

Distance-Based Partial Least Squares Analysis

Anjali Krishnan, Nikolaus Kriegeskorte, and Hervé Abdi

In
Abdi, H., Chin, W., Esposito Vinzi,
V., Russolillo, G., & Trinchera, L. (Eds.), 2013,
New Perspectives in Partial Least Squares
and Related Methods. New York: Springer Verlag

Abstract Distances matrices are traditionally analyzed with statistical methods that represent distances as maps such as Metric Multidimensional Scaling (MDS), Generalized Procrustes Analysis (GPA), Individual Differences Scaling (INDSCAL), and DISTATIS. MDS analyzes only one distance matrix at a time while GPA, INDSCAL and DISTATIS extract similarities between several distance matrices. However, none of these methods is predictive. Partial Least Squares Regression (PLSR) predicts one matrix from another, but does not analyze distance matrices. We introduce a new statistical method called Distance-based Partial Least Squares Regression (DISPLSR), which predicts one distance matrix from another. We illustrate DISPLSR with data obtained from a neuroimaging experiment, which explored semantic categorization.

Key words: Partial Least Squares, Regression, Correlation, Distance, MDS, DISTATIS

1 Introduction

Distance matrices are ubiquitous in the social sciences and several multivariate descriptive methods have been developed to analyze them. Specifically, methods such as metric multidimensional scaling (MDS), generalized Procrustes analysis (GPA), individual differences scaling (INDSCAL), and DISTATIS, all display distances as points on a map. MDS transforms a distance matrix into a cross-product matrix (akin to a variance-covariance matrix) in order to compute a suitable coordinate system such that the original distances between elements are represented as accurately as possible as Euclidean distances [1, 2]. GPA analyzes the similarities between more than two distance matrices obtained on the same elements [3]. INDSCAL analyzes multiple distance matrices that each store distances measured in the same observations [4]. DISTATIS analyzes multiple distance matrices simultaneously by computing an optimal compromise between all the distance matrices ([5–7]). While MDS

only analyzes one distance matrix at a time, and GPA, INDSCAL and DISTATIS analyze several distance matrices, none of these methods predicts one distance matrix from another.

Partial Least Squares Regression (PLSR) predicts a set of dependent variables (predictee) from a set of independent variables (predictor). PLSR belongs to the family of PLS methods which also includes Partial Least Squares Correlation (PLSC) (also called inter-battery-analysis [9], PLS-SVD [10, 11], inter-correlation analysis, canonical covariance analysis, [12], robust canonical analysis [13], or co-inertia analysis [14]), a correlation technique that analyzes associations between two matrices. While PLSC has mostly been used in neuroimaging research [8], PLSR has been applied in many fields such as Econometrics and Chemometrics [14, 16, 17]. The basis for PLSC and PLSR is the singular value decomposition (SVD) of a matrix [19, 20, 20]. PLSR requires an iterative application of the SVD, in order to find latent variables that model the independent variables and simultaneously predict the dependent variables. Each iteration of the SVD produces orthogonal (*i.e.*, uncorrelated) latent variables and corresponding regression weights for prediction. PLSR displays the latent variables in the form of maps, which describe the relation between the predictor and predictee.

There are a few methods that are related to PLSR and similar to DISPLSR. Sample-based Partial Least Squares analysis (SAMPLS), developed for comparative molecular field analysis [21], extracts latent variables from inter-sample distances (transformed into covariances) to predict a single response variable. A variation of SAMPLS was later developed [22] with the modification that inter-sample distances were calculated intrinsically (*i.e.*, distances were calculated for each pair of elements separately and then aggregated for the SAMPLS analysis). Distance-based Redundancy Analysis (DB-RDA), developed for ecological research [23], combines MDS and redundancy analysis (related to multiple linear regression with more than one dependent variable) to predict a set of dependent variables from a distance matrix. Multivariate Distance Matrix Regression (MDMR), developed for genomic research [24], assesses the relationship between a distance matrix and non-distance responses. This method was performed without using dimensional analysis and instead defined a modified F statistic similar to the F statistic in ANOVA (*i.e.*, the ratio between explained and unexplained variance). Although SAMPLS, DB-RDA, and MDMR are relevant methods to analyze distance matrices, they do not predict one distance matrix from another.

2 Methodology

In order to reduce the long computation time taken to iteratively derive latent variables in PLSR, a kernel PLSR algorithm was developed to condense large data matrices before the PLSR step [25]. Kernel PLSR computes association matrices (akin to variance-covariance matrices) for the predictor and predictee separately. The product of these association matrices, called the *kernel* matrix, is used to compute the

latent variables for the PLSR step. While PLSR does not analyze distance matrices, the structure of the association matrices in Kernel PLSR is comparable to the cross-product matrix generated by MDS for distances. Therefore, the properties of both Kernel PLSR and MDS were adapted to develop Distance-based Partial Least Squares Regression (DISPLSR) and Distance-based Partial Least Squares Correlation (DISPLSC). Both versions of DISPLS are discussed here with mathematical details and illustrations.

2.1 Distance-based partial least squares regression

The main algorithm for DISPLSR is derived from Kernel PLSR. All vectors in the Kernel PLSR algorithm can be computed with the eigenvalue decomposition of *kernel* matrices [26], which in turn are computed as the product of two association matrices. For Kernel PLSR, the association matrix for an $I \times J$ matrix \mathbf{X} is computed as $\mathbf{X}\mathbf{X}^T$ (i.e., $\mathbf{S}_\mathbf{X}$), and the association matrix for $I \times K$ matrix \mathbf{Y} is computed as $\mathbf{Y}\mathbf{Y}^T$ (i.e., $\mathbf{S}_\mathbf{Y}$). For DISPLSR, the data are in the form of distances: an $I \times I$ predictor distance matrix $\mathbf{D}_\mathbf{X}$ and an $I \times I$ predicted distance matrix $\mathbf{D}_\mathbf{Y}$. These distance matrices are converted into cross-product matrices $\mathbf{S}_\mathbf{X}$ and $\mathbf{S}_\mathbf{Y}$. In order to convert a distance matrix \mathbf{d} into a cross-product matrix we first define a mass vector, whose elements are all positive and whose sum is equal to 1. When the masses for the rows are equal, the value of each element is $\frac{1}{I}$. The masses are stored in a vector \mathbf{m} so that:

$$\mathbf{m}^T \mathbf{1} = 1 . \quad (1)$$

Then, an $I \times I$ conformable *centering* matrix is defined as:

$$\mathbf{E} = \mathbf{I} - \mathbf{1}\mathbf{m}^T , \quad (2)$$

where \mathbf{I} is a conformable identity matrix. The cross-product matrix is obtained by double-centering the rows and columns of the distance matrix as:

$$\mathbf{S} = -\frac{1}{2}\mathbf{E}\mathbf{D}\mathbf{E}^T . \quad (3)$$

Both $\mathbf{D}_\mathbf{X}$ and $\mathbf{D}_\mathbf{Y}$ are transformed into $\mathbf{S}_\mathbf{X}$ and $\mathbf{S}_\mathbf{Y}$, respectively, which are re-named as $\mathbf{S}_{\mathbf{X}_0}$ and $\mathbf{S}_{\mathbf{Y}_0}$ as the input for the first iteration of the DISPLSR algorithm. The first latent variable for the predictor is determined from the solution of this singular value decomposition problem:

$$(\mathbf{S}_{\mathbf{X}_0}\mathbf{S}_{\mathbf{Y}_0}) \mathbf{t}_1 = \delta_1 \mathbf{t}_1 , \quad (4)$$

where $\mathbf{S}_{\mathbf{X}_0}\mathbf{S}_{\mathbf{Y}_0}$ is the kernel matrix for the first iteration of DISPLSR, and δ_1 and \mathbf{t}_1 are respectively the first singular value and the first right singular vector of $\mathbf{S}_{\mathbf{X}_0}\mathbf{S}_{\mathbf{Y}_0}$. The first latent variable for the predictee is computed as:

$$\mathbf{u}_1 = \mathbf{S}_{Y_0} \mathbf{t}_1 . \quad (5)$$

Because the Kernel PLSR latent variables (*i.e.*, \mathbf{t} and \mathbf{u}) are calculated on the cross-product matrices, their lengths are not comparable with the latent variables that are computed in the original PLSR algorithm. Therefore, latent variables have to be rescaled to the original PLSR algorithm in order to get comparable prediction [25]. First, we scale \mathbf{u}_1 to \mathbf{u}_{temp} as:

$$\mathbf{u}_{\text{temp}} = \frac{\mathbf{u}_1}{\mathbf{t}_1^T \mathbf{u}_1} . \quad (6)$$

Then, we obtain the weights to rescale \mathbf{t}_1 as:

$$\mathbf{t}_{\text{weight}} = \mathbf{u}_{\text{temp}}^T \mathbf{S}_X \mathbf{u}_{\text{temp}} . \quad (7)$$

Finally, we rescale both \mathbf{t}_1 and \mathbf{u}_{temp} as:

$$\mathbf{t}_{\text{scaled}} = \mathbf{t}_1 \sqrt{\mathbf{t}_{\text{weight}}} , \quad (8)$$

and

$$\mathbf{u}_{\text{scaled}} = \mathbf{u}_{\text{temp}} \sqrt{\mathbf{t}_{\text{weight}}} . \quad (9)$$

Once the first latent variables have been computed, the matrices \mathbf{S}_{X_0} and \mathbf{S}_{Y_0} are deflated. Because we have now condensed the original matrices into cross-product matrices, the deflation is done directly on the cross-product matrices by multiplying them with an *updating* matrix \mathbf{G} (*i.e.*, \mathbf{G}_0 for the first iteration) computed as:

$$\mathbf{G}_0 = \mathbf{I} - \mathbf{t}_{\text{scaled}} \mathbf{t}_{\text{scaled}}^T . \quad (10)$$

The cross-product matrices \mathbf{S}_{X_0} and \mathbf{S}_{Y_0} are deflated as:

$$\mathbf{S}_{X_1} = \mathbf{G}_0 \mathbf{S}_{X_0} \mathbf{G}_0^T , \quad (11)$$

and

$$\mathbf{S}_{Y_1} = \mathbf{G}_0 \mathbf{S}_{Y_0} \mathbf{G}_0^T . \quad (12)$$

This process is continued until all the latent variables have been computed and rescaled to the original PLSR algorithm length. The latent variables of the predictor are stored in matrix \mathbf{T} and the latent variables of the predictee are stored in matrix \mathbf{U} . To derive the weights (*i.e.*, \mathbf{W}) and loadings (*i.e.*, \mathbf{C} and \mathbf{P}) for DISPLSR, the original matrices \mathbf{S}_X and \mathbf{S}_Y are projected onto the space of \mathbf{T} and \mathbf{U} . We correct for scaling with the pseudo-inverse of the square root of \mathbf{S}_X and \mathbf{S}_Y (both being square matrices) as:

$$\mathbf{W} = \left(\mathbf{S}_X^{\frac{1}{2}} \right)^+ \mathbf{S}_X \mathbf{U} . \quad (13)$$

$$\mathbf{C} = \left(\mathbf{S}_Y^{\frac{1}{2}} \right)^+ \mathbf{S}_Y \mathbf{T} , \quad (14)$$

$$\mathbf{P} = \left(\mathbf{S}_X \frac{1}{2} \right)^+ \mathbf{S}_X \mathbf{T}, \quad (15)$$

The weights \mathbf{W} and the loadings of \mathbf{S}_Y on \mathbf{T} , (*i.e.*, \mathbf{C}), are normalized so that the sum of squares equals one. The regression weights \mathbf{B}_{PLS} are computed as in Kernel PLSR:

$$\mathbf{B}_{\text{PLS}} = \mathbf{W} \left(\mathbf{P}^T \mathbf{W} \right)^{-1} \mathbf{C}^T. \quad (16)$$

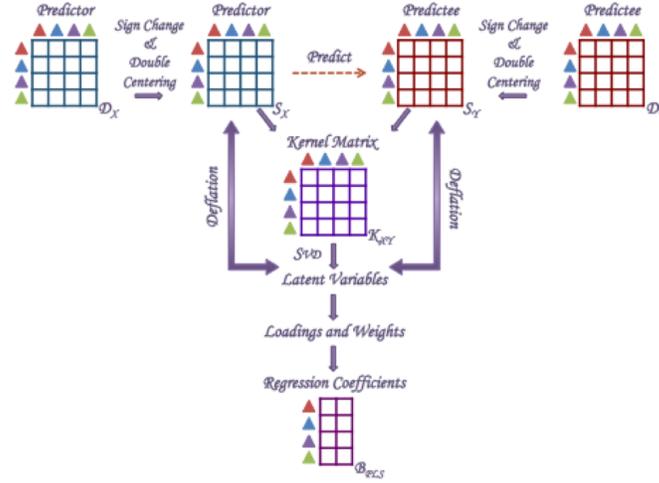


Fig. 1: Steps for DISPLS Regression: Compute cross-product matrices; compute kernel matrix; iteratively compute latent variables; compute weights, loadings and regression coefficients.

Figure 1 shows the steps involved to generate latent variables and regression weights in DISPLSR. Once the latent variables and regression weights have been extracted, the predicted similarity structure is computed as:

$$\widehat{\mathbf{S}}_Y = \frac{1}{2\sqrt{I}} \left(\mathbf{S}_X \mathbf{B}_{\text{PLS}} \mathbf{B}_{\text{PLS}}^T \mathbf{S}_X^T \right), \quad (17)$$

where $\frac{1}{2\sqrt{I}}$ is a scaling factor with I being the number of rows (and columns) of \mathbf{D}_X and \mathbf{D}_Y . The residual similarity structure is obtained by subtracting the predicted similarity structure from the original predictee:

$$\widetilde{\mathbf{S}}_Y = \mathbf{S}_Y - \widehat{\mathbf{S}}_Y. \quad (18)$$

The quality of the DISPLSR model is evaluated by the R_V coefficient, which is similar to a squared coefficient of correlation [27, 28], computed as:

$$R_V = \frac{\text{trace}\{\mathbf{S}_Y^T \widehat{\mathbf{S}}_Y\}}{\sqrt{(\text{trace}\{\mathbf{S}_Y^T \mathbf{S}_Y\}) (\text{trace}\{\widehat{\mathbf{S}}_Y^T \widehat{\mathbf{S}}_Y\})}}. \quad (19)$$

The original predictee is displayed on an MDS map. The regression and residual are projected as supplementary structures on this map to show the additivity of the regression model:

$$\mathbf{F}_{\text{sup}} = \mathbf{S}_{\text{sup}}^T \mathbf{F} \mathbf{\Lambda}^{-1}, \quad (20)$$

where \mathbf{F}_{sup} is the matrix of factor scores for the supplementary similarity structure (*i.e.*, regression or residual), \mathbf{F} is the matrix of factor scores for the predictee, \mathbf{S}_{sup} is the regression or residual similarity structure, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues of the predictee. Figure 2 shows the steps involved to compute and display the regression and residual in DISPLSR.

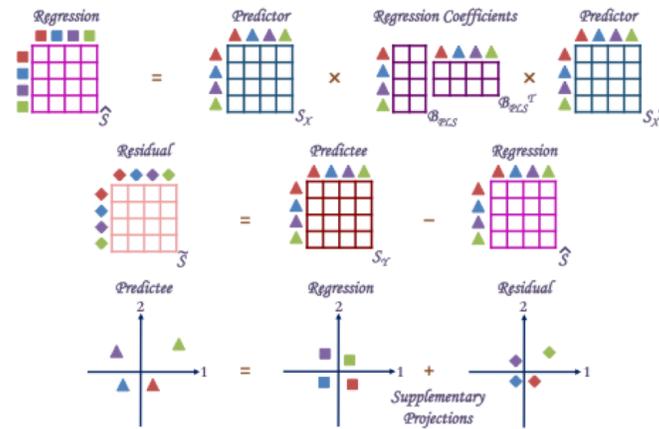


Fig. 2: Steps for DISPLS Regression: Compute regression; compute residual; project regression and residual as supplementary elements onto map of predictee

2.1.1 Cross-validation for DISPLSR

A permutation test on the R_V coefficient can be used for hypothesis testing. In a permutation test, a new data set—called a *permuted* sample—is obtained by randomly reordering the labels of the rows and columns of one distance matrix and leaving the other distance matrix unchanged. The DISPLSR model is then recomputed for the permuted sample to obtain a new R_V coefficient. This procedure is repeated for a large number of permuted samples, say 1,000 or 10,000. The set of all the R_V coefficients provides a sampling distribution of the R_V coefficient under the null hypothesis. The resulting null distribution of the R_V is conditional on the

original distance matrices (see [29] for more details). While the sampling distribution of the R_V coefficient for permuted positive semi-definite matrices has been documented [27, 30], the sampling distribution of the R_V coefficients for permuted *distance* matrices has not been fully explored.

2.2 Distance-based partial least squares correlation

The original application of DISPLSR as a predictive method is ideal only when there exist a clear predictor and a clear predictee. When both the distance matrices are dependent variables, a correlation technique will better capture the similarity between the two matrices. The idea behind Distance-based Partial Least Squares Correlation (DISPLSC; derived from PLSC [8]) is to extract the commonalities between two distance matrices. The difference between the algorithms of DISPLSR and DISPLSC is that for DISPLSC the latent variables are computed in one iteration of the SVD. Also, because there is no prediction step, the weights \mathbf{W} , loadings \mathbf{P} and \mathbf{C} , and regression weights \mathbf{B}_{PLS} are not computed for DISPLSC. Specifically, the kernel matrix K is given by:

$$K = \mathbf{S}_X \mathbf{S}_Y . \quad (21)$$

The SVD of K is given by:

$$K = \mathbf{U} \Phi \mathbf{V}^T , \quad (22)$$

where \mathbf{U} is the matrix of right singular vectors, \mathbf{V} is the matrix of left singular vectors, and Φ is the diagonal matrix of singular values. The latent variables for \mathbf{S}_X (and \mathbf{D}_X) are given by:

$$\mathbf{L}_{DX} = \mathbf{S}_X \mathbf{U} , \quad (23)$$

and describe the relationship between the distances in \mathbf{D}_X with respect to the distances in \mathbf{D}_Y . The latent variables for \mathbf{S}_Y (and \mathbf{D}_Y) are given by:

$$\mathbf{L}_{DY} = \mathbf{S}_Y \mathbf{V} , \quad (24)$$

and describe the relationship between the distances in \mathbf{D}_Y with respect to the distances in \mathbf{D}_X . Figure 3 shows the steps involved in DISPLSC.

Just like the quality of the DISPLSR model The quality of the DISPLSC model is evaluated by the R_V coefficient computed between \mathbf{L}_X and \mathbf{L}_Y (see Equation 19).

2.2.1 Cross-validation for DISPLSC

A permutation test on the R_V coefficient between the latent variables can be used for hypothesis testing. As mentioned earlier, a permuted sample is created by randomly reordering the labels of the rows and columns of one distance matrix, leaving the other distance matrix unchanged. The set of all the R_V coefficients computed from say 1,000 or 10,000 permutations, provides a sampling distribution of the R_V

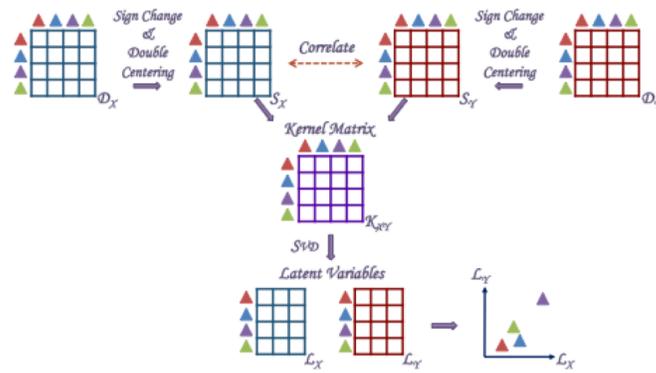


Fig. 3: Steps for DISPLS Correlation: Compute cross-product matrices; compute kernel matrix; compute latent variables.

coefficient under the null hypothesis, which is conditional on the original distance matrices [29].

3 Illustration

We illustrate DISPLSR with data from a neuroimaging experiment, which explored how the brain semantically categorizes objects [31]. Functional Magnetic Resonance Imaging (*fMRI*) measures were obtained from the inferior temporal cortex of four human participants while they viewed pictures of ninety two real-world objects. The similarity between patterns of brain activation elicited for each pair of pictures was measured as a correlation coefficient r (ranging from -1 to $+1$), and the correlation distance between these patterns of brain activation was quantified as $1 - r$. The authors derived a stimulus-by-stimulus correlation distance matrix from the *fMRI* data for each participant and averaged the distance matrices across all four participants to get a mean distance matrix. The authors also obtained an averaged stimulus-by-stimulus correlation distance matrix from single-cell data measured from the inferior temporal cortex of two monkeys [32]. In addition, the authors computed a stimulus-by-stimulus correlation distance matrix from the actual pictures used in the experiment. The pictures were modeled using Gabor filters that generate a computational model of images based on texture and shape, which closely mimics the Gabor model of neuronal function in the primary visual cortex of mammalian brains [33].

An MDS was first performed on each of the distance matrices to display the objects on a map (only the first and second dimensions are displayed). Figure 4a shows the original map of the stimuli as represented by a Gabor model of the pictures. On the whole we see no categorical structure defined by the Gabor model. Figure 4b

shows the original map of the stimuli as measured by *fMRI* from the human participants and Figure 4c shows the original map of the stimuli as measured by single-cell recordings from the monkeys. We see categories such as scenes, human faces, monkeys and other animals in these maps. The single-cell data from the monkeys were more robust than the *fMRI* data from human participants and revealed some of the categories more clearly.

Mds analysis for Gabor model of pictures, human *fMRI* & monkey single-cell data

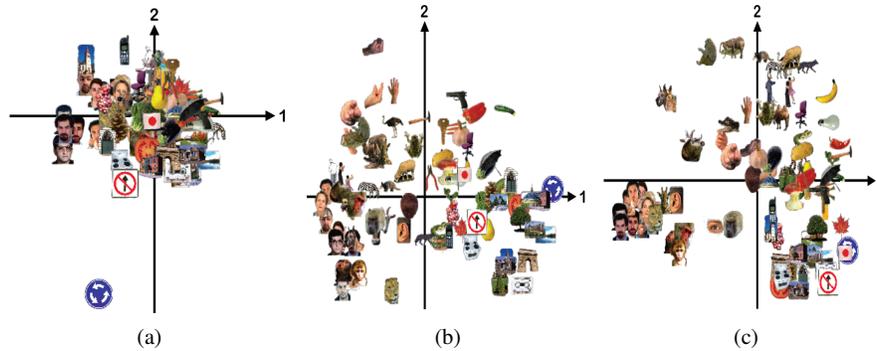


Fig. 4: (a) MDS map for the Gabor model of pictures; (b) MDS map for *fMRI* data from human participants; (c) MDS map for single-cell data from monkeys.

We performed three separate DISPLSR analyses for the data. The first DISPLSR analysis used the distances from the Gabor model of the pictures (Figure 4a) as the predictor and the distances from the *fMRI* data from the human participants (Figure 4b) as the predictee. Figure 5a shows the map of the stimuli represented by the *fMRI* data as *predicted* from the Gabor model of the pictures (which appears to be linear because the predictor and predictee might have only one dimension in common). If we compare this map with the original map of stimuli derived from the *fMRI* data (see Figure 4b), we see that DISPLSR predicts the face category, oblong objects and the roundabout on the first and second dimensions. If we subtract the map of the regression (Figure 5a) from the original map (Figure 4b), we obtain the residual map (Figure 5b), which shows what is unique to brain activation in the inferior temporal cortex (*i.e.*, cannot be predicted from the Gabor model of the pictures) such as semantic categories of faces (both human and monkey), animals, vegetables and scenes.

The R_V coefficient computed between the prediction and the original data from the human participants was equal to .08, this value—even though quite small—was statistically significant at $p < .001$ based on the permutation test. The DISPLSR analysis has modelled the information encoded in the brain as the sum of low-level information predicted by the Gabor model of the pictures and the high-level information unique to the brain.

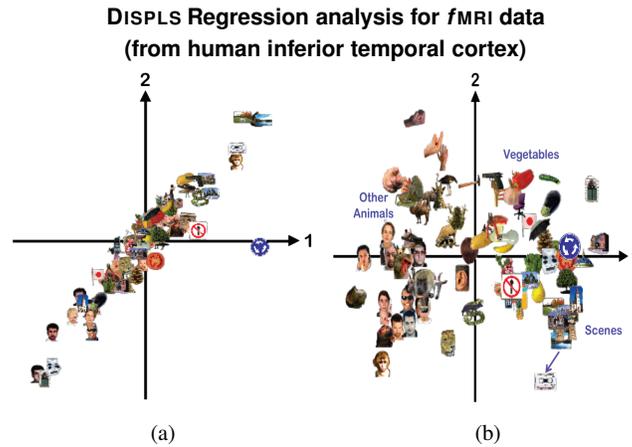


Fig. 5: (a) Regression map: Human *fMRI* data as *predicted* by the Gabor model of the pictures; (b) Residual map: Human *fMRI* data *not* predicted by the Gabor model of the pictures.

The second DISPLSR analysis used the distances from the Gabor model of the pictures (Figure 4a) as the predictor and the distances from single-cell data from the monkeys (Figure 4c) as the predictee. Figure 6a shows the map of the stimuli represented by the single-cell data as *predicted* from the Gabor model of the pictures. If we compare this map with the original map of stimuli derived from the single-cell data (see Figure 4c), we see the separation of oblong or rectangular objects from the circular objects on the first and second dimensions. If we subtract the map of the regression (Figure 6a) from the original map (Figure 4c), we obtain the residual map (Figure 6b), which isolates semantic categories (Figure 4c). Because the single-cell data are more robust, the residual did not isolate much of the categorical structure. The R_V coefficient between the prediction and the original data from the monkeys was not statistically significant based on a permutation test.

Because of the similarity between primate and human vision [34], we can reasonably speculate that some of the basic semantic categories in humans could be traced back to semantic categorization in monkeys. The third DISPLSR analysis used the residual from the DISPLSR analysis for the monkeys (Figure 6b) as the predictor and the residual from the DISPLSR analysis for the human participants (Figure 5b) as the predictee in order to predict the semantic categorical structure in humans from the semantic categorical structure in monkeys (after removing perceptual information modeled by the Gabor filters).

Figure 7a shows the map of the stimuli represented by the *fMRI* data (not predicted by Gabor model) from the humans as *predicted* from the single-cell data (not predicted by Gabor model) from the monkeys. The first dimension separates the basic categories of animate and inanimate objects, and the second dimension separates the non-human and human related categories. This could imply that these

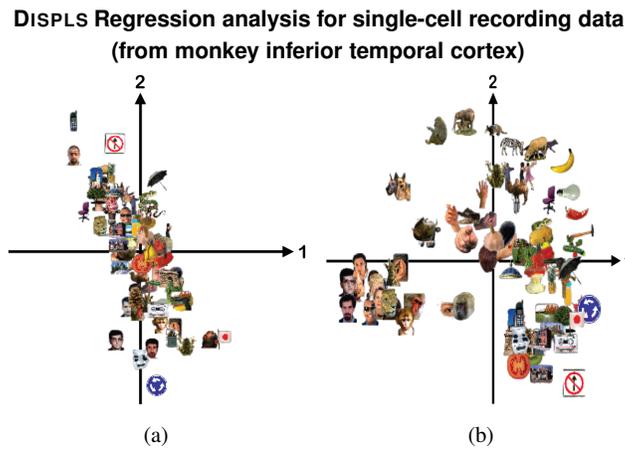


Fig. 6: (a) Regression map: Monkey single-cell data as *predicted* by the Gabor model of the pictures; (b) Residual map: Monkey single-cell data *not* predicted by the Gabor model of the pictures.

**DISPLS Regression analysis for human *fMRI* and monkey single-cell data
(not predicted by Gabor model)**

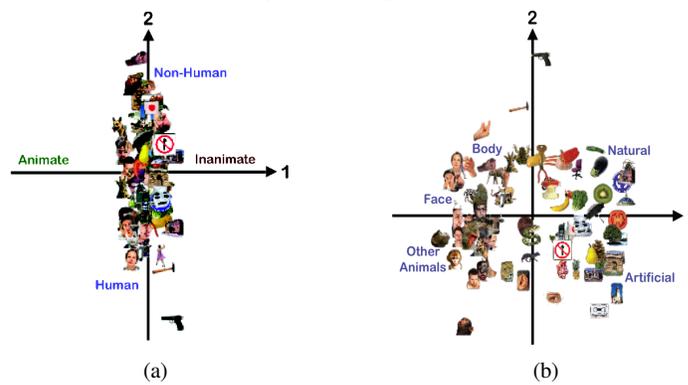


Fig. 7: (a) Regression map: Non-Gabor *fMRI* data from the human participants as *predicted* by the non-Gabor single-cell data from the monkeys; (b) Residual map: Non-Gabor *fMRI* data from the human participants *not* predicted by by non-Gabor single-cell data from the monkeys.

monkeys (who were accustomed to a human environment) and human participants share information about these categories. If we subtract the map of the regression (Figure 7b) from the original map (Figure 5b), we obtain the residual map (Figure 7b), which isolates semantic categories unique to human participants such as the natural and artificial objects. The R_V coefficient between the prediction and the residual data from the DISPLSR analysis for the human participants was not statistically significant based on a permutation test.

Lastly, we performed a DISPLS Correlation analysis to determine the commonalities between the data from the human participants and the monkeys. Figure 8a shows the first dimension of the latent similarity structure for both the human participants and the monkeys. We see that the first dimension separates the animate and inanimate objects. Figure 8b shows the second dimension of the latent similarity structure for both the human participants and the monkeys.

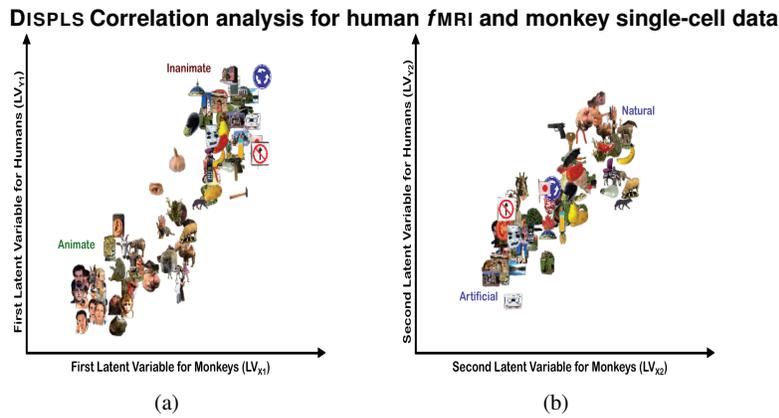


Fig. 8: (a) First dimension of latent similarity structures for monkeys and human participants; (b) Second dimension of latent similarity structures for monkeys and human participants.

We see that the second dimension separates the natural and artificial objects. It is important to note that each of these maps only depict one dimension and therefore objects that are better represented on other orthogonal dimensions might appear to be mis-categorized. The R_V coefficient of .75 between the latent variables for the human participants and the latent variables for the monkeys was found to be statistically significant at $p < .001$ based on a permutation test.

4 Discussion

DISPLSR is a regression method, and so the residual will contain the unexplained part of experimental variance, which may confound the results of a DISPLSR analysis. DISPLSR (in combination with DISTATIS) can also be used to describe the directional dependencies of three or more sets of variables for path modeling or multi-block analyses. The reliability of the DISPLSR maps can be tested with the bootstrap method to generate confidence intervals for the predictee, predictor and residual [35]. A limitation of DISPLSR is that bootstrap maps are only generated when the original data, which were used to derive the distances between categories of observations, are available. Although other techniques exist that find commonalities between multiple distance matrices (*e.g.*, GPA, INDSCAL, DISTATIS), it is worth exploring the potential of DISPLS Correlation (DISPLSC) to capture the information from one distance matrix in relation to the information from another distance matrix. More research is required to investigate the permutation of distance matrices in order to obtain a random sampling distribution, although a permutation test with the R_V coefficient for both DISPLSR and DISPLSC is currently possible. The prediction-based approach for distance matrices offered by Distance-based Partial Least Squares Regression (DISPLSR), and the association-based approach for distance matrices offered by Distance-based Partial Least Squares Correlation (DISPLSC) provide essential links between various domains of research.

References

1. H. Abdi, "Metric multidimensional scaling," in *Encyclopedia of Measurement and Statistics*, N. J. Salkind, ed., pp. 598–605, Thousand Oaks (CA): Sage, 2007.
2. W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika* **17**, pp. 401–419, 1952.
3. J. Gower and G. Dijksterhuis, *Procrustes Problems*, New York: Oxford University Press, 2004.
4. J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition," *Psychometrika* **35**, pp. 283–319, 1970.
5. H. Abdi, D. Valentin, A. J. O'Toole, and B. Edelman, "DISTATIS: The analysis of multiple distance matrices," in *Proceedings of the IEEE Computer Society: International Conference on Computer Vision and Pattern Recognition*, pp. 43–47, 2005.
6. H. Abdi, J.P. Dunlop, and L.A. Williams, "How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling DISTATIS," in *NeuroImage* **17**, pp. 89–95, 2009.
7. H. Abdi, D. Valentin, S. Chollet, and C. Chrea, "Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications," in *Food Quality and Preference*, **18**, pp. 627–640, 2007.
8. A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial least squares (PLS) methods for neuroimaging: A tutorial and review," *NeuroImage* **56**, pp. 455–475, 2011.
9. L.R., Tucker, "An inter-battery method of factor analysis," *Psychometrika* **23**, pp. 111–136, 1958.

10. F.L. Bookstein, P.L. Sampson, A.P. Streissguth, and H.M. Barr, "Exploring redundant measurements of dose and developmental outcome: New methods from the behavioral teratology of alcohol," *Developmental Psychology* **32**, pp. 404–415, 1996.
11. P.D. Sampson, A.P. Streissguth, H.M. Barr, and F.S. Bookstein, "Neurobehavioral effect of prenatal alcohol: Part II, partial least square analysis," *Neurotoxicology and Teratology* **11**, pp. 477–491.
12. A. Tishler, D. Dvir, A. Shenhar, and S. Lipovetsky, "Identifying critical success factors in defense development projects: A multivariate analysis," *Technological Forecasting and Social Change* **51**, pp. 151–171, 1996.
13. A. Tishler, and S. Lipovetsky, "Modelling and forecasting with robust canonical analysis: method and application," *Computers and Operations Research* **27**, pp. 217–232, 2000.
14. S. Dolédec, and D. Chessel, "Co-inertia analysis: an alternative methods for studying species-environment relationships," *Freshwater Biology* **31**, pp. 277–294.
15. H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *WIREs Computational Statistics* **2**, pp. 97–106, 2010.
16. H. Wold, "Soft modelling: The basic design and some extensions," in *Systems under indirect observation: Causality-structure-prediction Part II*, K. Jöreskog and H. Wold, eds., pp. 1–54, Amsterdam: North-Holland Publishing Company, 1982.
17. S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems* **58**, pp. 109–130, 2001.
18. H. Abdi, "Singular value decomposition (SVD) and generalized singular value decomposition (GSVD)," in *Encyclopedia of Measurement and Statistics*, N. Salkind, ed., pp. 907–912, Thousand Oaks (CA): Sage, 2007.
19. M. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, London, 1984.
20. H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*, New York, Springer, 2011.
21. B. L. Bush and R. B. Nachbar, Jr., "Sample-distance partial least squares: PLS optimized for many variables, with application to COMFA," *Journal of Computer-aided Molecular Design* **7**, pp. 587–619, 1993.
22. Y. C. Martin, C. T. Lin, C. Hetti, and J. DeLazzer, "PLS analysis of distance matrices to detect nonlinear relationships between biological potency and molecular properties," *Journal of Medicinal Chemistry* **38**, pp. 3009–3015, 1995.
23. P. Legendre and M. J. Anderson, "Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments," *Ecological Monographs* **69**, pp. 1–24, 1999.
24. M. A. Zapala and N. J. Schork, "Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables," *Proceedings of the National Academy of Sciences* **103**, pp. 19430–19435, 2006.
25. S. Rännar, F. Lindgren, P. Geladi, and S. Wold, "A PLS kernel algorithm for data sets with many variables and fewer objects. Part I: Theory and algorithm," *Journal of Chemometrics* **8**, pp. 111–125, 1994.
26. A. Höskuldsson, "PLS regression methods," *Journal of Chemometrics* **2**, pp. 211–228, 1988.
27. H. Abdi, "Congruence: Congruence coefficient, R_V coefficient and Mantel coefficient," in *Encyclopedia of Research Design*, N. Salkind, D. D.M., and B. Frey, eds., pp. 222–229, Thousand Oaks (CA): Sage, 2010.
28. H. Abdi, " R_V coefficient and congruence coefficient," in *Encyclopedia of Measurement and Statistics*, N. Salkind, ed., pp. 849–853, Thousand Oaks (CA): Sage, 2007.
29. E. J. Dietz, "Permutation tests for association between two distance matrices," *Systematic Zoology* **32**, pp. 21–26, 1983.
30. J. Josse, J. Pagès, and F. Husson, "Testing the significance of the R_V coefficient," *Computational Statistics & Data Analysis* **53**, pp. 82–91, 2008.
31. N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis connecting the branches of systems neuroscience," *Frontiers in Systems Neuroscience* **2**, p. doi:10.3389/neuro.06.004.2008, 2008.

32. R. Kiani, H. Esteky, K. Mipour, and K. Tanaka, "Object category structure in response patterns of neuronal population in monkey inferior temporal cortex," *Journal of Neurophysiology* **97**, pp. 4296–4309, 2007.
33. J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology* **14**, pp. 21–30, 2004.
34. G. Orban, D. van Essen, and W. Vanduffel, "Comparative mapping of higher visual areas in monkeys and humans," *Trends in Cognitive Sciences* **8**, pp. 315–324, 2004.
35. B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical Science* **1**, pp. 54–77, 1986.