Science Foundation under NSF grant IIS 2008202 and NSF 2008276. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research was also partially supported by the Oracle grant to the College of Science and Mathematics at the University of Massachusetts Boston.

References

[1] T. H. Dewan, "Societal impacts and vulnerability to floods in bangladesh and nepal," *Weather and Climate Extremes*, vol. 7, pp. 36–42, 2015.

[2] B. Rozenberg and S. H. Adrien, "Building the resilience of the poor in the face of natural disasters," *Int. Bank Reconstr. Dev*, 2017.

[3] H. L. Cloke and F. Pappenberger, "Ensemble flood forecasting: A review," *Journal of hydrology*, vol. 375, no. 3-4, pp. 613–626, 2009.

[4] J. P. Charba and F. G. Samplatsky, "High-resolution gfs-based mos quantitative precipitation forecasts on a 4-km grid," *Monthly weather review*, vol. 139, no. 1, pp. 39–68, 2011.

[5] S. Dravitzki and J. McGregor, "Predictability of heavy precipitation in the waikato river basin of new zealand," *Monthly Weather Review*, vol. 139, no. 7, pp. 2184–2197, 2011.

[6] N. Voisin, F. Pappenberger, D. P. Lettenmaier, R. Buizza, and J. C. Schaake, "Application of a medium-range global hydrologic probabilistic forecast scheme to the ohio river basin," *Weather and Forecasting*, vol. 26, no. 4, pp. 425–446, 2011.

[7] W. Palash, Y. Jiang, A. S. Akanda, D. L. Small, A. Nozari, and S. Islam, "A streamflow and water level forecasting model for the ganges, brahmaputra, and meghna rivers with requisite simplicity," *Journal of Hydrometeorology*, vol. 19, no. 1, pp. 201–225, 2018.

[8] F. Pappenberger, K. J. Beven, N. Hunter, P. Bates, B. T. Gouweleeuw, J. Thielen, and A. De Roo, "Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the european flood forecasting system (effs)," *Hydrology and Earth System Sciences*, vol. 9, no. 4, pp. 381–393, 2005.

[9] D. Demeritt, H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes, and M.-H. Ramos, "Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting," *Environmental Hazards*, vol. 7, no. 2, pp. 115–127, 2007.

[10] K. Marks and P. Bates, "Integration of high-resolution topographic data with floodplain flow models," *Hydrological processes*, vol. 14, no. 11-12, pp. 2109–2122, 2000.

[11] M.-u.-D. A. Mac Kirby, M. Mainuddin, W. Palash, E. Qadir, and S. M. Shah-Newaz, "Bangladesh integrated water resources assessment supplementary report: approximate regional water balances," 2014.

[12] M. M. Q. Mirza, R. A. Warrick, N. Ericksen, and G. Kenny, "Are floods getting worse in the ganges, brahmaputra and meghna basins?" *Global Environmental Change Part B: Environmental Hazards*, vol. 3, no. 2, pp. 37–48, 2001.

[13] W. Palash, M. Quadir, S. Shah-Newaz, M. Kirby, M. Mainuddin, A. Khan, and M. Hossain, "Surface water assessment of bangladesh and impact of climate change," *Bangladesh Integrated Water Resources Assessment Rep*, 2014.

[14] T. C. Mills, *Time series techniques for economists*. Cambridge University Press, 1990.

[15] F. Jakobsen, A. Z. Hoque, G. N. Paudyal, and S. Bhuiyan, "Evaluation of the short-term processes forcing the monsoon river floods in bangladesh," *Water international*, vol. 30, no. 3, pp. 389–399, 2005.

[16] K. J. Beven, *Rainfall-runoff modelling: the primer*. John Wiley & Sons, 2011.

[17] P. C. Young and K. J. Beven, "Data-based mechanistic modelling and the rainfall-flow non-linearity," *Environmetrics*, vol. 5, no. 3, pp. 335–363, 1994.

[18] P. C. Young, "Advances in real–time flood forecasting," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 360, no. 1796, pp. 1433–1450, 2002.

[19] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (lstm) networks," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 6005–6022, 2018.

[20] A. Minns and M. Hall, "Artificial neural networks as rainfall-runoff models," *Hydrological sciences journal*, vol. 41, no. 3, pp. 399–417, 1996.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[23] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[24] M. Dixon, D. Klabjan, and J. H. Bang, "Classification-based financial markets prediction using deep neural networks," *Algorithmic Finance*, vol. 6, no. 3-4, pp. 67–77, 2017.

[25] H. Ashouri, K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, "Persiann-cdr: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies," *Bulletin of the American Meteorological Society*, vol. 96, no. 1, pp. 69–83, 2015.