Using typical daily flow patterns and dry-weather scenarios for screening flow rate measurements in sewers

E. Piatyszek*, C. Joannis** and M. Aumond**

*Société Rhéa, 11 rue du Vieux Pont, 92000 Nanterre, France (E-mail: eric.piatyszek@emse.fr)
**Laboratoire Central des Ponts et Chaussées, Route de Bouaye, BP 4129, 44341 Bouguenais Cedex, France (E-mail: claude.joannis@lcpc.fr; michel.aumond@lcpc.fr)

Abstract Monitoring systems for measuring rainfall, as well as flow rates and pollutant quantities conveyed by sewers and/or discharged by stormwater overflow devices, have become a common feature in many municipal sewer services, in part spurred by recent regulatory requirements. However, the state of measurement conditions in sewer facilities does not always ensure reliable results. For this reason, it is essential that measured values be carefully screened prior to their use, since many sources of disruption capable of skewing data can be encountered. The present article describes a method for validating dry-weather data a posteriori. This method relies upon flow rate forecasts, a combination of standard daily wastewater flow profiles and an estimation of infiltration flows. Measurement results are then compared with this forecast and an appropriate series of statistical tests are run to detect all major data anomalies. A number of diagnostic rules are then applied in order to derive an initial interpretation of these anomalies and, in particular, to identify the influence of rainfall events.

Keywords Flow rate measurements; data screening; residual analysis; automated classification; sewerage

Introduction

The implementation of measurement devices within sewer systems has become widespread, spurred by the recent introduction of regulations requiring “self-monitoring” (December 22, 1994 decree). In general, such systems are based on a network of measurement stations for reading rainfall and flow rates, in conjunction (where applicable) with effluent quality assessments. The objective of self-monitoring may be combined with other objectives: ongoing sewer evaluation, calibration of design models, etc.

The state of measurement conditions within sewer facilities does not ensure a perfect reliability of results. For this reason, it is essential for measured values to be carefully screened prior to validation; this process entails determining the likelihood of the results in the light of all information and knowledge of the system/conditions available through past experience.

The present article describes an a posteriori validation method for flow rate data series. It involves an automated validation-assistance tool designed to alert the system operator to potentially “suspicious” data readings, thereby eliminating the need to examine large quantities of “ordinary” data.

Most researchers dealing with data validation for urban sewerage focused on rainfall data, and flow rate data during rain events (Jörgensen et al., 1998), (Maul-Kötter and Einfalt, 1998), (Boukris et al., 1999), (Szafnicki et al., 1994). Instead, our method uses typical dry-weather flow patterns as a reference. Such a reference can serve several purposes, including those associated with a system primarily aimed at identifying rainfall flow conditions. The very definition of a rainfall period (and with it the shift to other anomaly-detection models) proves to be somewhat arbitrary, yet may be sharpened by virtue of characterising rainfall by its effects, as defined with respect to the observed dry-weather values. Moreover, a dry-weather anomaly-detection procedure may enable the anomaly to be corrected even before the critical events occur.
The validation method featured herein is based on a comparison of measured values to a dry-weather forecast of flow rates comprising both foulwater and extraneous infiltration. As regards foulwater, the focus has been placed on catchments serving primarily residential uses. For these flows, the statistical modelling approach incorporated the concept of a standard daily wastewater flow profile. The pertinence of this concept was then studied during a preliminary phase by applying an automated classification procedure to a series of measurement results from which the infiltration component had already been removed. In this way, all objectively evaluated similarities in the families of daily hydrograms (as described by various types of indicators) could be highlighted.

Thanks to these initial results and by including an estimation of the infiltration component (Joannis, 1994), a method for detecting and identifying anomalies was developed. These two phases of processing measurement results were refined and tested on data generated by the monitoring system set up by the Nantes Metropolitan Area (France) Authority.

**Typology of daily wastewater hydrograms**

**Description of the method**

The objective of this method is to establish a set of standard daily wastewater flow patterns. On the basis of general or detailed descriptions of the daily hydrograms (net of the influence of parasite infiltration water, as obtained by subtracting the value of the minimum night-time flow rate), the number of distinct daily patterns is optimised so as to ascertain the “regular” variability in the results and then to detect, or even quantify, atypical deviations within a one-day time frame.

The descriptions employed for these daily profiles consist of the hour-by-hour flow rates (either in absolute terms or as a percentage of total daily volume) along with additional variables, such as the daily volume or dispersion of hourly flow rate values over the course of a single day.

The automated classification procedure involves combining onto a scale of distance (or another proximity-based criterion) all days with similar profiles. The selected criterion is the Euclidian distance between the “n”-variables descriptions of the various days. The method employed is known as “ascending hierarchical classification” or AHC, whereby groupings of days are assembled according to their relative proximity. At the outset of the calculation, each group corresponds to a separate day. Next, the two closest groups are combined in a successive sequence until obtaining a single group that contains all of the individual days. In this manner, a tree structure is ultimately derived, with each level of this structure corresponding to a specific classification. The most pertinent level is then chosen depending upon the evolution in the distances between classes.

Preliminary screening is performed in order to avoid expanding the number of classes, leaving groups of just a few days remaining at the advanced stages of the combinatorial process (i.e. composed of unclassifiable rainfall days and/or atypical days). The advantages of this screening step include preventing saturation of the statistics software package and enhancing the legibility of results.

The final phase is devoted to analysing the profiles obtained. It is preferable to confirm the results of the AHC by an interpretation of the classes created by the method. For wastewater flow rates, the criteria herein pertained to local economic and residential activity (workdays, holidays, weekends, vacation periods, etc.). Such an interpretation also makes it possible to assign *a priori* a given day to one standard profile, a step which proves useful when running the applications.
Application to the sewer system of the City of Nantes

This method was applied to the flow rate data generated from four measurement stations within the monitoring system set up for the City of Nantes’ sewerage service. The data presented correspond to a station located on the wastewater collection network, with an average daily dry-weather volume of 20,000 m³. At this particular site, hourly flow rate data is available over a continuous 561-day period, extending from February 1, 1998 through September 30, 1999.

Data pre-processing. The raw data (561 days) include rainfall days and days presenting other measurement or operating anomalies. Screening is carried out using the “maximum difference” criterion between a given hourly reading and the observed median value over the entire data series. In all, 291 days were selected. The number of disqualified days is not critical; this number does not jeopardise the assembly of subsequent data classes, provided no specific category of standard daily profiles has been systematically excluded. This screening process served to eliminate the rainfall days with an obvious impact on flow data: 95% of the selected days displayed a pluviometric reading of less than 1.5 mm of precipitation per day, whereas 29% of all days in the raw data series had readings above this threshold.

Automated classification. Several techniques for describing daily hydrograms were compared:

- raw flow rate data for each of 24 hour-long periods;
- flow rate as a percentage of the total daily volume of wastewater for each of 24 hour-long periods;
- same as above, with two added variables: the total daily volume and dispersion of flow rates throughout the day, as represented by the standard deviation of the 24 hourly values (expressed in absolute terms);
- daily description by means of 6 time periods, with the respective flow rates expressed as a percentage of total daily volume. A preliminary frequency distribution analysis has demonstrated that a 6-period representation should be sufficient;
- same as above, with two added variables: the total daily volume and dispersion of flow rates throughout the day, as represented by the standard deviation of the 24 hourly values (expressed in absolute terms).

The classification carried out on the basis of relative flow rates of wastewater over 24 hourly periods produces the most attractive results. As a matter of fact, the classifications derived from the other descriptions yielded an expanded number of classes, which means that the classes are only slightly differentiated and composed of very few days. It would seem that incorporating the daily volumes disturbs the classification process, due perhaps to the gradual changes in these volumes over time. The classification obtained is independent of the variations in daily wastewater volumes (whose standard deviation is of the order of 25% of the mean value).

So the classification is sensitive to the description of daily flow patterns. On the opposite it seems quite robust with respect the level chosen to “cut” the classification tree, as can be seen in Table 1.

<table>
<thead>
<tr>
<th>Level in classification tree (total number of classes)</th>
<th>25</th>
<th>50</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of classes with population &gt; 10</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mean population of classes with population &gt; 10</td>
<td>70.7</td>
<td>65.3</td>
<td>56.3</td>
</tr>
</tbody>
</table>
It can be seen that even at the beginning of the aggregation process (total number of classes = 75), 3 main groups are present, which grow as the process goes on.

Three typical patterns have been identified. The a posteriori analysis of these daily profiles in terms of calendar patterns reveals that 98% of the first class is composed of workdays (Monday to Friday). The second class contains 28% Saturdays and the third 46% Sundays. The three flow rate profiles, relative to total wastewater volume, for these standard days are presented in Figure 1. The primary characteristic of these three patterns is their time lapse over the 7 a.m.–5 p.m. period. The standard day exhibiting the earliest rise in flow rate corresponds to those days with intensive economic activity. Sunday, the least-active day of the week, happens to be the day showing the latest rise in flow rate. As for Saturdays, the rise in flow rate lies between weekdays and Sunday, with a relative flow rate profile over the second half of the day closely resembling that of Sundays. This discrimination between weekends and the five-day workweek appears to a similar extent at the three other measurement locations studied in Nantes (without the Saturday/Sunday distinction). The same type of time lapse is observed with respect to weekends.

**Conclusion**

The classification method applied herein has been intended to combine daily flow rate data series into a small number of homogeneous “standard” profiles, on the basis of a set of objective criteria. It then becomes possible to rationalise the number of dry-weather profiles. From the sites studied in the Nantes Metropolitan Area, available data has enabled the establishment two or three such standard wastewater profiles, as described by hourly flow rates and expressed as a percentage of total daily volume. These profiles could be identified a posteriori by means of very simple date-related criteria. By virtue of such a reference, various computations and operations thereby become possible: statistical tests to help validate the dry-weather data series, detection of rainfall events exerting an impact on sewer system operations, quantification of rainwater included in total sewage effluent, etc.

**Detection and diagnosis of anomalies during dry-weather periods**

**Description of the method**

*Model development.* This model is intended to forecast the dry-weather flow rate, which is composed of two components: a foulwater flow and an infiltration flow. For foulwater flows, the automatic classification results have allowed us to identify three scenarios as yielding the most meaningful description (working day, weekend, holiday).
During a first step, we choose, at the beginning of the period to analyse, dry weather days to build the initial scenarios (Figure 2). For each scenario, four dry-weather days are selected (each day is represented by 24 flow values—1 hour time step). Then, we extract foulwater flows and infiltration flow.

As regards infiltration flow, the model calculates the mean of the four most recent days. Other more complex scenarios, suited to rainy weather or to specific sewer network layouts, could also be designed.

During the validation process, the dry-weather scenarios are updated. When the observed daily flow data is consistent with the dry-weather scenario, the daily flow rate for this particular day is inserted into the group of profiles used to build the dry-weather scenario. This day then replaces the least recent day in the series. This procedure makes it possible to incorporate the changes in flow rate which are not integrated by the typology. The same updating sequence is followed for the infiltration flows as well. Moreover, the infiltration flow rate is re-estimated following rainfall events, when a single shift is sufficient for validating a daylong data series that seems to present anomalies.

Detection module. For each day to be validated, two data elements are known: the flow measurement for the given day, and the dry-weather (model) scenario. These two elements are described using a one-hour time step; two 24-value matrices can thus represent them. The subtraction of these two data then yields a signal, called a residual. This residual in fact acts as an anomaly-indicating variable, inasmuch as its value lies close to zero when the two data are relatively consistent with one another; in contrast, an inconsistency between the two data will result in a modification of this residual’s characteristics. More precisely, inconsistencies cause a modification to the residual mean; such modifications have been detected using a binary sequential probability ratio test (Wald, 1947). This test is quite robust (Humenik and Gross, 1990) and appears to be more efficient than the classical “filtered derivatives” detectors as it minimises the detection time lapse for a fixed false alarm rate (Basseville, 1988). Moreover it is possible to link the parameters of the test to both false alarm and missed alarm probabilities (Brunet et al., 1990), (Gross and Humenik, 1991). We associated Wald’s test with a simple threshold test (Piatyszek et al., 2000). The
speed of the simple threshold test is thereby preserved, while securing some robustness with respect to false-alarm occurrences thanks to Wald’s test.

Diagnostic module. Once a significant difference is detected at a node, it becomes necessary to seek the cause of this anomaly; potential causes may include:

- failure of the measuring device;
- atypical sewer network operations;
- a rainfall event (typical or exceptional hydrological parameters);
- a model scenario poorly adapted to the real-life situation.

In order to remove this ambiguity, focus is confined to the hydraulic behaviour at the measurement location where the anomaly has been identified. Simple rules are then introduced to complement the local hydraulic data in order to diagnose the cause of various inconsistencies between data measurements and the model scenario, due to:

- measurement errors, when the measured data are less than those of the scenario and, in particular, when these measurements are zero or negative;
- rainwater flows, when measured data are greater than those of the scenario and when available pluviometric information confirms the occurrence of a rainfall event;
- high-rate infiltration flows in combination with rainwater flows, when measured data are greater than those of the scenario and when this detected inconsistency follows rainfall events characterised by a total precipitation of at least 20 mm over a 72-hour period.

More precisely, the validation process is described in Figure 3: First, the procedure chooses the right scenario to validate the day, secondly, it compares the observed daily flow with the dry weather scenario chosen, thirdly, it diagnoses the cause of inconsistencies between these two informations, finally, if the daily flow is coherent with the scenario, it updates this scenario by replacing the oldest profile used to build this dry-weather scenario.

Example on the combined sewer network of the City of Nantes

The validation method described above was tested on data provided by Doppler flowmeters and generated from several locations along the combined sewer network of the City of Nantes. We present here the results obtained with data collected in 1998–1999, at the input of the main treatment plant, which processes flow rates averaging 100,000 m³/day in dry-weather conditions.

An example is given by Figure 4. This example displays the results of the validation
sequence on the data series between September 15, 1998 and September 29, 1998 (see figure below). The figure contains:

- in the upper part, the measured flow (thin black curve) and the model scenario (gray curve); and
- in the lower part, the state of the node. This state can fluctuate among three different values (‘OK’: consistency between measurements and the scenario, ‘–B’: detection of a negative deviation (measurement higher than the scenario), and ‘+B’: detection of a positive deviation (measurement less than the scenario)).

This figure clearly indicates those rainfall periods exerting an influence on flows (–B state). Moreover, the relatively short-lived –B state (of the order of one or two hours) has also been detected. A diagnosis carried out a posteriori on the basis of additional information has revealed that these variations represent tidal intrusions into the drainage network. Furthermore, a long +B state is detected at the end of this period: the measurement signal is null, which reflects failure of the Doppler sensor.

The rules-based diagnostic module currently enables us to explain the inconsistencies due to rainfall, including the associated delayed effects. The tidal intrusions could easily be treated in a similar fashion.

Figure 5 presents an evaluation of the detection and identification of anomalies at the same measurement site as cited previously. It can be observed that the days without any anomaly (in comparison with the dry-weather scenarios) are less frequent, yet rainfall remains the primary cause of deviation with respect to a dry-weather scenario; moreover, unidentified anomalies are relatively few in number. This distinction depends upon the assigned detection thresholds as well as upon the rules used for interpretation, which are still very basic. A certain proportion of the unidentified anomalies therefore would correspond with forecast inaccuracies and not with measurement errors.
Conclusion
This dry-weather flow rate validation procedure for sewer systems is intended to determine whether the flows observed at various points in the network are consistent with typical dry-weather flows.

This procedure consists of a detection module that checks the consistency between measurement data and the dry-weather model scenario. This scenario represents the forecasted flows through the operator’s conception of the network layout and within the hydrological conditions of a dry-weather period.

The differences between these two elements, or anomalies, may be due to a variety of causes (e.g. measurement error, an inaccurately identified scenario, an atypical network layout, a rainfall event). The search for these causes involves a diagnostic module based on a simple set of rules.

In the current state, detection is carried out automatically and allows locating in a reliable and systematic fashion those periods during which measurements deviate from the dry-weather scenario. In contrast, the diagnosis is not completely automated and the user is required to seek additional information in order to explain certain detected errors (e.g. tidal intrusions). However, it is hopeful that such a diagnosis can be streamlined in the future, by virtue of receiving systematic information on network operations.

Overall conclusion
Models designed to forecast dry-weather flow rates within a sewer network have hitherto been paid relatively little attention, whereas systems operation models under rainy conditions have been developed extensively. Indeed, these latter models boast a more deterministic feature with more immediately perceptible impacts. Nonetheless, relatively intricate statistical models do enable us to forecast the various components of dry-weather flows as well as detecting the presence of “suspicious” results. Further model developments will allow refining the sensitivity of this detection process; examples of such improvements are the combined use of two time scales (e.g. daily and hourly), and additions to the body of diagnostic rules.

Acknowledgements
This program has been supported by the Ministry of Environment and the City of Nantes.

References