# Managing a Million Data Points: A How-To Guide

**Kimberly A. Coffie[1], Sarah A. Crawford[1], David L. Lane**

[1]Hampton Roads Sanitation District, Virginia Beach, VA 23455
Email: kcoffie@hrsd.com

## ABSTRACT

The Hampton Roads Sanitation District (HRSD), serving the Tidewater Area of Virginia, has been tasked to maintain strict data reliability of its flow, pressure, and rain gauge monitoring sites. Per the Environmental Protection Agency (EPA) Administrative Order (AO), HRSD is required to maintain 75% data reliability for each individual meter during each monthly monitoring period and 90% data reliability for all data for each type of monitoring sensor or meter during qualifying wet weather events. HRSD has deployed 122 pressure monitoring sites, 175 flow monitoring sites and 66 rain gauge sites. Approximately 87,000 data points must go through a daily QA/QC process to ensure data validity and reliability for the flow meters and pressure sensors alone. The automation of data quality alerts through SQL statements has created a streamlined, efficient method to handle thousands of data points on a daily basis to identify system changes, system reaction, and instrument fouling.

**KEYWORDS:** Data Analysis, wastewater, SQL, Telog

## INTRODUCTION

The Hampton Roads Sanitation District (HRSD), serving the Tidewater Area of Virginia, has been tasked to maintain and document data reliability of its flow, pressure, and rain gauge monitoring sites. Per the Environmental Protection Agency (EPA) Administrative Order (AO), HRSD is required to maintain a data reliability of 75% for each individual meter during each monthly monitoring period and 90% data reliability for all data for each type of monitoring sensor or meter during qualifying wet weather events. HRSD currently has 122 sites monitoring pressure, 175 monitoring flow, 66 rain gauge sites and 24 groundwater monitoring sites. Currently, only those sites chosen to be used for the hydraulic model calibration are required to follow the data reliability guidelines outlined above, which include 46 flow meters and 75 pressure sensors. Additionally, wet well levels, pump run status, and pump revolutions per minute (rpm) are also recorded and utilized for Quality Control/Quality Assurance (QA/QC) purposes. Approximately 87,000 data points must go through a daily QA/QC process to ensure data validity and reliability for the flow meters and pressure sensors in support of the hydraulic model calibration.

Presently, HRSD employs one Data Analysis Manager, two full-time Data Analysts, two part-time Data Analysts and external consultants to perform QA/QC on each data point, including data from sites not being used for hydraulic model calibration. Due to the large amount of data points, the QA/QC process was initially a daunting manual task. In an effort to make this process more efficient, an automated process was developed to alert Analysts to data anomalies.

## THE SOLUTION

### Automated Alerts

Field data is managed using Telog Instruments, Inc® equipment and software. The database where this information resides has been programmed to generate automatic alerts through a series of SQL statements if the recorded raw data deviates from expected "normal" conditions. "Normal" conditions vary from a calculated 4-week average dry period, wet weather response, and "normal" meter function based on historical data.

The automation of "Alerts" was made possible with SQL functions that were previously programmed within the software as well as the data server by Telog®. The pre-programmed functions require the input of a SQL statement stating the alert criteria, as well as the interval compression. In addition, it was necessary to input new calculated measurements in which the alerts for that particular site would calculate from. More than 3,000 sensor and operational specific alerts have been inputted to assist and automate the QA/QC process. Each type of alert has been useful in identifying system changes, system reaction, and instrument fouling or failure. Table 1 provides the generic definition of each alert; however, certain sites may require an adjusted SQL statement based on the sites "normal" trend.

**Table 1. Alert Definitions**

| Alert | Purpose of Alert |
|---|---|
| Flow/Pressure_Deviation | Analysts are alerted when the flow or pressure exceeds ±40% of the dry week average on an hourly basis. Alerts are calculated based on a 4 week average dry period, which also takes into account, weekdays and weekends. |
| Flow_Downstream | Analysts are alerted when the flow at the site is greater than the flow at a downstream meter. |
| Flow_Upstream | Analysts are alerted when the flow at the site is less than the flow at an upstream meter. |
| Flow/Pressure_WW_Peak | Analysts are alerted when the flow or pressure at the site exceeds three times of the 4-week dry average when there has been greater than 0.1 inches of rain in the past 24-hours. |
| Flow/Pressure_Sensor | Analysts are alerted when the minimum, average, and maximum flow/pressure measurement is within 0.005% of each other. This alerts assists in evaluated sensor failure, sensor fouling, or sensor flat lining. |
| Rain_10Greater | Analyst are alerted when the rainfall at the site records 25% greater rainfall than neighboring rainfall sites after 0.3" inches of rain. |
| Rain_10Less | Analyst are alerted when the rainfall at the site records 25% less rainfall than neighboring rainfall sites after 0.3" inches of rain. |

Figure 1 displays an example of a Flow Deviation alert. This particular SQL statement alerts when the flow at the metering site deviates 40% from a calculated 4-week average dry period on an hourly basis and less than 0.1 inches of rain has occurred from the nearest rain gauge in the past 48-hours. The Flow_Deviation alert has been useful for identifying Inflow/Infiltration (I/I) effects, a recent calibration, a system diversion, and a system bypass. The same SQL statement is used for the Pressure_Deviation alert by exchanging flow for pressure.

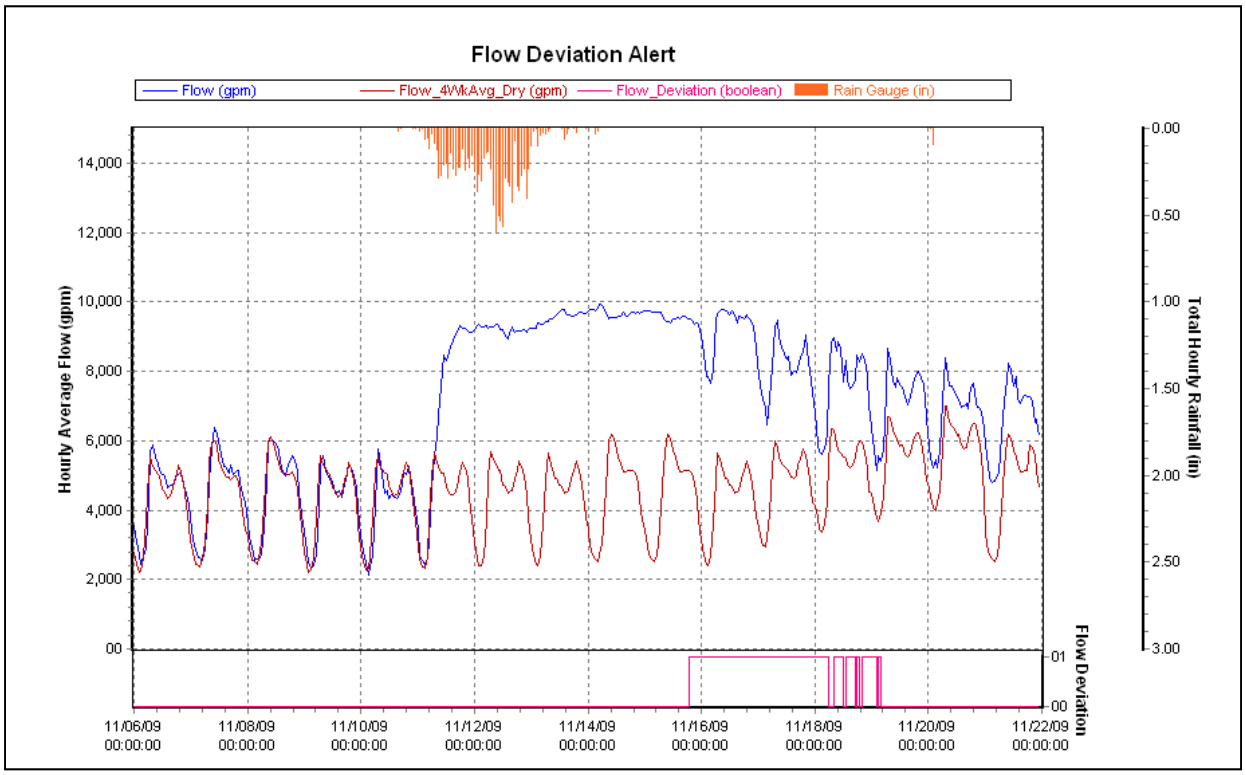


**Figure 1. Example of Flow_Deviation Alert Proceeded a Rainfall Event**

Figure 2 displays an example of a Pressure_Sensor alert. This particular SQL statement alerts when the minimum and maximum 2-minute values deviates from the average 2-minute reading by 0.005% on an hourly basis. The Pressure Sensor alert has been useful in identifying sensor fouling, a sensor reading at the maximum or minimum scale, and seasonal changes. The Flow Sensor alert is most useful in alerting when the meter is recording at the maximum or minimum scale. Figure 2 shows an example of a red-valve pressure sensor that became clogged with grease. Once the clog was removed by HRSD staff, the data returned to its valid state.

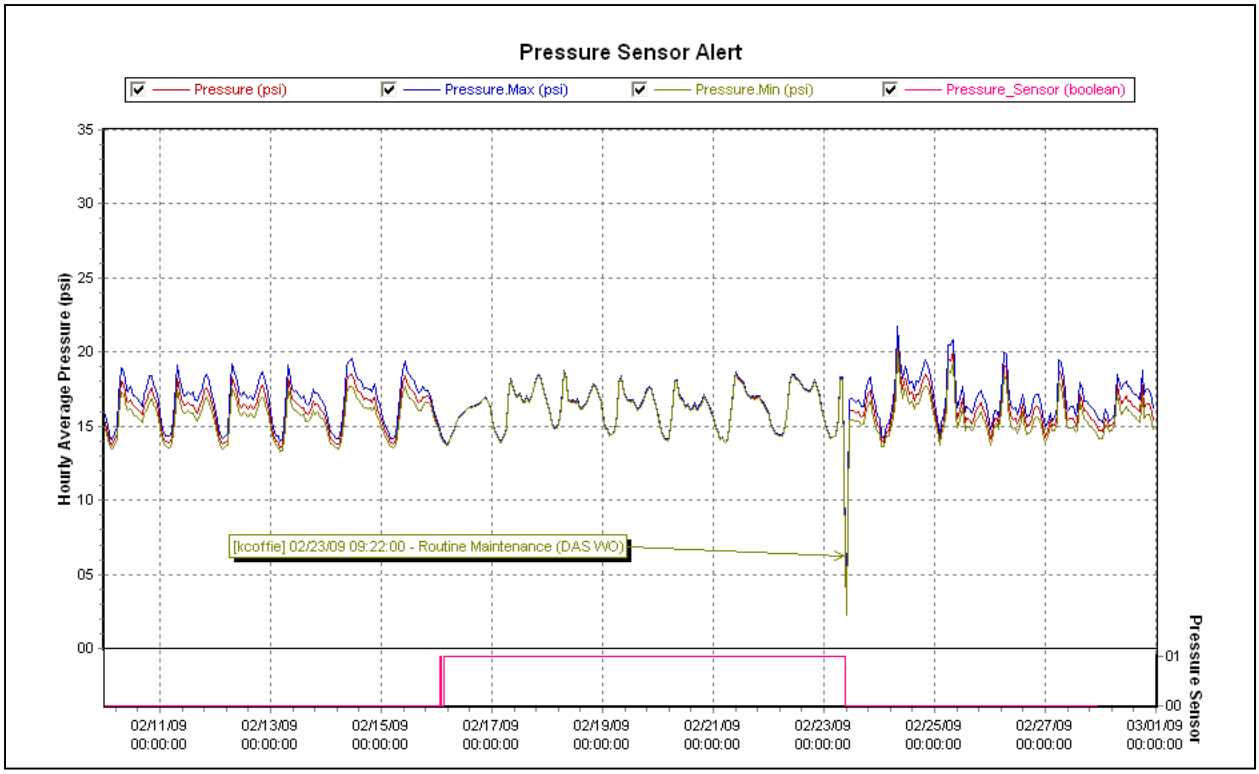


**Figure 2. Example of Pressure Sensor Alert**

However, automatic alerts were not practical for all sites, especially in reference to the Flow_Deviation alert. In some cases, pump station flows may not produce a "typical" diurnal curve and daily manual reviews may be necessary.

**Automated Work Orders**

Alerts inputted into the Telog® SQL server can be sent to necessary personnel by SMS text, email, or paging system. HRSD chose to have the automated alerts sent by email, which in turn were routed into the Numara Track-It!® software to prepare automated work orders. The automated work orders are reviewed and the anomaly is deemed either valid or invalid. See Table 2 for examples of valid and invalid data.

**Table 2. Accepted Identifiers for Valid and Invalid data.**

| Invalid due to: | Valid due to: |
|---|---|
| Site Maintenance | Data Adjusted |
| Sensor Drift | Normal Trend |
| Sensor Failure | Operational Change |
| Networking Issue | Seasonal Trend |
| Power Failure | User Specified |
| Site Constraint | |
| Unknown Causes | |

Approximately 2,000 alerts are received and resolved through the Numara Track-It!® software on a monthly basis.

**Flagging Invalid Data**

It was necessary to automate data reliability for each sensor for efficient reporting. Telog® made automation capable through their software in which the length of invalid data is compared against the time duration needed for reliability (i.e. – 1 month, 6 months, etc.). Figure 3 shows the capability of flagging data. In this example, the analyst chose the start and end time for the invalid data, inputted the timestamps into the Telog® server by simply highlighting the data, and has the capability to annotate the data. Below, invalid data is shown in the color yellow. From the timestamp highlighted, Telog® can calculate the percent of valid/invalid data for the selected timeframe.

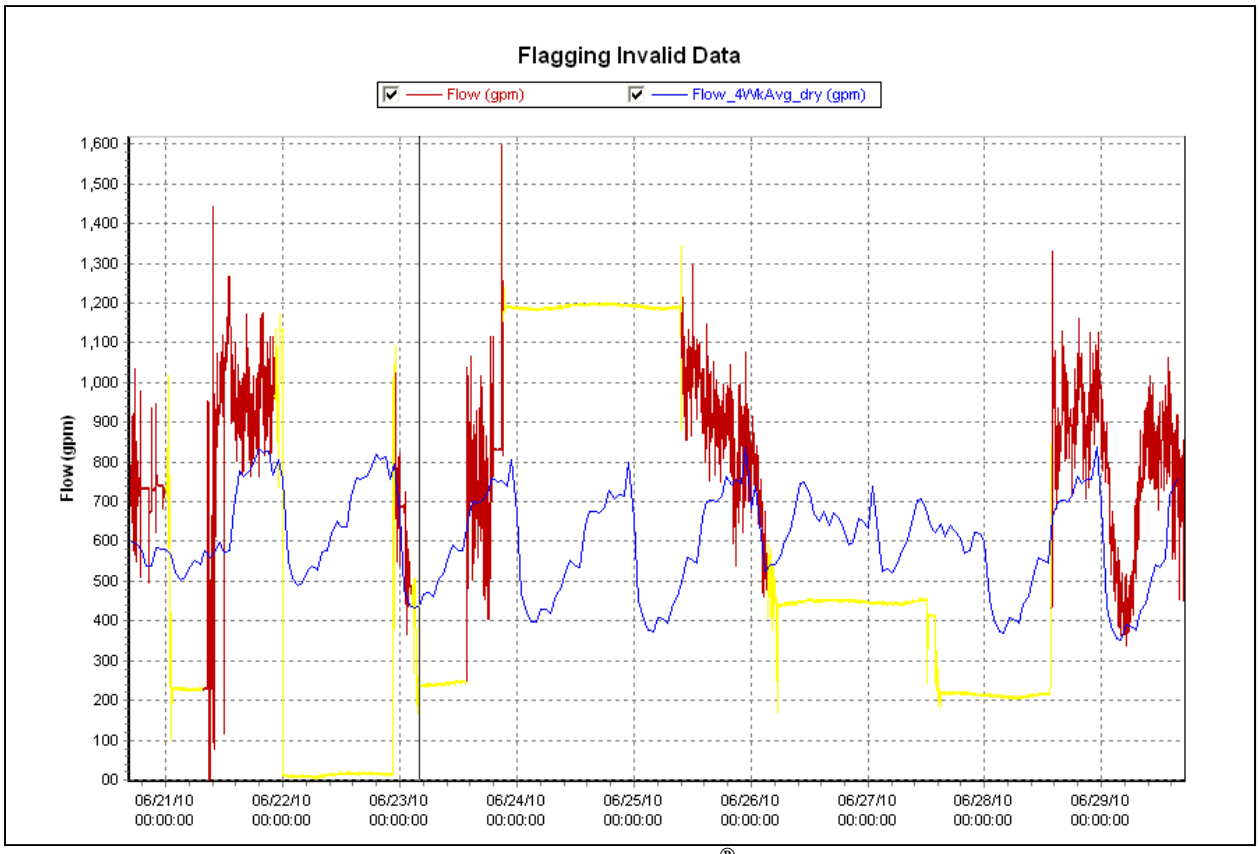


**Figure 3. – Flagging Invalid Data Through the Telog® Software**

## CONCLUSIONS

The automation of data quality alerts has created a streamlined, efficient method to handle thousands of data points on a daily basis to identify system changes, system reaction, and instrument fouling. Analyst did find it necessary to adjust a handful of alerts SQL statements due to “site specific” conditions and it is expected that alerts will continue to be adjusted as operational changes are made to the interceptor system. Additionally, an alert SQL statement may not be appropriate and a daily manual review of the site would be necessary. Review of the process has revealed increased efficiency of the data analysts and their ability to annotate data valid or invalid on a daily basis. The process first created by Telog®, Inc and continuously

perfected by their staff and HRSD staff, has become an innovative process to improve data quality and analysis of large wastewater interceptor systems and managing a million data points.

**ACKNOWLEDGEMENTS**
Hampton Roads Sanitation District (HRSD) Data Analysis Section Staff
Telog Instruments, Inc® staff with a special thank you to David Baginski and Daniel Drake