Potential of machine learning for estimating the impact of water efficient scenarios on solids accumulation in sewers

C. Harpaz[1], S. Russo[2], J.P. Leitão [3], R. Penn[1]*

[1]Department of Civil and Environmental Engineering, Technion Israel Institute of Technology, 3200 Haifa, Israel

[2] ETH Zürich, Ecovision Lab, Photogrammetry and Remote Sensing, Zürich, Switzerland

[3] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

*Corresponding author email: roni.penn@technion.ac.il

Highlights

- Three Machine Learning algorithms were implemented to predict sediment locations in sewers.
- The study was based on highly variable synthetic sewer networks with different water saving scenarios.
- In this study, Random Forest algorithm showed the highest accuracy (97%).

Introduction

In most affluent cities, centralized water and wastewater conveyance and treatment systems are widely established. However, in many cases, such investment-intensive systems may no longer provide the best solution (Eggimann et al., 2016). Attractive water efficient solutions to replace or complement existing centralized systems are widely being developed and promoted (Larsen et al., 2016). Alongside their contribution to more sustainable urban water systems, these systems can however negatively influence existing centralised systems. Of the expected and concerning effects are the accumulation of solids in sewers and their negative contribution to sewer blockages and formation of toxic gases (Liu et al., 2015; Penn 2017; Murali et al., 2021). Model predictions of solids deposition locations require setting up and simulating detailed sewer models. Many times, precise details of the networks layout and diurnal flow patterns are not available, limiting the applicability of using models' predictions for such hazards. Development of a simplified model predicting sediment accumulation in sewers will spare the need to conduct complex simulations; it may be used to analyse any existing and future systems, where transition from its traditional centralized systems is considered; it will potentially indicate where the most affected points in the system are expected to occur. A more detailed model can then be calibrated only for these sensitive locations and simulated for more precise predictions of the effects.

In this study, a detailed sewer model integrating stochastic generation of domestic wastewater streams and hydrodynamic simulations of flows was set up to investigate sewer networks characteristics that are more prone to sediment accumulation. 117 highly variable synthetic sewer networks were generated and simulated for various water efficient transition scenarios. Machine Learning (ML) was used to predict shear stresses from which sediment accumulation locations were derived. Three ML models were trained using hydrodynamic simulation results. Features used for their training included design parameters of the network, and graph theory characteristics. Features obtained from hydraulic simulations and detailed water usage patterns, which are usually difficult to obtain, were not used.

Methodology

Generation of synthetic sewer networks

117 highly variable synthetic separate sewer networks were generated using DrainNetGen (Muranho et al., 2016). These allow variability in the networks geometry, affecting solids accumulations. The networks comprise 50 - 150 nodes, slopes in the range of 3 – 10%, and a population density of 4,000, 15,000 and 23,000 residents/km^2. The wastewater discharged into the sewer was generated using a stochastic domestic wastewater generator (Penn et al., 2017). Then, the EPA Storm Water Management Model (SWMM) software (Rossman 2004) was used for the simulation of the sewer network hydraulics. Detailed properties of sewer flows allowed diurnal patterns of shear stress to be calculated and locations of sediment accumulations to be derived. Ten transition scenarios were simulated (Table 1); scenarios 2 to 6 correspond to a uniform reduction of wastewater flows through the day of 10, 30, 50, 70, and 90%. Scenarios 7 – 11 correspond to reduction of flow peaks to mimic a reduction mostly in these hours, resulting, for example, from greywater reuse (Penn et al., 2012). In these scenarios, a similar reduction in daily flows (12.9, 34.3, 54.5, 73.7, 91.6%) was targeted. These general scenarios enable the assessment of the performance of the system without considering specific transition technologies. Scenario 1 (baseline scenario) served as the scenario for comparison.

Table 1. Scenarios simulated.

Transition scenario	Flow remaining after reduction	Flow reduction type
0	Q	[Baseline]
1	0.9Q	Uniform reduction
2	0.7Q	Uniform reduction
3	0.5Q	Uniform reduction
4	0.3Q	Uniform reduction
5	0.1Q	Uniform reduction
6	0.9Q	Peak reduction
7	0.7Q	Peak reduction
8	0.5Q	Peak reduction
9	0.3Q	Peak reduction
10	0.1Q	Peak reduction

Table 2. Features and target for ML models

Feature	Units	Values range
Pipe diameter	m	0.188 - 0.593
Pipe slope	m/m	0.003 - 0.1
Aspect ratio	[-]	0.133 – 7.5
Flow remaining after reduction	Fraction of baseline flow (Q)	0.1 - 1
Flow reduction type	1 = Uniform reduce 0 = Peak reduce	Binary
Stream order	[-]	1-75
Density	1000 residents/km^2	4, 15, 23
Contributing residents	[-]	9-62,146
Node degree	[-]	1-5
Centrality features	[-]	0 - 4,790
Y = maximum shear stress (Target parameter)	Pa	0 - 100

Machine Learning

In this study, supervised ML regression algorithms, specifically Linear Regression (LR), Random Forest (RF), and Neural Network (NN), were applied for developing models with different characteristics to predict sediment accumulation locations. The objective function of the algorithms was the root mean square error (RMSE). The correlation between the algorithms complexity and prediction accuracy was examined. The ML algorithms were trained to learn to predict the maximum diurnal shear stress in each sewer pipe, from which sewer pipes were identified as prone to accumulate solids or not. This identification was based on a typical critical bed shear stress of 2 Pa, which ensures self-cleansing of the pipes (Vongvisessomjai et al. 2010). The models sensitivity to this value was analysed. The models were tested on unseen test data in order to evaluate their performance. The simulated data was a 196,878 x 16 matrix, which constitutes of 196,878 data examples, i.e., the pipes from all the networks and scenarios, and 16 columns, i.e., the feature values (Table 2). The target parameter was the maximum diurnal shear stress. In order to build the models, the data were split in two sub-sets: 80% used for training and 20% for testing. The regression algorithms accuracy was evaluated based on R-squared (R^2) and mean absolute error (MAE). Classification metric as F1-score and accuracy were used to evaluate the algorithms' capability of identifying pipes prone to solids accumulation.

Results and discussion

The relatively simple LR algorithm showed a poor performance with R^2 of 0.77 and MAE of 3.2 Pa, when validated using the test data set. R^2 of 0.96 and MAE of 1.25 Pa were observed for the RF. The NN regressor, which is the most complex algorithm, presented a very similar prediction accuracy of R^2 of 0.96 and MAE of 1.25 Pa. Most of the data points, i.e., simulated and predicted shear stresses, fell within the range of up to 10 Pa. A closer observation around this range showed that the RF algorithm provided the best predictions and lowest scattering around the expected value, making the RF slightly more accurate (Figure 1). The predictive accuracies, were of 92%, 97%, 95%, and F1-score of 0.84, 0.93, and 0.91 for the LR, RF, and NN algorithms, respectively.

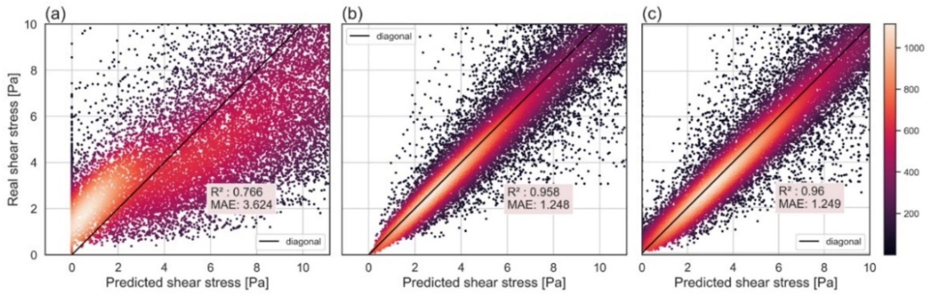

Figure 1. Results of the ML regression algorithms for the shear stress in the range of 0 – 10 Pa (a) linear regression (b) random forest (c) neural networks. The colours indicate the number of data points.

Conclusions and future work

In this work, highly variable synthetic sewer networks were generated and simulated for various water saving scenarios. Three ML models to predict sediment accumulation in sewers were developed using three algorithms: (i) Linear regression (LR), (ii) Random Forest (RF), and (iii) Neural Network (NN). The RF and NN models performed better than the LR model. The developed approach can be implemented on more sophisticated models of solids transport in sewers. Such models are currently under development by co-author Penn.

References

Eggimann S., Truffer B. and Maurer, M. (2016). The cost of hybrid waste water systems: A systematic framework for specifying minimum cost-connection rates. Water Research, 103, 472-484.

Larsen T. A., Hoffmann S., Lüthi C., Truffer B. and Maurer M. (2016). Emerging solutions to the water challenges of an urbanizing world. Science, 352(6288), 928-933.

Liu Y., Ni B.J., Ganiguè R., Werner U., Sharama K.R. and Yuan Z. (2015). Sulphide and methane production in sewer sediments. Water Research, 70, 350-359.

Murali, M. K., Hipsey, M. R., Ghadouani, A. and Yuan, Z. (2021). An investigation into the impacts of water demand management and decentralized water recycling on excess sewer sediment deposition. Journal of Environmental Management, 279

Muranho J., Simões N.E., Sá Marques A., Sousa J. and Leitão J.P. (2016). Customisation of SWMM for Automatic Generation of Urban Drainage Networks. In *CCWI 2016: Computer and Control for the Water Industry*, 7-9 November, Amsterdam, the Netherlands.

Penn R., Hadari M. and Friedler E. (2012). Evaluation of the effects of greywater reuse on domestic wastewater quality and quantity. Urban Water Journal, 9(3),137-148.

Penn R., Schütze M., Gorfine M. and Friedler, E. (2017). Simulation method for stochastically generating domestic wastewater discharge and the effect of greywater reuse on gross solid transport. Urban Water Journal, 14 (8), 846-852.

Refaeilzadeh P., Tang L., and Liu H., (2009). Cross-validation, in Encyclopedia of Database Systems. Springer, pp. 532–538

Rish, Irina., 2001. An empirical study of the naive Bayes classifier. *IJCAI workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22.

Rossman, L.A. (2004). Stormwater Management Model. User's Manual. SWMM5.0. Environment Protection Agency.

Vongvisessomjai N., Tingsanchali T., Babel M.S., 2010. Non-deposition design criteria for sewers with part-full flow. *Urban Water Journal*, 7 (1), 61 – 77