

Systems and Methods for Monitoring A Process Environment

Field of the Invention

[1] The invention relates to the field of process control systems, including industrial or other process control systems. In particular the invention provides methods and systems for monitoring process environments, predicting process outputs, and diagnosing abnormalities in process outputs, for the purposes of process environment control.

Background

[2] Process control systems seek to implement comprehensive and cost effective strategies to improve process efficiencies and to control product output quality – since losses arising from product quality abnormalities can be significant. In absence of continuous monitoring of process output quality, process deviations that are undetected during a manufacturing process may only be detected during product quality control testing. By the time process deviations are detected in this manner, significant manufacturing and manpower resources have already been expended and overheads incurred - which expenditures are irrecoverable. This problem is significant in every case where product output quality monitoring is implemented in a non-real time basis.

[3] Existing fault monitoring and diagnostics solutions for process environments are based on either first principles modelling or on data driven modelling approaches.

[4] First principles modelling involves describing relevant phenomena in terms of first principle representations, to predict an output state – and typically involves consideration of mass balance, energy balance, and reaction equations. The process environment is initially set up and test runs are based on theoretical calculations. Over time however, plant conditions undergo complex variations – and identifying and representing these variations using first principles modelling has been found to be complex, time consuming, and difficult to implement.

[5] Data driven modelling techniques on the other hand analyse the state of process environment parameters, and predict output states based on data corresponding to process

environment parameters. Data driven modelling has the advantage of reducing or eliminating reliance on detailed physical understanding of the underlying process, and on knowledge of material properties, geometry, and other characteristics of the process environment and its components. The underlying process model is created by developing a generic linear or non-linear process model, and fitting measured or simulated process environment data or plant data to the generic model through a learning process. The learning process involves the identification and embedding of relationships between the process variables into the model, which is optionally improved through an iterative process of minimizing an error function.

[6] Recent developments in soft sensor technology and in information technology systems have facilitated widespread and consistent collection and storage of data points corresponding to variable parameters within process environments. Various data driven approaches including principal component analysis (PCA), non-linear principal component analysis (NLPCA), dimension reduction approach (DRA) analysis, kernel principal component analysis (KPCA), support vector machine (SVM) analysis, partial least square (PLS) analysis, principal component regression (PCR) analysis, discriminant analysis, canonical variates analysis and artificial neural networks have been previously used to develop data driven modelling techniques. These prior approaches have however been found to present various drawbacks, including a lack of predictive ability, failure to capture non-linearity (i.e. failure to effectively isolate parameters contributing to an abnormal output state), ineffectiveness in accommodating load changes (i.e. increases / decreases in output demand), and exponential increases in computational resource requirements corresponding to the number of samples and choice of functions involved.

[7] There is accordingly a need for developing and implementing an effective data driven modelling approach that utilizes existing data repositories and sensor technologies for monitoring process environments, predicting process output states, and diagnosing abnormalities in process output states, for the purposes of process environment control

Summary

[8] The present invention provides methods and systems for monitoring process environments, predicting process outputs, and diagnosing abnormalities in process outputs, for the purposes of process environment control.

[9] The invention presents a method for monitoring a process output state of a process environment. The method comprises (i) receiving from one or more field devices, a first set of data identifying states of a set of n process environment parameters, (ii) identifying a set of m functions that satisfy two or more predefined conditions, (iii) generate a data model for determining a process output state of the process environment, wherein the data model is generated based on (a) the n process environment parameters, (b) the identified set of m functions, and (c) one or more coefficients, and wherein the generated data model discretely models at least (d) the effect of every k -combination of the n variable parameters, where $k = 1$, and (e) the effect of every k -combination of the n variable parameters, for at least one value of k such that $1 < k \leq n$, (iv) determining values for coefficients within the generated data model, wherein the determination of values for coefficients within the generated data model is based on (g) data representing recorded states of the n process environment parameters, and (h) data representing recorded process output states of the process environment, wherein the recorded process output states correspond to the recorded states of the n process environment parameters, (v) determining a process output state of the process environment based on a second set of data identifying respective states of the set of n process environment parameters, wherein the second set of data is input to the generated data model, and (vi) generating a notification identifying the determined process output state of the process environment.

[10] For every value of k such that $1 < k \leq n$, the generated data model may discretely model the effect of every k -combination of the n variable parameters.

[11] In an embodiment of the invention, the two or more predefined conditions satisfied by the set of m functions may comprise:

(i)

$$\int_a^b \varphi_f(x) dx = 0$$

...Condition (1)

where $1 \leq f \leq m$;

(ii)

$$\int_a^b \varphi_f^2(x) dx = 1$$

...Condition (2)

where $1 \leq f \leq m$; and

$$\int_a^b \varphi_f(x) \varphi_l(x) dx = 0$$

...Condition (3)

where (i) ($m \geq 2$), (ii) ($1 \leq f \leq m$), (iii) ($1 \leq l \leq m$), and (iv) ($f \neq l$).

[12] Identifying the set of m functions may include selecting the m functions from a set of defined functions, wherein selection of the m functions relies on measured or normalized upper and lower bounds of data representing states of one or more of the n process environment parameters.

[13] In an embodiment of the invention, a and b respectively may comprise measured or normalized lower and upper bounds of data representing states of one or more of the n process environment parameters.

[14] The method may additionally comprise optimizing the data model, wherein optimizing the data model comprises the steps of (i) determining a deviation between (a) the process output state that has been determined based on the second set of data as input to the generated data model, and (b) a measured process output state of the process environment, which process output states correspond to the respective states of the set of n process environment parameters identified by the second set of data, (ii) responsive to the deviation exceeding a predefined deviation limit (c) identifying a set of $m1$ functions that satisfy the two or more predefined conditions, wherein $m1 > m$, and (d) regenerating the data model based on (I) the n process environment parameters, (II) the

identified set of m functions, and (III) one or more coefficients, and (iii) determining values for coefficients within the regenerated data model.

[15] In an embodiment of the method, $m \leq 3$. In a specific embodiment, $m \leq 3$.

[16] The method may further comprise at least one of (i) ordering a plurality of k-combinations of the n process environment parameters, based on the respective contribution of each of said k-combinations to the process output state that is determined by the generated data model based on states of the n process environment parameters, and (ii) ordering a plurality of the n process environment parameters, based on the respective contribution of each of said plurality of the n process environment parameters to the process output state determined by the generated data model based on states of the n process environment parameters.

[17] The method may additionally include at least one of (i) selecting from among the ordered plurality of k-combinations or from among the ordered plurality of n process environment parameters, a sub-set of the ordered k-combinations or a sub-set of the n process environment parameters, wherein the selection is based on relative contributions of each k-combination or each process environment parameter, to the determined process output state, and (ii) including information identifying the selected sub-set of k-combinations or the selected sub-set of the n process environment parameters, in the generated notification identifying the determined process output state.

[18] The invention additionally provide a system for monitoring a process output state of a process environment.

[19] The system comprises a processor configured to (i) receive from one or more field devices, a first set of data identifying states of a set of n process environment parameters, and (ii) identify a set of m functions that satisfy two or more predefined conditions.

[20] The system includes a data model generator configured to (i) generate a data model for determining a process output state of the process environment, wherein the data model is generated based on (a) the n process environment parameters, (b) the identified set of m functions, and (c) one or more coefficients.

[21] The generated data model may discretely model at least (i) the effect of every k-combination of the n variable parameters, where $k = 1$, and (ii) the effect of every k-combination of the n variable parameters, for at least one value of k such that $1 < k \leq n$.

[22] The generated data model may determine values for coefficients within the generated data model, wherein the determination of values for coefficients within the generated data model is based on (i) data representing recorded states of the n process environment parameters, and (ii) data representing recorded process output states of the process environment – wherein the recorded process output states correspond to the recorded states of the n process environment parameters.

[23] The system may additionally include a process monitor configured to (i) determine a process output state of the process environment based on a second set of data identifying respective states of the set of n process environment parameters, wherein the second set of data is input to the generated data model, and (ii) generate a notification identifying the determined process output state of the process environment.

[24] In a system embodiment, for every value of k such that $1 < k \leq n$, the generated data model discretely models the effect of every k-combination of the n variable parameters.

[25] In a specific embodiment of the system, the two or more predefined conditions satisfied by the set of m functions may comprise:

(i)

$$\int_a^b \varphi_f(x) dx = 0$$

...Condition (1)

where $1 \leq f \leq m$;

(ii)

$$\int_a^b \varphi_f^2(x) dx = 1$$

...Condition (2)

where $1 \leq f \leq m$; and

$$\int_a^b \varphi_f(x) \varphi_l(x) dx = 0$$

...Condition (3)

where (i) ($m \geq 2$), (ii) ($1 \leq f \leq m$), (iii) ($1 \leq l \leq m$), and (iv) ($f \neq l$).

[26] In an embodiment, identifying the set of m functions may include selecting the m functions from a set of defined functions, wherein selection of the m functions relies on measured or normalized upper and lower bounds of data representing states of one or more of the n process environment parameters.

[27] In a specific system embodiment, a and b may respectively comprise measured or normalized lower and upper bounds of data representing states of one or more of the n process environment parameters.

[28] The data model generator may be configured to optimize the data model. Optimizing the data model may comprise the steps of (i) determining a deviation between (a) the process output state that has been determined based on the second set of data as input to the generated data model, and (b) a measured process output state of the process environment, which process output states correspond to the respective states of the set of n process environment parameters identified by the second set of data, (ii) responsive to the deviation exceeding a predefined deviation limit (c) identifying a set of $m1$ functions that satisfy the two or more predefined conditions, wherein $m1 > m$, and (d) regenerating the data model based on (I) the n process environment parameters, (II) the identified set of $m1$ functions, and (III) one or more coefficients, and (iii) determining values for coefficients within the regenerated data model.

[29] In a system embodiment, $m \leq 3$. In another embodiment, $m1 \leq 3$.

[30] The system may include a ranking generator configured for at least one of (i) ordering a plurality of k-combinations of the n process environment parameters, based on the respective contribution of each of said k-combinations to the process output state that is determined by the generated data model based on states of the n process environment parameters, (ii) ordering a plurality of the n process environment parameters, based on the respective contribution of each of said plurality of the n process environment parameters to the process output state determined by the generated data model based on states of the n process environment parameters.

[31] In an embodiment of the invention, the process monitor may be further configured to (i) select from among the ordered plurality of k-combinations or from among the ordered plurality of n process environment parameters, a sub-set of the ordered k-combinations or a sub-set of the n process environment parameters, wherein the selection is based on relative contributions of each k-combination or each process environment parameter, to the determined process output state, and (ii) include information identifying the selected sub-set of k-combinations or the selected sub-set of the n process environment parameters, in the generated notification identifying the determined process output state.

[32] The invention additionally provides a computer program product for monitoring a process output state of a process environment, comprising a non-transitory computer readable storage medium having program code stored therein, the program code comprising computer readable instructions for implementing one or more methods steps in accordance with the present invention.

Brief Description of the Accompanying Drawings

[33] Figures 1 and 2 illustrate method steps for generating and implementing a data model for monitoring a process environment.

[34] Figure 3 illustrates a method of calculating values for one or more coefficients corresponding to a data model in accordance with the present invention.

[35] Figure 4 illustrates a method for refining or improving a generated data model.

[36] Figure 5 illustrates a method for using a generated data model for ranking k-combinations of variable parameters based on their relevance to a process output state.

[37] Figure 6 illustrates an exemplary process environment configured to implement the methods of the present invention.

[38] Figure 7 illustrates an exemplary computing system in which various embodiments of the invention may be implemented.

Detailed Description

[39] Control system engineering refers to a domain of engineering that deals with architectures, mechanisms and algorithms for maintaining output of a specific system or process within a desired range. Automated control systems are used extensively in industry and for example can be achieved using programmable logic controllers (PLC(s)), or in the case of more complex systems using distributed control systems (DCS) or supervisory control and data acquisition systems (SCADA).

[40] Automated control systems rely on one or more controllers communicatively coupled to one or more field devices. Field devices (e.g. sensors, valves, switches, receivers and transmitters) are located within the process environment corresponding to the control system (for example, but not limited to, an industrial plant or system) and may be configured to perform physical or process control functions to control one or more components, processes or variables under observation within the process environment. Process controllers may be located within the process environment and are configured to receive signals from field devices, make control decisions, generate control signals and communicate with field devices.

[41] Existing control systems rely on data received from field devices within the process environment and on archival of such data. Data generated by controllers or field devices within a process environment are stored in a plurality of specifically designated databases (e.g. data historians or data silos).

[42] The invention relies on data received in real time from field devices and / or data from databases within a control system for monitoring process environments, predicting process output

states, and diagnosing abnormalities in process output states, for the purposes of process environment control. For the purposes of the invention, it would be that the term “process output state” may describe any output state of a process environment. In certain embodiments of the invention, a process output state may comprise a parameter value that represents one or more quality parameters related to an output or result from a process environment.

[43] Figure 1 illustrates a method for generating and implementing a data model for monitoring a process environment, in accordance with the present invention.

[44] For the purposes of the method illustrated in Figure 1, it is assumed that the process environment sought to be modelled and monitored comprises a set of n variable parameters. The n variable parameters may either comprise the entire set of variable parameters within the process environment, or alternatively may comprise a specific sub-set of variable parameters that desired to be monitored. In an embodiment of the invention, one or more of the n variable parameters may comprise machine condition parameters.

[45] Step 102 comprises selecting a set of m functions that satisfy two or more predefined conditions. The value of m may in an embodiment be an operator defined value. In an embodiment of the invention, the selected set of m functions satisfies at least the following predefined conditions:

Condition 1 :

$$\int_a^b \varphi_f(x) dx = 0$$

Condition 1 requires to be satisfied in every case where $1 \leq f \leq m$.

Condition 2 :

$$\int_a^b \varphi_f^2(x) dx = 1$$

Condition 2 requires to be satisfied in every case where $1 \leq f \leq m$.

It would be understood that in Condition 2, the superscript “2” associated with function $\varphi_f^2(x)dx$ in the above represents the exponent “2” i.e. a square of the function φ_f .

Condition 3 :

$$\int_a^b \varphi_f(x)\varphi_l(x)dx = 0$$

Condition 3 requires to be satisfied in every case where (i) ($m \geq 2$); (ii) ($1 \leq f \leq m$); (iii) ($1 \leq l \leq m$); and (iv) ($f \neq l$).

[46] In an embodiment of step 102, $m \leq 3$ i.e. the set of functions selected by the method comprises 3 or less functions.

[47] In an embodiment of the invention, the set of m functions may be selected from a larger set of defined functions, for example a library of functions retrieved from a repository.

[48] In addition to the requirement for satisfying one or more of the above conditions, selection of the set of m functions may additionally rely on identified values for a and b , which serve as the limits of integration for Conditions 1 to 3 above. In an embodiment of method step 102, values a and b correspond to upper and lower bounds for data values observed in respect of the n variable parameters. In a more specific embodiment, values a and b correspond to upper and lower bounds for normalized data values derived from raw (measured) data values observed in respect of the n variable parameters.

[49] In an illustrative example, assuming that the method derives normalized data values using the following Equation (1):

$$\text{Normalized Value} = \frac{(\text{raw value} - \text{minimum raw value})}{(\text{maximum raw value} - \text{minimum raw value})}$$

...Equation (1)

the corresponding upper bound a would be 1, while the lower bound b would be 0.

[50] Depending on the values identified for limits of integration a and b , method step 102 may accordingly identify or select m functions that satisfy two or more of Conditions 1 to 3 above.

[51] In an embodiment of the invention, the selected m functions are a set of orthogonal basis functions.

[52] Step 104 comprises generating a data model for calculating a process output state (corresponding to an output from the process environment comprising the n variable parameters) - wherein the data model is expressed in terms of one or more expressions that collectively model the effect of a plurality of (in a preferred embodiment, every) distinct k -combinations of the n variable parameters, where $1 \leq k \leq n$. In an embodiment of the invention the expressions used in the data model are expressed in terms of at least one or more of (i) an expression that models the effect of every distinct k -combination of the n variable parameters, where $k = 1$, (ii) an expression that models the effect of every distinct k -combination of the n variable parameters, for any one or more values of k , where $1 < k \leq n$ and (iii) an expression that models the effect of a k -combination of the n variable parameters, where $k = n$.

[53] For the purposes of the present invention, it would be understood that a k -combination of a set S is a subset of k distinct elements of S . In the present invention, assuming a set S comprises the n variable parameters, a k -combination of the set S would comprise a subset of k distinct variable parameters selected from among the n variable parameters. By way of example, assuming a process environment comprising n variable parameters, where the n variable parameters comprise a set of 3 variable parameters p_1, p_2 and p_3 :

- k -combinations where $k = 1$, would comprise (i) $\{p_1\}$ (ii) $\{p_2\}$ and (iii) $\{p_3\}$
- k -combinations where $k = 2$, would comprise (i) $\{p_1, p_2\}$ (ii) $\{p_1, p_3\}$ and (iii) $\{p_2, p_3\}$
- k -combinations where $k = 3$, would comprise (i) $\{p_1, p_2, p_3\}$

- The seven distinct sub-sets $\{p_1\}$, $\{p_2\}$, $\{p_3\}$, $\{p_1, p_2\}$, $\{p_1, p_3\}$, $\{p_2, p_3\}$ and $\{p_1, p_2, p_3\}$ collectively represent every distinct k -combination within the set of n (where $n = 3$) variable parameters, where $1 \leq k \leq n$.

[54] Likewise, assuming a process environment comprising n variable parameters, where the n variable parameters comprises a set of 4 variable parameters p_1 , p_2 , p_3 and p_4 :

- k -combinations where $k = 1$, would comprise (i) $\{p_1\}$ (ii) $\{p_2\}$ (iii) $\{p_3\}$ and (iv) $\{p_4\}$
- k -combinations where $k = 2$, would comprise (i) $\{p_1, p_2\}$ (ii) $\{p_1, p_3\}$ (iii) $\{p_1, p_4\}$ (iv) $\{p_2, p_3\}$ (v) $\{p_2, p_4\}$ and (vi) $\{p_3, p_4\}$
- k -combinations where $k = 3$, would comprise (i) $\{p_1, p_2, p_3\}$ (ii) $\{p_1, p_2, p_4\}$ (iii) $\{p_1, p_3, p_4\}$ and (iv) $\{p_2, p_3, p_4\}$
- k -combinations where $k = 4$, would comprise (i) $\{p_1, p_2, p_3, p_4\}$
- The fifteen distinct sub-sets $\{p_1\}$, $\{p_2\}$, $\{p_3\}$, $\{p_4\}$, $\{p_1, p_2\}$, $\{p_1, p_3\}$, $\{p_1, p_4\}$, $\{p_2, p_3\}$, $\{p_2, p_4\}$, $\{p_3, p_4\}$, $\{p_1, p_2, p_3\}$, $\{p_1, p_2, p_4\}$, $\{p_1, p_3, p_4\}$, $\{p_2, p_3, p_4\}$ and $\{p_1, p_2, p_3, p_4\}$ collectively represent every distinct k -combination within the set of n (where $n = 4$) variable parameters, where $1 \leq k \leq n$.

[55] Step 104 results in a data model expressed in terms of one or more expressions that collectively model the effect of every possible distinct k -combination of the n variable parameters within the process environment, where $1 \leq k \leq n$. The one or more expressions used to express the data model may comprise one or more of the n variable parameters, one or more of the m functions identified at step 102, and one or more coefficients. In an embodiment of the invention, for a set of input data values corresponding to the n variable parameters, the data model generated at step 104 outputs a data value representing a process output state of the process environment. Specific embodiments of step 104 (and data models resulting from step 104) are discussed in more detail subsequently in connection with Figure 2.

[56] Step 106 comprises calculating values for the one or more coefficients corresponding to expressions within the generated data model. Values of coefficients may be calculated based on (i) previously stored data values for the n variable parameters, and (ii) recorded data values for process

output states corresponding to the stored data values for the n variable parameters. Specific embodiments of step 106 are discussed in more detail subsequently in connection with Figure 3.

[57] Step 108 comprises refining or improving the data model. In an embodiment, the data model may be refined or improved by:

- measuring deviations between (i) process output states generated using previously stored data values for the n variable parameters and (ii) recorded data values for process output states corresponding to said stored data values for the n variable parameters
- responsive to the measured deviation exceeding a defined threshold, regenerating the data model by increasing the number of functions that are selected at step 102 (and that satisfy the predefined set of conditions) and using the increased number of selected functions for generating the data model at step 104.

[58] In certain embodiments of the invention, step 108 is an optional step. Specific embodiments of step 108 are discussed in more detail subsequently in connection with Figure 4.

[59] Step 110 comprises using the generated data model for ranking (i) one or more of the n variable parameters based on their significance or relevance to process output states of the process environment and / or (ii) one or more of the k -combinations of the n variable parameters based on the relevance or significance of the effect of said k -combination of the variable parameters to process output states of the process environment. Step 110 may in an embodiment additionally comprise determining relative contributions of each of the n variable parameters in respect of process output states of the process environment.

[60] By way of example, in a data model representing the effect of 4 variable parameters {p1, p2, p3 and p4} on process output states, step 110 may result in a ranking of the 4 variable parameters as p4, p2, p1 and p3 – in descending order of their significance or relevance to process output states of the process environment. In another example, in a data model representing the effect of 3 variable parameters {p1, p2 and p3}, step 110 may result in ranking the seven possible k -combinations of the variable parameters (i.e. {p2}, {p3}, {p1, p2}, {p1}, {p1, p2, p3}, {p2,p3}, and {p1, p3}), in

descending order of their significance or relevance to process output states of the process environment. In yet another example, in a data model representing the effect of 3 variable parameters {p1, p2 and p3}, step 110 may result in weighting and ranking the seven possible k-combination of the variable parameters as {p2 - 40%}, {p3 - 15%}, {p1,p2 - 15%}, {p1 - 10%}, {p1, p2, p3 - 10%}, {p2,p3 - 5%}, and {p1, p3 - 5%}, in descending order of the calculated weightages representing their significance or relevance to process output states of the process environment.

[61] In certain embodiments of the invention, step 108 is an optional step. Specific embodiments of step 110 are discussed in more detail subsequently in connection with Figure 5.

[62] Step 112 comprises receiving measured data values corresponding to the n variable parameters (for example from field devices or sensors within the process control system, or from one or more databases or data historians) and using the received data values in the generated data model to determine a process output state corresponding to the received data values of the n variable parameters under observation. In addition to determining a process output state, information derived from the data model may also be used to determine the respective contribution of individual parameters and / or one or more combinations of the individual parameters to the process output state. The information determined at step 112 may be used in a variety of different ways including for display or presentation to an operator of a process control system, or for the purposes of raising notifications, alarms or alerts in situations where a determined process output state exceeds an acceptable deviation from a desired process output state. In another embodiment, where the process environment does not measure process output states in real time, the data model may be used to calculate a process output state based on detected data values corresponding to the n variable parameters, and may display or notify an operator of the calculated process output state in real time, and optionally raise alarms or alerts in cases where the calculated process output state is outside an acceptable set of values. Specific embodiments of step 112 are discussed in more detail subsequently.

[63] Figure 2 illustrates an embodiment of the step of generating a data model (more generally described in step 104 of Figure 1). Step 202 comprises determining a first expression that models the effect of every k-combination of the n variable parameters, where $k = 1$. Step 204 comprises, for

every value of k such that $1 < k < n$, determining a corresponding k^{th} expression, which k^{th} expression models the effect of every k -combination of the n variable parameters. Step 206 comprises determining an n^{th} expression that models the effect of the k -combination of the n variable parameters, where $k = n$. Each of the n expressions determined at steps 202 to 206 are defined in terms of (i) one or more of the n variable parameters, (ii) one or more of the m functions identified at step 102 (of Figure 1) and (iii) one or more coefficients. Step 208 optionally comprises specifying at least one constant. Finally, step 210 comprises generating a data model based on the determined n expressions and optionally the at least one constant.

[64] While method steps 202 to 206 collectively discuss determining expressions that respectively model the effect of every k -combination of the n variable parameters, for every value of k such that $1 \leq k \leq n$, it would be understood that in another embodiment, the expressions used in the data model may include any one or more of (i) an expression that models the effect of every distinct k -combination of the n variable parameters, where $k = 1$, (ii) an expression that models the effect of every distinct k -combination of the n variable parameters, for any one or more values of k , where $1 < k \leq n$ and (iii) an expression that models the effect of a k -combination of the n variable parameters, where $k = n$.

[65] In an embodiment of Figure 2, the data model is generated based on a selected set of m functions that have been found to satisfy predefined Conditions 1 to 3 discussed above in connection with step 102. The data model generated in this embodiment includes:

- for every value of k such $1 \leq k < n$, a corresponding expression modelling the effect of every k -combination of the n variable parameters;
- an expression modelling the effect of a k -combination of the n variable parameters, where $k = n$; and
- at least one constant

[66] In an embodiment, the data model described in accordance with the teachings of Figure 2 may be expressed as:

$$f(x) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{i,j}(x_i, x_j) \dots \dots + f_{1,2 \dots n}(x_1, x_2, \dots x_n)$$

... Equation (2)

wherein

- i) f_0 is a constant
- ii) $\sum_{i=1}^n f_i(x_i)$ is an expression modelling the effect of every k-combination of the n variable parameters, where k =1;
- iii) $\sum_{1 \leq i < j \leq n} f_{i,j}(x_i, x_j)$ is an expression modelling the effect of every k-combination of the n variable parameters, where k =2;
- iv) $f_{1,2 \dots n}(x_1, x_2, \dots x_n)$ is an expression modeling the effect of the combination of all n variable parameters taken together.

[67] Although not specifically shown, it would be understood from the format of Equation (2) that the generated data model also includes expressions respectively modelling the effect of every k-combination of the n variable parameters, for each value of k where

$$2 < k \leq (n-1)$$

[68] It would accordingly be understood from the format of Equation (2) that the generated data model may include expressions respectively modelling the effect of every k-combination of the n variable parameters, for each value of k where

$$1 \leq k \leq n$$

[69] In an embodiment of the invention, every expression within Equation (2), other than constant f_0 , is expressed in terms of one or more of the n variable parameters, one or more of the m functions that have been selected at step 102, and one or more coefficients.

[70] In an embodiment of the invention $f_i(x_i)$ within the expression

$$\sum_{i=1}^n f_i(x_i)$$

in Equation (2) may be derived using the following Equation (3):

$$f_i(x_i) = \sum_{r=1}^m \alpha_{i,r} \varphi_r(x_i)$$

...Equation (3)

[71] Taking an illustrative example where the number of parameter variable $n = 3$, and the number of selected functions $m = 2$, and applying the above general expression for $f_i(x_i)$:

- $f_1(x_1) = \alpha_{1,1} \varphi_1(x_1) + \alpha_{1,2} \varphi_2(x_1)$
- $f_2(x_2) = \alpha_{2,1} \varphi_1(x_2) + \alpha_{2,2} \varphi_2(x_2)$
- $f_3(x_3) = \alpha_{3,1} \varphi_1(x_3) + \alpha_{3,2} \varphi_2(x_3)$

[72] In an embodiment of the invention $f_{ij}(x_i, x_j)$ within the expression

$$\sum_{1 \leq i < j \leq n} f_{i,j}(x_i, x_j)$$

in Equation (2) may be derived using the following relationship:

$$f_{i,j}(x_i, x_j) = \sum_{p=1}^{m'} \sum_{q=1}^{m''} \alpha_{i,j,p,q} \varphi_p(x_i) \varphi_q(x_j)$$

...Equation (4)

[73] It would be understood that m' and m'' in Equation (4) have the same value as m i.e. is equal to the number of functions selected at step 102.

[74] Taking an illustrative example where the number of parameter variables $n = 3$, and the number of selected functions $m = 2$, and applying the above general expression for $f_{ij}(x_i, x_j)$:

- $f_{1,2}(x_1, x_2) = \alpha_{1,2,1,2} \varphi_1(x_1) \varphi_2(x_2) + \alpha_{1,2,2,1} \varphi_2(x_1) \varphi_1(x_2)$
- $f_{1,3}(x_1, x_3) = \alpha_{1,3,1,2} \varphi_1(x_1) \varphi_2(x_3) + \alpha_{1,3,2,1} \varphi_2(x_1) \varphi_1(x_3)$
- $f_{2,3}(x_2, x_3) = \alpha_{2,3,1,2} \varphi_1(x_2) \varphi_2(x_3) + \alpha_{2,3,2,1} \varphi_2(x_2) \varphi_1(x_3)$

[75] In an embodiment of the invention $f_{ijh}(x_i, x_j, x_h)$ within the expression

$$\sum_{1 \leq i < j < h \leq n} f_{ijh}(x_i, x_j, x_h)$$

(which is included but not specifically shown in Equation (2)) may be derived using the following relationship:

$$f_{ijh}(x_i, x_j, x_h) = \sum_{p=1}^{m'} \sum_{q=1}^{m''} \sum_{r=1}^{m'''} \alpha_{i,j,h,p,q,r} \varphi_p(x_i) \varphi_q(x_j) \varphi_r(x_h)$$

...Equation (5)

[76] It would be understood that m' , m'' and m''' in Equation (4) have the same value as m i.e. is equal to the number of functions selected at step 102.

[77] Taking an illustrative example where the number of parameter variable $n = 3$, and the number of selected functions $m = 2$, and applying the above general expression for $f_{ijh}(x_i, x_j, x_h)$:

$$\begin{aligned}
f_{123}(x_1, x_2, x_3) &= \alpha_{1,2,3,1,1,1} \varphi_1(x_1) \varphi_1(x_2) \varphi_1(x_3) + \alpha_{1,2,3,1,2,1} \varphi_1(x_1) \varphi_2(x_2) \varphi_1(x_3) \\
&+ \alpha_{1,2,3,1,2,2} \varphi_1(x_1) \varphi_2(x_2) \varphi_2(x_3) + \alpha_{1,2,3,2,1,1} \varphi_2(x_1) \varphi_1(x_2) \varphi_1(x_3) \\
&+ \alpha_{1,2,3,2,2,1} \varphi_2(x_1) \varphi_2(x_2) \varphi_1(x_3) + \alpha_{1,2,3,2,1,2} \varphi_2(x_1) \varphi_1(x_2) \varphi_2(x_3) \\
&+ \alpha_{1,2,3,2,2,2} \varphi_2(x_1) \varphi_2(x_2) \varphi_2(x_3)
\end{aligned}$$

[78] It would be understood that similar determinations are possible for all expressions that respectively model the effect of every k-combination of the n variable parameters, for each value of k where $3 < k \leq n$.

[79] Figure 3 illustrates an embodiment of method step 106 of Figure 1 – comprising calculating values for the one or more coefficients corresponding to expressions within the generated data model. Step 302 comprises obtaining values for a set of recorded process output states, along with values for variable parameters (for the n variable parameters) corresponding to the recorded process output state values. Based on a sufficient number of process output state values and corresponding variable parameter values, step 304 determines specific values for each of the coefficients within the data model generated at step 104.

[80] Table 1 below illustrates an exemplary data table of a kind that may be retrieved from a plant historian or other database.

Table 1

<u>S No.</u>	<u>Time Stamp</u>	<u>Availability of Variable Parameter Data</u>	<u>Availability of Corresponding Output Quality Data</u>
1.	9.15 am	✓	✗
2.	9.16 am	✓	✓
3.	9.17 am	✓	✗
4.	9.18 am	✓	✓
5.	9.19 am	✓	✗

6.	9.20 am	✓	✓
7.	9.21 am	✓	✗
8.	9.22 am	✓	✓

[81] It will be observed from Table 1 that while variable parameter data corresponding to each of time stamps 9.15 am, 9.16 am, 9.17 am, 9.18 am, 9.19 am, 9.20 am, 9.21 am and 9.22 am are available, corresponding output quality data (i.e. a corresponding process output state) has only been recorded corresponding to time stamps 9.16 am, 9.18 am, 9.20 am and 9.22 am.

[82] Per step 304 of Figure 3, the recorded process output state values and corresponding variable parameter data values corresponding to time stamps 9.16 am, 9.18 am, 9.20 am and 9.22 am, may be used to determine specific values for each of the coefficients within the data model generated at step 104 of Figure 1.

[83] Figure 4 illustrates a specific embodiment of step 108 of Figure 1 – comprising refining or improving the data model.

[84] Step 402 comprises receiving a measured first process output state value and corresponding variable parameter values. Step 404 uses the received variable parameter values as inputs to the data model generated at step 104 (of Figure 1) to calculate a second process output state value. Step 406 comprises determining a deviation between the measured first process output state value and the calculated second process output state value. At step 408, based on a finding that a calculated deviation (between at least one measured process output state values and at least one corresponding calculated process output state value) exceeds a permissible deviation, the data model is regenerated, preferably by incrementing the value of m i.e. by increasing the number of selected functions that satisfy the two or more predefined conditions (see step 102 of Figure 1), and which selected functions are used to generate the data model. The permissible deviation may comprise any deviation limit, beyond which deviations between a measured process output value and a calculated process output value is unacceptable – and in an embodiment may comprise an operator defined deviation limit. It would be understood that coefficients for the regenerated data model may be determined in a manner similar to determination of coefficients for the initial data model i.e. values of coefficients may be calculated based on (i) previously stored data values for the n variable

parameters, and (ii) recorded data values for process output states corresponding to the stored data values for the n variable parameters.

[85] In an embodiment of the invention, subsequent to generation and / or refining of the data model, the resulting data model may be used to determine one or more process output states (e.g. values corresponding to one or more output quality parameter(s)) based on one or more instances of recorded variable parameter values, wherein the one or more instances of recorded variable parameter values do not have corresponding instances of recorded process output states or recorded values describing output quality parameters.

[86] Referring to the exemplary data table illustrated above at Table 1, it would be observed that while variable parameter data corresponding to each of time stamps 9.15 am, 9.16 am, 9.17 am, 9.18 am, 9.19 am, 9.20 am, 9.21 am and 9.22 am are available, corresponding output quality data (i.e. a corresponding process output state) is unavailable corresponding to time stamps 9.15 am, 9.17 am, 9.19 am and 9.21 am. Accordingly, subsequent to generation and / or refining of a relevant data model in accordance with the description in connection with one or more of Figures 1 to 4, the resulting data model may be used to determine one or more process output states corresponding to one or more of time stamps 9.15 am, 9.17 am, 9.19 am and 9.21 am – wherein the determined process output states are based on using the corresponding recorded variable parameter values as input values for the data model. Table 2 below illustrates an exemplary updated data table, wherein the data from Table 1 has been updated with calculated process output states corresponding to time stamps 9.15 am, 9.17 am, 9.19 am and 9.21 am.

Table 2

<u>S No.</u>	<u>Time Stamp</u>	<u>Availability of Variable Parameter Data</u>	<u>Availability of Corresponding Output Quality Data</u>
1.	9.15 am	✓	✓
2.	9.16 am	✓	✓
3.	9.17 am	✓	✓
4.	9.18 am	✓	✓

5.	9.19 am	✓	✓
6.	9.20 am	✓	✓
7.	9.21 am	✓	✓
8.	9.22 am	✓	✓

[87] Figure 5 illustrates a specific embodiment of step 110 of Figure 1 – comprising using the generated data model for ranking (i) one or more of the n variable parameters based on their significance or relevance to process output states of the process environment and / or (ii) one or more of the k-combinations of the n variable parameters based on the relevance or significance of the effect of said k-combinations of the variable parameters on process output states of the process environment.

[88] Step 502 of Figure 5 comprises calculating (based on measured variable parameter values, and coefficient values within the data model) a value for a plurality of (and in a specific embodiment, for each) k-combination of the n variable parameters within the data model – which value is derived based on the coefficient(s) and expressions corresponding to said k-combination, within the data model.

[89] Thereafter, step 504 comprises using the values calculated at step 502 to rank (i) one or more of the n variable parameters based on their respective significance or relevance to the value of a calculated process output state of the process environment and / or (ii) one or more of the k-combinations of the n variable parameters based on the respective relevance or significance of the effect of said k-combination of the variable parameters on a calculated process output state of the process environment. In an embodiment of the invention, the one or more variable parameters or one or more k-combinations may be ranked in ascending or descending order according to their respective contribution to the calculated process output state of the process environment.

[90] Step 506 comprises identifying a sub-set of the n variable parameters or the k-combination of the n variable parameters, based on their rankings. In an embodiment, step 506 comprises identifying a plurality of the highest ranked variable parameters or k-combinations. For example, step 506 may identify 3 variable parameters that have the highest rank or highest weightage in terms

of significance or relevance to the value of a calculated process output state. In another embodiment, step 506 may comprise identifying a plurality of the variable parameters or k-combinations that individually or collectively influence a process output state by at least a predefined value. For example, step 506 may identify a plurality of variable parameters or a plurality of k-combinations that individually or collectively influence a calculated process output state by 90% or more.

[91] Step 508 thereafter comprises initiating a system monitoring action based on the plurality of variable parameters or k-combinations of variable parameters identified at step 506. In an embodiment of the invention, step 508 comprises displaying or raising a system report, alert or alarm identifying (i) a process output state and (ii) the plurality of variable parameters or k-combinations of variable parameters that have been identified at step 506 - as being responsible for or particularly relevant to a process output state being reported. In an embodiment of the invention, responsive to detection of detected abnormality or fault condition in a process output state, state 508 comprises generating a system notification or alert identifying (ii) the plurality of variable parameters or k-combinations of variable parameters that have been identified at step 506 - as being responsible for or particularly relevant to the detected abnormality or fault condition.

[92] Figure 6 illustrates an exemplary process environment 600 configured to implement the methods of the present invention. Controller 602 is connected to an output state database (OSDB) 606a that is configured to store data values corresponding to measured output states (which in certain embodiments may include values corresponding to output quality parameters) and to a variable parameter database (VPDB) 606b configured to store data values corresponding to measured variable parameter values. In an embodiment of the invention, one or both of the OSDB and VPDB may consist of a plant historian. In another embodiment of the invention, one or both of the OSDB and VPDB may comprise a real time data feed from field devices within process environment 600. Controller 602 is also communicably coupled to one or more field devices 604 (which field devices may include sensors and actuators for effecting control of the process environment). Controller 602 is communicably coupled to a processor implemented data model generator 608 – which in an embodiment of the invention, is configured to generate a data model in accordance with one or more of the methods discussed above in connection with of Figures 1 and 2. The process environment 600 additionally includes a processor implemented process monitor 610

that is communicably coupled to controller 602 and data model generator 608 for monitoring the process environment, and generating, displaying and prioritising notifications, alerts or alarms in accordance with one or more of the method steps discussed above in connection with Figure 5. Process environment 600 may additionally include a processor implemented ranking generator 612 communicably coupled to one or more of data model generator 608, process monitor 610, databases 606a and 606b and controller 602 – wherein ranking generator 612 may be configured to rank (i) one or more of the n variable parameters based on their significance or relevance to process output states of the process environment and / or (ii) one or more of the k -combinations of the n variable parameters based on the relevance or significance of the effect of said k -combination of the variable parameters to process output states of the process environment (in accordance with step 504 of Figure 5).

[93] Ranking generator 612 may additionally be configured to identify a plurality of the highest ranked variable parameters or k -combinations. By way of illustrative example, we may take the case of a data model representing the effect of 3 variable parameters {p1, p2 and p3}, wherein step 110 of Figure 1 has resulted in weighting and ranking the seven possible k -combination of the variable parameters as {p2 - 40%}, {p3 - 15%}, {p1,p2 - 15%}, {p1 - 10%}, {p1, p2, p3 - 10%}, {p2,p3 - 5%}, and {p1, p3 - 5%}, in descending order of the calculated weightages representing their significance or relevance to process output states of the process environment. Assuming that ranking generator 612 is configured to identify 3 k -combinations that have the highest rank in terms of significance or relevance to the value of a calculated process output state, said ranking generator 612 would isolate the k -combinations {p2 - 40%}, {p3 - 15%} and {p1,p2 - 15%} as combinations of variable parameters that most significantly influence the calculated process output state.

[94] In another embodiment, ranking generator 612 may be configured to identify a plurality of variable parameters or k -combinations that individually or collectively influence a process output state by at least a predefined value. By way of illustrative example, we may take the case of a data model representing the effect of 3 variable parameters {p1, p2 and p3}, wherein step 110 of Figure 1 has resulted in weighting and ranking the seven possible k -combination of the variable parameters as {p2 - 40%}, {p3 - 15%}, {p1,p2 - 15%}, {p1 - 10%}, {p1, p2, p3 - 10%}, {p2,p3 - 5%}, and {p1, p3 - 5%}, in descending order of the calculated weightages representing their significance or relevance to process output states of the process environment. Assuming that ranking generator

612 is configured to identify k-combinations that individually or collectively are responsible for influencing a process output state by at least 90%, said ranking generator 612 would isolate the k-combinations {p2 - 40%}, {p3 - 15%}, {p1,p2 - 15%}, {p1 - 10%} and {p1, p2, p3 - 10%} as combinations of variable parameters that most significantly influence the calculated process output state.

[95] While not shown in Figure 6, it would be understood that process environment 600 may additionally include one or more additional processor implemented entities, including without limitation, filters for signal processing, a controller for data normalization, and a database of predefined functions from which the *m* functions may be selected. In various embodiments of the invention, process environment may comprise an industrial plant, a machine, or components within an industrial plant or machine.

[96] Figure 7 illustrates an exemplary computing system in which various embodiments of the invention may be implemented.

[97] The system 702 comprises at least one processor 704 and at least one memory 706. The processor 704 executes program instructions and may be a real processor. The processor 704 may also be a virtual processor. The computer system 702 is not intended to suggest any limitation as to scope of use or functionality of described embodiments. For example, the computer system 702 may include, but not be limited to, one or more of a general-purpose computer, a programmed microprocessor, a micro-controller, an integrated circuit, and other devices or arrangements of devices that are capable of implementing the steps that constitute the method of the present invention. Exemplary embodiments of a system 702 in accordance with the present invention may include one or more servers, desktops, laptops, tablets, smart phones, mobile phones, mobile communication devices, tablets, phablets and personal digital assistants. In an embodiment of the present invention, the memory 706 may store software for implementing various embodiments of the present invention. The computer system 702 may have additional components. For example, the computer system 702 includes one or more communication channels 708, one or more input devices 710, one or more output devices 712, and storage 714. An interconnection mechanism (not shown) such as a bus, controller, or network, interconnects the components of the computer system 702. In various embodiments of the present invention, operating system software (not shown) provides an

operating environment for various softwares executing in the computer system 702 using processor 704, and manages different functionalities of the components of computer system 702.

[98] The communication channel(s) 708 allow communication over a communication medium to various other computing entities. The communication medium provides information such as program instructions, or other data in a communication media. The communication media includes, but not limited to, wired or wireless methodologies implemented with an electrical, optical, RF, infrared, acoustic, microwave, Bluetooth or other transmission media.

[99] The input device(s) 710 may include, but is not limited to, a touch screen, a keyboard, mouse, pen, joystick, trackball, a voice device, a scanning device, or any another device that is capable of providing input to the computer system 702. In an embodiment of the present invention, the input device(s) 710 may be a sound card or similar device that accepts audio input in analog or digital form. The output device(s) 712 may include, but not limited to, a user interface on CRT, LCD, LED display, or any other display associated with any of servers, desktops, laptops, tablets, smart phones, mobile phones, mobile communication devices, phablets, personal digital assistants, printer, speaker, CD/DVD writer, or any other device that provides output from the computer system 702.

[100] The storage 714 may include, but is not limited to, magnetic disks, magnetic tapes, CD-ROMs, CD-RWs, DVDs, any types of computer memory, magnetic stripes, smart cards, printed barcodes or any other transitory or non-transitory medium which can be used to store information and can be accessed by the computer system 702. In various embodiments of the present invention, the storage 714 contains program instructions for implementing the described embodiments.

[101] In an embodiment of the present invention, the computer system 702 is part of a distributed network.

[102] The present invention may be implemented in numerous ways including as a system, a method, or a computer program product such as a computer readable storage medium or a computer network wherein programming instructions are communicated from a remote location.

[103] The present invention may suitably be embodied as a computer program product for use with the computer system 702. The method described herein is typically implemented as a computer program product, comprising a set of program instructions which is executed by the computer system 702 or any other similar device. The set of program instructions may be a series of computer readable codes stored on a tangible medium, such as a computer readable storage medium, for example, diskette, CD-ROM, ROM, flash drives or hard disk, or transmittable to the computer system 702, via a modem or other interface device, over either a tangible medium, including but not limited to optical or analogue communications channel(s) 708. The implementation of the invention as a computer program product may be in an intangible form using wireless techniques, including but not limited to microwave, infrared, bluetooth or other transmission techniques. These instructions can be preloaded into a system or recorded on a storage medium such as a CD-ROM, or made available for downloading over a network such as the Internet or a mobile telephone network. The series of computer readable instructions may embody all or part of the functionality previously described herein.

[104] While the exemplary embodiments of the present invention are described and illustrated herein, it will be appreciated that they are merely illustrative. It will be understood by those skilled in the art that various modifications in form and detail may be made therein without departing from or offending the spirit and scope of the invention as defined by the appended claims.

Claims:

1. A method for monitoring a process output state of a process environment, the method comprising:

receiving from one or more field devices, a first set of data identifying states of a set of n process environment parameters;

identifying a set of m functions that satisfy two or more predefined conditions;

generating a data model for determining a process output state of the process environment, wherein the data model is generated based on:

the n process environment parameters;

the identified set of m functions; and

one or more coefficients;

and wherein the generated data model discretely models at least:

the effect of every k -combination of the n variable parameters, where $k = 1$; and

the effect of every k -combination of the n variable parameters, for at least one value of k such that $1 < k \leq n$;

determining values for coefficients within the generated data model, wherein the determination of values for coefficients within the generated data model is based on (i) data representing recorded states of the n process environment parameters, and (ii) data representing recorded process output states of the process environment, wherein the recorded process output states correspond to the recorded states of the n process environment parameters;

determining a process output state of the process environment based on a second set of data identifying respective states of the set of n process environment parameters, wherein the second set of data is input to the generated data model; and

generating a notification identifying the determined process output state of the process environment.

2. The method as claimed in claim 1, wherein for every value of k such that $1 < k \leq n$, the generated data model discretely models the effect of every k -combination of the n variable parameters.

3. The method as claimed in claim 1, wherein the two or more predefined conditions satisfied by the set of m functions comprise:

(iii)

$$\int_a^b \varphi_f(x) dx = 0$$

...Condition (1)

where $1 \leq f \leq m$;

(iv)

$$\int_a^b \varphi_f^2(x) dx = 1$$

...Condition (2)

where $1 \leq f \leq m$; and

$$\int_a^b \varphi_f(x) \varphi_l(x) dx = 0$$

...Condition (3)

where (i) ($m \geq 2$), (ii) ($1 \leq f \leq m$), (iii) ($1 \leq l \leq m$), and (iv) ($f \neq l$).

4. The method as claimed in claim 1, wherein identifying the set of m functions comprises selecting the m functions from a set of defined functions, wherein selection of the m functions relies on measured or normalized upper and lower bounds of data representing states of one or more of the n process environment parameters.

5. The method as claimed in claim 3, wherein a and b respectively comprise measured or normalized lower and upper bounds of data representing states of one or more of the n process environment parameters.

6. The method as claimed in claim 1, comprising optimizing the data model, wherein optimizing the data model comprises the steps of:

determining a deviation between (i) the process output state that has been determined based on the second set of data as input to the generated data model, and (ii) a measured process output state of the process environment, which process output states correspond to the respective states of the set of n process environment parameters identified by the second set of data;

responsive to the deviation exceeding a predefined deviation limit:

identifying a set of $m1$ functions that satisfy the two or more predefined conditions, wherein $m1 > m$; and

regenerating the data model based on:

the n process environment parameters;

the identified set of $m1$ functions; and

one or more coefficients;

and

determining values for coefficients within the regenerated data model.

7. The method as claimed in claim 1 wherein $m \leq 3$.

8. The method as claimed in claim 6, wherein $m1 \leq 3$.

9. The method as claimed in claim 1, further comprising at least one of:

ordering a plurality of k-combinations of the n process environment parameters, based on the respective contribution of each of said k-combinations to the process output state that is determined by the generated data model based on states of the n process environment parameters; or

ordering a plurality of the n process environment parameters, based on the respective contribution of each of said plurality of the n process environment parameters to the process output state determined by the generated data model based on states of the n process environment parameters.

10. The method as claimed in claim 10, further comprising at least one of:

selecting from among the ordered plurality of k-combinations or from among the ordered plurality of n process environment parameters, a sub-set of the ordered k-combinations or a sub-set of the n process environment parameters, wherein the selection is based on relative contributions of each k-combination or each process environment parameter, to the determined process output state; and

including information identifying the selected sub-set of k-combinations or the selected sub-set of the n process environment parameters, in the generated notification identifying the determined process output state.

11. A system for monitoring a process output state of a process environment, the system comprising:

a processor configured to:

receive from one or more field devices, a first set of data identifying states of a set of n process environment parameters;

identify a set of m functions that satisfy two or more predefined conditions;

a data model generator configured to:

generate a data model for determining a process output state of the process environment, wherein the data model is generated based on

the n process environment parameters;

the identified set of m functions; and

one or more coefficients;

and wherein the generated data model discretely models at least:

the effect of every k -combination of the n variable parameters, where $k = 1$; and

the effect of every k -combination of the n variable parameters, for at least one value of k such that $1 < k \leq n$; and

determine values for coefficients within the generated data model, wherein the determination of values for coefficients within the generated data model is based on
(i) data representing recorded states of the n process environment parameters, and
(ii) data representing recorded process output states of the process environment, wherein the recorded process output states correspond to the recorded states of the n process environment parameters;

and

a process monitor configured to:

determine a process output state of the process environment based on a second set of data identifying respective states of the set of n process environment parameters, wherein the second set of data is input to the generated data model; and

generate a notification identifying the determined process output state of the process environment.

12. The system as claimed in claim 11, wherein for every value of k such that $1 < k \leq n$, the generated data model discretely models the effect of every k -combination of the n variable parameters.

13. The system as claimed in claim 11, wherein the two or more predefined conditions satisfied by the set of m functions comprise:

(iii)

$$\int_a^b \varphi_f(x) dx = 0$$

...Condition (1)

where $1 \leq f \leq m$;

(iv)

$$\int_a^b \varphi_f^2(x) dx = 1$$

...Condition (2)

where $1 \leq f \leq m$; and

$$\int_a^b \varphi_f(x) \varphi_l(x) dx = 0$$

...Condition (3)

where (i) ($m \geq 2$), (ii) ($1 \leq f \leq m$), (iii) ($1 \leq l \leq m$), and (iv) ($f \neq l$).

14. The system as claimed in claim 11, wherein identifying the set of m functions comprises selecting the m functions from a set of defined functions, wherein selection of the m functions relies on measured or normalized upper and lower bounds of data representing states of one or more of the n process environment parameters.

15. The system as claimed in claim 13, wherein a and b respectively comprise measured or normalized lower and upper bounds of data representing states of one or more of the n process environment parameters.

16. The system as claimed in claim 11, wherein the data model generator is configured to optimize the data model, and wherein optimizing the data model comprises the steps of:

determining a deviation between (i) the process output state that has been determined based on the second set of data as input to the generated data model, and (ii) a measured process output state of the process environment, which process output states correspond to the respective states of the set of n process environment parameters identified by the second set of data;

responsive to the deviation exceeding a predefined deviation limit:

identifying a set of $m1$ functions that satisfy the two or more predefined conditions, wherein $m1 > m$; and

regenerating the data model based on:

the n process environment parameters;

the identified set of $m1$ functions; and

one or more coefficients;

and

determining values for coefficients within the regenerated data model.

17. The system as claimed in claim 11 wherein $m \leq 3$.

18. The system as claimed in claim 16, wherein $m1 \leq 3$.

19. The system as claimed in claim 11, further comprising a ranking generator configured for at least one of:

ordering a plurality of k-combinations of the n process environment parameters, based on the respective contribution of each of said k-combinations to the process output state that is determined by the generated data model based on states of the n process environment parameters; or

ordering a plurality of the n process environment parameters, based on the respective contribution of each of said plurality of the n process environment parameters to the process output state determined by the generated data model based on states of the n process environment parameters.

20. The system as claimed in claim 19, wherein the process monitor is further configured to:

select from among the ordered plurality of k-combinations or from among the ordered plurality of n process environment parameters, a sub-set of the ordered k-combinations or a sub-set of the n process environment parameters, wherein the selection is based on relative contributions of each k-combination or each process environment parameter, to the determined process output state; and

include information identifying the selected sub-set of k -combinations or the selected sub-set of the n process environment parameters, in the generated notification identifying the determined process output state.

21. A computer program product for monitoring a process output state of a process environment, comprising a non-transitory computer readable storage medium having program code stored therein, the program code comprising computer readable instructions for:

receiving from one or more field devices, a first set of data identifying states of a set of n process environment parameters;

identifying a set of m functions that satisfy two or more predefined conditions;

generate a data model for determining a process output state of the process environment, wherein the data model is generated based on:

the n process environment parameters;

the identified set of m functions; and

one or more coefficients;

and wherein the generated data model discretely models at least:

the effect of every k -combination of the n variable parameters, where $k = 1$; and

the effect of every k -combination of the n variable parameters, for at least one value of k such that $1 < k \leq n$;

determining values for coefficients within the generated data model, wherein the determination of values for coefficients within the generated data model is based on (i) data representing recorded states of the n process environment parameters, and (ii) data representing recorded process output

states of the process environment, wherein the recorded process output states correspond to the recorded states of the n process environment parameters;

determining a process output state of the process environment based on a second set of data identifying respective states of the set of n process environment parameters, wherein the second set of data is input to the generated data model; and

generating a notification identifying the determined process output state of the process environment.

-x-