02 SIPIBEL dataset description

Jean-Luc BERTRAND-KRAJEWSKI[1*], Noémie PERNIN[1], Vivien LECOMTE[2]

1. Univ Lyon, INSA Lyon, DEEP, EA 7429, 11 rue de la physique, F-69621 Villeurbanne cedex, France.

2. Groupe de Recherche, d'Animation technique et d'Information sur l'eau (GRAIE), 66 boulevard Niels Bohr, F-69100 Villeurbanne, France.

* Corresponding author: jean-luc.bertrand-krajewski@insa-lyon.fr

Abstract

The Bellecombe pilot site – SIPIBEL – was created in 2010 to study the characterisation, treatability and impacts of hospital effluents in an urban wastewater treatment plant. This pilot site is composed of: i) the CHAL hospital, opened in February 2012, ii) the Bellecombe wastewater treatment plant, with two separate treatment lines allowing to fully separate the hospital wastewater and the urban wastewater, and iii) the Arve River as the receiving water body and a tributary of the Rhône River and the Geneva aquifer. The data base includes in total 48 439 values measured on 961 samples (raw and treated hospital and urban wastewater, activated sludge in aeration tanks, dried sludge after dewatering, river and groundwater, and a few additional campaigns in aerobic and anaerobic sewers) with 44 455 physico-chemistry values (including 15 pharmaceuticals and 14 related transformation products, biocides compounds, metals, organic micropollutants, etc.), 2 193 bioassay values (ecotoxicity), 1 682 microbiology values (including microorganisms and antibioresistance indicators) and 120 hydrobiology values. As a complement to the SIPIBEL data introduction paper (pre-print available in Zenodo), this paper describes the 10 csv files of the SIPIBEL data set, aiming to provide the user the necessary information to read and interpret the data.

Keywords

Hospital wastewater, urban wastewater, emerging contaminants, ecotoxicological risk assessment, wastewater treatment plant, sludge, receiving water body, groundwater.

1. INTRODUCTION

The full SIPIBEL dataset in Zenodo is composed of 12 files listed in Table 1. The column separator of the csv files is the semicolon ';'. In the following sections, each csv file is described.

File name	Content
01_SIPIBEL_data_paper.pdf	Presentation of SIPIBEL observatory and data set
02_SIPIBEL_dataset_description.pdf	This file
03_SIPIBEL_metadata_sampling_points.csv	List and description of sampling points
04_SIPIBEL_metadata_physico_chemistry_methods.csv	List and description of the analytical methods
05_SIPIBEL_metadata_bioassay_methods.csv	List and description of the bioassay methods
06_SIPIBEL_metadata_microbiology_methods.csv	List and description of the microbiology methods
07_SIPIBEL_metadata_hydrobiology_methods.csv	List and description of the hydrobiology methods
08_SIPIBEL_dataset_samples.csv	List and description of samples
09_SIPIBEL_dataset_physico_chemistry.csv	Full set of physico-chemistry data
10_SIPIBEL_dataset_bioassay.csv	Full set of bioassay data
11_SIPIBEL_dataset_microbiology.csv	Full set of microbiology data
12_SIPIBEL_dataset_hydrobiology.csv	Full set of hydrobiology data

Table 1 : List of the 12 files of the SIPIBEL data set available on Zenodo.

2. FILE 03_SIPIBEL_METADATA_SAMPLING_POINTS.CSV

This file contains 46 lines and 7 columns. The columns indicate successively:

1. The identification code for each sampling point, in the format SP_00xx with xx a two-digit code identifying each sampling point in both space and time (see maps and evolution of sampling points in file 01_SIPIBEL_data_paper.pdf).
2. The sampling point name, with a short description for the Bellecombe wastewater treatment plant where sampling conditions have changed over time due to modifications of the treatment lines (see file 01_SIPIBEL_data_paper.pdf).
3. The sample matrix, selected among raw wastewater, treated wastewater, surface freshwater, groundwater, combined sewer overflow, and sludge.
4. The sample origin, selected among hospital wastewater, urban wastewater, river water, and groundwater.
5. The longitude of the sampling point.
6. The latitude of the sampling point.
7. A comment if appropriate. 'NA' indicates there is no comment.

3. FILE 04_SIPIBEL_METADATA_PHYSICO_CHEMISTRY_METHODS.CSV

This file contains 760 lines and 16 columns. The columns indicate successively:

1. The identification code for each physico-chemistry analytical method, in the format MP_xxxx with xxxx a four-digit code.
2. The name of the physico-chemistry indicator.
3. The CAS (Chemical Abstract Service) number if appropriate. 'NA' indicates no CAS number is given.
4. The sample matrix, as given in file 03_SIPIBEL_metadata_sampling_points.csv (column 3).
5. The fraction to which the physico-chemistry analysis is applied, selected among 1 – Dissolved fraction, 2 - Whole water with no separation of liquid and SPM (Suspended Particulate Matter) phases, 5 – TSS (Total Suspended Solids) fraction, and 6 – Sludge.
6. The standard, guideline, or bibliographic reference for the physico-chemistry analytical method. For some cases, internal methods are used. The literature references (excluding standards and internal methods) are given below in section 12 'References'.
7. The limit of detection (LoD). 'NA' indicates no LoD is given.
8. The limit of quantification threshold LoQ 1 used to estimate low values. See detailed explanations in file 01_SIPIBEL_data_paper.pdf. 'NA' indicates no LoQ 1 values is given.
9. The limit of quantification threshold LoQ 2 used to estimate low values. See detailed explanations in file 01_SIPIBEL_data_paper.pdf. 'NA' indicates no LoQ 2 values is given.
10. The limit of quantification threshold LoQ 3 used to estimate low values. See detailed explanations in file 01_SIPIBEL_data_paper.pdf. 'NA' indicates no LoQ 3 values is given.
11. The limit of quantification threshold LoQ 4 used to estimate low values. See detailed explanations in file 01_SIPIBEL_data_paper.pdf. 'NA' indicates no LoQ 4 values is given.
12. The limit of quantification (LoQ). 'NA' indicates no LoQ is given.
13. The unit of the measured physico-chemistry indicator.
14. The uncertainty (in %) at LoQ for a 95% confidence level. 'NA' indicates no uncertainty is given.
15. The uncertainty (in %) at the measured concentration for a 95% confidence level. 'NA' indicates no uncertainty is given.
16. The type of analytical method used. 'NA' indicates no specific method is given.

4. FILE 05_SIPIBEL_METADATA_BIOASSAY_METHODS.CSV

This file contains 93 lines and 15 columns. The columns indicate successively:

1. The identification code for each bioassay method, in the format ME_0xxx with xxx a three-digit code.
2. The name of the bioassay.
3. The fraction to which the bioassay is applied, selected among 1 – Dissolved fraction and 2 - Whole water with no separation of liquid and SPM phases.
4. The standard, guideline, or bibliographic reference for the bioassay. The literature references (excluding standards and internal methods) are given below in section 12 'References'.

5. The exposure time of the bioassay. 'NA' indicates no exposure time is given.
6. The type of effect, selected among agonist, antagonist, cytotoxicity, DNA damage, growth inhibition, luminescence inhibition, mobility inhibition, mortality, mutation, and population growth inhibition.
7. The endpoint of the bioassay, selected among EC20 (20% effective concentration), EC50 (median effective concentration), growth inhibition, LD50 (lethal dose 50%), luciferase activity, micronuclei / well, mortality, number of revertant wells (/48), SOS induction factor, tail intensity, and viability.
8. The unit of the result. 'NA' indicates the absence of specific unit.
9. The test condition, selected among semi-static and static.
10. The type of specific treatment applied. 'NA' indicates no specific treatment is applied.
11. The use of freezing (Yes or No).
12. The sample preparation method, selected among settling, solid-phase extraction (SPE) using C18 columns and solid-phase extraction (SPE) using HLB columns. 'NA' indicates no specific sample preparation is applied.
13. The extraction solvent, selected among acetone, DMSO and methanol. 'NA' indicates no solvent is used.
14. The solvent used in the bioassay, selected among acetone, DMSO, and methanol. 'NA' indicates no solvent is used.
15. A comment if appropriate. 'NA' indicates there is no comment.

5. FILE 06_SIPIBEL_METADATA_MICROBIOLOGY_METHODS.CSV

This file contains 40 lines and 10 columns. The columns indicate successively:

1. The identification code for each microbiology method, in the format MM_00xx with xx a two-digit code.
2. The name of the microbiology test.
3. The fraction to which the microbiology test is applied, selected among 2 - Whole water with no separation of liquid and SPM phases, 5 – TSS fraction, and 6 – Sludge.
4. The endpoint of the test, selected among bacteria (16s rDNA), eggs, integrons all classes, integrons class 1, integrons class 2, integrons class 3, oocysts, relative abundance class 1, relative abundance class 2, relative abundance class 3, and relative abundance all classes. 'NA' indicates no specific endpoint.
5. The standard, guideline, or bibliographic reference for the microbiology test. For some cases, internal methods are used. The literature references (excluding standards and internal methods) are given below in section 12 'References'.
6. The limit of detection (LoD). 'NA' indicates no LoD is given.
7. The limit of quantification (LoQ). 'NA' indicates no LoD is given.
8. The unit of the result. 'NA' indicates the absence of specific unit.
9. The use of freezing (Yes or No).
10. The type of microbiology method used.

6. FILE 07_SIPIBEL_METADATA_HYDROBIOLOGY_METHODS.CSV

This file contains 5 lines and 4 columns. The columns indicate successively:

1. The identification code for each hydrobiology method, in the format HM_000x with x a one-digit code.
2. The name of the hydrobiology indicator. All indicators are evaluated on whole water with no separation of liquid and SPM phases.
3. The standard, guideline, or bibliographic reference for the test.
4. The type of hydrobiology method used.

7. FILE 08_SIPIBEL_DATASET_SAMPLES.CSV

This file contains 962 lines and 8 columns. The columns indicate successively:

1. The identification code for each sample, in the format Ech_xxxx with xxxx a four-digit code.
2. The identification code of the campaign to which the sample in column 1 belongs, in the format Camp_0xxx with xxx a three-digit code.

3. The corresponding sampling point given with its code SP_00xx (see file 03_SIPIBEL_metadata_ sampling_points.csv).
4. The start date and time of the campaign.
5. The duration (in h:mm) of the campaign. 'One-off' indicates instantaneous samples.
6. The water flow (in m^3/day) at the sampling point during the campaign. 'NA' indicates no water flow data is available or flow data is not relevant.
7. The water flow data provider (for information only). 'NA' indicates no provider is given or flow data is not relevant.
8. The sample quality, as explained in file 01_SIPIBEL_data_paper.pdf, which can be either 'CORRECT" if all quality tests are OK or 'UNCERTAIN' in other cases. Data with sample quality evaluated as 'INCORRECT' are not included in the dataset.

8. FILE 09_SIPIBEL_DATASET_PHYSICO_CHEMISTRY.CSV

This file contains 44 456 lines and 5 columns. The columns indicate successively:

1. The identification code of the sample Ech_xxxx (see file 08_SIPIBEL_dataset_samples.csv).
2. The identification code of the physico-chemistry method applied MP_xxxx (see file 04_SIPIBEL_metadata_ physico_chemistry_methods.csv).
3. The type of concentration value, selected among <LoD, <LoQ, and individual value (corresponding to measured values >LoQ).
4. The corrected value of the physico-chemistry indicator, which corresponds either to the raw measured value when no correction is needed or to the corrected final value when a correction is applied (e.g. to estimate values lower than LoQ – see detail in file 01_SIPIBEL_data_paper.pdf – or to account for blank samples with non-zero values).
5. The data quality, as explained in file 01_SIPIBEL_data_paper.pdf, which can be either 'CORRECT" if all data quality tests are OK or 'UNCERTAIN' in other cases. Values with data quality evaluated as 'INCORRECT' are not included in the dataset.

9. FILE 10_SIPIBEL_DATASET_BIOASSAY.CSV

This file contains 2 194 lines and 9 columns. The columns indicate successively:

1. The identification code of the sample Ech_xxxx (see file 08_SIPIBEL_dataset_samples.csv).
2. The identification code of the bioassay method applied ME_0xxx (see file 05_SIPIBEL_metadata_ bioassay_methods.csv).
3. The percentage of the sample submitted to the bioassay. 'NA' indicates cases where no percentage is applied.
4. The value indication (to be used jointly with the value given in column 5): '<' in column 4 indicates that the test value is lower than the number given in column 5. For example, '<' in column 4 and '15' in column 5 indicate that the test value is '<15'. Similarly, '>' in column 4 indicates that the test value is greater than the number given in column 5, and '=' in column 4 indicates that the test value is equal to the number given in column 5. 'NA' indicates that this value indication is not applicable.
5. The numerical value of the result the bioassay. In addition, '<LoD' indicates values lower than the bioassay LoD, 'NS' indicates a not-significant value (it can be interpreted as <LoQ).
6. The lower limit of the confidence interval of the result, if applicable. 'nm' indicates the lower limit is not given.
7. The upper limit of the confidence interval of the result, if applicable. 'nm' indicates the upper limit is not given.
8. The effect value reflects the intensity of the effect demonstrated by bioassay test on a given sample. This value is used in SIPIBEL to perform cross-statistics with physico-chemistry and microbiology data. A zero value indicates no effect. The effect value is determined as follows:
    *   For bioassays with *Daphnia magna*, *Brachionus calyciflorus* and *Pseudokirchneriella subcapitata*, for cytotoxicity assays and for genotoxicity assays (Comets assay, SOS Chromotest and Ames Test): the effect value corresponds to toxic units (e.g.: $UT_{50} = 100 / EC_{50}$). Special case: if the value (column 5) is "NS" (not significant) then the effect value is zero.

- For bioassays with *Heteroxypris incongruens* and for estrogenic and thyroid disrupting effect assays: the effect value corresponds to the value (column 5). Special case: if the value is <LoD then the effect value is zero.

9. The data quality, as explained in file 01_SIPIBEL_data_paper.pdf, which can be either 'CORRECT" if all data quality tests are OK or 'UNCERTAIN' in other cases. Values with data quality evaluated as 'INCORRECT' are not included in the dataset.

10. FILE 11_SIPIBEL_DATASET_MICROBIOLOGY.CSV

This file contains 1 680 lines and 7 columns. The columns indicate successively:

1. The identification code of the sample Ech_xxxx (see file 08_SIPIBEL_dataset_samples.csv).
2. The identification code of the microbiology method applied MM_00xx (see file 06_SIPIBEL_metadata_microbiology_methods.csv).
3. The numerical value of the result of the microbiology test. In addition, '<LoD' and '<LoQ' indicate values lower than the LoD and the LoQ, respectively.
4. The standard deviation, if applicable. '<LoD' and '<LoQ' indicate values lower than the LoD and the LoQ, respectively. 'NM' indicates the standard deviation is not given.
5. The final corrected value of the microbiology test, which corresponds either to the raw measured value when no correction is needed or to the corrected final value when a correction is applied.
6. The corrected standard deviation, which corresponds either to the raw standard deviation when no correction is needed or to the corrected value when a correction is applied.
7. The data quality, as explained in file 01_SIPIBEL_data_paper.pdf, which can be either 'CORRECT" if all data quality tests are OK or 'UNCERTAIN' in other cases. Values with data quality evaluated as 'INCORRECT' are not included in the dataset.

11. FILE 12_SIPIBEL_DATASET_HYDROBIOLOGY.CSV

This file contains 113 lines and 5 columns. The columns indicate successively:

1. The identification code of the sample Ech_xxxx (see file 08_SIPIBEL_dataset_samples.csv).
2. The identification code of the hydrobiology method applied HM_000x (see file 07_SIPIBEL_metadata_hydrobiology_methods.csv).
3. The interpretation of the result of the hydrobiology test, selected among very good, good, intermediate, and poor.
4. The value of the hydrobiology test.
5. The data quality, as explained in file 01_SIPIBEL_data_paper.pdf, which can be either 'CORRECT" if all data quality tests are OK or 'UNCERTAIN' in other cases. Values with data quality evaluated as 'INCORRECT' are not included in the dataset.

12. REFERENCES

Bergé A., Giroud B., Wiest L., Domenjoud B., Gonzalez-Ospina A. & Vulliet E. (2016). Development of a multiple-class analytical method based on the use of synthetic matrices for the simultaneous determination of commonly used commercial surfactants in wastewater by liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1450, 64-75. https://doi.org/10.1016/j.chroma.2016.04.078.

Jugan M.L., Lévy-Bimbot M., Pomérance M., Tamisier-Karolak S., Blondeau J.P. & Lévi Y. (2007). A new bioluminescent cellular assay to measure the transcriptional effects of chemicals that modulate the alpha-1 thyroid hormone receptor. *Toxicology in Vitro*, 21(6), 1197-1205. https://doi.org/10.1016/j.tiv.2007.03.020.

Jugan M.L., Oziol L., Bimbot M., Huteau V., Tamisier-Karolak S., Blondeau J.P. & Lévi Y. (2009). In vitro assessment of thyroid and estrogenic endocrine disruptors in wastewater treatment plants, rivers and drinking water supplies in the greater Paris area (France). *Sci. Total Environ.*, 407(11), 3579-3587. https://doi.org/10.1016/j.scitotenv.2009.01.027.

Kienzler A., Tronchère X., Devaux A. & Bony S. (2012). Assessment of RTG-W1, RTL-W1, and PLHC-1 fish cell lines for genotoxicity testing of environmental pollutants by means of a Fpg-modified comet assay. *Toxicology in Vitro*, 26(3), 500-510. https://doi.org/10.1016/j.tiv.2012.01.001.

Peysson W. & Vulliet E. (2013). Determination of 136 pharmaceuticals and hormones in sewage sludge using quick, easy, cheap, effective, rugged and safe extraction followed by analysis with liquid chromatography–time-of-flight-mass spectrometry. *Journal of Chromatography A*, 1290, 46–61. http://dx.doi.org/10.1016/j.chroma.2013.03.057.

Piegorsch W.W., Simmons S.J., Margolin B.H., Zeiger E., Gidrol X.M. & Gee P. (2000). Statistical modeling and analyses of a base-specific *Salmonella* mutagenicity assay. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*, 467(1), 11-19. https://doi.org/10.1016/S1383-5718(00)00019-X.

Stalder T. (2012). *Implication des effluents d'activités hospitalières et de la filière carnée sur la dissémination de l'antibiorésistance : dynamique des intégrons de l'émission au rejet [Involvement of the hospital effluents and the meat industry on the spread of antibiotic resistance: dynamics of the integrons from the source to the release]*. PhD thesis, Université de Limoges, Limoges, France, nr 74-2012, 194 p. (in French). Available at http://aurore.unilim.fr/ori-oai-search/advanced-search.html?submenuKey=advanced&userChoices%5Bsimple_all%5D.simpleValueRequestType=default&userChoices%5Bsimple_all%5D.simpleValue=stalder&search=true&menuKey=all (accessed 07 July 2021).

White P.A., Rasmussen J.B. & Blaise C. (1996). A semi-automated, microplate version of the SOS Chromotest for the analysis of complex environmental extracts. *Mutation Research - Environmental Mutagenesis and Related Subjects*, 360(1), 51-74. https://doi.org/10.1016/s0165-1161(96)90237-x.

Wiest L. & Vulliet E. (2017). *Développement des méthodes d'analyse de métabolites et produits de dégradation du diclofénac et du sulfamethoxazole [Development of analytical methods for metabolites and degradation products of diclofenac and sulfamethoxazole]*. Villeurbanne (France): GRAIE, Rilact Task 1.1 Deliverable 2, October 2017, 15 p. Available at http://www.graie.org/Sipibel/docs/docs-rilact/RILACT-T1.1-L2-ISA-Metabolites.pdf (accessed 07 July 2021).

Wiest L., Baudot R. & Vulliet E. (2016). Impact des rejets de médicaments sur le milieu récepteur : résultats des projets Sipibel et Irmise [Impact of pharmaceuticals discharges on the receiving environment: Sipibel and Irmise projects results]. *TSM*, 6, 12-21. https://doi.org/10.1051/tsm/201606012 (in French).