

# On the trade-off between compression efficiency and distortion of a new compression algorithm for multichannel EEG signals based on singular value decomposition

Giuseppe Campobello<sup>1</sup>, Giovanni Gugliandolo<sup>1</sup>, Angelica Quercia<sup>2</sup>, Elisa Tatti<sup>3</sup>, Maria Felice Ghilardi<sup>3</sup>, Giovanni Crupi<sup>2</sup>, Angelo Quartarone<sup>2</sup>, Nicola Donato<sup>1</sup>

<sup>1</sup> Department of Engineering, University of Messina, Contrada di Dio, S. Agata, 98166 Messina, Italy

<sup>2</sup> BIOMORF Department, University of Messina, AOU "G. Martino", Via C. Valeria 1, 98125, Messina, Italy

<sup>3</sup> CUNY School of Medicine, CUNY, 160 Convent Avenue, New York, NY 10031, USA

## ABSTRACT

In this article we investigate the trade-off between the compression ratio and distortion of a recently published compression technique specifically devised for multichannel electroencephalograph (EEG) signals. In our previous paper, we proved that, when singular value decomposition (SVD) is already performed for denoising or removing unwanted artifacts, it is possible to exploit the same SVD for compression purpose by achieving a compression ratio in the order of 10 and a percentage root mean square distortion in the order of 0.01 %. In this article, we successfully demonstrate how, with a negligible increase in the computational cost of the algorithm, it is possible to further improve the compression ratio by about 10 % by maintaining the same distortion level or, alternatively, to improve the compression ratio by about 50 % by still maintaining the distortion level below the 0.1 %.

**Section:** RESEARCH PAPER

**Keywords:** Biomedical signal processing; electroencephalograph (EEG); EEG measurements; near-lossless compression; singular value decomposition (SVD)

**Citation:** Giuseppe Campobello, Giovanni Gugliandolo, Angelica Quercia, Elisa Tatti, Maria Felice Ghilardi, Giovanni Crupi, Angelo Quartarone, Nicola Donato, On the trade-off between compression efficiency and distortion of a new compression algorithm for multichannel EEG signals based on singular value decomposition, Acta IMEKO, vol. 11, no. 2, article 30, June 2022, identifier: IMEKO-ACTA-11 (2022)-02-30

**Section Editor:** Francesco Lamonaca, University of Calabria, Italy

**Received** October 24, 2021; **In final form** February 22, 2022; **Published** June 2022

**Copyright:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Corresponding author:** Giuseppe Campobello, e-mail: [gcampobello@unime.it](mailto:gcampobello@unime.it)

## 1. INTRODUCTION

Since its invention by the German psychiatrist Hans Berger almost a century ago [1], electroencephalography (EEG) has continuously evolved, becoming a powerful and extensively used method that allows measuring safely and noninvasively the spatiotemporal dynamics of the brain activity with a high temporal resolution in the range of milliseconds, which enables detecting rapid changes in the brain rhythms [2]. The brain rhythms are the periodic fluctuations of human EEG, which are associated with cognitive processes, physiological states, and neurological disorders [3]. Hence, the use of EEG can range from basic research to clinical applications [4]. Among the various applications of EEG, it is worth highlighting its recent application in brain computer interface (BCI) research [5], [6].

EEG consists of a neurophysiological measurement of the electrical activity generated by the brain through multiple

electrodes placed on the scalp surface. EEG data are measured as the electrical potential difference between two electrodes: active and reference electrodes. At neurophysiological level, the electrical potential differences are mostly generated by the summation of both excitatory and inhibitory post-synaptic potentials in tens of thousands of cortical pyramidal neurons that are synchronously activated [7]. Hence, the brain sources of the electrical potentials recorded by EEG may be suited to an infinite number of configurations, thereby limiting the spatial resolution of scalp EEG. To overcome this drawback, several source localization methods have been proposed and their application with high-density EEG (HD-EEG) systems, such as 64-256 electrodes, can lead to a remarkable improvement in EEG spatial resolution [3], [8].

Many applications require that EEG systems have to record continuously for several days or even weeks and this might easily yield several gigabytes (GB) of generated data, which makes compression algorithms necessary for efficient data handling. As

an illustrative example, about 2.6 GB of data per day are generated by an EEG system recording the data from 64 electrodes with a sampling rate of 250 Hz and a 16-bit resolution. It should be mentioned that intracranial EEG recordings can generate even terabytes (TB) of data per day [9]. Therefore, EEG data need to be largely compressed to efficiently manage their storage. Furthermore, data compression is necessary also to reduce both transmission rate and power consumptions when telemonitoring EEG via wireless [10], [11]. For instance, wireless wearable EEG systems for long-term recordings should operate under a low-power budget, due to limitation on battery lifetime, and then the power consumption needs to be significantly reduced by compressing the data before transmission [12].

Various EEG compression algorithms have been developed to minimize the number of bits needed to represent EEG data by exploiting inter and/or intra-channel correlations of EEG signals.

EEG compression algorithms can be classified into two main categories: lossless and lossy compression [13], [14]. As the main goal of the compression algorithms is to reduce the size of the data, their performance is, typically, evaluated by using the compression ratio (CR), which is calculated as the ratio between the number of bits required to represent the original and compressed EEG data.

Generally, lossy compression enables superior compression performance compared to the lossless counterpart but it cannot guarantee a reconstruction of the exact original data from the compressed version. In such a case, the percent root-mean-square distortion (PRD) is used as indicator for assessment of the quality of the reconstructed signal, which is affected by the distortion introduced by the lossy compression. Typically, lossless compression algorithms are preferred in clinical practice to avoid diagnostic errors, since important medical information may be disregarded using lossy compression and, in addition, there is a lack of legislation and/or approved standards on lossy compression, making EEG reconstruction a more critical requirement than compression performance. On the other hand, the lossless compression approach has a limited impact on storage requirements for EEG applications. As a matter of fact, the use of the state-of-the-art lossless compression algorithms allows achieving typical compression ratios in the order of 2 or 3 [15]-[21].

On the other hand, EEG signals are of very small amplitude, typically in the order of microvolts ( $\mu\text{V}$ ), and thus they can be easily contaminated by noise and artifacts, which should be filtered to highlight and/or extract the actual clinical information [22], [23]. To accomplish this task, digital filters and denoising procedures based on wavelets, principal component analysis (PCA) and/or independent component analysis (ICA) are often used [24]-[28].

This enables the development of near-lossless PCA/ICA-based compression algorithms that can achieve much higher compression ratios than those obtained with lossless compression algorithms with a tolerable reconstruction distortion for the application of interest. Different near-lossless EEG compression schemes based on parallel factor decomposition (PARAFAC) and singular value decomposition (SVD) have been investigated and compared with wavelet-based compression techniques [29]. In most cases, PARAFAC leads achieving better compression performance but the maximum CR obtained with a PRD lower than 2 % was 4.96 [29]. A near-lossless algorithm able to obtain a CR of 4.58 with a PRD in the range between 0.27 % and 7.28 %, depending on the specific

dataset under study, has been proposed in [30]. More recently, a SVD-based compression scheme able to obtain 80 % data compression (i.e., CR = 5) with a PDR of 5 % has been reported in [31].

More recently, in [32] we proposed a near-lossless compression algorithm for EEG signals able to achieve a compression ratio in the order of 10 with a  $PRD < 0.01\%$ . In particular, the algorithm has been specifically devised for achieving a very low distortion in comparison to other state-of-the-art solutions. In this paper, we present an improved version of our previous algorithm and particular attention is given to achieve a good trade-off between compression efficiency and distortion.

The rest of this paper is organized as follows. In Section II, we briefly review SVD and describe our original algorithm proposed in [32]. In Section III, we illustrate the proposed algorithm. In Section IV, we present the experimental results obtained on a real-world EEG dataset. Finally, future works and our conclusions are drawn in Section V.

## 2. SINGULAR VALUE DECOMPOSITION

EEG signals are easily contaminated with artifacts and noise and, therefore, they need to be filtered before extracting the actual clinical information. For this purpose, SVD-based PCA and ICA techniques are commonly used.

In order to briefly review how SVD is exploited in this context, let us consider a high-density  $N$ -channel EEG system whose signals are sampled for a time interval  $T$  and at a rate of  $f_s$  samples per second (sp/s). In this case, we have  $M = T \cdot f_s$  samples per channel and thus an overall number of samples equal to  $N \cdot M$ . We assume that such as samples are represented by a  $N \times M$  matrix  $A$ .

It is known from the SVD theory that it possible to decompose a matrix  $A$  into three matrices  $U$ ,  $\Sigma$ , and  $V$ , such that  $A = U\Sigma V^T$ . In particular,  $\Sigma$  is a diagonal matrix whose diagonal elements, i.e.,  $\sigma_i$  with  $i \in [1, \dots, N]$ , are named singular values. Moreover, a rank  $k$  approximation of  $A$ , i.e.,  $A_k = U_k \Sigma_k V_k^T$ , exists which minimizes the norm  $\|A - A_k\|$  and that can be obtained by considering the submatrices  $U_k$  and  $V_k$ , given by the first  $k$  columns of  $U$  and  $V$ , respectively, and the leading principal minor of order  $k$  of  $\Sigma$ , i.e.,  $\Sigma_k$ , containing the first  $k < N$  singular values.

In the specific context of EEG, the desired rank  $k$ , and thus the number of singular values exploited for approximation, is chosen by clinicians, or other EEG experts, in order to reduce the effect of undesired artifacts and noise by keeping unaltered the clinical information.

In this case the actual clinical information is contained in  $A_k$  and, with the aim of reducing storage resources that are needed to store EEG samples, it is mandatory to encode the matrix  $A_k$  in the most efficient manner.

In [32] authors proposed a solution for the above problem by deriving the near-lossless compression algorithm, as reported in Figure 1.

The basic idea of the algorithm is to decompose the matrix  $A_k$  into two matrices,  $X_k$  and  $Y_k$ , such that  $A_k = X_k Y_k$ . In particular, the matrices  $X_k$  and  $Y_k$  can be obtained, as shown in step 2, by first evaluating the matrix  $S = \Sigma^{1/2}$  and then considering the first  $k$  columns of the matrix  $US$  and the first  $k$  rows of the matrix  $SV^T$ , i.e.,  $X_k = (US)[:, 1:k]$  and  $Y_k =$

- INPUTS: an integer number  $k < N$  and a matrix  $A$ , formed by  $N \times M$  EEG samples;
- OUTPUTS: integer matrices  $\tilde{X}_k$  and  $\tilde{Y}_k$  and the scale factor  $s = m_X \cdot m_Y$ .
- ALGORITHM:
  - 1) Use SVD to decompose  $A$  as  $A = U\Sigma V^T$
  - 2) Obtain the matrices  $S = \Sigma^{1/2}$ ,  
 $X_k = (US)[:, 1:k]$  and  $Y_k = (SV^T)[1:k, :]$
  - 3) Evaluate  $m_X = \max(|X_k|)$ ,  $m_Y = \max(|Y_k|)$
  - 4) Calculate  $s = m_X \cdot m_Y$ ,  
 $\tilde{X}_k = \text{round}(m_Y \cdot X_k)$  and  $\tilde{Y}_k = \text{round}(m_X \cdot Y_k)$

Figure 1. Illustration of the compression algorithm proposed in [32].

$(SV^T)[1:k, :]$  in Matlab-like notation. Successively (see step 3), maximum absolute values of the matrices  $X_k$  and  $Y_k$ , i.e.,  $m_X = \max(|X_k|)$  and  $m_Y = \max(|Y_k|)$ , are evaluated. Such values are used in the last step, i.e., step 4, to transform the floating-point matrices  $X_k$  and  $Y_k$  into two integer matrices,  $\tilde{X}_k$  and  $\tilde{Y}_k$ , on the basis of the following equations:

$$\begin{aligned} \tilde{X}_k &= \text{round}(m_Y \cdot X_k) \\ \tilde{Y}_k &= \text{round}(m_X \cdot Y_k). \end{aligned} \quad (1)$$

Note that the  $\text{round}()$  operator in above equations is the usual rounding operator, i.e., it rounds a floating point number to the nearest integer number.

It is worth observing that actual dimensions of the matrices  $A_k$ ,  $\tilde{X}_k$ , and  $\tilde{Y}_k$  are,  $N \times M$ ,  $N \times k$ , and  $k \times M$ , respectively. Thus, the number of elements in  $\tilde{X}_k$  and  $\tilde{Y}_k$  is lower than the number of EEG samples in the matrix  $A_k$ .

Therefore, the matrices  $\tilde{X}_k$  and  $\tilde{Y}_k$  can be considered as an alternative but compressed representation of the matrix  $A_k$ .

In particular, the expected compression ratio can be derived as follows.

Let us indicate with  $w$  the number of bits used to represent each EEG sample in  $A_k$ . Considering the actual dimensions of the matrices  $A_k$ , the overall number of bits needed for representing the matrix  $A_k$  is  $B_o = w \cdot N \cdot M$ .

In the same way, if we suppose that  $w + a$  is the maximum number of bits needed to represent the elements of  $\tilde{X}_k$  and  $\tilde{Y}_k$ , the overall number of bits needed for representing the compressed matrices  $\tilde{X}_k$  and  $\tilde{Y}_k$  is at most  $B_c = (w + a) \cdot (N + M) \cdot k$  and therefore the compression ratio can be evaluated as

$$CR = \frac{B_o}{B_c} = \frac{w \cdot M \cdot N}{(w + a) \cdot (N + M) \cdot k}. \quad (2)$$

In particular, when  $w + a \approx w$  and  $M \gg N$ , the expected compression ratio of the proposed algorithm can be approximated as  $CR \approx N/k$ . Therefore, a considerable compression can be achieved when  $N \gg k$ , i.e., in the case of high-density EEG systems with correlated signals.

For instance, in the case  $N = 256$  and  $k = 15$  we have  $CR \approx 17$  so that each GB of EEG data can be compressed and thus stored in less than 60 MB.

In their paper, authors proved that, given the matrices  $\tilde{X}_k$  and  $\tilde{Y}_k$  and the scale factor  $s = m_X \cdot m_Y$ , an effective approximation  $\tilde{A}_k$  of the matrix  $A_k$  is given by the following equation

- INPUTS: an integer number  $k < N$ , the matrix  $A$  formed by  $N \times M$  EEG samples **and a scale factor  $F$** ;
- OUTPUTS: integer matrices  $\tilde{X}_k$  and  $\tilde{D}_{Y_k}$  and the scale factor  $s = m_X \cdot m_Y$ .
- ALGORITHM:
  - 1) Use SVD to decompose  $A$  as  $A = U\Sigma V^T$
  - 2) Obtain the matrices  $S = \Sigma^{1/2}$ ,  
 $X_k = (US)[:, 1:k]$  and  $Y_k = (SV^T)[1:k, :]$
  - 3)  $m_X = \max(|X_k|)/F$ ,  $m_Y = \max(|Y_k|)/F$
  - 4) Calculate  $s = m_X \cdot m_Y$ ,  
 $\tilde{X}_k = \text{round}(m_Y \cdot X_k)$  and  $\tilde{Y}_k = \text{round}(m_X \cdot Y_k)$
  - 5) **Calculate the matrix**  
 $\tilde{D}_{Y_k} = [(\tilde{Y}_k[1, :])^T; \text{diff}(\tilde{Y}_k^T)]$

Figure 2. Illustration of the proposed compression algorithm.

$$\tilde{A}_k = \text{round}\left(\frac{\tilde{X}_k \tilde{Y}_k}{s}\right). \quad (3)$$

Basically, the above relation provides the reconstruction equation needed for decompression.

Experimental results reported in [32] have shown that the maximum absolute error  $MAE = |A_k - \tilde{A}_k|$  introduced by the above approximation is bounded by  $MAE \leq 2$ , that is a negligible error in comparison to the actual range of the original EEG samples, i.e.  $[-2^{w-1}, +2^{w-1} - 1]$ .

### 3. PROPOSED ALGORITHM

In this section, we slightly modify the previous algorithm with the aim of:

- 1) improving the compression ratio;
- 2) parameterizing the algorithm.

In particular, we derived a new version of the algorithm able to achieve different trade-offs between compression efficiency and distortion.

Basically, the new algorithm exploits the fact that consecutive values in the matrix  $\tilde{Y}_k$  are highly correlated. Therefore, a further reduction in the number of bits, and thus an increasing in the compression ratio, can be obtained by encoding the differences between consecutive values in  $\tilde{Y}_k$  instead of the matrix  $\tilde{Y}_k$  itself.

More precisely, let us introduce the matrix

$$\tilde{D}_{Y_k} = [(\tilde{Y}_k[1, :])^T; \text{diff}(\tilde{Y}_k^T)], \quad (4)$$

where  $\text{diff}()$  returns the matrix of differences along the first dimension. It is worth observing that the matrix  $\tilde{Y}_k$  can be exactly recovered from  $\tilde{D}_{Y_k}$  as

$$\tilde{Y}_k = \text{cumsum}(\tilde{D}_{Y_k})^T, \quad (5)$$

where  $\text{cumsum}()$  is the cumulative sum of elements along the first dimension.

Therefore, no further losses are introduced if, instead of the matrix  $\tilde{Y}_k$ , the matrix of differences  $\tilde{D}_{Y_k}$  is stored or transmitted.

On the basis of the previous observation, a new compression algorithm for EEG has been derived and can be summarized as shown in Figure 2. Note that, in comparison to the previous algorithm, we introduced a new step (see step 5), highlighted in bold for the sake of readability.

Reconstruction, i.e., decompression, can be easily achieved by obtaining  $\tilde{Y}_k$  with (5) and thus using again (3) to recover  $\tilde{A}_k$ .

Note that in the new algorithm we introduced a new input parameter, i.e., the scale factor  $F$ , which can be used to achieve different tradeoffs between compression efficiency and distortion. In particular, the factor  $F$  is exploited for reducing  $m_X$  and  $m_Y$  (see step 3 in Figure 2).

It is worth nothing that  $m_X$  and  $m_Y$  in the new and previous algorithms assume the same values when  $F = 1$ .

Therefore, intuitively and as confirmed in experimental results reported in the next section, we have no difference in the distortion achieved by the two algorithms when  $F = 1$ .

Instead, by choosing a value of  $F$  greater than 1, it is possible to achieve a greater compression ratio.

This can be easily justified by observing that, according to (1), by reducing  $m_X$  and  $m_Y$  we further reduce the dynamic range of the elements in the matrices  $\tilde{X}_k$  and  $\tilde{Y}_k$  and thus the number of bits that are needed for their representation.

Obviously, a greater compression ratio is obtained at the cost of a greater distortion.

Nevertheless, experimental results reported in the next section show that the proposed algorithm improves the compression ratio by about 10 % by achieving the same distortion level of our previous algorithm, i.e., 0.01 %. Moreover, a substantial increase in the compression ratio, up to 50 %, can be achieved by still maintaining the distortion level below the 0.1 %.

#### 4. MEASUREMENT-BASED RESULTS

The proposed compression algorithm is applied to a dataset containing real EEG signals, which have been preprocessed by EEG experts to denoise and remove artifacts. The EEG dataset under study has been provided by CUNY School of Medicine (New York, NY, USA). This dataset refers to awake EEG recordings from the research work published in [25]. In this study, Tatti et al. have investigated the role of beta oscillations (13.5-25 Hz) in the sensorimotor system in a group of healthy individuals. In this experiment, participants were asked to perform planar reaching movements (*mov* test). *Mov* test required the participants to reach a target, located at different distances and directions, that appeared on a screen in non-repeating and unpredictable order at 3 seconds interval. Participants were asked to make reaching movements by moving a cursor on a digitizing tablet with their right hand to targets appearing on the screen. The total testing time was approximately five to six minutes for each EEG recording (96 targets).

Each *mov* test was measured with a 256-channel high-density EEG system (HydroCel Geodesic Sensor Net, HCGSN, produced by Electrical Geodesic Inc., Eugene, OR, USA), amplified using a Net Amp 300 amplifier, and sampled at 250 Hz with 16-bit resolution using the Net Station software (version 5.0). EEG was noninvasively recorded using scalp electrodes and electrode-skin impedances were kept lower than 50 k $\Omega$ . The EEGLAB toolbox (v13.6.5b) for MATLAB (v.2016b) was used for off-line preprocessing of the gathered EEG data [33], [34]. The signal of each recording was first filtered using a finite impulse response (FIR) bandpass filter with a passband that extends from 1 Hz to 80 Hz and notch filtered at 60 Hz. Then, each recording was segmented in 4-seconds epochs and visually examined to remove sporadic artifacts and channels with poor signal quality. Moreover, ICA with PCA-based dimension reduction (max 108 components) was employed to identify

stereotypical artifacts (e.g., ocular, muscle, and electrocardiographic artifacts). Only ICA components with specific activity patterns and component maps characteristic of artefactual activity were removed. Electrodes with poor signal quality were reconstructed with spherical spline interpolation procedures, whereas those located on the cheeks and neck were excluded, resulting in 180 signals.

After the preprocessing, all signals were re-referenced to their initial average values and processed EEG data were exported in the European Data Format (EDF) [35] by means of the EEGLAB toolbox.

In particular, with the aim of evaluating the performance of the proposed algorithm, six EDF files, which are related to three subjects (labelled with the Subject Numbers SN\_M2, SN\_M4 and SN\_M5) and two sets of *mov* tests for each subject (“alltrials\_1” and “alltrials\_4”), have been tested.

Data range, number of samples, and a few other information of the above mentioned EDF files are reported in Table 1. It is worth observing that samples are represented with 16-bit integer numbers, i.e.,  $w = 16$ , and that the overall number of samples exploited for tests is more than  $80 \cdot 10^6$ .

In order to apply the proposed algorithm, each EDF file has been read and related data have been processed in blocks of  $N \times M$  samples, where  $N$  has been chosen equal to 180, i.e.,  $N$  coincides with the number of EEG channels remained after the preprocessing phase, and  $M$  has been fixed equal to 1,000. So that each block of samples represents 4 seconds of data recorded by the multichannel EEG systems.

The proposed compression algorithm, i.e., the algorithm in Figure 2, has been applied to each block and the average compression ratio have been evaluated according to the relation:

$$CR_F = \frac{1}{L} \sum_{i=1}^L \frac{B_{o,i}}{B_{c,i}}, \quad (6)$$

where  $L$  represents the number of blocks processed,  $B_{o,i}$  is the number of bits that are needed to represent the  $i$ -th block before compression, and  $B_{c,i}$  is the number of bits that are needed to represent the same block but after compression. Note that we use the subscript  $F$  to highlight the scale factor used for compression, e.g.,  $CR_2$  is the compression ratio achieved when  $F = 2$ . When the scale factor is not expressly stated we assume  $F = 1$ .

Subsequently, each compressed block has been reconstructed according to (5) and (3). Finally, distortion metrics, i.e.,  $PRD$  and  $MAE$ , have been evaluated on the whole EDF file using the following equations:

$$PRD = 100 \cdot \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^{M-L} (a_{ij} - \tilde{a}_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^{M-L} a_{ij}^2}} \quad (7)$$

$$MAE = \max_{i,j} |a_{ij} - \tilde{a}_{ij}|, \quad (8)$$

where  $\tilde{a}_{ij}$  are the integer values obtained after reconstruction and  $a_{ij}$  are the original samples.

In our experiments, we further evaluated the compression efficiency of the proposed algorithm with respect to the near-lossless compression algorithm proposed in [32]. In particular, compression efficiency ( $CE$ ) is here defined as:

Table 1. EDF files used as dataset.

| File name         | Channels | Duration (s) | Number of samples | Physical Range ( $\mu\text{V}$ ) | Data range      |
|-------------------|----------|--------------|-------------------|----------------------------------|-----------------|
| SN_M2_alltrials_1 | 180      | 344          | 15 480 000        | [-35.500, +30.314]               | [-32768, 32767] |
| SN_M2_alltrials_4 | 180      | 340          | 15 300 000        | [-48.550, +38.539]               | [-32768, 32767] |
| SN_M4_alltrials_1 | 180      | 328          | 14 760 000        | [-34.532, +34.756]               | [-32768, 32767] |
| SN_M4_alltrials_4 | 180      | 264          | 11 880 000        | [-41.100, +40.673]               | [-32768, 32767] |
| SN_M5_alltrials_1 | 180      | 324          | 14 580 000        | [-41.463, +38.867]               | [-32768, 32767] |
| SN_M5_alltrials_4 | 180      | 308          | 13 860 000        | [-41.347, +46.929]               | [-32768, 32767] |

$$CE = 100 \cdot \frac{CR_F - CR_0}{CR_0}, \quad (9)$$

where  $CR_0$  is the compression ratio obtained with the algorithm proposed in [32], i.e., the algorithm reported in Figure 1. Similarly, we use  $PRD_0$  and  $MAE_0$  for referring related distortion metrics.

Compression results achieved with the proposed compression algorithm by setting the scale factor  $F = 1$  are shown in Table 2.

More precisely, for each compressed file, we reported the number of singular values exploited for compression ( $k$ ), the compression ratio ( $CR_1$ ), values of the distortion metrics ( $PRD$  and  $MAE$ ), and the compression efficiency ( $CE$ ) of the proposed algorithm and, in brackets, corresponding compression results obtained with the algorithm proposed in [32], evaluated on the same files and considering the same number of singular values.

As it is possible to observe, the compression ratio achieved by the proposed algorithm when  $F = 1$  is near  $N/k$ , which confirms our analytical results reported in Section III.

Note that the  $PRD$  is less than 0.01 % for all EDF files tested. In particular,  $PRD$  values obtained with  $F = 1$  are the same values obtained in [32]. The same consideration can be extended to the  $MAE$ . This confirms that the two algorithms in Figure 2 and Figure 1 have the same performance in terms of distortion when  $F = 1$ .

However, by observing the results on compression efficiency (see the last column of Table 2), it is possible to conclude that, in comparison to the algorithm proposed in [32], the new one proposed in this paper is able to improve the compression ratio in a range between 7 % and 9 %. Moreover, the scale factor  $F$  introduced in the new algorithm provides the possibility to

Table 2. Compression ratio ( $CR$ ), percent root-mean-square distortion ( $PRD$ ), maximum absolute error ( $MAE$ ), and compression efficiency ( $CE$ ) of the proposed algorithm when  $F = 1$  (values of  $CR_0$  and  $PRD_0$  are reported in brackets).

| File name         | $k$ | $CR_1$         | $PRD$ (%)          | $MAE$ | $CE$ (%) |
|-------------------|-----|----------------|--------------------|-------|----------|
| SN_M2_alltrials_1 | 20  | 8.6<br>(7.9)   | 0.0065<br>(0.0065) | 2 (2) | 8.6      |
| SN_M2_alltrials_4 | 20  | 8.6<br>(7.9)   | 0.0063<br>(0.0063) | 2 (2) | 8.9      |
| SN_M4_alltrials_1 | 12  | 14.6<br>(13.6) | 0.0075<br>(0.0075) | 2 (2) | 7.4      |
| SN_M4_alltrials_4 | 12  | 14.4<br>(13.4) | 0.0069<br>(0.0069) | 2 (2) | 7.5      |
| SN_M5_alltrials_1 | 13  | 13.4<br>(12.5) | 0.0071<br>(0.0071) | 2 (2) | 7.2      |
| SN_M5_alltrials_4 | 13  | 13.2<br>(12.3) | 0.0069<br>(0.0069) | 2 (2) | 7.3      |

achieve even higher compression ratios, obviously at the cost of a greater distortion.

We investigated the trade-off between compression efficiency and distortion of the proposed algorithm by considering different values of the scale factor  $F$  within the range [1, 16].

In particular, we reported in Table 3 compression ratios ( $CR_F$ ), distortion metrics ( $PRD$  and  $MAE$ ), and compression efficiency ( $CE$ ) corresponding to  $F \in \{1, 2, 4, 8, 16\}$ .

As can be observed in Table 3, by increasing  $F$  it is possible to improve the compression ratio and thus the compression efficiency. In particular, by fixing  $F = 16$ , the proposed algorithm is able to improve the compression ratio by about the 50 % by maintaining the  $PRD$  below the 0.1 % threshold (in fact the  $PRD$  is at most equal to 0.081 % for all the files tested).

It is also worth noting that the  $MAE$  obtained in our experimental results is approximatively equal to  $2F$ .

Finally, we evaluated the distribution of absolute errors in recovered signals.

In Figure 3 we reported the cumulative distribution function (CDF) of the absolute errors, i.e., the probability  $P(|err| \leq x)$

Table 3. Compression results achieved with the proposed algorithm for different values of the scale factor  $F$ .

| File name         | $F$ | $CR_F$ | $PRD$ (%) | $MAE$ | $CE$ (%) |
|-------------------|-----|--------|-----------|-------|----------|
| SN_M2_alltrials_1 | 1   | 8.58   | 0.006     | 2     | 8.6      |
|                   | 2   | 9.23   | 0.010     | 4     | 16.8     |
|                   | 4   | 9.98   | 0.018     | 8     | 26.3     |
|                   | 8   | 10.87  | 0.035     | 16    | 37.6     |
|                   | 16  | 11.94  | 0.069     | 32    | 51.1     |
| SN_M2_alltrials_4 | 1   | 8.60   | 0.006     | 2     | 8.9      |
|                   | 2   | 9.25   | 0.010     | 4     | 17.1     |
|                   | 4   | 10.01  | 0.017     | 8     | 26.7     |
|                   | 8   | 10.91  | 0.033     | 17    | 38.1     |
|                   | 16  | 11.98  | 0.066     | 31    | 51.6     |
| SN_M4_alltrials_1 | 1   | 14.55  | 0.007     | 2     | 7.4      |
|                   | 2   | 15.68  | 0.012     | 5     | 15.7     |
|                   | 4   | 16.99  | 0.021     | 7     | 25.4     |
|                   | 8   | 18.53  | 0.041     | 15    | 36.8     |
|                   | 16  | 20.39  | 0.081     | 34    | 50.5     |
| SN_M4_alltrials_4 | 1   | 14.43  | 0.007     | 2     | 7.5      |
|                   | 2   | 15.53  | 0.011     | 4     | 15.7     |
|                   | 4   | 16.81  | 0.019     | 8     | 25.3     |
|                   | 8   | 18.33  | 0.036     | 16    | 36.6     |
|                   | 16  | 20.14  | 0.072     | 31    | 50.1     |
| SN_M5_alltrials_1 | 1   | 13.42  | 0.007     | 2     | 7.2      |
|                   | 2   | 14.45  | 0.011     | 5     | 15.4     |
|                   | 4   | 15.66  | 0.020     | 8     | 25.1     |
|                   | 8   | 17.08  | 0.039     | 20    | 36.4     |
|                   | 16  | 18.79  | 0.077     | 34    | 50.1     |
| SN_M5_alltrials_4 | 1   | 13.16  | 0.007     | 2     | 7.3      |
|                   | 2   | 14.16  | 0.010     | 4     | 15.5     |
|                   | 4   | 15.31  | 0.019     | 9     | 24.9     |
|                   | 8   | 16.67  | 0.037     | 16    | 36.0     |
|                   | 16  | 18.30  | 0.073     | 35    | 49.3     |

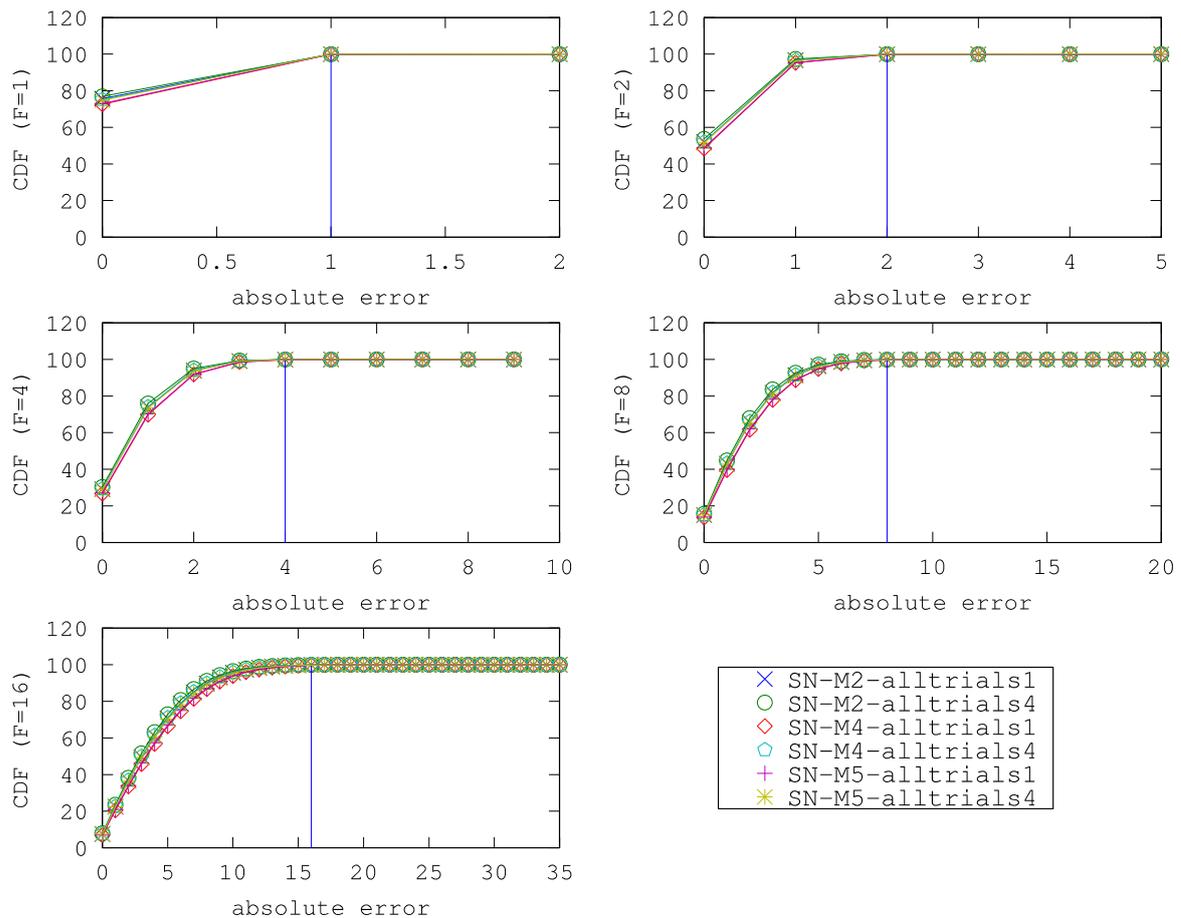


Figure 3. Cumulative distribution function (CDF) of the absolute errors obtained after reconstruction with the proposed near-lossless algorithm for different scale factors ( $F$ ). Vertical lines represent the condition  $P(|err| \leq F)$ .

that the absolute error ( $|err|$ ) is lower than a threshold  $x$ , achieved for different values of  $F$ .

As can be observed in Figure 3, the percentage of samples with an absolute error lower than  $F$  after reconstruction is near to 100 % for all the EDF files in the dataset. Note that the vertical lines in Figure 3 represent the condition  $P(|err| \leq F)$ .

Therefore, we can state that the scale factor  $F$ , which is needed as input in the proposed algorithm, can be fixed according to the desired MAE, i.e., for a given value of  $F$ , the MAE obtained after reconstruction will be, with high probability, within the range  $[F, 2F]$ .

## 5. CONCLUSIONS

In this paper, we developed and validated an improved version of a recently proposed near-lossless compression algorithm for multichannel EEG signals. The algorithm exploits the fact that SVD is usually performed on EEG signals for artifacts removal or denoising tasks. Experimental results, reported in this paper, show that the developed algorithm is able to achieve a compression ratio proportional to the number of EEG channels with a root-mean-square distortion less than 0.01 %. Moreover, with proper settings of input parameters, the compression ratio can be further improved up to 50 % by maintaining the distortion level below the 0.1 %. Moreover, the algorithm allows the desired maximum absolute error to be fixed a priori. It should be highlighted that, although an EEG dataset

has been considered as a case study, the proposed compression algorithm can be quite straightforwardly applied to different types of dataset. In a future work, we will further investigate performance of the proposed algorithm considering more extended datasets and other types of signals.

## REFERENCES

- [1] H. Berger, Uber Das Elektrenkephalogramm Des Menschen. Arch Psychiat Nervenkr (1929), pp. 527–570. DOI: [10.1007/BF01797193](https://doi.org/10.1007/BF01797193)
- [2] J. T. Koyazo, M. A. Ugwiri, A. Lay-Ekuakille, M. Fazio, M. Villari, C. Liguori, Collaborative systems for telemedicine diagnosis accuracy, Acta IMEKO 10 (2021) 3, pp. 192-197. DOI: [10.21014/acta\\_imeko.v10i3.1133](https://doi.org/10.21014/acta_imeko.v10i3.1133)
- [3] D. A. Pizzagalli, Electroencephalography and High-Density Electrophysiological Source Localization (Handbook of Psychophysiology, 3th Ed.), Cambridge University Press, 2007. DOI: [10.1017/CBO9780511546396](https://doi.org/10.1017/CBO9780511546396)
- [4] D. L. Schomer, F. L. Da Silva, Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Lippincott Williams & Wilkins, 2012.
- [5] N. Yoshimura, O. Koga, Y. Katsui, Y. Ogata, H. Kambara, Y. Koike, Decoding of emotional responses to user-unfriendly computer interfaces via electroencephalography signals, Acta IMEKO 6, (2017). DOI: [10.21014/acta\\_imeko.v6i2.383](https://doi.org/10.21014/acta_imeko.v6i2.383)
- [6] R. Abiri, S. Borhani, E. Sellers, Y. Jiang, X. Zhao, A Comprehensive Review of EEG-Based Brain-Computer Interface Paradigms. J. Neural Eng. 16 (2018), pp. 011001–011001.

- DOI: [10.1088/1741-2552/aaf12c](https://doi.org/10.1088/1741-2552/aaf12c)
- [7] P. L. Nunez, R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, USA, 2006, ISBN: 9780195050387.
  - [8] C. Lustenberger, R. Huber, High Density Electroencephalography in Sleep Research: Potential, Problems, Future Perspective. *Front. Neurol.* 3 (2012), pp. 77.  
DOI: [10.3389/fneur.2012.00077](https://doi.org/10.3389/fneur.2012.00077)
  - [9] B. H. Brinkmann, M. R. Bower, K. A. Stengel, G. A. Worrell, M. Stead, Large-scale electrophysiology: acquisition, compression, encryption, and storage of big data, *J. Neurosc. Meth.* 180, (2009), pp. 185–192.  
DOI: [10.1016/j.jneumeth.2009.03.022](https://doi.org/10.1016/j.jneumeth.2009.03.022)
  - [10] G. Gugliandolo, G. Campobello, P. P. Capra, S. Marino, A. Bramanti, G. D. Lorenzo, N. Donato, A Movement-Tremors Recorder for Patients of Neurodegenerative Diseases, *IEEE Trans Instrum Meas* 68 (2019), pp. 1451–1457.  
DOI: [10.1109/TIM.2019.2900141](https://doi.org/10.1109/TIM.2019.2900141)
  - [11] G. Campobello, A. Segreto, S. Zanafi, S. Serrano, An Efficient Lossless Compression Algorithm for Electrocardiogram Signals. In *Proceedings of the 26th European Signal Processing Conference, EUSIPCO 2018, Roma, Italy, September 3-7, 2018*, pp. 777–781.  
DOI: [10.23919/EUSIPCO.2018.8553597](https://doi.org/10.23919/EUSIPCO.2018.8553597)
  - [12] A. Casson, D. Yates, S. Smith, J. Duncan, E. Rodriguez-Villegas, Wearable electroencephalography, *IEEE EMBS Mag.* 29 (2010), pp. 44–56.  
DOI: [10.1109/memb.2010.936545](https://doi.org/10.1109/memb.2010.936545)
  - [13] G. Campobello, O. Giordano, A. Segreto, S. Serrano, Comparison of Local Lossless Compression Algorithms for Wireless Sensor Networks, *J Netw. Comput. Appl* 47 (2015), pp. 23–31.  
DOI: [10.1016/j.jnca.2014.09.013](https://doi.org/10.1016/j.jnca.2014.09.013)
  - [14] A. Nait-Ali, C. Cavaro-Menard, *Compression of Biomedical Images and Signals*, John Wiley & Sons, 2008, ISBN: 978-1-848-21028-8.
  - [15] N. Sriraam, Correlation Dimension Based Lossless Compression of EEG Signals, *Biomed. Signal Process. Control* 7 (2012), pp. 379–388.  
DOI: [10.1016/j.bspc.2011.06.007](https://doi.org/10.1016/j.bspc.2011.06.007)
  - [16] N. Sriraam, C. Eswaran, Lossless Compression Algorithms for EEG Signals: A Quantitative Evaluation, in *Proceedings of the IEEE/EMBS 5th International Workshop on Biosignal Interpretation, Tokyo Japan, September 6-8, 2005*, pp. 125–130.
  - [17] Y. Wongsawat, S. Orintara, T. Tanaka, K. R. Rao, Lossless Multi-Channel EEG Compression, in *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems, Island of Kos, Greece, May 21-24, 2006*, p. 4 pp. – 1614.  
DOI: [10.1109/ISCAS.2006.1692909](https://doi.org/10.1109/ISCAS.2006.1692909)
  - [18] G. Antonioli, P. Tonella, EEG Data Compression Techniques, *IEEE Trans Biomed. Eng.* 44 (1997), pp. 105–114.  
DOI: [10.1109/10.552239](https://doi.org/10.1109/10.552239)
  - [19] N. Sriraam, C. Eswaran, Performance Evaluation of Neural Network and Linear Predictors for Near-Lossless Compression of EEG Signals, *IEEE Trans. Inf. Technol. Biomed.* 12 (2008), pp. 87–93.  
DOI: [10.1109/TTTB.2007.899497](https://doi.org/10.1109/TTTB.2007.899497)
  - [20] G. Campobello, A. Segreto, S. Zanafi, S. Serrano, RAKE: A Simple and Efficient Lossless Compression Algorithm for the Internet of Things. In *Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos island, Greece, 28 August - 2 September, 2017*, pp. 2581–2585.  
DOI: [10.23919/EUSIPCO.2017.8081677](https://doi.org/10.23919/EUSIPCO.2017.8081677)
  - [21] K. Srinivasan, J. Dauwels, M. R. Reddy, A Two-Dimensional Approach for Lossless EEG Compression, *Biomed. Signal Process. Control* 6 (2011), pp. 387–394.  
DOI: [10.1016/j.bspc.2011.01.004](https://doi.org/10.1016/j.bspc.2011.01.004)
  - [22] N. Ille, P. Berg, M. Scherg, Artifact Correction of the Ongoing EEG Using Spatial Filters Based on Artifact and Brain Signal Topographies, *J. Clin. Neurophysiol* 19 (2002), pp. 113–124.  
DOI: [10.1097/00004691-200203000-00002](https://doi.org/10.1097/00004691-200203000-00002)
  - [23] R. J. Davidson, D. C. Jackson, C. L. Larson, *Human Electroencephalography (Handbook of Psychophysiology, 2nd Ed.)*, Cambridge University Press, 2000.
  - [24] N. Mammone, D. Labate, A. Lay-Ekuakille, F. C. Morabito, Analysis of Absence Seizure Generation Using EEG Spatial-Temporal Regularity Measures, *Int J Neural Syst* 22 (2012).  
DOI: [10.1142/s0129065712500244](https://doi.org/10.1142/s0129065712500244)
  - [25] E. Tatti, S. Ricci, A. B. Nelson, D. Mathew, H. Chen, A. Quartarone, C. Cirelli, G. Tononi, M. F. Ghilardi, Prior Practice Affects Movement-Related Beta Modulation and Quiet Wake Restores It to Baseline, *Front. Syst. Neurosci* 14 (2020), pp. 61.  
DOI: [10.3389/fnsys.2020.00061](https://doi.org/10.3389/fnsys.2020.00061)
  - [26] M. K. Islam, A. Rastegarnia, Z. Yang, Methods for Artifact Detection and Removal from Scalp EEG: A Review. *Neurophysiol. Clin.* 46 (2016), pp. 287–305.  
DOI: [10.1016/j.neucli.2016.07.002](https://doi.org/10.1016/j.neucli.2016.07.002)
  - [27] S. Casarotto, A. M. Bianchi, S. Cerutti, G. A. Chiarenza, Principal Component Analysis for Reduction of Ocular Artefacts in Event-Related Potentials of Normal and Dyslexic Children, *Clin. Neurophysiol.* 115 (2004), pp. 609–619.  
DOI: [10.1016/j.clinph.2003.10.018](https://doi.org/10.1016/j.clinph.2003.10.018)
  - [28] Z. Anusha, J. Jinu, T. Geevarghese, Automatic EEG Artifact Removal by Independent Component Analysis Using Critical EEG Rhythms, in *Proceedings of the 2013 IEEE International Conference Control Communication and Computing (ICCC), Trivandrum, Kerala, India, December 13-15, 2013*, pp. 364–367.  
DOI: [10.1109/ICCC.2013.6731680](https://doi.org/10.1109/ICCC.2013.6731680)
  - [29] J. Dauwels, K. Srinivasan, M. R. Reddy, A. Cichocki, Near-Lossless Multichannel EEG Compression Based on Matrix and Tensor Decompositions, *IEEE J. Biomed. Health Inform.* 17 (2013), pp. 708–714.  
DOI: [10.1109/TTTB.2012.2230012](https://doi.org/10.1109/TTTB.2012.2230012)
  - [30] L. Lin, Y. Meng, J. Chen, Z. Li, Multichannel EEG Compression Based on ICA and SPIHT, *Biomed. Signal Process. Control* 20 (2015), pp. 45–51.  
DOI: [10.1016/j.bspc.2015.04.001](https://doi.org/10.1016/j.bspc.2015.04.001)
  - [31] M. K. Alam, A.A. Aziz, S. A. Latif, A. Awang, EEG Data Compression Using Truncated Singular Value Decomposition for Remote Driver Status Monitoring, in *Proceedings of the 2019 IEEE Student Conference on Research and Development (SCORED), Universiti Teknologi PETRONAS (UTP), Malaysia, 15-17 October 2019*, pp. 323–327.  
DOI: [10.1109/SCORED.2019.8896252](https://doi.org/10.1109/SCORED.2019.8896252)
  - [32] G. Campobello, A. Quercia, G. Gugliandolo, A. Segreto, E. Tatti, M. F. Ghilardi, G. Crupi, A. Quartarone, N. Donato, An Efficient Near-lossless Compression Algorithm for Multichannel EEG signals, 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Neuchâtel, Switzerland, June 23-25, 2021.  
DOI: [10.1109/MeMeA52024.2021.9478756](https://doi.org/10.1109/MeMeA52024.2021.9478756)
  - [33] A. Delorme, S. Makeig, EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis, *J. Neurosci. Methods* 134 (2004), pp. 9–21.  
DOI: [10.1016/j.jneumeth.2003.10.009](https://doi.org/10.1016/j.jneumeth.2003.10.009)
  - [34] S. Makeig, S. Debener, J. Onton, A. Delorme, Mining Event-Related Brain Dynamics, *Trends Cogn. Sci.* 8 (2004), pp. 204–210.  
DOI: [10.1016/j.tics.2004.03.008](https://doi.org/10.1016/j.tics.2004.03.008)
  - [35] B. Kemp, A. Värri, A. C. Rosa, K. D. Nielsen, J. Gade, A Simple Format for Exchange of Digitized Polygraphic Recordings, *Electroencephalogr. Clin. Neurophysiol* 82 (1992), pp. 391–393.  
DOI: [10.1016/0013-4694\(92\)90009-7](https://doi.org/10.1016/0013-4694(92)90009-7)