

Real-time algorithm for changes detection in depth of anesthesia signals

Blinded Manuscript

Received: date / Accepted: date

Abstract This paper presents a real-time algorithm for changes detection in depth of anesthesia signals. A Page-Hinkley test (PHT) with a forgetting mechanism (PHT-FM) was developed. The samples are weighted according to their "age" so that more importance is given to recent samples. This enables the detection of the changes with less time delay than if no forgetting factor was used. The performance of the PHT-FM was evaluated in a two-fold approach. First, the algorithm was run offline in DoA signals previously collected during general anesthesia, allowing the adjustment of the forgetting mechanism. Second, the PHT-FM was embedded in a real-time software and its performance was validated online in the surgery room. This was performed by asking the clinician to classify in real-time the changes as true positives, false positives or false negatives. The results show that 69% of the changes were classified as true positives, 26% as false positives, and 5% as false negatives. The true positives were also synchronized with changes in the hypnotic or analgesic rates made by the clinician. The contribution of this work has a high impact in the clinical practice since the PHT-FM alerts the clinician for changes in the anesthetic state of the patient, allowing a more prompt action. The results encourage the inclusion of the proposed PHT-FM in a real-time decision support system for routine use in the clinical practice.

Keywords Adaptive systems · Data flow analysis · Change detection algorithms · Dynamic behavior

1 Introduction

The automatic detection of changes in physiological signals has been topic of research for some years now [1, 19, 31]. This paper presents a real-time algorithm to automatically detect changes in depth of anesthesia (DoA) signals.

The majority of the indices to quantify the DoA e.g. index of consciousness (IoC) [11], auditory evoked potentials (AEP) [29], spectral entropy (SE) [30] and bispectral index (BIS) [8,9], range between a value close to 100, corresponding to fully awake state, and 0, corresponding to electrocortical activity suppression during a fully asleep and unconscious anesthetized state. In general anesthesia, values between 40 and 50 show to be ideal to perform surgical procedures [16,25]. However, during the time course of a surgery, this target range may change depending on the surgical protocols that are being used or on the overall state of the patient. A trained clinician judges the current observations and interprets them with sensitivity to the context, and in comparison with previous observations, to provide a warning of potential deterioration in the anesthetic state, or to change the target values of the monitored physiological signals. Naturally, the thresholds for the clinician to trigger an alarm are dynamically dependent on all the available information and clinical environment. The automatic and correct identification of a change that needs to be alarmed can be considered a detection and decision problem.

As for the DoA, this detection has paramount importance in the adaptation of the drug doses needed to achieve an optimal DoA level while avoiding undesirable side-effects [18,27]. Changes in DoA signals may occur due to either intrinsic or extrinsic factors. Since most of the DoA indices, namely the BIS, report the patients' level of hypnosis (related to unconsciousness), and analgesia (lack of pain) [20], these are the intrinsic factors. Intubation, incisions or other painful stimuli are the extrinsic factors.

1.1 The Change Detection Problem

When monitoring data it is fundamental to know if the received data comes from the distribution observed so far. In non-stationary scenarios, due to the dynamic nature of data, some properties of the problem might change over time, namely the target concept on which data is obtained may shift from time to time. When some properties of the problem change over time, "old" observations (reflecting the behavior of nature in the past) become irrelevant to the current state. These dynamic processes are challenging and need to be addressed with appropriate change detection algorithms.

Changes may occur due to modifications in the context of learning (caused by changes in hidden variables) or in the intrinsic properties of the observed variables. Usually, literature considers two types of drift. The term *Concept Shift* refers to abrupt changes (for example, the patterns of costumers' buying preferences that may change with seasons), while the term *Concept Drift* is associated to gradual changes in the target concept (for example, small faults in parts of an industrial process can modify the quality of the product). A *Concept Drift* is more difficult to detect, and, at least in the initial phases of the change, it may be confused with noise.

Recently, several methods capable of dealing with change detection have been proposed [3,4,12,13,15,26]. Drifting concepts are often handled by time

windows or weighted samples according to their "age" or utility. Weighted samples are based on the idea that the importance of a sample should decrease with time (references about this approach can be found in [13,14,17]). When a time window is used, at each time step, only the samples inside that window are relevant. With the aim of detecting concept changes, some indicators (e.g. performance measures, data distribution, properties of the data, etc.) can be monitored over time [5,12,23,26].

Assuming that samples are independent and randomly generated according to an unknown distribution, to assess if a concept is shifting over time, it is necessary to perform tests in order to determine if there is a change in the generating distribution. To be efficient, a change detection algorithm must be able to forget outdated data, be single pass and run in efficient space, allowing constant updates in time and memory. Therefore, the main challenge of algorithms for change detection is a tradeoff between the robustness in the presence of noise and sensitivity to concept changes. The presence of noise and outliers constitute the main difficulties to change detection algorithms since they may increase the number of false alarms. Furthermore, such kind of change detection tests should detect only true changes with high probability, establishing a trade-off between false negatives and false positives. Specifically, these algorithms should:

- Be able to detect and react to drift.
- Not exhibit miss detections.
- Be resilience to false alarms (detect a change in stationary environments).
- Require few samples to detect a change after the occurrence of one.

1.2 Change Detection in BIS Signals: Challenges and Contributions

In the standard clinical practice, when the behavior of the BIS signal changes, the clinician manually adjusts the drug dosages to control the anesthetic state of the patient. A quick reaction is crucial since a delay in this adjustment may compromise the well-being and general condition of the patient [18], possibly leading to some undesirable side-effects (such as awareness experience during surgery, post-operative nausea and vomiting and muscle aches).

The detection of initial phases of smooth changes are as challenging for the clinicians as it is for the changes detection algorithms. The enhancement of the Page-Hinkley test (PHT) with a forgetting mechanism aims to overcome this difficulty. By giving more importance to recent samples, the initial phase of the change will be reinforced and most easily detected.

On the other hand, sudden changes are easily observed and detected. Hence, they do not pose great difficulties to change detection algorithms. However, these changes are the most critical ones because the behavior of the DoA signal changes abruptly and quickly. In these situations, it is of foremost importance to bring the DoA signal back to the desired range or level after the occurrence of the change. This fact reinforces the advantage of using an automatic change detection algorithm to alert the clinician to act promptly.

The first contribution of this work is the development of an algorithm to detect changes in the DoA signals of patients undergoing surgery.

The second contribution is the offline evaluation of the proposed algorithm using BIS signals of patients undergoing abdominal surgery. The main purpose of this validation is the adjustment of the forgetting mechanism. Even though being the DoA index most widely used, there still exists some controversy around the benefits of the use of the BIS in extensive clinical practice [21]. The reason for the use of BIS signals to assess the performance of the proposed change detection algorithm is bi-fold. First, BIS is the DoA index used in the surgery rooms the authors have collaboration with. Second, insights gained with this assessment are useful to extend this change detection algorithm to other DoA indices.

The third contribution of this paper is the real-time evaluation and test of the proposed method in a clinical environment.

The remainder of this paper is organized as follows. Section 2 describes the methodology to detect changes in the data. Section 3.1 and 3.2 show the results from the offline and online evaluations, respectively. Concluding remarks and possibilities for further research are presented in section 5.

2 Methods

2.1 Classification of the changes detected by the PHT-FM

Both for the offline and the online database, the changes detected by the proposed change detection algorithm were classified by a single clinician. A True Positive (TP) corresponds to a change that actually occurred; a False Positive (FP) is a wrongly detected change (also known as a type I error and as a false alarm); a False Negative (FN) correspond to a change is not detected, when in fact there exists one (also known as a type II error and as a miss detection). The True Negatives (TN) are the points where the algorithm did not detect a change and there was indeed no change. In this work the number of TN was not assessed because the TN are all the other data points in the signal that were not classified as TP, FP or FN. The PHT-FM changes detections that were not classified by the clinician are marked as nC. In most of the cases these detections were not classified because the clinician was not available at the time the change occurred.

2.2 Clinical data collection

In the clinical practice, hypnotics and analgesics are the drugs that mainly influence the DoA. In the real cases used on this paper, the hypnotic propofol and analgesic remifentanyl were intravenously administrated to induce and maintain the patient's DoA, which was manually controlled by the clinician who changed the drug doses according to clinical requirements, using as reference the patient's vital signs, e.g. the blood pressures and the heart rate,

and the BIS values. The BIS, hemodynamic parameters and drug rates were recorded with a frequency of $1/5 \text{ s}^{-1}$.

2.2.1 Offline database

For the offline evaluation, data previously collected from 22 patients undergoing abdominal surgery was used. The patients were 60 ± 15 years old, 76.8 ± 17.7 kg and 13 female, and the surgeries had an average duration of 144 ± 74 minutes.

The PHT-FM was run on each case of the database and the changes were classified afterwards. The changes that were followed by a modification in the rate of propofol and/or remifentanyl were classified as TP. Similarly, if no modification in drugs rates occurred after the detection, the change was, in general, considered to be a FP. However, it should also be that, during the time course of a surgery, when informed by the surgeons about the possibly painful procedures that are about to be performed, the clinician often avoids a possible increase in the DoA index value by anticipating the administration of a higher dose of the hypnotic and/or analgesic. Some of the detected changes might hence be consequence of an increase in the administered drugs rates. Due to this, the changes that were detected by the PHT-FM were also considered TP if some modification in the propofol and/or remifentanyl administered rates occurred right before the change.

2.2.2 Online database

Due to the encouraging results that were obtained with the offline evaluation, the PHT-FM was implemented in the software GALENO: Computer Aided System for Modeling, Monitoring and Control in Anesthesia [24], and run online in the surgery room, during 78 general anesthesia episodes, for different types of surgeries. The patients were 56 ± 14 years old, 71.8 ± 16.4 kg and 70 female, and the surgeries had an average duration of 123 ± 66 minutes.

The clinician was advised every time a change was detected by the PHT-FM and asked to classify it as TP, FP, and FN. The major advantage of this online assessment is that this classification is done on-site, and, consequently, based on the current observations and actions performed around the time when the change was detected. In general, cases where the rate of propofol and/or remifentanyl was modified after the change was detected were classified as TP. Similarly, in most of the cases, if no action by the anesthesiologist was taken after the detection of a change, it was considered to be a FP. Exceptions were when, for example, the patient was subject to painful stimuli of limited duration, caused by a known source of stimulation, such as small incisions or patient repositioning. In those cases the clinician did not usually react by increasing the propofol and/or remifentanyl rate when advised that a change had occurred because, the DoA index value would decrease to its previous value as soon as the stimulus ended. Those changes were, however, classified as TP.

2.3 Change detection algorithm

In order to detect changes in the BIS signals, the PHT was selected among other change detection algorithms because is easy to implement and the time required to evaluate the samples are reduced. The changes detection in the signals must be performed on the flow and the low computational complexity of the PHT makes it appropriated to use in this context. The PHT is a sequential adaptation of the detection of an abrupt change in the average of a Gaussian signal [2] and is commonly used to online detect changes in signal processing (see [5, 22, 23]).

To detect increases and decreases in data, the PHT consists of running two tests in parallel. It considers two cumulative variables U_T and L_T , defined as the cumulated difference between the observed values and their mean until the current moment:

$$U_T = \sum_{t=1}^T (x_t - \bar{x}_T - \delta) \quad L_T = \sum_{t=1}^T (x_t - \bar{x}_T + \delta)$$

where $t = 1 \dots T$ represents the sample step, x_t is the variable value at time t , $\bar{x}_T = 1/T \sum_{t=1}^T x_t$ is the current mean at time T and δ corresponds to the magnitude of changes that are allowed. The minimum and maximum value of these variables are also computed: $m_T = \min(U_t, t = 1 \dots T)$ and $M_T = \max(L_t, t = 1 \dots T)$, respectively. As a final step, the PHT runs two tests in parallel, one to detect increases and another to detect decreases. The first monitors the difference between the cumulative variable (U_T) and its minimum (m_T) and the second monitors the difference between the cumulative variable (L_T) and its maximum value (M_T).

Intending to reduce the time delay of the changes detected and since in swift and evolving environments "old" data is usually less important than recent one, this method was enhanced with a forgetting mechanism (PHT-FM). The samples are weighted according to their "age" so the PHT will focus more on recent samples, detecting the changes with low delay time. The resulting tests are the following:

<p>For increase cases:</p> $U_0 = 0$ $U_T = \frac{T-1}{T} U_{T-1} + (x_T - \bar{x}_T - \delta)$ $m_T = \min(U_t, t = 1 \dots T)$ $PH_U = U_T - m_T$	<p>For decrease cases:</p> $L_0 = 0$ $L_T = \frac{T-1}{T} L_{T-1} + (x_T - \bar{x}_T + \delta)$ $M_T = \max(L_t, t = 1 \dots T)$ $PH_L = M_T - L_T$
---	---

In these equations, the forgetting mechanism is the weight of the variables U_{T-1} and L_{T-1} in the update process. Since the ratio $\frac{T-1}{T}$ increases with time, the recent samples have more importance in the update process than the older ones. With this forgetting mechanism, while assigning more importance to

recent observations, the algorithm will be able to earlier detect both abrupt (sudden) and gradual (slow) changes.

At every instant the two PH statistics (PH_U and PH_L) are monitored and a change is reported whenever one of them is above a given threshold λ . When a change is detected, all the variables and index of samples are reseted and a new test is reinitialized.

The parameter δ is highly dependent of the characteristics of the signal under study. The value of this parameter is chosen to avoid false detections due to noise, taking into account the magnitude of changes that are allowed and should not trigger an alarm. The change threshold parameter is chosen considering a tradeoff between admissible false alarm rates and delay time detections. Therefore, increasing λ the algorithm will entail fewer false alarms but might miss some true changes.

The proposed change detection algorithm relies on the difference between the observed value and its current average. Whenever the referred difference increases continuously, eventually exceeding the user predefined limit (λ), the algorithm detects a change in the recorded BIS signal. The psedo-code for the PHT-FM is the following:

Page-Hinkley test with Forgetting Mechanism

Inputs:

```

Require:  $x(t)$  /* observation at example  $t$  */
Require:  $\delta$  /* admissible change*/
Require:  $\lambda$  /* change threshold*/
Ensure:  $change \in \{TRUE, FALSE\}$ 
...
  /* Initialize the variables */
   $S(0) \leftarrow 0$ ;  $SR(0) \leftarrow 0$ ;
   $U(0) \leftarrow 0$ ;  $m \leftarrow 0$ ;
   $L(0) \leftarrow 0$ ;  $M \leftarrow 0$ 
  /* Update the error estimator */
   $S(t) \leftarrow S(t-1) + x(t)$ 
   $U(t) \leftarrow \frac{T-1}{T}U(t-1) + x(t) - \frac{S(t)}{t} - \delta$ 
   $L(t) \leftarrow \frac{T-1}{T}L(t-1) + x(t) - \frac{S(t)}{t} + \delta$ 
   $m(t) \leftarrow \min(m, U(t))$ 
   $M(t) \leftarrow \max(M, L(t))$ 
  if  $PH_U = U(t) - m \geq \lambda$  then
     $change \leftarrow TRUE$  /* increase */
  else
     $change \leftarrow FALSE$ 
  end if

```

```

if  $PH_L = M - L(t) \geq \lambda$  then
    change  $\leftarrow TRUE$  /* decrease */
else
    change  $\leftarrow FALSE$ 
end if
...

```

Figure 1 illustrates how PHT-FM works. The upper plot shows the initial phase of a BIS signal used in this study. As it can be observed, two changes occur around minutes 14 and 17 due to a decrease in the signal. The bottom figure represents the evolution of the statistic test PH_L and the detection threshold (λ). The PHT-FM statistic test captures both the decreases presented at this stage of the signal. The λ parameter should guarantee that the algorithm, while being resilient to false alarms, can detect and react to changes as soon as they occur, reducing the detection delay time. By controlling this detection threshold parameter, a tradeoff between the false positive alarms and the mis detections is established. Regarding both thresholds, to the best of our knowledge, it is not possible to automatically derive them through the characteristics of the signal under study.

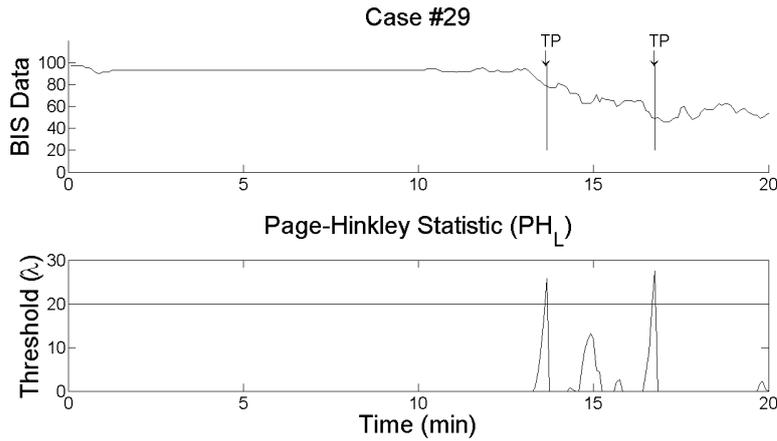


Fig. 1 The upper figure shows the initial phase of a BIS signal used in this study. The bottom figure represents the evolution of the PHT-FM statistic and the detection threshold λ .

2.3.1 Forgetting mechanism

A comparative analysis was performed in the database described in section 2.2.1 to assess the effect of using the forgetting mechanism. Figure 2 shows the mean delay time (in seconds) between accordant detections obtained by the

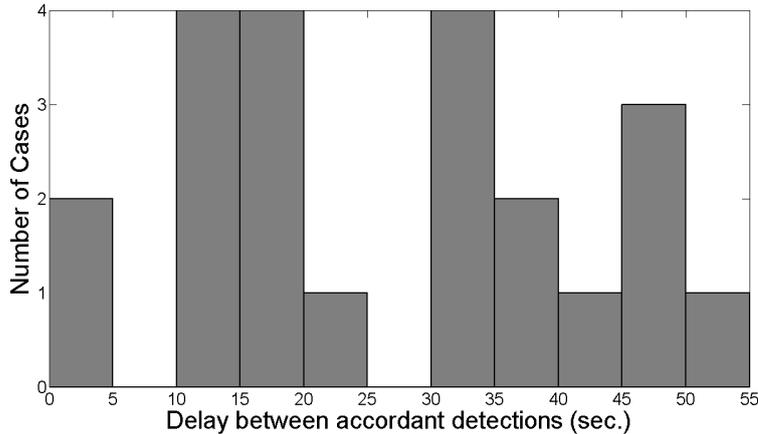


Fig. 2 Mean time delay (in seconds) between accordant detections obtained with the original PHT and with the enhanced PHT-FM, for all cases in the aforementioned database.

original PHT and by the PHT enhanced with a forgetting mechanism (PHT-FM), for all the cases in the database. When using the PHT-FM, in 50% of the cases, the clinician is advised with more than half a minute of advance in relation to the case where the PHT is used. This is the main advantage of the PHT-FM since reducing the delay time in detections gives more room to the clinician to decide based on that information. The result of this assessment supports the use of the proposed forgetting mechanism.

2.3.2 Algorithm input parameters

To adjust the algorithm input parameters, an analysis was previously conducted on 106 BIS signals collected from a patients under similar anesthetic procedures to the ones described in section 2.2 [7,6]. From that, the parameters λ and δ were set equal to 20 and 10, respectively. That was done considering the admissible false alarm rate and the magnitude of the permitted changes regarding the intrinsic characteristics of the signal under study. The sensibility of both parameters was also assessed by fixing one of the parameters and varying the other. The results obtained are shown in Figure 3. As it is clear in the upper plot of Fig. 3, where λ was equal to 20, values of δ below 10 give rise to an increasing number of FP. For values of δ above 10, the number of FN increases. The same behavior is found when δ is set to 10, and λ is made to vary between 15 and 25 (bottom plot of Fig. 3).

2.4 Evaluation metrics

Throughout this paper, the clinician's classifications of the changes detected by the proposed PHT-FM are considered as ground truth. Naturally, they are

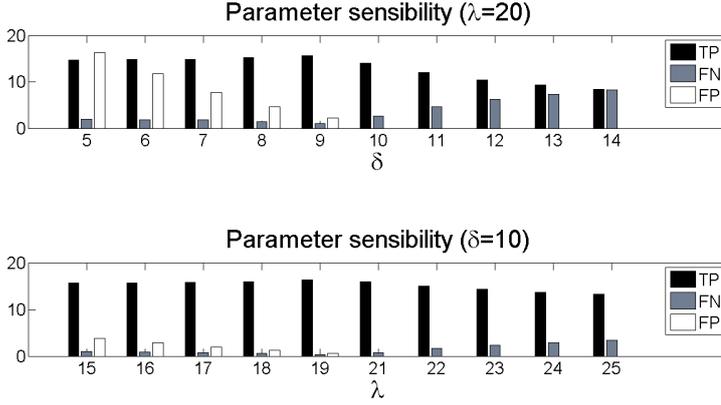


Fig. 3 Parameters sensibility.

highly dependent on the evaluation of a single clinician and are not 100% accurate. These validations allow computing quality metrics of the PHT-FM performance, such as *Precision* and *Recall* (also known as *Sensitivity*). The *Precision* measures the ratio between the correct detected changes (TP) and all the detected changes, TP+FP, and the *Recall* is defined as a ratio between the correct detected changes and all the occurred changes, TP+FN. For both quality metrics, the closer to 100% the most accurate is the change detection algorithm. Both metrics are closely related to the concepts of type I and type II errors: an algorithm with high recall has a low type II error rate, which means that it misses few changes detections. While an algorithm with high precision has a low type I error rate, which means that it is resilient to false alarms.

3 Results and discussion

3.1 Offline evaluation

Since the changes that occur in the BIS signal were not simulated, being instead consequence of changes coming from life operating settings, it was difficult to ascertain the exact times when the changes in the BIS signal occurred. Hence, a delay time evaluation could not be performed. A preliminary evaluation has however been performed scoring the offline algorithm detections. Regarding the quality measures, for all cases in the database a *Precision* of 87% and a *Recall* of 98% were obtained. Table 1 shows the confusion matrix obtained for the offline database. In spite of the high number of false alarms, this fact does not represent a major concern because clinician might be advised and then decide an action based upon his expertise and on the patient's vital

signals. Another important result of this evaluation is the low number of miss detections, which is an evidence that the PHT-FM is able to detect almost all the changes occurred in the BIS signals that induce the clinician to adjust the drug doses. It should also be pointed out that most part of the detections were classified as TP. These results supported the online implementation and evaluation of this algorithm, which results are presented in section 3.2.

Table 1 Confusion matrix obtained for the offline database.

		Real		Total
		Change	No Change	Total
Detected	Change	266	40	306
	No Change	4	X	4
Total		270	40	310

3.2 Online Evaluation

Figure 4 shows records of three patients from the database described in section 2.2.2. For each case, the upper plot shows the BIS signal and the detected changes by the PHT-FM (indicated by a vertical line and arrows pointing upwards if an increase in the BIS signal was detected and pointing downwards if a decrease in the BIS was detected).

As shown in Fig. 4 (a), the change around minute 6, consequence of the administration of the initial *bolus* of propofol is detected by the algorithm, as expected, and evaluated as TP by the clinician. The PHT-FM consistently detects the decreases of the BIS signal as a result of this propofol *bolus*, as it can also be observed later at minute 10, both validated as TP. Although a TP validation by the clinician, around minute 50 the algorithm detects two ascendant changes that were neither consequences nor followed by any clinical action. Around minute 65, the clinician validated as FP an algorithm detection. These situations intend to illustrate the difficulties that the problem under study poses to the development of change detection algorithms, namely the false positive detections due to noise present in the BIS signals. Later, around minute 70 a detection of an increase in the BIS, marked as TP, followed by the administration of a propofol *bolus* by the clinician is noticeable. As expected, after this *bolus* the BIS decreases which is detected by the PHT-FM (despite the first detected change not having been classified by the clinician, the second one is validated as TP and both are clearly a result of the *bolus* administration). This is one example where the online use of this algorithm may be advantageous: advised by the algorithm of this increase the clinician could act more promptly. The latest changes detections, classified as TP, were result of the end of drugs administration.

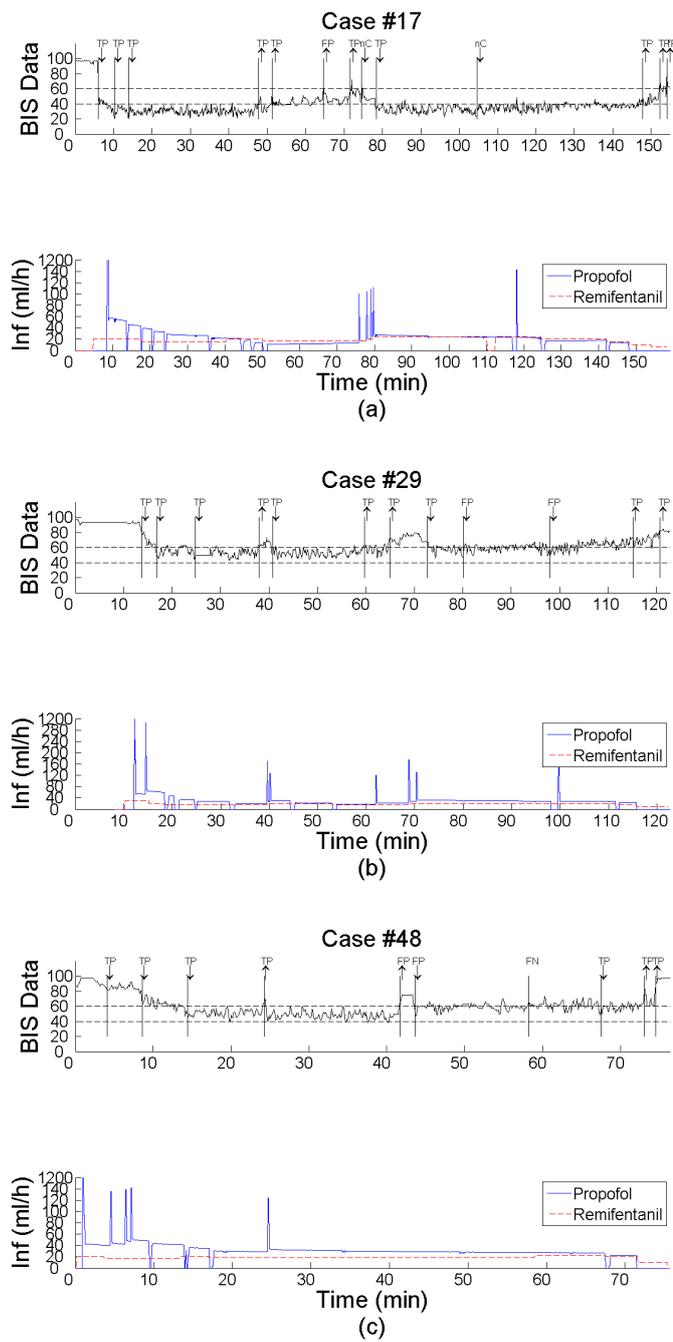


Fig. 4 Example of the detected changes in a BIS signal of three clinical cases in the database. The upper plots show the BIS signal, the detected changes (indicated by a vertical line and arrows) and the clinician's evaluation. The middle and bottom plots show the propofol and remifentanyl doses (ml/h).

Figure 4 (b) presents another clinical case. After the administration of the initial *bolus* of propofol the PHT-FM detected two decreases and later (around minute 25) another one as the result of the accommodation of the BIS to this dose (all classified by the clinician as TP). After a stable period with values within the range 40 to 60, the BIS signal reveals an increase which was detected by the algorithm. This change, evaluated as TP, contributed to an administration of a propofol *bolus* by the clinician. It is also noticeable, as a result of this administration, a decrease detection evaluated as TP. From minute 40 to minute 60 the BIS signal remains stable around a mean value of 45. Around minute 60 the algorithm detects another increase, classified as TP leading to an administration of a propofol *bolus*. However, this dosage was not enough to reduce the BIS signal which remain with an increasing behavior that was identified by the PHT-FM around minute 65. As a consequence, to avoid that the BIS increases, the clinician administrated another propofol *bolus*. After these administrations, the BIS recovered to the clinical reference range which was also detected by the PHT-FM (around minute 75). Later, it can be observed two false positive detections (the algorithm alarmed these changes without evident existence of any). It should be noted that the noisy level of these signals poses difficulties to this algorithm and often noise can be confused with initial phases of a change, alarming a change when the signal remains stable and raising the rate of false positives. The latest algorithm's detections (validated as TP) are a consequence of the end of drugs administration due to the end of the surgery.

Figure 4 (c) shows a clinical case where the PHT-FM missed a change, identified as FN (around minute 60). The first three detected decreases, classified as TP by the clinician, are the result of the accommodation of the BIS to the initial dose of propofol. Around minute 25 the PHT-FM detects an increase in the BIS. The administration of a propofol *bolus* follows this increase in order to maintain the BIS in the predefined target window. A rise in the BIS around minute 40 is also detected by the algorithm and succeeded by decrease in the signal despite any clinical action. These two detections were result of the incision and did not report a relevant variation in the BIS behavior and as a consequence were classified as FP. Around minute 70, a TP detection led to a decrease in the propofol administration. Approaching the end of the surgery, the drugs administration was set off and the algorithm, as desirable, detected two increases in the BIS signal (both validated as TP).

Table 2 Confusion matrix.

		Real		Total
		Change	No Change	Total
Detected	Change	660	255	915
	No Change	48	X	48
Total		708	255	963

Likewise the offline analysis, the *Precision* and the *Recall* were computed for all cases in this database, obtaining 72% and 93%, respectively. The high accuracy achieved by the proposed PHT-FM support its inclusion in real-time decision support system for routine use in the clinical practice. Table 2 shows the evaluation metrics corresponding to the online clinician validation. It must be stressed out that the number of false alarms are not a major concern for this problem. A false alarm will alert the clinician, who decides if one action should be taken, based upon his expertise and on the patient's vital signals. The miss detections might be more problematic. If the PHT-FM misses a change that occurred in the DoA signal, and if the clinician could not detect it by looking at the available monitors, the drugs dosages will not be adjusted in order to provide more comfort to the patient. Both situations, can disturb and/or potentially mislead the clinician but does not cause immediate wrong action.

Since the change detection algorithm is implemented online is not possible to avoid the presence of noise and sensor faults. Therefore, it should be pointed out, that the clinician's validations were only taken into account if the quality of the BIS signal was greater or equal to 50%. Along with this, some of the detected changes by the algorithm were not classified by the clinician (those were not considered in the these measurements).

4 Limitations of the proposed decision support system

The presence of noise and outliers in the collected signals of BIS and the outliers in those are the main limitations of the proposed PHT-FM to incorporate in a decision support system.

The corruption of the BIS signals by noise is as challenging for the application domain as for the change detection algorithms. For the application domain, the noise can confuse the clinician in the observation of the signal behavior, disturbing and potentially confusing his judgments. While for change detection algorithms, the noise can be easily confused with a initial phase of a change. The noise in the signals could be addressed by performing some filter strategy before presenting the signal to the change detection algorithm. However, this was not scope of this work, but it is a task to which is it necessary to devote some effort in future research. Figure 5 shows a typical BIS signal collected in the surgery room, illustrating the level of noise present in the BIS signals. Specially, after minute 50, along with decreases and increases in the BIS behavior, it is also noticeable the noise present reflected in the high variation associated with those.

Outliers and sensor faults also pose difficulties to both domains and were not addressed in this work. Sensor faults, which are impossible to avoid, often causes outliers and/or missed data.

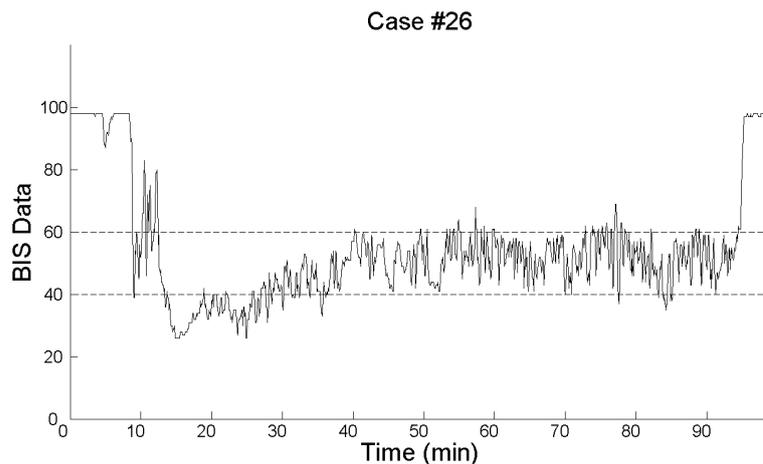


Fig. 5 Example of a BIS collected in the surgery room with a typical noise level.

5 Conclusions

The Page-Hinkley Test with a forgetting mechanism (PHT-FM) was implemented and tuned to detect changes in BIS signals from patients under general anesthesia. The developed PHT-FM algorithm consistently reveals the increasing and decreasing behaviors of the BIS signals under study.

The online evaluations obtained so far support the incorporation of this change detection algorithm in a robust and reliable online decision support system based on DoA monitoring for general anesthesia procedures.

It should be noted that the environment of the application and the specific features of the BIS signal, namely the high level noise present in the measurements, point to further improvements of this detection algorithm. The development of a dedicate online filter to smooth the BIS signals is a task to be addressed in the near future to enrich the detection algorithm results.

Due to lack of reliable sensors to directly and quantitatively measure the level of analgesia, and considering the controversy around the use of BIS to quantify the patients' DoA, further research on other vital signals such as the electrocardiogram, blood pressure and heart rate should be carried out to improve the decision support system. Results should naturally also be correlated with the ones obtained using BIS signals, or other DoA signals, since the extension of the PHT-FM to them is straightforward.

6 Acknowledgments

The work of Raquel Sebastião and Margarida M. Silva is supported by FCT - Fundação para a Ciência e a Tecnologia - under the PhD Grants SFRH/BD/41569/2007 and SFRH/BD/60973/2009, respectively.

This work is also funded by the ERDF through the Programme COMPETE and by FCT through - projects GALENO - Modeling and Control for Personalized Drug Administration (PTDC/SAU-BEB/103667/2008) and KDUDS - Knowledge Discovery from Ubiquitous Data Streams (PTDC/EIA-EIA/098355/2008).

The authors gratefully acknowledge Dr. Simão Esteves and his colleagues from Hospital Geral de Santo António (HGSA), Centro Hospitalar do Porto, Portugal and Dr. Manuel Seabra and his colleagues from Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Portugal, for their participation in the data collection and collaboration in the results analysis.

References

1. Ansermino J.M., .: An Evaluation of a Novel Software Tool for Detecting Changes in Physiological Monitoring *Anesthesia & Analgesia*, vol:108(3), pages 873–880 (2009).
2. Basseville M., Nikiforov I.: Detection of abrupt changes: theory and applications. Prentice-Hall Inc, Englewood Cliffs, N.J. (1987).
3. A. Bifet and R. Gavaldà. Adaptive Learning from Evolving Data Streams. *Proceedings of the 78th International Symposium on Intelligent Data Analysis: Advances in Intelligent Data Analysis VIII*, pages 249–260 (2009).
4. Gama J., *et al.*: Learning with drift detection. In A. L. C. Bazzan and S. Labidi (Eds.), *Advances in Artificial Intelligence - SBIA 2004*, vol. 3171, *Lecture Notes in Computer Science*, pages 286–295 (2004).
5. Gama J., *et al.*: Issues in evaluation of stream learning algorithms. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 329–338 (2009).
6. Gambús P.L., *et al.*: Modeling the effect of propofol and remifentanil combinations for sedation-analgesia in endoscopic procedures using an Adaptive Neuro Fuzzy Inference System (ANFIS). *Anesthesia & Analgesia*, vol:112(2), pages 331–339 (2011).
7. Gambús P.L., *et al.*: Modelling the interaction of propofol and remifentanil by means of an adaptive neuro fuzzy inference system (ANFIS). *European Journal of Anaesthesiology*, vol:23, pages 123 (2006).
8. Gan T.J., *et al.*: Bispectral index monitoring allows faster emergence and improved recovery from propofol, alfentanil and nitrous oxide anaesthesia. *Anesthesiology*, vol. 87(4), pages 808–15 (1997).
9. Glass P.S., *et al.*: Bispectral analysis measures sedation and memory effects of propofol, midazolam, isoflurane, and alfentanil in healthy volunteers. *Anesthesiology*, vol. 86(4), pages 836–47 (1997).
10. Hartl C., *et al.*: Change point detection and meta-bandits for online learning in dynamic environments. In *Conférence Francophone sur l'apprentissage automatique*, pages 237-250, Cepadues (2007).
11. Jensen E.W., *et al.*: Validation of the Index of Consciousness (IoC) during sedation/analgesia for ultrasonographic endoscopy. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5552–5555 (2008).
12. Kifer D., *et al.*: Detecting change in data streams. *Proceedings of the Thirtieth International Conference on Very Large Data Bases*, pages 180–191 (2004).
13. Klinkenberg R.: Learning drifting concepts: Example selection vs. example weighting. *Intelligent Data Analysis*, vol:8(3), pages 281–300 (2004).
14. R. Klinkenberg and T. Joachims. Detecting concept drift with support vector machines. *Proceedings of ICML-00, 17th International Conference on Machine Learning*, pages 487–494 (2000).
15. Koychev I.: Gradual forgetting for adaptation to concept drift. *Proceedings of ECAI Workshop Current Issues in Spatio-Temporal Reasoning*, pages 101–106 (2000).

16. Luginbuhl M., *et al.*: Detection of awareness with the bispectral index: two case reports. *Anesthesiology*, vol:96(1), pages 241–243 (2002).
17. M. Maloof and R. Michalski. Selecting examples for partial memory learning. *Machine Learning*, vol:41, pages 27–52 (2000).
18. Mashour G., *et al.*: A novel electronic algorithm for detecting potentially insufficient anesthesia: implications for the prevention of intraoperative awareness. *Journal of Clinical Monitoring and Computing*, vol:23(5), pages 273–277 (2009).
19. Melek W.W., *et al.*: Comparison of trend detection algorithms in the analysis of physiological time-series data. *IEEE Transactions on Biomedical Engineering*, vol:52(4), pages 639–651 (2005).
20. Minto C.F., *et al.*: Response surface model for anesthetic drug interactions *Anesthesiology*, vol:92, pages 1603–13 (2000).
21. Monk T.G., *et al.*: Does depth of anesthesia monitoring improve postoperative outcomes? *Current Opinion in Anesthesiology*, vol:24, pages 665–669 (2011).
22. Mouss H., *et al.*. Test of Page-Hinkley, an approach for fault detection in an agro-alimentary production system. *Proceedings of the 5th Asian Control Conference*, pages 815–818 (2004).
23. Page E.S.: Continuous inspection schemes. *Biometrika*, vol:41(1-2), pages 100–115 (1954).
24. Paz L.A., *et al.*. Integrated design system for monitoring digital processing and control in anesthesia. American Society of Anesthesiologists Annual Meeting 2011, abstract A1169 (2011).
25. Rampil I.J.: A primer for EEG signal processing in anesthesia. *Anesthesiology*, vol:89(4), pages 980–1002 (1998).
26. Sebastião R., *et al.*: Monitoring incremental histogram distribution for change detection in data streams. In Gaber, M.M., Vatsavai, R.R., Omitaomu, O.A., Gama, J., Chawla, N.V., Ganguly, A.R. (eds.), *Knowledge Discovery from Sensor Data*, Volume 5840 of *Lecture Notes on Computer Science*, pages 25–42 (2010).
27. Selbst S.M.: Adverse sedation events in pediatrics: a critical incident analysis of contributing factors. *Pediatrics*, vol:105(4), pages 864–865 (2000).
28. Silva M.M., *et al.*: Total mass TCI driven by parametric estimation. *Proceedings of the IEEE Mediterranean Conference on Control and Automation*, pages 1149–1154 (2009).
29. Struys M.M., *et al.*: Performance of the ARX-derived auditory evoked potential index as an indicator of anesthetic depth: a comparison with bispectral index and hemodynamic measures during propofol administration. *Anesthesiology*, vol:96(4), pages 803–816 (2002).
30. Viertiö-Oja H., *et al.*: Description of the entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module. *Acta Anaesthesiol Scand*, vol:48(2), pages 154–161 (2004).
31. Yang P., *et al.*: Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE Transactions on Biomedical Engineering*, vol:53(11), pages 2211–2219 (2006).