

EFFECTS OF HEAVY-TAILED DEMAND MODEL UNCERTAINTY ON WATER DISTRIBUTION SYSTEM SIMULATION

Bernhard Jonathan Sattler¹, Siew Ann Cheong², Andrea Tundis³, Jonas Joerin⁴, Peter F. Pelz⁵

^{1,3} Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR), Sankt-Augustin, Germany

² School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

⁴ Future Resilient Systems, Singapore-ETH Centre, Singapore, Singapore

⁵ Chair of Fluid Systems, Technical University of Darmstadt, Darmstadt, Germany

¹*bernhard.sattler@dlr.de*

ABSTRACT

Simulation of Water Distribution Systems (WDSs) is used to evaluate WDS management to ensure the security of water supply. Many such simulations rely on assumptions of demand uncertainty. In this paper, we investigate which probability distributions adequately describe demand uncertainty and how the choice of a distribution affects the simulation results. To identify distributions, we first decompose water demand data of District Metered Areas of a city into demand trends, daily patterns, and residual uncertainty using the LOESS algorithm. Residuals are heavy-tailed, typically fitting a local log-normal distribution, but occasionally aligning better with a log-t distribution. We then assess the operational impact of the identified demand uncertainty by simulating the L-Town benchmark network subject to log-normal and log-t-distributed uncertainty using the WNTR Python package. The simulation results are evaluated based on technical KPIs. The results show that log-t-distributed uncertainty leads to worse simulated WDS performance on these KPIs, indicating that the inadequate use of normal or log-normal distributions could overestimate the WDS performance. Our findings highlight the importance of selecting appropriate uncertainty distributions for stochastic WDS optimization.

Keywords: Water distribution system, demand uncertainty, heavy-tails

INTRODUCTION

Water Distribution Systems (WDSs) ensure the continuous supply of potable water to citizens and thereby serve the 6th sustainable development goal defined by the United Nations. A common tool for the management of these WDSs is the computational simulation of WDS models, which relies on water demand (WD) data as an input. Since future WD is uncertain, knowledge of the statistical properties of WD can be used to directly simulate WDSs subjected to such uncertain WD, or to validate behavioural models of WD in specific scenarios, as previously outlined [1]. Both of these approaches build on the assumption that statistical properties of past WD are an adequate description of the uncertainty in future WD. One statistical property that has been observed are heavy-tails (HTs) of frequency distributions of WD [2]. HTs refer to the empirical distributions of WD showing a higher frequency of values far from the mean of the distribution than a normal distribution. Such distributions with HTs

appear in many systems and are important for extreme events and risk management [3]. Despite the potential relevance for the management of WDSs, to our knowledge there is no in-depth analysis of HTs in WD and its impacts on WDSs. To address this research gap, we analyse District Metered Area (DMA) level WD data and simulate a benchmark WDS to investigate (i) if HTs can be observed in DMA-level WD data, (ii) how models of WD can account for these HTs, and (iii) how these HTs impact a WDS. We show that (i) HTs can be observed in WD and its hourly changes through tail estimation, that (ii) these HTs can be modelled by decomposing the data into a daily pattern, a slower varying trend, and heavy-tailed model residuals, which we consider the uncertainty in the WD. Finally, we (iii) simulate a WDS subject to the observed uncertainty and evaluate the impact using key performance indicators (KPIs).

METHODS

Describing heavy tails

To discover HTs in the frequency distributions of WD time series, we used a data set of hourly water meter readings over a period of three years from ten DMAs labelled A to J [4]. Corresponding weather data and data from days with incomplete or double readings were not regarded. The distribution of a variable X is commonly considered heavy-tailed, if its kurtosis, i.e. the fourth moment of the distribution, exceeds that of a normal distribution. However, kurtosis can lack robustness for empirical distributions [5]. We therefore first consider only the tails by assuming that HTs follow a power-law described by

$$p(|X - \mu|) \propto |X - \mu|^{-\alpha}, \quad (1)$$

with the probability (or frequency for empirical distributions) p and a real tail exponent $\alpha > 1$. Smaller α indicate a higher probability of extreme events. We used a maximum likelihood estimate based on [6] for α . The part of the distribution where $x < \mu$ is called the negative tail and $x > \mu$ is called the positive tail. The analysis was performed on the frequency distributions of each time series, referred to as the raw data, and their differentiated forms, defined as the changes between hourly values.

Isolating the heavy tails

While the aforementioned method enables an estimation of the tail behaviour, the data is still dominated by the periodicities and trends observed in WD, which are not the focus of our analysis. To isolate the short-term stochastic uncertainty from the time series, we first assume that the data can be described by

$$Q_t = \bar{Q}_d * P_h * \varepsilon_t. \quad (2)$$

Here, Q_t denotes the WD at hour t of the time series, $\bar{Q}_d$ is the average daily demand subject to slowly varying trend, P_h denotes the daily usage pattern for each hour h of the day, and ε_t denotes an error term, which captures the uncertainty of the time series. The use of a multiplicative average demand with a daily pattern is in line with common assumptions [7]. A logarithmic transform yields an additive relationship, which can be separated using the LOESS algorithm [8], as shown in [9]. The LOESS-parameters of the low pass, the seasonal smoothing, and the trend smoothing were set to 25, 7, and 47 respectively, in line with the recommendations in [8]. The resulting decomposition yields a residual time series, to which we fitted a t-distribution. This distribution is defined through its location, scale, and degrees of freedom ν . For $\nu \rightarrow \infty$ it approximates a normal distribution where the location and scale

are given by the mean and variance respectively. For smaller values of ν , the distribution displays heavier tails. Furthermore, ν allows a direct calculation of σ , κ , and α . We then tested if deviations from normality arise from a normal distribution with time-varying variance or a distribution with persistent HTs, using a normality test based on [10] applied to each 48-value window at a 5% confidence level. This approach allows us to separate the time series of residual uncertainty into two distinct regimes of normal and non-normal residuals. For the latter we then again estimated the parameters for the t-distribution.

Simulating the effects of the heavy tails

Using the obtained ν for the t-distribution, we simulated the behaviour of a WDS subjected to both normal and t-distributed WD uncertainty and compared the simulation results on technical KPIs of the network. The simulation was performed using the Water Network Tool for Resilience (WNTR) [7] and the L-Town network [11]. For each distribution, 100 Monte-Carlo runs were performed. In each run, stochastic errors were sampled from the respective uncertainty distribution and added to the WD of the original L-Town network, which was then simulated. In the original L-Town network, each junction is assigned one of three demand patterns for the quarter-hourly WD over one week, with one pattern showing a constant WD. As the uncertainty distributions were obtained from hourly data, all original patterns were averaged to hourly values. Stochastic model error was added to each non-constant pattern, with all junctions sharing a pattern receiving the same error, as in the original network. This is consistent with the DMA-level identification of uncertainty. Based on [12], the water service, energy surplus, pressure, flow velocity, and flow directional changes were identified as critical for a WDS. Lacking operational standards for the L-Town network, we assessed simulation results using the following KPIs: total energy use, energy surplus measured by the Todini Index [13] with the minimum pressure set to 2.7 bar, mean pressure, mean absolute flow velocity, and number of flow directional changes in the network. KPI distributions of the two Monte-Carlo-sets were compared using a KS test and quartiles.

RESULTS AND DISCUSSION

Our analysis shows multimodal properties of the raw time series for most DMAs and for the hourly changes of some of the DMAs, as presented in Figure 1 for DMA E and its hourly changes. Multimodal distributions exhibit multiple local maxima (modes), which in this case are due to the daily, weekly and yearly seasonality in the WD. The tail estimation and distribution fit to the LOESS residuals for DMA E are shown in Figure 1, and the tail exponents and distribution parameters are displayed in Table 1. We see consistently heavier positive than negative tails in the raw time series, while hourly changes exhibit more symmetric tails. Furthermore, higher values of α indicate that HTs are less pronounced in some DMAs, e.g. DMA H. The LOESS decomposition yields unimodal uncertainty distributions for most DMAs. Only DMA A exhibits a second mode in the positive tail, indicating a second distinct outflow behaviour. While such additional modes, called Dragon Kings, can be relevant for risk management [14], this specific second mode is likely caused by flushing events rather than WD by users [4]. For all DMAs, the uncertainty distributions obtained by the LOESS are better approximated through a t-distribution than a normal distribution.

As the analysis was performed on logarithmic data, the uncertainty distributions transform into log-normal and log-t-distributions respectively. This is in line with the positive tails in the raw data generally appearing heavier-tailed than the negative tails. The normality tests of the LOESS residuals indicate that the residuals of some DMAs are mostly normally distributed with bursts of local HTs, while other DMAs exhibit HTs at most times. The histograms for the normally and t-distributed share of the residuals of DMA I are presented in Figure 2, the degrees of freedom for the LOESS residuals and for the separated residuals, and the share of normal residuals are presented in Table 1.

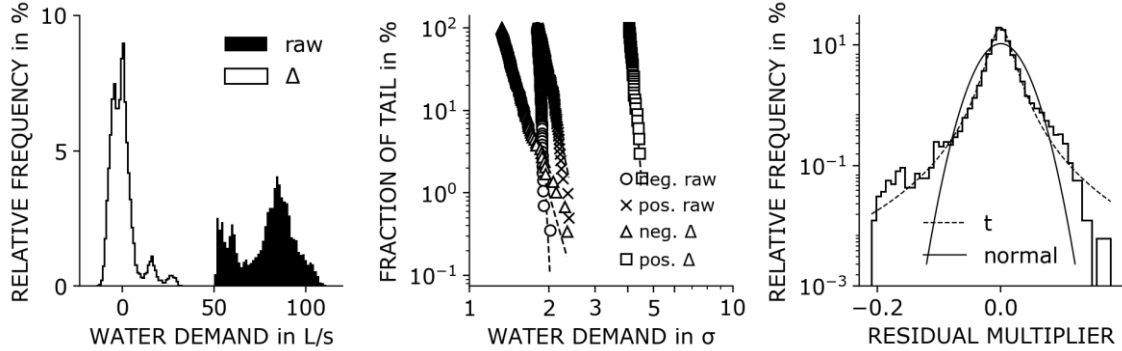


Figure 1.: Relative frequency of hourly data (“raw”) and hourly changes (“ Δ ”) (left), complementary cumulative density function of the tails (centre), and relative frequency of LOESS-residuals with t - and normal distributions fits (right).

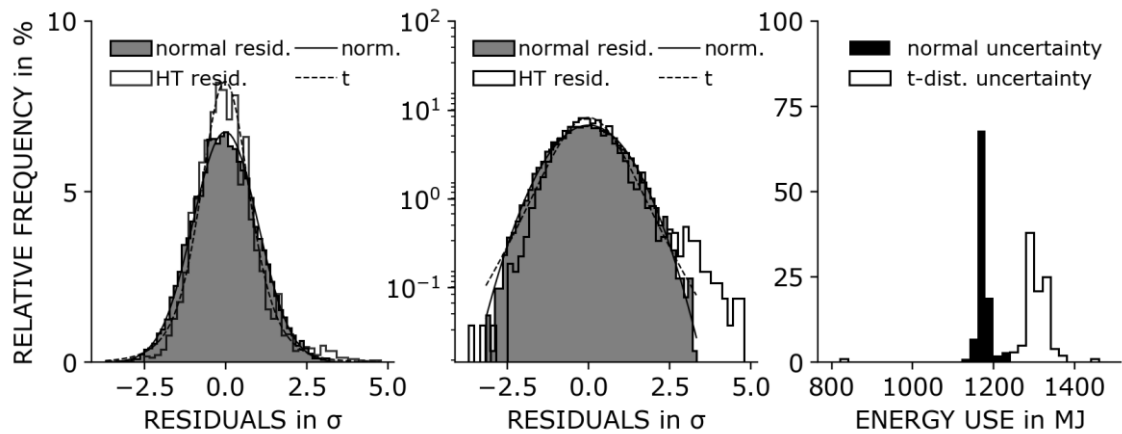


Figure 2.: Linear (left), and logarithmic histogram (centre) of the separated residual regimes, and the simulated energy use by the two regimes (right).

The simulation was performed sampling from a $\mathcal{N}(0, 0.1)$ and $t(3, 0, 0.1)$ distribution respectively, following the lowest degrees of freedom reported in Table 1. The distributional properties of the KPIs obtained from the simulation of the L-Town network subject to the two different demand uncertainty distributions are presented in Table 2. The KS test results (p-values are all 0) show statistically significant differences between the results of normal and t -distributed errors for all KPIs.

The t -distributed uncertainty yields higher energy use, higher number of flow changes, lower energy reserve measured by the mean Todini Index, lower mean pressures, higher mean flow velocities, and a broader range of values for all these KPIs. This shows that log- t -distributed model uncertainty generates a worse performance of the network compared to the log-normal distribution. These results indicate that relying on normal or log-normal distributed

uncertainty for the simulation of WDS could underestimate output uncertainty in WDS operation. However, the results of this study are limited due to the combination of demand data and WDS topology from different geographical contexts. Furthermore, exploring other decomposition methods, data and simulations of higher resolution, and further KPI analysis may deepen insights into the relevance of HTs in WD. The presented results should not be considered a definitive evaluation of the operational risks of a WDS, but highlight the importance of extreme value WD uncertainty for WDS and indicate a potential research gap for WD models for rare extreme deviations from normal operations.

Table 1. Tail exponents for the raw time series data, the hourly changes, and distributional properties of the residual time series after LOESS-decomposition for all DMAs A to J.

DMA	α_{raw} (neg. pos. tail)	$\alpha_{\Delta h}$ (neg. pos. tail)	ν_{LOESS}	$\nu_{LOESS-separated}$	normally distributed residuals in %
A	40.3 4.6	2.8 2.8	3.0	2.8	38.5
B	52.0 14.3	3.1 4.3	3.7	3.5	78.6
C	71.5 6.4	3.2 5.0	5.8	4.8	80.6
D	17.1 9.3	13.3 11.1	13.9	7.9	88.6
E	62.4 16.4	12.5 39.4	2.0	2.5	19.1
F	24.8 7.5	5.8 3.7	9.8	7.1	88.0
G	26.6 14.5	8.0 14.9	3.5	3.8	57.9
H	69.6 28.6	9.8 9.4	3.4	3.2	39.8
I	35.7 7.6	4.7 4.5	7.8	3.3	86.6
J	11.1 7.0	6.7 5.9	11.0	6.1	88.0

Table 2. Quartiles of the KPI distributions resulting from the Monte-Carlo runs.

	total energy use in MJ		mean Todini Index		total no. of flow changes		mean pressure in m		mean flow velocity in $cm\ s^{-1}$	
	$\mathcal{N}$	t	$\mathcal{N}$	t	$\mathcal{N}$	t	$\mathcal{N}$	t	$\mathcal{N}$	t
Q_0	1127	828	0.441	0.313	1618	1277	46.26	42.41	9.57	10.54
Q_1	1160	1286	0.446	0.435	1702	1914	46.28	46.12	9.69	10.74
Q_2	1176	1313	0.448	0.437	1739	1984	46.28	46.13	9.72	10.83
Q_3	1177	1330	0.449	0.439	1782	2050	46.28	46.15	9.77	10.90
Q_4	1226	1458	0.453	0.445	1880	2210	46.30	46.18	9.91	13.23

CONCLUSIONS

In this paper, we showed that DMA-level WD time series are heavy-tailed. These HTs can be modelled through uncertain hourly fluctuations of WD. Finally, our simulations show that HTs could have a statistically significant impact on the KPIs relevant to WDS management. Our results draw attention to the importance of adequate modelling of extreme value uncertainty for the simulation of WDS. Future work should explore the robustness of these results and investigate suitable measures for WDS management considering these findings.

REFERENCES

- [1] B. J. Sattler, J. Friesen, A. Tundis and P. F. Pelz, "Modeling and validation of residential water demand in agent-based models: A systematic literature review," *Water*, vol. 15, no. 3, p. 579, 2023.
- [2] T. E. Downing, R. Butterfield, B. Edmonds, J. W. Knox, S. Moss, B. Piper and E. Weatherhead, "CCDeW: Climate Change and Demand for Water.," *Research Rep. Prepared for Stockholm Environment Institute*, 2003.
- [3] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, pp. 207-211, 2005.
- [4] S. Alvisi and e. al., "The Battle of Water Demand Forecasting (BWDF)," *Proceedings of the 3rd International WDSA-CCWI Joint Conference, Ferrara, Italy*, pp. 1-4, 2024.
- [5] T.-H. Kim and H. White, "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, vol. 1, no. 1, pp. 56 - 73, 2004.
- [6] J. Alstott, E. Bullmore and D. Plenz, "powerlaw: a Python package for analysis of heavy-tailed distributions," *PloS one*, vol. 9, no. 1, p. e85777, 2014.
- [7] K. Klise, R. Murray und T. Haxton, „An overview of the Water Network Tool for Resilience (WNTR),“ in *Proceedings of the 1st International WDSA/CCWI Joint Conference*, Kingston, Ontario, Canada, 2018.
- [8] R. B. Cleveland, W. S. Cleveland, J. E. McRae und I. Terpenning, „STL: A seasonal-trend decomposition,“ *Journal of Official Statistics*, Bd. 9, Nr. 1, pp. 3 - 73, 1990.
- [9] C. Koycegiz, „Seasonality effect on trend and long-term persistence in precipitation and temperature time series of a semi-arid, endorheic basin in Central Anatolia, Turkey,“ *Journal of Water & Climate Change*, Bd. 15, Nr. 5, 2024.
- [10] R. D'Agostino and P. E. S., "Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$," *Biometrika*, vol. 60, no. 3, pp. 613 - 622, 1973.
- [11] S. G. Vrachimis, D. G. Eliades, R. Taormina, Z. Kapelan, A. Ostfeld, S. Liu, M. Kyriakou, P. Pavlou, M. Qiu and M. M. Polycarpou, "Battle of the Leakage Detection and Isolation Methods," *Journal of Water Resources Planning and Management*, vol. 148, no. 12, p. 04022068, 2022.
- [12] B. J. Sattler, J. Stadler, A. Tundis, J. Friesen and P. F. Pelz, "A framework for the simulation-based selection of social models for socio-technical models of infrastructures using technical requirements analysis," in *22nd International Conference on Modeling and Applied Simulation, MAS 2023*, 2023.
- [13] E. Todini, "Looped water distribution networks design using a resilience index based heuristic approach," *Urban Water*, vol. 2, no. 2, pp. 115-122, 2000.
- [14] D. Sornette and G. Ouillon, "Dragon-kings: mechanisms, statistical methods and empirical evidence," *The European Physical Journal Special Topics*, vol. 205, no. 1, pp. 1-26, 2012.