Monitoring of SARS-CoV-2 in wastewater: what normalisation for improved understanding of epidemic trends?

Charlotte Sakarovitch, Olivier Schlosser, Sophie Courtois, Cecile Proust-Lima, Joanne Couallier, Agnes Petrau, Xavier Litrico, Jean-Francois Loret

Journal of Water & Health

IWA PUBLISHING

Journal of Water and Health Vol 20 No 4, 712 doi: 10.2166/wh.2022.012

Monitoring of SARS-CoV-2 in wastewater: what normalisation for improved understanding of epidemic trends?

Charlotte Sakarovitch [a,*], Olivier Schlosser [b], Sophie Courtois [b], Cécile Proust-Lima [c], Joanne Couallier [a], Agnès Pétrau [d], Xavier Litrico [e] and Jean-François Loret [b]

[a] SUEZ, LYRE, 15 av Léonard de Vinci, 33600 Pessac, France
[b] SUEZ, CIRSEE, 38 rue du Président Wilson, 78230 Le Pecq, France
[c] Université de Bordeaux, INSERM, Bordeaux Population Health Center, UMR1219, F-33000 Bordeaux, France
[d] SUEZ Rivages Pro Tech, Technopôle Izarbel, 2 Allée Théodore Monod, 64210 Bidart, France
[e] SUEZ, CB21, 16 Place de l'Iris, 92040 Paris La Défense, France
*Corresponding author. E-mail: charlotte.sakarovitch@suez.com

ABSTRACT

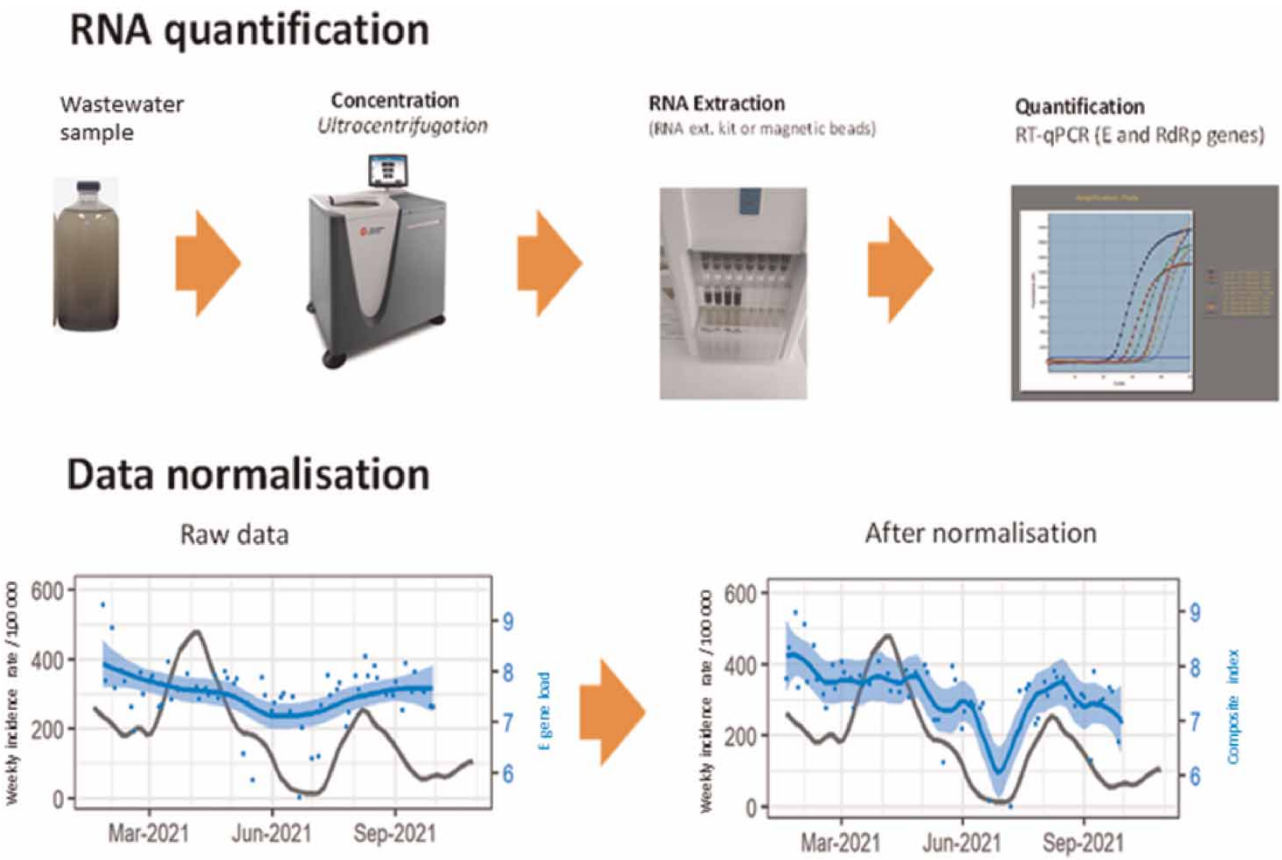
SARS-CoV-2 RNA quantification in wastewater has emerged as a relevant additional means to monitor the COVID-19 pandemic. However, the concentration can be affected by black water dilution factors or movements of the sewer shed population, leading to misinterpretation of measurement results. The aim of this study was to evaluate the performance of different indicators to accurately interpret SARS-CoV-2 in wastewater. Weekly/bi-weekly measurements from three cities in France were analysed from February to September 2021. The concentrations of SARS-CoV-2 gene copies were normalised to the faecal-contributing population using simple sewage component indicators. To reduce the measurement error, a composite index was created to combine simultaneously the information carried by the simple indicators. The results showed that the regularity (mean absolute difference between observation and the smoothed curve) of the simple indicators substantially varied across sampling points. The composite index consistently showed better regularity compared to the other indicators and was associated to the lowest variation in correlation coefficient across sampling points. These findings suggest the recommendation for the use of a composite index in wastewater-based epidemiology to compensate for variability in measurement results.

Key words: COVID-19, normalisation, SARS-CoV-2, wastewater-based epidemiology, wastewater monitoring

HIGHLIGHTS

- SARS-CoV-2 RNA quantification in wastewater shows high variability and need to be normalized.
- The performance of various normalized indicators, including a composite index was assessed via two criteria: regularity and correlation with the incidence rate.
- The composite index showed the best performance.
- Except for the F-specific RNA bacteriophage normalisation index, correlation of all indicators with the incidence rate was substantial.

GRAPHICAL ABSTRACT



INTRODUCTION

The COVID-19 pandemic affecting the world since the end of 2019 has generated a scientific activity unprecedented in the history of mankind, and has reactivated the interest for wastewater epidemiology (Medema *et al.* 2020a; Polo *et al.* 2020). The genome of SARS-CoV-2, the coronavirus causing the disease, is indeed excreted in large quantities and during several weeks in the faecal material of infected individuals (Cheung *et al.* 2020; Gupta *et al.* 2020; Lescure *et al.* 2020; van Doorn *et al.* 2020; Wölfel *et al.* 2020; Zheng *et al.* 2020). It has been demonstrated that this virus is rapidly inactivated in wastewater (Varbanov *et al.* 2021; Wurtzer *et al.* 2021), and several studies report an absence of detection of viable SARS-CoV-2 in this medium (Döhla *et al.* 2020; Rimoldi *et al.* 2020; Albert *et al.* 2021; Westhaus *et al.* 2021; Robinson *et al.* 2022). In contrast, the viral genome is sufficiently stable in this medium (Varbanov *et al.* 2021; Wurtzer *et al.* 2021) to make it a good indicator of the spread of the disease in the population. This approach has been successfully used in the past to monitor the evolution of different infectious diseases, notably to survey the circulation of the poliovirus (WHO 2003).

With the objective of a large-scale surveillance of the circulation of infectious agents, considering viral detection in wastewater presents clear advantages, in comparison with a surveillance based on clinical testing, since it enables to cover large populations with a single analysis, takes into account pre-symptomatic and asymptomatic individuals, and offers a rapid answer enabling an early detection of the epidemic trends, potentially a few days ahead of more conventional health indicators (Medema *et al.* 2020b; Nemudryi *et al.* 2020; Bibby *et al.* 2021; D'Aoust *et al.* 2021a). For these reasons, the European Commission published in March 2021 a recommendation encouraging the European Member States to establish a wastewater surveillance system for SARS-CoV-2, focused primarily on sanitation systems of cities with over 150,000 inhabitants (CE 2021).

However, SARS-CoV-2 RNA concentration in wastewater can be affected by black water dilution factors in the sewer and movements of the population connected to the sewer shed due to commuting or vacations, thus leading to misinterpretation

of measurement results. To take these factors into account, it is possible to correct the quantification results by using indicators reflecting the amount of population connected to the sanitation system. This approach is commonly referred to as 'normalisation'. In its document, the European Commission primarily recommends to normalise the quantification results by the population and the wastewater flow, as already used successfully by some authors (Rusiñol *et al.* 2021) and applied, e.g. in The Netherlands (https://coronadashboard.government.nl/landelijk/rioolwater). Additional normalisation means by using cross-assembly phage or pepper mild mottle virus is also recommended in this document, since both viruses have been described as good indicators of human faecal contamination and have been used for such purpose (Rosario *et al.* 2009; Park *et al.* 2020).

With the aim of rapidly detecting epidemic trends, it is however necessary to use normalisation means based on parameters that (i) reflect the amount of population connected to the sanitation system, (ii) are simultaneously measured on the water sample used to determine the concentration of viral RNA and (iii) are fast and easy to measure, and ideally cheap in order to enable a high frequency of analysis. With this aim, we compared in this study the following approaches:

- A normalisation by the population and the wastewater flow, as recommended by the European Commission.
- A normalisation based on the quantification of F-specific RNA bacteriophages. These phages are routinely quantified in our laboratory, together with SARS-CoV-2 RNA, to verify the recovery of viral genetic material. They are excreted by humans and usually present in large concentrations in urban wastewaters (McMinn *et al.* 2017).
- A normalisation based on the quantification of ammoniacal nitrogen. This parameter is routinely analysed as part of the regulatory monitoring, and ammonium has previously been shown to be a good indicator of population changes, and as such applicable for normalisation purposes in wastewater-based epidemiology (Been *et al.* 2014). More recently, it has been applied to normalise SARS-CoV-2 concentration in a monitoring programme in England (Sweetapple *et al.* 2021).

We also considered a composite index combining the information carried by all the indicators generated by the previous approaches. Indeed, each SARS-CoV-2 RNA indicator is subject to substantial random measurement error and combining the indicators into a composite score could thus reduce the measurement uncertainty (von Oertzen *et al.* 2010).

MATERIALS AND METHODS

Study sites and sample collection

This study used the data produced in the framework of the COVID-19 wastewater surveillance programme driven by the SUEZ for communities in France (https://www.suez.fr/fr-fr/actualites/covid-city-watch). Three French cities (named A, B and C) presenting a minimum of 6-month follow-up were selected for this study. The sewer system is combined at Sites A and C, leading to a significant increase in collected flows during rain events, whereas the sewer system at site B is separate and presents a much more constant flow (flow distributions are presented in Supplementary Material, Figure S2 and Table S2).

The sampling period covered the third epidemic wave (from March to June 2021) and fourth epidemic wave (from July to September 2021) that affected France (sites A and B), or only the third wave (site C). The sampling frequency ranged from twice a week to once every 2 weeks depending on the site and period (Table 1). The samples consisted of 24 h flow-proportional composite samples collected with automated samplers, preserved and transported at 5 °C, and delivered to the laboratory for analysis within 24 h. The sampling points were located either at the inlet of the wastewater treatment plant (A.1, A.2, B.1, B.2 and C.2) or on the collection network (C.1) (Table 1).

Health surveillance data

The weekly incidence rate (the total number of positive tests performed in a 7-day rolling period divided by the number of inhabitants) was extracted from the French institute for public health surveillance database (https://geodes.santepubliquefrance.fr/) at the metropolis level for sites A and B and at the department level for site C. In this national database, the new cases are reported by sampling date and residential address. During all the study period, the tests were free of charge for French residents. The date of entry in our database was the end-term of the 7-day period.

Table 1 | Sites and sampling points analysed

Site	Sampling point	First sample	Last sample	Sampling frequency	Total samples analysed	Population connected
A	1	Jan 18	Sept 29	2/week	72	152,000
	2	Jan 18	Sept 29	2/week	70	45,000
B	1	Feb 3	Aug 30	1–2/week	41	426,000
	2	Feb 3	Aug 30	1–2/week	41	182,000
C	1	Jan 18	June 28	1/week	24	10,000
	2	Jan 18	June 28	1/week	24	36,000
	2	July 5	Sept 27	1/2 weeks	7	36,000

Quantification of SARS-CoV-2 and normalisation parameters

For each wastewater sample, the concentration of RdRp-IP4 gene, E gene and F-specific RNA bacteriophage GGII genome were quantified by the same laboratory (details are provided in Supplementary Material A1). Genomes were quantified using one-step qRT-PCR assays.

F-specific RNA bacteriophages were used as internal standards for our analytical method to estimate the recovery rate of viruses. The reason for using these phages rather than crAssphage or PMMoV as recommended by the European Commission is that our method was developed in 2020, at the beginning of the pandemic, 1 year before the publication of the EU recommendation of March 2021, so we could not take into account this recommendation at that time.

For each sampling point, the size of the population connected to the sewer was estimated using data from the National Institute of Statistics and Economics Studies (INSEE, https://www.insee.fr). The wastewater daily flow rate was measured at each sampling point on each sampling day.

The instruments used for the measurement of ammonium concentrations were those operated routinely and differed from site to site. For site A, the ammonium concentration was measured by the Hach Lange LCK303 cuvette test using the spectrophotometer Hach Lange DR 2800. The measuring range extends from 2 to 47 mg/L of $N\text{-}NH_4$. For site C, the ammonium concentration was measured by the Nanocolor cuvette test using the spectrophotometer Nanocolor UV/VIS II. The measuring range extends from 4 to 80 mg/L of $N\text{-}NH_4$. These two alternative methods were validated according to the ISO 8466-1 and DIN 38402 A51. For site B, the post-steam distillation titrimetric method was used, according to the NF T90-015-1 standard. This method is suited for concentrations higher than 4 mg/L.

Normalised indicator candidates

Three quantifications of SARS-CoV-2 in wastewater were considered: RdRp-IP4 gene RNA concentration, E gene RNA concentration and the arithmetic mean between the two gene concentrations (referred to as RdRp, E and mean hereafter). For each of them, three different adjustments for the amount of faecal matter or shedding population were investigated in addition to the raw concentration:

$$\text{Viral load per day per inhabitants (gc/day/inhabitants)} = [\text{SARS} - \text{CoV} - 2\ \text{RNA}](\text{gc/L}) \times \text{dailyflow}(\text{m}^3/\text{d}) \times 1000(\text{L/m}^3)/\text{Population(inh)};$$

referred to as RdRp load, E load and Mean load.

$$\text{Phage normalisation index (without unit)} = [\text{SARS} - \text{CoV} - 2\ \text{RNA}](\text{gc/L})/[\text{Phages}](\text{gc/L}),$$

referred to as RdRp_phages; E_phages and Mean_phages.

$$\text{Ammoniacal nitrogen normalisation index (gc/N-NH}_4\text{ mg)} = [\text{SARS} - \text{CoV} - 2\ \text{RNA}](\text{gc/L})/[\text{Ammoniacal nitrogen}](\text{mg/L});$$

referred to as $RdRp\_NH_4$, E_NH_4 and $Mean\_NH_4$.

This resulted in 12 indicators, referred to as simple indicators. All quantities were log-transformed to approach a normal distribution.

In addition, we developed a composite index to take into account the information from all the previous indicators simultaneously. This composite indicator had to be a simple formula (simple enough to be implemented in a spreadsheet) and to be the same for every site (allowing to be used on any new site starting the surveillance without the need to learn a new model on

specific historical data or to maintain different models for each site). After a first analysis of the 12 simple indicators, the three normalisation indexes using phages were excluded from our composite index model due to high variability (see 'Results' section). Furthermore, since the mean of RdRp-IP4 and E genes is a combination of the other indicators, we did not consider the three related simple indicators.

We considered a latent process mixed model (Proust-Lima *et al.* 2013) to analyse the repeated measures of the selected six simple indicators (the log-transformed of RdRp-IP4 gene concentration or E gene concentration and the log-transformation of their normalised values according to flow per inhabitant and NH_4 concentration) assuming that they all measured the same underlying quantity: the virus quantity in wastewater (see Supplementary Material A2 for details on the statistical model).

The composite score $\hat{\Lambda}(t)$ was derived as a weighted sum of the six simple indicators noted $Y_k(t)$ for indicator k ($k = 1$–6) and time t according to the following formula:

$$\hat{\Lambda}(t) = \frac{1}{K}\sum_{k=1}^{6}\frac{Y_k(t) - \hat{\eta}_{1k}}{\hat{\eta}_{2k}}$$

where $\hat{\eta}_{1k}$ and $\hat{\eta}_{2k}$ were optimised using the latent process mixed model.

An Excel spreadsheet is provided in the appendix to calculate this composite index (see Supplementary Material A0). The formula parameters have been calculated so as to be used generically.

To avoid overfit and over-optimistic results, we assessed the performance of the composite score in a leave-one-site-out cross-validation (Figure 1).

Measurement data analysis

The concentrations of SARS-CoV-2 RNA, bacteriophages, N-NH_4 and daily flow rates were described for all sites and related sampling points. In a first step, the association between the weekly incidence rate (at the end-term of the 7-day period) and each of the 12 simple indicators (log-transformed values) and the composite index was estimated using the Pearson correlation coefficient. Only the days with all 12 indicators available were used to calculate the correlation. To inform on the uncertainty of estimates and enable comparisons, the correlation coefficient r was associated with its 95% confidence interval. Concentration values below the limit of detection (LOD) or limit of quantification (LOQ), were omitted from calculation of correlation coefficients due to their uncertainty. For visualisation in figures and for the composite index statistical model calibration, these concentrations were substituted by values equal to half the corresponding limit.

Assessment of indicator performance

We compared the performance of the indicators according to the following two criteria: (i) their regularity (since the virus concentration in wastewater is expected to change smoothly over time) and (ii) their association with the weekly incidence rates.

For each sampling point and each indicator, the LOESS (locally estimated scatterplot smoothing) method was applied to summarise the observations into a smooth curve over time. The selection of the best fitting LOESS smoothed curve was based

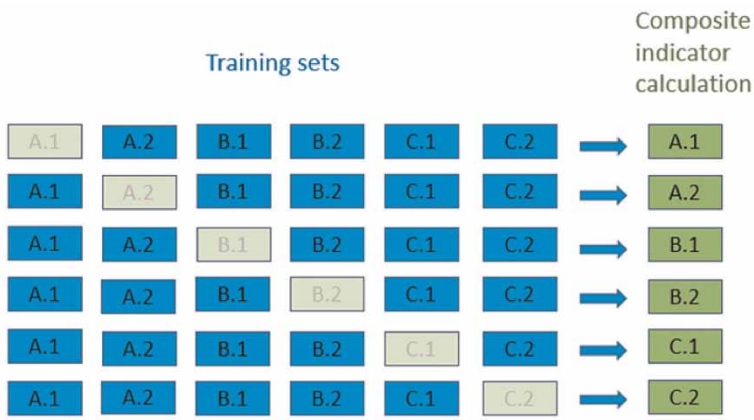


Figure 1 | Schematic representation of cross-validation for the composite index calculation.

on the Akaike information criterion (AIC). The regularity was assessed through the mean absolute difference (MAD) between the observed values and the AIC-selected LOESS smoothed curve. A smaller value of MAD indicates a smaller measurement variability. Since all the indicators did not vary within the same scale, we rescaled the MAD (rMAD) to obtain values that can be compared, by dividing each MAD by the range of the indicator smooth curve. To assess the potential gain of using the composite index over the simple indicators, we computed the percentage of rMAD reduction (rMADred) defined as the difference between the rMAD of the simple indicator and of the composite indicator, divided by the rMAD of the simple indicator. We considered that $-10\% < \text{rMADred} \leq 10\%$ means regularity performances similar to those of the composite index, $10\% < \text{rMADred} \leq 30\%$ is considered irregular and $\text{rMADred} > 30\%$ very irregular.

The association with clinical cases was assessed through the strength of the correlation (the Pearson correlation coefficient) between the AIC-selected LOESS smoothed curve and the weekly incidence rate curve.

RESULTS

Descriptive analysis

Health surveillance data

Figure 2 presents the evolution of the weekly incidence rate at the three sites, during the study period.

The magnitude of each wave was different from one site to the other. Site A experienced a severe second wave in October/November 2020 with a weekly incidence rate greater than 1,000 per 100,000, that was twice that of the third wave and four times that of the fourth wave. Sites B and C had a similar third wave but differed in the magnitude of waves 2 and 4.

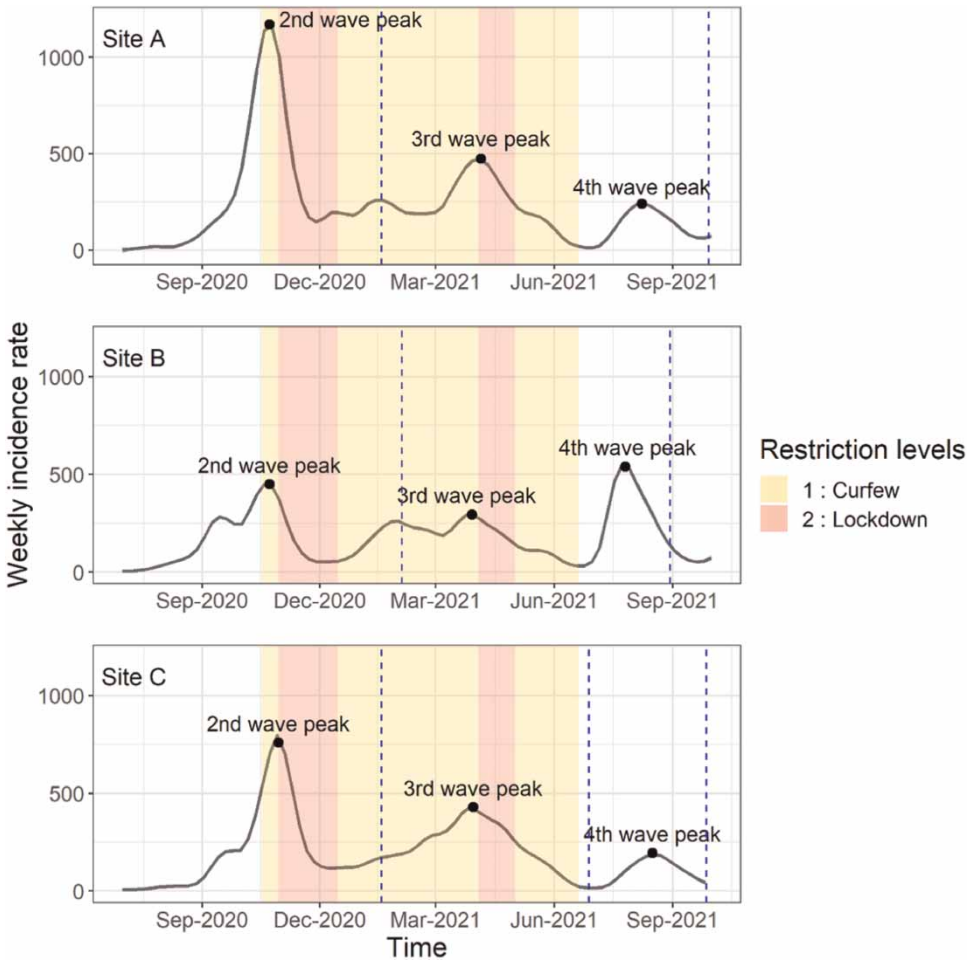


Figure 2 | Evolution of the weekly incidence rate per 100,000 inhabitants associated with sites A and B (metropolis level) and site C (department level). The periods delimited by the dashed lines are the periods of sampling in wastewater for each site.

Quantification of SARS-CoV-2 and normalisation parameters

Over the sampling period, RdRp and E genes were quantified in 94–100% and 79–100% of the samples, respectively. RdRp gene concentration reached the level of 10^6 gc/L on sites A and C and the level of 10^7 gc/L on site B. The highest level of E gene concentration was 10^6 gc/L on sites A and B, and 10^5 gc/L on site C (Figure 3 and Table S1). Bacteriophages RNA was quantified in all samples, with concentrations between 10^7 and 10^9 gc/L orders of magnitude (Figure S1 and Table S1 in Supplementary Material).

The order of magnitude of the flow rate at the inlet of the sewage treatment plants was 10^4 and 10^2 m^3/day at the collection network sampling point (C.1) (Table S2). At sites A and C, the daily flow measurement showed large variations between dry periods and high rainfall events (measured flows at the 99th percentile represented, respectively, about seven and nine times the median flow, during rain events). Conversely, at site B, which is equipped with a separate sewer system, the factor of variation did not exceed 2 (Figure S2).

The median concentration of N-NH_4 in wastewater ranged between 28 and 62 mg/L, depending on the sampling point (Table S3). Measurements at sites A and C showed minimum values much lower than at site B, mainly due to the dilution effect in combined sewer pipes during high rainfall events (Figure S3).

Correlations between the weekly incidence rate and measurement data-based indicators

The analysis of the correlation between the weekly incidence rate and indicators based on measurement data shows some common trends between sites:

At all sampling points except C.1, the weekly incidence rate and the raw concentration were significantly correlated (all $P \leq 0.003$), both for RdRp gene and E gene (Figure 4). The correlation was also statistically significant for the viral load, the N-NH_4 normalisation index and the composite index. The correlation was consistently lower with the phage normalisation index.

The strength of the correlation did not significantly differ between the four indicators, i.e. the raw concentration, the viral load, the N-NH_4 normalised index and the composite index. On the other hand, the strength of the correlation was most often lower for the phage normalisation index as compared to the four other indicators. For both sampling points on each site, no substantial difference in the strength of correlation was observed between targeted genes, whatever the indicator.

On site C, the correlation between the weekly incidence rate and the wastewater indicators clearly differed between the two sampling points. At C.2, all indicators except E_phages significantly correlated with the weekly incidence rate. On the other hand, at the sampling point C.1, where only 19 samples with all indicators available were analysed, the correlation coefficients were low with large confidence intervals.

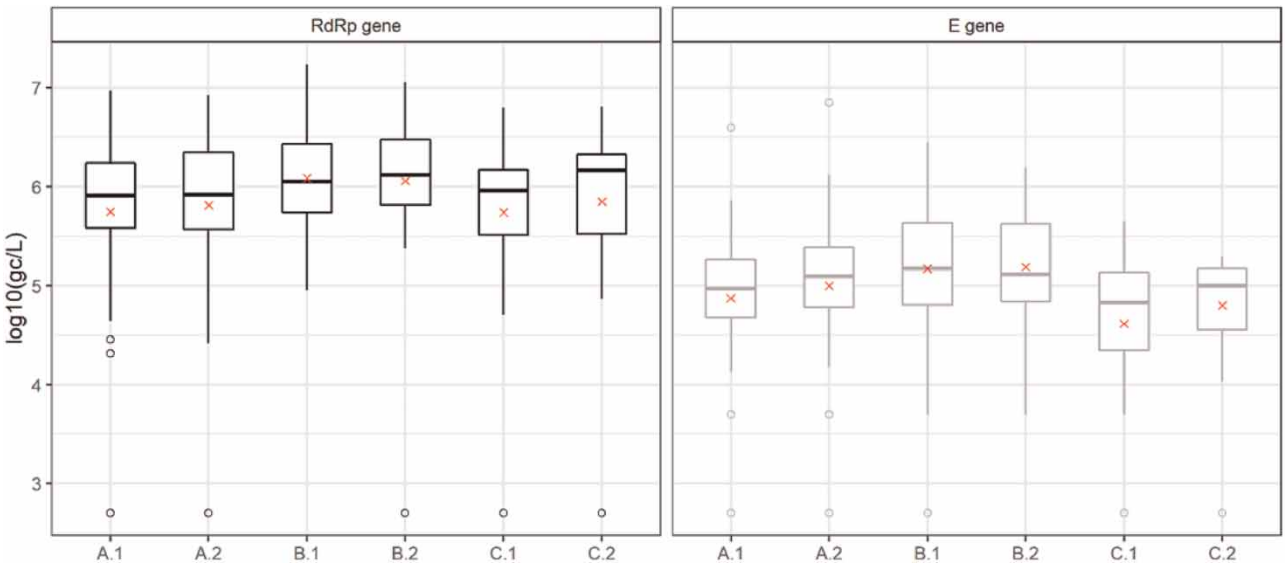


Figure 3 | Distribution of the RdRp gene and E gene concentrations in wastewater at different sampling points.

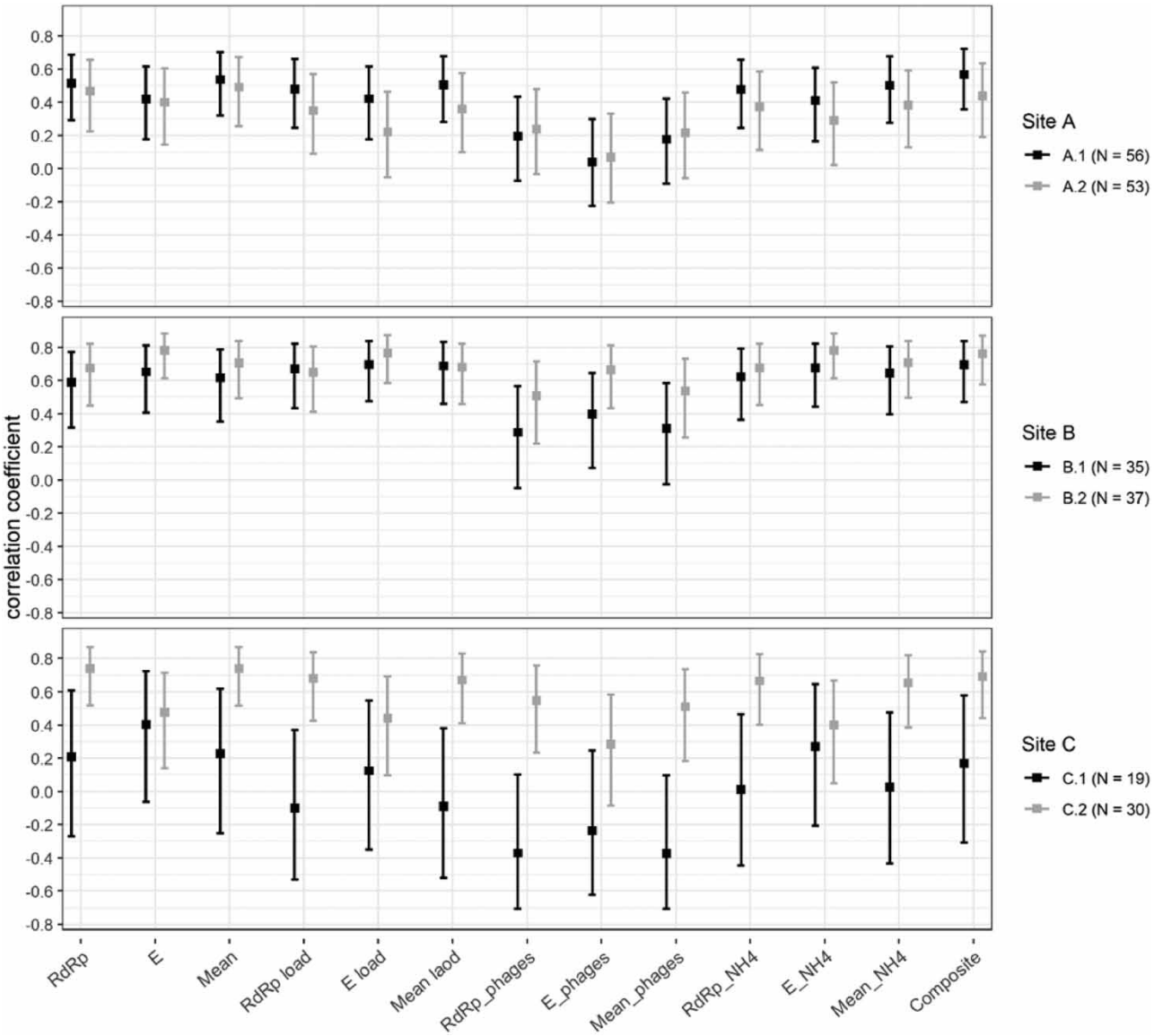


Figure 4 | Pearson correlation coefficient (*r*) and 95% confidence interval between the weekly incidence rate and the wastewater indicators.

Comparison of wastewater indicator performance

Indicator regularity

As shown by Figure 5 where the sampling points A.2 and B.1 were taken as examples, the profile of the smoothed curve and the dispersion of the observation points around the smoothed curve is different depending on the selected indicator. The graphs for all the other sites are available in Supplementary Material (Figures S4–S6).

As indicated by the rMAD displayed in Figure 6, the regularity of the indicators differed substantially across indicators and across sampling points. The composite index showed the overall best regular performance (mean rMAD = 0.12) with low between-site variability (standard deviation = 0.017).

The comparison of the regularity of the simple indicators with that of the composite index (Figure 7) further underlined the variability of the performance of indicators across sampling points: each indicator is very irregular (rMADred>30%) for at least one sampling point and 'irregular' for most of them; E load and E_NH_4 are 'regular' for B.1, B.2 and C.1 (rMADred≤10%) but 'very irregular' for A.1 and A.2 (rMADred>30%); Mean_NH_4 is 'regular' for B.2 but 'very irregular' for A.2.

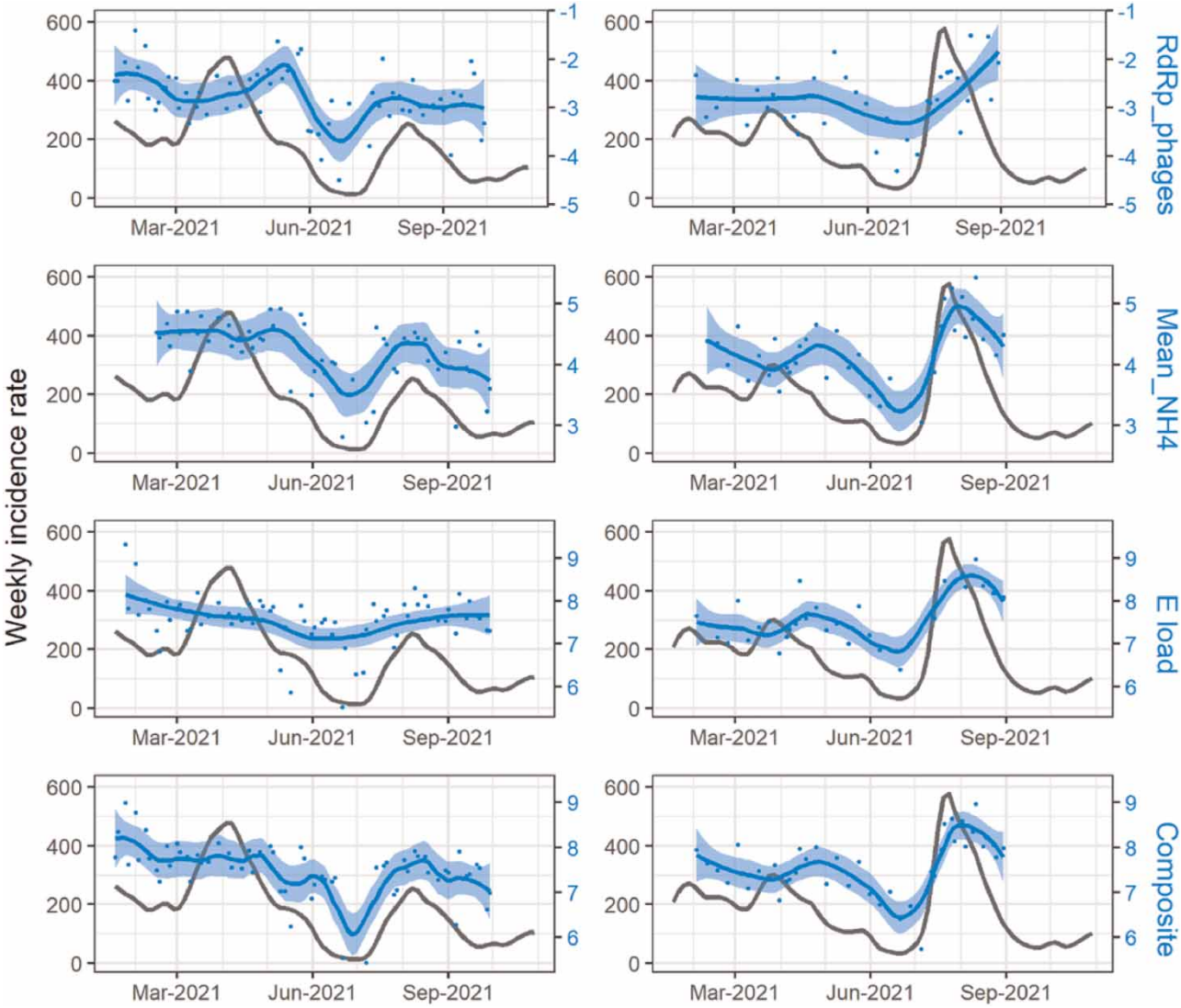


Figure 5 | Indicator observations (blue points) with the associated AIC-selected LOESS smoothed curve (blue line). The black curve is the corresponding weekly incidence rate per 100,000 inhabitants. The left column is for the sampling point A.2 and the right column is for the sampling point B.1. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wh.2022.012.

At C.2, the simple indicators Mean and Mean load outperformed the regularity of the composite index by, respectively, 30 and 20%. Since the sampling frequency was very low at C.2 during the last 3 months of the observation period (only one sample every 2 weeks), the smoothed curve constructed from these points might be misleading and results from that site should be interpreted with caution.

Association of the weekly incidence rate with the LOESS curve-based indicators

Pearson correlation strength used to assess the association between the weekly incidence rate curve and the AIC-selected LOESS smoothed curve is indicated in Figure 8.

Except for the normalisations using phages, correlation coefficients were higher than 0.5 (with $P < 0.0001$) for all indicators and all sites and higher than 0.6 for all indicators except E and E load on site A. For the composite index, the correlation coefficients were among the highest.

Applying the LOESS method increased the strength of the correlation for almost all indicators. The variability between sampling points measured by the standard deviation ranged between 0.04 and 0.18 (Figure 8). The standard deviation was

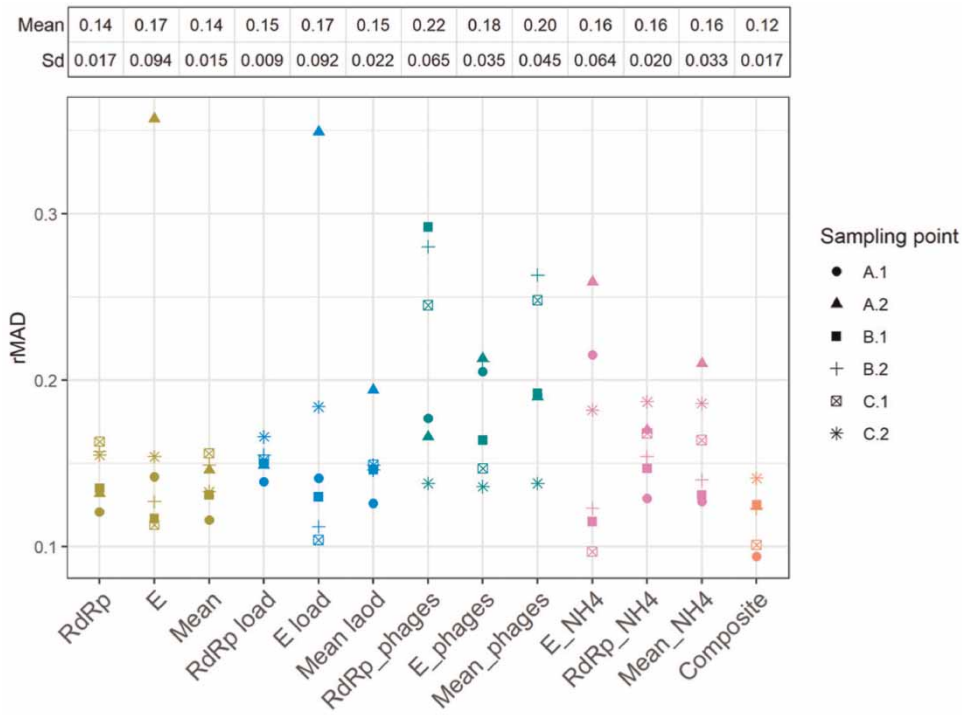


Figure 6 | Rescaled mean absolute difference between the observed values and the AIC-selected LOESS smoothed curve (rMAD) of each indicator for each sampling point. Mean, standard deviation and relative standard deviation for each indicator across sampling points.

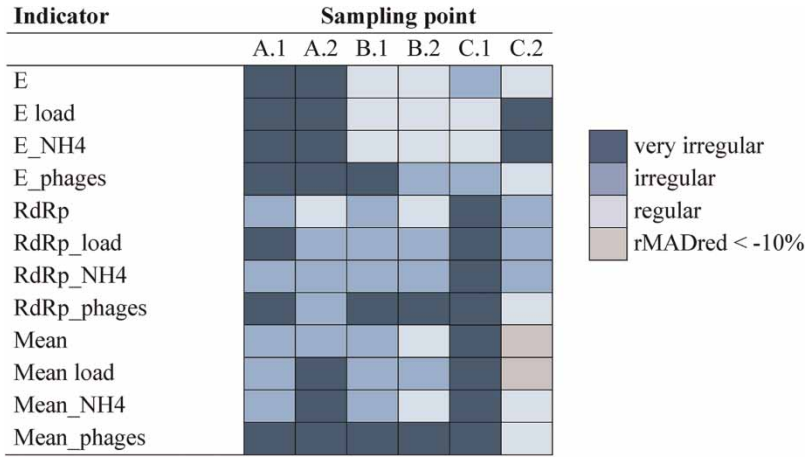


Figure 7 | Normalised indicator regularity performance classification using rMAD reduction (rMADred) obtained by the use of the composite index – light blue: rMADred≤10%, regularity performances alike to the composite index; regular blue: 10%<rMADred≤30% irregular indicator and dark blue: rMADred>30% very irregular indicator. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wh.2022.012.

the lowest for the composite index. Nevertheless, for some of the other indicators such as RdRp, RdRp load, RdRp_NH_4, Mean, Mean load and Mean_NH_4, the standard deviation was too close to represent a significant difference.

DISCUSSION

SARS-CoV-2 RNA quantification in wastewater presents a good potential for monitoring the COVID-19 epidemics, but it shows high variability due to several factors responsible for noise in measurement, such as analysis uncertainty and variation

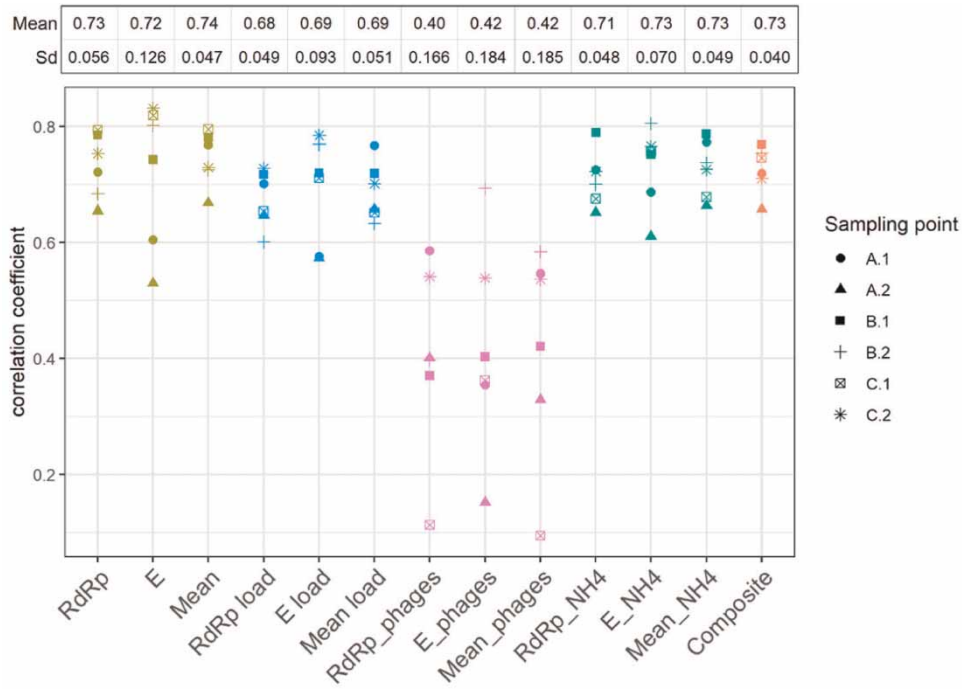


Figure 8 | Pearson correlation coefficient between each LOESS-based indicator and the corresponding weekly incidence rate for each sampling point. Mean, standard deviation and relative standard deviation for each indicator across sampling points.

in the total amount of human faecal matter in wastewater (Medema *et al.* 2020a; Crank *et al.* 2021; Zhu *et al.* 2021). In this work, we compared the performance of 13 indicators derived from RdRp and E genes quantification as candidates for the monitoring of the COVID-19 epidemics. At most sampling points, gene concentration and derived simple indicators were significantly, and moderately, correlated with the weekly incidence rate. However, between sampling points, variations in the strength of the correlation were high, even for the composite index. Applying the LOESS smoothing method to the measurement results and derived indicators in order to focus on the underlying signals substantially increased the strength of the correlation. The composite index was then associated with the lowest standard deviation estimates between the sampling points; however, the difference with other indicators was too small to be significant. Importantly, the performance comparison indicated that the composite index showed the best regularity (i.e. smallest errors) as compared to simple indicators that showed unstable regularity across sampling points.

We chose to rely on two complementary metrics for comparing the performance of different indicators and normalisation methods: regularity and correlation strength with weekly incidence rate.

By quantifying the distance between the observations and the smoothed curve of virus concentration with time, the regularity metric assesses the amount of measurement error in the indicator observations. Indeed, since the quantity of virus in wastewater is the result of the spread of the infection within the population, it is expected to change smoothly with time. Consequently, the further the observations from the smoothed curve, the poorer the indicator performs at estimating the 'true' variations with time we are looking for. Except for C.2 which should be interpreted with caution due to the low sampling frequency, the composite index was the only indicator that consistently showed better regularity, whatever the sampling point. Each simple indicator was 'very irregular' according to rMADred for at least one sampling point and 'irregular' for several. These variabilities highlight differences in simple indicator regularity performance between sites and the interest in having an indicator that consistently demonstrated the best regularity performance, whatever the site. This evidence indicates the value of using the composite index as a wastewater indicator.

The second metric we used to assess the performance of the wastewater indicator was the strength of the correlation of the AIC-selected LOESS smoothed curve with the weekly incidence rate. SARS-CoV-2 RNA is shed in faeces by infected individuals but the proportion of faecal shedders among them and the duration and magnitude of faecal shedding after infection are not very well known (Jones *et al.* 2020; van Doorn *et al.* 2020; Li *et al.* 2021). There is still uncertainty on

the averaged excretion rate distribution, especially since it might depend on several factors such as the severity of the infection, the type of variant and the vaccination status (Kitajima *et al.* 2020; Wölfel *et al.* 2020; Bibby *et al.* 2021; Li *et al.* 2021; Miura *et al.* 2021; Zhu *et al.* 2021). Nevertheless, some studies provided evidence for a maximum excretion quantity within 1 week after infection (Cavany *et al.* 2021; Hoffmann & Alsing 2021; Miura *et al.* 2021). Considering that the cases detected through recorded laboratory tests in France are within a week after the infection, we assumed that the weekly incidence rate and the concentration of SARS-CoV-2 RNA in wastewater were synchronously correlated. Except for normalisation with phages, the correlation strength was good for all indicators at each site ($r>0.6$). In addition, as mentioned above, the composite index was the most stable indicator across sampling points with the lowest variation in correlation coefficient.

These findings demonstrated a lowest performance for the phage normalisation index. Regularity was low or very low, correlation was either not statistically significant or substantially lower than for the other indicators. Some hypotheses may be raised. First, the phage normalisation index is affected by biomolecular analysis uncertainty for both SARS-COV-2 and phages, resulting in higher total uncertainty compared to the other indicators. Second, even if F-specific RNA bacteriophages have been found in large quantities in raw sewage, in the order of 10^3–10^5 plaque forming units/mL (Havelaar *et al.* 1990; Cole *et al.* 2003; Blanch *et al.* 2006), studies demonstrated that their prevalence in human faeces was low, in the order of 10–20% (Grabow *et al.* 1995; Schaper *et al.* 2002). In addition, measurement of F-specific RNA bacteriophages in the wastewater samples in our study indicated large variations in phage load over time, up to 2 log (results not shown). These findings do not support the use of F-specific RNA bacteriophages as a reliable quantitative marker of the amount of faecal matter in wastewater samples. As mentioned in the 'Materials and Methods' section, F-specific RNA bacteriophages were selected before the publication of the EU recommendations. CrAssphage and PMMoV are other phages and viruses that have been proposed as potential useful markers for normalisation of SARS-CoV-2 quantification in wastewater (Wu *et al.* 2020; Heijnen *et al.* 2021; Wilder *et al.* 2021). Although F-specific RNA bacteriophages may present a higher variability in wastewater compared to crAssphage and PMMoV, the analytical uncertainty related with their quantification by PCR is probably in the same order. To date, there is no or limited evidence of improvement of results with this approach (Ai *et al.* 2021; D'Aoust *et al.* 2021a; Feng *et al.* 2021).

Although the strength of correlation with incidence rate is an important first-intention performance criterion to assess wastewater indicators (D'Aoust *et al.* 2021b; Sweetapple *et al.* 2021), it suffers several unavoidable flaws. First, the incidence rate estimation depends on the behaviour toward screening and this behaviour evolves with time as illustrated by the positivity rate of tested people (this rate varied between 1% to almost 10% in France during the study period). Therefore, for instance, if asymptomatic infected people are encouraged to be widely tested or not, if people are vaccinated and more often asymptomatic, if it is necessary to get a negative test to access some entertainment, then for the same level of the epidemic the incidence rate estimated by health authorities might vary. Second, the use of metropolis or departmental incidence rate is one of the limitations in this study. Within each site A and B, the two sampling points corresponded to two different populations, which may be associated to two different incidence rates. These specific incidence rates were unknown, and correlations involved sampling point-specific wastewater indicators and the metropolis incidence rate. This feature may have led to inaccuracy and uncertainty in correlation coefficient estimates. For site C, correlations involved the departmental incidence rate, which may have not been representative of the local incidence of infection.

Another limitation of this study is the relative low size of samples at each sampling point. Although the use of one sample per week is frequent in this field, this limitation led to large uncertainty in the correlation coefficient estimates and made it difficult to find differences in correlation strength between wastewater indicators. More samples are needed to improve this analysis.

CONCLUSION

Except for normalisation with bacteriophages, we observed a substantial association between simple wastewater indicators and the weekly incidence rate of SARS-CoV-2 infection. However, the strength of this correlation greatly varied between the sampling points, making it difficult to identify the best performing indicator. We proposed a composite index that was consistently more regular than the simple indicators across the six sampling points, and was highly correlated to the weekly incidence rate and showed one of the lowest between-sampling points variability in correlation coefficient. The

composite index can be easily computed in any new sampling site based on observations of RdRp and E gene concentrations, N-NH_4 and daily flow per inhabitant (spreadsheet provided in Supplementary Material).

These findings highlight the interest of using the composite index in wastewater-based epidemiology to mitigate variability in measurement results and facilitate identification of trends in the evolution of SARS-CoV-2 RNA concentration in wastewater. Further studies should be performed to confirm these findings. Furthermore, the strong correlations observed between the indicators after smoothing and the weekly incidence rate indicates that the wastewater surveillance of COVID-19 is very valuable. Normalisation using this composite index and curve smoothing may improve the understanding of epidemic trends. Further efforts to model the complex relation between the number of new cases and the wastewater indicators should be conducted to enhance indicator relevance and interpretation of short-term changes in SARS-CoV-2 RNA quantification in wastewater.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ru, P., Pan, X., Bohrerova, Z. & Lee, J. 2021 Wastewater SARS-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States. *The Science of the Total Environment* **801**, 149757. https://doi.org/10.1016/j.scitotenv.2021.149757.

Albert, S., Ruíz, A., Pemán, J., Salavert, M. & Domingo-Calap, P. 2021 Lack of evidence for infectious SARS-CoV-2 in feces and sewage. *European Journal of Clinical Microbiology & Infectious Diseases*. https://doi.org/10.1007/s10096-021-04304-4.

Been, F., Rossi, L., Ort, C., Rudaz, S., Delémont, O. & Esseiva, P. 2014 Population normalization with ammonium in wastewater-based epidemiology: application to illicit drug monitoring. *Environmental Science & Technology* **48** (14), 8162–8169. https://doi.org/10.1021/es5008388.

Bibby, K., Bivins, A., Wu, Z. & North, D. 2021 Making waves: plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research* **202**, 117438. https://doi.org/10.1016/j.watres.2021.117438.

Blanch, A. R., Belanche-Muñoz, L., Bonjoch, X., Ebdon, J., Gantzer, C., Lucena, F., Ottoson, J., Kourtis, C., Iversen, A., Kühn, I., Mocé, L., Muniesa, M., Schwartzbrod, J., Skraber, S., Papageorgiou, G. T., Taylor, H., Wallis, J. & Jofre, J. 2006 Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. *Applied and Environmental Microbiology* **72** (9), 5915–5926. https://doi.org/10.1128/AEM.02453-05.

Cavany, S., Bivins, A., Wu, Z., North, D., Bibby, K. & Perkins, T. A. 2021 *Inferring SARS-CoV-2 RNA Shedding Into Wastewater Relative to Time of Infection (P. 2021.06.03.21258238)*. https://doi.org/10.1101/2021.06.03.21258238.

CE 2021 *Commission Recommendation of 17.3.2021 on A Common Approach to Establish A Systematic Surveillance of SARS-CoV-2 and Its Variants in Wastewaters in the EU. European Commission*.

Cheung, K. S., Hung, I. F. N., Chan, P. P. Y., Lung, K. C., Tso, E., Liu, R., Ng, Y. Y., Chu, M. Y., Chung, T. W. H., Tam, A. R., Yip, C. C. Y., Leung, K.-H., Fung, A. Y.-F., Zhang, R. R., Lin, Y., Cheng, H. M., Zhang, A. J. X., To, K. K. W., Chan, K.-H., Yuen, K.-Y. & Leung, W. K. 2020 Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: systematic review and meta-analysis. *Gastroenterology* **159** (1), 81–95. https://doi.org/10.1053/j.gastro.2020.03.065.

Cole, D., Long, S. C. & Sobsey, M. D. 2003 Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. *Applied and Environmental Microbiology* **69** (11), 6507–6514. https://doi.org/10.1128/AEM.69.11.6507-6514.2003.

Crank, K., Chen, W., Bivins, A., Lowry, S. & Bibby, K. 2021 Contribution of SARS-CoV-2 RNA shedding routes to RNA loads in wastewater. *Science of The Total Environment* **806**, 150376. https://doi.org/10.1016/j.scitotenv.2021.150376.

D'Aoust, P. M., Graber, T. E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Servos, M. R., Srikanthan, N., MacKenzie, M., Figeys, D., Manuel, D., Jüni, P., MacKenzie, A. E. & Delatolla, R. 2021a Catching a resurgence: increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *The Science of the Total Environment* **770**, 145319. https://doi.org/10.1016/j.scitotenv.2021.145319.

D'Aoust, P. M., Mercier, E., Montpetit, D., Jia, J. J., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Langlois, M. A., Servos, M. R., MacKenzie, M., Figeys, D., MacKenzie, A. E., Graber, T. E. & Delatolla, R. 2021b Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Research* **2021** (188), 116560. doi: 10.1016/j.watres.2020.116560.

Döhla, M., Wilbring, G., Schulte, B., Kümmerer, B. M., Diegmann, C., Sib, E., Richter, E., Haag, A., Engelhart, S., Eis-Hübinger, A. M., Exner, M., Streeck, H. & Schmithausen, R. M. 2020 SARS-CoV-2 in environmental samples of quarantined households. *medRxiv*.

Feng, S., Roguet, A., McClary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G. & McLellan, S. L. 2021 Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin Communities. *ACS ES&T Water* **1** (8), 1955–1965. https://doi.org/10.1021/acsestwater.1c00160.

Grabow, W. O. K., Neubrech, T. E., Holtzhausen, C. S. & Jofre, J. 1995 Bacteroides fragilis and Escherichia coli bacteriophages: excretion by humans and animals. *Water Science and Technology* **31** (5), 223–230. https://doi.org/10.1016/0273-1223(95)00270-W.

Gupta, S., Parker, J., Smits, S., Underwood, J. & Dolwani, S. 2020 Persistent viral shedding of SARS-CoV-2 in faeces – a rapid review. *Colorectal Disease: The Official Journal of the Association of Coloproctology of Great Britain and Ireland* **22** (6), 611–620. https://doi.org/10.1111/codi.15138.

Havelaar, A. H., Pot-Hogeboom, W. M., Furuse, K., Pot, R. & Hormann, M. P. 1990 F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. *The Journal of Applied Bacteriology* **69** (1), 30–37. https://doi.org/10.1111/j.1365-2672.1990.tb02908.x.

Heijnen, L., Elsinga, G., de Graaf, M., Molenkamp, R., Koopmans, M. P. G. & Medema, G. 2021 Droplet digital RT-PCR to detect SARS-CoV-2 signature mutations of variants of concern in wastewater. *Science of The Total Environment* **799**, 149456. https://doi.org/10.1016/j.scitotenv.2021.149456.

Hoffmann, T. & Alsing, J. 2021 *Faecal Shedding Models for SARS-CoV-2 RNA Amongst Hospitalised Patients and Implications for Wastewater-Based Epidemiology (p. 2021.03.16.21253603)*. https://doi.org/10.1101/2021.03.16.21253603.

Jones, D., Baluja, M. Q., Graham, D., Corbishley, A., McDonald, J., Malham, S., Hillary, L., Connor, T., Gaze, W., Moura, I., Wilcox, M. & Farkas, K. 2020 *Fecal Shedding of SARS-CoV-2 and Its Potential Role in Person-To-Person Transmission and the Environment-Based Spread of COVID-19*. https://doi.org/10.20944/preprints202007.0471.v1.

Kitajima, M., Ahmed, W., Bibby, K., Carducci, A., Gerba, C. P., Hamilton, K. A., Haramoto, E. & Rose, J. B. 2020 SARS-CoV-2 in wastewater: state of the knowledge and research needs. *Science of The Total Environment* **739**, 139076. https://doi.org/10.1016/j.scitotenv.2020.139076.

Lescure, F.-X., Bouadma, L., Nguyen, D., Parisey, M., Wicky, P.-H., Behillil, S., Gaymard, A., Bouscambert-Duchamp, M., Donati, F., Le Hingrat, Q., Enouf, V., Houhou-Fidouh, N., Valette, M., Mailles, A., Lucet, J.-C., Mentre, F., Duval, X., Descamps, D., Malvy, D., Timsit, JF., Lina, B., Van-der-Werf, S. & Yazdanpanah, Y. 2020 Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *The Lancet Infectious Diseases* **6**, 697–706. https://doi.org/10.1016/S1473-3099(20)30200-0.

Li, X., Zhang, S., Shi, J., Luby, S. P. & Jiang, G. 2021 Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. *Chemical Engineering Journal* **415**, 129039. https://doi.org/10.1016/j.cej.2021.129039.

McMinn, B. R., Ashbolt, N. J. & Korajkic, A. 2017 Bacteriophages as indicators of fecal pollution and enteric virus removal. *Letters in Applied Microbiology* **65** (1), 11–26. https://doi.org/10.1111/lam.12736.

Medema, G., Been, F., Heijnen, L. & Petterson, S. 2020a Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: opportunities and challenges. *Current Opinion in Environmental Science & Health* **17**, 49–71. https://doi.org/10.1016/j.coesh.2020.09.006.

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. 2020b Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in The Netherlands. *Environmental Science & Technology Letters* **7** (7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357.

Miura, F., Kitajima, M. & Omori, R. 2021 Duration of SARS-CoV-2 viral shedding in faeces as a parameter for wastewater-based epidemiology: re-analysis of patient data using a shedding dynamics model. *Science of The Total Environment* **769**, 144549. https://doi.org/10.1016/j.scitotenv.2020.144549.

Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K. K., Wilkinson, R. & Wiedenheft, B. 2020 Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. *Cell Reports Medicine* **1** (6), 100098. https://doi.org/10.1016/j.xcrm.2020.100098.

Park, G. W., Ng, T. F. F., Freeland, A. L., Marconi, V. C., Boom, J. A., Staat, M. A., Montmayeur, A. M., Browne, H., Narayanan, J., Payne, D. C., Cardemil, C. V., Treffiletti, A. & Vinjé, J. 2020 Crassphage as a novel tool to detect human fecal contamination on environmental surfaces and hands. *Emerging Infectious Diseases* **26** (8), 1731–1739. https://doi.org/10.3201/eid2608.200346.

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D. L., Singer, A. C., Graham, D. W. & Romalde, J. L. 2020 Making waves: wastewater-based epidemiology for COVID-19–approaches and challenges for surveillance and prediction. *Water Research* **186**, 116404. https://doi.org/10.1016/j.watres.2020.116404.

Proust-Lima, C., Amieva, H. & Jacqmin-Gadda, H. 2013 Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *British Journal of Mathematical and Statistical Psychology* **66** (3), 470–487. doi: 10.1111/bmsp.12000.

Rimoldi, S. G., Stefani, F., Gigantiello, A., Polesello, S., Comandatore, F., Mileto, D., Maresca, M., Longobardi, C., Mancon, A., Romeri, F., Pagani, C., Cappelli, F., Roscioli, C., Moja, L., Gismondo, M. R. & Salerno, F. 2020 Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. *Science of The Total Environment* **744**, 140911. doi:10.1016/j.scitotenv.2020.140911.

Robinson, C. A., Hsieh, H.-Y., Hsu, S.-Y., Wang, Y., Salcedo, B. T., Belenchia, A., Klutts, J., Zemmer, S., Reynolds, M., Semkiw, E., Foley, T., Wan, X., Wieberg, C. G., Wenzel, J., Lin, C.-H. & Johnson, M. C. 2022 Defining biological and biophysical properties of SARS-CoV-2 genetic material in wastewater. *Science of The Total Environment* **807**, 150786. https://doi.org/10.1016/j.scitotenv.2021.150786.

Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J. & Breitbart, M. 2009 Pepper mild mottle virus as an indicator of fecal pollution. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.00410-09.

Rusiñol, M., Zammit, I., Itarte, M., Forés, E., Martinez Puchol, S., Gironés, R., Borrego, C., Corominas, L. & Bofill-Mas, S. 2021 Monitoring waves of the COVID-19 pandemic: inferences from WWTPs of different sizes. *Science of The Total Environment* **787**, 147463. https://doi.org/10.1016/j.scitotenv.2021.147463.

Schaper, M., Jofre, J., Uys, M. & Grabow, W. O. K. 2002 Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South Africa and Spain. *Journal of Applied Microbiology* **92** (4), 657–667. https://doi.org/10.1046/j.1365-2672.2002.01600.x.

Sweetapple, C., Wade, M. J., Grimsley, J. M. S., Bunce, J. T., Melville-Shreeve, P. & Chen, A. S. 2021 Population normalisation in wastewater-based epidemiology for improved understanding of SARS-CoV-2 prevalence: A multi-site study. *MedRxiv*, 2021.08.03.21261365. https://doi.org/10.1101/2021.08.03.21261365.

van Doorn, A. S., Meijer, B., Frampton, C. M. A., Barclay, M. L. & de Boer, N. K. H. 2020 Systematic review with meta-analysis: SARS-CoV-2 stool testing and the potential for faecal-oral transmission. *Alimentary Pharmacology & Therapeutics* **52** (8), 1276–1288. https://doi.org/10.1111/apt.16036.

Varbanov, M., Bertrand, I., Philippot, S., Retourney, C., Gardette, M., Hartard, C., Jeulin, H., Duval, R. E., Loret, J.-F., Schvoerer, E. & Gantzer, C. 2021 Somatic coliphages are conservative indicators of SARS-CoV-2 inactivation during heat and alkaline pH treatments. *Science of The Total Environment* **797**, 149112. https://doi.org/10.1016/j.scitotenv.2021.149112.

von Oertzen, T., Hertzog, C., Lindenberger, U. & Ghisletta, P. 2010 The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology* **63** (Pt 3), 627–646. doi: 10.1348/000711010X486633.

Westhaus, S., Weber, F. A., Schiwy, S., Linnemann, V., Brinkmann, M., Widera, M., Greve, C., Janke, A., Hollert, H., Wintgens, T. & Ciesek, S. 2021 Detection of SARS-CoV-2 in raw and treated wastewater in Germany – suitability for COVID-19 surveillance and potential transmission risks. *Science of The Total Environment* **751**, 141750. doi: 10.1016/j.scitotenv.2020.141750.

WHO 2003 *Guidelines for Environmental Surveillance of Poliovirus Circulation (WHO/V & B/03.03)*. World Health Organization. Available from: https://apps.who.int/iris/handle/10665/67854.

Wilder, M. L., Middleton, F., Larsen, D. A., Du, Q., Fenty, A., Zeng, T., Insaf, T., Kilaru, P., Collins, M., Kmush, B. & Green, H. C. 2021 Co-quantification of crAssphage increases confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas. *Water Research X* **11**, 100100. https://doi.org/10.1016/j.wroa.2021.100100.

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C. & Wendtner, C. 2020 Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469.

Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A. D., Erickson, T. B., Chai, P. R., Thompson, J. & Alm, E. J. 2020 SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *mSystems* **5** (4), e00614–20. https://doi.org/10.1128/mSystems.00614-20.

Wurtzer, S., Waldman, P., Ferrier-Rembert, A., Frenois-Veyrat, G., Mouchel, J. M., Boni, M., Maday, Y., Marechal, V. & Moulin, L. 2021 Several forms of SARS-CoV-2 RNA can be detected in wastewaters: implication for wastewater-based epidemiology and risk assessment. *Water Research* **198**, 117183. https://doi.org/10.1016/j.watres.2021.117183.

Zheng, S., Fan, J., Yu, F., Feng, B., Lou, B., Zou, Q., Xie, G., Sha, Lin, Wang, R., Yang, X., Chen, W., Wang, Q., Zhang, D., Liu, Y., Gong, R., Ma, Z., Lu, S., Xiao, Y., Gu, Y., Zhang, J., Yao, H., Xu, K., Lu, X., Wei, G., Zhou, J., Fang, Q., Cai, H., Qiu, Y., Sheng, J., Chen, Y. & Liang, T. 2020 Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *BMJ* **369**, m1443. https://doi.org/10.1136/bmj.m1443.

Zhu, Y., Oishi, W., Saito, M., Kitajima, M. & Sano, D. 2021 Early warning of COVID-19 in Tokyo via wastewater-based epidemiology: how feasible it really is? *Journal of Water and Environment Technology* **19** (3), 170–183. https://doi.org/10.2965/jwet.21-024.

First received 14 January 2022; accepted in revised form 25 February 2022. Available online 14 March 2022