Title. Recent advances on Regularized Generalized Canonical Correlation Analysis. Applications to deciphering the chronic effects of urban pollution exposure.

Summary. In contrast to standard data that is structured by a single 'individuals × variables' data matrix, multi-source data are characterized by multiple and heterogeneous sources of information, interconnected, potentially of high dimension. In addition, each source of information may also have a complex structure (e.g. tensor structure). Multi-source data can be found in multiomics field (e.g metabolomics, microbiome, exposome, etc...) where the main objective is to extract relevant variables within massive amounts of variables spread across different sources, reducing dimensionality, summarizing information in a comprehensible way and displaying it for interpretation purposes. The need to analyze conjointly the data by taking into account their natural structure appears to be essential but requires the development of new statistical techniques. Regularized Generalized Canonical Correlation Analysis (RGCCA) framework [1] has been proposed as a general framework for multi-source data. RGCCA continually evolve to accommodate the increasingly complex structures of the data to be analyzed and this paper presents the latest advances of the RGCCA framework [2, 3]. These latest developments are illustrated on data from an observational study on the chronic effects of urban pollution exposure.

[1] Tenenhaus, A., & Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika*, *76*(2), 257.
[2] Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. (2017). Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. *Psychometrika*, *82*(3), 737-777.
[3] Gloaguen A, Philippe C, Frouin V, Gennari G, Dehaene-Lambertz G, Le Brusquet L, Tenenhaus A, (2020) Multiway generalized canonical correlation analysis, Biostatistics, kxaa010, https://doi.org/10.1093/biostatistics/kxaa010