Principal component analysis of raw domestic sewage

A. Breton*, **, C.Vialle*, C. Albasi***, M. Montréjaud-Vignoles*, C. Vignoles****, P. Gallien** and C. Sablayrolles*

* Laboratoire de Chimie Agro-industrielle (LCA), Université de Toulouse, INRA, INPT, Toulouse, France
(E-mail: *claire.vialle@ensiacet.fr)*
** ITREN, Traitement, épuration des eaux usées résiduaires, 82 600 Verdun sur Garonne, France
*** Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France
**** Veolia Eau ; Direction Régionale Sud-Ouest Générale des Eaux, 22 avenue Marcel Dassault, BP 5873, 31506 Toulouse, France

Abstract
This study aims to characterize raw domestic wastewater directly at the exit of individual dwellings before any pretreatment. For this purpose, wastewater sampling campaigns were conducted from six installations (family houses) during seven consecutive days. Eight parameters to characterize wastewater were determined: chemical oxygen demand, biochemical oxygen demand, nitrogen, phosphorus, ammonium, nitrate, nitrite, suspended solids. The principal component analysis (PCA) was used to exploit the results. Significant differences in concentrations were observed from one family to another, but also within the same family between the two wastewater campaigns. The volume of wastewater treated relate to population equivalent is 1.5 smaller than those conventionally used in the design of on-site wastewater treatment. The PCA showed that several parameters of physicochemical characterization of wastewater were interrelated, as chemical oxygen demand, biochemical oxygen demand and total suspended solids. Some parameters were also related to the use of water in homes, as the number of shower and ammonium ion concentration.

Keywords
On-site wastewater treatment; Physicochemical quality; Principal component analysis; Wastewater

INTRODUCTION

Wastewater treatment is an essential part of today's society care because pollution discharge is constantly increasing with the development of human activities. Wastewater or 'sewage', highly polluted, may have a high environmental impact and pose a health risk, which is the reason why this discharge must be treated to preserve valuable water resources. Hence, a network of wastewater treatment plants (WWTP) has been established to process the discharges and thus reduce pollution. However, certain zones (especially in rural areas) are not connected to this WWTP network. In that case, the only alternative is to set up an on-site sewage facility (OSSF). In France, about 13 million people are in this category (Lakel et al. 2006). Now the European Water Framework Directive (Directive 2000/60/EC) implemented to protect the aquatic environment, requires Member State to improve the wastewater treatment in all areas. In that context, OSSF in France have to be composed of pretreatment and treatment systems which allow to reduce wastewater pollution in order to respect the limit values indicated in the French, 7 September 2009 decree (Decree September 7, 2009). In general, on-site system facility was composed of a septic tank which received all wastewater from the house and in which a pretreatment of wastewater was performed. The pretreatment system was followed by a treatment system, using soil infiltration or another system, in order to continue reduction of wastewater pollution. After treatment, treated wastewater was discharged to an outlet or infiltrated in soil.

Flows and global pollution loads arriving in WWTP are quite well known. Even if there are daily and seasonal fluctuations, the high number of users connected to the sanitation system permits to

smooth the quality of wastewater. That is another consideration when only one house is considered. On-site wastewater treatment system received all wastewaters from this house. The domestic wastewater is composed of “black water” from toilet and “grey water” from laundry, kitchen, bathroom and is variable in quality and quantity. There are actually few data concerning quality of the actual discharge from individual dwellings although they would be useful for on-site treatment. However, it is important to characterize the physicochemical parameters (biological and chemical oxygen demand (BOD_5; COD), total suspended solids (TSS), nitrogen parameters (TKN; NH_4^+), total phosphorus (tot-P)…) of this household wastewater effluent to be able to correctly size the installation and allow proper treatment before the discharge into the natural environment.
The application of principal component analysis (PCA) allows to analyse large datasets. PCA identify groups and sets of variables with similar properties and may allow us to simplify our description of observations by finding the structure or patterns in chaotic or confusing datasets.
The aim of this study is to characterize the household effluents of individual dwellings that is to say influents from OSSF before their entry into the pre-treatment system, and to look for any possible correlations between the different parameters studied.

MATERIALS AND METHODS

Sample collection and analysis

Six sites in the south west of France have been equipped with an on-site sewage facility (OSSF) for this study. On site details are given in Table 2. The wastewater treatment system are sized to absorb 6 or 7 population equivalents (PE), where 1 PE corresponds to 60 g/day BOD_5 discharged per day and per person (Directive 91/271/CEE). All the installations studied have been constructed on the same plan: a buffer tank for raw wastewater, a septic tank, a treatment system, a drainfield or an outlet.

Table 2. Information on the six sites monitored

Site	A	B	C	D	E	G
Occupancy	3 adults 1 child 2 babies	2 adults 1 teenager	2 adults 2 teenagers	2 adults	5 adults	2 adults
Type	Main residence					Second home

Samples have been taken from the house wastewater inflow before entry into the septic tank, every 24 h for a whole week. Two sampling campaigns have been conducted at the six places. By using a swiveling T piece on the inflow pipe, wastewater from the house could be either directed into the septic tank for treatment, or for 24 h into the 200 L buffer tank upstream, for sampling. The latter wastewater was then transferred to a $1m^3$ tank. This system permits to collect all the wastewater from the house over 24 h. The volume discharged over 24 h was evaluated, and then wastewater was homogenized and sampled (Figure 1).

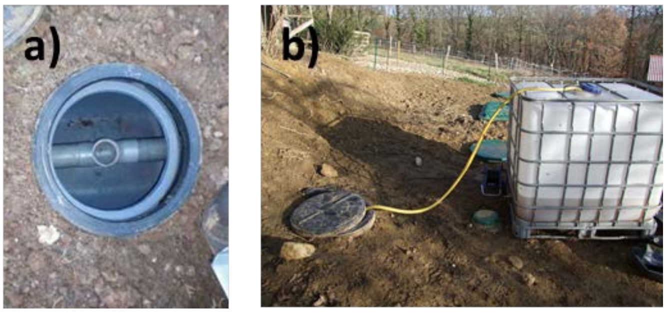

Figure 1: Raw wastewater sampling system. a) Buffer tank with tee tube – b) 1 m^3 tank

The wastewater was sampled as follows. The contents of the 1 m^3 tank, thoroughly mixed, were sampled via a tap into a bucket in three steps, a third at a time, and mixing up the waters in between to ensure homogeneity. Three, one liter sample flasks, were then filled up in three steps from the bucket in a similar manner. COD, BOD_5, TKN, NH_4^+, tot-P, TSS were determined. All analyses have been made within 24 h.

The inhabitants of the houses have been given a questionnaire at the beginning of each campaign, to list information concerning their water and products usage over the seven-day sampling period. These provided data on the numbers of toilet flushes, machine washes, meals and showers taken.

Principal component analyses

When there are a large number of samples characterized by numerous parameters, it is useful to employ multivariate statistical analyses to characterize the results obtained in turn. Principal component analysis (PCA) has been made using XL stat. software. The main idea of PCA is to represent all the data in a reduced space without losing any of the information, in order to allow easier interpretation of the data series. PCA transforms the variables linked to each other into fresh variables, called principal components, which are not themselves correlated. This enables analysis of the relations between the variables and the grouping of individuals. To use PCA, the variables and the individuals must be defined. The representation of the variables in the new space formed by the principal components, allows possible relations between these variables to be determined. Similarly, the representation of the individuals in the new space formed by the principal components, allows determination of whether the individuals are dispersed or not.

RESULTS AND INTERPRETATIONS

Parameters analyses results

The results are shown in figure 2 as Box Plots. These give a graphical representation of the data spread. They show the average (+), the median (-), the maximum and minimum values of the data series as spots, plus the 1st and 3rd quartile (box extremities). Box Plots highlight extreme values (minimum and maximum) of the data series, e.g. a data series with maximum values very different from the rest of the series, will have a Box Plot which is very elongated towards the high values. The sampling campaigns are identified on the graphs by the number ‘1’ or ‘2’ next to the site name. The French, 7 September 2009 decree describes agreement for treatment performance evaluation of OSSF. The horizontal lines on the graphs are the maximum and minimum values for the different parameters COD, BOD_5 and SS, as stipulated for raw wastewater under this agreement. Concerning volume, the horizontal line corresponds to the value recognized as reference. For the parameters NTK and tot-P, horizontal lines refers to maximum and minimum value from the assessment protocol established by a French national agency (AFSSET, 2009).

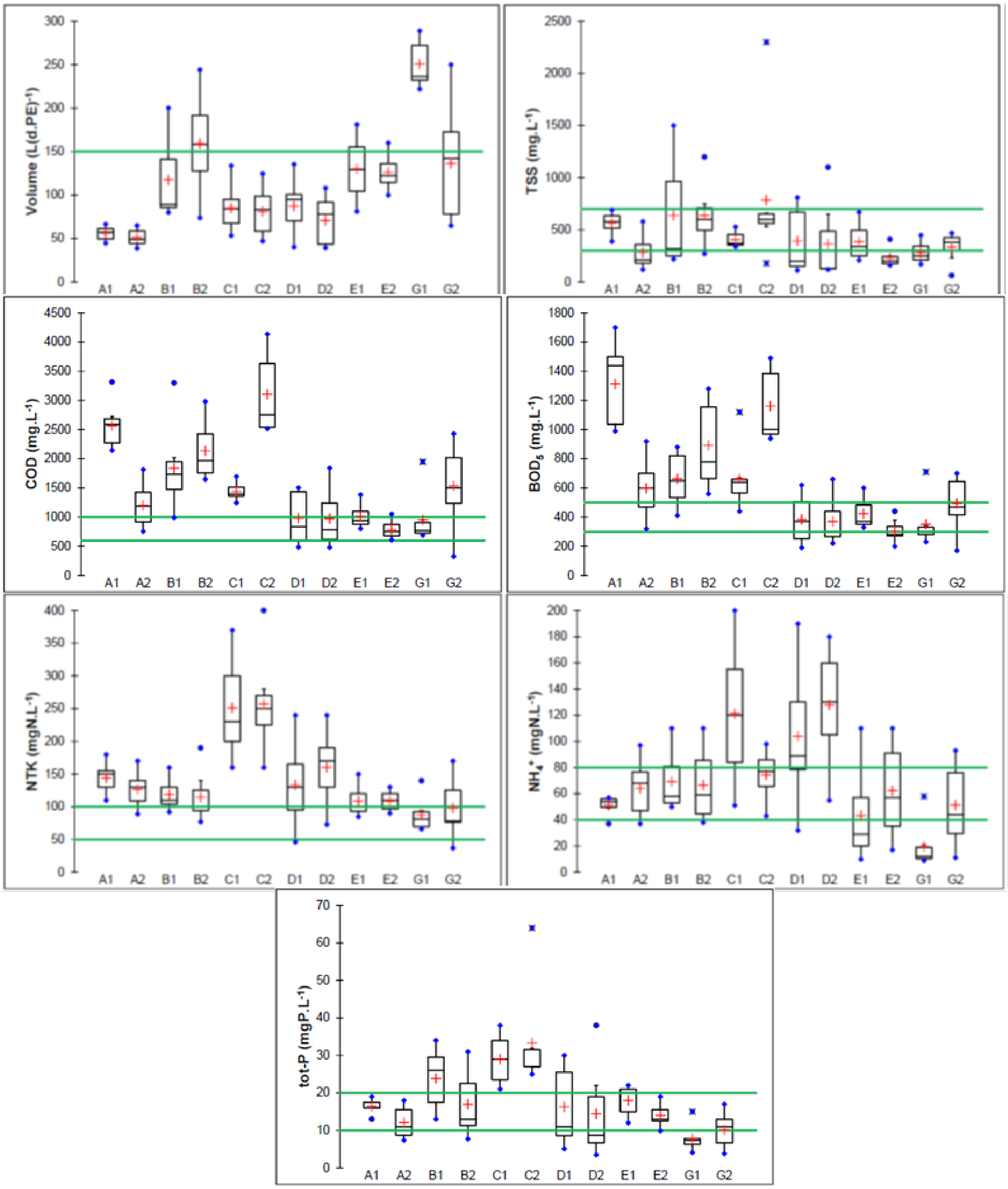

Figure 2. Volume and concentrations for the parameters studied in the raw wastewater

The volumes of wastewater to be treated per person equivalent (PE) are on average lower (approaching 90 L/d/PE) than the 150 L/d/PE generally accepted, except for site G at the 1st campaign and site B at the 2nd campaign. The results of the analyses show wide variations between different samples taken over a week at the same site. For example, for the site B, the COD content varies between 991 and 3 302 mg L^{-1} during the sample week in the first campaign. A large variability of the levels for the different parameters tested is also observed from one site to another. The average TSS values found are within the range stipulated under the 7 September 2009 decree and are consistent with a study conducted in France (Cauchi and Vignoles, 2011). However, the average COD and BOD_5 values for sites A, B and C for both campaigns are above these limit values and higher than those found in literature (Cauchi and Vignoles 2011, Devi and Dahiya 2008). Average TKN values for both campaigns are also above the limit values in an assessment protocol (AFSSET, 2009). For sites A and C, values for TSS, COD and BOD_5 between the 1^{st} and 2^{nd} campaigns are variable. The site A results show a drop in the concentrations of these three parameters which were divided by 2 between the 1^{st} and the 2^{nd} campaign. Whereas for site C there is an increase in values for these same parameters in the 2^{nd} campaign, values were doubled. For sites B, D and E, the average values for TSS, COD and BOD_5 remain about the same order of magnitude over both campaigns.

As a conclusion, individual dwelling discharges sampled before the entry into septic tank show a wide variability not only of physico-chemical quality but also of the quantity of water released. The concentrations observed show that assumed quality standards about OSSF seem inappropriate. The levels found in raw waters especially for COD, BOD_5 and NTK are often indeed higher than those indicated in the reference texts. In addition, water volumes to be treated (around 90 L/d/PE) are lower than the volume used to assess OSSF performance (around 100 L/d/PE).

Daily loads per inhabitant estimation

Daily loads per inhabitant were calculated using equation 1.

Equation 1. Daily load calulation

$$Daily\ load = \frac{[X]_i \times Volume_i}{N_i}$$

$[X]_i$ = concentration of the parameter X on the date I; $Volume_i$ = Volume measured on the date I ; N_i = number of person in the house on the date i

Cumulative frequency for COD, BOD_5 and TSS are shown in figure 2. To obtain these graphs, all the results were classified from the lower value to the higher and the frequency (number of time the value is measured) was calculated for each value. The cumulative frequency was then plotted as a function of the measured concentrations. On these graphs the full arrow corresponds to the median and the dotted arrow to the mean value of the data set. Daily loads per inhabitant was estimated from regulatory limits (AFSSET, 2009; Decree September 7, 2009) using a daily volume of 150 L per PE (45-75 g de DBO_5 ; 90-150 g de DCO ; 60-90 g de MES). Blue rectangles on graphs permit to delimit this value range.

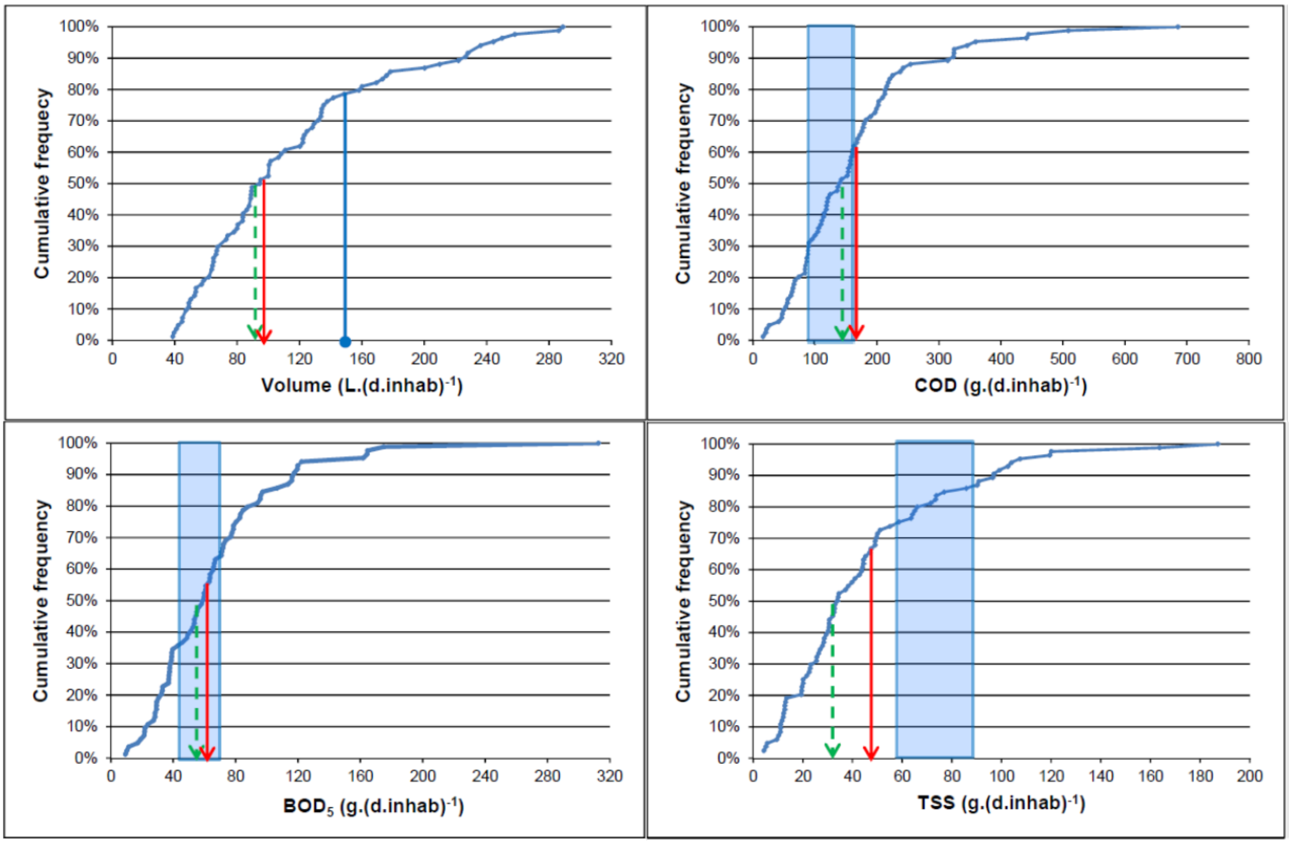

Figure 3. Daily loads per inhabitant

As already highlighted, measured volumes are 1.5 times less than reference volumes for an inhabitant per day. Concerning BOD_5 and COD , concentrations were higher than those used to test OSSF (Decree September 7, 2009). When daily loads were calculated, mean BOD5 load is equal to 67 g and COD load about 162 g what corresponds to admitted value for a PE. However, a value of 50 g of TSS per inhabitant was calculated for mean daily load. This value is lower than the admitted range for TSS of 60 g - 90 g per person equivalent. In fact, measured and reference concentrations of TSS are in the same order of magnitude but the raw waste water volume measured in the field is less important.

Downstream these raw waters, OSSF must be able to treat a “common” daily load (about 60 g BOD_5 per person) for a water discharged volume 1.5 times less than estimated volume used to design these facilities. On the field, pollutants concentrations to be treated are higher than those OSSF were tested for. These observations must be taken into account to avoid malfunctions because of more concentrated effluents.

Correlation investigations by PCA

Principal component analysis has been used to better explain this large data set.

Two campains comparison. Firstly, a PCA has been made for the 7 samples taken at the six sites, for each of the campaigns, with parameters: volume (L), COD ($mgO_2.L^{-1}$), BOD_5 ($mgO_2.L^{-1}$), TSS ($mg.L^{-1}$), NH_4^+ ($mgN.L^{-1}$), P_T ($mgP.L^{-1}$). Each data set is thus represented by 252 values. 4 shows the results obtained for the six study sites for each campaign.

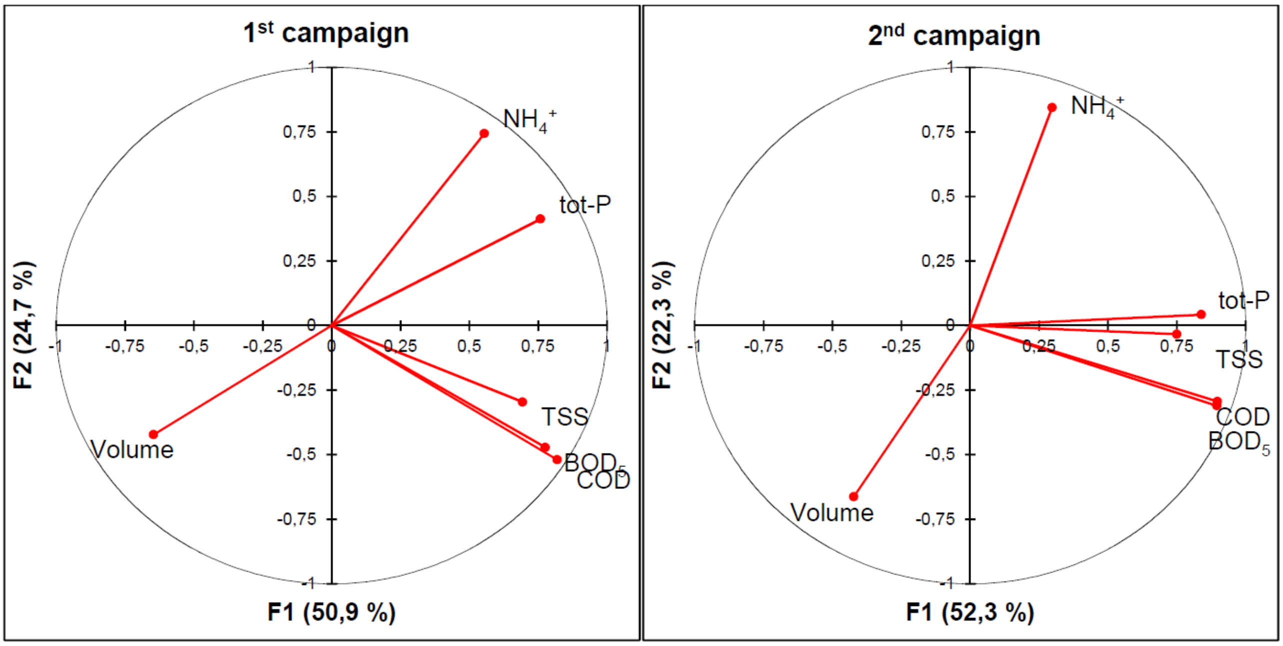

Figure 4. Representation of the variables in the plane formed by the first 2 principal. (6 sites – 2 samplings)

The first two principal components represent 75.6 % of the initial inertia for the 1st campaign and 74.6 % for the 2nd campaign. These percentages are satisfactory to describe the initial data set. This representation for each campaign differs, notably for the phosphorus. For both campaign, There are three distinct groups: (i) a group composed of ammonium ions, called the « nutrient » group, (ii) a group composed of TSS, of the COD and of the BOD_5, called the « organic » group, (iii) a group composed of the volume. The nutrient and organic groups form an angle of 90°, which indicates that they are not correlated between themselves. The volume in these two PCA is not correlated with the « organic » group. Conversely, in the PCA for the 1st campaign, the volume forms an angle of 180° with the « nutrient » group, and with the ammonium ions in the 2nd campaign PCA, showing that it is negatively correlated with this group, i.e. when the volume increases the levels of ammonium ions and total phosphorus decrease. Variable representation for each campaign differ in particular concerning total phosphorus. The phosphorus which was correlated with the ammonium ions in the 1st campaign appears in the « organic » group made up of the TSS, the COD, and the BOD_5 in the 2nd campaign. This difference in correlation shows that the occupants have changed some of their habits between the two campaigns.

Secondly, figure 5 represents for each campaign the individuals, which are the samples taken for each of the sites, in the plan formed by the F1 and F2 previously described. In order to identify each sample, the individuals are noted as follow: Aa-M, with A: letter of the site, a: number of the campaign (a=1st, b=2nd), M: day of the week (M: Monday, Tu: Tuesday, W: Wednesday, Th: Thursday, F: Friday, Sa: Saturday, Su: Sunday).

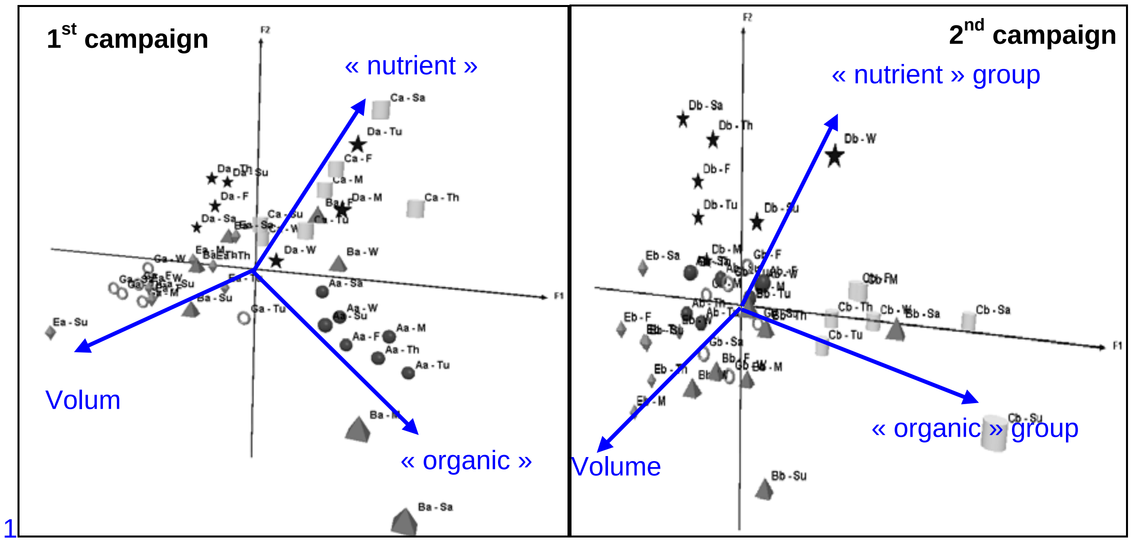

Figure 5. Representation of the individuals in the plane formed by the first 2 principal components F1-F2 of the PCA's for the 1st and 2nd campaigns on the 6 sites studied

The variability of the concentrations found between the sites on 3, is reflected in the representation of the individuals in the F1-F2 plane, for both campaigns. Three sites stand out from the others for the 1st campaign; sites A and B which are orientated towards and with the « organic » group made up of the TSS, the COD and the BOD_5, and site C which is orientated towards the « nutrient » group made up of the ammonium ions and the phosphorus. Sites E and G are orientated towards the volume axis. The site C discharge is characterized by high levels of nitrogen and phosphorus, whereas those from sites A and B have high concentrations of TSS, COD and BOD_5. For the 2nd campaign the site breakdown is not so clear-cut. Site A is no longer directed towards the « organic » group, and the plots representing it are spread around the center. The plots for site C are orientated towards the « organic » group made up of the BOD_5, the TSS, the COD and the P_T.

As a conclusion, the representation of the individuals does not allow a typical family's general behavior over a week, to be given. The orientations taken by the different individual site samples clearly show the behavioral differences regarding water use, from one family to another. The PCA has brought confirmation concerning observations made on the initial data, i.e. wide variability in concentrations between sites and within one house.

Combination of numerous campaigns. A PCA has then been made by pooling the data for both campaigns on sites A, B, C, D, E and G, and campaigns on ten other sites. The PCA has been undertaken using 185 sets of samples with the parameters: volume (L), COD ($mgO_2.L^{-1}$), BOD_5 ($mgO_2.L^{-1}$), TSS ($mg.L^{-1}$), NH_4^+ ($mgN.L^{-1}$), P_T ($mgP.L^{-1}$). This gives a dataset representing 1 110 values, and figure 6-a) shows the resulting PCA.

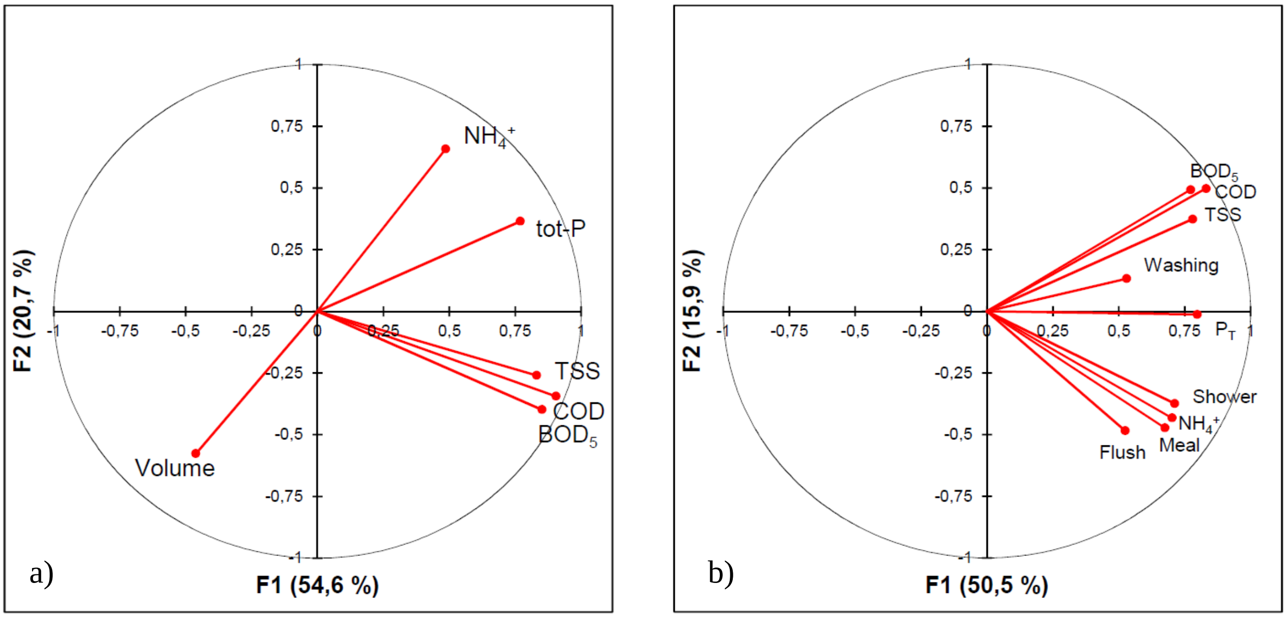

Figure 6. Representation of the variables in the plane formed by the first 2 principal components (16 sites – 185 samplings): a) Water quality parameters b) Water usage indicators added

The distribution pattern of the variables resembles that obtained with the results of the PCA from the two separate campaigns. By increasing the initial dataset, the percentage of the total inertia explained by the PCA has increased and the three types of behavior detected forward are also observed.

The more the data set increases, the more parameter correlations are significant. The explained variability of the dataset increases with the number of data considered indicating that the observed correlations are more and more reliable. To characterize raw waste water, at least one parameter from each group previously identified must be followed. Determining a parameter of each group will permit to identify other parameters tendencies in wastewater before the entry into septic tank . This monitoring would permit to identify by which type of behavior is characterized a site.

Correlation between analysed parameters and water usage. A PCA has also been made with the aim of correlating the values obtained for the different parameters, with the data collected from the inhabitant questionnaires given at the time of sampling. The two campaigns conducted on the 6 sites have been used. Following a study where the volumes collected daily and the concentrations found for the different parameters were included, the results obtained were all highly influenced by the volume parameter, notably the data from the questionnaires. Therefore it has been decided to work with the daily load (g/d) for the parameters TSS, COD, BOD_5, P_T, NH_4^+, and to this end, the obtained concentrations have been multiplied by the daily volume measured. Figure 6-b) shows the results obtained for this PCA.

The first two principal components represent 66.4 % of the initial inertia. Correlations between water quality parameters and water usage indicators can be seen. The ammoniums ions are correlated with the number of showers, meals and toilet flushes. It seems coherent as higher concentrations of NH_4^+ are expected in “black” water from toilets. The phosphorus and the number of machine washes made do not seem to be correlated with other parameters. The TSS, COD and the BOD_5 are not correlated with the data derived from the occupants’ daily behavior.

CONCLUSIONS

Two wastewater inflow sampling campaigns have been undertaken on 5 principal residences and on a second home. The results obtained for the concentrations of the different parameters analyzed, show that the characteristics of the wastewaters discharged by the occupants are very variable. There are wide variations from one habitation to another but also from one set of samples to another over one week from the same habitation. Household water use and individual's behavior are difficult to exploit with a view to obtain a wastewater characteristics 'benchmark' model. In addition, discharge volume was lower 1.5 less than reference value of 150 L admitted for an inhabitant. However organic load (COD and BOD_5) obtained for the 6 sites were of the same magnitude as reference organic loads per equivalent person. As a result effuents in the field appear more concentrated than effluents used to design and test OSSF which as to be taken into account to avoid disfunctioning of these facilities.

Principal component analysis were undertaken on the different parameters monitored and sets of samples taken. Results have shown three types of raw wastewater behavior are detected: (i) "organic" group (COD, BOD_5, TSS), (ii) "nutrient" group (NH_4^+, P_T); (iii) Volume. Different behavior were identified. Some discharges were characterized by high organic parameters concentrations and others by high concentrations in ammonium. At least one parameter of each group must be followed in order to characterize raw domestic wastewater. PCA results also highlighted that organic parameters do not seem to be correlated with studied water usage indicators. The ammonium parameter is influenced by the number of meals and showers taken and by the number of toilets flushing during the sampling period.

ACKNOWLEDGEMENT
This work was supported by Veolia Eau (France). We would like to thank Christophe Cambon, Guillaume Busset and Olivier Leon for their technical support.

REFERENCES

AFSSET, 2009. Protocole d'évaluation technique pour les installations d'assainissement non collectif dont la charge est inférieure ou égale à 20 équivalents habitants. ANSES.

Decree September 7 2009 Technical requirements for sewerage facilities receiving a load of organic pollution than or equal to 1.2 kg/day of BOD_5 (modified by decree March 7th 2012; fixant les prescriptions techniques applicables aux installations d'assainissement non collectif recevant une charge brute de pollution organique inférieure ou égale à 1,2 kg/j de DBO_5). Available on: http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20120425&numTexte=3&pageDebut=07348&pageFin=07349

Devi, R., Dahiya, R.P. 2008 COD and BOD removal from domestic wastewater generated in decentralised sectors. Bioresource Technology 99(2), 344-349.

Directive 91/271/CEE (1991) Council directive of 21 May 1991 concerning urban waste water treatment. Available on:http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1991:135:0040:0052:EN:PDF

Directive 2000/60/EC 2000 Directive of the European Parliament and of the council of 23 October 2000 establishing a framework for Community action in the field of water policy. Available on: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:327:0001:0072:EN:PDF

Lakel, A., Meyronneinc, A., Lafforgue, A., Six, A., Oustric, A., Fernandes-Allonier, A.-S., Mouline, B., Berard, C., Vignoles, C., Bonnet-Langagne, F., Wanert, F., Venel, G., Ferrere, G., Carvallo, G., Philip, H., Bonnard, J., Braconnie, J., Lesavre, J., Dautais, J.-P., Dubourg, L., Tournaire, M., Richard, P., Felon, T. and Pujol, R. 2006 Rehabilitation of on-site wastewater treatment for single houses (Réhabilitation des installations d'assainissement non collectif des maisons individuelles), cahier 5370 « e-cahier du CSTB », CSTB/ASTEE, 40 pp.