The Riemannian Potato Field: A Tool for Online Signal Quality Index of EEG

Quentin Barthélemy, Louis Mayaud, David Ojeda, Marco Congedo

The Riemannian Potato Field:
a tool for online Signal Quality Index of EEG

Quentin Barthélemy, Louis Mayaud, David Ojeda and Marco Congedo

Abstract—**Electroencephalographic (EEG) recordings are contaminated by instrumental, environmental, and biological artifacts, resulting in low signal-to-noise ratio. Artifact detection is a critical task for real-time applications where the signal is used to give a continuous feedback to the user. In these applications, it is therefore necessary to estimate online a signal quality index (SQI) in order to stop the feedback when the signal quality is unacceptable.**

In this article we introduce the Riemannian Potato Field (RPF) algorithm as such SQI. It is a generalization and extension of the Riemannian Potato, a previously published real-time artifact detection algorithm, whose performance is degraded as the number of channels increases. The RPF overcomes this limitation by combining the outputs of several smaller potatoes into a unique SQI resulting in a higher sensitivity and specificity, regardless of the number of electrodes. We demonstrate these results on a clinical dataset totalizing more than 2200 hours of EEG recorded at home, that is, in a non-controlled environment.

Index Terms—**EEG, signal quality index, online, Riemannian geometry, artifact detection.**

I. Introduction

Electroencephalography (EEG) is a non-invasive technology measuring the brain electrical potentials at several electrode sites. It has become a widely-used technology for the monitoring of brain activity in applications such as brain-computer interfaces (BCI) [1] and neurofeedback (NFB) [2]. Unfortunately, EEG recordings are contaminated by several types of artifacts that are typically not generated by the brain electrical functioning. These can be biological, instrumental or environmental artifacts and often have a strong amplitude that lowers the signal-to-noise ratio (SNR) dramatically, making most applications unpractical. Consequently, measuring signal quality and denoising artifacted epochs are crucial steps for any application relying on EEG data analysis [3], [4]. This is specially important for real-time applications, such as BCI or NFB, where the system behavior fully depends on the current data quality and the presence of artifacts can disturb the feedback stream [5]–[7]. In this context, the task is significantly more complex: it cannot be done retrospectively as in offline analysis, by a human operator or an automatic offline algorithm; denoising can only rely on automated pre-calibrated

Q. Barthélemy, L. Mayaud and D. Ojeda are with Mensia Technologies SA, 130, rue de Lourmel, 75015 Paris, France (e-mail: q.barthelemy@gmail.com, lm@mensiatech.com, do@mensiatech.com).

M. Congedo is with GIPSA-Lab, CNRS, Grenoble-Alpes University, Grenoble Institute of Technology, 38402 Saint Martin d'Hères, France (e-mail: marco.congedo@gipsa-lab.grenoble-inp.fr).

approaches that should be more generalizable as compared to an offline strategy. While some techniques have proven efficacious and reliable to correct some types of stationary artifacts including eye blinks [8], most types of noise affecting EEG remain transient and non-stationary (such as muscle activities and body movements) making *correction* strategies, specifically for real-time applications, only partly reliable. In this context, it is necessary to fall back on a *detection* strategy, informing of the data quality. Detecting artifacts in real-time allows to temporarily suspend the EEG processing, therefore preventing the extraction of information on noisy EEG. Such technique can also be used to send a real-time warning to the user, possibly reinforcing behaviors leading to good data quality.

Due to headset installation and impedance check, the beginning of the EEG recordings often contains artifacts. To not initialize a real-time detection method on noisy EEG, the capability of detecting artifacts must be transferred from another recording. One of the technical aspects of any algorithm applied to EEG is that these recordings typically have many sources of variabilities [9] from: *i)* the system (electrodes type, amplifiers and acquisition devices), *ii)* the subject (personal brain anatomy and circadian state), and *iii)* the session setup (instrumental setups, impedances change [10], shifts in electrodes position, montage errors and all problems related to a non-controlled environment [11], [12]). These variabilities hinder the generalization of a model learned on a recording to another, even if it is the same user. Known in general under the name of transfer learning [13], many methods have been proposed to cope with this problem [14], [15]. In EEG, transfer learning may be session-to-session and/or subject-to-subject.

In a nutshell, an efficient artifact detection approach needs to: *i)* measure the quality of EEG, using a signal quality index (SQI), *ii)* detect when this quality is not acceptable, *i.e.*, when EEG is too noisy, *iii)* be able to transfer a model of acceptable quality from another recording. Moreover, the SQI must be a continuous measure, *e.g.*, a real value in $[0, 1]$, that characterizes the quality of the EEG along a continuum ranging from 1 (clean) to 0 (noisy). A complete range of values gives a measure more informative than a binary threshold. For example, summarizing the quality of a recording by an averaged SQI, integrating the intensity of artifacts is more precise than the percentage of rejected data.

The *Riemannian Potato* (RP) [16] is a recent multivariate approach to detect artifacts using covariance matrices processed by Riemannian geometry [17]. It provides a measure of dispersion (standardized z-score) using the distribution of distances between covariance matrices and a reference of clean
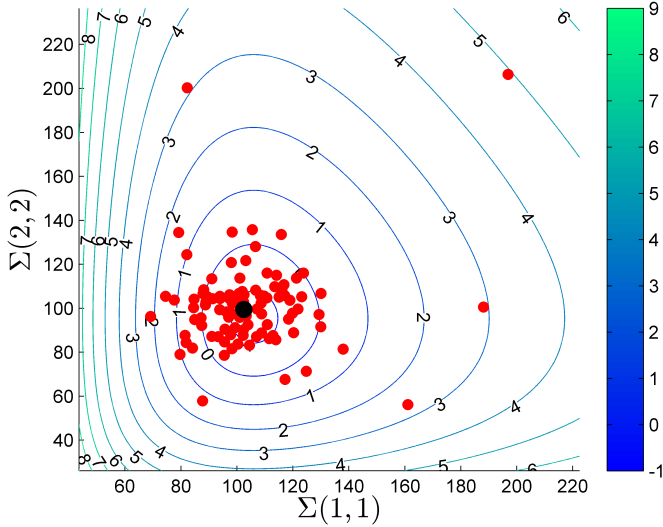
Fig. 1: 2D projection of the z-score map of a Riemannian potato, for 100 simulated 2×2 matrices Σ (in red) and their reference matrix (in black). The colormap defines the z-score and a chosen isocontour z_{th} defines the potato.

EEG. The RP rejects epochs whose covariance matrices lie out of a region of acceptability defined by z-score threshold, a region referred to as the *potato* because its crooked aspect is induced by the non-linearity of the Riemannian manifold. Even if it is not bounded, the z-score could be considered as a SQI, as illustrated in Fig. 1. Due to its simplicity and efficiency, the RP has been intensively used for *i)* online artifact rejection for P300-based BCI spellers [18]–[20] and games [21], *ii)* offline rejection before the statistical analysis of cognitive assessments [22], [23], *iii)* offline rejection before the classification of psychiatric disorders [24], [25] or motor imagery BCI [26], *iv)* offline rejection in the source space after applying a blind source separation (BSS) [27], *v)* rejection for a reliable estimation of class centers of a SSVEP-based BCI, online updated [28], or offline estimated and then used for an offline [29], [30] or for an online [31] classification. However, in all these cases, the SQI is used as a binary output resulting from the z-score thresholding and its continuous range of values is not fully exploited.

Belonging to anomaly detection [32], other approaches exist for detecting artifacts in EEG [33]–[36]. Most of them process block-online or offline instead of online (see [8], [37] for details on these distinctions), or apply BSS and give one index by source to remove artifacted ones [38]–[41]. A supplementary advantage of the online RP is that, since it is based on covariance matrices, it uses the same features used for classification in the new generation of BCI paradigms [17]: motor-imagery [42], event-related potentials [43], and SSVEP [31]. Specifically, the use of some specific Riemannian metrics improves transferability of learning, removing variabilities modeled by spatial filtering due to their congruence invariance [17] and reducing the risk of overfitting since spatial filters are very specific to subjects and sessions [44].

The RP algorithm does have some limitations. First, its sensitivity and specificity tend to drop as the number of channels increases. Second, its online initialization for real-time

applications requires an acquisition beginning of good quality. This article introduces the *Riemannian Potato Field* (RPF), a generalization of the Riemannian potato. The RPF overcomes these problems by using several small potatoes in parallel, each one designed to address a specific artifact, considerably increasing its specificity. In this setting, each potato focuses on a specific subset of channels and on a frequency band, which are known to be affected by this specific artifact. The output of each potato is then combined in a single SQI.

To address the initialization issues, a semi-dynamic mode is introduced, where the algorithm is initialized with a clean recording and updates itself with the current recording whenever its quality is considered sufficient. We show that even for a small number of electrodes, RPF is more sensitive and specific than RP.

After presenting the useful elements of Riemannian geometry in Section II, the Riemannian potato is detailed in Section III and several improvements are listed in Section IV. Section V introduces the Riemannian potato field, addressing the limitations of the classic potato. EEG data used as validation is presented in Section VI. Section VIII compares the different covariance normalizations for cross-session transfer learning, whereas Section VII compares RP and RPF for artifact detection. The last sections discuss the limitations of the method and conclude the article.

II. PROCESSING OF COVARIANCE MATRICES

This section presents briefly the geometry of covariance matrices that captures the multivariate second-order statistics of the EEG. Readers interested in a deeper introduction may refer to the primer [17] and references therein. We expand this introduction by discussing the technical elements related to the *online* implementation of this method.

A. Geometry of Covariance Matrices

Let $X \in \mathbb{R}^{C \times N}$ denote an epoch of EEG signal, recorded on C channels/electrodes and on N temporal samples. If this signal is centered (*e.g.*, after a band-pass filtering), its covariance matrix $\Sigma \in \mathbb{R}^{C \times C}$ is estimated as:

$$\Sigma = \frac{1}{N-1} X X^T \ . \tag{1}$$

This estimation is computationally simple and fast. However, N must be much greater than C to have a reliable estimator of the true covariance (see [31] for comparison of estimators).

Covariance matrices are symmetric positive definite (SPD), *i.e.*, they have strictly positive eigenvalues. For this reason, covariance matrices are confined in a subspace of the Euclidean space, that is, the cone defined by the Cauchy-Schwarz inequality, stating that the product of two variances is greater than or equal to the square of their covariance [17]. These matrices are therefore processed in a Riemannian manifold $\mathcal{M}_C$, of dimension $m = C(C+1)/2$, having a dedicated geometry [45], [46]. Using the eigenvalue decomposition of Σ:

$$\Sigma = U \operatorname{diag}(\lambda_1, \ldots, \lambda_C) U^T \ ,$$

where $\lambda_1, \ldots, \lambda_C$ are the eigenvalues and U the matrix of eigenvectors of Σ, the unique symmetric square root $\Sigma^{\frac{1}{2}}$ and symmetric inverse square root $\Sigma^{-\frac{1}{2}}$ are defined as:

$$\Sigma^{\frac{1}{2}} = U \operatorname{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_C^{\frac{1}{2}}) U^T ,$$
$$\Sigma^{-\frac{1}{2}} = U \operatorname{diag}(\lambda_1^{-\frac{1}{2}}, \ldots, \lambda_C^{-\frac{1}{2}}) U^T .$$

The affine-invariant Riemannian (AIR) distance between Σ_1 and Σ_2 is defined as [47]:

$$\delta_R(\Sigma_1, \Sigma_2) = \| \operatorname{Log}(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}) \|_F = \left(\sum_{c=1}^{C} \log^2 \lambda_c \right)^{\frac{1}{2}} ,$$

where λ_c, $c = 1, \ldots, C$, are the eigenvalues of $\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}$. The geometric mean of $i = 1, \ldots, I$ matrices Σ_i can be defined as the matrix that minimizes the dispersion of Σ_i on the manifold, *i.e.*, the sum of squared distances:

$$\bar{\Sigma} = \arg \min_{\Sigma \in \mathcal{M}_C} \sum_{i=1}^{I} \delta_R^2(\Sigma_i, \Sigma) .$$

This mean is more robust to outliers than the Euclidean mean [17], but is negatively affected by ill-conditioned input matrices [48]. It has no closed form and therefore has to be computed iteratively, for example through a gradient descent algorithm [49] or by the more efficient fixed-point algorithm given in [50].

The AIR distance has many properties [45, Chap 9], specially the invariance under congruent transformation:

$$\delta_R(\Sigma_1, \Sigma_2) = \delta_R(W^T \Sigma_1 W, W^T \Sigma_2 W) ,$$

for any invertible matrix $W \in \mathbb{R}^{C \times C}$. This is particularly interesting for EEG analysis, since it shows that applying a full-rank spatial filter does not change the Riemannian distance between covariances [17]. Spatial filters are subject and session-dependent [44], [51], [52], while AIR distance offers cross-session and cross-subject transfer capabilities [17].

B. Online Implementation

Let us consider an EEG recording $\mathcal{X} \in \mathbb{R}^{C \times \mathcal{N}}$, in an online setting where new data become available sample-by-sample or block-by-block [8]. To meet the computational requirements of an online implementation, the recording is cut in overlapping epochs, resulting in a sliding window. Epochs have size N, and the interval between two consecutive epochs is ΔN, with $N > \Delta N$. At time n, we can extract the epoch $X = \mathcal{X}(1 \ldots C, n - N + 1 \ldots n) \in \mathbb{R}^{C \times N}$, and then the covariance matrix Σ is computed with Eq. (1).

III. ORIGINAL RIEMANNIAN POTATO

In this section, we present the original Riemannian potato [16] for online artifact detection.

A. Principle of the RP

The basic idea of the RP is to represent clean EEG by estimating a reference covariance matrix $\bar{\Sigma}$ and a measure of dispersion (z-score), and then to reject all epochs whose covariance matrices are too far from this reference according to an objective statistical criterion. To do so, for epoch number t, the Riemannian distance between the current covariance matrix Σ_t and the reference $\bar{\Sigma}_{t-1}$ is computed as $d_t = \delta_R(\Sigma_t, \bar{\Sigma}_{t-1})$. Then, the z-score z_t of this distance is computed as:

$$z_t = \frac{d_t - \mu_{t-1}}{\sigma_{t-1}} , \tag{2}$$

where the mean μ and the standard deviation σ are computed using the Riemannian distances between previous clean matrices and the reference matrix:

$$\mu_t = \frac{1}{t} \sum_{\tau=1}^{t} d_\tau , \tag{3}$$

$$\sigma_t = \sqrt{\frac{1}{t} \sum_{\tau=1}^{t} (d_\tau - \mu_t)^2} . \tag{4}$$

This z-score models the distribution of covariance matrices considered as clean and provides a z-score threshold z_{th} (typically equal to 2.5, rejecting around 0.6% of data under Gaussian assumption) defining the hull of acceptability. The Riemannian potato then rejects epochs whose covariance matrices lie out of this region of acceptability. Thus, the output z-score of the Riemannian potato can be used as a SQI for EEG signals.

In a typical adaptive implementation of this method, the clean EEG model parameters ($\bar{\Sigma}_{t-1}$, μ_{t-1}, and σ_{t-1}) are updated whenever the z-score z_t is lower than the threshold z_{th}, that is, when the current EEG epoch is considered as clean / non-artifacted. This update goes as follows.

First, the reference matrix $\bar{\Sigma}_{t-1}$ is updated, using the current clean matrix Σ_t:

$$\bar{\Sigma}_t = \bar{\Sigma}_{t-1}^{\frac{1}{2}} \left(\bar{\Sigma}_{t-1}^{-\frac{1}{2}} \Sigma_t \bar{\Sigma}_{t-1}^{-\frac{1}{2}} \right)^{\alpha} \bar{\Sigma}_{t-1}^{\frac{1}{2}} , \tag{5}$$

where a coefficient $\alpha \in [0, 1]$ defines the speed of adaptation or learning rate. Different learning strategies are possible in online implementations: α can be constant (giving an exponential moving average), or be a function such as $1/t$ (giving a cumulative moving average). Note that this online averaging is equivalent to $\bar{\Sigma}_t = (1 - \alpha)\bar{\Sigma}_{t-1} + \alpha\Sigma_t$ in the Euclidean geometry [17].

Second, from Eq. (3) and (4), the mean and the standard deviation of Riemannian distances are updated online as:

$$\mu_t = (1 - \beta)\mu_{t-1} + \beta d_t , \tag{6}$$

$$\sigma_t = \sqrt{(1 - \beta)\sigma_{t-1}^2 + \beta(d_t - \mu_t)^2} , \tag{7}$$

where the second learning rate β is usually chosen equal to α.

B. Online Initialization

A hyper-parameter T_{init} defines the number of covariance matrices (or equivalently, the recording duration) used for

initialization of the region of acceptability, modeling clean EEG. During this period, the EEG is preferably clear of any artifacts so as to provide an accurate estimate for the mean and the distribution of distances to it, which will greatly influence the artifact detection performance [16]. Otherwise, the model of clean EEG will be badly initialized and the Riemannian potato will be inefficient to reject artifacts. The calibration method to initialize the region of acceptability is shown in Algorithm 1, used here with $T = T_{\text{init}}$.

Algorithm 1 : $\bar{\Sigma}, \mu, \sigma = $ RP_Calibration $(\{\Sigma_t\}_{t=1}^T)$

1: Estimation of the mean $\bar{\Sigma}$ from the matrices $\{\Sigma_t\}_{t=1}^T$
2: Estimation of the T distances to the reference $\bar{\Sigma}$
3: Estimation of the mean of distances μ with Eq. (3)
4: Estimation of the standard deviation σ with Eq. (4)

C. Summary

The Riemannian potato is summarized in Algorithm 2, and its implementation is illustrated on the left part of Fig. 2. By modeling clean EEG as a unimodal distribution of covariance matrices thanks to $\bar{\Sigma}$, μ and σ, this one-class classifier is a simple, white-box, parameter-free, adaptive, deterministic, and it is not specific to a particular kind of artifacts. Moreover, it relies on simple concepts that can be illustrated graphically, increasing further its wide and fast adoption by the community. To illustrate it, Fig. 1 shows how 100 simulated 2×2 covariance matrices can be viewed as spatial coordinates contained in a hyper-cone [17], from which a reference can be computed (in black) and isolines of z-scores (colormap) that can easily be translated into a likelihood of being a clean epoch. Due to its popularity, the RP algorithm has been implemented in different languages Python[1], Matlab[2], OpenViBE[3], and NeuroRT Studio[4].

Algorithm 2 : $z_t = $ RP_Online $(\Sigma_t, T_{\text{init}})$

1: **initialization** : $z_{\text{th}} \leftarrow 2.5$, $\mathcal{T} \leftarrow \emptyset$
2: **if** $t < T_{\text{init}}$
3:     $\mathcal{T} \leftarrow \mathcal{T} \cup \{t\}$
4: **if** $t = T_{\text{init}}$
5:     $\bar{\Sigma}_{T_{\text{init}}}, \mu_{T_{\text{init}}}, \sigma_{T_{\text{init}}} \leftarrow $ RP_Calibration $(\{\Sigma_t\}_{t \in \mathcal{T}})$
6: **if** $t > T_{\text{init}}$
7:     Riemannian distance $d_t \leftarrow \delta_R(\Sigma_t, \bar{\Sigma}_{t-1})$
8:     Distance z-score $z_t \leftarrow (d_t - \mu_{t-1})/\sigma_{t-1}$
9:     **if** $z_t \leq z_{\text{th}}$
10:         Update of matrix mean $\bar{\Sigma}_t$ with Eq. (5)
11:         Update of distances mean μ_t with Eq. (6)
12:         Update of standard deviation σ_t with Eq. (7)

---

[1] http://github.com/alexandrebarachant/pyRiemann
[2] http://github.com/alexandrebarachant/covariancetoolbox
[3] http://code.google.com/p/openvibe-gipsa-extensions
[4] http://www.mensia.com

IV. IMPROVED RIEMANNIAN POTATO

This section reviews several enhancements of the Riemannian potato that have been introduced in the literature, followed by some improvements proposed by this paper.

A. Robust Mean by Offline Outlier Removal

Even though the Riemannian mean is more robust than the Euclidean one [17], it is still sensitive to outliers. This limitation can easily be addressed by computing a robust mean, where the Riemannian potato is used as an outlier removal during an offline averaging [29], [30]. First, mean matrix and z-scores are computed. Then, outliers above the threshold are discarded, and finally, mean matrix is recomputed only on the remaining matrices. This procedure is described in Algorithm 3, returning the robust mean.

Offline outlier removal is useful to estimate reliable centers of classes for classification, but results are better when a potato is applied for each class [26], [29]–[31], rather than on all matrices [24], [25] that can lead to the removal of matrices close to the center of their class but far from the mean of all matrices [53]. While the performance improvement brought by outlier removal in the reference estimate has not necessarily lead to significant improvements in controlled environments [26], [31], it certainly contributed positively in more adverse settings [11], [12]. A related approach is the trimmed geometric mean [54], more robust than the geometric median, where the outlier removal criterion is changed from a z-score threshold to a percentage of the most distant matrices.

Algorithm 3 can be used iteratively [31]. At iteration k, input is defined as $\{\Sigma_t\}_{t \in \mathcal{T}_{k-1}}$, where $\mathcal{T}_{k-1}$ is the output indices set of the previous iteration $k - 1$. The number of iterations K can be fixed [31] (typically 3-5) or could be determined by a convergence/stopping criterion.

Algorithm 3 : $\{z_t\}_{t=1}^T, \bar{\Sigma}, \mathcal{T} = $ RP_Offline $(\{\Sigma_t\}_{t=1}^T)$

1: **initialization** : $z_{\text{th}} \leftarrow 2.5$, $\mathcal{T} \leftarrow \emptyset$
2: $\bar{\Sigma}, \mu, \sigma \leftarrow $ RP_Calibration $(\{\Sigma_t\}_{t=1}^T)$
3: **for** $t \leftarrow 1, \ldots, T$
4:     Distance z-score $z_t \leftarrow (d_t - \mu)/\sigma$
5:     **if** $z_t \leq z_{\text{th}}$
6:         $\mathcal{T} \leftarrow \mathcal{T} \cup \{t\}$
7: Estimation of the mean $\bar{\Sigma}$ from remaining matrices $\{\Sigma_t\}_{t \in \mathcal{T}}$

B. Geometric Statistics

All statistics defined in Eq. (2), (3) and (4) are *arithmetic*, and are not optimal because Riemannian distances are not normally distributed. Since distances empirically follow a (non-negative) highly right-skewed distribution, they can be

modeled by a log-normal or chi-squared distribution, and *geometric* statistics are better descriptors [17], [45]:

$$\mu_t = \exp\left(\frac{1}{t}\sum_{\tau=1}^{t}\log(d_\tau)\right) \ , \tag{8}$$

$$\sigma_t = \exp\left(\sqrt{\frac{1}{t}\sum_{\tau=1}^{t}(\log(d_\tau/\mu_t))^2}\right) \ , \tag{9}$$

$$z_t = \frac{\log(d_t/\mu_t)}{\log(\sigma_t)} \ . \tag{10}$$

The geometric z-scores of distances are more normally distributed than arithmetic ones[5]. Several offline RP have been used with these geometric statistics [20], [31], replacing Eq. (2), (3) and (4) respectively by Eq. (10), (8) and (9) in Algorithms 1 and 3.

Moreover, like Eq. (6) and (7), geometric statistics can be updated online as:

$$\mu_t = \exp\left((1-\beta)\log(\mu_{t-1}) + \beta\log(d_t)\right) \ , \tag{11}$$

$$\sigma_t = \exp\left(\sqrt{(1-\beta)(\log(\sigma_{t-1}))^2 + \beta(\log(d_t/\mu_t))^2}\right) . \tag{12}$$

Consequently, we shall replace Eq. (6) and (7) respectively by Eq. (11) and (12) in Algorithm 2.

C. Offline Initialization

An important limitation of the original RP was its online calibration (see Section III-B) relying on the fact that the first few seconds of the recording had to be clear of artifacts. This use case is particularly problematic because the beginning of EEG recordings are typically more prone to various sources of contaminations from electrode adjustment to changes of the skin electrochemical balance after the electrode setup. It is therefore not desirable for Algorithm 2 to be initialized that way as this would hinder its performance.

In order to be able to reject artifacts present from the beginning of the recording, reference matrix $\bar{\Sigma}_0$, mean μ_0 and standard deviation σ_0 can be computed on a previously recorded signal, using Algorithm 1, and then given as initial parameters to the online RP. This procedure is very similar to Algorithm 2, but steps 1 to 6 do not exist anymore since the potato is already initialized, and output z-score can be computed from $t = 1$. In this case, the value of T_{init}, used in update equations, defines the importance given to the initialization, with a high value for a confident calibration. Since this initialization is made offline, the outlier removal described in Section IV-A can be used to estimate a robust center of the potato.

In other words, the online Riemannian potato, described in Algorithm 2, can be used online in three modes:

- dynamic: online initialization and potato update (as originally published [16] and described in Algorithm 2),
- semi-dynamic: offline initialization and potato update (described in this paragraph),
- static: offline initialization and no potato update (steps 9 to 12 of Algorithm 2 are never executed, equivalent to $\alpha = \beta = 0$).

[5]This will be important in the following, notably in Eq. (15).

D. Covariance Normalization for Transfer Learning

In order to increase the efficiency of the artifact detection of the semi-dynamic or static RP used with a transferred initialization, covariance matrices can be normalized to reduce remaining sources of variability. As previously noted, transfer learning is a challenge in EEG and Riemannian geometry is inherently insensitive to some session and subject variabilities thanks to the congruence invariance property [17]. Other sources of variability, for instance related to electrode impedances, can also be addressed by pre-processing steps applied to the input covariance matrices such as normalization.

Covariance matrices can be either trace-normalized [55] as follow:

$$\Sigma \leftarrow \frac{1}{\text{trace}(\Sigma)}\Sigma \ = \frac{N-1}{\text{trace}(XX^T)}\Sigma = \frac{N-1}{\|X\|_F^2}\Sigma \ , \tag{13}$$

or alternatively determinant-normalized [20], [56] as:

$$\Sigma \leftarrow \frac{1}{\det(\Sigma)^{1/C}}\Sigma \ , \tag{14}$$

providing a unit hyper-volume of the feature space, with C the number of channels. Such normalizations can be applied for cross-session as well as for cross-subject transfer learning.

E. Generic RP on several recordings

Algorithm 1 describes how to calibrate a RP on a single recording, but it does not well estimate the region of acceptability when we want to calibrate a generic RP on several recordings, coming from different sessions and/or different subjects. Adding a superscript $s = 1,\ldots,S$ to index the recordings, the offline initialization of the generic RP (GRP) is described in Algorithm 4. The reference covariance matrix is the geometric mean of the reference covariance matrices $\bar{\Sigma}^s$ of the recordings, and the mean μ (*resp.* standard deviation σ) of distances is the arithmetic mean of means μ^s (*resp.* standard deviations σ^s).

Algorithm 4 : $\bar{\Sigma}, \mu, \sigma = \text{GRP\_Calibration} \left(\{\{\Sigma_t^s\}_{t=1}^{T^s}\}_{s=1}^{S}\right)$

1: **for** $s \leftarrow 1,\ldots,S$
2: $\quad \bar{\Sigma}^s, \mu^s, \sigma^s = \text{RP\_Calibration} \left(\{\Sigma_t^s\}_{t=1}^{T^s}\right)$
3: Estimation of $\bar{\Sigma}$ as the mean of matrices $\{\bar{\Sigma}^s\}_{s=1}^{S}$
4: Estimation of μ as the mean of $\{\mu^s\}_{s=1}^{S}$
5: Estimation of σ as the mean of $\{\sigma^s\}_{s=1}^{S}$

V. FROM ONE POTATO TO A POTATO FIELD

The original Riemannian potato is usually applied on all the C channels, which gives a feature space of dimension $m = C(C+1)/2$. When the dimension C of the covariance matrix becomes too large, any artifact generating an important variation in one dimension disappears in the noise of all other dimensions. The distance between the current covariance matrix Σ_t and the reference matrix $\bar{\Sigma}_t$ does not increase significantly and, consequently, such artifact is not detected.

This phenomenon, related to the curse of dimensionality, can occur for a headset with $C = 8$ channels giving a space of dimension $m = 36$. It is furtherly emphasized for
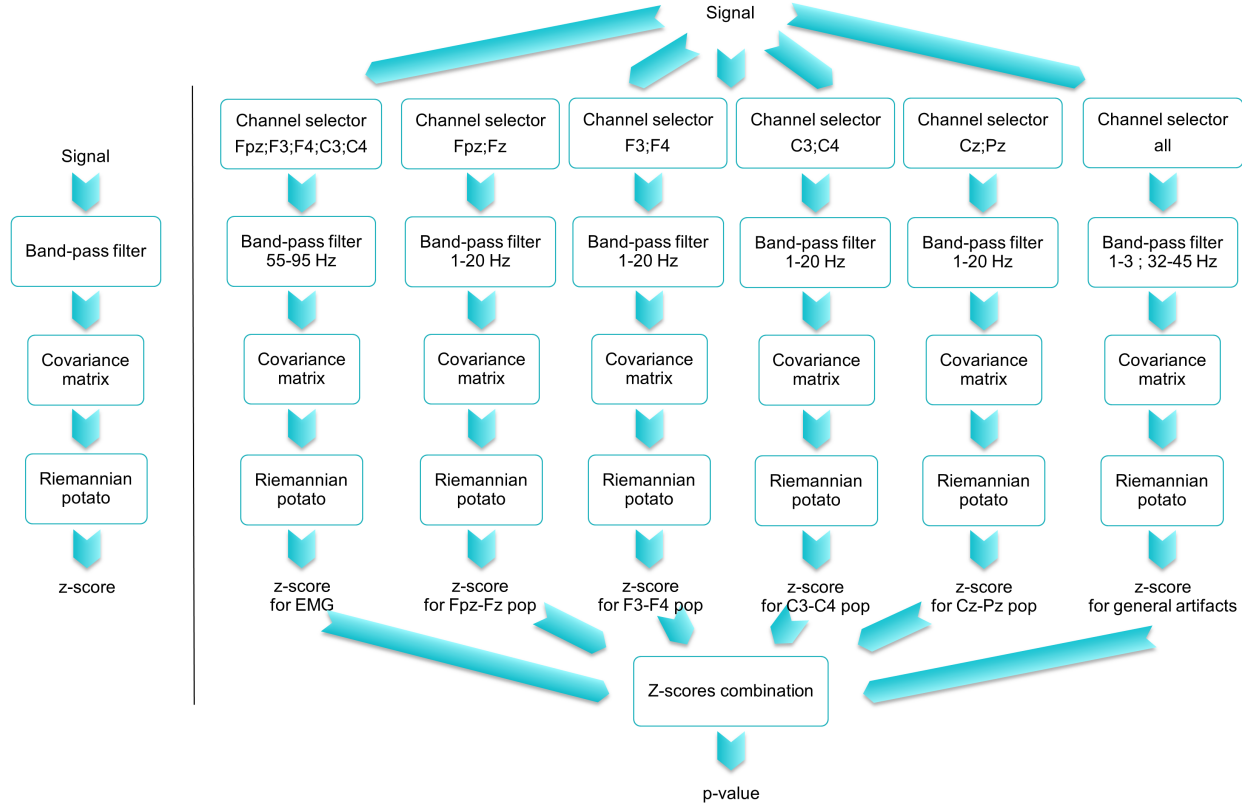
Fig. 2: Illustration of the difference between implementations of the Riemannian potato (left) and the Riemannian potato field (right).

headsets with $C = 16, 32, 64$ and more, giving dimensions $136, 528$ and 2080, respectively. To overcome this limitation, the Riemannian potato field defines and combines several potatoes of smaller dimensions.

A. Principle of the RPF

The RPF consists in using several potatoes of low dimension that are placed in parallel, each one of them being designed to capture specific artifact typically affecting specific spatial areas (*i.e.*, to a subset of channels) and/or specific frequency bands. Then, the output z-scores of these small potatoes are combined into a single SQI. Depending on the available channels, different potatoes are defined to detect all types of potential artifacts [4], [36], [57].

For ocular artifacts, a potato is applied to frontal electrodes with a signal filtered in low frequencies. For muscular artifacts, several potatoes can be defined with the signal filtered in high frequencies: a potato applied on temporal electrodes for each side, left and right, to detect jaw clenching and swallowing; a potato applied on frontal electrodes to detect forehead tension and a potato applied on occipital electrodes to detect neck tension. For intermittent electrode disconnections ("electrode pops"), a potato is applied on each pair of channels, with a signal filtered in low frequencies. For other artifacts contaminating all channels (*e.g.*, head motion artifacts), the classic potato can be applied in one or several frequency bands, excluding target bands (in BCI or NFB, modulations of target frequency bands should not be detected as artifacts). In this lower dimensional framework, variations caused by artifacts produce a more significant increase of potato distances and derived z-scores leading to a greatly increased sensitivity.

After defining a set of J potatoes composing the RPF, their output z-scores are combined in a single p-value using the right-tail Fisher's method [58] where for z-scores z_j, $j = 1, \ldots, J$, their associated p-values p_j are defined as:

$$p_j = 1 - \text{cdf}_{\mathcal{N}_{0,1}}(z_j) \;, \tag{15}$$

where $\text{cdf}_{\mathcal{D}}$ denotes the cumulative distribution function of $\mathcal{D}$ and $\mathcal{N}_{0,1}$ the normal distribution of mean 0 and standard deviation 1. The Fisher's method combines the p-values p_j as [58]:

$$q = -2 \sum_{j=1}^{J} \log(p_j) \;, \quad q \sim \chi^2_{2J} \;, \tag{16}$$

where χ^2_{2J} is the chi-square distribution with $2J$ degrees of freedom. The final p-value is thus:

$$p = 1 - \text{cdf}_{\chi^2_{2J}}(q) \;, \quad p \sim \mathcal{U}_{0,1} \;, \tag{17}$$

where $\mathcal{U}_{0,1}$ is the uniform distribution between 0 and 1. In the RPF, $p \in [0, 1]$ is the SQI, combining the output z-scores of the Riemannian potatoes.

This SQI can be applied to different use cases. For quality assessment, it can be averaged across all epochs of a recording so as to give an averaged SQI, which would provide a more representative value than the percentage of clean epochs. For artifact detection, a threshold p_{th} (empirically set to 0.01) is applied on p: an EEG epoch is considered as normal/clean

when p is higher than threshold p_{th}, and abnormal/artifacted when p is lower than p_{th}. Essentially, p is the probability to reject the null hypothesis "clean EEG".

B. Design of the Field

Since potatoes of the field are dedicated to specific spatial areas and/or frequency bands, the Riemannian potato field must be designed for each headset, considering its specific channel localizations. In this work, we illustrate with a headset with $C = 8$ electrodes, placed according to the international 10-20 system [59]: Fpz, F3, Fz, F4, C3, Cz, C4 and Pz.

1) Potato for muscular artifacts: Electromyographic (EMG) activities, such as jaw clenching, swallowing, chewing, shivering, forehead wrinkling, neck contraction, and facial muscles typically contaminate electrodes closest to muscles and affect high frequencies [60] with a spectral peak at around 70Hz. To detect these muscular artifacts, peripheral electrodes are selected: Fpz, F3, F4, C3 and C4, and signals band-pass filtered from 55 to 95Hz.

2) Potato for ocular artifacts: Electrooculographic (EOG) activities such as eye-blinks and eyes movements (horizontal and vertical) mostly contaminate frontal electrodes with low frequencies (inferior to 20 Hz) [61], [62]. To detect these ocular artifacts, frontal electrodes are selected: Fpz, F3, Fz and F4, and band-pass filtered from 1 to 20 Hz.

3) Potato for temporary electrode contact loss ("pops"): Electrode pops are generated by a temporary contact loss: it can occur in each electrode and typically provokes a discontinuity in the recording followed by a typical oscillatory period generated by the impulse response of various analogous and digital filtering steps, eventually converging to normal functioning. This phenomenon results in a sudden drop in the covariance of this channel with any other, thereby affecting the off-diagonal term of any bivariate covariance matrix including the disconnected channel. Consequently, any two-by-two pairs of electrode band-pass filtered around 1 and 20 Hz would effectively detect these events: Fpz with Fz, F3 with F4, C3 with C4, and Cz with Pz.

4) Potato for general artifacts: Many other artifacts can contaminate EEG affecting more broadly or less specifically channels and frequency. Movement artifacts for instance and cardiac contaminations are such examples. To detect these general artifacts, all the channels are filtered in the non-target frequency bands. In fact, BCI or NFB applications aim at the modulation specific brain oscillations, it is thus advisable to exclude such frequencies from the general potato so as to prevent desirable brain modulations to be detected as artifactual. Consequently, non-target frequencies must be chosen depending on the application. Since the data on which we validate this techniques comes from a NFB experiment in which subjects where asked to modulate their theta to beta ratio (TBR) or sensorimotor rhythms (SMR), we decided to exclude them and refrain the general potato to the following bands: 1 to 3Hz and 32 to 45 Hz, that are then concatenated in a spatio-frequential covariance matrix [17].

C. General Remarks

In summary, the Riemannian potato field introduced in this section is meant to be more sensitive to the variations of each dimension by using small potatoes dedicated to the detection of specific artifacts. The selection of channels and frequency bands must be adapted to each headset and each application. Similarly to the RP, the RPF can be used for offline averaging as well as for online detection (with three modes of initialization: dynamic, semi-dynamic and static, see Section IV-C).

The previous list of potatoes is not exhaustive and can be easily extended depending on which electrodes are available. It could also be completed to capture more artifacts. For example, a bi-variate potato on Fp1 and Fp2 will capture horizontal eyes movements at low frequencies.

The cost of the algorithm is dominated by the computation of eigen decomposition required for *i)* the distance to the reference of each potato in the field, and *ii)* the update of the each reference matrix. The complexity of an eigenvalue-eigenvector decomposition is cubic with the number of electrodes C and therefore is dominated by the largest potato of the RPF. This is a fast computation even for $C = 100$ on modern computers, however, in practice when C is large and large potatoes are used, dimensions can be reduced (by PCA), in order to keep the computational cost always very low.

The RPF, containing $J = 6$ potatoes, is illustrated on the right part of Fig. 2. For research reproducibility, plugins of RP and RPF are available in software NeuroRT Studio 3 (Mensia Technologies, Paris, France), an extended and certified version of the OpenViBE software [63], which was initially designed for real-time analysis of EEG signals, BCI and virtual reality.

VI. EXPERIMENTAL MATERIALS

The dataset used to benchmark the methods introduced in this article were collected during the NEWROFEED clinical trial [64]. Children suffering from Attention Deficit with or without Hyperactive Disorder (ADHD) aged 7 to 13 years old were included in this trial after giving informed assent and collecting consent from a legal guardian. In total 191 children were included in the study over 12 clinical centers across Europe. Amongst them, 113 were randomized to a Mensia Koala intervention and received approximately 46 thirty-minute-long EEG-NFB sessions, more than 80% of which happened in a non-controlled environment (typically at home). During such session, patients were instructed to modulate a feedback that was computed from their instantaneous EEG activity, namely, TBR or SMR.

The EEG was recorded with a sampling rate of 512 Hz using an Eego8 headset (ANT/Eemagine, Berlin, Germany) with $C = 8$ electrodes placed according to the international 10-20 system [59] at Fpz, F3, Fz, F4, C3, Cz, C4 and Pz, with reference and ground located at right and left earlobes, respectively.

During electrode setup, users where instructed with photographs on how to set the cap right, and an interactive real-time visualization of impedances required each of them to not exceed $30\,\text{k}\Omega$. After DC and powerline removal by digital
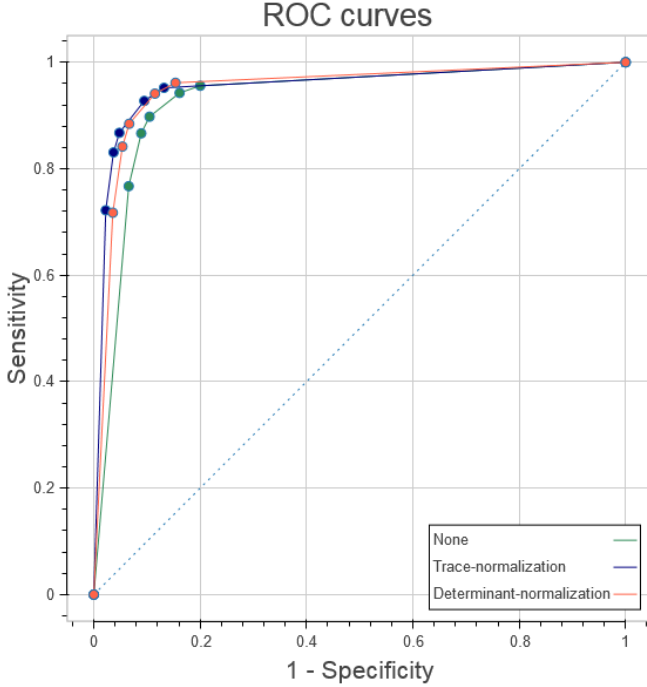
Fig. 3: The Receiver Operating Characteristic (ROC) curves for the RPF showing the Sensitivity as a function of 1-Specificity for three covariance normalization: none (green), trace-normalization (blue) and determinant-normalization (red). For these ROC curves, transferred RPF are evaluated using the true labels from self-calibrated RPF.
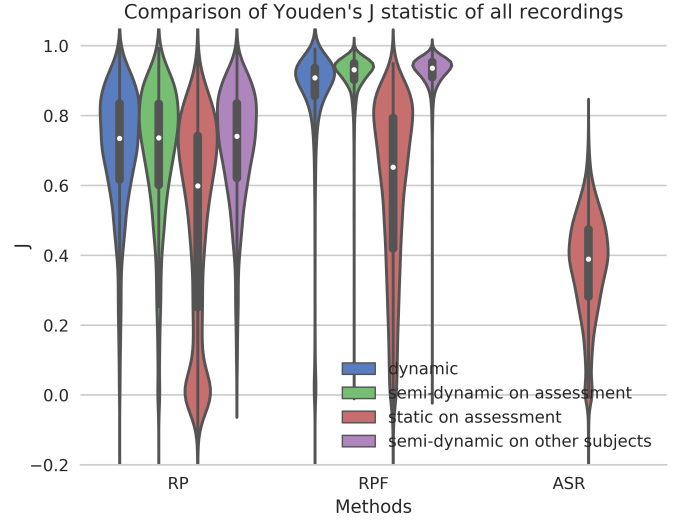

Fig. 4: Distributions of Youden's J statistics for the RP, RPF and ASR operating at different level of transfer learning and different modes of update: dynamic (blue), semi-dynamic calibrated on first assessment (green), static calibrated on first assessment (red), and semi-dynamic calibrated on other subjects (purple).

filtering, a two-step strategy was applied for real-time artifact management: *i)* correction of strong and unavoidable artifacts in eyes-open EEG, *i.e.* eye-blinks, and then *ii)* detection of remaining artifacts by the RPF described in Fig. 2. Correction of eye-blinks is performed by an online BSS method based on the approximate joint diagonalization of Fourier cospectra (AJDC) from 1 to 21 Hz [8], explaining why there is no potato dedicated to ocular artifacts in the RPF of Fig. 2. The blink source is monitored in real-time and a threshold allows a local removal of blinks [65], avoiding to remove residual neural activity when there is no blink. Detection of remaining artifacts by the online RPF[6] ensures NFB training was performed on high SNR epochs.

The goal of the following two sections is to validate the covariance normalization for transfer learning and then the artifact detection introduced in the methods above.

VII. VALIDATION OF COVARIANCE NORMALIZATION FOR CROSS-SESSION TRANSFER

To validate the cross-session transfer ability of the semi-dynamic RPF, four subjects were randomly selected from the database and it was manually verified that their recordings were mostly composed of good-quality epochs. Using an iterative offline outliers removal, the robust Riemannian mean of each of those recordings is expected to well estimate the

---

[6]Online RPF has been used with a cross-session transfer for the initialization of the potatoes, with a semi-dynamic mode and with geometric statistics. Covariance matrices have been computed on sliding windows of 2 s every 0.125 s, and have been trace-normalized. Parameters values were $\alpha = \beta = 1/t$, $T_{\text{init}} = 50$, $z_{\text{th}} = 2.5$, $p_{\text{th}} = 0.01$.

reference covariance of clean EEG, since most of the recording was clean. In this context, we believe that a RPF applied on the recording used for calibration is a good automated "silver-standard" for the assessment of transfer learning efficacy of this method, that is calibrated on a different reference recording and therefore exposed to additional sources of variability. To summarize, the following modes of operation were compared in this section: *i)* a *transferred RPF*, a RPF calibrated on a reference session and then applied to detect artifacts on other sessions (of the same subject), and *ii)* a *self-calibrated* RPF, a RPF calibrated on a session and then retrospectively applied to detect artifacts on this same session; it is our silver standard defining the "true" labels.

The validation is carried using a leave-one-session-out cross-validation procedure, during which the comparison is repeated for each session. At each iteration, the transferred RPF is calibrated on the left-out session and applied pseudo-prospectively to the others so as to be compared with the silver standard (self-calibrated RPF). This procedure is repeated so that each session eventually serves once as a reference and contributes exactly $S - 1$ times to the assessment of the performance, where S is the total number of sessions by subject. The Receiver Operating Characteristic (ROC) curves are computed by considering several operating points with thresholds on the combined Fishers' p-values.

The ROC curve is plotted in Fig. 3 for the three covariance normalizations: none (green), trace-normalization (blue), and determinant-normalization (red) with associated Areas Under the ROC curves (AUCs) at 0.929, 0.952, and 0.948, respectively. This experiment shows that covariance normalization improves cross-session transfer learning, and that the two normalization strategies behave similarly.

TABLE I: Table of statistical significance for pairwise tests. A *** symbol indicates that the p-value is smaller than 0.001. The direction of the arrows indicates which of the two methods performs best: → for the method in the row and ↑ for the method in the column.

dynamic RP	semi-dynamic RP on assessment	static RP on assessment	semi-dynamic RP on other subjects	dynamic RPF	semi-dynamic RPF on assessment	static RPF on assessment	semi-dynamic RPF on other subjects	static ASR on assessment
dynamic RP								
*** →	semi-dynamic RP on assessment							
*** ↑	*** ↑	static RP on assessment						
*** →	*** →	*** →	semi-dynamic RP on other subjects					
*** →	*** →	*** →	*** →	dynamic RPF				
*** →	*** →	*** →	*** →	*** →	semi-dynamic RPF on assessment			
*** ↑	*** ↑	*** →	*** ↑	*** ↑	*** ↑	static RPF on assessment		
*** →	*** →	*** →	*** →	*** →	*** →	*** →	semi-dynamic RPF on other subjects	
*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	*** ↑	static ASR on assessment

VIII. Validation of the RPF

In order to evaluate the performance of the RPF, one needs to compare its performance to a reliable "truth". Manual annotation are often considered as "ground truth" or "gold standard" but suffers two major limitations: *i)* its implementation is often unpractical for large volume of data and *ii)* it is hard to capture inter- and intra-expert variability unless asking several annotations per observation, which falls back into the first limitation. We believe that large amount data of relatively high quality coupled with the natural robustness of the Riemannian methods opens an opportunity for a reliable automated silver standard. In order not to bias the standard towards any particular method, it will consist in a majority voting between the two suggested methods plus a univariate Gaussian model of normality, each technique being calibrated on the whole data of each subject (using a generic calibration for RP and RPF), which accounts for the larger source of variability [9].

The univariate Gaussian model of normality is simply comparing the instantaneous power metrics to their normal univariate distribution computed from the whole subject's data. The metrics are the electromagnetic log-power of $50\,\text{Hz}$ (band-pass filtering between 47 to $53\,\text{Hz}$) summed across channels, the log-power in high frequencies (high-pass filtering from 40 and notch of $50\,\text{Hz}$ and its harmonics), and the log-power in low frequencies (band-pass filtering between 0.2 to $0.55\,\text{Hz}$). The model was calibrated by computing the distribution of each feature and defining thresholds of acceptance as $[Q_1 - 1.5\text{IQR}, Q_3 + 1.5\text{IQR}]$, with Q_1 and Q_3 the first and third quartile and IQR the inter-quartile range. This was done under the hypothesis that the distributions are normal with zero-kurtosis and potential artifacts in the calibration set would mostly affect the right-hand side of the distribution as artifacts often display higher energy. All these univariate features were finally combined using a logical 'OR' operation requiring any of the three indicators to trigger a bad quality for the resulting epoch to be classified as noisy.

Nine methods are then compared to this self-calibrated majority voting standard. Eight methods derive from the Cartesian product of the following two parameters: method (RP or RPF) and update mode (dynamic, semi-dynamic initialized on assessment (the subject's first available recording), semi-dynamic initialized on all other subjects (using a generic calibration), and static initialized on assessment). These Riemannian techniques are compared to another online method called Artifact Subspace Reconstruction (ASR) [66], originally designed for artifact correction, but which can be easily used for artifact detection[7] and can be considered as online static[8].

In terms of criteria for the evaluation of the methods performance, one has to steer away from traditional statistics. Indeed, in order to account for both the sensitivity and the specificity of the methods, a Youden's J statistic [67] was preferred and computed as $J = \text{sensitivity} + \text{specificity} - 1 \in [-1, 1]$, giving an equal contribution from false positives and false negatives (which is crucial in our case, where classes are unbalanced). The J statistic was finally computed for each NFB session and the resulting distributions are plotted in Fig. 4. Results visually suggest that the RPF (middle) outperforms RP (left) and ASR (right). The statistical significance of these differences was assessed using a paired non-parametric Wilcoxon signed-rank T test. The test was applied on each pair of methods and multiple testing was corrected with a T-min permutation test correcting efficiently the family-wise error rate [68]. The significances of these 36 pairwise tests are reported in Table I.

Before studying the results, we must mention that 75 recordings have been excluded from this experiment, because the beginning of these recordings was excessively artifacted and dynamic initialization failed (due to numerical errors in

---

[7]Detection uses only the right part of R operator in [66].

[8]The look-ahead window has been removed to be actually online. It is said static, because matrix $\bar{M}$ is initialized offline, but never updated thereafter.

the eigen decomposition solver) whereas the semi-dynamic and static modes simply reject these epochs. The fact that it was necessary to remove these recordings is a first indication of the superiority of the semi-dynamic and static modes over the dynamic one.

Table I shows that static RP and RPF are consistently inferior to their dynamic and semi-dynamic counterparts, which can easily be explained by their lack of adaptability to local distributions. This is also true for ASR whose inner model $\bar{M}$ is never updated [66]. This result suggests that it is important for EEG models to be adaptive possibly because the brain electric activity is notably non-stationary but also possibly due to unrelated technical factors such as the long-term changes in electrode impedance. For all other initialization modes, RPF outperforms RP. For RP and RPF, semi-dynamic modes are better than dynamic ones. Furthermore, the results show that the semi-dynamic mode calibrated on all other subjects is better than when calibrated on the first assessment, which will be discussed in the next section. To conclude, these results validate the interest of the RPF and the semi-dynamic initialization.

IX. DISCUSSION

The RPF presented and validated in this article, though superior in performance, has some limitations. First, a RP which selects all channels filtered in the same frequency band has very few parameters, unlike RPF. In fact, RPF embeds *a priori* knowledge on artifacts, defining specific channels and frequency bands for each potato. For this reason, RPF must be re-designed for each headset, considering the specific channels localizations and their potential contaminations. Moreover, to output a valid p-value (*i.e.*, uniformly distributed between 0 and 1), Fisher's [58] or Stouffer's [69] p-combination procedures assume that all input p-values are independent. This assumption does not necessarily hold when potatoes overlap. However, the goal of this method is to provide a value for SQI, and it is not important[9] if the value of the threshold is not equivalent to the percentage of rejected data (which is the true meaning of a p-value). In any case, the threshold has to be empirically found in order to separate noisy from clean EEG.

Second, the validation in Section VIII is not based on a gold standard. Since manual validation is operator dependent, a multi-evaluation would be required for a confident artifact labelling, but this is not feasible on the large volume that we used for validation. In addition to that, it is our opinion that manual annotation would also constitute a silver standard as the operator has a limited accuracy, a limited attention span, and some variability. A common alternative is to use a normality model, which is done in the experiment in an Euclidean way using univariate statistics, and in a Riemannian way using RP and RPF (both are used to avoid to introduce a bias in the voting). These models are calibrated on all data of each subject, since inter-subject variabilities are higher than inter-session ones [9], [15], and then combined through a vote. These models are applied on their training data however, the

[9]It is always possible to resample this value into a true p-value using a labelled database.

risk of overfitting is limited by the large quantity of data available and because models are generative classifiers (unlike discriminative ones, more likely to overfitting).

The results given in Table I support the fact that the RPF offers increased performance compared to the RP. However, this result only grazes the potential of the RPF because the EEG recordings only had $C = 8$ channels. With more channels, the limitations of RP will have a stronger detrimental impact, while RPF will show even better performances. Finally, regarding the best mode of calibration of the RP and RPF; the RPF calibrated on other subjects was consistently better as compared to any other calibration strategy, including a subject-specific approach. This approach combines all data from other recordings, which compensates for its non-specificity (with respect to the target subject), in contrast to a calibration based on the first assessment of the target subject, containing significantly less data but very specific. This opens the prospect of a calibration-free device using a cross-subject transfer [70].

Finally, our strategy for real-time artifacts management is in two steps, as described in Section VI: correction of unavoidable artifacts (as blinks), and then detection of remaining artifacts to suspend processing and to send a warning feedback teaching the user how to avoid artifacts. We believe that this strategy is better than a silent online denoising like ASR [66], which does not help the user to improve the quality of acquired data. Moreover, RPF could be applied to magnetoencephalographic (MEG) recordings, contaminated by specific artifacts which could be captured by specific potatoes.

This method is implemented in a ISO-13485 [71] medical device currently prescribed in Europe. The NEWROFEED clinical trial that collected the data used for this work applied this method in real-time and provided the user with a real-time feedback on its signal quality. This feature greatly contributed to the acquisition of data, whose quality was found comparable to that collected in a clinical environment under the supervision of a trained expert.

X. CONCLUSION

After reviewing the Riemannian potato for the estimation of a signal quality index (SQI), this paper has introduced the Riemannian potato field to overcome the lack of sensitivity when EEG recordings have many channels, combining several small potatoes dedicated to specific artifacts. We have also proposed a semi-dynamic calibration mode that addresses the situation when the onset of an acquisition is too noisy, and a covariance normalization approach to improve the transfer learning capacities.

Online detection methods, RP, RPF and ASR, have been compared on the largest EEG dataset collected in the home environment. The results unequivocally support the use of the RPF as a better artifact detection method than the RP, offering excellent transfer learning ability. The results of this article validate the new approach for SQI, *i.e.*, the semi-dynamic RPF, trace-normalized for optimized transfer learning capacities. RPF is appropriate for recording EEG in a non-controlled environment for out-of-the-lab applications; an open field in NFB and BCI [11], [12].

The limitations of the method presented in this paper can be addressed with the following perspectives. A first extension could be a multimodal modelization of the distribution of covariance matrices. Currently, each potato has a unique reference center. To capture the variabilities not encoded by the affine-invariance property, we could estimate several centers for each potato, using Riemannian clustering techniques like K-means, Mean-shift [72] or mixture of Riemannian Gaussian distributions [73]. Another perspective is the cross-system transfer learning, to cope with the inter-system variabilities, since current RPF must be re-designed for each headset contrary to RP. A last prospect is to use free-access EEG datasets [74] to build and share reference covariance centers of clean EEG, in order to standardize the artifact detection process across studies.

References

[1] F. Lotte, L. Bougrain, and M. Clerc, *Electroencephalography (EEG)-based Brain-Computer Interfaces*. Wiley Encyclopedia of Electrical and Electronics Engineering, 2015.

[2] M. Arns, J. M. Batail, S. Bioulac, M. Congedo, C. Daudet, D. Drapier, T. Fovet, R. Jardri, M. Le Van Quyen, F. Lotte, D. Mehler, M.-F. J. A., D. Purper-Ouakil, and F. Vialatte, "Neurofeedback: One of today's techniques in psychiatry?" *Encephale*, vol. 43, pp. 135–145, 2017.

[3] K. Sweeney, T. Ward, and S. McLoone, "Artifact removal in physiological signals - practices and possibilities," *IEEE Trans Inf Technol Biomed*, vol. 16, pp. 488–500, 2012.

[4] J. Urigüen and B. Garcia-Zapirain, "EEG artifact removal - state-of-the-art and guidelines," *J Neural Eng*, vol. 12, p. 031001, 2015.

[5] D. Montgomery, "Change: Detection and modification," *Appl Psychophysiol Biofeedback*, vol. 26, pp. 215–226, 2001.

[6] L. H. Sherlin, M. Arns, J. Lubar, H. Heinrich, C. Kerson, U. Strehl, and M. B. Sterman, "Neurofeedback and basic learning theory: implications for research and practice," *J Neurother*, vol. 15, pp. 292–304, 2011.

[7] K. Paluch, K. Jurewicz, J. Rogala, R. Krauz, M. Szczypińska, M. Mikicin, A. Wroóbel, and E. Kublik, "Beware: Recruitment of muscle activity by the EEG-neurofeedback trainings of high frequencies," *Front Hum Neurosci*, vol. 11, p. 119, 2017.

[8] Q. Barthélemy, L. Mayaud, Y. Renard, D. Kim, S. Kang, J. Gunkelman, and M. Congedo, "Online denoising of eye-blinks in electroencephalography," *Neurophysiol Clin*, vol. 47, pp. 371–391, 2017.

[9] A. Melnik, P. Legkov, K. Izdebski, S. M. Kärcher, W. D. Hairston, D. P. Ferris, and P. König, "Systems, subjects, sessions: To what extent do these factors influence EEG data?" *Front Hum Neurosci*, vol. 11, p. 150, 2017.

[10] E. Kappenman and S. Luck, "The effects of electrode impedance on data quality and statistical significance in ERP recordings," *Psychophysiology*, vol. 47, pp. 888–904, 2010.

[11] R. Leeb, S. Perdikis, L. Tonin, A. Biasiucci, M. Tavella, A. Molina, A. Al-Khodairy, T. Carlson, and J. del R. Millan, "Transferring brain-computer interfaces beyond the laboratory: Successful application control for motor-disabled users," *Artif. Intell. Med.*, vol. 59, pp. 121–132, 2013.

[12] V. Mihajlović, B. Grundlehner, R. Vullers, and J. Penders, "Wearable, wireless EEG solutions in daily life applications: What are we missing?" *IEEE J Biomed Health Inform*, vol. 19, pp. 6–21, 2015.

[13] S. Pan and Q. Yang, "A survey on Transfer Learning," *IEEE Trans Knowl Data Eng*, vol. 22, pp. 1345–1359, 2010.

[14] F. Lotte and C. Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Trans Biomed Eng*, vol. 58, no. 2, pp. 355–362, 2011.

[15] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces," *IEEE Trans Biomed Eng*, vol. 65, pp. 1107–1116, 2018.

[16] A. Barachant, A. Andreev, and M. Congedo, "The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry," in *Proc. TOBI Workshop IV*, 2013, pp. 19–20.

[17] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, pp. 1–20, 2017.

[18] F. Schettini, "Brain-computer interfaces for augmented communication: Asynchronous and adaptive algorithms and evaluation with end users," Ph.D. dissertation, Università degli studi di Bologna, 2014.

[19] L. Mayaud, S. Filipe, L. Pétégnief, O. Rochecouste, and M. Congedo, "Robust brain-computer interface for virtual keyboard (RoBIK): Project results," *IRBM*, vol. 34, pp. 131–138, 2014.

[20] L. Mayaud, S. Cabanilles, A. Van Langhenhove, M. Congedo, A. Barachant, S. Pouplin, S. Filipe, L. Pétégnief, O. Rochecouste, E. Azabou, C. Hugeron, M. Lejaille, D. Orlikowski, and D. Annane, "Brain-computer interface for the communication of acute patients: a feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device," *Brain-Computer Interfaces*, vol. 3, pp. 197–215, 2016.

[21] A. Andreev, A. Barachant, F. Lotte, and M. Congedo, *Brain-Computer Interfaces 2: Technology and Applications*. John Wiley; Sons, 2016, ch. 14 - Recreational Applications of OpenViBE: Brain Invaders and Use-the-Force, pp. 241–257.

[22] G. Borghini, P. Aricò, F. Ferri, I. Graziani, S. Pozzi, L. Napoletano, J. Imbert, G. Granger, R. Benhacene, and F. Babiloni, "A neurophysiological training evaluation metric for air traffic management," in *Int Conf IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014.

[23] F. Taya, Y. Sun, G. Borghini, P. Aricò, F. Babiloni, A. Bezerianos, and N. Thakor, "Training-induced changes in information transfer efficiency of the brain network: A functional connectome approach," in *Int IEEE / EMBS Conf Neural Engineering (NER)*, 2015, pp. 1028–1031.

[24] F. Alimardani, R. Boostani, and M. Taghavi, "Classification of BMD and schizophrenic patients using geometrical analysis of their EEG signal covariance matrices," in *Int Conf Telecommunications and Signal Processing (TSP)*, 2015.

[25] F. Alimardani, R. Boostani, and B. Blankertz, "Presenting a spatial-geometric EEG feature to classify BMD and schizophrenic patients," *Int J Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 5, pp. 79–85, 2016.

[26] ——, "Weighted spatial based geometric scheme as an efficient algorithm for analyzing single-trial EEGS to improve cue-based BCI classification," *Neural Netw*, vol. 92, pp. 96–76, 2017.

[27] C. Escolano, M. Navarro-Gil, J. Garcia-Campayo, M. Congedo, and J. Minguez, "The effects of individual upper alpha neurofeedback in ADHD: An open-label pilot study," *Appl Psychophysiol Biofeedback*, vol. 39, pp. 193–202, 2014.

[28] E. Massart and S. Chevallier, "Inductive means and sequences applied to online classification of EEG," in *Int Conf on Geometric Science of Information (GSI)*, 2017, pp. 763–770.

[29] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, Y. Hamam, and E. Monacelli, "From Euclidean to Riemannian Means: Information Geometry for SSVEP Classification," in *Int Conf on Geometric Science of Information (GSI)*, 2015, pp. 595–604.

[30] E. K. Kalunga, S. Chevallier, and Q. Barthélemy, "Data augmentation in Riemannian space for Brain-Computer Interfaces," in *ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins 2015)*, 2015.

[31] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Monacelli, and Y. Hamam, "Online SSVEP-based BCI using Riemannian geometry," *Neurocomputing*, vol. 191, pp. 55–68, 2016.

[32] V. Chandola, A. Banerjee, and A. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv*, vol. 41, pp. 1–58, 2009.

[33] I. Daly, F. Pichiorri, J. Faller, V. Kaiser, A. Kreilinger, R. Scherer, and G. Müller-Putz, "What does clean EEG look like?" in *Int Conf IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 3963–3966.

[34] V. Lawhern, W. Hairston, and K. Robbins, "DETECT: a MATLAB toolbox for event detection and identification in time series, with applications to artifact detection in EEG signals," *PLoS One*, vol. 8, pp. 1–13, 2013.

[35] H. Sagha, S. Perdikis, J. del R. Millán, and R. Chavarriaga, "Quantifying electrode reliability during brain-computer interface operation," *IEEE Trans Biomed Eng*, vol. 62, pp. 858–864, 2015.

[36] M. Mannan, M. Kamran, and M. Jeong, "Identification and removal of physiological artifacts from electroencephalogram signals: A review," *IEEE Access*, vol. 6, pp. 30 630–30 652, 2018.

[37] S. Chevallier, E. K. Kalunga, Q. Barthélemy, and F. Yger, *Brain Computer Interfaces Handbook: Technological and Theoretical Advances.* CRC Press, 2018, ch. 19 - Riemannian classification for SSVEP based BCI : offline versus online implementations, pp. 371–396.

[38] H. Nolan, R. Whelan, and R. Reilly, "FASTER: Fully automated statistical thresholding for EEG artifact rejection," *J Neurosci Methods*, vol. 192, pp. 152–162, 2010.

[39] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, pp. 229–240, 2011.

[40] I. Daly, R. Scherer, M. Billinger, and G. Müller-Putz, "FORCe: Fully online and automated artifact removal for brain-computer interfacing," *IEEE Trans Neural Syst Rehabil Eng*, vol. 23, pp. 725–736, 2015.

[41] M. Chaumon, D. Bishop, and N. Busch, "A practical guide to the selection of independent components of the electroencephalogram for artifact correction," *J Neurosci Methods*, vol. 250, pp. 47–63, 2015.

[42] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans Biomed Eng*, vol. 59, no. 4, pp. 920–928, 2012.

[43] A. Barachant and M. Congedo, "A Plug&Play P300 BCI using information geometry," *arXiv preprint arXiv:1409.0107*, 2014.

[44] R. Tomioka, K. Aihara, and K.-R. Müller, "Logistic regression for single trial EEG classification," in *NIPS*, vol. 19, 2007, pp. 1377–1384.

[45] M. Congedo, "EEG source analysis," Grenoble University, Tech. Rep. HDR, 2013.

[46] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in Brain-Computer Interfaces: a review," *IEEE Trans Neural Syst Rehabil Eng*, vol. 25, pp. 1753–1762, 2017.

[47] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J Matrix Anal Appl*, vol. 26, no. 3, pp. 735–747, 2005.

[48] L. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *J Neural Eng*, vol. 15, p. 031005, 2018.

[49] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans Med Imaging*, vol. 23, no. 8, pp. 995–1005, 2004.

[50] M. Congedo, A. Barachant, and E. Koopaei, "Fixed point algorithms for estimating power means of positive definite matrices," *IEEE Trans Signal Process*, vol. 65, pp. 2211–2220, 2017.

[51] B. Reuderink and M. Poel, "Robustness of the common spatial patterns algorithm in the BCI-pipeline," Univ. Twente, Tech. Rep., 2008.

[52] M. Grosse-Wentrup, C. Liefhold, K. Grasmann, and M. Buss, "Beamforming in noninvasive brain-computer interfaces," *IEEE Trans Biomed Eng*, vol. 56, pp. 1209–1219, 2009.

[53] E. K. Kalunga, S. Chevallier, and Q. Barthélemy, "Using Riemannian geometry for SSVEP-based brain computer interface," Tshwane University of Technology & Université de Versailles Saint-Quentin, Tech. Rep., 2015.

[54] T. Uehara, M. Sartori, T. Tanaka, and S. Fiori, "Robust averaging of covariances for EEG recordings classification in motor imagery brain-computer interfaces," *Neural Comput*, vol. 29, pp. 1631–1666, 2017.

[55] E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Process Lett*, vol. 3, pp. 248–250, 1996.

[56] T. Kawatani, "Handwritten Kanji recognition with determinant normalized quadratic discriminant function," in *IEEE Int Conf on Pattern Recognition*, 2000, pp. 343–346.

[57] M. Fatourechi, A. Bashashati, R. Ward, and G. Gary, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin Neurophysiol*, vol. 118, pp. 480–494, 2007.

[58] R. Fisher, *Statistical Methods for Research Workers.* Oliver and Boyd, 1925.

[59] H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalogr Clin Neurophysiol*, vol. 10, pp. 371–375, 1958.

[60] I. Goncharova, D. McFarland, T. Vaughan, and J. Wolpaw, "EMG contamination of EEG: spectral and topographical characteristics," *Clin Neurophysiol*, vol. 114, pp. 1580–1593, 2003.

[61] D. Hagemann and E. Naumann, "The effects of ocular artifacts on (lateralized) broadband power in the EEG," *Clin Neurophysiol*, vol. 112, pp. 215–231, 2001.

[62] M. Iwasaki, C. Kellinghaus, A. Alexopoulos, R. Burgess, A. Kumar, Y. Han, H. Lüders, and R. Leigh, "Effects of eyelid closure, blinks, and eye movements on the electroencephalogram," *Clin Neurophysiol*, vol. 116, pp. 878–885, 2005.

[63] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, "OpenViBE: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments," *Presence*, vol. 19, pp. 35–53, 2010.

[64] Mensia Technologies, "Effectiveness of a Personalized Neurofeedback Training Device (ADHD@Home) in Attention-Deficit / Hyperactivity Disorder (Newrofeed)," http://clinicaltrials.gov/ct2/show/study/NCT02778360, 2016.

[65] M. Delpozo-Baños and C. Weidemann, "Localized component filtering for electroencephalogram artifact rejection," *Psychophysiology*, vol. 54, pp. 608–619, 2017.

[66] T. Mullen, C. Kothe, Y. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans Biomed Eng*, vol. 62, pp. 2553–2567, 2015.

[67] W. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, pp. 32–35, 1950.

[68] P. Westfall and S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* Wiley-Blackwell, 1993.

[69] S. Stouffer, E. Suchman, S. DeVinney, L.C.; Star, and R. Williams, *The American Soldier, Vol.1: Adjustment during Army Life.* Princeton University Press, 1949.

[70] E. K. Kalunga, S. Chevallier, and Q. Barthélemy, "Transfer learning for SSVEP-based BCI using Riemannian similarities between users," in *Proc. Eur. Signal Process. Conf. EUSIPCO*, 2018, pp. 1685–1689.

[71] International Organization for Standardization, "ISO 13485:2003 Medical devices - Quality management systems - Requirements for regulatory purposes," https://www.iso.org/standard/36786.html, 2003.

[72] R. Subbarao and P. Meer, "Nonlinear mean shift over Riemannian manifolds," *Int J Comput Vis*, vol. 84, pp. 1–20, 2009.

[73] S. Said, L. Bombrun, Y. Berthoumieu, and J. Manton, "Riemannian Gaussian distributions on the space of symmetric positive definite matrices," *IEEE Transactions on Information Theory*, vol. 63, pp. 2153–2170, 2017.

[74] V. Jayaram and A. Barachant, "MOABB: trustworthy algorithm benchmarking for BCIs," *J Neural Eng*, vol. 15, p. 066011, 2018.