Assessing Infiltration and Exfiltration on the Performance of Urban Sewer Systems

Contract number : EVK1-CT-2000-00072 APUSS
Project homepage : http://www.insa-lyon.fr/Laboratoires/URGC-HU/apuss

DELIVERABLE 1.3

Standard Operation Procedure (SOP):

Quantification of Infiltration by the Analysis of Pollutant Time Series as an Intrinsic Tracer

O. Kracht
EAWAG

March 2005

Czech Technical University in Prague
Faculty of Civil Engineering

Swiss Federal Institute
for Environmental Science
and Technology

Standard Operation Procedure (SOP):

Quantification of Infiltration by the Analysis of Pollutant Time Series as an Intrinsic Tracer

Number: -

Version: 2

No. pages: 19

Author: *Oliver Kracht*
oliver.kracht@eawag.ch
EAWAG
Environmental Engineering
Ueberlandstrasse 133
CH-8600 Duebendorf
Switzerland

Date: **28 March 2005**

Changed at: **28 March 2005**

Valid from: **28 March 2005**

Table of Content

1 General Principles

1.1 Introduction

The following description gives a compendium for the quantification of infiltration into sewers by the analysis of pollutant time series as an intrinsic tracer (**Pollutant Time Series Method**). Detailed background information on the underlying theory of the method and the required boundary conditions are described in a scientific paper text that was presented at the 4th International Conference on Sewer Processes and Networks (Kracht and Gujer, 2004). This text is attached as a separate file.

The method is based on a combined analysis of measured time series of pollutant concentrations and discharged wastewater. It is suited to quantify the infiltration into a sewer system on catchment or subcatchment scale, where a continuous discharge of wastewater can be assured. A minimum amount of wastewater flow is required for the disturbance free operation of the measuring devices, which may be critical during minimum night flow. Furthermore, predominant types of industrial effluents should be excluded, as this may hinder a regular data analysis.

1.2 Content of the example packages

The example packages are meant for an exemplification of the described data analysis. They consist of the following files:

1) AQUASIM example package:

- Sewer_Infiltration.aqu

(Sewer_Infiltration.aqu is an AQUASIM system definition file. The example data sets are contained in the file.)

2) R-Script example package:

- APUSS_ChemHydSep.r
- APUSS_ChemHydSep_biblio.r
- COD_RL.txt
- Q_RL.txt

(APUSS_ChemHydSep.r and APUSS_ChemHydSep_biblio.r are R-Script files. The file name will be extended by a suffix indicating the version number. Only "ChemHydSepMCS.R" is foreseen to be edited by the user, as it contains the model input definitions. COD_RL.txt and Q_RL.txt are example ASCI-text files, containing a COD and discharge time series respectively.)

1.3 Terminology

The amount of discharged **wastewater** in a sewer generally shows a characteristic diurnal behaviour. This **hydrograph** is composed of a variable volume of real **foul sewage** and a certain quantity of parasitic **infiltration**.

$$Q_{\text{Wastewater}} = Q_{\text{Foul Sewage}} + Q_{\text{Infiltration}}$$ **Equation 1**

Infiltration is groundwater or other types of **extraneous water** that enters the sewer system through defective pipes (cracks and fissures), pipe joints, couplings, manholes and house connections. In this text we do not distinguish this type of "undeliberate" infiltration from extraneous water stemming from creeks and drainages, which were intentionally connected to the sewer system.

An often used expression is the amount of infiltration as a fraction of the total discharge of wastewater in the sewer (**infiltration ratio**):

$$X_{\text{Infiltration}} = \frac{Q_{\text{Infiltration}}}{Q_{\text{Wastewater}}}$$ **Equation 2**

1.4 Principles of the pollutant time series method

The fraction of infiltrating water is determined from a combined analysis of measured time series of pollutant concentrations and discharged wastewater. The data analysis uses a mixing model describing the concentration of pollutants (C) in the wastewater in dependency of the quantity of wastewater flow (Q) and time (t) (equation 3). The employed parameter set contains variables to consider time dependencies of the infiltration rate as well as temporal fluctuations of the pollutant concentration in the foul sewage (equations 4 and 5):

$$C_{Wastewater,model} = \frac{(Q_{Wastewater} - Q_{Infiltration}) \cdot C_{Foul-Sewage}}{Q_{Wastewater}} \quad \text{(eq. 3)}$$

with: $C_{Foul-Sewage} = f(t, Q_{Foul-Sewage})$ (eq. 4) and: $Q_{Infiltration} = Q_{Baseflow} + Q_{Interflow}(t)$ (eq. 5)

The parameters defining $Q_{Infiltration}$ are subsequently estimated by fitting a modelled time series of pollutant concentrations to the measured data.

1.5 Site description: example data set "Rümlang (CH)"

The R-Scripts are distributed with an example data set that has been derived from a measurement campaign conducted in the village of Rümlang in the fall of 2003. Rümlang is a commune of about 5'400 inhabitants, located to the north-eastern boarder of the agglomeration of Zurich. The total length of its sewer system amounts to 23.1 km. Rümlang has a mixed infrastructure with no predominant type of industry. COD_RL.txt and Q_RL.txt are example ASCI-text files, containing a COD and discharge time series respectively. The in-line measurements were conducted in a trunk sewer that connects the village to the regional treatment plant. Users may exchange the example files with their own data.

2 Methodological Description

The underlying theory of the pollutant time series method is described in detail in the scientific paper text "Quantification of infiltration into sewers based on time series of pollutant loads" (Kracht and Gujer, 2004), which is attached as a separate file. **This text is regarded to be a principal part of this SOP.**

3 Use of the AQUASIM model file "Sewer_Infiltration.aqu"

This chapter describes the use of the AQUASIM model file "Sewer_Infiltration.aqu". It explains the concept of the script and gives the user the necessary information to run it on his own set of data. For general information on the use of AQUASIM please refer to its manual (Reichert, 1998).

The script is filled with an example data set, that is meant to be overwritten by the users own data.

3.1 Model description

In the following the single elements of the model are described first. Afterwards a short description of the principle procedure for conducting a parameter estimation for the quantification of infiltration is given.

3.1.1 Concentration in the infiltrating water

C_Inf is a "Constant Variable".

The concentration in the infiltrating water is assumed to be **constant** over time. In the case of COD (chemical oxygen demand) the assumption C_Inf = 0 is expected to be a good approximation.

3.1.2 Concentration in the foul sewage

C_FS is a "Formula Variable" that is defined by multiple "Constant Variables":

1) In the most basic model case the concentration in the foul sewage is assumed to be **constant** over time. In this case it is simply:

C_FS = C_FS_a

possible extensions are:

2) The Concentration in the foul sewage is **depending on the quantity of the foul sewage discharge:**

C_FS = C_FS_a + C_FS_b * Q_FS + C_FS_c * Q_FS^2

It is recommended to start with the basic case C_FS = C_FS_a. For this C_FS_b and C_FS_c are simply assigned the value 0 and are not included in the parameter estimation. If required both parameters can then successively be added to the model. Pay attention to the identifiability of these parameters.

3) The Concentration in the foul sewage is **depending on the time of day:**

C_FS = C_FS_amp * sin(C_FS_freq * (t-C_FS_phase) * 2 * pi)

It is recommended to start with a value C_FS_amp = 0 and at first not to incorporate C_FS_amp to the parameter estimation (this is the basic case with a time invariant C_FS). If required this part can then later be added to the model as follows: C_FS_freq is recommended to be assigned the value 1 (= "one time every day") and not being included in the parameter estimation. Appropriate start values must be assigned to C_FS_amp and C_FS_phase, which are both activated for the parameter estimation. Pay attention to the identifiability of these parameters.

(Attention: If C_FS_amp is activated for parameter estimation, but it is still being assigned a value of 0, the algorithm fails. Set C_FS_amp to an appropriate initial value > 0 !)

The complete definition for the concentration in the foul sewage is finally:

C_FS = C_FS_a + C_FS_b * Q_FS + C_FS_c * Q_FS^2 + C_FS_amp * sin(C_FS_freq * (t-C_FS_phase) * 2 * pi)

To assure comparability between the different APUSS data sets, I recommend assigning the following limits for the parameter estimation:

C_FS_phase: min = 0 and max = 1
C_FS_amp: min = 0 (max is not relevant)

3.1.3 Time spans that are considered for the parameter estimation (fit_times)

fit_times is a "Real List Variable (t)". It defines which spans of the measured time series C_WW_measured und Q_WW_measured are used for the parameter estimation:

fit_times = 1: this time span is considered for the parameter estimation.
fit_times = 0: this time span is excluded from the parameter estimation.

Example: In the example „Sewer_Infiltration.aqu" the time span from day 18,60 to day 18,85 is excluded from the parameter fit for the reason of a short breakdown of the flow measurement unit. The data after day 26,7 are also excluded because of the initiation of a heavy rain event.

3.1.4 Modelled concentration in the wastewater

C_WW is a "Formula Variable". The modelled concentration in the wastewater is derived from a mixing of foul sewage and infiltration:

C_WW = (C_FS * Q_FS + C_Inf * Q_Inf) / Q_WW

C_WW is calculated for the whole modelled time span and can therefore be used for graphical representations.

In contrast, the „formula variable" **C_WW_fit** is the modelled wastewater concentration that is used for the parameter fit. For fit_times = 1 it is automatically set to C_WW_fit = C_WW. In case of fit_times = 0 it is C_WW_fit = 0, therefore these periods have no influence on the parameter estimation.

(It is also possible to use C_WW_fit instead of C_WW for the graphical representations. It is then more clearly visualized which spans of time were used for the parameter estimation.)

3.1.5 Measured concentration in the wastewater

C_WW_measured is a "Real List Variable (t)".

New data are imported by the use of the "Read" function (use tab or comma separated text files). The time argument must be a number that is strictly monotonic increasing from row to row. Unfortunately it is not possible to read in „real" time and date formats. We preferred the definition

of a time step of one to represent one day. Pay attention to give an appropriate number of digits to not coarsen the temporal resolution.

Remark: C_WW_measured_out is a "Formula Variable". It is only used for the data export. This showed to be helpful for further data processing in spread sheet programs, as it provides a time series with equal time steps (which might not be the case with your input data). It is simply: C_WW_measured_out = C_WW_measured.

3.1.6 Measured wastewater discharge (quantity)

Q_WW_measured is a "Real List Variable (t)".

Calculated from this, the "Formula Variable" Q_WW is the corrected measured wastewater discharge:

Q_WW = Q_WW_measured * Q_WW_systerr_b + Q_WW_systerr_a

The two auxiliary variables for the representation of possible systematic measurement are described at point "Auxiliary variables 2".

Remarks:

1) It can be the case that several alternative time series Q_WW_measured_1, Q_WW_measured_2, Q_WW_measured_3 etc. are available (i.e. from alternative measuring principles). In this case it is easy to set "Q_WW = Q_WW_measured_1" or "Q_WW = Q_WW_measured_2" or "Q_WW = Q_WW_measured_3" etc. This definition of Q_WW allows narrowing down the amount of necessary changes to one single entry. Thus it can be avoided to edit every single formula, where Q_WW occurs.

2) Q_WW is helpful for further data processing in spread sheet programs instead of using Q_WW_measured directly (in analogy to C_WW_measured_out).

3.1.7 Modelled extraneous discharge (quantity of infiltration)

Q_Inf is a "Formula Variable" that is defined by multiple "Constant Variables":

1) In the most basic model case the quantity of infiltration is assumed to be constant over time (**baseflow**):

Q_Inf = Q_baseflow

possible extensions are:

2) We additionally introduce a virtual linear reservoir that additionally causes an exponential receding discharge component (**interflow**) after rain events:

Q_Inf = Q_baseflow + Q_interflow

with: Q_interflow = Q_0_interflow * exp (-k_interflow * (t-t_0_interflow))

k_interflow: recession constant that defines the shape of the receding interflow hydrograph. Note that unit of k_interflow is defined as [1/days].

Q_0_interflow: Discharge from the virtual interflow reservoir at the point in time t_0_interflow.

To open the possibility to account for the influence of multiple rain events within the time of investigation, a set of multiple Q_interflow_i is available (Q_interflow_1, Q_interflow_2 and Q_interflow_3).

It is simply: $Q\_interflow = \sum Q\_interflow\_i$

Usually t_0_interflow_i is manually set to an arbitrary point of time in between the start and the end of the corresponding rain event (The end of a rain event would mean in this case the point in time when all inflow from direct surface runoff into the sewer has ceased). Note the fact that t_0_interflow_i is generally not activated in the parameter estimation, as it can not be estimated independently from Q_0_interflow and t_0_interflow. The estimated Q_0_interflow and t_0_interflow are as the case may be purely virtual values, not necessarily occurring in the real time series. However, the only important demand on these parameters is to adequately describe the discharging behaviour of the interflow reservoir during the investigated span of time.

It is recommended to initially set all Q_0_interflow_i to zero and therewith not include an interflow component in the model. If required this part can successively be added later. Pay attention to the identifiability of these parameters.

3.1.8 Modelled wastewater discharge (optional use)

Q_WW_modelled is a “Formula Variable” that is defined as:

Q_WW_modelled = ((Q_Inf+Q_FS)-Q_WW_systerr_a)/Q_WW_systerr_b

This variable is an additional option that allows for setting a second fit target

Q_WW_modelled =!= Q_WW_measured

as it is defined in the parameterfit type “extended”. This variable is **not required** for the parameter estimation with the by default parameterfit type “basic”.

3.1.9 Auxiliary variables 1

t is a “Program Variable” that refers to the programs internal time argument “Time”. It is required to make the time argument available for the definitions of the “Formula Variables”.

Residulas_C_WW is a “Formula Variable” that calculates the difference between C_WW and C_WW_measured. A visualization of these residuals can give a first control for the adequacy of the model structure. Residulas_C_WW can be exported for further statistical analyses and tests.

3.1.10 Auxiliary variables 2

To offer the possibility for a basic evaluation of the influence of **systematic measurement errors** on the estimated parameter values, four auxiliary variables are introduced. These variables can be set manually in example to the maximum assumption for a systematic measurement error and therewith allow to investigate the maximum influence on the estimated parameters values that can be expected.

Q_WW_systerr_a: offset error in the wastewater flow measurements. This variable is assigned the value 0, if no offset error is assumed.
Q_WW_systerr_b: constant relative error in the wastewater flow measurements. This variable is assigned the value 1, if no relative error is assumed.

(These two variables are introduced to the model by defining the variable Q_WW to Q_WW = Q_WW_measured * Q_WW_systerr_b + Q_WW_systerr_a.)

C_WW_systerr_a: offset error in the wastewater concentration measurements. This variable is assigned the value 0, if no offset error is assumed.
C_WW_systerr_b: constant relative error in the wastewater concentration measurements. This variable is assigned the value 1, if no relative error is assumed.

(These two variables are introduced to the model by redefining the variable C_WW to C_WW = (((C_FS * Q_FS+C_Inf * Q_Inf) / Q_WW) - C_WW_systerr_a) / C_WW_systerr_b.)

3.2 Principle procedure for conducting a parameter estimation

1) Read in the data for Q_WW_measured and C_WW_measured.

2) Start the parameterfit type “basic” with a simple model first: In example with only C_FS_a and Q_baseflow beeing active for the parameter estimation. (Pay attention to have correctly filled the entry in the “Initial Time”-field that is accessible by the “Edit Calculation for Parameter Estimation” option.)

   -> start the parameter estimation routine

   Attention, possible pitfall: The other relevant parameters (C_FS_b, C_FS_c, C_FS_amp and Q_interflow_i) must be set to the value 0. Otherwise they are active in the model, even if they are not active for the parameter estimation.

3) Set the appropriate “Output Steps” and “Initial Time” in the “Edit Calculation Definition” of the simulation definition “calc 1”. “Initialize” the simulation”, then start “Simulation”.

4) Evaluate the results and the quality of the parameter fit critically:
   a) Use the build in graphic representations.
   b) Have a look at the text file “xxy.fit” that is automatically produced by AQUASIM. Besides the estimated values of the parameters also estimated standard errors and a correlation matrix of the parameters are given, which should be your decisive factors to judge the identifiability of the parameters.
   Remark: Standard errors and correlation matrix are only calculated when using the “secant” algorithm. In case of convergence problems, it can be helpful to perform a preliminary parameter estimation with the “simplex” algorithm first and then to redo the parameter estimation again with the “secant” algorithm. If the “secant” algorithm can not calculate standard errors, this indicates a bad identifiabilty of the parameter set.

5) You can now continue to refine the model, by successively activating more of the offered parameters in the parameter estimation. Control the model refinements by returning to point 4).
   Remark: The use of a large quantity of parameters expands the flexibility of the model. This improves the parameter fit and reduces the sum of squared residuals. However, as a matter of fact the identifiabilty of the individual parameters will be impaired and standard errors

increase. Therefore it is advisable to not exaggerate in building too complex models on a rather limited set of measured data. The extend of the identifiable parameter set will also depend on the quality of the measured input data.

6) Finally judge your results critically (point 4) and 5)). Time series data can now be exported to the spread sheet compatible text file "xxy.lis" by the use of the "Output_data" graphics. Estimated model parameters and statistical information are finally stored to the "xxy.fit" text file.

3.3 Implemented graphics for visualization of the results

Some standard graphical representations are implemented in the script and can be accessed by the "View Results" option:

The chart "Input_data" displays the raw input data series C_WW_measured and Q_WW_measured. As these series are not altered by interpolation or smoothing they are useful for a fast overview.

"Fit_times" informs about the periods that have chosen to be considered for the parameter estimation.

"Concentrations" displays the measured and modelled concentrations C_WW_measured, C_WW_fit, C_FS and C_Inf.

"Discharge" displays the measured and modelled discharges Q_WW, Q_Inf, Q_baseflow and Q_FS.

"Residuals" visualises the differences between C_WW and C_WW_measured.

"Output_data" is not a real graphical representation, but rather intended for the data export by the "list to file" functionality. It is used to export the relevant time series data to a spread sheet compatible text file "xxy.lis". "Output_data" contains time series of the variables Q_WW, Q_baseflow, Q_interflow, C_WW_measured_out, C_WW and C_FS. However, remind that all other graphics have the possibility to be directly exported to a text file in the same way.

3.4 Remarks on the uncertainties of the estimated parameter values

The use of a frequentistic parameter estimation and error approximation has certain implications on the interpretation of the results that should shortly be summarized here:

The model parameters p_i that are represented by the means of constant variables are estimated by minimizing the sum of the squares of the weighted deviations between measurements and calculated model results.

$$\chi^2(p) = \sum_{k=1}^{n} \left(\frac{y_{k,measured} - y_k(p)}{\sigma_{y_{k,measured}}} \right)^2$$

Where $y_{k,measured}$ is the measured value at the point in time k of the time series and $y_k(p)$ is the corresponding model result. n is the total number of data points. The standard deviations $\sigma_{yk,measured}$ are used as the weighing factors.

The covariance matrix Cov(p) of the parameter estimates is derived by the use of a linear approximation of the corresponding parameter estimation functions. The algorithm that is employed calculates these derivatives using the finite difference approximation.

$$\frac{\partial p_i}{\partial y_k} \approx \frac{p_i\left(y_k + \Delta y_k\right) - y_k}{\Delta y_k}$$

Where Δy_k is chosen to be 1 % of the standard error $se(y_k)$.

The approximated standard errors $se(p_j)$ of the estimated parameter values p_i are then derived from the diagonal elements of the covariance matrix Cov(p) by

$$se\left(p_i\right) = \sqrt{Cov\left(p\right)_{i,i}}$$

In consequence two aspects must be pointed out:

1) This approximation of standard errors for the estimated parameter values (only) takes into account the **random errors** (statistical scattering) that are assigned to the measured data. It must be underlined that the influence of **systematic measurement errors** on the estimated parameter values is not automatically calculated by the script. Systematic errors are not included in the standard errors of the parameter estimates that are given in the "xxy.fit" file! It is foreseen to supply a Mont Carlo facility (work of the next weeks) to allow for an automated calculation of the "total accuracy of measurement".

2) Depending on your model definitions (this means the parameters = constant variable you have chosen to be active in the model) the model results (C_WW) might show a distinct non linear dependency on the parameter values. The employed linear approximation therefore limits the validity of the error estimation to a relatively small surrounding around the estimated parameter values within the parameter space. This problem will also be overcome by the foreseen Mont Carlo facility.

The relevance of the uncertainty (of the estimated parameter values) that is stemming from possible systematic errors which are embedded in the measurements of C_WW and / or Q_WW will also depend on the intended use of the examination findings.
As an example the over- or underestimation of Q_WW will of course lead to an over- or underestimation of Q_Infiltration. However, if the demanded result of the examination is not the absolute Q_Infiltration, but rather the fraction of infiltration in relation to the totally discharged wastewater volume (X_Infiltration), the larger part of the error will in turn be crossed out. Nevertheless a certain part of error contribution will remain, due to the nonlinear behaviour of the model and its parameter estimation functions.

3.5 Practical hints for working with AQUASIM

- It is important to check if AQUASIM has really adjusted the parameters to the optimum. Always restart the parameter fit to assure that the sum of squared residuals (Chi2) does not reduce further. If required repeat this procedure, until the sum of squared residuals does not change anymore. Similarly it is always advantageous to restart the parameter estimation with different initial conditions to control and confirm the first estimate.

- In case of a repeated data export (after a new simulation …), AQUASIM will not overwrite the text file "xxy.lis": The new data will rather be annexed to the old file. To avoid this, you should give a new name to the new file or delete the old file before doing a new data export. However, unlike the "xxy.lis" file, the "xxy.fit" file is overwritten whenever a new parameter estimation is performed.

- After having effectually finished a parameter fit, AQUASIM does not automatically perform a new simulation with the most recent parameter set. Therefore it is necessary to initialize and start the simulation again manually, to provide the actual data for the graphical representations and the data export.

4 Use of the R-Script file "APUSS_ChemHydSep.r"

This chapter describes the principal use of the R-Script files "APUSS_ChemHydSep.r" and "APUSS_ChemHydSep_biblio.r".

4.1 General Remarks

The algorithms for data analysis have been programmed in the R language (Ihaka et al, 1996) and packed in the libraries APUSS_ChemHydSep.r (front end for user defined entries) and APUSS_ChemHydSep_biblio.r (general library of underlying functions). All libraries and code examples are available for public use. As we see our role in the provision of a thorough functionality instead of a user-friendly software design, this implementation of the code relies completely on the user interface of R. As we provided no GUI (Graphical User Interface), we recommend the use of convenient editors (e.g. WinEdt or SciViews-R).

4.2 Working with R

It is not necessary to have an understanding of R for the execution of the data analysis script. However, some basic knowledge on R's data structure and file handling is helpful for the data analysis.

The binary distribution of R comes with a documentation that is stored in the \doc folder. More useful documentation for a beginner can be found on www.r-project.org in the Documentation\Contributed section. Help on specific problems can be sought at the R-newsgroup (see www.r-project.org). In general, every S-Plus documentation is also valid for R.

4.3 Principal use of the R-Script files

A data analysis with the supplied R-Script files will generally consist of the following steps:

1) Copy the four files ChemHydSepMCS.R, ChemHydSepMCS_biblio.R, COD_RL.txt and Q_RL.txt to a folder on your hard disk (the "working directory").

   Users may exchange the example data files COD_RL.txt and Q_RL.txt by files containing their own data. Remark: both files have to be recorded with the same time steps (same time data in the first column).

2) Open the file ChemHydSepMCS.R with a text editor (e.g. WinEdt or SciViews-R).

   You will now need to modify the User Input 1 to 21, according to your data set and requirements. All user inputs are explained in the file itself.

   For a first trial, it is recommend to not change these entries and perform a parameter estimation with the provided example data set. However, you must at least modify "User Input 1": the working directory.

3) Start the parameter estimation by copying the content of ChemHydSepMCS.R to the R console.

4.4 FAQ's on the R-script file ChemHydSepMCS.R:

This chapter aims to summarize answers to some of the most frequently asked questions (FAQ's) that were asked by users during the test runs of the scripts within the APUSS project.

4.4.1 Output of statistical parameters

Statistical information about the estimated total amount of infiltration during the considered span of time is obtained by pasting the lines:

```
# Output of statistical parameters

summary(Total_inf_MCS)                    #Total infiltration for the considered span of time
summary(X_Total_inf_MCS)                  #Infiltration ratio for the considered span of time

quantile(Total_inf_MCS,probs = c(0.025,0.975))      #95 % confidence interval total infiltration
quantile(X_Total_inf_MCS,probs = c(0.025,0.975))    #95 % confidence interval infiltration ratio
```

Remark: The automatically generated graphs somehow seem to overwrite the statistical results: If this happens, simply paste again these lines.

4.4.2 Generating an output file

A result file ("results.txt) for external postprocessing of the curves describing the hydrograph separation can be produced with the following code:

```
Results <- cbind(t,t(Q_baseflow),t(Q_infiltration),Q_ww_measured)
write.table(Results, file = "results.txt")
```

Remark: The part that produces a result file

```
#Results <- cbind(t,t(Q_baseflow),t(Q_infiltration),Q_ww_measured)
#write.table(Results, file = "results.txt")
```

is for the moment commented out. That means: the character # causes the lines to be not considered as an executable code. To generate a result file: just remove the # characters and paste the lines again.

4.4.3 Time units and intervals

The time unit is days: The unit 1 equals one day.

In order to generate the example data set, we have measured every minute. The data were then interpolated with a time step of 0.004 days. This results in a data point beeing available every 86400*0.004 seconds.

4.4.4 Data range considered for the parameter estimation

The estimation of the 8 model parameters $Q_{baseflow}$, $Q_{0,interflow}$, k_{rec}, *a, b, c, A and "phase"* is based on the whole range of the measured data sets *"input_file_Q_ww"* and *"input_file_COD_ww"*.

In practice, the control of the subset of data that is considered for the parameter estimation is done by preprocessing the two input files (COD and Q): the whole series contained in these two files will be the basis for t the parameter estimation. To exclude certain time spans, these parts need to be deleted from the read in files.

4.4.5 Data range considered for the calculation of totals

The date range that is considered for the calculation of totals ("Total Amount of Infiltration", "Average Infiltration Ratio") is specified by "t_start" and "t_end" (User Input 4 and User Input 5):
Based on the estimates for $Q_{baseflow}$, $Q_{0,interflow}$ *and* k_{rec} the total volume of infiltration that was discharged within the span of time between t_start and t_end is integrated. This result is made available by the R-Script under the output variables name "Total_inf_MCS". Analogous the output variable "X_Total_inf_MCS" relates the volumes of infiltration to the total amount of wastewater discharge within this considered span of time.

4.4.6 Consideration of input uncertainties with respect to the two measured variables

To quantify the effect of input uncertainties (stemming from possible systematic errors embedded in the two measured variables) on our estimates, a Monte Carlo Simulation step is included in the R-Script:
For both measured variables ($COD_{wastewater}$ and $Q_{wastewater}$) a hypothetical constant offset error α and a relative error β is assumed:

$$COD_{wastewater,measured} = \alpha + \beta \cdot COD_{wastewater,real}$$
$$Q_{wastewater,measured} = \alpha + \beta \cdot Q_{wastewater,real}$$

The number of Monte Carlo Simulation runs to be performed is specified in User Input 21: "n.MCS". The assumed statistical key parameters for the probability distributions of these error terms must be specified in User Input 18 ("syst_errors_means"), 19 ("syst_errors_ranges") and 20 ("syst_errors_stdvs"). Details about the format of these inputs can be found in the scripts embedded comments.

According to these specifications, the parameter estimation is repeated "n.MCS" times. Each of these estimations is based on a newly drawn random sample for the error parameters. As an intermediate result we obtain a number of "n.MCS" sets of estimates for $Q_{baseflow}$, $Q_{0,interflow}$ *and* k_{rec} . From these sets a number of "n.MCS" simmulation results for the two output variables "Total_inf_MCS" and "X_Total_inf_MCS" is calculated.

The statistical key parameters of the probability distributions of "Total_inf_MCS" and "X_Total_inf_MCS" are made available by the following R-code:

```
summary(Total_inf_MCS)
summary(X_Total_inf_MCS)
quantile(Total_inf_MCS,probs = c(0.025,0.975
quantile(X_Total_inf_MCS,probs = c(0.025,0.975))
```

The distribution of "Total_inf_MCS" and "X_Total_inf_MCS" is also graphically displayed in the two small histograms that are part of the implemented standard graphical output **(**figure 1).

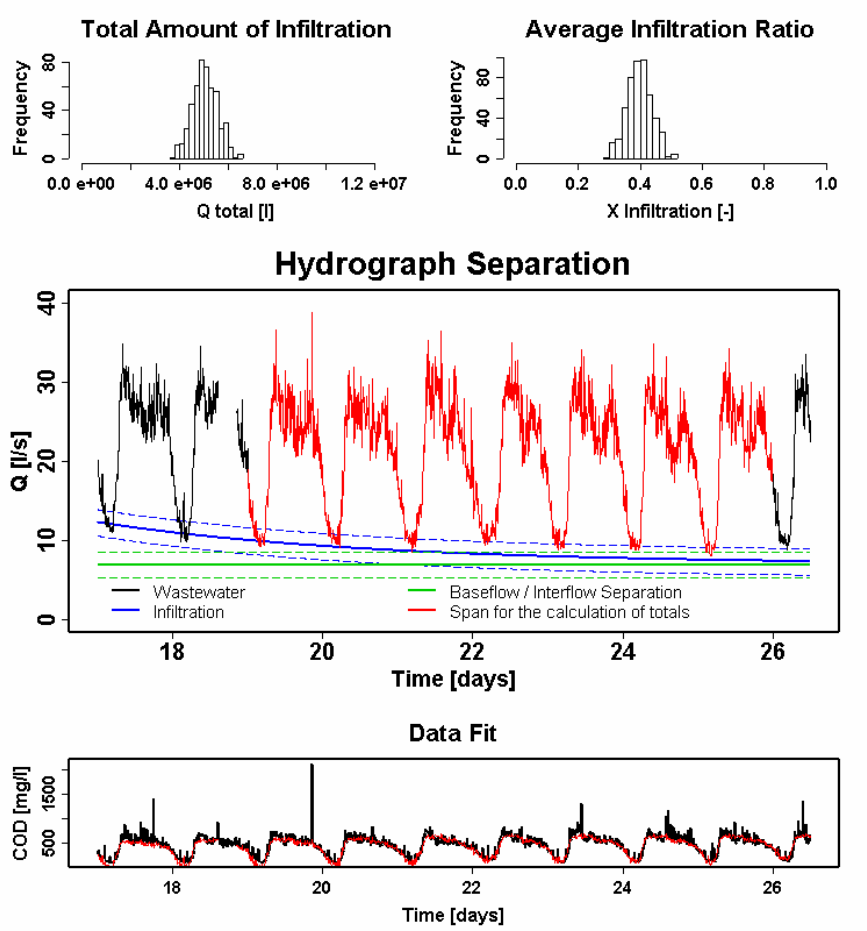

Figure 1: Standard graphical output of the R-script file ChemHydSepMCS.R

4.4.7 graphical control of fit quality

For a first control of model structure,a plot is generated automatically that compares the modeled COD time series to the measured data **(**figure 1).

4.4.8 Control on the model definitions

From version 1.1 or higher, the model definition can easily be controlled by the "User Inputs" 11 to 17: If an initial value (a number) is entered here, the parameter will be considered in the model and a parameter value will be estimated. Alternatively the entry "NA" excludes the model part that is described by this parameter.

In example:

```
Q_baseflow.ini = 40                #User Input 11
Q_0_interflow.ini = NA             #User Input 12
k_rec.ini = NA                     #User Input 13
COD_fs_a.ini = 700                 #User Input 14
COD_fs_b.ini = 10                  #User Input 15
COD_fs_amp.ini = NA                #User Input 16
COD_fs_phase.ini = NA              #User Input 17
```

Would mean: The model consists of a constant baseflow (Q_baseflow) and a polynomic description for the COD_foul sewage (COD_fs_a and COD_fs_b). Interflow and time dependency of COD_foul sewage (COD_fs_amp, COD_fs_phase)are assumed to be not present.

(Whilst performing the model structure selection, it is recommended to set the number of Monte Carlo Simulations down to n=3, to save computing time.)

4.4.9 Direct output of the individual parameter estimates

The 8 model parameters $Q_{baseflow}$, $Q_{0,interflow}$, k_{rec}, a, b, c, A and "phase"are fitted from the start values given in "User Inputs" 11 to 17. To obtain information about the final estimates (i.e. the final values that have been found for Input 11 to 17) you need to paste the following lines of the R-Sript to the R-console:

```
# Output of identified model parameters
estimates_MCS                    #Matrix of estimates from all MCS runs
summary(estimates_MCS)           #Statistical summary of estimates from all MCS runs
```

4.4.10 Save and / or load an R-workspace image

It can be useful to save an image of the R-workspace after having conducted the script calculations. This allows accessing all variables later by reloading them to R (without the time consuming need to conduct all R calculations again):

```
save.image(file = "imagefile.Rdata", compress = TRUE)
load("imagefile.Rdata")
```

4.4.11 Excluding parts of an R-script

In the script code, the prefixing character # causes a line to be not considered as executable code (commenting out). If you want to include lines of the script that are for the moment commented out you need to remove the # characters. (This offers a possibility to include or exclude parts of a script from being executed.)

5 References

Ihaka, Ross and Gentleman, Robert (1996): R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics,* **5**(3), 299--314.

Kracht O., Gujer W. (2004): Quantification of infiltration into sewers based on time series of pollutant loads. *Proceedings of the 4th International Conference on Sewer Processes and Networks*, Funchal, Madeira, Portugal, 22-24 November, 293-300.

Reichert, P. (1998): AQUASIM 2.0 – User Manual. Technical report, Swiss Federal Institute for Environmental Science and Technology (EAWAG), Dübendorf, Switzerland