Supplements to "Accelerating Permutation Testing in Voxel-wise Analysis through Subspace Tracking: A new plugin for SnPM"

Felipe Gutierrez-Barragan[a], Vamsi K. Ithapu[a], Chris Hinrichs[a], Camille Maumet[e], Sterling C. Johnson[d,c], Thomas E. Nichols[e], Vikas Singh[b,a], and the Alzheimer's Disease Neuroimaging Initiative [1]

[a] *Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA.*
[b] *Department of Biostatistics & Med. Informatics, University of Wisconsin-Madison, Madison, WI, USA.*
[c] *Department of Medicine, University of Wisconsin-Madison, Madison, WI, USA.*
[d] *William S. Middleton Veteran's Affairs Hospital, Madison, WI, USA.*
[e] *Department of Statistics, The University of Warwick, UK.*
http://felipegb94.github.io/RapidPT/

Corresponding Author(s): Felipe Gutierrez-Barragan (fgutierrez3@wisc.edu)

1. Expanded Section 4.4 from the main paper

Here we provide some analysis of the proposed algorithm. We give two results which show that as long as the variance of the residual is below a certain level, we can recover the distribution of the sample maximum. Recall from section 3 of the main paper, that for low-rank matrix completion methods to be applicable, we must assume that the permutation matrix $\mathbf{T}$ can be decomposed into a low-rank component plus a high-rank residual matrix $\mathbf{S}$: $\mathbf{T} = \mathbf{UW} + \mathbf{S}$. Here, $\mathbf{U}$ is a $v \times r$ orthogonal matrix that spans the $r \ll \min(v, L)$–dimensional column subspace of $\mathbf{T}$, and $\mathbf{W}$ is the corresponding coefficient matrix. Subsuming the shift into the coefficients, we can then treat the residual $\mathbf{S}$ as a random matrix whose entries are I.I.D zero-mean Gaussian with variance σ^2. We arrive at our first result by analyzing how the low-rank portion of $\mathbf{T}$'s singular values spectrum interlaces with the contribution coming from $\mathbf{S}$ by treating $\mathbf{T}$ as a low-rank perturbation of a random matrix. If this low-rank perturbation is sufficient to dominate the eigenvalues of the random matrix, then $\mathbf{T}$ can be recovered with high fidelity at a low sampling rate Balzano et al. (2010); He et al. (2012). Consequently, we can estimate the distribution of the maximum as well — this is shown by our second result.

Since the eigenvalues of $\mathbf{T}\mathbf{T}^T$ are the squared singular values of $\mathbf{T}$, rather than analyzing the singular value spectrum of $\mathbf{T}$ directly, we can analyze the eigenvalues of $\mathbf{T}\mathbf{T}^T$ using a recent result from Benaych-Georges and Nadakuditi (2011). This is important because in order to ensure recovery of $\mathbf{T}$, we require that its singular value spectrum should approximately retain the shape of $\mathbf{UW}$'s. More precisely, we require that for some $0 < \delta < 1$,

$$|\tilde{\phi_i} - \phi_i| < \delta\phi_i \qquad i = 1, \ldots, r; \qquad \tilde{\phi_i} < \delta\phi_r \qquad i = r + 1, \ldots, v \tag{1}$$

where ϕ_i and $\tilde{\phi_i}$ are the singular values of $\mathbf{UW}$ and $\mathbf{T}$ respectively. Recall that in this analysis, $\mathbf{T}$ is considered to be a perturbation of $\mathbf{UW}$. Theorem 1.1 relates the rate at which eigenvalues are perturbed, δ, to the parameterization of $\mathbf{S}$ in terms of σ^2.

The theorem's principal assumption also relates σ^2 inversely with the number of columns of the testing matrix $\mathbf{T}$ which is just the number of permutations L. Note, however, that the process may be split up between several matrices $\mathbf{T}_i$, and the results can then be combined. For purposes of applying this result in practice we may then choose a number of columns L which gives the best bound. Theorem 1.1 also assumes that the number of permutations L is greater than the number of voxels v, which is a difficult regime to explore empirically. Thus,

[1] Data used in preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

our numerical evaluations cover the case where $L < v$, while Theorem 1.1 covers the case where L is larger. From the definition of $\mathbf{T}$, we have,

$$\mathbf{T}\mathbf{T}^T = \mathbf{U}\mathbf{W}\mathbf{W}^T\mathbf{U}^T + \mathbf{S}\mathbf{S}^T + \mathbf{U}\mathbf{W}\mathbf{S}^T + \mathbf{S}\mathbf{W}^T\mathbf{U}^T \tag{2}$$

We first analyze the change in eigenvalue structure of $\mathbf{S}\mathbf{S}^T$ when perturbed by $\mathbf{U}\mathbf{W}\mathbf{W}^T\mathbf{U}^T$ (which has r non-zero eigenvalues). The influence of the cross-terms, $\mathbf{U}\mathbf{W}\mathbf{S}^T$ and $\mathbf{S}\mathbf{W}^T\mathbf{U}^T$, is addressed later. We have the following result,

Theorem 1.1 (Perturbation of eigenvalues). *Denote the r non-zero eigenvalues of $\mathbf{Q} = \mathbf{U}\mathbf{W}\mathbf{W}^T\mathbf{U}^T \in \mathbb{R}^{v\times v}$ by $\lambda_1 \geq \lambda_2 \geq, \ldots, \lambda_r > 0$; and let $\mathbf{S}$ be a $v \times t$ random matrix such that $\mathbf{S}_{i,j} \sim \mathcal{N}(0, \sigma^2)$, with unknown σ^2. As $v, L \to \infty$ such that $\frac{v}{L} \ll 1$, the eigenvalues $\tilde{\lambda_i}$ of the perturbed matrix $\mathbf{Q} + \mathbf{S}\mathbf{S}^T$ will satisfy*

$$|\tilde{\lambda_i} - \lambda_i| < \delta\lambda_i \qquad i = 1, \ldots, r; \qquad \tilde{\lambda_i} < \delta\lambda_r \qquad i = r+1, \ldots, v \tag{$\star$}$$

for some $0 < \delta < 1$, whenever $\sigma^2 < \frac{\delta\lambda_r}{L}$

Proof. The first half of the proof emulates Theorem 2.1 from Benaych-Georges and Nadakuditi (2011). Consider the matrix $\mathbf{X} = \sqrt{t}\mathbf{S}$. By the structure of $\mathbf{S}$, each entry of $\mathbf{X}$ is i.i.d.Gaussian with zero–mean and variance $\sigma^2 t$. Let $\mathbf{Y} = \frac{1}{t}\mathbf{X}\mathbf{X}^T$ and denote its ordered eigenvalues as $\gamma_i, i = 1, \ldots, v$ (large to small). Consider the random spectral measure

$$\mu_v(A) = \tfrac{1}{v}\#\{\gamma_i \in A\}, \qquad A \subset \mathbb{R}$$

The Marchenko–Pastur law Marčenko and Pastur (1967) states that as $v, t \to \infty$ such that $\frac{v}{t} \leq 1$, the random measure $\mu_v \to \mu$, where $d\mu$ is given by

$$d\mu(a) = \tfrac{1}{2\pi\sigma^2 t\gamma a}\sqrt{(\gamma_+ - a)(a - \gamma_-)}\mathbf{1}_{[\gamma_-,\gamma_+]}da$$

where $\gamma = \frac{v}{t}$. Here $\mathbf{1}_{[\gamma_-,\gamma_+]}$ is an indicator function that is non–zero on $[\gamma_-, \gamma_+]$. $\gamma_\pm = \sigma^2 t(1 \pm \sqrt{\gamma})^2$ are the extreme points of the support of μ. It is well known that the extreme eigenvalues converge almost surely to $\gamma_\pm$ Edelman (1988). Since $v, t \to \infty$ and $\gamma = \frac{v}{t} \ll 1$, the length of $[\gamma_-, \gamma_+]$ is much smaller than the values in it. Hence we have,

$$\gamma_\pm \sim \sigma^2 t(1 \pm 2\sqrt{\gamma}) \quad ; \quad \sqrt{(\gamma_+ - a)(a - \gamma_-)} \ll a$$

and the new $d\mu(a)$ is given by

$$\begin{aligned} d\mu(a) &= \frac{\sqrt{(\sigma^2 t(1+2\sqrt{\gamma}) - a)(a - \sigma^2 t(1-2\sqrt{\gamma}))}}{2\pi\gamma\sigma^4 t^2}\mathbf{1}_{[\sigma^2 t(1-2\sqrt{\gamma}),\sigma^2 t(1+2\sqrt{\gamma})]}da \\ &= \frac{1}{2\pi\gamma\sigma^4 t^2}\sqrt{4\gamma\sigma^4 t^2 - (a - \sigma^2 t)^2}\mathbf{1}_{[\sigma^2 t(1-2\sqrt{\gamma}),\sigma^2 t(1+2\sqrt{\gamma})]}da \end{aligned}$$

The form we have derived for $d\mu(a)$ shares some similarities with $d\mu_X(x)$ in Section 3.1 of Benaych-Georges and Nadakuditi (2011). The analysis in Benaych-Georges and Nadakuditi (2011) takes into account the phase transition of extreme eigen values. This is done by imitating a time–frequency type analysis on compact support of extreme spectral measure i.e. using Cauchy transform. For our case, the Cauchy transform of $\mu(a)$ is

$$G_\mu(z) = \frac{1}{2\gamma\sigma^4 t^2}\left(z - \sigma^2 t - sgn(z)\sqrt{(z - \sigma^2 t)^2 - 4\gamma\sigma^4 t^2}\right)$$
$$\text{for} \quad z \in (\infty, \sigma^2 t(1 - 2\sqrt{\gamma})) \cup (\sigma^2 t(1 + 2\sqrt{\gamma}), \infty)$$

Since we are interested in the asymptotic eigen values (and $\gamma \ll 1$), $G_\mu(\gamma_\pm)$ and the functional inverse $G_\mu^{-1}(\theta)$ are

$$G_\mu(\gamma_+) = \tfrac{1}{\sigma^2 t\sqrt{(\gamma)}} \;\; ; \;\; G_\mu(\gamma_-) = -\tfrac{1}{\sigma^2 t\sqrt{(\gamma)}} \;\; ; \;\; G_\mu^{-1}(\theta) = \sigma^2 t + \tfrac{1}{\theta} + \gamma\sigma^4 t^2\theta$$

Hence, the asymptotic behavior of the eigen values of perturbed matrix $\mathbf{Q}+\mathbf{S}\mathbf{S}^T$ is (observing that $\mathbf{S}\mathbf{S}^T=\mathbf{Y}$ and $\mathbf{Q}$ has r non–zero positive eigen values)

$$\begin{aligned} \tilde{\lambda_i}(i=1,\ldots,r) &\approx \begin{cases} \lambda_i+\sigma^2 t+\frac{\gamma\sigma^4 t^2}{\lambda_i} & \text{for} \quad \lambda_i>\gamma\sigma^2 t \\ \gamma\sigma^2 t & \text{else} \end{cases} \qquad (*) \\ \tilde{\lambda_i}(i=r+1,\ldots,v) &\approx \sigma^2 t(1-2\sqrt{(\gamma)}) \end{aligned}$$

With $\tilde{\lambda_i}, i=1,\ldots,v$ in hand, we now bound the unknown variance σ^2 such that $(\star)$ is satisfied. We only have two cases to consider,

$$(1) \quad \lambda_i>\gamma\sigma^2 t\ ,\ i=1,\ldots,r \qquad (2) \quad \lambda_i\le\gamma\sigma^2 t\ ,\ i=k,\ldots,r \text{ (for some } k\ge 1)$$

We constrain the unknown σ^2 such that case (2) does not arise. Substituting for $\tilde{\lambda_i}$'s from $(*)$ in $(\star)$, we get,

$$\sigma^2 t+\tfrac{\gamma\sigma^4 t^2}{\lambda_i}<\delta\lambda_i \quad ; \quad \lambda_i>\gamma\sigma^2 t \quad ; \quad \sigma^2 t(1-2\sqrt{(\gamma)})<\delta\lambda_r$$

These inequalities will hold when $\sigma^2<\frac{\delta\lambda_r}{t}$ (since $\gamma=\ll 1$, $\delta<1$ and $\lambda_1\ge\lambda_2\ge,\ldots,\lambda_r$).

□

Note that the missing cross-terms will not change the result of Theorem 1.1 drastically, because $\mathbf{U}\mathbf{W}$ has r non-zero singular values and hence $\mathbf{U}\mathbf{W}\mathbf{S}^T$ is a low-rank projection of a low-variance random matrix, and this will clearly be dominated by either of the other terms. Having justified the model $\mathbf{T}=\mathbf{U}\mathbf{W}+\mathbf{S}$, the following theorem shows that the empirical distribution of the max null statistic approximates the true distribution.

Theorem 1.2. *Let $m_l=\max_i\mathbf{T}_{i,l}$ be the maximum observed test statistic at permutation trial l, and similarly let $\hat{m}_l=\max_i\hat{\mathbf{T}}_{i,l}$ be the maximum reconstructed test statistic. Further, let the maximum reconstruction error be ϵ, such that $|\mathbf{T}_{i,t}-\hat{\mathbf{T}}_{i,t}|\le\epsilon$. Then we have,*

$$Pr\left[m_t-\hat{m}_t-(b-\hat{b})>k\epsilon\right]<\frac{1}{k^2}$$

where b is the bias term described in Section 4.1 *of the main paper, and $\hat{b}$ is its estimate from the training phase.*

Proof. Recall that there is a bias term in estimating the distribution of the maximum which must be corrected for this is because var$(\hat{S})$ underestimates var(S) due to the bias/variance tradeoff. Let b be this difference:

$$b=\mathbb{E}_t\left[\max_i\mathbf{T}_{i,l}\right]-\mathbb{E}_t\left[\max_i\hat{\mathbf{T}}_{i,l}\right].$$

Further, recall that we estimate b by taking the difference of mean sample maxima between observed and reconstructed test statistics over the training set, giving $\hat{b}$, which is an unbiased estimator of b — it is unbiased because a difference in sample means is an unbiased estimator of the difference of two expectations.

Let $\delta_l=m_l-\hat{m}_l$. To show the result we must derive a concentration bound on δ_l, which we will do by applying Chebyshev's inequality. In order to do so, we require an expression for the mean and variance of δ_l. First, we derive an expression for the mean. Taking the expectation over l of $m_l-\hat{m}_l$ we have,

$$\begin{aligned} \mathbb{E}_t\left[m_l-\hat{m}_l\right] &= \mathbb{E}_t\left[\max_i\mathbf{T}_{i,l}-\max_i\hat{\mathbf{T}}_{i,l}-\hat{b}\right] \\ &= \mathbb{E}_t\left[\max_i\mathbf{T}_{i,l}\right]-\mathbb{E}_t\left[\max_i\hat{\mathbf{T}}_{i,l}\right]-\hat{b} \\ &= b-\hat{b} \end{aligned}$$

where the second equality follows from the linearity of expectation.

Next, we require an expression for the variance of δ_l. Let i be the index at which the maximum observed test statistic occurs for permutation trial l, and likewise let j be the index at which the maximum reconstructed test statistic occurs. Thus we have,

$$\begin{aligned} \mathbf{T}_{i,l} &\leq \hat{\mathbf{T}}_{i,l} + \epsilon && \leq \hat{\mathbf{T}}_{j,l} + \epsilon \\ \mathbf{T}_{i,l} &\geq \hat{\mathbf{T}}_{j,l} && \geq \hat{\mathbf{T}}_{j,l} - \epsilon, \end{aligned}$$

and so we have that

$$|m_l - \hat{m}_l| < 2\epsilon$$

and so

$$\text{var}(m_l - \hat{m}_l) \leq \epsilon^2.$$

Applying Chebyshev's bound,

$$\Pr\left[m_l - \hat{m}_l - (b - \hat{b}) > k\epsilon\right] < \frac{1}{k^2}$$

which completes the proof.

□

2. Supplementary Results

The supplementary results for this paper contain a more exhaustive set of plots for each accuracy and runtime performance benchmark discussed in the main document. We first present the accuracy benchmarks: Kullback-Leibler divergence (KL-divergence), p-values and t-statistics, and the resampling risk. Then we present the serial and parallel speedups of RapidPT over SnPM and NaivePT. Finally, we show how RapidPT and SnPM scale as the number of permutations or the dataset size increases. The scaling results are presented for various sets of hyperparameters that led to the low KL-Divergences shown in section 2.1.

2.1. Kullback-Leibler Divergence

Figure 1 contains the colormaps of the KL-Divergence between the max null distributions of RapidPT and SnPM.

2.2. P-Value Spectrums

Figure 2 contains the t-statistic vs. p-value plots obtained from the max null distributions of RapidPT, NaivePT, and SnPM.

2.3. Resampling Risk

Figure 3 contains the resampling risk plots.

2.4. Statistical Significance Maps on the ADNI Datasets

Figures 4 and 5 show the results from SnPM and RapidPT for a given set of hyperparameters. The overlays in the brain images correspond to the null hypotheses (voxels) rejected at an $\alpha = 0.05$. The colormap of the overlays represents the value of the test statistics. The tables in the figures are a summary of the brain regions whose null hypothesis was rejected and the associated statistics.

2.5. Serial Speedups over SnPM

Figure 6 contains the colormaps of the speedup of RapidPT over SnPM running on a serial environment (1 core).

2.6. Parallel Speedups over SnPM

Figure 7 contains the colormaps of the speedup of RapidPT over SnPM running on a parallel environment (16 cores).

2.7. Parallel Speedups over NaivePT

Figure 8 contains the colormaps of the speedup of RapidPT over NaivePT running on a parallel environment (16 cores).

2.8. Permutation Scaling

Figure 9 and 10 show how RapidPT and SnPM scale as the number of permutations increases.

2.9. Dataset Scaling

Figure 11 and 12 show how RapidPT and SnPM scale as the number of permutations increases.

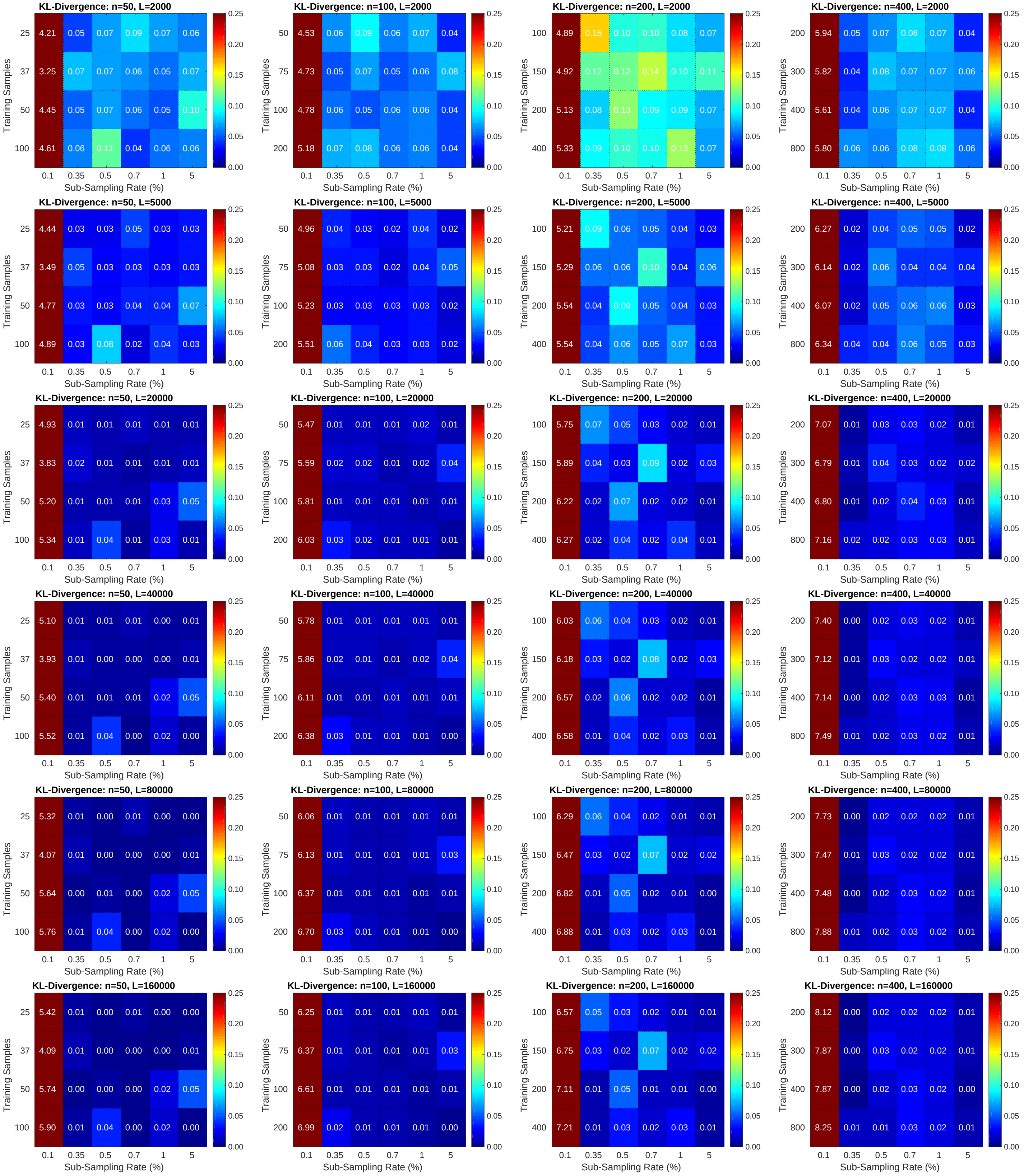


Figure 1: Colormap of the KL-Divergence between the max null distributions of RapidPT and SnPM.Each colormap is associated to a run on one of the datasets and a fixed number of permutations. The resulting KL-Divergence from 24 hyperparameter combinations is displayed on each colormap. Columns 1, 2, 3, and 4 of this figure are associated to the 50, 100, 200, and 400 subject datasets, respectively. The colormaps for $L = 10,000$ are omitted because they are shown in the main document. Refer to the results section in the main paper for a visual illustration of what a low KL-Divergence between two distributions look like.

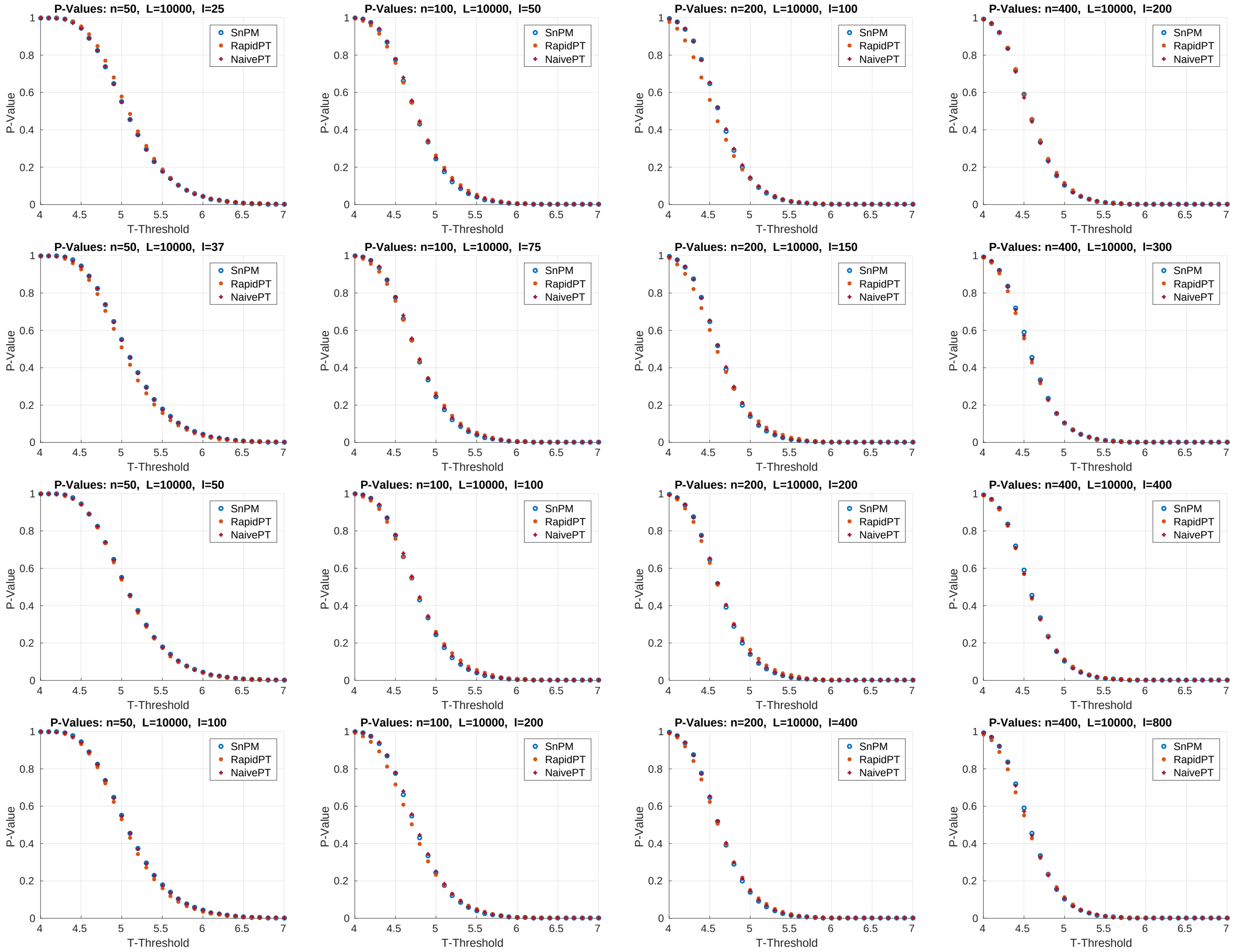


Figure 2: p-values for SnPM, RapidPT, and NaivePT. Columns 1, 2, 3, and 4 of this figure are associated to the 50, 100, 200, and 400 subject datasets, respectively. Rows 1, 2, 3, and 4 are associated to runs using $l = \frac{n}{2}$, $l = \frac{3n}{4}$, $l = n$, $l = 2n$, respectively. The other hyperparameters were fixed to: $\eta = 0.35\%$ and $L = 10000$.

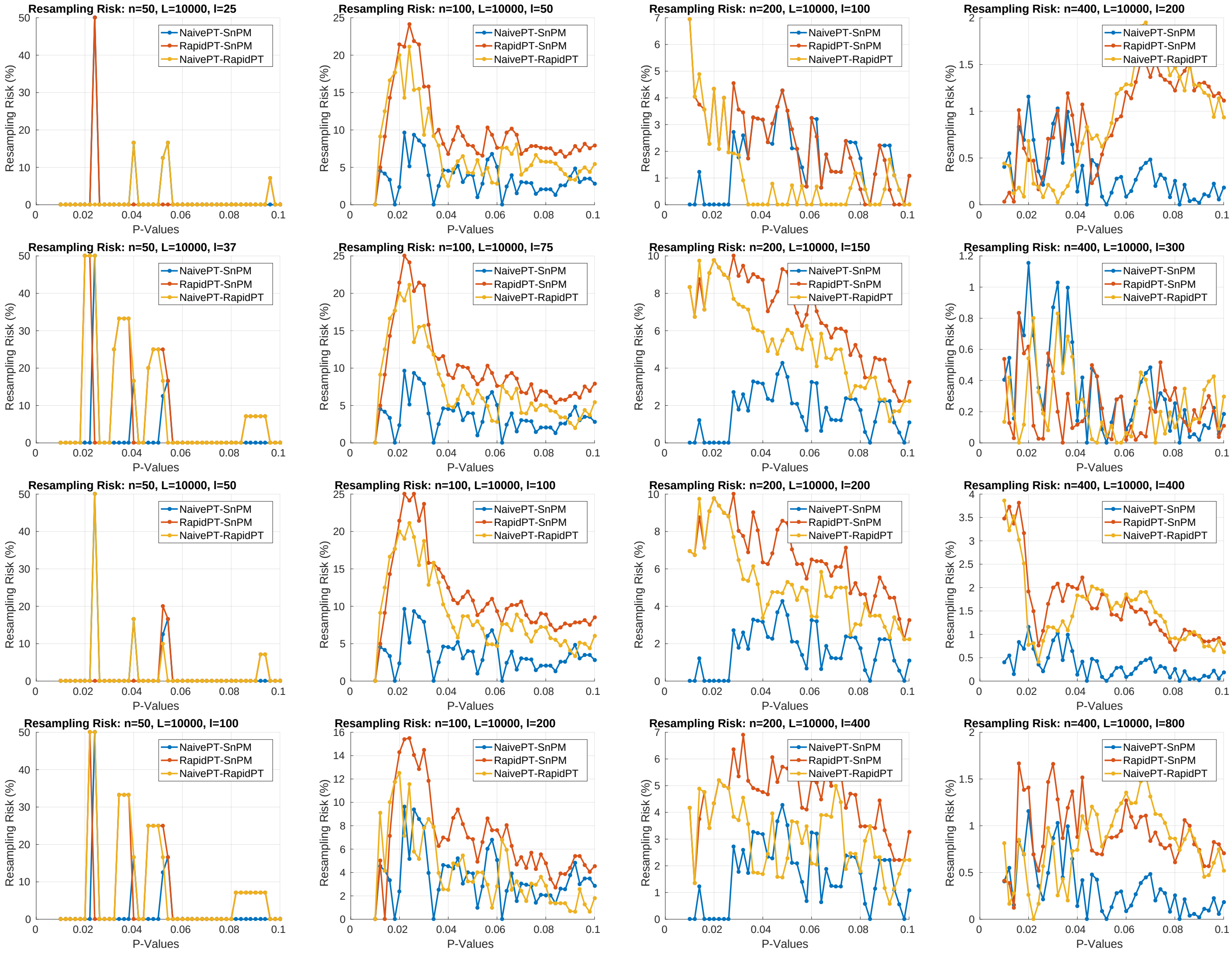
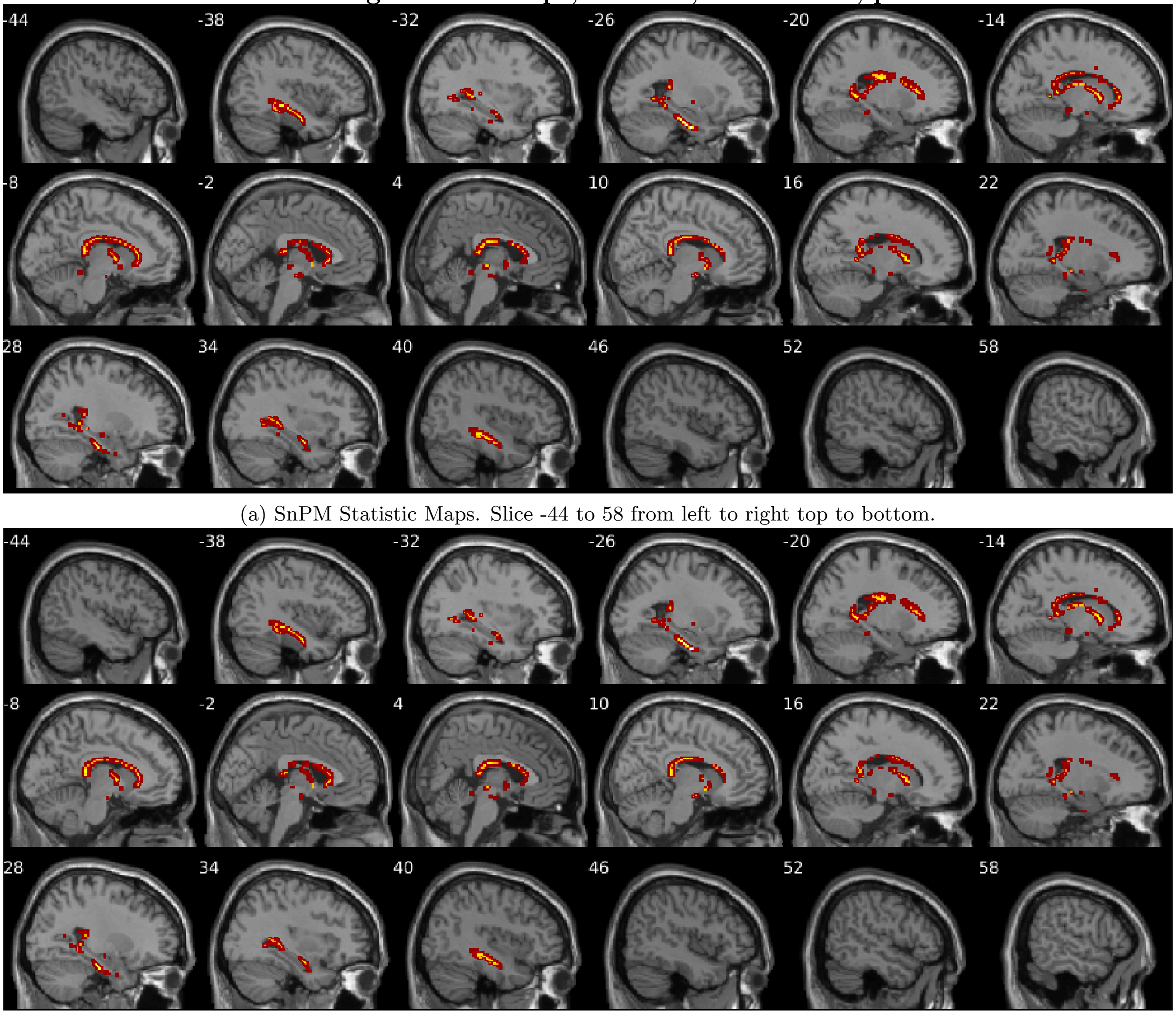


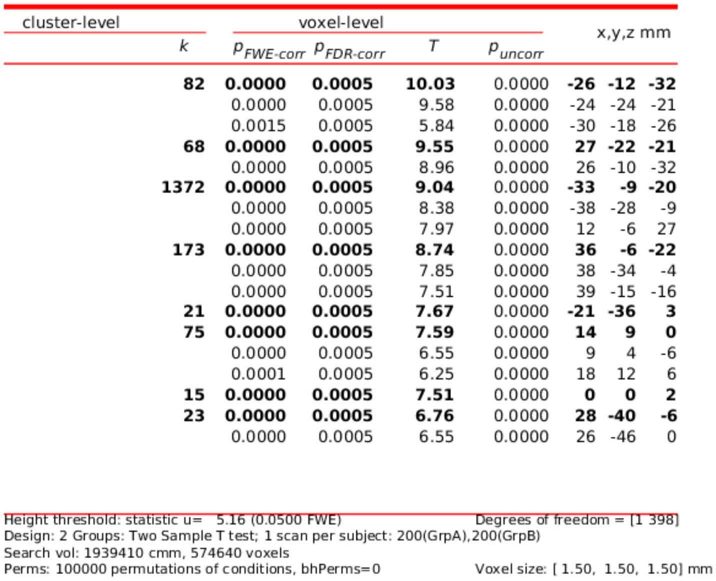
Figure 3: Resampling risk comparison between SnPM, RapidPT, and NaivePT. Columns 1, 2, 3, and 4 of this figure are associated to the 50, 100, 200, and 400 subject datasets, respectively. Rows 1, 2, 3, and 4 are associated to runs using $l = \frac{n}{2}$, $l = \frac{3n}{4}$, $l = n$, $l = 2n$, respectively. The fixed hyperparameters were: $\eta = 0.35\%$ and $L = 10000$.

ADNI Statistical Significance Maps, n = 400, L = 100000, p-value = 0.05



(a) SnPM Statistic Maps. Slice -44 to 58 from left to right top to bottom.

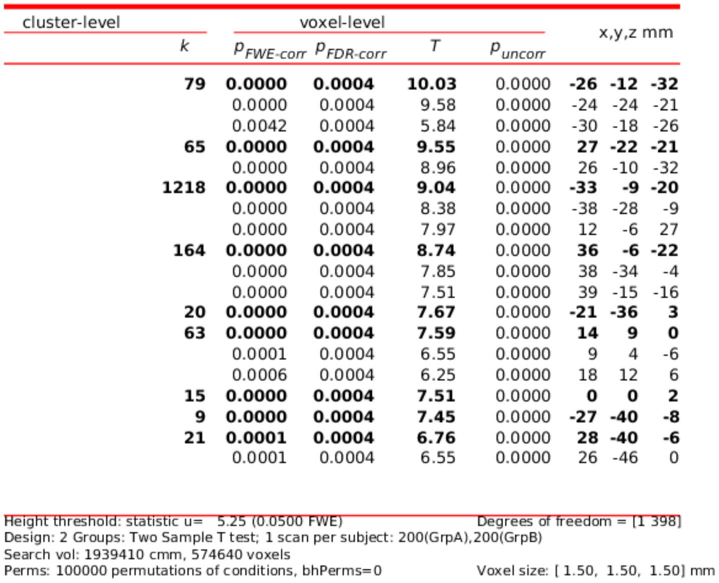
(b) RapidPT Statistic Maps. Slice -44 to 58 from left to right top to bottom. $\eta = 0.5\%$,$l = 200$.



cluster-level	voxel-level				
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	T	p_{uncorr}	x,y,z mm
82	**0.0000**	**0.0005**	**10.03**	0.0000	**-26 -12 -32**
	0.0000	0.0005	9.58	0.0000	-24 -24 -21
	0.0015	0.0005	5.84	0.0000	-30 -18 -26
68	**0.0000**	**0.0005**	**9.55**	0.0000	**27 -22 -21**
	0.0000	0.0005	8.96	0.0000	26 -10 -32
1372	**0.0000**	**0.0005**	**9.04**	0.0000	**-33 -9 -20**
	0.0000	0.0005	8.38	0.0000	-38 -28 -9
	0.0000	0.0005	7.97	0.0000	12 -6 27
173	**0.0000**	**0.0005**	**8.74**	0.0000	**36 -6 -22**
	0.0000	0.0005	7.85	0.0000	38 -34 -4
	0.0000	0.0005	7.51	0.0000	39 -15 -16
21	**0.0000**	**0.0005**	**7.67**	0.0000	**-21 -36 3**
75	**0.0000**	**0.0005**	**7.59**	0.0000	**14 9 0**
	0.0000	0.0005	6.55	0.0000	9 4 -6
	0.0001	0.0005	6.25	0.0000	18 12 6
15	**0.0000**	**0.0005**	**7.51**	0.0000	**0 0 2**
23	**0.0000**	**0.0005**	**6.76**	0.0000	**28 -40 -6**
	0.0000	0.0005	6.55	0.0000	26 -46 0

Height threshold: statistic u= 5.16 (0.0500 FWE) Degrees of freedom = [1 398]
Design: 2 Groups: Two Sample T test; 1 scan per subject: 200(GrpA),200(GrpB)
Search vol: 1939410 cmm, 574640 voxels
Perms: 100000 permutations of conditions, bhPerms=0 Voxel size: [1.50, 1.50, 1.50] mm

(c) SnPM Report



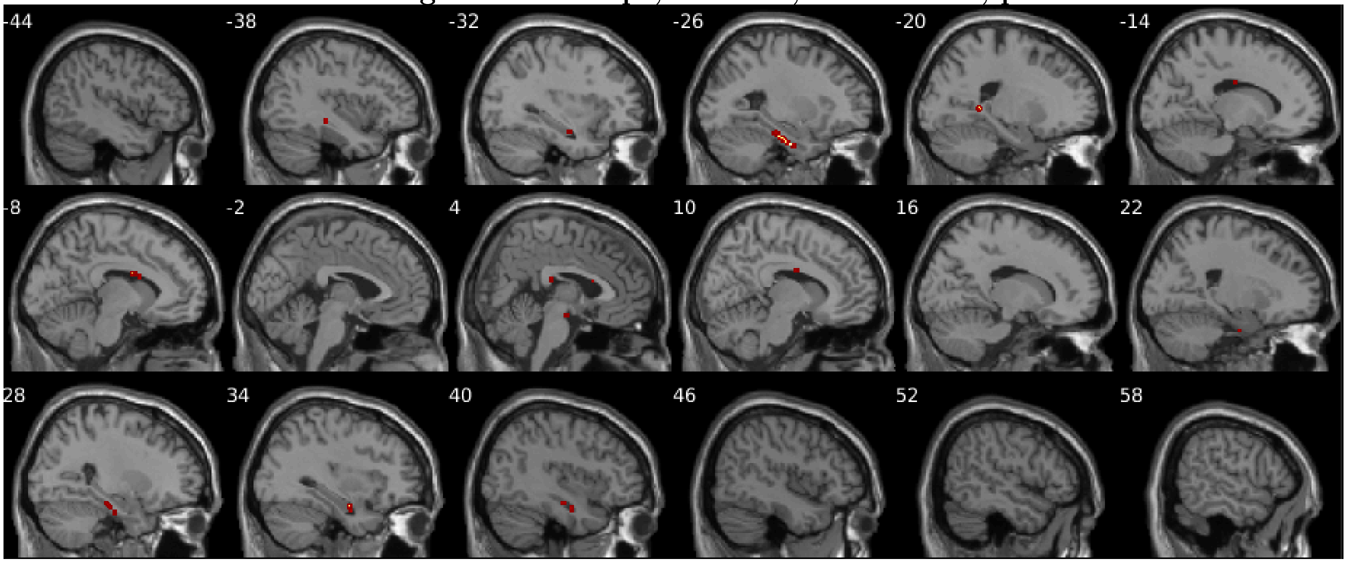
cluster-level	voxel-level				
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	T	p_{uncorr}	x,y,z mm
79	**0.0000**	**0.0004**	**10.03**	0.0000	**-26 -12 -32**
	0.0000	0.0004	9.58	0.0000	-24 -24 -21
	0.0042	0.0004	5.84	0.0000	-30 -18 -26
65	**0.0000**	**0.0004**	**9.55**	0.0000	**27 -22 -21**
	0.0000	0.0004	8.96	0.0000	26 -10 -32
1218	**0.0000**	**0.0004**	**9.04**	0.0000	**-33 -9 -20**
	0.0000	0.0004	8.38	0.0000	-38 -28 -9
	0.0000	0.0004	7.97	0.0000	12 -6 27
164	**0.0000**	**0.0004**	**8.74**	0.0000	**36 -6 -22**
	0.0000	0.0004	7.85	0.0000	38 -34 -4
	0.0000	0.0004	7.51	0.0000	39 -15 -16
20	**0.0000**	**0.0004**	**7.67**	0.0000	**-21 -36 3**
63	**0.0000**	**0.0004**	**7.59**	0.0000	**14 9 0**
	0.0001	0.0004	6.55	0.0000	9 4 -6
	0.0006	0.0004	6.25	0.0000	18 12 6
15	**0.0000**	**0.0004**	**7.51**	0.0000	**0 0 2**
9	**0.0000**	**0.0004**	**7.45**	0.0000	**-27 -40 -8**
21	**0.0001**	**0.0004**	**6.76**	0.0000	**28 -40 -6**
	0.0001	0.0004	6.55	0.0000	26 -46 0

Height threshold: statistic u= 5.25 (0.0500 FWE) Degrees of freedom = [1 398]
Design: 2 Groups: Two Sample T test; 1 scan per subject: 200(GrpA),200(GrpB)
Search vol: 1939410 cmm, 574640 voxels
Perms: 100000 permutations of conditions, bhPerms=0 Voxel size: [1.50, 1.50, 1.50] mm

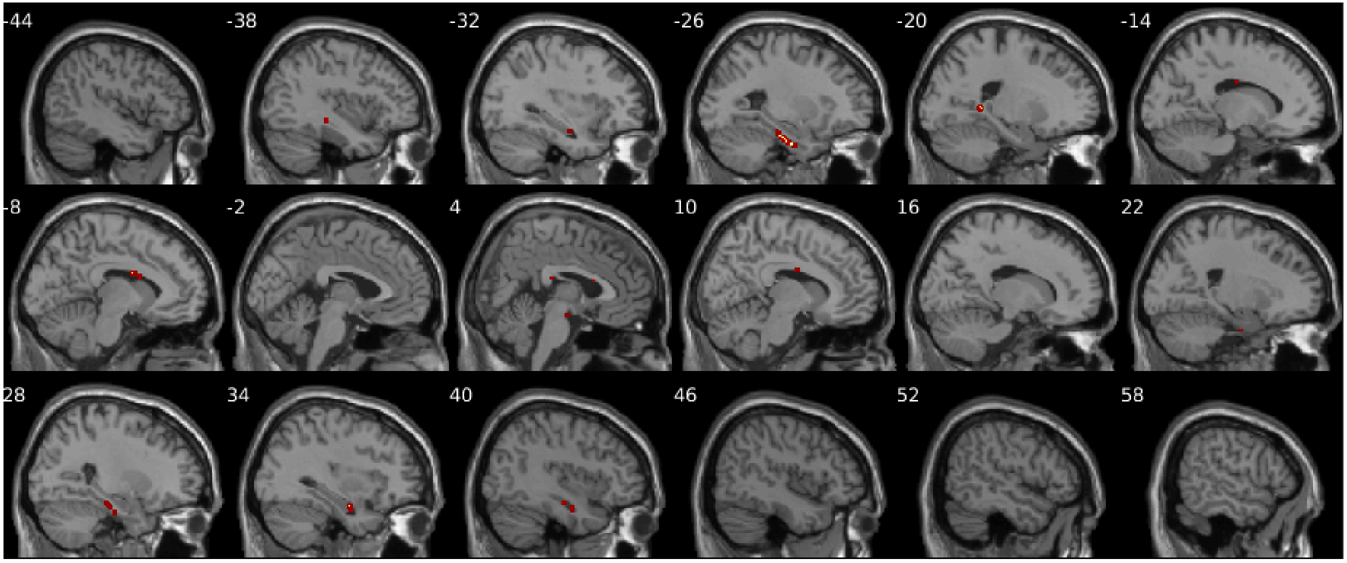
(d) RapidPT Report

Figure 4: Thresholded FWER corrected statistical maps at ($\alpha = 0.05$). The images show the test statistics for which the null was rejected in SnPM (top) and RapidPT (bottom). The tables show a numerical summary of the images. The columns refer to: k - cluster size, $p_{FWE-corr}$ - corrected p-values, T - max cluster t-statistic. $p_{FWE-corr}$ that appears as 0.0000 are $1e^{-5}$. These results were obtained from a run with the hyperparameters specified in the title.

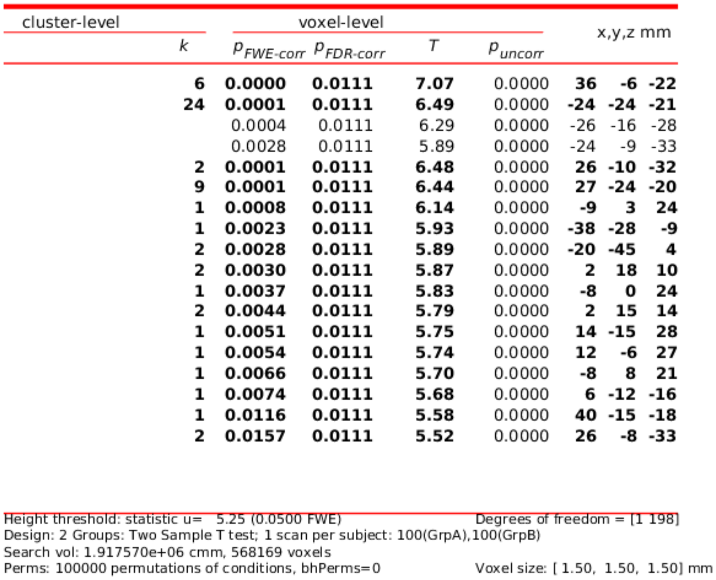
ADNI Statistical Significance Maps, n = 200, L = 100000, p-value = 0.05



(a) SnPM P-Map. Slice -44 to 58 from left to right top to bottom.



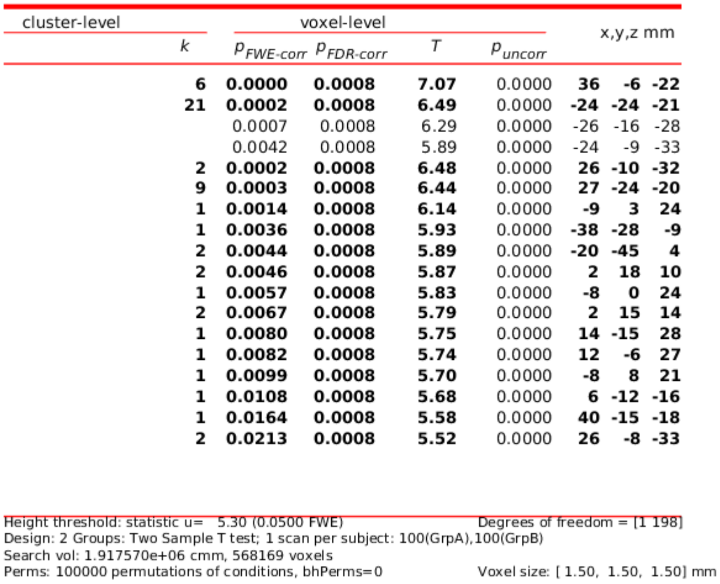
(b) RapidPT P-Map. Slice -44 to 58 from left to right top to bottom. $\eta = 0.5\%, l = 200$.



cluster-level	voxel-level				x,y,z mm
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	T	p_{uncorr}	
6	**0.0000**	**0.0111**	**7.07**	0.0000	**36 -6 -22**
24	**0.0001**	**0.0111**	**6.49**	0.0000	**-24 -24 -21**
	0.0004	0.0111	6.29	0.0000	-26 -16 -28
	0.0028	0.0111	5.89	0.0000	-24 -9 -33
2	**0.0001**	**0.0111**	**6.48**	0.0000	**26 -10 -32**
9	**0.0001**	**0.0111**	**6.44**	0.0000	**27 -24 -20**
1	**0.0008**	**0.0111**	**6.14**	0.0000	**-9 3 24**
1	**0.0023**	**0.0111**	**5.93**	0.0000	**-38 -28 -9**
2	**0.0028**	**0.0111**	**5.89**	0.0000	**-20 -45 4**
2	**0.0030**	**0.0111**	**5.87**	0.0000	**2 18 10**
1	**0.0037**	**0.0111**	**5.83**	0.0000	**-8 0 24**
2	**0.0044**	**0.0111**	**5.79**	0.0000	**2 15 14**
1	**0.0051**	**0.0111**	**5.75**	0.0000	**14 -15 28**
1	**0.0054**	**0.0111**	**5.74**	0.0000	**12 -6 27**
1	**0.0066**	**0.0111**	**5.70**	0.0000	**-8 8 21**
1	**0.0074**	**0.0111**	**5.68**	0.0000	**6 -12 -16**
1	**0.0116**	**0.0111**	**5.58**	0.0000	**40 -15 -18**
2	**0.0157**	**0.0111**	**5.52**	0.0000	**26 -8 -33**

```
Height threshold: statistic u=   5.25 (0.0500 FWE)                Degrees of freedom = [1 198]
Design: 2 Groups: Two Sample T test; 1 scan per subject: 100(GrpA),100(GrpB)
Search vol: 1.917570e+06 cmm, 568169 voxels
Perms: 100000 permutations of conditions, bhPerms=0               Voxel size: [ 1.50,  1.50,  1.50] mm
```

(c) SnPM Report



cluster-level	voxel-level				x,y,z mm
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	T	p_{uncorr}	
6	**0.0000**	**0.0008**	**7.07**	0.0000	**36 -6 -22**
21	**0.0002**	**0.0008**	**6.49**	0.0000	**-24 -24 -21**
	0.0007	0.0008	6.29	0.0000	-26 -16 -28
	0.0042	0.0008	5.89	0.0000	-24 -9 -33
2	**0.0002**	**0.0008**	**6.48**	0.0000	**26 -10 -32**
9	**0.0003**	**0.0008**	**6.44**	0.0000	**27 -24 -20**
1	**0.0014**	**0.0008**	**6.14**	0.0000	**-9 3 24**
1	**0.0036**	**0.0008**	**5.93**	0.0000	**-38 -28 -9**
2	**0.0044**	**0.0008**	**5.89**	0.0000	**-20 -45 4**
2	**0.0046**	**0.0008**	**5.87**	0.0000	**2 18 10**
1	**0.0057**	**0.0008**	**5.83**	0.0000	**-8 0 24**
2	**0.0067**	**0.0008**	**5.79**	0.0000	**2 15 14**
1	**0.0080**	**0.0008**	**5.75**	0.0000	**14 -15 28**
1	**0.0082**	**0.0008**	**5.74**	0.0000	**12 -6 27**
1	**0.0099**	**0.0008**	**5.70**	0.0000	**-8 8 21**
1	**0.0108**	**0.0008**	**5.68**	0.0000	**6 -12 -16**
1	**0.0164**	**0.0008**	**5.58**	0.0000	**40 -15 -18**
2	**0.0213**	**0.0008**	**5.52**	0.0000	**26 -8 -33**

```
Height threshold: statistic u=   5.30 (0.0500 FWE)                Degrees of freedom = [1 198]
Design: 2 Groups: Two Sample T test; 1 scan per subject: 100(GrpA),100(GrpB)
Search vol: 1.917570e+06 cmm, 568169 voxels
Perms: 100000 permutations of conditions, bhPerms=0               Voxel size: [ 1.50,  1.50,  1.50] mm
```

(d) RapidPT Report

Figure 5: Thresholded FWER corrected statistical maps at ($\alpha = 0.05$). The images show the test statistics for which the null was rejected in SnPM (top) and RapidPT (bottom). The tables show a numerical summary of the images. The columns refer to: k - cluster size, $p_{FWE-corr}$ - corrected p-values, T - max cluster t-statistic. $p_{FWE-corr}$ that appears as 0.0000 are $1e^{-5}$. These results were obtained from a run with the hyperparameters specified in the title.

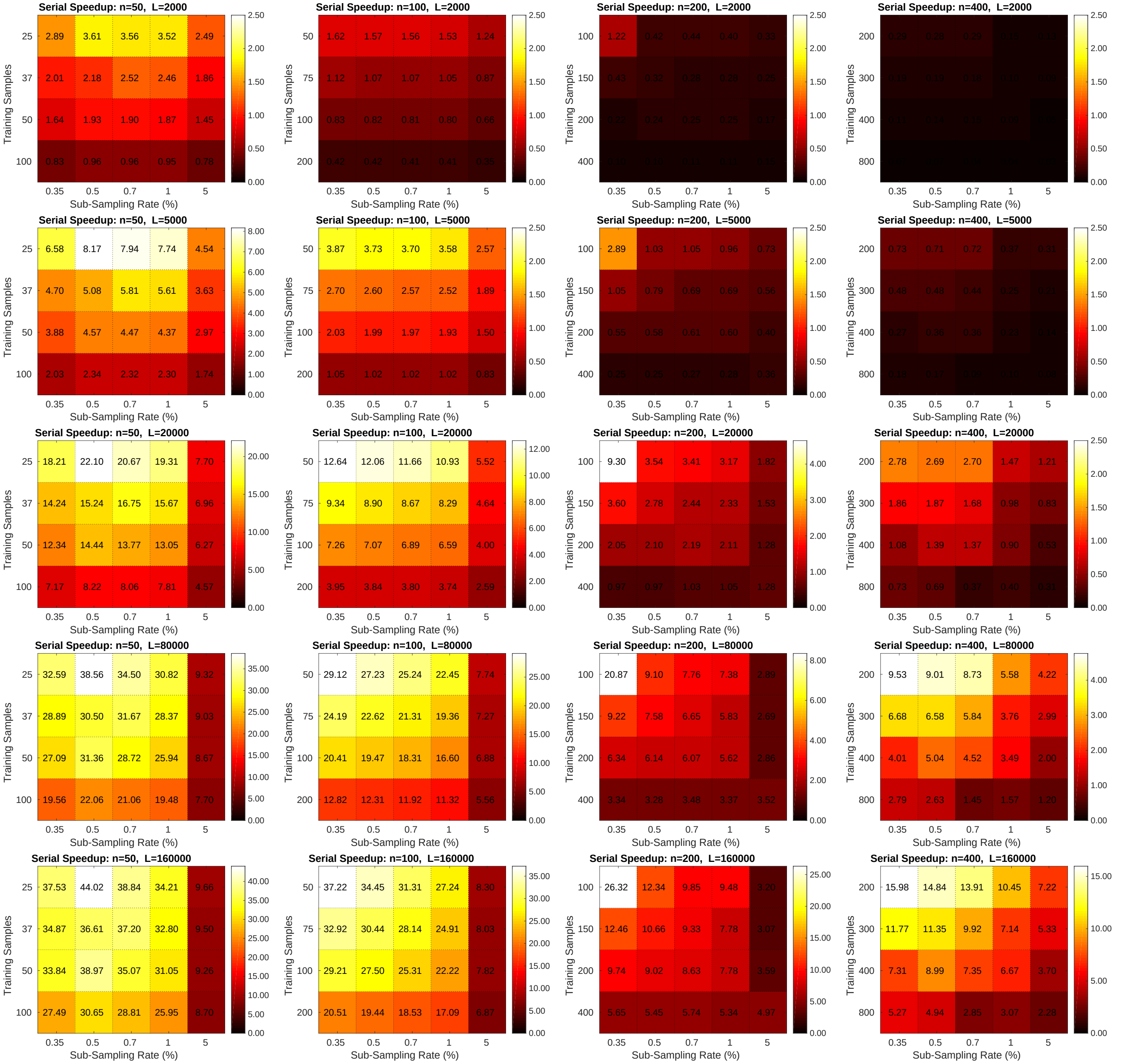


Figure 6: Colormaps of the speedup gains of RapidPT over SnPM running on a single core. Each colormap corresponds to a run with a given dataset and a fixed number of permutations, and displays 20 different speedups resulting from the hyperparameter combinations. Columns 1, 2, 3, and 4 of this figure are correspond to the 50, 100, 200, and 400 subject datasets, respectively. The colormaps for $L = 10,000$ and $L = 40,000$ are omitted because they are shown in the main document.

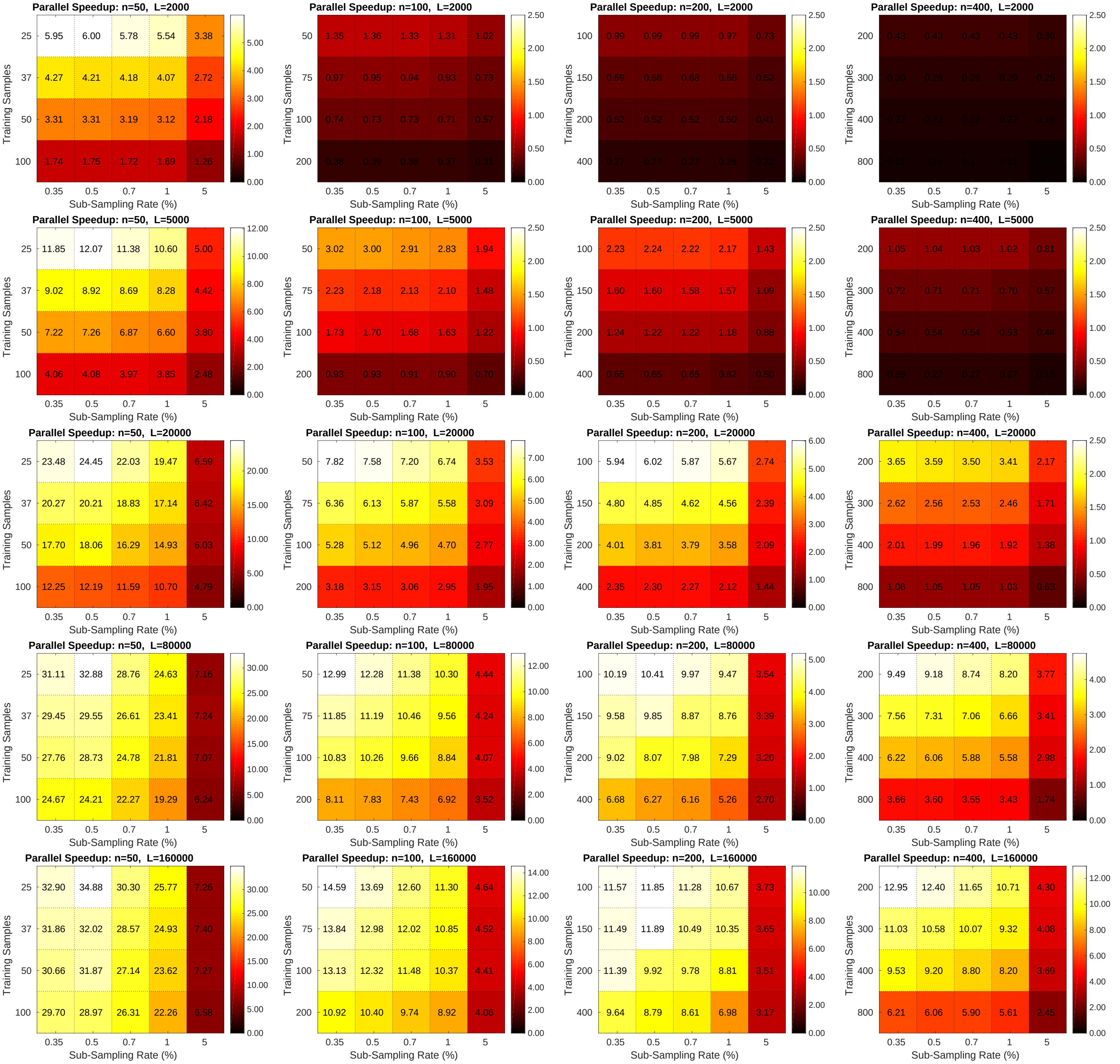


Figure 7: Colormaps of the speedup gains of RapidPT over SnPM running on a parallel environment with 16 cores as described in the main document. Each colormap corresponds to a run with a given dataset and a fixed number of permutations, and displays 20 different speedups resulting from the hyperparameter combinations. Columns 1, 2, 3, and 4 of this figure are correspond to the 50, 100, 200, and 400 subject datasets, respectively. The colormaps for $L = 10,000$ and $L = 40,000$ are omitted because they are shown in the main document.

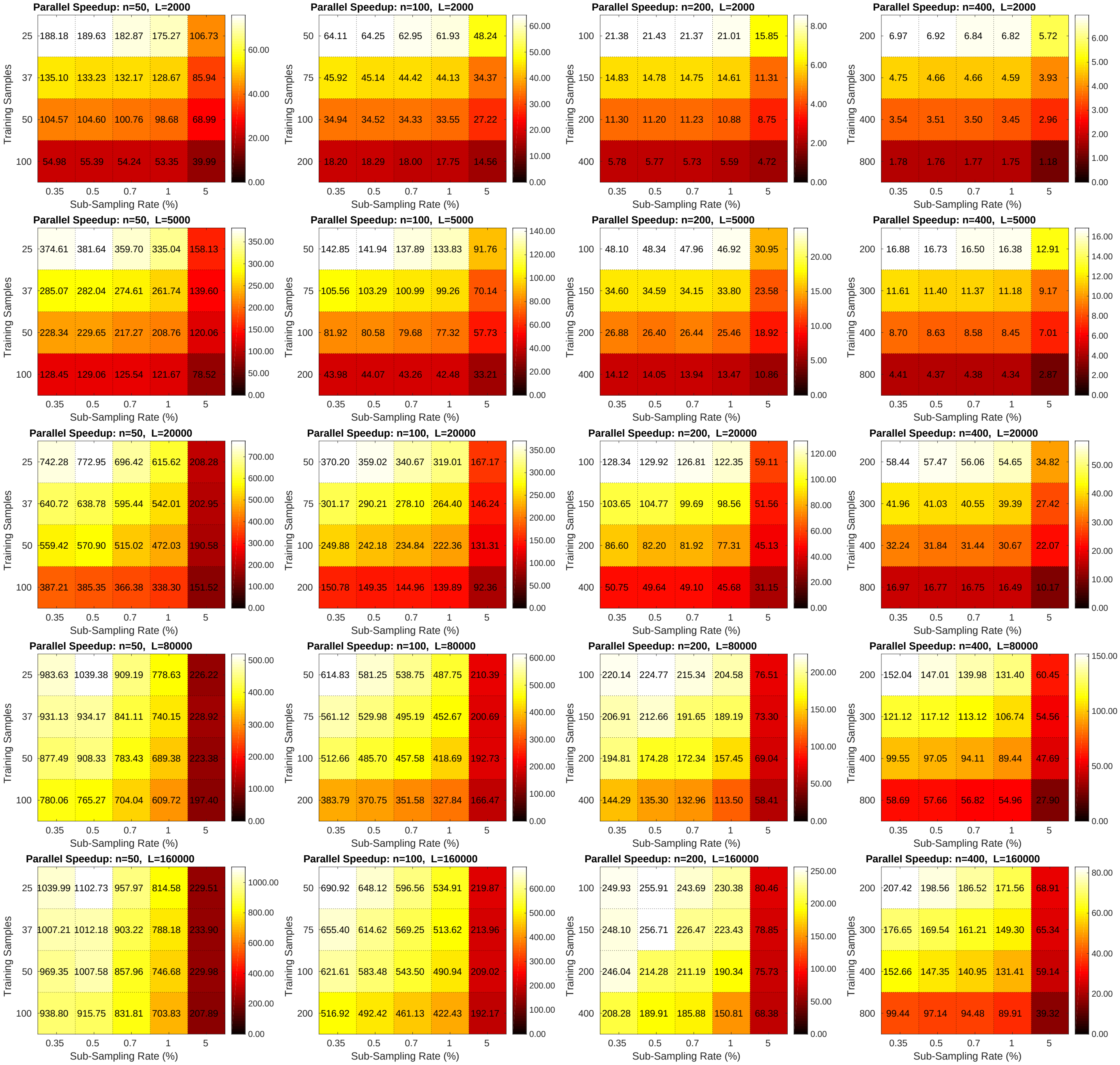


Figure 8: Colormaps of the speedup gains of RapidPT over NaivePT running on a parallel environment with 16 cores as described in the main document. Each colormap corresponds to a run with a given dataset and a fixed number of permutations, and displays 20 different speedups resulting from the hyperparameter combinations. Columns 1, 2, 3, and 4 of this figure are correspond to the 50, 100, 200, and 400 subject datasets, respectively. The colormaps for $L = 10,000$ and $L = 40,000$ are omitted because they are shown in the main document.

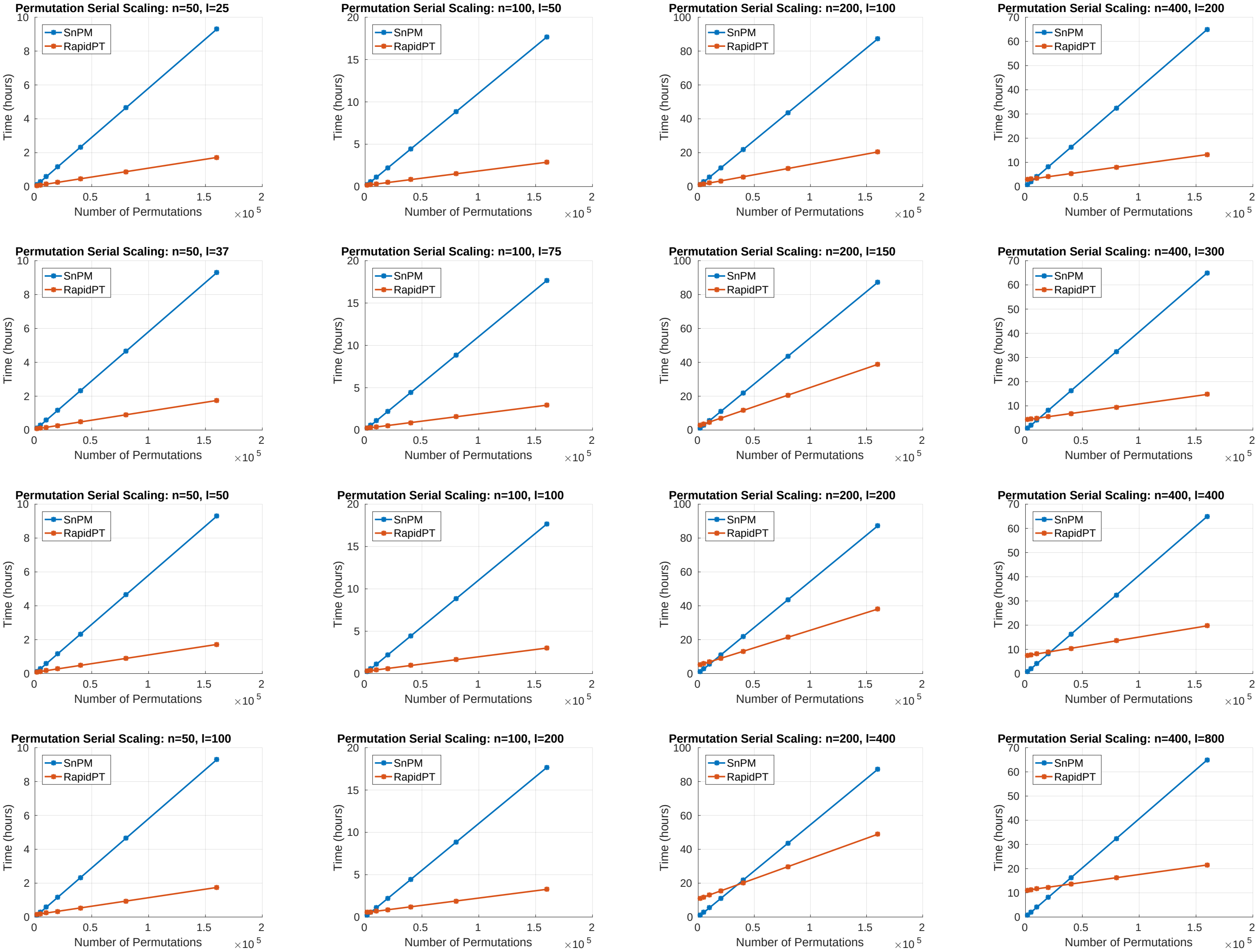


Figure 9: Effect of the number of permutations on the runtime performance of RapidPT and SnPM running on a single core. Columns 1, 2, 3, and 4 of this figure are associated to the 50, 100, 200, and 400 subject datasets, respectively. Rows 1, 2, 3, and 4 are associated to runs using $l = \frac{n}{2}$, $l = \frac{3n}{4}$, $l = n$, $l = 2n$, respectively. The sub-sampling rate hyperparameter was fixed to: $\eta = 0.35\%$.

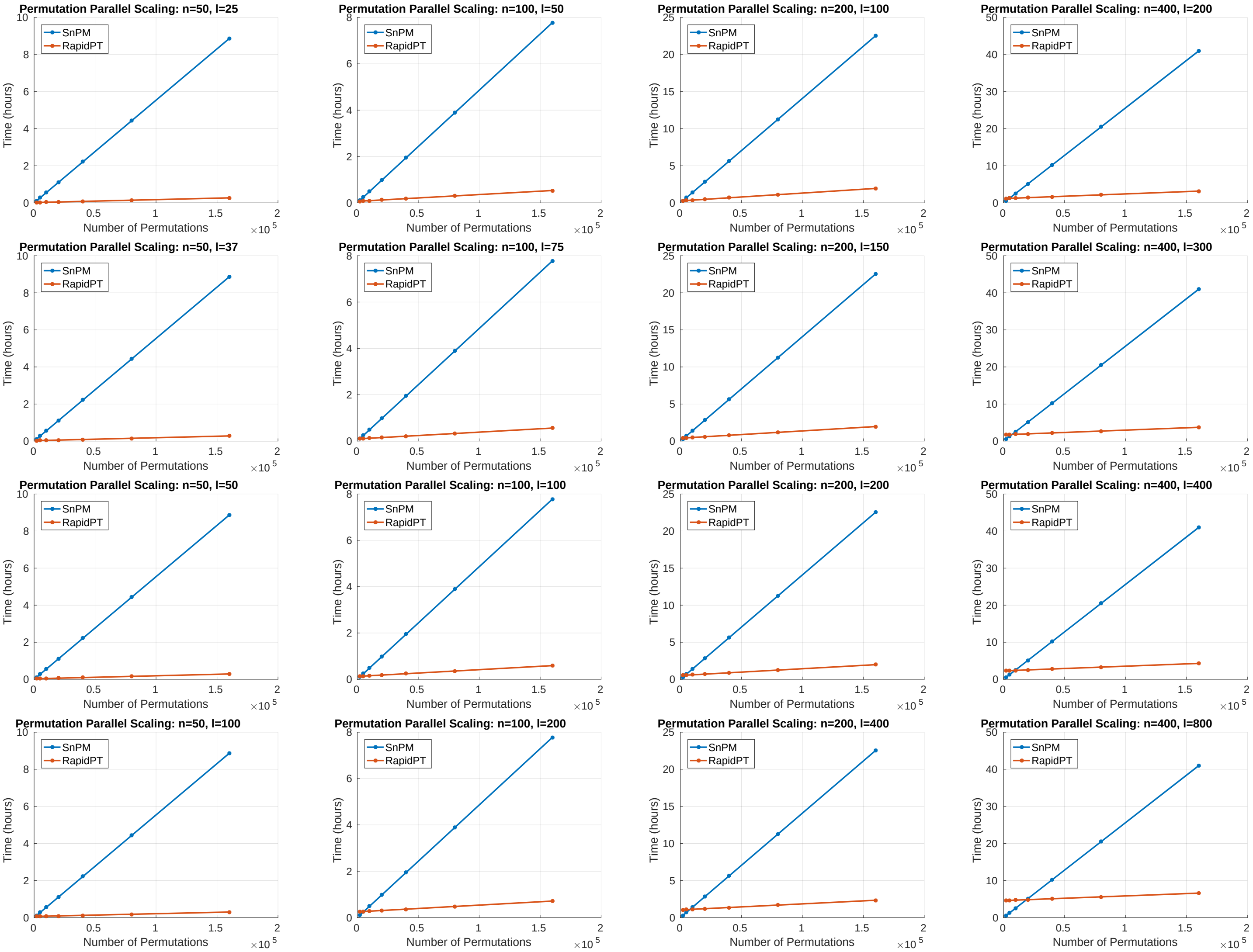


Figure 10: Effect of the number of permutations on the runtime performance of RapidPT and SnPM running on 16 cores. Columns 1, 2, 3, and 4 of this figure are associated to the 50, 100, 200, and 400 subject datasets, respectively. Rows 1, 2, 3, and 4 are associated to runs using $l = \frac{n}{2}$, $l = \frac{3n}{4}$, $l = n$, $l = 2n$, respectively. The sub-sampling rate hyperparameter was fixed to: $\eta = 0.35\%$.

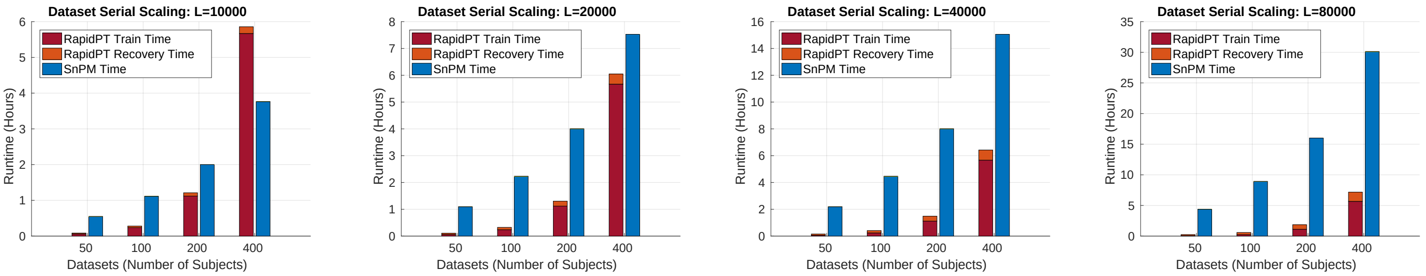


Figure 11: Effect of the dataset size on the runtime of RapidPT and SnPM running on a single core. The overall measured time of RapidPT is the result of the total time spent on the training phase and the recovery phase. Columns 1, 2, 3, and 4 are associated to runs using $L = 10000$, $L = 20000$, $L = 40000$, $L = 80000$, respectively.The hyperparameters used were: $\eta = 0.35\%$ and $l = n$.

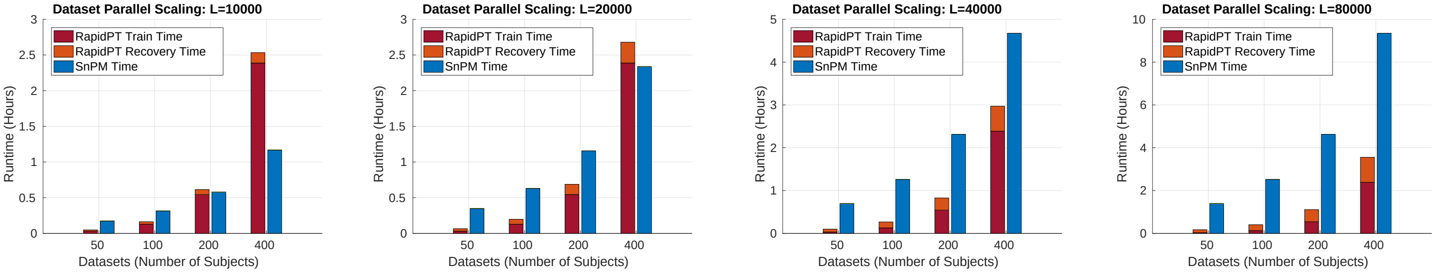


Figure 12: Effect of the dataset size on the runtime of RapidPT and SnPM running on 16 cores. The overall measured time of RapidPT is the result of the total time spent on the training phase and the recovery phase. Columns 1, 2, 3, and 4 are associated to runs using $L = 10000$, $L = 20000$, $L = 40000$, $L = 80000$, respectively.The hyperparameters used were: $\eta = 0.35\%$ and $l = n$.

3. Sub-Sampling Rate in Practice

3.1. Minimum Sub-Sampling Rate Estimate for Experiments

Our experiments showed that as long as we selected a sub-sampling rate, $\eta \geq 0.35\%$, we could accurately recover the max null distribution. This value follows from the discussion on low-rank matrix completion in section 3 and the analysis of recovery guarantees in section 5. In particular, the authors of Candès and Tao (2010) and others have shown that with a *sufficient* number of entries one can recover the orthogonal basis of the row space as well as the expansion coefficients. The sufficient number of entries is on the order of $rlog(d)$, where r is the column space rank and d is the ambient dimension. In our case, the column rank space of $\mathbf{T}$ is bounded by the number of subjects (i.e $r \leq n$), as discussed in the paper. In general the ambient dimension of $\mathbf{T}$ is v. Equation brings these numbers together and defines an estimate of the minimum sub-sampling rate should be in practice.

$$\eta v \geq nlog(v) \tag{3}$$

$$\eta \geq \frac{nlog(v)}{v} \tag{4}$$

In our experiments, the number of subjects and number of voxels for each dataset were:

Number of Subjects and Number of Voxels: (n,v)			
(50,547783)	(100,558295)	(200,568086)	(400,574640)

Table 1: Number of subjects and number of voxels per subject on each dataset.

Therefore, the corresponding estimate of the minimum sub-sampling rate would be:

Minimum sub-sampling rate estimate: η_{min}			
$\approx$ 0.1206%	$\approx$ 0.2370%	$\approx$ 0.4665%	$\approx$ 0.9231%

Table 2: Number of subjects and number of voxels per subject on each dataset.

Evidently, the bound that we have established in equation (3) is conservative because as shown in our results we were able to recover the maxnull for all of the datasets with $\eta \geq 0.35\%$. However, it is a useful bound that is incorporated into RapidPT to avoid inexperienced users from using an η that will lead to incorrect results.

3.2. A practical example

Figure 13 illustrates the effect of η on the KL-Divergence of the maxnull distributions of RapidPT vs. SnPM. The data used for this example is a subset of the data used for non-regressing testing of SnPM Maumet (2017). The data is composed of 14 subjects ($n = 14$) and 1000 voxels ($v = 1000$) per subject. Therefore, from equation (3) we find that the minimum sub-sampling rate should be roughly $\eta_{min} \approx 4.2\%$. Once we have sub-sampled a sufficient number of entries we are able to accurately recover the maxnull, and in fact, sub-sampling more entries will not improve significantly the KL-Divergence as shown in figure 13.

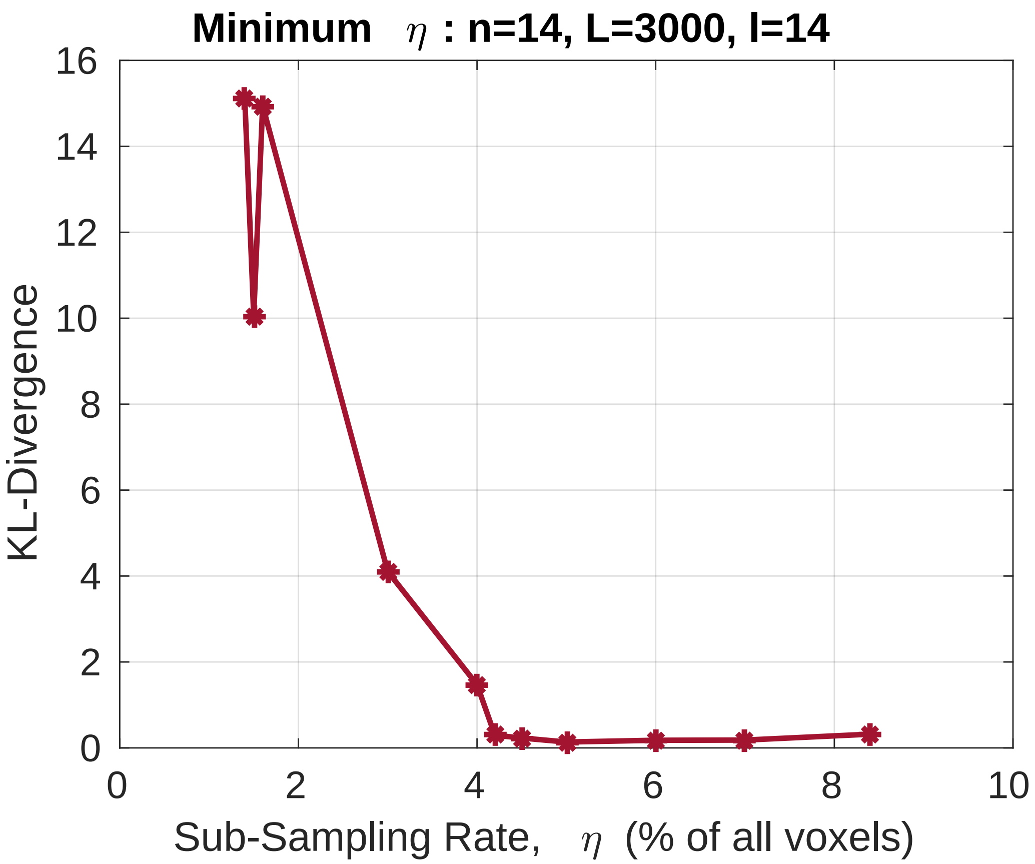


Figure 13: An experiment showing the effect of the sub-sampling rate (η) on the KL-Divergence.

References

Balzano, L., Nowak, R., Recht, B., Sept 2010. Online identification and tracking of subspaces from highly incomplete information. In: Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on. pp. 704–711.

Benaych-Georges, F., Nadakuditi, R. R., 2011. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics 227 (1), 494–521.

Candès, E., Tao, T., 2010. The power of convex relaxation: Near-optimal matrix completion. IEEE Trans. Inf. Theor. 56 (5), 2053–2080.
URL `http://dx.doi.org/10.1109/TIT.2010.2044061`

Edelman, A., 1988. Eigenvalues and condition numbers of random matrices. SIAM Journal on Matrix Analysis and Applications 9 (4), 543–560.

He, J., Balzano, L., Szlam, A., June 2012. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 1568–1575.

Marčenko, A. A., Pastur, L. A., 1967. Distribution of eigenvalues for some sets of random matrices. Sbornik: Mathematics 1 (4), 457–483.

Maumet, C., 2017. Snpm test data. Available online at https://github.com/SnPM-toolbox.