EventTracker for Real-time Sensitivity and input variable selection

Presented by: Alireza Mousavi, Brunel University

MOTIVATION

1. Systems need to respond to externally generated stimuli within a finite-specified period (i.e. **Real-Time)** → *Quick Response*

2. Producing time-critical accurate knowledge about the **state of the system** still remains a major challenge → *Innovative*

3. The knowledge is critical to the <u>safety</u> and the <u>integrity</u> of the operators and equipment → *Economical Gain*

4. Industrial systems' capability to capture data (SCADA) and flexibility to adjust to changing system requirements → *Existing Technological Capabilities*
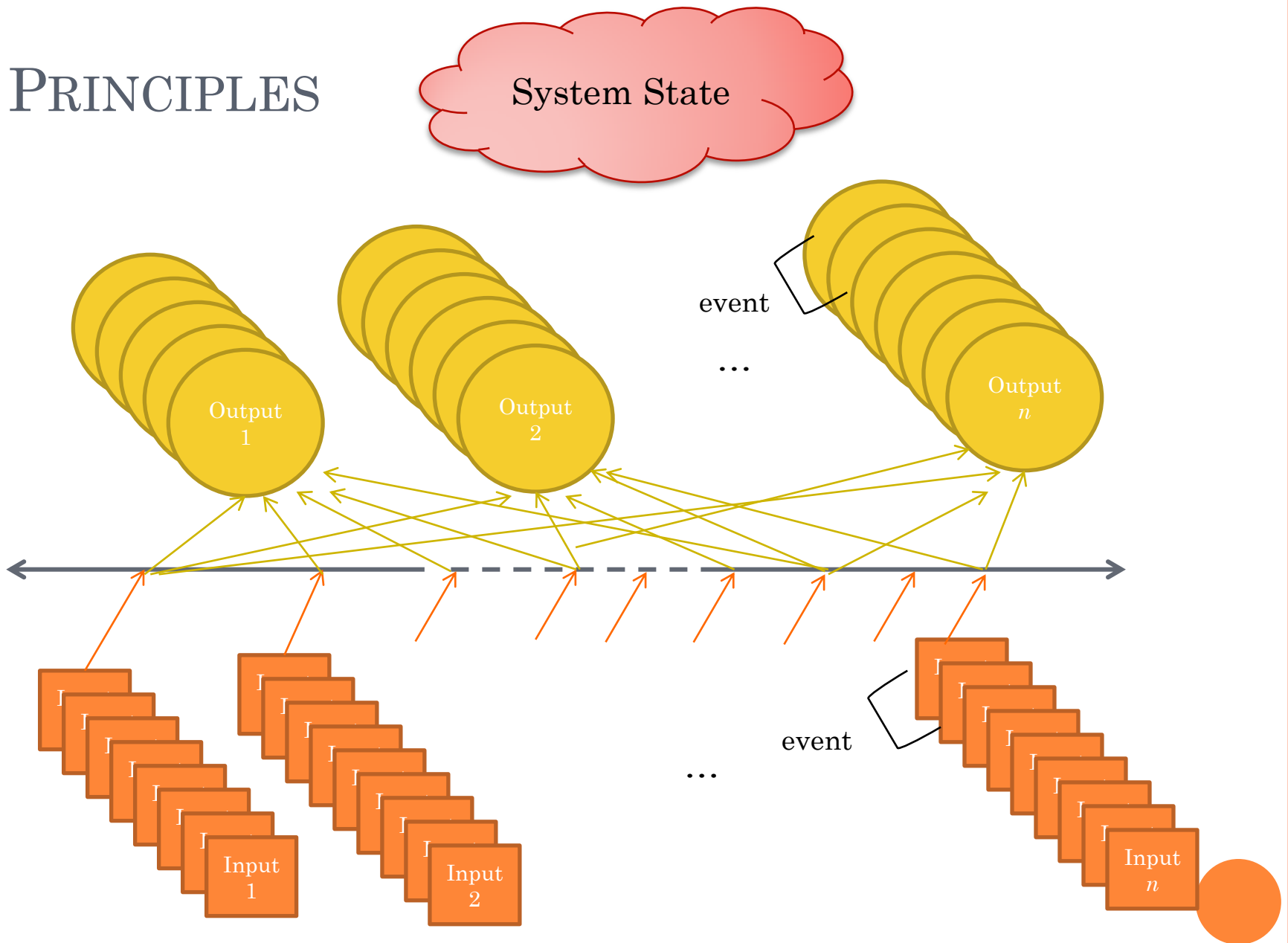
All the inputs have an effect on all outputs **unless proven otherwise**

SENSITIVITY ANALYSIS AND IVS TECHNIQUES

The analytical relationship between I/O:

- Determine the impact of changes in one variable in relation to others:

  - Analytical models (i.e. differential analysis, coupled/decoupled direct, and Green's Function)

  - Where no direct trans. physical models exist → statistical methods (i.e. Fourier Amplitude Sensitivity, Morris, Monte-Carlo, Latin hypercube → Rely on available historical data

  - The most comparable is the Entropy method → due to sporadic data – ***Not Really***
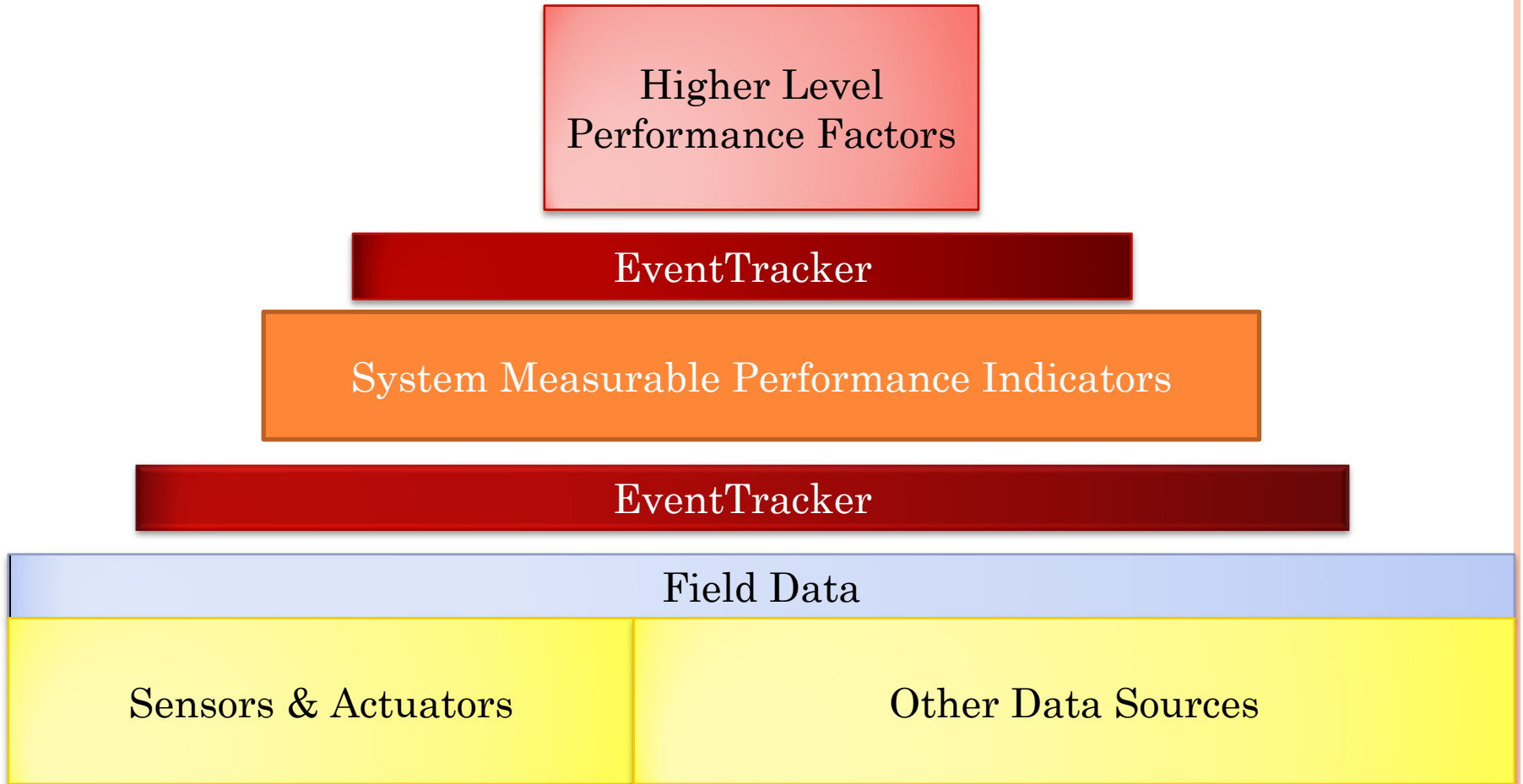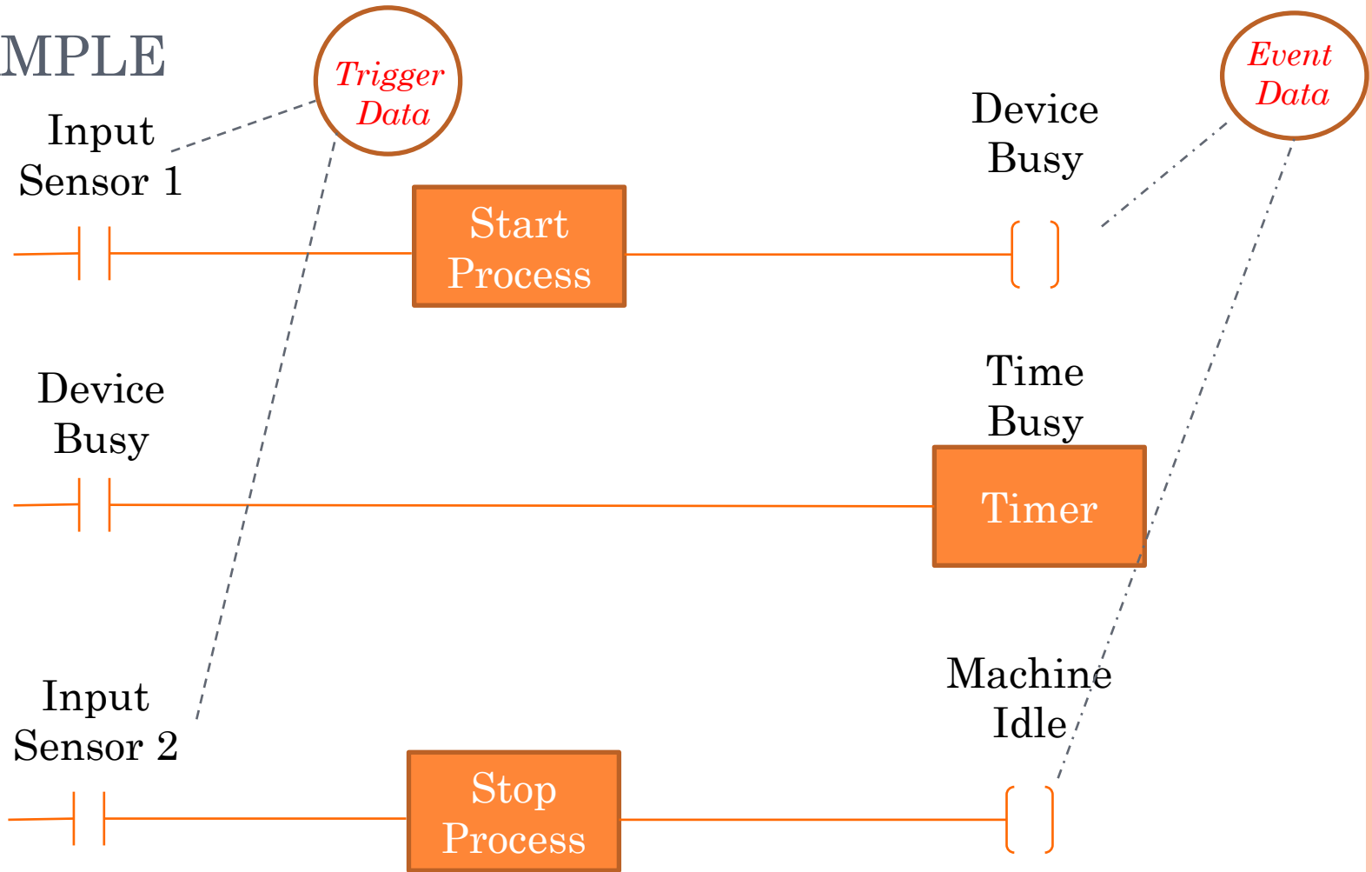
METHODOLOGY

- Lorenz's Butterfly Effect!

- Principle of cause-effect → Occurrence Matrix

- Causes (Trigger Data) – Actual Events (Event Data)

- Scan rates (pending the importance and volatility)

- Allows for discretisation (discrete event framework) → State Change is an Event

- The mapping of low-level system input (i.e. sensors) to high level system *Indicators*

STRUCTURE

Higher Level Performance Factors

EventTracker

System Measurable Performance Indicators

EventTracker

Field Data

Sensors & Actuators

Other Data Sources

EXAMPLE

Input
Sensor 1

Trigger
Data

Device
Busy

Event
Data

| Start Process |

| Timer |

Device
Busy

Time
Busy

Input
Sensor 2

Machine
Idle

| Stop Process |

$$Machine\ Utilisation = \frac{Machine\ Busy}{Machine\ Available}$$

Event
Data

Causal Relationship in Search Slot
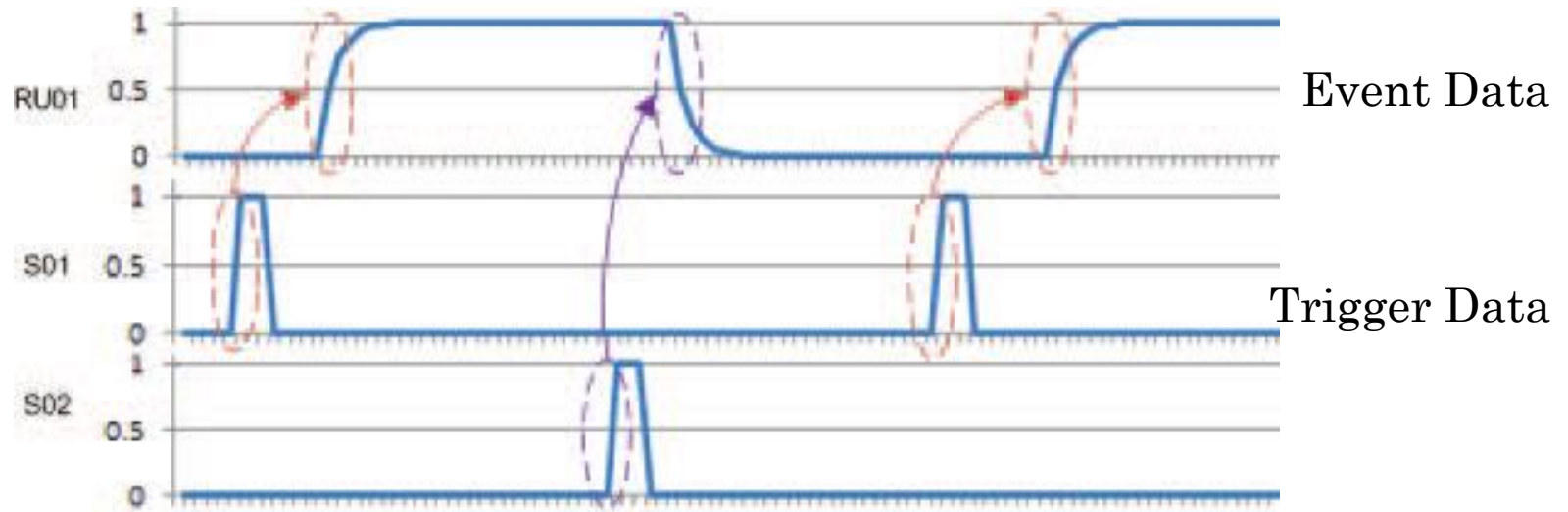
PARAMETERS

1. *Search slot (SS):* a fixed time in which batches of TD and ED are captured (system dependent)

2. *Analysis Span (AS):* the time period that a sensitivity analysis is conducted (AS contains a number of SS)

3. *Event Threshold (ET):* expressed as a percentage is the level of change in the value of an event determines whether a genuine event has occurred or not (e.g. change in flow rate)

4. *Trigger threshold (TT):* expressed as a percentage is the level of change in the value of a trigger determines whether a genuine event has occurred or not (e.g. sensor high/low)

THE ALGORITHM

1. Stepwise Scan & TT detection   2: Two-way Matching
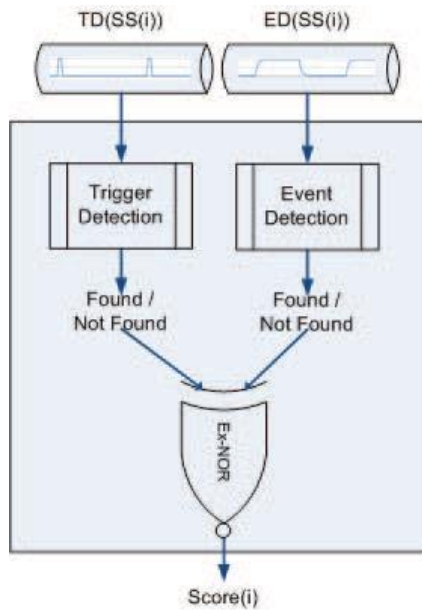
Input 1	Input 2	Output
0	0	+1
0	1	-1
1	0	-1
1	1	+1

THE ALGORITHM CONTINUED

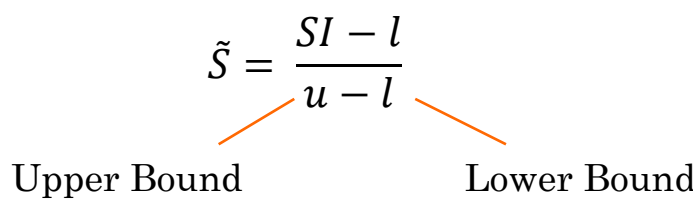3. Summation of two-way matching (Sensitivity Index at time *t*)

$$SI_{(t)} = \sum_{1}^{n} Search\ Slot\ Scores$$

Where *n is number of SS in a AS*

4. Normalisation:

$$\tilde{S} = \frac{SI - l}{u - l}$$

Upper Bound   Lower Bound

AN EXAMPLE FOR 10 SEARCH SLOT

1. Detected Events

2. Two-Way Matching Scores

Search Slot	0	1	2	3	4	5	6	7	8	9	10
ED	*	*		*		*	*	*		*	*
TD1			*			*			*	*	*
S1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1
SI1	-1	-2	-3	-4	-3	-2	-3	-4	-5	-4	-3
SIn1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TD2	*				*	*	*	*		*	
S2	1	-1	1	-1	-1	1	1	1	1	1	-1
SI2	1	0	1	0	-1	0	1	2	3	4	3
SIn2	1.00	1.00	1.00	0.67	0.33	0.33	0.67	0.75	0.80	0.80	0.75
TD3		*		*		*		*		*	
S3	-1	1	1	1	1	1	-1	1	1	1	-1
SI3	-1	0	1	2	3	4	3	4	5	6	5
SIn3	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ED is most sensitive to TD3 and least to TD1

3. Summation of two-way matches

4. Normalised sensitivity index

Search Slot	0	1	2	3	4	5	6	7	8	9	10
ED	*	*		*		*	*	*		*	*
TD1			*			*			*	*	*
S1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1
SI1	-1	-2	-3	-4	-3	-2	-3	-4	-5	-4	-3
SIn1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TD2	*				*	*	*	*		*	
S2	1	-1	1	-1	-1	1	1	1	1	1	-1
SI2	1	0	1	0	-1	0	1	2	3	4	3
SIn2	1.00	1.00	1.00	0.67	0.33	0.33	0.67	0.75	0.80	0.80	0.75
TD3		*		*		*		*		*	
S3	-1	1	1	1	1	1	-1	1	1	1	-1
SI3	-1	0	1	2	3	4	3	4	5	6	5
SIn3	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

l is the lowest SI in a column and u is the highest in the column:
Col 0: l =-1 , u = 1 and SI is the summation of all S1 to S10 as we move to next slots (red line)

$$Col0 \begin{vmatrix} SIn1 = \dfrac{-1-(-1)}{1-(-1)} = 0 \\[2mm] SIn2 = \dfrac{1-(-1)}{1-(1-1)} = 1 \\[2mm] SIn3 = \dfrac{-1-(-1)}{1-(-1)} = 0 \end{vmatrix}$$

$$Col1 \begin{vmatrix} SIn1 = \dfrac{-2-(-2)}{0-(-2)} = 0 \\[2mm] SIn2 = \dfrac{0-(-2)}{0-(-2)} = 1 \\[2mm] SIn3 = \dfrac{0-(-2)}{0-(-2)} = 1 \end{vmatrix}$$

$$\cdots$$

$$Col7 \begin{vmatrix} SIn1 = \dfrac{-4-(-4)}{4-(-4)} = 0 \\[2mm] SIn2 = \dfrac{2-(-4)}{4-(-4)} = 0.75 \\[2mm] SIn3 = \dfrac{4-(-4)}{4-(-4)} = 1 \end{vmatrix}$$

GRAPHICAL REPRESENTATION
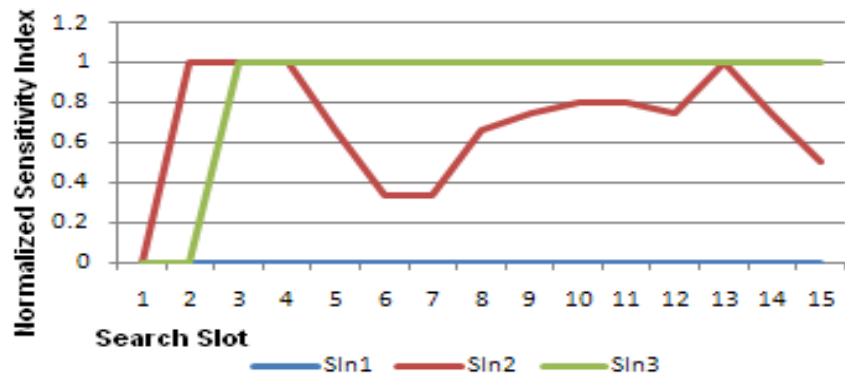
Normalized sensitivity indices as in Table II
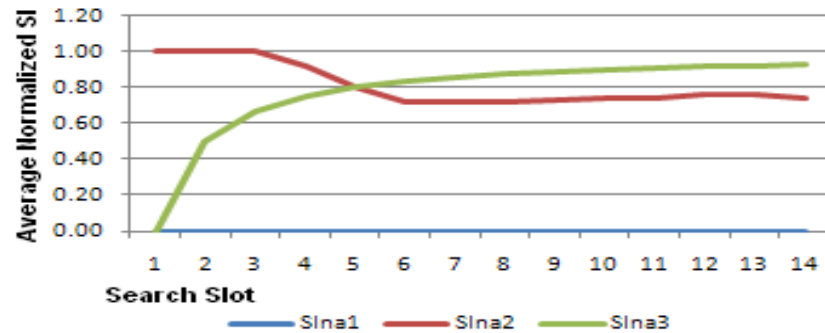
Averaged normalized sensitivity indices as in Table II

EFFICIENCY AND VALIDITY TESTS

The key objective of EventTracker is to record all events, but filter out the least important ones so that the optimum number of data points can be transferred to the data processing unit.

But highlight the potential interrelationships and not compromise quality

EFFICIENCY AND VALIDITY TESTS CONT.

Step 1: Establish the maximum number of least important trigger data and eliminate them from the analysis.

Step 2: Find the optimum SS, ET and TT values.

Step 3: Compare the results with a scenario where the total number of TDs have been included in the analysis (i.e. no reduction in data load).

- This last step was used to validate the filtering process in steps 1 and 2.

EFFICIENCY AND VALIDITY – INDUSTRIAL EXAMPLE

In order to determine the maximum number of the least important TD series, EventTracker was configured to read 28 TD series (see paper for details) and generate sensitivity indices for the four ED series. ET and TT values of 50% and a SS period of 5 seconds were used in this example.
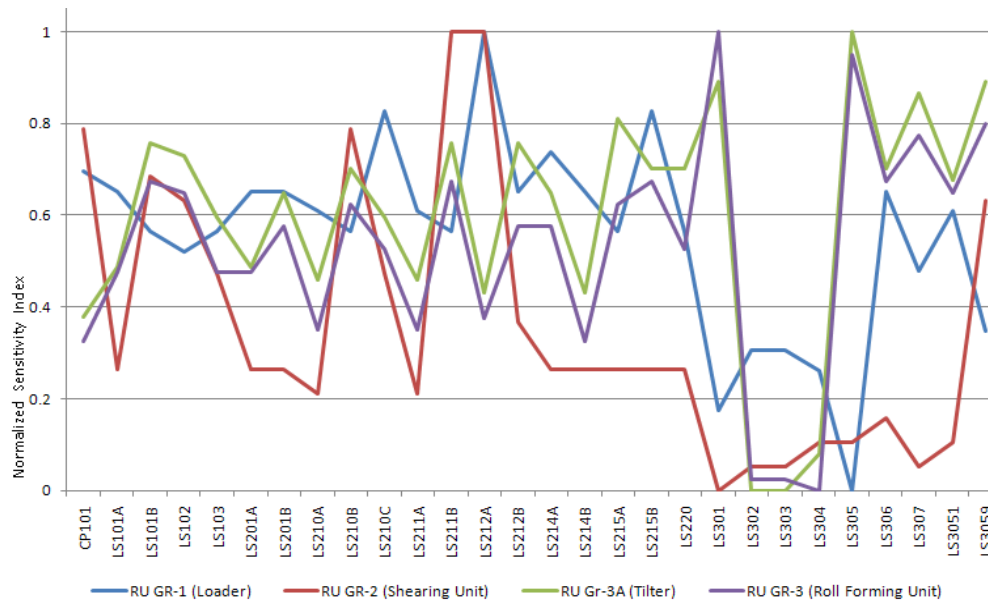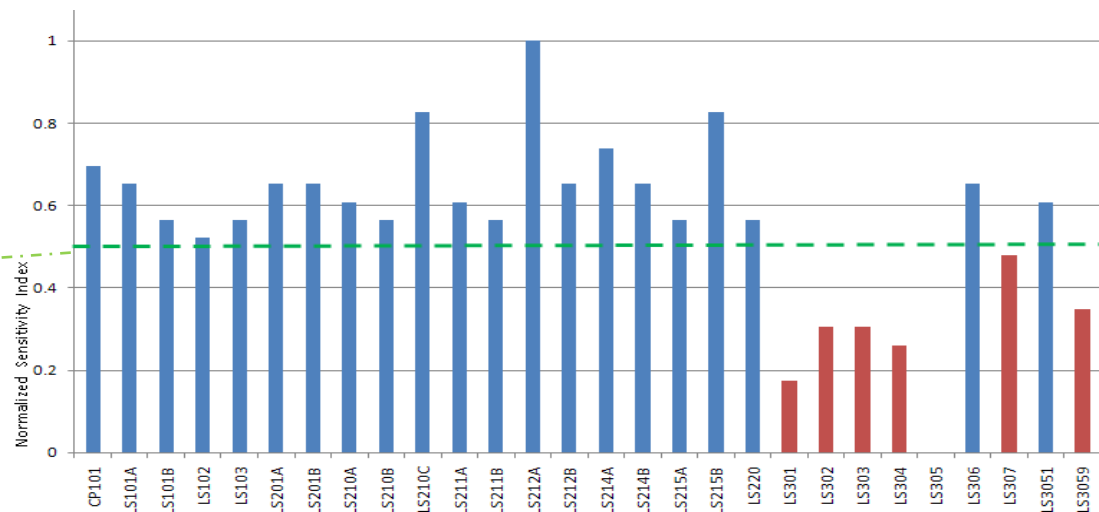
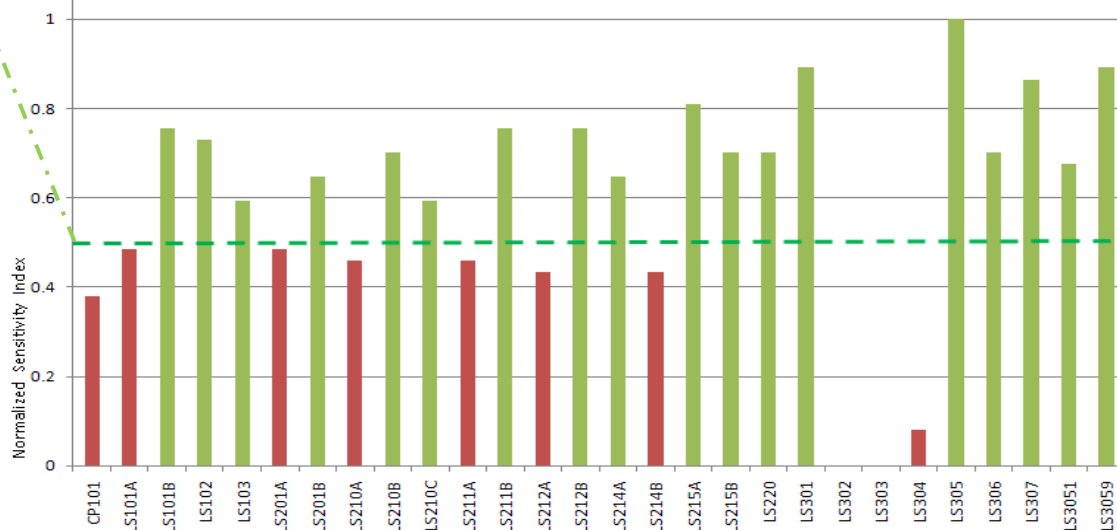EventTracker Sensitivity Indices of 4 EDs with respect to 28 TDs

ELIMINATION PROCESS

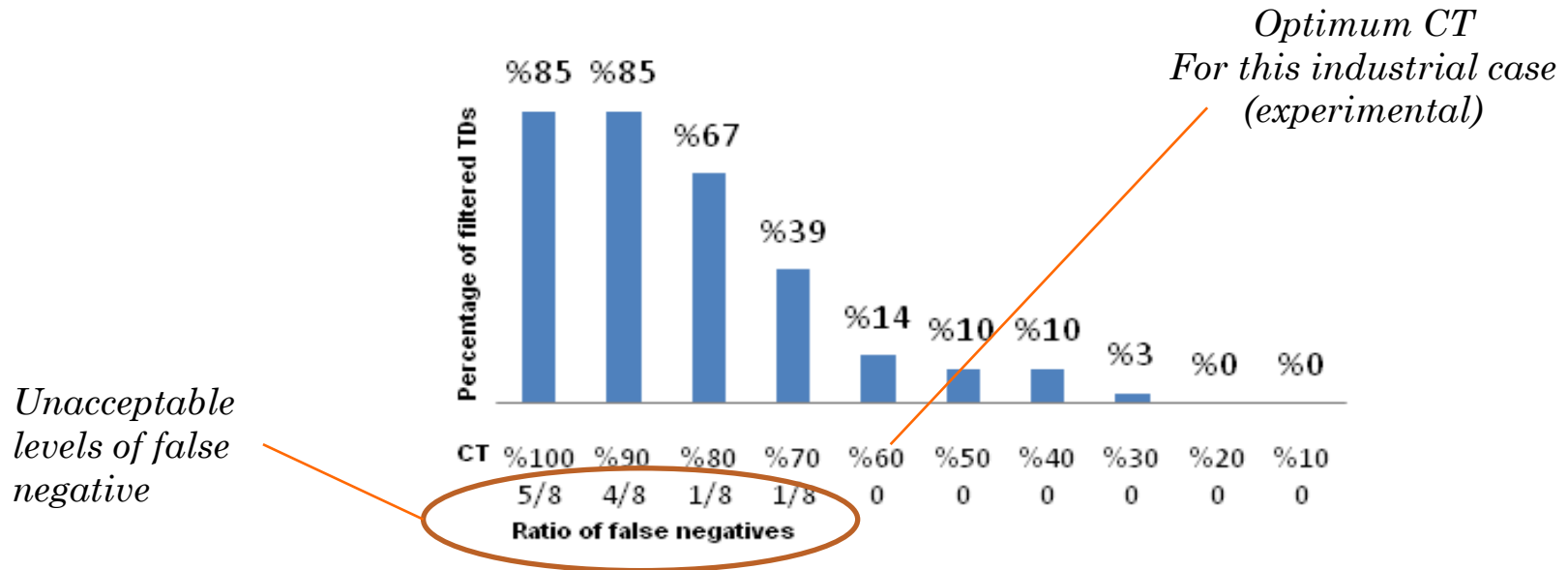Normalized sensitivity indices of RUGR1 (ED1) with respect to 28 TDs

Normalized sensitivity indices of RUGR3-A (ED3) with respect to 28 TDs
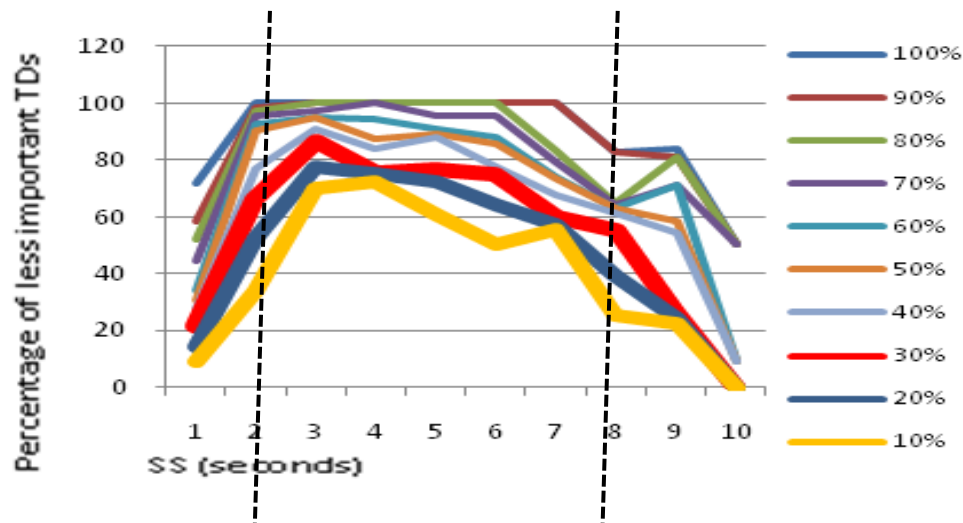
HANDLING FALSE NEGATIVES

Percentage of filtered TDs per CT and Ratio of False Negative

RESULT

- 4 Sensors eliminated no compromise on quality of data
- Inclusion of all input data for the analysis of system performance (improved calculation time by 50%)
- The optimal scan rate was calculated for this specific case to be between 2-8 seconds

Percentages of less important TDs per percentage of SS with different CR values

ANOTHER APPLICATION – IVS FOR DISASTER PREVENTION

- A case study in the Deep Drilling Industry (Deep Drilling Disaster Prevention)
- Demonstrates the ability of a recently proposed real-time input SA and VS technique
- Generating accurate disaster knowledge support system for the purpose of Disaster Prevention
- Unique solution for dealing with time-critical and unaware situations.
- The emphasis on — *"unawareness"* is to insist that in such cases there is no historical knowledge of the system behaviour.
- In this particular case a drilling machine drilling deep into a geophysical unknown terrain

Deep drilling

- 100 Variables input time series

- 10 Sensors

- 6 Systems States

- 100 *msec* Triggers

$$F_{Hook} = F_A + F_W \pm F_F + \varepsilon$$
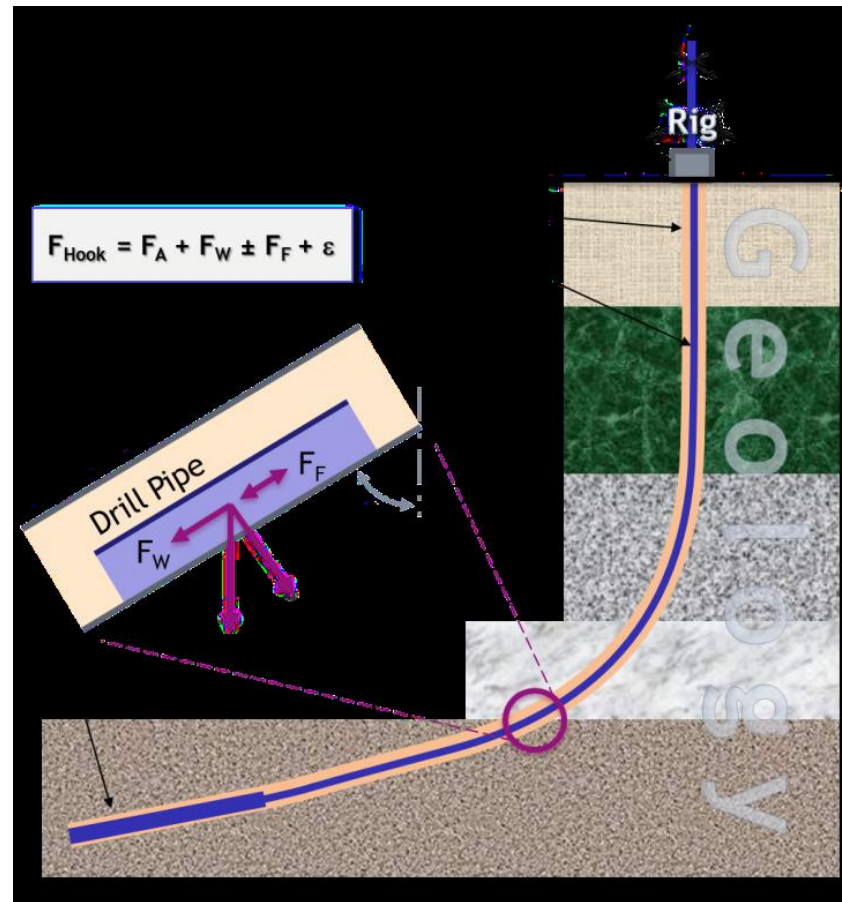
Sources

Tavakoli S., Mousavi A. and Polsad S. (2013), Input Variable Selection in Time-critical Knowledge Integration Applications: A Review, Analysis, and Recommendation Paper, *accepted in Advanced Engineering Informatics.* (June 2013).

Tavakoli S. and Mousavi A. (2013), "Event Tracking for Real-Time Unaware Sensitivity Analysis (EventTracker)", *IEEE Trans. On Knowledge and Data Engineering,* Vol:25, No. 2, pp. 348-359, DOI: 10.1109/TKDE.2011.240.