EventiC: A Real-Time Unbiased Event-Based Learning Technique for Complex Systems

Morad Danishvar, *Member, IEEE*, Alireza Mousavi, *Senior Member, IEEE*, and Peter Broomhead

***Abstract*—An improved method for the real time sensitivity analysis in large scale complex systems is proposed in this paper. The method borrows principles from the event tracking of interrelated causal events and deploys clustering methods to automatically measure the relevance and contribution made by each input event data (ED) on system outputs. The ethos of the proposed event modeling (EM) technique is that the behavior or the state of a system is a function of the knowledge acquired about events occurring in the system and its wider operational environment. As such it builds on the theoretical and the practical foundation for the engineering of knowledge and data in modern and complex systems. The proposed EM platform EventiC filters noncontributory ED sources and has the potential to include information that was initially thought irrelevant or simply not considered at the design stage. The real-time ability to group and rank relevant input–output ED in order of its importance and relevance will not only improve the data quality, but leads to an improved higher level of mathematical formulization in the modern complex systems. The contribution of the approach to systems' modeling is in the automation of data analysis, control, and plant process modeling. EventiC has been validated as the monitoring and the control system for a cement factory. In addition to the previously known parameters, the proposed EventiC identified new influential parameters that were previously unknown. It also filtered 18% of the input data without compromising the data quality or the integrity. The solution has improved the quality of input variable selection and simplify plant control strategies.**

***Index Terms*—Clustering, control, event modeling (EM), hardware-in-the-loop (HiL), input variable selection (IVS), real-time systems, sensitivity analysis (SA), sensors, and actuators.**

I. Introduction

THE CONVENTIONAL approach to the modeling and the design of physical systems typically relies on a known set of linear/nonlinear differential equations or analytical models that describe the physical behavior of the given system. Knowledge of the input variables (i.e., the excitation[1] parameters) and their impact on the model's performance (i.e., the output) requires domain knowledge, know-how, and effort by highly qualified experts. Clearly, any improvement in the process of input data selection and analysis will result in economic benefit to the stakeholders.

Observations made in the manufacturing, automotive, and aerospace industries reveal that the process of evaluating and the accuracy and validating the correctness of models in real time is an expensive and extremely time-consuming process. For example, the accuracy of hardware-in-the-loop (HiL) models relies on experts. Most practitioners depend on a process of trial and error and/or costly destructive and nondestructive testing to increase the accuracy of a solution. Fig. 1 shows one interpretation of this modeling process.

Within this context a method that automates and integrates the process of data acquisition and the analysis of raw data in near real-time is timely. In the proposed process, the acquisition of large scale data and its organization in the form of interrelationships and clusters of relevance takes place in the lower layer of the interface between the physical system and the higher level information framework. As such the proposed method could be considered as the linkage between the engineering of the physical systems and the higher level data modeling system. One major difference between the proposed event modeling (EM) and more traditional data modelling methods is that in traditional methods a state vector is expressed as a series of known input data representing more complex information (e.g., $V_N = [x_1, \ldots, x_n]$) about the output. Any subsequent operations on the vector are based on the assumption that the vector is a true representation of the known data series related to the output of the system. In contrast, the proposed EM technique, EventiC, makes no such assumption about the input data in terms of the association between system parameters. It treats data as an "unknown" collection of the information that needs to be organized prior to any formal representation of the information.

Typically, in EM, the process begins with the definition of an event vector $E_S = [e_1, \ldots, e_s]$, where event E is expressed by all observable events that occurred at a given instant in the sample space S. In the case of the EventiC algorithm that definition becomes $E_N = [e_1, \ldots, e_n]$, where e_1 to e_n, $N \leq S$ and N is the total number of events in the state space S, input events are true representations of E. By reverting back to the actual values of the variables the state vector, $V_n = [x_1, \ldots, x_n]$

[1]Excitation: The changes in parameters resulting into an event.

Manuscript received November 28, 2016; revised August 4, 2017; accepted November 10, 2017. This work was supported by the European Union's Horizon 2020 Research and Innovation Program under Grant 723906. This paper was recommended by Associate Editor A. H. Tan. *(Corresponding author: Morad Danishvar.)*

M. Danishvar and A. Mousavi are with the College of Engineering, Design and Physical Sciences, Department of Electronic and Computer Engineering, Brunel University London, London UB8 3PH, U.K. (e-mail: morad.danishvar@gmail.com; ali.mousavi@brunel.ac.uk).

P. Broomhead, deceased, was with the College of Engineering, Design and Physical Sciences, Department of Electronic and Computer Engineering, Brunel University London, London UB8 3PH, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2017.2775666

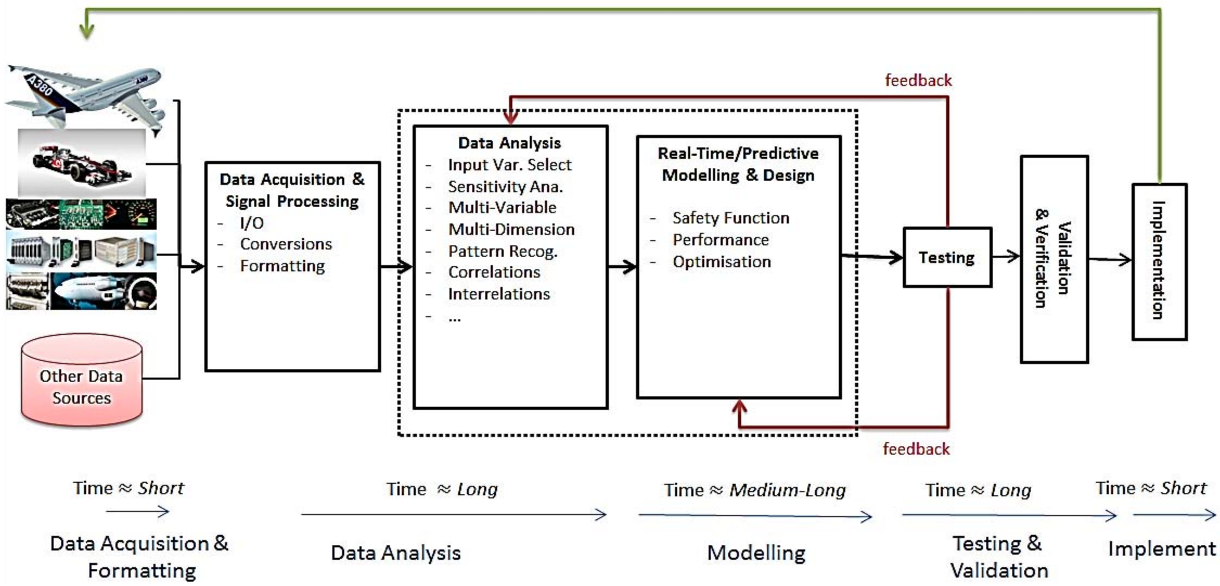


Fig. 1. Current design project, problem solving, performance measurement, and optimization process.

can be considered as a more reliable state vector in preparation for subsequent operations such as transfer functions, inferential models and other forms of data manipulation, and knowledge representation.

EventiC achieves this by: 1) interpreting changes in the value of input-output (I/O) data at the given event level; 2) detecting if I/O events coincide; and 3) grouping I/O events as related event. This processing happens in real-time during a specified time interval, known as the scan rate whose duration can potentially range from microseconds to seconds, and so on. At each scan a matrix of I/O coincidence is produced, which is similar in concept to the recording of a clip in a film. A time span for the recording is determined to generate sufficient frames to give statistical confidence. The impact of an input on output is calculated as the number of coincidence in the time span. Once the relationship between the inputs and the outputs and their weightings is established for the purpose of modeling and control we revert back to the actual value of the inputs and the outputs. The translation of system parameters to events and the grouping of relevant I/O events in near real-time is a novel approach in the understanding and the processing of large scale data/signals.

An industrial case study is presented that demonstrates the application of EventiC in the data analysis phase of a systems modeling and control optimization exercise. Here, the application of EventiC at the preliminary automation and the data analysis stage significantly reduced the time required to perform the system modeling, design, and validation. Fig. 2 depicts the new modeling platform.

When faced with the challenge of building a real-time data modeling method, EventiC is a novel data and the knowledge engineering platform that meets the challenges of data modeling and analysis in modern day complex systems. The proposed method endeavors to create a logical and yet simple foundation for management of the interrelationships and the dynamics of the components present within the embedded and related systems and their operating environment. Its sole purpose is as a tool by which to build the first rung on the ladder for understanding the causal relationships that exist between a system and its operational environment, as the system state and the boundaries change. As a method, EventiC is able to evaluate in near real-time the impact of every relevant event on the performance, stability, and overall system behavior.

In the following sections, we first discuss the underpinning theory of the proposed data and the knowledge engineering approach. Second, we introduce the most relevant input variable selection (IVS) and sensitivity analysis (SA) methods available, and subsequently, discuss the proposed EventiC method and its application in an industrial case study. Finally, a comparison is made between the proposed EventiC and EventTracker [1] methods using the same case study. This latter activity helps the authors establish the applicability of the two techniques in real-time data and knowledge engineering for industrial applications.

The challenges of understanding and interpreting the state (the being) and the behavior of a physical entity (i.e., system) have fascinated philosophers, systems theorists, and engineers. The underpinning idea of the proposed method is based on Descartes' philosophy of "Discours de la Methode," that is "break down every problem into as many separate elements as possible" [2], and then reassemble them to form an eco-system of causality of the smallest units. We also borrow from the concept of "coincidentia oppositorum" (or the "fight among parts") attributed to the 15th century thinker Nicholas of Cusa [3], [4] and interpret the concept as the causal interrelationships of parts in the whole.

To express this in the language of engineering, we try to identify the sources of excitation driven by events; measure the influence such events have on the input variables and in turn the effect that inputs have on the behavior, the stability, and the safety of the given system.

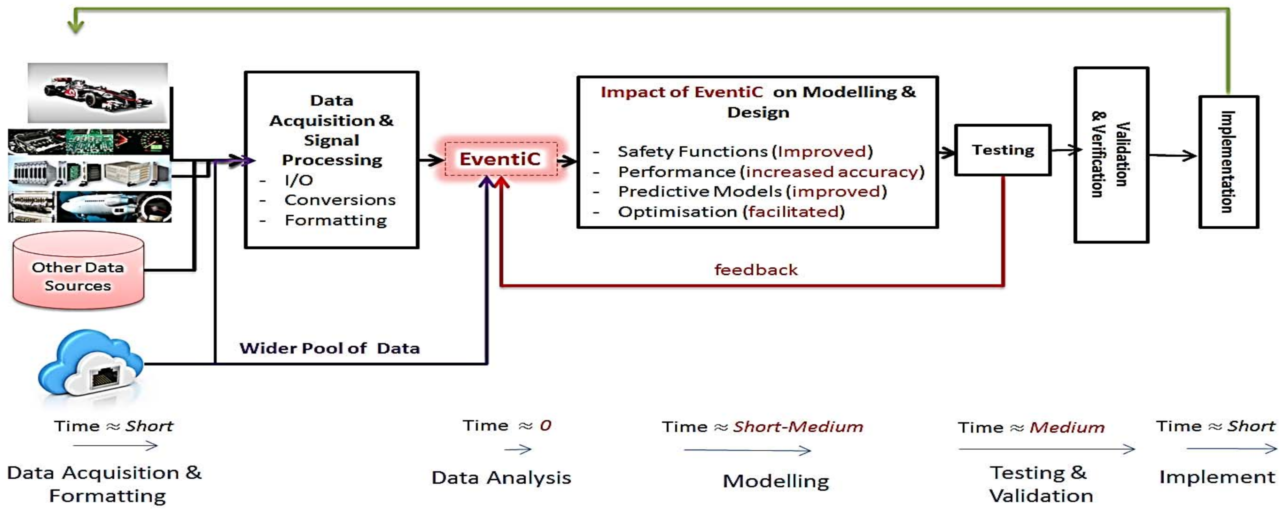


Fig. 2. Design project, problem solving, performance measurement, and optimization processes with EM embedded.

The scientific and technological challenge is to be able to assemble a system (process) definition, i.e., “character equation” that accurately represents the complex system in a timely fashion. The current theoretical approaches to systems and control that have successfully implemented Descartes’ Discours de la Méthode, have isolated individual systems with near perfect character functions. Especially, in the realms of physical (phenomenological) models (i.e., Newtonian, Kirchhoff’s circuit laws, etc.), numerical and analytical models (finite element analysis, statistical inference models, etc.), and in instances where processes become difficult to explain using character functions, then the heuristic models [neural networks, genetic algorithms (GAs), fuzzy inference, etc.] have been employed. In all cases, we can observe significant achievements. The principle taken by these approaches is based on finding a perfect solution from the outset that encompasses the knowledge about all the excitation parameters. As such the control system for complex processes becomes a multitude of independent isolated problems and the challenge itself becomes significant. Normally the above methods rely on a historical record of events and data. At the time of writing there are very few solutions that can interpret the data in real-time and respond to the excitation in an appropriate optimal manner.

Typically, the problem of timeliness has been solved by the integration of controllers (e.g., microcontrollers, programmable logic controllers, etc.) into the system, which usually manifests itself in the form of a supervisory control and data acquisition (SCADA) System. Such systems are recorders, managers, and archivists of data, and add little value beyond that fundamental functioning. Effectively across all the layers, such real-time subsystems conduct little in the way of raw data processing and interpretation, the interpretation and analysis of data is delegated to higher level subsystems [e.g., digital controllers, neural networks, fuzzy controllers, artificial intelligence (AI) techniques, or simply direct human intervention]. The inherent operating tardiness of these subsystems means that by the time they have learned the response to an event pattern, the system has moved on and there only hope of reacting to similar repeated incidents is if the system is able to recall cached historical related data, only then can the above models find the appropriate response to the excitation. Borrowing a quote from von Bertalanffy [2] “Problems must be intuitively seen and recognized before they can be formalized mathematically.” Mathematical formalism may very well impede rather than expedite the exploration of very real problem. We consider the visualization of the observable world to be the key to solving such complex problems and suggest that the proposed EventiC method takes a logical step forward in achieving this visualization.

II. Related Work

An important factor that facilitates data interpretation and information modeling is an appreciation of the effect that system inputs have on each output at the time of their occurrence. In the literature, methods that facilitate this interpretation have generally been referred to as an IVS in the engineering domain and within mathematics as SA. The purpose of IVS techniques is to maximize the quality of data acquisition and interpretation. In this context, input variables determine performance parameters and the cause-effect relationship between the input variables and performance parameters generates knowledge about the system. Measures taken to minimize the cost of data acquisition and its subsequent interpretation could arguably be interpreted as SA [1], [5]–[9]. The purpose of data-filtering and SA is to measure the true impact of each system input on each output. Due to the uncertainties in such relationships finding a true and faithful mathematical representation can be challenging. Finding a suitable SA method according to [1] first requires an awareness of the relationships that exist between input and output variables. This can be achieved in a number of ways, through analytical and numerical methods, sample-based and statistical models, or heuristic methods. However, all of the

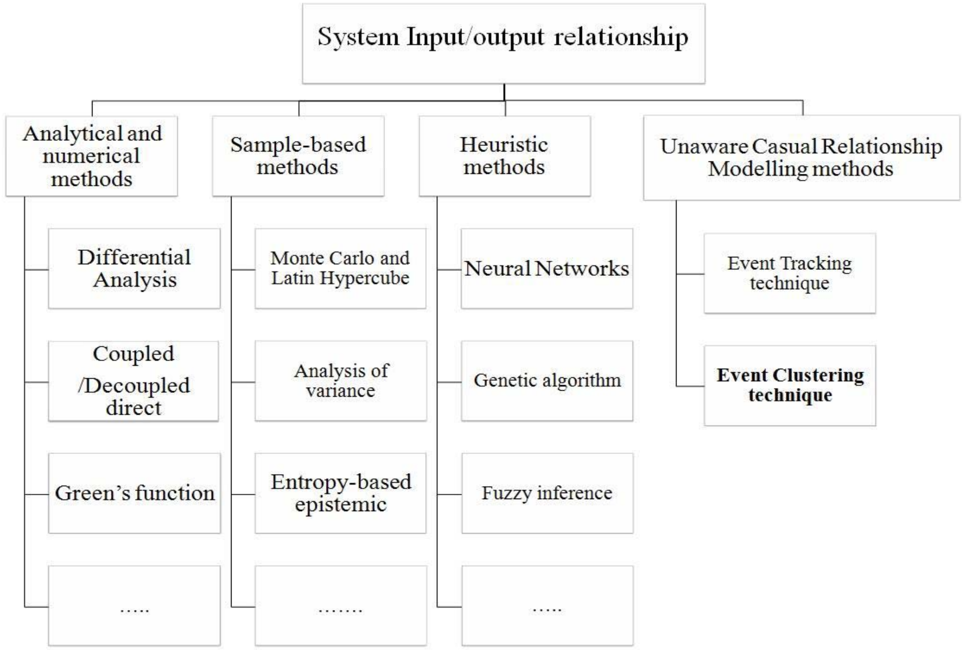


Fig. 3. Different IVS methodologies.

aforementioned techniques have a dependency on accurate historical data or alternatively rely on the knowledge of a domain expert who may not necessarily always be available.

Fig. 3 categorizes the various IVS and SA methodologies. Next, we present a brief review of the latest and most relevant analytical, statistical, and heuristic IVS and SA methods.

A. Analytical and Numerical Methods

Analytical and numerical methods measure the impact of changes in a variable on others by means of mathematical equations. Among the more popular analytical methods are differential analysis and Green's function.

1) Direct Differential Analysis: Differential analysis, also referred to as the direct method, is structured on the behavior of a base-case model scenario, where all parameters set equal to their mean value. Differential SA is based on partial differentiation of the aggregated model. When an explicit differential equation describes the modeled relationship, the sensitivity coefficient for a particular independent variable is derived from the partial derivative of the dependent variable with respect to the independent variable [11]. Methods such as Neumann expansion and perturbation methods [12] can help in to extracting these coefficients by approximating differential equations. However, complex and nonlinear relationships between system variables cannot always be guaranteed to exist in this type of analysis.

2) Green's Function: In the case of Green's function, the sensitivity equations of a model are obtained by differentiating the model equations. The sensitivity equations are then solved by constructing an auxiliary set of Green's functions. This method minimizes the number of differential equations for solving the SA and replaces them with integrals that can be easily calculated [13].

The concept of Green's function stems from the fact that the total output of a linear time-invariant system can be formulated by a summation of terms that adds all outputs of the system for all single points [14].

The disadvantages of Green's function are that it is limited to linear and time-invariant systems and its ability to only work with ordinary differential equations that govern dependent variables with respect to independent variables. In real applications it is often difficult to separate independent variables from dependent variables. Additionally, working on one variable at a time for multidimensional systems can be computationally expensive [4], [10].

B. Sampling-Based Methods

Sampling-based methods do not depend on model equations. Such methods try to establish the relationship that exists between inputs and outputs of a system using direct measurements obtained from observation of the system. These measurements normally take place at a specified interval during the process and are the result of multiple runs of the system model. Factors influencing the adoption of sampling methods can be attributed to the lack of accurate established analytical models capable of representing the behavior of the system, a lack of expertise in identifying the relationships between system parameters, or variation in system configuration and parameters result that results in them being too complex for accurate numerical and analytical methods [10]. Sampling-based SA methods tend to establish a model equation by identifying statistical features in the data series of two variables.

The main shortcoming of these methods is their reliance on historical data, the reliability of the prediction decreases when the time for data collection and its interpretation are limited.

For example, Cloke *et al.* [16] applied a model to 1280 sample values of 20 input parameters. All iterations of the model required an execution time of between 2 and 52 h per set of samples. The period required for the complete execution was approximately 46 days. Such cases illustrate the shortcoming of sampling-based analysis when applied to volatile systems that require timely analysis and response. Some of the widely used sampling-based sensitivity/uncertainty analysis methods are: Monte Carlo and Latin Hypercube Sampling.

The Monte Carlo method is one of the most widely applied techniques for uncertainty analysis. The method involves generating a set of random samples from the distribution of inputs and then running the model until such time as a statistically significant distribution of outputs is obtained. Problems such as optimization and simulation are frequently solved using the Monte Carlo simulation. For the interested reader, a wide range of literature describing the methodology, tools, and the applicability of the Monte Carlo method is available in [10]. Typically, this method requires a large number of samples and/or model runs which in itself can limit the applicability to simpler models. In the case of computationally intensive models, the time and resources required can be prohibitively expensive. In order to mitigate this computational overhead, some efficiency can be achieved by using the modified Monte Carlo method, a method that is more efficient in sampling from the input distribution [17], [18].

The Latin hypercube sampling method [19] is one such widely used variant of the standard Monte Carlo method. In this method, the range of probable values for each uncertain input parameter are divided into intervals of equal probability and result in the whole parameter space being partitioned into cells of equal probability. They are sampled in an "efficient" manner such that each parameter is sampled the once from each of its possible intervals. The advantage of this approach is that random samples are generated from the full range of possible values, thus giving an insight into the extremes of the output probability distribution.

One of the main challenges faced when applying the Monte Carlo methods to time-critical applications is in terms of the effort required to estimate the distribution of input variables prior to sample generation, this in itself can be computationally expensive when dealing with a large number of input variables.

C. Heuristic-Based Methods

Here, the IVS process is based on heuristic methods that in themselves normally rely on the knowledge of system experts. This knowledge often manifests itself in the form of experience, engineering and modeling expertise, or special algorithms. For example, fuzzy inference models, GA, AI (e.g., artificial neural network), or principal component analysis [20] fall in this category.

The strength of heuristic methods in solving complex data modeling and control systems is well established and recorded in industry and in the literature. For example, cement factories worldwide are being controlled by the direct knowledge of expert kiln operators. Fuzzy control of cement kilns has been one of the first successful applications of fuzzy control in the industry. The expert knowledge has a direct impact on identifying the fuzzy inference rules that optimize the key performance indicators in the manufacturing process. The advantages and shortcomings of expert reliant systems have been discussed in some detail in [21] and [22]. To reduce the reliance on the need for domain expert input, which can at times be time consuming and prone to variation, automatic AI-based learning methods have been used. The AI techniques examine the pattern of the acquired data and develop the necessary knowledge for measurement or optimization plans. GA techniques are also considered as one of the heuristic methods that derive knowledge from a known set of data points (genomes) and use the principles of random mutation and filtering of unwanted genes. A GA can be built with arbitrary flexibility and can be successfully trained using any combination of input variables [23]. GA's have been effectively used in the automating of IVS processes [24].

What distinguishes the proposed event modeling techniques from other automated data pattern and the knowledge derivation techniques is its simplicity and the speed at which it extracts all available data from the system domain, and then converts and processes the necessary information in near real-time. There is no reliance on a set of predefined rules such as good or bad genes, historical trends, or investigation of long-term patterns. More importantly, unlike heuristic methods, the EventiC technique does not rely on any prejudgment of the data relevancy, that is, normally a characteristic of expert interference and in that respect it is an unbiased method.

III. Event Clustering Data Grouping Technique

Event clustering [25], [26] is a technique based on the assumption that the state of a system during its lifetime can be broken down into a series of consecutive discrete events triggered by changes in the state of input variables (sensors and actuators). In real-time applications, this can help in associating important events with the performance indicators of the system. However, it is important to appreciate that discrete unbiased events imply that the system is not influenced by the history of previous events.

A. Data Clustering Methods for Big Data

Clustering is a class of unsupervised learning or semiunsupervised [37] methods where objects are grouped into a set of disjointed classes, known as clusters, such that objects within each class have close similarity. The goal of data clustering, also referred to as cluster analysis, is to find the natural grouping within a set of patterns or events. A review of the literature on clustering techniques reveals that despite the large number of algorithms used across a variety of applications it is not easy to decide on the most appropriate algorithm for a given data set with respect to satisfying both the computation efficiency and quality of the solution. Fahad *et al.* [38] have conducted a comprehensive survey of the most utilized clustering algorithms for big data. They found three criteria on which to classify the strengths and weaknesses of clustering algorithms. Namely, the volume of data, the velocity of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

data flows in real-time systems, and the variety of data types are the three major criteria to consider when considering the clustering of big data. These three Vs (volume, velocity, and variety) are the core aspects and characteristics of big data that have to be taken into account when selecting an appropriate clustering algorithm.

In summary, the literature review on existing clustering algorithms for the big data indicates that large quantities of memory and time are required for the data clustering. That the algorithms are complex to implement and the vast majority is designed to operate on historic datasets. In early experiments performed on various clustering techniques reported in the literature and evaluated on a laboratory devised platform showed that the rank order clustering (ROC) method has the most potential for grouping data in real-time. This method could handle a large volume and variety of data sources with the excellent efficiency and the effectiveness. More importantly, ROC has shown that it is capable of handling large data sets in real-time, thus, fulfilling the most important aspect of the proposed solution.

B. Rank Order Clustering

The ROC method introduced by King [27] used matrix manipulation to rearrange the row and the columns of a matrix in an iterative manner. Ultimately, given a finite number of steps, the method results in a matrix whose rows and columns are arranged in order of decreasing value. It is an effective algorithm in determining the occurrence of clusters in a block diagonal format. The application of this approach is limited by the assumption that groups of data are similar and will be placed into mutually exclusive categories.

In the cluster analysis method, data values considered "similar" in accordance with a "similarity criteria" can be replaced by a new value representing the group (clumping) or assigned a unique type of label (partitioning) [28], [29]. The proposed event clustering technique uses this approach to build a cause-effect grouping of system input events (originating from sensor/actuations) and output events (performance indicators of the system).

C. Basic Concepts and Parameters

The basic parameters of the proposed event clustering method are borrowed from [1] (also see [25], [26]). As such a quick cross reference is presented.

1) Discrete Event Systems: A discrete event system (DES) is defined by the disparate occurrence of events in a specified time span. In this context an event is any change in system state. The state of the system changes, when changes in input variables lead to a change in system outputs. Therefore, in DES, only the attributes that represent the occurrence of an event are considered.

2) Trigger Data and Event Data: Any input variable whose value transition registers an event is defined as trigger data (TD) in the DES. The series of data that represent the state of the system at any given time is described as event data (ED).

3) Trigger and Event Thresholds: The fluctuations in the TD and ED series that are interpreted as triggers are determined by comparison with the trigger threshold (TT) and event threshold (ET) values. TT and ET are, respectively, expressed as a percentage of the TD and ED value ranges for a given time span. They are determined by system domain experts.

4) Actual Value of Data: Is expressed as the value or state of the actual data at a given time instance. This data can be expressed in a binary, integer, or decimal format.

D. Assumptions of the Proposed Method

The event clustering method is based on the following assumptions.

1) The delay between EDs and the corresponding TDs is negligible (for all intent and purposes instantaneous).
2) The definition of trigger and ETs are prespecified and determined by system domain experts.
3) The EDs and TDs series are homogeneous since the heterogeneous nature of the systems parameters have now been translated into the homogenous ED.

E. EventiC Algorithm

In the context of real-time event driven systems, the proposed method is based on the assumption that changes to input variables may potentially trigger events. Every single or combination of events may subsequently result in a change to the system state. The proposed event clustering method describes variables and system states in terms of a collection of events.

The algorithm was developed with the real-time operation in mind. The design of the event-driven incidence matrix (EDIM) is based on sorting rows by the inputs and the columns by the outputs. Incidence matrix elements take the values 0 or 1. The value is 1 when both or neither of the input/output ED is triggered, otherwise it is 0. This operation is similar to a logical exclusive-NOR (XNOR) operator. XNOR acts as a function that measures the correlation between inputs and outputs.

1) Trigger-Event Detection: Equation (1) shows the relationship between each event triggered by input_t and input_{t-1} with respect to changes in the output. Each change to the output in a given timespan can be expressed as an event and the positive value of the inputs as triggers, and then output can be defined as Event ED. Both Input_t and Input_{t-1} can be considered as TD

$$\begin{aligned} &\text{if } \left(\mathrm{Input}_t - \mathrm{Input}_{t-1}\right) \geq \theta \xrightarrow{\text{Trigger}} \mathrm{TD}_t \\ &\text{if } \left(\mathrm{Output}_t - \mathrm{Output}_{t-1}\right) \geq \Psi \xrightarrow{\text{Event}} \mathrm{ED}_t. \end{aligned} \tag{1}$$

Fig. 4 shows that within each time span, input/output pairs are detected and used to generate the elements of the incidence matrix. The ROC method is applied to the incidence matrix and the weighted rows and columns are clustered in the upper-left part of the EDIM. The resulting EDIM shows the ranked relevance of each input to the output. The application of the ROC method results in clusters of the most relevant group of input ED (sensors and actuators) against output ED (plant performance indicators).

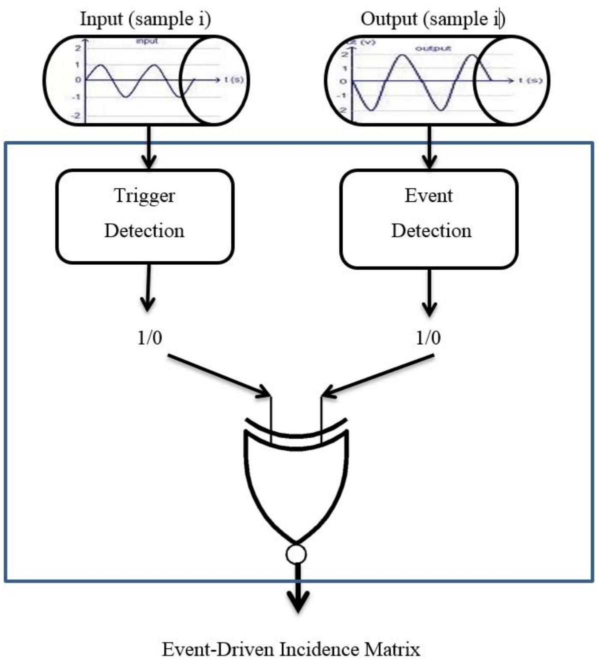


Fig. 4. Trigger and event detection functionality on each time scale.

2) Implementation of the ROC Algorithm: In this section, a step-by-step implementation of ROC method will be presented. A weight for each row i and column j (in the m by n matrix) are calculated using the following algorithm [28].

Step 1: Build a coincidence matrix of TDs (rows) and process output EDs (column).

Step 2: Populate the model's TDs/EDs coincidence matrix with binary weighed values resulting from application of the exclusive NOR function.

Step 3: Assign a binary weight $\mathrm{BW}_j = 2^{m-j}$ to each column j of the incidence matrix.

Step 4: Determine the Decimal Equivalent (DE) of the binary value of each row i using

$$\mathrm{DE}_i = \sum_{j=1}^{m} 2^{m-j} a_{ij}. \tag{2}$$

Step 5: Rank the rows in decreasing order of their DE values. Rearrange the rows based on this ranking. If no rearrangement is necessary, stop; otherwise go to step 4.

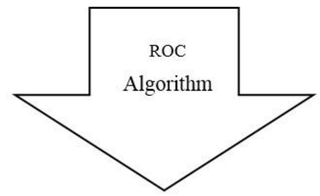
Step 6: For each rearranged row of the incidence matrix, assign binary weight $\mathrm{BW}_i = 2^{n-i}$.

Step 7: Determine the decimal equivalent of the binary value of each column j using the formula

$$\mathrm{DE}_j = \sum_{i=1}^{m} 2^{n-i} a_{ij}. \tag{3}$$

Step 8: Rank the columns in decreasing order of their DE values. Rearrange the columns based on this ranking. If no rearrangement is necessary, stop; otherwise go to step 7.

TDs/EDs	1	2	3	4	5
1	0	1	1	0	1
2	0	1	1	0	1
3	0	1	1	0	1
4	1	0	0	1	0
5	0	1	1	0	1
6	1	0	0	1	0
7	1	0	0	1	0
8	0	1	1	0	1



TDs/EDs	1	4	2	3	5
4	1	1	0	0	0
6	1	1	0	0	0
7	1	1	0	0	0
1	0	0	1	1	1
2	0	0	1	1	1
3	0	0	1	1	1
5	0	0	1	1	1
8	0	0	1	1	1

Fig. 5. Input triggered/ED incidence matrix before (up) and after (down) implementation of the ROC algorithm.

For example, as shown in Fig. 5 after implementing the ROC algorithm TD4, TD6, and TD7 are related to the ED1 and ED4.

3) Sample Scan Size: The sample size (i.e., the number of samples used to build the incident matrix) is determined by domain experts. While there is no theoretical upper bound on the sample size taken from the data series, typically in terms of a lower bound then usually some 250 samples are used. The data is then processed using the EventiC algorithm to build the incidence matrix. A new output matrix is generated for each scan of the system; the normalized weight of each input variable acts as the coefficient of the system outputs. Fig. 6 shows four sample scans and their analysis operations in four sequential sample slots.

IV. Industrial Case Study

By the way of additional illustration and explanation, a case study is presented that addresses the application of the proposed real-time event clustering technique in the cement manufacturing industry. The cement manufacturing process is one of the most challenging industries in terms of environmental impact, energy consumption, and raw material usage. This paper examines the optimization of kiln operation, the plant in question needs to be more reactive and predictive in order to improve the quality and efficiency of its operation without impacting on production throughput. One of the most important outputs in cement production is the formation of clinker that results from a sintering process carried out in the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

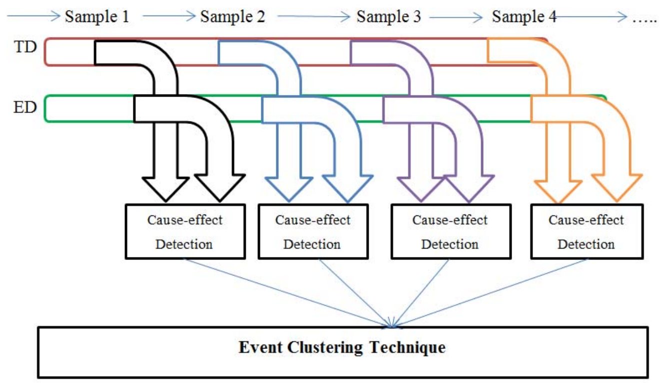


Fig. 6. Trigger event detection functionality over whole sampling time.

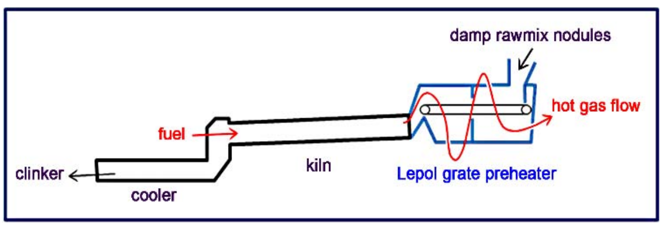


Fig. 7. Kiln system diagram including preheater and cooler.

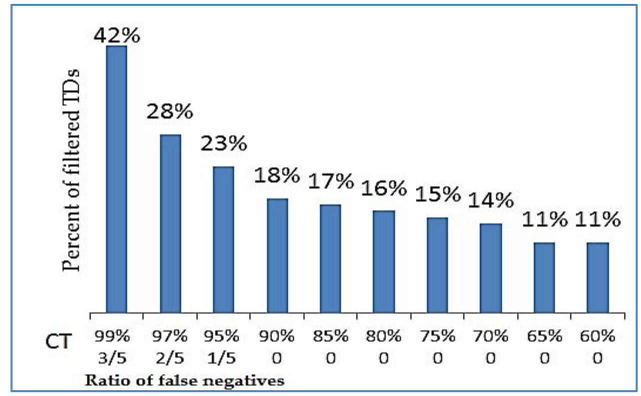


Fig. 8. Percentage of filtered TDs per CT and ratio of false negatives.

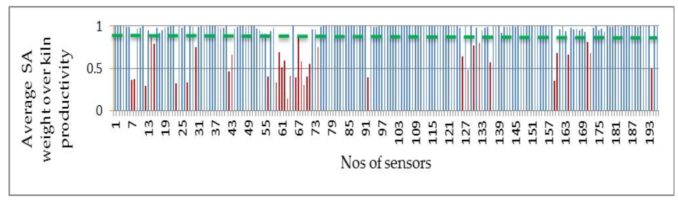


Fig. 9. Kiln production rate SA with respect to 196 inputs over 1 month sampling snapshots with 90% CT threshold.

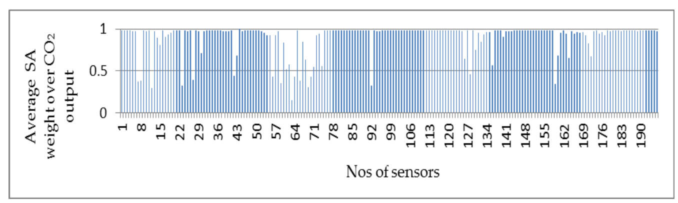


Fig. 10. Kiln's CO_2 emission SA with respect to 196 inputs over 1 month sampling snapshots.

kiln. The complete process is shown in Fig. 7, and includes the kiln, preheater, and cooler subsystems.

In the implementation, EventiC was used for: 1) providing production and operation managers with information about the effects of the causal relations that exist between input events and production outputs; 2) for providing production engineers with the necessary knowledge about the optimal state of the production process and machine behavior; and 3) providing the process optimizer and the decision aid system with accurate information about the relationship that exists between key performance indicators and the actual shop-floor control parameters.

The raw data source is the cement plant SCADA system. The total number of data points coming from the shop floor is generated by 196 sources that represent data from the kiln (i.e., the sensors, the actuators, and the control parameters connected to the SCADA system). EventiC translates this data into input TD for the purpose of controlling the kiln and its peripheral equipment. The output ED is collected from sensors and counters that measure the production rate (kiln output is defined as the volume of satisfactory products), energy consumption (kiln temperature is directly related to energy consumption), and CO_2 emission. The data sampling rate of the system is set at 1 per min (the scan rate of the SCADA system). The event modeling process was conducted over a 1 month production period and represented some 43 000 data samples.

At stage two of the EventiC algorithm, the data tables for the plant control and monitoring system, where the cause-effect relationship between data triggers (changes in sensors outputs) and events (changes in system outputs) are measured. A new output matrix is generated on each scan of the system. At stage three the average sensitivity of each input on output parameters is calculated.

At stage four, a cut-off (CT) threshold is applied to filter the less important input variables in the data set. Its value is in the range $0 \leq \text{CT} \leq 1$. For example, when $\text{CT} = 0.8$, all inputs with an average SA weighting of less than 0.8 (that is 80%) are filtered out. A false negative test was used to check whether any important inputs were accidentally discounted. The false negative assessment took the form of a simple experiment that varied the CT value in the range 0.99 to 0.6; the results are shown in Fig. 8 which shows the percentage of filtered TDs measured in the experiment with respect to different CT values and the ratio of false negatives. Figs. 9–11, respectively, show the EventiC sensitivity measures for the three key performance indicators of: 1) plant production rate; 2) CO_2 emission; and 3) kiln temperature. These results were obtained with a CT value of 0.90 (90%), and based on this value, 18% of triggered inputs (36 TDs) have been filtered out as false negatives (the red bars in Fig. 9), leaving 82% of input sensors readings as the input variables representing the state of the kiln.

Table I lists nine of the key input and their corresponding impact on kiln production rates. The input variables shown in bold represent the inputs that have the greatest impact

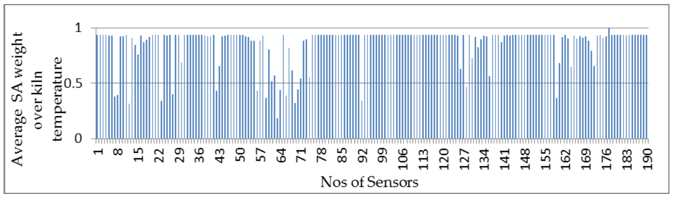


Fig. 11. Kiln's temperature SA with respect to 196 inputs over one month sampling snapshots.

TABLE I
AVERAGED SA WEIGHT OF SELECTED NUMBER OF KILN'S INPUT DATA OVER KILN PRODUCTION RATE

Input Name	Sensitivity Level of kiln Production Rate	Subjective Importance level with CT=90%
Kiln temperature	92 %	High
CO output	63 %	Moderate
I/h return in Kiln	90%	High
Kiln fan	98%	High
CO_2 output	97%	High
Motors pulls material from kiln	92%	High
Injected 0_2 to Kiln	37%	Low
Injected NO_2 to Kiln	54%	Moderate
Injected SO_2 to kiln	36%	Low

on the production rate. The weighing mechanism discussed previously is based on the number of times the input-output ED coincided during the duration of the analysis period consisting of some 43 000 data points in this instance.

The state vector (also see Section I) for the production rate of the kiln started from $V_S = [x_1, \ldots, x_S]$, where the event vector is expressed as $E_S = [e_1, \ldots, e_{196}]$ and $S = 196$. By applying EventiC the event vector is rationalized to $E_N = [e_1, \ldots, e_5]$, the state vector can be presented as $V = [x_1, \ldots, x_5]$. By referring to the actual control system and in consultation with experts in the cement plant, the algorithm successfully reduced the original number of input parameters used to control the measurement of kiln output from nine parameters to five parameters without compromising the quality of relevant control parameters. This result demonstrates a major reduction in dimensionality (degree of freedom) of the problem statement and its complexity. Figs. 8 and 9 show the false negative and CT threshold process.

A. EventiC and Real-Time Plant Control

The proposed event clustering technique is used as an IVS method, a filter whose aim is to provide timely high quality input data to the higher level optimization, autonomous or intelligent systems in the control hierarchy. In the case of the cement plant and its kiln operations, we demonstrate how the method functions as a stability/optimization tool for process control.

One interesting observation is that the maximum production rate was not the result of a single cluster of system inputs (i.e., system settings). Fig. 12 shows the five highest plant production rates of 9.68–9.76 tons/h were achieved over the analysis time span (i.e., one month of production). These five maximum production rates and their corresponding input clusters (i.e., plant settings) are listed in Table II(a). The data shows that in practice it is possible to maintain the highest production rate with five different system settings (i.e., five alternative ways to achieve the same output), a fact that was previously unknown to the plant engineering team. In terms of simple production economics the data series shows that the control system is capable of operating at the minimum cost whilst still achieving maximum production throughput. Solution 1 [row 1 in Table II(a)] demonstrates a solution that maximizes production throughput while maintaining the lowest kiln temperature (less energy consumption). Solution 2 [row 2 in Table II(a)] shows a solution that minimizes CO_2 emission while maximizing production throughput. Table II(b) and (c) shows data relating to the input variable clusters, energy consumption, and CO_2 emission of the kiln.

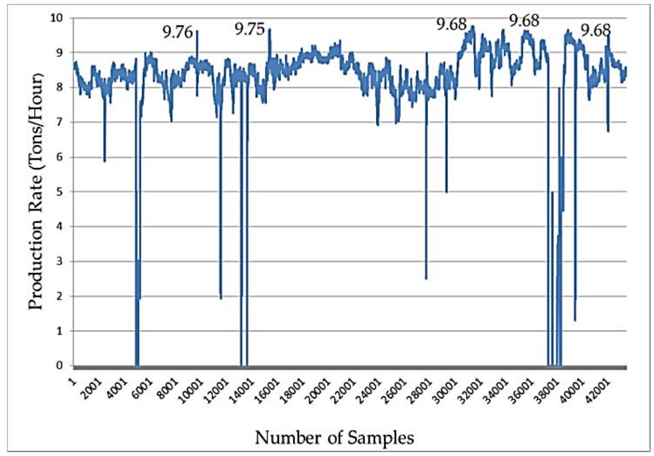


Fig. 12. Actual value of production rate and its maximum points over one month sampling.

Knowledge relating to the clustering of input variables and their impact on a given performance indicator allows timely intervention to be made by controllers in maintaining stability and optimizing performance. Knowledge of the relationship between key systems parameters (i.e., control inputs) and performance parameters (i.e., output) allows plant engineers to employ alternative solutions to a given problem in a timely manner and provides a degree of flexibility where alternative optimum solutions can be chosen if the current solution fails. EventiC provides the ability to return the system to stable/optimal operating condition in near real-time using an alternative solution that can potentially impact on open and closed loop control systems efficiency [30].

In the next section, we demonstrate the application of EventiC in production quality control of the cement plant.

1) Cement Quality As Key Performance Indicator: The quality of the cement produced is directly related to the kiln temperature. The best quality cement is produced at a kiln temperature of 1550 °C, the quality drops to medium at 1450 °C, and to the lowest passable quality at approximately 1350 °C [32].

Table II(a) shows system setting where the kiln temperature on average is approximately 1350 °C. Not surprisingly with

TABLE II

(a) Five Alternatives Cluster of Input Variables in Maximum Kiln's Production Rate. (b) Five Alternatives Cluster of Input Variables in Minimum Kiln's Energy Consumption. (c) Five Alternatives Cluster of Input Variables in Minimum Kiln's CO_2 Emission. (d) Five Alternatives Cluster of Input Variables for Medium Quality of Cement

(a)

	Kiln Production Rate (Tons /Hour) Kiln output	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO_2 output (%)	Motors pulls material from Kiln (on/off)
Solution 1: Minimum Kiln Temperature	9.76	**1274**	374.05	1	15.47	1
	9.75	1355	370.86	1	15.50	1
Solution 2: Minimum CO_2 emission	9.68	1368.3	365.34	1	**14.52**	1
	9.68	1369.65	371.15	1	15.55	1
	9.68	1371.7	369.05	1	15.47	1

(c)

	Kiln Production Rate (Tons /Hour) Kiln output	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO_2 output (%)	Motors pulls material from Kiln (on/off)
Solution1: Minimum Kiln temperature	8.10	**1283.3**	383.28	1	0.037	1
	8.57	1371.3	389.68	1	0.037	1
	8.09	1283.7	383.4	1	0.037	1
Solution 2: Maximum Kiln productivity	**9.34**	1395.7	379	1	0.037	1
	9.27	1315	357.5	1	0.055	1

(b)

	Kiln Production Rate (Tons /Hour) Kiln output	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO_2 output (%)	Motors pulls material from Kiln (on/off)
	2.8	1053.3	98.39	1	2.69	1
	3.47	1053.3	80.08	1	3.34	1
Solution1: Minimum CO_2 emission	5.00	1053.3	81.79	1	**0.111**	1
	8.45	1053.3	799.84	1	12.19	1
Solution 2: Maximum Kiln productivity	**8.97**	1053.3	159.6	1	12.07	1

(d)

	Kiln Production Rate (Tons/Hour)	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO_2 output (%)	Motors pulls material from Kiln (on/off)
	8.18	1454	338	1	14.26	1
	8.2	1446	340	1	14.07	1
	8.36	1441	319	1	13.52	1
Minimum CO2 emission and Maximum production rate	**8.95**	1452	319	1	**12.52**	1
	8.86	1439	342	1	13.69	1

such a low temperature, the quality of cement produced with this cluster of inputs is at the low end of the quality range. Table II(d) shows a cluster of inputs, where the kiln temperatures are in the medium cement quality range (approximately 1450 °C), and the EventiC algorithm optimizes the best set of inputs to meet maximum production throughput and minimum CO_2 emission.

2) Environmental Impact As Key Performance Indicator: Due to high demand for cement from the construction sector, the cement industry is likely to remain a major emitter of green house gases in the foreseeable future. The current existing methods for pollutant emission reduction within the industry do not appear capable of offsetting such growth [31], [32]. Emission control is a major challenge for the industry where the focus is on CO_2 emissions. Key factors influencing CO_2 emission levels are lime production, cement kiln dust, and fuel combustion patterns. This is further compounded by the use of hybrid fuel systems that are typically fueled by natural gas, coal, coke, oil, or organic material—each with its own specific burning profile and emissions levels. Statutory volumes of CO_2 emissions are regulated by the host countries regulations and international standards. The application of EventiC has given rise to a scenario where the system settings have delivered minimum CO_2 emission while achieving maximum production throughput.

B. Detection of the Unknown Factors Affecting the Behavior of the System

EventiC not only functions as an intelligent event recorder but also as middleware between the plant and its operating environment that facilitates the preliminary data and the knowledge construction. By providing information about events, information that was not necessarily available to engineers at the outset of design and modeling process, can lead to a fundamentally shift in the perception of system boundaries,

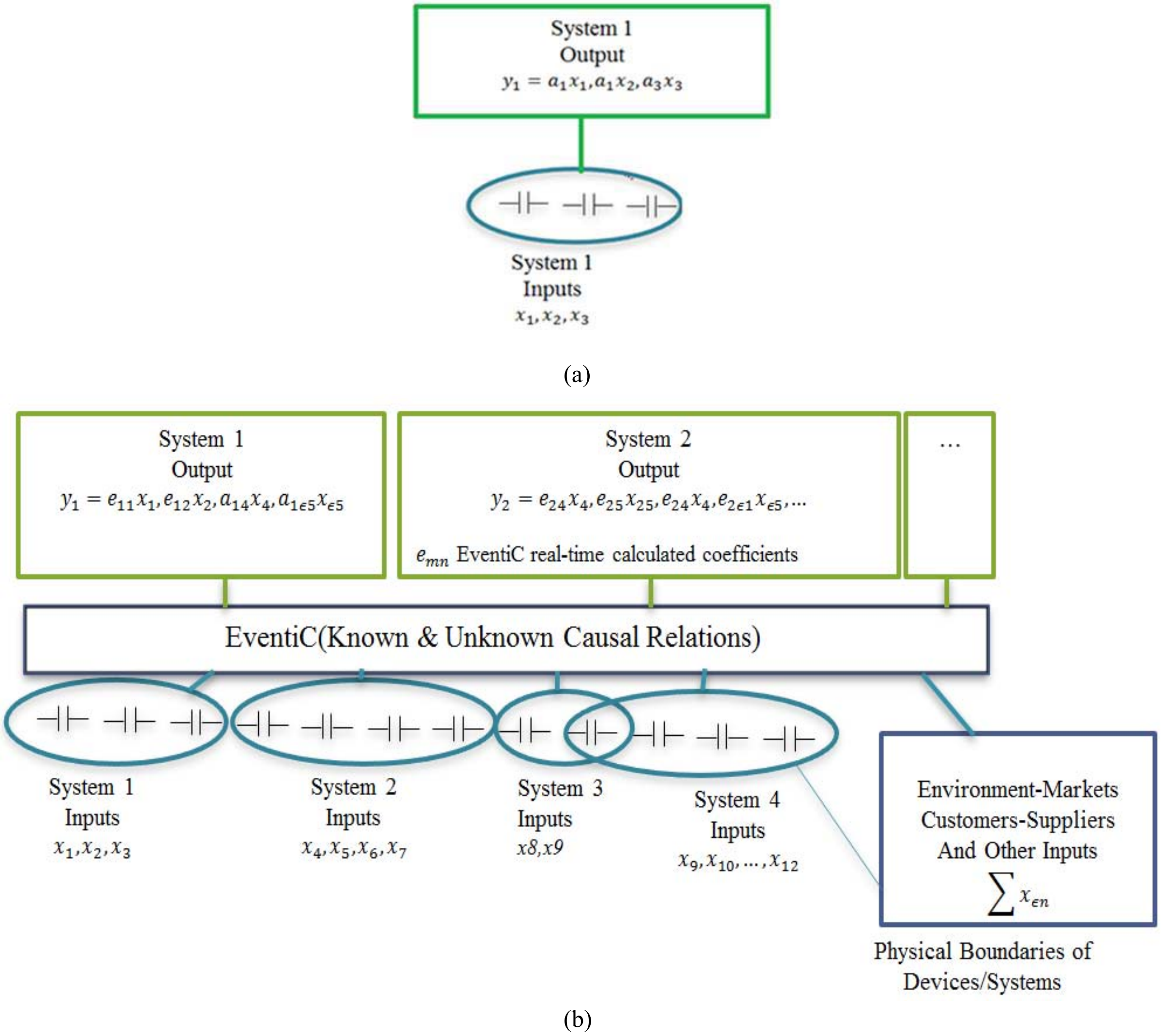


Fig. 13. (a) Input/output relationship in current systems. (b) Input/output interrelationship via EventiC.

where there is a move from a rigid system boundary to one that is more dynamic.

Fig. 13(a) illustrates the current approach to complex system architecture were large systems are broken down into isolated and abstract smaller subsystems for the purposes of simplification and control. In the modern complex systems, the principle of isolation is now becoming less feasible. Given system designers were able to analyze a greater range of potential influences then they would be in a better position to define more accurate models.

EventiC is based on the premise of managing the interrelationships and the internal dynamics of the components within the ecosystem of an embedded system and its environment. Automatically achieving this leads to a reduction in the time required for detection, classification, and analysis of known and previously unknown input data.

Assuming that all inputs having an influence on the system are related to the system outputs, the method finds potentially unintuitive and complex relationships that were unlikely to be identified by conventional systems analysis alone. Fig. 13(b) shows how EventiC could be used to integrate isolated systems together in order to detect potentially unknown factors in predictive models. This feature will allow engineers to build more effective, safer, and responsive systems that become part of the volatile environment in which they function.

For example in our experiment, we observed that the "motor speed used to pull material from the kiln" had a 92% sensitivity impact on the kiln's production rate. In most cement-related literature, the impact of the motor speed is completely ignored. One of the major advantages of the EventiC methodology is in the recognition of such previously unknown/undetected influencing parameters.

V. Assessment of Efficiency and Validity of the EventiC in Comparison With EventTracker Sensitivity Analysis Techniques

The EventTracker [1] methodology is an event-based SA technique that relates field data to the performance and the process parameters. The algorithm is discussed at some length in [1], but in brief the algorithm consists of setting parameters for ED and associated event trigger rules. The second step is to produce the input/output coincidence matrix based on the specified system scan rate. The third step is to extract sensitivity indices for the parameters at the specified interval (set by system engineer). The fourth step is to generate a normalized sensitivity index (SI) for each parameter in the analysis span (contiguous scan intervals). The fifth step is to filter out the unimportant inputs, by defining a CT threshold. For example, any input with an SI value less than 0.6 may not be considered to have a significant influence on the output, and therefore, can be ignored. The final step is the validation and verification of the results using a false-negative testing process. It should be noted that the EventTracker employs pair-wise event coincidence analysis. In order to validate the EventiC SA technique, the same data series was also analyzed using EventTracker [1]. The rational for choosing EventTracker over other SA methods for this comparison is that they are similar in nature and operate on real-time data, and moreover, the methods do not rely on the availability of statistically reliable or the homoscedasticity of historical data. The outcome from their deployment shows that the results of the SA are very similar, but there is a difference in the CT thresholds as shown in Fig. 14.

The experiments discussed in Section IV reveal that with a CT of 90%, 18% of TDs (36 TDs) are filtered out and the percentage of false negatives drops to 0. Fig. 14 also shows that to meet the same percent of filtering (i.e., 18%) then in the case of EventTracker, its CT needs to be set at 87%. The results also shows that with full set of 196 TDs, the EventiC algorithm took 40 s to calculate system output, whereas when using only 160 TDs the time taken to achieve the same result was 28 s representing a reduction of 30% in computation time. This computational time saving for EventTracker algorithm is some 35%. In terms of the computational effort on a personal computer (Intel Core i7 CPU & 4 GB RAM) the average CPU utilization for EventiC remained at 55% during the run-time, whereas for EventTracker was 65%.

The key difference between EventiC and EventTracker is that EventTracker reveals the correlations between multiple input and single output (i.e., one-to-one relationship), but EventiC reveals many to one and many to many correlations between inputs and outputs. This allows data analyst to visualize the relations between groups of input and output from the outset. Grouping could potentially reveal new insight into the internal relationship between the members on input series or output series.

VI. Application of EventiC in Expert Systems—Future Work

The complexity of kiln operation does not readily lend itself to the deployment of classical control architectures; as such our industrial partner uses an alternative expert system for plant performance optimization that employs fuzzy controllers. Traditionally, fuzzy controller inference rules are defined by domain experts with knowledge of cement production. However, recently such control systems have employed neural network and genetic programming to automatically extract fuzzy control inference rules [33]–[36]. These methods mainly rely on historical data derived from experience or machine learning algorithms.

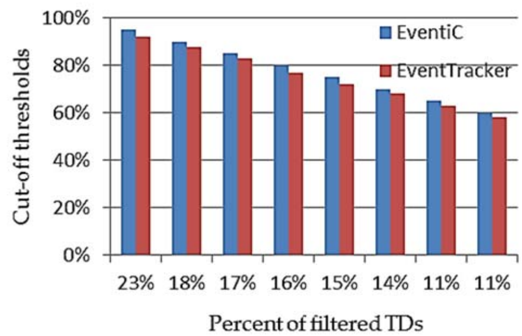


Fig. 14. EventiC and EventTracker CT versus percent of filtered TDs.

The authors suggest that the proposed method based on causal relationship modeling and parameter weighing mechanism could be used to extract fuzzy control inference rules and be utilized as a cost effective alternative for fuzzy controller IVS. In such scenarios, the EventiC would reside on top of a typical data acquisition system (e.g., SCADA) and translate the data into cause-effect event models. It would link the set of events (inputs) to the set of performance variables (outputs). The fuzzy inference rules and the parameters would be made available to the plant's fuzzy controller layer.

VII. Conclusion

This paper proposes a novel SA methodology for large scale real-time data analysis and modeling. The authors believe that the EventiC can be used as a robust real-time unbiased IVS technique for embedded control systems. When implemented as an architectural fusion that combines embedded monitoring and control with the proposed IVS method, then the result is a data modeling platform capable of real-time operation. This will allow control engineers and system designers to build more adaptable and responsive systems that exhibit optimal performance and system stability. The technique will not only yield improvements in the design of systems for process manufacturing but also has the potential to be applied into other domains such as aerospace, automotive, and smart metering. As a platform the method is capable of providing a foundation for sharing and integrating multiple users across various applications and resulting in the creation of cyber-physical systems that understands the effect of known and previously *unknown* inputs.

In this paper, the authors have demonstrated the key feature of a real-time event modeling method and demonstrated its ability to rapidly generate an EDIM that measures the degree of influence inputs have on system outputs. Moreover, the proposed real-time event modeling method does not require prior knowledge of the analytical or statistical relationship that may well exist between system input and output variables. The authors believe the method goes some way to removing the logical boundaries of isolation that exists in complex systems and replaces it with the principle that every data input effects the system output unless proven otherwise.

The application of the proposed method for kiln control in the cement industry is presented. The kiln is equipped with 196 sensors that are scanned at 1 min intervals during the duration of the cement processing. The data acquired provided EventiC with sufficient information to optimize the number of relevant input variables and provide accurate knowledge of the systems state over a period of one month's operation (some 43 000 observations). The results showed that 18% of TDs had very little effect on kiln productivity and as such could be totally ignored without impacting on kiln output.

To validate the proposed EventiC SA method, the results are presented to industry experts who collectively have experience in the design, control, and operation of over 40 cement manufacturing sites worldwide. The algorithm will shortly be implemented in a number of plants for the purpose of automating input data selection and helping to generate control and optimization rules. An additional observation is that the computational efficiency of the EventiC method has improved performance by some 10% in comparison with the earlier EventTracker [1].

The proposed EventiC method has the potential to act as a platform for data analysis in HiL simulation systems, where the quality of input data plays a major role in the accuracy of control and safety models.

Acknowledgment

The authors would like to dedicate this paper to the late P. Broomhead (coauthor) who recently passed away. May his soul rest in peace.

References

[1] S. Tavakoli, A. Mousavi, and P. Broomhead, "Event tracking for real-time unaware sensitivity analysis (EventTracker)," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 2, pp. 348–359, Feb. 2013.

[2] L. von Bertalanaffy, "The history and status of general systems theory," *Acad. Manag. J.*, vol. 15, no. 4, pp. 407–426, 1972.

[3] J. Hoff, *The Analogical Turn: Rethinking Modernity With Nocholas of Cusa, Part II: Nocholas of Cusa and the Genealogy of Modern Concepts of Space and Autonomy*. Grand Rapids, MI, USA: William B Eerdmans, 2012.

[4] W. Minnebo and S. Stijven, "Empowering knowledge computing with variable selection," Ph.D. dissertation, Dept. Comput. Sci. Math., Univ. at Antwerp, Antwerp, Belgium, 2011.

[5] Z. Zi, "Sensitivity analysis approaches applied to systems biology models," *IET Syst. Biol.*, vol. 5, no. 6, pp. 336–346, 2011.

[6] E. Fock, "Global sensitivity analysis approach for input selection and system identification purposes—A new framework for feedforward neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 8, pp. 1484–1495, Aug. 2014.

[7] J. Lv, H. Hu, K.-Y. Cai, and T. Y. Chen, "Adaptive and random partition software testing," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 12, pp. 1649–1664, Dec. 2014.

[8] S. Li, K. Deng, and M. Zhou, "Sensitivity analysis for building energy simulation model calibration via algorithmic differentiation," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 905–914, Apr. 2017.

[9] A. Saltelli, K. Chan, and E. M. Scott, *Sensitivity Analysis.* Chichester, U.K.: Wiley, 2008.

[10] S. Tavakoli, A. Mousavi, and S. Posland, "Input variable selection in time-critical knowledge integration applications: A review, analysis, and recommendation paper," *Adv. Eng. Informat.*, vol. 27, no. 4, pp. 519–536, 2013.

[11] R. Caruana and V. R. De Sa, "Benefiting from the variables that variable selection discards," *J. Mach. Learn. Res.*, vol. 3, pp. 1245–1264, Mar. 2003.

[12] A. Buonomo and A. Lo Schiavo, "Nonlinear distortion analysis via perturbation method," *Int. J. Circuit Theory Appl.*, vol. 38, no. 5, pp. 515–526, 2010.

[13] S. S. Isukapalli, "Uncertainty analysis of transport-transformation models," Ph.D. dissertation, Chem. Biochem. Eng., Graduate School, Rutgers Univ., New Brunswick, NJ, USA, 1999.

[14] G. Beylkin, C. Kurcz, and L. Monzón, "Fast algorithms for Helmholtz Green's functions," *Proc. Roy. Soc. A Math. Phys. Eng. Sci.*, vol. 464, no. 2100, pp. 3301–3326, 2008.

[15] E. Volkova, B. Iooss, and F. Van Dorpe, "Global sensitivity analysis for a numerical model of radionuclide migration from the RRC 'Kurchatov Institute' radwaste disposal site," *Stochastic Environ. Res. Risk Assesment*, vol. 22, no. 1, pp. 17–31, 2006.

[16] H. L. Cloke, F. Pappenberger, and J.-P. Renaud, "Multi-method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes," *Hydrol. Processes*, vol. 22, no. 11, pp. 1660–1674, 2008.

[17] C. Andrieu, A. Doucent, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 72, no. 3, pp. 269–342, 2010.

[18] Y. Liu, M. Y. Hussaini, and G. Ökten, "Optimization of a Monte Carlo variance reduction method based on sensitivity derivatives," *Appl. Numerical Math.*, vol. 72, pp. 160–171, Oct. 2013.

[19] S. C. Hora and J. C. Helton, "A distribution-free test for the relationship between model input and output when using Latin hypercube sampling," *Rel. Eng. Syst. Safety*, vol. 79, no. 3, pp. 333–339, 2003.

[20] H. Jeong *et al.*, "A heuristic method of variable selection based on principal component analysis and factor analysis for monitoring in a 300 kW MCFC power plant," *Int. J. Hydrogen Energy*, vol. 37, no. 15, pp. 11394–11400, 2012.

[21] M. Fallahpoor, A. Fatehi, B. N. Araabi, and M. Azizi, "A supervisory fuzzy control of back-end temperature of rotary cement kilns," in *Proc. Control Autom.*, Seoul, South Korea, 2007, pp. 429–434.

[22] G. Noshirvani, A. Fatehi, B. Araabi, M. Shirvani, and M. Azizi, "Comparison of rotary cement kiln identified models," in *Proc. IEEE Int. Conf. Control Autom. (ICCA)*, 2009, pp. 1290–1295.

[23] R. May, G. Dandy, and H. Maier, "Review of input variable selection methods for artificial neural networks," in *Artificial Neural Networks—Methodological Advances and Biomedical Applications*. Rijeka, Croatia: InTech, 2011, pp. 19–44.

[24] R. Madhanagopal and R. Chandrasekaran, "Selecting appropriate variables for DEA using genetic algorithm (GA) search procedure," *Int. J. Data Envelop. Anal. Oper. Res.*, vol. 1, no. 2, pp. 28–33, 2014.

[25] M. Danishvar, A. Mousavi, and P. Sousa, "Event-clustering for improved real time data modelling and sensitivity analysis," in *Proc. IEEE Conf. Control Appl. (CCA)*, Antibes, France, Oct. 2014, pp. 1801–1806.

[26] M. Danishvar, A. Mousavi, P. Sousa, and R. Araujo, "Event-clustering for real-time data modeling," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Madison, WI, USA, 2013, pp. 362–367.

[27] J. R. King, "Machine-component grouping in production flow analysis: An approach using a rank order clustering algorithm," *Int. J. Prod. Res.*, vol. 18, no. 2, pp. 213–214, 1980.

[28] T. Ghosh and P. Dan, "Effective clustering method for group technology: A short communication," *e-J. Sci. Technol.*, vol. 6, pp. 23–28, 2011.

[29] M. P. Groover, *Automation, Production Systems, and Computer-Integrated Manufacturing*, 4th ed. Singapore: Pearson Educ., 2014.

[30] H. Farahmand *et al.*, "Hybrid mutation particle swarm optimisation method for available transfer capacity enhancement," *Int. J. Elect. Power Energy Syst.*, vol. 42, no. 1, pp. 240–249, 2008.

[31] M. B. Ali, R. Saidur, and M. S. Hossain, "A review on emission analysis in cement industries," *Renew. Sustain. Energy Rev.*, vol. 15, no. 5, pp. 2252–2261, 2011.

[32] N. Pardo, J. A. Moya, and A. Mercier, "Prospective on the energy efficiency and CO2 emissions in the EU cement industry," *Energy*, vol. 36, no. 5, pp. 3244–3254, 2011.

[33] J. Mendes, R. Seco, and R. Araujo, "Automatic extraction of the fuzzy control system for industrial processes," in *Proc. IEEE 16th Conf. Emerg. Technol. Factory Autom.*, Toulouse, France, 2011, pp. 1–8.

[34] C. J. Carmona, P. Gonzalez, M. J. del Jesus, and F. Herrera, "NMEEF-SD: Non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 958–970, Oct. 2010.

[35] C.-M. Lin and H.-Y. Li, "Intelligent hybrid control system design for antilock braking systems using self-organizing function-link fuzzy cerebellar model articulation controller," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 6, pp. 1044–1055, Dec. 2013.

[36] M. Prasad *et al.*, "Soft-boosted self-constructing neural fuzzy inference network," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 3, pp. 584–588, Mar. 2017.

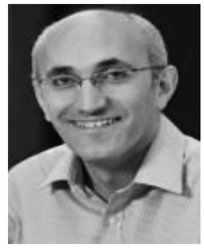
[37] W. Sheng, S. Chen, G. Xiao, J. Mao, and Y. Zheng, "A biometric key generation method based on semisupervised data clustering," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 9, pp. 1205–1217, 2015.

[38] A. Fahad *et al.*, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.

Morad Danishvar (M'14) received the B.Sc. and M.Sc. degrees in electronic engineering and the Ph.D. degree in systems engineering in 2015 from Brunel University London, London, U.K.

He is a Research Fellow with the System Engineering Research Group, Brunel University London. His current research interests include real-time systems modeling, middlewares, and data engineering. He is currently on the EU Horizon 2020 Z-Factor Project which focuses on zero defects in manufacturing systems.

Dr. Danishvar is a member of IET.



Alireza Mousavi (SM'13) received the B.Sc. degree in industrial engineering and the Ph.D. degree in system engineering in 1998 from Brunel University London, London, U.K.

He is a Reader of Systems Engineering and Computing with Brunel University London. His current research interests include smart supervisory control and data acquisition systems applied to real-time systems modeling and optimization. The key areas of application are in stochastic modeling, ontology alignment, and sensor networks.

Dr. Mousavi is a member of the Institute of Engineering and Technology.

Peter Broomhead received the B.Sc. and M.Sc. degrees in computer science.

He was a Production Technologist and a Lecturer in Computing with Brunel University London, London, U.K. His research interests included simulation modeling and analysis of manufacturing systems and parallel and practical applications in distributed computing. He recently passed away. May his soul rest in peace.