SUPPLEMENTARY MATERIAL

DESCRIPTION OF ADDITIONAL SUPPLEMENTARY FILES (DATASETS 1-6)

File name: Dataset S1 (DS1_DataCollection.xlsx)

Description: Data about sample collection. This document includes the dates when each sample was collected in the two locations considered: the WWTP Bens and the University Hospital of A Coruña (CHUAC). They all correspond to 24-h composite samples. Additional 2-h samples were collected during some of those days (marked as "each 2h").

File name: Dataset S2 (DS2_BensWeather.xlsx)

Description: Daily observations at the meteorological station of Coruña-Bens, for the period March 1st – May 31st, 2020, obtained from the Galician Meteorology Agency, MeteoGalicia (https://www.meteogalicia.gal/observacion/estacionshistorico/historico.action?idEst=14010). It contains measures for daily rainfall (L/m2), average humidity (%), and average temperature (ºC).

File name: Dataset S3 (DS3_COVIDCases.xlsx)

Description: Cumulative and active number of COVID-19 cases in the metropolitan area of A Coruña and from the health area A Coruña – Cee for the period March 1st – May 31st, 2020. It contains: date; ma_cumulative_cases, which is the cumulative number of confirmed COVID-19 cases in the metropolitan area according to the records provided by the General Directorate of Public Health (Automous Government of Galicia) of individual patients with domicile in one of the following municipalities: A Coruña, Arteixo, Cambre, Oleiros, Culleredo; ha_reported_active_cases, which is the number of active COVID-19 cases in the health area reported by SERGAS (Galician Health Service) and extracted from https://galiciancovid19.info/; ma_real_active_cases, corresponds to the series of real active cases based on the estimated daily official COVID-19 cases in the metropolitan area of A Coruña; and ICU_patients, which is the number of ICU patients with COVID-19 at University Hospital of A Coruña (CHUAC).

File name: Dataset S4 (DS4_Flow.xlsx)

Description: Two-minute flow measurements ($m^3 \cdot s^{-1}$) at WWTP Bens for the period January 1st – May 14th. It includes the following columns: date, hour, UR-MC-01, UR-MC-02, UR-MC-03. The last three columns are different flow measurements, namely: UR-MC-01 corresponds to the flow meter at the first pumping line; UR-MC-02 corresponds to the flow meter at the second pumping line; and UR-MC-03 corresponds to the flow meter found in the auxiliary rainwater pumping. For this article, the total flow rate pumped from the raw water well at WWTP Bens is computed as the sum of the three measurements (UR-MC-01 + UR-MC-02 + UR-MC-03).

File name: Dataset S5 (DS5_ViralLoad24h.xlsx).

Description: Viral load and Cq values obtained from the 24-h samples, measured in number of SARS-CoV-2 RNA copies/L. The measurements for six RT-PCR replicates are included. The data contains: date, number of copies/L and location (WWTP_Bens or CHUAC). Measurements marked with an asterisk (*) in the last column mean that the RT-PCR failed for at least one replication. In those cases, the missing values along the replications were imputed with the sample mean of the non-missing values of the replications. When all the replications were missing, they were imputed as the lowest value of all the observed measurements along the whole dataset.

File name: Dataset S6 (DS6_ViralLoad2h.xlsx).

Description: Viral load and Cq values obtained from the 2-h samples, measured in number of SARS-CoV-2 RNA copies/L. The measurements for three RT-PCR replicates are included. The data contains: date, hour of the day, number of copies/L and location (WWTP_Bens or CHUAC). Measurements marked with an asterisk (*) in the last column mean that the RT-PCR failed for at least one replication. In those cases, the missing values along the replications were imputed with the sample mean of the non-missing values of the replications. When all the replications were missing, they were imputed as the lowest value of all the observed measurements along the whole dataset. The document includes the measurements for three days in the case of WWTP Bens (May 5th, 6th, and 11th, 2020) and for one day in the case of CHUAC (May 12th, 2020).

STRAIGHT PATTERN DONE WITH THE HUMAN 2019-nCoV RNA STANDARD (EVAg) FOR VIRAL LOAD QUANTIFICATION THROUGH RT-qPCR ASSAYS

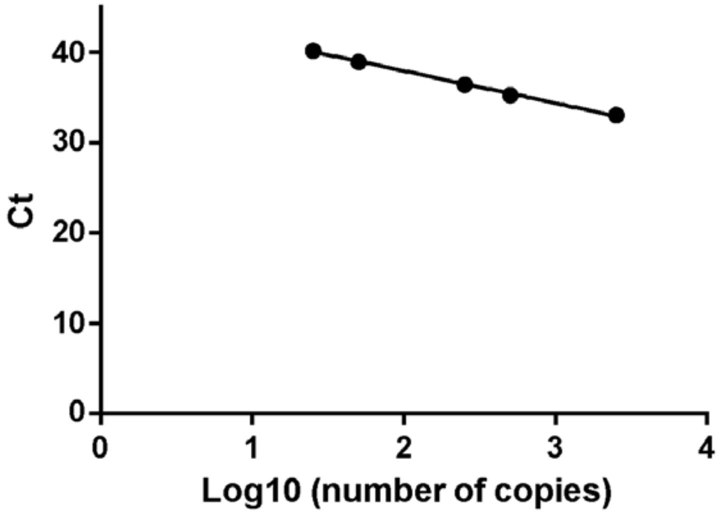

Figure S1. Straight pattern done with the Human 2019-nCoV RNA standard (EVAg) for viral load quantification through RT-qPCR assays. Linear regression analysis showed a calibration curve of y= -3.5657x + 42.588, R^2= 0.9984.

DETERMINATION OF THE LOWER LIMIT OF VIRAL LOAD DETECTED BY RT-qPCR

The limit of detection (LOD) of the RT-qPCR was analyzed using serial dilutions of Human 2019-nCoV RNA standard (EVAg). The Cq values and detection rates are shown in Table S1, where the LOD was established in 25 copies *per* reaction. The low values of coefficient of variation (CV) suggest that limit of quantification (LOQ) is between 25 and 15 copies *per* reaction. The slope of the standard curve was -3.56, the R^2 was 0.9984 and the amplification efficiency was above 90%, calculated according to the recommendations of the MIQE guidelines (Bustin et al., 2009).

Table S1. Lower limit of detection of SARS-CoV-2.

Total number of copies/reaction	Detected/tested (%)	C_q Mean (SD)	CV
2500	10/10 (100)	33.09 (1.48)	4.47
500	10/10 (100)	35.01 (0.52)	1.46
250	10/10 (100)	36.48 (0.50)	1.37
50	9/10 (90)	38.99 (0.93)	2.39
25	10/10 (100)	40.20 (1.16)	2.88
15	ND	-	-

ND: amplification not detected
SD: standard deviation
CV: coefficient of variation

NONPARAMETRIC FITTING OF VIRAL LOAD OVERTIME

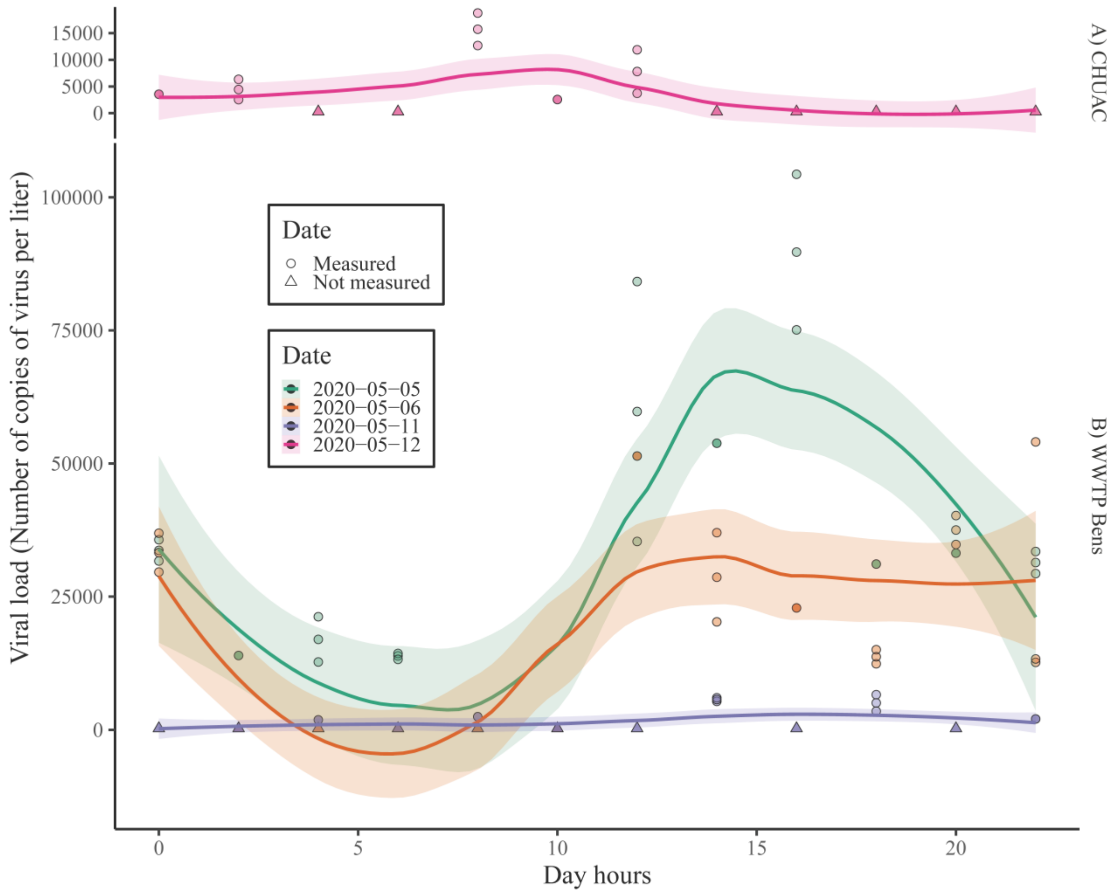

Figure S2. Viral load trend during the day at CHUAC and WWTP Bens. A) Viral load with respect to the hour of the day in CHUAC during May 12, 2020 and nonparametric LOESS fitted model (span parameter equal to 0.75) with 95% confidence interval. B) Viral load with respect to the hour of the day in WWTP Bens for three different days in May, and nonparametric models (span parameter equal to 0.75) with 95% confidence interval fitted to the data of each day separately.

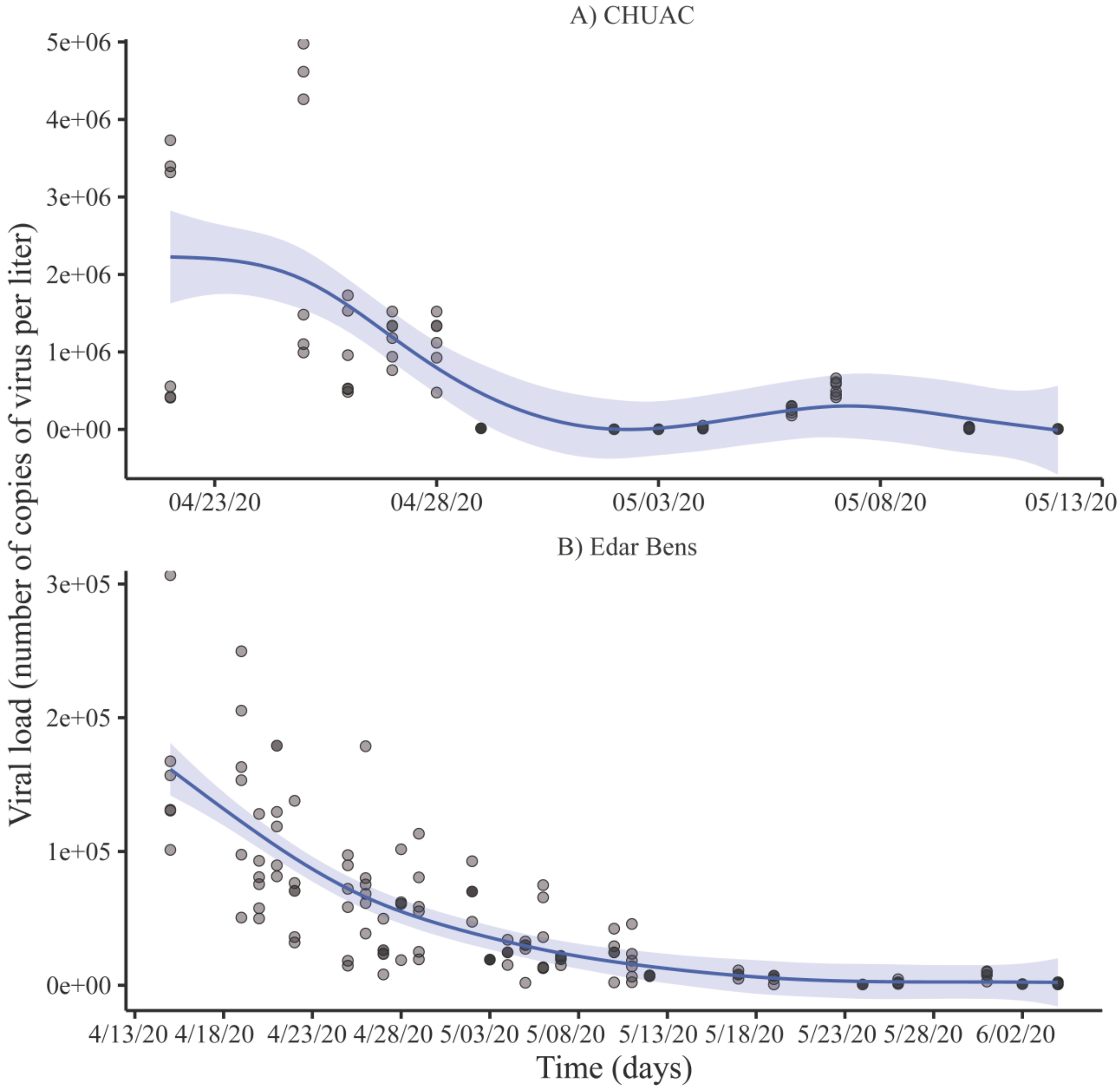

Figure S3. Date effect in the viral load using a nonparametric estimator (GAM) A) at CHUAC and B) at WWTP Bens.

MULTIVARIATE LINEAR MODEL SELECTION USING THE R^2 MAXIMIZATION CRITERION FOR ESTIMATING THE NUMBER OF COVID-19 CASES

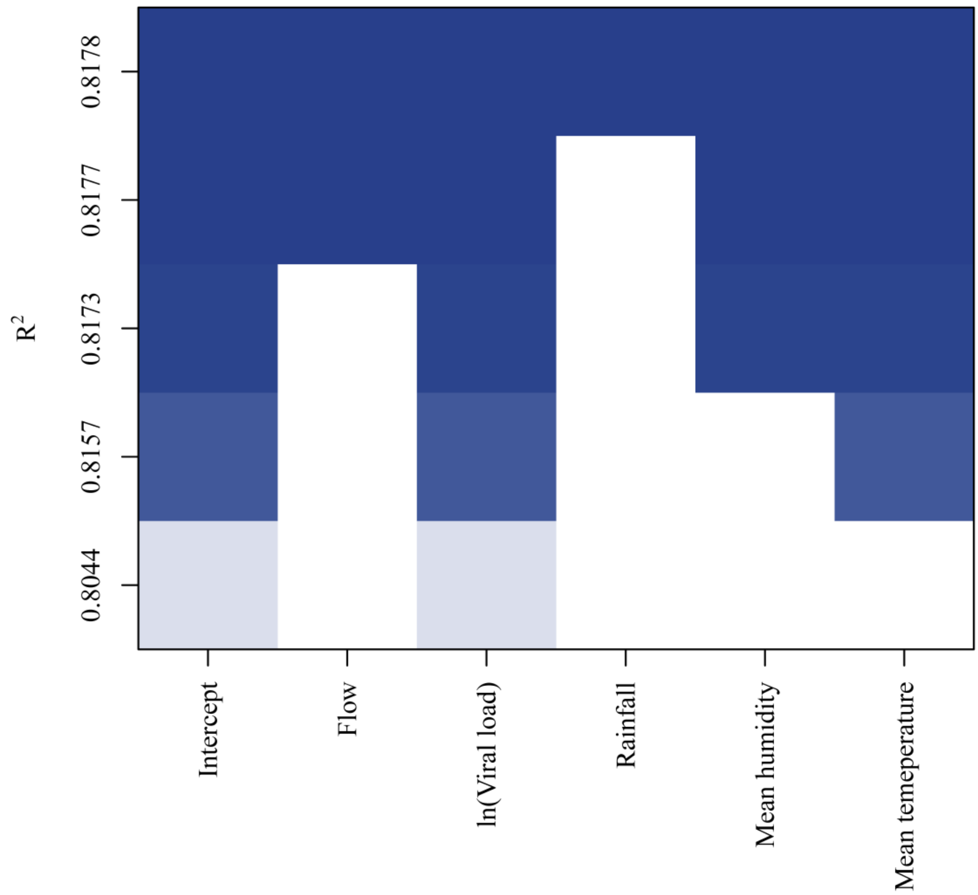

Figure S4. Multivariate linear model selection using the R^2 maximization criterion. Each row corresponds with the best model using from one to five predictors. The color of the row is darker for higher values of R^2. The best option to explain the nº of COVID-19 cases is a simple linear model as a function of the ln(Viral load). In fact, the R^2 hardly change when adding Flow, Rainfall, Humidity and Temperature variables.

FLOW STUDY BEFORE, DURING AND AFTER LOCKDOWN

METHODOLOGY

Two-minute data of volume of pumped water in the entrance of the WWTP in the period January 1st – May 4th 2020 were collected. This period was split into four time intervals corresponding to: (a) regular conditions (January 1st – March 13th), (b) initial alarm state period (March 14th –26th), (c) strict lockdown period (March 27th – April 9th), and (d) lockdown de-escalation (April 10th – May 3rd). All these data are recovered in Dataset S4. The mean hour flow at the WWTP has been computed and a multivariate exploratory data analysis was performed. Box-plots for flow at every hour have been computed using the data in every time period a-d. They were plotted as a function of the hour of the day. Exploratory functional data analyses have been also performed. Since the raw flow curves are exceedingly noisy, local polynomial methods (Fan and Gijbels, 1996) have been used to obtain smooth curves. A direct plug-in method (Ruppert *et al.*, 1993) has been used to choose the smoothing parameter. The collection of smoothed daily curves has been analyzed and the deepest curve (Fraiman and Muniz, 2001) among every time period a-d has been computed.

RESULTS

Data at WWTP Bens showed a clear daily trend (**Figure S5**), with the highest values between 12:00 and 18:00 in period A (before the lockdown). Then, as periods B, C and D passed, a reduction in the level and the variability of the flow was observed. The estimated time for the wastewater to reach the WWTP Bens along the network was between 1.5 h and 3 h, depending on the source in the metropolitan area. Therefore, when interpreting **Figure S5**, it must be taken into account that a peak between 12:00 and 18:00 reflects greater human or industrial activity at least 1.5 h beforehand. In addition, the daily two-minute flow curves performed were exceptionally noisy (**Figure S6A**), so these curves were smoothed (**Figure S6B**) to study their real shape. When considering the entire daily curves and grouping them into the four time intervals (a-d), clear patterns were shown (**Figure S6C**). These patterns were even clearer when plotting central curves (the deepest curves) within every group (**Figure S6D**).

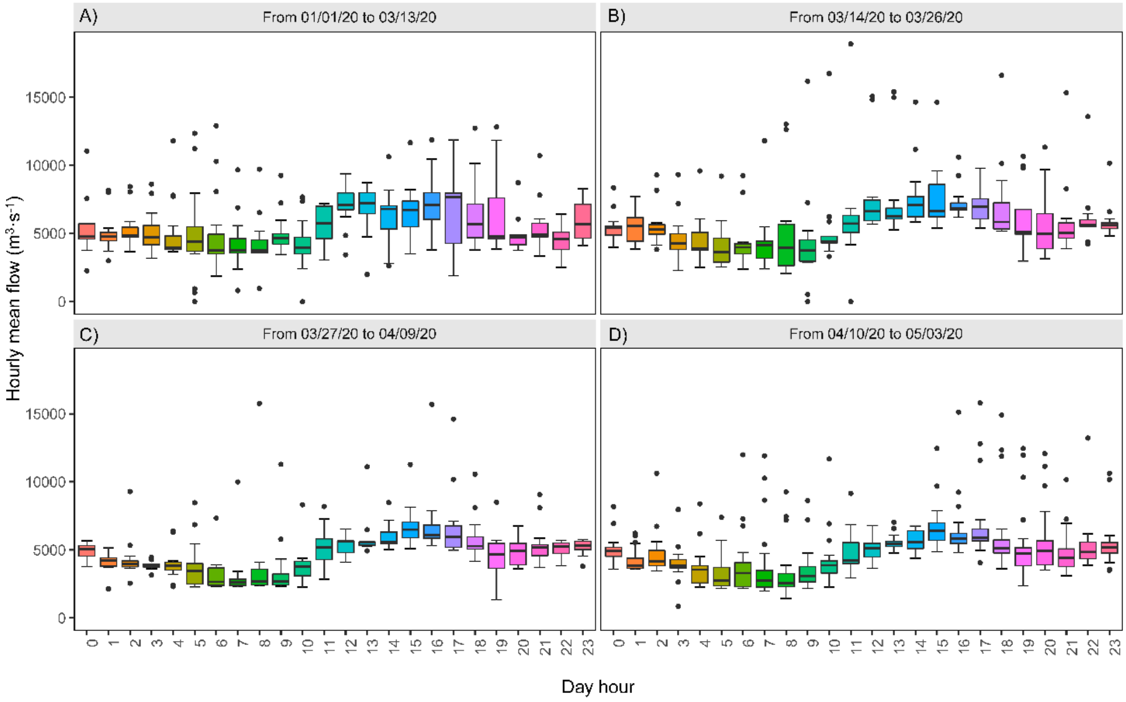

Figure S5. Hourly mean flow box-plots for the 24 h in the day corresponding to the four time intervals, A-D, considered during the lockdown period in the metropolitan area of A Coruña. A) Regular conditions (January 1st – March 13th). B) Initial alarm state period (March 14th –26th). C) Strict lockdown period (March 27th – April 9th). D) Lockdown de-escalation (April 10th – May 3rd).

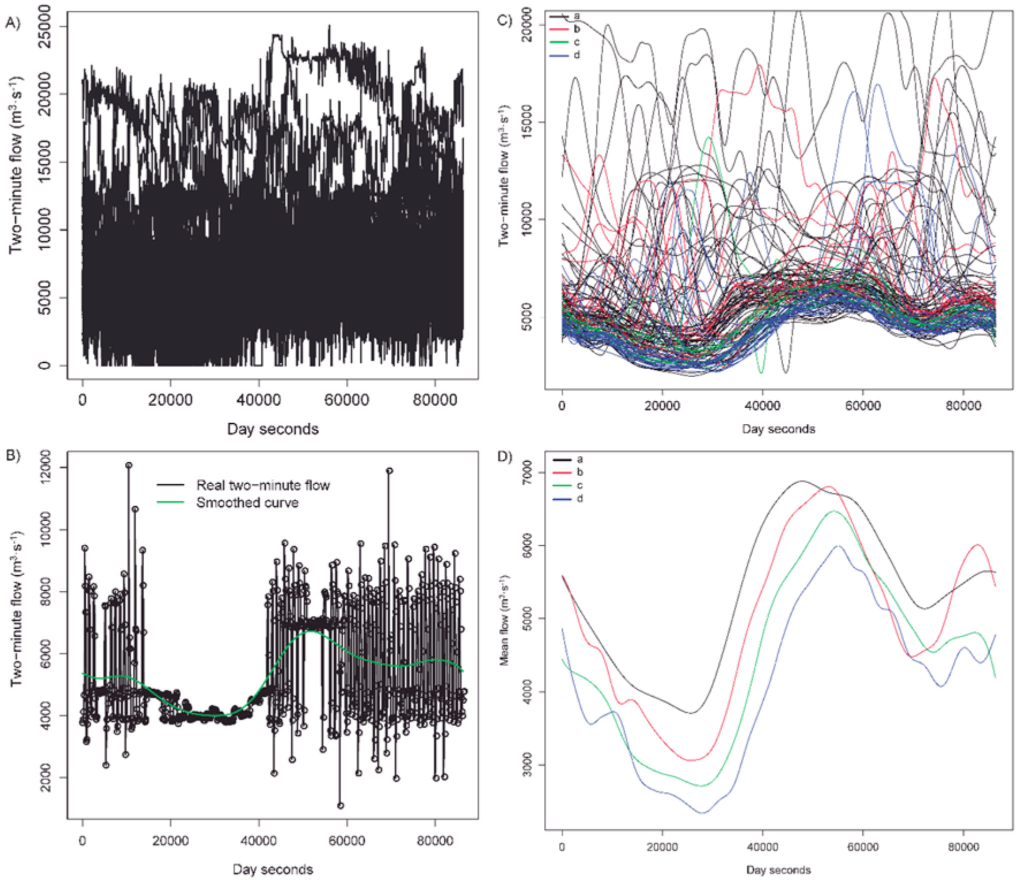

Figure S6. Exploratory flow study during the COVID-19 epidemic in the metropolitan area of A Coruña. A) Set of daily two-minute flow curves for the period January 1st – May 3rd. B) One two-minute flow curve (black) and its smoothed version (green) using a local linear fit. C) Smooth daily two-minute flow curves for the four time intervals considered, a (black), b (red), c (green) and d (blue) during the lockdown period. D) Deepest daily flow smoothed curves for every time interval, a (black), b (red), c (green) and d (blue), during the lockdown period.

DISCUSSION

An initial study of flows was made to analyze their variability, which could have influenced the concentration of the virus in the wastewater (**Figures S2** and **S3**). For instance, it was expected that on rainy days, the viral load detected at the entrance of WWTTP Bens would be less than for dry days, with the same number of COVID-19 cases. Therefore, a study of flow rates was first carried out at the wastewater inlet of the treatment plant. The daily flow analysis showed that the usage of the sewage network changed over time, especially when comparing the pre-pandemic period with the three phases during the lockdown. The mean flow curves exhibited higher levels before the lockdown and their levels decreased as time passed. In addition, the mean flow curves tended to shift to the right when moving from A to B, C, and D periods (**Figure S3**). This is probably due to the change in habits related to the restrictions to work activity, the confinement in people's homes and the increasing paralysis of the economic activity during the state of alarm in Spain.

REFERENCES

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., Wittwer, C. T.. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 2019;55(4): 611–622.

Fan, J. & Gijbels, I. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. (Taylor & Francis, 1996).

Ruppert, D., Sheather, S. J., Wand, M. P. & Management, A. G. S. o. *An Effective Bandwidth Selector for Local Least Squares Regression*. (Australian Graduate School of Management, University of New South Wales, 1993).

Fraiman, R. & Muniz, G. *Trimmed Means for Functional Data*. (Universidad de San Andrés, 2001).