White Paper

OptiRamp® Leak Detection

Pipeline Monitoring with Advanced Analytics

Vadim Shapiro
Dmitriy Khots, Ph.D.

Table of Contents

Introduction

The *OptiRamp*® Leak Detection Submodule (LDM) is a software package within the *OptiRamp* Advanced Dispatch Control System (ADCS) suite of tools designed to provide Statistics & Control, Inc., (S&C) customers with an internally based Computational Pipeline Monitoring (CPM) system. The LDM incorporates state-of-the-art statistical and alarming algorithms and adheres to best practices recommended by the American Petroleum Institute (API).

S&C engineered *OptiRamp* LDM to align with API RP 1130. In particular, as a CPM system, LDM uses a hybrid method of leak detection, utilizing both volume/mass balances and statistical algorithms to improve accuracy. It also ensures reliability and robustness through redundant and independent detection approaches. LDM integrates with all existing systems and process signals through its connectivity with *OptiRamp* ADCS.

The ultimate purpose of LDM is to provide plant and pipeline operators with an accurate, reliable, and robust estimate of leak occurance and magnitude. It is therefore designed to adapt to various customer requirements and transportation system (pipeline) configurations. Additionally, LDM allows the end user to manually adjust detection parameters and thresholds so that all federal and local regulations are satisfied.

OptiRamp LDM utilizes a blend of dynamic simulation, material balance models, predictive modeling, and statistical analysis to ascertain the probability of leaks in a transportation system—such as a pipeline—in a timely fashion. The submodule then provides audio and visual alarming to the operator based on parametrically defined thresholds.

During initial set up as well as throughout deployment, *OptiRamp* LDM uses online and off-line testing to evaluate leak detection accuracy, thereby ensuring continuous reliable and robust support of corporate and regulatory guidelines. Operators can also initiate testing on demand.

API Alignment

According to API RP 1130, any CPM system must possess the following four characteristics: Reliability, Sensitivity, Accuracy, and Robustness. Reliability is the system's ability to function within a predetermined operating envelope. Sensitivity is the relationship between leak magnitude and time required for the system to detect the leak. Accuracy is the system's ability to minimize type I (false positives) and type II (false negatives) errors. Robustness is the system's ability to function outside of its normal operating envelope, e.g., during transient conditions or in the presence of bad data.

OptiRamp LDM integration with *OptiRamp* ADCS ensures system reliability through accurate commodity release alarming with dynamic thresholds, the ability to trend pressure values at each measuring device using proprietary archiving algorithms, timely commodity release detection, and database query times of under five seconds using Firebird database software.

LDM allows the user to select system sensitivity thresholds and uses the *OptiRamp* Modeling Submodule and dynamic simulation to build sensitivity curves for every threshold setting. Figure 1 shows an example of sensitivity curves given two operating thresholds, S_1 and S_2.

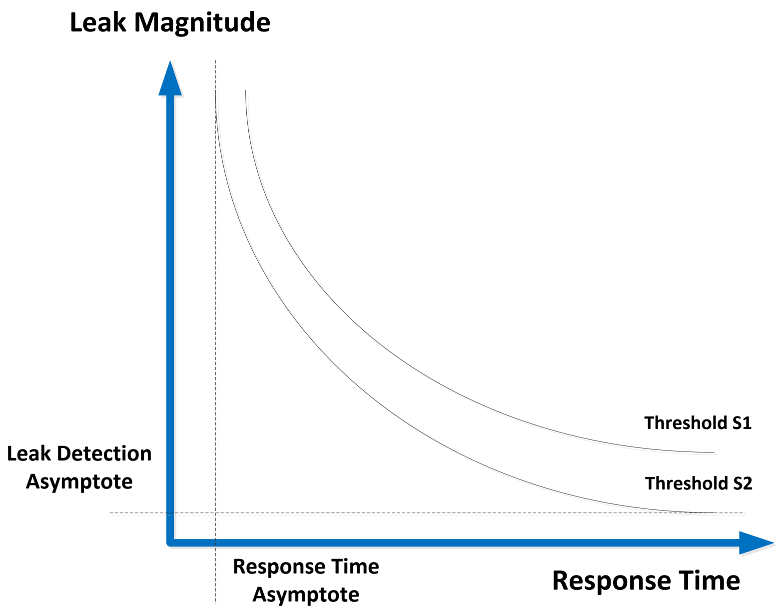

Figure 1. *OptiRamp* LDM sensitivity curves

The thresholds determine asymptote locations for both leak magnitude and detection response time. These locations indicate that there is a minimum detectable leak magnitude as well as minimum attainable response times to identify the leak. S&C engineers will work with the pipeline operators to determine these values during commissioning.

To satisfy process accuracy requirements, LDM treats leak detection as a binary classification problem where there can only be four unique outcomes, which are summarized in Figure 2.

	Leak Exists	Leak Does Not Exist
Leak Detected Alarm	True Positive	False Positive Type **I** Error
Leak not Detected	False Negative Type **II** Error	True Negative

Figure 2. *OptiRamp* LDM accuracy

Accuracy is defined according to equation (1).

$$Accuracy = 1 - (Type\ I\ Error\ Rate + Type\ II\ Error\ Rate) \quad (1)$$

To ensure the highest possible accuracy, *OptiRamp* LDM is equipped with a leak probability analyzer that is driven by statistical methods, such as logistic regressions and neural networks. Furthermore, LDM is able to identify the transportation segment where the leak occurred as long as measuring devices exist at both ends of the segment.

OptiRamp LDM is a robust CPM system equipped with data simulation and filtering algorithms that allow it to function during communication outages and data failures. LDM is able to provide alarming (albeit with a degraded measure of accuracy) during abnormal operating conditions and transient processes. Accuracy degradation is estimated as a function of signal loss. The system allows operators to manually override signal values when signals are unavailable.

The biggest advantage of using *OptiRamp* LDM is the built-in redundancy that creates an independent leak detection confirmation via a suite of algorithms that are constantly retrained on new data and on the occurrence of actual leaks. For example, logistic regressions and artificial neural network models are automatically retrained using a moving window approach that minimizes the misclassification rate as well as maximizes the Kolmogorov–Smirnov (KS) statistic values.

OptiRamp LDM is equipped with self-learning and self-tuning models that require minimal engineering involvement to maintain. The submodule includes a comprehensive suite of testing tools that are used during system installation as well as throughout utilization to ensure continuous pipeline monitoring. The system testing processes are described in detail later in this paper.

LDM provides alarming mechanisms that correspond to API best practices as both audio and visual signals that are directly available on the S&C-designed operator HMI system. Additionally, customized reporting provides a history of any alarms and alarm dispositions and a complete audit trail of changes/actions taken by pipeline operators.

Leak Detection Algorithm

The *OptiRamp* LDM general leak detection algorithm and the statistical methods used in detection are described in this section. The algorithm consists of the following steps:

1. Construct material balance equations for every segment of the transportation network.
2. Extract the history of past leaks and/or simulate leaks to generate historical events to be used as targets for predictive modeling.
3. Transform input variables to normalize continuous variables, eliminate outliers, and to create binary dummies for discrete variables.
4. Create process data train/validation/test samples.
5. Build predictive models using logistic regressions and artificial neural networks with a binary target (leak occurred or not) to estimate the propensity for leaks to occur based on all of the transformed process variables in the train sample. Models will have a sound fit when the validation misclassification rate is as small as possible or the KS statistic is as high as possible.
6. Use test sample performance (either minimal misclassification rate or maximal KS statistic) to select the best model.
7. Use model output (leak probability at every given time scan) to generate a prioritized list of potential leaks to verify with an external CPM system in order to tune accuracy and determine the probability cutoff for acceptable type I and type II errors.
8. Retrain the model in online/off-line modes based on parametrically defined time frames to sustain a continuous champion/challenger approach, where the current (champion) model is checked at every time scan against a voting system of challenger models to ensure sustained model lift and to prevent model decay.
9. Provide output to the transportation network operator HMI as a set of audio and visual alarms along with the classification and severity probability of each alarm (e.g., data failure, irregular operating conditions, and probable leak).
10. Retrain the models again based on operator confirmation/rejection of detected leaks.

The basic principle of material balances states that *Input* = *Output* + *Accumulation*. As is described in the *OptiRamp* Material Balance Reconciliation Submodule white paper, the material balance with time delay is given by equation (2).

$$\sum_i X_i(t) = \sum_j Y_j(t+\tau_j) + \sum_p Z_p(t), \tag{2}$$

where X_i are a set of inflows, Y_j are the lagged outflows with τ_j being corresponding delays, and Z_p are the accumulated quantities.

It is always best to start predictive modeling with a set of known targets. Thus, if a record of past leaks exists in the DCS database, this dataset is imported into the LDM data archive. However, such history may not exist for certain networks (e.g., new construction). Having the material balance equations allows LDM to simulate leaks by creating unbalanced process operating modes, which is achieved by introducing random change factors $\delta_{X_i}(t)$, $\delta_{Y_j}(t)$, and $\delta_{Z_p}(t)$ for inflows, outflows, and accumulated quantities, respectively. LDM then calculates expected values for all other process variables specific to the same time scans when the change was introduced. Thus, the LDM simulation submodule generates a dataset that can be fed into the predictive modeling submodule.

OptiRamp LDM utilizes all modeling techniques available in the *OptiRamp* Modeling Submodule. Two primary modeling techniques that are especially useful for binary response prediction are logistic regressions and artificial neural networks. LDM uses maximum likelihood to estimate logistic regression parameters β, as shown in equation (3).

$$ln\left(\frac{p}{1-p}\right) = P(\beta, X), \tag{3}$$

where *P(β,X)* is an m^{th}-degree polynomial with *k* independent variables *x* and *p = E(Y\X)* is the probability of leak *Y* occurring given the values of independent variable vector *X*. Note, *Y* is either 0 or 1.

LDM uses the multilayer perceptron (MLP) implementation of artificial neural networks for highly nonlinear processes. The MLP network consists of individual neurons. Each artificial neuron receives *n* weighted inputs that are summarized and transferred to the neuron output.

Figure 3 displays the structure of the individual neuron with three inputs. The composition of the combination and transfer function constitute the activation function. The MLP model uses the hyperbolic tangent as the activation function given by equation (4).

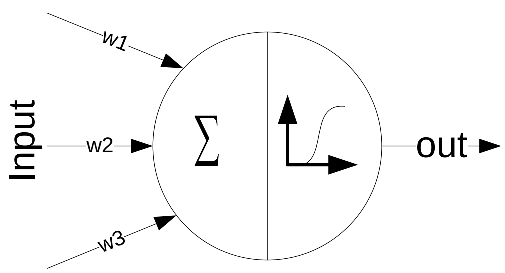

Figure 3. Artificial neuron with three inputs

$$h_j = tanh\left(b_j + \sum_i W_{ij} x_i\right), \tag{4}$$

where *b* and *w* are the estimates/weights and *j* is the number of hidden units in the network. Note that if the activation function is the identity, then the artificial neural network simply becomes the logistic regression.

The MLP weights are initially chosen at random and then adjusted so that the error function shown in equation (5) is minimized on the training set:

$$E = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - F(x_i)\right), \tag{5}$$

where N is the sample size, y_i is the leak indicator variable (1 or 0), and $F(x_i)$ is the MLP output. The only downside to using artificial neural networks over logistic regressions is the possibility of model overfit. Hence, it is imperative to use the test samples to validate model results.

LDM evaluates predictive model goodness-of-fit using two primary methods: misclassification rate and KS statistic. The misclassification rate is given by equation (6).

$$Misclassification\ Rate = Type\ I\ Error\ Rate + Type\ II\ Error\ Rate \tag{6}$$

This model fit statistic can only be calculated when a probability cutoff is established, i.e., when the model is used as a classifier. For ranked list output, the appropriate goodness-of-fit statistic is associated with the Kolmogorov–Smirnov test of the difference between two cumulative distributions. The KS statistic is calculated according to equation (7):

$$KS = max_Q\left(\frac{\sum_i Y_i}{N} - \frac{\sum_i 1-Y_i}{M}\right), \tag{7}$$

where the *max* is taken over all possible probability partitions Q, Y_i is a binary leak indicator, N is the number of observations with a leak, and M is the number of observations without a leak.

The challenger models are created by modifying parameters used in logistic regressions or neural networks. In logistic regressions, nonlinear factors and transformations are applied to the input variables along with different feature selection algorithms (forward, backward, and stepwise). With artificial neural networks, two parameters provide the necessary set of challenger models: number of hidden layers and number of neurons.

Alarming

OptiRamp LDM provides three types of alarms: data failure, irregular operating conditions, and probable leak. The data failure alarms typically correspond to communication outages (i.e., when a specific measuring device stops sending signal values) or to incorrect signals (i.e., when there is a systematic error with the measuring device). Gross error detection algorithms provided in the *OptiRamp* Material Balance Reconciliation Submodule are applied to determine one of four general data failure situations: bias, drifting, precision degradation, and complete failure. Figure 4 provides visual examples of such data failures. Thus, LDM will generate four types of data failure alerts. Additionally, the system will communicate any corrective action taken along with the alarm to ensure CPM system robustness.

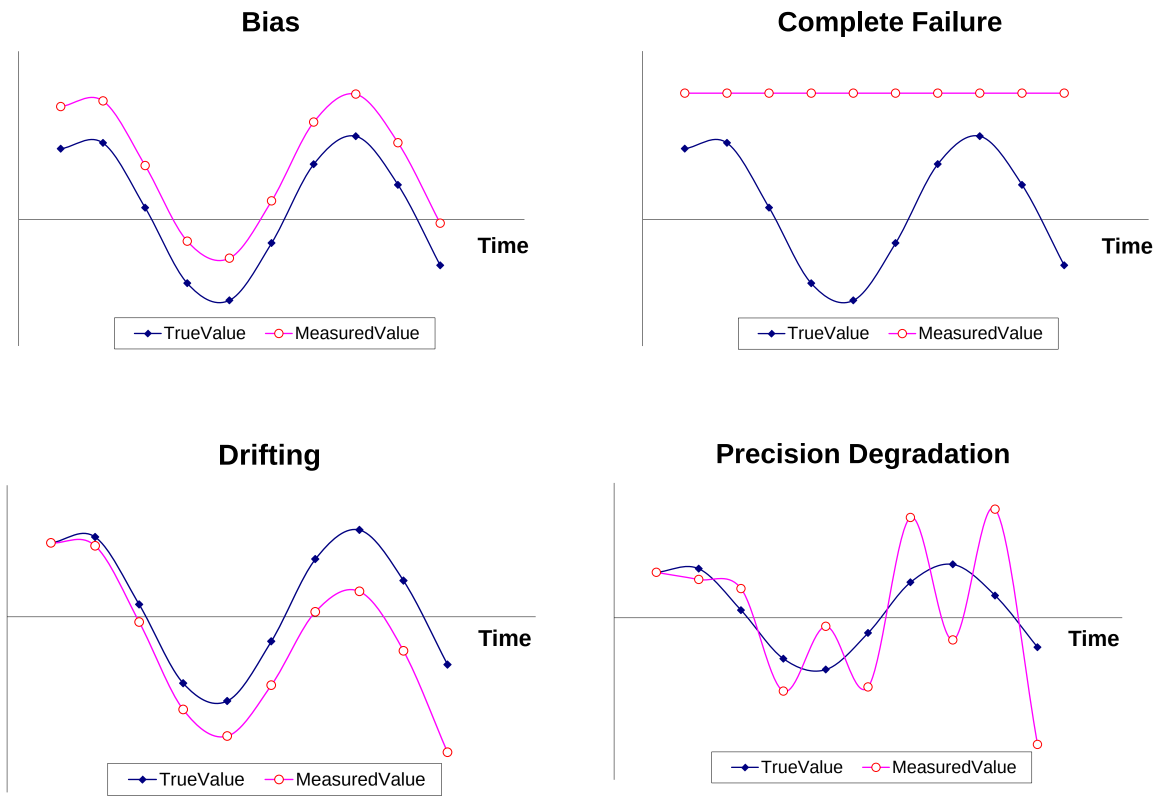

Figure 4. Data failure types

Irregular operating condition alarms are issued whenever a signal has exceeded its normal operating conditions but the LDM is not certain whether a data failure or a leak has occurred. This irregular condition occurs whenever hypothesis testing using the Generalized Likelihood Ratio (GLR), Tjao-Biegler, and neural network tests are inconclusive and whenever predicted leak occurrence probability has not reached a predetermined threshold but a six sigma test shows that a certain data element is outside the normal range. Figure 5 provides a conceptual example where an irregular operating condition alarm may be generated.

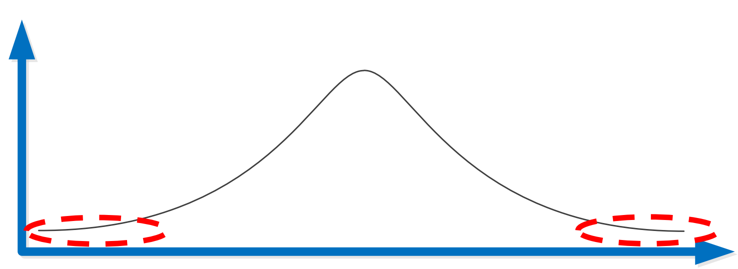

Figure 5. Irregular operating conditions alarm

Depending on system settings, *OptiRamp* LDM can either issue a probable leak alarm whenever the predicted probability surpasses a predetermined cutoff or it can generate an entire list of probable leaks that need to be cross-referenced against an external CPM system output. Some of the possible external systems could be hydrocarbon sensors, emissions detectors, or vapor sensors. The alarm will also contain metadata regarding the segment location where a leak has been predicted to occur.

OptiRamp LDM is supplied with ergonomic displays that make the visual alarms stand out on the screen. It is recommended that the HMI PC has good speakers so that operators can hear the audible alarm.

System Testing

S&C engineers will come on-site to test the *OptiRamp* LDM according to the rigors recommended by API RP 1130. All information about the tests will be stored in the LDM database, including the date; test time and duration; technical reasons for the test, method, location, and description of commodity withdrawn (real or simulated); operating conditions at the time of the test; and details of any alarms generated during the test.

The advantage of using *OptiRamp* LDM is that it has built-in simulated, cost-effective test algorithms. LDM is ready for tests as soon as the *OptiRamp* Material Balance Reconciliation Submodule has built the material balance equations. In fact, the entire process of simulated tests is very similar to the process of simulating leaks for the purpose of building the predictive models described earlier. The idea is to create an unbalanced operating mode and measure the time it takes for LDM to respond as well as check the predicted probability values.

This type of testing determines the system sensitivity curves discussed in the API Alignment section given a set of threshold values. It also enables the engineers to calculate Type I and Type II errors and, therefore, measure system accuracy. Additionally, *OptiRamp* LDM utilizes *OptiRamp* Simulation to determine the input/output transfer functions and to simulate transient processes, thereby ensuring system robustness.

S&C recommends retesting the CPM system whenever any of the following occurs: major transportation network or software configuration changes or installation of additional features, abnormal pipeline operating conditions, new versions of CPM software, instrument and measurement device additions or changes, or data infrastructure updates (e.g., DCS or SCADA updates).

About Statistics & Control, Inc.

S&C—an engineering consulting and technology company headquartered in West Des Moines, IA—solves complex challenges for customers through its unique technology and its highly seasoned team of professionals. The company has a global portfolio spanning the energy, oil and gas, utility, and digital oil field industry sectors. S&C provides clients with turbomachinery control solutions that easily integrate with the existing system as well as *OptiRamp*® solutions, which focus on process and power analytics to optimize processes and, in turn, reduce costs and increase reliability. S&C also provides consulting, dynamic system studies, modeling, automation, training and OTS, and support services.

Statistics & Control, Inc.
4401 Westown Pkwy, Suite 124
West Des Moines, IA 50266 USA
Phone: 1.515.267.8700
Fax: 1.515.267.8701