Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Regional and temporal differences in the relation between SARS-CoV-2 biomarkers in wastewater and estimated infection prevalence – Insights from long-term surveillance

Björn Helm a,⁎, Michael Geissler b,1, Robin Mayer a, Sara Schubert c,d, Reinhard Oertel c, Roger Dumke b, Alexander Dalpke b,e, Ali El-Armouche c,f, Bertold Renner c, Peter Krebs a

a Institute of Urban and Industrial Water Management, Technische Universität Dresden, Helmholtzstrasse 10, 01069 Dresden, Germany
b Institute of Medical Microbiology and Virology, University Hospital Carl Gustav Carus, Technische Universität Dresden, Fetscherstrasse 74, 01307 Dresden, Germany
c Institute of Clinical Pharmacology, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Fetscherstrasse 74, 01307 Dresden, Germany
d Institute of Hydrobiology, Technische Universität Dresden, Helmholtzstrasse 10, 01069 Dresden, Germany
e University Heidelberg, Institute of Medical Microbiology and Hygiene, Heidelberg, Germany
f Institute of Pharmacology and Toxicology, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Fetscherstrasse 74, 01307 Dresden, Germany

HIGHLIGHTS
• Regression models relate SARS-CoV-2 biomarkers in wastewater to infection prevalence.
• Model components, parametrization and performance time and site-specific.
• Sampling and data evaluation more challenging for small wastewater catchments.
• Log-log transformed input data and lag/lead-time enable satisfactory model performance.
• Diagnostic testing efficiency variable between and throughout epidemic waves.

ABSTRACT
Wastewater-based epidemiology provides a conceptual framework for the evaluation of the prevalence of public health related biomarkers. In the context of the Coronavirus disease-2019, wastewater monitoring emerged as a complementary tool for epidemic management. In this study, we evaluated data from six wastewater treatment plants in the region of Saxony, Germany. The study period lasted from February to December 2021 and covered the third and fourth regional epidemic waves. We collected 1065 daily composite samples and analyzed SARS-CoV-2 RNA concentrations using reverse transcription-quantitative polymerase chain reaction (RT-qPCR). Regression models quantify the relation between RNA concentrations and disease prevalence. We demonstrated that the relation is site and time specific. Median loads per diagnosed case differed by a factor of 3–4 among sites during both waves and were on average 45% higher during the third wave. In most cases, log-log-transformed data achieved better regression performance than non-transformed data and local calibration outperformed global models for all sites. The inclusion of lag/lead time, discharge and detection probability improved model performance in all cases significantly, but the importance of these components was also site and time specific. In all cases, models with lag/lead time and log-log-transformed data obtained satisfactory goodness-of-fit with adjusted coefficients of determination higher.
1. Introduction

In December 2019, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was discovered as cause for Coronavirus disease-2019 (COVID-19). SARS-CoV-2 has spread globally and is an ongoing challenge for public health. Assessing the dynamic of SARS-CoV-2 prevalence in the human population provides crucial information for timely introduction and cancellation of epidemiological control measures. In clinical laboratories, investigation of respiratory specimens is the most important procedure for confirming an infection with this pathogen. Unfortunately, the multifaceted clinical symptoms of COVID-19, ranging from many asymptomatic cases to life-threatening manifestations, hamper the surveillance based on the detection of the pathogen among individuals in a given population. Besides the occurrence in the respiratory tract of infected persons, early studies reported the excretion of SARS-CoV-2 with feces (Chen et al., 2020; Wölfel et al., 2020), this offers the possibility of detecting the specific RNA in wastewater. Relating to the clinical course of infection, presence of RNA in stool samples is time- and patient-dependent but can reach concentrations up to $7 \log_{10}$ genome copies per gram of feces (Schmitz et al., 2021). According to the current knowledge, stool-excreted viruses are not infectious (Pedersen et al., 2022), but RNA remains detectable for a relatively long time in wastewater (Ahmed et al., 2020; Chin et al., 2020). Thus, Wastewater-Based Epidemiology (WBE) offers an additional tool to monitor the prevalence of SARS-CoV-2 infections at local or even national level (Shah et al., 2022).

Testing for virus in clinical specimens of many individuals is expensive and time-consuming. In contrast, WBE has the advantage of providing an integrated picture of the prevalence in the catchment area of a wastewater treatment plant (WWTP). The captured population includes an uncertain connection to the sampling point), as well local influence factors including sampling conditions (catchment flow and concentration variability, sample storage and treatment), virus detection (used procedures for virus enrichment, RNA preparation, reverse-transcription real-time PCR and RNA quantification), test strategy (test rate, positivity rate, reporting efficiency in the population connected to the sampling point), as well local influence factors including wastewater constituents (PCR inhibitors, variable dilution), uncertain population size and time of transport in sewers to the WWTP (Bertels et al., 2022; Kumar et al., 2022). Additionally, the time period between the start of virus shedding and the onset of symptoms in infected persons is approximately 6 days, followed by a decrease of excreted viruses (Cavany et al., 2022; McMahan et al., 2023; Wu et al., 2022). Combined with delays in reporting clinically confirmed cases, this results in a time shift of peaks of concentration of SARS-CoV-2-specific RNA in wastewater and number of positive medical tests (D’Aoust et al., 2021; Barua et al., 2021; Ho et al., 2022; Kuhn et al., 2022; Shah et al., 2022; Wurtzer et al., 2022; Kumar et al., 2022). The suitability of long-term use of WBE as a complementary tool to monitor the prevalence of SARS-CoV-2 is generally accepted. Nevertheless, interpretation of wastewater surveillance data to derive future trends of the number of clinical cases remains challenging (Nagarkar et al., 2021; Fitzgerald et al., 2021; Melvin et al., 2021; Cluzel et al., 2022; Vallejo et al., 2022; Wade et al., 2022; Xiao et al., 2022). This is especially the case for predicting hospitalizations associated with epidemic waves (Aberi et al., 2021; Galani et al., 2022; Petala et al., 2022; Sangsanont et al., 2022). Li et al. (2021) provided a framework for assessing uncertainties in prevalence estimation by WBE.

Regression analysis provides a framework for exploring the relation between a response variable and potential explanatory features. It further supports the evaluation of parameter significance, model performance and uncertainty. Previous studies explored the effect of shedding dynamics (Wu et al., 2022; Cavany et al., 2022), lag/lead time (Barua et al., 2021; Galani et al., 2022; Ho et al., 2022; Kuhn et al., 2022) and flow and/or population normalization (Aberi et al., 2021; Nagarkar et al., 2021). Xiao et al. (2022) examined the relation between wastewater and clinical cases throughout two epidemic waves at one catchment. Aberi et al. (2021) and Fernandez-Cassí et al. (2021), among others, compared the relation between several sites. However, to our knowledge, candidate regression components were not previously evaluated in a multi-site study comprising separate epidemic waves.

Quantitative analysis of micropollutants in wastewater by liquid chromatography – tandem mass spectrometry (LC-MS/MS) is a well-established application in WBE. Multivariate substance concentrations support the identification of modelling processes, flow regimes as well as residence times in wastewater systems (reviewed by Warner et al., 2019). Different pharmaceuticals and human consumption products or their main metabolites were frequently used as so-called indicator substances or tracers for micropollution in the aquatic environment and especially for wastewater flow modelling (Buerge et al., 2008; Warner et al., 2019). In the context of prevalence estimation, WBE of ubiquitous pharmaceuticals supports the estimation of the population connected to the wastewater system (O’Brien et al., 2014; Baz-Lomba et al., 2016; Westhaus et al., 2021) and thus population normalization (Thomas et al., 2017).

In this study, we investigated six WWTPs of different size in the state of Saxony, south-east Germany, for ten months with a standardized sampling and investigation protocol for detection of SARS-CoV-2. The sampling period included two epidemic waves, allowing a deeper analysis of the quantitative relation between virus concentration in wastewater and the reported clinical prevalence in the catchment area of the corresponding WWTPs during these events. We investigated the impact of input data transformation, flow and population normalization, lag/lead time and detection probability at low incidence rates on model performance and uncertainty. Furthermore, we explored the influence of sampling conditions (catchment size, type of wastewater system, sampling frequency) on regression results and estimated the dynamic proportion of undiagnosed cases by wastewater analysis. The following compounds validated flow conditions and load estimation at the investigated WWTPs: carbamazepine (anticonvulsant), gabapentin (anticonvulsant), and metoprolol (beta blocker), as well as cotinine (predominant metabolite of nicotine).

2. Methods

2.1. Study area and data collection

The samples for this study were collected at the inflow of six wastewater treatment plants (WWTP) in the state of Saxony, Germany. Table S1 concludes the basic properties of the connected areas and sampling characteristics. The catchments of the WWTPs each collect the wastewater of one to 13 municipalities. The municipal populations range from 855 to 561,000 residential inhabitants according to the 2021 census update (Statistical
Office of the Free State of Saxony, 2022) and settlement structures cover rural to urban conditions. Catchment names were abbreviated because specific names were deemed irrelevant and incomplete with regard to connected municipalities. Varying proportions of combined and separate sewer systems drain the wastewater catchments. Samples were collected from each site as volume or time proportional 24 h composites. The sampling period included February to December 2021. Sampling took place two to seven times per week, preferably on dry weather days. Samples were stored refrigerated at 4 °C and transported to the laboratory for analysis within three days. WWTP operators provided information on catchment properties, flow data and chemical analysis data according to their internal surveillance monitoring. The number of samples is markedly lower at the two smallest plants due to operational constraints in the work procedure of these WWTPs.

2.2. Virus concentration, RNA extraction, SARS-CoV-2 detection and virus quantification

Enumeration of SARS-CoV-2 gene copies was carried out as described recently (Dumke et al., 2021). Briefly, centrifuged raw wastewater (45 ml) was concentrated by polyethylene glycol precipitation (PEG) and sediment was diluted in phosphate-buffered saline (400–600 μl). An aliquot (200 μl) was used to extract the RNA (RNasey; Qiagen, Hilden, Germany) according to the manufacturer’s recommendations. Eluted RNA (50 μl) was further treated to remove PCR inhibitors (Zymo Research, Irvine, CA, USA). Extracts were subjected to quantitative real-time PCR using the RealStar kit (Altona Diagnostics, Hamburg, Germany) which amplifies the E and S gene of SARS-CoV-2. All samples were tested in duplicate in a QuantStudio5 cycler (Thermo, Waltham, MA, USA). Positive (included in the kit), negative (water) and extraction controls (internal control) were carried out in any run. The result of each run was inspected by the same person and crossing thresholds (Ct) values below 40 were considered positive. Quantification of SARS-CoV-2 genome copies was performed using standard curves (10-fold dilutions) of a synthetic RNA standard (Wuhan strain; Twist Bioscience, San Francisco, CA, USA).

The limit of detection (LOD) of the virus concentration and detection procedure was estimated by spiking wastewater samples (45 ml) from different treatments plants with aliquots of clinical samples (nasopharyngeal swabs, frozen/thawed once) of patients who tested positive for SARS-CoV-2. Genome copies in PEG-enriched wastewaters (ten replicates per test) were compared with those in spiking material to calculate the 95% probability of detection.

To determine the variability of the result of virus enumeration, 10 samples from different treatment plants were mixed and divided into three aliquots. The same steps of virus concentration, RNA preparation and virus detection were performed in parallel in these samples. The precision of virus enumeration was evaluated through coefficient of variation.

To compare the efficacy of PEG precipitation further, 25 samples from all six sites were concentrated in parallel by ultrafiltration. After carefully mixing of samples, Centricron plus-70 columns (10 kDa, MerckMillipore, Burlington, MA, USA) were used to process 45 ml of centrifuged wastewater according to the recommendation of the manufacturer. RNA extraction, virus detection and enumeration were as previously described.

2.3. Preparation, extraction and quantification of pharmaceuticals by SPE-LC–MS/MS

In order to characterize the wastewater of different sized WWTPs, pharmaceutical tracers and/or indicator substances typically found in municipal catchment areas were analyzed. The three ubiquitous pharmaceuticals, carbamazepine (anticonvulsant), gabapentin (anticonvulsant), and metoprolol (beta blocker), as well as cotinine (nicotine metabolism product), were analyzed by solid phase extraction (SPE) and LC-MS/MS according to Rossmann et al. (2014) and Gurke et al. (2015). A detailed method description can be found in the Supplementary Material Tables S2 and S3. Concentrations were transformed to inhabitant specific loads and subsequently used for validating plausibility and representativeness of wastewater samples during the third and fourth epidemic wave.

2.4. Determination of estimated SARS-CoV-2 prevalence

In Germany, data on diagnosed COVID-19 cases is legally reported at the district level. The Federal Robert Koch Institute compiled and provided data through a data hub (Robert Koch Institut, 2022a). Additionally, the Free State of Saxony provided daily updates of newly reported cases from the previous seven days (seven-day incidence) at the municipality level through an online information portal (Saxon State Ministry for Social Affairs and Social Cohesion, 2022). This data is not permanently available; however updates were stored daily and merged into a municipality level time series. We used seven-day incidence data from both data sources as a proxy for infection prevalence. Diagnosed case numbers and prevalences of the WWTP catchments were aggregated from connected municipalities and normalized by 2021 residential population. In cases where not all inhabitants are connected to the WWTP, the epidemiological numbers were reduced in proportion to the connection rate.

We adapted the semi-empirical model proposed by Chiu and Ndeffo-Mbah (2021), in order to estimate the potential contribution of undiagnosed cases on wastewater-based prevalence. The model accounts for test rate and positivity rate in diagnosed cases to infer the prevalence in the untested population. In Germany, positivity rates are published weekly at the states level, based on reporting from a subset of 76 surveillance laboratories (Robert Koch Institut, 2022b). The contributing Saxony surveillance laboratories conducted 24% of all diagnostic tests in the state during the period covered in this study. We aligned weekly test positivity rates and seven-day prevalences and, referring to Chiu and Ndeffo-Mbah (2021), derive:

\[
prev_{wT,pos}(t) = prev_{wT}(t)^p \cdot pos(t)^{1-p} \cdot D^k
\]

Per capita infection prevalence of non-tested population \(\text{pop}_{wT,pos}(t)\) results from the weighted geometric mean of per capita prevalence among the tested \(prev_{wT,pos}(t)\) and test positivity rate \(pos(t)\). The power parameter \(k\) is restricted between 0 and 1, and represents testing selectivity. The limit of \(k = 1\) reflects random testing among the population, with equal prevalences in tested and non-tested populations. Values of \(k\) below one correspond to selective testing of infected subjects.

2.5. Regression model components

The estimation of prevalence in WBE assumes that the number of infected individuals (including asymptomatic, pre-symptomatic, and post-symptomatic cases), who excrete SARS-CoV-2 biomarkers in the service area of a WWTP, relates to the concentration of genome copies in the wastewater. Based on the experiences in the present study (compare Section 3.1), we focused on the concentration of E gene copies of SARS-CoV-2 for further calculations. Eq. (2) provides a conceptual model for prevalence back-calculation:

\[
prev_{wT}(t) = \frac{c(t) \cdot Q(t)}{E(1 - D) \cdot \text{Pop}}
\]

Eq. (2) includes \(prev_{wT}\) as infection prevalence in the (tested and non-tested) population \(\text{pop}\), the concentration \(c(t)\) of SARS-CoV-2 RNA copies in the wastewater and \(Q(t)\) the daily inflow volume at the sampling point. \(E\) is the mean daily excretion rate per infected person, and \(D\) is the decay coefficient due to decomposition in the wastewater system. The components \(E\) and \(D\) are subject to pronounced uncertainty. Both parameters vary over time and depend on epidemiological and site-specific boundary conditions.

In analogy to the seven-day incidence rate, as a proxy for prevalence, we used the moving average of gene copy concentration of a date and the six previous days to represent \(c(t)\). This is also in line with a mean shedding duration of 6.7 days, as estimated by (Fernandez-Cassí et al., 2021). While seven-day accumulation was previously applied in studies (e.g. Weidhaas...
et al., 2021; Fernandez-Cassi et al., 2021; Xiao et al., 2022), other studies have reported longer aggregation periods (e.g. Westhaus et al., 2021; Amereh et al., 2022), and (Galani et al., 2022) found four days aggregation suitable.

If $E$, $D$, $Pop$ and $Q$ are considered sufficiently constant in time, they can be combined in a single scaling factor. In this case, Eq. (2) reduces to a simple linear regression with zero offset. Prevalence and concentration both intensities; they are non-negative and show increasing variance with increasing values. Log transformation improves homoscedasticity and normality of the residuals and prevents negative back-transformed variables. Similarly, Xiao et al. (2022) convoluted log-log transformed clinically reported cases and wastewater gene concentration with a beta distributed transfer function. Eq. (2) implies that the observed load $cQ$ relates proportionally to the excreted non-decayed RNA load. Previous studies have successfully correlated total load and diagnosed cases in the sewershed (e.g. Westhaus et al., 2021; Weidhaas et al., 2021) in multisite studies; however, we argue that this overestimates the model confidence because population size acts as a hidden covariate on explained and explanatory variables.

Load transfer between days, due to travel time or retention dynamics in the sewer, may cause deviations from this relation. SARS-CoV-2 biomarkers are predominantly excreted by stool (Li et al., 2021), and can be partly transported in the solid fraction of wastewater (Westhaus et al., 2021). As a result, mobilization during increased flow could contribute to load variability. Thompson et al. (2011) proposed the evaluation of power-law relationships between concentration and discharge in order to identify export mechanisms for solute and particulate substances at the catchment scale. We extended this conceptual approach by relating prevalence to population normalized discharge $q(t)$ and concentration in an additive regression model. Combining log-transformation and discharge dependency, we derive:

$$\ln (prev(i)) = m' + n \cdot \ln (c(i)) + a \cdot \ln (q(i)) \mid prev(i) = m \cdot c(i)^n \cdot q(i)^a$$

Eq. (3) is a generalized form of Eq. (2). Parameter $m$ accounts for excretion and reduction, while $m'$denotes the log-transformed expression of $m$. Parameter $n$ represents the concentration effect and $a$ represents the discharge effect. For $n = a = 1$ prevalence relates linearly to SARS-CoV-2-specific RNA load. For $a = 0$ discharge yields no effect.

Previous studies reported a temporal load of wastewater concentrations by several days compared to hospitalizations and diagnostic testing (e.g. Aberi et al., 2021; Ho et al., 2022; Peccia et al., 2020). We evaluated time lag through cross-correlation and used a shift of $\pm$ seven days around the lagged concentration with maximum correlation as transformed regression with increasing complexity and two to five parameters.

The component structure provides a set of candidate regression models with increasing complexity and two to five parameters.

$$p(Det) = \frac{1}{1 + \exp - (\alpha Prev + b)}$$

Incorporating all candidate components, the regression equation extends to:

$$Prev = (1 + \exp - (a + b \cdot Prev)) \cdot m \cdot c_0 \cdot n \cdot q^a$$

The component structure provides a set of candidate regression models with increasing complexity and two to five parameters.

2.6. Model fitting, testing and evaluation

All models were fitted using ordinary least-squares optimization in the R programming environment. Significance of regression parameters were tested against two-sided t-statistics. The difference in distribution location was tested according to the Wilcoxon rank sum test. Difference in regression parameters was tested with Tukey's significant difference test. The significance of difference in correlation coefficients was tested according to Dunn and Clark's test.

Prediction intervals were evaluated at the 95 % prediction confidence level. Prediction intervals were preferred over confidence intervals in order to account for the uncertainty related to observations beyond the constituent data set. Relative bounds of the prediction interval were calculated as mean relative difference of upper and lower prediction range over best fit. The goodness of fit performance was evaluated by the coefficient of determination $R^2$ or the adjusted coefficient of determination ($\text{adj}R^2$) in case of models with multiple parameters and relative mean squared error (RMSE). Akaike Information Criterion (AIC) served as an additional evaluation parameter for the comparison of models with varying complexity. Both parameters, $R^2$ and RMSE indicators were used as measures of similarity in time-tagged data series and for comparison among sites.

Models with more than four explanatory variables partly showed non-convergent solutions in the least squares optimization. For these models, latin hypercube sampling generated 3000 parameter combinations that were iteratively tested against the observed data. Parameter combinations with the best model performance were subsequently used as initial values in the regression model optimization. This extended approach resulted in all cases in convergent solutions.

3. Results and discussion

3.1. Evaluation of SARS-CoV-2 detection and sampling plausibility

The LoD of the RNA quantification method was estimated in eight independent experiments with SARS-CoV-2-negative raw waters of different WWTPs. After spiking with virus-positive patient material, samples were PEG enriched, 1:4 diluted with non-spiked PEG concentrates of the same sample and RNA was prepared as described. Overall, 34 dilutions were investigated. PCR determined SARS-CoV-2 concentration of the spiking material in ten replicates per dilution, then the positivity rate of the results was derived. Assuming a positivity of 95 %, an LoD of 58 E gene copies in 45 ml of raw water was determined (Fig. S1) corresponding to approximately 1300 copies per liter. This result is relatively low in comparison with detection limits reported in recent inter-laboratory comparison studies (Ahmed et al., 2022; Pecson et al., 2021) and allows a sensitive determination of SARS-CoV-2 in WWTPs sampled. Regarding the Ct value, a coefficient of variation of 1.0 % for the E gene and 0.7 % for the S gene after concentration and RNA preparation was calculated, resulting in a variation of the concentrations of the E gene and S gene copies of 18.9 % and 14.2 %, respectively (data not shown). Finally, PEG precipitation was compared with a frequently described ultrafiltration method (Centricon Plus-70, 10 kDa). Using parallel enriched raw wastewaters ($n = 24$) of the six WWTPs sampled in the present study, ultrafiltration resulted in equal or lower concentrations of SARS-CoV-2-specific gene copies in 23 samples corresponding to an average of 53.8 ± 27.4 % (E gene) and 58.4 ± 32.5 % (S gene) of those after precipitation. Only one sample exhibited a higher mean concentration (125 %) after ultrafiltration. These findings confirm the general suitability of ultrafiltration with Centricon columns for the enrichment of SARS-CoV-2 in wastewater (Boogaerts et al., 2021; Forés et al., 2021) but the higher efficacy of PEG precipitation.

Fig. 1 demonstrates the consistency and sensitivity of the E and S genes during both epidemic waves. The two genes showed a linear relation throughout the observed range. Higher concentrations for the E gene in the same sample prevailed. Only 1 % of samples during the third wave and 4 % of samples during the fourth wave had higher S than E gene concentrations. During the third wave, the S gene could be quantified in 68 % of all samples, while E gene was above the limit of quantification in 76 % of samples. Quantification probability was much higher in the fourth
wave with 85 % for the S gene and 94 % for the E gene. During both waves, all samples with S gene quantification also had quantifiable E gene concentrations. Coefficients of determination scored at 0.75 for the third wave and 0.86 for the fourth wave. The slope parameter was highly significant in both models. During the third wave, 0.46 S genes (CI 0.41–0.50) corresponded to one E gene. During the fourth wave, the slope was 0.62 (CI 0.60–0.64). Slopes differed with high significance in both waves. In contrast, differences among sampling sites were not systematic, and only in a few occasions significant.

Further, standardized pharmaceutical analysis was used to assess representativeness and plausibility of wastewater samples. Data of small WWTPs Els and Mor showed high variances compared to WWTPs with higher population equivalent and routine sampling, e.g. Dre with 702,000 PE (Table S1). Therefore, measured concentrations and loads of chemical compounds and virus particles in both WWTPs seem prone to impairment. Frequency distributions of concentration and load data for WWTPs Ann, Che, Dre, and Pla were evaluated for each compound in four epidemic phases between February and December 2021 (see Fig. S2). Population normalized loads of all four substances showed similar median values across all sites but differences among the epidemic phases. Gabapentin and Cotinine loads decreased continuously during the study period while Carbamazepin and Metoprolol did not exhibit systematic changes. The similarity of loads across sites confirmed the applicability of pharmaceutical tracers for plausibility validation of samples. In contrast, fluctuations in connected population cannot explain the magnitude of variation in load during the epidemic phases. Temporally resolved information on pharmaceutical prescription and consumption would support the estimation of connected population from load measurements (Baz-Lomba et al., 2016).

3.2. Diagnosed incidence and estimated total prevalence

We defined the periods 15.02.2021–15.06.2021 as the main phase of the third epidemic wave and 01.10.2021–31.12.2021 as the main phase of the fourth epidemic wave in Saxony. The upper panel of Fig. 2 shows the temporal course of the epidemic waves at all sampling sites during the study period.

The catchments with lower populations (Ann, Els, Mor) exhibit the higher short-term variability, higher peak incidence rates and higher deviation from the state level characteristics. Table 1 quantifies these observations. The temporal variability was quantified through the root mean squared error of the four-day delayed time series (RMSE4). This parameter expresses the mean difference between the incidence at a time step and four days later, or the mean error from not updating the incidence during four days. Coefficient of determination (R2) and the root mean squared error (RMSE4) between local and Saxony incidence rates determine the similarity between epidemic phases. While the state-wide incidence rate consistently explained over 90 % of the variance of sites Dre and Che, <50 % variance were covered for sites Ann, Els and Mor during the third epidemic wave. Moreover, the third wave was characterized by lower peak incidence rates and stronger variations among sites. Peak incidences ranged from 182 to 1248, an almost sevenfold difference. An early, pronounced peak at Els site caused the lowest agreement with Saxony-wide data in the data set. In contrast, prolonged high incidences at Ann site led to the second lowest R2 in the data set. The largest sites, Che and Dre, exhibited the best agreement to the Saxony-wide data set over all four descriptive properties.

The lower panel of Fig. 2 illustrates the course of diagnosed and total prevalence as well as the proportion of diagnosed cases for an exemplary selectivity power parameter k = 0.8. Despite the differing dynamics of diagnosed prevalence in both epidemic waves, the estimated total prevalence showed a similar proportional regime. Lower test and positivity rates outside of the epidemic waves resulted in a lower estimated proportion of diagnosed cases. The diagnosed proportion increased during phases of high prevalence without a systematic shift in time. For the selected power parameter, the lowest diagnosed proportion was around 30 % during the low incidence phase in July 2021, while the highest diagnosed proportion co-occurred with the peak of the fourth epidemic wave. According to state-wide diagnostic testing statics, the proportion of unreported cases should not differ significantly during both epidemic waves.

3.3. Local vs. global regression

Several strategies for data stratification were applied in order to assess the suitability of a general regression relation between the E gene concentration and diagnosed seven-day incidence. The basic model of Eq. (3) without time lag and discharge components provides a baseline. The upper two panels of Fig. 3 set contrast on the relation during the third versus the fourth epidemic wave. For the third epidemic wave, sites Che and Dre consistently had the lowest incidence rates per gene concentration. In contrast, Pla and Els exhibited the highest incidence rates per concentration. The relations were more complex for the fourth wave. At incidence rates below 100 diagnosed cases per 100,000 inhabitants, a similar order of sites occurred as during the third wave. However, at incidence rates above 500 diagnosed...
cases per 100,000 inhabitants, sites Mor and Ann provided the highest gene concentrations per incidence rate. Despite a similar range of gene concentrations, the range of observed seven-day incidences during the third wave (7–996 cases per 100,000 inhabitants) was lower than during the fourth wave (25–2931 cases per 100,000 inhabitants). Model parameters and performance indicators are concluded in the supplementary material (Table S4), where the model is indicated as 2–1.

All regression models provided a highly significant relation between virus concentrations and incidences. In almost all cases, the slope parameter $m$ was significantly different from zero. During the third epidemic wave, lower significance levels were associated with low $m$ values at sites Che, Mor and Pla. In contrast, during the fourth wave, high $m$ values at the smaller sites Ann, Els and Mor, were less or not significant, as a consequence of higher parameter standard errors. The power parameter $n$ was highly significant for all models and ranged from 0.15 to 0.99, indicating a less than linear relation between gene concentration and incidence. The fourth epidemic wave resulted in higher power parameters at all sites. In fact, larger catchments Che, Dre and Pla obtained close to linear regression relations during the fourth wave. To some extent, parameters $m$ and $n$ compensated each other, and sites with lower slopes exhibited higher power parameters.

Generally, the goodness-of-fit (measured as $aR^2$) was better for the third epidemic wave than for the fourth. Larger sites (Che, Dre, Pla) achieved better agreement than smaller sites for both epidemic waves. The global calibration over all sites outperformed the local calibration only for site Ann during the third and fourth epidemic wave and for site Els during the fourth wave. In all other cases local calibrations performed better.

In most cases, local calibration functions resulted in narrower prediction intervals than the global calibration. This observation is illustrated by the lower panel of Fig. 3 for site Dre. Additionally, Table S4 gives the full details on the ranges of relative prediction intervals. The ranges are calculated as mean relative difference between the fitted regression relation and the upper and lower 95% prediction boundary. Regression models for the third epidemic wave tended towards narrower prediction intervals than for the fourth wave. For all sites except Els during the third and fourth wave and Ann during the fourth wave, local regression resulted in more compact predictions than global regression. The superior prediction confidence of local models stems from the higher overall variance in the global data set: the range of concentrations at a specific incidence rate is much higher across all sites than at each individual site. This is not compensated by the larger number of samples from all sites combined.

The daily viral gene load at the sampling points is represented in the middle panel of Fig. 3. Load values are normalized by corresponding seven-day diagnosed case numbers. Loads exhibited an approximately log-normal distribution behavior and differed among sites and epidemic waves. Table S5 concludes an overview of interquartile ranges and median values. Across all sites, the third wave had a 45% higher median load. With a median load of approximately 47 gigacopies per diagnosed person per day, site Ann consistently had the highest load per diagnosed case values in both epidemic waves. Difference in location according to Wilcoxon rank sum test was only moderately significant for site Ann but highly significant for all other sites. Median loads per diagnosed case differed by a factor of 3–4 among sites during both waves. Fernandez-Cassi et al. (2021)

Table 1
Descriptive properties of the diagnosed seven-day incidence rates at the sampling sites and in overall Saxony (SN) during the third and fourth epidemic wave. The data in the table includes peak incidence (PI), root mean squared error of the four-day delayed time series (RMSE$_t$), coefficient of determination to the Saxony time series ($R^2_{SN}$) and root mean squared error of each site vs. Saxony time series (RMSE$_{SN}$), all units except $R^2_{SN}$, are diagnosed cases per 100,000 inhabitants.

| Site | Third epidemic wave | Fourth epidemic wave |
|------|---------------------|----------------------|
|      | PI | RMSE$_t$ | $R^2_{SN}$ | PI | RMSE$_t$ | $R^2_{SN}$ |
| Ann  | 393 | 50 | 0.19 | 107 | 2927 | 292 | 0.85 | 143 |
| Che  | 363 | 39 | 0.94 | 55 | 1236 | 115 | 0.96 | 112 |
| Dre  | 182 | 18 | 0.98 | 48 | 1574 | 143 | 0.96 | 71 |
| Els  | 1248 | 130 | 0.01 | 384 | 1742 | 184 | 0.67 | 97 |
| Mor  | 384 | 82 | 0.44 | 136 | 2479 | 265 | 0.88 | 125 |
| Pla  | 429 | 41 | 0.74 | 89 | 1511 | 158 | 0.83 | 11 |
| SN   | 269 | 26 | 1470 | 265 | 158 | 1158 | 136 |

Fig. 2. Upper: Diagnosed seven-day incidence of COVID-19 infections February – December 2021 at the sampling sites and in overall Saxony (SN). Lower: Diagnosed, estimated total prevalence, and proportion of diagnosed cases of COVID-19 infection in Saxony February – December 2021.
Fig. 3. Diagnostic plots for the regression relation during the third (left) and fourth (right) epidemic wave in Saxony. Upper – relation of E gene concentration ($c_{E\text{ gene}}$) and diagnosed seven-day incidence rate ($\text{incidence}_{7d}$), middle – density plot of E gene load ($l_{E\text{ gene}}$) per diagnosed person, lower – time series for site Dre of seven-day incidence (red line), E gene concentration (black dots, void dots below quantification limit, black line seven-day moving average), and 95% prediction interval of seven-day incidence rate according to global (light grey) and local (dark grey) regression relation.
reported similar site-specific differences among several Swiss towns. While sites Che and Dre had similar viral loads to Ann during the third wave, values were significantly lower during the fourth wave. Sites Els, Mor, and Pla had markedly lower loads than the other three sites and a less pronounced difference among the epidemic waves. Site Els, in particular, exhibited a much lower load and stronger variability (higher relative interquartile range) than all other sites. Xiao et al. (2022) observed similar loads per diagnosed case during two epidemic waves. In contrast, our results, except for site Ann, suggest significant differences among epidemic waves.

Differences among the epidemic waves might be caused by the predomination of distinct SARS-CoV-2 variants. The regional Corona information portal (Saxon State Ministry for Social Affairs and Social Cohesion, 2022) reported mostly variant B.1.1.7 (Alpha) during the third epidemic wave, while the fourth wave was mainly caused by variant B.1.617.2 (Delta). The different composition of biomarker genes (compare Section 3.1) could be similarly caused by the distinct variants. In order to substantiate this claim, further research could examine temporal and variant-specific differences in excreted biomarker loads.

Overall, the epidemic waves showed markedly different features in regression relations and distributional patterns. The combined effect of higher short-term variability in the incidence rates (see Section 3.2) and lower sampling frequency at the smallest sites likely contributed to their lower goodness-of-fit and wider prediction intervals. Time proportional sampling and the lower total number of shedding events at the same incidence rates could result in a systematic underrepresentation of the real daily mean concentration, as already demonstrated (Ort et al., 2010) for household chemicals. Lower loads per diagnosed case at the two smallest sites Els and Mor support this argument. Kumar et al. (2022) argued that hydraulic residence time and associated decay of SARS-CoV-2 RNA increases uncertainties and decreases the expected proportion of recovered RNA material. In contrast, our study resulted in higher RNA recovery per infected case for larger catchments, with longer hydraulic residence time. This substantiates that biased sampling from smaller catchments potentially overcompensated the losses from decay processes in larger catchments.

3.4. Regression model components

3.4.1. Lag-component

The lag component represents the temporal shift between the seven-day incidence rate and the seven-day mean E gene concentration. Results of the cross-correlation evaluation are concluded in Table 2, where positive values indicate lead of the wastewater data. Predominantly, sites showed a lead of wastewater data. Shorter time shifts, with one day lag to six days lead, occurred during the third epidemic wave. Three sites (Che, Els and Mor) exhibited no lead or even a one-day lag of incidence rates. The fourth wave caused two to twelve days of lead time. Generally, the difference in shifted versus unshifted correlation coefficients of sites with longer lead times were significantly higher than sites with shorter lead time. There was no apparent site-specific pattern in time shift. Intermediately sized catchments Ann and Pla ranked among the longest leads during both epidemic waves. In contrast, small sites Els and Mor had the shortest lead times. As discussed in Section 3.3, lower sampling frequency and accuracy in small systems as well as the higher short-term variability of prevalence rates might increase the probability of misrepresented event courses. Despite the prevailing lead of wastewater data, the pronounced differences among sites and between both waves support the assumption that most of the lead time results from lags in detection and reporting of diagnosed cases. As such, we confirmed the range of previously reported lead times between two to four days (Hillary et al., 2021) and up to ten to fourteen days (Agrawal et al., 2021). In a review of worldwide studies, (Kumar et al., 2022) established a range of five to fourteen days lead time.

Xiao et al. (2022) identified longer lead times of wastewater data during the earlier epidemic wave and argued that improved testing capabilities in later stages compensated this lead. In contrast, our results suggest a shorter lead of wastewater data during the earlier wave. This potentially hints at differing responses to the epidemic events in different countries or due to different epidemic progressions. In the period between onset and peak of the third epidemic wave, incidence numbers increased by 25 cases per 100,000 inhabitants and week in Saxony. Across all sites, the increase ranged from 17 to 184 cases per 100,000 inhabitants and week. The onset to peak increase of the fourth epidemic wave developed much more pronounced, with 165 cases per 100,000 inhabitants and week in Saxony and 143 to 317 cases per 100,000 inhabitants and week among the sites. Hence, longer lead times of wastewater data were associated with steeper increases of diagnosed incidences.

3.4.2. Discharge interaction

Two different submodels were tested in order to assess the impact of discharge on prevalence estimation. In submodel 2–2, the inhabitant-specific load (cQ/Pop) served as independent variable, while in submodel 3–1 concentration and inhabitant-specific discharge represented separate explanatory variables. Regression parameters, parameter significance and model performance are concluded in Table S4. The regression performance and parameter significance for the global and local models were mostly similar to the concentration-based models (compare Section 3.3). This is consistent with Nagarkar et al. (2021), while most other studies used either load or concentration-based models.

When discharge serves as a separate explanatory variable, its impact is not significant in most of the models. Only site Che, during the fourth epidemic wave, had a highly significant contribution of the discharge parameter. In most cases, the discharge parameter obtained positive values smaller than one, representing higher predicted prevalence rates at higher discharges and the same E gene concentrations, but a less than linear contribution. Despite the low significance of the discharge parameter, this indicated that dilution dynamics did not completely determine in the overall relation between E gene concentration and incidence.

The inclusion of discharge as an additional explanatory variable did not significantly improve model performance, assessed through change in aR². Models of site Pla deviate from this statement, with moderately (during the third wave) to highly significant (during the fourth wave) improvements in aR² for submodel 3–1. The subordinate influence of discharge on model performance is potentially explained by the lower variance of inhabitant-specific discharge when compared to gene concentration. While the coefficients of variation of the E gene concentration were 100 % and 110 % during the third and fourth epidemic wave, discharge variation coefficients were only 23 % and 42 % during the same periods.

3.4.3. Detection probability

Detection probability was estimated by logistic regression between seven-day diagnosed incidence and success in detection of the E gene. The E gene concentration was 0–1 transformed and the logistic function
(Eq. (4)) was fitted by generalized least-squared estimation. Results are concluded in Table S6. Across all sites, the third epidemic wave was characterized by a higher proportion of values below LoD. The proportion ranged from 18 to 62% during the third wave and from two to 35% during the fourth wave. The smallest sites Els and Mor had the highest below LoD proportion among sites. As discussed in previous sections, higher dynamics of the epidemic events and systematic underrepresentation in sampling of a low number of excreting individuals might be reasons for these results.

Sites with a higher proportion of values below LoD mostly obtained higher significance levels of the logistic regression parameters and a higher goodness-of-fit, as measured by Mc Fadden’s pseudo coefficient of determination. Larger catchments obtained 50% and 95% detection probability levels at lower diagnosed incidence rates than smaller catchments. For the third epidemic wave, sites Che, Dre, and Pla incidence rates around 60 diagnosed cases per 100,000 inhabitants corresponded to a 50% detection probability in wastewater. In contrast, the same sites had over 50% detection probability even at zero diagnosed incidence rates. This finding supports the assumption that the fourth epidemic wave witnessed a larger proportion of undiagnosed cases. Sites Els and Mor had significantly higher 50% detection probabilities, beyond 100 diagnosed cases per 100,000 inhabitants during the third epidemic wave. In contrast, during the fourth wave, 50% detection probabilities reduced to 85 and 11 diagnosed cases per 100,000 inhabitants. Additionally, the steepness parameter of the logistic regression at the small and intermediate sites tended to be lower, resulting in even more pronounced differences in detection at the 95% probability level.

3.4.4. Combined model components: performance versus complexity

The significance and performance of regression models among the sites and epidemic waves was systematically assessed through models with increasing numbers of parameter. In line with principles of information theory, we assessed model complexity with the number of parameters in the model. Two estimators of prediction error with inherent penalty components for the number of parameters served as indicators for model quality or performance. In total, nine different model versions with two to five parameters were derived and 108 models were optimized and compared. The following list shall support overview and tracing of the results:

Model 2–0 relates non-transformed seven-day diagnosed incidence to the seven-day moving average of E gene concentration. Regression parameters and performance indicators are concluded in Table S4.

Model 2–1 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration. Regression parameters and performance indicators are concluded in Table S4.

Model 2–2 relates seven-day diagnosed incidence to the seven-day moving average of E gene daily load per 100,000 inhabitants. Regression parameters and performance indicators are concluded in Table S4.

Model 3–1 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration and discharge per inhabitant. Regression parameters and performance indicators are concluded in Table S4.

Model 3–2 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration and lag/lead time. Regression parameters and performance indicators are concluded in Table S7.

Model 3–3 relates seven-day diagnosed incidence to the seven-day moving average of E gene daily load per 100,000 inhabitants and lag/lead time. Regression parameters and performance indicators are concluded in Table S7.

Model 4–1 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration and detection probability. Regression parameters and performance indicators are concluded in Table S7.

Model 4–2 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration, discharge per inhabitant and lag/lead time. Regression parameters and performance indicators are concluded in Table S7.

Model 5 relates seven-day diagnosed incidence to the seven-day moving average of E gene concentration, detection probability and lag/lead time. Regression parameters and performance indicators are concluded in Table S7.

The comparison of models 2–0 and 2–1 validated the assumptions stated in Section 2.5. Results are concluded in Table S4. In twelve out of 14 models (including the global models for all sites) the log-log transformed models obtained a better goodness-of-fit, and in nine cases narrower 95% prediction intervals. Due to the differing probability densities of the response variables, $r^2$ cannot be statistically compared but, in most cases, improvements in $r^2$ were higher than 0.2. In five cases the lower bound of the non-transformed prediction interval stretched into the negative number range. Furthermore, the offset parameter $m$ of model 2–0 requires justification. We found ranges from 48 to 584. If this parameter is interpreted as the incidence rate at zero E gene concentration, it would suggest a low detection sensitivity. On the other hand, if $m$ is fixed at zero, regression performance further deteriorates. Similarly, Aberi et al. (2021) stated offset parameters (denoted as Const. in their work) with a range from $-279$ to $573$ for four different sites, which are difficult to interpret deterministically. Due to the conceptual difficulties and limited comparability with transformed data models, we excluded model 2–0 from the following evaluation.

The panels of Fig. 4 visualize the model performance in relation to the number of parameters included in the regression models. The model quality indicators $r^2$ (upper) and AIC (lower) of models 2–1 to 5 were grouped according to the number of parameters. In addition, the local polynomial regression curves, depicted as solid lines, highlight the relation between performance indicator and number of parameters. A general improvement in $r^2$ with an increasing number of parameters occurred in both epidemic waves. Improvements were more pronounced during the fourth wave and in models that included lag/lead time (3–2, 3–3, 4–2, 5). This is consistent as $r^2$ inefar E gene concentration had a longer temporal lead for most sites during this period. In contrast, models 3–1 and 4–1 without lag/lead time often resulted in minor model improvements or even a setback of $r^2$ due to a higher adjustment term without proportional improvement in the variance explained. Among the sites with short or no lag/lead time (Mor, Dre, Che and Els during the third wave) time shift contributed less to model improvement. The different submodels showed minor shifts in the optimal lag/lead time. As already discussed in Section 3.4.2, models with explanatory inhabitant specific E gene load (models 2–2, 3–3) showed minor differences in $r^2$ when compared to E gene concentration (models 2–1, 3–2) as input. Out of the 24 submodels, load based regression performed better in 13 cases. In four cases the improvement was highly significant.

In some cases, models with discharge as an independent explanatory variable (models 3–1, 4–2) improved performance. The comparison of models 2–1 and 3–1 was discussed in Section 3.4.3. Discharge parameter $o$ was significant at sites Ann and Pla (both waves), Dre (third wave), Che (fourth wave) in model 4–2. In 14 out of the 24 comparisons, $o$ improved $r^2$. In six cases this improvement was highly significant. Models with detection probability parameters (4–1 and 5) consistently improved $ar^2$ when compared to the corresponding models without this component (2–1 and 3–2, respectively). The improvement was highly significant in 15 out of 24 comparisons. Overall, in three cases each, model 3–3 and 4–2 achieved highest $r^2$, while in six cases model 5 performed best. In tendency, large catchments (Che, Dre, Pla) and catchments with a higher proportion of separate sewer systems (Che, Mor) demonstrated better model performance. This is conceptually plausible because of a lower flow variation and fewer potential spillages from the separate system. Furthermore, as discussed in Section 3.3, a higher recovery of viral particles seems likely due to higher total number of sheds in large catchments and more representative sampling.

Model performance with regard to AIC was more ambiguous. Differences in AIC were less pronounced for models of the third epidemic
wave. During the fourth wave, models with lag/lead term strongly reduced AIC. Apparently, the detection probability $a$ and $b$ parameters decreased the log-likelihood term for some submodels (site Dre during the third wave, sites Che, Pla during both waves). Therefore, model 5 achieved lowest AIC in four cases but was outperformed by 4–2 and 2–2 in three cases each, as well as 3–2 in two cases. When comparing both model performance and selection indicators, AIC highlighted differences among the models more and appeared more rigorous in penalizing models with a higher number of parameters.

Fig. 5 exemplifies the course of the regression relations for site Ann. For the third wave, disproportionally low concentrations of E gene at incidences below 100 cases per 100,000 inhabitants, lead to a systematic overestimation of the non-transformed linear model. The three models with time lag components (3–2, 4–2 and 5) represent higher incidences at high concentrations; this is in line with the higher power parameters $n$, for these models (compare Table S7). Generally, all models have power parameters significantly smaller than one (0.32–0.57) resulting in decreasing slopes with increasing concentrations. The jitter in model 4–2 is caused by the influence of covariable $q$ that is not projected into the plotting plane. For the fourth wave, the difference in the model expression between shifted and non-shifted regression models is more apparent. This is partly explained by the shorter lead time (6–7 days during the third waves vs. 8–10 days during the fourth wave), but also due to the more volatile epidemic course (compare RMSE in Table 1). Consequently, power parameters of the non-shifted models were systematically lower (0.46–0.65) than for models with lead time (0.74–1.12). Overall, the log-log transformed model provided a robust framework for the functional relations of the different epidemic waves. This translated to the temporal representation of the regression models in the lower panels of Fig. 5. Models 2–1, 3–2 and 4–2 were selected for representation because they provided a good range of performance at site Ann in both waves. The improved match in temporal dynamics and magnitude of models 3–2 and 4–2, as well as narrower prediction ranges, are apparent.

3.4.5. Back-estimation of testing efficiency

The regression models optimize the least squared difference between observed and predicted data. Yet, in the case of diagnosed incidence, the observed value represents at best the real prevalence, when all infected cases are diagnosed. Variable infection duration, incomplete testing selectivity and efficiency, as well as an unknown proportion of asymptomatic or paucisymptomatic but infected individuals lead to a likely systematic underestimation of prevalence through diagnostic testing. The time series of the fourth epidemic wave at sites Ann (Fig. 5) and Dre (Fig. 3) exemplarily cover early onsets and preliminary peaks in E gene concentration. Other studies have introduced total number of tests as an additional regression variable (Aberi et al., 2021) or epidemiological modelling (Fernandez-Cassi et al., 2021) to estimate real prevalence. Necessary data for these approaches is not published for Germany at the municipality level. In order to counter this underestimation, we introduced an additional scaling factor (or extended the existing slope parameter $m$), so that the diagnosed cases are mostly surpassed by the wastewater estimate. The rationale behind this approach assumes that diagnosed cases are at all times lower than real prevalence. We then evaluated the ratio of diagnosed incidence to estimated prevalence in order to approximate testing efficiency.
Fig. 6 demonstrates the application of the scaled slope parameter for sites Che and Mor during the fourth epidemic wave. Similar to sites Ann and Dre, preliminary onsets and peaks appear through the regression models. Furthermore, the scaling revealed an underestimation of diagnosed peak incidence rates. In order to evaluate estimated testing efficiencies systematically, both epidemic waves were subsetted in onset, peak and recession phases. For sites Che and Mor these corresponded to: onset 15.10.2021 to 15.11.2021, peak 15.11.2021 to 15.12.2021, recession 15.12.2021 to 31.12.2021. For sites Els, Mor and Pla, the onset phase of the third epidemic wave was not covered due to the later start of sampling.

The scaling factors and testing efficiencies are concluded in Table 3. Overall, testing efficiencies were comparable in both epidemic waves. This confirmed the estimation based on state wide testing statistics (compare Section 3.2). Xiao et al. (2022) identified significant differences in testing efficacy during different epidemic waves; their study covered an earlier phase of the epidemic and might capture progress in epidemic management.

Fig. 5. Fitted models and prediction uncertainty of regression models for site Ann during the third (left) and fourth (right) epidemic wave. Upper: scatter plot of E gene concentration (cE gene) and diagnosed seven-day incidence rate (7d), with unshifted (grey dots) and optimally shifted (black dots) concentrations. Lower: time series of seven-day incidence (red line), seven-day moving average of E gene concentration (black line), fitted regression models (colored lines) and 95 %-prediction interval (colour shaded ribbons) of seven-day incidence rate according to regression models 2–1, 3–2 and 4–2.

Fig. 6. Time series of the seven-day incidence (red line), seven-day moving average of E gene concentration (black line), fitted and scaled regression model 3–2 (purple line) for sites Che (left) and Mor (right) during the fourth epidemic wave.
Nevertheless, there are notable differences in the course of testing efficiency. During the third epidemic wave, testing seemed more efficient in the onset and peak periods. In contrast, the fourth wave witnessed lowest efficiencies during the onset. As argued in Section 3.4.1, increase of incidence rates of the epidemic fourth wave was steeper. Both aspects might extend the reasoning behind the longer lead times of this wave. Early cases were less efficiently identified through testing, yet appeared in the wastewater signal. Furthermore, in tendency, testing efficiencies were higher in large catchments (Che, Dre, Pla) than in smaller catchments.

4. Conclusions

With this paper, we presented a systematic evaluation of candidate regression models for the relation between SARS-CoV-2 biomarkers in wastewater and infection prevalence, for six sites in Germany. The evaluation demonstrated, that the relation is specific to site and epidemic wave. Thus, a general model results in lower model performance and higher uncertainty.

The log-log transformed regression proved more robust than untransformed linear regression between biomarker concentration and diagnosed incidence. Although conceptually more consistent, load-based models did not perform better than concentration-based models. Thus, model reliability is not greatly affected if discharge information is unavailable at the sampling point. Modelling components for discharge interaction, detection probability and lag/lead time can contribute significantly to model improvement. Again, their importance and parameter expressions are site and epidemic wave specific. Although these insights may be discouraging for a straightforward use of the regression models, a satisfactory model performance, with an adjusted coefficient of determination higher than 0.5, was in all cases achieved with a three parameter model including gene concentration and lag/lead time.

Differences among the epidemic waves are potentially caused by the predomination of distinct SARS-CoV-2 variants. In tendency, sites with a higher number of connected inhabitants and a higher proportion of separate sewer systems enabled higher recovery of viral particle load per infected inhabitant and better performance of regression models. Both aspects hint at sampling and data evaluation challenges for smaller systems with a higher flow and shedding variability. Additionally, uncertainties about the connected population are not covered in this study and should be addressed in future works.

We estimated prevalence through a scaling of the fitted regression models and derived testing efficiency as ratio between incidence rate and prevalence. Results showed that the overall efficiency was similar in both epidemic waves but varied during the onset, peak and recession phase. The fourth epidemic wave had low testing efficiency in the onset phase and correspondingly long lead-time of the wastewater data. This finding supports the potential of wastewater-based epidemiology for the early warning of epidemic outbreaks, especially if clinical diagnostic testing is not widely performed.

Table 3

| Site     | Ann | Che | Dre | Els | Mor | Pla |
|----------|-----|-----|-----|-----|-----|-----|
| Third wave |     |     |     |     |     |     |
| Scaling factor | 1.60 | 1.50 | 1.40 | 1.80 | 1.65 | 1.60 |
| Onset t.e. [%] | 64 | 87 | 88 | – | – | – |
| Peak t.e. [%] | 69 | 79 | 84 | 58 | 65 | 83 |
| Recession t.e. [%] | 78 | 68 | 65 | 63 | 63 | 77 |
| Mean t.e [%] | 71 | 78 | 82 | 60 | 64 | 80 |
| Fourth wave |     |     |     |     |     |     |
| Scaling factor | 1.95 | 1.45 | 1.55 | 2.40 | 1.60 | 1.70 |
| Onset t.e. [%] | 28 | 63 | 62 | 44 | 65 | 58 |
| Peak t.e. [%] | 63 | 80 | 81 | 58 | 73 | 76 |
| Recession t.e. [%] | 85 | 89 | 92 | 88 | 93 | 88 |
| Mean t.e [%] | 59 | 77 | 78 | 63 | 77 | 74 |

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no competing interests that could influence this study.

Ethics application (BO-EK-383072021) has been confirmed by the Ethic Committee of Technische Universität Dresden, which is registered as institutional review board (IRB0001473) at the Office of Human Research Protection.

Acknowledgements

We appreciate the support and cooperation of wastewater system operators Zweckverband Wasser Abwasser Vogtland, Abwasserzweckverband Oberes Zschopau- und Sehmatal, eins energie in Sachsen and Stadtentwässerung Dresden GmbH. Robert Koch Institute and Saxon State Ministry for Social Affairs and Social Cohesion provided open access epidemiological data. Further, we are thankful for special technical support of N. Engl (Clinical Pharmacology) and A. Skupin (Medical Microbiology and Virology) and language revision by S. Healy.

Project “Wastewater-CoV-2-Tracking” (Abwasser-CoV-2-Tracking: Abwasserspektion zur Bestimmung des SARS-CoV-2-Infektionsgrads der Bevölkerung: Aufbau eines flächendeckenden Warn- und Trackingssystems für Pandemien) was thankfully funded by SMWK - State Ministry of Science and Cultural Affairs of Saxony via Sächsische Aufbaubank (developmental bank Saxony; FKZ: 100535976).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.159358.

References

Aberi, P., Arabzadeh, R., Insam, H., Markt, R., Mayr, M., Kreuzinger, N., Rauch, W., 2021. Quest for optimal regression models in SARS-CoV-2 wastewater based epidemiology. Int. J. Environ. Res. Public Health 18, 10778. https://doi.org/10.3390/ijerph182010778.

Agrawal, S., Onschler, L., Lackner, S., 2021. Long-term monitoring of SARS-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in southern Germany. Sci. Rep. 11, 5372. https://doi.org/10.1038/s41598-021-84914-2.

Ahmed, W., Bertrich, P.M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., Gyawali, P., Korajkic, A., Riddell, S., Sherchan, S.P., Simpson, S.L., Sirkankancha, K., Symonds, E.M., Verhagen, R., Vasan, S.S., Kitajima, M., Bivins, A., 2020. Decay of SARS-CoV-2 and surrogated murren hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. Environ. Res. 191, 110902. https://doi.org/10.1016/j.envres.2020.110902.

CRediT authorship contribution statement

B. Helm: Main Funding acquisition & Conceptualization, Data Regression and Modelling, Writing - Original Draft, Visualization, Supervision, M. Geissler: Methodology, Validation & Analysis of qPCR, microbiological Data Curation, R. Mayer: hydrological and wastewater related Data Curation, sampling and organizing of wastewater and wastewater related data, S. Schubert: Ethics Application, Methodology & Analysis of pharmaceuticals, Writing - Original Draft, Visualization of pharmaceutical data abstract, R. Oertel: Methodology, Validation & Analysis of pharmaceuticals, Writing - Original Draft, Visualization of pharmaceutical data, Supervision, R. Dumke: Methodology & Analysis of qPCR, Methodology, Validation of qPCR, microbiological Data Curation, Writing - Original Draft, Visualization of microbiological data, Supervision, A. Dalpke: Funding acquisition & Conceptualization, Project administration, Ali El-Armouche: Funding acquisition & Conceptualization, Project administration, Bertold Renner: Ethics Application, Supervision, Peter Krebs: Funding acquisition & Conceptualization, Project administration.

Data availability

Data will be made available on request.

Acknowledgements

We appreciate the support and cooperation of wastewater system operators Zweckverband Wasser Abwasser Vogtland, Abwasserzweckverband Oberes Zschopau- und Sehmatal, eins energie in Sachsen and Stadtentwässerung Dresden GmbH. Robert Koch Institute and Saxon State Ministry for Social Affairs and Social Cohesion provided open access epidemiological data. Further, we are thankful for special technical support of N. Engl (Clinical Pharmacology) and A. Skupin (Medical Microbiology and Virology) and language revision by S. Healy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.159358.

References

Aberi, P., Arabzadeh, R., Insam, H., Markt, R., Mayr, M., Kreuzinger, N., Rauch, W., 2021. Quest for optimal regression models in SARS-CoV-2 wastewater based epidemiology. Int. J. Environ. Res. Public Health 18, 10778. https://doi.org/10.3390/ijerph182010778.

Agrawal, S., Onschler, L., Lackner, S., 2021. Long-term monitoring of SARS-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in southern Germany. Sci. Rep. 11, 5372. https://doi.org/10.1038/s41598-021-84914-2.

Ahmed, W., Bertrich, P.M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., Gyawali, P., Korajkic, A., Riddell, S., Sherchan, S.P., Simpson, S.L., Sirkankancha, K., Symonds, E.M., Verhagen, R., Vasan, S.S., Kitajima, M., Bivins, A., 2020. Decay of SARS-CoV-2 and surrogated marine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. Environ. Res. 191, 110902. https://doi.org/10.1016/j.envres.2020.110902.
Wang, Y., Liu, P., Zhang, H., Ibaraki, M., VanTassell, J., Geith, K., Cavallo, M., Kann, R., Saber, M.J., Lo Jacomo, A., Armenise, E., Brown, M.R., Bunce, J.T., Cameron, G.J., Fang, Z., Thomas, K.V., Amador, A., Baz-Lomba, J.A., Reid, M., 2017. Use of Mobile device data to better estimate dynamic population size for wastewater-based epidemiology. Environ. Sci. Technol. 51, 11363–11370. https://doi.org/10.1021/acs.est.7b02538.

Thompson, S.E., Basu, N.B., Lascurain, J., Aubeneau, A., Rao, P.S.C., 2011. Relative dominance of hydrologic versus biogeochemical factors on solute export across impact gradients. Water Resour. Res. 47. https://doi.org/10.1029/2010WR009605.

Vallejo, J.A., Trigo-Tasende, N., Rambo-Feliu, S., Conde-Pérez, K., López-Orióna, Á., Barbeito, I., Vaamonde, M., Tarrio-Saavedra, J., Reif, R., Ladrà, S., Rodiño-Janeiro, B.K., Naser-Ali, M., Cid, Á., Veiga, M., Acevedo, A., Lamora, C., Bou, G., Cao, R., Pouza, M., 2022. Modeling the number of people infected with SARS-COV-2 from wastewater viral load in North-west Spain. Sci. Total Environ. 811, 152334. https://doi.org/10.1016/j.scitotenv.2021.150060.

|B. Helm et al. Science of the Total Environment 857 (2023) 159358|

Warner, W., Licha, T., Nöddler, K., 2019. Qualitative and quantitative use of micropollutants as source and process indicators. A review. Sci. Total Environ. 686, 75–89. https://doi.org/10.1016/j.scitotenv.2019.05.385.

Weidhaas, J., Aanderud, Z.T., Roper, D.K., VanDerslice, J., Gaddis, E.B., Ostermiller, J., Hoffman, K., Jamal, R., Heck, P., Zhang, Y., Torgersen, K., Laan, J.V., LaCross, N., 2021. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. Sci. Total Environ. 775, 145790. https://doi.org/10.1016/j.scitotenv.2021.145790.

Westhaus, S., Weber, F.-A., Schivvy, S., Linnemann, V., Brinkmann, M., Widera, M., Greve, C., Janke, A., Hölker, H., Wüstges, 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. Sci. Total Environ. 751, 141750. https://doi.org/10.1016/j.scitotenv.2020.141750.

Wölfl, R., Corman, V.M., Gugmemos, W., Seilmäier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., Hölscher, M., Blecker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C., Wendtner, C., 2020. Virological assessment of hospitalized patients with COVID-19. Nature 581 (7809), 465–469.

Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M.A., Bushman, M., Chai, P.R., Duvalliet, C., Erickson, T.B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W.P., Huang, K.H., Lee, W.L., Matus, M., McElroy, K.A., Nagler, J., Rhode, S.F., Santillana, M., Tucker, J.A., Wurtzer, S., Xiao, A., Zhang, J., Moniz, K., Rhode, S.F., Thompson, J., Alam, E.J., 2022. SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. Sci. Total Environ. 805, 150121. https://doi.org/10.1016/j.scitotenv.2021.150121.

Wurzler, S., Waldman, P., Levert, M., Almayrac, J.L., Charpentier, C., Mannsda, S., Gillon-Ritz, M., Mouchel, J.M., Maday, Y., Boni, M., OBEPINE Consortium, AP-HP Virol-ogist Group, Marechal, V., Moxlin, L., 2022. SARS-CoV-2 genome quantification in wastewater at regional and city scale allows precise monitoring of the whole outbreaks dynamics and variants spreading in the population. Sci. Total Environ. 810, 152213. https://doi.org/10.1016/j.scitotenv.2021.152213.

Xiao, A., Wu, F., Bushman, M., Zhang, J., Imakaev, M., Chai, P.R., Duvalliet, C., Endo, N., Erickson, T.B., Armas, F., Arnold, B., Chen, H., Chandra, F., Ghaeli, N., Gu, X., Hanage, W.F., Lee, W.L., Matus, M., McElroy, K.A., Moniz, K., Rhode, S.F., Thompson, J., Alam, E.J., 2022. Metrics to relate COVID-19 wastewater data to clinical testing dynamics. Water Res. 212, 118070. https://doi.org/10.1016/j.watres.2022.118070.