Reviewer 1: 1) Relation to regression strategies within person: In BCI applications, samples are epochs from a subject. The focus is on linear signal changes (evoked response) or the powers (induced response). Our method could be applied in that context, with ‘subjects’ replaced by ‘epochs’ and a constant mixing matrix across subjects (leadfield from individual anatomy). In fact, the Riemannian methods were historically developed there. We will include a sentence in the paper.

2) Include block diagram: This a good idea. We will include one summarizing the different computations.

3 & 5) Missing comparisons: Extensions to other pipelines are interesting and we have computed them (See figure on the right). Y-axis depicts spatial filtering. “Identity” denotes its absence (using ‘Ledoit-Wolf’-regularized covariance for geometric) and “random” for arbitrary spatial filters. We will detail a few points: i) Using supervised spatial filtering (aka SPoC) is critical to the log. diag. method but has little effect on the geometric method (expected due to the affine invariance). ii) The Riemannian method using Wasserstein distance does not need spatial filtering as it directly applies to rank-deficient matrices.

4) significance levels: Rejecting a null-hypothesis that differences are due to chance would require several independent datasets. Instead, we computed uncertainty estimates of paired differences using 100 Monte Carlo splits, 10% test set size. The rate of improvements in data samplings over the id. log. diag. baseline (in Fig. 3) is quantified by the percentage of splits with lower errors: sup. log. diag. 73%, id. Wass. 85%, unsup. geom. 96%, biophys. 95%. This suggests that improvements over the baseline will carry over to new data. The results will be included in the text.

6) discussion on limitations of the work: Primarily, our method falls behind the gold-standard based on biophysics (See Fig. 3). Another limitation is the low interpretability of the Riemannian methods, whereas inspection is readily available for other methods (See Fig. 4). We will clarify these 2 points in the discussion.

Reviewer 2: More evaluations of geometric distance on Cam-CAN: please see response 3) to Rev. 1.

More evaluations on toy data: To test a suboptimal scenario on toy data, we simulate data where each subject has a mixing matrix deviating from a reference ($A_i = A + E_i$, where entries of $E_i$ are i.i.d. $N(0, \sigma^2)$). We also evaluate the interaction between the spatial filters and the projection method (See figure on the right). This shows: i) only when $\sigma = 0$, supervised + log. diag. and Riemann reach perfect performance, ii) else, geometric Riemann is uniformly better, and indifferent to filtering iii) Wasserstein, despite model mismatch, outperforms supervised + log. diag. with high $\sigma$. We will add this figure in the paper.

Authors fail to discuss the basis of the different generative models: We will add a sentence and references showing that log linearity relationships are standard for such brain signals [7 eq. (2)] or [Buzsaki & Mizuseki, Nat Neuro Rev] for evidence of log-linear links between signal power and cognition. Square root linearity is not commonly used for M/EEG modeling. It is included for its formal relationship with Wasserstein metric that can naturally cope with rank reduced data. We found it interesting that this Riemannian metric is consistent with this particular generative model.

Model inspection results unnecessary: We will replace it with block diagram asked by Rev.1 and move Fig. 4 to appendix.

Missing references to recent works: We will include the suggested references and nuance the claims in the discussion.

Sections 2.1 and 2.2 are dense: We will do our best to reword those parts. In particular we will complement them with a more intuitive introduction to manifolds that we developed over the past months when interacting with practitioners.

No code was provided: Code for toy simulations was actually provided in the supplementary materials. Code for Cam-CAN methods and experiments will be fully released upon paper’s acceptance.

This paper introduces the first benchmark on MEG-based age prediction for the Cam-CAN dataset Thanks for this encouraging comment. Our work was motivated by open issues in clinical neuroscience research and we are currently working together with medical doctors on translating the proposed method into the clinic.

Reviewer 3: The discussion of pros and cons of the method could be increased: see response 6) to Rev.1.

A leave one-subject out cross-validation would have made a more realistic assessment of the performance Results from neuroscience suggest that leave-one-out CV (LOO) may yield too optimistic scores (Varoquaux et al. 2016, NIMG, Woo et al. Nat Neuro Rev). Leave-one-subject-out should be helpful with multiple samples from each subject. Applying LOO, we find lower errors but higher variance: id. + log. diag.: 9.11 +/- 7.00, sup. + log. diag.: 8.65 +/- 6.72, id. + Wass.: 8.29 +/- 6.63, unsup. geom.: 8.02 +/- 6.02. We will include the results in the appendix.