

Basic influent sewage quality reflects sewershed characteristics in Tokyo city

Zhongyu Guo^a, Takayuki Hatakeyama^a, Chihiro Yoshimura^{ib}^{a,*}, Tingting Wang^b and Yuta Hatano^a

^a Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Meguro-Ku, Tokyo 152-8552, Japan

^b Graduate School of Science, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464-8602, Japan

*Corresponding author. E-mail: yoshimura.c.aa@m.titech.ac.jp

 CY, 0000-0002-2680-8038

ABSTRACT

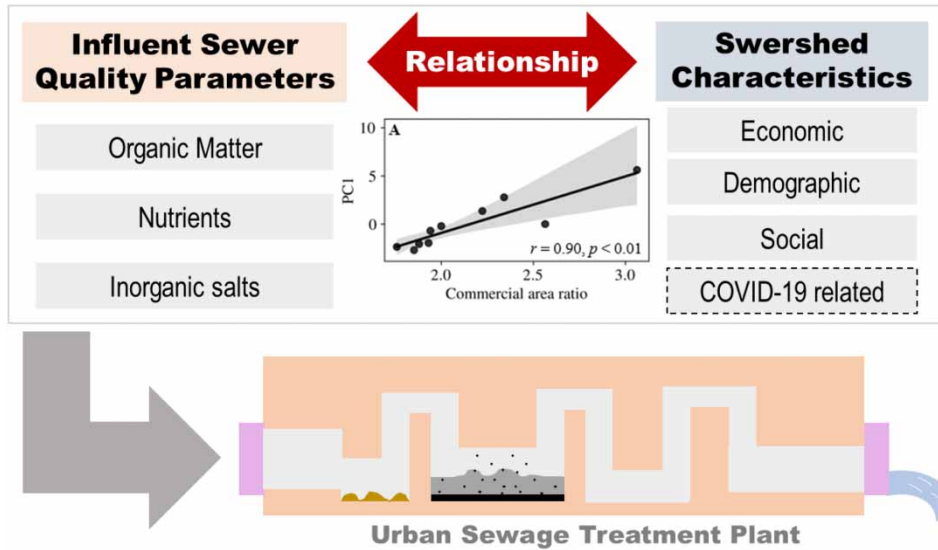
Sewage comprises multifarious information on sewershed characteristics. For instance, influent sewage quality parameters (ISQPs) (e.g., total nitrogen (TN)) are being monitored regularly at all treatment plants. However, the relationship between ISQPs and sewershed characteristics is rarely investigated. Therefore, this study statistically investigated relationships between ISQPs and sewershed characteristics, covering demographic, social, and economic properties in Tokyo city as an example of a megacity. To this end, we collected ISQPs and sewershed characteristic data from 2015 to 2020 in 10 sewersheds in Tokyo city. By principal component analysis, spatial variability of ISQPs was aggregated into two principal components (89.8% contribution in total), indicating organics/nutrients and inorganic salts, respectively. Concentrations of organics/nutrients were significantly correlated with the population in sewersheds (daytime population density, family size, age distribution, etc.). Inorganic salts are significantly correlated with land cover ratios. Finally, a multiple regression model was developed for estimating the concentration of TN based on sewershed characteristics ($R^2=0.97$). Scenario analysis using the regression model revealed that possible population movements in response to the coronavirus pandemic would substantially reduce the concentration of TN. These results indicate close relationships between ISQPs and sewershed characteristics and the potential applicability of big data of ISQPs to estimate sewershed characteristics and vice versa.

Key words: COVID-19, land use, lifestyle, population distribution, regression model, sewage

HIGHLIGHTS

- Influent sewage quality parameters (ISQPs) were classified into organics, nutrients, and inorganic salts.
- Organics and nutrients are mainly correlated to population information.
- Inorganic salts are significantly correlated to land-use characteristics.
- ISQPs concentration can be estimated by sewershed characteristics.
- Scenario analysis indicated total nitrogen concentration will decrease during/after COVID-19.

GRAPHICAL ABSTRACT



INTRODUCTION

The sewerage system is an essential infrastructure, playing a vital role in the community's sustainable development. It transports sewage from various sources in a sewershed to a centralized treatment plant or disposal area through a network of pipelines, pumping stations, and other ancillary facilities. Thus, the sewage quality serves as a marker of water use and society in a sewershed. This means that each sewershed community possibly shows its own 'sewage fingerprint,' which reflects the characteristics of the community such as lifestyle and socioeconomic activities (Thomas & Reid 2011). Therefore, monitoring the influent sewage quality could allow us to obtain information on multifarious characteristics of a community, which might be difficult and costly to obtain by other approaches covering a whole community.

In fact, previous studies have widely reported that analyses of chemical residues in influent sewage can estimate a community's consumption of those chemicals or drugs (Zuccato *et al.* 2008; Castiglioni *et al.* 2013; Choi *et al.* 2019; O'Brien *et al.* 2019a; Zhang *et al.* 2019). This approach has been used in illicit drug control or the estimation of pharmaceutical consumption in recent years (Thomaidis *et al.* 2016). They have also reported that some chemicals show tight relations with sewershed characteristics such as social, demographic, and economic information, making it feasible to estimate several sewershed characteristics based on chemical residues in sewage. For example, the sewershed population has shown strong correlations with the daily mass load of homovanillic and vanillylmandelic acids (Pandopulos *et al.* 2021) and other biomarkers (Hou *et al.* 2021; O'Brien *et al.* 2014). Those cases imply the great potential of sewage quality for estimating sewershed characteristics.

In addition to chemicals in influent sewage, general influent sewage quality parameters (ISQPs) such as $\text{NH}_4\text{-N}$, biochemical oxygen demand (BOD), and total nitrogen (TN) are also possibly meaningful markers to estimate some sewershed characteristics. For example, the sewershed population has been successfully estimated based on a load of chemical oxygen demand (COD), BOD, total phosphorus (TP), and $\text{NH}_4\text{-N}$ concentrations (Van Nuijs *et al.* 2011; Rico *et al.* 2017; Tschärke *et al.* 2019). Importantly, ISQPs have been measured regularly basically in all the treatment plants, resulting in big data. The method of analyzing ISQPs has been standardized, resulting in sufficient accuracy for comparisons across regions. Thus, no additional major cost is required to utilize ISQPs, which is another benefit. Those properties contrast with monitoring chemical compounds in influent sewage, which requires complex procedures that are expensive and time-consuming (Ort *et al.* 2010). Therefore, it is worthwhile to investigate the potential of big data of ISQPs as sewage markers of sewershed characteristics.

Only a few studies, however, have attempted to investigate the relationship between ISQPs and sewershed characteristics (e.g., Been *et al.* 2014; Lin *et al.* 2019). For such purposes, ISQPs can be analyzed using their concentration and load. As briefly reviewed above, the load-based analysis of influent sewage has been commonly applied in the world. In contrast,

the direct application of influent concentration without calculating load is not common. *Been et al. (2014)* have shown that $\text{NH}_4\text{-N}$ concentration in influent sewage is correlated with the population and can be applied to estimate the population size of the target treatment area. Furthermore, such direct application of concentration of ISQPs is possibly meaningful in case we compare several sewersheds having different populations and socioeconomic activities in a region. To date, however, the relation has been determined only between the concentration of ISQPs and the sewershed population. Relationships between ISQPs and other sewershed characteristics, such as land use, health condition, and economy, have not been explored. Therefore, further detailed investigation into the relationships between ISQPs and sewershed characteristics might be valuable to obtain a comprehensive overview of how to estimate sewershed characteristics utilizing the readily available data of ISQPs.

Tokyo city is the political, economic, and cultural hub of Japan, and its sewerage system served over 9.49 million people until 2019 (*Japan Statistics Bureau 2020*). Various sewershed characteristics and ISQP data have been archived as public information and can potentially serve as big data, and thus the relationship between ISQPs and sewershed characteristics is possibly explored. Therefore, we aimed to (1) investigate the relationship of concentration of ISQPs with regional and temporal sewershed characteristics and (2) establish regression models describing the relationship in Tokyo city, as a case study of a megacity. The sewershed characteristics in this research covered social, demographic, economic, and climate conditions. We also estimated the shifts in ISQPs in scenario analysis considering different responses in population dynamics to COVID-19 in the city. These results help decision-makers devise appropriate plans in sewerage management in accordance with predicted shifts of the society in Tokyo. In addition, the derived regression can be a reference for analyzing the relationship in other regions.

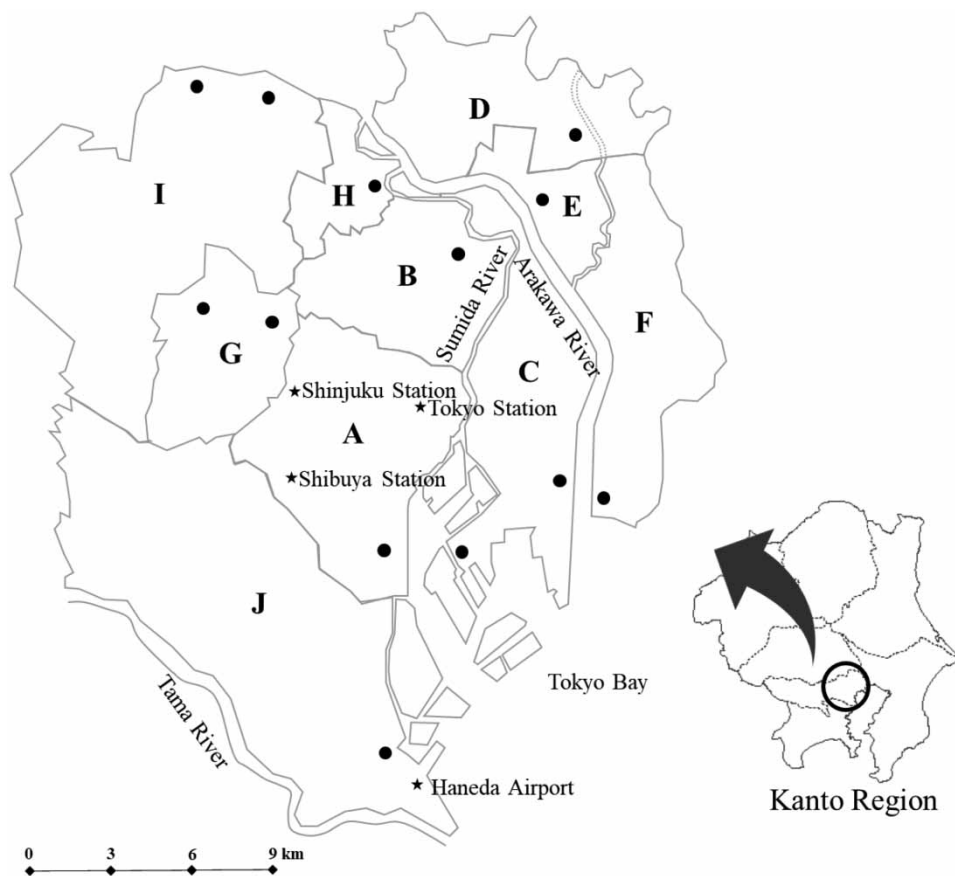


Figure 1 | The 10 sewersheds with a total of 13 WWTPs (black points) in Tokyo city, Japan.

METHODS

Study area and data

The study area was Tokyo city, located in the south of the Kanto region (Figure 1), Japan. This city is the political, economic, and cultural center of Japan and consists of 23 special-ward areas as the national capital. The sewerage system in 23 specific ward areas is covered by 10 sewersheds, with 13 wastewater treatment plants (WWTPs) in total (Figure 1), serving a population of about 9.49 million in 2019 (Japan Statistics Bureau 2020).

The monthly ISQP data from April 2015 to March 2020 were provided by the Bureau of Sewerage, the Tokyo Metropolitan Government. The sewage quality is measured at each WWTP once a month, minimizing the effect of rainwater (i.e., basically sampled on dry days). On each sampling day, a certain volume of the influent sewage was collected at 2 h intervals over a 24-h period, and these volumes were mixed as a composite sample for water-quality determination. The monitored parameters were concentrations of BOD, COD measured with the oxidant of potassium permanganate, suspended solids (SS), ignition loss (IL), total solid (TS), dissolved solid (DS), chloride ion (Cl^-), TN, ammonia nitrogen (NH_4^+), TP, and phosphate-phosphorus (PO_4^{3-}). The concentrations of NH_4^+ and PO_4^{3-} were expressed in terms of N and P masses, respectively (hereafter denoted by $\text{NH}_4\text{-N}$ and $\text{PO}_4\text{-P}$). Besides these concentration data, we obtained the flow rate of each trunk line (w_i) for sewage discharge.

We also collected the information on sewershed characteristics for the 23 wards in Tokyo. As summarized in Table 1, these data contain population, land use, health and income, meteorological information, economic information, and monthly

Table 1 | Collected data on the sewershed characteristics of Tokyo city

Category	Regional characteristics	Year	Data sources
Population	Daytime and nighttime population	2015	Tokyo Metropolitan Government Bureau of General Affairs
	Number of people per household		
	Age composition	2019	
Land use	Land-use ratio	2016	Tokyo Metropolitan Government Bureau of Urban Development
	Middle-high-rise rate		
	Building structure ratio		
Health and income	Average life expectancy	2015	Ministry of Health, Labor and Welfare Tokyo Metropolitan Government Bureau of Social Welfare and Public Health
	Metabolic syndrome rate	2017	
	Smoking rate		
	Annual income	2019	
Meteorological information	Average atmospheric pressure	2015–2020	Japan Meteorological Agency
	Total precipitation		
	Daily average temperature		
	Average humidity		
	Average wind speed		
	Sunshine duration		
	Cloud cover		
Economic information	Consumer confidence index	2015–2020	Cabinet Office
Monthly expenditure information per household	Consumption expenditure	2015–2020	Ministry of Internal Affairs and Communications
	Food expenditure		
	Dwelling expenditure		
	Light, heating, and sewer expenditure		
	Healthcare expenditure		
	Education and recreation expenditure		
Other	Other miscellaneous expenditure		
	The ratio of separated sewer	2020	Tokyo Metropolitan Government Bureau of Sewerage

expenditure information per household for the target period of 5 years from April 2015 to March 2020. For the meteorological information, the average of the meteorological data observed during the first to fifth days of each month was applied to further analyses, considering that the ISQPs were measured at the beginning of each month.

Data analysis

For the sewersheds that receive sewerage via two or more trunk lines, we calculated the flow-weighted average for each ISQP using Equation (1), by which the weighted average concentration of ISQPs was calculated with the flow rate of the trunk line as the weight.

$$X = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (1)$$

where X is the flow-weighted average concentration of ISQP in each treatment area, x_i is the concentration of ISQP of the trunk line, w_i is the flow rate of the trunk line, and n is the number of the trunk line connected to each WWTP.

The ward-base data about the characteristics were converted into sewershed-base data by taking the area-weighted average:

$$R = \frac{\sum_{i=1}^n W_i r_i}{\sum_{i=1}^n W_i} \quad (2)$$

where R is the representative sewershed characteristics in a sewershed, r_i is the regional data of a ward, W_i is the ward area, and n is the number of wards contained by a sewershed.

The relationship between the spatial sewershed characteristics and ISQPs of each sewershed was analyzed using multivariate statistical methods. First, principal component analysis (PCA) was applied to elucidate the overall variation in the ISQPs. Pearson's correlation analysis was then used to examine the relationship between the major principal components and the regional characteristics. For ISQPs, the 5-year annual averages from 2015 to 2019 were subjected to these statistical analyses under the assumption that the distribution of regional characteristics in Tokyo city does not show major changes in these 5 years. Prior to the correlation analysis, the normality of all the variables was tested by the Shapiro–Wilk test, whereas logarithmic transformation was applied if necessary. In case normality was not confirmed, we omitted those variables in the correlation analysis.

In addition, considering the seasonality of water use and socioeconomic property, correlation analysis was applied to investigate the relationship between temporal variations of ISQPs and sewershed characteristics on a monthly basis. The analyzed sewershed characteristics included meteorological information, economic information, and monthly expenditure information per household. The monthly variation ratio of the ISQPs (V) was indicated by the following ratio:

$$V = \frac{x_t}{\bar{x}} \quad (3)$$

where V is the monthly variation ratio, x_t is the monthly average sewage quality, and \bar{x} is the annual average sewage quality. R program (version 3.6.3) was used for all those statistical analyses.

Regression models and scenario analyses

Regression models were constructed using multiple linear regression as a model that describes the relationship between sewershed characteristics and ISQPs. Based on the data from 2015 to 2020, the significant explanatory variables ($p < 0.05$) were determined for the equation below by the forward–backward stepwise selection method.

$$y = \alpha_0 + \sum_{i=1}^n \alpha_i x_i \quad (4)$$

where y is the sewage quality (mg/L), x is the explanatory variable (regional characteristics and the ratio of a separate sewer system in each sewershed) (Tokyo Metropolitan Government Bureau of Sewerage 2021), α is the partial regression coefficient, and n is the number of explanatory variables.

Those models were then applied to the scenario analyses, in which the ISQPs were estimated for three population scenarios for the near future in Tokyo city. The scenarios combined the ‘impact of the spread of COVID-19 infection’ and ‘future population forecast’ projected by the Tokyo Metropolitan Government Bureau of General Affairs (<https://www.toukei.metro.tokyo.lg.jp/kyosoku/ky-index.htm>). The population scenarios are (I) the baseline case in which the population is not affected by COVID-19 (i.e., the same as the period from 2015 to 2020), (II) the case in which the number of commuting people in the Kanto region decreases because of the prevalence of teleworking (both inflow and outflow populations decrease by 20% compared with the baseline), and (III) the case in which people flow out of Tokyo city and restrictions on the movement of people continue (the nighttime population in the city decreases by 10% and the inflow and outflow populations decrease by 30% compared with the baseline). Microsoft Excel 2010 was used for the multiple regression and scenario analyses.

RESULTS

Overall relation of ISQPs to sewershed characteristics

The results from PCA revealed that the first two principal components contributed 63.4 and 26.4% (in total, 89.8%) to the total variance (Figure 2, Supplementary Table S1). The loadings of IL, TN, $\text{NH}_4\text{-N}$, and TP were relatively high for PC1, followed by BOD, COD, and $\text{PO}_4\text{-P}$. The DS and Cl^- showed high loadings on PC2, whereas TS and SS showed similar loadings for PC1 and PC2. Each of these principal components identified distinct characteristics of two sewersheds (A and C in Figure 2), where we confirmed the relatively high concentration of ISQPs indicating organics and nutrients (e.g., $\text{NH}_4\text{-N}$, TN, and IL) and indicating inorganic salts (e.g., Cl^- , DS, and TS), respectively.

The correlation analysis showed significant correlations of scores of PC1 and PC2 with several sewershed characteristics about population, land use, and health and income information (Table 2). The sewershed characteristics showed a higher number of significant correlations with PC1 than PC2, in particular for indicators of population. In addition, no significant

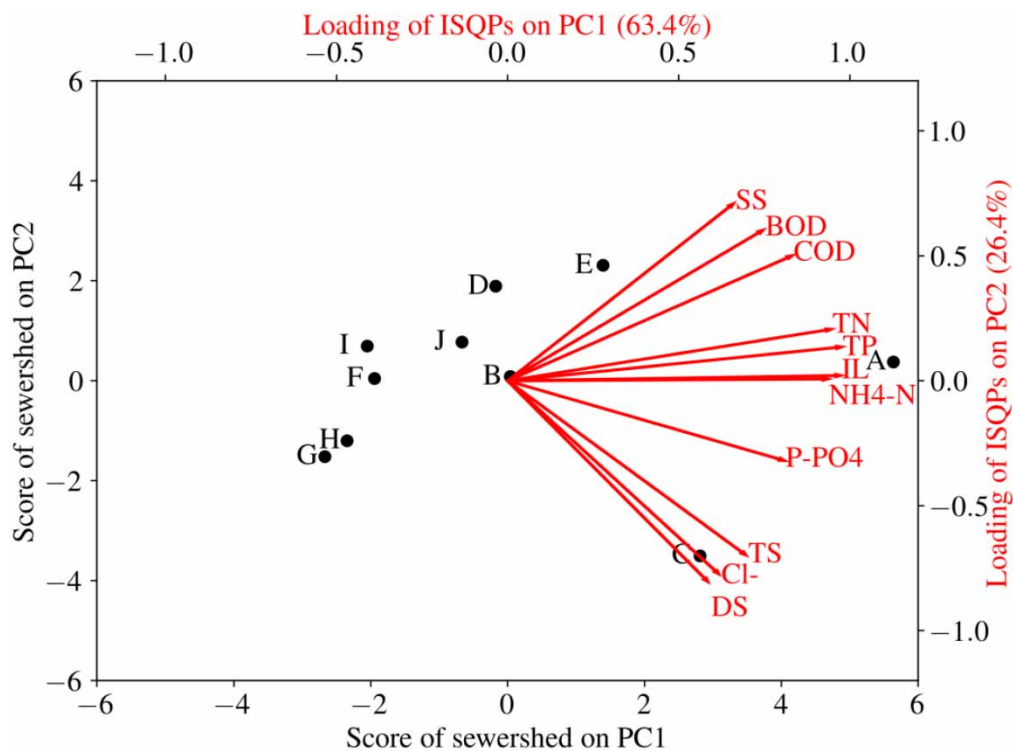


Figure 2 | Biplot of the PCA results on ISQPs for 10 studied sewersheds (based on 5-year averages from 2015 to 2019).

Table 2 | Correlation coefficients between regional characteristics and the scores of PC1 and PC2

	Regional characteristics	PC1	PC2
Population	Daytime population	0.54	0.04
	Daytime population density	0.85**	0.24
	Residential/nighttime population density	0.03	-0.01
	Number of people per household	-0.62*	-0.54
	Young population density (age 0–14 years)	-0.06	0.07
	Working-age population density (age 15–64 years)	0.13	0.47
	Elderly population density (age >65 years)	-0.31	0.40
	Young population ratio (age 0–14 years)	-0.11	-0.77**
	Working-age population ratio (age 15–64 years)	0.63*	0.35
	Elderly population ratio (age >65 years)	-0.66*	-0.06
Land use	Public area ratio	0.71*	0.27
	Commercial area ratio	0.90**	0.02
	Residential area ratio	-0.51	0.69*
	Industrial area ratio	-0.07	-0.76*
	Park ratio	0.02	-0.69*
	Unused land ratio	0.61	-0.34
	Road ratio	0.35	-0.16
	Agricultural area ratio	-0.67*	0.08
	Water area ratio	-0.06	-0.82**
	Wilderness ratio	-0.78*	-0.30
Health and income	Metabolic syndrome rate	-0.52	-0.42
	Smoking rate	-0.51	-0.33
	Inadequate sleep rate	-0.05	0.31
	Life expectancy (male)	0.30	0.39
	Life expectancy (female)	0.29	0.46

*0.01 < p < 0.05, **p < 0.01.

correlation was observed between either PCs and the indicator of health and income (metabolic syndrome rate, smoking rate, inadequate sleep rate, and life expectancy).

A significant positive correlation was observed between the daytime population density with PC1 (Supplementary Figure S1A), although no correlation was found with the residential population density or the so-called nighttime population density (Table 2). The number of people per household was negatively correlated with PC1 (Supplementary Figure S1B). Negative correlations were also observed for the elderly population ratio (age >65 years) with PC1 (Supplementary Figure S1C), while the working-age population ratio (age 15–64 years) showed different positive correlations with PC1. However, only one negative correlation was found between PC2 and population information, which is the young population age ratio (age 0–14 years) (Supplementary Figure S1D).

With respect to land use, a positive correlation was observed between the public area ratio and PC1 (Table 2). A positive correlation was also found between the commercial area ratio and PC1 (Supplementary Figure S2A). Furthermore, negative correlations were observed for agricultural ratio and wilderness ratio with PC1 (Supplementary Figure S2B). On the contrary, only one positive correlation was observed for PC2, which was the residential area ratio. Meanwhile, industrial area ratio, park ratio, and water area ratio were negatively correlated with PC2, whereas water area ratio is strongly correlated with PC2 (Supplementary Figure S2C and S2D).

Monthly ISQPs and sewershed characteristics

The correlation analyses for time-series data (monthly basis) for each sewershed showed no or weak correlations between ISQPs and economic indicators or expenditure per household (see Supplementary Tables S2–S4 for details). Regarding meteorological information, significant negative correlations between three ISQPs (i.e., BOD, TN, and TS) and the monthly average air temperature were confirmed in several sewersheds (Table 3).

Based on this correlation analysis, concentrations of BOD, TN, and TS showed significant negative correlations in six, eight, and five sewersheds, respectively (Table 3 and Supplementary Figures S3–S5), whereas no positive correlation was

Table 3 | Correlation coefficients of BOD, TN, and TS with the monthly average air temperature ($n=60$)

Sewershed	BOD	TN	TS
A	-0.65*	-0.51*	-0.46*
B	-0.46*	-0.37*	-0.04
C	0.09	-0.26*	-0.34*
D	0.26*	-0.35*	0.06
E	-0.44*	-0.49*	-0.37*
F	-0.55*	-0.48*	-0.37*
G	-0.24	-0.34*	-0.21
H	0.20	-0.12	-0.42*
I	-0.48*	-0.32*	-0.16
J	-0.15	-0.01	0.04

*0.01 < p < 0.05. Note that the temperature on the 15th of each month was adopted to represent its monthly condition in this analysis.

observed. Of those three water-quality parameters, BOD showed a higher range of correlation coefficients than others overall. Among the sewersheds, the correlation coefficient of the sewershed A, located at the center of Tokyo city, was the largest (Table 3).

Scenario analyses on shifts in population and ISQPs

Based on the multiple regression analysis, the ISQP estimation models were constructed as shown in Supplementary Tables S5 and S6. Unary regression models were built for ISQPs mainly in PC1 except for the case of TN, where a bivariate regression model was constructed, and the coefficient of determination (R^2) ranged from 0.42 (BOD) to 0.97 (TN). In addition, bivariate regression models were constructed for ISQPs mainly in PC2 using the scores of PC1 and PC2 (Supplementary Table S7). These scores of PC1 and PC2 were obtained from PCA of land-use data (detailed in Supplementary Table S7) because more correlations were found between PC2 and land-use information (Table 2). The corresponding R^2 values of those models were 0.68, 0.76, and 0.76 for Cl^- , TS, and DS, respectively.

Then, the regression model of TN (Equation (5)) was chosen as the target ISQP in the following scenario analysis because of the highest R^2 among the regression models.

$$W_{TN} = 5.44 \times 10^{-4}x_1 + 3.63 \times 10^{-2}x_2 + 18.34 \quad (5)$$

where W_{TN} is the concentration of TN (mg/L), x_1 is daytime population density, and x_2 is the ratio of the separate sewer system (Supplementary Table S8).

As shown in Figure 3, the scenario analysis predicted the TN concentration in each treatment area in 2040 when the Japanese population would decrease by 11.5% (Japan Statistics Bureau 2020). When the movement of people is restrained in the future (scenario II), TN would decrease in the areas, mainly containing many commercial establishments (Figure 3(a)–(c) and 3(e)) by 3.2–6.7%. If the population outflow in the Tokyo wards would progress (scenario III), meaning the decrease in residential population, TN would decrease in all sewersheds by 2.1–14.0%.

DISCUSSION

Organics and nutrients

Based on the loadings of each ISQP on PC1, the PC1 can be regarded as the component representing concentrations of organics and nutrients. The positive correlation of the daytime population density with the organics and nutrients was strong and significant ($r=0.85$, $p<0.01$) (Table 2 and Supplementary Figure S1A), which was mainly because these organics and nutrients originate from waste that is related to individual activities. Taking TN as an example (a major factor for PC1), 75% of TN was derived from human waste (i.e., wastewater from the toilet), and only 25% of TN was derived from domestic wastewater (i.e., wastewater other than toilet drainage, including kitchen, laundry, and shower) according to the Tokyo Metropolitan Government Bureau of Environment (2020). Such correlation is likely due to the fact that sewerage water is a mixture of different

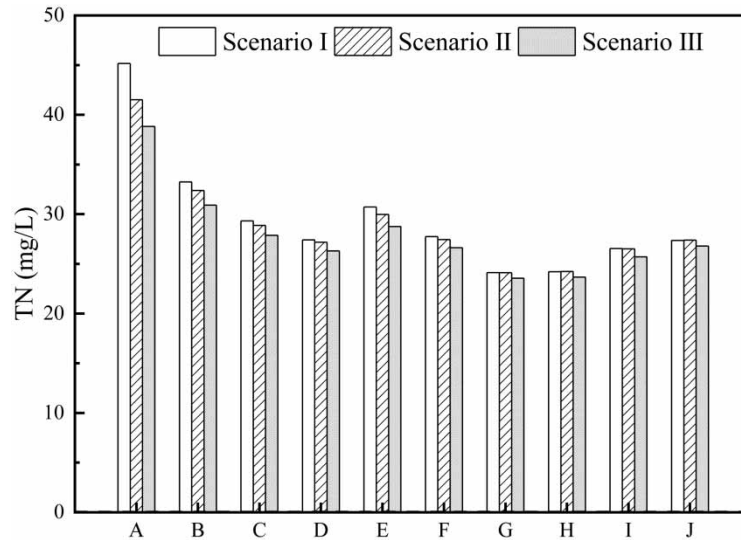


Figure 3 | Prediction of TN in 2040 based on the population forecast by the Tokyo Metropolitan Government for the scenarios: (I) the baseline value, (II) the value obtained by reducing the inflow and outflow populations by 20% from the baseline value, and (III) the value obtained by reducing the nighttime population by 10% and the inflow and outflow populations by 30% from the baseline value.

types of wastewaters, and their mass proportion varies spatially and temporally. Therefore, ISQPs possibly show some variations depending on the mass proportion of drainage types.

Meanwhile, the correlation of PC1 with residential/nighttime population density can be neglected ($r=0.03$, Table 2), inferring that sewerage quality is more closely related to daily population density than residential/nighttime population density, being reflected in concentrations of organics and nutrients. This is probably because nighttime is a period when sewage use was not as active as daytime or the sewage in nighttime is not as spatially variable as daytime. The presented results confirmed that the sewerage plan should be based on the daytime population rather than the nighttime population, although currently, the planned population for sewerage in Tokyo city was based on the nighttime population (Supplementary Table S9).

The high correlation coefficient between PC1 and the daytime population density possibly allows us to estimate the daytime population in sewersheds in Tokyo city. Such estimations of the daytime population might be valid only if the difference in wastewater loading from sectors other than domestic use is minor. Thus, further studies are required to investigate those loadings from major sectors and to verify the correlation in other cities. Nevertheless, it is worth mentioning that $\text{NH}_4\text{-N}$ has been successfully applied to estimate the population size in the metropolitan area of Lausanne in Switzerland (Been *et al.* 2014). The concentration of $\text{NH}_4\text{-N}$ can also be utilized to estimate the population size in Tokyo city, as the factor loading of $\text{NH}_4\text{-N}$ for PC1 is very high (Figure 2) and population size can be shown by the product of population density and sewershed area.

Another estimation of population size has been attempted based on human urine biomarkers such as acesulfame and caffeine (Rico *et al.* 2017). In those cases, the estimation is based on pollutant load. In contrast, our case in Tokyo found the unique property of concentration of ISQPs in relation to sewershed population, and this approach is simpler than the commonly reported approach based on chemical biomarkers and loads, although further study is required to validate the relationships.

In addition, the negative correlation between the number of people per household and PC1 (Table 2 and Supplementary Figure S1B) indicated that the larger the number of people per household, the better the influent sewage quality. This trend is clear in Supplementary Figure S6, where TN (a major factor for PC1) is taken as an example. This is probably related to the fact that solitary people tend to prefer developed commercial areas for their residence, where influent sewage quality is relatively deteriorated (Supplementary Figure S2A), and families tend to live in areas outside such commercial areas, as is also shown in Supplementary Table S10. In addition, the positive correlation of PC1 with the working-age population ratio suggested that influent sewage quality might show a higher concentration of pollutants in sewersheds, where the ratio of working-age people is higher. The negative correlation of PC1 with the elderly-age population ratio supports this inference as elderly people stay at home with their families (Supplementary Figure S1C).

Considering the positive correlation of PC1 with the commercial area ratio and public area ratio, and the negative correlation of PC1 with the agricultural area ratio and the wilderness ratio (Table 2), we may conclude that the concentration of organics and nutrients tended to be higher in urban areas with many commercial facilities, whereas it is low in the outskirts of the city, where agriculture and natural vegetation is relatively dominant. People in a megacity tend to gather in commercial areas during the daytime for business, shopping, and leisure (Li *et al.* 2016), which drastically increases the population density, whereas such a diurnal change in population density seems minor in the outskirts. Such different dynamics in daytime population density raise a load of human waste and, thus, the concentration of organics and nutrients increases. This is also a possible reason for land-use types that were not directly related to organic matter/nutrient concentrations, except for commercial land. Rather, we may infer that the land use in each sewershed affects the movement of people, which is an important factor in determining the concentrations of organic matter and nutrients. Regarding health and annual income information, no significant correlation was found with PC1, indicating that health and income indicators are not as relevant as population or land use to organics and nutrients.

Regarding the seasonal shifts, the negative correlation between the average monthly temperature (meteorological information) and concentrations of BOD and TN indicated that the influent sewage quality tended to decrease in summer and increase in winter in some sewersheds in the Tokyo wards (Table 3, Supplementary Figures S2–S4). This result is consistent with a previous study reporting that BOD removal efficiency exhibited a seasonal variation, where the removal efficiency was high in the hot summer and low in the cold winter (Griffin *et al.* 1999). We also confirmed that summer is not the season showing significantly higher water consumption compared with other seasons in Japan (Tokyo Metropolitan Government Bureau of Waterworks 2019), which means that the dilution of ISQPs may play a negligible role. Thus, the decrease of BOD was likely caused by the tendency that high temperature promotes biological degradation within sewer pipes and ultimately lowers BOD before reaching the treatment plant. In fact, the correlation of BOD with the average monthly temperature was stronger than that of TN and TS (Table 3 and Supplementary Figure S3–S5), indicating that temperature would influence BOD removal within sewer pipes more than TN and TS. In addition, the correlation coefficient of one sewershed (A) located in the center of Tokyo city was the highest among all sewersheds, while no significant correlation was observed in some sewersheds. Thus, the seasonality of lifestyle also seems to depend on the sewersheds.

Inorganic salts

Based on the loadings of ISQPs on PC2, PC2 can be considered as a component indicating inorganic salts although its score is negatively correlated to salt concentration. The negative correlation of PC2 with the young population ratio (Table 2 and Supplementary Figure S1D) is significant ($p < 0.01$). Besides, no correlation was found between PC2 and population information, meaning that inorganic salts were weakly associated with regional characteristics indicating human life.

If we consider that inorganic salts are inversely correlated to PC2 (Figure 2), the negative correlation of inorganic with the residential area ratio and its positive correlation with the industrial area (Table 2 and Supplementary Figure S2D) indicated that the concentration of inorganic salts tends to increase in areas used for nonresidential purposes. Landfill leachate can be one of the factors causing the increase in inorganic salt in areas designed for purposes other than housing. According to the Center for Material Cycles and Waste Management Research in Japan (Ogata 2014), 80% of general waste was treated by incineration, and $\text{Ca}(\text{OH})_2$ was added to absorb the acid gas generated during this process. The sewersheds along Tokyo Bay are dominated by reclaimed lands, where the treated solid waste was brought and properly filled up together with natural soil. It is also known that the leachate from the reclaimed land is collected and treated before being drained into the sewerage (Tokyo Metropolitan Government Bureau of Environment 2021). Thus, after rainwater fell on the landfill site, a large number of salts from slaked lime in the incineration residue would be rinsed into the sewer. Although the leachate is treated in advance, it might flow into the nearby WWTP containing a high concentration of inorganic salts.

In addition, concentrations of inorganic salts (TS, Cl^- , and DS) were high in the sewersheds along the coast. The intrusion of seawater with high salinity might also play a role. The study of WWTPs in Yokohama indicated that the Cl^- and Na^+ concentrations in the WWTP located near coastal areas were higher than those in other WWTPs far from the coast (Ishii & Mochizuki 1993). For instance, assuming that the concentration of Cl^- in seawater is 19,170 mg/L (Yang & Pignatello 2017) and the concentration of domestic wastewater is 53 mg/L (from the presented data in a sewershed far from Tokyo Bay), the mixing model estimated the maximum of 1.71% of the sewerage being penetrated seawater based on the highest

Cl^- concentration in a sewershed right next to Tokyo Bay (381 mg-Cl/L). The result indicates the possibility of seawater penetration to sewerage although it is a primitive estimation. Thus, it can be inferred that seawater may penetrate the groundwater in some low-altitude coastal areas. On the contrary, as the Tokyo sewerage system has a history of more than 130 years and some pipelines have cracks, it was also possible that seawater penetrates such damaged pipelines in Tokyo wards (Morikawa 2018). Besides, wastewater from factories and development sites could be responsible for the high concentration of inorganic salts. As shown in Supplementary Table S11, a strong positive correlation was observed between the number of ceramic product manufacturing industries or steel industries and inorganic salts (TS, DS, and Cl^-). These manufacturing industries handle rocks and minerals, and substances such as ionic crystals may dissolve in wastewater. The current discharge standards for sewerage in the Tokyo wards specify no regulation restricting the salt concentration, such as the Cl^- concentration or total DS. Thus, the untreated substances flow into the sewerage with no change from factories and development sites, increasing the concentration of inorganic salts.

Potential applications of ISQPs and challenges

As the multiple regression analysis results show (Supplementary Tables S5 and S6), the high regression coefficients for most ISQPs, namely, IL, TN, $\text{NH}_4\text{-N}$, TP, $\text{PO}_4\text{-P}$, TS, Cl^- , and DS imply the potential application of these models to estimate future ISQPs. The characteristics representing sewershed population were used as variables for ISQPs mainly indicating PC1 (IL, TN, $\text{NH}_4\text{-N}$, TP, and $\text{PO}_4\text{-P}$), whereas characteristics indicating land use were used for ISQPs representing PC2 (TS, Cl^- , and DS). However, the regression coefficients for BOD, COD, and SS, which basically stand for organics, were low (Supplementary Table S5). The models employed the number of people per household as the variable which appeared to be correlated with PC1 and included BOD, COD, and SS. This suggests that the further elucidation of relevant processes is necessary for the valid estimation of organics concentration in influent sewage.

As a result of scenario analysis (Figure 3), TN concentration would decrease in the areas mainly containing commercial establishments (A, B, C, and E) in the case that the movement of people is restrained in the future (scenario II). If the population outflow in Tokyo city progressed (scenario III, reducing residential population), TN concentration would decrease in all sewersheds. Even after the end of the COVID-19 situation, the style of telecommuting is expected to continue to a certain degree (Magnavita *et al.* 2021; Abulibdeh 2020) and thus the population in Tokyo city would not increase. Therefore, the society is likely to take a trajectory in between scenarios II and III in the near future, and the TN concentration would continue to decrease, whereas the decreasing rate would differ depending on the regional characteristics of each sewershed.

Herein, the scenario analysis revealed the potential effects of population change on TN concentration in the future as an example of ISQPs. The presented regression models are simple with two variables at most. Thus, the present approach could be helpful for decision-makers to estimate the potential change of ISQPs in the future and design appropriate sewage management plans. However, ISQPs regarding inorganic salts were found to have relations to land-use information, which are more complex for understanding actual processes than population-related ones. Thus, additional efforts should be made to collect and analyze the relevant data for further understanding. Although this study found the correlations taking Tokyo city as an example of a well-developed megacity, the correlations between ISQPs and sewershed characteristics could be similar in other megacities such as Shanghai, New York, and London. However, it is also worth noting that the correlations in Tokyo may not be representative of other cities with a small population or a different degree of development, although further investigations are required for such a generalization.

The RNA of SARS-CoV-2 has been detected frequently in wastewater (Medema *et al.* 2020; Randazzo *et al.* 2020; Wurtzer *et al.* 2020), making it feasible to conduct wastewater-based epidemiology to understand the infectious situation (O'Brien *et al.* 2019b; Aguiar-Oliveira *et al.* 2020; Hart & Halden 2020; Wurtzer *et al.* 2020). For example, based on the concentration of SARS-CoV-2 in wastewater, one can back-calculate the number of patients with COVID-19 (i.e., virus count per day per patient) (Hart & Halden 2020). At the same time, the load is possibly affected by the conditions of a sewershed, such as the daytime population, economic situation, and land use, as shown by ISQPs in our analysis. Therefore, the integration of ISQPs and pathogens monitored at WWTPs enables us to better estimate the infection status in a sewershed rather than using solely pathogen data. For instance, the number of patients can be combined with the daytime population, which helps us better estimate the actual ratio of patients, infectious risk, and potential spread of infection. To date, the way of integrating these data and empirical relationships has not been developed; thus, further investigations should be explored for the development of data integration and its applications.

CONCLUSIONS

The study investigated the relationship between ISQPs and sewershed characteristics in Tokyo city, and the multiple linear regression model was applied to the scenario analysis considering the COVID-19 pandemic. Based on PCA, ISQPs were categorized into PC1 indicating concentrations of organics and nutrients, and PC2 indicating concentrations of inorganic salts. The concentrations of organics and nutrients were found to significantly correlate with the population distribution in sewersheds, such as the daytime population density, family size, and age distribution. Inorganic salts showed clear correlation coefficients with land-use characteristics. These results indicate the close relationships between sewerage quality and sewershed characteristics and the applicability of a huge dataset of ISQPs to estimate sewershed characteristics and vice versa, as exemplified by the scenario analysis considering the COVID-19 pandemic.

ACKNOWLEDGEMENT

We appreciate Prof. Tatsuo Omura for his insightful discussions on this research.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Abulibdeh, A. 2020 Can COVID-19 mitigation measures promote telework practices? *J. Lab. Soc.* **23** (4), 551–576.
- Aguiar-Oliveira, M. L., Campos, A., Matos, A. R., Rigotto, C., Sotero-Martins, A., Teixeira, P. F. P. & Siqueira, M. M. 2020 Wastewater-based epidemiology (WBE) and viral detection in polluted surface water: a valuable tool for COVID-19 surveillance—a brief review. *Int. J. Environ. Res. Public Health* **17** (24), 9251.
- Been, F., Rossi, L., Ort, C., Rudaz, S., Delemont, O. & Esseiva, P. 2014 Population normalization with ammonium in wastewater-based epidemiology: application to illicit drug monitoring. *Environ. Sci. Technol.* **48** (14), 8162–8169.
- Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Hernandez, F., Reid, M., Ort, C., Thomas, K. V., Van Nuijs, A. L. N., De Voogt, P. & Zuccato, E. 2013 Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. *Environ. Sci. Technol.* **47** (3), 1452–1460.
- Choi, P. M., Tschärke, B., Samanipour, S., Hall, W. D., Gartner, C. E., Mueller, J. F., Thomas, K. V. & O'Brien, J. W. 2019 Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. *Proc. Natl. Acad. Sci. U.S.A.* **116** (43), 21864–21873.
- Griffin, D. M., Bhattarai, R. R. & Xiang, H. J. 1999 The effect of temperature on biochemical oxygen demand removal in a subsurface flow wetland. *Water Environ. Res.* **71** (4), 475–482.
- 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.
- Hou, C. Z., Chu, T. T., Chen, M. Y., Hua, Z. D., Xu, P., Xu, H., Wang, Y. M., Liao, J. & Di, B. 2021 Application of multi-parameter population model based on endogenous population biomarkers and flow volume in wastewater epidemiology. *Sci. Total Environ.* **759**, 143480.
- Ishii, A. & Mochizuki, M. 1993 Investigation of Salinity in Relation to the Effective Use of Treated Sewage Water (in Japanese). Available from: https://www.city.yokohama.lg.jp/kurashi/machizukuri-kankyo/kasen-gesuido/gesuido/shori/sokutei/chosa/sonota.files/0060_20180827.pdf (accessed 30 December 2020).
- Japan Statistics Bureau 2020 *Japan Statistical Yearbook*. Available from: <https://www.stat.go.jp/data/nenkan/70nenkan/zenbun/jp70/book/book.pdf> (accessed 30 December 2020).
- Li, M. Y., Shen, Z. J. & Hao, X. H. 2016 Revealing the relationship between spatio-temporal distribution of population and urban function with social media data. *Geojournal* **81** (6), 919–935.
- Lin, W. T., Zhang, X. H., Tan, Y. Z., Li, P. & Ren, Y. 2019 Can water quality indicators and biomarkers be used to estimate real-time population? *Sci. Total Environ.* **660** (10), 603–610.
- Magnavita, N., Tripepi, G. & Chiorri, C. 2021 Telecommuting, off-time work, and intrusive leadership in workers' well-being. *Int. J. Environ. Res. Public Health* **18** (7), 3330.
- 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.
- Morikawa, N. 2018 Countermeasures against aging of sewage pipe. Annual Technical Survey of the Tokyo Metropolitan Sewerage Authority 42, 4-1-1.
- O'Brien, J. W., Thai, P. K., Eaglesham, G., Ort, C., Scheidegger, A., Carter, S., Lai, F. Y. & Mueller, J. F. 2014 A model to estimate the population contributing to the wastewater using samples collected on census day. *Environ. Sci. Technol.* **48** (1), 517–525.

- O'Brien, J. W., Choi, P. M., Li, J. Y., Thai, P. K., Jiang, G. M., Tschärke, B. J., Mueller, J. F. & Thomas, K. V. 2019a Evaluating the stability of three oxidative stress biomarkers under sewer conditions and potential impact for use in wastewater-based epidemiology. *Water Res.* **166**, 115068.
- O'Brien, J. W., Grant, S., Banks, A. P. W., Bruno, R., Cater, S., Choi, P. M., Covaci, A., Crosbie, N. D., Gartner, C., Hall, W., Jiang, G. M., Kaserzon, S., Kirkbride, K. P., Lai, F. Y., Mackie, R., Marshall, J., Ort, C., Paxman, C., Pichard, J., Thai, P., Thomas, K. V., Tschärke, B. & Muellier, J. F. 2019b A national wastewater monitoring program for a better understanding of public health: a case study using the Australian Census. *Environ. Int.* **122**, 400–411.
- Ogata, Y. 2014 *Landfill Leachate-A Comparison Between Japan and Southeast Asia*. Available from: <https://www-cycle.nies.go.jp/magazine/mame/201403.html> (accessed 30 December 2020).
- Ort, C., Lawrence, M. G., Rieckermann, J. & Joss, A. 2010 Sampling for pharmaceuticals and personal care products (PPCPs) and illicit drugs in wastewater systems: are your conclusions valid? A critical review. *Environ. Sci. Technol.* **44** (16), 6024–6035.
- Pandopoulos, A. J., Bade, R., Tschärke, B. J., O'Brien, J. W., Simpson, B. S., White, J. M. & Gerber, C. 2021 Application of catecholamine metabolites as endogenous population biomarkers for wastewater-based epidemiology. *Sci. Total Environ.* **763**, 142992.
- Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simon, P., Allende, A. & Sanchez, G. 2020 SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Res.* **181**, 115942.
- Rico, M., Andrés-Costa, M. J. & Picó, Y. 2017 Estimating population size in wastewater-based epidemiology. Valencia metropolitan area as a case study. *J. Hazard Mater.* **323** (A), 156–165.
- Thomaidis, N. S., Gago-Ferrero, P., Ort, C., Maragou, N. C., Alygizakis, N. A., Borova, V. L. & Dasenaki, M. E. 2016 Reflection of socioeconomic changes in wastewater: licit and illicit drug use patterns. *Environ. Sci. Technol.* **50** (18), 10065–10072.
- Thomas, K. V. & Reid, M. J. 2011 What else can the analysis of sewage for urinary biomarkers reveal about communities? *Environ. Sci. Technol.* **45** (18), 7611–7612.
- Tokyo Metropolitan Government Bureau of Environment 2020 Measures Against Domestic Wastewater (in Japanese). Available from: https://www.kankyo.metro.tokyo.lg.jp/water/pollution/life_drainage/life_drainage.html (accessed 30 December 2020).
- Tokyo Metropolitan Government Bureau of Environment 2021 *Tokyo Metropolitan Government Waste Landfill Pamphlet*. Available from: https://www.kankyo.metro.tokyo.lg.jp/data/publications/resource/pamphlet_list.files/english_pamphlet2021.pdf (accessed 7 April 2022).
- Tokyo Metropolitan Government Bureau of Sewerage 2021 Spread of Public Sewerage in 23 Wards (in Japanese). Available from: <https://www.gesui.metro.tokyo.lg.jp/living/a2/spread/> (accessed 7 April 2022).
- Tokyo Metropolitan Government Bureau of Waterworks 2019 Sewerage System Report. Chapter 4.6: Water Allocation (in Japanese). Available from: https://www.waterworks.metro.tokyo.lg.jp/suido/jigyogaiyou/nenpou_r01.html (accessed 30 December 2020).
- Tschärke, B. J., O'Brien, J. W., Ort, C., Grant, S., Gerber, C., Bade, R., Thai, P. K., Thomas, K. V. & Mueller, J. F. 2019 Harnessing the power of the census: characterizing wastewater treatment plant catchment populations for wastewater-based epidemiology. *Environ. Sci. Technol.* **53** (17), 10303–10311.
- Van Nuijs, A. L., Mougél, J. F., Tarcomnicu, I., Bervoets, L., Blust, R., Jorens, P. G., Neels, H. & Covaci, A. 2011 Sewage epidemiology—a real-time approach to estimate the consumption of illicit drugs in Brussels, Belgium. *Environ. Int.* **37** (3), 612–621.
- Wurtzer, S., Marechal, V., Mouchel, J. M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J. L. & Moulin, L. 2020 Time course quantitative detection of SARS-CoV-2 in Parisian wastewaters correlates with COVID-19 confirmed cases. *MedRxiv*. <https://doi.org/10.1101/2020.04.12.20062679>.
- Yang, Y. & Pignatello, J. J. 2017 Participation of the halogens in photochemical reactions in natural and treated waters. *Molecules* **22**, 1684.
- Zhang, Y. Z., Duan, L., Wang, B., Du, Y. L., Cagnetta, G., Huang, J., Blaney, L. & Yu, G. 2019 Wastewater-based epidemiology in Beijing, China: prevalence of antibiotic use in flu season and association of pharmaceuticals and personal care products with socioeconomic characteristics. *Environ. Int.* **125**, 152–160.
- Zuccato, E., Chiabrando, C., Castiglioni, S., Bagnati, R. & Fanelli, R. 2008 Estimating community drug abuse by wastewater analysis. *Environ. Health Perspect.* **116** (8), 1027–1032.

First received 9 January 2022; accepted in revised form 19 May 2022. Available online 1 June 2022