SCIENTIFIC REPORTS

OPEN

Recursive Exponentially Weighted N-way Partial Least Squares Regression with Recursive-Validation of Hyper-Parameters in Brain-Computer Interface Applications

Andrey Eliseyev[1], Vincent Auboiroux [1], Thomas Costecalde[1], Lilia Langar[2], Guillaume Charvet[1], Corinne Mestais[1], Tetiana Aksenova[1] & Alim-Louis Benabid[1]

A tensor-input/tensor-output Recursive Exponentially Weighted N-Way Partial Least Squares (REW-NPLS) regression algorithm is proposed for high dimension multi-way (tensor) data treatment and adaptive modeling of complex processes in real-time. The method unites fast and efficient calculation schemes of the Recursive Exponentially Weighted PLS with the robustness of tensor-based approaches. Moreover, contrary to other multi-way recursive algorithms, no loss of information occurs in the REW-NPLS. In addition, the Recursive-Validation method for recursive estimation of the hyper-parameters is proposed instead of conventional cross-validation procedure. The approach was then compared to state-of-the-art methods. The efficiency of the methods was tested in electrocorticography (ECoG) and magnetoencephalography (MEG) datasets. The algorithms are implemented in software suitable for real-time operation. Although the Brain-Computer Interface applications are used to demonstrate the methods, the proposed approaches could be efficiently used in a wide range of tasks beyond neuroscience uniting complex multi-modal data structures, adaptive modeling, and real-time computational requirements.

The Brain-Computer Interface (BCI) measures and processes the brain's neural activity to provide a patient with a non-muscular communication pathway to external devices such as robotic arms, wheelchairs, or exoskeletons[1–4]. BCI systems could be used to rehabilitate and improve the quality of life of individuals with severe motor disabilities[5,6]. To register the neural activity, electroencephalography (EEG)[1,7], electrocorticography (ECoG)[8–11], microelectrodes array[12,13], and magnetoencephalography (MEG)[14–16], etc. are applied in BCIs.

The high dimensionality of the data is unique to neural activity processing[17]. Partial Least Squares (PLS) regression[18,19] and Principal Component Analysis (PCA)[20] are particularly appropriate for high dimension tasks. In contrast to PCA, PLS considers both the exploratory and response variables by projecting them to the low dimensional space of latent variables. Statistical properties are considered for both vector[21] and multi-way (tensor)[22] versions of PLS. Due to its efficient dimensional reduction technique PLS-family methods are often used for BCIs[11,23–25].

While most state-of-the-art approaches for a BCI decoder identification are vector oriented[11,25–31], it was shown[32] that multi-way (tensor) data processing could significantly improve the quality of analysis because it considers the intrinsic nature of the data under analysis. The tensor-based approaches allow simultaneous treatment of several domains (modalities), e.g., temporal, frequency, and spatial, for BCI applications. Moreover, the influence of each modality on the prediction model could be studied[33]. In addition, different modalities could

[1]Univ. Grenoble Alpes, CEA, LETI, CLINATEC, MINATEC Campus, 38000, Grenoble, France. [2]Centre Hospitalier Universitaire Grenoble Alpes, 38700, La Tronche, France. Correspondence and requests for materials should be addressed to A.E. (email: eliseyev.andrey@gmail.com)

be independently penalized to introduce special properties to the modeling tasks, e.g., a sparse solution in the spatial domain for selection of the most informative electrodes subset[34]. Finally, as demonstrated previously[35], the tensor-based methods could improve the noise-robustness versus the vector-oriented approaches that is of great importance for BCI tasks. To date, multi-way methods were efficiently applied in several BCI systems[36–39].

The variability of the neuronal signals requires regular recalibration of the BCI systems[35,40,41]. While full system recalibration is computational and time consuming, adaptive adjustment of the previously calibrated model could successfully process mild changes in the brain signals. Different adaptive methods were proposed and shown to be efficient for both EEG-based[41–45], ECoG-[35,46] and microelectrode array-based[12] BCI systems.

Adaptation is efficient in the case of long periods of signal observation because identifies the model without storing it in the memory over entire training sets. Moreover, continuous adjustment of the predictive model allows reaction to changes in the subject's behavior with minimal delay. This is critically important for the training process[47,48].

Several vector-oriented PLS-family recursive algorithms, e.g. Recursive PLS (RPLS)[49,50] and Recursive Exponentially Weighted PLS (REW-PLS)[51,52] were proposed for adaptive modeling. Adaptive PARAFAC tensor-based algorithms were suggested recently[53,54]. Recursive Multi-way PLS (RNPLS) was proposed[35] as a generalization of the RPLS algorithm to the case of tensor-input/tensor-output variables. However, a common shortcoming of both RPLS and RNPLS methods consists in only part of informative features being stored in each subsequent iteration. This could lead to a loss of quality in the predictive model. On the other side, although the REW-PLS algorithm does not have this shortcoming, it is vector-oriented and cannot efficiently treat the multi-modal data.

Here, we propose a Recursive Exponentially Weighted Multi-way PLS (REW-NPLS) algorithm that is a generalization of the REW-PLS method to the tensor case. Similar to the REW-PLS algorithm, the REW-NPLS does not suffer from a loss of data. Moreover, the REW-NPLS inherits computational efficiency of the REW-PLS, which, as it was demonstrated in[52], has superior computational performance in comparison to RPLS-based methods. This is especially important for real-time applications. Concurrently, the REW-NPLS approach has all the advantages of tensor-based methods.

All PLS-family methods are iterative and require estimation of the iterations number (factors number) hyper-parameter. Although there is no theoretically-justified algorithm for choice of the optimal value, heuristic approaches are effective in the practical applications, e.g., in the cross-validation procedure[50,55,56]. However, cross-validation is proposed for offline-oriented tasks. It is computationally time-consuming and requires simultaneous access to the entire training set, which is not always possible for online tasks due to the memory limitations. Moreover, offline estimation of hyper-parameters does not allow their adjustment in time. Thus, the possibility to do online estimation and adaptation of the hyper-parameters are important for online tasks. Here, we proposed the Recursive-Validation procedure for hyper-parameters estimation and adjustment, which is particularly well-suited for data flow tasks. The method was integrated to the REW-NPLS algorithm and was compared to the standard "offline" cross-validation approach. The comparison demonstrates equivalent prediction accuracies for both online and offline methods.

The methods proposed here were studied with applications in brain signal decoding, however, they also could be used in a wide range of tasks beyond neuroscience where online data flows of complex structure are processed: e.g., images and video sequences analysis and adaptive monitoring of complex industrial processes.

Methods

Generic PLS and NPLS regressions.
Ordinary Partial Least Squares regression[18] is an iterative procedure to estimate a linear relationship between a vector of independent (input) variables and a vector of dependent (output) variables on the basis of observation matrices $\mathbf{X} \in \mathbb{R}^{N \times n}$ and $\mathbf{Y} \in \mathbb{R}^{N \times m}$: $\mathbf{Y} = \mathbf{XB} + \mathbf{D}$, where $\mathbf{B} \in \mathbb{R}^{n \times m}$ is the matrix of linear coefficients, and $\mathbf{D} \in \mathbb{R}^{N \times m}$ is the matrix of noise. To build the model, the observations are iteratively projected into the low dimensional spaces of latent variables while trying to explain the maximum of variance of $\mathbf{X}$ and $\mathbf{Y}$ simultaneously: $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$, $\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$; where $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_F] \in \mathbb{R}^{N \times F}$ and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_F] \in \mathbb{R}^{N \times F}$ are the matrices of the latent variables (scores), $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_F] \in \mathbb{R}^{n \times F}$ and $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_F] \in \mathbb{R}^{m \times F}$ are the matrices of the loading vectors (projectors), $\mathbf{E}$ and $\mathbf{F}$ are residual matrices, and F is the number of iterations (factors). The PLS approach is particularly well suited for high dimensional data due to its efficient dimensional reduction technique.

Multi-way Partial Least Squares (NPLS) regression is an algorithm in the PLS-family that is adopted to the case of tensor variables[32]. Tensors (multi-way arrays) are higher-order generalization of vectors and matrix data representation: vectors are tensors of order one, and the matrix is a second order tensor. In this paper, a tensor is denoted as $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \ldots \times I_n}$, where n represents the order of the tensor. Detailed information about the tensors is described[57]. NPLS combines the robustness of PLS regression with the ability to preserve the structure of the data, which is lost in vector-oriented approaches. For independent and dependent tensors of observation $\underline{\mathbf{X}} \in \mathbb{R}^{N \times I_1 \times \ldots \times I_n}$ and $\underline{\mathbf{Y}} \in \mathbb{R}^{N \times J_1 \times \ldots \times J_m}$, the NPLS iteratively constructs the linear relationship by projecting them to the space of latent variables similar to the PLS way: $\underline{\mathbf{X}} = \sum_{f=1}^{F} \mathbf{t}_f \circ \mathbf{w}_f^1 \circ \ldots \circ \mathbf{w}_f^n + \underline{\mathbf{E}}$, $\underline{\mathbf{Y}} = \sum_{f=1}^{F} \mathbf{u}_f \circ \mathbf{q}_f^1 \circ \ldots \circ \mathbf{q}_f^m + \underline{\mathbf{F}}$, $\mathbf{u}_f = \mathbf{T}_f \mathbf{b}_f$. Here, the operator "$\circ$" is the outer product[57], $\mathbf{T}_f = [\mathbf{t}_1, \ldots, \mathbf{t}_f] \in \mathbb{R}^{N \times f}$ and $\mathbf{U}_f = [\mathbf{u}_1, \ldots, \mathbf{u}_f] \in \mathbb{R}^{N \times f}$ are the matrices of the latent variables after $f = 1, \ldots, F$ iterations, $\mathbf{w}_f^i \in \mathbb{R}^{I_i}$ and $\mathbf{q}_f^j \in \mathbb{R}^{J_j}$ are the projection vectors, $\mathbf{b}_f$ is the vector of the linear coefficients, and $\underline{\mathbf{E}}$ and $\underline{\mathbf{F}}$ are the residual tensors.

Recursive PLS regression.
Several recursive PLS-based algorithms were proposed to treat online data flows without complete recalibration of the model[35,49,50,58–60]. Helland *et al.*[49] presented the Recursive PLS algorithm,

where the old data from the matrices $\mathbf{X}_{t_0} \in \mathbb{R}^{N_0 \times n}$ and $\mathbf{Y}_{t_0} \in \mathbb{R}^{N_0 \times m}$, collected before the moment t_0, are captured by their loading matrices $\mathbf{P}_{t_0}^T \in \mathbb{R}^{F \times n}$ and $\mathbf{Q}_{t_0}^T \in \mathbb{R}^{F \times m}$. The new matrices $\mathbf{X}_{t_1} \in \mathbb{R}^{N_1 \times n}$ and $\mathbf{Y}_{t_1} \in \mathbb{R}^{N_1 \times m}$ are concatenated to $\mathbf{P}_{t_0}^T$ and $\mathbf{Q}_{t_0}^T$, respectively. Then, the PLS model is recalibrated, and the new loading matrices $\mathbf{P}_{t_1}^T \in \mathbb{R}^{F \times n}$ and $\mathbf{Q}_{t_1}^T \in \mathbb{R}^{F \times m}$ are prepared for use in the next iteration of the algorithm. Thus, the method always keeps the size of the analyzed data by packing the data into the loading matrices $\mathbf{P}$ and $\mathbf{Q}$ of constant size. The main shortcoming of the RPLS consists in the possible loss of information if all of the data are not retained in the matrix of latent variables. Thus, appropriate choice of the number of factors F is of crucial importance.

Qin[50] proposed an algorithm where the hyper-parameter F is estimated by the cross-validation procedure applied for the blocks of data: the whole data are divided on the blocks followed by leave-one-block-out cross-validation. However, application of the cross-validation procedure requires storing in memory the entire data set. Moreover the cross-validation procedure is not possible for online tasks. In addition, the hyper-parameters could be non-stationary and time-varying for some tasks. Despite this, if the value of the hyper-parameter F is predefined in some way, then the Recursive PLS method is widely used due to relatively small memory requirement. Forgetting factor[59], nonlinearity[50], as well as neural networks[60] were coupled with the Recursive PLS to model time-varying process. Recursive N-way PLS (RNPLS)— a generalization of the RPLS algorithm to the case of the tensor-input/tensor-output variables—was proposed in[35].

Contrary to Qin's approach[50], Dayal and MacGregort[51,52] proposed Recursive Exponentially Weighted PLS regression, which allows online treatment of data flow without loss of information. The algorithm is extremely fast in comparison with other recursive PLS-based approaches and is of great importance for real-time applications. Moreover, the integrating forgetting factor allows one to exponentially discount past data.

Unlike Qin's RPLS, in the REW-PLS, the old data from the matrices $\mathbf{X}_{t_0} \in \mathbb{R}^{N_0 \times n}$ and $\mathbf{Y}_{t_0} \in \mathbb{R}^{N_0 \times m}$ were captured by the matrices $\mathbf{C}_{\mathbf{XX}}^{t_0} = \mathbf{X}_{t_0}^T \mathbf{X}_{t_0} \in \mathbb{R}^{n \times n}$ and $\mathbf{C}_{\mathbf{XY}}^{t_0} = \mathbf{X}_{t_0}^T \mathbf{Y}_{t_0} \in \mathbb{R}^{m \times m}$, which are used for model calibration. When the new data is available, the covariance matrices are updated: $\mathbf{C}_{\mathbf{XX}}^{t_1} = \lambda(\mathbf{C}_{\mathbf{XX}}^{t_0}) + \mathbf{X}_{t_1}^T \mathbf{X}_{t_1}$ and $\mathbf{C}_{\mathbf{XY}}^{t_1} = \lambda(\mathbf{C}_{\mathbf{XY}}^{t_0}) + \mathbf{X}_{t_1}^T \mathbf{Y}_{t_1}$. Here, λ ($0 \leq \lambda \leq 1$) is a forgetting factor. In the case of $\lambda = 1$ past data cannot be discounted. There is no loss of information in the method because the calibrated model is completely identified by the covariance matrices. The detailed description of the fast model calibration algorithm based on the covariance matrices is reported[52]. However, as for all the PLS-family methods, the hyper-parameter F should be predefined for the REW-PLS algorithm. The question of the efficient choice of F—especially for online applications—was not considered by the authors of the REW-PLS algorithm.

Recursive Exponentially Weighted NPLS. The Recursive Exponentially Weighted NPLS is a generalization of the REW-PLS to the tensor data. On the first step, the REW-NPLS algorithm receives the tensors of observation $\underline{\mathbf{X}} \in \mathbb{R}^{N \times I_1 \times \ldots \times I_n}$ and $\underline{\mathbf{Y}} \in \mathbb{R}^{N \times J_1 \times \ldots \times J_m}$. Similar to the REW-PLS approach, the covariance tensors, denoted as: $\underline{\mathbf{C}}_{\mathbf{XX}} = \underline{\mathbf{X}} \times_1 \underline{\mathbf{X}} \in \mathbb{R}^{(I_1 \times \ldots \times I_n) \times (I_1 \times \ldots \times I_n)}$ and $\underline{\mathbf{C}}_{\mathbf{XY}} = \underline{\mathbf{X}} \times_1 \underline{\mathbf{Y}} \in \mathbb{R}^{(I_1 \times \ldots \times I_n) \times (J_1 \times \ldots \times J_m)}$ are estimated. Here, the k-mode tensor product is denoted as "$\times_k$"[57]. Similar to the NPLS method, the covariance tensors are used in the PARAFAC-based tensor decomposition procedure[32] to estimate a set of projectors $\{\mathbf{w}_f^1 \in \mathbb{R}^{I_1}, \ldots, \mathbf{w}_f^n \in \mathbb{R}^{I_n}\}_{f=1}^F$ as well as the tensor of the prediction coefficients $\underline{\mathbf{B}} = (b_{i_1,\ldots,i_n,j_1,\ldots,j_m}) \in \mathbb{R}^{(I_1 \times \ldots \times I_n) \times (J_1 \times \ldots \times J_m)}$; here, F is a hyper-parameter representing the number of factors.

The application of the model to the new explanatory variables tensor $\underline{\mathbf{X}}^{\text{Test}} \in \mathbb{R}^{N_{\text{Test}} \times I_1 \times \ldots \times I_n}$ for prediction of the response variables tensor $\underline{\hat{\mathbf{Y}}}^{\text{Test}} \in \mathbb{R}^{N_{\text{Test}} \times J_1 \times \ldots \times J_m}$ was done similarly to the NPLS approach:

$$\hat{\underline{\mathbf{Y}}}_{l,j_1,\ldots,j_m}^{\text{Test}} = \sum_{i_1,\ldots,i_n} \underline{\mathbf{X}}_{l,i_1,\ldots,i_n}^{\text{Test}} \underline{\mathbf{B}}_{i_1,\ldots,i_n,j_1,\ldots,j_m}, \quad l = 1, \ldots, N_{\text{Test}}. \tag{1}$$

For simplicity, we will denote it as $\hat{\underline{\mathbf{Y}}}^{\text{Test}} = \underline{\mathbf{X}}^{\text{Test}} \mathbf{B}$.

To update the previously calibrated model, represented by the set of $\underline{\mathbf{C}}_{\mathbf{XX}}, \underline{\mathbf{C}}_{\mathbf{XY}}, \{\mathbf{w}_f^1, \ldots, \mathbf{w}_f^n\}_{f=1}^F$, and $\mathbf{B}$, with the newly available data batch $\underline{\mathbf{X}}^{\text{New}} \in \mathbb{R}^{L \times I_1 \times \ldots \times I_n}$ and $\underline{\mathbf{Y}}^{\text{New}} \in \mathbb{R}^{L \times J_1 \times \ldots \times J_m}$, and the forgetting factor λ ($0 \leq \lambda \leq 1$), the covariance tensors are recalculated as:

$$\underline{\mathbf{C}}_{\mathbf{XX}}^{\text{New}} = \lambda \underline{\mathbf{C}}_{\mathbf{XX}} + \underline{\mathbf{X}}^{\text{New}} \times_1 \underline{\mathbf{X}}^{\text{New}}, \tag{2}$$

$$\underline{\mathbf{C}}_{\mathbf{XY}}^{\text{New}} = \lambda \underline{\mathbf{C}}_{\mathbf{XY}} + \underline{\mathbf{X}}^{\text{New}} \times_1 \underline{\mathbf{Y}}^{\text{New}}. \tag{3}$$

The new projectors $\{\mathbf{w}_f^{1,\text{New}}, \ldots, \mathbf{w}_f^{n,\text{New}}\}_{f=1}^F$ are derived from the new covariance tensors by the PARAFAC-based tensor decomposition procedure with the previous projectors $\{\mathbf{w}_f^1, \ldots, \mathbf{w}_f^n\}_{f=1}^F$ as an initial approximation. The $\underline{\mathbf{B}}^{\text{New}}$ is then defined. A graphical representation of the REW-NPLS method is shown in Fig. 1. A detailed description of the algorithm is given in Appendix A.

The mean-centering and rescaling of each column of the observation data should be done as in the PLS-family methods. In a time-varying process, the centering and rescaling parameters could change with time and should be continuously updated. The procedure for the matrix case is reported[51]. The same approach could be applied for the tensors because all parameters are estimated independently for each coordinate.

At moment t, tensors of observation $\underline{\mathbf{X}}_{t-1} \in \mathbb{R}^{N_{t-1} \times I_1 \times \ldots \times I_n}$ and $\underline{\mathbf{Y}}_{t-1} \in \mathbb{R}^{N_{t-1} \times J_1 \times \ldots \times J_m}$ are updated with newly available tensors $\underline{\mathbf{X}}_t \in \mathbb{R}^{N_t \times I_1 \times \ldots \times I_n}$ and $\underline{\mathbf{Y}}_t \in \mathbb{R}^{N_t \times J_1 \times \ldots \times J_m}$. Taking into account the forgetting factor λ, the effective number of observation at t is $N^{\text{eff}(t)} = \lambda N_{t-1} + N_t = \lambda N^{\text{eff}(t-1)} + N_t$.

The sum and the sum of squares of the elements along the observation modality are estimated as:
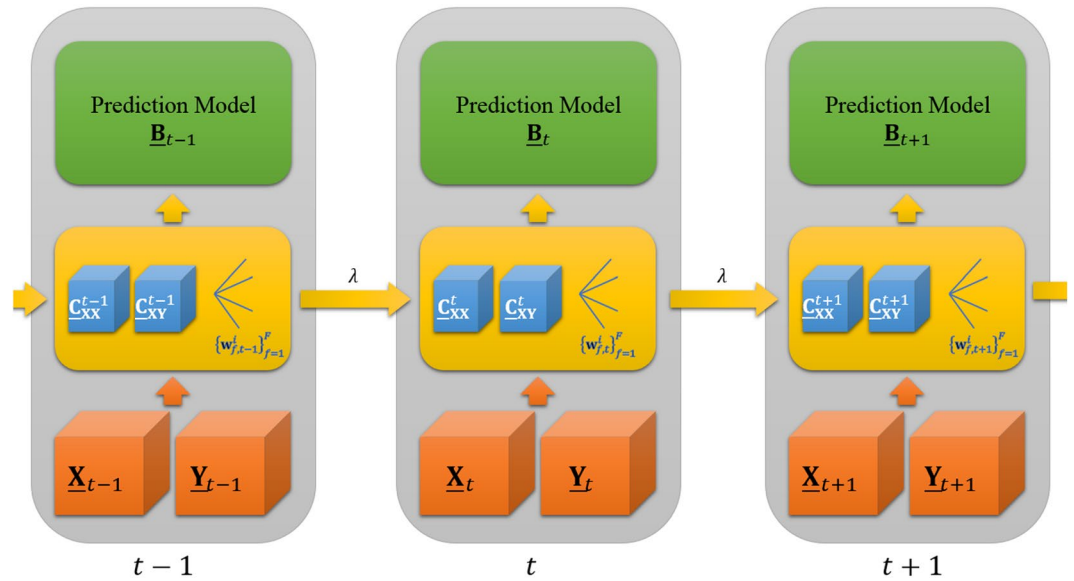
REW-NPLS scheme

Figure 1. The REW-NPLS scheme. When the new tensors of observation $\{\underline{\mathbf{X}}_t, \underline{\mathbf{Y}}_t\}$ are available, their details are then combined with the previously captured covariance tensors $\{\underline{\mathbf{C}}_{XX}^{t-1}, \underline{\mathbf{C}}_{XY}^{t-1}\}$, and the set of projection vectors $\{\mathbf{w}_{f,t-1}^1, \ldots, \mathbf{w}_{f,t-1}^n\}_{f=1}^F$ with forgetting factor λ. Then the prediction model, represented by the regression coefficients $\underline{\mathbf{B}}_t$, is generated.

$$\underline{\mathbf{S}}^{\mathrm{eff}(t)} = \lambda \underline{\mathbf{X}}_{t-1} \times_1 1^{N_{t-1}} + \underline{\mathbf{X}}_t \times_1 1^{N_t} \in \mathbb{R}^{I_1 \times \ldots \times I_n}, \tag{4}$$

$$\underline{\mathbf{SS}}^{\mathrm{eff}(t)} = \lambda \underline{\mathbf{X}}_{t-1}^{*2} \times_1 1^{N_{t-1}} + \underline{\mathbf{X}}_t^{*2} \times_1 1^{N_t} \in \mathbb{R}^{I_1 \times \ldots \times I_n}, \tag{5}$$

where "$\underline{\mathbf{X}}^{*2}$" represents the element-wise square of the tensor $\underline{\mathbf{X}}$. The effective mean and the standard deviation along the first modality of the tensors at time t are:

$$\underline{\boldsymbol{\mu}}_1 \mathbf{X}^{\mathrm{eff}(t)} = \underline{\mathbf{S}}^{\mathrm{eff}(t)} / N^{\mathrm{eff}(t)} \in \mathbb{R}^{I_1 \times \ldots \times I_n}, \tag{6}$$

$$\underline{\boldsymbol{\sigma}}_1 \mathbf{X}^{\mathrm{eff}(t)} = \sqrt{\frac{\underline{\mathbf{SS}}^{\mathrm{eff}(t)} - (\underline{\mathbf{S}}^{\mathrm{eff}(t)})^{*2} / N^{\mathrm{eff}(t)}}{N^{\mathrm{eff}(t)} - 1}} \in \mathbb{R}^{I_1 \times \ldots \times I_n}. \tag{7}$$

For simplicity of notation, we omit the indexes and write $\underline{\boldsymbol{\mu}}\mathbf{X}$ and $\underline{\boldsymbol{\sigma}}\mathbf{X}$. The tensors $\underline{\boldsymbol{\mu}}\mathbf{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_m}$ and $\underline{\boldsymbol{\sigma}}\mathbf{Y} \in \mathbb{R}^{J_1 \times \ldots \times J_m}$ are identified similarly.

To apply the centered model to the non-centered testing data $\underline{\mathbf{X}}^{\mathrm{Test}}$, equation (1) should be rewritten:

$$\hat{\underline{\mathbf{Y}}}^{\mathrm{Test}} = \underline{\mathbf{X}}^{\mathrm{Test}} \widetilde{\underline{\mathbf{B}}} + \underline{\mathbf{Y}}_0, \tag{8}$$

where $\widetilde{\underline{\mathbf{B}}} = (b_{i_1,\ldots,i_n,j_1,\ldots,j_m} \underline{\boldsymbol{\sigma}}\mathbf{Y}_{j_1,\ldots,j_m} / \underline{\boldsymbol{\sigma}}\mathbf{X}_{i_1,\ldots,i_n}) = \underline{\mathbf{B}}\underline{\boldsymbol{\sigma}}\mathbf{Y} / \underline{\boldsymbol{\sigma}}\mathbf{X}, \underline{\mathbf{Y}}_0 = \underline{\boldsymbol{\mu}}\mathbf{Y} - \underline{\boldsymbol{\mu}}\mathbf{X}\widetilde{\underline{\mathbf{B}}}$.

A detailed description of the normalization procedure is given in Appendix B.

Determination of the appropriate number of factor F is a common task for all PLS-family methods. While the cross-validation approach[55] is widely used for this purpose, it cannot be efficiently applied for online modeling. Here, we propose a procedure for recursive estimation of the optimal number of factors (F^*) for REW-NPLS algorithm. The main idea exploits the newly available data as a testing set for the currently available models before the models are updated.

At time t, the new tensors of observation $\underline{\mathbf{X}}_t$ and $\underline{\mathbf{Y}}_t$ are available. The current prediction model $\underline{\mathbf{B}}_{t-1}^{F_{\max}}$ resulted from the previous data and is computed for the maximum considered number of factors $F_{\max}$. The distinctive property of the REW-NPLS algorithm consists in the ability to generate all intermediate models $\{\underline{\mathbf{B}}^1, \ldots, \underline{\mathbf{B}}^{F_{\max}-1}\}$ during the computation of $\underline{\mathbf{B}}^{F_{\max}}$, see Appendix A. All models defined in the previous step $\{\underline{\mathbf{B}}_{t-1}^1, \ldots, \underline{\mathbf{B}}_{t-1}^{F_{\max}}\}$ could be applied to the newly available tensor of observation $\underline{\mathbf{X}}_t$ that results in a set of predictions $\{\hat{\underline{\mathbf{Y}}}_t^1, \ldots, \hat{\underline{\mathbf{Y}}}_t^{F_{\max}}\}$. The error of each prediction could be estimated: $\{e_t^1, \ldots, e_t^{F_{\max}}\}$, where $e_t^f = \gamma e_{t-1}^f + \mathrm{ERROR}(\hat{\underline{\mathbf{Y}}}_t^f, \underline{\mathbf{Y}}_t)$, with the forgetting factor γ ($0 \leq \gamma \leq 1$). The optimal number of factors, corresponding to the moment t is defined as
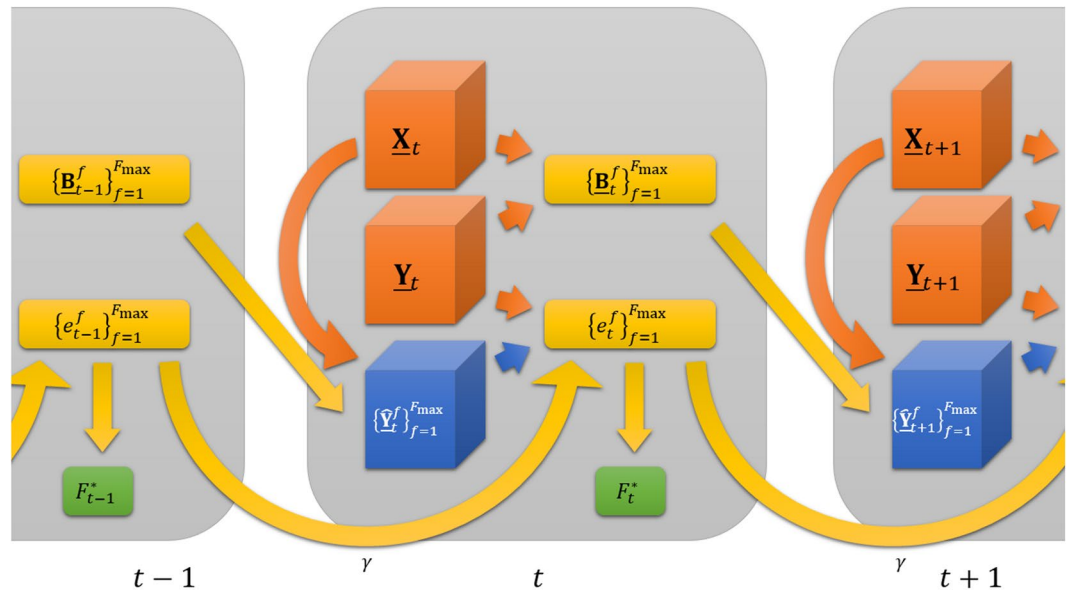
Figure 2. The Recursive-Validation scheme for the number of factors hyper-parameter (F) identification. When the new tensors of observation $\{\underline{\mathbf{X}}_t, \underline{\mathbf{Y}}_t\}$ are available, the previously defined models $\{\underline{\mathbf{B}}_{t-1}^1, \ldots, \underline{\mathbf{B}}_{t-1}^{F_{\max}}\}$ are tested on the new data and resulted in $\{\underline{\widehat{\mathbf{Y}}}_t, \ldots, \underline{\widehat{\mathbf{Y}}}_t^{F_{\max}}\}$. The prediction errors $\{e_t^f\}_{f=1}^{F_{\max}}$ are estimated considering the previous errors with the forgetting factor γ. The number of factors $F_t^* = \mathrm{argmin}_f \{e_t^f\}_{f=1}^{F_{\max}}$ correspond to the minimal error at the current moment t; this is considered to be optimal for the current moment.

$$F_t^* = \mathrm{argmin}_f \{e_t^f\}_{f=1}^{F_{\max}}, \tag{9}$$

and the optimal model is

$$\underline{\mathbf{B}}_t^* = \underline{\mathbf{B}}_t^{F_t^*}. \tag{10}$$

The detailed description of the Recursive-Validation procedure is given in Appendix C. Its graphic representation is shown in Fig. 2.

Experiments in Monkeys. *Data description.* To validate the proposed approaches, the publicly available dataset is considered (http://neurotycho.org/epidural-ecog-food-tracking-task). The data are recorded and distributed by the Laboratory for Adaptive Intelligence, BSI, RIKEN. All procedures were performed in accordance with protocols approved by the RIKEN ethics committee. The experiments consisted in recording of epidural ECoG signals simultaneously with continuous 3D trajectories of a Japanese macaque's right wrist, elbow, and shoulder[25,61]. The experiments were carried out in two monkeys denoted as "B" and "C". Overall, 20 recordings (10 for each monkey) are provided[11]. The length of each recording was approximately 15 minutes. To record the hand motions (shoulders, elbows, and wrists), an optical motion capture system (Vicon Motion System, Oxford, UK) with a sampling rate of 120 Hz was applied. The ECoG signals were recorded from 64 electrodes (Blackrock Microsystems, Salt Lake City, UT, USA) with a sampling rate of 1 kHz, implanted in the epidural space of the left hemisphere (Fig. 3A). Each experiment showed the monkeys pieces of food, and they were trained to receive it with the right hands. The location of the food was random at a distance of about 20 cm from the monkeys. A scheme of the experiment is shown in Fig. 3B.

Feature extraction and tensors formation. An input data feature tensor $\underline{\mathbf{X}}$ was formed from the ECoG epochs containing 1 second of the signal taken continuously with a time step of 100 ms. Each ECoG epoch was mapped to the spatial-temporal-frequency space by continuous wavelet transform (CWT). The complex Morlet wavelet was chosen as a mother wavelet[25,38,62,63]. The frequency band from 10 to 150 Hz with step 10 Hz was chosen[11]. Absolute values of the wavelet coefficients were decimated along the temporal modality 100 times by averaging 10 sliding windows each 100 ms long. A detailed description of the feature extraction procedure has been reported[24]. The tensor of the output variables $\underline{\mathbf{Y}}$ was formed from the corresponding 3D coordinates of the monkey's right wrist, elbow, and shoulder. Figure 3C represents the scheme of the data preparation procedure.

According to[63], each 15-minute recording from the available dataset of 20 files was split into non-overlapping training (~10 first minutes) and test (~5 last minutes) subsets; see Fig. 3D:
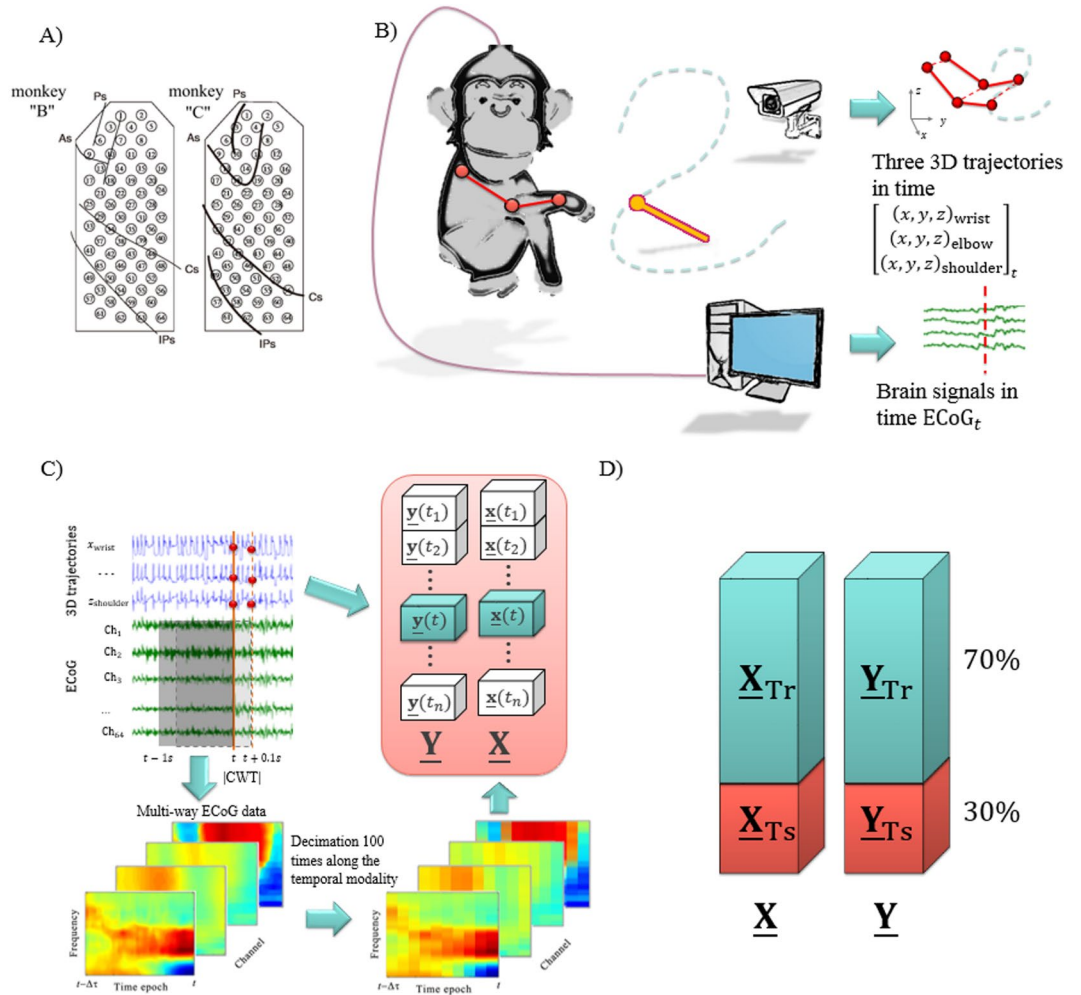
Figure 3. ECoG-based primate experiments. (**A**) 64 electrodes were implanted in the epidural space of the left hemisphere of two Japanese macaques denoted as monkey "B" and "C". Ps: principal sulcus; As: arcuate sulcus; Cs: central sulcus; IPs: intraparietal sulcus[11]. (**B**) The scheme of the experiment. The monkey is following the food with its right hand. Monkey's ECoG activity is recorded simultaneously with 3D coordinates of the right wrist, elbow, and shoulder. (**C**) For each time *t*, to form the explanatory variable $\underline{x}(t) \in \mathbb{R}^{15 \times 10 \times 64}$, the ECoG signal from 64 channels is mapped by the Continuous Wavelet Transform with 15 frequencies (10, 20, …, 150 Hz). Then, the absolute values of the wavelet coefficients are decimated 100 times along the temporal modality, i.e., 1000 points, representing 1 second; these are decimated to 10 points. The response variable $\underline{y}(t) \in \mathbb{R}^{3 \times 3}$ is formed from the corresponding 3D coordinates (*x*, *y*, *z*) of monkey's wrist, elbow, and shoulder. The next epoch was taken with a time step of 100 ms. (**D**) Data tensors are split into non-overlapping training (70%) and test (30%) sub-tensors.

$$\underline{X}^{\text{recoding}} = \begin{bmatrix} \underline{X}^{\text{recoding}}_{\text{training}} \in \mathbb{R}^{7000 \times 15 \times 10 \times 64} \\ \underline{X}^{\text{recoding}}_{\text{test}} \in \mathbb{R}^{3000 \times 15 \times 10 \times 64} \end{bmatrix},$$

(11)

$$\underline{Y}^{\text{recoding}} = \begin{bmatrix} \underline{Y}^{\text{recoding}}_{\text{training}} \in \mathbb{R}^{7000 \times 3 \times 3} \\ \underline{Y}^{\text{recoding}}_{\text{test}} \in \mathbb{R}^{3000 \times 3 \times 3} \end{bmatrix}.$$

(12)

The models provided by the different methods are calibrated on the training tensors and validated on the corresponding test tensors.

Experiments in humans. *Data description.* The experiments were carried out by the Clinatec team (CEA-LETI-CLINATEC®, Grenoble, France) and recorded the MEG signals in four healthy subjects (25–43 years old) with no known neurological or psychiatric problem. This study was approved by Comité de protection des personnes Sud-Est V ethics committee (Clinical Trial NTC02574026, AFFSAPS 2010-A00421-38). All experiments were performed in accordance with relevant guidelines and regulations. Informed consent was obtained
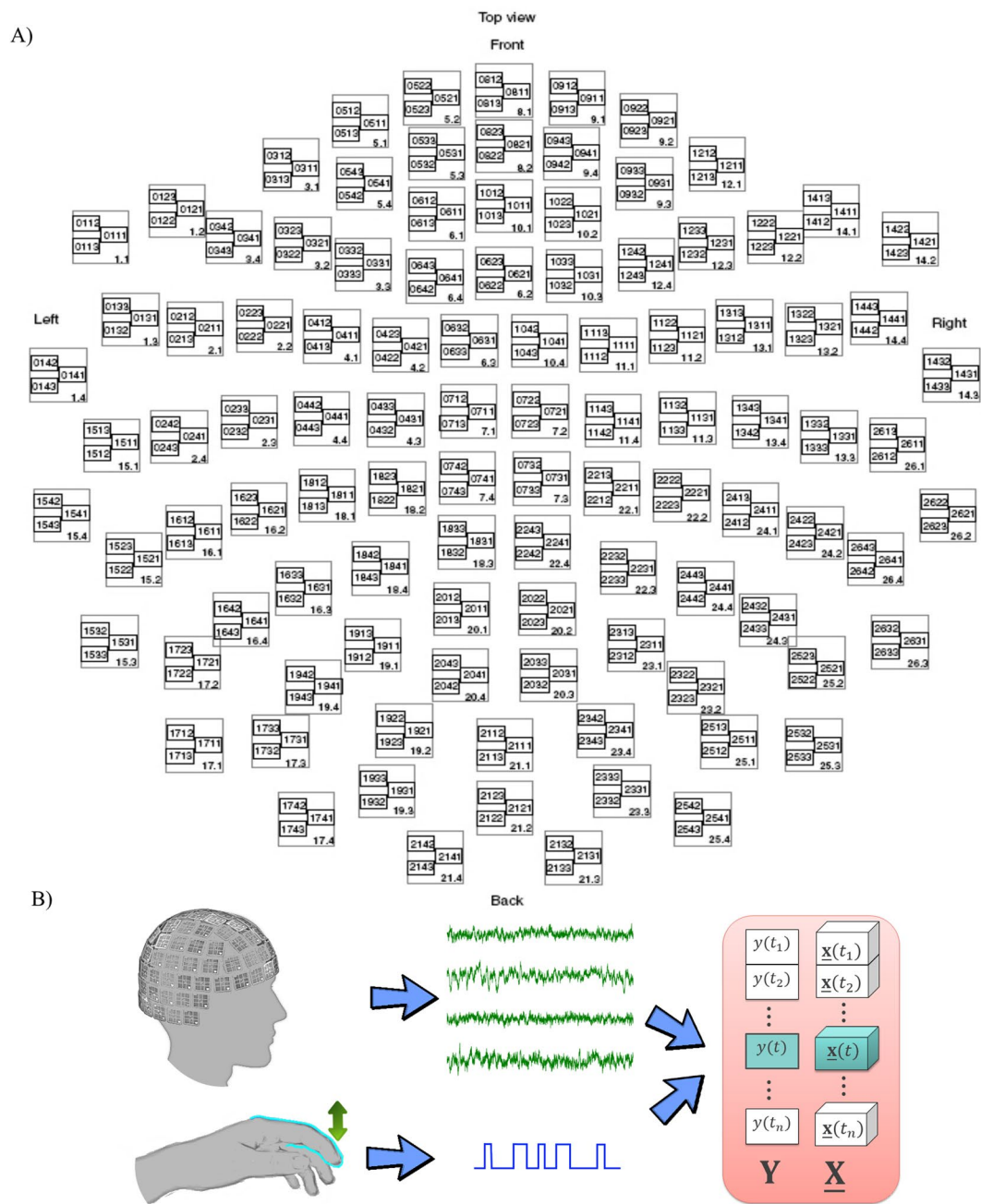
Figure 4. MEG-based experiments in human. (**A**) Elekta Neuromag electrodes scheme. (**B**) The scheme of the experiment. The subject voluntarily performs up/down movements of the index finger. For each time moment t, to form the explanatory variable $\underline{\mathbf{x}}(t) \in \mathbb{R}^{20 \times 10 \times 306}$, the MEG signal from 306 channels is mapped by the Continuous Wavelet Transform with 20 frequencies (5, 10, …, 100 Hz). Then, the absolute values of the wavelet coefficients are decimated 100 times along the temporal modality. The response variable $\mathbf{y}(t) \in \{0, 1\}$ is formed from the corresponded down/up position of the index finger at the moment t. The next epoch was taken with a time step of 100 ms.

from all participants. During the experiments, each subject lifted his/her index finger without any external cue or conditional stimulus. The binary position of the finger ("up" or "down") was registered by a laser beam. Only one finger was moved during each session. The sessions varied between 5 and 20 minutes and provided 90 to 520 self-paced finger movements. The brain activity recordings were made in a magnetically shielded room using 306-channel MEG devices (Elekta Neuromag, Helsinki, Finland). The device provides MEG data recorded from 102 triple sensor units (one magnetometer and two orthogonal planar gradiometers) with a sampling rate of 1 kHz (Fig. 4A).
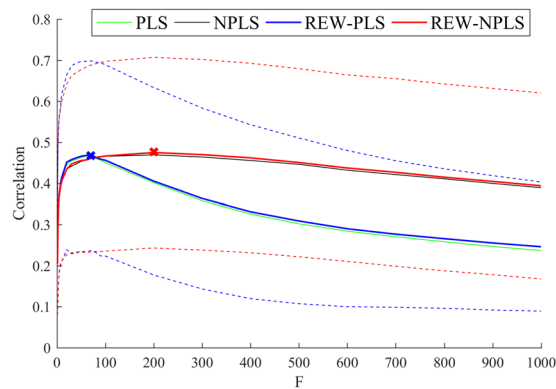
Figure 5. Prediction accuracy and robustness to the factors number parameter (F). The maximal prediction accuracy is represented by the crosses in the figure and is almost equivalent for all the approaches. However, in comparison with vector-oriented methods (PLS, REW-PLS), the tensor-based approaches (NPLS, REW-NPLS) demonstrate better robustness to variability of F. Dotted lines represent the standard deviation of the corresponding correlations.

Method	Optimal Factors Number (F^*)	Correlation (r^*)
PLS	70	0.47 ± 0.23
REW-PLS	70	0.47 ± 0.23
NPLS	200	0.47 ± 0.23
REW-NPLS	200	0.48 ± 0.23

Table 1. The optimal values of the factors number F^* and the values of the corresponding correlations $r^* = r(F^*)$ on the test data for the comparison methods.

Feature extraction and tensors formation. Similarly to the experiments in monkeys, an input data feature tensor **X** was formed from the MEG epochs containing 1 second of the signal taken continuously with a time step of 100 ms. Each MEG epoch, was mapped to the spatial-temporal-frequency space by CWT based on the complex Morlet wavelet. The frequency band from 5 to 100 Hz with step 5 Hz was used. The binary vector of the output variables **Y** was formed from the corresponding position of the index finger. Figure 4B represents the scheme of the experiments and the data preparation procedure.

Four subjects participated in the experiments. All four sessions were recorded with each subject: 2 with left and 2 with right index movements. Overall, 16 recordings were prepared. Each recording was split on two non-overlapping training (70%) and test (30%) subsets.

Prediction performance evaluation. To predict performance evaluation, several criteria have been applied in BCI experiments[64]. In this paper, we use Pearson correlation (r) between predicted and observed response because it is sufficiently informative and could be easily computed and interpreted. It is commonly used in BCIs to assess decoders[11,25,62,63].

Results

Experiments in monkeys.
Here we consider vector- and tensor-oriented approaches (PLS and NPLS) as well as their recursive modification (REW-PLS and REW-NPLS). For the recursive approaches, the training sets were split on the non-overlapping 10-second long batches with 100 epochs per batch and about 70 batches per training set. The recursive models (REW-PLS and REW-NPLS) were adjusted each time when the new batch arrived. The non-recursive models (PLS and NPLS) were calibrated just once on the entire training set. For the recursive approaches, a forgetting factor $\lambda = 1$ was chosen, i.e., there was no discounting of past data batches.

Prediction accuracy and robustness. The methods were compared based on their prediction accuracy and robustness to the value of the number of factors. For each approach, namely, PLS, NPLS, REW-PLS, and REW-NPLS, the correlation between predicted and observed trajectories is represented in Fig. 5 and is averaged over recordings, hand's joints (shoulder, elbow, wrist), and coordinates. The optimal values of the factor numbers that provide the maximum of averaged correlations and the corresponding correlations are given in Table 1.

Recursive-Validation of factors number. To assess the quality of the Recursive-Validation procedure in estimating the number of factors F, the prediction results were compared to the REW-NPLS with an optimal number of factors $F^* = 200$ (see Table 1). To model the online data flow, the training dataset was split on seventy 10-second non-overlapping batches (100 epochs per batch). When the new data-batch arrives, the model was adjusting according to the new data and was then tested on entire testing set. For the Recursive-Validation procedure, the
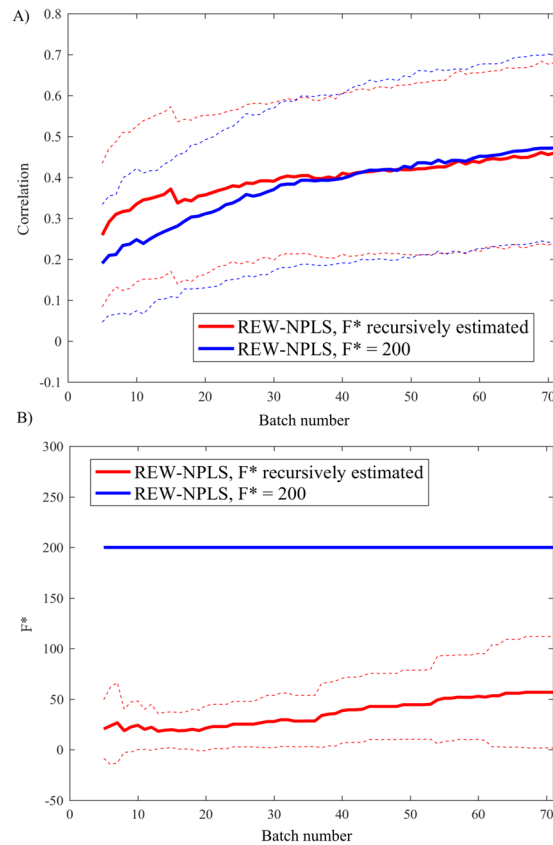
Figure 6. Comparison of the prediction correlation for REW-NPLS; the preliminary estimate of optimal number of factors ($F^* = 200$) vs. the Recursive-Validation of the factors number. The size of each data-batch is equal to 10 seconds (100 epochs), and 70 batches are available. The data are averaged through 20 recordings and 9 coordinates. The maximum number of factors $F_{max} = 300$ is taken in the Recursive-Validation approach. (**A**) The method with Recursive-Validation of F^* significantly outperforms the one with optimal $F^* = 200$ until 25 batches are analyzed. The difference then became insignificant (ANOVA test, significance level $\alpha = 0.05$). (**B**) When 70 batches are treated, the dynamically estimated number of factors is $F^*_{RV} = 57 \ll F^* = 200$, whereas the difference in prediction accuracy is insignificant. Dotted lines represent standard deviations.

maximal number of tested factors was $F_{max} = 300$ (Fig. 6A). The REW-NPLS with dynamic adaptation of F significantly outperformed the REW-NPLS with an optimal value of factors ($F = F^*$) until the number of batches reached 25. The difference became insignificant (ANOVA test, significance level $\alpha = 0.05$). Figure 6B demonstrates the optimal number of factors (F^*_{RV}) estimated by the Recursive-Validation procedure as a function of the number of analyzed batches. When the complete training tensor was analyzed (all 70 batches), the dynamically-estimated optimal number of factors was $F^*_{RV} = 57 \ll F^* = 200$; the prediction correlations are comparable.

Modality influences analysis. Tensor-based approaches facilitate study of the resulting models in different modalities[33]. The predictive models are identified by the REW-NPLS with Recursive-Validation of F^* and are compared to the model identified on the entire training data with an optimal number of factors $F^* = 200$; see Fig. 7. The models are represented by their averaged influences[65] in spatial, temporal, and frequency modalities. Each influence is computed as the weight of sum of the absolute values of the model's coefficients along the corresponding modality. The weights are averaged over the recordings and coordinates. For the case of the Recursive-Validation of F^*, the models are shown for batch numbers 10, 25, 40, 55, and 70 (complete dataset).

Experiments in humans. *Recursive-Validation vs. Cross-Validation.* To evaluate the proposed Recursive-Validation procedure, the number of factors estimated by this approach was compared to the number of factors estimated by the standard 10-fold cross-validation procedure. The number of factors was taken from the interval $F \in [1, \ 20]$. Table 2 demonstrates the optimal number of factors F^* estimated by the methods as well as the value of the corresponding correlation $r^* = r(F^*)$ on the test sets. The difference in the resulting correlations was not significant for either left- or right indexes (ANOVA test, significance level $\alpha = 0.05$).

Modality influences analysis. The predictive models were identified by the REW-NPLS with Recursive-Validation of F^* on the complete training set and are shown in Fig. 8. The models are represented by their averaged influence
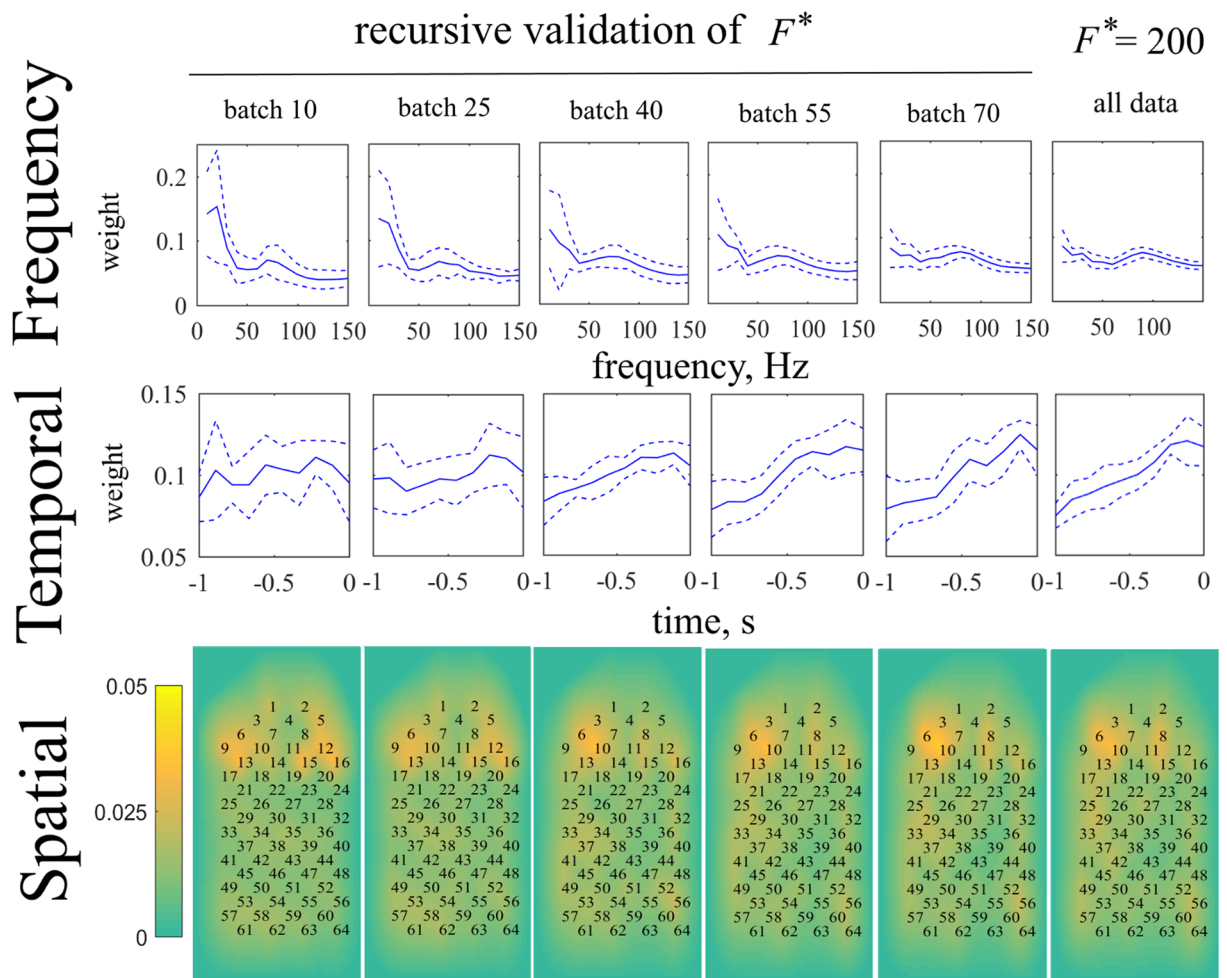
Figure 7. The influences on the predictive models of the elements in frequency, temporal, and spatial modalities identified by the REW-NPLS ($F^* = 200$) on the complete training set, and the REW-NPLS with Recursive-Validation of F^*. The batch numbers 10, 25, 40, 55, and 70 (complete dataset) demonstrate the model's variability over time. The modalities influences are averaged over 20 recordings and 9 coordinates. Dotted lines represent the standard deviation of the corresponding results.

Method	Finger	Optimal Factors Number (F^*)	Correlation (r^*)
10-fold Cross-Validation	Left	6 ± 3	0.26 ± 0.10
	Right	6 ± 2	0.27 ± 0.08
Recursive-Validation	Left	5 ± 2	0.28 ± 0.10
	Right	5 ± 2	0.29 ± 0.10

Table 2. The optimal values of the factors number F^* and the corresponding correlations $r^* = r(F^*)$ for the comparison methods. The results are averaged over 8 recordings for the left- and right fingers.

in frequency, temporal, and spatial modalities. The weights are averaged over 4 subjects (2 recording per subject) for left and right finger self-paced movements.

Implementation

The crucial point for practical application of signal-processing approaches in BCI systems is a possibility for of the real-time computations. All methods proposed here are implemented in Matlab® (The MathWorks, Inc., Natick, US) and can function in real-time on a standard computer (Intel Xenon E5-2620, 2 GHz, RAM 32 Gb). Due to real-time functioning requirements, blocks of applications and adaptations of the model are separated and implemented in two independent processes while communicating through shared memory. The Model Application Block provides a 10-Hz decision rate, i.e. control command is generated each time that the new 100 ms data block
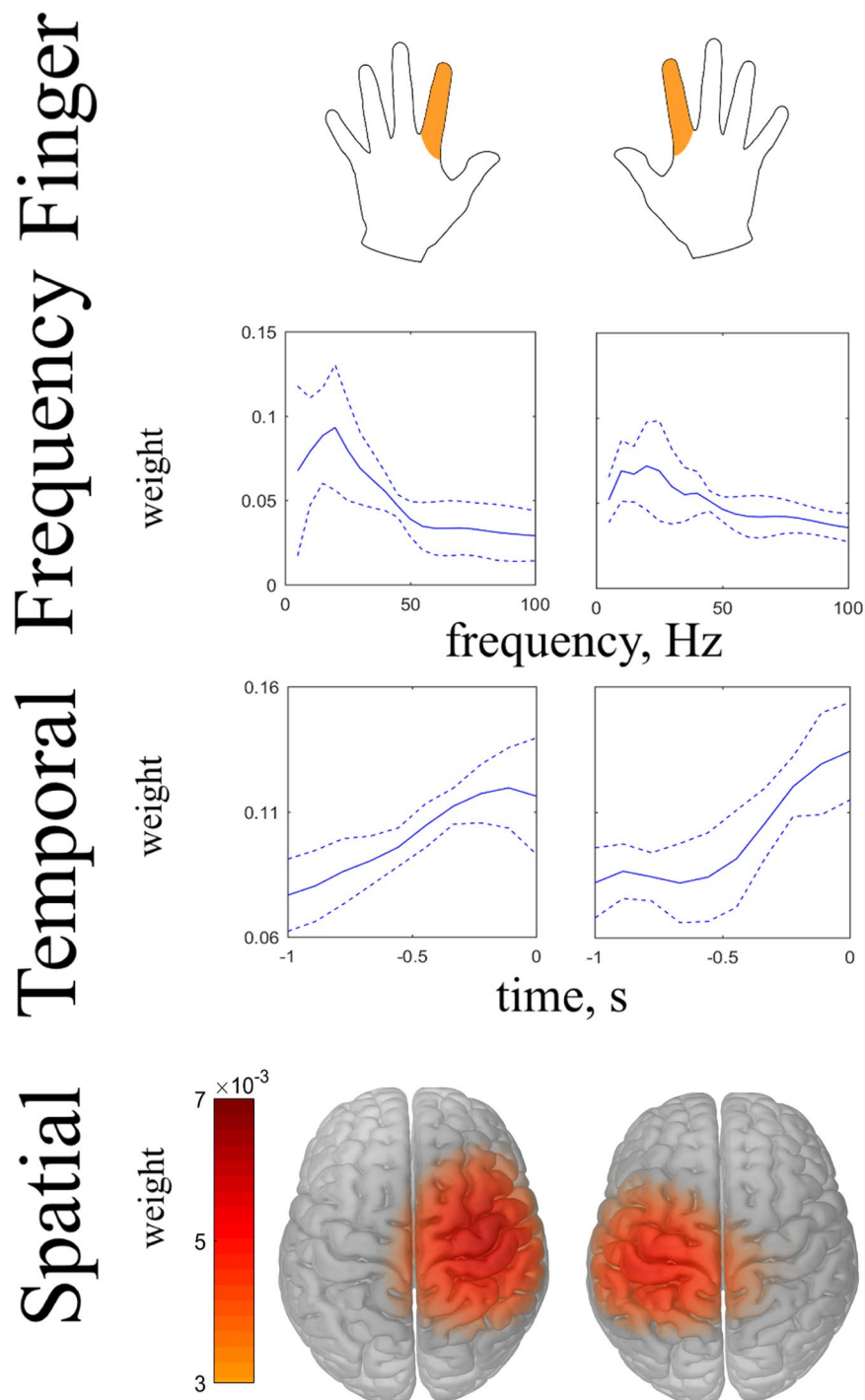
Figure 8. The influences on the predictive models of the elements in frequency, temporal, and spatial modalities identified by the REW-NPLS with Recursive-Validation of F^* on the complete training set. The modalities influence was averaged over 8 recordings for the left- and right fingers. Dotted lines represent the standard deviation of the corresponding results.

arrives. The Model Adaptation Block is waiting when the new data stack of 10 seconds is acquired. The model then is updated with newly available data (REW-NPLS algorithm with Recursive-Validation of the factors number).

When the updated model is prepared, it replaces the current model in the Model Application Block. For the Model Application Block, the processing time for generation of a control command with a new 100 ms data block (sampling rate 1 kHz, 64 channels, 15 analyzed frequencies) is 37 ± 8 ms. Updating the model with 10 seconds of data stack in the Model Adaptation Block takes 3.41 ± 0.06 s. Thus, both blocks meet the real-time functioning requirements (37 ms $\ll$ 100 ms and 3.41 s $\ll$ 10 s). The scheme of the system is given in Fig. 9.
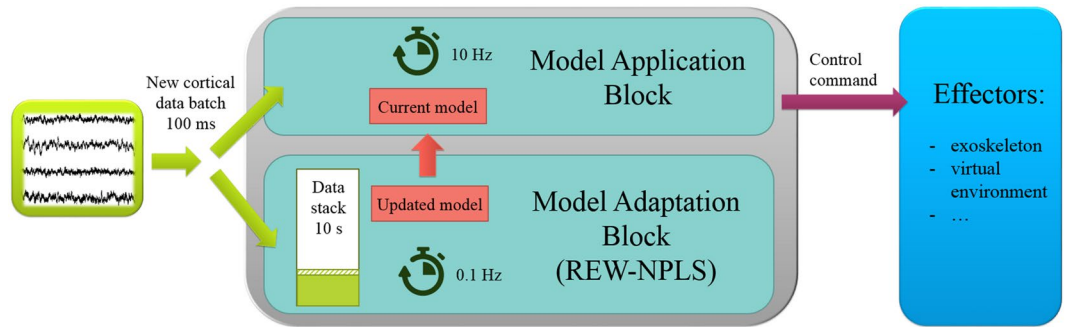
Figure 9. Implementation scheme of the real-time operating system. The control command for an external device is generated every 100 ms by the Model Application Block. The model is adjusted every 10 seconds by the Model Adaptation Block based on the REW-NPLS algorithm with Recursive-Validation of the factor number.

Discussion

Decoding of the neural activity for BCI systems is a challenging task[1]. The huge dimension, considerable data variability in time, as well as real-time computational requirements impose essential limitations on real-life applications of BCIs[66,67].

The paper proposes the REW-NPLS algorithm for recursive analysis of multimodal data. While the standard vector-oriented approaches applied in BCI do not consider the structure of the analyzed data[11,68], the tensor-based methods are commonly not adjusted for online computations and require complete recalibration when the new data are available[38,69]. Previously[35], the RNPLS algorithm allows online treatment of multi-modal data, however, only part of information is extracted from arriving data batches. This could lead to a loss in the quality of the prediction model. The REW-NPLS method unites both tensor-based structures of NPLS with the possibility of efficient online treatment of data without information loss inherited from the REW-PLS.

The crucial parameter of the PLS-family methods is the number of iterations (number of factors, F) used in the model for projections into the space of latent variables[70]. The robustness of the methods in the case of F's suboptimal values is very important because there are mostly heuristic algorithms for selection of the optimal value of F. The task is especially critical for online data flows when the values of the parameters are unfixed and could vary widely with time.

In the current paper, the proposed REW-NPLS approach was studied for the robustness of its prediction accuracy to the suboptimal values of the number of factors on the ECoG data. The recursive approaches (REW-NPLS, REW-PLS) do not concede to the corresponding non-recursive approaches (NPLS, PLS) according to their prediction quality result; Fig. 5 and Table 1. Moreover, the recursive approaches are slightly better and this could be explained by the additional stability provided by the batch-wise normalization. In comparison with the tensor-based methods (NPLS, REW-NPLS), the vector-oriented approaches (PLS, REW-PLS) require fewer factors to achieve best correlation: $F^*_{\text{REW−PLS}} = 70, r^*_{\text{REW−PLS}} = 0.47 \pm 0.23, F^*_{\text{REW−NPLS}} = 200, r^*_{\text{REW−NPLS}} = 0.48 \pm 0.23$. However, the prediction accuracy of the vector-oriented approaches drops less than 10% only when the number of factors varying in the range $20 \leq F \leq 100$; the tensor-oriented methods remain within 90% of the optimal correlation in the range $20 \leq F \leq 600$. The prediction accuracy results provided by the tensor-oriented approaches significantly outperform the vector-oriented ones for $F \geq 200$ (ANOVA test, significance level $\alpha = 0.05$).

The cross-validation procedure[56] is a state-of-the-art approach to estimate the number of factor hyper-parameters for the PLS-family methods, but it is time and memory intensive. Moreover, its direct application to online data flows is impossible. In addition, any predefined and fixed value of the hyper-parameter could be inappropriate because it should be regularly re-estimated to follow possible data variations over time. The Recursive-Validation procedure proposed here allows online estimates of the parameter values without significant additional computational time expenses.

Figure 6A demonstrates the prediction accuracy of the REW-NPLS model (identified with a predefined optimal number of factors $F^* = 200$) compared to the REW-NPLS coupled with Recursive-Validation procedure for estimation of factor numbers. Both approaches showed equivalent prediction accuracies as follows from the experiments. Moreover, when the number of batches for model calibration is less than 25, the use of constant F^* leads to overfitting in small training sets. However, the Recursive-Validation procedure adjusts F^* value to optimize the prediction accuracy. Figure 6B demonstrates that the number of factors estimated by the Recursive-Validation procedure is significantly less than F^* as estimated offline on the entire dataset. This additionally reduces the possibility of overfitting.

When all 70 batches are processed, $F^*_{\text{RV}} = 57 \ll F^* = 200$, the difference in prediction accuracies is insignificant (ANOVA test, significance level $\alpha = 0.05$). Table 2 compares cross-validation and Recursive-Validation approaches on the MEG data set. The difference in the estimated optimal values of the factor numbers F^*_{RV} and F^*_{CV}, as well as the corresponding values of the correlations on the test sets, are statistically insignificant (ANOVA test, significance level $\alpha = 0.05$). Moreover, in comparison with the cross-validation method, the number of the factors provided by the Recursive-Validation approach is slightly smaller ($F^*_{\text{RV}} = 5$ vs. $F^*_{\text{CV}} = 6$), whereas the

correlation is essentially the same ($r_{RV}^* = 0.29$ vs. $r_{CV}^* = 0.27$) for both left- and right index fingers movements. This could be explained by additional robustness of the Recursive-Validation procedure.

For tensor-based approaches, Influence Analysis can assess the weights of the temporal, frequency, and spatial modalities on the predictive model. Figure 7 represents the evaluation of the REW-NPLS model with Recursive-Validation of F^* in time and its comparison with REW-NPLS model calibrated on the complete dataset with an optimal number of factors F^* on the ECoG data.

For the REW-NPLS algorithm with Recursive-Validation, the models identified on batches 10, 25, 40, 55, and 70 (corresponding to 15%, 36%, 57%, 79%, and 100% of the training set) are considered. For the frequency modality, the weight distributions demonstrate the same behavior for all the models: the maximum peaks are at 10, 30, and 90 Hz. In the temporal modality, the weights of the elements tend to increase until 100 ms before the motion moment for both the online and offline models. In the spatial modality, for the REW-NPLS model, the set of the most relevant electrodes (10, 6, 8, and 15) is defined on the batch number 40 and does not change after; however, the weights of the electrodes become more equilibrate with time. For the offline model, the same electrodes 10, 6, 8, and 15 are the most informative.

Figure 8 represents applications of the Influence Analysis to the models identified by the REW-NPLS algorithm with the Recursive-Validation of F^* in the MEG experiments. The results are prepared for the left and right finger self-paced movements and are averaged over 8 recordings carried out in 4 subjects. In the frequency modality, there is a maximum around 20 Hz for both fingers. The frequency bands above 50 Hz are less informative. This corresponds to the quality of the MEG-data recording system. In the temporal modality—similar to the case of the ECoG-based experiments—the weights of the elements tend to increase until the motion moment. In the spatial modality, the localization of the informative electrodes for the left and right finger correspond to the motor zone of the cortex[71]. For the left finger, the most informative electrodes are 1143, 1112, 1133, 1132, and 1142; whereas for the right finger, the most informative electrodes are 0442, 0433, 0712, 0443, and 0423 (according to the Elekta Neuromag system notation, see Fig. 4A).

The main limitation of the proposed REW-NPLS approach is its memory consumption, because it stores the covariance tensors in the memory. At the same time, the memory used by the algorithm is constant and does not change over time. The limitation of the Recursive-Validation algorithm is the need to compute the model for F_{max} number of factors. This could be significantly greater than the optimal number F_{RV}^*. However, F_{max} can decrease with time if $F_{RV}^* \ll F_{max}$ is constantly observed.

In parallel to a set of neuroscience tasks, such as BCIs, fMRI analysis, etc., the proposed REW-NPLS and the Recursive-Validation algorithms could be applied to a wide range of tasks where adaptive modeling of high dimension tensor flows is necessary. Examples include dynamic analysis of images and video sequences, adaptive monitoring of complex industrial processes, etc.

Perspectives

The next step of the study is application of the proposed methods implemented in real-time operating software in the clinical BCI in tetraplegic subjects. Online adjustment provided by the REW-NPLS algorithm allows application of the closed-loop paradigm when the user immediately receives feedbacks from the system to his/her movement imaginations reflected in the brain activity. This will allow adaptation of the human behavior in parallel with adjustment of the BCI decoder. This should result in efficiency of the BCI system. Within the framework of the CEA-LETI-CLINATEC® BCI project, the fully-implantable device WIMAGINE®[72] for chronic measurement and wireless transmission of ECoG data is currently developed. The full body exoskeleton EMY®[73] is designed to let the tetraplegic subjects possibility control the exoskeleton fragments in real-time. The clinical protocol named "Brain Computer Interface: Neuroprosthetic Control of a Motorized Exoskeleton" was authorized from the French regulatory agencies to start with five patients including bilateral implantation of 2 WIMAGINE® implants per patient (https://clinicaltrials.gov/ct2/show/NCT02550522).

References

1. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. & Vaughan, T. M. Brain-computer interfaces for communication and control. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology* **113**, 767–791 (2002).
2. Donoghue, J. P. Bridging the brain to the world: a perspective on neural interface systems. *Neuron* **60**, 511–521, https://doi.org/10.1016/j.neuron.2008.10.037 (2008).
3. Benabid, A. L. *et al*. Deep brain stimulation: BCI at large, where are we going to? *Progress in brain research* **194**, 71–82, https://doi.org/10.1016/B978-0-444-53815-4.00016-9 (2011).
4. Pfurtscheller, G. *et al*. The hybrid BCI. *Frontiers in neuroscience* **4** (2010).
5. Bouton, C. E. *et al*. Restoring cortical control of functional movement in a human with quadriplegia. *Nature* **533**, 247–250 (2016).
6. Chaudhary, U., Birbaumer, N. & Ramos-Murguialday, A. Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology* (2016).
7. Birbaumer, N. *et al*. A spelling device for the paralysed. *Nature* **398**, 297–298 (1999).
8. Wang, W. *et al*. An electrocorticographic brain interface in an individual with tetraplegia. *PloS one* **8**, e55344, https://doi.org/10.1371/journal.pone.0055344 (2013).
9. Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G. & Moran, D. W. A brain–computer interface using electrocorticographic signals in humans. *Journal of neural engineering* **1**, 63 (2004).
10. Schalk, G. & Leuthardt, E. C. Brain-computer interfaces using electrocorticographic signals. *IEEE reviews in biomedical engineering* **4**, 140–154, https://doi.org/10.1109/RBME.2011.2172408 (2011).
11. Shimoda, K., Nagasaka, Y., Chao, Z. C. & Fujii, N. Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in Japanese macaques. *Journal of neural engineering* **9**, 036015, https://doi.org/10.1088/1741-2560/9/3/036015 (2012).
12. Hochberg, L. R. *et al*. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* **485**, 372–375, https://doi.org/10.1038/nature11076 (2012).
13. Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S. & Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).

14. Mellinger, J. *et al.* An MEG-based brain–computer interface (BCI). *NeuroImage* **36**, 581–593 (2007).
15. Ohata, R., Ogawa, K. & Imamizu, H. Single-trial prediction of reaction time variability from MEG brain activity. *Scientific reports* **6**, 27416 (2016).
16. Kauhanen, L. *et al.* EEG and MEG brain-computer interface for tetraplegic patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **14**, 190–193 (2006).
17. Muller, K., Anderson, C. W. & Birch, G. E. Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* **11**, 165–169 (2003).
18. Geladi, P. & Kowalski, B. R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **185**, 1–17, https://doi.org/10.1016/0003-2670(86)80028-9 (1986).
19. De Jong, S. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems* **18**, 251–263 (1993).
20. Lee, J. A. & Verleysen, M. *Nonlinear Dimensionality Reduction*. (Springer Publishing Company, Incorporated, 2007).
21. Helland, I. S. On the structure of partial least squares regression. *Communications in statistics-Simulation and Computation* **17**, 581–607 (1988).
22. Zhang, X. & Li, L. Tensor Envelope Partial Least-Squares Regression. *Technometrics*, 1–11 (2017).
23. Cichocki, A., Zdunek, R., Phan, A. H. & Amari, S.-i. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. (John Wiley & Sons, 2009).
24. Eliseyev, A. & Aksenova, T. Stable and artifact-resistant decoding of 3D hand trajectories from ECoG signals using the generalized additive model. *Journal of neural engineering* **11**, 066005 (2014).
25. Chao, Z. C., Nagasaka, Y. & Fujii, N. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys. *Frontiers in neuroengineering* **3**, 3, https://doi.org/10.3389/fneng.2010.00003 (2010).
26. Rakotomamonjy, A., Guigue, V., Mallet, G. & Alvarado, V. In *artificial Neural Networks: Biological Inspirations–ICANN 2005* 45–50 (Springer, 2005).
27. Schlögl, A., Lee, F., Bischof, H. & Pfurtscheller, G. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of neural engineering* **2**, L14 (2005).
28. Vidaurre, C., Krämer, N., Blankertz, B. & Schlögl, A. Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Networks* **22**, 1313–1319 (2009).
29. Kayser, J. *et al.* Event-related brain potentials (ERPs) in schizophrenia for tonal and phonetic oddball tasks. *Biological psychiatry* **49**, 832–847 (2001).
30. Zhao, Q., Zhang, L., Cichocki, A. & Li, J. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*. 2656–2659(IEEE).
31. Scherer, R., Müller, G. R., Neuper, C., Graimann, B. & Pfurtscheller, G. An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *Biomedical Engineering, IEEE Transactions on* **51**, 979–984 (2004).
32. Bro, R. Multiway calibration. Multilinear PLS. *Journal of Chemometrics* **10**, 47–61 (1996).
33. Bro, R. *Multi-way analysis in the food industry: models, algorithms, and applications*, Københavns Universitet'Københavns Universitet', LUKKET: 2012 Det Biovidenskabelige Fakultet for Fødevarer, Veterinærmedicin og NaturressourcerFaculty of Life Sciences, LUKKET: 2012 Institut for FødevarevidenskabDepartment of Food Science, 2012 Institut for Fødevarevidenskab, 2012 Kvalitet og TeknologiDepartment of Food Science, Quality & Technology, (1998).
34. Eliseyev, A. *et al.* L1-penalized N-way PLS for subset of electrodes selection in BCI experiments. *Journal of neural engineering* **9**, 045010, https://doi.org/10.1088/1741-2560/9/4/045010 (2012).
35. Eliseyev, A. & Aksenova, T. Recursive N-way partial least squares for brain-computer interface. *PloS one* **8**, e69962, https://doi.org/10.1371/journal.pone.0069962 (2013).
36. Cichocki, A. *et al.* Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine* **32**, 145–163 (2015).
37. Hou, M. & Chaib-draa, B. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 6205–6209 (IEEE).
38. Zhao, Q. *et al.* Higher order partial least squares (HOPLS): a generalized multilinear regression method. *IEEE transactions on pattern analysis and machine intelligence* **35**, 1660–1673, https://doi.org/10.1109/TPAMI.2012.254 (2013).
39. Cong, F. *et al.* Tensor decomposition of EEG signals: a brief review. *Journal of neuroscience methods* **248**, 59–69 (2015).
40. Kaufmann, T., Völker, S., Gunesch, L. & Kübler, A. Spelling is just a click away–a user-centered brain-computer interface including auto-calibration and predictive text entry. *Frontiers in neuroscience* **6**, 72 (2012).
41. Vidaurre, C., Sannelli, C., Müller, K.-R. & Blankertz, B. Machine-learning-based coadaptive calibration for brain-computer interfaces. *Neural computation* **23**, 791–816 (2011).
42. Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P. & Müller, K.-R. Towards adaptive classification for BCI. *Journal of neural engineering* **3**, R13 (2006).
43. Vidaurre, C., Schlogl, A., Cabeza, R., Scherer, R. & Pfurtscheller, G. A fully on-line adaptive BCI. *IEEE Transactions on Biomedical Engineering* **53**, 1214–1219 (2006).
44. Sykacek, P., Roberts, S. J. & Stokes, M. Adaptive BCI based on variational Bayesian Kalman filtering: an empirical evaluation. *IEEE Transactions on biomedical engineering* **51**, 719–727 (2004).
45. Wolpaw, J. R. & McFarland, D. J. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17849–17854 (2004).
46. Li, Y., Koike, Y. & Sugiyama, M. In *Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference on*. 1–5 (IEEE).
47. Willett, F. R., Suminski, A. J., Fagg, A. H. & Hatsopoulos, N. G. Improving brain–machine interface performance by decoding intended future movements. *Journal of neural engineering* **10**, 026011 (2013).
48. Shimada, S., Hiraki, K., Matsuda, G. & Oda, I. Decrease in prefrontal hemoglobin oxygenation during reaching tasks with delayed visual feedback: a near-infrared spectroscopy study. *Cognitive brain research* **20**, 480–490 (2004).
49. Helland, K., Berntsen, H. E., Borgen, O. S. & Martens, H. Recursive algorithm for partial least squares regression. *Chemometrics and intelligent laboratory systems* **14**, 129–137 (1992).
50. Qin, S. J. *Recursive PLS algorithms for adaptive data modeling. Computers & Chemical Engineering* **22**, 503–514 (1998).
51. Dayal, B. S. & MacGregor, J. F. Recursive exponentially weighted PLS and its applications to adaptive control and prediction. *Journal of Process Control* **7**, 169–179 (1997).
52. Dayal, B. & MacGregor, J. F. Improved PLS algorithms. *Journal of chemometrics* **11**, 73–85 (1997).
53. Nion, D. & Sidiropoulos, N. D. Adaptive algorithms to track the PARAFAC decomposition of a third-order tensor. *IEEE Transactions on Signal Processing* **57**, 2299–2310 (2009).
54. Nguyen, V.-D., Abed-Meraim, K. & Linh-Trung, N. Second-order optimization based adaptive PARAFAC decomposition of three-way tensors. *Digital Signal Processing* **63**, 100–111 (2017).
55. Devijver, P. A. & Kittler, J. *Pattern recognition: A statistical approach*. (Prentice hall, 1982).
56. Kohavi, R. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* 1137–1143 (Morgan Kaufmann Publishers Inc., Montreal, Quebec, Canada, 1995).
57. Kolda, T. G. & Bader, B. W. Tensor decompositions and applications. *SIAM review* **51**, 455–500 (2009).

58. Wang, X., Kruger, U. & Lennox, B. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Engineering Practice* **11**, 613–632 (2003).

59. Vijaysai, P., Gudi, R. & Lakshminarayanan, S. Identification on demand using a blockwise recursive partial least-squares technique. *Industrial & engineering chemistry research* **42**, 540–554 (2003).

60. Li, C., Ye, H., Wang, G. & Zhang, J. A recursive nonlinear PLS algorithm for adaptive nonlinear process modeling. *Chemical engineering & technology* **28**, 141–152 (2005).

61. Nagasaka, Y., Shimoda, K. & Fujii, N. Multidimensional recording (MDR) and data sharing: an ecological open research and educational platform for neuroscience. *PloS one* **6**, e22561, https://doi.org/10.1371/journal.pone.0022561 (2011).

62. van Gerven, M. A., Chao, Z. C. & Heskes, T. On the d*ng* **9**ecoding of intracranial data using sparse orthonormalized partial least squares. *Journal of neural engineeri*, 026017, https://doi.org/10.1088/1741-2560/9/2/026017 (2012).

63. Eliseyev, A. & Aksenova, T. Penalized Multi-Way Partial Least Squares for Smooth Trajectory Decoding from Electrocorticographic (ECoG) Recording. *PloS one* **11**, e0154878 (2016).

64. Schlogl, A., Kronegg, J., Huggins, J. E. & Mason, S. G. 19 Evaluation Criteria for BCIResearch. *Toward brain-computer interfacing* (2007).

65. Cook, R. D. & Weisberg, S. Residuals and influence in regression. (1982).

66. Marathe, A. R. & Taylor, D. M. The impact of command signal power distribution, processing delays, and speed scaling on neurally-controlled devices. *Journal of neural engineering* **12**, 046031, https://doi.org/10.1088/1741-2560/12/4/046031 (2015).

67. Miller, K., Zanos, S., Fetz, E., Den Nijs, M. & Ojemann, J. Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. *The Journal of neuroscience* **29**, 3132–3137 (2009).

68. Wang, Z. *et al.* Decoding onset and direction of movements using electrocorticographic (ECoG) signals in humans. *Frontiers in neuroengineering* **5** (2012).

69. Eliseyev, A. *et al.* Iterative N-way partial least squares for a binary self-paced brain-computer interface in freely moving animals. *Journal of neural engineering* **8**, 046012, https://doi.org/10.1088/1741-2560/8/4/046012 (2011).

70. Li, B., Morris, J. & Martin, E. B. Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **64**, 79–89 (2002).

71. Woolsey, C. N., Erickson, T. C. & Gilson, W. E. Localization in somatic sensory and motor areas of human cerebral cortex as determined by direct recording of evoked potentials and electrical stimulation. *Journal of neurosurgery* **51**, 476–506 (1979).

72. Mestais, C. *et al.* WIMAGINE®: Wireless 64-channel ECoG recording implant for long term clinical applications. *IEEE TNSRE* **23** (2015).

73. Morinière, B., Verney, A., Abroug, N., Garrec, P. & Perrot, Y. In *Intelligent Robots and Systems* (*IROS*), *2015 IEEE/RSJ International Conference on* 5333–5338 (2015).

Acknowledgements

Author Contributions

A.L.B. conceived the project. A.E., G.C., C.M., T.A., and A.L.B. conceived the task. V.A., T.C., and L.L. conducted the experiments. A.E. developed and implemented the methods and scripts. All authors were involved in writing and reviewing the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-16579-9.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.