

Figure 1: R^2 scores \uparrow for different methods on simulated data. Performance is measured across 5 source domains and 1 target domain, with shifts controlled by ξ (0 to maximum). Data are generated 100 times, with 5 sensors and 300 covariance matrices per domain. The generation follows the classical instantaneous mixing model: $\mathbf{x}_i(t) = \mathbf{A}\boldsymbol{\eta}_i(t)$ where \mathbf{x}_i are the observed time-series, $\boldsymbol{\eta}_i$ are the underlying signal of the neural generators and \mathbf{A} is the mixing matrix whose columns are the observed spatial patterns of the neural generators. y follows a log-linear model $y_i = \beta_0 + \sum_{\ell=1}^d \beta_\ell \log(p_{\ell i})$ where $p_{\ell i}$ is the variance of the ℓ -th element of the underlying signal $\boldsymbol{\eta}_i$ as proposed in [34, 45, 46]. (A) A shift is applied on the covariance matrices: $\Sigma \mapsto \mathbf{B}_k^\xi \Sigma \mathbf{B}_k^\xi$ with $\mathbf{B}_k \in \mathbb{S}_d^{++}$ and k denotes the domain number (source or target). (B) A shift is applied on the variances: $p_{\ell i} \mapsto p_{\ell i}^{1+k\xi}$ where k is the domain number in $\llbracket 0, K \rrbracket$. The target domain is randomly selected, with the remaining domains as sources. Thus, the distribution of y is shifted per domain because of the log-linear relationship. (C) Both shifts are applied simultaneously.

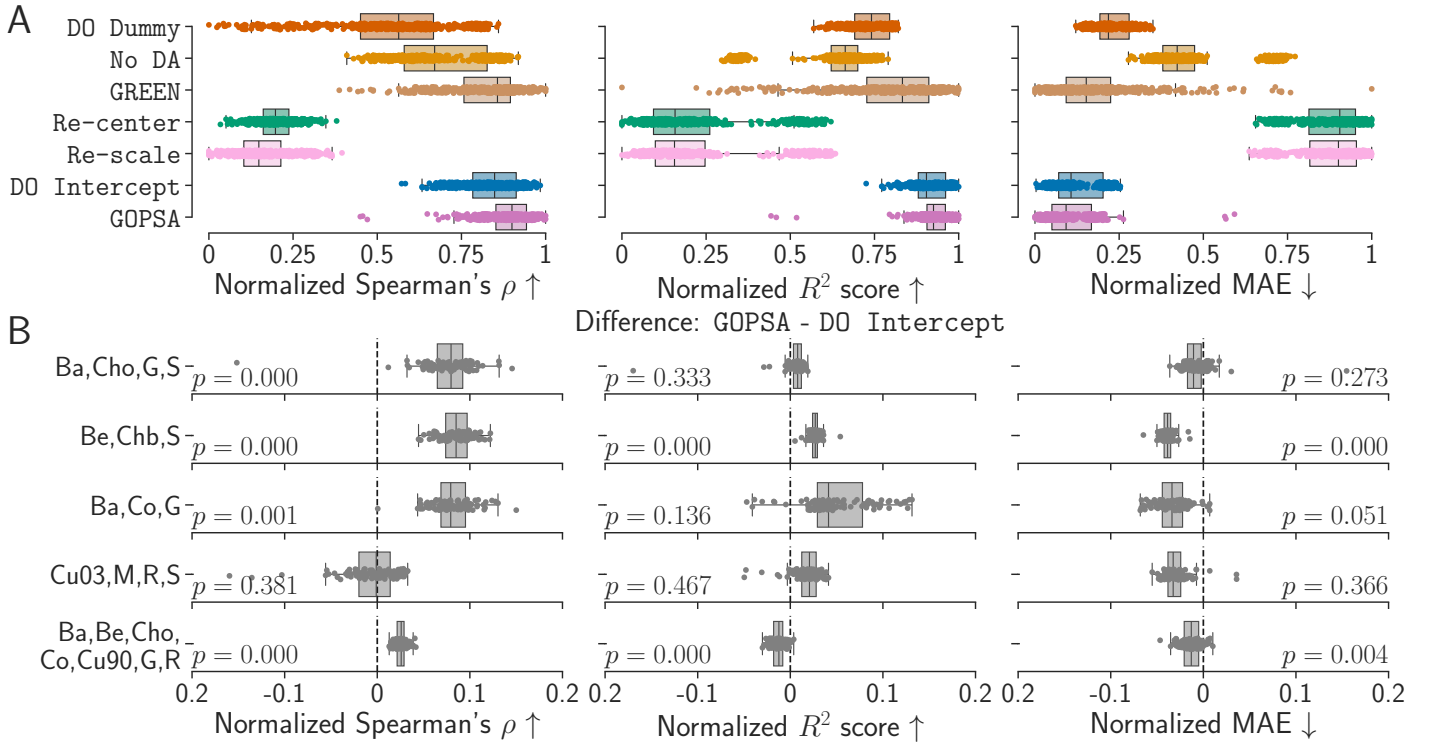


Figure 2: **Normalized performance of the different methods on several source-target combinations for three metrics: Spearman's $\rho \uparrow$ (left), R^2 score \uparrow (middle) and Mean Absolute Error \downarrow (right).** As a large variability in the score values was present between the site combinations, we applied a min-max normalization per combination to set the minimum score across all methods to 0 and the maximum score to 1. (A) Boxplot of the concatenated results for the three normalized scores. One point corresponds to one split of one site combination. (B) Boxplots of the difference between the normalized scores of GOPSA and DO Intercept. A row corresponds to one site combination, one point corresponds to one split. For each plot, the associated results of Nadeau's & Bengio's corrected t-test [35] are displayed. A p-value lower than 0.05 indicates a significant difference between the two methods. Ba: Barbados, Be: Bern, Chb: CHBMP (Cuba), Co: Columbia, Cho: Chongqing, Cu03: Cuba2003, Cu90: Cuba90, G: Germany, M: Malaysia, R: Russia, S: Switzerland