

# Probabilistic Intraday Wastewater Treatment Plant Inflow Forecast Utilizing Rain Forecast Data and Sewer Network Sensor Data

Björn Sonnenschein<sup>1</sup> and Florian Ziel<sup>2</sup>

<sup>1</sup>Hochschule ruhr west

<sup>2</sup>University of Duisburg-Essen

August 23, 2023

# Probabilistic Intraday Wastewater Treatment Plant Inflow Forecast Utilizing Rain Forecast Data and Sewer Network Sensor Data

Björn Sonnenschein<sup>1,2</sup>, Florian Ziel<sup>1</sup>

<sup>1</sup>Environmental Economics, esp. Economics of Renewable Energy, Universität Duisburg Essen, Essen, Germany

<sup>2</sup>Institute of Business Administration, Hochschule Ruhr West, Mülheim an der Ruhr, Germany

## Key Points:

- We propose a model for probabilistic wastewater treatment plant inflow prediction utilizing rain forecast data to handle rain events.
- Non-linear dynamics within the sewer system are approximated in a linear framework using rectified linear units.
- Our results indicate similar performance to less intelligible non-linear models, but rain forecast accuracy constitutes a limitation.

---

Corresponding author: Björn Sonnenschein, [bjoern.sonnenschein@hs-ruhrwest.de](mailto:bjoern.sonnenschein@hs-ruhrwest.de)

**Abstract**

Forecasting of wastewater treatment plant inflow dynamics constitutes an enabler technology for wastewater treatment process optimization using model predictive control. However, accurate inflow prediction is still challenging, especially for strong rainfall events, where complex system dynamics and missing information on future rainfall represent limiting factors. We propose a seasonal probabilistic time series model for modelling the short-term wastewater inflow accurately while providing quantification of forecast uncertainty. To ensure suitability for practical implementation, the unconstrained parameters of the predictive distribution are modelled as linear functions of the input variables in the framework of Generalized Additive Models for Location Scale and Shape. Non-linear effects are approximated by Rectified Linear Units, accounting for buffering within the sewer network and flow-dependent catchment response time. In addition to water level measurements from within the sewer network and rain rate measurements, rain forecasts are incorporated as exogenous regressors, where historical rain forecasts are used for model calibration. The model performance is evaluated on historical data from a German wastewater treatment plant using deterministic and probabilistic scoring rules. We benchmark against an autoregressive time series model and a long short-term memory artificial neural network. Our results show that the proposed model unites the benefits of high prediction accuracy of the neural network and enhanced intelligibility of the autoregressive model, but accurate real-time rain forecasts are mandatory for successful real-world implementation.

**Plain Language Summary**

The wastewater treatment process accounts for a large proportion of a wastewater treatment plant's energy consumption. Here, an accurate prediction of future wastewater inflow is an important tool for the optimization of energy efficiency. The potential for optimization is particularly large for rain events. However, especially in such situations, inflow prediction is most challenging. We developed a specialized computer algorithm for inflow prediction that utilizes weather forecasts in order to produce reliable inflow predictions in such situations. The developed algorithm is a machine learning model that learns the knowledge needed to perform accurate predictions from historical data. Although capable machine learning models are generally complex, we aimed to design the algorithm in such a way that it is understandable and hence trustworthy for practitioners so that it can easily be used in real-world application. We tested the algorithm and our results show that it is possible to generate precise wastewater inflow predictions, but very accurate and quickly available weather forecasts are mandatory, which is an interesting objective for further research.

**1 Introduction**

Wastewater inflow rate and composition represent central quantities for the operative control of wastewater treatment plants (WTP) (Newhart et al., 2019; Robecke, 2015; Leitão et al., 2006). In particular, the biological treatment process is sensitive to the inflow dynamics. The oxygen aeration rate in aerobic treatment constitutes a central control parameter that must be adjusted hours in advance of strong changes in inflow dynamics (Newhart et al., 2019; Makinia & Zaborowska, 2020). However, the aeration process is the dominant energy consumer in a WTP (Drewnowski et al., 2019). Thus, operating with safe margins to account for quick changes in inflow dynamics while ensuring compliance with legal limits of WTP outflow pollutant concentration leads to non-optimal energy efficiency (Bodik & Kubaska, 2013; Z. Li et al., 2019), particularly for strong rain events that result in a vast increase in WTP inflow rate (inflow) with a leading spike of pollutant concentration (first-flux phenomena). Here, accurate predictions of future inflow characteristics play a crucial role as input quantities for model predic-

tive control schemes and decision-making processes. Thus, these predictions contribute to enhancing the efficiency of WTP operations and also support the enhancement of climate resilience with respect to the increased risk of heavy precipitation events associated with climate change (D. Zhang et al., 2018; Q. Zhang et al., 2019; Z. Li et al., 2019; Stentoft et al., 2019; Nissen & Ulbrich, 2017). Furthermore, active sewer network control utilizing forecasted flow properties can be employed to optimize the usage of a network's buffering capacity to prevent overflows and WTP overloads and hence enhance wastewater system resilience (D. Zhang et al., 2018; Seggelke et al., 2013; Garofalo et al., 2017; Svensen et al., 2021).

However, it has been shown that predicting inflow dynamics is particularly challenging for strong rainfall events (Pedersen et al., 2016). Whereas inflow dynamics are dominated by diurnal seasonality in case of dry weather (Leitão et al., 2006; Q. Zhang et al., 2019), exogenous information on future rain is required for successful inflow prediction for strong rain events with a sufficiently long forecast horizon (Pedersen et al., 2016; Q. Zhang et al., 2019). Yet, utilizing exogenous information for inflow prediction is challenging due to the complex hydrodynamical properties of wastewater sewer networks (Q. Zhang et al., 2019).

Regarding suitability for implementation into existing WTP control structures, inflow forecasting methodology is subject to further constraints in addition to the challenges posed by complex system dynamics: Low computational demand is necessary in order to generate predictions quickly enough for real-time control. In addition, the ability to quantify forecast uncertainty represents an important aspect for sophisticated decision-making (Stentoft et al., 2019; Svensen et al., 2021), rendering probabilistic forecast models favourable over deterministic methods. Furthermore, it has been suggested in personal communication with practitioners and in the literature (Vaughan & Wallach, 2020) that black-box-models like artificial neural networks face a generally rejective attitude from practitioners due to security concerns for critical infrastructure. Thus, interpretability and intelligibility represent important aspects to foster trust in new methodology and hence facilitate practical utilization. Finally, applicability to different wastewater sewer networks without manual calibration and re-modelling to account for the specific network structure is advantageous due to low cost and time consumption.

Hydrodynamic, machine learning and hybrid approaches for inflow rate and composition prediction exist in the literature. Machine learning models utilize algorithms to automatically learn patterns and relationships in data, allowing for efficient and scalable modeling of complex dynamics without relying on explicit expert knowledge (Bishop, 2006) and are therefore a current topic of research for intraday inflow forecasting. Used input variables vary throughout the literature with a focus on autoregressive properties of inflow and rain measurements: Zhou et al. (2019) implemented a random forest model for generating probabilistic forecasts of daily inflow using climate data as exogenous regressors, where historical measurements were used as oracles substituting climate forecasts. D. Zhang et al. (2018) used a Long Short-Term Memory (LSTM) to predict future inflow for load-balancing between multiple WTPs and improved buffering capacity usage within the sewer network. Pedersen et al. (2016) demonstrated the utilization of a simple rainfall-runoff model (lumped reservoir model) for predicting the WTP inflow rate's response on rainfall events, proposing a Bayesian update of estimated model parameters, but did not incorporate rain forecasts. Heinonen et al. (2013) utilized rain radar forecasts as input for a hydraulic model to predict intraday WTP inflow. Wang et al. (2019) combined Convolutional artificial neural networks (ANN) and LSTM ANN in order to forecast WTP inflow chemical oxygen demand (COD) from temperature, pH, NH<sub>3</sub>-N, inflow rate and COD data. El Ghazouli et al. (2021) predicted the inflow using an autoregressive ANN with real-time and predicted water consumption as well as infiltration flows as exogenous variables. Langeveld et al. (2017) developed an empirical model for inflow quality prediction, modelling different water quality processes for

individual inflow dynamics regimes. F. Li and Vanrolleghem (2022) utilized a multi-objective genetic algorithm to train an LSTM network with respect to influent average behaviour and variability.

In this article, we propose an inflow forecasting model designed with the research goal to develop a machine learning method for probabilistic intra-day inflow forecasting with a forecast horizon of 2.5 hours and a focus on rain events. Here, measurement data from within the sewer network alongside rain forecast data shall be utilized while aiming to provide a balance of intelligibility and prediction accuracy suitable for practical implementation. Furthermore, the method should be designed not to require detailed knowledge of the particular sewer system to which it is applied or reliance on hydrodynamical simulations. Instead, it should generalize to allow coupling with physically-based methods as model inputs, such as a hydraulic surface-runoff model. This flexibility is intended to enhance the model’s applicability across various scenarios while maintaining its core data-driven characteristics. We state the following major contributions of this research:

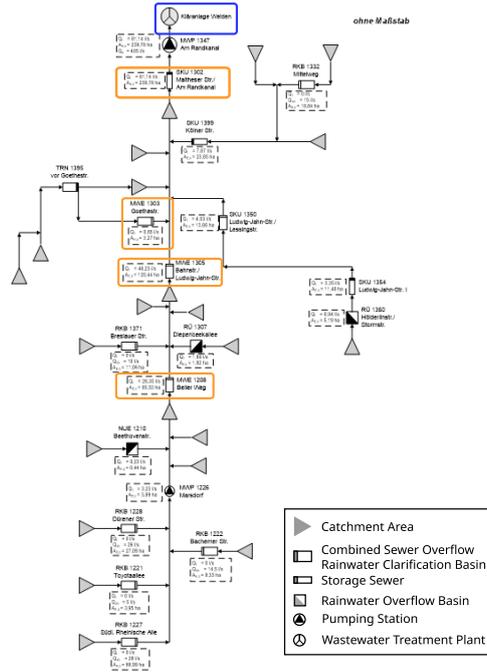
1. To our knowledge, it is the first work on a distributional machine learning inflow forecasting model that is trained on historical rain forecasts to facilitate forecasting of intraday flow dynamics after rain events.
2. Due to the probabilistic approach, the generated forecasts yield detailed information on forecast uncertainty from the full predictive distribution.
3. We demonstrate that, by modelling the unconstrained distribution parameters as linear functions of the input variables, the non-linear relationships in the data resulting from the system dynamics after rain events can be approximated so well that the inflow forecast uncertainty is dominated by the rain forecast uncertainty.
4. The proposed model exhibits good adaptability to different wastewater networks without extensive expert knowledge because the central hyperparameters are tunable based on data statistics and automated parameter search. Furthermore, the utilization of LASSO regularization improves robustness against model overspecification.

The model was trained with and evaluated on data from a combined sewer network located in the district Weiden of Cologne, Germany, covering a relatively small catchment area of  $1.01 \times 10^7 \text{ m}^2$  with approximately 52300 residents and a population equivalent (PE) of approximately 80 000 PE. The network’s layout is depicted in Figure 1. The remainder of the paper is organized as follows: Important system properties considered for the model design and the developed forecasting model are described in Section 2. The used dataset is described in Section 3. Experimental results are given in Section 4 and a concluding discussion is provided in Section 5.

## 2 Materials and Methods

We propose a machine learning model for generating distributional inflow forecasts from time series data. The model, which is described in Section 2.2, was designed to approximate non-linear relationships arising from the system’s physical properties that are described in Section 2.1. Our procedure for generating forecasts using our model specification (equation 2) consisted of the following steps:

1. Model inputs were measured inflow, measured water levels, measured rain rates and future rain rates (rain forecasts or, in part of our evaluation, historical rain rates as oracles).
2. The model hyperparameters were tuned as described in Section 2.3.2 by iterative re-calibration on the training data in order to adjust the model specification to the data of the investigated sewer network. Optionally, knowledge of system prop-



**Figure 1.** Layout of the analyzed sewer network, highlighting water level sensors (orange) and the rain gauge (blue) employed as exogenous regressors in our model.

erties can be induced in this step by explicitly defining nodes of linear splines for individual inputs where behavior changes are expected. We performed the tuning on a point-forecast model for computational performance and used the best-fit model specification and coefficients as initial values for the distributional forecast model training the next step.

3. The model coefficients were estimated using the training data as described in Section 2.3.1.
4. Predictions on the test dataset were generated using the calibrated distributional forecast model.

The sewer network’s properties we considered relevant for the model design are described in Section 2.1 and the model is specified in Section 2.2.

## 2.1 Properties of Considered Sewer Network

The proposed model was designed to address two dominant characteristics of sewer networks that simple autoregressive models struggle to capture because of the emergence of non-linear effects. These characteristics include the wastewater buffering capacity within the sewer network and the relationship between flow velocity and flow rate.

Buffering by storage elements represents a common design principle of sewer networks in order to alleviate the sewer network’s limited hydraulic capacity in case of strong rainfall events (Butler et al., 2018). Hence, a strong increase in network inflow does not result in an equivalent increase in future WTP inflow. In the case of the studied sewer network, the inflow rate is limited to approximately 700l/s after a rain event and remains at this level until the buffering elements are depleted. A storage tank sewer (tank sewer) (Butler et al., 2018) located upstream from the WTP was identified as an important element in the studied sewer network because its state of depletion was expected to pro-

vide valuable information on the expected time of inflow decrease after a rain event. In the remainder of this paper, the three phases of rising inflow, near-constant inflow near the hydraulic capacity limit and the following decline of inflow are referred as rise, high and decline.

The dependency of the flow velocity  $V$  on the flow rate follows from the Manning equation (Hager, 2010) that delivers a first-order approximation for the open surface flow regime. In particular, the hydraulic radius contributes to  $V$  and depends on the water level and the pipe cross-section (Hager, 2010). Accordingly, the catchment response time (Giani et al., 2021) and hence the temporal dependency structure between measurements in the network and future inflow are not constant but depend on the system’s hydrodynamic state.

## 2.2 Model

In this section, the time series data from one input data source (e.g. one water level sensor location or rain station) is referred as measurement variable whereas derived quantities from the input data are referred as features. The matrix  $\mathcal{X}$  corresponds to the measured data, where the column vectors are the individual measurement variables’ time series. Hence, a measured value of measurement variable  $v$  at time  $t$  is referred as  $\mathcal{X}[v, t]$ .

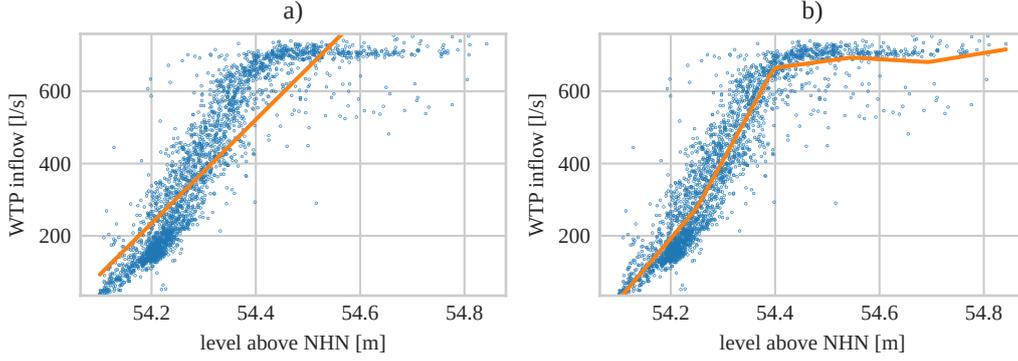
We propose a Generalized Additive Model for Location, Scale and Shape (GAMLSS) (Stasinopoulos & Rigby, 2008) for probabilistic inflow prediction. Within the GAMLSS framework, the response variable’s probability density function (PDF)  $f(Y|\mathbf{x}_p)$  is conditioned on a vector  $\mathbf{x}_p \in \mathbb{R}^{N_p}$  of  $N_p$  realized explanatory features. With a PDF with  $N_p$  parameters indexed by parameter index  $p \in \{1, \dots, N_p\}$ , each PDF parameter  $\Theta_p$  is a sum of  $J_p$  non-parametric functions  $S_{p,j}$  and a linear component with parameter vector  $\beta_p$ , mapped to the PDF’s domain using a monotonic link function  $g_p$ :

$$g_p(\Theta_p) = \mathbf{x}_p^T \cdot \beta_p + \sum_{j=1}^{J_p} S_{p,j}(\mathbf{x}_p[j]) \quad (1)$$

The proposed model only incorporates the linear component in favour of model interpretability and computational efficiency. Hence  $S_{p,j}$  simply represent the identity and are therefore omitted. Our formulation of  $\mathbf{x}_p$  incorporates features corresponding to the autoregressive structure of the data generating process and seasonal effects. We designed the autoregressive features to approximate the non-linear effects described in Section 2.1 and to model interactions between measurement variables. The resulting expression for  $g_p(\Theta_p)$  is:

$$\begin{aligned} g_p(\Theta_p) &= \sum_{v \in V} \sum_{l \in L_{p,v}} \sum_{\tau \in T_{p,v}} \beta_{p,v,l,\tau}^{ar} \min(\mathcal{X}[v, t-l], \tau) \\ &+ \sum_{a \in S_a} \sum_{w \in S_w} \sum_{d \in S_d} \beta_{p,a,w,d}^{se} \phi_a(t) \mathbf{1}_w(t) \mathbf{1}_d(t) \end{aligned} \quad (2)$$

Here, the first term corresponds to the autoregressive component, and the second term represents the seasonal component. The coefficients  $\beta_{p,v,l,\tau}^{ar}$  and  $\beta_{p,a,w,d}^{se}$  were estimated from the data in the model training process, as described in Section 2.3.  $V$ ,  $L_{p,v}$  and  $T_{p,v}$  are sets of measurement variables, lags and thresholds.  $S_a$ ,  $S_w$  and  $S_d$  are seasonal indices indicating the phase of annual, weekly and diurnal seasonality with corresponding indicator functions  $\mathbf{1}_w(t)$  and  $\mathbf{1}_d(t)$  to express seasonal dummies (Hylleberg et al., 1993).  $\phi_a(t)$  denotes B-Spline (Ziel et al., 2016) basis functions for modelling a smooth change of diurnal and weekly patterns throughout the year. The individual components of equation 2 are motivated and explained further in the following.



**Figure 2.** Illustration of linear regression without (a) and with (b) approximation of non-linear relationship by linear splines using five clipped versions of the explanatory variable with evenly spaced thresholds as regressors. The shown data is the measured WTP inflow versus a water level measurement from within the sewer network from 15 minutes in the past that is also depicted in figure 5.

### 2.2.1 Autoregressive component

As described in Section 2.1, measurements of rain and water levels provide information on the future inflow. The according temporal dependency structure was modelled as an autoregressive process (Hamilton, 2020):  $g_p(\Theta_p)$  at time  $t$  was formulated as linear combination of past realizations of the input variables with coefficients  $\beta_{p,v,l}$ :

$$g_p(\Theta_p) = \sum_{v \in V} \sum_{l \in L_{p,v}} \beta_{p,v,l} \mathcal{X}[v, t-l] \quad (3)$$

Here,  $L_{p,v}$  denotes sets of lags determining which past time steps of a each variable  $v$  are used.

### 2.2.2 Non-linear effects

Non-linear effects of individual lagged variables were modelled by piecewise linear approximation using Rectified Linear Units (ReLU) (Ziel et al., 2016; Nair & Hinton, 2010). A set of thresholds  $T_{p,v}$  was defined for each PDF parameter and variable, dividing the variable's domain into intervals contributing with individual slopes. The according features were constructed as clipped values,  $\max(\mathcal{X}[v, t-l], \tau)$ , for each threshold  $\tau \in T_{p,v}$ . In order to include completely linear behaviour, the threshold  $\tau = -\infty$  was included. The choice of appropriate thresholds was part of the hyperparameter tuning procedure described in Section 2.3. The resulting auto-regressive term from equation 3 with linear approximation for each considered lag of each measurement variable is:

$$g_p^{ar}(\Theta_p) = \sum_{v \in V} \sum_{l \in L_{p,v}} \sum_{\tau \in T_{p,v}} \beta_{p,v,l,\tau} \min(\mathcal{X}[v, t-l], \tau), \quad (4)$$

The approximation method is illustrated in Figure 2 for the rain event data from Figure 5. By modelling the non-linear effects of each lag, changes of the autoregressive dependency structure throughout different regimes of the network state, in particular the buffering and the flow-dependent dwell-time (see Section 2.1), can be approximated. As a simplified example, consider a level measurement, where a higher level will cause a more immediate effect on inflow than a lower level due to the shorter catchment response time. In this case, the coefficient  $\beta_{p,v,l,\tau}$  would be estimated significantly differently from zero

for accordingly large  $\tau$  at short lags, whereas a value near zero would be found for the same  $\tau$  at long lags. However, it is important to note that the actual relations between coefficients in the trained model are more complex in general, for example encoding derivatives through finite differences.

### 2.2.3 Seasonal effects

As described in Section 3.1, the inflow exhibited diurnal patterns varying throughout the week and the year. The diurnal and weekly patterns were modelled using seasonal dummies as artificial features with value 1 for each time  $t$  corresponding to the represented daytime and weekday and 0 else. As components of the feature vector, these seasonal dummies represent a time-dependent intercept by multiplication with their corresponding model coefficients. With the data frequency of 15 min yielding 96 phases a day and three modelled diurnal patterns (working day, Saturday, Sunday),  $96 \cdot 3 = 288$  seasonal dummies were incorporated for one annual season.

As described in Section 3.1, the diurnal patterns varied throughout the year. The according annual seasonality was modelled using separate diurnal and weekly seasonal dummies for a set  $S_a$  of four annual seasons. By weighting the resulting  $96 \cdot 3 \cdot 4$  seasonal dummies with quadratic B-Spline (Ziel et al., 2016) basis functions  $\phi_a(t)$  for each annual season  $a \in S_a$ , a smooth periodic transition between the annual seasons' patterns is realized. The resulting seasonal component of transformed PDF parameter  $g_p(\Theta_p)$  is:

$$g_p^{se}(\Theta_p) = \sum_{a \in S_a} \sum_{w \in S_w} \sum_{d \in S_d} \beta_{p,a,w,d}^{se} \phi_a(t) \mathbf{1}_w(t) \mathbf{1}_d(t) \quad (5)$$

Here, the seasonal dummies are expressed through a product of indicator functions  $\mathbf{1}_w(t) \mathbf{1}_d(t)$ , which is one if  $t$  corresponds to the diurnal phase  $d$  and the weekly phase  $w$ .

### 2.2.4 Rain cumulation and interaction

Due to the system's hydrodynamic properties, we expected the short-term accumulated rain rate values to influence inflow dynamics. Hence, for the measured and forecasted rain rate, the sum of the  $N$  previous values was incorporated as artificial measurement variable, where  $N = 6$  was found to provide the best fit with respect to the Bayesian information criterion (BIC), which is described in Section 2.3.

The effect of rain was expected to depend on the current inflow situation due to the system's limited hydraulic capacity. In particular, the effect of rain was expected to decrease for high inflow rates. This was modelled as a two-way interaction term of inflow at  $t - 1$  and rain rate by using the product of the two variables as artificial measurement variable. In experiments where rain forecasts were used instead of rain oracles, addition of two-way interaction of all rain forecast steps further improved the forecast. The piecewise linear approximation described in Section 2.2.2 enabled handling non-linear effects of the interaction.

### 2.2.5 Choice of PDF and link functions

As PDF, Johnson's SU (JSU) (Johnson, 1949) distribution was chosen due to its versatility for modeling non-normally distributed data, while still providing an intuitive parameterization:

$$f(Y = y | \mathbf{x}_p) = \frac{\delta}{\lambda \sqrt{2\pi}} \frac{1}{\sqrt{1 + \left(\frac{y-\xi}{\lambda}\right)^2}} \exp^{-\frac{1}{2}(\gamma + \delta \sinh^{-1}\left(\frac{y-\xi}{\lambda}\right))^2} \quad (6)$$

The JSU represents a transformation of a normal distribution with standard deviation  $\lambda > 0$  and mean  $\xi$ . It provides parameters  $\gamma$  and  $\delta > 0$  to control the skewness and kurtosis.

The following link functions were found to provide favourable results in our experiments:

$$g_1^{-1}(\gamma) = 3 \left( \frac{1}{1 + \exp^{-\gamma}} - 0.5 \right) \quad (7)$$

$$g_2^{-1}(\xi) = \xi \quad (8)$$

$$g_3^{-1}(\delta) = \frac{3}{1 + \exp^{-\delta}} + 0.05 \quad (9)$$

$$g_4^{-1}(\lambda) = \frac{\log(\exp^{5\lambda} + 1)}{5} \quad (10)$$

Here, Equations 7 and 9 represent shifted and scaled sigmoid functions (Han & Moraga, 1995) and equation 10 corresponds to the Softplus as implemented in the PyTorch framework (Paszke et al., 2019).

### 2.3 Calibration and Hyperparameter Tuning

Training the model given in equation 2 involved the estimation of the coefficients  $\beta_{p,a,w,d}^{se}$ ,  $\beta_{p,v,l,\tau}^{ar}$  from training data and the choice of hyperparameters.

#### 2.3.1 Estimation of Model Coefficients

The coefficient estimation process involved two consecutive steps to facilitate regularization while limiting computational demands during the iterative hyperparameter tuning process. First, a point forecast model was fitted using LASSO regularization (Tibshirani, 1996) to prevent overfitting and perform feature selection. The estimated coefficients were then used as starting values for the coefficients of the location PDF parameter  $\xi$  in the probabilistic model's optimization process minimizing a penalized-likelihood (Wasserman, 2004) loss function.

With LASSO, the estimate  $\hat{\beta}$  of a parameter vector  $\beta$  is obtained by minimizing the mean squared error penalized by the parameter vector's l1-norm. With the realized response variable  $\mathbf{y}$  and the design matrix  $\mathbf{X}$  for  $N$  samples, the LASSO estimator is:

$$\hat{\beta} = \min_{\beta} \left( \frac{1}{N} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 - \lambda_{\text{lasso}} \|\beta\|_1 \right) \quad (11)$$

An important property of the LASSO estimator is the ability to perform feature selection by setting individual parameters to zero. By construction, LASSO selects randomly among sets of highly correlated features (Xiao & Sun, 2019). Thus, regarding the model interpretation, an excluded parameter does not necessarily correspond to low explanatory value of the corresponding feature without considering the data's correlation structure.

The selected features were used for training the probabilistic model, under the assumption that the features selected for the point forecast model were also a viable selection for the probabilistic model. The coefficients for the PDF parameter  $\xi$  were initialized with the LASSO estimate. The coefficients for  $\gamma, \lambda$  were initialized as zero, as this resulted in higher consistency of parameter estimation than random initialization in our experiments and  $\xi$  corresponds to the JSU distribution's mean at those starting conditions. The PDF parameter  $\delta$  was set as a trainable constant, and hence not conditioned on the model inputs, because instabilities were observed in learning and forecasting otherwise. As  $\delta$  has strong influence on the JSU distribution's higher moments that give high weight to extreme values, its estimation is sensitive to outliers (Barnett et al., 1994).

Parameter estimation for the probabilistic model was performed using ADAM gradient descent (Kingma & Ba, 2014), minimizing a penalized-likelihood (Wasserman, 2004) loss function

$$L = -\log(\mathcal{L}(\beta_\gamma, \beta_\xi, \beta_\delta, \beta_\lambda|\mathcal{X})) + \alpha \cdot \text{MSE} \quad (12)$$

In equation 12, the likelihood  $\mathcal{L}$  is penalized with the mean square error MSE, weighted with a constant  $\alpha = 10$ , as our experiments showed that the penalty term improved prediction accuracy with respect to the MSE without reducing probabilistic accuracy, as evaluated using the energy score (see Section 2.4). The parameter vectors for the individual PDF parameters were optimized independently in a cyclic scheme with independent instantiations of the ADAM optimizer, which improved the convergence rate compared to simultaneous optimization.

### 2.3.2 Hyperparameter Tuning

Hyperparameters for our model include the choice of thresholds  $T_{p,v}$  for the linear splines, lags  $L_{p,v}$  and the regularization parameter  $\lambda_{\text{lasso}}$ .  $\lambda_{\text{lasso}}$  was tuned by an exponential grid search with respect to the Bayesian information criterion (BIC) (Stoica & Selen, 2004) for each model configuration. The BIC provides a measure for the balance between model complexity and in-sample prediction accuracy and can be used for optimizing the out-of-sample prediction performance based on tuning with in-sample data. It penalizes the negative in-sample likelihood  $\mathcal{L}$  with the parameter count  $k$  and the number of observations  $n$ :

$$\text{BIC} = k \log(n) - 2 \log(\mathcal{L}) \quad (13)$$

The utilization of feature selection with LASSO facilitated the definition of the thresholds and lags as densely populated sequences without incorporation of extensive expert knowledge for exact definition of each threshold and lag for each variable. The thresholds for each variable were defined as equidistant sequences with variable-specific number of steps, and with minimum and maximum values determined as follows: For rain measurement variables, the smallest measured rain rate from any rain event was chosen as minimum and the 0.95-quantile was defined as maximum. For other measurement variables, the smallest measured value and the 0.99-quantile were chosen as minimum and maximum values respectively. Results for different numbers of thresholds are provided in Section 4. The lags were chosen as the past six timesteps,  $L_{p,v} = \{1, 2, 3, 4, 5, 6\}$ , because using more lags did not yield an improvement with regard to the BIC.

## 2.4 Forecast study design

The proposed model's forecasting performance was evaluated on in-sample and out-of-sample data. In-sample evaluation indicates how well the model learns from the training data, while out-of-sample evaluation gauges its generalization to new data. The dataset covering three years was split into an in-sample dataset comprising two consecutive years and an out-of-sample dataset for the subsequent year. This division ensured the maximum utilization of training data while still encompassing all annual seasons in the out-of-sample data.

In order to analyse the effects of water level, realized and future rain rate as well as of the individual model components described in Section 2.2, different model configurations were implemented and calibrated with the procedure described in Section 2.3. In addition to the model's overall prediction performance over the complete dataset, an analysis for different inflow situations was of interest with regard to the practical use cases described in Section 1. Hence, individual analysis was performed for subsets of the data representing dry weather periods and the rise, high and decline phases after rain events.

### 2.4.1 Scoring Rules

An appropriate choice of scoring rules with respect to the studied forecasting problem is important to quantify the prediction performance with respect to the realized (truth) values (Ziel & Berk, 2019). The probabilistic forecasts generated by the proposed model can be reduced to point forecasts by calculating the mean  $\hat{\mathbf{y}}_t \in \mathbb{R}^H$  for all steps of the forecast horizon  $H$ , with  $t$  denoting the time of the first forecast horizon step, so that the commonly used (Bjerregård et al., 2021) scoring rules root-mean-square error (RMSE)

$$\text{RMSE} = \frac{1}{N} \sum_{t=1}^N \sqrt{\frac{1}{H} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2} \quad (14)$$

and mean absolute error (MAE)

$$\text{MAE} = \frac{1}{NH} \sum_{t=1}^N \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_1 \quad (15)$$

can be utilized with respect to the truth  $\mathbf{y}_t \in \mathbb{R}^H$ . Here,  $N$  is the number of evaluated datapoints.

However, while those point forecast scores provide good interpretability, the information on the predictive distribution is discarded. Here, probabilistic scoring rules are available in order to quantify the predicted distribution's correspondence to the unknown distribution of the underlying data generating process. We used the energy score (Gneiting & Raftery, 2007) as a probabilistic scoring rule. The energy score is a strictly proper scoring rule that can be estimated from  $M$  independently sampled trajectories  $\tilde{\mathbf{y}}_{t,m}$  of the inflow over the forecast horizon with:

$$ES = \frac{1}{M} \sum_{i=1}^M \|\mathbf{y}_t - \tilde{\mathbf{y}}_{t,m}\|_2^\beta - \frac{1}{2M^2} \sum_{i=1}^M \sum_{j=1}^M \|\tilde{\mathbf{y}}_{t,i} - \tilde{\mathbf{y}}_{t,j}\|_2^\beta \quad (16)$$

where  $\beta = 1$  is chosen in accordance to common literature (Ziel & Berk, 2019).

### 2.4.2 Statistical Significance Test

We employed the Diebold-Mariano (DM) test (Mariano, 2002) to assess the statistical significance of a forecast  $A$  exhibiting better accuracy than forecast  $B$ . The DM test is applicable to the present multivariate setting with possible dependency structure between the individual forecast steps (Ziel & Berk, 2019). The DM test resembles an asymptotic test against the null-hypothesis that the mean score difference  $\bar{\Delta}_{A,B}$  is zero with test statistic

$$t_{\text{DM}} = \frac{\bar{\Delta}_{A,B}}{\hat{\sigma}(\bar{\Delta}_{A,B})} \quad (17)$$

with the estimated standard deviation  $\hat{\sigma}$ . The results from the forecast study were tested with a significance level of 5%.

### 2.4.3 Sampling

The probabilistic predictions were generated by sampling  $M = 80$  trajectories for each time  $t$  representing the last measurement time before the respective forecast horizon. Each trajectory was constructed using Monte Carlo (Wasserman, 2004) simulation by sampling each step  $h$  of the forecast horizon consecutively. For each forecast step, the sampled values of the trajectory's previous steps were used as inflow measurement variables for lags  $l < h$  referring to future and hence unobserved values. As a result, the constructed samples represented the inflow's autoregressive dependency structure throughout the forecast horizon.

#### 2.4.4 Benchmarks

In order to benchmark the proposed model’s performance, two benchmark models were implemented: First, an autoregressive process with exogenous regressors and seasonal effects (Hamilton, 2020) (SARX) was implemented as a linear point-forecast model that is widely used in time-series analysis. Second, a distributional ANN (Marcjasz et al., 2023) was used as black-box benchmark model trading computational efficiency and interpretability for the ability to learn highly non-linear relationships from the data (Yu et al., 2019). For all benchmarks, the lags  $L = \{1, \dots, 6\}$  were chosen in order to condition the prediction on a history window equivalent to the proposed model. Also, the future rain oracles as well as the water level measurements were used as exogenous regressors, so that the same information was utilized as with the proposed model.

As a linear model, the SARX model is easy to interpret and shows low computational demand. The model was implemented according to the formulation used by the statsmodels (Seabold & Perktold, 2010) python library:

$$y_t = \beta_0 + \sum_{l \in L} \beta_l^{\text{ar}} \mathcal{X}[w, t - l] \sum_{v \in V} \beta_v^{\text{exo}} \mathcal{X}[v, t - 1] \sum_{s \in S} \beta_s^{\text{season}} d_s \quad (18)$$

In equation 18,  $\mathcal{X}[w, t - l]$  refers the inflow (represented by feature index  $w$ ) at lag  $l$ ,  $\mathcal{X}[w, t - 1]$  is the last measurement of feature  $v$  from the set  $V$  of exogenous regressor indices and  $d_s$  is a seasonal dummy for diurnal seasonality. Probabilistic samples were generated for each step in the forecast horizon by sampling  $M = 80$  samples from a normal distribution with the point estimate used as mean and the in-sample residuals’ standard deviation.

In the distributional ANN model, an LSTM was used to predict the unconstrained  $\gamma, \xi, \lambda$  parameters of the predictive JSU distribution. The architecture of an LSTM enables selective memorization and forgetting of information throughout lagged data, making it suitable for modelling multivariate time series data that exhibits non-linear relationships (Hochreiter & Schmidhuber, 1997). The PDF parameter  $\delta$  was set as a trainable constant like for our model as described in Section 2.3.1. For modelling of daily, weekly and annual seasonal patterns, time information was encoded as features using both, sine and cosine functions with position in day, week and year as phase. Hyperparameter tuning for the LSTM model’s dropout rate and unit count was performed using the Optuna (Akiba et al., 2019) Python package.

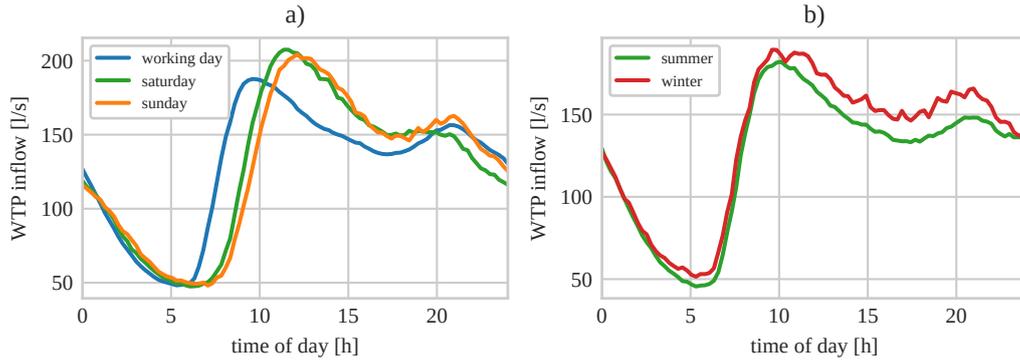
### 3 Data

The dataset used covered the period from 1st January 2017 to 31st December 2019 and consisted of the following subsets. All data was available with a frequency of 15 minutes except radar raster data with 5 minutes.

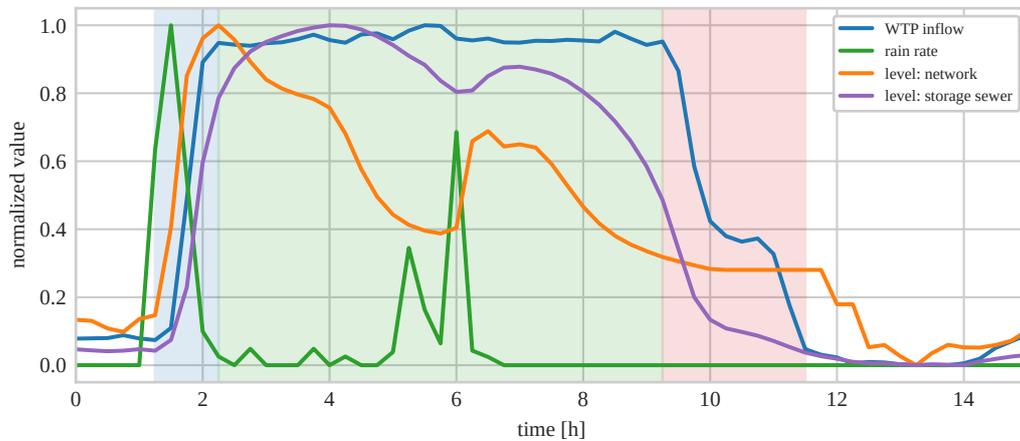
1. WTP inflow measurement (inflow)
2. Local rain rate measurement from one station located near the WTP within the sewage water catchment area of Weiden (rain rate)
3. Water level measurements from four locations spatially distributed throughout the sewer system (water level), including the tank sewer located near the WTP
4. Rainfall radar raster images covering Germany, provided by the DWD opendata service (Wetterdienst, 2022)

#### 3.1 Sewer System and Rain Measurement Data

Measured inflow minimum, maximum, average and standard deviations were 28.2 l/s 825.7 l/s 177.8 l/s and 151.1 l/s . During dry weather periods, water level and inflow rate exhibited pronounced diurnal and weekly seasonal patterns. In addition, an-



**Figure 3.** Diurnal median inflow profile for weekly (a) and annual (b) pattern.

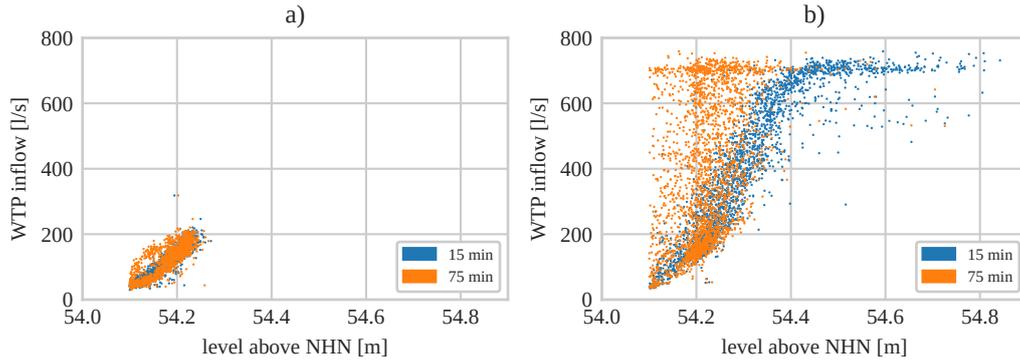


**Figure 4.** Example strong rain event followed by phases of rising (blue background), constant (green background) and declining (red background) inflow with corresponding response of measurement variables.

nual seasonality was observed. Diurnal, weekly and annual seasonalities are shown in Figure 3.

Figure 4 illustrates the system dynamics after strong rain events: due to the short catchment response time, the surface runoff dynamics and the broad spatial distribution of the rain event, inflow starts to rise steeply with a short delay of 15 to 30 minutes after rainfall onset. Simultaneously, buffering within the network occurs. The water level in a buffering element at a remote location within the network starts to rise immediately, whereas the water level in the tank sewer near the WTP starts to rise later. A proportion of the water from the quickly depleting remote buffering elements arrives at the tank sewer, causing a longer phase of water level increase until depletion starts. Hence, after the hydraulic capacity limit at the WTP has been reached, the inflow remains approximately constant until the buffering elements are depleted and the inflow then decreases quickly towards dry weather levels.

Figure 5 shows inflow measurements plotted against past water level data from a remote location within the network with lags of 15 minutes and 75 minutes for each, dry weather periods and rise. Two important system properties explained in Section 2.1 are



**Figure 5.** Inflow versus a level measurement from within the sewer network from 15 and 75 minutes in the past for dry weather (a) and after strong rain events (b). Some non-linearity can be observed particularly after rain events. The shown measurements cover a period from 1st January 2017 to 31st December 2019.

emergent: first, a non-linear relationship between water level and inflow is observed especially after strong rain events due to limited hydraulic capacity, buffering and complex flow dynamics. The non-linear effects are less pronounced in the dry weather regime, indicating that this regime can be modelled well with a linear regression model. Second the inflow measurement correlates well with the water level from 75 minutes before during dry weather periods. This indicates that the water level measurements might provide information on future inflow due to the long catchment response time for relatively large forecast horizons. After rain events, in contrast, a pronounced relation is observed for the 15-minute lag, but much less for 75 minutes, indicating the effects of buffering and regime change in flow dynamics with short catchment response time.

### 3.2 Rain Forecast Data

Historical rain radar forecast data with sufficiently high frequency of at least 15 minutes, as required for model training and calibration, was not available. Hence, radar forecasts have been constructed in analogy to the DWD RADVOR product (Winterrath & Rosenow, 2014): The dense optical flow between subsequent raster images has been estimated using the Robust Local Optical Flow (RLOF) (Senst et al., 2012) algorithm. The optical flow was then used to interpolate the 5-minute frequency data to 1-minute resolution in order to prevent fencing artefacts. The resulting raster images were temporally averaged to 15-minute frequency. For the resulting 15-minute averaged raster images, RLOF was calculated between subsequent images. The resulting optical flow was then used to extrapolate the 15-minute raster images to forecasts with 15-minute frequency with a forecast horizon of 2.5 hours. From the constructed raster image forecasts, a time series was generated for each forecasting step by spatial aggregation over the network's catchment area for each image of the corresponding forecasting step. The absolute physical dimension of the rain rate was not preserved throughout the process, which is, however, not relevant for the machine learning inflow forecasting model.

### 3.3 Preprocessing

The water level data exhibited frequent sharp local minima, indicating invalid sensor data. For each measurement station, an individual threshold was defined by visual examination of the data. Datapoints below the given thresholds were replaced by the

last valid measurement. In addition, prolonged periods of constant values were present in water level data. According to personal communication, those datapoints represented missing sensor data. The corresponding datapoints were marked as invalid and were excluded from this research. Rain rate data was clipped by a threshold of 0.085 mm/min because our experiments indicated that very high rain rate values did not provide additional information due to the system’s limited hydraulic capacity, while causing a reduction of model stability by constituting a sparsely populated extreme data regime.

Due to daylight saving time, a one-hour shift in diurnal inflow patterns was observed during summer time. One hour of interpolated measurements was inserted at the beginning of summer time and one hour of data was removed at the end of summer time in order to achieve consistency with respect to the UTC time for straight-forward modelling of the seasonal patterns.

## 4 Results

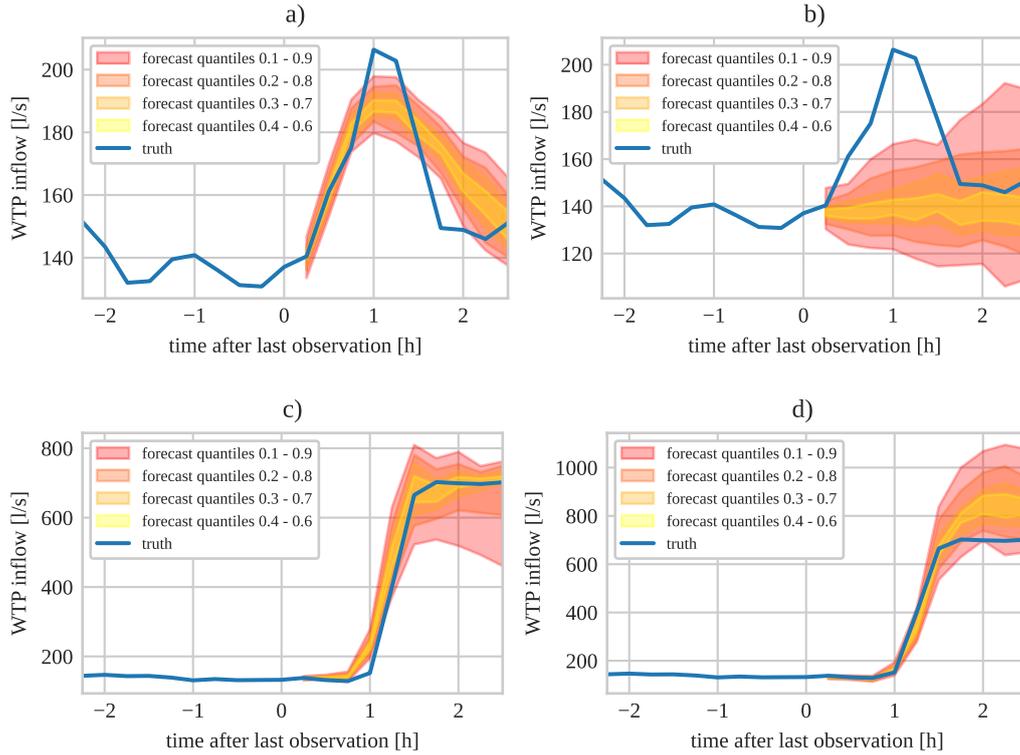
In Section 2, we proposed a machine learning model for probabilistic short-term WTP inflow rate prediction utilizing rain forecast data and water level measurements from the sewer network. In this section, our results of the forecast study described in Section 2.4 are provided and discussed. In Section 4.1, the proposed model’s ability to learn the data dependency structure arising from the system properties is examined without the further inaccuracy added by the uncertainty of rain forecasts by using historical rain values as perfect (oracle) rain predictions. The influence of rain forecast inaccuracy in a real-world scenario is then investigated in Section 4.2 using historical rain forecasts constructed according to Section 3.2. We focus on out-of-sample scores in order to analyse the performance in a real-world scenario and in-sample results are provided in the supporting information. Hence, score values mentioned in this section refer to out-of-sample values. Relative improvements refer to the energy score.

### 4.1 Forecast Accuracy with Rain Oracle

Scores over all data and the subsets dry weather, rise and decline are provided in Table 1. The results indicate that the inflow during dry weather conditions can be forecasted accurately by a simple model configuration because the process is dominated by seasonal effects. In case of rain events, however, the approximation of non-linear effects by splines and the utilization of future rain as well as network water level data yielded a significant enhancement of prediction quality. In the following, the effects of the different model design aspects described in Section 2.2 are discussed for dry weather periods and the rise, high and decline phases after rain events..

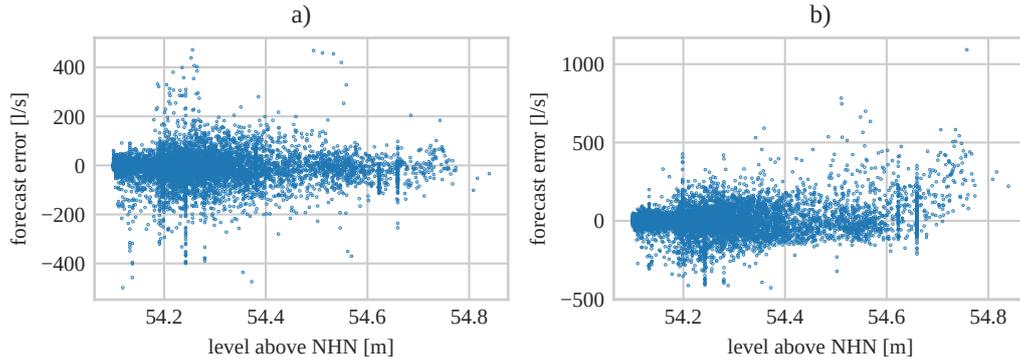
The energy score for dry weather periods without approximation of non-linear effects, using only the lagged inflow and seasonal terms as regressors is 37.461/s. However, the prediction accuracy for rise after rain events is low with an energy score of 527.111/s due to the lack of information on the effects of future and past rain.

Utilization of water level data resulted in only a small improvement of 8.28 % for rising inflow after a rain event, but a larger improvement of 19.16 % for all data, indicating that the catchment response time after strong rainfall events became so short that the level measurements did not provide information on imminent inflow rise events. A considerable improvement of 21.88 % is found for the decline phase. For the decline phase, we have identified the water level measurement from the tank sewer located before the WTP as most important level measurement. Its water level provided direct information on the buffering state, as anticipated in Section 2.1. A large improvement by 26.67 % is observed for dry weather periods, indicating that exogenous level measurements did indeed provide usable information on future inflow within the catchment response time. In Figure 6, an example is shown where level measurements enabled the prediction of



**Figure 6.** Probabilistic inflow forecast by the proposed model with (a) and without (b) utilization of level measurements from within the network. The observed inflow spike is not explained by a rain event. (c) and (d) show a probabilistic inflow forecast by the proposed model for a strong rain event, with and without the approximation of non-linear effects by linear splines, respectively.

a temporary inflow increase that is not explained by a measured rain event. We assume a low-intensity and spatially constrained rain event that did not cover the rain measurement station as a possible explanation for this peak. Due to the moderate increase of flow rate within the network, the catchment response time was still sufficiently long to enable forecasting using water level measurements. For the successful prediction of the rise phase after rain events, our results show that information on future rain is mandatory. For this phase, incorporation of measured rain data yielded an improvement over the model with inflow and water level measurements of 17.62%. A further improvement by 29.43% was achieved using future rain information. Still, high forecast inaccuracy is observed during the transition from the steep inflow increase after a rain event to the high inflow phase when the hydraulic capacity limit is reached. In this regime, the inflow was often under- or overestimated, depending on the temporal structure of the associated rain event. A visual example is depicted in Figure 6. The incorporation of linear splines for the approximation of non-linear behaviour resulted in an improvement over all data of 16.49% for 3 thresholds per measurement variable and a further improvement of 8.29% for 15 thresholds. For the rise phase after rain events, the improvements are 21.82% and 14.76%. The aforementioned overestimation and fluctuations in predicted inflow after inflow rise are reduced, indicating that the non-linear effects were approximated well by the proposed model, as shown in Figure 6.



**Figure 7.** Inflow forecast error versus level measurement from within the network. The best-fit-configuration (a) does not exhibit the overestimation of inflow for high water levels found in the predictions of a model configuration without approximation of non-linear effects by linear splines (b).

Figure 7 shows the residuals as a function of a water level measurement with and without linear splines. For high water levels, overestimation of future inflow is present without linear splines, indicating that the non-linear relationship shown in Figure 7 was not modelled well. In contrast, this regime is represented well when linear splines were used. Furthermore, Figure 7 illustrates heteroscedasticity in the data and hence the expedience of conditioning the full predictive distribution on the features. There is also an improvement of 12.51 % and 10.26 % for dry weather periods. Here, it is important to note that this does not necessarily indicate that pronounced non-linear effects were present in this regime. Without linear splines, if a feature exhibits linear and non-linear relationships with the inflow for different regimes, accuracy in linear regimes is sacrificed for performance over the whole data range.

Both, two-way interaction terms of rain with inflow and cumulation of rain over the last 1.5 hours did not yield an improvement of prediction accuracy if studied separately. However, utilization of both, where the cumulated rain was used for interaction with inflow, resulted in a considerable improvement for rise after rain events by 14.83 % and also for inflow decline after the following phase near the hydraulic capacity limit by 7.40 % . The behaviour of different model configurations described above with respect to energy score is also found for the RMSE and MAE. This indicates that the proposed model was capable to quantify the forecast uncertainty reasonably well.

The proposed model without linear splines outperformed the SARX benchmark model that used the same inputs by 32.01 %. It is assumed that the proposed model's advantage originates from the modelling of weekly and annual seasonality and in particular from conditioning the full predictive distribution on the data, as the RMSE difference of 14.10 % is significantly smaller . The best-fit configuration of the proposed model outperformed the SARX by 51.88 % for all data and 53.49 % for rise after rain events due to the effects described above.

The LSTM benchmark model performed similar to the proposed model regarding both, point forecast and probabilistic scores. Over all data, the proposed model achieved a small, but statistically significant improvement of 2.94 % over the LSTM. The LSTM results show superior performance for rise after rain events by 6.19 % , indicating that the LSTM is suited well for modelling the strongly non-linear dependency structures as well as possibly complex interactions.

**Table 1.** Out-of-sample scores for different configurations of the proposed model with rain oracles.  $N_{\tau,W}, N_{\tau,n}, N_{\tau,r}, N_{\tau,i}$ : Number of linear splines for inflow, network, rain and interactions.  $R_f, R_c$ : Whether rain forecasts and rain cumulation are used. Bold and italic: Best scores of all models and proposed model configurations if statistical significance is given.

data	Model	$N_{\tau,W}$	$N_{\tau,n}$	$N_{\tau,r}$	$N_{\tau,i}$	$R_f$	$R_c$	RMSE [l/s]	MAE [l/s]	Energy [l/s]
all	GAMLSS	1	-	-	-	no	no	35.15	29.88	89.35
all	GAMLSS	1	1	-	-	no	no	28.92	24.24	72.27
all	GAMLSS	1	1	1	-	no	no	28.17	23.62	68.38
all	GAMLSS	1	1	1	-	yes	no	24.12	20.50	56.45
all	GAMLSS	3	3	3	-	yes	no	20.19	16.71	47.14
all	GAMLSS	15	15	5	-	yes	no	18.53	15.24	43.23
all	GAMLSS	15	15	5	-	yes	yes	18.39	15.12	42.72
all	GAMLSS	15	15	5	3	yes	no	18.57	15.25	43.58
all	GAMLSS	15	15	5	3	yes	yes	<b>17.24</b>	<b>14.24</b>	<b>39.95</b>
all	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	28.08	23.92	83.03
all	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	17.96	14.88	41.16
<hr/>										
dry	GAMLSS	1	-	-	-	no	no	16.28	14.08	37.46
dry	GAMLSS	1	1	-	-	no	no	11.85	9.96	27.47
dry	GAMLSS	1	1	1	-	no	no	10.48	8.67	24.71
dry	GAMLSS	1	1	1	-	yes	no	11.04	9.11	25.74
dry	GAMLSS	3	3	3	-	yes	no	9.73	8.04	22.52
dry	GAMLSS	15	15	5	-	yes	no	8.77	7.22	20.21
dry	GAMLSS	15	15	5	-	yes	yes	8.52	6.99	19.65
dry	GAMLSS	15	15	5	3	yes	no	8.81	7.25	20.31
dry	GAMLSS	15	15	5	3	yes	yes	<b>8.31</b>	<b>6.82</b>	<b>19.08</b>
dry	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	14.42	12.24	51.91
dry	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	8.48	6.94	19.46
<hr/>										
rise	GAMLSS	1	-	-	-	no	no	193.54	168.04	527.11
rise	GAMLSS	1	1	-	-	no	no	183.17	158.86	483.46
rise	GAMLSS	1	1	1	-	no	no	159.87	136.01	398.28
rise	GAMLSS	1	1	1	-	yes	no	119.87	104.00	281.08
rise	GAMLSS	3	3	3	-	yes	no	94.54	79.96	219.73
rise	GAMLSS	15	15	5	-	yes	no	81.75	68.11	187.29
rise	GAMLSS	15	15	5	-	yes	yes	81.28	67.61	186.44
rise	GAMLSS	15	15	5	3	yes	no	82.01	68.31	189.94
rise	GAMLSS	15	15	5	3	yes	yes	<b>69.96</b>	<b>58.82</b>	<b>159.52</b>
rise	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	130.59	112.94	342.96
rise	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	<b>67.26</b>	<b>56.05</b>	<b>149.65</b>
<hr/>										
decline	GAMLSS	1	-	-	-	no	no	116.26	100.39	259.18
decline	GAMLSS	1	1	-	-	no	no	88.07	76.11	202.46
decline	GAMLSS	1	1	1	-	no	no	81.59	71.42	189.32
decline	GAMLSS	1	1	1	-	yes	no	69.76	61.45	165.86
decline	GAMLSS	3	3	3	-	yes	no	61.03	52.04	142.20
decline	GAMLSS	15	15	5	-	yes	no	58.54	49.97	134.94
decline	GAMLSS	15	15	5	-	yes	yes	57.74	49.36	134.08
decline	GAMLSS	15	15	5	3	yes	no	59.03	50.30	136.63
decline	GAMLSS	15	15	5	3	yes	yes	<b>54.16</b>	<b>46.60</b>	<b>124.95</b>
decline	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	71.34	62.82	179.01
decline	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	55.55	47.44	124.92

## 4.2 Forecast Accuracy with Rain Forecasts

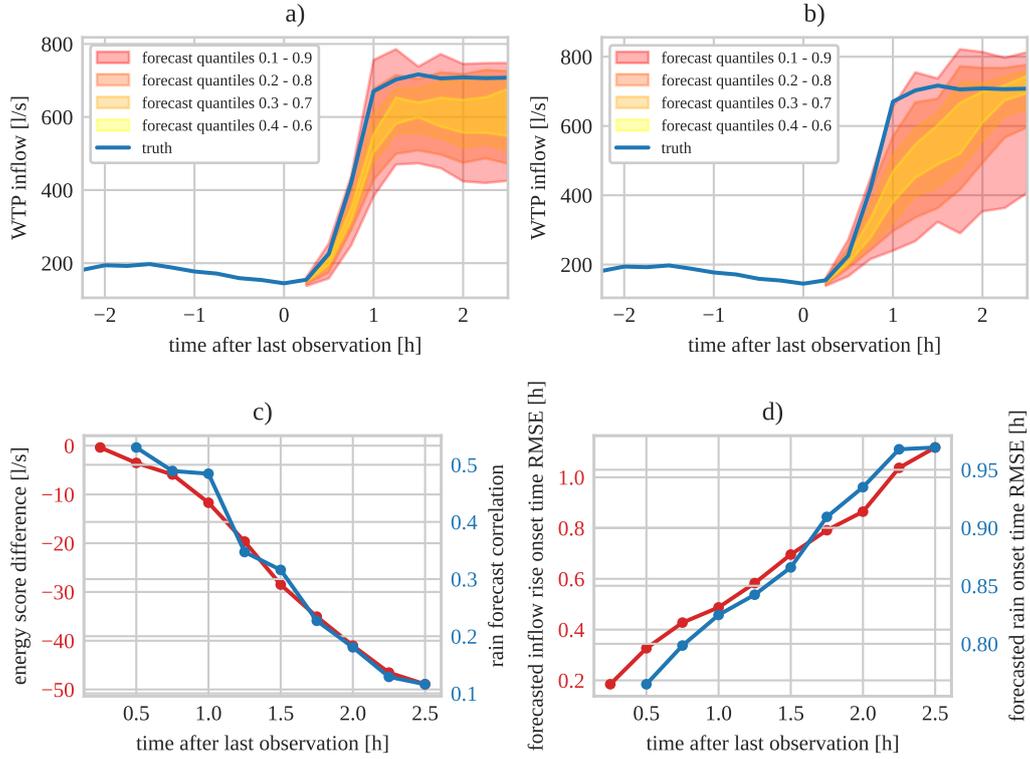
The forecast study results for the models with forecasts used for training and prediction are discussed in the following. Out-of-sample scores over all data and the subsets dry weather, rise and decline are provided in Table 2. The patterns found in the results with oracles, as described in Section 4.1, are also evident in the results with rainforecasts. However, the forecast accuracy of our best-fit model with rain forecasts is reduced significantly by 26.66% in comparison to our best-fit model with rain oracles. The inferior performance is primarily found for the phases after a rain event, indicating that the rain forecast accuracy can be identified as the limiting factor here. However, a minor increase in RMSE by 6.39% is also found for dry weather, which may also be attributed to the rain forecast accuracy, as false rain events are predicted in some situations. Figure 8 shows a comparison of rain forecasts using rain oracles and rain forecasts for an exemplary rain event, illustrating the impact of rain forecast inaccuracy for a real-world use case. Furthermore, the impact of the rain forecast error on the energy score and on the prediction of the inflow rise starting time after rain events is visualized as a function of forecast time step. The energy score disadvantage increases approximately linearly with forecast time, mirroring the decrease in correlation between the rain forecast and actual measured rain. Similarly, the RMSEs of predicted rain onset times and predicted inflow rise times exhibit an approximately linear trend. Notably, the RMSE of inflow rise start times is significantly lower than that of rain onset times, except for long forecast times. This discrepancy may suggest that the model has learned to offset some of the errors in the rain forecasts. Also, information from level measurements within the network can be leveraged, particularly for short forecast times, thereby enhancing the predictive accuracy.

Considering the benchmarks, the LSTM results show a more pronounced advantage over the GAMLSS for later steps in the forecast horizon than with rain oracles, whereas the GAMLSS still performed better for short forecast horizons. As the rain forecasts were constructed as time series from radar raster forecast, they exhibited uncertainty in the temporal domain as well as in magnitude of rain rate. Thus, we suggest that the LSTM was able to extract more predictive information from the relations in the temporal domain and the magnitude of the rain forecasts, as forecasts for the full forecast horizon were presented as features to the inflow forecast models. Furthermore, whereas the GAMLSS and LSTM showed far superior performance to the SARX model for rain oracles, the relative performance difference between the GAMLSS and LSTM in the one hand and the SARX in the other hand is smaller. This supports the finding that the rain forecast accuracy represented the main limiting factor for the inflow forecast accuracy.

## 5 Discussion

We implemented and evaluated a machine learning model for probabilistic intraday inflow prediction utilizing water level measurements from within the network, rain measurements and future rain as exogenous variables. The model was specified with respect to dominant properties of sewer networks, with a focus on non-linear effects arising from flow-dependent catchment response time and buffering of water within the network.

Our results show that the incorporation of exogenous variables provides a significant benefit for the prediction of inflow. In particular, reliable information on future rain is mandatory to achieve a prediction of the dynamics found for strong rain events with sufficient accuracy for practical use. Also, data from sensors spatially distributed throughout the network can provide valuable information to understand and predict the system behaviour. However, our results indicate that non-linear effects prevent accurate inflow prediction using a basic linear time series model like the evaluated SARX benchmark model. Using approximation via linear splines, those relationships were represented suc-



**Figure 8.** Probabilistic forecast of inflow after a rain event by the proposed model with (a) rain oracles and (b) rain forecasts. In this particular example, prediction using rain forecasts is relatively accurate, as the onset time of inflow rise is predicted well, but prediction is still significantly worse than with rain oracles. (c) depicts the average energy score disadvantage when using rain forecasts versus rain oracles for all rain events (red), alongside the Pearson correlation coefficient between forecasted and measured rain rates (blue), as functions of forecast time. Larger negative values indicate a larger disadvantage. (d) shows the RMSE of predicted WTP inflow rise onset times after rain events, using rain forecasts (red), and the RMSE of rain onset times in the rain forecasts, compared to actual measurements (blue), both as functions of forecast time.

**Table 2.** Out-of-sample scores for different configurations of the proposed model with rain forecasts.  $N_{\tau,w}, N_{\tau,n}, N_{\tau,r}, N_{\tau,i}$ : Number of linear splines for inflow, network, rain and interactions.  $R_f, R_c$ : Whether rain forecasts and rain cumulation are used. Bold and italic: Best scores of all models and proposed model configurations if statistical significance is given.

data	Model	$N_{\tau,w}$	$N_{\tau,n}$	$N_{\tau,r}$	$N_{\tau,i}$	$R_f$	$R_c$	RMSE [l/s]	MAE [l/s]	Energy [l/s]
all	GAMLSS	15	15	5	3	yes	yes	<i>21.58</i>	<i>17.54</i>	<i>50.60</i>
all	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	30.38	25.64	91.15
all	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	21.66	17.64	<b>49.14</b>
dry	GAMLSS	15	15	5	3	yes	yes	<i>8.92</i>	<i>7.33</i>	<i>20.30</i>
dry	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	15.29	12.91	57.06
dry	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	<b>8.85</b>	<b>7.20</b>	<b>20.03</b>
rise	GAMLSS	15	15	5	3	yes	yes	<i>102.35</i>	<i>84.32</i>	<i>240.92</i>
rise	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	144.99	124.41	381.28
rise	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	<b>96.76</b>	<b>80.45</b>	<b>219.74</b>
decline	GAMLSS	15	15	5	3	yes	yes	<i>66.82</i>	<i>57.07</i>	<i>152.10</i>
decline	SARX	n.a.	n.a.	n.a.	n.a.	yes	n.a.	79.63	69.93	199.86
decline	LSTM	n.a.	n.a.	n.a.	n.a.	yes	n.a.	65.64	55.95	<b>146.78</b>

cessfully while still resembling a linear model for the unconstrained PDF parameters, providing interpretability and intelligibility.

Our model generates probabilistic forecasts, providing full information on the forecasted probability distribution of future inflow. Resembling a machine learning method, the proposed model exhibits good adaptability properties as it does not rely on extensive specification of expert knowledge on the target system. Although knowledge of some of the studied system’s key properties like the watershed lag time may guide the a priori choice of included measurement variables and model hyperparameters like the lags and thresholds, feature selection and regularization using LASSO relaxes the demand for exact specification. Furthermore while the model has been designed considering physical network properties, the resulting inclusion of linear splines as a modeling technique offers versatility in capturing other non-linear relationships. Therefore, incorporating additional input variables into the model is a straightforward process through retraining.

Our results using rain oracles indicate that the proposed model as well as the probabilistic LSTM used as benchmark are capable of handling the system dynamics so well that practical application seems promising within the context of WTP process control optimization. However, high accuracy of rain forecast data is crucial to exploit the potential of the studied machine learning methods. Here, the probabilistic LSTM showed superior performance for long forecast horizons if our inaccurate baseline rain forecasts were used instead of the perfect rain oracles. This comes at the price of reduced interpretability and intelligibility, though, and the proposed GAMLSS showed marginally better accuracy for short forecast horizons. Although both, the LSTM and our GAMLSS exhibited superior performance to the SARX, we therefore consider the handling of the complex information from inaccurate rain forecasts for long forecast horizons a present

limitation of our model and propose further tuning with respect to the used rain forecasting method’s properties as a subject of further research. Furthermore, the following relevant aspects can be identified for future research:

1. Probabilistic rain forecast information (Sønderby et al., 2020) hold strong potential as input quantity of probabilistic inflow models because information on the rain forecast predictive distribution can be propagated.
2. We utilized spatially aggregated rain information as a single time series. Spatially resolved measured rain rates and rain forecast information may represent an interesting aspect of future research, especially for wastewater sewer networks with large spatial extent, where local features of rain events as well as longer catchment response time may provide additional information.
3. While this research focused on the prediction of inflow rates, the adaption of our model for forecasting also pollutant load, which represents a further relevant input for WTP optimization methodology, is of interest.
4. Regarding changing rainfall dynamics due to climate change, our results, in line with the literature, demonstrate that a data-driven model can be trained on relatively short historic time horizons, considering the gradual changes caused by climate change. To further facilitate effective adaptation to changing structures resulting from climate change, online learning schemes can be employed.
5. Incorporating physical models, such as rainfall-runoff models, as model inputs may be beneficial for explicitly modeling system dynamics where possible and desirable. This approach can reduce the complexity of the relationships that our model must approximate. Furthermore, this may be advantageous with regard to the previous point on climate change, as the dynamics within the sewer network do not change significantly with increasing rainfall strength once the network’s hydraulic capacity limit is reached. Therefore, changes in rainfall-runoff dynamics might dominate the deviations from the learned dynamics.
6. Within the scope of our study, data and resources for building a well-calibrated hydrodynamic benchmark model were not accessible. We consider situations where the adaptation of hydrodynamic models for a particular sewer network is not feasible or desired as primary use cases for data-driven models. Nevertheless, benchmarking data-driven models, such as ours, against a hydrodynamic model can be of interest. As our results demonstrate, the prediction uncertainty of our data-driven inflow forecast model and LSTM is primarily driven by the inaccuracy of rain forecasts. Here, it is an interesting research question whether data-driven methods show an advantage in adapting to and compensating for some uncertainties or suboptimal calibration of rain forecasts.

As a conclusion, our results show that exogenous information on future rain and the network’s hydrodynamic state is crucial for inflow prediction with a focus on rain events and the information can be utilized within an intelligible probabilistic machine learning forecast model suitable for practical implementation. However, accurate probabilistic rain forecasts with suitably short runtimes are required for successful application.

## Acronyms

**WTP** Wastewater treatment plant

**GAMLSS** Generalized additive model for location, scale and shape

**SARX** Seasonal autoregressive exogenous model

**ANN** Artificial neural network

**ReLU** Rectified linear unit

**LSTM** Long short-term memory

**RMSE** root-mean-square error

**MAE** Mean absolute error

**BIC** Bayesian information criterion

## 6 Open Research

The inflow, level measurement and rain measurement time series data used for model calibration and evaluation in the study are available at Zenodo via <https://doi.org/10.5281/zenodo.69926944> with Creative Commons Attribution 4.0 International license (Sonnenschein & Ziel, 2022c). The time series data of spatially aggregated rain radar forecasts used for model calibration and evaluation in the study are available at Zenodo via <https://doi.org/10.5281/zenodo.6992694> with Creative Commons Attribution 4.0 International license (Sonnenschein & Ziel, 2022b). The model implementation and the code for the generation of the results, tables and figures is available at Zenodo via <https://doi.org/10.5281/zenodo.6997701> with (Sonnenschein & Ziel, 2022a). Rain radar raster data is available on the Open Data Server of Deutscher Wetterdienst via [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/5\\_minutes/radolan/reproc/2017\\_002/bin/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/5_minutes/radolan/reproc/2017_002/bin/) Wetterdienst (2022).

## References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD international conference on knowledge discovery and data mining*.
- Barnett, V., Lewis, T., et al. (1994). *Outliers in statistical data* (Vol. 3) (No. 1). Wiley New York.
- Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4) (No. 4). Springer.
- Bjerregård, M. B., Møller, J. K., & Madsen, H. (2021). An introduction to multivariate probabilistic forecast evaluation. *Energy and AI*, 4, 100058.
- Bodik, I., & Kubaska, M. (2013). Energy and sustainability of operation of a wastewater treatment plant. *Environment Protection Engineering*, 39(2), 15–24.
- Butler, D., Digman, C., Makropoulos, C., & Davies, J. W. (2018). *Urban drainage*. Crc Press.
- Drewnowski, J., Remiszewska-Skwarek, A., Duda, S., & Łagód, G. (2019). Aeration process in bioreactors as the main energy consumer in a wastewater treatment plant. review of solutions and methods of process optimization. *Processes*, 7(5), 311.
- El Ghazouli, K., El Khattabi, J., Shahrour, I., & Soulhi, A. (2021). Wastewater flow forecasting model based on the nonlinear autoregressive with exogenous inputs (narx) neural network. *H2Open Journal*, 4(1), 276–290.
- Garofalo, G., Giordano, A., Piro, P., Spezzano, G., & Vinci, A. (2017). A distributed real-time approach for mitigating cso and flooding in urban drainage systems. *Journal of Network and Computer Applications*, 78, 30–42.
- Giani, G., Rico-Ramirez, M. A., & Woods, R. A. (2021). A practical, objective, and robust technique to directly estimate catchment response time. *Water Resources Research*, 57(2), e2020WR028201.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477), 359–378.
- Hager, W. H. (2010). *Wastewater hydraulics: Theory and practice*. Springer Science & Business Media.
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

- Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International workshop on artificial neural networks* (pp. 195–201).
- Heinonen, M., Jokelainen, M., Fred, T., Koistinen, J., & Hohti, H. (2013). Improved wet weather wastewater influent modelling at viikinmäki wwtp by on-line weather radar information. *Water science and technology*, *68*(3), 499–505.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.
- Hylleberg, S., Jørgensen, C., & Sørensen, N. K. (1993). Seasonality in macroeconomic time series. *Empirical Economics*, *18*(2), 321–335.
- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Langeveld, J., Van Daal, P., Schilperoort, R., Nopens, I., Flameling, T., & Weijers, S. (2017). Empirical sewer water quality model for generating influent data for wwtp modelling. *Water*, *9*(7), 491.
- Leitão, R. C., Van Haandel, A. C., Zeeman, G., & Lettinga, G. (2006). The effects of operational and environmental variations on anaerobic wastewater treatment systems: A review. *Bioresource technology*, *97*(9), 1105–1118.
- Li, F., & Vanrolleghem, P. A. (2022). An influent generator for wrrf design and operation based on a recurrent neural network with multi-objective optimization using a genetic algorithm. *Water Science and Technology*, *85*(5), 1444–1453.
- Li, Z., Zou, Z., & Wang, L. (2019). Analysis and forecasting of the energy consumption in wastewater treatment plant. *Mathematical Problems in Engineering*, 2019.
- Makinia, J., & Zaborowska, E. (2020). *Mathematical modelling and computer simulation of activated sludge systems*. IWA publishing.
- Marcjasz, G., Narajewski, M., Weron, R., & Ziel, F. (2023). Distributional neural networks for electricity price forecasting. *Energy Economics*, 106843.
- Mariano, R. S. (2002). Testing forecast accuracy. *A companion to economic forecasting*, *2*, 284–298.
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml*.
- Newhart, K. B., Holloway, R. W., Hering, A. S., & Cath, T. Y. (2019). Data-driven performance analyses of wastewater treatment plants: A review. *Water research*, *157*, 498–513.
- Nissen, K. M., & Ulbrich, U. (2017). Increasing frequencies and changing characteristics of heavy precipitation events threatening infrastructure in europe under climate change. *Natural Hazards and Earth System Sciences*, *17*(7), 1177–1190.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Pedersen, J. W., Lund, N. S., Borup, M., Löwe, R., Poulsen, T. S., Mikkelsen, P. S., & Grum, M. (2016). Evaluation of maximum a posteriori estimation as data assimilation method for forecasting infiltration-inflow affected urban runoff with radar rainfall input. *Water*, *8*(9), 381.
- Robecke, U. (2015). *Modellierung des energieverbrauchs für belüftungssysteme auf kläranlagen mittels künstlicher neuronaler netze*. Verein zur Förderung des Instituts IWAR der TU Darmstadt eV.

- Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference*.
- Seggelke, K., Löwe, R., Beeneken, T., & Fuchs, L. (2013). Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of wilhelmshaven. *Urban Water Journal*, *10*(5), 330–341.
- Senst, T., Eiselein, V., & Sikora, T. (2012). Robust local optical flow for feature tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(9), 1377–1387.
- Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., ... Kalchbrenner, N. (2020). Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*.
- Sonnenschein, B., & Ziel, F. (2022a). *Distributional wastewater treatment plant inflow forecast* [Software]. doi: <https://doi.org/10.5281/zenodo.6997701>
- Sonnenschein, B., & Ziel, F. (2022b). *Spatially aggregated rain radar forecast for the koeln weiden, germany* [Dataset]. doi: <https://doi.org/10.5281/zenodo.6992869>
- Sonnenschein, B., & Ziel, F. (2022c). *Wastewater treatment plant inflow rate, network water levels and rain rate for koeln weiden, germany* [Dataset]. doi: <https://doi.org/10.5281/zenodo.6992694>
- Stasinopoulos, D. M., & Rigby, R. A. (2008). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, *23*, 1–46.
- Stentoft, P. A., Guericke, D., Munk-Nielsen, T., Mikkelsen, P. S., Madsen, H., Vezzaro, L., & Møller, J. K. (2019). Model predictive control of stochastic wastewater treatment process for smart power, cost-effective aeration. *Ifac-papersonline*, *52*(1), 622–627.
- Stoica, P., & Selen, Y. (2004). Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, *21*(4), 36–47.
- Svensen, J. L., Falk, A., Madsen, H., Niemann, H., Poulsen, N., Mikkelsen, P., ... Ocampo-Martinez, C. (2021). Model predictive control in urban systems.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.
- Vaughan, J. W., & Wallach, H. (2020). A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence*.
- Wang, Z., Man, Y., Hu, Y., Li, J., Hong, M., & Cui, P. (2019). A deep learning based dynamic cod prediction model for urban sewage. *Environmental Science: Water Research & Technology*, *5*(12), 2210–2218.
- Wasserman, L. (2004). *All of statistics : a concise course in statistical inference*. Springer.
- Wetterdienst, D. (2022). *Deutscher wetterdienst open data portal. historische radolan-raster der niederschlagshöhe* [Dataset]. Retrieved 2022-04-20, from [https://opendata.dwd.de/climate.environment/CDC/grids\\_germany/](https://opendata.dwd.de/climate.environment/CDC/grids_germany/)
- Winterrath, T., & Rosenow, W. (2014). The radar-based precipitation nowcasting system radvor of deutscher wetterdienst for the support of meteorological and hydrological alert systems. In *8th european conference on radar in meteorology and hydrology, germany* (pp. 379–390).
- Xiao, H., & Sun, Y. (2019). On tuning parameter selection in model selection and model averaging: a monte carlo study. *Journal of Risk and Financial Management*, *12*(3), 109.
- Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, *31*(7), 1235–1270.
- Zhang, D., Hølland, E. S., Lindholm, G., & Ratnaweera, H. (2018). Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer. *Journal of Hydrology*, *567*, 792–802.
- Zhang, Q., Li, Z., Snowling, S., Siam, A., & El-Dakhkhni, W. (2019). Predictive models for wastewater flow forecasting based on time series analysis and

- artificial neural network. *Water Science and Technology*, *80*(2), 243–253.
- Zhou, P., Li, Z., Snowling, S., Baetz, B. W., Na, D., & Boyd, G. (2019). A random forest model for inflow prediction at wastewater treatment plants. *Stochastic Environmental Research and Risk Assessment*, *33*(10), 1781–1792.
- Ziel, F., & Berk, K. (2019). Multivariate forecasting evaluation: On sensitive and strictly proper scoring rules. *arXiv preprint arXiv:1910.07325*.
- Ziel, F., Croonenbroeck, C., & Ambach, D. (2016). Forecasting wind power—modeling periodic and non-linear effects under conditional heteroscedasticity. *Applied Energy*, *177*, 285–297.