White Paper

OptiRamp® Continuous Emission Modeling System

Advanced Analytics for a Safe Environment

Vadim Shapiro
Dmitriy Khots, Ph.D.

Table of Contents

Introduction

OptiRamp® Continuous Emission Modeling System (CEMoS) is a suite of tools designed by S&C to work in conjunction with Continuous Emission Monitoring Systems (CEMS) and in particular with extractive CEMS to ensure regulatory emission requirements outlined by the United States Environmental Protection Agency's (EPA) Code of Federal Regulations (CFR) Title 40 Protection of Environment, parts 60 and 75. *OptiRamp* CEMoS integrates with any legacy extractive CEMS as long as the legacy system is fully compliant across its three major components: sample transport and conditioning, sample analysis, and data acquisition and storage.

OptiRamp CEMoS identifies mathematical dependencies between various process variables and the turbine/boiler exhaust gas composition produced by the extractive CEMS or by laboratory tests. These dependencies are known as the CEMoS models, which are generated utilizing *OptiRamp* CEMoS integration with the *OptiRamp* Modeling Submodule. The *OptiRamp* Real-Time Optimization (RTO) Submodule then uses the CEMoS models to predict how the exhaust gas composition will respond to changes in the operating mode. Ultimately, the RTO Submodule finds the optimal operating mode based on the minimum exhaust criteria.

The advantages of using *OptiRamp* CEMoS are that it ensures complete regulatory compliance, environmental safety, and plant economic feasibility, as well as provides an additional layer of redundancy to existing extractive CEMS, 99.9% uptime, and integration with the plant's HMI for comprehensive alarming and reporting. *OptiRamp* CEMoS is also equipped with a self-testing submodule that builds a separate set of models to allow for automatic system recalibration in case accuracy drifts are detected. S&C further recommends that the *OptiRamp* CEMoS go through the third-party Relative Accuracy Test Audit (RATA) according to standard CEMS part 75 requirements, including the nine-run RATA on a semiannual basis.

Regulatory Considerations

The EPA is the federal body that governs the requirements for all CEMS. The CFR 40 (Protection of Environment) part 60 (Standards of Performance for New Stationary Sources) and part 75 (Continuous Emission Monitoring) are the two federal regulations that guide the design of all CEMS implementations. Part 60 predominantly deals with emission guidelines and performance standards across various industries and production processes. Part 75 specifically deals with continuous emission monitoring, which is the foundation for the U.S. “cap and trade” program.

These regulations apply to the processes that occur in the extractive CEMS, including continuous sampling of the exhaust gas from a processing point; sample filtering and conditioning (typically dependent on the nature of the gas); selecting gas analysis techniques, such as spectroscopic absorption, luminescence, electroanalysis, electrochemical analysis, and paramagnetism; and data storage and quality management (e.g., handling of missing data).

S&C’s design of the *OptiRamp* CEMoS aligns with these requirements by providing a reliable and accurate service, even when the extractive CEMS is off-line or unavailable (because emission values can now be predicted based on other process variables, i.e., these predicted values can be used in place of missing data). In particular, if a monitoring system is not installed and real-time data are not available for model construction, *OptiRamp* CEMoS can use manually entered lab test analysis data. Data can be obtained once a week, once a month, or even less frequently. CEMoS can still calibrate the model with this type of data.

Additionally, CEMoS provides redundant analyses to ensure accuracy as well as a substantial suite of tools for model testing. The algorithmic approach to accuracy and redundancy testing is described in detail below.

CEMoS Architecture

The core of *OptiRamp* CEMoS architecture consists of three components: emission measurement data produced and stored by the extractive CEMS (or manually entered laboratory test data), the model built by the *OptiRamp* Modeling Submodule that relates emission measurements to the process operating mode, and the optimal operating mode calculated by the *OptiRamp* RTO Submodule that minimizes the emissions while keeping the client’s operation economically feasible. The architecture is illustrated in Figure 1.

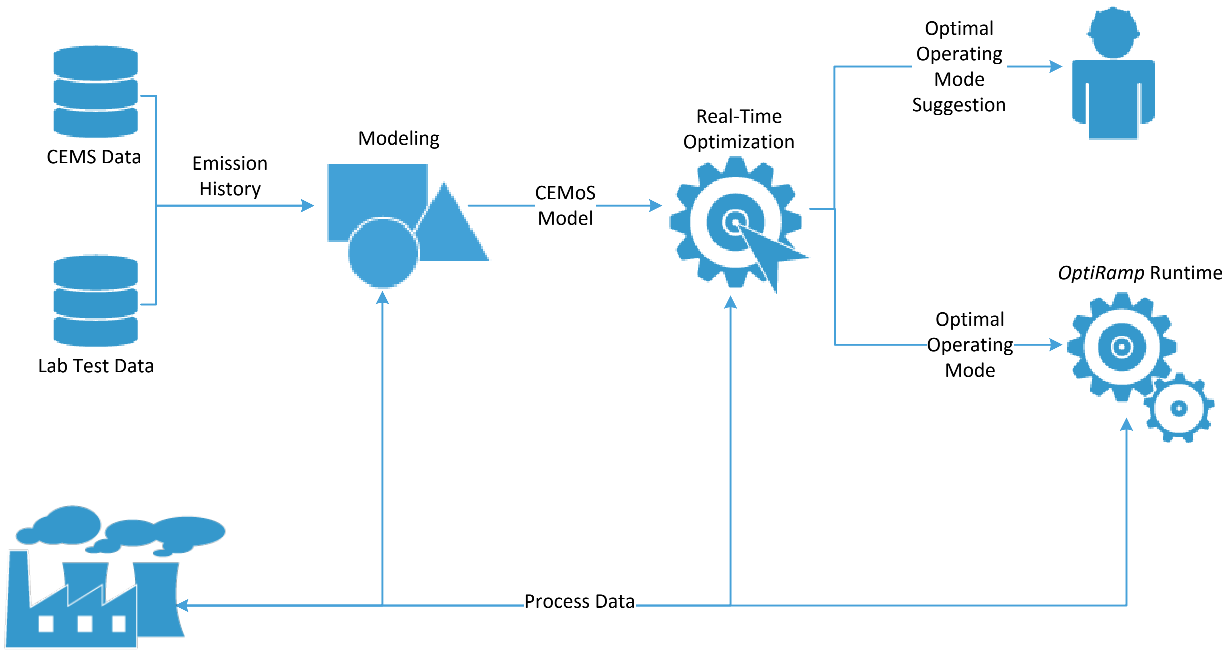

Figure 1. *OptiRamp* CEMoS architecture

Data

OptiRamp CEMoS uses three major data sources: extractive CEMS data (preferred), laboratory testing data (whenever extractive CEMS is off-line or does not exist), and process data. CEMoS applies signal processing techniques to filter and transform input data to prepare it for modeling. These techniques are described in detail in the *OptiRamp* Modeling Submodule white paper. For reference, certain examples that are particularly useful in a CEMS setting are provided.

In situations where the model has to use human-entered data (e.g., laboratory test data), data are often not clean or are missing. CEMoS is equipped to deal with such situations with the following tools: time series extrapolation, statistical imputation (using the mean or the median to replace missing values), and tree imputation methods (where auxiliary models are built to find relationships between missing data and other process variables). In extreme cases where the proportion of observations with missing values is too high, the variable is converted into a binary indicator according to equation (1), which presents additional predictive information to the model.

$$x_t = \begin{cases} 1 | \; x = missing \\ 0 | \; x \neq missing \end{cases}, \qquad (1)$$

where x is the original variable. CEMoS also automatically tests all data for unusual observations, or outliers, using the six sigma approach. Given m observations of an independent variable, $x(t_1), \dots, x(t_m)$, its standard deviation σ is calculated using equation (2).

$$\sigma = \sqrt{\frac{\sum_{j=1}^{m}\left(x(t_j)-\bar{x}\right)^2}{m}}, \tag{2}$$

where $\bar{x} = \frac{\sum_{j=1}^{m} x(t_j)}{m}$ is the sample mean. Then an outlier, xO, is defined according to equation (3).

$$x(t_O) \equiv xO \text{ iff } x(t_O) > \bar{x} + \delta \cdot \sigma \text{ or } x(t_O) < \bar{x} - \delta \cdot \sigma, \tag{3}$$

where $\delta > 0$ is a sufficiently large parameter.

If any outliers are found, CEMoS transforms the data using max normal transformations, such as the natural logarithm or power transformations, thereby removing the biases introduced by these outliers.

Model

OptiRamp CEMoS primarily uses two modeling techniques to predict continuous proportions of exhaust gas components of interest. These techniques are polynomial regression using ordinary least squares (OLS) fitting methodology and artificial neural networks utilizing multilayer perceptron (MLP) structure. The system trains the model in online mode using 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.

Given the target variable y (exhaust gas component proportion) and a set of n independent potential predictor variables $x_1, \dots, x_n$ (process variables) with m observations measured in a time interval T, the polynomial regression algorithm estimates coefficient vector $\boldsymbol{\beta}$ that minimizes equation (4).

$$S = \sum_{j=1}^{m}\left(y(t_j) - P\left(\boldsymbol{\beta}, x_1(t_j), \dots, x_n(t_j)\right)\right)^2 \tag{4}$$

This minimization is achieved by calculating partial derivatives of S with respect to each element $\beta_i \in \boldsymbol{\beta}$, with $i = 1, \dots, M$ given by equation (5).

$$\frac{\partial S}{\partial \beta_i} = 2\sum_{j=1}^{m}\left(y(t_j) - P\left(\boldsymbol{\beta}, x_1(t_j), \dots, x_n(t_j)\right)\right) \cdot \frac{\partial\left(y(t_j) - P\left(\beta, x_1(t_j), \dots, x_n(t_j)\right)\right)}{\partial \beta_i} \tag{5}$$

Then, partial derivatives are set to zero, resulting in a system of linear equations shown in equation (6) that is solved using Cramer's rule.

$$\begin{cases} \frac{\partial S}{\partial \beta_1} = 0 \\ \vdots \\ \frac{\partial S}{\partial \beta_M} = 0 \end{cases} \tag{6}$$

The resulting solution (if it exists) is the vector $\boldsymbol{\beta}$ that minimizes the distance from the fitted values of *y* to actual values of *y* and can be used for processing by the optimization submodule.

For artificial neural networks, the algorithm utilizes an MLP structure with the back propagation algorithm. The initial model input is a set of potential normalized predictor variables, $x_1, \dots, x_n$, along with their weights, $w_{1j}, \dots, w_{nj}$, chosen at random for each neuron. Each neuron processes the inputs to produce an output that is the weighted sum of the inputs provided by equation (7).

$$h_j = \sum_i w_{ij} x_i \qquad (7)$$

Each neuron's output is then activated using an activation function *F*, which could be a hyperbolic tangent or a logistic sigmoid function. The output is then shown by equation (8).

$$\mathrm{o_j} = \mathrm{F}\left(\sum_\mathrm{i} \mathrm{w_{ij}} \mathrm{x_i}\right) \qquad (8)$$

Next, each neuron's output becomes an input for the next neuron in the network, with corresponding weights assigned at random for the first run. Ultimately, the network's total output is the activated linear sum of output from all neurons from the last hidden layer, which can be expressed by equation (9).

$$\hat{\mathrm{y}} = \mathrm{F}\left(\sum_{\mathrm{i,j}} \mathrm{v_{ij}} \mathrm{g}(\mathrm{x_i})\right), \qquad (9)$$

where *j* runs across all units in the network, v_{ij} are the weights associated with the output from neurons in the last hidden layer, and functions *g* are a composite of all prior activation functions.

Figure 2 displays this process for an individual neuron with three inputs.

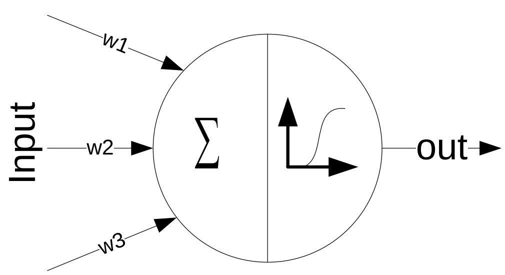

Figure 2. Artificial neuron with three inputs

The next step of the algorithm evaluates the cost function shown in equation (10) across all observations in the training dataset.

$$\mathrm{C} = \frac{1}{M} \sum_{p=1}^{M} \left(y(t_p) - \hat{y}(t_p) \right)^2 \qquad (10)$$

The algorithm then adjusts all of the weights in the network so that *C* is minimized, which is accomplished with the gradient descent algorithm. The algorithm output is the set of weights, $\boldsymbol{W}$, across the entire network that are used by the optimization submodule.

OptiRamp CEMoS varies model parameters to ensure the most accurate output. The best model is selected using the generalized R^2 goodness-of-fit criterion tested on multiple cross-validation samples. Specifics of this selection are described in the System Testing section.

Optimization

OptiRamp CEMoS is fully integrated with the *OptiRamp* RTO Submodule and thus has access to all mathematical optimization routines available in the RTO Submodule. For operating mode optimization problems, CEMoS allows the plant operator to select decision variables (e.g., unit start and stop times, product flow, etc.), constraints (e.g., fuel gas, operating times, product flow, exhaust gas composition proportions, etc.), and the objective function (e.g., fuel gas, exhaust gas composition proportions, etc.). Note that depending on the optimization problem, a single attribute could be used either as a decision variable, a constraint, or an objective function.

The most common optimization problems involving emissions involve finding the optimal operating mode that either minimizes exhaust gas component proportions (e.g., SO_2, CO_2, etc.) subject to plant output constraints or maximizes plant output subject to specific exhaust gas component values. Thus, the exhaust gas component proportion is the objective function in the former case and a constraint in the latter case. In either case, it is crucial to have a function that accurately describes the interdependence of the exhaust gas component proportion with other process variables. *OptiRamp* CEMoS solves this problem using the modeling techniques already described.

As an example, if y is set to be the proportion of component of interest and $\hat{y}$ is the modeled value, then the first-case optimization problem can be stated according to equations (11) through (13).

$$\mathbf{min}\ \boldsymbol{\hat{y} = f(x_1, \dots, x_n)} \tag{11}$$

subject to

$$c\_min_{x_i} \le x_i \le c\_max_{x_i}, \tag{12}$$

$$|O(x_1, \dots, x_n) - Out| \le \varepsilon, \tag{13}$$

where $c\_min_{x_i}$ is the lower bound for every x_i, $c\_max_{x_i}$ is the upper bound for every x_i, $O(\cdot)$ is the function that relates plant output to process variables, Out is the required output value, and $\varepsilon > 0$ is a sufficiently small parameter. The second-case optimization problem is then stated according to equations (14) through (16).

$$\mathbf{max}\ \boldsymbol{\hat{O} = O(x_1, \dots, x_n)} \tag{14}$$

subject to

$$c\_min_{x_i} \le x_i \le c\_max_{x_i}, \tag{15}$$

$$|f(x_1, \dots, x_n) - Exhaust| \leq \delta, \quad (16)$$

where $Exhaust$ is the required exhaust gas component proportion value and $\delta > 0$ is a sufficiently small parameter.

CEMoS chooses the appropriate optimization solution method depending on the nature of the objective and constraint functions as well as decision variables. For example, in load allocation problems where decision variables are binary (unit shutdown/start-up), genetic algorithms will be used to find the optimal combination. The *OptiRamp* CEMoS optimization submodule recalculates the optimal operating mode at every time scan to accommodate for environment, process, and business changes, thereby ensuring the optimal solution for current conditions.

Alarming

The *OptiRamp* CEMoS provides the following alarms: data failure and emission outside of normal. 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* Leak Detection Submodule are applied to determine one of four general data failure situations: bias, drifting, precision degradation, or complete failure. Readers are encouraged to review The Leak Detection Submodule white paper for detailed examples of such data failures. Thus, CEMoS will generate four types of alerts based on the data failure type. Additionally, the system will communicate any corrective action taken along with the alarm.

CEMoS applies six sigma process control techniques to generate alarms whenever abnormal emission values are detected. Figure 3 shows that an alarm will be generated whenever the proportion of the exhaust gas component of interest is in the zones indicated by red dashes.

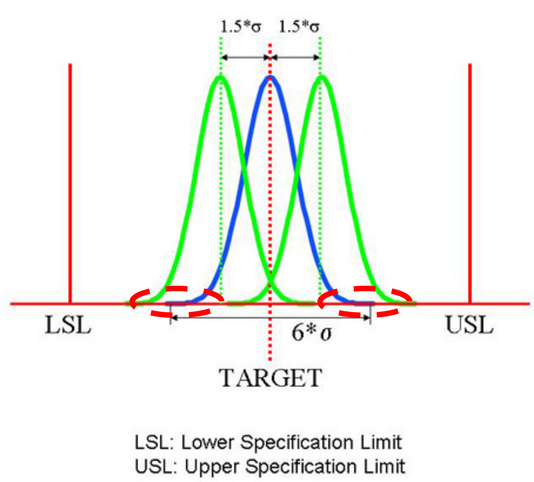

Figure 3. Abnormal value detection

One of the major benefits of the CEMoS models is the fact that not only are the measured values being tracked for normal behavior but also the predicted exhaust gas component proportions are tracked. Therefore, S&C is able to predict whether emissions are outside the norm and, more importantly, when it will happen. Thus, plant operators can take proactive corrective actions to ensure compliance before any undesirable outcomes occur.

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

System Testing

OptiRamp CEMoS is equipped with self-tuning and redundancy algorithms that ensure the most accurate emission predictions. A brief synopsis of these proprietary algorithms is provided.

Let y be the exhaust gas component being predicted by CEMoS, $x_1, \dots, x_n$ be the set of n uncorrelated and possibly transformed process variables that were deemed as predictive by the CEMoS model, and let $R^2_{champion}$ be the goodness-of-fit statistic associated with the current model in production (champion model), f. In the most general form, $R^2_{champion}$ is defined by equation (17).

$$R^2 = 1 - \left(\frac{L(0)}{L(\hat{\theta})}\right)^{2/m}, \qquad (17)$$

where $L(0)$ is the likelihood of an intercept-only model (i.e., where no x_i's are significant), $L(\hat{\theta})$ is the likelihood that at least one x_i is a significant predictor, and m is the sample size. For example, in the case of ordinary least squares regression, $R^2_{champion}$ is given by equation (18).

$$R^2_{champion} = R^2_{adj} = 1 - \frac{\sum_j (y_j - f_j)^2}{\sum_j (y_j - \bar{y})^2} \frac{m-1}{m-n-1} \qquad (18)$$

At every time increment, CEMoS will generate a set of challenger models by varying the following parameters:

1. The number of attributes used in the model, n
2. The sampling time frame, T
3. The number of cross-validation samples (bootstrapping), B
4. The degree of the regression, p
5. The number of interaction terms in the regression, Q
6. The number of hidden layers in the artificial neural network, l
7. The activation function in the artificial neural network, L
8. The number of neurons in each layer, lN

Each challenger model goodness-of-fit is evaluated against the champion on cross-validation samples. The champion model remains whenever $\left|R^2_{challenger} - R^2_{champion}\right| \leq \varepsilon$, where $\varepsilon > 0$ is a sufficiently small parameter. The challenger model succeeds the champion whenever $\left(R^2_{challenger} - R^2_{champion}\right) > \varepsilon$. This technique ensures accuracy of the champion model as well as provides redundant solutions.

S&C also recommends that the existing CEMS and CEMoS be subject to third-party RATA testing at least on a semiannual basis.

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