https://doi.org/10.71573/dw764v68

Diffusion-based Time Series Forecasting for Sewerage Systems

Nicholas A. Pearson[1,*] https://orcid.org/0009-0008-4285-6831, *Francesca Cairoli*[1] https://orcid.org/0000-0002-6994-6553, *Luca Bortolussi*[1] https://orcid.org/0000-0001-8874-4001, *Davide Russo*[2] *& Francesca Zanello*[2]

[1]*University of Trieste, Department of Mathematics, Informatics and Geosciences, Trieste, Italy*
[2] *Idrostudi srl, Trieste, Italy*
**Corresponding author email:* nicholasandrea.pearson@phd.units.it

Abstract

In light of the impacts of climate change and urbanisation, data-driven urban water management (UWM) has been proposed as a means of transforming sewerage and drainage systems, aiming to achieve resilient and sustainable served territories. The increased collection and utilization of data offers significant potential benefits; however, missing data and data quality pose challenges for their subsequent use in calibrating and validating computational models, evaluating infrastructures performance and implementing automatic control solutions in the field.

We introduce a novel deep learning approach that harnesses the power of generative artificial intelligence to enhance the accuracy of contextual forecasting in sewerage systems. By developing a diffusion-based model that processes multivariate time series data, our system excels at capturing complex correlations across diverse environmental variables, enabling robust predictions even during wet weather periods. To strengthen the model's reliability, we further calibrate its predictions with a conformal inference technique, tailored for probabilistic time series data, ensuring that the resulting prediction intervals are statistically reliable and cover the true target values with a desired confidence level. Our empirical tests on real sewerage system data confirm the model's capability to deliver reliable contextual predictions, maintaining accuracy even under severe weather conditions.

Highlights

- We propose a novel methodology for contextual forecasting on sewerage system time series
- We employ a diffusion-based generative model to obtain accurate imputations for the time series
- We obtain predictive intervals with statistical coverage guarantees using conformal prediction

Introduction

The growth of urban centres, both in terms of extension and population size, has brought increased stress on sewerage and drainage systems. These infrastructures are essential for local communities but rely on ageing networks which struggle to cope with the effects of increased population in urban areas and widespread soil sealing combined with extreme weather events, with the latter becoming increasingly more frequent and intense due to climate change (IPCC, 2023). Additional challenges arise from environmental concerns, regulatory and institutional changes, and public awareness. These pressing issues emphasize the urgent need for innovative solutions in the management of urban wastewater infrastructure in order to ensure more sustainable services. Digital solutions, such as networks of sensors, smart equipment, real-time digital twins, data analytics and advanced computational tools enable water utilities to be better prepared for this varying context and support the identification of optimal adaptation strategies to changing climate and demographics. Establishing a monitoring network is fundamental for capturing the current hydraulic functioning of sewerage and stormwater networks through collected time

series, which are also essential for calibrating and validating computational models that support all phases of urban wastewater and drainage systems management, from design to monitoring and control.
The data-driven Urban Water Management (UWM) approach has the potential to expand the functionality of today's sewerage and drainage systems (Eggimann et al., 2017, Fu et al., 2022, Sweetapple et al., 2023). However, several challenges related to data-driven approaches need to be addressed - for instance data collected by sensors and measuring devices may be inaccurate, partly unavailable, or affected by anomalies and technical failures (Bertrand-Krajewski et al., 2021). Human experts can identify problematic behaviours upon visual inspection of raw or pre-processed time series but performing this task on a large scale is often impractical as it is extensively time consuming, especially due to the growing amount and diversity of data. More complex tasks, such as accurately imputing missing data, pose an even greater challenge without the aid of an automated system. Furthermore, having the ability to perform short to medium term forecasts on network behaviour may be an invaluable support asset for operators on the field, especially during wet weather events when urban pluvial flooding and other undesirable events are more likely to occur.

This work focuses on using generative AI as a valuable asset to support data driven decision making in wastewater systems management by leveraging time series harvested from monitoring campaigns of sewerage networks. To this end, we developed a diffusion model that processes time series data (Tashiro et al., 2021) to deliver probabilistic predictions on the network's behaviour and effectively provides support when addressing the aforementioned problems, producing prediction intervals which offer high accuracy and statistically valid coverage guarantees.

Methodology

The dataset consists of multivariate time series recording the variations of various physical parameters, including sewage level and speed, obtained from sensors placed in specific network locations. Precipitation intensity measurements derived from rain gauges positioned in proximity to the selected measurement sites are also included. The main objective of this work is to perform *contextual forecasting*, a predictive approach that generates forecasts for selected target variables by combining information from their historical values with both historical and currently available information of other variables. In this framework, predictions are grounded on the historical values of the target variables and further refined by incorporating contextual information from other variables to enhance accuracy. This approach is particularly beneficial when the individual series are highly correlated. Contextual forecasting can be framed as a time series imputation problem, i.e. the task of estimating the missing values of a time series given its observed ones. This can be achieved by considering the final values of the features of interest as missing values to be imputed, while utilizing the entirety of the others as context for the prediction.

With their work on Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation (CSDI), Tashiro et al. (2021) proposed a novel approach that leverages Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) specifically for the task of time series imputation. DDPMs are a type of deep generative models trained in a two-step procedure. In the initial forward process, existing data is gradually corrupted by iteratively adding small amounts of random noise. In the ensuing backwards process, a neural network learns a denoising procedure that iteratively and gradually removes noise, eventually generating new data starting from pure random noise initializations.

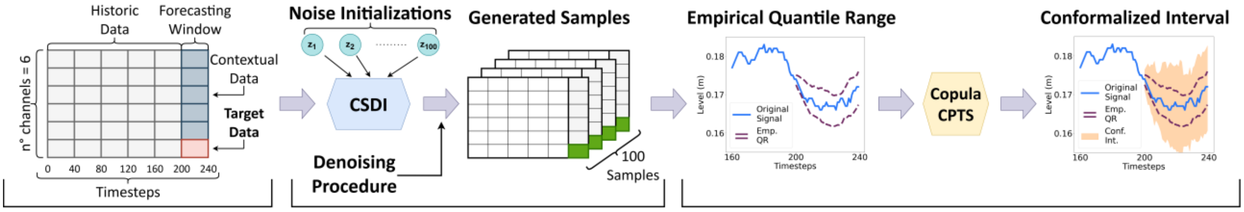
At training time CSDI learns an estimate for the conditional probability distribution of the missing values of the time series conditioned on the observed ones. Although this distribution cannot be expressed in closed form, it is possible to sample from via the denoising procedure. To approximate this distribution, CSDI employs a self-supervised procedure where various portions of the observed values are masked or obscured, and the model then attempts to reconstruct them given the remaining observations. The choice of which parts to mask plays an essential role in defining what data patterns the model will be able to reconstruct in the inference phase. Ideally, the masking strategy should align with both the objective of the target application and the expected missing data patterns. When dealing with contextual forecasting, an appropriate choice could be a random number of final values of the time series for one or more features, also referred to as channels. This increases model flexibility as, if trained correctly, a single model can impute the targeted data for any combination of concurrently masked channels.

At inference time, the model generates multiple predictions by repeatedly sampling from the learned conditional probability distribution, transforming different random noise initializations into plausible imputations. These samples collectively represent an ensemble of possible predictions for the target data which are consistent with the observed conditioning data. Starting from this set, we can compute summary statistics such as the mean or median of the predictive distribution. In addition, we can define an empirical quantile range by deriving two empirical quantiles from the set of sampled predictions. This approach provides a predictive interval for each time step within the target segment of the multivariate time series, characterizing the spread of plausible outcomes and offering a more robust output compared to single-point predictions.

However, predictive intervals constructed purely from empirical quantiles do not necessarily satisfy predetermined coverage requirements. Given an arbitrary confidence level $1 - \alpha$, the objective is to obtain confidence intervals that contain, on average, the true value with a probability greater than $1 - \alpha$. In this setting, the coverage requirement extends to the entire target sequence, requiring that the observed values fall within the respective predictive intervals simultaneously across all time steps within the target window. To achieve this, we employ conformal prediction (CP), a flexible two-step approach that adjusts the output of any model, transforming it into predictive intervals which offer statistically rigorous coverage guarantees.

In the initial calibration phase of CP, a non-conformity score is computed for each observation of a held-out calibration dataset. These scores quantify the differences between the predicted values and their respective ground-truth. The $(1-\alpha)^{th}$ largest non-conformity score is then selected and used as a critical value that, when added to and subtracted from new predictions during deployment, provides intervals that entail the desired coverage probability. The choice of which non-conformity score is used, combined with the quality of predictions by the original model, directly influence the size of the intervals, which we aim to minimize to provide meaningful insights. CP requires no retraining of the original model and simply operates as a wrapper around it, requiring minor adjustments to adapt to the predictive model at hand.

Our probabilistic model over time series data calls for CopulaCPTS (Sun and Yu, 2023), a copula-based approach for efficient CP on multivariate time series. We adapt CopulaCPTS to handle the probabilistic output of CSDI by employing the non-conformity score from Romano et al. (2019). This non-conformity score is specifically tailored for predictive outputs in the form of empirical quantile ranges rather than single point estimates, making it suitable for our probabilistic framework. The calibration phase of the CopulaCPTS procedure yields a set of adjustment terms – one for each time step within the target segment - that are applied to the individual intervals of the empirical quantile range. These adjustments can either expand or shrink the original intervals, providing predictive intervals which offer coverage guarantees. The resulting intervals are, on average, guaranteed to concurrently cover the true values of the time series over the full prediction window with a probability greater than the desired confidence level.



a) Multivariate Time Series b) Prediction Phase c) Conformal Prediction Phase

Figure 1. Pipeline for the proposed methodology. (a) Structure of the multivariate time series, which is provided as input to the CSDI model in the prediction phase. (b) Starting from 100 noise initializations, the CSDI model generates as many sample

trajectories, from which empirical quantile ranges are extracted. (c) These ranges are then adjusted using CopulaCPTS, providing intervals which are guaranteed to cover the true values of the time series over the full predictive window with the desired confidence level.

Figure 1 illustrates the complete pipeline of the proposed methodology, which includes the contextual forecasting setup for a multivariate time series, the extraction of empirical quantile ranges during the prediction phase, and their final adjustment using CopulaCPTS. For a more detailed theoretical description of the proposed methodology, we refer the reader to Pearson et al. (2025).

Case Study

Our case study draws on data from a five-year in-sewer monitoring campaign conducted in two neighbouring municipalities located in the North-West of Italy. The mainly combined sewer system that serves the selected municipalities is part of a 2’700 km long wastewater collection system with a catchment of approximately 260 km^2. The main objective of the monitoring campaign was the characterisation of the hydraulic dynamics both in dry and wet weather and the estimation of inflows and infiltration in each municipality. To this end, flowmeters were installed along the boundaries of the municipalities and rain gauges were placed to monitor rainfall events occurrence.

The case study catchment, visible in Figure 2 (left), spans an area of approximately 11 km^2 and is covered by a 160 km long combined sewerage network, serving around 60’000 inhabitants. The dataset derived from this catchment spans 4 years, from 2020 to 2023, and includes sewer level observations recorded at six-minute intervals from nine area-velocity sensors placed in-line within or on the boundaries of the selected municipalities. The dataset also contains time series representing rainfall intensity collected every minute by a rain gauge installed in the study area. All of the time series included in this study have been previously validated by human experts.

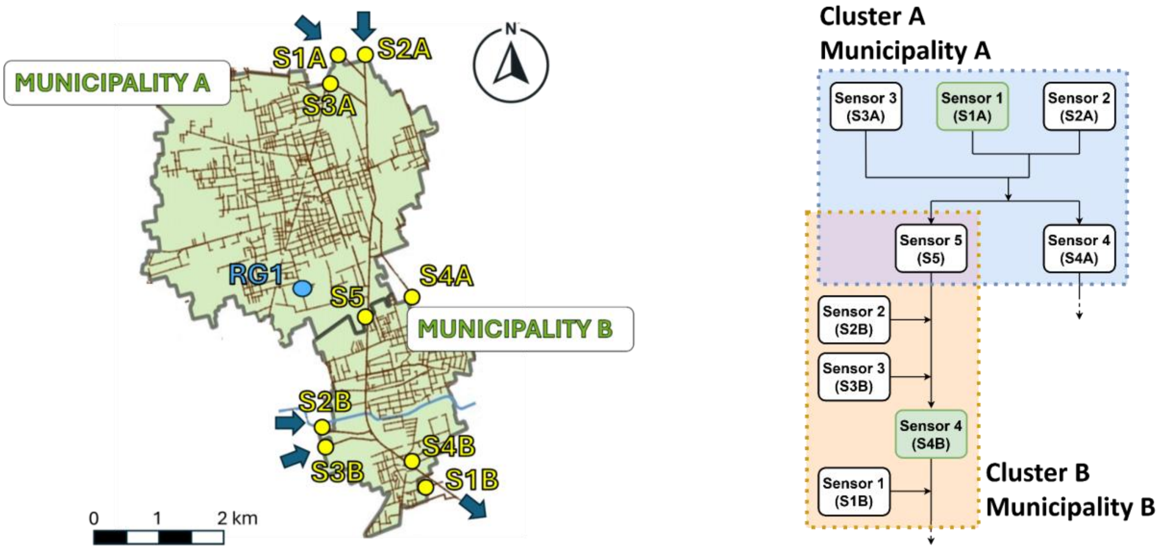


Figure 2. (left) Map view of the case study sewerage collection system, with pipes (brown lines), catchment (light green areas), river crossing (blue line) and location of the in-line sewer metering stations (yellow dots) and of the rain gauge (light blue dot). The blue arrows show the main direction of the flow. (right) Schematic representation of the hydraulic connectivity within and between the measurement zones comprising Cluster A and Cluster B. Arrows indicate the direction of flow through the network. Shaded areas correspond to different municipalities. Sensors highlighted in green correspond to the target time series evaluated in the experimental section.

To adapt the available monitoring data to our multivariate time series framework, we focus on groups of level measurements derived from sensors placed in locations that are hydraulically connected within the wastewater collection system. The two groups of metering stations, referred to as Cluster A (Municipality A) and Cluster B (Municipality B) in the remainder, consist of level readings derived from measurement zones located within the same municipality and linked by sewer conduits. This grouping approach allows us to capture dependencies among level measurements taken from the sensors belonging to each cluster.

Figure 2 (right) illustrates the hydraulic connections and measurement zone arrangement within and between the two clusters. The data of each cluster is further enriched with concurrent rain intensity measurements taken from rain gauges located in proximity to the acquisition devices of each group.
As a result, each cluster produces a multivariate time series with six channels, five corresponding to sewage level measurements and one to rainfall intensity, sampled every six minutes over a 24-hour period, yielding a sequence with 240 timesteps. Figure 3 provides a graphical representation of the multivariate time series considered in this study, both through a conceptual schema and a detailed example, with the contextual forecasting targeting strategy visible in both. In the left panel it is possible to observe the response of the level measurements to rainfall events, described by the rain intensity time series, highlighting how precipitation events induce similar variations for all sensor channels.

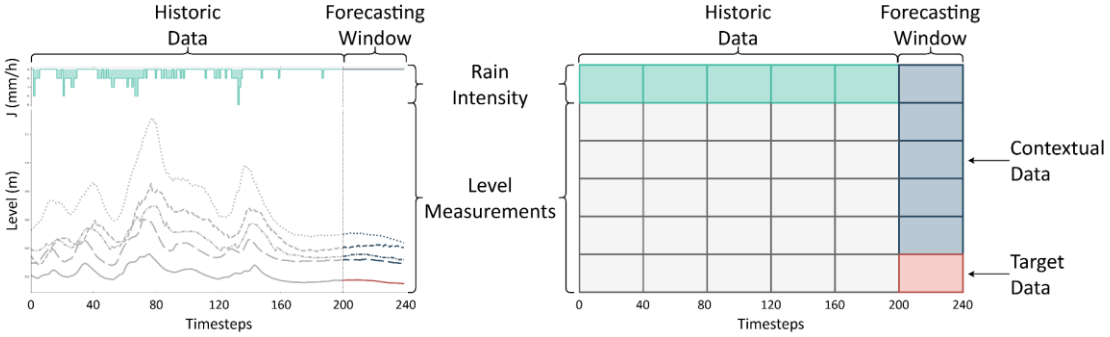


Figure 3. Detailed (left) and conceptual (right) representations of the multivariate time series and contextual forecasting targeting strategy used in this study. In both representations the teal shaded area represents historic rain intensity measurements, historical level measurements are shown in grey, contextual information in blue and target data in red. The right panel shows rain intensity measurements in the top part, while in the bottom part each line represents the level measurements recorded by different sensor of the same cluster, all over a 24-hour period.

Results and discussion

We present experimental results focusing on the imputation of a single channel at the time per cluster specifically Sensor 1 from Cluster A and Sensor 4 for Cluster B, providing estimates for a four-hour time window to emulate the frequency of data transmission from the dataloggers connected to the sensors. Held-out cluster-specific calibration datasets are initially partitioned based on weather patterns to provide condition-specific adjustments to the predicted empirical quantile ranges. This partition is necessary given that the recorded time series generally presents values which are significantly higher during rainfall events compared to dry periods. Combining series with values of such different scales would have negatively affected the conformal prediction procedure, resulting in intervals which are over-conservative in dry conditions while being insufficiently wide during rainfall events.

A single CSDI model is used to generate 100 samples for two cluster specific test datasets, each of which contains approximately 600 time series. Although the same model is used to generate samples regardless of weather condition or cluster, the test sets are further partitioned using the same criteria as for the calibration datasets, resulting in approximately 550 observations per group being classified as dry conditions while 50 series as wet conditions. The test set partition does not impact sample generation but allows for more meaningful result comparisons and for the application of condition-specific adjustments in the CP step.

We use the median as an aggregative statistic for the predictive distribution and evaluate its accuracy with respect to the original level value using traditional metrics such as the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Table 1 shows systematic differences between the error rates in dry and wet conditions across both clusters. This is expected as the wet occurrences are less numerous within the dataset and additionally present more heterogeneous dynamics within each event, making them harder to accurately predict.

Given a confidence level $1-\alpha=0.9$, we retrieved the symmetric empirical quantiles $q_{0.05}$ and $q_{0.95}$ from the set of samples, constructing a quantile range which was further adjusted leveraging the CopulaCPTS procedure. The resulting conformalized intervals provide a 90% guarantee of containing the true values across all timesteps simultaneously. Table 1 also shows that the conformalized intervals provide a significant increase in coverage compared to the empirical quantile ranges for both clusters and weather conditions. These improvements require only minor adjustments to the quantile ranges, with smaller corrections needed in dry conditions. Although they still achieve similar coverage levels in all conditions, the conformalized intervals for the sensor of Cluster A appear to be moderately narrower than those for Cluster B. This is particularly noticeable in dry conditions and suggests that the set of samples generated in this scenario is less accurate, as is also visible from the higher error rates in Table 1. The lower accuracy can be attributed to noisier data for the sensor of Cluster B and, as a consequence, the empirical quantiles derived for it require larger adjustments to reach the desired level of confidence. Nevertheless, when evaluated next to the scale of the target values, all of the widths of the intervals in the different scenarios remain within acceptable bounds for practical applications.

Figure 4 proposes a set of examples illustrating the output of the proposed methodology for the two clusters in both dry and wet weather conditions. Blue lines represent the target level measurement, while grey lines correspond to the level readings from the conditioning sensors. The teal shaded area represents recorded rain intensity measurements. The orange shaded area corresponds to the conformalized intervals while the purple dashed lines represent the bounds of the empirical quantile ranges. The conformalized intervals consistently correct instances of mis-coverage by the empirical quantile range in both dry conditions (panels a and f) and wet conditions (panels g and h). These examples show instances in which the blue line falls outside the purple dashed bounds but remains within the orange shaded area, demonstrating the superior coverage provided by the conformalized intervals which envelop the target trajectories over the full predictive window. It is also evident from Figure 4 that the adjustments needed to achieve the desired level of coverage are noticeably smaller in dry conditions than in wet periods. Consistent with Table 1, the figure further shows that smaller corrections are needed for the level measurements taken from Sensor 1 of Cluster A compared to Sensor 4 of Cluster B, a difference which is particularly apparent in dry conditions and likely reflects the marginally noisier time series of the latter. Figure 4 also reveals that, in wet conditions, the predictive model successfully exploits the correlations existing between level measurements from different sensors, providing accurate predictions even during complex rainfall events. It is also visible that throughout the predictive window the widths of the empirical quantile range, and consequently also those of the conformalized interval, exhibit greater variability during wet periods than in dry conditions. Both intervals expand notably in the early phases of rainfall events, reflecting the greater uncertainty associated to this phase, before gradually converging towards the ground truth in the latter stages of the event (panels d and g). In contrast, under dry conditions, both intervals maintain relatively constant widths throughout the predictive period.

Table 1. Comparison of various results for two sensors belonging respectively to Cluster A and Cluster B under different weather conditions. The MAE and MAPE error metrics are reported to evaluate the quality of the generated samples (lower values are preferable). Coverage percentages and interval widths are reported to compare empirical quantile ranges against conformalized intervals, along with their per-step average adjustments. Additionally, average target values with standard deviation are shown for each scenario to facilitate result interpretation.

		Error Metrics		Empirical QR		Conformalized Intervals		
	Avg. Target Value (±std)	**Avg. MAE**	**Avg. MAPE**	**Coverage (%)**	**Avg. Width**	**Coverage (%)**	**Avg. Width**	**Avg. Correction**
Cluster A – Sensor 1 – Dry	0.155 (±0.021)	0.0026	1.6431	14.10	0.0079	88.38	0.0179	0.0050
Cluster A – Sensor 1 – Wet	0.284 (±0.160)	0.0218	6.2571	27.27	0.0718	87.88	0.2012	0.0647
Cluster B – Sensor 5 – Dry	0.155 (±0.032)	0.0033	2.1275	08.07	0.0090	87.60	0.0247	0.0079
Cluster B – Sensor 5 – Wet	0.446 (±0.361)	0.0305	6.2797	17.14	0.0922	88.57	0.3259	0.1169

Lastly, it is important to note that the adjustments performed on the empirical quantiles strictly depend on multiple factors, including the cluster and sensor for which an imputation is being performed and current weather condition. Changing any of these parameters requires using the correct combination-specific adjustment coefficients or, if not already available, repeating the CopulaCPTS process to compute them. Given a set of generated samples for a sufficiently large calibration set, the computational cost of

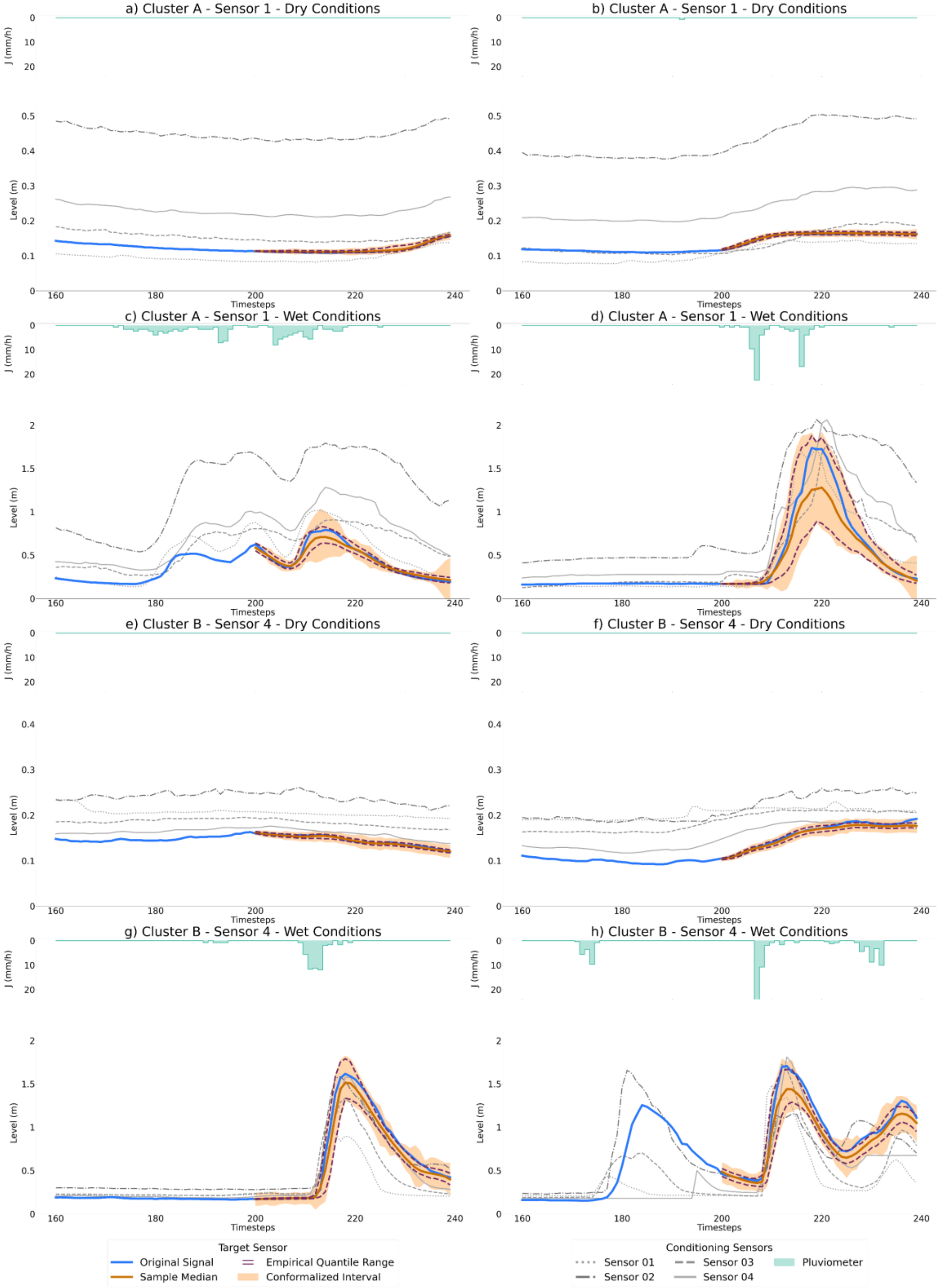


Figure 4. Comparison of model outputs for Sensor 1 of Cluster A (a-d) and Sensor 4 of Cluster B (e-h) across different weather conditions. All examples show the final 8-hours of a 24-hours period for which a 4-hour imputation is performed. Each panel shows precipitation intensity (top) and level measurements from the sensors under consideration (bottom). In dry conditions, the empirical quantile ranges coalesce close to the ground truth, requiring only minor adjustments to reach the desired coverage. In wet conditions larger adjustments are required.

this operation is modest, with the fitting of adjustment coefficients taking only a few seconds, which can then be applied instantly to any new empirical quantile range under the same conditions. Although results have only been presented for two clusters, this methodology has been successfully tested on over ten clusters from the same sewerage network as above, for varying forecasting horizons and on data from other wastewater systems with similar results overall.

Conclusions and future work

Urban sewerage and stormwater systems are a crucial component of urban infrastructure. Progresses in analytics and sensing, transmission, computing and data management have recently paved the way for information technology to play a greater role in UWM.

This paper has presented a novel application of diffusion-based generative models to perform contextual forecasting in the domain of sewerage systems. Our approach operates on multivariate time series of level measurements collected from various sensors placed throughout wastewater networks, exploiting correlations existing between the individual time series to provide accurate predictions. Combined with a conformal prediction procedure, the model produces prediction intervals which offer statistically valid coverage guarantees. Experimental results show that our model maintains robust performance in both dry and wet weather conditions, providing accurate predictive intervals even under challenging conditions. This work offers a strong starting point for more complex objectives such as the inpainting of missing data, data validation, near-real time anomaly detection and pure forecasting, proving its potential use in improving urban wastewater and drainage systems management through data-driven decision making.

Acknowledgements

This work has been supported by Idrostudi srl, who is funding the PhD scholarship of Nicholas A. Pearson. The authors also wish to thank Idrostudi srl for providing the operational datasets that enabled the experimental evaluation presented in this work and for their constant support and interest in this project.

This work has been partially supported by the PNRR project iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS_00000043).

References

Bertrand-Krajewski, J. L., Clemens-Meyer, F., & Lepot, M. (2021). Metrology in urban drainage and stormwater management: Plug and pray. *IWA Publishing*.

Eggimann, S., Mutzner, L., Wani, O., Schneider, M., Spuhler, D., Moy de Vitry, M., Beutler, P. & Maurer, M. (2017). The Potential of Knowing More: A Review of Data-Driven Urban Water Management. *Environmental Science & Technology*. 51. 10.1021/acs.est.6b04267.

Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler D. (2022), The role of deep learning in urban water management: A critical review, *Water Research*, 223,118973

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.

Intergovernmental Panel on Climate Change (IPCC). (2023). Cities, Settlements and Key Infrastructure. In *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 907–1040). chapter, Cambridge: Cambridge University Press.

Pearson, N. A., Zanello, F., Russo, D., Bortolussi, L., & Cairoli, F. (2025). CoCAI: Copula-based Conformal Anomaly Identification for Multivariate Time-Series. *arXiv preprint arXiv:2507.17796*.

Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. *Advances in neural information processing systems*, *32*.

Sun, S. H., & Yu, R. (2023). Copula Conformal prediction for multi-step time series prediction. In *The Twelfth International Conference on Learning Representations*.

Sweetapple, C., Webber, J., Hastings, A., & Melville-Shreeve, P. (2023). Realising smarter stormwater management: A review of the barriers and a roadmap for real world application. *Water Research*, *244*, 120505.

Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems*, *34*, 24804-24816.

van der Werf, J (2023). ICUD 2024 Extended Abstract Template, TU Delft