

Multiparametric modeling of the ineffective efforts in assisted ventilation within an ICU

I. G. Chouvarda¹ · D. Babalis² · V. Papaioannou³ · N. Maglaveras^{1,4} · D. Georgopoulos²

Received: 15 April 2014 / Accepted: 4 June 2015 / Published online: 17 June 2015
© International Federation for Medical and Biological Engineering 2015

Abstract In the context of assisted ventilation in ICU, it is of vital importance to keep a high synchronization between the patient's attempt to breath and the assisted ventilation event, so that the patient receives the ventilation support requested. In this work, experimental equipment is employed, which allows for unobtrusive and continuous monitoring of a multiple relevant bioparameters. These are meant to guide the medical professionals in appropriately adapting the treatment and fine-tune the ventilation. However, synchronization phenomena of different origin (neurological, mechanical, ventilation parameters) may occur, which vary among patients, and during the course of monitoring of a single patient, the timely recognition of which is challenging even for experts. The dynamics and complex causal relations among bioparameters and the ventilation synchronization are not well studied. The purpose of this work is to elaborate on a methodology toward modeling the

ventilation synchronization failures based on the evolution of monitored bioparameters. Principal component analysis is employed for the transformation into a small number of features and the investigation of repeating patterns and clusters within measurements. Using these features, nonlinear prediction models based on support vector machines regression are explored, in terms of what past knowledge is required and what is the future horizon that can be predicted. The proposed model shows good correlation (over 0.74) with the actual outputs, constituting an encouraging step toward understanding of ICU ventilation dynamic phenomena.

Keywords ICU · Assisted ventilation · SVM regression · Nonlinear model · PCA cluster

Electronic supplementary material The online version of this article (doi:10.1007/s11517-015-1328-1) contains supplementary material, which is available to authorized users.

✉ I. G. Chouvarda
ioannach@auth.gr

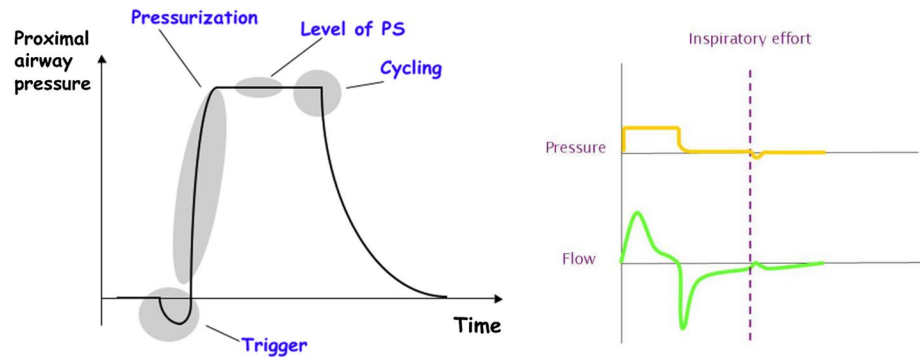
- ¹ Lab of Medical Informatics, School of Medicine, Faculty of Health Sciences, Aristotle University, PB 323, Thessaloniki 54124, Greece
- ² Department of Intensive Care Medicine, Medical School, University Hospital of Heraklion, University of Crete, Heraklion, Crete, Greece
- ³ Department of Intensive Care Medicine, Alexandroupolis University Hospital, Democritus University of Thrace, Dragana, Greece
- ⁴ Institute of Applied Biosciences, CERTH, Thessaloniki, Greece

1 Introduction

The intensive care unit (ICU) is a challenging environment, as a multitude of patient biosignals is continuously monitored. These physiological and intervention-related data, continuously displayed to experts, have to be combined with clinical and laboratory data in order to early recognize unusual, unstable and deteriorating or improving trends, alarming events, and adapt interventions accordingly.

It has to be taken into account that a human, even expert, would find it difficult to interpret more than 3–4 variables at any given time. Thus, the human factor steps in with inadequacy in situations where sound critical decisions cannot simply be made without risks unless clinicians are assisted by ICT tools. These critical expectations raise the need to support clinicians in analyzing and interpreting this multitude of data with tools which can handle fast changing multi-dimensional functions of variables and display useful

Fig. 1 Assisted ventilation. **a** Mechanical support, as expressed by pressure, follows patient effort both in terms of timing and magnitude, so patient successfully triggers the ventilator, **b** problematic patient–ventilator coupling. The inspiratory effort does not trigger the ventilator and results in an ineffective inspiratory effort



information in an easy-to-understand manner, which will facilitate fast and timely decisions. Currently, most ICUs in Europe are not necessarily equipped with these facilities to critically save lives.

Along these lines, the ultimate aim of this work is to set the basis for the management and analysis of large and complex ICU data. In this respect, a series of engineering challenges emerge in the ICU setting, including (a) preprocessing of signals to verify data and reduce false alarms; (b) processing of multiparameter data to continuously track physiological states and early detect critical events, including physiological and ventilation-related events such as sepsis [15, 16], significant hypotension and tissue hypoxia development, readiness to liberate from mechanical ventilation, or changes in ventilation synchronization; and (c) interpretation of findings according to pathophysiological models of critical illness. A worth-mentioning effort in the direction of ICU–DSS is Artemis [2], a platform tested in neonatal ICU that supports automated or clinician-driven knowledge discovery of new relationships between physiological data stream events and latent medical conditions, as well as refinement of existing analytics.

In the context of mechanical ventilation, multiparametric signal processing has presented some success, for example, in predicting successful weaning from mechanical ventilation (MV) [15]. Other efforts included decision support in ICU ventilation actions and the control of mechanical ventilation. An ICU decision support system, based on fuzzy logic and expert knowledge, has been proposed [12] concerning mechanical ventilation options with respect to different disease states. In [9], computerized physiological models and utility/penalty functions were employed as separate factors of the system. The case of assisted (as opposed to controlled) ventilation is now receiving more attention, due to the evidence of benefits [13]. In controlled ventilation, there is no need for patient's respiratory effort, nor for patient–ventilator coupling, as ventilation is forced. In assisted ventilation, it is patient effort that triggers the ventilator and there is need for patient–ventilator coupling. In some cases of triggering, patient's effort does not trigger

the ventilator at all. The basic concept in assisted ventilation is depicted in Fig. 1.

An emerging problem in this context that deserves more attention is the analysis of ventilation synchronization (between patient breath attempt and ventilation event), and the respective continuous prediction of synchronization failure, i.e., dysynchrony, based on the sequence of physiological and ventilation-related events under monitoring.

The problem of ventilation synchronization has been reported before, for example, as regards incidences of ineffective triggering [8]. Studies have shown that patients with frequent ineffective patient–ventilator synchronization, i.e., ineffective efforts (IEs), have worse outcome, i.e., mortality and duration of mechanical ventilation (MV). Ineffective triggering (IT) is common, but factors affecting IT vary considerably, among factors related to the patient condition (secretions, suctioning, atelectasis, fluid status, VAP, sepsis, patient positioning, coma, delirium, sedatives) and factors related to the ventilation system (ventilation mode & settings).

As previously reported [1], most patients have small (5 min) periods with high presence of IEs, sometimes with events highly concentrated in time. However, up to now, it is yet to be answered whether the cumulative effect of IE exposure (IEs with respect to breaths in 24 h) or the temporal patterns of IEs (periods of high IEs in 24 h) play a different role. Ineffective triggering of the ventilator is frequent but highly variable among patients and during the course of mechanical support for each patient. Studying these variable conditions throughout the course of MV, in a real life setup, presents a real challenge with a high impact, as IEs can be immediately treatable upon warning, either by changing ventilator's parameters (e.g., pressure rising time, level of pressure assist, expiratory threshold) or by changing medication (e.g., sedatives). In this work, a methodology for reliable compact description as well as estimation of this asynchrony based on short past observation periods in correlation with physiological parameters is investigated. Among the challenges pertaining to this problem and the proposed approach are:

- a. The related physiological parameters are typically not available by noninvasive methods. In our case, an experimental device is employed making available a multitude of ventilation and physiological parameters noninvasively. A basic principle for such analysis is that these signals cannot really be regarded independently, as they express systems and their interactions in stable or deteriorating states. The use of principal component transformation is proposed, previously successful in biosignal multiparametric analysis [17] and in ventilator variability analysis [5].
- b. The ventilation-related parameters are difficult to interpret by non-experts or in a 24-h workload. Automated short-term prediction of dysynchrony based on the monitored parameters is proposed, employing data-driven modeling based on SVM nonlinear regression [6, 19, 21].
- c. There is little (or at least not conclusive) clinical evidence on how different parameters effect on dysynchrony and how in turn this affects clinical outcome. The proposed approach is based on a large number of data and proposes a causal model predicting the number of IEs based on the multiparametric monitoring ICU data.
- b. PEEP and P01: intrinsic PEEP (positive end-expiratory pressure) and the strength of neural input or stimuli for respiration.
- c. Vent_Ti, Vent_Ttot, Patient_Ttot and younesratio: inspiratory and total time of a ventilation cycle and of patient cycle. $\text{younesratio} = \text{Vent_Ti} / \text{Patient_Ttot}$.
- d. Vt and BR: respiration volume and breathing rate.
- e. Peak_effort and Pressure_assist: maximum patient-generated pressure and the assist pressure produced from the ventilator.
- f. Cycling off: delay time (in sec) between opening of expiration valve and end of patient effort.
- g. Central_apnea, Auto_trigger and Secretion. Apnea (absence of patient breathing), automated triggering of ventilator without patient breathing and high presence of secretion.
- h. IEs between two respirations.

More detailed description of some characteristic parameters of the ventilation monitoring setup is presented in Table s1 of the supplementary file.

In the current setup, data were provided from the ICU clinic of the University Hospital of Heraklion, Crete (<http://icu.med.uoc.gr/>), from 110 patients, with one up to three 24-h recordings during the first week on assisted MV per patient. Treating physicians had no access to these data, which were later downloaded and analyzed offline, leading to 2946 h of recording, with median valid recording duration/patient: 22.8 h (range 3.4–69.7 h) and 4,456,537 valid breaths. The study was approved by the human studies subcommittee, and informed consent was obtained from surrogates. The study did not involve any change from usual clinical practice, and informed consent was related only to permission to obtain and use (anonymously) the data.

In the raw data, each sample consists of a breathing event and related physiological and ventilator parameters. Some details with respect to the initial data are included in the supplementary file Figure s1. This sequence of event tuples was transformed to sampled signals at 60 s, which expresses the number of IE events per 60 s and their resampled–interpolated values for each of the other bioparameters. Very short files (less than 120 samples, i.e., 2 h long) were excluded from the dataset, and for the rest, missing and erroneous values (outliers) were replaced by mean value. Eventually, 178 cases remained, with 173,743 observations total (at 60 s/sample). An example of recorded parameters time series is depicted in Fig. 2.

2 Methods

The methodology presented in the following subsections aims at describing multiparametric patterns related to IEs and at predicting the number of IEs based on the multiparametric ICU data. Specifically, the method includes preprocessing and homogenization of data, PCA-based data transformation to reduce dimensionality and extract robust features, investigation of potential clusters within the dataset to reveal observation clusters and within subject dynamics and application of a model that can predict the future output based on the previous known conditions.

2.1 Data description and preprocessing

In our assisted ventilation setup employing the PVI monitor [10, 24], the ventilator data per patient under ventilation support consist of continuous 24-h recordings of ventilation events streamed data (about 1 sample per second), including the ineffective ventilation synchronization efforts, onsets of inspiration efforts, central apneas, mechanical compliance, timings and pressures and secretions. The 17 parameters included in this analysis are:

- a. Kv, Kf, Kfv: estimations of mechanical elastance and resistance of the respiratory system and their ratio.

2.2 Principal component analysis, feature extraction and descriptive model

Taking into account that the monitored bioparameters are expected to present some degree of correlation, principal

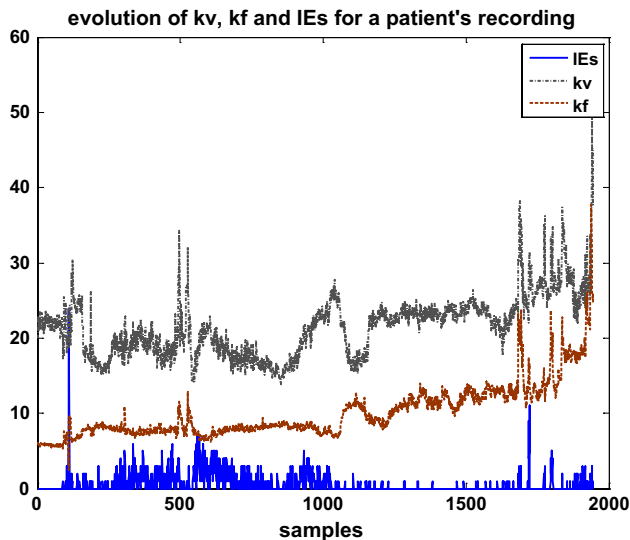


Fig. 2 An example of parameter evolution of parameters during a patient's recording. Kv and Kf stand for mechanical properties of the lungs, elastance and resistance. It can be seen that their variation patterns may at some points precede the onset of IEs

component analysis (PCA) was applied [17], in order to decrease the dimensionality in this vector of parameters and convert the observations set into a new set consisting of linearly uncorrelated variables of usually lower dimension. The number of principal components that explained 95 % of the variance was kept. The set of parameter values used as input for the PCA consisted of $\langle \text{parameter}_i, \text{dparameter}_i \rangle$, for $i = 1, \dots, 17$, where dparameter denotes the derivative in time of a parameter. The reason for including derivatives is that not only absolute values, but also their change with time conveys important information in this problem and needs to be taken into account during pattern analysis. This resulted in an initial dataset of 34 parameters per observation.

As some of the variables were actually binary, having values 0 or 1, corresponding to the occurrence or not of an event (e.g., central apnea, secretion, autotrigger), a choice was made to handle separately the continuous and binary datasets. The reason is twofold: (a) computationally to avoid any bias introduced by the binary variables and their obvious lack of normality and (b) physiologically to separate continuous gray zone intertwined phenomena from direct drivers of ventilator asynchrony (e.g., autotrigger). Thus, the data set was split in two subsets: (a) subset A in which all categorical parameters had zero value, and (b) subset B, including the observations where at least one of the categorical values had value 1. PCA was applied separately for each of the two subsets. Before PCA, data were normalized (subtraction of mean value, division by standard deviation).

PCA allowed us to extract two pieces of information, (a) which initial parameters were the most important along the main principal components and (b) how were the observations clustered in the principal component space. The latter also allowed exploring the transition among clusters in each single case. Clustering of observations took place via k means (with Euclidean distance and 10 repetitions), and the optimal number of clusters was decided based on the mean silhouette (measure of intra-group similarity as compared to similarity with other groups). The transition among clusters within the same subject was employed as a means to study the dynamics in each subject's recordings.

2.3 Predictive modeling

The purpose of modeling was to formulate a prediction method that would estimate the vector of the future n ineffective efforts index $Y_{\text{in}} = (y(i), \dots, y(i + n - 1))$ at times $i, i + 1, \dots, i + n - 1$, $\text{InEf}(i)$, based on the k previous values of the feature vector X , of the previous values of Y , and their derivatives, as expressed in Eq. 1.

$$Y_{\text{in}} = f(X(i - 1), X(i - 2), \dots, X(i - k), y(i - 1), \dots, y(i - k), dy(i - 1), \dots, dy(i - k)) \quad (1)$$

where the feature vector X_i is based on the l first principal components, as described in the previous section. In other words, the model was meant to test if k previous input–output values can predict n future output values Y , where Y denotes the ineffective effort (IE). In this work, maximum k was 5, i.e., a maximum history of 5 min of previous values was employed, and the maximum n was also 5, i.e., the number of future prediction window was 5 min. Predicting IEs in a short-term window is expected to provide an opportunity for timely preventing/mitigating the IE adverse impact.

A nonlinear SVM regression model was adopted in order to implement the function f , following a least-squares support vector machine (LS-SVM) model implementation. LS-SVM supports regression problems and is computationally efficient. The main idea behind LS-SVM is described by [19], and the implementation in MATLAB was adopted from the toolbox (<http://www.esat.kuleuven.ac.be/sista/lssvmlab>).

The nonlinear kernel function employed was the radial basis function (RBF), which was chosen, rather than a linear kernel, so that the potential nonlinearity in this complex modeling problem could be better described. The LS-SVM model parameters used were gam , the regularization parameter, determining the trade-off between the training error minimization and smoothness of the estimated function and sig , a kernel function parameter. For each run, the model parameters were fine-tuned via simplex method with leave-one-out-validation on a smaller subset of the full training set.

In order to assure robustness, the following procedure was repeated ten times. In each run, a random set of cases was picked to create the training set, consisting of 10–11 recordings summing up a total length of no less than 8000 observations. SVM model was trained based on these data, for the combinations $\{(k = 1, n = 1), (k = 2, n = 1,2), (k = 3, n = 1,2,3), (k = 4, n = 1,2,3,4), (k = 5, n = 1,2,3,4,5)\}$. The rest of the unseen cases were used as test set. It has to be noted that using not only unseen observations, but unseen patient cases (full records) is considered as a more realistic testing scenario, ensuring that the model was not built on data neighboring to the ones it was tested on. The measures used to assess goodness of fit between the estimated regression model Y and the actual IEs, in each run, include mean absolute error, mean error, standard deviation of the error, $\text{Rs}q = 1 - \frac{\text{error}^2}{\text{sum}(\text{IE} - \text{mean})^2}$, and linear correlation coefficient between real and predicted values. The average of these measures for all ten runs was eventually considered and compared among the different (k,n) combinations. Additionally, as a preliminary indicator of this approach's potential to contribute toward ICU alarming mechanisms, the classification accuracy when setting the cutoff threshold = 2, i.e., when attempting detection of times with less than 2 IEs (practically clean) against the rest, was assessed.

3 Results

3.1 PCA analysis and clustering

In the investigation of PCA transformation, only the time points with $\text{IE} \geq 1$ were employed. This resulted in 17,645 samples in subset A and 3299 samples in subset B (apnea, secretion, autotrigger events). The PCs explaining 95 % of the variance were selected, and in this manner, subset A was described with 5 PCs, and so did subset B, which now could represent the initial 34 intertwined variables.

Figure 3 presents a biplot with the PC loadings and scores of all observations.

The parameters mostly expressed in each PC, according to the parameter loadings, are depicted in Table 1, in an ordered manner (only five most important loadings shown here). It can be observed that in each PC, the parameters present some relevance. For example, in subset A, PC1 involves parameters related to rhythms (predominantly tachypnea) and event timings, while PC2 involves more physiological parameters, related to mechanical/neurological properties of the patient (increased expiratory resistance), PC3 ventilation over-assistance, etc. It is also worth noting that some derivative features (dx) are important, which justified the initial option to include them in the dataset.

Based on the PCA scores, k means separated subset A observations (17,645) in three clusters ($\text{NA}_i, i = 1 \dots 3$), with sizes 10,989, 4678 and 1978. The mean silhouette value of this clustering was 0.3681. Similarly, subset B (3299) resulted in five clusters ($A_i, i = 1 \dots 5$) with sizes 117, 1301, 376, 1128 and 377. The mean silhouette was 0.5209 in this case. Detailed figures of clusters are available in the supplementary file, Figure s2. For each cluster, the mean values of the PC scores and the mean values of the initial bioparameters are summarized in the supplementary file Tables s2 and s3, respectively.

As the observations for the generation of PCA transform came from multiple cases and patients, it was considered important to investigate whether these clusters corresponded to different patients, each patient belonging to a single cluster, or the clusters vary on an intra-patient and inter-patient basis, reflecting a more non-uniform and non-stationary situation. Analysis proved the second hypothesis (each patient case contains many clusters, and this distribution of IE time in clusters varies among patients), as can be seen in the visual representation of the percentage of the IE time per case spent in each cluster, depicted in Fig. 4a. This is further illustrated in the example presented in Fig. 4b. For a single patient with COPD exacerbation, sequential time segments with IEs may stay in the same cluster. This could suggest that a single phenomenon may pertain (e.g., not fully addressed by treatment). As seen in Fig. 4b, while the first IEs relate to cluster NA1 (increased expiratory resistance, slow breathing), following there is persistence in the second PC cluster (NA2), related to tachypnea (see also Table s2 and Table s3). There are also numerous transitions among clusters, in this case, spontaneous transitions from NA2 to NA3 (ventilation over-assistance), to A1, A2 and A4 (related to secretion and ventilation over-assistance), which point at cases where a phenomenon not properly addressed builds up other physiological phenomena, or the treatment of a problem in a wrong manner would create other problems, all potentially building up reasons for IEs. Other examples with different dynamics are available in the supplementary file, Figure s4. It has to be noted that here observations without IEs were not taken into account, so exact temporal neighborhood is not strictly preserved. This example is characteristic of the complexity in predicting IEs and also attributing them to a specific causality, being of mechanical nature, or wrong ventilation timings/pressures, etc.

3.2 Prediction of future IEs

The IE prediction employed a feature set consisting of (a) PC features (signals*coefficients) based on the PCA coefficients that were produced by subset A (see previous section), and (b) previous IE and dIE. An initial small-scale

Table 1 Parameters with highest loadings per PC, in the first five PCs

Subset A	Subset B
PC1. PatientTtot, VentTot, breathrate, VentTi, Peepi <i>predominantly tachypnea—or the opposite</i>	PC1. AutoTrigger, Secretion, PeakEffort, Younesratio, DAutoTrigger <i>autotrigger because of secretion</i>
PC2. Kfv, kf, Peepi, P01, VentTi <i>predominantly increased expiratory resistance</i>	PC2. VentTi, VentTot, DVentTi, breathrate, Cyclingoff <i>apnea</i>
PC3. Younesratio, Vt, PeakEffort, Dyounesratio, Cyclingoff <i>ventilation over-assistance</i>	PC3. Dkfv, kf, kfv, kv, DPeepi <i>change in lung mechanics because of secretion</i>
PC4. Dyounesratio, DVentTi, kf, kfv, Younesratio <i>predominantly change in lung mechanics</i>	PC4. Vt, kv, Peepi, PressureAssist, DPeakEffort <i>ventilation ventilation over-assistance</i>
PC5. Dkfv, Dkv, kf, Cyclingoff, DCyclingoff <i>predominantly bad cycling off in ventilation</i>	PC5. Dyounesratio, DPressureAssist, kf, breathrate, VentTot <i>change in expiratory resistance</i>

In italics, potential phenomena that may be related with variations in these parameters

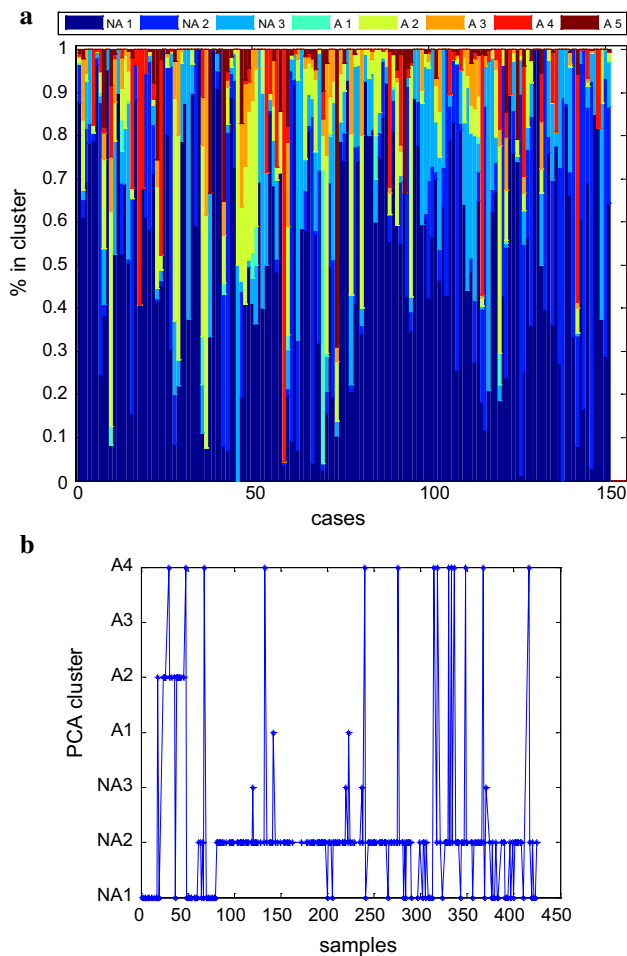


Fig. 4 **a** Stacked view of the percentage of each PCA cluster per case, i.e., for each case (*bar*), considering only times with IE, the percentage of time spent in each of the clusters. NA1-3: cluster 1–3 of subset A, A1–5: cluster 1–3 of subset (discrete events of apnea, high secretion, autotrigger). **b** For a single patient case, a sequential view of the samples with IE, and the cluster to which each sample belongs, suggests the type of phenomenon that may take place during each IE

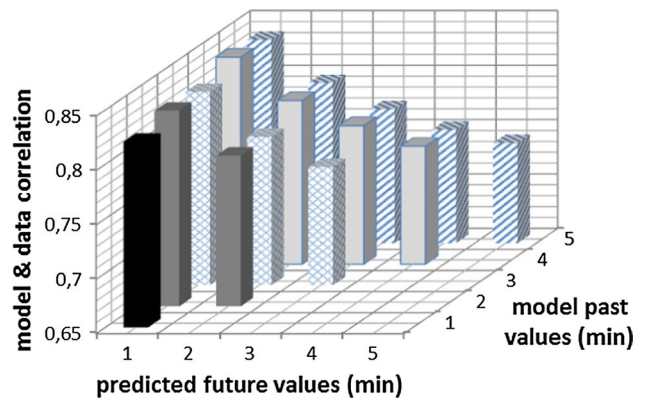


Fig. 5 Linear correlation between real and predicted values, for various model orders and future prediction windows

The performance measures are summarized in Table 2. It can be seen that the standard deviation among runs is small. All performance metrics decrease with future prediction window, however, not dramatically.

Figure 6 depicts a sequence of real and predicted IEs, as well as the predicted versus real IEs scattergram. Although the latter (Fig. 6b–c) shows a tendency for underestimation of actual spike values (e.g., predicted spikes have a lower value, and fitted line is below $y = x$), the general tendency for high/low IEs is preserved, in the sense that areas with rare IEs and areas with dense IEs are visible.

Finally, as regards the ability to predict periods of practically no IEs from the periods with IEs, by setting a threshold on the predicted values (class 1 = no IEs if $IEs \leq 2$, class 2 = $IEs > 2$), average specificity was around 93 % in all cases. This means that the proportion of no IEs which are correctly identified as is high, i.e., periods of no IEs are predicted, even for future windows of 5 min. Table 3 presents results on sensitivity, which ranges from 67 to 58 %. While this is not a perfect outcome, it highlights the

Table 2 SVM prediction performance expressed as mean \pm standard deviation among the ten runs

Measure	Model order/future prediction window				
	1/1	2/2	3/3	4/4	5/5
Mean abs error	0.84 \pm 0.06	0.90 \pm 0.06	0.96 \pm 0.10	0.95 \pm 0.06	0.98 \pm 0.06
Mean error	-0.06 \pm 0.09	-0.06 \pm 0.11	-0.06 \pm 0.14	-0.02 \pm 0.12	-0.06 \pm 0.14
Standard deviation of error	1.72 \pm 0.07	1.86 \pm 0.04	1.96 \pm 0.10	1.97 \pm 0.06	2.02 \pm 0.07
Rsqr	0.67 \pm 0.02	0.61 \pm 0.02	0.57 \pm 0.04	0.56 \pm 0.03	0.54 \pm 0.04
Correlation coefficient	0.82 \pm 0.01	0.79 \pm 0.01	0.76 \pm 0.03	0.76 \pm 0.02	0.74 \pm 0.03

potential for generating a prediction scheme that could predict and alarm for future events in ICU environment.

4 Discussion

The complexity of monitoring in an ICU environment is a challenging task with major impacts. In the case of assisted ventilation and the evaluation of the interaction (or synchronization) between the ventilation device and the patient respiration dynamics, a continuous multichannel recordings of ineffective effort events have to be employed in a way to provide valuable input to the clinicians, toward achieving optimal patient–ventilator synchrony [4]. This flow of continuous, multiparametric and correlated data, with complex causal relations, reflecting the dynamic interaction between ventilator and patient’s physiology, is of vital importance for the understanding of ventilation dynamics [24], toward recognition of ventilation dysynchrony problems and application of intervention adjustments, e.g., in ventilatory parameters or sedative therapy. This approach, via predicting IE events and shedding light on the related phenomena with clustering, is expected to contribute in avoiding overexposure to IEs with the respective short-term effects (dyspnea, hypercapnia, discomfort, sleep fragmentation, muscle overload) and long-term effects (increased duration of mechanical ventilation, shorter ventilator-free survival, increased length of stay, lower likelihood of home discharge, unsuccessful weaning and discomfort). Thus, it is considered as an example where modeling the dynamic procedure can lead to the prediction and prevention of critical events, as well as timely intervention and adaptation of therapy for each patient during the ICU stay.

The proposed methodology contributes to the understanding of ICU ventilation dynamics, toward a framework for ICU biosignal and bioparameter analysis, extending the initial work [7]. A basic concept proposed in this work is the use of a multiparametric approach, to investigate the dynamic and non-uniform phenomena occurring in assisted ventilation. In order to deal with multiple intertwined parameters, as well as the importance of their temporal evolution, the observed parameters and their derivatives are transformed via PCA into a more compact set of

5 PCs. These group together parameters that are relevant overall, for example, PC1 in subset A relates to inspiratory and total time of ventilation and of patient cycle (e.g., possible tachypnea-related event). Additionally, based on Table s2 and Table s3, the PCA clusters could link to specific and recognizable phenomena related to ventilation dysynchrony, for example, in subset A, (a) cluster 1 concentrates in PC1 (negative) and PC2 (positive) and presents high kvf and low breathing rate, which may be interpreted as predominantly increased expiratory resistance, leading to increased expiratory times and low breathing rate, (b) cluster 2 concentrates mainly along PC1 and could relate to tachypnea, and (c) cluster 3, with mainly negative PC2, negative PC1 and positive PC3, also presenting high breathing rate and Peepi, probably relates to ventilation over-assistance. Yet, a wider clinical interpretation and evaluation are on the way to further and systematically link the proposed PCs with their actual clinical relevance.

It has to be noted that at the moment, this PCA procedure is performed separately under the presence or not of specific events (e.g., apnea) which are expressed with categorical values. There is a potential physiological reason for separating specific events that directly affect synchrony (e.g., an incidence of apnea is directly linked to lack of breath and lack of coupling with assisted ventilation), from the other ‘gray zone’ time points where multiple factors contribute to the variation of synchrony. Besides that, these event parameters being categorical cannot be normalized and handled as the other continuous ones. However, a method to handle in a unified manner the continuous and categorical parameters and thus the whole set, with or without the presence of specific important events, would be a preferable future option.

Additionally, while the addition of parameter derivatives in the initial feature set proved a successful decision, regarding their role in the most important PCs, one could better tackle these phenomena of temporal evolution in multiple scales by introducing a wavelet analysis before PCA [20, 22]. This might then lead to parameters at different scales being together expressed in PCs and thus to the different scales of interwoven phenomena. Multiscale PCA has been applied in different biosignal feature extraction tasks [23].

Fig. 6 **a** Real and predicted IEs overlapped sequences, for test data, model order 3 and future window 5. **b, c** Real versus predicted IEs scattergram. X-axis: Real IE. Y-axis: predicted IEs (y-axis). Model order and future window (1,1) and (3,3) as in **(b)** and **(c)**, respectively

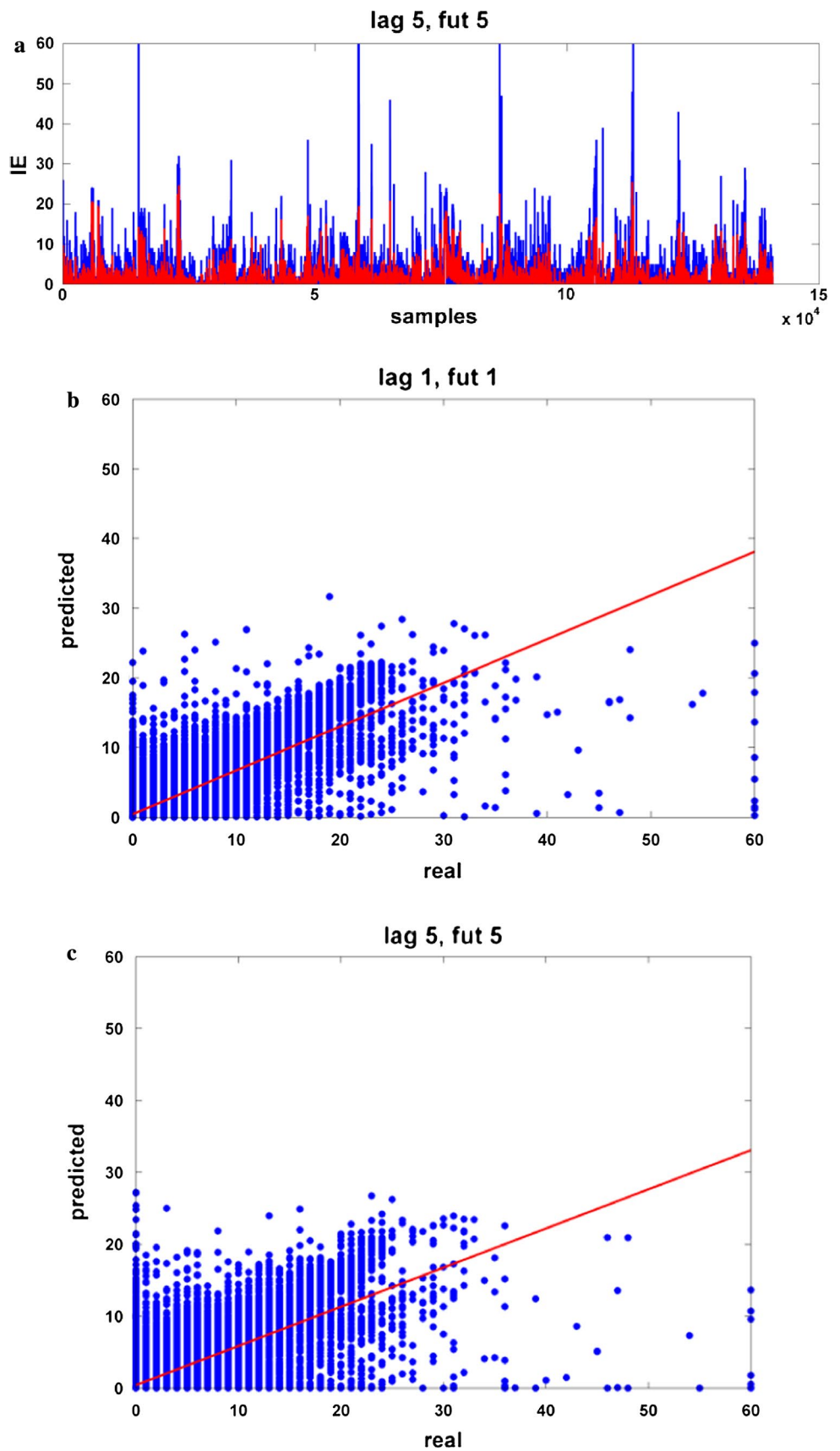


Table 3 Model sensitivity in separating future IEs from no IEs

Previous inputs	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
<i>Future IEs</i>					
1	67.20	67.22	67.48	67.59	67.90
2		62.12	62.71	63.44	63.72
3			60.66	61.08	61.26
4				59.16	59.16
5					57.96

A second important part of the proposed approach includes the dynamics of the event. After clustering the observations, based on the PC scores, a fragmentation in 8 clusters is illustrated, expressing different states, and potentially of different origin (patient or ventilator factors). More importantly, upon studying these clusters and their transitions in each patient case, one can see the great variability of the synchrony phenomena, being non-stationary within each patient case and non-uniform among them. Thus, clustering together cases [7] is a less realistic option than clustering segments of cases. In other words, one important aspect could be the segmentation of signals and the study of causality in these segments. Additional analysis can be investigated based on different signal transformations and similarity methods, e.g., based on wavelet transform that has already shown promising results in the ICU biosignal analysis [16] and other complex multiparametric problems [18]. Upon further elaboration on different clusters and transitions among clusters, as well as their physiological characterization, a means to predict the ventilation deterioration and its potential cause and alarm the health professionals accordingly could be considered.

More advanced physiological characterisations could be achieved if the ventilation measurements based on the PVI monitor were combined with other noninvasive or minimally invasive physiological measurements, for example heart rate and respiration rate variability [15] and electrical impedance tomography [14].

The proposed work has also investigated the problem of predicting future IEs and concluded that this is possible up to a degree and based on short history. Long history, i.e., history window much longer than future prediction window, is not performing better. Besides the consequences of long history in increasing the feature set and the training needs, it is possible that the time scales of temporal correlations and causalities are varying. In addition, the PC coefficients were based only on subsetA (the observations without apnea, secretion, autotrigger events). Again, employing wavelet-based parameters in a future more sophisticated model might help investigating what are the best history options. Furthermore, although the SVM-based approach proved credible, various modeling methods can be tested

via continuous system approaches (state space, nonlinear autoregressive, Volterra models) [11]. In any case, the proposed modeling approach reached a good level of prediction, even for up to 5 min.

Besides regression, results illustrated the value of predicting a low/high IE classification. In our case, with the simple thresholding of the regression output, specificity is high, yet sensitivity would need to be improved, provided a more optimized classification scheme is put in place. A future direction would be to provide a classification scheme for low to high or high to low transitions, i.e., event onsets and offsets.

Combining the clustering and prediction methods, especially in time windows where IEs raise from very low to some significant value, can lead to ICU applications with: (a) warnings and alerts in a timely fashion [3], (b) presentation of the necessary information that can guide the clinician on what is the most probable reason of dysynchrony at that specific time and what ventilation parameter needs to be changed, or what other intervention needs to take place, as corrective action. Alerts can be based on the IEs predictive models. Guiding interventions require depicting and visualizing the etiology, and PCA clustering constitutes a preliminary contribution in this direction, along with a rule-based system for the actions according to etiology.

5 Conclusion

In conclusion, various methodological aspects are tackled in this work, along with their future directions for improvement. The assisted ventilation bioparameters and their derivatives are transformed via PCA in five components, and their distribution in clusters is investigated. Furthermore, these transformed features are then employed in short-term prediction ineffective ventilation efforts via a nonlinear autoregressive exogenous model (nonlinear svm regression), where the model relates the current value of a time series with its past values and past values of the influencing series. Based on the encouraging results, a detailed clinical validation and interpretation with respect to the pathophysiology and known progress of each patient is of major importance as a future step. The impact of such analysis lies in the optimized management of assisted ventilation, toward not only understanding the mechanisms and patterns of assisted ventilation events, but also attempting short-term predictions of problematic synchronizations and alerting the clinicians accordingly.

There is a big technological potential in improving monitoring of ventilation bioparameters, both as regards sensing and multiparametric analysis. Incorporating such complex biosignal processing scenarios in an ICU setup is indeed interesting and challenging, and relevant efforts are

now being tested worldwide [2]. ICU intelligence can be a case for big data analytics technologies toward large-scale new knowledge discovery and toward real-time performance. Furthermore, these approaches may present a wider interest for cases of analysis in large and variable datasets, beyond ICU.

References

- Babalís D, Lilitsis E, Akoumianaki E, Prinianakis G, Georgopoulos D, Kondili E (2011) Incidence of ineffective efforts in mechanically ventilated critically ill patients. *Intensive Care Med* 37(1) Suppl S6 (Abstract no 0005)
- Blount M, Ebling MR, Eklund JM, James AG, McGregor C, Percival N, Smith KP, Daby S (2010) Real-time analysis for intensive care: development and deployment of the artemis analytic system. *Eng Med Biol Mag IEEE* 29(2):110–118
- Burns SM (2003) Working with respiratory waveforms: how to use bedside graphics. *AACN Clin Issues*. 14(2):133–144
- Burns SM (2004) Continuous airway pressure monitoring. *Crit Care Nurse* 24(6):70–74
- Castro RR, Magini M, Pedrosa S, Sales AR, Nóbrega AC (2011) Principal components analysis to evaluate ventilatory variability: comparison of athletes and sedentary men. *Med Biol Eng Comput* 49(3):305–311
- Chouvarda I, Mendez MO, Rosso V, Bianchi AM, Parrino L, Grassi A, Terzano M, Maglaveras N, Cerutti S (2011) Predicting EEG complexity from sleep macro and microstructure. *Physiol Meas* 32(8):1083–1101
- Chouvarda IG, Babalís D, Papaioannou V, Maglaveras N, Georgopoulos D (2014) Predicting the ineffective efforts in ICU assisted ventilation via multiparametric models. In: *IFMBE Proceedings of the international conference on health informatics*, vol 42, pp 224–227
- Georgopoulos D (2012) Ineffective efforts during mechanical ventilation: the brain wants, the machine declines. *Intensive Care Med* 38:738–740
- Karbing DS, Allerød C, Thomsen LP, Espersen K, Thorgaard P, Andreassen S, Kjærgaard S, Rees SE (2012) Retrospective evaluation of a decision support system for controlled mechanical ventilation. *Med Biol Eng Comput* 50(1):43–51
- Kondili E, Alexopoulou C, Xirouchaki N, Vaporidi K, Georgopoulos D (2010) Estimation of inspiratory muscle pressure in critically ill patients. *Intensive Care Med* 36(4):648–655
- Leontaritis IJ, Billings SA (1985) Input-output parametric models for non-linear systems. Part I: deterministic non-linear systems. *Int J Control* 41:303–328
- Lozano-Zahonero S, Gottlieb D, Haberthür C, Guttman J, Möller K (2011) Automated mechanical ventilation: adapting decision making to different disease states. *Med Biol Eng Comput* 49(3):349–358
- McMullen SM, Meade M, Rose L, Burns K, Mehta S, Doyle R, Henzler D, for the Canadian Critical Care Trials Group (CCCTG) (2012) Partial ventilatory support modalities in acute lung injury and acute respiratory distress syndrome—a systematic review. *PLoS ONE* 7(8):e40190
- Muders T, Luepschen H, Putensen C (2010) Impedance tomography as a new monitoring technique. *Curr Opin Crit Care* 16(3):269–275
- Papaioannou VE, Chouvarda I, Maglaveras N, Dragoumanis C, Pneumatikos I (2011) Changes of heart and respiratory rate dynamics during weaning from mechanical ventilation: a study of physiologic complexity in surgical critically ill patients. *J Crit Care* 26(3):262–272
- Papaioannou VE, Chouvarda IG, Maglaveras NK, Pneumatikos IA (2012) Temperature variability analysis using wavelets and multiscale entropy in patients with systemic inflammatory response syndrome, sepsis and septic shock. *Crit Care* 16(2):R51
- Ringnér M (2008) What is principal component analysis? *Nat Biotechnol* 26:303–304
- Rouyer T, Fromentin JM, Menard F, Cazelles B, Briand K, Pianet R, Planque B, Stenseth NC (2008) Complex interplays among population dynamics, environmental forcing, and exploitation in fisheries. *Proc Natl Acad Sci* 105:5420–5425
- Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J (2002) Least squares support vector machines. *World Scientific, Singapore* (ISBN 981-238-151-1)
- Takiguchi T, Ariki Y (2006) Robust feature extraction using Kernel PCA. In: *Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP '06)*, pp 509–512
- Wenqi Y, Widmer N, De Micheli G (2011) Personalized modeling for drug concentration prediction using Support Vector Machine. In: *4th International conference on biomedical engineering and informatics (BMEI)*, vol 3, pp 1505–1509
- Xie S, Lawniczak AT, Krishnan S, Lio P (2012) Wavelet Kernel principal component analysis in noisy multiscale data classification *ISRN computational mathematics*, vol 2012, Article ID 197352, 13 pp. doi:10.5402/2012/197352
- Xie S, Jin F, Krishnan S, Sattar F (2012) Signal feature extraction by multi-scale PCA and its application to respiratory sound classification. *Med Biol Eng Comput*. 50(7):759–768
- Younes M, Brochard L, Grasso S et al (2007) A method for monitoring and improving patient: ventilator interaction. *Intensive Care Med* 33(8):1337–1346