AIMS Mathematics, 8(7): 16790–16824.
DOI: 10.3934/math.2023859
Received: 09 September 2022
Revised: 17 April 2023
Accepted: 17 April 2023
Published: 15 May 2023

http://www.aimspress.com/journal/Math

Research article

Robust smoothing of left-censored time series data with a dynamic linear model to infer SARS-CoV-2 RNA concentrations in wastewater

Luke Lewis-Borrell[1], Jessica Irving[1], Chris J. Lilley[1], Marie Courbariaux[2], Gregory Nuel[3], Leon Danon[4], Kathleen M. O'Reilly[5], Jasmine M. S. Grimsley[1], Matthew J. Wade[1,6,*] and Stefan Siegert[7]

[1] UK Health Security Agency, Nobel House, Smith Square, London SW1P 3JR, UK

[2] Obépine/SUMMIT, Sorbonne University, 75005 Paris, France

[3] Stochastics and Biology Group, Probability and Statistics (LPSM, CNRS 8001), Sorbonne University, Campus Pierre et Marie Curie, 4 Place Jussieu, 75005, Paris, France

[4] Department of Engineering Mathematics, University of Bristol, Ada Lovelace Building, University Walk, Bristol BS8 1TW, UK

[5] Centre for Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

[6] School of Engineering, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

[7] Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter EX4 4QE, UK

* **Correspondence:** Email: matthew.wade@ukhsa.gov.uk.

Abstract: Wastewater sampling for the detection and monitoring of SARS-CoV-2 has been developed and applied at an unprecedented pace, however uncertainty remains when interpreting the measured viral RNA signals and their spatiotemporal variation. The proliferation of measurements that are below a quantifiable threshold, usually during non-endemic periods, poses a further challenge to interpretation and time-series analysis of the data. Inspired by research in the use of a custom Kalman smoother model to estimate the true level of SARS-CoV-2 RNA concentrations in wastewater, we propose an alternative left-censored dynamic linear model. Cross-validation of both models alongside a simple moving average, using data from 286 sewage treatment works across England, allows for a comprehensive validation of the proposed approach. The presented dynamic linear model is more parsimonious, has a faster computational time and is represented by a more flexible modelling framework than the equivalent Kalman smoother. Furthermore we show how the use of wastewater data, transformed by such models, correlates more closely with regional case rate positivity as published by the Office for National Statistics (ONS) Coronavirus (COVID-19) Infection Survey.

The modelled output is more robust and is therefore capable of better complementing traditional surveillance than untransformed data or a simple moving average, providing additional confidence and utility for public health decision making.

La détection et la surveillance du SARS-CoV-2 dans les eaux usées ont été développées et réalisées à un rythme sans précédent, mais l'interprétation des mesures de concentrations en ARN viral, et de leurs variations spatio-temporelles, pose question. En particulier, l'importante proportion de mesures en deçà du seuil de quantification, généralement pendant les périodes non endémiques, constitue un défi pour l'analyse de ces séries temporelles. Inspirés par un travail de recherche ayant produit un lisseur de Kalman adapté pour estimer les concentrations réelles en ARN de SARS-CoV-2 dans les eaux usées à partir de ce type de données, nous proposons un nouveau modèle linéaire dynamique avec censure à gauche. Une validation croisée de ces lisseurs, ainsi que d'un simple lissage par moyenne glissante, sur des données provenant de 286 stations d'épuration couvrant l'Angleterre, valide de façon complète l'approche proposée. Le modèle présenté est plus parcimonieux, offre un cadre de modélisation plus flexible et nécessite un temps de calcul réduit par rapport au Lisseur de Kalman équivalent. Les données issues des eaux usées ainsi lissées sont en outre plus fortement corrélées avec le taux d'incidence régional produit par le bureau des statistiques nationales (ONS) Coronavirus Infection Survey. Elles se montrent plus robustes que les données brutes, ou lissées par simple moyenne glissante, et donc plus à même de compléter la surveillance traditionnelle, renforçant ainsi la confiance en l'épidémiologie fondée sur les eaux usées et son utilité pour la prise de décisions de santé publique.

Keywords: dynamic linear model; wastewater-based epidemiology; COVID-19; time series; left-censoring; Bayesian inference
Mathematics Subject Classification: 62D20, 62F15, 92-10

1. Introduction

The COVID-19 pandemic has exerted huge and unprecedented pressure on public health resources globally. Cross-sectional surveys to establish disease prevalence are likely to be financially unsustainable in the long term and rely heavily on continued cooperation from the public [1]. Wastewater monitoring to detect and quantify SARS-CoV-2 viral RNA shed by infected individuals in the population, and to indicate infection prevalence, was adopted relatively early in the course of the pandemic across a number of countries [2, 3], expanding to 67 countries by mid-2022 [4]. Although the demographic coverage and utility of wastewater monitoring varies across adopters of this approach, the method is generally less intrusive, relatively unbiased in terms of its demographic and epidemiological coverage, and costs significantly less per capita than clinical testing programmes (e.g. $148 per individual PCR test cf. $300 per wastewater sample representing larger populations [5]). Wastewater surveillance is thus, arguably, an alternative, or at least complementary, approach to clinical testing programmes.

Wastewater-based epidemiology has been used in some areas of public health for decades [6], but is a relatively novel tool for emerging pathogens. Applications include tracking viral dynamics to monitoring chemical exposures and prescription drug consumption [7, 8]. As wastewater sampling for the detection and monitoring of SARS-CoV-2 has been developed and applied at an unprecedented

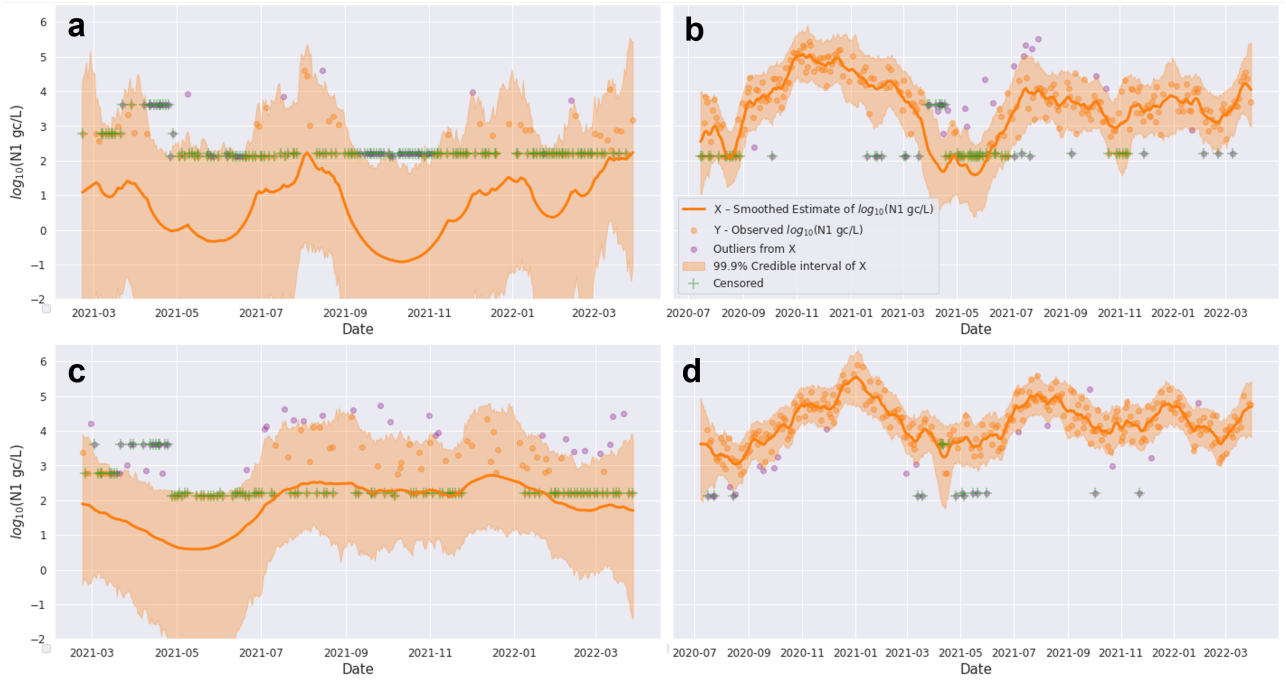

Figure 1. Observed concentrations of SARS-CoV-2 N1 gc/L ($\log_{10}$ transformed, orange points) plotted over time at four sites in England, along with corresponding fit of the proposed dynamic linear model (orange line). The shaded orange area around the lines represents the 99.9% credible intervals for the estimated underlying state X, with observed $\log_{10}$(N1 gc/L) values outside of these intervals classified as outliers to X. Plots **a** and **c** are examples of sites with large fitted τ parameter values (measurement noise, scale parameter in t-distribution). Plots **b** and **d** are sites with small fitted ν parameter values (probability of outliers, degrees of freedom in t-distribution). **a.** Site name: Burton-on-Trent, ν: mean = 5.26, sd = 1.4, $\hat{r}$ = 1.00, τ: mean = 1.25, sd = 0.24, $\hat{r}$ = 1.00. Date range: 21/02/2021 - 30/03/2022. **b.** Site name: Lincoln, ν: mean = 2.25, sd = 0.28, $\hat{r}$ = 1.00, τ: mean = 0.422, sd = 0.04, $\hat{r}$ = 1.00. Date range: 15/07/2020 - 30/03/2022. **c.** Site name: Alfreton, ν: mean = 5.91, sd = 1.43, $\hat{r}$ = 1.00. τ: mean = 1.36 , sd = 0.16, $\hat{r}$ = 1.00. Date range: 22/02/2021 - 28/03/2022. **d.** Site name: London Beckton, ν: mean = 2.28, sd = 0.27, $\hat{r}$ = 1.00, τ: mean = 0.29, sd = 0.03, $\hat{r}$ = 1.00. Date range: 08/07/2020 - 30/03/2022.

pace, uncertainty remains when interpreting the measured viral RNA signals and their spatiotemporal variation. Variation in the underlying sample, sampling method, and testing, due in part to lack of standardisation, as well as systematic variability in space and time in the measurement environment (e.g. sewersheds), can result in a large degree of noise in the observed signal [9]. Sampling frequency is typically dependent on cost constraints, resulting in sparse and irregularly sampled data. Furthermore, wastewater measurements are typically left-censored if they fall below certain analytical thresholds, such as the limit of detection (LOD), the lowest concentration at which viral RNA is detectable with a given probability (typically 95%); and the limit of quantification (LOQ), the lowest concentration at which viral RNA can be reliably measured with a predefined accuracy. Methods to handle measurements that fall below these limits (e.g. statistical methods, imputation, and scalar or zero

replacement) are not standardised and depend on the interpretation of the data (for example, if low values do not impact interpretation they may be omitted from downstream analysis), and information available to the analysts [10, 11].

There are several approaches to infer wastewater concentration from noisy, censored, incomplete time series measurements. One common and straightforward approach for denoising time series is to calculate a moving average (MA). However, MA are sensitive to outliers and missing values. There is ambiguity about the most appropriate window-size, and whether to calculate a weighted or ordinary average. Uncentred MAs also operate with a lag, where larger windows create larger lags, delaying reactivity of surveillance in time-dependent operations. Lastly, an MA estimate can never be smaller than the censoring threshold, which leads to biased estimates.

State-space methods model observed data as functions of latent, unobserved stochastic processes and can better account for missing data, observational noise, and censoring. Recently, others have proposed state-space methods to infer viral concentrations from wastewater time series. The underlying "true" viral concentration at time X_t is modelled as a first-order auto-regressive (AR1) process [12,13]. To account for measurement noise and outliers in the observations, measurements Y_t are assumed to be equal to X_t plus an independent mean-zero Gaussian observation error. To account for outliers, from time to time Y_t is assumed to be replaced by an independent Uniformly distributed random variable that is unrelated to X_t. Left-censoring is accounted for by capping Y_t at the (known) limit of quantification. Using a Kalman Filter and numerical approximations, the state variable X_t is inferred from observation data Y_t to produce a smoothed estimate of viral concentration, with outliers removed, that can extend below the known limit of quantification [12].

In this paper, we propose and test a simpler, more realistic, and more flexible state-space model. Our latent variable is modelled by a first order random walk (RW1) instead of an AR1 process, which reduces the number of model parameters. Instead of randomly replacing observations by a random number, our model generates outliers by assuming observation errors from a heavy-tailed t-distribution. This has the benefit that observations classified as "outliers" can still be informative about viral concentrations.

Our model is implemented in the Stan modelling language [14], which allows for fast Bayesian inference and straightforward extensions of the model.

2. Methods

2.1. Wastewater data source

Untreated influent samples were collected from sewage treatment plant sites across England at a frequency of four days a week by the Environmental Monitoring for Health Protection (EMHP) programme, led by the UKHSA. The sampling strategy provides coverage of approximately 40 million people across England. Samples were analysed for SARS-CoV-2 RNA by quantifying the number of copies of the nucleocapsid gene (N1) using RT-qPCR. Concentrations under the limit of detection were assigned a value of -4, to be handled during the data processing pipeline depending on the use case. Only sites sampled 30 times or more (around seven weeks' worth of data) were included; median sample count across sites was 145, ranging from 31 to 323.

2.2. Flow normalisation

Extraneous sources of flow, such as heavy rainfall, snow melt, or groundwater ingress into sewers, may dilute wastewater and impact estimates of SARS-CoV-2 RNA concentration. Studies have indicated that the effect of dilution in most cases are minor, but in periods of high dilution events, normalisation is critical [15]. The normalisation approach applied by the English wastewater surveillance programme mitigates this by adjusting measured SARS-CoV-2 concentrations to consider flow. The model is based on the assumption that flow F_t at time t is not directly observable. Instead, information about flow is obtained by observing the correlation of concentrations ρ_{ti} of different markers i (orthophosphate and ammonia nitrogen). The model assumes:

$$\log F_t \sim \text{Normal}\left(0, \lambda^2\right) \tag{2.1}$$

$$\log x_{ti} \sim \text{Normal}\left(\mu_i, \sigma_i^2\right) \tag{2.2}$$

$$\therefore \log \rho_{ti} = \log x_{ti} - \log F_t \tag{2.3}$$

where λ^2 is the flow variance, x_{ti} is the load of marker i at time t, μ_i and σ_i^2 are the mean and variance of the load of marker i (all in log space). $\langle \log F_t \rangle$ is fixed at 0 to identify the model. Using multiple markers jointly to estimate flow variability can improve the accuracy of estimates [9, 16].

2.3. A left-censored dynamic linear model

In our model, the (unobserved) viral concentration signal X_t is modelled as a first-order random walk (RW1) process

$$X_t = X_{t-1} + \sigma\epsilon_t \tag{2.4}$$

where $\epsilon_t \sim N(0, 1)$ is an independent and identically distributed normal random variable for $t = 2, ..., n$. The measured concentrations Y_t^* are modelled by adding independent measurement noise to X_t:

$$Y_t^* = X_t + \tau\epsilon_t' \tag{2.5}$$

where the independent measurement error $\epsilon_t' \sim t_\nu$ has a Student t-distribution with ν degrees of freedom. The actually observed, censored data, are modelled by truncating Y_t^* at the known censoring threshold ℓ_t:

$$Y_t = \max(Y_t^*, \ell_t) \tag{2.6}$$

As samples are taken only four times a week, the vector of measurements $\mathbf{Y}$ contains data observed at a subset $\mathcal{T}$ of all n available time points. We infer the viral concentration X_t from $\mathbf{Y}$ by Bayesian inference [17], i.e. by calculating the posterior distributions of the latent state $X_1, \ldots, X_n$ and hyperparameters σ, τ and ν, conditional on $\mathbf{Y}$. The posterior distribution is given by:

$$\begin{aligned} & p(X_1, \ldots, X_n, \sigma, \tau, \nu|\mathbf{Y}) \\ & \propto \left[\prod_{t\in\mathcal{T}} p(Y_t|X_t, \tau, \nu)\right] \\ & \times p(X_1, \ldots, X_n|\sigma) \\ & \times p(\tau)p(\sigma)p(\nu) \end{aligned} \tag{2.7}$$

The first line on the right hand side of Eq 2.7 is determined by the distribution of independent measurement errors, and left-censoring, of the Y_t. The second term is determined by the RW1 time series model for the X_t. The distributions $p(\tau)$, $p(\sigma)$, and $p(\nu)$ in the last line are prior hyperparameter distributions: we specify uninformative uniform prior distributions for $\tau > 0$ and $\sigma > 0$, and a left-truncated Normal prior for ν, with prior expectation 3, prior variance 1, and truncated at 2. The parameters of the truncated Normal prior for ν were selected by simulation and based on subjective judgements about the likely magnitude of measurement errors. The (multiplicative) proportionality constant in Eq 2.7 is inferred by Markov-Chain Monte-Carlo (MCMC) using the Stan software [14].

The hyperparameters τ and ν of the measurement process can be interpreted as measurement error variance (larger τ's correspond to noisier measurements), and the tendency to generate outliers (smaller ν's generate greater deviations from measured viral concentrations). Posterior estimates of these parameters are thus interesting for diagnostic purposes, e.g. to identify anomalous sites.

2.4. Dynamic linear model (DLM)

The DLM was implemented in the open-source programming language Stan [14], which provides efficient sampling of probabilistic models via MCMC and other inference algorithms. Code specifying the model is provided in Supplementary Information Figure S17. MCMC convergence statistics for the fit examples shown in Figure 1 can also be found in the SI (Figure S4–S8, Tables S1–S4).

2.5. 10-fold cross-validation

10-fold cross-validation was performed on the data across the 286 sites that had at least 30 samples. For each iteration:

- Raw SARS-CoV-2 N1 gc/L (with no normalisation for flow) was used with a $\log_{10}$ transformation.
- Data were randomly split (90%/10%) into training and test sets
- Pre-existing missing values (days when samples were expected but were not collected) were included in the training set but not in the test set.
- The censoring threshold was set to a single value $\log_{10}(133.0$ gc/L) for simplicity. In reality the limit of quantification will vary across samples.
- Fit DLM, KS and MA models to training data (details below)
- generate estimates (MA) and posterior samples (DLM, KS) of $Y_{\text{pred},}t$ at times t that were left out during training

$Y_{\text{test},}t$ (the left-out observation data) are then compared to $Y_{\text{pred},}t$ inferred with the three methods, via mean squared error (MSE) and interval coverage.

$Y_{\text{pred},}t$ for the DLM were generated by using Stan to sample from the joint posterior distribution of $X_1, \ldots, X_n$ and hyperparameters σ, ν, τ, inferred from the training data. We then inferred posterior predictive samples $Y_{\text{pred},}t$ at times t left out during training by adding t-distributed measurement errors to posterior samples of X_t, and applying censoring if the sampled observation was below the censoring threshold.

$Y_{\text{pred},}t$ for KS was generated by taking the fitted parameters τ, p_{outlier}, $\mu_{\text{X-test}}$, $\sigma_{\text{X-test}}$. To get a posterior distribution on X_t 4000 samples were generated from a normal distribution with

$$X_t \sim \text{Normal}(\mu_{\text{X-test}}, \sigma_{\text{X-test}}) \tag{2.8}$$

To get O_t 4000 samples were generated from a binomial with

$$O_t \sim \text{Binomial}(1, p_{\text{outlier}}) \tag{2.9}$$

To get uncensored observations Y_{outliers} 4000 samples were taken from a uniform distribution

$$a = \min(Y_{\text{train}}) - 2\text{SD}_Y \tag{2.10}$$
$$b = \max(Y_{\text{train}}) + 2\text{SD}_Y \tag{2.11}$$
$$Y_{\text{outliers},}t \sim Unif(a, b) \tag{2.12}$$

To get $Y_{\text{pred},}t$ X_t is passed to a Normal distribution with scale τ

$$Y_{\text{pred},}t \sim \text{Normal}(X_t, \tau) \tag{2.13}$$

Simulating outliers in Y_{pred} was done by

$$Y_{\text{pred},}t = \begin{cases} Y_{\text{out},}t, & \text{if } O_t = 1. \\ Y_{\text{pred},}t, & \text{otherwise.} \end{cases} \tag{2.14}$$

Finally, $Y_{\text{pred},}t$ is censored at some limit l

$$Y_{\text{pred},}t = \max(Y_{\text{pred},}t, l) \tag{2.15}$$

2.6. Forward prediction

We then further validated model performance by testing how well it is able to predict 10 samples (2.5 weeks) ahead. We refit the model on all samples for all sites minus the final 10 samples, and then predict the left-out samples.

2.7. Exploratory analysis of ν and τ Outputs from the DLM

Analyses were performed using R statistical software (Version 4.1.3) to establish whether the DLM is more likely to observe data variability - characterised by ν and τ outputs - at sites that show greater concentrations of SARS-CoV-2 RNA in wastewater. For this purpose we regressed the median gene copies per litre (gc/L) obtained over all time ($\log_{10}$ transformed) against mean ν or τ, controlling for the standard deviation of ν and τ, respectively. We obtained the residuals from these linear regression models to identify sites where the ν and τ outputs from the model vary in excess of what is accountable to median gc/L and the posterior standard deviation of ν or τ. Residuals for both models were then mapped to the Lower Layer Super Output Area (LSOA) for a given site using the Simple Features (*sf*) package in R [18] (Figure 4).

3. Results and discussion

3.1. Model fit: Simulated data

To test the output of the DLM, we first simulated data by generating a random walk with variance parameter σ^2 to model the underlying state, which was then sampled with measurement error

parameters τ and ν. Exact values are provided in the code. Any values below a predefined limit l are set to the value of l. These synthetic data were then fit with the DLM. Supplementary Figure S1 shows that the underlying state X-true is tracked rather well by the X-smoothed estimate and lies within the inferred credible intervals, demonstrating that the model can reliably recover the underlying state from noisy observations in a synthetic dataset.

3.2. Model fit: Real data

We fit the DLM to data from 286 sewage treatment works across England, restricted to sites with greater than 30 samples present. Each site is sampled four times a week. Figure 1 shows a range of fitted sites selected, based upon their estimated parameters τ and ν, to illustrate model behaviour at the extreme ends of the spectrum, i.e. low ν and low τ (Figures 1b and 1d) or high ν and high τ (Figures 1a and 1c). Sites with high parameter values typically show low levels of SARS-CoV-2 (N1 gene gc/L, the target used to approximate viral concentration in the sample) recovery and more frequent censoring. More censoring leads to more estimation uncertainty (wider credible intervals) as less information is available to constrain viral concentration estimates. Conversely, sites with low parameter values generally correspond to high levels of SARS-CoV-2 recovery and less censoring, therefore providing more information and tighter credible intervals. Supplementary Figure S2 shows a strong positive correlation between ν and τ, and Figures 4b and S3b show a strong negative correlation between ν and τ in relation to the site's median viral RNA concentration ($\log_{10}$ (N1 gc/L)), respectively. We note that our model seems to produce realistic estimates of viral concentration during long periods of censoring, and on days where observations are missing entirely.

3.3. Model validation

Model performance was assessed by comparing the MSE produced by the DLM, KS and a seven-day centered MA over 10 folds of cross-validation (see Methods). The MA represents a simple way to remove noise from data, and is used here as a benchmark for comparison. All three models generated comparable MSE per site (Figure 2). However, the DLM and KS can estimate viral concentrations below the censoring threshold and, therefore, provide additional information on value for applications, such as case prevalence estimation (see Applications section). In addition both the DLM and KS provide useful parameters for quantifying uncertainty and outliers within the data (DLM: $\sigma/\tau/\nu$, KS: $\sigma/\tau/p_{\text{outlier}}$). This is particularly useful to identify sites generating unexpected data. So, while an MA scores equally well in terms of the MSE, the smoothing methods still confer additional advantages. A boxplot of the pairwise MSE differences, shown in Supplementary Figure S6, shows that the differences are not consistently better or worse for the DLM when compared to the KS or MA models.

As MSE assesses the accuracy of a single point estimate of the predictive distribution, it cannot inform on the reliability of the whole model distribution. In Figure 2, the coverage frequency of prediction intervals was used to characterise the reliability of the predictive distribution. Coverage frequency assesses how well the fitted model represents the variability of the data by analysing to what extent the observations could pass as a random sample from the predictive distribution. If observations and samples from the predictive distribution are statistically indistinguishable, we should expect a 90% chance that the observation is included in a 90% prediction interval derived from the predictive distribution. See Methods for information on calculating coverage. Figure 2b shows mean coverage

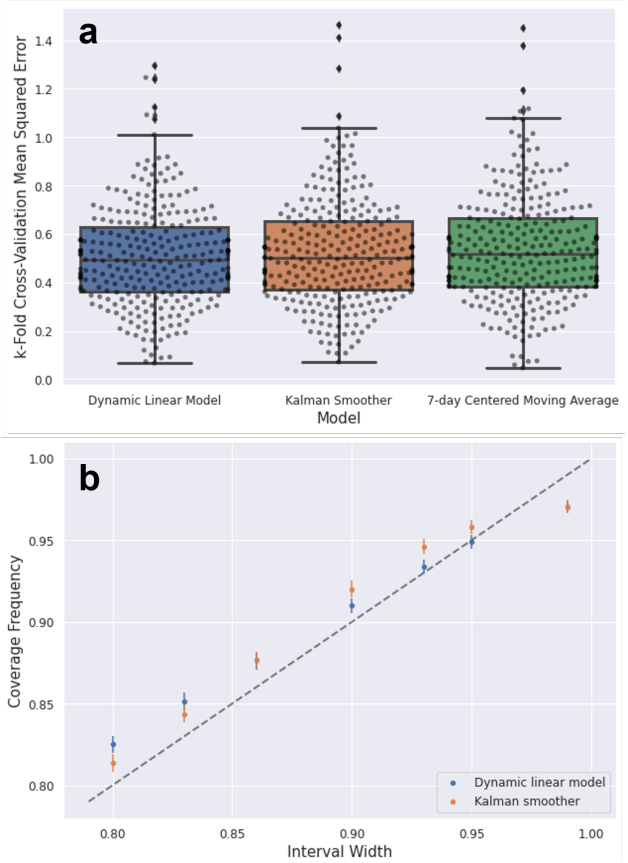

Figure 2. 10-fold cross-validation: comparison between a DLM, KS and seven-day centered MA. 286 sites with at least 30 samples were used. See methods for further details on k-fold CV methodology. **a.** Boxplots of MSE, with dots representing each site used in the CV. **b.** Calibration plot showing coverage of different intervals for both the DLM and KS (moving averages do not generate intervals of the predictive distribution). Each point represents mean coverage frequency for all 286 sites for that interval width; error bars show two times the standard error of the mean.

frequencies across all sites. For nominal interval widths between 0.8 and 0.95, the KS coverage frequencies lie above the dashed line indicating that the model intervals are slightly wider than the true interval and are thus slightly under-confident. For the DLM, the coverage is too wide below nominal values of 90% and appears more reliable between 0.90 and 0.95 than KS. Both models appear over-confident at nominal values above 95%. For additional information on the distributions coverage values see Figures S4 and S5.

Cross-validation was also performed for forward prediction by removing the last 10 samples and predicting them with either the DLM or KS. Figures S9 and S10 show that both the DLM and KS perform equally well at forecasting up to 10 days of samples.

The DLM performed equally as well as the KS in cross-validation, but with greater parsimony: we removed the Bernoulli outlier functionality, and autoregressive and offset parameters (η and δ), to specify a simpler model. By providing full Bayesian posterior information, the DLM offers

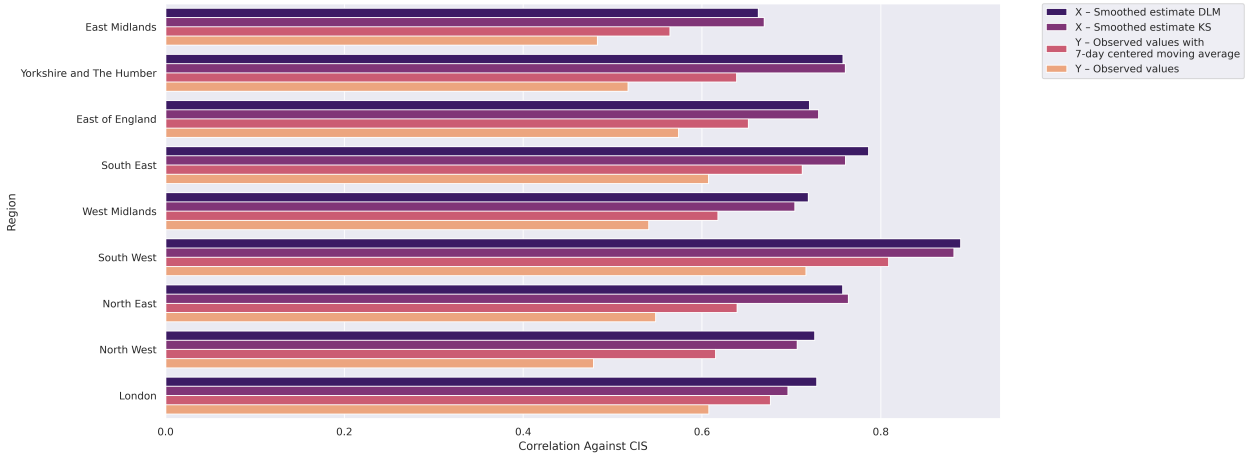

Figure 3. A bar chart showing the correlation against Coronavirus (COVID-19) infection survey (CIS) for: X – smoothed estimate DLM; sites in a give region are smoothed using the Dynamic Linear Model and then aggregated with a mean average; X – smoothed estimate KS; sites in a give region are smoothed using the Kalman Smoother and then aggregated with a mean average; Y – Observed $\log_{10}$ (N1 gc/L) with a 7-day centered moving average; sites are aggregated with a mean average and then the rolling average applied; and Y – Observed $\log_{10}$ (N1 gc/L); sites aggregated with a mean average and no additional transformation is applied. DLM and KS smoothed estimates improve the correlation of wastewater measurements against CIS in every region in England. All data used in this analysis were flow normalised prior to any additional manipulation, see methods for more information on flow normalisation.

more information on the distributions of all the parameters in the model, thereby facilitating greater quantification of model uncertainty. Furthermore, the Stan framework offers flexibility for modification of the underlying state model (e.g. AR(2) random walk) or the addition of autoregressive parameters, if desired. The MCMC inference algorithm provided in Stan also allowed the model to be estimated more than 10x faster than the Kalman Smoother: the mean runtime of the DLM for each fold in 10-fold cross-validation of 10 sites was 14 seconds compared to 155 seconds with the KS, although with known parameters the prediction speed by the KS is much improved. Results of the test are provided in Supplementary Table SS5, however in both cases the run times are small enough that we believe the difference is of little practical significance. The speed difference that is of more practical relevance (although difficult to quantify) is that our model was written in a general purpose modelling framework and so is easier to maintain, modify and adapt than the handcrafted R code of the Kalman Smoother. On the other hand, only the Kalman Smoother is able to quantify the probability of a given sample being an outlier and, therefore, this model will be more desirable for specific use cases. The DLM can only inform on whether a given sample lies outside of a predefined interval of the estimated underlying state, as shown in Figure 1.

4. Applications

4.1. Comparison of smoothed estimate with Coronavirus (COVID-19) infection survey

Work by multiple groups has shown that SARS-CoV-2 gc/l concentrations in wastewater measurements can track case prevalence ('positivity rate', the percentage of people who have tested positive for COVID-19 on a polymerase chain reaction (PCR) test at a point in time) [19–21]. In England, the latter has been measured by the Office for National Statistics' Coronavirus (COVID-19) Infection Survey (CIS), a randomised household survey that provides an estimate of disease prevalence at sub-regional, regional and national levels [22]. Therefore, smoothed estimates of $\log_{10}$(N1 gc/L) from a DLM or KS can be compared with flow-normalised raw estimates to establish which correlates more strongly with $\log_{10}$(CIS prevalence%) over time. Figure 3 compares correlations of CIS with (i) flow-normalised $\log_{10}$(N1 gc/L), (ii) flow-normalised $\log_{10}$(N1 gc/L) with a 7-day centered MA, and (iii) flow normalised $\log_{10}$(N1 gc/L) smoothed estimate of X for all nine English regions between 1st September 2020 and 1st March 2022. This time range includes a period in which wastewater RNA concentration rates decoupled from clinical measures of disease prevalence, of which the cause is unknown [23]. It is worth noting that this relationship is likely not deterministic, i.e. they are not equivalent and are subject to their own spatiotemporal variation and uncertainty that would manifest in significant changes in the ratio of the measures. Such observations have not been limited to England, and the cause is likely to have multiple factors, both epidemiological (i.e., changes in viral shedding distribution as circulating virus variants emerge and evolve) and metrological (e.g., degree of clinical testing coverage can be demographically biased; laboratory sensitivity can vary significantly with virus concentration method employed for wastewater analysis) [24–26].

Smoothed wastewater concentration rates using a DLM or KS correlate more strongly with CIS positivity rate than raw or averaged rates (Figure 3). The enhanced correlation performance of the DLM and KS is likely due to both models' ability to generate data from below the censoring limit. This assertion is supported by the comparison between the smoothed estimates improvement in correlations verses the averaged $\log_{10}$(N1 gc/L), which according to the MSE cross-validation should perform equally well. The key difference being that the DLM and KS infer values below the censored limit, thus we attribute at least part of the increase in correlation to this aspect of the models. Figure S9 provides a time series comparison of $\log_{10}$(CIS prevalence%), $\log_{10}$(N1 gc/L) , and smoothed estimates. Smoothed estimates show a specific advantage over raw $\log_{10}$(N1 gc/L) during times of low case prevalence. Using a simple sensitivity analysis to exclude the period in which wastewater concentration rates diverged from case rates to train the models, we find the same results (Figures S10–S11). DLM-smoothed rates therefore better complement CIS data, providing additional useful information for public health decision makers.

4.2. Exploration of fitted ν and τ parameters

The DLM provides two useful parameters for a given site: the extent to which outliers are observed (ν), with smaller values indicating greater frequency and size of outlier values, and the amount of measurement noise at a given site (τ), with larger values indicating noisier measurements. Figure 4a shows the geographical distribution of ν values for fitted sites mapped to each Lower Layer Super Output Area (LSOA) in England. There is some evidence of localised behaviour, with areas of

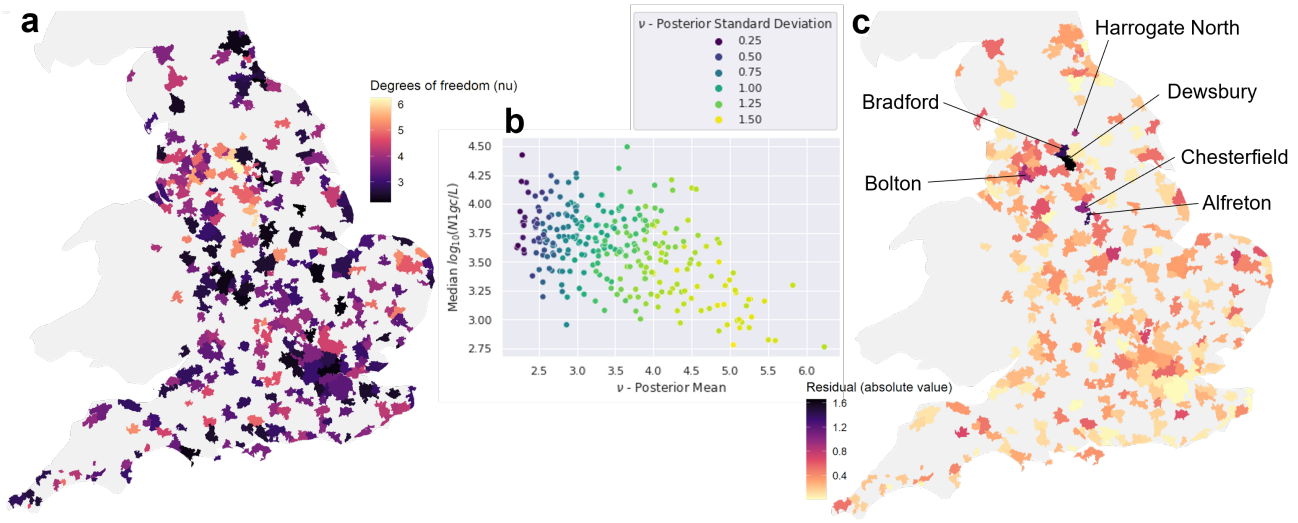

Figure 4. Exploratory analysis of fitted site parameter ν (degrees of freedom). **a.** Local Layer Super Output Area (LSOA) map of the posterior mean of ν. Smaller values of ν indicate heavier tails in the t-distribution and thus a stronger probability of outliers at a given site. **b.** Scatterplot showing the relationship between the posterior mean of ν, median flow-normalised viral RNA concentration $\log_{10}$(N1 gc/L), and the posterior standard deviation of ν. **c.** LSOA map of residuals from multivariable linear regression fit (posterior mean of $\nu \sim$ median($\log_{10}$(N1 gc/L)) + SD_ν). Values closer to 0 indicate that observed variance in ν is better explained by a linear combination of the posterior standard deviation of ν and median $\log_{10}$(N1 gc/L), while larger residuals indicate there are likely other factors driving the propensity of outliers. The approach may be useful to highlight sites or geographical areas with abnormal results. Six sites with the largest residuals are shown.

large ν in the North and East, and low values found in the West and London regions. However, interpretation of this map is challenging as ν is strongly related to median($\log_{10}$(N1 gc/L)) and the quality of fit, quantified here as the posterior standard deviation of ν (SD_ν (Figure 4b). To account for these relationships and draw more insight from the ν parameter we performed a multivariable linear regression analysis where median($\log_{10}$(N1 gc/L)) was regressed onto the mean of the posterior of ν, controlling for the standard deviation of the posterior of ν (see Methods). Figure 4c plots the absolute values of the regression model residuals (see Figure S12 for distribution of residuals); sites with the highest absolute residuals (i.e., the most variance not explained by either median($\log_{10}$(N1 gc/L)) or quality of fit) are clustered in the North West. We repeat this analysis for τ in Figure S3; again sites with the largest absolute residuals are concentrated in the North West, with additional large residuals seen in the South and East of England. Observed non-linearity is potentially attributable to high levels of censorship at low levels of median($\log_{10}$(N1 gc/L)). Future analyses should explore this suggestion, potentially with a censored regression model.

Further examples of sites with high and low parameter values are provided in Figures S13–S16.

5. Conclusions

We show that use of a Bayesian Dynamic Linear Model is a viable method for smoothing left-censored wastewater SARS-CoV-2 measurement data. Handling outliers through a t-distribution, rather than through an independent Bernoulli distribution, as applied in a previously published Kalman Smoother [12], is likely to more directly relate to the underlying state to be recovered. While the DLM and KS perform equivalently with mean squared error under cross-validation, the proposed DLM is more parsimonious (fewer model parameters), has a faster computational time, and is implemented in a more flexible modelling framework, allowing for easier modifications. Additionally, the DLM produces two site-specific parameters, ν and τ, which are able to highlight sites with variable performance. This can be useful when assessing sampling strategies applied at scale (e.g. national or regional surveillance). Sites identified as providing inconsistent, noisy, or low information data may be removed from multi-site monitoring campaigns, for example.

The smoothed data, using our method, more closely correlate with regional infection survey data (CIS) than untransformed raw measurements. Wastewater data, smoothed in this fashion, are therefore more robust, capable of better complementing traditional surveillance, and providing additional confidence and utility for public health decision making.

Nevertheless, our approach has some limitations. The limit of censorship was set to a single value during cross-validation $\log_{10}(133.0\ \text{gc/L})$, for simplicity. In reality this limit can vary across samples. From September 2021 SARS-CoV-2 RNA measurements from English wastewater diverged from reported clinical data where it had been previously tracking it. The reason why has still not been established but is potentially attributable to differential shedding rates between variants. Our sensitivity analyses reported in the Supplementary material found this does not impact the performance of our model.

Acknowledgments

The United Kingdom Government (Department of Health and Social Care) funded the sampling, testing, and data analysis of wastewater in England. Obépine funded the work of Marie Courbariaux and provided the R code of the modified Kalman Smoother and support to use it.

Conflict of Interest

The authors declare no conflicting interests in this paper.

References

1. F. Balloux, *Mass COVID testing and sequencing is unsustainable – here's how future surveillance can be done*, The Conversation, 2022. Available from: https://theconversation.com/mass-covid-testing-and-sequencing-is-unsustainable-heres-how-future-surveillance-can-be-done-177404.
2. M. J. Wade, D. Jones, A. Singer, A. Hart, A. Corbishley, C. Spence, et al., *Wastewater COVID-19 monitoring in the UK: summary for SAGE*, 2020. Available from:

https://www.gov.uk/government/publications/defrajbc-wastewater-covid-19-monitoring-in-the-uk-summary-19-november-2020.

3. A. Bivins, D. North, A. Ahmad, W. Ahmed, E. Alm, F. Been, et al., Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight Against COVID-19, *Environ. Sci. Technol.,* **54** (2020), 7754–7757. https://doi.org/10.1021/acs.est.0c02388
4. UC Merced Researchers, *COVIDPoops19: Summary of Global SARS-CoV-2 Wastewater Monitoring Efforts*, 2022. Available from: https://www.arcgis.com/apps/dashboards/c778145ea5bb4daeb58d31afee389082.
5. H. R. Safford, K. Shapiro, H. N. Bischel, Wastewater analysis can be a powerful public health tool - if it's done sensibly, *Proc. Natl. Acad. Sci.,* **119** (2022), e2119600119. https://doi.org/10.1073/pnas.2119600119
6. D. A. Larsen, H. Green, M. B. Collins, B. L. Kmush, Wastewater monitoring, surveillance and epidemiology: a review of terminology for a common understanding, *FEMS Microbes,* **2** (2021), xtab011. https://doi.org/10.1093/femsmc/xtab011
7. N. Sims, B. Kasprzyk-Hordern, Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level, *Environ. Int.,* **139** (2020), 105689. https://doi.org/10.1016/j.envint.2020.105689
8. M. Huizere, T. L. ter Lak, P. de Voogt, A. P. van Wezel, Wastewater-based epidemiology for illicit drugs: A critical review on global data, *Water Res.,* **207** (2021), 117789. https://doi.org/10.1016/j.watres.2021.117789
9. M. J. Wade, A. Lo Jacomo, E. Armenise, M. R. Brown, J. T. Bunce, G. J. Cameron, et al., Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes, *J. Hazard. Mater.,* **424** (2022), 127456. https://doi.org/10.1002/essoar.10507606.1
10. M. A. Cohen, P. B. Ryan, Observations Less than the Analytical Limit of Detection: A New Approach, *JAPCA,* **39** (1989), 328–329. https://doi.org/10.1080/08940630.1989.10466534
11. D. R. Helsel, Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it, *Chemosphere*, **65** (2006), 2434–2439. https://doi.org/10.1016/j.chemosphere.2006.04.051
12. M. Courbariaux, N. Cluzel, S. Wang, V. Maréchal, L. Moulin, S. Wurtzer, et al., A Flexible Smoother Adapted to Censored Data With Outliers and Its Application to SARS-CoV-2 Monitoring in Wastewater, *Front. Appl. Math. Stat.,* **8** (2022), 836349. https://doi.org/10.3389/fams.2022.836349
13. S. Wurtzer, P. Waldman, M. Levert, N. Cluzel, J. L. Almayrac, C. Charpentier, et al., 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. Tot. Environ.,* **810** (2022), 152213. https://doi.org/10.1016/j.scitotenv.2021.152213
14. Stan Development Team, *Stan Modeling Language User's Guide and Reference Manual*, Version 2.30, 2022.

15. C. Sweetapple, M. J. Wade, J. M. S. Grimsley, J. T. Bunce, P. Melville-Shreeve, A. S. Chen, Dynamic population normalisation in wastewater-based epidemiology for improved understanding of the SARS-CoV-2 prevalence: a multi-site study, *J. Water Health,* (2023), in press. https://doi.org/10.2166/wh.2023.318

16. A. L. Rainey, S. Liang, J. H. Bisesi Jr., T. Sabo-Attwood, A. T. Maurelli, A multistate assessment of population normalization factors for wastewater-based epidemiology of COVID-19, *PLOS ONE,* **18** (2023), e0284370. https://doi.org/10.1371/journal.pone.0284370

17. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin *Bayesian Data Analysis*, 3rd Ed. CRC Press, 2013. https://doi.org/10.1201/b16018

18. E. Pebesma, Simple Features for R: Standardized Support for Spatial Vector Data, *The R Journal,* **10** (2018), 439–446. https://doi.org/10.32614/RJ-2018-009

19. M. Morvan, A. Lo Jacomo, C. Souque, M. J. Wade, T. Hoffmann, K. Pouwels, et al., An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence, *Nat. Commun.,* **13** (2022), 4313. https://doi.org/10.1038/s41467-022-31753-y

20. C. S. McMahan, S. Self, L. Rennert, C. Kalbaugh, D. Kriebel, D. Graves, et al., COVID-19 wastewater epidemiology: a model to estimate infected populations, *Lancet Planet. Health,* **5** (2021), e874–881. https://doi.org/10.1016/S2542-5196(21)00230-8

21. X. Li, J. Kulandaivelu, S. Zhang, J. Shi, M. Sivakumar, J. Mueller, et al., Data-driven estimation of COVID-19 community prevalence through wastewater-based epidemiology, *Sci. Total Environ.,* **789** (2021), 147947. https://doi.org/10.1016/j.scitotenv.2021.147947

22. Office for National Statistics, *Coronavirus (COVID-19) Infection Survey: methods and further information*, 2022. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#coronavirus-covid-19-infection-survey.

23. UK Health Security Agency, *EMHP wastewater monitoring of SARS-CoV-2 in England: 15 July 2020 to 30 March 2022*, 2022. Available from: https://www.gov.uk/government/publications/monitoring-of-sars-cov-2-rna-in-england-wastewater-monthly-statistics-15-july-2020-to-30-march-2022/emhp-wastewater-monitoring-of-sars-cov-2-in-england-15-july-2020-to-30-march-2022.

24. G. Vogel, Signals from the sewer, *Science,* **375** (2022), 1100–1104. https://doi.org/10.1126/science.adb1874

25. A. Xiao, F. Wu, M. Bushman, J. Zhang, M. Imakaev, P. R. Chai, et al., Metrics to relate COVID-19 wastewater data to clinical testing dynamics, *Water Res.,* **212** (2022), 118070. https://doi.org/10.1016/j.watres.2022.118070

26. P. M. D'Aoust, X. Tian, S. Tasneem Towhid, A. Xiao, E. Mercier, N. Hegazy, et al., Wastewater to clinical case (WC) ratio of COVID-19 identifies insufficient clinical testing, onset of new variants of concern and population immunity in urban communities, *Sci. Total Environ.,* **853** (2022), 158547. https://doi.org/10.1016/j.scitotenv.2022.158547

Supplementary Information

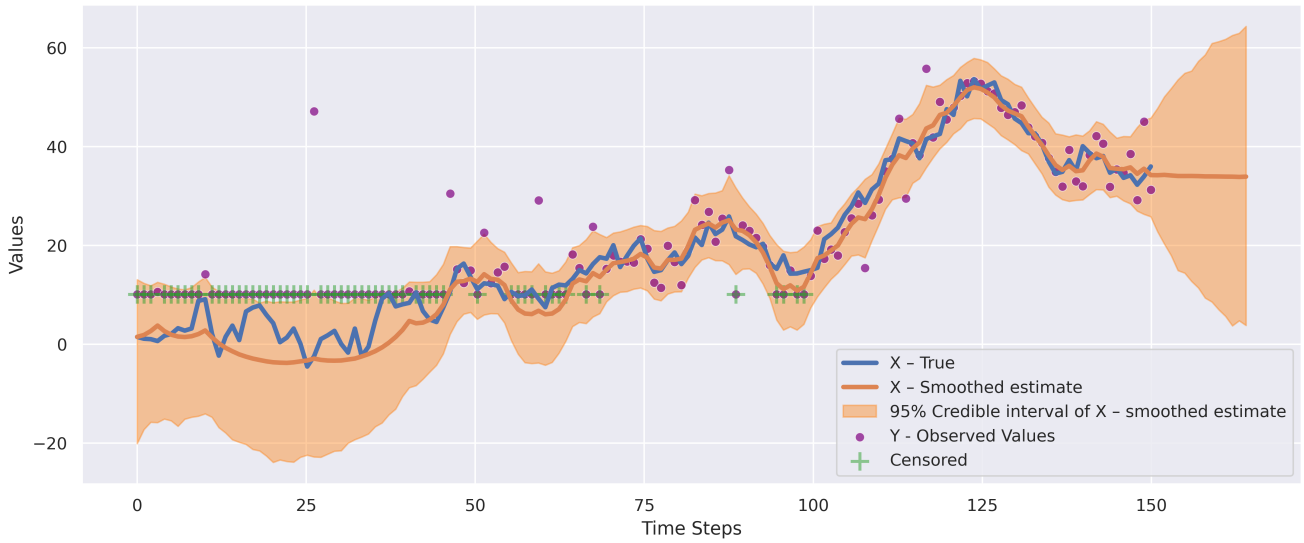

Figure S1. Fit of the proposed dynamic linear model on synthetic data (see Methods in the main text for a description of how these data were generated). In blue is the underlying true process (X - True), in a real-world situation X would be latent and unobservable. X is sampled giving observed values shown in purple (Y – Observed values), some of which are censored at some limit l (green crosses). The fit of the dynamic linear model is shown in orange with the bold line being the mean of the posterior of the estimated X (X – smoothed estimate) and light orange is the 95% Bayesian credible interval of the posterior of estimated X. Note at time step n > 150 the model is predicting X for 14 time steps forward, which is why the credible intervals expand over this period.

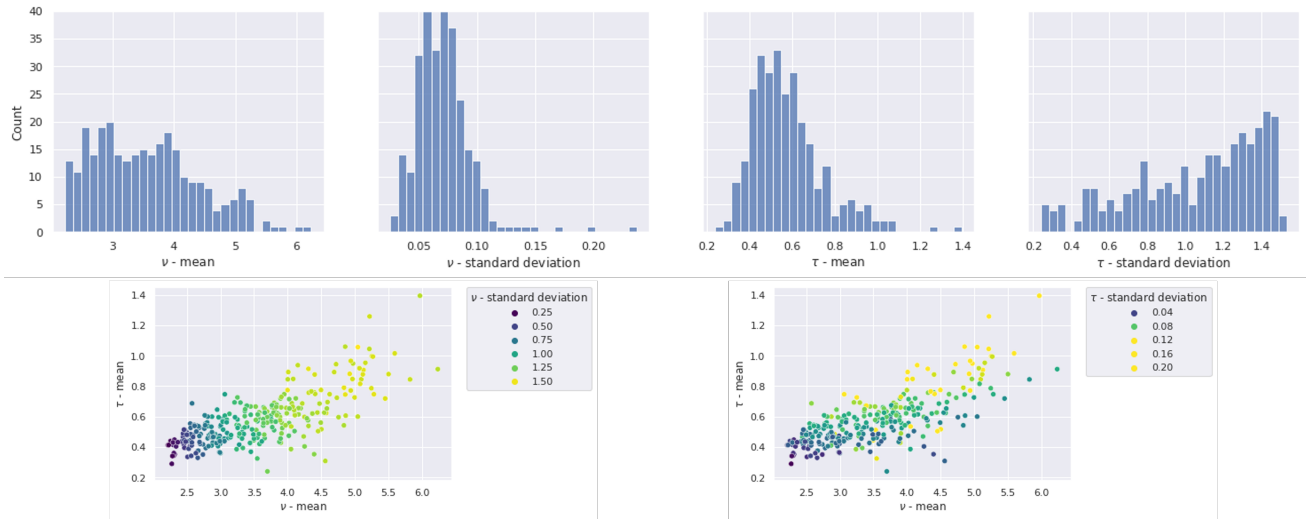

Figure S2. Posterior means and standard deviations (and the relationships between them) of fitted site parameters ν and τ.

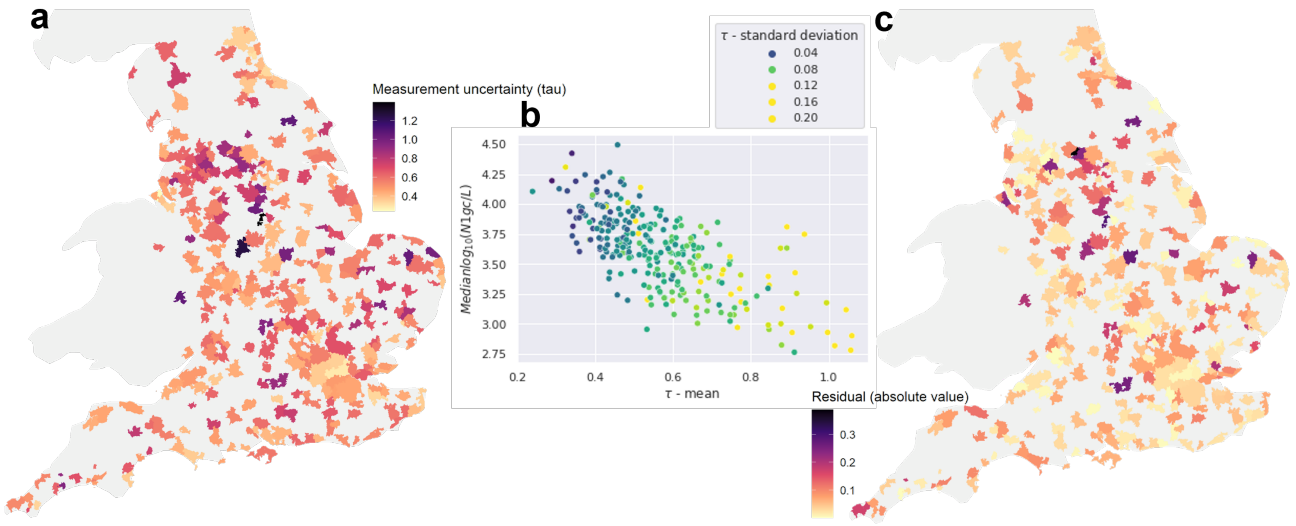

Figure S3. Exploratory analysis of fitted site parameter τ (measurement noise). **a.** Local Layer Super Output Area (LSOA) map of the posterior mean of τ. Larger values of τ indicate a wider t-distribution and thus more measurement noise at a given site. **b.** Scatterplot showing the relationship between the posterior mean of τ, median flow-normalised viral RNA concentration $\log_{10}$(N1 gc/L), and the posterior standard deviation of τ. **c.** LSOA map of residuals from multivariable linear regression fit (posterior mean of $\tau \sim$ median($\log_{10}$(N1 gc/L)) + SD_τ). Values closer to 0 indicate that observed variance in τ is better explained by a linear combination of the posterior standard deviation of τ and median $\log_{10}$(N1 gc/L), while larger residuals indicate there are likely other factors driving the fitted τ parameter.

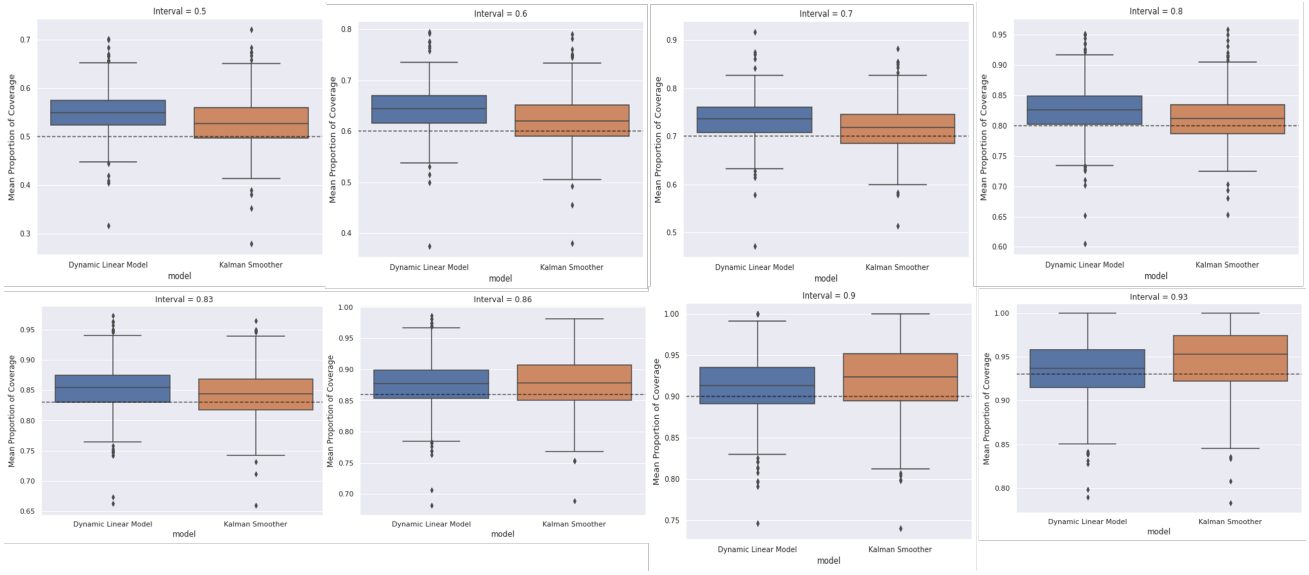

Figure S4. Mean proportion of coverage for the dynamic linear model versus the Kalman smoother at eight different intervals (0.50 to 0.93). The dashed line is the expected proportion of coverage at this interval. Boxplot medians above and below this line indicate overestimation and underestimation of the uncertainty respectively. Both models consistently overestimate; their coverage proportions for each interval do not differ significantly.

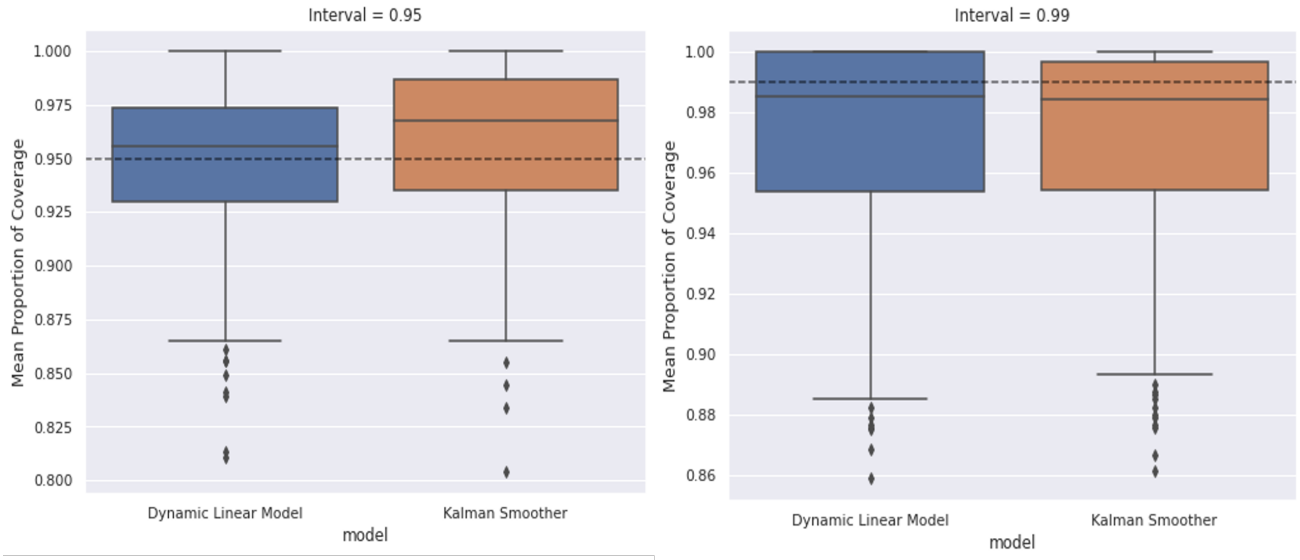

Figure S5. Mean proportion of coverage for the dynamic linear model versus the Kalman smoother at two different intervals (0.95 and 0.99). The dashed line is the expected proportion of coverage at this interval. Boxplot medians above and below this line indicate overestimation and underestimation of the uncertainty respectively. The DLM more closely matches the expected value at a lower interval of 0.95, although their confidence intervals overlap.

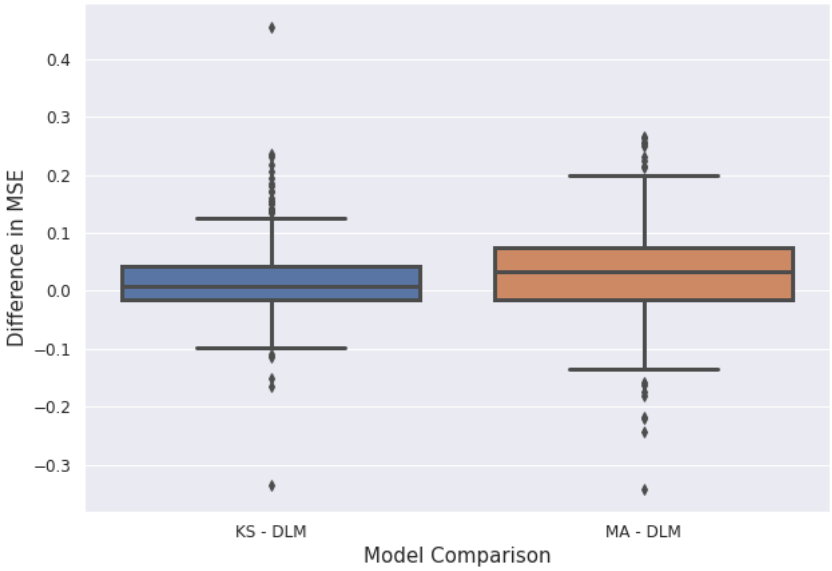

Figure S6. Pairwise differences in mean squared error (MSE) between Kalman Smoother and DLM (blue) and between simple moving average and DLM (orange). Although the median MSE is positive in both cases representing slight superiority of the DLM, this should not be considered significant.

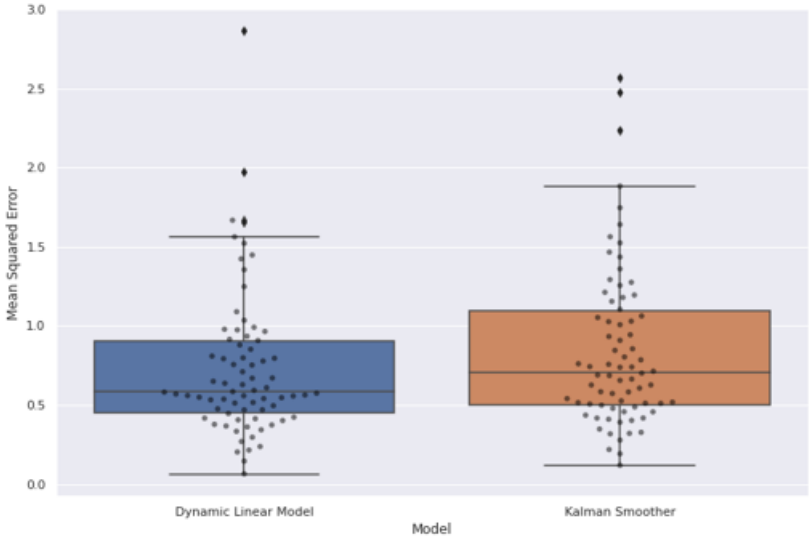

Figure S7. Predictions over a period of growth: 15th June 2021 to 4th July 2021. 20-sample forward prediction cross-validation on all sites with more than 30 samples for 74 sites. Each site is a point. The dynamic linear model (left) produces a lower MSE in 10-fold cross-validation than the Kalman smoother, although the confidence intervals of the two models overlap.

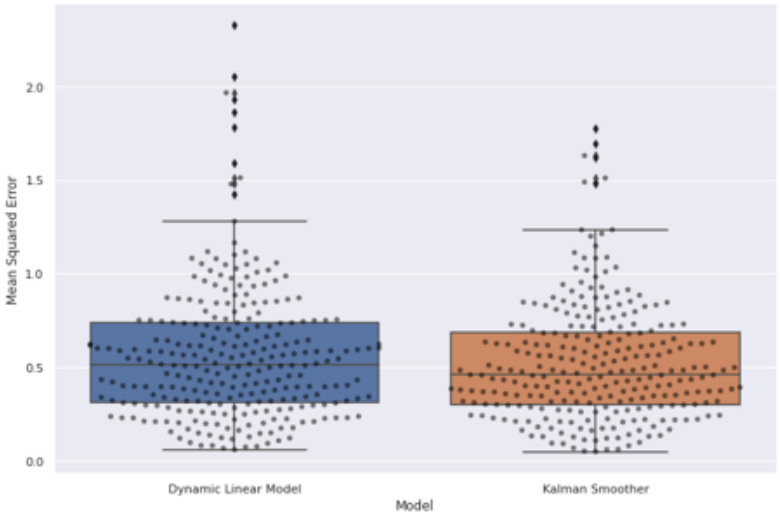

Figure S8. Predictions over a stable period of high wastewater concentrations: 1st January 2021 to 20th January 2021. 20-sample forward prediction cross-validation on 268 sites with more than 30 samples and data available until 20th January 2021. Each site is a point. The Kalman smoother (right) produces a slightly smaller lower MSE in 10-fold cross-validation than the dynamic linear model, although the confidence intervals of the two models overlap.

Figure S9. Regional time-series plot of smoothed estimate of wastewater concentration (gc/L) (red) and case prevalence established via the Coronavirus (COVID-19) Infection Survey (green). Raw log-10 N1 gc/L over time is provided in blue. These time series plots cover the period in which wastewater concentrations diverged from case rates in September 2021 onwards.

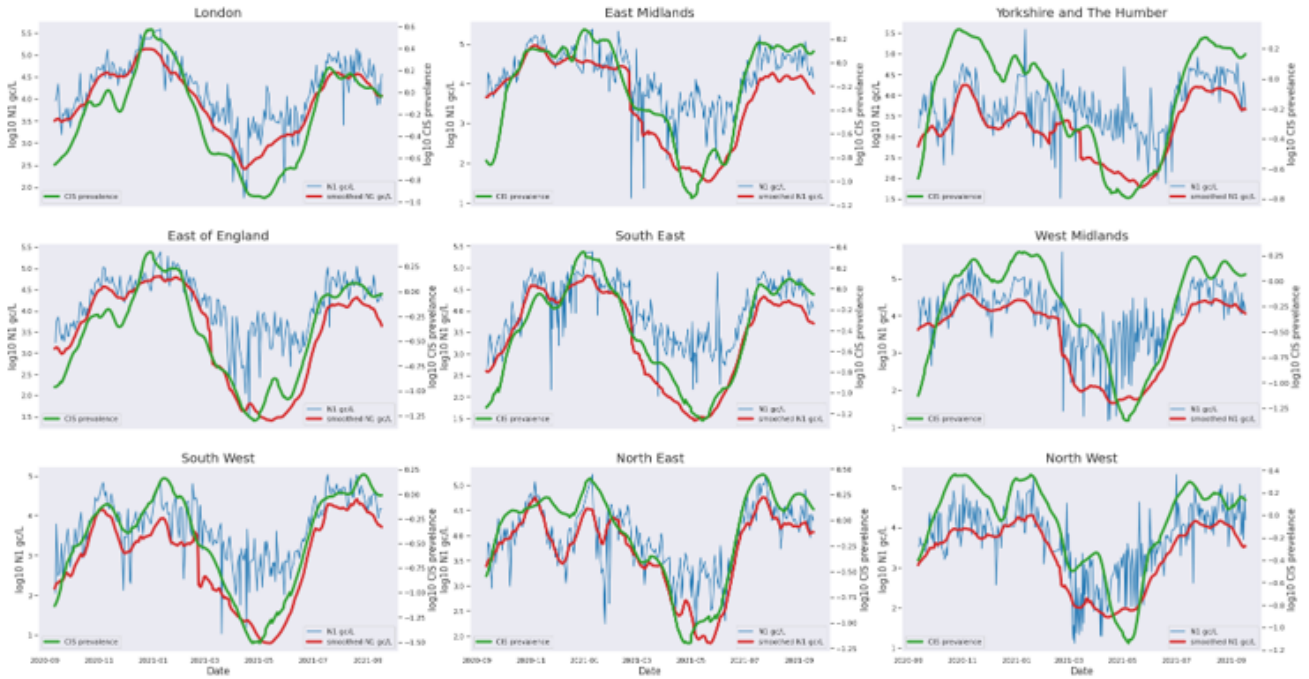

Figure S10. To assess the sensitivity of the model to the divergent period beyond 19th September 2021 the model was refit to exclude the period where wastewater concentrations diverged from case rates in this period (253 sites).

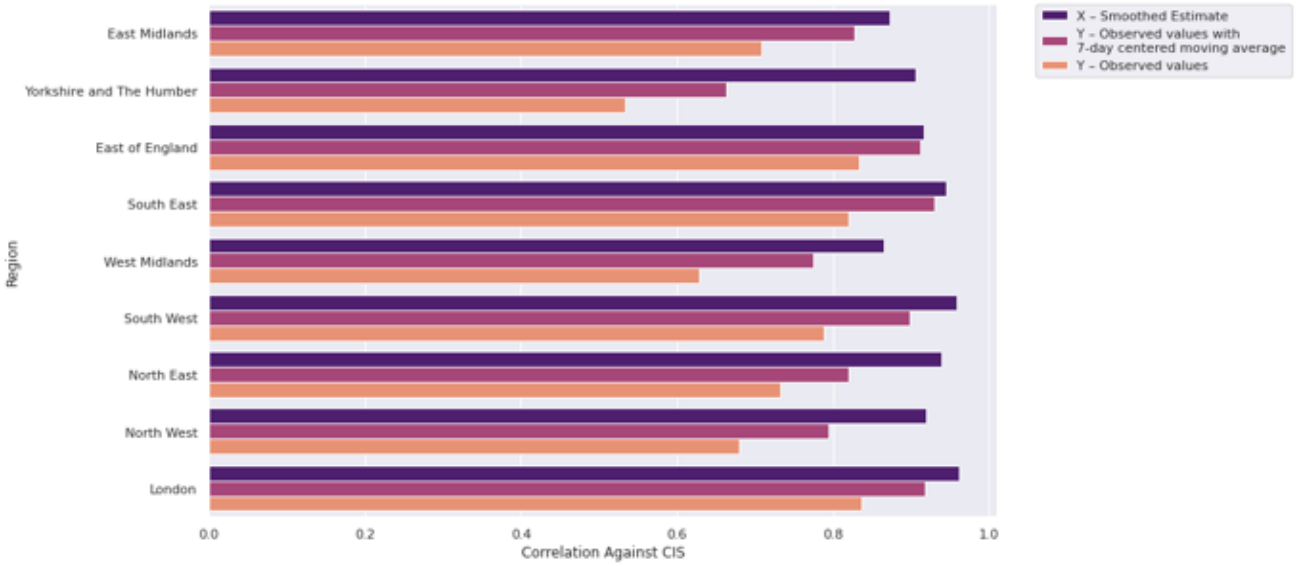

Figure S11. Region-level correlations of raw (orange), averaged (pink) and DLM-smoothed (purple) wastewater concentration data with Coronavirus (COVID-19) Infection Survey (CIS) data. In all regions, to varying degrees, smoothing time series data via the proposed DLM improves correlations between wastewater and CIS data. Note that this chart shows, and the smoothing model is trained on, data up until 17th September 2021, before the period where wastewater concentration trends stopped tracking case prevalence.

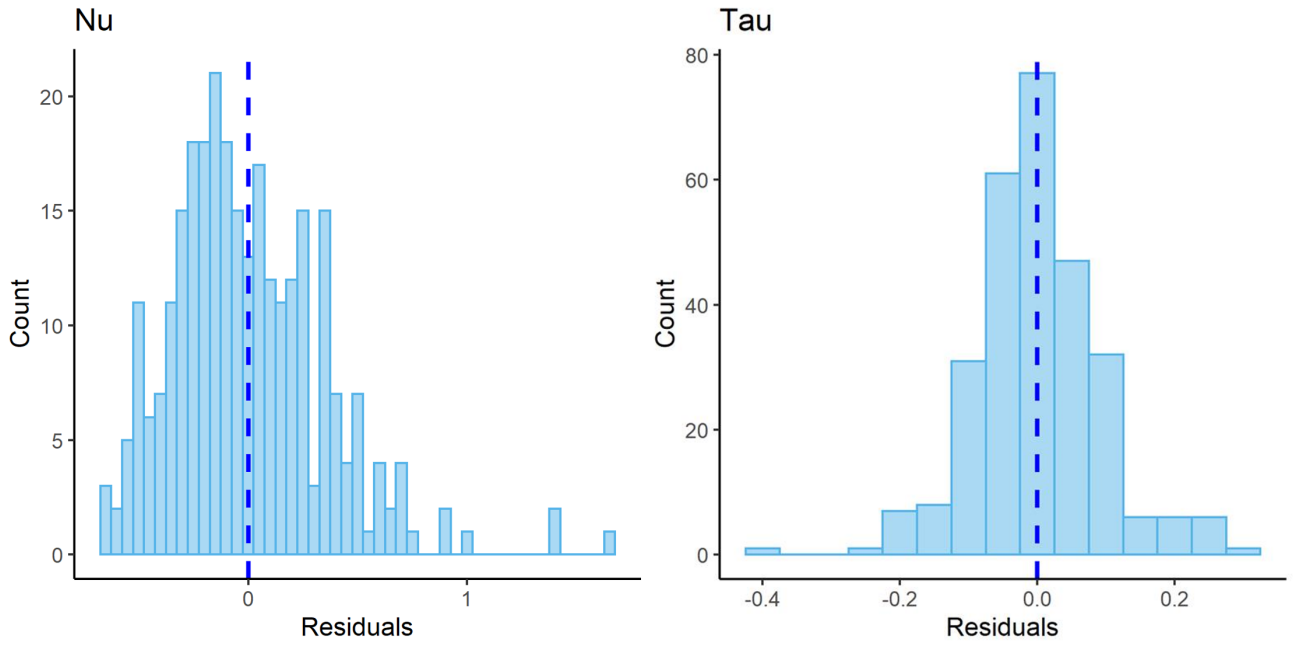

Figure S12. Histograms showing the distribution of residuals for ν and τ from multivariable linear regression.

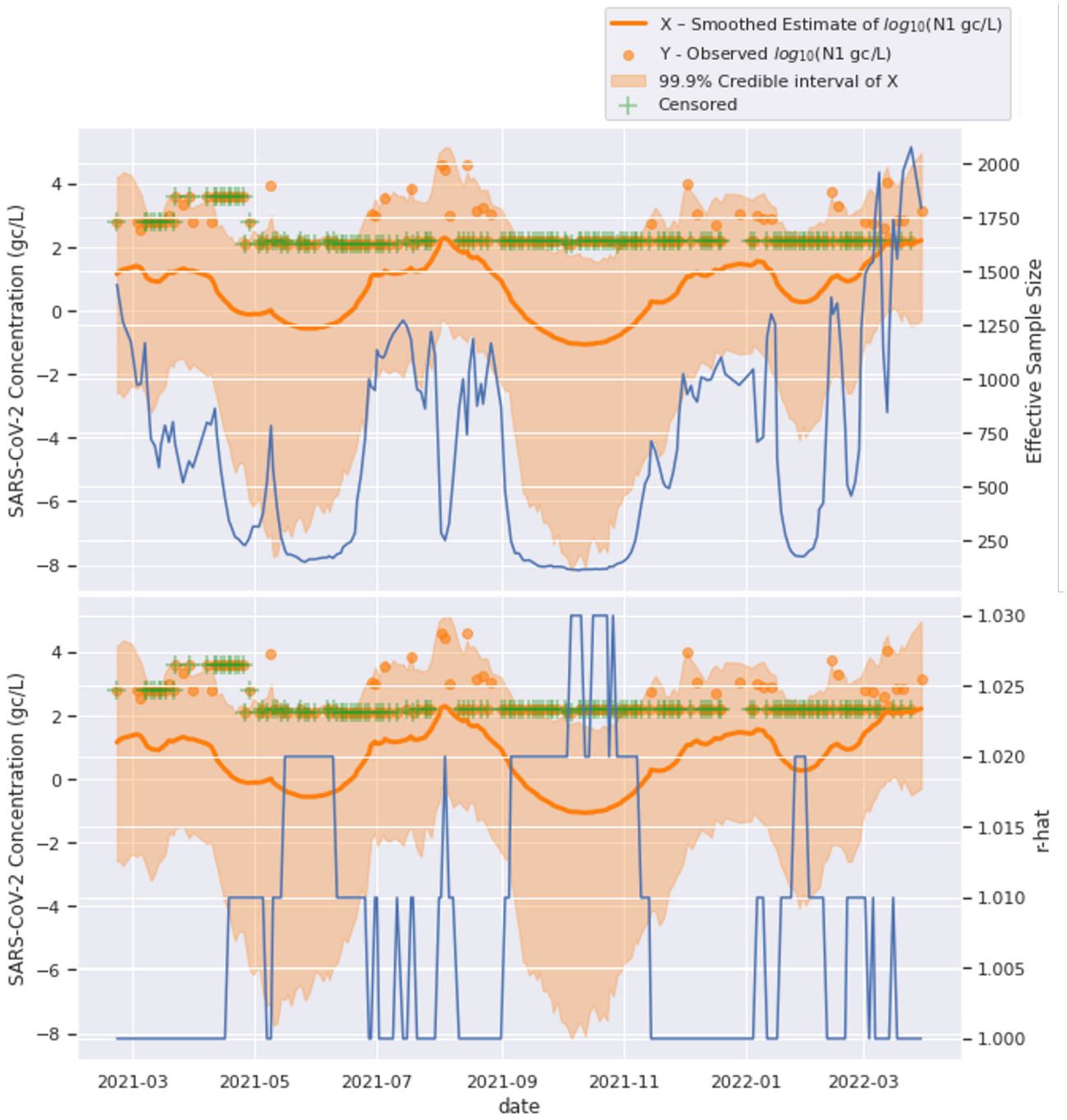

Figure S13. Example of a site with high ν and high τ: Burton upon Trent. The blue lines, Effective Sample Size (ESS, upper) and $\hat{r}$ (lower) correspond to the secondary Y-axes on the right. ESS and $\hat{r}$ are efficiency and convergence diagnostic statistics for Markov Chains.

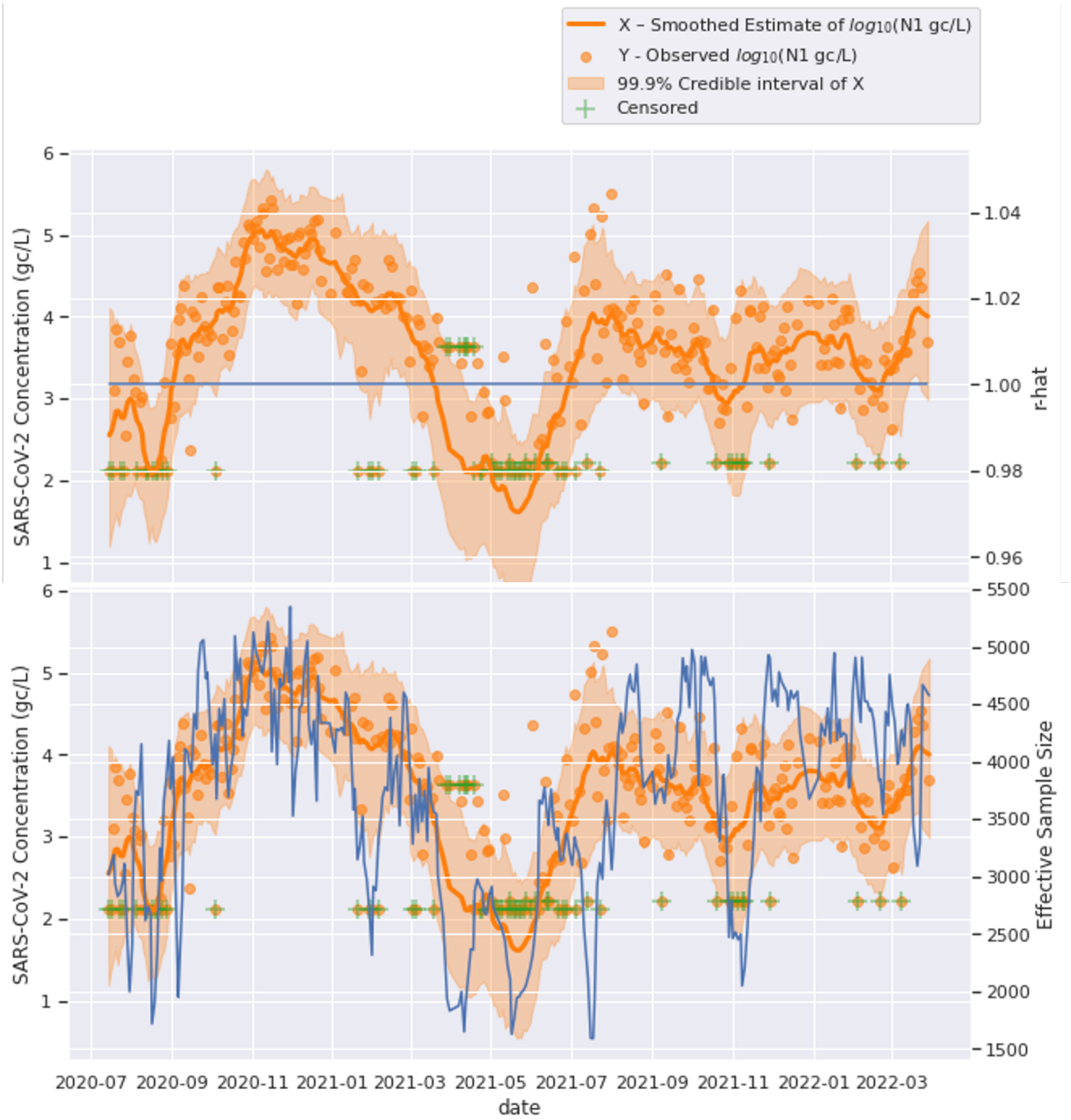

Figure S14. Example of a site with low ν and low τ: Lincoln. The blue lines correspond to the secondary Y-axes on the right.

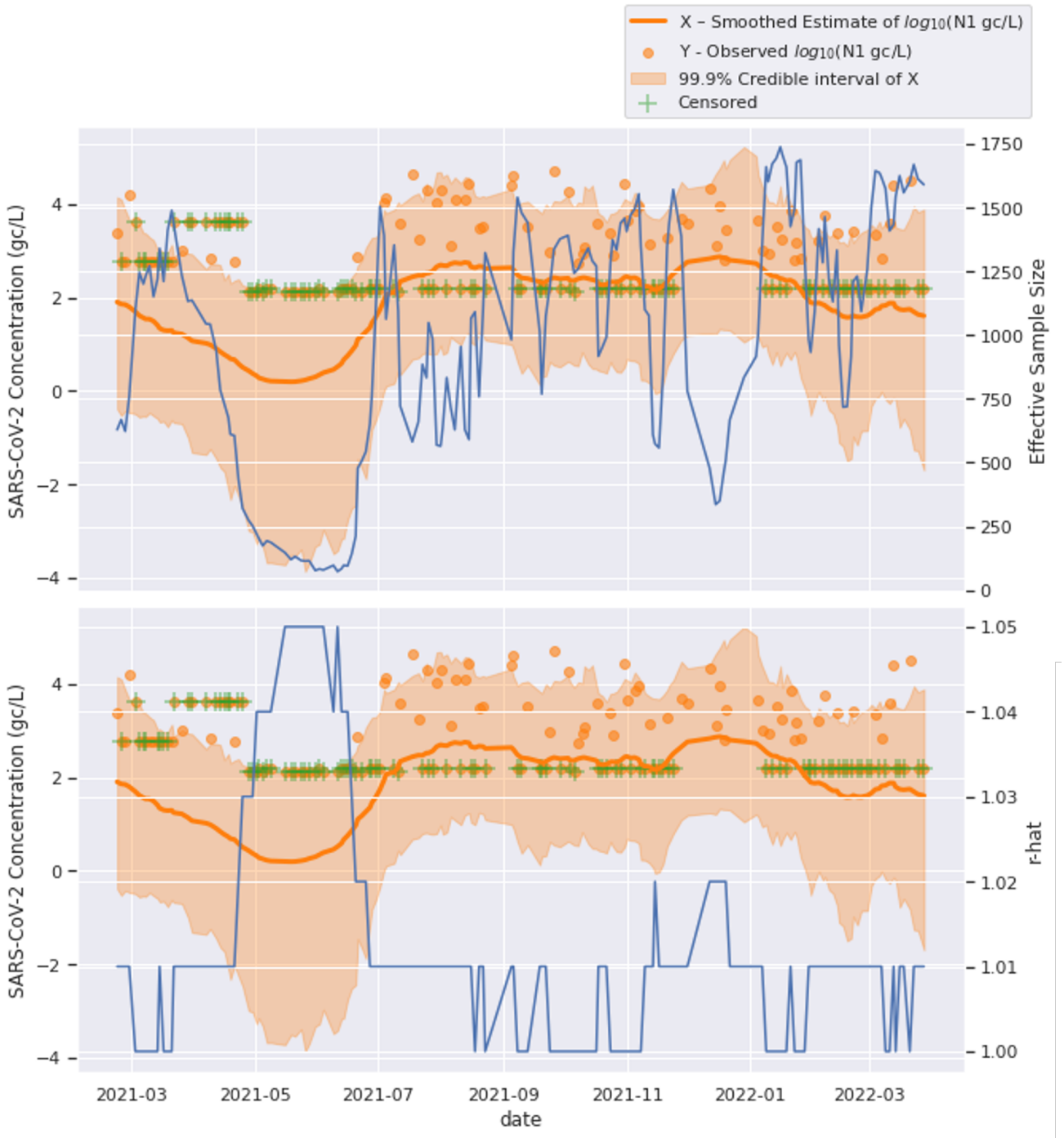

Figure S15. Example of a site with high ν and high τ: Alfreton. The blue lines correspond to the secondary Y-axes on the right.

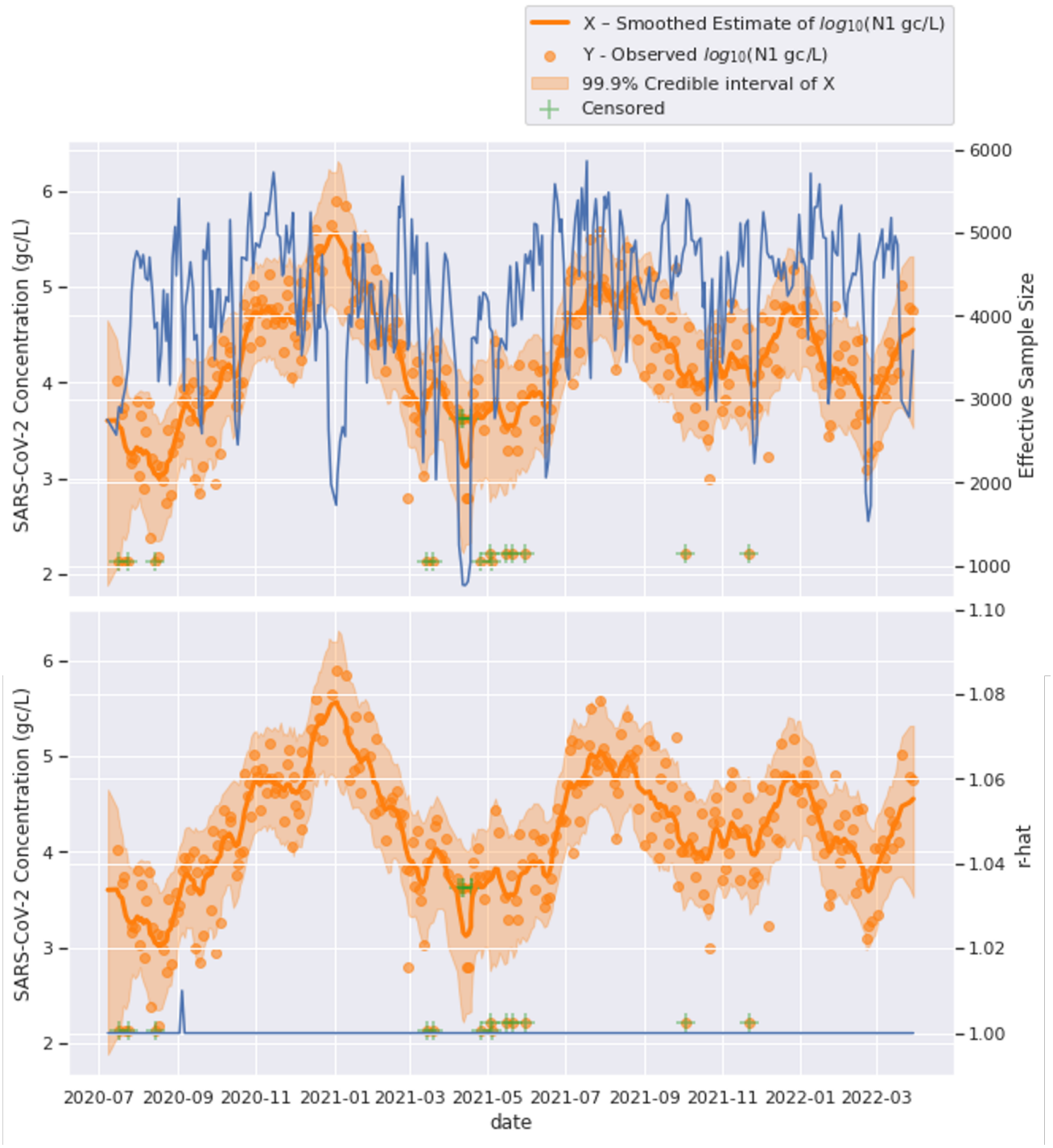

Figure S16. Example of a site with low ν and low τ: Beckton. The blue lines correspond to the secondary Y-axes on the right.

```
data {
  int<lower=0> N_obs;  // no. of uncensored obs without nans
  int<lower=0> N_cens; // no. of censored obs
  int<lower=0> N;      // total no. of obs
  real y_obs[N_obs];   // vector of uncensored obs
  int i_obs[N_obs];    // indices of uncensored obs
  int i_cens[N_cens];  // indices of censored obs
  real l;              // constant censoring threshold
}


parameters {
  real<upper=l> y_cens[N_cens]; // values of censored obs
                                // (with hard upper limit l)
  real X[N];                    // state variable
  real<lower=0> sigma;          // evolution variance
  real<lower=0> tau;            // obs noise scale parameter
  real<lower=2> nu;             // obs noise degrees of freedom
}

model {
  for (i in 2:N) {              // state evolution
    X[i] ~ normal(X[i-1], sigma);
  }
  for (i in 1:N_obs) {          // uncensored observations
    y_obs[i] ~ student_t(nu, X[i_obs[i]], tau);
  }
  for (i in 1:N_cens) {         // censored observations
    y_cens[i] ~ student_t(nu, X[i_cens[i]], tau);
  }
  nu ~ normal(3, 2);             // informative prior on d.o.f.
}
```

Figure S17. Stan model code specifying the proposed dynamic linear model.

Table S1. MCMC convergence statistics for Burton upon Trent. SD: Standard Deviation, HDI: Highest Density Interval, MCSE: Markov Chain Standard Error, ESS: Effective Sample Size.

	Mean	SD	HDI 3%	HDI 97%	MCSE Mean	MCSE SD	ESS Bulk	ESS tail	$\hat{r}$
σ	0.394	0.155	0.150	0.677	0.023	0.016	42.0	154.0	1.08
τ	1.251	0.241	0.851	1.731	0.011	0.008	523.0	906.0	1.00
ν	5.260	1.397	2.667	7.737	0.022	0.015	3687.0	2701.0	1.00

Table S2. MCMC convergence statistics for Lincoln. SD: Standard Deviation, HDI: Highest Density Interval, MCSE: Markov Chain Standard Error, ESS: Effective Sample Size.

	Mean	SD	HDI 3%	HDI 97%	MCSE Mean	MCSE SD	ESS Bulk	ESS tail	$\hat{r}$
σ	0.189	0.027	0.140	0.240	0.002	0.001	206.0	447.0	1.01
τ	0.422	0.039	0.356	0.502	0.001	0.001	2849.0	3718.0	1.00
ν	2.254	0.276	2.000	2.737	0.005	0.003	3155.0	2741.0	1.00

Table S3. MCMC convergence statistics for Alfreton. SD: Standard Deviation, HDI: Highest Density Interval, MCSE: Markov Chain Standard Error, ESS: Effective Sample Size.

	Mean	SD	HDI 3%	HDI 97%	MCSE Mean	MCSE SD	ESS Bulk	ESS tail	$\hat{r}$
σ	0.242	0.097	0.057	0.409	0.014	0.010	44.0	22.0	1.13
τ	1.355	0.159	1.064	1.650	0.005	0.003	1046.0	2012.0	1.00
ν	5.911	1.43v0	3.477	8.745	0.025	0.017	3157.0	2205.0	1.00

Table S4. MCMC convergence statistics for London Beckton. SD: Standard Deviation, HDI: Highest Density Interval, MCSE: Markov Chain Standard Error, ESS: Effective Sample Size.

	Mean	SD	HDI 3%	HDI 97%	MCSE Mean	MCSE SD	ESS Bulk	ESS tail	$\hat{r}$
σ	0.153	0.024	0.110	0.198	0.002	0.001	193.0	358.0	1.01
τ	0.288	0.026	0.237	0.333	0.001	0.000	1569.0	3334.0	1.00
ν	2.283	0.273	2.000	2.758	0.005	0.003	2761.0	2587.0	1.00

Table S5. Mean run times in seconds for model estimation of each fold in a 10-fold cross-validation. N Train: mean number of train datapoints. Tests were performed on Amazon AWS Sagemaker ml.t2.2xlarge notebook instance.

Site Code	N Train	Mean DLM run time	Mean KS run time
UKENAN_AW_TP000004	199	14.8	169.9
UKENAN_AW_TP000012	203	10.7	135.8
UKENAN_AW_TP000015	203	16.1	174.1
UKENAN_AW_TP000016	206	13.5	166.6
UKENAN_AW_TP000023	202	16.4	155.3
UKENAN_AW_TP000026	192	11.2	131.4
UKENAN_AW_TP000028	203	18.7	184
UKENAN_AW_TP000029	202	15.1	160.2
UKENAN_AW_TP000037	205	14.8	126.6
UKENAN_AW_TP000041	201	12.4	149.4
mean	201.6	14.37	155.33

Table S6. Date ranges for all wastewater treatment sites used in study. Showing site code, minimum date (date min), maximum date (date max), site reporting name.

ww_site_code	date_min	date_max	site_reporting_name
UKENNE_YW_TP000095	06/07/2020	30/03/2022	Hull
UKENTH_TWU_TP000054	08/07/2020	30/03/2022	London (Deepham)
UKENSW_SWS_TP000058	08/07/2020	27/03/2022	Plymouth
UKENTH_TWU_TP000010	08/07/2020	25/03/2022	Aylesbury
UKENTH_TWU_TP000013	08/07/2020	30/03/2022	Basingstoke
UKENTH_TWU_TP000014	08/07/2020	30/03/2022	London (Beckton)
UKENTH_TWU_TP000015	08/07/2020	30/03/2022	London (Beddington)
UKENSW_SWS_TP000031	08/07/2020	30/03/2022	St Ives and Penzance
UKENNW_UU_TP000076	08/07/2020	30/03/2022	Lancaster
UKENTH_TWU_TP000084	08/07/2020	30/03/2022	London (Hogsmill Valley)
UKENMI_ST_TP000222	08/07/2020	30/03/2022	Leicester
UKENNW_UU_TP000012	08/07/2020	30/03/2022	Barrow-in-Furness
UKENTH_TWU_TP000125	08/07/2020	30/03/2022	London (Riverside)
UKENSO_SW_TP000030	08/07/2020	30/03/2022	Maidstone and Aylesford
UKENSO_SW_TP000025	08/07/2020	30/03/2022	Chatham
UKENNW_UU_TP000110	08/07/2020	24/03/2022	Liverpool (Sandon)
UKENMI_ST_TP000156	08/07/2020	30/03/2022	Birmingham (Minworth)
UKENNW_UU_TP000095	08/07/2020	30/03/2022	Wirral
UKENSO_SW_TP000011	08/07/2020	30/03/2022	New Forest
UKENSO_SW_TP000001	08/07/2020	30/03/2022	Southampton
UKENNE_NU_TP000055	15/07/2020	30/03/2022	Washington
UKENMI_ST_TP000020	15/07/2020	30/03/2022	Barston
UKENMI_ST_TP000074	15/07/2020	30/03/2022	Derby
UKENNW_UU_TP000078	15/07/2020	30/03/2022	Leigh
UKENAN_AW_TP000200	15/07/2020	30/03/2022	Norwich
UKENAN_AW_TP000210	15/07/2020	30/03/2022	Peterborough
UKENMI_ST_TP000163	15/07/2020	30/03/2022	Nottingham
UKENSW_WXW_TP000004	15/07/2020	30/03/2022	Bristol
UKENNE_NU_TP000030	15/07/2020	30/03/2022	Horden
UKENNE_YW_TP000082	15/07/2020	30/03/2022	Bradford
UKENAN_AW_TP000161	15/07/2020	30/03/2022	Lincoln
UKENMI_ST_TP000068	15/07/2020	25/03/2022	Coventry
UKENSW_WXW_TP000092	15/07/2020	30/03/2022	Trowbridge
UKENTH_TWU_TP000113	15/07/2020	30/03/2022	London (Mogden)
UKENTH_TWU_TP000103	15/07/2020	30/03/2022	Luton
UKENNW_UU_TP000019	15/07/2020	30/03/2022	Bolton

ww_site_code	date_min	date_max	site_reporting_name
UKENAN_AW_TP000063	15/07/2020	30/03/2022	Colchester
UKENNE_YW_TP000098	15/07/2020	30/03/2022	Leeds
UKENNE_YW_TP000107	15/07/2020	30/03/2022	Dewsbury
UKENNW_UU_TP000011	01/10/2020	30/03/2022	Barnoldswick
UKENNE_YW_TP000119	08/02/2021	30/03/2022	Doncaster (Sandall)
UKENNE_NU_TP000012	10/02/2021	30/03/2022	Middlesbrough
UKENNE_NU_TP000031	10/02/2021	30/03/2022	Newcastle
UKENNE_NU_TP000003	10/02/2021	30/03/2022	Newton Aycliffe
UKENNE_NU_TP000051	10/02/2021	30/03/2022	Darlington
UKENNE_YW_TP000057	15/02/2021	30/03/2022	Sheffield (Blackburn Meadows)
UKENNE_NU_TP000019	17/02/2021	18/02/2022	Consett
UKENNE_YW_TP000094	17/02/2021	30/03/2022	Huddersfield
UKENTH_TWU_TP000139	17/02/2021	30/03/2022	Swindon
UKENNW_UU_TP000097	17/02/2021	30/03/2022	Northwich
UKENTH_TWU_TP000133	17/02/2021	28/03/2022	Slough
UKENTH_TWU_TP000126	17/02/2021	30/03/2022	Harlow
UKENTH_TWU_TP000122	17/02/2021	25/03/2022	Reading
UKENNE_NU_TP000020	17/02/2021	30/03/2022	Cramlington
UKENNE_NU_TP000054	17/02/2021	21/02/2022	Bishop Auckland
UKENTH_TWU_TP000102	17/02/2021	30/03/2022	London (Long Reach)
UKENNE_NU_TP000009	17/02/2021	30/03/2022	Billingham
UKENMI_ST_TP000050	19/02/2021	30/03/2022	Checkley
UKENNE_YW_TP000029	19/02/2021	30/03/2022	York
UKENNE_YW_TP000063	20/02/2021	30/03/2022	Wakefield
UKENNW_UU_TP000026	20/02/2021	30/03/2022	Bury
UKENNW_UU_TP000070	20/02/2021	30/03/2022	Kendal
UKENMI_ST_TP000099	21/02/2021	30/03/2022	Gloucester
UKENMI_ST_TP000100	21/02/2021	29/03/2022	Walsall
UKENMI_ST_TP000130	21/02/2021	30/03/2022	Leek
UKENMI_ST_TP000137	21/02/2021	30/03/2022	Loughborough
UKENMI_ST_TP000184	21/02/2021	25/03/2022	Telford
UKENNW_UU_TP000100	21/02/2021	30/03/2022	Penrith
UKENNW_UU_TP000050	21/02/2021	30/03/2022	Fleetwood
UKENMI_ST_TP000152	21/02/2021	30/03/2022	Melton Mowbray
UKENMI_ST_TP000242	21/02/2021	30/03/2022	Worksop
UKENMI_ST_TP000207	21/02/2021	30/03/2022	Stoke-on-Trent
UKENMI_ST_TP000180	21/02/2021	30/03/2022	Stourbridge and Halesowen
UKENMI_ST_TP000164	21/02/2021	30/03/2022	Nuneaton
UKENNW_UU_TP000116	21/02/2021	30/03/2022	Stockport

ww_site_code	date_min	date_max	site_reporting_name
UKENMI_ST_TP000036	22/02/2021	23/03/2022	Brancote
UKENNW_UU_TP000139	22/02/2021	30/03/2022	Workington
UKENMI_ST_TP000241	22/02/2021	30/03/2022	Worcester
UKENTH_TWU_TP000033	23/02/2021	30/03/2022	Camberley
UKENSW_SWS_TP000050	24/02/2021	30/03/2022	Newquay
UKENSW_SWS_TP000064	24/02/2021	30/03/2022	Sidmouth
UKENSO_SW_TP000096	24/02/2021	30/03/2022	Hailsham
UKENMI_ST_TP000062	24/02/2021	30/03/2022	Birmingham (Coleshill)
UKENTH_TWU_TP000050	24/02/2021	30/03/2022	Crawley
UKENSO_SW_TP000091	24/02/2021	30/03/2022	Bexhill
UKENTH_TWU_TP000159	24/02/2021	30/03/2022	Oxford
UKENSO_SW_TP000084	24/02/2021	30/03/2022	Scaynes Hill
UKENSO_SW_TP000083	24/02/2021	30/03/2022	Worthing
UKENSO_SW_TP000090	24/02/2021	30/03/2022	Littlehampton and Bognor
UKENSO_SW_TP000020	24/02/2021	30/03/2022	Tonbridge
UKENSO_SW_TP000082	24/02/2021	30/03/2022	Lewes
UKENSO_SW_TP000081	24/02/2021	30/03/2022	Burgess Hill
UKENSO_SW_TP000021	24/02/2021	30/03/2022	Tunbridge Wells
UKENNW_UU_TP000124	25/02/2021	28/03/2022	Warrington
UKENSW_WXW_TP000023	26/02/2021	30/03/2022	Chippenham
UKENSO_SW_TP000016	26/02/2021	30/03/2022	Isle of Wight
UKENNW_UU_TP000047	26/02/2021	30/03/2022	Ellesmere Port
UKENSW_SWS_TP000010	26/02/2021	30/03/2022	Camborne
UKENMI_ST_TP000120	26/02/2021	30/03/2022	Kidderminster
UKENSW_WXW_TP000005	26/02/2021	30/03/2022	Bath
UKENSW_WXW_TP000100	26/02/2021	30/03/2022	Weston-super-Mare
UKENSW_WXW_TP000044	28/02/2021	30/03/2022	Clevedon and Nailsea
UKENMI_ST_TP000167	01/03/2021	30/03/2022	Oswestry
UKENTH_TWU_TP000154	02/03/2021	30/03/2022	Witney
UKENMI_ST_TP000091	03/03/2021	30/03/2022	Evesham
UKENTH_TWU_TP000012	03/03/2021	25/03/2022	Banbury
UKENMI_ST_TP000178	03/03/2021	28/03/2022	Retford
UKENMI_ST_TP000139	03/03/2021	30/03/2022	Ludlow
UKENMI_ST_TP000147	03/03/2021	30/03/2022	Market Drayton
UKENMI_ST_TP000186	03/03/2021	28/03/2022	Scunthorpe
UKENMI_ST_TP000017	03/03/2021	30/03/2022	Malvern
UKENMI_ST_TP000256	03/03/2021	30/03/2022	Cheltenham
UKENTH_TWU_TP000021	05/03/2021	30/03/2022	Radlett
UKENTH_TWU_TP000116	05/03/2021	30/03/2022	Newbury
UKENAN_AW_TP000004	08/03/2021	30/03/2022	Anwick

ww_site_code	date_min	date_max	site_reporting_name
UKENAN_AW_TP000254	08/03/2021	30/03/2022	Sudbury
UKENAN_AW_TP000293	08/03/2021	30/03/2022	Wisbech
UKENAN_AW_TP000116	08/03/2021	30/03/2022	Grimsby
UKENAN_AW_TP000261	08/03/2021	30/03/2022	Thetford
UKENAN_AW_TP000286	08/03/2021	30/03/2022	Daventry
UKENAN_AW_TP000051	08/03/2021	30/03/2022	Chalton
UKENAN_AW_TP000041	08/03/2021	30/03/2022	Buckingham
UKENAN_AW_TP000028	08/03/2021	30/03/2022	Brackley
UKENAN_AW_TP000107	08/03/2021	30/03/2022	Northampton
UKENAN_AW_TP000055	08/03/2021	30/03/2022	Chelmsford
UKENAN_AW_TP000067	08/03/2021	30/03/2022	Corby
UKENAN_AW_TP000069	08/03/2021	30/03/2022	Milton Keynes
UKENAN_AW_TP000037	08/03/2021	30/03/2022	Wellingborough
UKENAN_AW_TP000023	08/03/2021	30/03/2022	Boston
UKENAN_AW_TP000026	08/03/2021	30/03/2022	Bourne
UKENAN_AW_TP000078	08/03/2021	30/03/2022	Diss
UKENAN_AW_TP000082	08/03/2021	30/03/2022	Downham Market
UKENAN_AW_TP000096	08/03/2021	30/03/2022	Felixstowe
UKENAN_AW_TP000106	08/03/2021	30/03/2022	Grantham
UKENAN_AW_TP000016	08/03/2021	30/03/2022	Bedford
UKENAN_AW_TP000015	08/03/2021	30/03/2022	Beccles
UKENAN_AW_TP000012	08/03/2021	30/03/2022	Barton-upon-Humber
UKENAN_AW_TP000077	08/03/2021	30/03/2022	Breckland
UKENAN_AW_TP000029	08/03/2021	27/03/2022	Braintree
UKENTH_TWU_TP000123	10/03/2021	30/03/2022	Reigate
UKENAN_AW_TP000237	10/03/2021	30/03/2022	Soham
UKENSW_WXW_TP000086	10/03/2021	30/03/2022	Taunton
UKENAN_AW_TP000194	10/03/2021	30/03/2022	Newmarket
UKENAN_AW_TP000047	10/03/2021	30/03/2022	Bury St. Edmunds
UKENSW_WXW_TP000096	10/03/2021	30/03/2022	Wellington
UKENSW_WXW_TP000057	10/03/2021	30/03/2022	Minehead
UKENSW_WXW_TP000077	10/03/2021	30/03/2022	Shepton Mallet
UKENAN_AW_TP000224	10/03/2021	30/03/2022	Saffron Walden
UKENAN_AW_TP000222	10/03/2021	30/03/2022	Royston
UKENTH_TWU_TP000019	12/03/2021	30/03/2022	Bicester
UKENAN_AW_TP000060	15/03/2021	30/03/2022	Shefford
UKENAN_AW_TP000154	15/03/2021	30/03/2022	Kings Lynn
UKENNE_YW_TP000076	15/03/2021	30/03/2022	Driffield
UKENNE_YW_TP000112	15/03/2021	30/03/2022	Chesterfield
UKENNE_YW_TP000026	15/03/2021	30/03/2022	Malton

UKENSW_SWS_TP000025	22/02/2021	30/03/2022	Falmouth
UKENSW_SWS_TP000045	22/02/2021	30/03/2022	Liskeard
UKENSW_SWS_TP000051	22/02/2021	30/03/2022	Newton Abbot
UKENMI_ST_TP000233	22/02/2021	30/03/2022	Wigston
UKENSW_SWS_TP000056	22/02/2021	30/03/2022	Plymouth (Camels Head)
UKENSW_SWS_TP000055	22/02/2021	30/03/2022	Par
UKENSW_SWS_TP000059	22/02/2021	30/03/2022	Plympton
UKENNW_UU_TP000129	22/02/2021	30/03/2022	Whaley Bridge
UKENSW_SWS_TP000074	22/02/2021	30/03/2022	Tiverton
UKENMI_ST_TP000003	22/02/2021	28/03/2022	Alfreton
UKENSW_SWS_TP000075	22/02/2021	30/03/2022	Torquay
UKENMI_ST_TP000018	22/02/2021	30/03/2022	Wolverhampton
UKENAN_AW_TP000148	08/03/2021	30/03/2022	Jaywick
UKENAN_AW_TP000160	08/03/2021	30/03/2022	Letchworth
UKENAN_AW_TP000169	08/03/2021	30/03/2022	Louth
UKENAN_AW_TP000170	08/03/2021	30/03/2022	Lowestoft
UKENAN_AW_TP000172	08/03/2021	30/03/2022	Mablethorpe
UKENAN_AW_TP000176	08/03/2021	30/03/2022	March
UKENAN_AW_TP000177	08/03/2021	30/03/2022	Market Harborough
UKENAN_AW_TP000308	08/03/2021	30/03/2022	Tilbury
UKENAN_AW_TP000307	08/03/2021	30/03/2022	Southend-on-Sea
UKENAN_AW_TP000201	08/03/2021	30/03/2022	Oakham
UKENAN_AW_TP000303	08/03/2021	30/03/2022	Basildon
UKENAN_AW_TP000296	08/03/2021	30/03/2022	Witham
UKENAN_AW_TP000242	08/03/2021	30/03/2022	Spalding
UKENAN_AW_TP000248	08/03/2021	30/03/2022	Stamford
UKENAN_AW_TP000253	08/03/2021	30/03/2022	Stowmarket
UKENNE_YW_TP000061	15/03/2021	30/03/2022	Bridlington
UKENNE_YW_TP000131	15/03/2021	30/03/2022	Pontefract
UKENNE_YW_TP000102	17/03/2021	30/03/2022	Barnsley
UKENNE_YW_TP000096	17/03/2021	30/03/2022	Keighley
UKENNE_YW_TP000133	17/03/2021	30/03/2022	Doncaster (Thorne)
UKENMI_ST_TP000208	19/03/2021	30/03/2022	Stroud
UKENNW_UU_TP000133	21/03/2021	30/03/2022	Wigan
UKENNW_UU_TP000103	21/03/2021	30/03/2022	Rochdale
UKENNW_UU_TP000067	21/03/2021	30/03/2022	Hyde
UKENNW_UU_TP000037	21/03/2021	25/03/2022	Congleton
UKENSW_WXW_TP000074	24/03/2021	30/03/2022	Salisbury

UKENSW_WXW_TP000075	24/03/2021	30/03/2022	Shaftesbury
UKENSW_WXW_TP000018	24/03/2021	30/03/2022	Chard
UKENSO_SW_TP000107	24/03/2021	30/03/2022	Chichester
UKENSO_SW_TP000002	24/03/2021	30/03/2022	Lymington and New Milton
UKENSO_SW_TP000004	24/03/2021	30/03/2022	Portsmouth and Havant
UKENSO_SW_TP000006	24/03/2021	30/03/2022	Andover
UKENSO_SW_TP000033	24/03/2021	30/03/2022	Canterbury
UKENSO_SW_TP000032	24/03/2021	30/03/2022	Sittingbourne
UKENSO_SW_TP000008	24/03/2021	30/03/2022	Fareham and Gosport
UKENSO_SW_TP000026	24/03/2021	30/03/2022	Ashford
UKENSO_SW_TP000013	24/03/2021	30/03/2022	Eastleigh
UKENNW_UU_TP000027	24/03/2021	30/03/2022	Carlisle
UKENSW_WXW_TP000085	24/03/2021	30/03/2022	Blandford Forum
UKENNW_UU_TP000062	26/03/2021	27/03/2022	Maghull
UKENNW_UU_TP000018	26/03/2021	30/03/2022	Blackburn
UKENTH_TWU_TP000039	26/03/2021	14/03/2022	Chesham
UKENSW_WXW_TP000111	26/03/2021	30/03/2022	Yeovil
UKENTH_TWU_TP000047	26/03/2021	30/03/2022	Cirencester
UKENTH_TWU_TP000055	26/03/2021	30/03/2022	Didcot
UKENTH_TWU_TP000073	26/03/2021	28/03/2022	Guildford
UKENNW_UU_TP000024	26/03/2021	30/03/2022	Burnley
UKENMI_ST_TP000141	29/03/2021	30/03/2022	Lydney
UKENTH_TWU_TP000004	31/03/2021	28/03/2022	Alton
UKENTH_TWU_TP000106	31/03/2021	30/03/2022	St Albans
UKENTH_TWU_TP000023	31/03/2021	21/03/2022	Bordon
UKENSW_WXW_TP000012	07/04/2021	30/03/2022	Bridport
UKENMI_ST_TP000060	07/04/2021	30/03/2022	Telford South
UKENSW_WXW_TP000038	07/04/2021	30/03/2022	Bournemouth (Central)
UKENSO_SW_TP000027	07/04/2021	30/03/2022	Hythe
UKENSW_WXW_TP000084	07/04/2021	30/03/2022	Swanage
UKENSO_SW_TP000028	07/04/2021	30/03/2022	Dover and Folkestone
UKENMI_ST_TP000143	09/04/2021	30/03/2022	Mansfield
UKENSO_SW_TP000022	05/05/2021	30/03/2022	“Ramsgate, Sandwich and Deal”

ww_site_code	date_min	date_max	site_reporting_name
UKENNE_NU_TP000046	21/05/2021	30/03/2022	Hartlepool
UKENSW_SWS_TP000067	26/05/2021	30/03/2022	Menagwins
UKENSW_SWS_TP000033	26/05/2021	30/03/2022	Helston
UKENSW_SWS_TP000005	26/05/2021	30/03/2022	Bodmin Sc.Well
UKENTH_TWU_TP000155	04/06/2021	25/03/2022	Woking
UKENAN_AW_TP000071	09/06/2021	30/03/2022	Cromer
UKENAN_AW_TP000280	09/06/2021	30/03/2022	Wells-next-the-Sea
UKENAN_AW_TP000247	09/06/2021	30/03/2022	Stalham
UKENAN_AW_TP000219	09/06/2021	30/03/2022	Reepham
UKENAN_AW_TP000128	09/06/2021	30/03/2022	Hunstanton
UKENAN_AW_TP000191	11/06/2021	30/03/2022	Needham Market
UKENNE_NU_TP000028	21/06/2021	30/03/2022	Sunderland
UKENNW_UU_TP000113	30/07/2021	30/03/2022	Skelmersdale
UKENNW_UU_TP000104	04/08/2021	27/03/2022	Rossendale
UKENNW_UU_TP000032	13/08/2021	30/03/2022	Chorley
UKENNW_UU_TP000034	16/08/2021	30/03/2022	Clitheroe
UKENNE_YW_TP000039	18/08/2021	30/03/2022	Scarborough
UKENNW_UU_TP000068	20/08/2021	30/03/2022	Hyndburn
UKENSW_SWS_TP000016	13/10/2021	30/03/2022	Bideford
UKENSW_SWS_TP000073	13/10/2021	30/03/2022	Tavistock
UKENNE_NU_TP000004	05/11/2021	30/03/2022	Durham (Barkers Haugh)
UKENNE_NU_TP000048	05/11/2021	30/03/2022	Houghton-le-Spring
UKENNE_NU_TP000007	17/11/2021	30/03/2022	Durham (Belmont)
UKENNE_NU_TP000039	28/11/2021	30/03/2022	MARSKE REDCAR
UKENNW_UU_TP000017	20/12/2021	30/03/2022	Birkenhead
UKENNW_UU_TP000023	20/12/2021	30/03/2022	Bromborough
UKENNW_UU_TP000066	22/12/2021	30/03/2022	Huyton and Prescot
UKENAN_AW_TP000056	05/01/2022	30/03/2022	Clacton-on-Sea and Holland-on-Sea
UKENAN_AW_TP000306	05/01/2022	30/03/2022	Basildon (Vange)
UKENAN_AW_TP000289	05/01/2022	30/03/2022	Wickford
UKENAN_AW_TP000221	05/01/2022	30/03/2022	Rochford
UKENAN_AW_TP000305	05/01/2022	30/03/2022	Canvey Island
UKENAN_AW_TP000052	05/01/2022	30/03/2022	Ipswich (Chantry)
UKENAN_AW_TP000084	09/01/2022	30/03/2022	Dunstable
UKENNE_YW_TP000126	10/01/2022	30/03/2022	Hemsworth and South Elmsall
UKENNE_YW_TP000054	10/01/2022	30/03/2022	Rotherham
UKENNE_YW_TP000075	10/01/2022	30/03/2022	Bingley
UKENNE_YW_TP000137	12/01/2022	30/03/2022	Castleford
UKENNE_YW_TP000073	14/01/2022	30/03/2022	Mexborough and Conisbrough

ww_site_code	date_min	date_max	site_reporting_name
UKENAN_AW_TP000115	08/03/2021	30/03/2022	Great Yarmouth
UKENAN_AW_TP000127	08/03/2021	30/03/2022	Haverhill
UKENAN_AW_TP000139	08/03/2021	30/03/2022	Huntingdon
UKENAN_AW_TP000143	08/03/2021	30/03/2022	Ingoldmells
UKENAN_AW_TP000144	08/03/2021	30/03/2022	Ipswich
UKENNW_UU_TP000102	21/02/2021	30/03/2022	Preston
UKENMI_ST_TP000056	21/02/2021	30/03/2022	Burton on Trent
UKENMI_ST_TP000225	22/02/2021	30/03/2022	Warwick
UKENSW_SWS_TP000002	22/02/2021	30/03/2022	Barnstaple
UKENMI_ST_TP000199	22/02/2021	28/03/2022	Spernal
UKENSW_SWS_TP000022	22/02/2021	30/03/2022	Ernesettle and Saltash
UKENSW_SWS_TP000024	22/02/2021	30/03/2022	Exmouth
UKENMI_ST_TP000182	22/02/2021	28/03/2022	Rugby
UKENNE_YW_TP000141	15/03/2021	30/03/2022	Sheffield (Woodhouse Mill)
UKENNE_YW_TP000008	15/03/2021	30/03/2022	Colburn
UKENNE_YW_TP000015	15/03/2021	30/03/2022	Harrogate North
UKENNE_YW_TP000030	15/03/2021	30/03/2022	Northallerton
UKENNE_YW_TP000056	15/03/2021	30/03/2022	Beverley
UKENAN_AW_TP000050	15/07/2020	30/03/2022	Cambridge
UKENTH_TWU_TP000100	15/07/2020	30/03/2022	Wycombe
UKENSW_WXW_TP000101	15/07/2020	30/03/2022	Weymouth
UKENTH_TWU_TP000052	15/07/2020	30/03/2022	London (Crossness)

AIMS Press