Wastewater data from individual homes: Quantitative and qualitative measurements

Vivien Dubois, Eva Falipou, Claire Lauvernet, Catherine Boutin

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Wastewater data from individual homes: Quantitative and qualitative measurements

Vivien Dubois [a,*], Eva Falipou [a], Claire Lauvernet [b], Catherine Boutin [a]

[a] *INRAE REVERSAAL, 69625 Villeurbanne Cedex France*
[b] *INRAE RIVERLY, 69625 Villeurbanne Cedex France*

A R T I C L E   I N F O

A B S T R A C T

The data article is related to the generation of raw wastewater from 15 single-family dwellings. Such a dataset is rare. They are complicated to set up due to the technical difficulty of deriving a representative sample of this kind of pollution, composed of suspended solids potentially very coarse in size, compounded by the administrative difficulty of performing these measurements on private property. The data were obtained by means of two distinct and complementary monitoring campaigns: i) continuous measurement of the volumes discharged by three of the 15 dwellings during more than one year, and ii) characterization, in terms of both quality and quantity, of more than 300 raw wastewater samples discharged over 24 consecutive hours by all 15 houses during weekly periods of seven consecutive days, which deliberately included weekend days. Their acquisition methodologies are described. The validation's methodologies, also described, include a R script. The data analysis can be used for i) designing and operating onsite and small wastewater treatment systems and ii) characterising the strictly domestic pollution. Therefore, this data article is associated with the manuscript "Quantification and qualification of the urban domestic pollution discharged per households and per resident".The hydraulic values obtained with continuous measurement are available in two units: in l.s^{-1} as primary data and in l.h^{-1} as secondary data. They can be reused at any

* Corresponding author.
  E-mail address: vivien.dubois@inrae.fr (V. Dubois).

time step. These elements could be incorporated into future research or innovations related to single-family dwellings. In a general point of view, such data could be introduced in databases used for Life Cycle Assessment. These values can also be used in sanitation (collection, design and operation) and be useful to better specify the potential for the recovery of domestic wastewater.

Specifications Table

Subject	Water Science and Technology
Specific subject area	Wastewater treatment and sanitation
Type of data	Table, figure
How the data were acquired	**Volume data (campaign i)**: Water height at a pumping station measured by a piezometer-type pressure sensor (ATM/N range 0-1 m, manufactured by STS), connected to a data acquisition unit (Minilog B RDL10, built by Endress Hauser).
	Raw wastewater quantity and quality data (campaign ii):
	Method 1: collection of all the raw wastewater discharged from the house for 24 hours within a 1,000-liter graduated tank fed by a pumping station and placed in a shaded spot. Homogenization using a drum pump (B2 Vario Niro, manufactured by Lutz). Sampling for physicochemical analysis. Daily volume determined by reading the water level in the storage tank.
	Method 2: sampling of the wastewater in sync with its rate of inflow, downstream of the existing pumping station. Sampling upon each pump activation stroke by virtue of a water sensor-presence detector that triggered a response from the refrigerated automatic sampler. Daily volume acquired according to two distinct protocols: a) by multiplying the number of pump strokes by the average pumped volume (*Method 2a*) and b) actual knowledge of the pumped volumes (*Method 2b*).
	Parameters analysed: COD, BOD_5, SS, KN, NH_4^+-N and TP. The standards implemented to analyse these parameters were dependent on each laboratory and have been listed in **supplementary material**.
Data format	Raw, filtered, analysed
Description of data collection	The dwelling selection criteria are as follows:
	• serving as a main residence,
	• being used solely for household activities(without any connection to the farming or agrofood processing sector,
	• featuring a conventional plumbing installation, devoid of any dry toilets or use of recycled water inside.
	These selection criteria are not stringent. The dwelling selection was based primarily on technical feasibility criteria relative to the requisite monitoring in private space.
Data source location	The 15 houses are distributed throughout the French national territory:
	campaign i):
	• H_1 in the Rhône department (69)
	• H_{10} in the Saône-et-Loire department (71)
	• H_{11} in the Drôme department (26)
	campaign ii):
	• H_1 in the Rhône department (69)
	• H_2, H_3, H_4, H_6 and H_8 in the Côtes-d'Armor department (22)
	• H_5, H_7, H_{13}, H_{14} and H_{15} in the Tarn department (81)
	• H_9 in the Meuse department (55)
	• H_{10} in the Saône-et-Loire department (71)
	• H_{11} in the Drôme department (26)
	• H_{12} in the Eure-et-Loir department (28)

Data accessibility	The data of the campaign i) and ii) have been deposited to Data INRAE. The open source format is .csv file or zip form.
	Repository name: Continuous measurement of the volumes discharged from three French single-family homes
	Data identification number: doi:10.15454/HMTSXZ
	Direct URL to data: https://doi.org/10.15454/HMTSXZ
Related research article	V. Dubois, E. Falipou, C. Boutin, *Quantification and qualification of the urban domestic pollution discharged per household and per resident (2022)*, Water Science and Technology, 85 (5) 1484-1499. https://doi.org/10.2166/wst.2022.064.

Value of the Data

- These data are useful as they characterize simultaneously the volumes and concentrations of raw wastewater discharged by households with a known number of residents.
- This dataset is rare due to both the technical difficulty involved in conducting a sampling program representative of actual pollution and the administrative difficulty of performing these measurements on private property.
- The sanitation community can directly benefit from these data.
- The data can be used to design onsite and small wastewater treatment systems.
- These data may be used for characterising the strictly domestic pollution.
- The dataset could be useful in further research on the recovery of domestic wastewater.
- The dataset could also enrich the databases used for Life Cycle Assessment.

1. Data Description

The dataset given in the supplementary data corresponds to: i) three series of hourly volumes measured for at least 21 months and ii) 302 daily quantitative and qualitative characterizations of raw wastewater, sampled so as to represent the discharge over 24 consecutive hours. The three houses H_1, H_{10} and H_{11} underwent two monitoring campaigns i) and ii), while the other 12 houses, i.e. H_2 to H_9 and H_{12} to H_{15}, were solely involved in campaign ii).

The composition of the resident families was known and varied across households, from a single person to families composed of several adults (2 or 3), along with a variable number of children, from none to four. The maximum number of residents in any single house was six. The residents' ages were not known with any precision; only the distinction between adult and child, based simply on age recorded as younger or older than 18, was made. The 15 houses were sorted by increasing number of adult residents, then within each adult number category by increasing number of children residents, thus leading to the numbering system chosen for these 15 houses. The term "resident" refers to any individual who typically occupies, as a primary place of residence, the dwelling unit. Hence, a child attending boarding school during the week is a "resident" despite not being systematically present when pollution measurements are performed. Conversely, guests staying at the unit are not counted as "residents" even though the additional on-site activity generated by their stay and/or presence may be included in measurements.

All the parameters of the different households are summarised in the Table 1 below:

Table 1
Parameters of the different households.

Households	French department	Number of resident		24 h composite sample	
		Adults	Children	Campaign i) (storage tank)	Campaign ii) (2a: number of pump stroke, 2b: pumped volume)
H1	69 (Rhône)	1	0	1	
H2	22 (Côtes-d'Armor)	2	0		2a
H3	22 (Côtes-d'Armor)	2	0		2a
H4	22 (Côtes-d'Armor)	2	0		2a
H5	81 (Tarn)	2	0	1	
H6	22 (Côtes-d'Armor)	2	1		2a
H7	81 (Tarn)	2	1	1	
H8	22 (Côtes-d'Armor)	2	2		2a
H9	55 (Meuse)	2	2		2a
H10	71 (Saône-et-Loire)	2	3	1	
H11	26 (Drôme)	2	4		2a
H12	28 (Eure-et-Loir)	2	4		2b
H13	81 (Tarn)	3	1	1	
H14	81 (Tarn)	3	1	1	
H15	81 (Tarn)	3	2	1	

The dwelling selection criteria are not extremely stringent; the dwelling selection was based primarily on technical feasibility criteria relative to the requisite monitoring in private space.

The data file contains, under the "Comments" tab: the locations of the 15 houses, the number of adults and children residents, the data acquisition methodologies for the two distinct but complementary monitoring campaigns, as well as the chemical analysis methods. For the continuous monitoring of raw wastewater volumes (campaign i), the pumping station surface areas are also given in the "Comments" file. The precise measurement dates and validated hourly flow rates are presented in three "Volumes" files individualized by houses H1, H10 and H11. For campaign ii), the "Raw wastewater characterization" file lists: house references, precise measurement dates, values of the validated measurement concentrations, and daily volumes.

24h composite samples (campaign ii) were acquired during weekly periods of seven consecutive days, which deliberately included weekend days. These sampling campaigns were conducted between one and five times, thereby covering several seasons. The Table 1 shows the number of weekly monitoring and the number of 24h composite samples available.

For the continuous monitoring of raw wastewater volumes (campaign i), the acquisition periods, began in November 2013 for the two sites H1 and H10 and later, in February 2015, for H11. Lasting at least 23 months, the durations are ranging from 1 year, 8 months to 2 years, 2 months depending on the dwelling and T the periods of missing data, resulting from technical issues (e.g. empty batteries, defective water height recorder devices). The number of continuous daily monitoring are 611 d, 533 d and 432 d respectively for dwellings H1, H10 and H11.

The water heights data recorded during the continuous monitoring are deposited in the Data INRAE repository as primary data (see [1]).

The volumes data may be processed with respect to various time steps, i.e. day, week, month, season and year. For example, it is possible to generate the average daily hydrographs of produced raw wastewater, by hourly interval, and in so doing reflect the household activity taking place inside the dwelling. Fig. 1, based on 25,584 data points, displays the hydrograph of house H10, at a daily scale, in the form of 24 box plots representing the median, 1st and 3rd quartiles, and upper bound. The left panel of the figure has been produced with the validated complete dataset, while the right panel was generated using a dataset that excluded all statistically outlying values (i.e. beyond the upper bound), thus reducing the number of data points by 9 %. This result does in fact correspond to a close-up image of the left part of the graph with respect to amplitude: [0 - 100 L.h-1]. Fig. 2 shows the specific consumption each day of the week, as visualized using the seven average daily hydrographs for the seven days of the week.
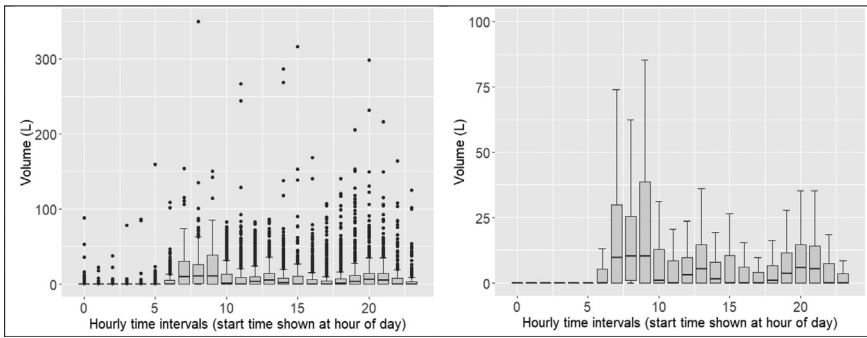
Fig. 1. Average daily hydrographs (volume in L / hourly interval) covering the entire data acquisition period of house H_{10}: with the validated complete dataset (left), after removal of the statistically outlying values (right).
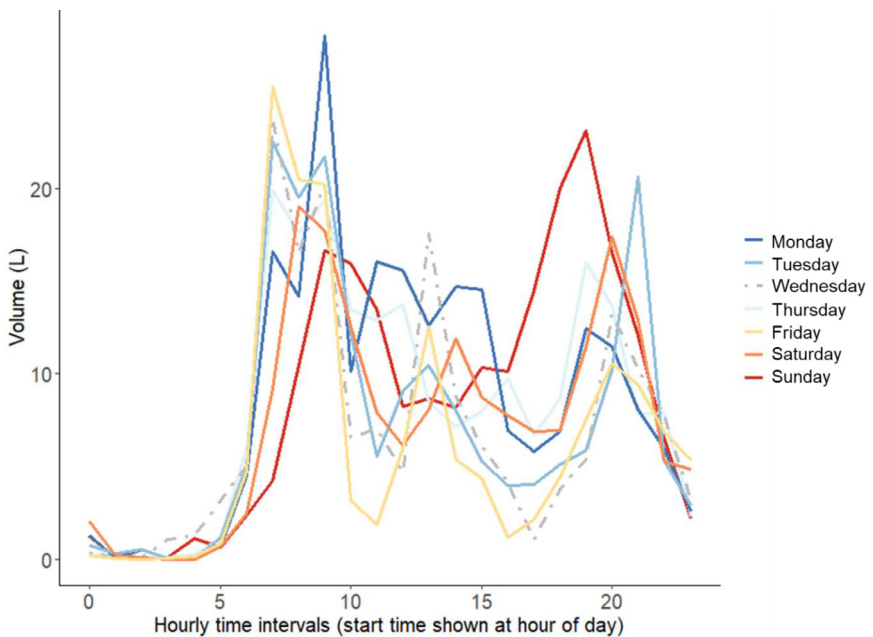
Fig. 2. Average daily hydrographs of raw wastewater, covering the entire data acquisition period, by day of week for house H_{10}.

Fig. 3 schematically depicts the two distinct modes of data acquisition used for the characterization of the raw wastewater.

Wastewater composition data are numerous, and the ratios calculated between the various chemical forms serve to confirm the strictly domestic nature of the wastewater to be treated. An example presented in Fig. 4 refers to the 100 TP/COD ratio. Since phosphorus mainly originates from human activity alone, or from washing, this ratio may be used to trace a given daily household activity. Fig. 4 indicates the average of the 1 to 5 values of the 100 TP/COD ratio available by day of the week as well as by house. This evaluation applies to the 13 households able to provide the seven daily data points that enable comparison at the weekly scale. Table 2.
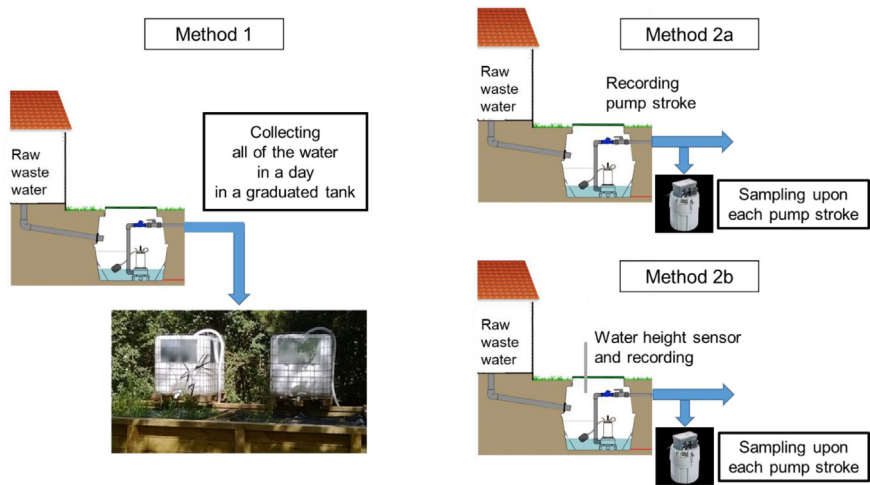
Fig. 3. Schematic diagram of the data acquisition methods for characterizing raw wastewater.
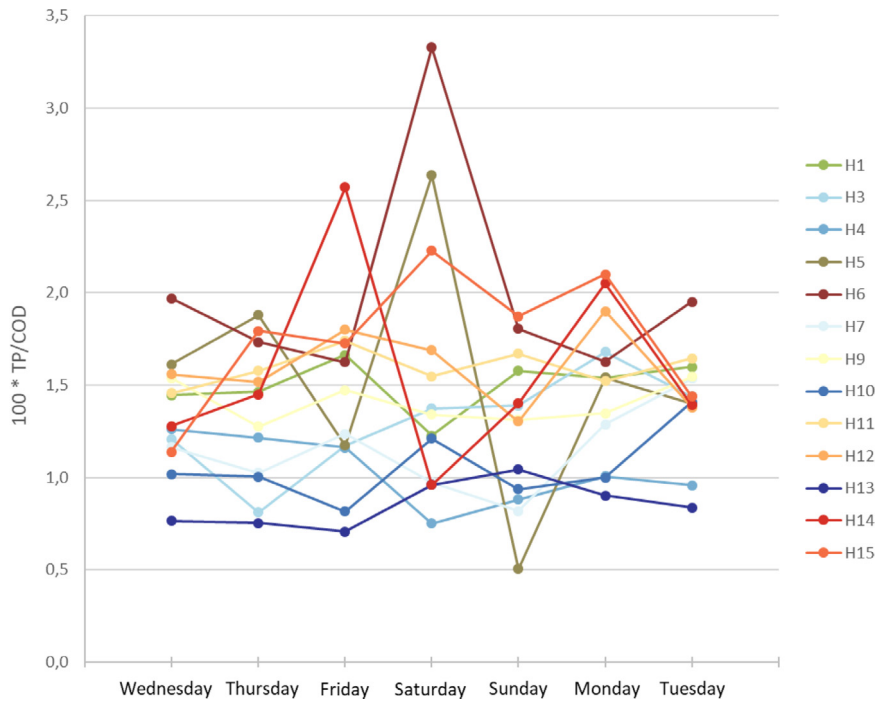

Fig. 4. Weekly trend in the average daily 100 TP / COD ratios of the effluent from 13 households.

Table 2
Campaign ii) monitoring conducted in each house.

House	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}
Number of weekly monitorings	5	2	3	3	2	4	2	2	5	4	5	5	4	1	2
Number of 24h composite samples available	34	10	17	18	14	26	14	8	35	25	30	29	21	7	14

2. Experimental Design, Materials and Methods

For a better understanding you can refer to Table 1.

2.1. Continuous monitoring of raw wastewater volumes (campaign i)

Quantitative monitoring was performed by means of load flow measurements:

- a piezometer-type pressure sensor measured the water height at a pumping station, controlled by float switches.
- the pressure sensor was connected to a data acquisition unit that recorded water height values at a given frequency.

The pump station was cylindrical in shape. Its effective surface area, which remained constant, was measured by calculating the ratio between the volume discharged by the pump during emptying and the height variation induced by this emptying step. These water heights were recorded every minute (see [1]) and multiplied by the calculated unit surface area to get the flow rates in volume per minute.

The probe detected amplitude variations of at least 1 mm. The validation step detects which small amplitude of water corresponds to either the raw wastewater contribution or the emptying sequences. These were episodes of slight turbulence at the surface water potentially caused by falling condensation droplets. As such, all deviations with an absolute value of 1 mm were retained, provided that they were the direct consequence of other deviations equal to or greater than 1 mm. The other deviations, grouped with the zero deviations, were systematically deleted. A script written in R language computing this data processing is available in a specific supplementary material. Based on validated water height values, it was possible to calculate the flow rates for each hour defined in the GMT+1 time zone during winter and in GMT+2 during summer (see [2]).

2.2. Characterization, in terms of both quantity and quality, of the raw wastewater (campaign ii)

The data were collected from September 2010 to March 2012 for five houses (H5, H7, H13, H14 and H15) and from September 2013 to February 2017 for the ten others (H1, H2, H3, H4, H6, H8, H9, H10, H11, H12). Samples were performed during weekly periods of seven consecutive days, which deliberately included weekend days. Depending on the particular situation, these sampling campaigns were conducted between one and five times, thereby covering several seasons.

The data were recorded according to two distinct modes.

- Method 1
  - collecting all the raw wastewater discharged from the house for 24 hours within a 1,000-liters graduated tank, fed by a pumping station and placed in a shaded spot
  - homogenizing the volume collected using a drum pump (B2 Vario Niro, manufactured by Lutz)
  - pumping a portion of the wastewater to a bucket

○ homogenizing the portion of wastewater
○ bottling for laboratory analysis
○ emptying, rinsing with clean water and draining the graduated tank before being placed back into use.

The daily volume was determined by reading the water level in the storage tank, which had previously been marked for every 50-liter increment.

- Method 2: flow-sensitive sampling
  ○ maintaining submerged the extraction hose in the pipe connected to the wastewater treatment plant. The end of this extraction hose is not fitted with a screen
  ○ emptying the pipe of old wastewater during 15 seconds after starting the pump.
  ○ simultaneously, homogenizing the volume of the lift station
  ○ then starting the wastewater sampling at each pump activation stroke by virtue of a water sensor-presence detector that triggered a response from the refrigerated automatic sampler. The end of the extraction hose, not fitted with a screen, was submerged in the pipe connected to the wastewater treatment plant

The daily volume was collected using two different procedures. The first one (method 2a) consists in multiplying the number of pump strokes by the average volume pumped. The other method (method 2b) is based on actual knowledge of the pumped volumes.

Both methods have been used here. Since the raw wastewaters are extremely fresh and contain much solid matter, the storage of all wastewater produced during a given day was preferred, namely as a means of guaranteeing that all suspended solids would be taken into account. Denied the possibility of installing and feeding the two storage tanks and provided the presence of a pumping station, method 2 was selected. Accordingly, method 1 was adopted for seven households (H1, H5, H7, H10, H12, H13 and H14) and Method 2 for the eight others (H2, H3, H4, H6, H8, H9, H11 and H15). This chosen method is mentioned in the data set.

The parameters analysed are COD, BOD5, SS (Carbon parameters), KN, NH4+-N and TP (Nutrients parameters) (see [2]). The physicochemical analyses were carried out within Cofrac-certified laboratories or facilities with equivalent credentials, in strictly respecting the mandated analysis timeline. The different analytical standards used by each analytical laboratory and for each parameter are given in the supplementary material.

The method for validating the physicochemical data required the following:

- the strict adherence to sample storage conditions and analysis time. Sixteen samples from house H9 collected at the end of the week were kept in the freezer. A column in the «Raw wastewater characterization » line in the supplementary material indicates this conservation mode.
- a comparison of the concentrations of nitrogen forms: verification that the NH4+-N concentrations were indeed less than the KN concentrations. Otherwise, the set of parameters corresponding to this sampling was excluded from the dataset.

Ethics Statements

Informed consent of all participants has been obtained.

CRediT Authors Statement

Vivien Dubois: Conceptualization, Methodology, Software, Writing – original draft preparation, Reviewing and Editing; **Eva Falipou:** Data curation, Methodology, Validation, Software, Writing – original draft preparation; **Claire Lauvernet:** Data curation, Software, Writing – review & editing; **Catherine Boutin:** Supervision, Methodology, Validation, Writing – review & editing.

Declaration of Competing Interest

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

Data Availability

Continuous measurement of the volumes discharged from three French single-family homes (Original data) (Dataverse).

References

[1] V. Dubois, Continuous measurement of the volumes discharged from three French single-family homes, Data INRAE (2021), doi:10.15454/HMTSXZ.
[2] V. Dubois, E. Falipou, C. Boutin, Quantification and qualification of the urban domestic pollution discharged per household and per resident, Water Sci. Technol. 85 (5) (2022) 1484–1499, doi:10.2166/wst.2022.064.