Data Validation Helps Catch Real-World Problems

by **Ken Cessac and Ardis Bartle, PGAS Systems,** Stafford, TX

(This is the last of a trilogy of articles on Data Validation in Gas Measurement. The November issue covered the reasons that data validation is not used more frequently in gas measurement systems, configuration of validation rules, the issues with maintaining validation rules, and how a good measurement system makes itself "user-friendly" in regards to validation.)

This article focuses on how data validation can reduce the amount of time needed to solve gas measurement problems. Validation once was a luxury in chart-processing organizations, but has become a necessity in the age of electronic flow meters.

Validation can be one of measurement's most valuable tools in identifying problems as well as efficiently using both the field technician's and analyst's time. The problem facing today's gas measurement departments is to find methods to validate the large amount of volume-related data with minimal resources and available time. Some of the most common problems faced by the measurement analyst are:

- Missing data,
- Meter freeze,
- Inconsistent flow data, and
- Configuration errors.

Looking for and detecting these problems can be a complicated and time-consuming process if done manually. The amount of data an analyst is required to review and validate has increased dramatically with the use of electronic flow computers for gas measurement. The older chart recorders provided a simple graphical recording of the flow variables, which typically include seven days of flow data on a single chart. A trained analyst could quickly detect problems by performing a simple visual review of these charts. This process of visually reviewing data to detect problems becomes very difficult with flow computers due to the large amount of information associated with the devices. In addition to the large amount of volume data, the flow computer provides gas quality and configuration information that must be reviewed and validated. The use of automated validation with exception processing can provide a solution to this problem. Some examples of flow data problems are discussed here.

Missing Data

One of the biggest problems associated with the use of flow computers is missing data. Flow computer failures, communication outages, or corrupt data can all lead to data missing within the host measurement system. A company with as few as 100 flow computers logging hourly data would create close to 72,000 volume records each month. The process of manually reviewing thousands of records looking for missing data can be a tedious task. Manually searching for missing flow data require the analyst to compare each volume record's production start and end dates to the previous and following records, verifying that there are no breaks in the data. Look at the example of records with missing data reproduced here.

records in a fraction of the time, usually seconds. When missing records are detected, the system would write an exception record indicating which meter has missing data and the date range of the missing flow data. By using this type of automated validation, the analyst only needs to review the exception log and address the meters that have missing data in place of reviewing all data for all meters. Therefore, the time required to search for missing data could be reduced from days to minutes.

Meter Freeze

A meter freeze is the result of an ice block forming inside the meter's gauge lines. This problem usually occurs on production orifice meters during the winter months and will result in invalid differential pressure and sometimes invalid static pressure readings from the meter. When a meter freeze occurs, the differential values may lock at a fixed

Example of Data with Missing Records

Production Start	Production End	DP	Static	Temp	MCF	DTH
12/02/2002 09:00	12/02/2002 10:00	26.80	149.65	57.09	16.5	16.9
12/02/2002 10:00	12/02/2002 11:00	9.50	147.39	59.02	9.2	9.4
12/02/2002 11:00	12/02/2002 12:00	7.46	148.59	64.77	8.5	8.8
12/02/2002 12:00	12/02/2002 13:00	16.74	151.58	62.12	12.6	12.9
12/02/2002 13:00	12/02/2002 14:00	15.00	152.55	61.35	12.4	12.7
12/02/2002 14:00	12/02/2002 15:00	29.77	153.74	60.96	17.4	17.9
12/02/2002 15:00	12/02/2002 16:00	30.70	153.73	60.47	17.8	18.3
12/02/2002 16:00	12/02/2002 17:00	30.13	154.17	59.89	17.7	18.1
12/02/2002 17:00	12/02/2002 18:00	40.31	155.23	59.20	20.4	20.9
12/02/2002 18:00	12/02/2002 19:00	52.06	158.05	58.73	23.5	24.1
12/02/2002 20:00	12/02/2002 21:00	63.24	166.27	58.46	26.6	27.3
12/02/2002 21:00	12/02/2002 22:00	45.26	167.57	58.15	22.3	22.9
12/02/2002 22:00	12/02/2002 23:00	39.17	163.72	57.88	20.8	21.3
12/03/2002 00:00	12/03/2002 01:00	32.63	165.44	57.69	18.6	19.1
12/03/2002 01:00	12/03/2002 02:00	2.89	159.08	55.29	5.4	5.6
12/03/2002 02:00	12/03/2002 03:00	7.20	152.50	55.70	8.6	8.8
12/03/2002 03:00	12/03/2002 04:00	11.41	152.54	56.49	10.8	11.1
12/03/2002 04:00	12/03/2002 05:00	10.26	151.90	56.40	10.2	10.5
12/03/2002 05:00	12/03/2002 06:00	16.76	152.45	56.74	13.1	13.5
12/03/2002 06:00	12/03/2002 07:00	25.51	152.48	57.29	16.2	16.6
12/03/2002 07:00	12/03/2002 08:00	22.46	151.21	57.42	15.1	15.5
12/03/2002 08:00	12/03/2002 09:00	29.02	150.57	58.05	17.2	17.6

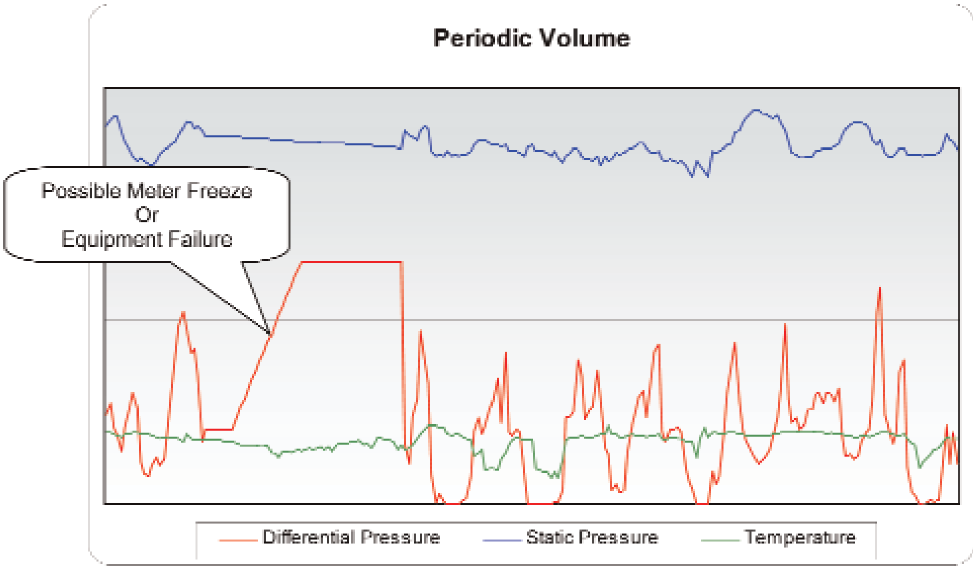
Missing 19:00 to 20:00 Record

Missing 23:00 to 00:00 Record

If we assume that an analyst could manually perform this review of the production dates in as little as one second per record, it would then take an estimated 20 hours to review the above-mentioned 72,000 volume records. An automated missing data validation process could perform a missing data check of the same 72,000 volume

reading or the flow pattern may change and begin following ambient temperature changes.

A knowledgeable analyst performing a manual review of the chart recordings could detect a meter freeze problem. However, this measurement problem is considerably more difficult to detect with data from flow computers. A flow



computer normally reports hourly averages for the different flow variables. Averages tend to filter out sudden flow changes, making it harder to detect if a problem exists, or the exact time a problem occurred. See the Periodic Volume Chart shown here.

Some of the different validation methods that could help detect this problem are Fixed Limit, Repeating Value and Deviation From Historical Average validation checks.

If the ice block occurs in only one of the two differential gauge lines, then the differential pressure value will usually be over or under range. The Fixed Limit low- and high-range checks will detect values that are outside the normal instruments' operating range.

If an ice block occurs with both gauge lines, then the differential pressure will either lock at the current value or vary as ambient temperature changes occur.

A Repeating Value validation check compares the current records data values against the meter's previous records data values. If one or more of the flow variables are exactly the same as a previous record, then the record is flagged for analyst review.

The Deviation From Historical Average validation compares the current flow values against the historical average and flags the values that deviate by more than a user-defined limit. This check will help detect flow values that suddenly increased or decreased from the meter's normal flow pattern.

Inconsistent Data

Another measurement problem associated with the use of flow computers is inconsistent configuration information between the field flow computer and the host measurement system. For example, a flow computer may log an upstream static pressure value when the static tap location is physically connected to the downstream tap. A second issue closely related is the flow computer's calculation of the flow extension. Some flow computers report a flow extension that was calculated using an upstream static pressure value when the flow computer is configured to use a downstream static pressure.

The host measurement system must support these different variations in the flow data in order to accurately process the measurement data. If the host system is not configured to process these data inconsistencies, then the result can be volume calculation errors. The difference between an upstream and downstream static pressure value can be as little as one to two pounds per square inch. This small change can result in volume calculation errors near 1 percent, depending on the static pressure value.

One method of detecting these types of data inconsistencies is for the host measurement system to perform a Recalculated Volume Comparison validation. The host would recalculate the volume for each record received from the flow computer using the host system's configuration information. This recalculated volume can then be compared to the flow computer's originally calculated volume. If the difference between the host's calculated volume and the flow computer's original volume is more than a user-defined limit, then the record can be flagged for analyst review.

Configuration Errors

Configuration errors can be the result of the host measurement system missing events from the flow computer or a field technician entering an invalid value into the flow computer. Whatever the cause, the result can be volume calculation errors. Some common configuration errors are an invalid atmospheric pressure value set in the field flow computer or invalid orifice plate size within the host measurement system. Next to the gas quality values, the orifice plate size is one of the most commonly changed configuration parameter.

Most flow computers can only store the 100 most recent alarms and operator events. If the flow computer generates excessive alarms, the result can be operator events being purged from the flow computer's memory. If this happens, the host measurement system does not receive the event about the configuration change and is now out of sync with the flow computer. An automated validation process can be used to detect these configuration problems.

A Configuration Comparison validation can be used to find anomalies between the flow computer and host measurement systems configuration. This validation method performs a comparison between the host system's configuration and the flow computer's configuration. Discrepancies between the host and flow computer can then be flagged for analyst review.

The Recalculated Volume Comparison validation is an additional method to further validate the configuration information. This method recalculates the volume for each record received from the flow computer using flow variables within the volume record and the host measurement system configuration information. This recalculated volume can then be compared to the flow computer's originally calculated volume. If the difference between the host calculated volume and the flow computer's original volume is more than a user-defined limit, the record can then be flagged for analyst review.

Conclusion

This final part of the trilogy listed just a few of the many problems that can occur with measurement data along with some of the validation methods that can implemented within the host measurement system to detect these problems. Some of the measurement problems described could be detected by manual review while others — such as data inconsistencies — would be nearly impossible to detect without some form of automated validation. A good validation system must not only detect the measurement problems, but must also provide a method for notifying the users of the problems without overwhelming them with thousands of repeating exceptions. A good exception viewer will provide the user with methods to group and analyze the exception results. ***P&GJ***