Environmental Science
Water Research & Technology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Environ. Sci.: Water Res. Technol.*, 2025, **11**, 2973

Lognormal distributions capture site-specific variability in enteric virus concentrations in wastewater

Chaojie Li,[a] Tamar Kohn,[a] Shotaro Torii,[a] Htet Kyi Wynn,[a] Alexander J. Devaux,[b] Charles Gan,[b] Timothy R. Julian [bcd] and Émile Sylvestre *[ef]

As more data on virus concentrations in influent water from wastewater treatment plants (WWTPs) becomes available, establishing best practices for virus measurements, monitoring, and statistical modelling can improve the understanding of virus concentration distributions in wastewater. To support this, we assessed the temporal variability of norovirus, adenovirus, enterovirus, and rotavirus concentrations in influent water across multiple WWTPs in Switzerland, the USA, and Japan. Our findings demonstrate that the lognormal distribution accurately describes temporal variations in concentrations for all viruses at all sites, outperforming the gamma and Weibull distributions, which fail to capture high variability. However, notable differences in variability and uncertainty were observed across systems, underscoring the need for site-specific assessments. Using lognormal parameters, we identified optimal monitoring frequencies that balance cost-effectiveness and precision. For most sites, weekly monitoring was sufficient to estimate the annual average concentration of enteric viruses within a 95% confidence interval of 0.5 $\log_{10}$. We further examined the mechanistic basis of the lognormal distribution, highlighting processes that drive its prevalence and shape the behavior of its upper tail. By integrating these insights, this study provides a statistical foundation for optimizing virus monitoring frameworks and informing public health interventions targeting wastewater systems.

Received 25th March 2025,
Accepted 3rd October 2025

DOI: 10.1039/d5ew00286a

rsc.li/es-water

Water impact

This study shows that virus concentrations in wastewater follow predictable patterns, helping explain why they change over time. The proposed approach enables affordable, reliable monitoring and accurate health risk estimates. Its use can support site-specific strategies to track and manage viruses in wastewater, advancing public health protection through improved risk assessment and wastewater-based epidemiology.

1 Introduction

Enteric viruses, such as norovirus, rotavirus, and enterovirus, can enter wastewater systems through fecal discharge, posing public health risks when wastewater is inadequately managed. Infected individuals shed these viruses into sewage *via* toilets, sinks, and other discharge points. Combined sewers may also receive viruses from stormwater runoff contaminated with human and animal feces. Wastewater treatment plant (WWTP) processes can inactivate or remove viruses before discharging treated water into the environment, but their efficacy varies by treatment processes and virus type.[1] Conventional treatment plants typically employ primary (physical) and secondary (biological) treatment steps, which together reduce enteric virus concentrations by approximately 1–3 $\log_{10}$.[2,3] This reduction may not always be sufficient to manage infection risks for people exposed to treated wastewater.[4] Quantitative microbial risk assessment (QMRA) can help inform infection risks to support the design and implementation of risk reduction strategies, but reliable virus concentration data is required to provide accurate risk estimates. Establishing best practices for virus measurements, monitoring, and statistical modelling can

[a] *Laboratory of Environmental Virology, School of Architecture, Civil & Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[b] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland*
[c] *Swiss Tropical and Public Health Institute, Allschwil, Switzerland*
[d] *University of Basel, Basel, Switzerland*
[e] *Department of Water Management, Delft University of Technology, Stevinweg 1, Delft, 2628 CN, The Netherlands. E-mail: E.Sylvestre@tudelft.nl*
[f] *KWR Watercycle Research Institute, P.O. Box 1072, 3430 BB Nieuwegein, Netherlands*

improve understanding of virus concentration distributions and strengthen the application of QMRA in wastewater management.

Quantitative PCR (qPCR) and digital PCR (dPCR) methods are increasingly used to monitor viruses in wastewater due to their sensitivity and specificity. These molecular methods introduce uncertainties related to the accuracy of quantitative results.[5] Factors such as adsorption to particles, PCR inhibition from organic matter, and losses during sample concentration and nucleic acid extraction can affect the reliability of measurements.[6] These uncertainties can be accounted for by correcting measurements based on the recovery rate of spiked surrogate viruses, such as murine hepatitis virus (MHV)[7] or non-enveloped viruses like MS2 and PhiX174.[8] However, even though spiked viruses are routinely used as process controls, this data is rarely applied to correct concentration estimates for enteric viruses.[5]

The value of virus concentration data sets also depends on the frequency of sample collection. Monitoring intervals, which can range from daily to monthly intervals,[9] should be guided by monitoring objectives, such as characterizing temporal variability in virus concentrations over a year or detecting short-term peaks indicative of viral outbreaks. High-frequency monitoring can be impractical in many contexts due to the significant costs and expertise required for virus analyses.[10] Therefore, cost-effective monitoring strategies and statistical models that balance precision and practicality are essential to improve the utility of virus monitoring programs.

Previous studies have combined enteric virus concentration data from multiple WWTPs into pooled datasets to derive aggregate estimates for use in QMRA.[11–13] Because each observation is given equal weight, these pooling approaches do not model between-WWTP heterogeneity and can therefore mask important site-specific differences driven by local shedding patterns, population size, sampling design, and analytical methods. Statistical approaches have been developed to select the most suitable parametric distributions for site-specific temporal variations of *E. coli* and protozoan pathogens.[14–16] These approaches have not been systematically adapted for enteric viruses in wastewater, where distinct challenges, such as higher variability and the influence of virus-specific shedding patterns, must be addressed. By applying these methods and tailoring them to enteric viruses, this study aims to fill this gap and provide a stronger statistical foundation for monitoring and assessing exposure to enteric viruses in wastewater.

In this study, we investigated temporal variations in enteric virus concentrations in influent wastewater to advance monitoring strategies. Using original data sets from two municipal WWTPs in Switzerland, as well as literature data reported for five WWTPs in the USA[1] and one in Japan,[17] we characterized temporal variability for multiple virus types in diverse geographic contexts. We proposed candidate parametric distributions to model the data and applied information criteria to evaluate their suitability. Additionally, we assessed the impact of incorporating sample-specific analytical recovery rates and varying monitoring frequencies on model accuracy.

2 Material and methods

2.1 Rationale for sites selection and virus monitoring strategies

The aim of this study was to develop and demonstrate statistical models for describing virus concentration distributions in wastewater influent. To support this aim, we selected datasets from eight WWTPs that together span multiple virus types, geographic regions, and monitoring strategies, and that provided sufficient temporal resolution for statistical modelling. Data sets from eight wastewater treatment plants (WWTPs) were selected for this study. Two WWTPs in Switzerland—the STEP Vidy WWTP in Lausanne and the ARA Werdhölzli WWTP in Zurich—provided original data, while previously published data were obtained from Matsushima, Japan (Kazama *et al.*, 2017),[17] and five WWTPs in California, USA [Los Angeles County Sanitation Districts (LACSD), City of Los Angeles Sanitation and Environment (LASAN), Orange County Sanitation District (OCSD), City of San Diego (SD), and San Francisco Public Utilities Commission (SFPUC)] (Pecson *et al.*, 2022).[8] These sites were chosen to represent various geographical locations and population sizes (Table 1).

Only data quantified by molecular methods—(RT-)qPCR or RT-dPCR—with documented process controls and recoveries were included to maintain methodological consistency. All data sets also fulfilled the following criteria: (i) ≥12 months of routine sampling, (ii) quantification of at least one target enteric virus, and (iii) sufficient sample size (≥10 positive detections) for parametric modelling. The viruses measured included adenovirus, enterovirus, norovirus, and rotavirus at Lausanne; enterovirus and norovirus at Zurich; norovirus at Matsushima; and adenovirus, enterovirus, and norovirus at the five WWTPs in California, USA. Certain studies involved extended monitoring campaigns (*e.g.*, the multi-year, weekly sampling in Matsushima), whereas others had higher-frequency monitoring (*e.g.*, the every-two-day sampling at ARA Werdhölzli). Analyzing data under these different monitoring regimes allowed us to examine whether the modelling framework proposed in this work holds across both short-term, high-frequency datasets and longer-term monitoring programs.

2.2 Sampling and quantification of enteric viruses

2.2.1 Original data from Swiss WWTPs. For the WWTP in Lausanne, a total of ninety-six 24-hour composite wastewater influent samples were collected monthly between November 2018 and October 2019. Detailed sampling processing procedures, nucleic acid extraction methods, and (RT-)qPCR quantification are described in Li, Sylvestre.[18] In brief, approximately 600 ml of raw sewage was first pre-filtered and further concentrated using a centrifugal filter unit, resulting in a final concentrate of about 2 ml. Nucleic acids were extracted using the QIAamp Viral RNA Mini kit and stored at −80 °C prior to qPCR/RT-qPCR quantification of adenovirus,

Table 1 Locations of municipal wastewater treatment plants (WWTPs) and corresponding virus types analyzed in influent wastewater at these sites

Site location	Population served by WWTP	Virus types	Sample type	Monitoring frequency & duration	Quantification method	Study
Lausanne, Switzerland (STEP Vidy)	220 000	Adenovirus Enterovirus Norovirus Rotavirus	24 hour composite	Monthly/1 year	(RT-)qPCR	Original data
Zurich, Switzerland (ARA Werdhölzli)	471 000	Enterovirus Norovirus	24 hour composite	Every 2 days/1 year	RT-dPCR	Original data
Matsushima, Japan	9600	Norovirus	Grab	Weekly/3 years	RT-qPCR	Kazama *et al.* (2017)[17]
California, USA (5 sites)	300 000–4 000 000	Adenovirus Enterovirus Norovirus	Grab	Every 2 to 3 weeks/1 year	(RT-)qPCR	Pecson *et al.* (2022)[8]

enterovirus, norovirus, and rotavirus. All qPCR/RT-qPCR assays were carried out in duplicate, with DNA gBlocks used as standards for absolute quantification. The primer and probe sequences are reported in Li *et al.* (2023).[18]

For the WWTP in Zurich, 180 24-hour composite influent samples were collected between February 1, 2021, and January 29, 2022. After collection, the wastewater samples were shipped on ice and stored at 4 °C for up to eight days before processing. The collected samples were processed either by protocol 1 (ultrafiltration followed by RNA extraction using QIAamp viral RNA Mini kit, used before November 10, 2021)[19] or protocol 2 (total nucleic acid extraction, samples collected after November 30, 2021).[20] Samples collected between November 10 and November 30, 2021, were processed by both protocols to establish a conversion factor that corrects for inter-method differences (see SI, Fig. S3). To adjust for differences in method performance, a correction factor was applied to measurements from protocol 1. Specifically, data from protocol 1 were adjusted by multiplying by 1.17 for EV and 2.65 for NoV GII. The extracted RNA was stored at −80 °C for up to 1.5 years before being measured on dPCR.

A duplex RT-dPCR assay for human enterovirus (EV) and norovirus genogroup II (NoV GII) RNA quantification was optimized by adapting the thermal cycling conditions and primer and probe concentrations from previously described RT-qPCR assays.[21–24] Ten-fold diluted extracts were used as RNA templates. The assay was performed in 12 μL reaction mixtures using the QIAcuity OneStep Advanced Probe kit (QIAGEN) and 8.5k 96-well Nanoplates (QIAGEN). The duplex RT-dPCR mixture contained 3 μL of 4× OneStep Advanced Probe Master Mix, 0.12 μl of OneStep RT Mix, 1.5 μl of GC Enhancer, 1000 nM forward primer (EV: 5′-CCTCCGGCCCCTGAATG-3′, NoV GII: 5′-ATGTTC AGRTGGATGAGRTTCTCWGA-3′), 1000 nM reverse primer (EV: 5′-ACCGGATGGCCAATCCAA-3′, NoV GII: 5′-TCGACGCCATCTTC ATTCACA-3′), 500 nM probe (EV: 5′-HEX-CGGAACCGACTACTTT GGGTGTCCGT-BHQ1-3′, NoV GII: 5′-FAM-AGCACGTGGGAGGG CGATCG-TAMRA-3′), 3 or 4 μl of template RNA, and DNase/RNase free water. All RT-dPCR reactions were performed in duplicate. The nanoplate was loaded onto the QIAcuity One, 2-plex Device (Qiagen). The thermal cycles include an RT step at 50 °C for 60 min, 95 °C for 5 min for enzyme activation, and followed by 45 cycles of denaturation (95 °C for 15 s) and annealing/extension (at 60 °C for 60 s). In each run, a negative control (no template) and a positive control (*i.e.*, synthetic DNA gBlock® containing the target sequence, Integrated DNA Technologies, Coralville, IA, USA) were included. Quantities were expressed as genome copies (GC μL^{-1}) per reaction. The RT-dPCR assays were performed using automatic settings for the threshold and baseline. Samples showing >5-fold difference between the duplicated reactions were excluded (3 out of 180 samples) from further analysis. Further quality control data for the duplex RT-dPCR assay is given in the SI. The completed dMIQE (digital Minimum Information for Publication of Quantitative Digital PCR Experiments) checklist for this RT-dPCR is available in Huisman, Scire.[19]

To assess the recovery efficiency for 91 wastewater samples collected at the Zurich WWTP, murine hepatitis virus (MHV) was spiked into replicate aliquots of 50 ml wastewater at a concentration of approximately 1×10^6 GC per 50 ml^{-1}. The preparation of MHV stock solutions is described in Fernandez-Cassi, Scheidegger.[25] MHV was determined using the primers and probes described in Fernandez-Cassi, Scheidegger,[25] but using a protocol modified for RT-dPCR and a Naica System (Stilla Technologies, Villejuif, France), using the qScript XLT 1-Step RT-PCR kit (QuantaBio, Beverly, Massachusetts, United States).

2.2.2 Literature data from WWTPs in the USA and Japan. The methods used for sample collection, processing, and quantification are summarized below; full methodological details are provided in the original studies cited.

In Pecson, Darby,[8] 4 L grab samples of raw wastewater were collected at each site every two to three weeks between December 2019 and January 2021. Commercial laboratories analyzed samples using standardized methods. Upon receipt, they were refrigerated at 4 °C and processed within 72 hours of collection. Norovirus (genogroups I and II) and enterovirus were quantified using RT-qPCR, and adenovirus was quantified using qPCR. Nucleic acid extraction was performed with the Zymo Quick-DNA/RNA Viral kit, and assays were run in triplicate with the average gene copies reported. To evaluate analytical recovery rates, MS2 and PhiX174 were spiked into 1000 ml of the sample at approximately 10^8 plaque-forming units (PFU). A minimum recovery efficiency threshold of 1% was set for all matrix spikes.

In Kazama, Miura,[17] 1 L grab samples of raw wastewater were collected weekly from a single WWTP between 2013 and 2016. Samples were transported to the laboratory on ice and stored in a deep freezer on the same day as collection. Norovirus genogroups I and II were quantified using RNA extraction with the QIAamp Viral RNA Mini kit, followed by RT-qPCR. Murine norovirus (MNV) was spiked in each sample as a whole-process control and quantified by qPCR. A minimum recovery efficiency threshold of 1% was applied to all matrix spikes.

For all data sets, concentrations of norovirus genogroups I and II were summed to analyze and compare the overall distribution of norovirus in wastewater across locations.

2.3 Statistical modelling of temporal variations in virus concentrations

2.3.1 Model formulation. To model the variability and uncertainty in virus concentrations, parametric probability distributions were fitted to the monitoring data. The analysis assumed stationarity, meaning the mean, variance, and autocorrelation of concentrations remain constant over time. This approach estimates the annual distribution of enteric virus concentrations, as this is typically required to calculate annual infection risks in QMRA. If seasonal or outbreak-driven variability were to be explicitly considered, separate distributions for specific periods, such as winter and summer, would be necessary.

Two mixed Poisson distributions, the Poisson lognormal distribution (PLN) and the Poisson gamma distribution (PGA), previously used for QMRA,[15,16] were used to model concentration variability using virus count and processed water volume data. When only virus concentrations were reported, virus counts were back-calculated using reported concentrations and processed water volume data. In this framework, the Poisson distribution accounts for the uncertainty associated with the spatial (random) dispersion of viruses in the water sample, and the continuous distribution (gamma or lognormal) characterizes temporal variation in virus concentrations.[16]

Virus concentrations in wastewater are expected to follow a lognormal distribution due to multiplicative processes such as shedding, decay, aggregation, and disaggregation, which, according to the central limit theorem, result in a normal distribution for the logarithm of the concentrations.[26] The probability function of the Poisson lognormal distribution for a virus count x is:

$$P_{\mathrm{PLN}}(x) = \int_0^{\infty} \frac{(cV)^x \exp(-cV)}{x!} \frac{1}{\sigma c\sqrt{2\pi}} \exp\left\{-\frac{[\ln(c)-\mu]^2}{2\sigma^2}\right\} \mathrm{d}c \quad (1)$$

where V is the sample volume, c is the arithmetic mean virus concentration in the sample, μ is the mean of the natural logarithm of the virus concentration and σ is the standard deviation of the natural logarithm of the virus concentration.

The gamma distribution, on the other hand, has a thinner upper tail compared to the lognormal, reflecting a more rapid decline in probability for high concentrations. This makes it a suitable candidate for modelling virus concentrations in systems where high concentrations are constrained by physical limits, such as dilution in wastewater systems, or a biological limit, such as shedding saturation. The probability function of the Poisson gamma distribution for a virus count x is given by:

$$P_{\mathrm{PGA}}(x) = \int_0^{\infty} \frac{(cV)^x \exp(-cV)}{x!} \left[\frac{1}{\gamma\Gamma(\delta)} \left(\frac{c}{\gamma}\right)^{\delta-1} \exp\left(-\frac{c}{\gamma}\right)\right] \mathrm{d}c \quad (2)$$

where y and δ are parameters of a gamma distribution which has a mean of $y\delta$ and a variance of δy^2.

Three continuous probability distributions—lognormal, gamma, Weibull—were also used to model the temporal variations in reported concentrations. Although these models do not account for the discrete nature of microbial counts, they offer a simpler alternative to mixed Poisson models. We assessed whether these continuous models could adequately approximate the distributions generated by the more complex mixed Poisson distributions. The probability function of the Weibull distribution for a reported virus concentration c is:

$$P_{\mathrm{WB}}(c) = \frac{r}{\lambda}\left(\frac{c}{\lambda}\right)^{r-1} \exp - \left(\frac{c}{\lambda}\right)^{r} \quad (3)$$

where r is the shape parameter and λ is the scale parameter.

2.3.2 Model implementation. For statistical inference, a frequentist method (maximum likelihood estimator) and Bayesian methods (Markov chain Monte Carlo) were applied to estimate parameters of the lognormal, gamma, and Weibull distribution, while only the Bayesian method was used to estimate parameters of the mixed Poisson models. Bayesian inference was carried out as described by Sylvestre *et al.*[15] Briefly, three Markov chains were run for 10^6 iterations after a burn-in phase of 10^4 iterations. The Brooks–Gelman–Rubin scale reduction factor was used to monitor the convergence of the four chains.[27] The effective sample size (ESS), the ratio of the sample size to the amount of autocorrelation in the Markov chains, was evaluated to ensure that the entire posterior distribution was explored.[28] The priors selected for the lognormal distribution parameters were: a diffuse uniform prior for the mean with a uniform distribution with a lower of −10 and upper of 10 and an exponential prior for the standard deviation with an exponential distribution of 0.1. Previous work has shown that variations in the hyperparameter values of this exponential prior (between 0.1 and 3.0) have a negligible effect on the behavior of the upper tail of the distribution when the proportion of non-detects is low,[29] as is the case for the enteric virus data in the present study. All statistical computations were performed using R (v4.1.2).[30]

2.3.3 Model comparison. Temporal variations in sewage virus concentrations were represented with complementary cumulative distribution functions (CCDF) curves as these curves help visualize the behavior of the upper tail of the distributions.[31] Goodness-of-fit criteria were specified to compare the performance of each model. Continuous distributions were compared using the deviance information

criterion (DIC)[32] and the Akaike information criterion (AIC).[33] The two mixed Poisson models were compared using the marginal deviance information criterion (DICm).[34] For each one of these information criterion, a lower value indicate a better performance. A difference in DIC or AIC or DICm of at least three points was considered meaningful.[35]

2.4 Incorporation of analytical recovery rates into the model

Analytical recovery rates of other viruses that were spiked into the wastewater were also evaluated for their ability to inform inter-sample variation in recovery rates. For datasets from Pecson, Darby,[8] we divided sample-specific enteric virus concentrations by the average of the sample-specific recovery rates of MS2 and PhiX174, following the approach used by Pecson, Darby.[8] To evaluate the impact of the recovery rates on the distributional form, the PLN model was fitted to data with and without sample-specific correction for the recovery rate.

2.5 Influence of virus monitoring frequencies on model predictions

Using the Zurich dataset, we evaluated how the monitoring frequency affects predictions of virus concentrations modelled by the PLN distribution over a period of a year. To achieve this, we randomly sub-sampled the dataset to represent different monitoring intervals: monthly, every two weeks, and weekly measurements. We compared the CCDF curves for each sub-sampled dataset with those derived from the original dataset.

Furthermore, we quantified how the monitoring frequency affects the uncertainty of the arithmetic mean virus concentration predicted by the PLN model. Following Olsson,[36] the 100(1 − α)% confidence interval (CI) of the arithmetic mean of a lognormally distributed variable is approximated by:

$$[L, U] = \exp\left(\overline{Y} + \frac{s_Y^2}{2} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{s_Y^2}{n} + \frac{s_Y^4}{2(n-1)}}\right) \quad (4)$$

where $\bar{Y}$ and s_Y^2 are, respectively, the sample mean and the unbiased sample variance on the natural-log-transformed concentrations, n is the number of samples, and $z_{1-\frac{\alpha}{2}}$ is the standard-normal quantile (1.96 for a 95% CI). Eqn (4) makes explicit the dependence of CI width on the sample size n and the log-scale variance. In our case, we approximate s_Y^2 using the model-estimated variance parameter σ^2 of the fitted lognormal distributions. To facilitate comparison of the uncertainty across datasets, we expressed the 95% CI width on a $\log_{10}$ scale:

$$\text{Width}_{95} = \log_{10}(U) - \log_{10}(L) \quad (5)$$

We then plotted Width_{95} against n to visualise how increasing monitoring frequency reduces uncertainty in the mean concentration.

3 Results

3.1 Variation in concentrations of viruses across locations

The time series data for enteric viruses from the eight WWTPs are shown in Fig. 1. Norovirus shows a clear temporal pattern with higher concentrations in winter (blue-shaded areas) in Zurich (panel B), and Matsushima, Japan (panel D). In contrast, the WWTP in Lausanne (panel A) shows relatively consistent concentrations throughout the year, while the five WWTPs in California, USA (panel C) display high variability without clear temporal trends. In winter 2019, rotavirus peaks in Lausanne, and adenovirus peaks in California, USA. Adenovirus and enterovirus concentrations are more sporadic in Lausanne and California, showing occasional peaks without a seasonal trend.

Table 2 presents the arithmetic mean concentration and coefficient of variation (CV) for each virus type across Switzerland, Japan, and the USA. Arithmetic mean norovirus concentrations in Japan and Switzerland (approx. 1.0×10^4 GC ml^{-1}) are 10 to 100 higher than in the USA. CVs below 1.0 for Swiss sites indicate relatively stable concentrations, while CVs higher than 2.5 in Japan and the USA suggest much greater variability. Adenovirus shows high CVs across all USA sites (up to 2.1), indicating substantial fluctuations compared to Lausanne (CV of 1.0).

3.2 Performance of parametric distributions in predicting temporal variations

Fig. 2 presents CCDFs for norovirus concentrations at four WWTPs, with mixed Poisson distributions fitted to count and volume data (panel A) and continuous distributions fitted to reported concentrations (panel B). The complete CCDF curves for all datasets are shown separately in Fig. S4 (discrete mixed Poisson distributions) and Fig. S5 (continuous distributions).

In panel A, the PLN and PGA distributions are compared across norovirus datasets. The PLN distribution closely fits concentrations across all locations. In contrast, the PGA distribution underestimates peak concentrations in the Japan (Matsushima) and USA (OCSD) datasets, though it performs well for the Lausanne and Zürich datasets. The USA (OCSD) dataset highlights a limitation of the PGA when an extreme outlier are present—in this case, a maximum measured concentration 1.8 $\log_{10}$ higher than the second-highest value. The fitting process forces the gamma distribution to adjust its parameters to accommodate this extreme value. However, due to the thinner tail of the gamma compared to the lognormal, this adjustment distorts the overall fit, causing deviations in how the model represents lower concentrations.

Panel B shows the CCDFs of three continuous distribution lognormal, gamma, and Weibull distributions fitted to reported concentrations. As for the mixed Poisson distributions, the lognormal distribution generally provides a good fit across all datasets, while the gamma and Weibull distributions underestimate peak concentrations in Japan and the USA.

Those visual comparisons are supported by statistical comparisons using information criteria (Tables S1 and S2).

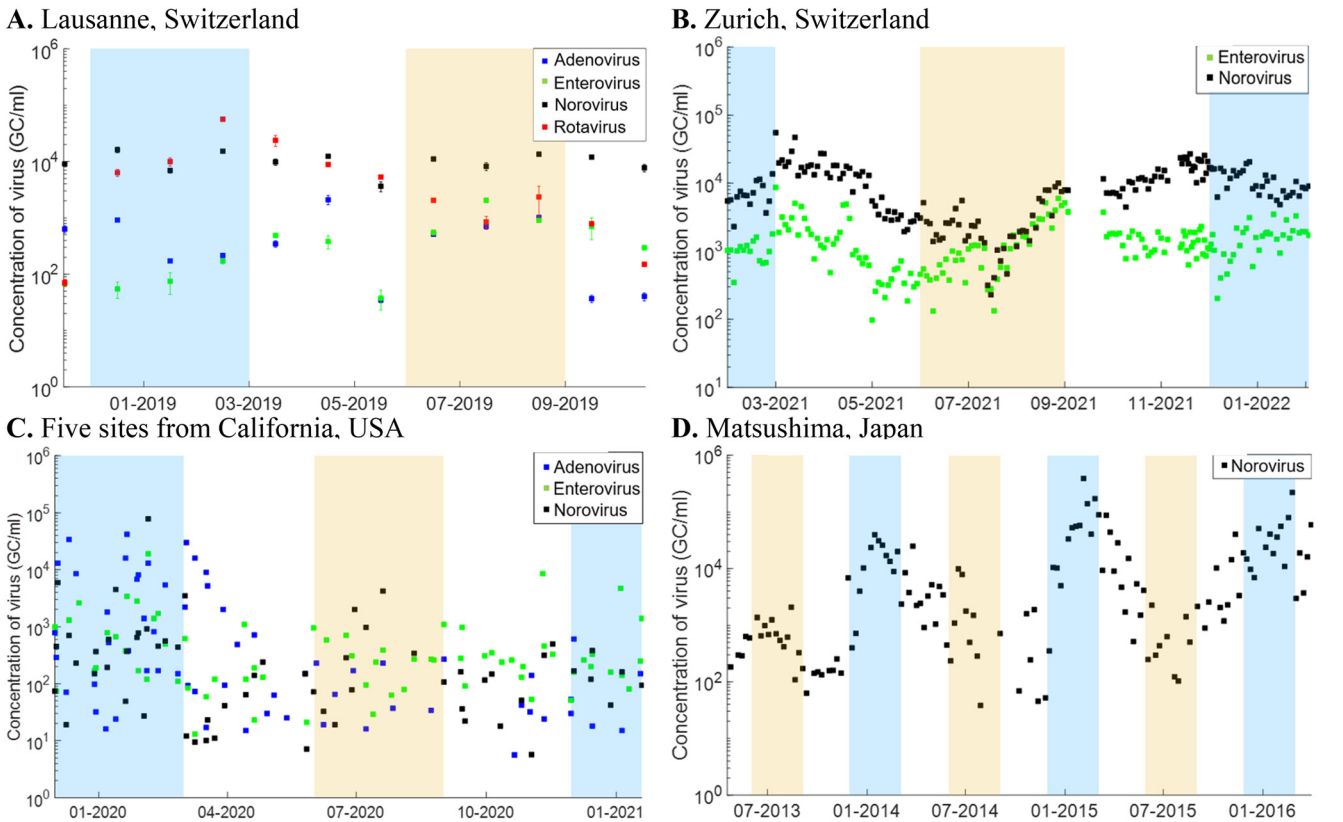

Fig. 1 Concentrations of enteric viruses in influent water from wastewater treatment plants (WWTPs) in genome copies (GC) per millilitre (ml) wastewater in Switzerland, USA, and Japan. Blue shades indicate winter seasons, and orange shades indicate summer seasons.

Table 2 Sample arithmetic mean concentration in genome copies (GC) per milliliter (ml), coefficient of variation, and the sigma parameter of the lognormal distribution of adenovirus, enterovirus, norovirus, and rotavirus in influent water from wastewater treatment plants located in Switzerland, Japan, and USA

Virus	Location of wastewater treatment plants	Sample size	Sample mean concentration (GC ml^{-1})	Coefficient of variation	Lognormal parameter (σ)
Adenovirus	Lausanne, Switzerland	12	5.6×10^2	1.0	1.5
	California, USA – LACSD	14	5.1×10^2	1.9	2.2
	California, USA – LASAN	14	6.7×10^2	1.8	1.9
	California, USA – OCSD	11	2.1×10^3	1.8	2.4
	California, USA – SD	11	9.5×10^2	2.1	2.2
	California, USA – SFPUC	13	1.6×10^3	2.1	2.6
Enterovirus	Lausanne, Switzerland	12	4.8×10^2	1.1	1.4
	Zurich, Switzerland	180	1.5×10^3	0.8	0.8
	California, USA – LACSD	10	1.8×10^2	0.6	1.5
	California, USA – LASAN	17	2.1×10^2	1.0	0.6
	California, USA – OCSD	13	5.1×10^2	2.6	1.6
	California, USA – SD	14	2.2×10^2	2.0	1.2
	California, USA – SFPUC	17	3.5×10^2	0.9	1.2
Norovirus	Lausanne, Switzerland	12	1.0×10^4	0.3	0.4
	Zurich, Switzerland	180	1.0×10^4	0.8	0.9
	Matsushima, Japan	131	1.8×10^4	2.5	2.1
	California, USA – LACSD	8	2.9×10^2	1.5	1.6
	California, USA – LASAN	14	1.1×10^2	1.3	1.0
	California, USA – OCSD	11	2.0×10^3	3.1	2.4
	California, USA – SD	15	3.7×10^2	1.6	2.0
	California, USA – SFPUC	13	3.4×10^2	2.5	1.7
Rotavirus	Lausanne, Switzerland	12	9.7×10^3	1.6	2.2

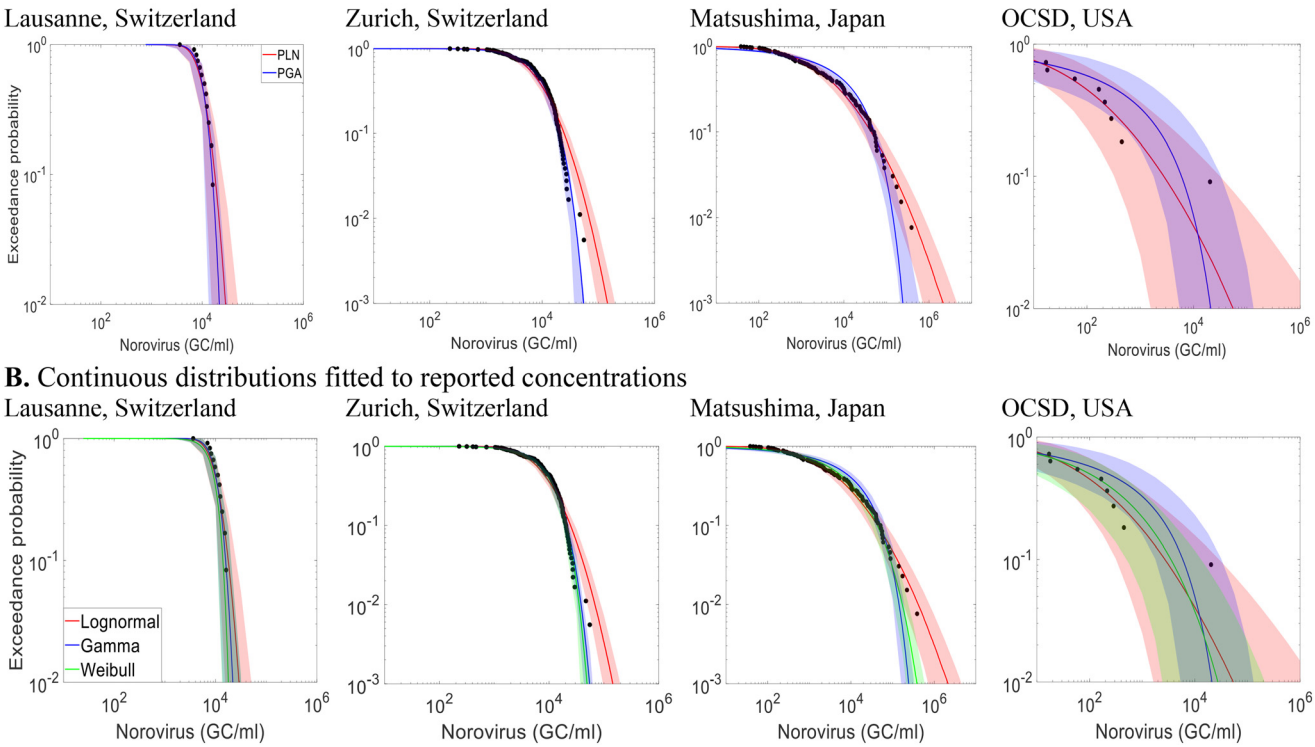

Fig. 2 Complementary cumulative distribution functions (CCDFs) of the mixed Poisson distributions and the continuous distributions of norovirus concentrations in four wastewater treatment plants. Black dots represent observations and bands represent the 95% uncertainty interval. In the upper row, red curves represent PLN distributions, and blue curves represent PGA distributions; in the lower row, red curves represent lognormal distributions, blue curves represent gamma distributions, and green curves represent Weibull distributions.

The PLN and lognormal distributions yield lower DICm and AIC values, respectively, indicating better overall performance than PGA, gamma, and Weibull distributions.

3.3 Impact of sample-specific recovery rates on distributional form

Fig. 3 illustrates recovery rate distributions and CCDFs of virus concentrations (adenovirus, norovirus, and enterovirus) at selected USA WWTPs, with and without sample-specific recovery corrections (using MS2 or PhiX174 as controls). The CCDFs highlight how recovery corrections impact the variability and uncertainty in predicted virus concentrations. The full set of recovery-adjusted CCDFs is presented in Fig. S6.

The impact of recovery rate correction on concentration distributions can depend on the distribution of the recovery data. For LACSD, where the recovery rate distribution is symmetric, with a mode of 50% and a mean of 48%, applying the recovery correction results in a horizontal shift in the CCDF, increasing concentrations without changing variability. The SFPUC and LASAN exhibit right-skewed recovery rate distributions, with modes around 20% and means around 40%. For SFPUC, recovery adjustment also leads to a horizontal shift in the upper tail of the distribution without affecting variability and uncertainty. However, for LASAN, the recovery correction does not produce a horizontal shift in the upper tail; instead, it reduces variability and uncertainty by compensating for previously underestimated low concentrations.

For the Zurich data set, recovery rates estimated using MHV process control data (Fig. 4) had a mean of 6%, ranging from 2% to 20%. These low recovery rates, when used to adjust enteric virus concentrations, increased the estimated concentration by up to two orders of magnitude. Applying the MHV recovery correction also results in a horizontal shift in the CCDF, increasing concentrations without changing variability.

3.4 Influence of virus monitoring frequency on distributional fit

Fig. 5 presents CCDFs of enterovirus and norovirus concentrations from sub-sampled data at the ARA Werdhölzli WWTP in Zurich, with monitoring frequencies set to monthly, every two weeks, and weekly. Increasing the monitoring frequency from monthly to every two weeks substantially reduces 95% uncertainty intervals for both viruses. This reduction results in an accurate and less uncertain prediction of norovirus peak concentrations but an underestimation of enterovirus peak concentrations. Further increasing the

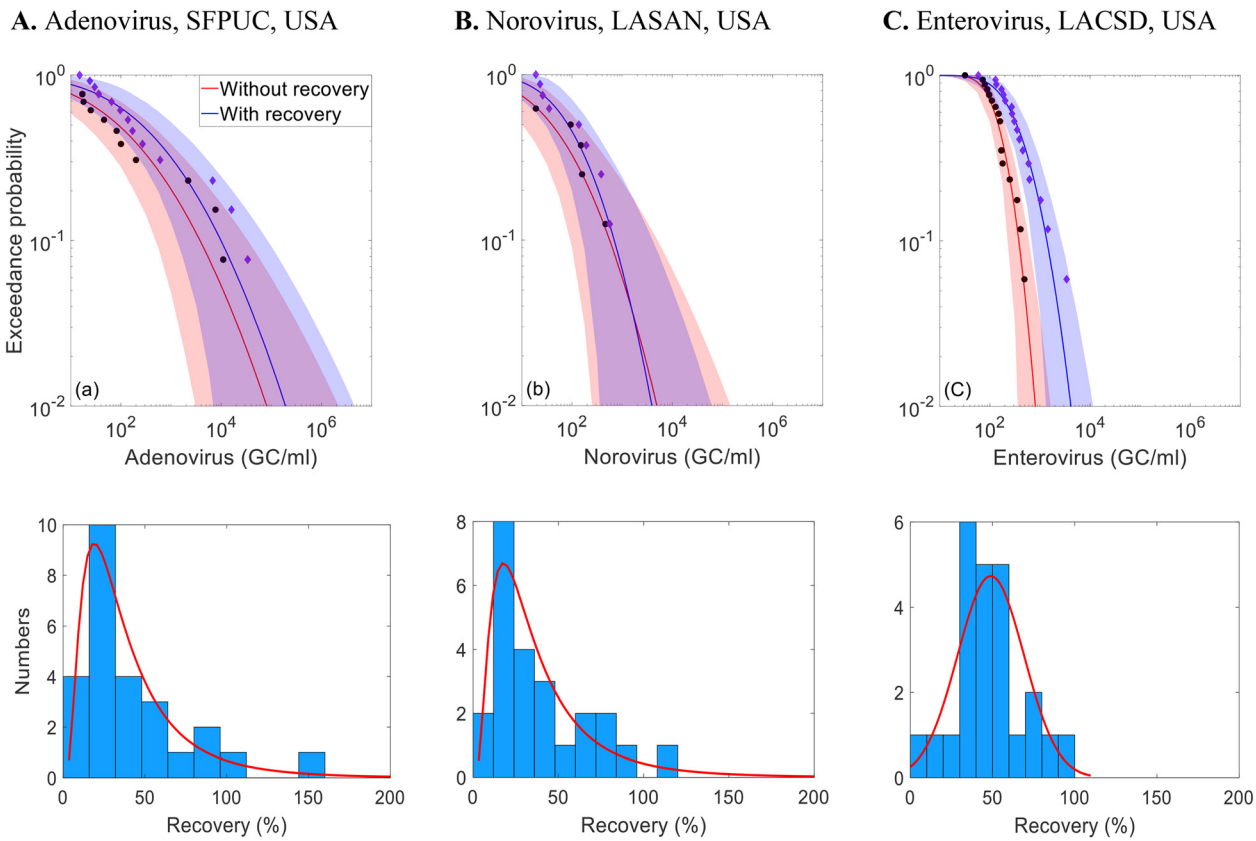

Fig. 3 Complementary cumulative distribution functions (CCDFs) of enteric virus concentrations with and without correcting for sample-specific recovery rates for three wastewater treatment plants (WWTPs) from California, USA. Black dots represent observations and bands represent the 95% uncertainty interval. Histograms illustrate the average analytical recovery rates of MS2 and PhiX174 measured in wastewater samples collected for enteric virus measurements for the three WWTPs from California, USA.

frequency from bi-weekly to weekly provides a minor additional reduction in uncertainty but peak concentrations were accurately predicted for both viruses. It can also be noted that, as sampling frequency increases, the PLN and PGA distributions converge in their ability to predict the variability.

Fig. 6 illustrates the relationship between sample size (*i.e.*, the number of samples collected at a specific monitoring frequency over a set duration) and the uncertainty of the arithmetic mean, represented by the width of the 95% CI in log scale, for different standard deviation (σ) of the lognormal distribution. At a σ of 1.5, approximately 16 samples are required to reduce the 95% CI width on the arithmetic mean to below 1.0 $\log_{10}$. For the Zurich WWTP dataset, where σ is 0.97 and the sample size is 180, the 95% CI width of the mean is well below 0.3 $\log_{10}$. The values of σ for other WWTPs range from 0.7 to 2.6 (Table 2), reflecting varying levels of concentration variability across sites. This indicates that WWTPs with higher σ values, reflecting greater variability in virus concentrations, require more frequent sampling to achieve a precise mean estimate. In contrast, for sites with lower σ values, the variability is smaller, so fewer samples are needed to achieve the same level of precision, allowing for less frequent sampling without compromising the reliability of the mean estimate.

4 Discussion

In this study, we assessed the temporal variability of enteric virus concentrations in influent wastewater from multiple WWTPs across Switzerland, the USA, and Japan. The data set and associated models support models reliant on estimates of distributions of pathogen concentrations in wastewater influent, for example exposure assessment models within a QMRA framework. By evaluating original and literature-derived datasets, we observed clear temporal patterns in norovirus concentrations, with fluctuations ranging up to five orders of magnitude over a year. An increase in norovirus was present in winter months in Zurich, Switzerland, and Matsushima, Japan, consistent with findings from other studies.[37,38] While adenovirus and rotavirus also peaked in winter in Lausanne, Switzerland, and the USA, more extensive datasets would be necessary to investigate seasonality at these locations. Our study revealed that although fluctuations in enteric virus concentrations may appear sporadic in time series analyses, they can be governed by common underlying generative processes described by simple parametric distributions. These distributions must be carefully chosen to accurately predict peak concentrations, as these peaks play a critical role in applications like QMRA.[31,39]

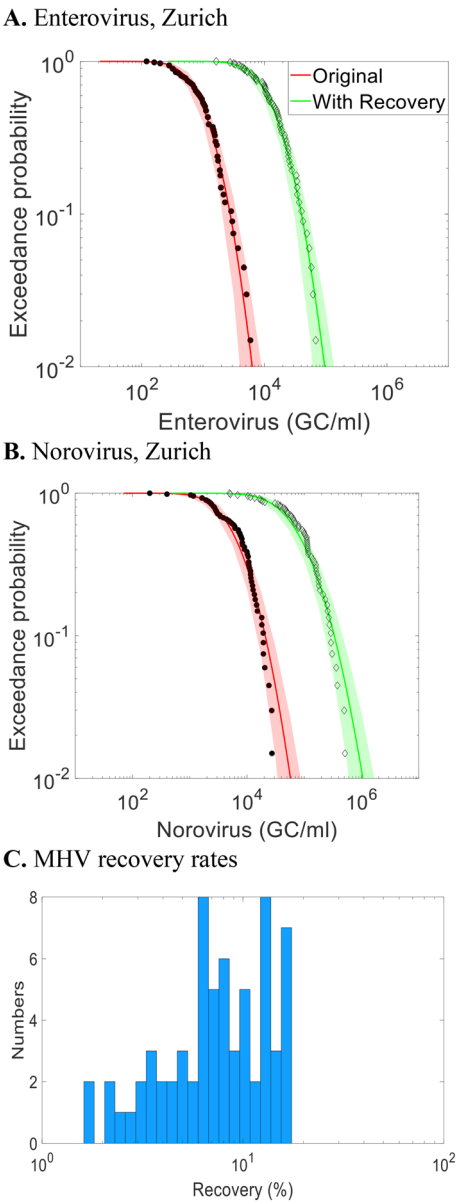

Fig. 4 Complementary cumulative distribution functions (CCDFs) of enterovirus and norovirus with raw measurement data (red curves) for ARA Werdhölzli WWTP in Zurich, sample-specific correction with MHV recovery rates (green curves). Dots represent observations and bands represent the 95% uncertainty interval. Histogram of analytical recovery rates of MHV process control data measured in wastewater samples collected for enteric virus measurements at this WWTP.

Our findings demonstrate that the lognormal distribution is well-suited for modelling the variability of enteric virus concentration across different locations and virus types. Posterior predictive checks and model comparison with deviance information criteria supported this selection. In contrast, gamma and Weibull distributions often fail to capture the high variability observed, especially at CV values above 2.5. This limitation is expected because these distributions lack the necessary skewness to model extreme variability, as their skewness plateaus with increasing CV, unlike the lognormal distribution.[40] The absence of clear multimodality in the empirical distributions suggests that seasonality may be adequately represented by a single lognormal distribution fitted to annual data. However, we note that most of our data sets spanned only one year, which limits the ability to assess seasonality. Longer-term, higher-resolution datasets are needed to confirm this and to assess whether seasonal analyses could improve model performance.

The prevalence of the lognormal distribution indicates that multiplicative processes shape the variability in virus concentrations and suggests the absence of strict constraints (*e.g.*, physical or biological limits) on maximum virus concentrations under monitored conditions. The variability likely arises from the way enteric viruses are shed by infected people within communities. Enteric virus densities in the stools of infected individuals vary widely; for example, ranging from 10^5 to 10^{11} genome copies per gram for norovirus.[41,42] When aggregated, these highly variable virus loads result in wastewater concentrations that follow lognormal distributions. The upper tail of the distribution may be shaped by rare individuals or clusters shedding exceptionally high virus loads. Localized outbreaks may further impact the upper tail by causing sudden increases in virus shedding in communities. Peak concentrations may be amplified in smaller WWTPs with lower dilution capacities. In our study, norovirus peaks were higher at the WWTP in Matsushima, Japan (serving ~9600 people) compared to the ARA Werdhölzli WWTP in Zurich, Switzerland (serving ~471 000 people). It is worth highlighting that not all microorganisms in wastewater follow a lognormal distribution in concentrations. For instance, the concentrations of *Escherichia coli* in influent wastewater from Swiss WWTPs follow a gamma distribution.[43] This distinction is likely due to the lower variability of *E. coli* densities in stools, which typically range from 10^6 to 10^9 colony-forming units (CFU) per gram,[2] combined with relatively constant shedding rates.

Accurately estimating the distribution of virus concentrations requires acknowledging that observations are subject to various measurement uncertainties, including sampling uncertainty, detection limits, and analytical recovery rates. Discrete mixed Poisson distributions are better candidate models than continuous distributions, as they can incorporate uncertainty associated with non-detects into parametric analyses.[44] In addition, discrete mixture distributions—which directly model virus counts within a sample—could be further developed to address uncertainties related to virus aggregation in wastewater.

When adjusting concentrations for recovery, the reliability of surrogate viruses depends on how closely their morphology and behavior in wastewater align with those of the target virus. In our study, the model enveloped virus MHV and the non-enveloped bacteriophages MS2 and PhiX174 had variable recovery estimates, ranging from 1 to 100%. On average, MHV recovery rates were around 1.0 $\log_{10}$ lower than those of MS2 and PhiX174, likely due to the enveloped structure of MHV. Enveloped viruses have unique adsorption, aggregation, and recovery behaviors in raw

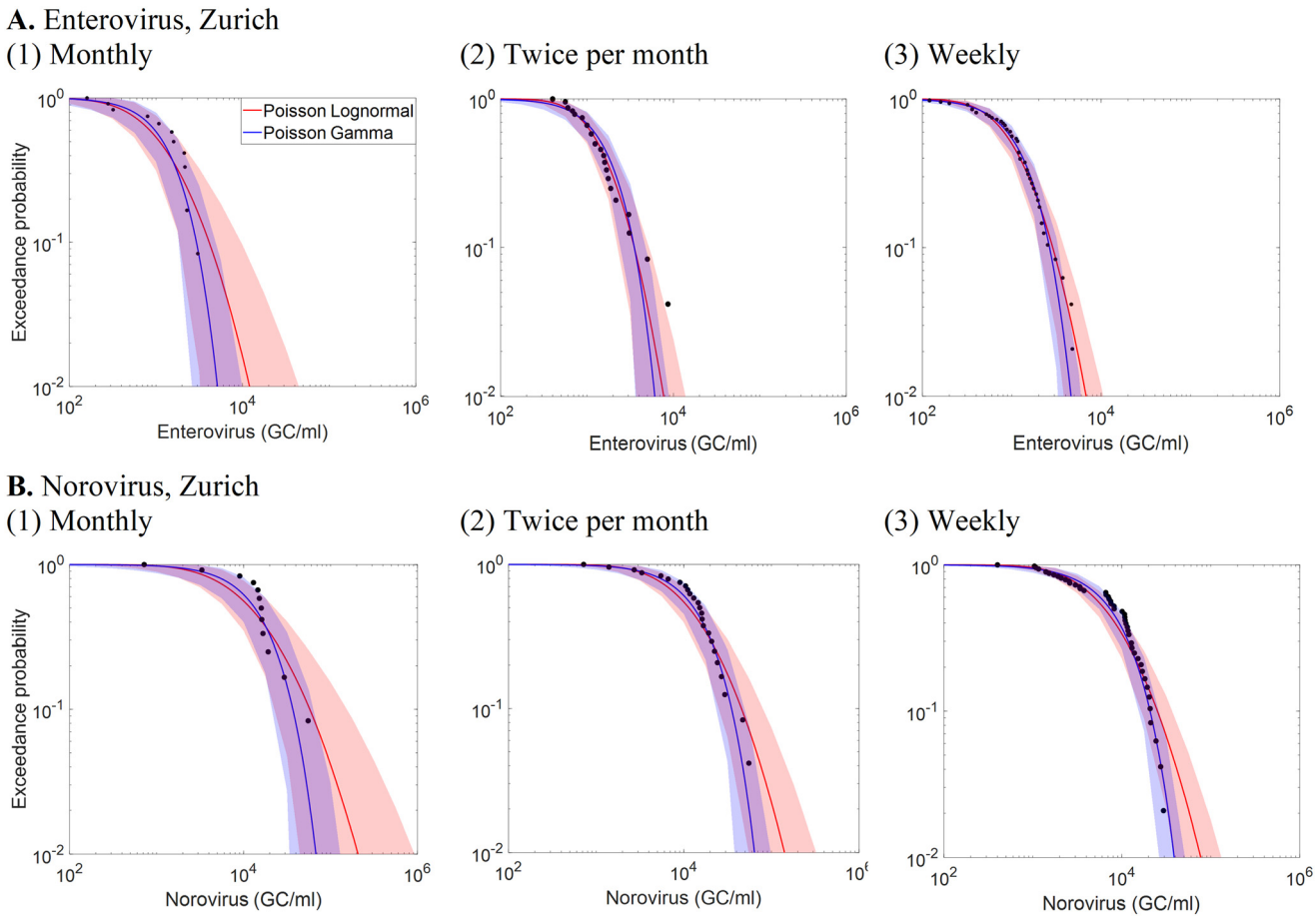

Fig. 5 Complementary cumulative distribution function (CCDFs) of sub-sampled (top) enterovirus and (bottom) norovirus measurement data from ARA Werdhölzli WWTP in Zurich, Switzerland.

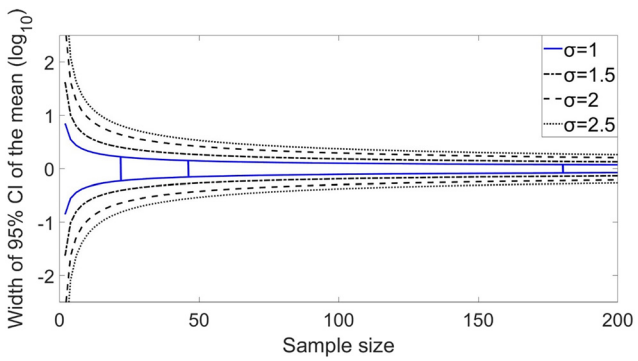

Fig. 6 Relationship between sample size (*i.e.*, the number of samples collected at a specific monitoring frequency for a specific duration and the width of 95% the confidence interval of the arithmetic mean virus concentration in $\log_{10}$ scale for distinct values of the σ parameter of the lognormal distribution. The blue curves represent the relationship for the σ value of 1.0 estimated for norovirus and enterovirus at the ARA Werdhölzli WWTP in Zurich. The vertical lines show the actual sample size (n = 180), a sample size four times smaller (n = 45, similar to weekly monitoring), and a sample size eight times smaller (n = 23; similar to monitoring twice a month).

wastewater, with studies indicating that they are more strongly associated with solids and are more susceptible to inactivation during recovery processes compared to non-enveloped viruses.[7] To improve surrogate selection, evaluating correlations between the recovery rates of these surrogates and those of naturally occurring enteric viruses could provide insights into which surrogates are most reliable for assessing virus concentration distributions.

Our analysis of high-frequency monitoring data from Zurich WWTP demonstrated that increasing the virus monitoring frequency from monthly to bi-weekly significantly reduced the uncertainty of the modelled distribution of concentrations. Further increasing to weekly sampling, however, provided minimal additional reductions in uncertainty but improved the prediction of peaks for enterovirus. For lognormally distributed concentrations, the optimal monitoring frequency varies based on the standard deviation of the natural logarithm of the concentration (σ). For the Zurich WWTP, σ is approximately 1.0 for norovirus and enterovirus, but for other WWTPs in this study, σ values up to 2.6 were found. In these cases, more frequent monitoring may be needed to adequately estimate the distribution for QMRA. Developing adaptive monitoring strategies that start with high-frequency sampling (*e.g.*, weekly)

to assess concentration variability (σ) and then adjusting the frequency as the estimated σ stabilizes could make monitoring efforts more cost-efficient and tailored to site-specific conditions.

Our findings highlight that site-specific differences in mean concentrations, variability, and distribution shapes of enteric viruses can be substantial. Local shedding patterns, population size, or outbreak dynamics may shape the underlying generative processes at each site. At the same time, methodological factors, such as sampling intervals, sample volumes, and analytical recovery rates, also influence measured concentrations. Although all viruses here were measured using PCR-based methods with process controls and recovery spikes, caution is warranted when comparing results across different WWTPs, as these factors may also impact reported variability. This underscores the importance of carefully modelling site-specific distributions, which can complicate meta-analysis. Although pooling data from multiple sites into one single distribution, as proposed by Darby, Olivieri,[12] can offer insights for developing risk-based pathogen treatment requirements in potable reuse; this approach can mask site-specific uncertainty by treating variability as random noise around a single mean. Consequently, sites with distinctly higher (or lower) concentrations may be missed, and peak concentrations, often the key driver in risk assessment, may be underestimated. Random-effects meta-analysis addresses this limitation by allowing each data set to have its own distribution nested within a broader population distribution, preserving site-specific variability.[45] Further development of meta-analytic models for microbial risk assessment would help refine how we estimate and manage virus concentration variability across diverse WWTPs.

5 Conclusions

Based on our analysis of enteric virus datasets from eight WWTPs across Switzerland, Japan, and the USA, we conclude:

- Enteric virus concentrations in wastewater influent vary widely by virus type and location, with average norovirus concentrations ranging from 1.1×10^2 to 1.8×10^4 genome copies per liter across sites, emphasizing the need to preserve site-specific variability rather than relying solely on aggregated estimates.
- The lognormal distribution accurately predicted peak virus concentrations and outperformed gamma and Weibull distributions.
- Recovery-corrected concentrations using MS2, PhiX174, and MHV matrix spike data were successfully incorporated into parametric models.
- Weekly monitoring over a year should be sufficient to estimate the annual average concentration within a 95% confidence interval of 0.5 $\log_{10}$ at most sites. High variability sites ($\sigma > 2$) may need more frequent monitoring to achieve accurate estimates of distributions.

Conflicts of interest

There are no conflicts to declare.

Data availability

Original data on enteric virus and process control concentrations from Swiss wastewater sites, as well as R scripts for statistical analyses, are available on Zenodo: **https://zenodo.org/records/17307946**.

Supplementary information is available. See DOI: **https://doi.org/10.1039/d5ew00286a**.

Acknowledgements

This work was supported by the Swiss National Science Foundation (grant no. 31003A_182468). Data from ARA Werdhölzli in Zurich, Switzerland, were obtained with support from the Swiss Federal Office of Public Health. ES was funded by postdoctoral fellowships from the Natural Sciences and Engineering Research Council of Canada (558161-2021) and the Fonds de Recherche du Québec Nature et Technologies (303866). Thanks to Franziska Böni, Xavier Fernandez-Cassi, Anina Kull, Elyse Stachler, and Blanche Wies for supporting the sample collection and processing.

References

1 E. R. Blatchley III, W. L. Gong, J. E. Alleman, J. B. Rose, D. E. Huffman and M. Otaki, *et al.*, Effects of wastewater disinfection on waterborne bacteria and viruses, *Water Environ. Res.*, 2007, **79**(1), 81–92.
2 M. Kitajima, B. C. Iker, I. L. Pepper and C. P. Gerba, Relative abundance and treatment reduction of viruses during wastewater treatment processes—identification of potential viral indicators, *Sci. Total Environ.*, 2014, **488**, 290–296.
3 C.-M. Zhang, L.-M. Xu, P.-C. Xu and X. C. Wang, Elimination of viruses from domestic wastewater: requirements and technologies, *World J. Microbiol. Biotechnol.*, 2016, **32**, 1–9.
4 I. Xagoraraki, Z. Yin and Z. Svambayev, Fate of viruses in water systems, *J. Environ. Eng.*, 2014, **140**(7), 04014020.
5 S. Petterson, R. Grøndahl-Rosado, V. Nilsen, M. Myrmel and L. J. Robertson, Variability in the recovery of a virus concentration procedure in water: implications for QMRA, *Water Res.*, 2015, **87**, 79–86.
6 D. Li, H.-c. Shi and S. C. Jiang, Concentration of viruses from environmental waters using nanoalumina fiber filters, *J. Microbiol. Methods*, 2010, **81**(1), 33–38.
7 Y. Ye, R. M. Ellenberg, K. E. Graham and K. R. Wigginton, Survivability, partitioning, and recovery of enveloped viruses in untreated municipal wastewater, *Environ. Sci. Technol.*, 2016, **50**(10), 5077–5085.
8 B. M. Pecson, E. Darby, R. Danielson, Y. Dearborn, G. Di Giovanni and W. Jakubowski, *et al.*, Distributions of waterborne pathogens in raw wastewater based on a 14-month, multi-site monitoring campaign, *Water Res.*, 2022, **213**, 118170.

9 M. R. Daelman, B. De Baets, M. C. van Loosdrecht and E. I. Volcke, Influence of sampling strategies on the estimated nitrous oxide emission from wastewater treatment plants, *Water Res.*, 2013, **47**(9), 3120–3130.
10 G. La Rosa and M. Muscillo, Molecular detection of viruses in water and sewage, *Viruses in food and water*, Elsevier, 2013, pp. 97–125.
11 S. E. Eftim, T. Hong, J. Soller, A. Boehm, I. Warren and A. Ichida, *et al.*, Occurrence of norovirus in raw sewage–a systematic literature review and meta-analysis, *Water Res.*, 2017, **111**, 366–374.
12 E. Darby, A. Olivieri, C. Haas, G. Di Giovanni, W. Jakubowski and M. Leddy, *et al.*, Identifying and aggregating high-quality pathogen data: a new approach for potable reuse regulatory development, *Environ. Sci.: Water Res. Technol.*, 2023, **9**(6), 1646–1653.
13 S. P. Nappier, T. Hong, A. Ichida, A. Goldstone and S. E. Eftim, Occurrence of coliphage in raw wastewater and in ambient water: a meta-analysis, *Water Res.*, 2019, **153**, 263–273.
14 É. Sylvestre, J.-B. Burnet, P. Smeets, G. Medema, M. Prévost and S. Dorner, Can routine monitoring of E. coli fully account for peak event concentrations at drinking water intakes in agricultural and urban rivers?, *Water Res.*, 2020, **170**, 115369.
15 É. Sylvestre, M. Prévost, P. Smeets, G. Medema, J. B. Burnet and P. Cantin, *et al.*, Importance of Distributional Forms for the Assessment of Protozoan Pathogens Concentrations in Drinking-Water Sources, *Risk Anal.*, 2021, **41**(8), 1396–1412.
16 C. N. Haas, J. B. Rose and C. P. Gerba, *Quantitative microbial risk assessment*, John Wiley & Sons, 2014.
17 S. Kazama, T. Miura, Y. Masago, Y. Konta, K. Tohma and T. Manaka, *et al.*, Environmental surveillance of norovirus genogroups I and II for sensitive detection of epidemic variants, *Appl. Environ. Microbiol.*, 2017, **83**(9), e03406–e03416.
18 C. Li, É. Sylvestre, X. Fernandez-Cassi, T. R. Julian and T. Kohn, Waterborne virus transport and the associated risks in a large lake, *Water Res.*, 2023, **229**, 119437.
19 J. S. Huisman, J. Scire, L. Caduff, X. Fernandez-Cassi, P. Ganesanandamoorthy and A. Kull, *et al.*, Wastewater-based estimation of the effective reproductive number of SARS-CoV-2, *Environ. Health Perspect.*, 2022, **130**(5), 057011.
20 S. Nadeau, A. J. Devaux, C. Bagutti, M. Alt, E. I. Hampe and M. Kraus, *et al.*, Influenza transmission dynamics quantified from wastewater, *medRxiv*, 2023, preprint, DOI: **10.1101/2023.01.23.23284894**.
21 T. Kageyama, S. Kojima, M. Shinohara, K. Uchida, S. Fukushi and F. B. Hoshino, *et al.*, Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR, *J. Clin. Microbiol.*, 2003, **41**(4), 1548–1557.
22 F. Loisy, R. Atmar, P. Guillon, P. Le Cann, M. Pommepuy and F. Le Guyader, Real-time RT-PCR for norovirus screening in shellfish, *J. Virol. Methods*, 2005, **123**(1), 1–7.
23 S. Monpoeho, A. Dehee, B. Mignotte, L. Schwartzbrod, V. Marechal and J.-C. Nicolas, *et al.*, Quantification of enterovirus RNA in sludge samples using single tube real-time RT-PCR, *BioTechniques*, 2000, **29**(1), 88–93.
24 Y.-L. Tsai, M. Sobsey, L. Sangermano and C. Palmer, Simple method of concentrating enteroviruses and hepatitis A virus from sewage and ocean water for rapid detection by reverse transcriptase-polymerase chain reaction, *Appl. Environ. Microbiol.*, 1993, **59**(10), 3488–3491.
25 X. Fernandez-Cassi, A. Scheidegger, C. Bänziger, F. Cariti, A. T. Corzon and P. Ganesanandamoorthy, *et al.*, Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high, *Water Res.*, 2021, **200**, 117252.
26 A. Andersson, Mechanisms for log normal concentration distributions in the environment, *Sci. Rep.*, 2021, **11**(1), 16418.
27 A. Gelman and K. Shirley, Inference from simulations and monitoring convergence, *Handbook of markov chain monte carlo*, 2011, vol. 6, pp. 163–174.
28 R. E. Kass, B. P. Carlin, A. Gelman and R. M. Neal, Markov chain Monte Carlo in practice: a roundtable discussion, *Am. Stat.*, 1998, **52**(2), 93–100.
29 É. Sylvestre, J. B. Burnet, S. Dorner, P. Smeets, G. Medema and M. Villion, *et al.*, Impact of Hydrometeorological Events for the Selection of Parametric Models for Protozoan Pathogens in Drinking-Water Sources, *Risk Anal.*, 2021, **41**(8), 1413–1426.
30 R Core Team, *R: A language and environment for statistical computing*, Citeseer, 2020.
31 P. W. Smeets, *Stochastic modelling of drinking water treatment in quantitative microbial risk assessment*, IWA Publishing, 2010.
32 D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. Linde, The deviance information criterion: 12 years on, *J. R. Stat. Soc., B: Stat. Methodol.*, 2014, **76**(3), 485–493.
33 H. Akaike, Akaike's information criterion, *International encyclopedia of statistical science*, Springer, 2025, pp. 41–42.
34 A. Quintero and E. Lesaffre, Comparing hierarchical models via the marginalized deviance information criterion, *Stat. Med.*, 2018, **37**(16), 2440–2454.
35 A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, *Bayesian Data Analysis*, 2013.
36 U. Olsson, Confidence intervals for the mean of a log-normal distribution, *J. Stat. Educ.*, 2005, **13**(1), DOI: **10.1080/10691898.2005.11910638**.
37 K. Farkas, M. Marshall, D. Cooper, J. E. McDonald, S. K. Malham and D. E. Peters, *et al.*, Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses, *Environ. Sci. Pollut. Res.*, 2018, **25**, 33391–33401.
38 B. Prevost, F. S. Lucas, A. Goncalves, F. Richard, L. Moulin and S. Wurtzer, Large scale survey of enteric viruses in river and waste water underlines the health status of the local population, *Environ. Int.*, 2015, **79**, 42–50.
39 É. Sylvestre, Systematic assessment of microbial risks associated with hydrometeorological events for drinking water safety management, *PhD thesis*, Ecole Polytechnique, Montreal (Canada), 2020.
40 E. Vargo, R. Pasupathy and L. M. Leemis, Moment-ratio diagrams for univariate distributions, *Computational Probability Applications*, 2017, pp. 149–164.

41 P. Teunis, F. Sukhrie, H. Vennema, J. Bogerman, M. Beersma and M. Koopmans, Shedding of norovirus in symptomatic and asymptomatic infections, *Epidemiol. Infect.*, 2015, **143**(8), 1710–1717.
42 R. L. Atmar, A. R. Opekun, M. A. Gilger, M. K. Estes, S. E. Crawford and F. H. Neill, *et al.*, Norwalk virus shedding after experimental human infection, *Emerging Infect. Dis.*, 2008, **14**(10), 1553.
43 S. Conforti, A. Holschneider, É. Sylvestre and T. R. Julian, Monitoring ESBL-Escherichia coli in Swiss wastewater between November 2021 and November 2022: insights into population carriage, *mSphere*, 2024, **9**(5), e00760-23.
44 A. H. S. Chik, P. J. Schmidt and M. B. Emelko, Learning something from nothing: the critical importance of rethinking microbial non-detects, *Front. Microbiol.*, 2018, **9**, 2304.
45 J. P. Higgins, S. G. Thompson and D. J. Spiegelhalter, A re-evaluation of random-effects meta-analysis, *J. R. Stat. Soc., A: Stat. Soc.*, 2009, **172**(1), 137–159.