

# Is multidimensional scaling suitable for mapping the input respiratory impedance in subjects and patients?

Clara M. Ionescu, Jose Tenreiro Machado, Robin De Keyser

## A B S T R A C T

This paper presents the application of multidimensional scaling (MDS) analysis to data emerging from noninvasive lung function tests, namely the input respiratory impedance. The aim is to obtain a geometrical mapping of the diseases in a 3D space representation, allowing analysis of (dis)similarities between subjects within the same pathology groups, as well as between the various groups. The adult patient groups investigated were healthy, diagnosed chronic obstructive pulmonary disease (COPD) and diagnosed kyphoscoliosis, respectively. The children patient groups were healthy, asthma and cystic fibrosis. The results suggest that MDS can be successfully employed for mapping purposes of restrictive (kyphoscoliosis) and obstructive (COPD) pathologies. Hence, MDS tools can be further examined to define clear limits between pools of patients for clinical classification, and used as a training aid for medical traineeship.

### Keywords:

Respiratory impedance  
Lung function  
Multidimensional scaling  
Geometrical mapping  
Respiratory disorders  
Dendrograms  
Clustering

## 1. Introduction

Multidimensional scaling (MDS) is a family of statistical techniques which attempt to discover the *hidden structure* in the available data [8]. MDS uses a matrix of proximities among the objects as input and produces an  $N \times N$  mapping matrix of the output, given  $N$  objects for mapping. In other words, the MDS technique provides a geometric interpretation to (dis)similarity data and is a natural tool for mapping data sets in a low dimensional space. Usually, by minimizing a loss function calculated for different possible configurations, a set of coordinates can be assigned to the envisaged objects, providing a functional meaning to the geometry of the map. The resulting map, or embed-

ding, places objects that have similar attributes close to each other.

Various examples from technical and non-technical fields of applications can be found to use MDS as a mapping tool. The MDS has been applied to pilot performance data obtained during simulated air-to-air combat [27], where an adequate performance metric was developed to describe the complex interaction between position advantage and energy management.

Multidimensional scaling is also aimed to represent high dimensional data in a low dimensional space with preservation of the similarities between data points [12]. This reduction in dimensionality is crucial for analyzing and revealing the genuine structure hidden in the data. For noisy data, dimen-

sion reduction can effectively diminish the effect of noise on the embedded structure. For large data sets, dimension reduction can effectively overcome the information retrieval complexity. Thus, MDS techniques are used in many applications of data mining and gene network research [30,31].

From the field of non-technical applications, the area of medicine seems to suit most the applicability of MDS, that is, the analysis of biomedical data in general. For medical image analysis, it is important to take advantage of the full range of information presented in an image, thus one has to consider distance and shape attributes [21,32]. Separation between left and right brain sulci was developed using MDS for a new geometric representation [21]. Topography of functional brain spaces and cortico-cortical interactions was implemented through MDS [13], in order to transform anatomical space so that the distance between cortical areas is directly related to their functional connectivity. Furthermore, the MDS has mapped the relationship between the odor descriptors to determine whether they might reveal an underlying structure in human odor perception. Some efforts to use an information theoretic approach suggested that, in fact, there might be an overall organization to the space of human olfactory perception [20]. Similarly, MDS has been useful to provide an automatic method for classification of EEG waveforms, in order to objectively detect changes in EEG recordings [14], with results in agreement with visual examination by trained physicians.

Bearing in mind the success of previously reported results of MDS in various medical applications, we propose the use of MDS tools to provide a geometrical mapping of data from lung function tests in healthy subjects and in patients with respiratory disorders. One of the most common features extracted from lung function tests is air-pressure and air-flow variations during forced breathing or during breathing at rest. The data used in this application has been acquired in lung function tests during breathing at rest, namely using the forced oscillation technique [25]. We aim to analyze six sets of patient data: healthy adults, adults with chronic obstructive pulmonary disease, adults with kyphoscoliosis, healthy children, children with asthma and children with cystic fibrosis. The objective of this paper is to determine whether or not MDS tools are a valuable mean to attain clustering of data for classification between respiratory disorders in clinical practice.

This paper is organized as follows. Section 2 introduces the materials and methods employed in the paper. The respiratory impedance data from patients in frequency domain is presented, along with information upon the various sets of data, biometric details of the population sets and measurement protocol. A short description of the MDS technique employed here

is given in the same section. Section 3 presents the results of the MDS analysis and discusses the geometrical interpretation of the emerged results, along with the potential of MDS for clinical use. Finally, a conclusion section summarizes the main outcome of this paper.

## 2. Materials and methods

### 2.1. Patients

There are six sets of patients available for analysis. We have both children, as well as adult subjects, for each class with healthy and respiratory disorders. The biometric data and pathology details are given in Tables 1 and 2 and in the remainder of this sub-section. As a general remark, it should be noted that each specific group was tested in a single location and that both team and device employed to perform the lung function tests was always the same.

The **healthy adult group** evaluated in this study consists of 21 Caucasian volunteers (students) without a history of respiratory disease, whose lung function tests were performed in our laboratory. Table 1 presents their biometric parameters and the data was validated as healthy using prediction values [29,26].

**Chronic obstructive pulmonary disease (COPD)** denotes any disorder that persistently obstructs the bronchial airflow [1]. However, it mainly involves two related diseases: chronic bronchitis and emphysema. Both cause chronic obstruction of air flowing through the airways and in and out of the lungs. The obstruction is irreversible and progresses (becomes worse) over time. The COPD group under study consisted of 21 Caucasian patients, diagnosed and under observation at the “Leon Danielo” Hospital in Cluj-Napoca, Romania. The patients were former coal miners from the Petrosani area in Romania. Their biometric and spirometric parameters are given in Table 1.

**Kyphoscoliosis (KS)** is a disease of the spine and its articulations, mostly beginning in childhood [24]. The deformation of the spine characteristically consists of a lateral displacement or curvature (scoliosis) or an antero-posterior angulation (kyphosis) or both (kyphoscoliosis). The angle of the spinal curvature called the *angle of Cobb* determines the degree of the deformity and consequently the severity of the restriction. This study was approved by the local Ethics Committee of the University Hospital Gent-Belgium and informed consent was obtained from all volunteers before inclusion in the study. The study involved 9 adults diagnosed with kyphoscoliosis (Cobb angle = 75°) and their corresponding biometric and spirometric values are given in Table 1.

**Table 1 – Biometric and spirometric parameters of the investigated adults subjects. Values are presented as mean  $\pm$  standard deviation; % pred: predicted according to the asymptomatic males of the present study; VC: vital capacity; FEV<sub>1</sub>: forced expiratory volume in 1 s. KS, kyphoscoliosis; COPD, chronic obstructive pulmonary disease.**

Adults	Healthy (n = 21)	KS (n = 9)	COPD (n = 21)
Age [years]	27 $\pm$ 2	62 $\pm$ 10	64 $\pm$ 3
Height [m]	1.73 $\pm$ 0.27	1.55 $\pm$ 0.08	1.73 $\pm$ 0.14
Weight [kg]	69 $\pm$ 5.6	63 $\pm$ 15	79 $\pm$ 12
VC [% pred]	–	33 $\pm$ 14	84 $\pm$ 12
FEV <sub>1</sub> [% pred]	–	31 $\pm$ 11	38 $\pm$ 6

**Table 2 – Biometric and spirometric parameters of the investigated children subjects. Values are presented as mean  $\pm$  standard deviation; % pred: predicted according to the asymptomatic males of the present study; VC: vital capacity; FEV<sub>1</sub>: forced expiratory volume in 1 s; FEF: forced expiratory flow; CF: cystic fibrosis.**

Children	Healthy (n = 16)	Asthma (n = 19)	CF (n = 9)
Age [years]	9 $\pm$ 1	11 $\pm$ 4	14 $\pm$ 6
Height [m]	1.35 $\pm$ 0.05	1.40 $\pm$ 0.2	1.49 $\pm$ 0.15
Weight [kg]	32 $\pm$ 6	36 $\pm$ 15	40 $\pm$ 11
FEF/VC [% pred]	–	85 $\pm$ 31	86 $\pm$ 9
FEV <sub>1</sub> /VC [% pred]	–	97 $\pm$ 1.2	95 $\pm$ 0.9

The measurements on the **healthy children** were performed at the St. Vincentius Basis School in Zwijnaarde, Belgium; the biometric details are given in Table 2. The children had no history of pulmonary disease, and were selected using a specific questionnaire. The questionnaire verified the absence of dyspnoea, chronic cough, wheeze in the chest, etc. The measurements performed on healthy children have been verified with predicted values [10].

**Asthma** denotes a pulmonary disease in which there is obstruction to the flow of air out of the lungs, but the obstruction is usually reversible and, between attacks of asthma, the flow of air through the airways is usually good [6]. Asthma is caused by chronic (ongoing, long-term) inflammation of the airways, making them highly sensitive to various triggers. Such triggers are usually: indoor and outdoor allergens, indoor and outdoor dust, exercise. Asthma can be controlled using specific medication (inhaled steroids). The study was performed at the University Hospital Antwerp-Belgium, including 19 asthmatic children whose corresponding biometric and spirometric values are given in Table 2.

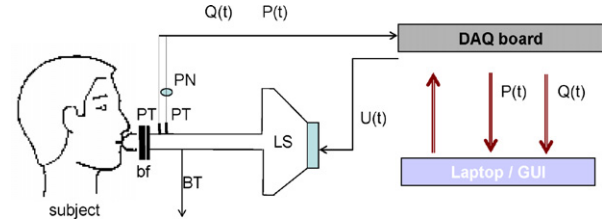
**Cystic fibrosis (CF)** is one of the most common severe genetic diseases, is characterized by the production of abnormal secretions, leading to mucous build-up, and persistent infections and inflammation in a variety of organs [4,11]. Inflammation and infection also cause injury and structural changes to the lungs, leading to a variety of symptoms and eventually to respiratory failure. The study was performed at the University Hospital Antwerp and involved 9 children diagnosed with cystic fibrosis whose biometric and spirometric values are given in Table 2.

In all measurements, drop-out criteria were: (i) technically biased measurements (swallowing, coughing, glottis closure); (ii) fatigue and therefore reduced ability to breathe spontaneously; and (iii) irregular breathing period. All subjects and patients were in stable physical conditions at the time of the evaluation.

Written and/or oral consent was obtained from all participants, and in case of children, from both children and their parents. Further selection of the participants was performed by oral/written questionnaire ruling out any other respiratory disease than the one envisaged for the study at the time of measurement or in the four weeks prior to the test. The results are obtained from the data for the participants to whom these inclusion criteria applied.

## 2.2. Lung function testing: the Forced Oscillation Technique

The impedance was measured using the Forced Oscillation Technique (FOT) standard setup, commercially available,



**Fig. 1 – A schematic overview of the device with patient performing the FOT lung function test. LS, loudspeaker; BT, bias-tube; PN, pneumotachograph; PT, pressure transducer; bf, biological filter with mouthpiece; Q, flow; P, pressure; U(t), driving signal (test input).**

assessing respiratory mechanics from 4 to 48 Hz in steps of 2 Hz. The subject is connected to the typical setup from Fig. 1 via a mouthpiece, suitably designed to avoid flow leakage at the mouth and dental resistance artifact. The oscillation pressure in most recent FOT devices is generated by a loudspeaker (LS) connected to a chamber, namely the SPH-165KEP MONACOR, with a range from 3 to 1000 Hz. The LS is driven by a power amplifier fed with the oscillating signal generated by a computer, namely a HP Pavilion dv1000 with a Pentium M processor, 1.5 GHz, 512 MB with 266 MHz SDRAM. The movement of the LS cone generates a pressure oscillation inside the chamber, which is applied to the patient's respiratory system by means of a flexible respiratory tube of 1 m length and 2 cm diameter, connecting the LS chamber and the bacterial filter (bf). A side opening (BT) of the main tubing allows the patient to decrease total dead space re-breathing (i.e. 40 ml). This bias tube exhibits high impedance at the excitation frequencies to avoid the loss of power from the LS pressure chamber [2]. During the measurements, the patient wears a nose clip and keeps the cheeks firmly supported to reduce the artifact of upper airway shunt. Pressure and flow are measured at the mouthpiece, respectively by means of (i) a pressure transducer (PT) and (ii) a pneumotachograph (PN) plus a differential pressure transducer (PT). The high precision pressure transducers are BSDX0050D4D, with a bipolar pressure range from 0 to 1 kPa, accuracy of 0.004 kPa and a common mode rejection ratio of 80 dB over the frequency range of interest. The working range is a peak-to-peak size between 0.1 and 0.3 kPa, in order to ensure optimality, patient comfort and stay within a narrow range in order to assume linearity [25]. The flow is measured using a Hans Rudolph pneumotachograph, covering a range of 0–400 l/min (6.6 l/s), 4830B series, with a dead space volume between 0 and 6.66 ml. Averaged measurements from 3 to 5 technically acceptable tests (coherence values higher than 0.8) were taken into consideration for further signal process-

ing. The signals are acquired using a PCMCIA DAQ card 4026E series, 12 bit from National Instruments, sampled every 1 ms. All patients were tested in the sitting position, with cheeks firmly supported and elbows resting in the table.

### 2.3. Respiratory impedance

One of the most common non-parametric representations of the impedance  $Z_r$  is obtained assuming a linear dependence between the breathing and superimposed oscillations at the mouth of the patient [15]. The spectral representation of  $Z_r$  is a fast, simple and fairly reliable evaluation.

Since the excitation signal is designed such that it is not correlated with the breathing of the patient, one can estimate the respiratory impedance as in:

$$Z_r(j\omega) = \frac{S_{PU_g}(j\omega)}{S_{QU_g}(j\omega)} \quad (1)$$

whereas the  $P$  corresponds to pressure (its electrical equivalent is voltage) and  $Q$  corresponds to air-flow (its electrical equivalent is current), the respiratory impedance  $Z_r$  can be defined as their spectral (frequency domain) ratio relationship, with  $S_{ij}(j\omega)$  the cross-correlation spectra between the various input-output signals,  $\omega$  is the angular frequency and  $j = (-1)^{1/2}$ . From the point of view of the forced oscillatory experiment, the signal components of respiratory origin ( $U_r$ ) have to be regarded as pure noise for the identification task [15].

### 2.4. Multidimensional scaling (MDS)

There are a manifold of techniques available to cluster data for classification purposes. These vary in optimization algorithms, speed of convergence, visualization technique and complexity. However, the MDS-based algorithms have lower complexity and faster convergence when compared to other variants [5]. If a large number data points are available, methods for high-dimensional data may be more suitable than MDS, e.g. self-organizing maps [19]. They are based on neural networks and rely on feature analysis to reduce dimensionality. In this paper, the dimension of our data is relatively low, hence the self-organizing feature maps are not employed. Moreover, they do not preserve distance information (i.e. only topological). Other variants are Sammon's mapping [28] and curvilinear principle analysis [17]. The authors of [19] showed that a MDS-based tool is preferable for its trade off between complexity and ability to partition datasets. Moreover, MDS can process many types of data: nonnegative or negative, frequencies, correlations, ratings, etc. [3] and can optimally transform the data for better results. The main property of MDS that will be explored in this paper is that the distances between the points can be directly interpreted.

MDS is a generic name for a family of algorithms that construct a configuration of points in a low dimensional space from information about inter-point distances measured in high dimensional space [3]. The new geometrical configuration of points, which preserves the proximities of the high dimensional space, facilitates the perception of data's underlying structure and often makes it much easier to understand. The problem addressed by MDS can be stated as follows:

given  $n$  items in a  $m$ -dimensional space and an  $n \times n$  matrix of proximity measures among the items, MDS produces a  $p$ -dimensional configuration  $X$ ,  $p \leq m$ , representing the items such that the distances among the points in the new space reflect, with some degree of fidelity, the proximities in the data. The proximity measures the (dis)similarities among the items, and in general, it is a distance measure: the more similar two items are, the smaller their distance is. The Minkowski distance metric provides a general way to specify distance for quantitative data in a multidimensional space:

$$d_{ij} = \left( \sum_{k=1}^m w_k |x_{ik} - x_{jk}|^r \right)^{1/r} \quad (2)$$

where  $m$  is the number of dimensions,  $x_{ik}$  is the value of dimension  $k$  for object  $i$  and  $w_k$  is a weight. For  $w_k = 1$ , with  $r=2$ , the metric equals the Euclidean distance metric, while  $r=1$  leads to the city-block (or Manhattan) metric. In practice, the Euclidean distance metric is generally used, but there are several other definitions that can be applied, including for binary data [8]. Typically MDS is used to transform the data into two or three dimensions for visualizing the result to uncover data's hidden structure, but any  $p < m$  is also possible [7]. The geometrical representation obtained with MDS is indeterminate with respect to translation, rotation, and reflection [12]. There are two forms of MDS, namely the metric MDS and the nonmetric MDS. The metric MDS uses the actual values of dissimilarities, while nonmetric MDS effectively uses only their ranks. Metric MDS assumes that the dissimilarities  $\delta_{ij}$  calculated in the original  $m$ -dimensional data and distances  $d_{ij}$  in the  $p$ -dimensional space are related as follows:

$$d_{ij} \approx f(\delta_{ij}) \quad (3)$$

where  $f$  is a continuous monotonic function. Metric (scaling) refers to the type of transformation  $f$  of the dissimilarities and its form determines the MDS model. If  $d_{ij} = \delta_{ij}$  (it means  $f=1$ ) and a Euclidean distance metric is used we obtain the classical (metric) MDS. In metric MDS the dissimilarities between all objects are known numbers and they are approximated by distances. Thus objects are mapped into a low dimensional space, distances are calculated, and compared with the dissimilarities. Then objects are moved in such way that the fit becomes better, until an objective function (called stress function in the context of MDS) is minimized. In nonmetric MDS, the metric properties of  $f$  are relaxed but the rank order of the dissimilarities must be preserved. The transformation function  $f$  must obey the monotonicity constraint  $d_{ij} < d_{rs} \Rightarrow f(d_{ij}) = f(\delta_{rs})$  for all objects. The advantage of nonmetric MDS is that no assumptions need to be made about the underlying transformation function  $f$ . Therefore, it can be used in situations that only the rank order of dissimilarities is known (ordinal data). Additionally, it can be used in cases where there is incomplete information. In such cases, the configuration  $X$  is constructed from a subset of the distances, and, at the same time, the other (missing) distances are estimated by monotonic regression. In nonmetric MDS it is assumed that  $d_{ij} \approx f(\delta_{ij})$ , therefore  $f(\delta_{ij})$  are often referred as the disparities [22,18] in contrast to the original dissimilarities  $\delta_{ij}$ , on the one hand, and the dis-



tances  $d_{ij}$  of the configuration space on the other hand. In this context, the disparity is a measure of how well the distance  $d_{ij}$  matches the dissimilarity  $\delta_{ij}$ . There is no rigorous statistical method to evaluate the quality and the reliability of the results obtained by an MDS analysis. However, there are two methods often used for that purpose: the Shepard plot and the stress. The Shepard plot is a scatter plot of the dissimilarities and disparities against the distances, usually overlaid with a line having unitary slope. The plot provides a qualitative evaluation of the goodness of fit. On the other hand, the stress value gives a quantitative evaluation. Additionally, the stress plotted as a function of dimensionality can be used to estimate the adequate  $p$ -dimension. When the curve ceases to decrease significantly, the resulting “elbow” may correspond to a substantial improvement in fit.

### 2.5. Data analysis with MDS

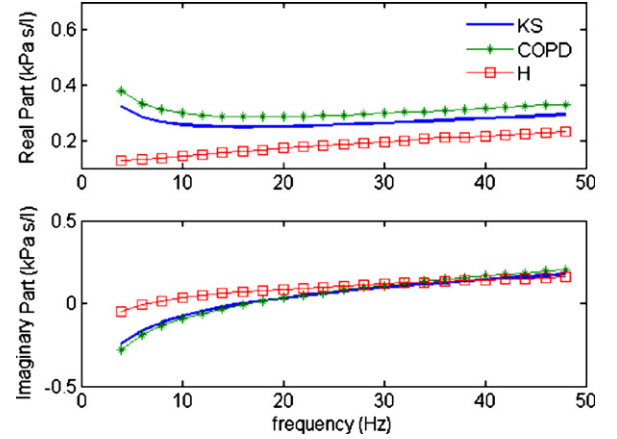
In order to obtain the MDS mapping, the respiratory impedances from all of the above groups have been calculated using relation (1), for the 4–48 Hz frequency interval in increments of 2 Hz. As a result, the impedance for each patient consists of a complex vector of real and imaginary parts, with 23 frequency points. Consequently, the distance between the real parts of the impedance  $Re$ , respectively the imaginary parts of the impedance  $Im$ , between various patients, can be calculated with some distance relations. We propose two such relations, as following:

$$D_1 = \sqrt{\sum_{k=1}^{M1,M2} [(Re_1 - Re_2)^2 + (Im_1 - Im_2)^2]_k} \quad (4)$$

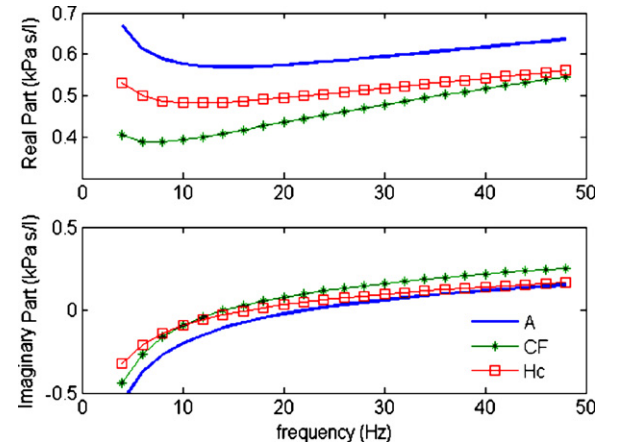
and

$$D_2 = \sqrt{\frac{\sum_{k=1}^{M1,M2} [(Re_1 - Re_2)^2 + (Im_1 - Im_2)^2]_k}{\sum_{k=1}^{M1,M2} [(Re_1 + Re_2)^2 + (Im_1 + Im_2)^2]_k}} \quad (5)$$

where both sums run for all distances between the patients in each group ( $M1$ , respectively  $M2$  are the total number of patients in each group used for calculating distance measurement). The  $(\aleph M1 + \aleph M2) \times (\aleph M1 + \aleph M2)$  symmetric MDS matrix is constructed with the values resulting from the calculus of the corresponding distances, where  $\aleph$  is the cardinal of the data set. The matrix can be visualized as a three dimensional plot, which takes an  $n \times n$  distance matrix  $D$ , and returns an  $n \times p$  configuration matrix  $Y$  [23]. Rows of  $Y$  are the coordinates of  $n$  points in  $p$ -dimensional space for some  $p < n$ . When  $D$  is a Euclidean distance matrix, the distances between those points are given by  $D$ . The variable  $p$  is the dimension of the smallest space in which the  $n$  points whose interpoint distances are given by  $D$  can be embedded. One can specify  $D$  as either a full dissimilarity matrix form of  $D$ , or in upper triangle vector form (such as, e.g. the output by PDIST in Matlab). A full dissimilarity matrix must be real and symmetric, and have zeros along the diagonal and positive elements everywhere else. A dissimilarity



**Fig. 2 – Values for the respiratory impedance in adults with KS: kyphoscoliosis, COPD: chronic obstructive pulmonary disease and in H: healthy adults.**



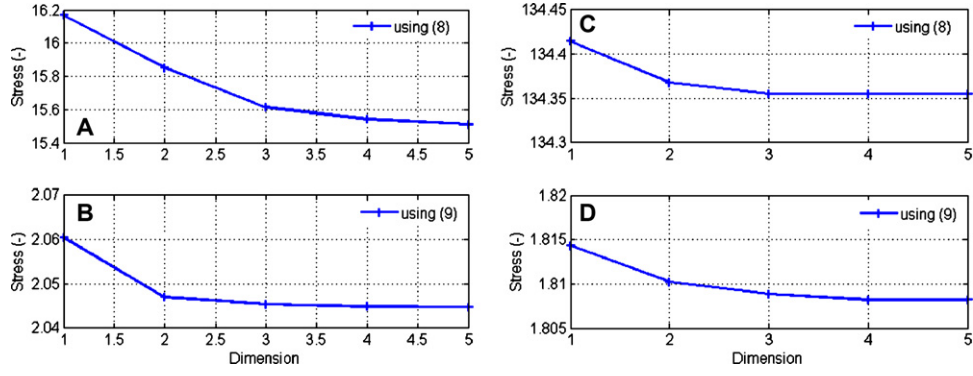
**Fig. 3 – Values for the respiratory impedance in children with A: asthma, CF: cystic fibrosis and in Hc: healthy children.**

ilarity matrix in upper triangle form must have real, positive entries. One can also specify  $D$  as a full similarity matrix, with ones along the diagonal and all other elements less than 1. In this application we use the dissimilarity matrix form for  $D$ .

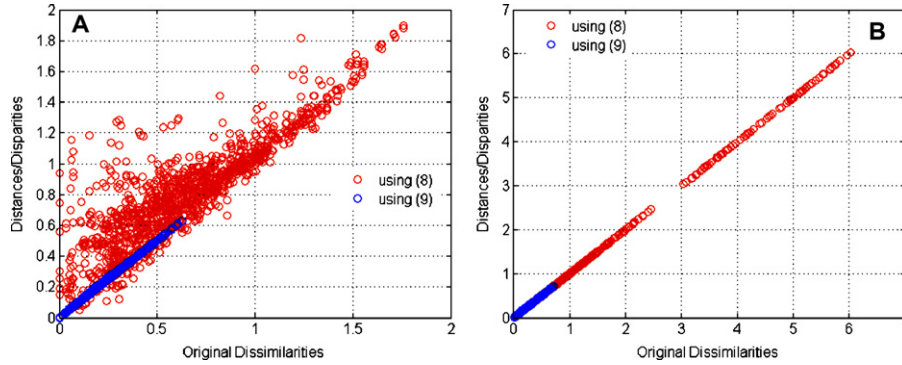
## 3. Results

Illustrative examples of the respiratory impedance in adults with COPD, kyphoscoliosis and healthy, are given in Fig. 2, by means of respective real and imaginary parts obtained with (1). Similarly, the respiratory impedance values in children are given in Fig. 3. Notice differences in real and imaginary parts, leading to differences in the MDS scaling matrix.

Before proceeding to the MDS plots, first we shall present the analysis for stress and Shepard plots. The stress plots are always decreasing and they have a *knee*, i.e. a change in the velocity of decreasing. That knee is usually adopted as an engineering compromise for deciding the number of dimensions to adopt in the MDS plot. In the stress plots we have for the



**Fig. 4 – Stress plots. The adult dataset using (4) in plot A and using (5) in plot B. The children dataset using (4) in plot C and using (5) in plot D.**



**Fig. 5 – Shepard plots for  $n=3$  with adult dataset (plot A) and with children dataset (plot B).**

adults in Fig. 4A, for (4), stress plot “knee” at  $n=3$  and in Fig. 4B for (5), stress plot “knee” at  $n=2$ . This suggests that (5) is “less demanding” in what concerns the MDS plot dimension, hence a low-dimension MDS plot. For children, in Fig. 4C we have that for (4), the stress plot “knee” is at  $n=3$ , while in Fig. 4D for (5) the stress plot “knee” is at  $n=4$ . In this case, the opposite situation arises, that (5) is more demanding. For the sake of uniformity, we shall choose  $n=3$  for all MDS plots.

The Shepard plots are depicted in Fig. 5A and B for (4) and (5), in adults and children, respectively. A Shepard plot is considered valid when the dots follow the  $45^\circ$  line. For example, if the dots present a low dispersion but following a curve (instead of the  $45^\circ$  line), it suggests that some non-linear relation exists. Usually, the Shepard plots are better the higher the dimension, but “stabilize” for the  $n$  value obtained from the stress plot. That is why the stress and the Shepard plots are usually “redundant”. From both figures we conclude that (5) delivers better dispersion than (4).

With the impedance data at hand, one can now assess the information from the MDS plots for  $n=3$ . The mapping obtained with the MDS for relations (4) are given in Figs. 6 and 7 for the comparison between healthy adults and those with COPD and KS; respectively in Figs. 8 and 9 for the comparison between healthy children and those with asthma and CF. Similar results are obtained with relation (5).

In Fig. 10, one can observe the MDS plot using all three groups of adults (healthy, COPD, KS) in one matrix. A clear separation between these groups is obtained using (4), but not

when using (5). The choice of distance metrics is therefore important when more than two groups are compared.

Further on, the mapping obtained with the MDS for relations (4) is given in Fig. 11, respectively for (5) in Fig. 12, for all three children groups (healthy, asthma and cystic fibrosis).

Dendrograms are often used for displaying relationships among clusters. A dendrogram shows the multidimensional distances between objects in a tree-like structure. Objects which are closest to each other in the multidimensional data space are connected by a horizontal line, forming a cluster which can be regarded as a “new” object. The new cluster and the remaining original data are again searched for the closest pair, and so on. The distance of the particular pair of objects (or clusters) is reflected in the height of the horizontal line. Therefore, for comparison with MDS, using the same measure (4), the corresponding dendrograms for adults and children are given in Figs. 13 and 14, respectively. From all possible variations, the dendrogram using unweighted average method provided reasonable clustering results. It is worth noticing that when applied in paired-sets, the dendrogram clustering is more homogeneous and can be easily applied to clinical use.

#### 4. Discussion

The frequency where the imaginary part of the impedance  $Im(Z_r)=0$  is called the resonance frequency  $F_{res}$  and it depends

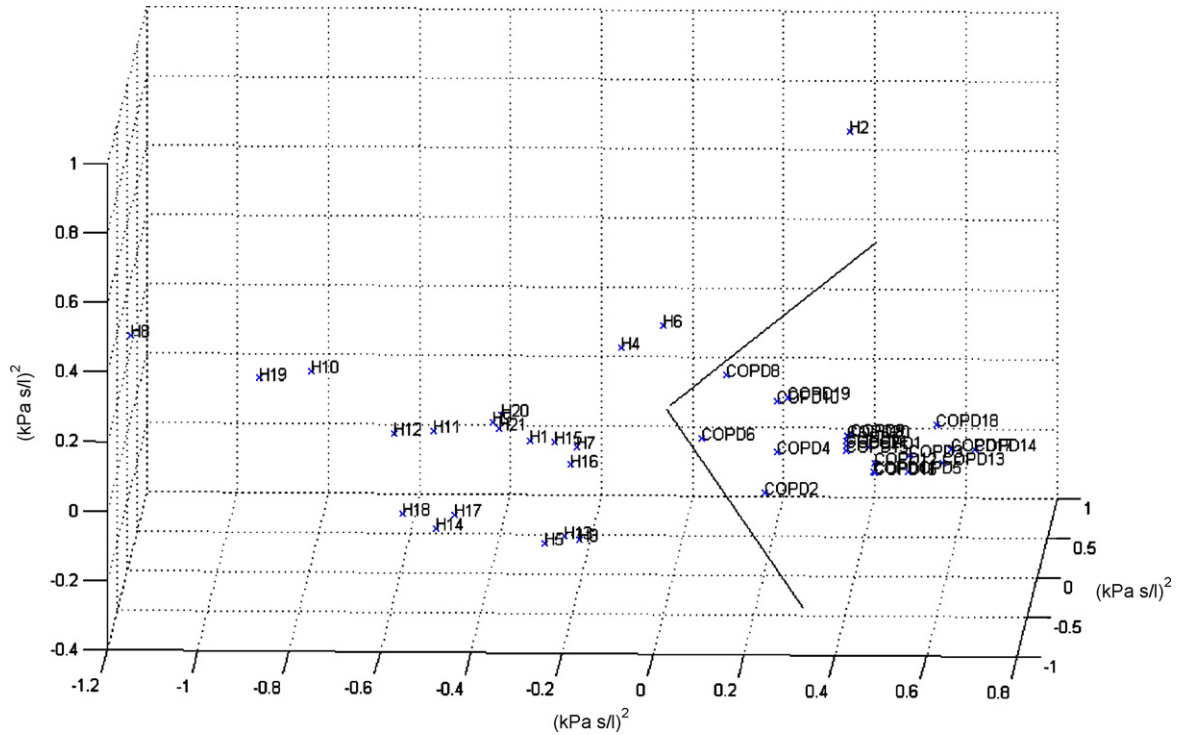


Fig. 6 – In adults: MDS plot, using expression (4), for healthy (H) vs chronic obstructive pulmonary disease (COPD), providing a perfect separation between the groups.

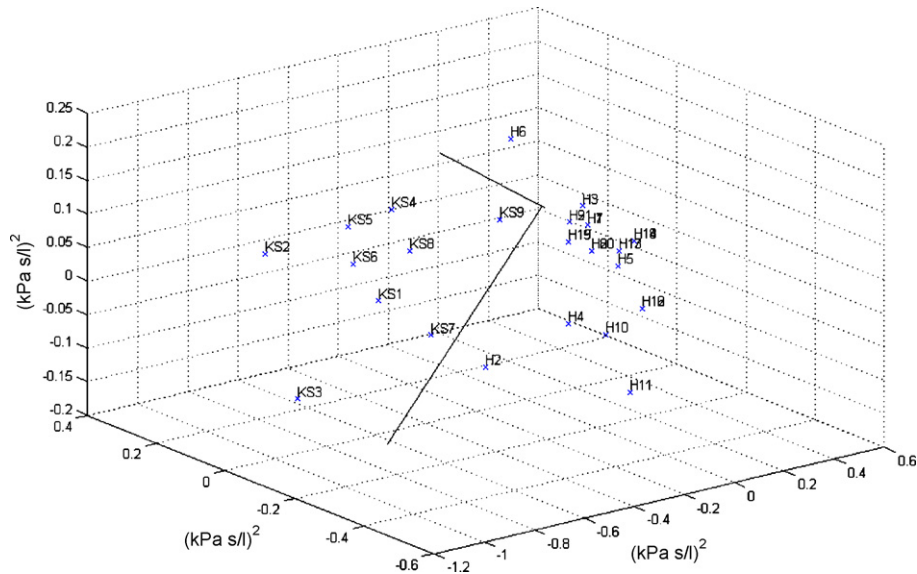
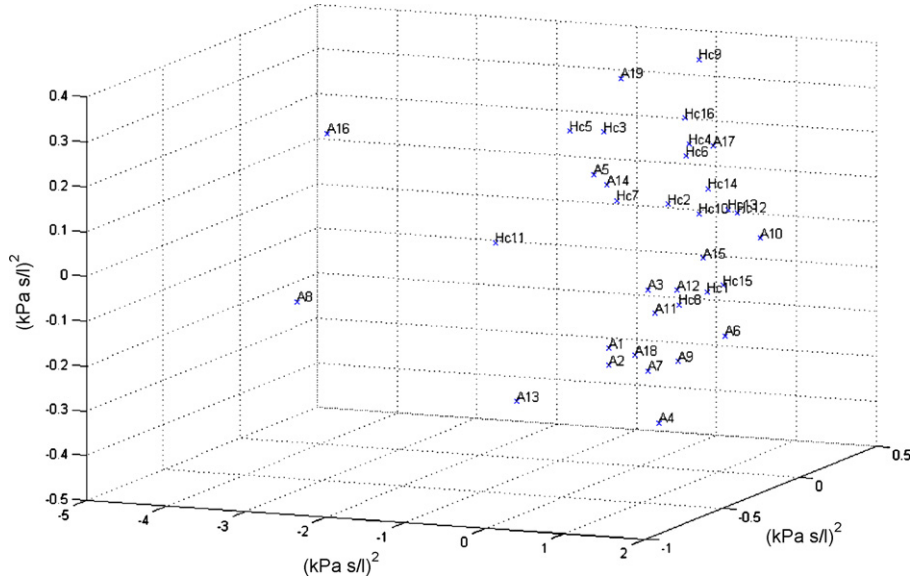


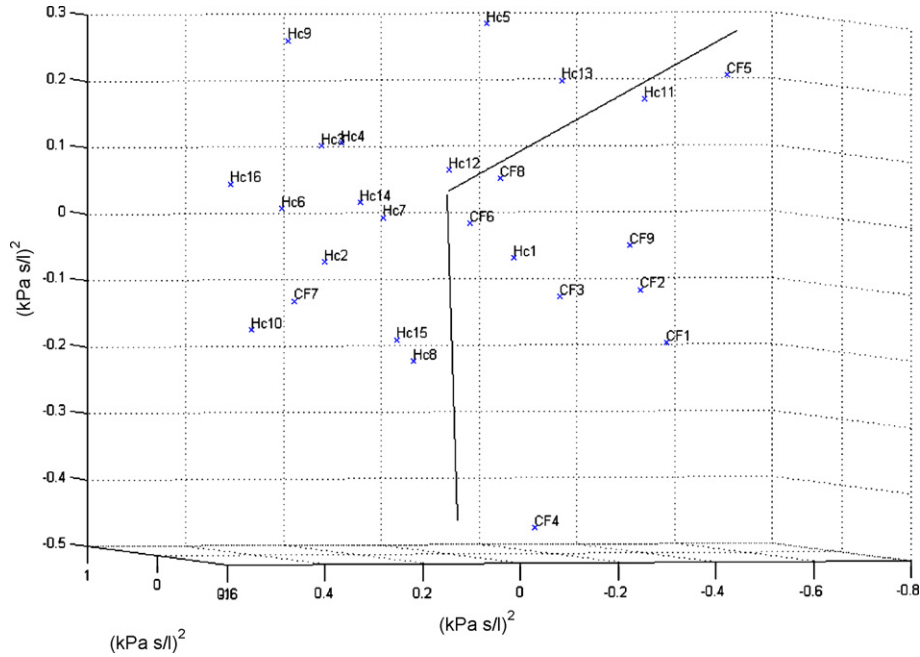
Fig. 7 – In adults: MDS plot, using expression (4), for healthy (H) vs kyphoscoliosis (KS), providing a perfect separation between the groups.

Table 3 – Resonant frequency ( $F_{res}$ ) in Hz and real part of impedance at 6 Hz (R6) in kPa s/l; values are given as mean  $\pm$  standard deviation; values in brackets indicate the corresponding 95% confidence intervals.

	$F_{res}$	R6
Healthy adults	10.48 $\pm$ 3.56 (8.75, 13.87)	0.13 $\pm$ 0.05 (0.11, 0.16)
COPD	20.58 $\pm$ 8.98 (11.89, 30.27)	0.33 $\pm$ 0.07 (0.29, 0.36)
Kyphoscoliosis	15.01 $\pm$ 2.08 (12.80, 18.02)	0.28 $\pm$ 0.06 (0.23, 0.33)
Healthy children	21 $\pm$ 5.9 (17.81, 24.18)	0.49 $\pm$ 0.06 (0.46, 0.53)
Asthma	22.94 $\pm$ 12.29 (17.01, 28.87)	0.61 $\pm$ 0.37 (0.42, 0.8)
Cystic fibrosis	15.75 $\pm$ 4.71 (11.80, 19.69)	0.38 $\pm$ 0.08 (0.32, 0.45)



**Fig. 8 – In children: MDS plot, using expression (4), for healthy (Hc) vs asthma (A); not a clear separation due to controlled asthma medication prior to the lung function test.**



**Fig. 9 – In children: MDS plot, using expression (4), for healthy (Hc) vs cystic fibrosis (CF); relatively good separation, with 1–2 outliers.**

on the balance between the different kinds of mechanical properties (elastic, inertial). This then allows for differentiating between healthy and pathologic cases, since the resonance frequency changes significantly from typically 8 Hz for a healthy adult to 14 Hz for a patient with mild airway obstruction, and about 20 Hz in cases of severe obstruction [25]. The values obtained for the respiratory impedance data and resonant frequency have been compared with reported values from the literature. We found that the values were very close to those reported in literature, namely for healthy adults [25];

for COPD [1]; for asthmatic children [6]; for cystic fibrosis [4,11]; for healthy children [10] and for kyphoscoliosis [24]. Moreover, the impedance values and shape were significantly close to those identified with several parametric models, such as in [16,9]. Additionally, we also looked at the real part of impedance values at 6 Hz, i.e.  $R_6$ , which were also close to the reported and expected values from literature. Table 3 enumerates all averaged values and their standard deviation values for each group of patients discussed in this paper, for  $F_{res}$  and  $R_6$ .





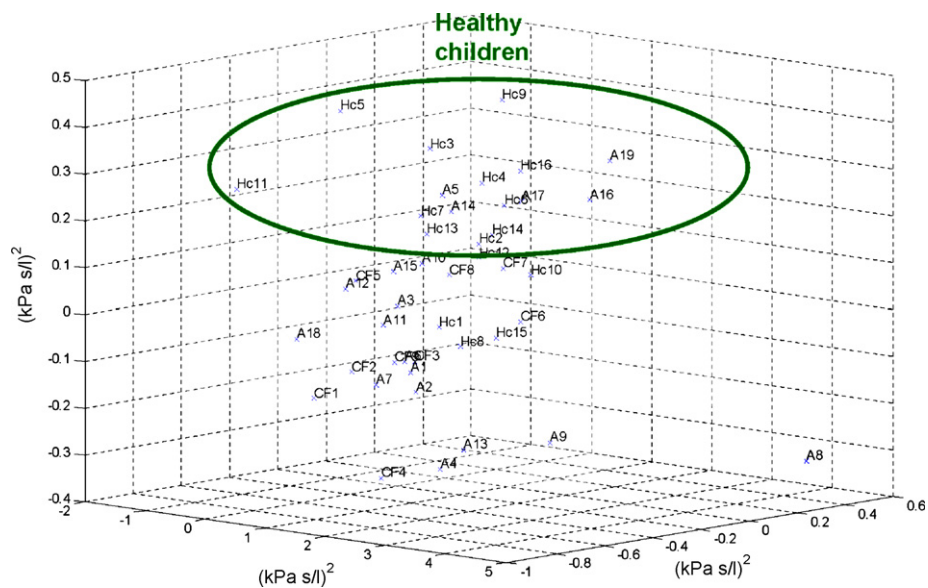


Fig. 12 – The MDS results for the second sum, in children, thus using (5).

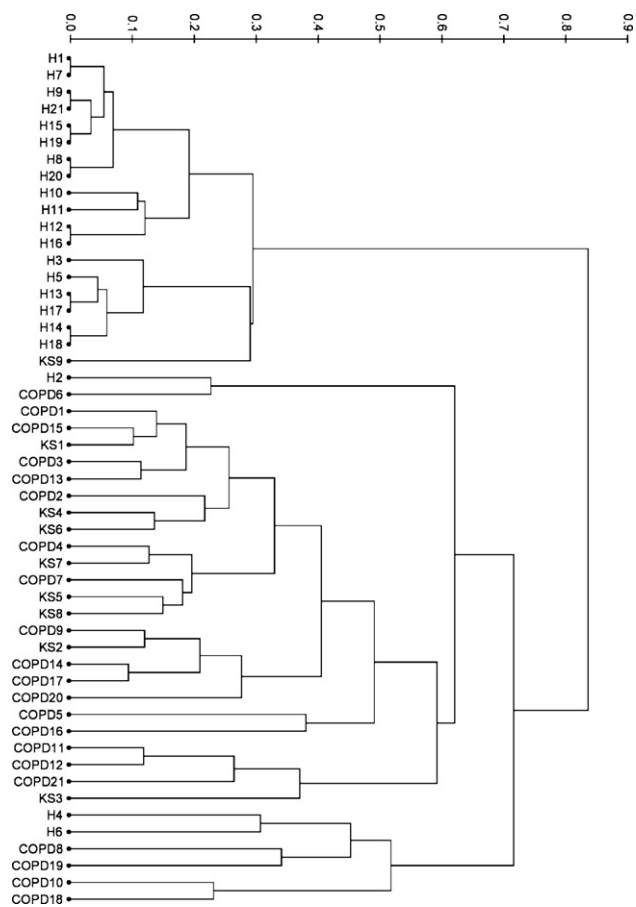


Fig. 13 – The dendrogram tree clustering for the adult dataset using (4), in healthy (H), chronic obstructive pulmonary disease (COPD) and kyphoscoliosis (KS).

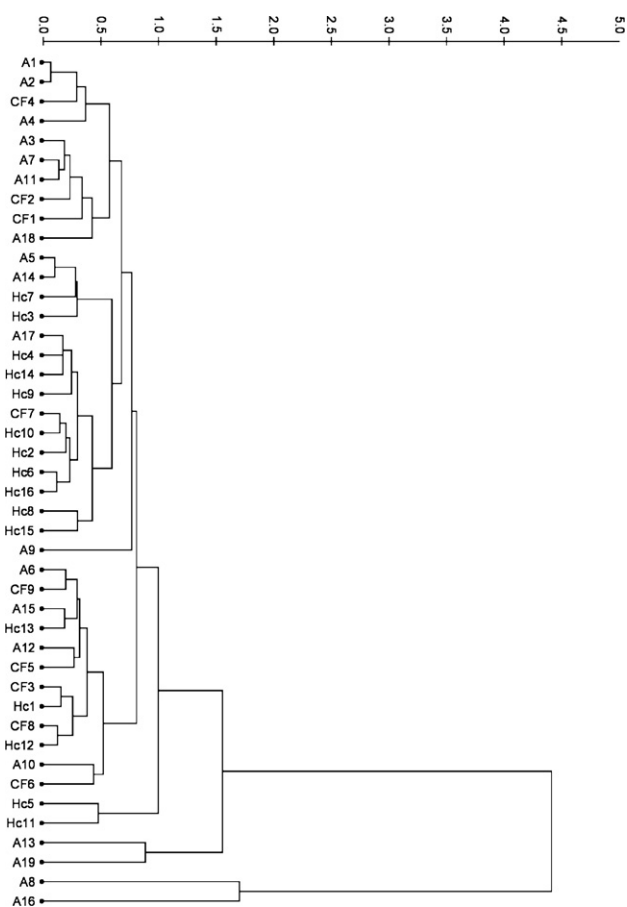


Fig. 14 – The dendrogram tree clustering for the children dataset using (4), in healthy (Hc), asthma (A) and cystic fibrosis (CF).

is reduced due to obstruction to air escape. Thus, the  $FEV_1/FVC$  ratio will be reduced. In restrictive lung disease, the  $FEV_1$  and FVC are equally reduced due to fibrosis or other lung pathology (not obstructive pathology). Thus, the  $FEV_1/FVC$  ratio should be approximately normal. From the above, we observe that the spirometric values given in Tables 1 and 2 confirm the a priori diagnosis.

#### 4.1. Adult dataset

**Chronic pulmonary emphysema** literally means excess air in the lungs [1]. It results from three major pathophysiological events in the lung:

- chronic infection, caused by inhaling smoke or other substances that irritate the bronchi and bronchioles;
- the infection, the excess of mucus, and inflammatory edema of the bronchiolar epithelium together cause chronic obstruction of smaller airways;
- the obstruction of the airways makes it especially difficult to expire, causing entrapment of air in the lungs (barrel chest effect) and over-stretching the alveoli.

The bronchiolar obstruction causes increased airway resistance and results in greatly increased work of breathing.

On the other hand, **kyphoscoliosis** is a respiratory failure resulted from a deformation of the spine, as a combination effect of scoliosis and kyphosis [24]. In the sense of a deformity, it is the pathological curving of the spine, where parts of the spinal column lose some or all of their normal profile. This causes a bowing of the back, seen as a slouching back and breathing difficulties. Severe cases can cause great discomfort and even lead to death. As a result of these deformities at the spinal level, the thorax cannot perform its normal function, leading to changes in airway resistance and total lung compliance. It is therefore classified as a restrictive disease.

From Fig. 10 we notice that the decision of using one or another distance form is important for the mapping representation. With the sum given by (4) the separation between the groups is possible, whereas using (5) it becomes impossible to strictly separate between the groups. One can clearly distinguish separated high density nuclei for each of the three groups: healthy, COPD and KS. This means that it is possible to classify between these groups by means of MDS, given the suitable distance measure. Some COPD outliers are present, denoting marked progress of the obstructive disease, whereas majority of the KS objects lie fairly close to the nuclei of COPD. The reason for this similitude is that, although being a restrictive disease, KS affects in a similar manner the airway resistance. In spite of having different origins, airway resistance increases in both COPD and KS. In a similar manner, the compliance is lower in both COPD and KS. Nevertheless, different balance between these mechanical properties will place an object in the MDS map closer or further from the nucleus of the group. This is usually the case with significant pathologic restrictions, hence a more pronounced manifestation of the disease. Our results suggest that MDS is able to distinguish between restrictive (KS) and obstructive (COPD) pathologies, for clinical classification purposes. Moreover, the correspond-

ing dendrogram tree from Fig. 13 supports this conclusion, by similar clustering.

#### 4.2. Children dataset

**Asthma** is characterized by spastic contraction of the bronchioles, which causes extremely difficult breathing [6]. The usual cause is bronchial hyper-responsiveness towards a variety of specific and a-specific stimuli. As a result of the irritants, the allergic person has a tendency to produce a high amount of antibodies, which attach to specific cells in the bronchioles and small bronchi. As a result of the antibodies reaction with the irritant, some substances are released (e.g., histamine).

**Cystic fibrosis** is an inherited disease characterized by an abnormality in the glands that produce sweat and mucus [11]. Cystic fibrosis affects various systems in children and young adults, including the following: respiratory system, digestive system and the reproductive system.

In children, from Figs. 11 and 12 we can observe that some of the asthma patients were fairly close to the healthy subjects nucleus. The underlying reason for this result is that these asthma patients were controlled with specific medication, with intake prior to the lung function test. Hence, no significant differences from the healthy group can be seen, i.e. false 'healthy points'. The other asthma patients, who had exacerbations and a pronounced respiratory restriction, are either partially controlled, or did not take the medication prior to the exam. As far as cystic fibrosis is concerned, due to the nature of the disease, which affects in general the organism and not only the respiratory system, a clear separation could not be made from the other objects in the MDS map. Although well grouped, the nuclei of the CF lie within the surface of the other two groups (healthy children and asthma). Our results suggest that MDS may not be able to distinguish between pathologies with similar manifestations (both asthma and cystic fibrosis show the presence of mucus, etc.). And that in children, the difficulty of diagnostic remains more pronounced than in adults. This conclusion is supported in Fig. 14 depicting the corresponding dendrogram tree.

## 5. Conclusions

In this paper an MDS analysis has been performed on respiratory impedance data from six groups of patients: healthy, asthmatic and cystic fibrosis in children, respectively healthy, chronic obstructive pulmonary disease and kyphoscoliosis in adults. The results suggest the usefulness of using MDS for mapping purposes, describing (dis)similarities within groups of patients and between groups of patients.

For the adult database, our findings suggest that MDS is able to distinguish between restrictive (KS) and obstructive (COPD) pathologies. In children, MDS may not be able to distinguish between pathologies with similar manifestations (both asthma and cystic fibrosis show the presence of mucus, etc.).

The MDS mapping is, in general, a useful tool to characterize the nature of the disease compared to healthy subjects and if large data is available, it might reveal more homogeneous clustering for several pathologies and several degrees of respiratory diseases (e.g. follow up studies). Given specific

areas for typical pathologic results, one can verify the status (distance from reference nucleus) of the patient within the group.

The MDS can be further investigated for detection of separation areas for various pathologies; however, since the current paper might serve as a proof-of-concept, the number of investigated patients must be significant in order to provide statistically significant results for clinical use.

## Acknowledgments

The authors would like to thank Miguel Lima for his helpful comments about the adoption of MDS tools. C. Ionescu gratefully acknowledges the students from Ghent University and the children from primary school in Zwijnaarde who volunteered to perform lung function testing in our laboratory. We also acknowledge the technical assistance provided at University of Pharmacology and Medicine – “Leon Daniello” Cluj, Romania to measure COPD diagnosed patients, University Hospital Gent, Belgium, to measure KS-diagnosed patients and University Hospital Antwerp, Belgium to measure asthma and CF diagnosed children.

## REFERENCES

- [1] P.J. Barnes, Chronic obstructive pulmonary disease, *NEJM Med. Prog.* 343 (2) (2000) 269–280.
- [2] M. Birch, D. MacLeod, M. Levine, An analogue instrument for the measurement of respiratory impedance using the forced oscillation technique, *Phys. Meas.* 22 (2001) 323–339.
- [3] I. Borg, P. Groenen, *Modern Multidimensional Scaling—Theory and Applications*, 2nd edition, Springer-Verlag, New York, 2005.
- [4] S. Brennan, G. Hall, F. Horak, A. Moeller, P. Pitrez, A. Franzmann, S. Turner, N. de Clerck, P. Franklin, K. Winfield, E. Balding, S. Stick, P. Sly, Correlation of forced oscillation technique in preschool children with cystic fibrosis with pulmonary inflammation, *Thorax* 60 (2005) 159–163.
- [5] M. Bronstein, A. Bronstein, R. Kimmel, I. Yavneh, Multigrid multidimensional scaling, *Numer. Linear Algebra Appl.* 13 (2006) 149–171.
- [6] W. Busse, R. Lemanske, Asthma, *New Engl. J. Med.* 344 (5) (2001) 350–362.
- [7] M. Carreira-Perpinan, A review of dimension reduction techniques, Technical report CS-96-09, Department of Computer Science, University of Sheffield, 1997.
- [8] T. Cox, M. Cox, *Multidimensional Scaling*, 2nd edition, Chapman Hall/CRC, 2001, ISBN 1-58488-094-5.
- [9] B. Diong, A. Rajagiri, M. Goldman, H. Nazerah, The augmented RIC model of the human respiratory system, *Med. Biol. Eng. Comput.* 47 (2009) 395–404.
- [10] E. Duiverman, J. Clement, K. Van de Woestijne, H. Neijens, A. van den Bergh, K. Kerrebijn, Forced oscillation technique: reference values for resistance and reactance over a frequency spectrum of 2–26 Hz in healthy children aged 2.3–12.5 years, *Clin. Respir. Physiol.* 21 (1985) 171–178.
- [11] A. Elizur, C. Cannon, T. Ferkol, Airway inflammation in cystic fibrosis, *Chest* 133 (2) (2008) 489–495.
- [12] I. Fodor, A survey of dimension reduction techniques, Technical report, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, 2002.
- [13] K. Friston, C. Frith, P. Fletcher, P. Liddle, R. Frackowiak, Functional topography: multidimensional scaling and functional connectivity in the brain, *Cerebral Cortex* 6 (March/April) (1996) 156–164.
- [14] J.L. Hernandez, R. Biscay, J.C. Jimenez, P. Valdes, R. Grave de Peralta, Measuring the dissimilarity between EEG recordings through a nonlinear dynamical system approach, *Int. J. Bio-Med. Comput.* 38 (1995) 121–129.
- [15] C. Ionescu, R. De Keyser, A novel parametric model for the human respiratory system, in: *Proceedings of the IASTED International Conference on Modelling and Simulation*, USA, Acta Press, Anaheim, 2003, pp. 246–251.
- [16] C. Ionescu, R. De Keyser, Parametric models for characterizing respiratory input impedance, *J. Med. Eng. Technol.* 32 (4) (2008) 315–324.
- [17] J. Lee, A. Lendase, N. Donkers, M. Verleysen, A robust nonlinear projection method, in: *Proceedings of the ESANN’2000, 8th European Symposium on Artificial Neural Networks*, M.D-Facto Public, Bruges, Belgium, 2000, pp. 13–20.
- [18] J. Leeuw, M. Patrick, Multidimensional scaling using majorization: SMACOF, *R. J. Stat. Softw.* 31 (i03), Available online: <http://www.jstatsoft.org/v31/i03/paper>.
- [19] J. Li, Visualization of high-dimensional data with relational perspective map, *Inform. Visual.* 3 (2004) 45–59.
- [20] M. Mamlouk, C. Chee-Ruiter, U. Hofmann, J. Bower, Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps, *Neurocomputing* 52–54 (2003) 591–597.
- [21] M. Mani, A. Srivastava, C. Barillot, The labeling of cortical sulci using multidimensional scaling, Presented at MICCAI Workshop, Manifolds in Medical Imaging: Metrics, Learning and Beyond, NY, USA, 2008, 9 pp.
- [22] W. Martinez, A. Martinez, *Exploratory Data Analysis with MATLAB*, Chapman Hall/CRC Press, UK, 2005, ISBN 1-58488-366-9.
- [23] *Mathworks – Matlab: Statistics Toolbox*; v. 7.2, 2006, <http://www.mathworks.com/products/statistics/>.
- [24] F. McCool, D. Rochester, Non-muscular diseases of the chest wall, in: A. Fishman (Ed.), *Fishman’s Pulmonary Disease and Disorders*, vol. II, McGraw Hall Medical, NY, 2008, pp. 1541–1548.
- [25] E. Oostveen, D. Macleod, H. Lorino, R. Farré, Z. Hantos, K. Desager, F. Marchal, The forced oscillation technique in clinical practice: methodology, recommendations and future developments, *Eur. Respir. J.* 22 (2003) 1026–1041.
- [26] H. Pasker, M. Peeters, P. Genet, N. Nemery, K. Van De Woestijne, Short-term ventilatory effects in workers exposed to fumes containing zinc oxide: comparison of forced oscillation technique with spirometry, *Eur. Respir. J.* 10 (1997) 523–529.
- [27] D. Polzella, G. Reid, Multidimensional scaling analysis of simulated air combat maneuvering performance data, *Aviat. Space Environ. Med.* (February) (1989) 141–144.
- [28] W. Sammon, A nonlinear mapping algorithm for data structure analysis, *IEEE Trans. Comput.* C18 (1969) 401–409.
- [29] P. Quanjer, Referentiewaarden, Longfunctie Onderzoek, Leuven, Garant, Ed. Demedts, Decramer, 1998, pp. 27–37 (in Dutch).
- [30] J. Tzeng, H. Shing-Lu, W. Hsiung-Li, Multidimensional scaling for large genomic data sets, *BMC Bioinformatics* 9 (2008) 179, Available online (Open Access): <http://www.biomedcentral.com/1471-2105/9/179>.
- [31] S. Vernon, E. Unger, I. Dimulescu, M. Rajeevan, W. Reeves, Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome, *Disease Markers* 18 (2002) 193–199.
- [32] D. Welchew, D. Honey, T. Sharma, T. Robbins, E. Bullmore, Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnection disorder, *Neuro-Image* 17 (2002) 1227–1239.