Prediction of hospitalisations based on wastewater-based SARS-CoV-2 epidemiology

Hannes Schenk [a], Petra Heidinger [d], Heribert Insam [b], Norbert Kreuzinger [c], Rudolf Markt [b,h], Fabiana Nägele [b], Herbert Oberacher [g], Christoph Scheffknecht [e], Martin Steinlechner [g], Gunther Vogl [f], Andreas Otto Wagner [b], Wolfgang Rauch [a,*]

[a] Unit of Environmental Engineering, University of Innsbruck, Technikerstraße 13, Innsbruck 6020, Austria
[b] Department of Microbiology, University of Innsbruck, Technikerstraße 25d, Innsbruck 6020, Austria
[c] Institute of Water Quality and Resource Management at TU Wien, Karlsplatz 13, Vienna 1040, Austria
[d] Austrian Centre of Industrial Biotechnology, Krenngasse 37, Graz 8010, Austria
[e] Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4, Bregenz 6900, Austria
[f] Institut für Lebensmittelsicherheit, Veterinärmedizin und Umwelt, Kirchengasse 43, Klagenfurt 9020, Austria
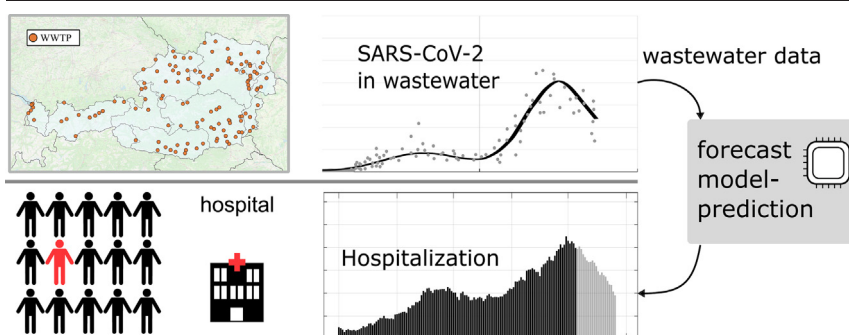[g] Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Müllerstraße, 44, Innsbruck 6020, Austria
[h] Department of Health Sciences and Social Work, Carinthia University of Applied Sciences, St. Veiter Straße, 47, Klagenfurt 9020, Austria

HIGHLIGHTS

- Wastewater-based epidemiological monitoring of 123 treatment plants in Austria
- Hospitalisation forecast is feasible by exploiting wastewater signal's time lead.
- Usage of higher numbers of wastewater data results in improved forecasts.
- Forecast models are sensitive to variant types and require recalibration.

GRAPHICAL ABSTRACT

ARTICLE INFO

ABSTRACT

Wastewater-based epidemiology is widely applied in Austria since April 2020 to monitor the SARS-CoV-2 pandemic. With a steadily increasing number of monitored wastewater facilities, 123 plants covering roughly 70 % of the 9 million population were monitored as of August 2022. In this study, the SARS-CoV-2 viral concentrations in raw sewage were analysed to infer short-term hospitalisation occupancy. The temporal lead of wastewater-based epidemiological time series over hospitalisation occupancy levels facilitates the construction of forecast models. Data pre-processing techniques are presented, including the approach of comparing multiple decentralised wastewater signals with aggregated and centralised clinical data. Time-lead quantification was performed using cross-correlation analysis and coefficient of determination optimisation approaches. Multivariate regression models were successfully applied to infer hospitalisation bed occupancy. The results show a predictive potential of viral loads in sewage towards Covid-19

hospitalisation occupancy, with an average lead time towards ICU and non-ICU bed occupancy between 14.8-17.7 days and 8.6–11.6 days, respectively. The presented procedure provides access to the trend and tipping point behaviour of pandemic dynamics and allows the prediction of short-term demand for public health services. The results showed an increase in forecast accuracy with an increase in the number of monitored wastewater treatment plants. Trained models are sensitive to changing variant types and require recalibration of model parameters, likely caused by immunity by vaccination and/or infection. The utilised approach displays a practical and rapidly implementable application of wastewater-based epidemiology to infer hospitalisation occupancy.

1. Introduction

The worldwide pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late 2019 in Wuhan, China (Zhu et al., 2021; Wölfel et al., 2020). The disease is strongly contagious and often causes fever, respiratory difficulties and may lead to long lasting effects (Tene et al., 2022). Coronavirus disease (Covid-19) is primarily transmitted through airborne droplets in proximity to infected individuals (Mostaghimi et al., 2022). According to the World Health Organization, in February 2023, over 662 million confirmed cases and over 6 million deaths have been observed (WHO, WHO Coronavirus (COVID-19) Dashboard, Online).

The prevention of a steady increase in Covid-19 caused hospitalizations and an expanding demand for public health infrastructure required the introduction of SARS-CoV-2 control measures comprising non-pharmaceutical interventions such as mobility restrictions and public health and social interventions. A globally widespread decline in adherence to non-pharmaceutical interventions has been observed as a consequence of pandemic fatigue (Petherick et al., 2021). Pandemic mitigation strategies based on effective vaccines allow progressive reduction of interventions (Oliu-Barton et al., 2021).

Strategies to monitor the ongoing health crisis rely on epidemiological surveillance, with the aim of revealing cases and clusters before the further escalation of infections overwhelms the public health care system. The main pillar worldwide for tracking cases and monitoring pandemic development is based on clinical testing. However, clinical testing suffers from limitations such as reporting bias, testing fatigue, and difficulty in detecting asymptomatic virus carriers (Thompson et al., 2020; Daughton, 2020). As a second pillar, wastewater-based surveillance provides a non-invasive surveillance tool for tracking pandemic development. Some of the first major contributions in the field were made by Medema et al. (2020), Ahmed et al. (2020) (among others (La Rosa et al., 2020; Lodder and de Roda Husman, 2020)), providing a proof-of-concept by using established methods of wastewater-based epidemiology (WBE) in the context of SARS-CoV-2 detection. Since then, the research community has steadily gained interest in new applications of WBE with international contributors (Lundy et al., 2021).

The detection of coronavirus in wastewater is due to the presence of viruses in feces and sputum of infected individuals (Cevik et al., 2021; Wyllie et al., 2020; Kashi et al., 2020). As the viral load is introduced into the sewer, the ribonucleic acid of SARS-CoV-2 is transported in conduit pipes and subsequently detected in the inlet sewage of wastewater treatment plants (WWTP) (Crank et al., 2022). The shedding dynamics of infected individuals varies widely. A meta-analysis by Cevik et al. (2021), comprising 79 studies observing SARS-CoV-2 infections, inferred a mean duration of 17 days (± 0.8, CI 95 %) of viral respiratory and feces shedding. The viral load is shed before symptom onset (He et al., 2020). However, the literature does not provide consent regarding the peak timing of viral load shedding, ranging from days before symptom onset to weeks after (Cevik et al., 2021).

WBE provides benefits as a surveillance tool. Firstly, WBE provides a cost-effective SARS-CoV-2 surveillance framework in a population. The measurement approach is analogous to polymeric chain reaction analysis in clinical swab tests but is applied to a community-wide wastewater composite sample. Secondly, the potential of WBE is its ability to confirm viral infections, even if no evidence of infection is obtained by clinical testing (Kitajima et al., 2020). Furthermore, wastewater surveillance provides an estimation of the degree of viral circulation independent of clinical diagnosis. Newly emerging variants of concern and their dominance over existing variants can be detected and quantified (Wurtzer et al., 2021; Amman et al., 2022). The limiting factors are represented by lack of standardized procedures for viral quantification and for trends representation (Li et al., 2021; Hart and Halden, 2020). Despite these limitations, WBE is successfully applied as a complementary surveillance strategy.

One of the main driving forces of WBE is the potential to detect case prevalence before clinical testing. This temporal lead in our reference system may be caused by the occurrence of a viral load in excrement before symptom onset in infected individuals (Peccia et al., 2020). A second factor that explains the temporal lead is the weakness of individual testing in local communities (Greenwald et al., 2021). A well-established and rapid infrastructure for WBE quantification and a result-reporting framework is critical for exploiting the time lead over clinical data (Bibby et al., 2021). The average time frame of data accessibility for this study was between three and four days after measurement. According to Aberi et al. (2021), a time lead of 2–7 days of viral load in wastewater over the incidence signals was observed, depending on the time period and site. Vaughan et al. (2023) identified challenges with machine learning approaches for time-series forecasting by using wastewater-based epidemiological data. The factors influencing forecast quality are complex and interrelated. Limiting factors for WBE forecasts models are low sampling frequency and low (or zero) values for the target concentrations (Vaughan et al., 2023).

This study aimed to present data processing techniques for WBE SARS-CoV-2 measurements to infer short-term hospitalisation occupancy. As proposed by Galani et al. (2022), clinical cases hospitalised can be predicted using SARS-CoV-2 wastewater surveillance data, using regression analysis. Similar approaches are employed herein, where the novelty lies in the application of predicting hospitalisation occupancy at a federal state level by combining multiple time series from different WWTPs. This includes data preparation for the mismatch between a multitude of viral load time series and aggregated hospitalisation time series within each region of Austria. In this work, multivariate regression models were applied together with regularisation methods, outlining the qualitative resemblance of the wastewater signal with hospitalised cases. Regularized regression penalizes model coefficients to prevent overfitting. This allows to use multiple regressor time-series and leads to better prediction results with unseen validation data. In addition to the lead-time quantification analysis and the construction of the forecast models, the forecast accuracy as a function of the abundance of wastewater data was investigated.

2. Materials and methods

In this section, the quantity and type of data used in the investigation are presented, as well as the methods and analysis procedures applied. Data pre-processing techniques are adopted from Rauch et al. (2022) and explained in the following subsections. As this study builds on a nationwide Austrian WBE endeavour, approximately 70 % of Austria's population is monitored. The abundance of epidemiological data from wastewater expresses the quantity of viral genetic material circulating at local catchment sites. The investigation of this work starts with a data evaluation of the viral concentrations obtained from the laboratories. Therefore, the scientific focus is not on the methodical detection procedures and sampling details but on subsequent data treatment for hospitalisation forecasting purposes.

The underlying sampling and quantification procedures are outlined in Markt et al. (2022, 2021) and Daleiden et al. (2022).

2.1. Austrian WBE surveillance data

In Austria, wastewater surveillance in the context of the Covid-19 pandemic started in April 2020 and has been extended to over 120 WWTPs, covering approximately 70 % of the population. Fig. 1 shows the map of WBE monitoring locations in Austria, as well as a histogram outlining the size of the WWTPs by means of their design capacity. The plant in Vienna has the highest design capacity in Austria, able to serve 4 million people, the median design capacity of plants used in this work is 44,000 people equivalent. The monitored plants have been selected based on their importance for national surveillance by a governmental panel.

Composite samples are collected with a measurement frequency of 2 to 3 times per week. Samples were concentrated and purified using polyethylene glycol-precipitation or a direct capture method (Promega, Madison, WI, USA). Reverse transcription real-time polymerase chain reaction assays were used to determine viral concentrations in the processed wastewater samples (Markt et al., 2022; Daleiden et al., 2022). Furthermore, complementary pollutant concentrations from wastewater surveillance are used as population size markers (chemical oxygen demand (COD), ammonium NH4-N, and total nitrogen N) (Arabzadeh et al., 2021). The influent volume flow rate Q is required for outlier detection. Wastewater facilities in all nine federal states of Austria are represented, with Vienna being an exception because only one large-scale WWTP serves a population of 1.9 million inhabitants.

The abundance of data processed in this study is the result of several monitoring programs with contributions and funding from different entities. In this work a total of 15,656 data points are utilised, gathered in the period between April 2020 and August 2022. The following list provides an overview of the participating programs and the corresponding key data:

- Covid-Nachweis im Abwasser, Universität Innsbruck Förderkreis 1669 (8 WWTPs, 27.07.2020 - 01.11.2020, 130 data points)
- Coron-A project of the Austrian Federal Ministry of Agriculture, Forestry, Regions and Water Management and Austrian Federal Ministry of Education, Science and Research (51 WWTPs, 23.04.2020 - 30.08.2021, 1528 data points)
- Monitoring Kärnten (11 WWTPs, 01.06.2021 - 31.10.2021, 439 data points)
- Monitoring Vorarlberg (4 WWTPs, 01.03.2021 - 29.08.2021, 627 data points)
- SARS-CoV-2 Wastewater Monitoring Program of the Austrian Federal Ministry of Education, Science and Research (108 WWTPs, 09.01.2021 - 31.08.2022, 10,563 data points)

- National SARS-CoV-2 Wastewater Monitoring Program of the Austrian federal Ministry of Social Affairs, Health, Care and Consumer Protection (24 WWTPs, 16.01.2022 - 31.08.2022, 1459 data points)
- Tyrolean SARS-CoV-2 Wastewater Monitoring Program (9 WWTPs, 27.09.2021 - 11.01.2022, 755 data points)
- TU Wien WBE (31 WWTPs, 01.07.2021 - 30.08.2021, 155 data points)

2.2. Data processing

The quantity of SARS-CoV-2 genetic material in wastewater follows a variety of uncertain processes, such as viral shedding dynamics of infected individuals, in-sewer degradation, sampling biases, and laboratory quantification. To compensate for inherent measurement noise, data preprocessing is required. The laboratory results of the real-time polymerase chain reaction analysis provided the SARS-CoV-2 concentration in genome copies per millilitre of wastewater. Extreme discharge conditions following periods of high precipitation in combined sewer systems result in outliers and dilution errors. Rauch et al. (2021) suggest a simple and practical approach, which is to regard measurements as outliers and exclude them, if Q exceeds the 90 percentile of the long term recorded inflow data. This approach requires at least one year of data recording for the sound estimation of quantiles and is used in this study.

Population-size marker normalisation is applied to compensate for population fluctuations within a catchment area by referencing SARS-CoV-2 virus titer c_v to a measured surrogate concentration c_m (Arabzadeh et al., 2021). The population-normalised load L_v is computed as:

$$L_v = \frac{c_v f_m}{c_m}. \tag{1}$$

A variety of population size markers are applicable to WBE (Picó and Barceló, 2021). The population-size marker compounds investigated in this work are NH4-N, COD, and N_{tot}. Standard loads of $f_{NH4\text{-}N} = 8$, $f_{COD} = 120$ and $f_{Ntot} = 12$ with the unit g/(Pe d) (grams per person equivalent per day) were used for population referencing. As recommended by Arabzadeh et al. (2021), NH4-N is the most reliable parameter for SARS-CoV-2 signal normalisation, and is prioritised over COD and N_{tot}. Only if there are insufficient data - because of a particular WWTP not measuring/reporting the compound, the population-size marker normalisation is switched to N_{tot} and COD. Similar to the viral load itself, the population size marker also shows a degree of irregularity. Occasionally, surrogate compounds are scarcely registered or show an impractical quantity. These irregularities are compensated for using the following two procedures: First, a 10-
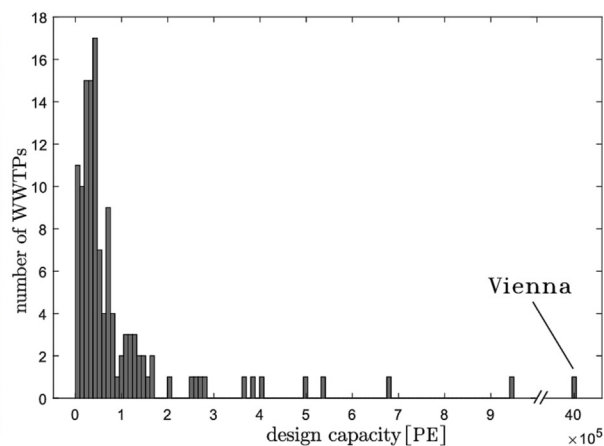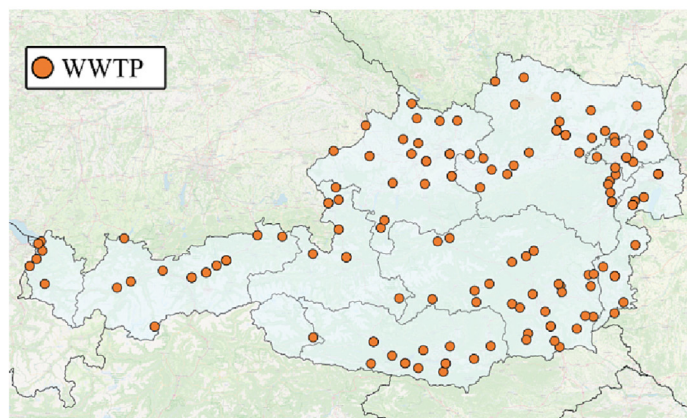

Fig. 1. Map of Austria with locations of the monitored WWTPs (left). Histogram outlining the design capacities of the WWTPs (right).

percentile filter was applied, cutting the highest and lowest 5 % of population-size marker values. As for viral concentrations, this procedure is only recommended for long time series with sufficient data. Second, the population-size marker outlier values were replaced with the median of the respective population-size marker concentration. This pragmatic approach is necessary with the high amount of epidemiological data processed in this study, which does not allow manual correction.

In contrast to the high spatial measurement resolution provided by the 123 WWTPs, severe Covid-19 cases were hospitalised in central public health facilities. The time-series provided by the Austrian Federal Ministry of Social Affairs, Health, Care, and Consumer Protection are the total hospitalisation occupancy within each federal state. This creates a comparability problem, with aggregated time-series of hospitalised cases at the federal state level and locally distributed wastewater signals on the other. To compare the aggregated hospitalisation time series, all wastewater measurements within a federal state must be aggregated and averaged. Therefore, the daily weighted averages of viral load levels in a federal state were computed. The weights correspond to the design capacity of the respective WWTPs, prioritising large plants over smaller ones. The design capacity of each WWTP is an available parameter that represents the size of the catchment area. In principle, the preferred weighting factor is the catchment population. However, this information is unavailable. This regional pooling procedure provides a solution to the comparability problem of aggregated hospitalisation occupancy in a federal state and multiple monitored WWTPs within.

The regression models trained in this study require equally spaced samples in the time series. $L_{v,weight}$ is a scattered time-series resulting from WBE measurements that are not gapless on a daily basis. To distribute the time-series data equally, up- and downsampling approaches can be used. Both upsampling (i.e. deriving daily values) and downsampling (i.e. deriving weekly values) techniques can be used to address this issue (Rauch et al., 2022). In this study, the upsampling approach on a daily basis is performed by linearly interpolating gaps before applying data smoothing.

After outlier detection, population-size marker normalisation, and regional pooling, the time series was characterised by a substantial amount of noise. Therefore, data filtering techniques are applied to reduce the signal noise and provide a mechanism to obtain the underlying information of the signal. According to Arabzadeh et al. (2021), the Friedman-Super-Smoother and Spline techniques are best suited for WBE measurements of SARS-CoV-2. Spline smoothing is a common and easily applied curve-fitting tool that can be manually tuned to optimally fit the data. Friedman-Super-Smoother is a non-parametric curve-fitting estimator that

is best operated on time series with substantial length. For shorter time series (<50 data points), Spline provides more flexibility towards high-variance data, where the Friedman-Super-Smoother risks overfitting the data. The benefit of data smoothing is the reduction of signal noise. In this special case, the ability to influence volatility characteristics is suitable for forecasting hospitalisation occupancy. A potential disadvantage of data smoothing is the loss of information, which may cause significant data points to be undervalued. As an example, Fig. 2 shows the time-line of the daily weighted average viral loads in the federal state of Vorarlberg, as well as the signal smoothed with Friedman-Super-Smoother, combining the information of the six WWTPs. the other federal states show similar dynamics.

2.3. Lead time analysis

In this work, the lead time of the SARS-Cov-2 wastewater signal over hospitalisation occupancy is exploited to infer short-term demand for public health services. Hospitalisation occupancy refers to the number of cases in intensive care units (ICU) and non-ICUs at a given point in time. Lead time quantification was performed on historical data on three different clinical time-lines: occupancy of ICU, non-ICU, and the sum of both, the hospitalisations. Clinical data are available on a daily basis and spatially aggregated in the nine federal states of Austria (Bundesministerium für Soziales, 2022). Algorithmically, the lead time τ, where the signals are best aligned is determined by the argument of the maximum, as in

$$\tau = \underset{t \in \mathbb{N}}{\operatorname{argmax}} \big(ICU(t) \odot f_t \{ L_{v,weight}(t) \} \big), \tag{2}$$

where f_t denotes the shift operator displacing the time series and $\odot$ denotes the operation of choice, i.e. cross-correlation function (CCF).

The lead time between the wastewater signal as compared to the ICU, non-ICU, and the total hospitalisation occupancy was determined by two methods. First, the CCF-analysis determines the optimal lead time based on Pearson's correlation coefficient. Second, the coefficient of determination R^2 optimisation in an OLS framework is performed to determine the optimal lead-time τ. A comparison between these two metrics indicated the optimal alignment of the signals. Two separate analyses with respect to the prevailing Covid-19 variant type were performed. First, from June to December 2021, where the delta variant is dominant, and second from January to July 2022, where the omicron variant is dominant (Amman et al., 2022).
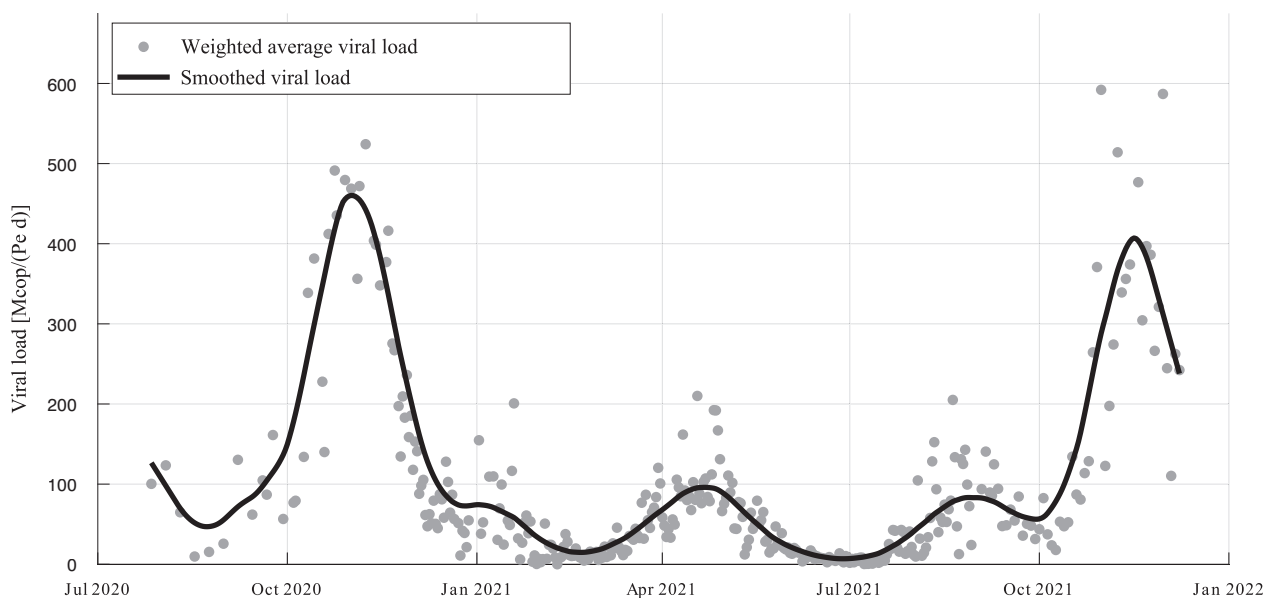

Fig. 2. Scattered data of the daily averaged viral load signal of 6 WWTPs in the federal state of Vorarlberg and the smoothed signal over time.

2.4. Regression analysis, validation and forecast

The lead time of WBE measurements allows the construction of forecast models. When comparing epidemiological hospitalisation time series over the course of the SARS-CoV-2 pandemic, it can be seen that the relationship between viral load levels in wastewater and hospitalisation occupancy shifts over time. Hospitalisation rates are trending downward because of effective vaccination and immunity to infection. Viral load levels in wastewater are affected by in-sewer temperature and variant types, unpredictably changing the shedding behaviour of infected people (Hart and Halden, 2020). Although a qualitative resemblance between hospitalisation occupancy and wastewater viral load is evident, a systematic deviation over time can be recognised. In this study, hospitalisation occupancy was predicted using multivariate regression models, aiming to optimally capture the transient process. Multivariate regression models were trained on the following predictor variables: WBE viral load signal, vaccinations, Covid-19 deaths, positivity rate, and public health and social intervention stringency index (Bundesministerium für Soziales, 2022; Hale et al., 2021). Vaccination data, Covid-19 deaths and the positivity rate are available, such as hospitalisation occupancy, aggregated at the federal state level. The vaccination time series was used in the cumulative form of total vaccinations up to a particular point in time. The public health and social intervention stringency index uses data from the Oxford Covid-19 Government Response Tracker for the comparison of government policies at a national level. Incidence data were not utilised in the regression models because they show a dependency on local test strategies and are spatially and temporally inconsistent.

Significance tests were performed in an ordinary least squares (OLS) framework to evaluate the p-values of the predictor variables in reference to the hospitalisation time series. Varying significance levels were detected, depending on the combination and number of predictor variables used. The majority of models detected a significant relationship ($p < 0.05$) between the response variable and the predictors: WBE viral load signal, Covid-19 deaths and positivity rate. Not significant predictor variables indicate a dissimilarity with the response variable but improve the regression models by compensating for the systematic deviation between the viral load signal in wastewater and hospitalisation occupancy.

The forecast prediction is constructed by taking advantage of the time lead of WBE data over the hospitalisation time series. By shifting the viral load signal for exactly this time, the differential results can be extended into the forecast period. This technique deviates from traditional forecasting methods, where predictions are based on past and present data without explicitly transposing the data in time. Therefore, an extensive effort of this study is to quantify the lead time, as explained in the previous section. Regression analysis estimates the relationship between the predictor variables encompassed in the design matrix X and response vector Y. The regression techniques applied in this study were OLS and support vector regression (SVR). OLS determines the regression parameters by minimising the sum of the squared residuals, where the residuals are the difference between

the estimated linear function and observed response variable (Holland and Welsch, 1977). Linear regression can be formalised using.

$$Y = X\beta \tag{3}$$

where β is the coefficient vector of the model. The least-squares coefficients can be determined by

$$\beta = (X^T X)^{-1} X^T Y. \tag{4}$$

SVR, on the other hand, minimizes the L2-norm of the coefficients, where the error term ε constrains the model within a margin (Rivas-Perea et al., 2013). The objective function of SVR is given by:

$$\min \frac{1}{2} ||\alpha||^2$$
$$\text{subject to: } |y_i - \langle \alpha, x_i \rangle - \beta| \le \varepsilon \tag{5}$$

To test the two regularisation methods against overfitting, ridge and lasso regularisations were implemented. These regularisation methods penalise regression coefficients to avoid overfitting and simplify the model (Boser et al., 1992; Tibshirani, 1996). Ridge regression penalizes all coefficients using the factor l, which is determined via cross-validation. Lasso regularisation shrinks all the coefficients by an absolute value. These regression and regularisation methods provide a total of six regression models, comprising the two models OLS and SVR without regularisation and an additional four from the combination with ridge (ri) and lasso (la) regularisation. Five-fold cross-validation was employed to ensure the credibility of the trained models and to test the fitness and resilience of the data partitioning. Additionally, cross-validation determines the degree of penalisation in both regularisation methods.

To evaluate the most appropriate forecast model, the validation procedure assessed the prediction accuracy in the forecast timeframe. The rolling origin procedure was employed to increase the robustness of the model selection (Tashman, 2000). With this technique, the origin of the regression timeframe remained constant, whereas the length of the training period increased progressively. Fig. 3 outlines the rolling origin forecast procedure performed in this study.

For each model, 30 rolling progression steps were computed, starting at a fixed origin and consisting of a minimum of 90 time-steps. This ensures a good resemblance of the model performance over a range of different forecast circumstances, comprising an increase in hospitalisation cases, tipping point, and levelling of cases. The rolling origin procedure is particularly important for estimating the general performance of models. The model results varied depending on the state of the pandemic. Determining the forecast errors with rolling origin cross-validation ensures that a variety of pandemic circumstances with varying forecast difficulties. In addition to the absolute and relative forecast errors, multiple performance indicators are used to evaluate model fitness, such as the root mean square error (RMSE), the coefficient of determination R^2 and its multivariate adjustment
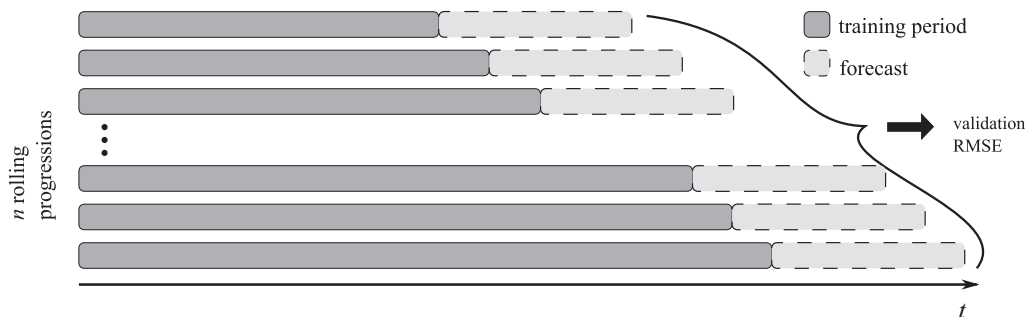

Fig. 3. Rolling origin forecast cross-validation principle.
Figure adopted from Oliveira and Ramos (2019).

R^2_{adj}, as well as the Akaike and Bayesian information criteria (AIC & BIC). The formulas for the metrics are presented in Appendix A.

3. Result and discussion

In this study, hospitalisation occupancy was predicted using multivariate regression models based on epidemiological data. Viral load levels in wastewater and hospitalisation occupancy data are closely related to Covid-19 prevalence and are highly correlated.

3.1. Lead time results

The lead time τ between the viral load in the wastewater and the hospitalised SARS-CoV-2 cases was determined using the methods discussed in Section 2.3. Fig. 4 shows the optimal lead time τ between the wastewater signal and the three clinical time-series, ICU, non-ICU, and the sum of both (Hosp). For each federal state, CCF-analysis and R^2 optimisation routine were applied. The Austrian mean is displayed at the bottom right of the table. The CCF-analysis tends to compute lower lead-time estimates than the R^2 optimisation routine. A high variance in lead times among federal states can be observed. This behaviour may be caused by differences in the case reporting procedure and size differences of the regions. Likewise, it is possible that patients are transferred between the ICU and non-ICU, distorting the lead-time quantification. Regardless of these limitations, the quality of the constructed forecasts was not compromised because the forecasting period was not disproportionately high.

The results in Fig. 4 show that the average lead time of ICU occupancy is higher than that of non-ICU occupancy. As the sum of both signals, Hosp positions itself between the other two, as expected. This result can be explained by the mean difference in length of stay (LOS) between ICU and non-ICU patients in the hospital. Chiam et al. (2021) derives a mean LOS of 12.3 days for ICU patients and 5.7 days for non-ICU SARS-CoV-2 patients. Similar results were obtained by Vekaria et al. (2021) with a mean LOS of 12.9 days for ICU patients and 8.5 days for non-ICU patients. In general, the higher the LOS, the larger the lead time between the WBE signal and hospitalisation.

From a conceptual perspective, predicting hospitalisation admission rather than occupation is preferred; however, admission data are not publicly available to train the models. Nonetheless, an estimate for hospitalisation admission from occupancy can be formalised as

$$C_a = f_t \left\{ \frac{C_o}{LOS} \right\}. \tag{6}$$

C_a and C_o denote the clinical admission/occupancy time series and f_t is the shift operator, transposing the argument in time by $t = -LOS/2$. This procedure is a simple estimate that disregards the statistical distribution of LOS for patients, as well as dynamic effects. Assuming that the occupancy signal lags the admission signal by half of the respective LOS, the mean lead time in the delta variant time-frame of the SARS-CoV-2 wastewater signal towards clinical admissions resulted in 11.4 days-time lead for ICU and 8.1 days for non-ICU cases. For the omicron variant time frame, a mean lag time of 8.5 days for ICU and 5.1 days for non-ICU admission was observed. This result resembles the estimates from Galani et al. (2022), predicting a time lead towards hospital admissions of 8 days and 9 days for ICU admission. It is important to note that although a clear time delay can be detected in a retrospective analysis, the concurrent lead time over clinical data is controversial in the literature (Bibby et al., 2021; Olesen et al., 2021). A rapid and well-established WBE sampling and analysis infrastructure is key to exploiting the temporal lead.

By evaluating the time differentials between the delta and omicron variant time frames, it can be seen that, on average, a smaller lead time occurs between the signals in the omicron time-frame. From the Austrian mean, the detected decline in time lead for ICU occupancy was 2.9 days and 3.0 days for non-ICU occupancy. The unavailable entries in Fig. 4 are caused by a lack of significant correlations between the two time-series of ICU occupancy and viral load levels in the omicron variant time-frame. This is likely caused by the low number of ICU cases in the omicron timeframe in combination with integer-valued numbers of the response variable. The relatively flat and discrete ICU curve does not resemble the features of the viral load signal, and thus shows sporadic no significant resemblance of the signals.
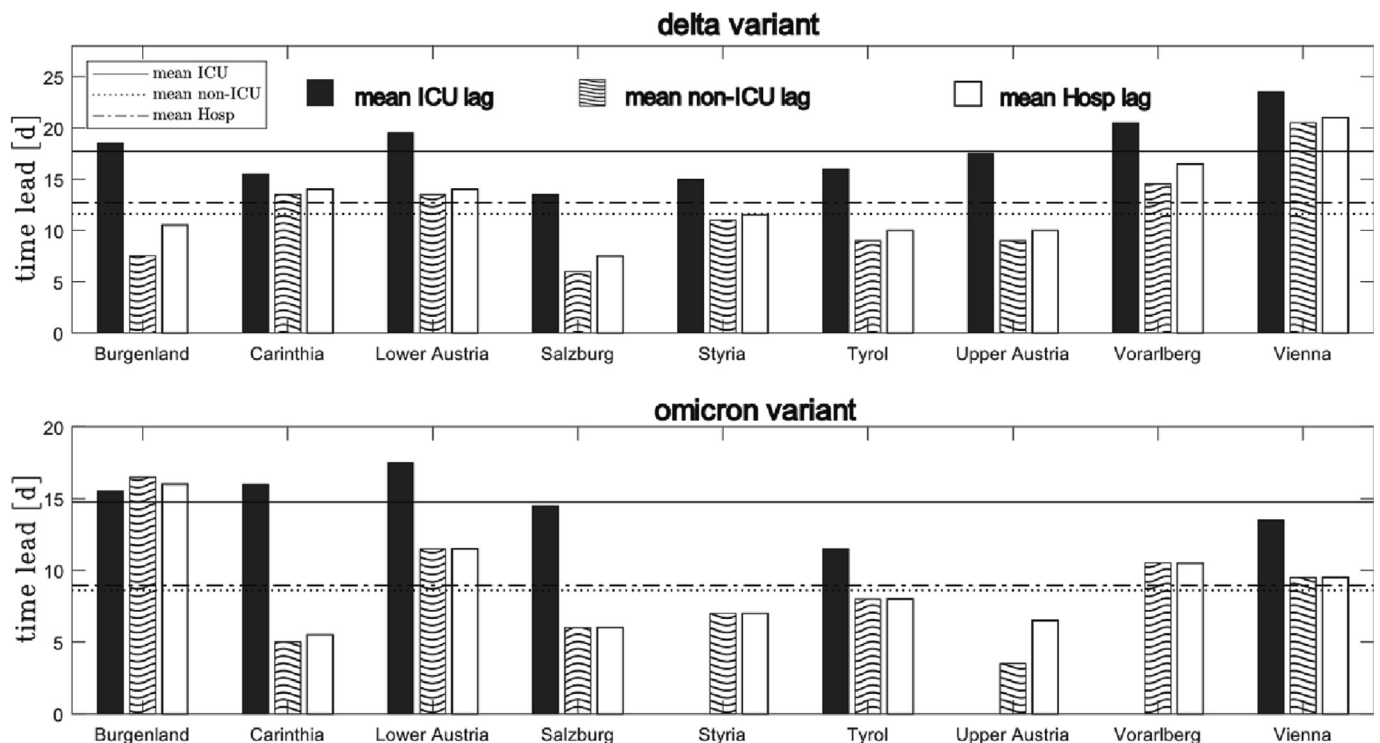

Fig. 4. Lead time quantification results in the delta and omicron variant time-frame.
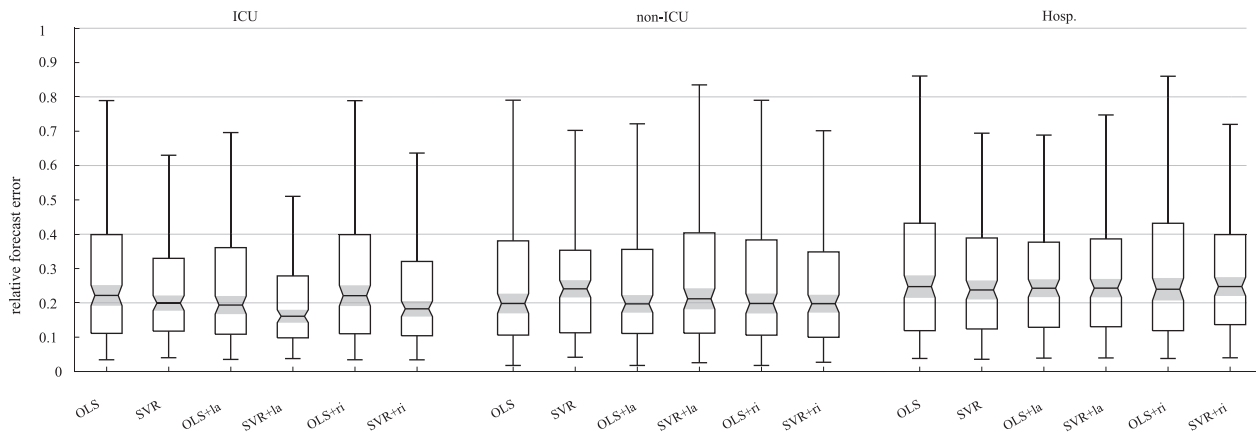
Fig. 5. Boxplot of the relative forecast error at the end of the forecast period, predicting the time-series ICU, non-ICU and total hospitalizations in the delta time-frame.

3.2. Forecast results

The goodness-of-fit and performance indicators are presented in this section. Fig. 5 displays boxplots of the relative forecast error at the end of the forecast period between June and December 2021 (delta variant). Each box-plot comprises 270 values from nine federal states with 30 different forecast periods resulting from the rolling-origin procedure. It can be seen that the mean relative error of the majority of the models is between 0.15 and 0.25. This result demonstrates a consistent forecast of hospitalisation, ICU, and non-ICU occupancy from WBE regression models with a mean forecast error between 15 % and 25 % in a forecast horizon of approximately two weeks ahead. SVR outperforms OLS regression by a small margin. No significant increase in model performance can be detected by ridge or lasso regularisation, indicating that overfitting is not a concern for the trained models. The relative forecast errors in the delta timeframe are shown in Fig. 5. The same approach for the omicron time-frame produces infeasible high relative errors, owing to low levels of hospitalisation occupancy leading to inconsistently high errors caused by division by small numbers (occasionally by 0). However, in absolute numbers, the forecasts in the omicron timeframe produce a similar quality of results.

Tables 1 and 2 display the performance indicators of the trained models in the two respective periods. The spatial aggregation of the data is again performed on a regional scale, and the displayed values are the averages of the nine federal states. The AIC and BIC metrics have varying levels regarding the three response variables, owing to the amplitude of the different time series. Therefore, these metrics can only be compared between different models that predict the same response variable. Lower values of AIC and BIC indicate a superior quality of fit. A throughout the lower AIC/BIC can be seen for ICU over the non-ICU and hospitalisation analysis because the amplitude of the signal is an order of magnitude lower.

The overall lowest AIC and BIC values were attained by SVR compared to the other models. Regularisation has no significant or slightly negative impact on model performance. The coefficients of determination show similar behaviour. Discrete-time series with a small amplitude, such as ICU occupancy, are prone to cause a higher relative model error. This is caused by the discrete characteristic and coarseness of the response variable on the one hand and the languid nature of the smoothed viral load signal on the other. This problem of granularity is emphasised for small federal states, where periods of ICU occupancy are characterised by levels ranging between 0 and 5.

By comparing the results in Tables 1 and 2 it can be seen that the Akaike and Bayesian information criteria show similar performance. A decrease in R^2 was visible, especially for the ICU response variable. The poor R^2 performance in Table 2 can be explained by the comparatively low ICU occupancy levels in the omicron timeframe, preventing high model fitness with real-valued predictions. Integer time series with small amplitudes are intrinsically more difficult to predict using real-valued models because of the higher relative influence of random errors in the training data. Although the relative errors are high when forecasting ICU occupancy with small amplitudes, the absolute error for practical forecasting purposes is not higher than that for signals with a high amplitude.

Fig. 6 shows two forecast examples for Vienna and Vorarlberg. The bars indicate non-ICU occupancy, where the dark colour indicates the training period and the light grey bars indicate the forecast period. The black dotted line indicates the model prediction. A close agreement between the model prediction and actual hospitalisation occupancy can be seen. With the size difference between the two federal states, different non-ICU patient levels can be observed, as well as a slightly better forecast in Vienna due to the previously mentioned reasons for coarseness in small regions such as Vorarlberg.

Table 1
Results of the performance indicators of the applied models in the delta variant time-frame.

Delta	OLS			SVM			OLS + la		
	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.
AIC	10.394	15.797	16.509	9.632	14.840	15.210	10.802	16.146	16.802
BIC	15.726	21.098	21.748	14.964	20.140	20.449	16.134	21.446	22.042
R^2	0.845	0.855	0.883	0.877	0.913	0.937	0.774	0.805	0.851
R^2_{adj}	0.841	0.852	0.880	0.875	0.911	0.935	0.769	0.801	0.848
	SVM + la			OLS + ri			SVM + ri		
	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.
AIC	10.796	16.091	16.826	10.394	15.797	16.509	10.581	15.761	16.484
BIC	16.128	21.391	22.066	15.726	21.098	21.748	15.913	21.062	21.724
R^2	0.756	0.801	0.828	0.845	0.855	0.883	0.770	0.839	0.860
R^2_{adj}	0.751	0.796	0.824	0.841	0.852	0.880	0.765	0.835	0.857

Table 2

Results of the performance indicators of the applied models in the omicron variant time-frame.

Omicron	OLS			SVM			OLS + la		
	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.
AIC	9.974	17.075	16.963	7.637	15.923	15.963	10.388	18.118	17.999
BIC	14.334	21.722	21.610	11.996	20.570	20.610	14.747	22.764	22.646
R^2	0.339	0.683	0.670	0.768	0.877	0.854	0.254	0.540	0.517
R^2_{adj}	0.317	0.674	0.661	0.761	0.873	0.850	0.230	0.527	0.504
	SVM + la			OLS + ri			SVM + ri		
	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.	ICU	Non-ICU	Hosp.
AIC	11.014	18.580	18.310	9.974	17.075	16.963	10.534	17.773	17.521
BIC	15.374	23.227	22.957	14.334	21.722	21.610	14.894	22.420	22.168
R^2	0.015	0.412	0.414	0.339	0.683	0.670	0.127	0.514	0.515
R^2_{adj}	−0.018	0.396	0.398	0.317	0.674	0.661	0.099	0.501	0.502

3.3. Wastewater data abundance and forecast accuracy

As an additional analysis of the hospitalisation forecasts, an increase in forecast accuracy as a function of the available wastewater data was performed. With 123 actively monitored wastewater treatment plants, Austria monitors extensively (average of 15 WWTPs per federal state, excluding Vienna). Multiple monitored plants in each federal state provide considerable data to train the regression models. The trained regression models in this work utilise all available wastewater data. In this analysis, forecasts are constructed with varying amounts of wastewater data for each federal state to determine the resilience of the applied method to less available WBE data. Fig. 7 shows the average forecast accuracy of the six trained regression models on 16 randomly selected days as a function of the utilised wastewater data. The scattered data display the relative forecast error, the grey lines outline the regression trends over the utilised wastewater data, and the dotted black line indicates the trend. It is evident that a trend towards better forecast prediction is achieved when more wastewater data are used, which supports the findings of Vaughan et al. (2023). This result justifies the inclusion of more plants in the surveillance system and suggests an increase in model fitness when more data are included.

4. Conclusion

In this study, we show the similarity of viral load signals obtained by WBE with hospitalisation occupancy by means of multivariate regression analysis. Hospitalisation occupancy at the federal state level was predicted based on several monitored wastewater treatment plants within one region. The aggregation of dispersed WBE measurements within a region increases the correlation quality of the signals and the performance of the regression analysis. The approach of utilising a multitude of WWTPs within a federal state to forecast hospitalisation demand at a regional level was applied successfully. Forecast predictions were constructed by exploiting the temporal lead of wastewater data over a clinical hospitalisation time series. Consistent forecasts with a mean relative error between 15 and 25 % on a forecast time frame of approximately two weeks is feasible. Regression analysis is carried out on all nine regions of Austria, and the size influences of the federal states are discussed. Encountered challenges are caused by changing Covid-19 variant types, causing systematic deviations in the viral concentration in relation to hospitalisations. Newly emerging variants likely cause a change in the shedding dynamics of infected people and, therefore, influence the epidemiological time series. Low numbers in the ICU time
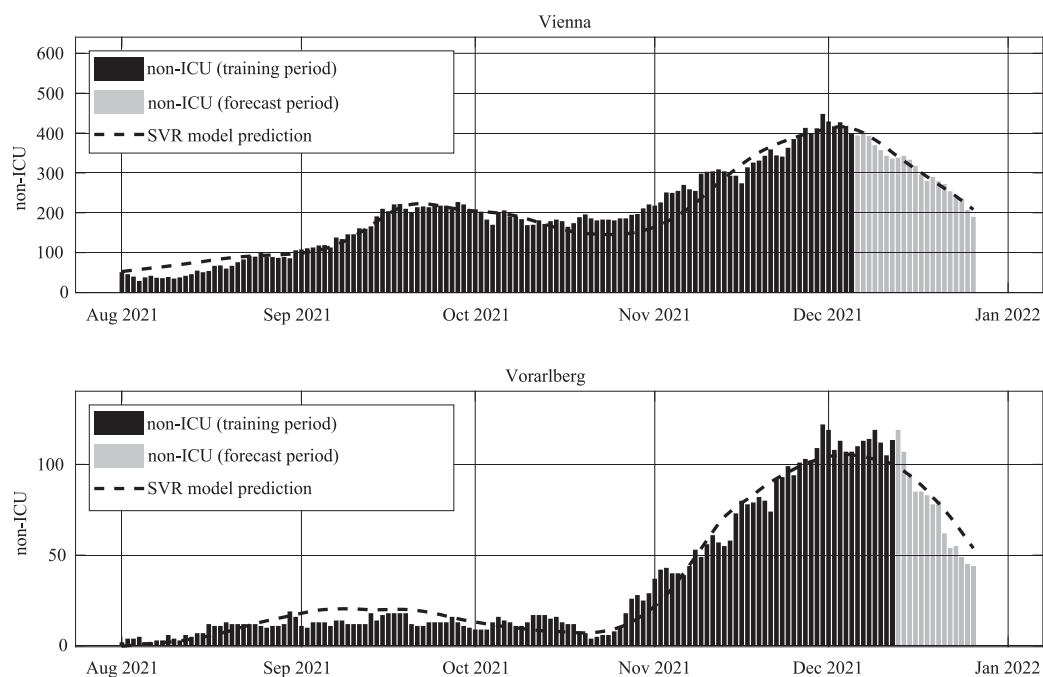

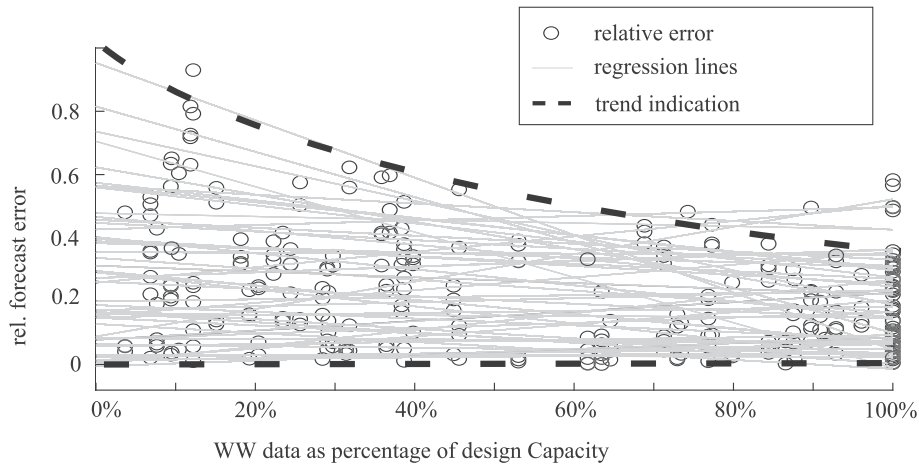**Fig. 6.** Forecast examples of non-ICU occupancy in Vienna and Vorarlberg.

Fig. 7. Forecast error decreases with an increase in WW data.

series and a high degree of granularity impede the optimal model construction during the dominance of the omicron variant.

The results support a strong qualitative resemblance between wastewater-based epidemiologically derived signals and SARS-CoV-2 hospitalisation occupancy. Support vector regression and ordinary least square regression models were successfully trained based on viral load levels in raw sewage and epidemiological time-series, such as vaccination data, Covid-19 deaths, positivity rate, and the public health and social intervention stringency index. Therefore, ridge and lasso regularisation had no benefit on the forecast performance. The models were examined with various performance indicators, such as the root mean square error, Akaike and Bayesian information criteria, and the standard and adjusted coefficient of determination. The time lead of the SARS-CoV-2 viral load signal in wastewater and the clinical hospitalisation time series facilitate the direct prediction of hospitalisation occupancy. The time lead quantification analysis suggests a feasible mean forecast horizon of 10.1 days for non-ICU occupation and 16.2 days for ICU occupation. These results demonstrate the short-term predictive potential of WBE SARS-CoV-2 surveillance and allow for direct application of WBE.

CRediT authorship contribution statement

Hannes Schenk: Writing – original draft, Data curation, Software, Formal analysis, Visualization. **Petra Heidinger:** Resources. **Heribert Insam:** Funding acquisition, Writing – review & editing. **Norbert Kreuzinger:** Funding acquisition. **Rudolf Markt:** Funding acquisition, Writing – review & editing. **Fabiana Nägele:** Funding acquisition, Resources. **Herbert Oberacher:** Data curation, Writing – review & editing, Funding acquisition. **Christoph Scheffknecht:** Resources. **Martin Steinlechner:** Data curation, Writing – review & editing, Funding acquisition. **Gunther Vogl:** Resources. **Andreas Otto Wagner:** Funding acquisition, Resources. **Wolfgang Rauch:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Project administration.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Performance indicators

The performance metrics used to quantify the model fitness in this work are the root mean square error (RMSE), the coefficient of determination R^2, its multivariate adjustment R^2_{adj}, and the Akaike and Bayesian information criteria (AIC and BIC). The RMSE metric is calculated as follows:

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{\Sigma_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{A 1}$$

where y_t denotes the dependent variable and $\hat{y}$ is the model prediction on day t. The relative RMSE is computed by dividing the RMSE by the mean of the response variable within the timeframe of interest.

$$\text{rRMSE} = \frac{RMSE}{\overline{Y}}. \tag{A 2}$$

The coefficient of determination quantifies the variation in the dependent variable that is predictable by the independent variables and qualifies as a measure of goodness of fit. R^2 is calculated as follows:

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}}, \tag{A 3}$$

where SSR denotes the sum of squares of residuals and SST denotes the total sum of squares. The related R^2_{adj} considers the number of predictor variables p and is defined as.

$$R^2_{adj} = 1 - \frac{\text{SSR}}{\text{SST}}\left(\frac{n-1}{n-p}\right) \tag{A 4}$$

The information criteria AIC and BIC are tools for model selection, designed according to the principle of Occam's razor. The criteria were defined as follows:

$$AIC = 2p - 2\ \ln\left(\hat{L}\right), \tag{A 5}$$

$$BIC = \ln(n)\,p - 2\ \ln\left(\hat{L}\right), \tag{A 6}$$

where $\ln(\hat{L})$ denotes the maximum logarithmic likelihood function under the assumption of normally distributed errors.

References

Aberi, P., et al., 2021. Quest for optimal regression models in SARS-CoV-2 wastewater based epidemiology. Int. J. Environ. Res. Public Health 18 (20), pp. https://doi.org/10.3390/ijerph182010778.

Ahmed, W., et al., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Sci. Total Environ. 728, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764.

Amman, F., et al., 2022. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. Nat. Biotechnol. https://doi.org/10.1038/s41587-022-01387-y.

Arabzadeh, R., Grünbacher, D.M., Insam, H., Kreuzinger, N., Markt, R., Rauch, W., 2021. Data filtering methods for SARS-CoV-2 wastewater surveillance. Water Sci. Technol. 84 (6), 1324–1339. https://doi.org/10.2166/wst.2021.343.

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 Res. 202, 117438. https://doi.org/10.1016/j.watres.2021.117438.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT '92, Pittsburgh, Pennsylvania, United States, pp. 144–152.

Bundesministerium für Soziales, 2022. Gesundheit, Pflege und Konsumentenschutz (BMSGPK). https://www.data.gv.at/katalog/dataset/846448a5-a26e-4297-ac08-ad7040af20f1 September.

Cevik, M., Tate, M., Lloyd, O., Maraolo, A.E., Schafers, J., Ho, A., 2021. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. Lancet Microbe 2 (1), e13–e22. https://doi.org/10.1016/S2666-5247(20)30172-5.

Chiam, T., et al., 2021. Hospital length of stay among COVID-19-positive patients. J. Clin. Transl. Res. 7 (3), 377–385.

Crank, K., Chen, W., Bivins, A., Lowry, S., Bibby, K., 2022. Contribution of SARS-CoV-2 RNA shedding routes to RNA loads in wastewater. Sci. Total Environ. 806 (Pt 2), 150376. https://doi.org/10.1016/j.scitotenv.2021.150376.

Daleiden, B., et al., 2022. Wastewater surveillance of SARS-CoV-2 in Austria: development, implementation, and operation of the Tyrolean wastewater monitoring program. J. Water Health 20 (2), 314–328. https://doi.org/10.2166/wh.2021.218.

Daughton, C.G., 2020. Wastewater surveillance for population-wide Covid-19: the present and future. Sci. Total Environ. 736, 139631. https://doi.org/10.1016/j.scitotenv.2020.139631.

Galani, A., et al., 2022. SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. Sci. Total Environ. 804, 150151. https://doi.org/10.1016/j.scitotenv.2021.150151.

Greenwald, H.D., et al., 2021. Tools for interpretation of wastewater SARS-CoV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay Area. Water Res. X 12, 100111. https://doi.org/10.1016/j.wroa.2021.100111.

Hale, T., et al., 2021. A global panel database of pandemic policies (Oxford COVID-19 government response Tracker). Nat. Hum. Behav. 5 (4), 529–538. https://doi.org/10.1038/s41562-021-01079-8.

Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges. Sci. Total Environ. 730, 138875. https://doi.org/10.1016/j.scitotenv.2020.138875.

He, X., et al., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. 26 (5), 672–675. https://doi.org/10.1038/s41591-020-0869-5.

Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. Commun. Stat. Theory Methods 6 (9), 813–827. https://doi.org/10.1080/03610927708827533.

Kashi, A.H., de La Rosette, J., Amini, E., Abdi, H., Fallah-Karkan, M., Vaezjalali, M., 2020. Urinary viral shedding of COVID-19 and its clinical associations: a systematic review and meta-analysis of observational studies. Urol. J. 17 (5), 433–441. https://doi.org/10.22037/uj.v16i7.6248.

Kitajima, M., et al., 2020. SARS-CoV-2 in wastewater: state of the knowledge and research needs. Sci. Total Environ. 739, 139076. https://doi.org/10.1016/j.scitotenv.2020.139076.

La Rosa, G., et al., 2020. First detection of SARS-CoV-2 in untreated wastewaters in Italy. Sci. Total Environ. 736, 139652. https://doi.org/10.1016/j.scitotenv.2020.139652.

Li, X., Zhang, S., Shi, J., Luby, S.P., Jiang, G., 2021. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chem. Eng. Journal (Lausanne, Switzerland : 1996) 415, 129039. https://doi.org/10.1016/j.cej.2021.129039.

Lodder, W., de Roda Husman, A.M., 2020. SARS-CoV-2 in wastewater: potential health risk, but also data source. Lancet. Gastroenterol. Hepatol. 5 (6), 533–534. https://doi.org/10.1016/S2468-1253(20)30087-X.

Lundy, L., et al., 2021. Making waves: collaboration in the time of SARS-CoV-2 - rapid development of an international co-operation and wastewater surveillance database to support public health decision-making. Water Res. 199, 117167. https://doi.org/10.1016/j.watres.2021.117167.

Markt, R., Mayr, M., Peer, E., Wagner, A.O., Lackner, N., Insam, H., 2021. Detection and stability of SARS-CoV-2 fragments in wastewater: impact of storage temperature. Pathogens (Basel, Switzerland) 10 (9). https://doi.org/10.3390/pathogens10091215.

Markt, R., et al., 2022. Detection and abundance of SARS-CoV-2 in wastewater in Liechtenstein, and the estimation of prevalence and impact of the B.1.1.7 variant. J. Water Health 20 (1), 114–125. https://doi.org/10.2166/wh.2021.180.

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. 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. Environ. Sci. Technol. Lett. 7 (7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357.

Mostaghimi, D., Valdez, C.N., Larson, H.T., Kalinich, C.C., Iwasaki, A., 2022. Prevention of host-to-host transmission by SARS-CoV-2 vaccines. Lancet Infect. Dis. 22 (2), e52–e58. https://doi.org/10.1016/S1473-3099(21)00472-2.

Olesen, S.W., Imakaev, M., Duvallet, C., 2021. Making waves: defining the lead time of wastewater-based epidemiology for COVID-19. Water Res. 202, 117433. https://doi.org/10.1016/j.watres.2021.117433.

Oliu-Barton, M., et al., 2021. Elimination versus mitigation of SARS-CoV-2 in the presence of effective vaccines. Lancet Glob. Health https://doi.org/10.1016/S2214-109X(21)00494-0.

Oliveira, J.M., Ramos, P., 2019. Assessing the performance of hierarchical forecasting methods on the retail sector. Entropy (Basel, Switzerland) 21 (4). https://doi.org/10.3390/e21040436.

Peccia, J., et al., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat. Biotechnol. 38 (10), 1164–1167. https://doi.org/10.1038/s41587-020-0684-z.

Petherick, A., et al., 2021. A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. Nat. Hum. Behav. 5 (9), 1145–1160. https://doi.org/10.1038/s41562-021-01181-x.

Picó, Y., Barceló, D., 2021. Identification of biomarkers in wastewater-based epidemiology: main approaches and analytical methods. TrAC Trends Anal. Chem. 145, 116465. https://doi.org/10.1016/j.trac.2021.116465.

Rauch, W., et al., 2021. Datenbehandlung in der SARS-CoV-2 Abwasserepidemiologie. KA - Korrespondenz Abwasser, Abfall 7.

Rauch, W., Schenk, H., Insam, H., Markt, R., Kreuzinger, N., 2022. Data modelling recipes for SARS-CoV-2 wastewater-based epidemiology. Environ. Res. 214 (Pt 1), 113809. https://doi.org/10.1016/j.envres.2022.113809.

Rivas-Perea, P., Cota-Ruiz, J., Chaparro, D.G., Venzor, J.A., Carreón, A.Q., Rosiles, J.G., 2013. Support vector Machines for Regression: a succinct review of large-scale and linear programming formulations. IJIS 03 (01), 5–14. https://doi.org/10.4236/ijis.2013.31002.

Tashman, L.J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. Int. J. Forecast. 16 (4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0.

Tene, L., Bergroth, T., Eisenberg, A., David, S.S.B., Chodick, G., 2022. "Risk factors, health outcomes, healthcare services utilization, and direct medical costs of patients with long COVID", international journal of infectious diseases : IJID : official publication of the international society for. Infectious Diseases 128, 3–10. https://doi.org/10.1016/j.ijid.2022.12.002.

Thompson, J.R., et al., 2020. Making waves: wastewater surveillance of SARS-CoV-2 for population-based health management. Water Res. 184, 116181. https://doi.org/10.1016/j.watres.2020.116181.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. 58 (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Vaughan, L., et al., 2023. An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data. Sci. Total Environ. 858 (Pt 1), 159748. https://doi.org/10.1016/j.scitotenv.2022.159748.

Vekaria, B., et al., 2021. Hospital length of stay for COVID-19 patients: data-driven methods for forward planning. BMC Infect. Dis. 21 (1), 700. https://doi.org/10.1186/s12879-021-06371-6.

WHO, WHO Coronavirus (COVID-19) Dashboard. [Online]. Available: https://covid19.who.int/.

Wölfel, R., et al., 2020. Virological assessment of hospitalized patients with COVID-2019. Nature 581 (7809), 465–469. https://doi.org/10.1038/s41586-020-2196-x.

Wurtzer, S., et al., 2021. SARS-CoV-2 genome quantification in wastewaters at regional and city scale allows precise monitoring of the whole outbreaks dynamics and variants spreading in the population. Sci. Total Environ., 152213 https://doi.org/10.1016/j.scitotenv.2021.152213.

Wyllie, A.L., et al., 2020. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. N. Engl. J. Med. 383 (13), 1283–1286. https://doi.org/10.1056/NEJMc2016359.

Zhu, Y., et al., 2021. Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. Sci. Total Environ. 767, 145124. https://doi.org/10.1016/j.scitotenv.2021.145124.