Dated: January 16, 2025

Spectral normative modeling of brain structure

Sina Mansour L.[1,2,✉], **Maria A. Di Biase**[2,3,4], **Hongwei Yan**[1], **Aihuiping Xue**[1], **Narayanaswamy Venketasubramanian**[5,6], **Eddie Chong**[6], **Aaron Alexander-Bloch**[7,8,9], **Christopher Chen**[6,10], **Juan Helen Zhou**[1,11,12], **B.T. Thomas Yeo**[1,11,12,13,14,15,⚖,✉], and **Andrew Zalesky**[2,16,⚖,✉]

[1] *Centre for Sleep & Cognition & Centre for Translational Magnetic Resonance Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[2] *Systems Neuroscience Lab, Department of Psychiatry, The University of Melbourne, Parkville, Victoria, Australia*
[3] *Stem Cell Disease Modelling Lab, Department of Anatomy and Physiology, The University of Melbourne, Parkville, Victoria, Australia*
[4] *Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, USA*
[5] *Raffles Neuroscience Centre, Raffles Hospital, Singapore*
[6] *Memory Aging and Cognition Centre, National University Health System, Singapore, Singapore*
[7] *Brain-Gene Development Laboratory, Lifespan Brain Institute at Children's Hospital of Philadelphia and University of Pennsylvania, Philadelphia, PA, United States*
[8] *Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA United States*
[9] *Department of Psychiatry, University of Pennsylvania, Philadelphia, PA United States*
[10] *Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[11] *Department of Electrical and Computer Engineering, National University of Singapore, Singapore*
[12] *Integrative Sciences and Engineering Programme (ISEP), National University of Singapore, Singapore*
[13] *Department of Medicine, Healthy Longevity Translational Research Programme, Human Potential Translational Research Programme & Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[14] *N.1 Institute for Health, National University of Singapore, Singapore*
[15] *Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, United States*
[16] *Department of Biomedical Engineering, The University of Melbourne, Parkville, Victoria, Australia*
⚖ *These authors contributed equally to this work.*

✉ *E-mail: sina.mansour.lakouraj@gmail.com; thomas.yeo@nus.edu.sg; azalesky@unimelb.edu.au*

Abstract: Normative modeling in neuroscience aims to characterize interindividual variation in brain phenotypes and thus establish reference ranges, or brain charts, against which individual brains can be compared. Normative models are typically limited to coarse spatial scales due to computational constraints, limiting their spatial specificity. They additionally depend on fixed regions from fixed parcellation atlases, restricting their adaptability to alternative parcellation schemes. To overcome these key limitations, we propose *spectral normative modeling* (SNM), which leverages brain eigenmodes for efficient spatial reconstruction to generate normative ranges for arbitrary new regions of interest. Benchmarking against conventional counterparts, SNM achieves a 98.3% speedup in computing accurate normative ranges across spatial scales, from millimeters to the whole brain. We demonstrate its utility by elucidating high-resolution individual cortical atrophy patterns and characterizing the heterogeneous nature of neurodegeneration in Alzheimer's disease. SNM lays the groundwork for a new generation of spatially precise brain charts, offering substantial potential to drive advances in individualized precision medicine.

Keywords: Normative Modeling; Brain Charts; Graph Signal Processing; Brain Eigenmodes

1 Introduction

Normative modeling aims to estimate reference ranges of normative population-wide variation in a phenotype of interest[1–3]. Individuals can be benchmarked to an established normative range for their age and sex to determine whether they fall outside a critical healthy range[4,5]. Normative modeling is a major goal in neuroscience and such models have been established for numerous whole-brain and regional phenotypes, such as cortical thickness or volume[6]. Prior research has demonstrated how such methods can accurately model the heterogeneous nature of these deviation patterns in brain structure[4,6]. Consequently, normative techniques hold significant promise in advancing precision medicine[7–14].

Normative brain charts for magnetic resonance imaging (MRI) phenotypes can in principle be established at the highly detailed spatial resolution at which the MRI scan is acquired. This would enable highly localized and spatially specific inference about an individual's deviations in a cortical phenotype. However, thus far, concerns about computational feasibility and quality control of higher-resolution phenotypes have hindered the development of efficient normative charts with high spatial precision. Conventional normative approaches are designed to estimate ranges for a single phenotypic summary statistic, such as mean cortical thickness averaged across the whole cortex[2,15,16]. To reach higher spatial specificity, normative brain charting studies typically repeat model-fitting for regional summary statistics defined over a predetermined brain atlas[6,17].

Alternatively, a computationally burdensome exhaustive repetition of conventional techniques over all voxels/vertices has been used to produce norms at high spatial resolution[8,13,18]. These approaches are inherently sensitive to noise at the level of single voxels, and their computational intractability becomes even more pronounced when population-wide normative models are trained using data collated from several large-scale imaging biobanks. Moreover, with current normative modeling approaches, determining a reference range for a new spatial region of interest requires fitting a separate normative model anew, a process often hindered by limited access to the original training data that is typically unavailable to end-users. Establishing a methodological framework that efficiently alleviates such limitations will enable a more principled and efficient mapping of brain charts at high spatial resolution compared to the current brute force approach.

Developing high-resolution normative models is challenging, particularly due to the high dimensionality of the feature space, i.e. hundreds of thousands of vertices on the cortical surface mesh. Moreover, spatial dependencies across vertices/voxels undermine independence assumptions and further complicate the development of models that accurately explain high-resolution statistical interdependencies[19–21]. As such, finding an appropriate low-dimensional encoding of high-resolution cortical information may enable the development of computationally tractable techniques for high-resolution normative models.

Through recent advances in brain signal processing, eigenmodes constructed from the brain's geometry and connectivity have yielded promising basis functions that can summarize phenotypic variations on the cortical surface[22–24]. As such, eigenmodes can provide a solution to high-resolution normative modeling of brain phenotypes. Spatial variation in a cortical phenotype can be captured using a lower dimensional graph spectral embedding[25,26]. We exploit this parsimonious eigenmode basis to establish normative models on the coefficients of cortical phenotypes expressed in this lower dimensional latent space. By formulating a method that relates the normative range of an arbitrary region of interest to eigenmode normative ranges, we develop a computationally efficient method that simultaneously estimates normative ranges over multiple spatial granularities.

In this work, we introduce spectral normative modeling (SNM) as a novel method to establish normative reference brain charts that are independent of any spatial resolution or parcellation atlas. We explain how high-dimensional brain phenotypes (e.g., cortical thickness) can be summarized by a concise representation based on brain eigenmodes and detail the mathematical framework to reconstruct reference normative brain models at arbitrary spatial scales. We evaluate the performance of the proposed model relative to existing approaches to demonstrate its success in providing efficient and accurate multi-scale individual-level insights that advance the frontiers of precision normative assessments. Finally, we demonstrate the practical utility of SNM in characterizing individual deviations in cortical thickness linked to cognitive impairments in Alzheimer's disease (AD). This emphasizes the success of SNM in generalizing to unseen samples and elucidating individual heterogeneity in cortical atrophy.

2 Results

2.1 SNM: An Efficient Framework for Multi-Scale Normative Modeling

The fundamental characteristics of many engineered and natural systems can be modeled by their structural eigenmodes, which offer a simplified yet powerful means of capturing the system's behavior. Recent advances in neuroscience have similarly demonstrated that brain eigenmodes provide a parsimonious basis for characterizing cortical information[22,25,27]. By leveraging such basis sets, we aim to enhance both the efficiency and spatial versatility of conventional normative models. In particular, we use brain connectivity eigenmodes, which naturally extend across both cortical hemispheres while preserving key neuroanatomical landmarks, such as homotopic symmetries and the separation of cortical lobes. This is especially advantageous given that pathological brain alterations often propagate along the brain's structural network[28–34], making connectivity eigenmodes well suited to capture normative deviations induced by mechanisms of axonal propagation. However, it is important to note that the methodological advances introduced in this work are not limited to connectivity eigenmodes and can be readily generalized to any orthonormal basis set for information reconstruction.

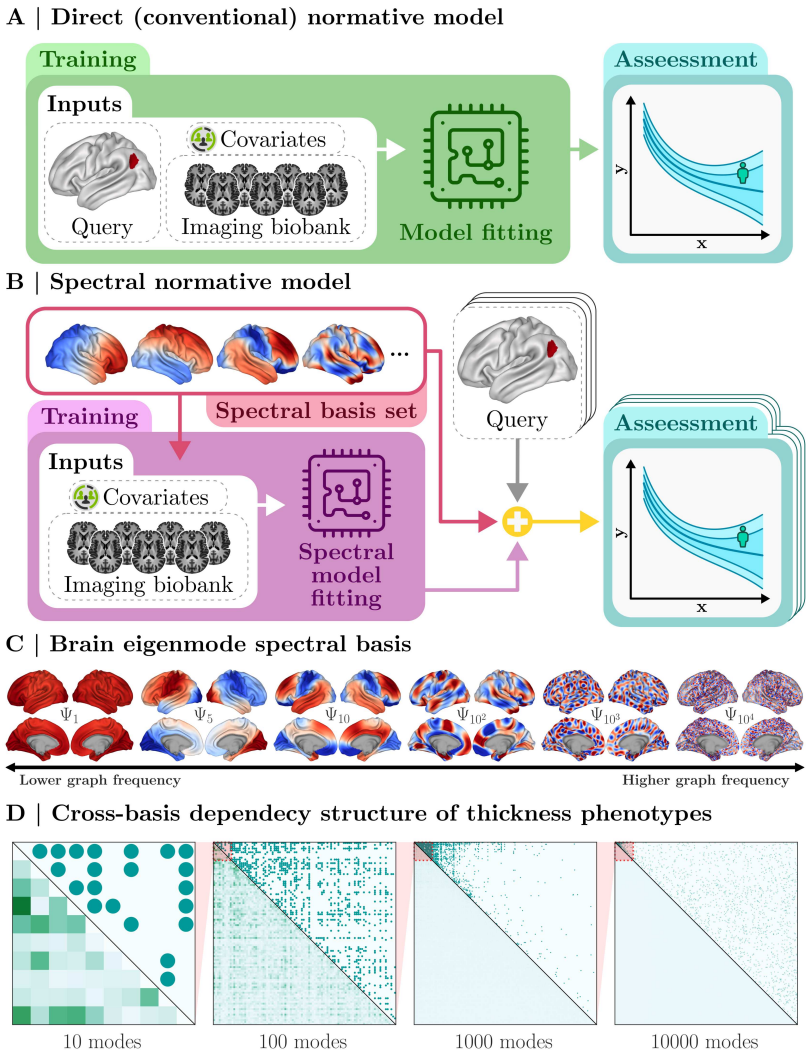


Figure. 1. Schematic Comparison of Spectral Normative Modeling (SNM) with Conventional Normative Models. (A) Design diagram for conventional (direct) normative models, which require a predetermined spatial query for model training. **(B)** Design diagram of SNM, which alleviates the need for predefined spatial queries. Instead, SNM enables assessments of multiple, arbitrary, a posteriori-defined spatial queries, allowing for greater flexibility. **(C)** Cortical projections of brain connectivity eigenmodes used in SNM. Eigenmodes across a range of graph frequencies are shown, illustrating how higher frequencies capture increasingly finer spatial details. **(D)** Cross-basis correlation magnitudes of cortical thickness phenotypes encoded onto the eigenmode basis set. The sparsity of cross-mode dependencies is demonstrated through nested heatmaps, with the lower triangles representing correlation magnitudes and the upper triangles indicating suprathreshold eigenmode pairs at a correlation threshold of $\rho = 0.25$.

Conventional normative modeling typically aims to infer the normative range of a phenotype, such as cortical thickness, for a fixed region of interest defined by a spatial query (Figure 1A). These approaches, referred to here as direct normative models, rely on predetermined spatial queries for each region under investigation. In contrast, SNM learns the normative distribution of a high-resolution phenotype using its graph spectral encoding on the eigenmode spatial basis set (Figure 1B). As a result, normative ranges can be rapidly computed across multiple spatial queries without refitting the model for each query (see Section 4.8 for details).

In this work, SNM utilizes connectivity eigenmodes derived from the random-walk Laplacian decomposition of high-resolution structural connectivity[35] ($L_{rw}\psi_i = \lambda_i\psi_i$, see Section 4.7 Brain Signal Reconstruction). These eigenmodes, shown in Figure 1C, are ordered according to graph spectral frequencies, with higher frequencies capturing increasingly finer spatial details. Importantly, cortical thickness phenotypes exhibit a sparse cross-basis dependency structure when encoded on the eigenmode basis (Figure 1D). This sparsity contributes to the computational tractability of SNM, particularly when incorporating a larger number of eigenmodes.

2.2 Eigenmodes Reconstruct Cortical Thickness and Normative Query Maps

We first demonstrate the accuracy with which brain eigenmodes encode information represented across typical brain maps, including individual cortical thickness phenotypes, and three families of spatial queries commonly used in normative models. This tests the utility of eigenmodes as a low-dimensional basis for reconstructing brain maps, while also gauging the appropriate number of eigenmodes to be included in SNM.

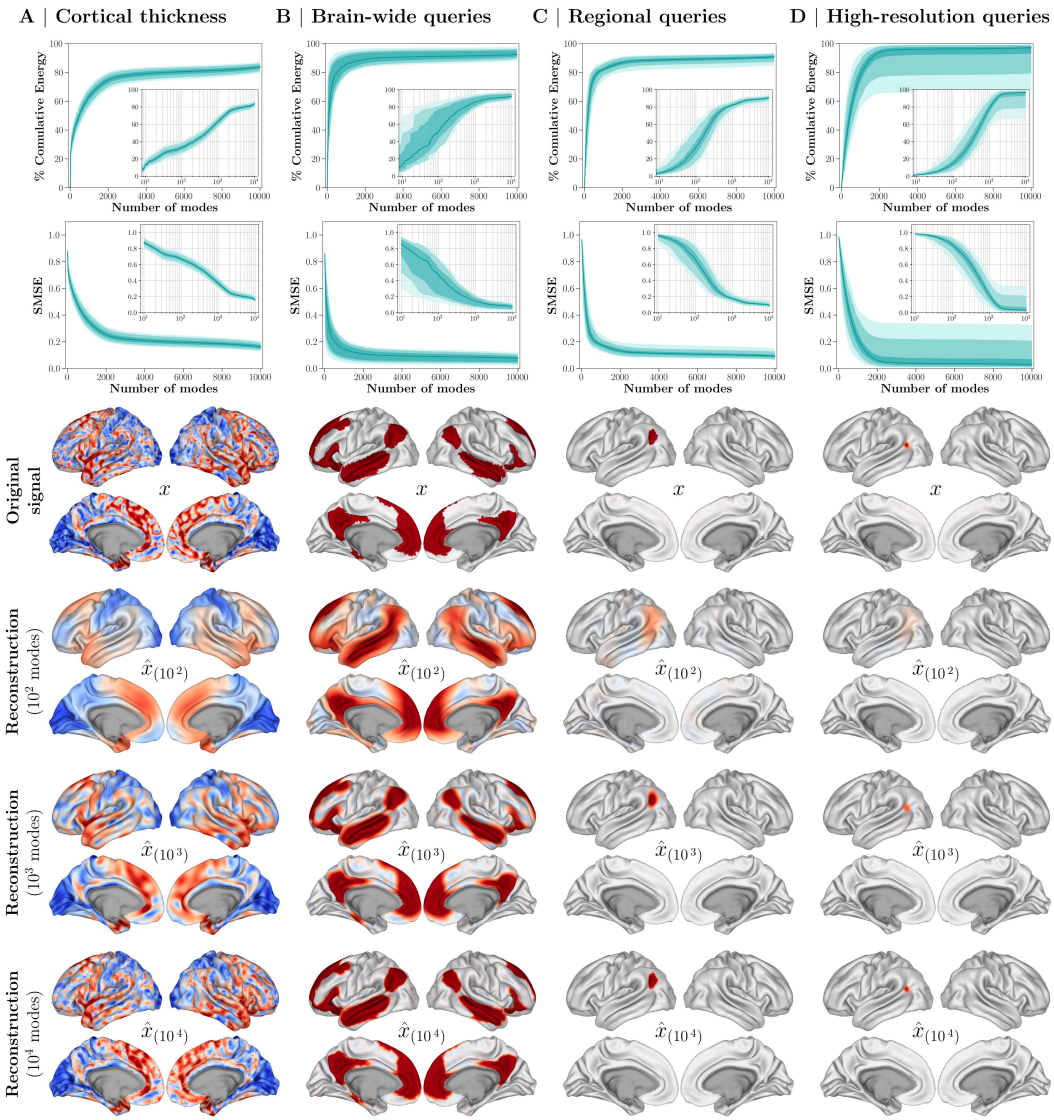


Figure. 2. Cortical Signal Reconstruction Accuracy. The columns display eigenmode reconstruction accuracies for **(A)** individual participant's cortical thickness phenotypes, as well as **(B)** brain-wide, **(C)** regional, and **(D)** high-resolution spatial queries. In the shaded line plots, the lines represent the median across all observations, while the shades indicate the [25, 75], [5, 95], and [1, 99] percentiles. For cortical thickness, the data comprises observations from 2,473 participants. For spatial queries, the data respectively includes a total of 25, 200, and 400 different brain-wide, regional, and high-resolution regions of interest. The first and second rows respectively show the cumulative energy and standardized mean square error (SMSE) as a function of the number of low-frequency eigenmodes used for spectral reconstruction (logarithmic x-axis for the insets). The third row illustrates one exemplary brain map from each category, while the last three rows show the same map reconstructed using 100, 1,000, and 10,000 eigenmodes, respectively. The cortical thickness projections (first column) display the thickness for a single exemplary individual.

As shown in Figure 2A, the reconstruction accuracy of individual cortical thickness phenotypes improves as more eigenmodes are included. While as few as 400 low-frequency modes ($\sim 0.7\%$ of the total dimensions) capture 50% of the signal energy in cortical thickness maps, incremental inclusion of up to 2000 modes ($\sim 3.4\%$ of the total) captures $> 80\%$ of signal energy, lowering reconstruction error to $SMSE < 0.2$ (see Section on evaluation metrics). Improvements in accuracy beyond 2000 modes are marginal and plateau. Cortical projections of a participant's thickness map visually demonstrate that low-frequency modes primarily capture spatially smooth, global features, while higher-frequency modes add finer spatial details to the reconstruction.

Similarly, eigenmodes parsimoniously reconstruct various types of spatial query maps used in multi-scale normative modeling, including whole-brain, regional, and high-resolution queries. Whole-brain queries represent regions of interest spanning a large extent of the entire cortex, such as functional networks; regional queries correspond to parcels from a brain parcellation; and high-resolution queries are focused on a specific brain vertex (see Figure 2, and Section 4.9 Model Evaluation for details on these categories).

Fewer modes are needed to accurately reconstruct whole-brain maps than for regional or high-resolution maps. Including the first 100 modes ($\sim 0.2\%$ of the total) captures around 52.6%, 33.3%, and 14.9% of the signal energy for whole-brain, regional, and high-resolution maps, respectively. By using 1000 modes ($\sim 1.7\%$ of the total), approximately 85.5%, 82.1%, and 78.1% of the signal energy is captured for the respective query families, after which reconstruction error reduces to $SMSE < 0.2$ and plateaus. Cortical projections confirm that including higher frequency modes (above $K = 1000$) yields minimal visual improvements, mainly refining sharp transitions such as region borders. Supplementary visualizations are provided to further explore reconstruction residuals and assess the sensitivity of these findings to spatial granularity and cortical asymmetry (see Supplementary Figures S.12, S.5, S.8).

2.3 SNM Achieves Direct Model Accuracy with Adequate Modes

We leverage eigenmodes to approximate the normative ranges of any region of interest from its spectral encoding. SNM utilizes normative models trained on the spectral coefficients of cortical phenotypes (eigenmode loadings) and their cross-dependency structure to estimate the normative range for arbitrary normative queries (see Section 4.8.3 Spectral Normative Model (SNM) for methodological details). We evaluate SNM's performance (with different numbers of modes: $k = \{10, 10^2, 10^3, 10^4\}$) in reconstructing normative ranges for brain-wide, regional, and high-resolution thickness queries. These results are compared to a direct model trained on the exact cortical phenotype observed for each respective query.

Unlike the direct model, SNM relies solely on low-dimensional spectral approximations and is trained without prior information about the spatial extent of the query. Both models were trained on the same sample of 1,978 healthy individuals, covering a wide age range from 5 to 95 years (see Section 4.1 Brain Imaging Data and Supplementary Table S.1 for details). Model performance was assessed on an independent held-out sample of 495 individuals. Performance was evaluated using mean absolute error (MAE) to measure central tendency and mean standardized log-loss (MSLL) to assess the accuracy in modeling both normative means and deviations (see Section 4.9 Model Evaluation for details). Both models operate under the assumption that, after adjusting for covariate effects, the population distribution of each cortical thickness phenotype conforms to a Gaussian distribution; $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ (see Section 4.8 Normative Modeling Framework).

We evaluated the performance of SNMs with varying numbers of modes against direct models, testing both approaches across three spatial scales of normative queries: brain-wide, regional, and high-resolution. As shown in Figure 3, while SNM with only 10 modes shows substantially inferior performance across most queries, compared to the direct model. With as few as 100 modes, SNM yields valid normative estimates for the majority of brain-wide signals. However, most regional and high-resolution normative ranges require a higher number of modes for accurate normative estimation. Indeed, including at least 1000 modes in SNM (less than 2% of the total signal dimensionality) yields normative estimates that match the performance of direct normative models across all spatial scales. Adding more modes—up to 10,000—yields only marginal improvements, suggesting that the first $\sim$ 1000 modes sufficiently capture the bulk of normative cortical thickness information. Supplementary analyses further explore the spatial distribution of SNM performance across brain regions, as well as

the effects of query symmetry and granularity (see Supplementary Section A.3 Cortical projections of normative performance metrics and Supplementary Section A.4 Sensitivity Analyses). These findings underscore the robustness of SNM in reliably inferring normative ranges across different spatial scales, particularly when using at least 1000 modes.

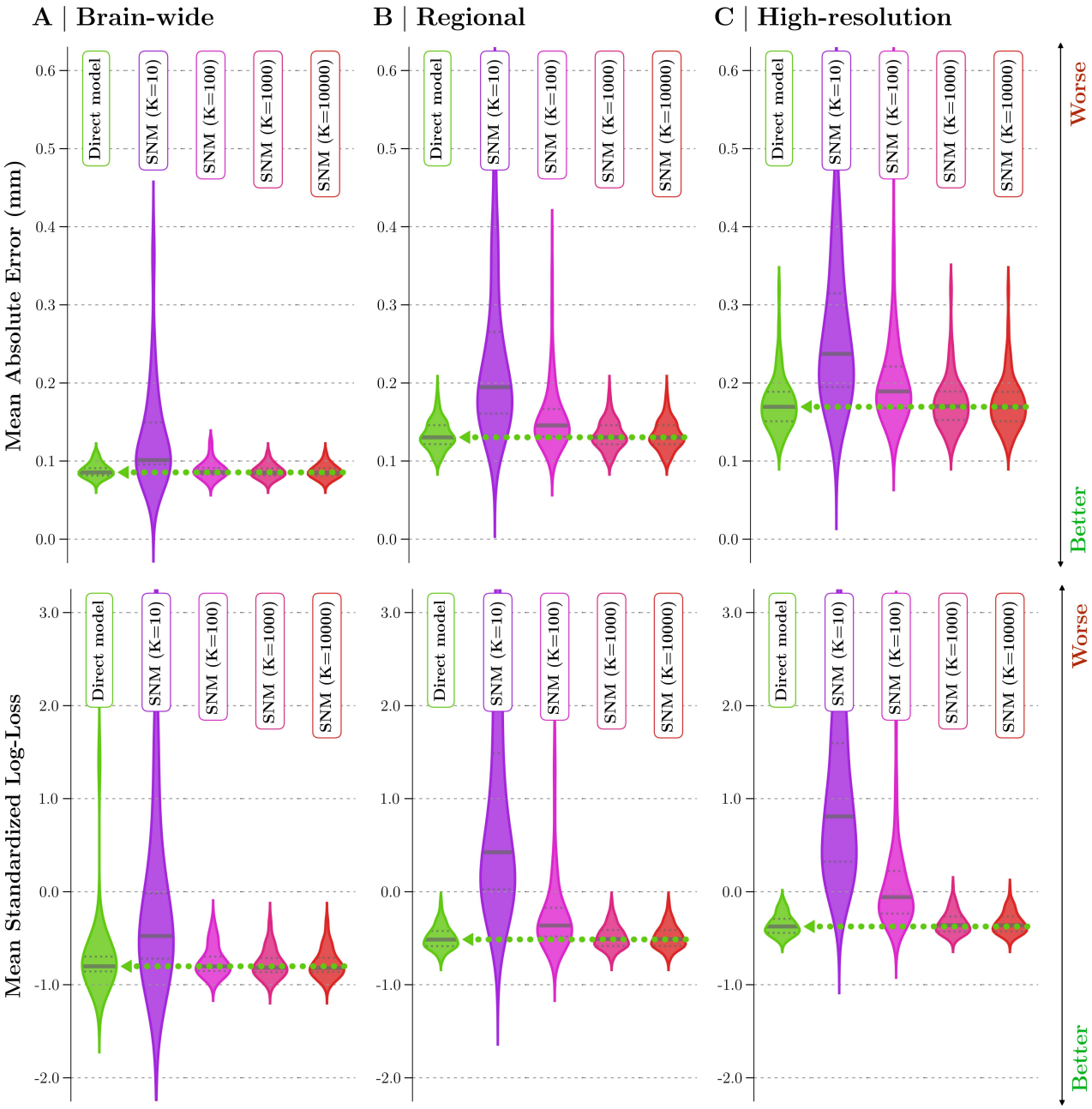


Figure. 3. SNM Normative Performance. This figure compares performance metrics of the SNM at various number of modes ($k = 10, 10^2, 10^3, 10^4$) against a direct normative model. Performance is evaluated across three scales of **(A)** brain-wide, **(B)** regional, and **(C)** high-resolution spatial queries. The rows display performance in modeling the mean (mean absolute error, MAE, top row) and the overall shape of the normative distribution (mean standardized log-loss, MSLL, bottom row). Lower values indicate better performance for both metrics. Green violin plots represent the direct model (benchmark), and SNM performance is shown in shades from purple to red for different numbers of modes. The distributional variation in the violin plots illustrates the performance variability across different spatial queries within each spatial scale. Solid and dashed lines mark the median and first/third quartiles, with a green arrow denoting the direct model's median for reference. In all evaluations, SNMs with at least 1000 modes achieve performance comparable to the direct model.

2.4 SNM Enables High-Resolution Normative Brain Charting

SNM not only achieves accuracy comparable to direct models but does so with significantly lower computational costs. This presents notable practical advantages, such as efficiently computing high-resolution normative ranges for cortical characteristics and deriving individualized normative estimates.

Figure 4 highlights a practical application in which SNM is used to derive high-resolution normative ranges of cortical thickness and chart individual thickness deviations. By utilizing SNM with $K = 1000$ modes, we derive normative estimates of cortical thickness across all cortical vertices using high-resolution queries centered on each respective vertex. This yields smooth high-resolution normative estimates of cortical thickness distribution (mean and standard deviation, as illustrated in Figure 4A), adjusted for age, sex, and site effects (see ???).

Notably, these estimates are computed in a fraction of the time required by the direct approach (Figure 4B). This efficiency improvement encompasses both steps of model training and assessment (i.e. generating normative centiles for a specific set of covariates). SNM significantly outperforms the direct model in terms of computational efficiency. Specifically, SNM eliminates the need to train separate models over tens of thousands of cortical vertices, a process that would otherwise require months to complete. Instead, by modeling the norms across the leading 1,000 eigenmode coefficients, the training time is reduced to less than 2 days. This similarly affects assessment times, reducing the time required to infer high-resolution ranges of a particular set of covariates (age, sex, and site) from minutes to seconds. Crucially, SNM can produce estimates for any arbitrary region of interest, a capability that is impossible with the direct model. As shown in Figure 4C, these high-resolution assessments enable the inference of personalized thickness deviations in individual participants, elucidating specific spatial patterns where cortical thickness deviates from healthy norms.

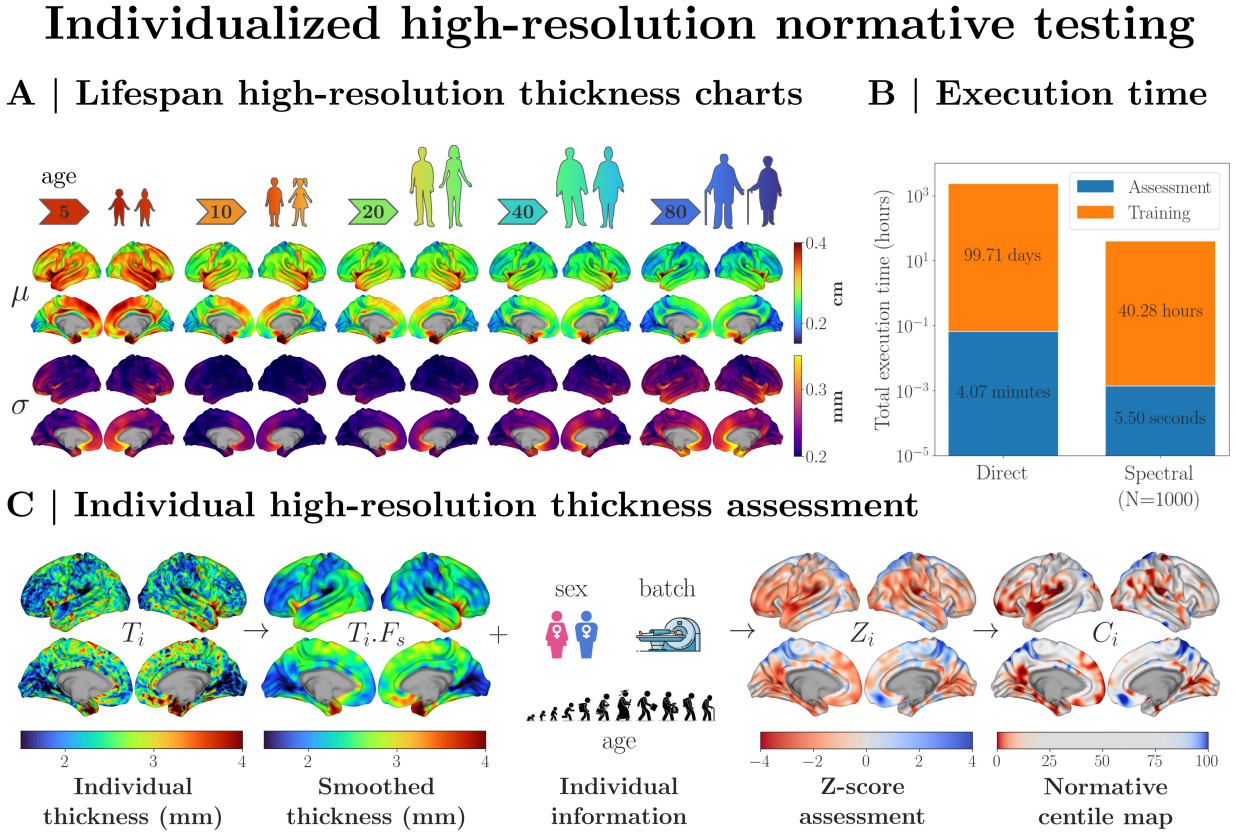


Figure. 4. Application of Spectral Normative Assessments for Personalized High-resolution Normative Testing. **(A)** SNM can extract high-resolution lifespan charts of healthy cortical thickness changes. The model can provide estimates of normative vertex-wise thickness distribution moments (mean and deviation). **(B)** Execution times for training and assessments of SNM with 1,000 modes are compared to a hypothetical implementation using separate vertex-wise direct models. Times are displayed on a logarithmic scale due to the magnitude of differences; on a linear scale, the bar indicating SNM's performance would be nearly imperceptible due to its significantly smaller execution time. **(C)** The high-resolution normative charts can be used for personalized assessments of individual brain scans. The cortical projections represent an exemplary individual from the test sample. Individual thickness values are smoothed using a selected kernel, and the high-resolution moments estimated by SNM are used to create individualized normative maps, indicating deviations quantified via high-resolution Z-scores or centile maps.

2.5 SNM Uncovers Cortical Signatures of Atrophy in Alzheimer's Disease

Next, we demonstrate the practical utility of SNM by applying it to an independent clinical sample. Specifically, we analyze deviations in cortical thickness in an independent imaging dataset comprising three elderly cohorts of individuals with no cognitive impairment (healthy controls, HC, N=132), mild cognitive impairments (MCI, N=202), and Alzheimer's Disease (AD, N=208) (see Section 4.1 Brain Imaging Data and Supplementary Section A.1.2 MACC Clinical Data for details about these cohorts). To leverage the transfer learning capabilities of SNM, we utilize the model trained on the large healthy dataset and fine-tune only the harmonization parameters. Consequently, the model is adapted to this independent dataset and can identify deviations in cortical thickness at high spatial resolution (i.e., vertex-wise z-scores, as described in the previous section).

These deviation maps offer a powerful tool for investigating the normative thickness changes associated with AD. By spatially comparing deviation maps between healthy individuals and those diagnosed with AD, we uncover a cortical signature of structural atrophy in AD (Figure 5A). This demonstrates reduced cortical thickness in several neocortical regions spanning the temporal, parietal, and frontal lobes, indicating widespread atrophy associated with AD. While most abnormal differences indicate a thinner cortex in AD, we also find abnormally thick gray matter in certain visual areas along the lingual gyrus and cuneus. We also assess the relationship between cortical atrophy and cognitive impairment by examining Mini-Mental State Examination (MMSE) scores in relation to individualized normative assessment maps. Our results show significant effects linking cognitive impairment to widespread cortical thinning. These effects are highest in temporal regions but span several cortical areas (Figure 5B).

Findings of Figure 5A,B are consistent with previous studies and are therefore not claimed as a novel achievement of SNM but are reported to demonstrate its validity in replicating established findings. SNM can additionally generate individualized predictive cortical atrophy biomarkers through extreme value statistics. Specifically, we quantify the number of cortical vertices exhibiting extreme thinning for each individual ($z < -1.96$, see Section 4.10.2 Linking Atrophy to Cognitive Impairment for details). This extreme value statistic, termed Extremely Thin Vertex Count (ETVC), serves as a robust biomarker for cognitive impairment in AD. Namely, we find that individuals with more

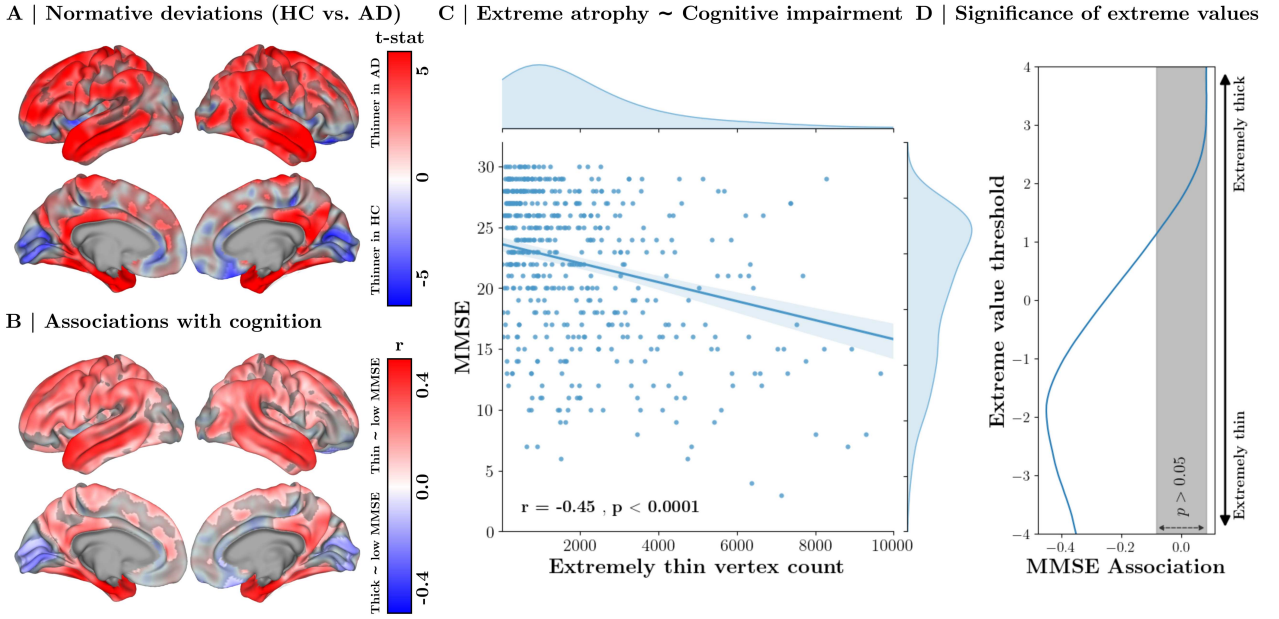


Figure. 5. Cortical Signature of Atrophy in Alzheimer's Disease and Its Cognitive Correlates. High-resolution deviation maps were used to compute the cortical signature of atrophy in an elderly clinical cohort and assess its ability to predict cognitive impairments that are associated with AD. **(A)** Group-level normative differences between HC and AD. **(B)** Vertex-wise associations between normative deviation z-scores and cognitive performance (MMSE). **(C)** The ETVC metric, quantifying extreme atrophy, can predict cognitive performance in the clinical cohort. **(D)** Comparison of z-score thresholds for ETVC reveals that vertices with extreme atrophy provide the highest predictive power for cognitive impairment. Cortical projections highlight significant regions (dimmed for non-significant voxels) at $\alpha = 5\%$ after FDR correction. Abbreviations: HC: Healthy Cohort, AD: Alzheimer's Disease, MMSE: Mini-Mental State Examination, ETVC: Extremely Thin Vertex Count, FDR: False Discovery Rate.

pronounced cortical thinning (higher ETVC) are at a significantly higher risk for cognitive impairment ($r = -0.45, p < 0.0001$, Figure 5C).

Supplementary evaluations indicate that this association is reproducible within the AD cohort ($r = -0.31, p < 0.0001$), but not within the healthy ($r = -0.005, p = 0.96$) or MCI ($r = -0.10, p = 0.14$) cohorts (see Supplementary Section A.6 Within-group Cognitive Associations). As shown in Figure 5D, we find that regions exhibiting extreme atrophy ($-3 < z < -2$) contribute more significantly to the predictive power of ETVC biomarker compared to regions with mild atrophy ($-2 < z < 0$) or regions with hypertrophy ($z > 0$). In supplementary analyses, we find an inferior predictive power for the normative z-score of mean cortical thickness compared to the high-resolution ETVC biomarker (see Section A.7 Brain-wide Cognitive Associations).

2.6 SNM Highlights Heterogeneity in Individual Cortical Atrophy

While group-level analyses provide valuable insights into the average patterns of cortical atrophy associated with AD, they unavoidably obscure important interindividual variability. Group means can mask unique patterns of atrophy that are highly personalized, which may hold crucial implications for understanding disease mechanisms and tailoring clinical interventions. Complementing the group-level evaluations presented in the previous section, SNM enables the investigation of individualized deviation patterns. This facilitates a deeper exploration of the heterogeneity in cortical atrophy among individuals with AD. As illustrated in Figure 6A, two individuals with similar demographic characteristics, extreme thinning profiles, and identical diagnoses can exhibit markedly distinct cortical thinning patterns, with minimal spatial overlap ($r = 0.05, p_{spin} = 0.24$). Importantly, SNM allows these deviation maps to be projected back into each individual's native brain space, whether in surface or volumetric format, providing clinically interpretable atrophy maps that can potentially aid in personalized cortical atrophy assessments.

Differences in atrophy patterns may reflect distinct mechanisms through which pathology associates with normative deviations. To illustrate this, we simulated four hypothetical scenarios (Figure 6B), each showing differences between 60 individual normative deviation maps. Pairwise differences between maps were quantified by Euclidean distance to measure interindividual variability. In the first scenario, deviations occur in a random, non-systematic manner (Figure 6B, left). In the second scenario, deviations are tightly linked to diagnostic categories (Figure 6B, middle left, rows ordered by diagnosis group). In the third scenario, deviations are linked to disease severity, with individuals of similar severity exhibiting similar atrophy patterns (Figure 6B, middle right, rows ordered by severity). In the final scenario, increasing severity leads to greater heterogeneity in deviations (Figure 6B, right, rows ordered by severity).

We next compute the empirical interindividual difference matrix (Figure 6C) and sort it by diagnostic group (HC, MCI, and AD) and, within each group by cognitive performance (as a marker of symptom severity). We find that healthy controls and MCI patients exhibit relatively smaller differences in deviation patterns. However, individuals with AD show significantly greater variability, both when compared to healthy and MCI cohorts, as well as amongst themselves. These findings suggest that AD is associated with increasingly divergent cortical atrophy patterns across individuals, highlighting the heterogeneous nature of individual atrophy in AD.

To further elucidate the heterogeneity in deviation patterns, we embed the high-dimensional normative assessments into a lower-dimensional latent space that maintains the difference structure (Figure 6D, see Section 4.10.3 Examining Interindividual Heterogeneity for details). This reveals that healthy individuals form a dense central cluster, characterized by more homogeneous deviation maps. In contrast, the MCI group shows a more dispersed distribution, with greater variability in deviation patterns. The AD cohort exhibits the highest degree of deviation from the healthy cluster, with substantial diversity in the direction of deviations across individuals. This underscores various ways in which AD-diagnosed individuals deviate from normative cortical structure. When averaging deviation patterns within local neighborhoods of this landscape (cortical projections in Figure 6D, left), we observe distinct subgroups characterized by negligible (top), localized (middle), or widespread (bottom) atrophy (see Supplementary Section A.8 Heterogeneity Landscape Subgroups to further explore these local neighborhood projections). Exemplary cortical projections in Figure 6D (right) reiterate that AD-diagnosed individuals vary not only in the severity of cortical atrophy but also in the spatial

distribution of these deviations.

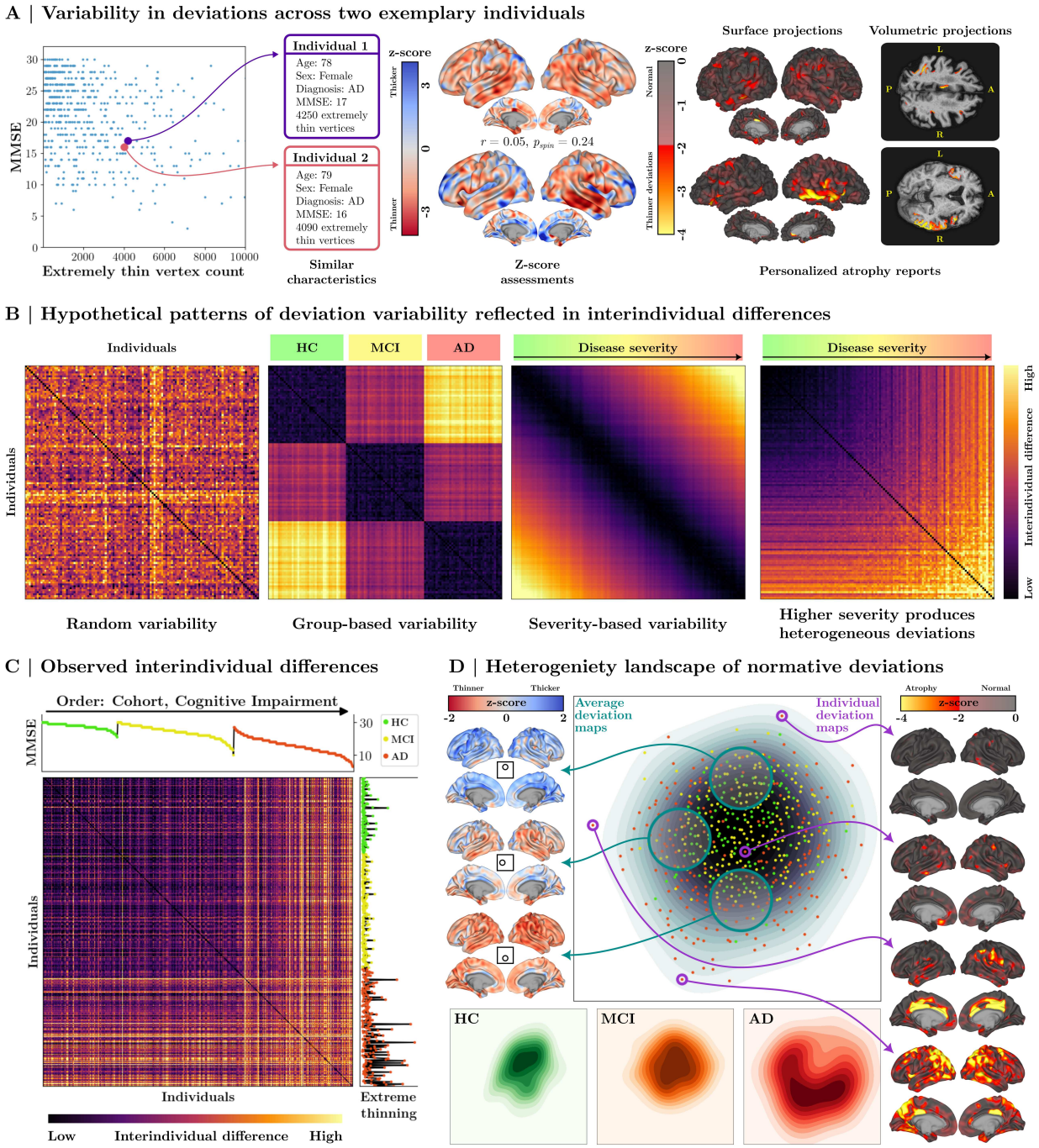


Figure. 6. Heterogeneity Landscape of Atrophy Associated with Alzheimer's Disease. Individualized high-resolution deviation maps were used to assess the extent of heterogeneity in AD-related normative deviations. **(A)** Examples of two individuals with similar age, sex, cognitive scores, and diagnosis, who nevertheless display markedly different patterns of cortical atrophy. These maps can be overlaid onto native scans as summary reports to assist assessments of cortical atrophy. **(B)** Interindividual differences in normative deviations are quantified using Euclidean distance between assessment maps. Four hypothetical heatmaps illustrate how the structure of interindividual differences depends on the underlying deviation mechanisms. Deviations can be completely random and unrelated to severity (left), strictly delineated by diagnostic groups (middle left), uniformly progressive across disease stages (middle right), or display heterogeneous divergence from norms (right). **(C)** An empirical interindividual difference matrix is computed from the clinical cohort's deviation maps, with individuals sorted by clinical diagnosis and, within each diagnosis, by cognitive performance (from high to low MMSE). This distance matrix suggests that AD-diagnosed individuals exhibit heterogeneous normative deviations. **(D)** High-resolution normative assessments are projected onto a 2-dimensional landscape, while preserving interindividual difference structure. The central scatter plot displays the distribution of individuals from HC (green), MCI (yellow), and AD (red) cohorts within this landscape, with density plots (bottom) highlighting regions predominantly occupied by each cohort. Cortical projections (left) show average deviation maps for three exemplary local areas of this landscape, while exemplary individual deviation maps are shown (right) for four AD-diagnosed individuals.

3 Discussion

In this manuscript, we introduce SNM, a novel method to enhance the efficiency and versatility of normative brain charts. Our results demonstrate that SNM improves spatial specificity, reduces computational complexity, and maintains accuracy comparable to conventional, direct assessments across a broad spectrum of spatial normative queries. SNM shows considerable promise for precision brain charting, advancing the capabilities of existing approaches in modeling trajectories, analyzing group-level pathological deviations, and generating individualized profiles of brain anomaly characteristics.

3.1 Computational Efficiency

While direct models require re-training for each new query, SNM can provide accurate normative estimates for novel, a posteriori-defined queries without additional training. This results in significant time efficiency improvements. For example, as shown in Figure 4, high-resolution SNM assessments can be applied to different scales of spatial smoothing within seconds, without retraining the model. In contrast, direct models require retraining for every vertex when applying an alternative smoothing, a process that can take several months. SNM hence offers several orders of magnitude improvement in execution time, reducing it from months to seconds. This efficiency is particularly advantageous for computing vertex-wise normative maps and studying psychopathological deviations across large samples.

3.2 Spatial Versatility

Another key advantage of SNM is its ability to estimate a wide range of normative queries. The model can theoretically reconstruct an infinite number of queries through linear combinations of eigenmode basis functions, effectively spanning all signals that can be approximated by the low-pass graph filter. We demonstrate that this encompasses conventional normative queries used to chart cortical phenotypes. Additionally, our supplementary evaluations show that this extends to assessing other norms, such as cortical asymmetry, which can help detect abnormalities in cortical lateralization. Importantly, this means that a single pre-trained spectral normative model can generate regional normative charts for various brain parcellation schemes without additional cost. Given the multitude of existing brain atlases[36–45], each offering particular advantages, SNM provides the flexibility to choose an atlas that best fits any specific research needs.

In addition to accommodating flexible regional queries, our results demonstrate that spectral normative models can generate high-resolution normative maps for smooth spatial queries, adhering to the low-pass reconstruction criterion. This means that if no regional brain atlas is preferred, the model can alternatively produce normative estimates at the resolution of acquired brain images (voxels or vertices). This is particularly useful for individual charting applications. By enabling efficient atlas-free normative assessment of individual brain characteristics, SNM provides high-resolution maps that pinpoint localized spatial deviations. This increased resolution can enhance the power and sensitivity to study heterogeneity in pathology-driven deviations from healthy norms[1]. Finally, this spatial versatility also enables on-demand assessment of individualized normative queries. For instance, by identifying patient-specific functional networks through individualized parcellation techniques[46] and subsequently performing SNM on those individualized regional queries, we can directly assess whether particular behavioral differences are linked to underlying neuroanatomical deviations, offering a valuable tool for personalized clinical evaluation.

3.3 Accessibility Benefits

The concept of an a posteriori-defined query is particularly compelling, as it enables the use of large-scale datasets while safeguarding the privacy of the original data sources. Conventionally, utilizing normative assessments for a user's specific spatial query required access to an extensive set of imaging biobanks to train the normative model. Traditionally, performing normative assessments for arbitrary spatial queries required direct access to extensive imaging biobanks to train the normative model. In contrast, SNM decouples model training from query assessment, thereby eliminating the need for end-users to access the training data itself (see Figure 1).

Access to many large-scale imaging biobanks is subject to time-consuming data access procedures[47–50]. This limits the feasibility of community-wide normative modeling studies that require multiple datasets, as users may lack the necessary permissions to train the model from scratch. SNM overcomes this hurdle by enabling the creation of adaptable and openly accessible normative charts for brain development. Once a model is trained, posterior spectral estimates of normative moments can be shared, allowing assessments using the pre-trained model. This approach democratizes access to large-scale normative models and facilitates broader adoption in both research and clinical settings, particularly for end-users with limited data access.

As demonstrated by our evaluations on an independent clinical cohort, SNM can generalize effectively to previously unseen data from new sites, even with relatively small sample sizes, while appropriately modeling site-specific effects. This highlights SNM's ability to leverage transfer learning, applying knowledge gained from large-scale cohorts to smaller datasets[51,52]. Furthermore, SNM's architecture is well-suited for federated learning implementations, allowing the model to be trained on decentralized datasets across multiple sites[53]. These features position SNM as an efficient tool for studying clinical populations, enhancing the practical utility of normative modeling in situations where data collection and sharing is constrained.

3.4 Clinical Applications

In addition to systematic benchmarking and performance evaluations, we demonstrated SNM's practical applications for high-resolution atrophy detection in a clinical cohort. Neurodegenerative mechanisms, such as amyloid-β plaque deposition and tau-related neurofibrillary tangles, are known to be associated with cortical atrophy in AD [54–57]. Given that signs of cortical atrophy can emerge up to a decade before AD symptoms appear [58,59], SNM's high-resolution assessments can facilitate early detection and characterization of neurodegeneration, enabling timely intervention. Our approach shows that a pre-trained SNM can produce individualized, high-resolution deviation maps, presenting a valuable biomarker for neurodegeneration that translates normative modeling insights into clinical applications[60].

Our analysis supports previous findings by reproducing hallmark group-level cortical thinning patterns of AD, with the strongest reductions in cortical thickness centered on the temporal lobe (temporal pole, superior, middle, and inferior temporal gyri) [57,61–66]. These effects also extend to regions of the neocortex, including the posterior cingulate gyrus [64,67], supramarginal and angular gyri [61,62,64,65], superior parietal lobule, superior, middle, and inferior frontal gyri [61–64,68], posterior cingulate cortex [64,66,68], and the precuneus [57,61,62,64,68]. Additionally, other less established anomalies emerged, such as increased cortical thickness in certain visual and orbitofrontal areas[69] that are potentially linked to compensatory neurodegenerative mechanisms[70]. Notably, our evaluations indicate that extreme atrophy quantified by ETVC at a higher threshold is a stronger predictor of cognitive impairment than mild atrophy, underscoring the utility of high-resolution normative assessment.

Our evaluations also emphasize the heterogeneity in individual cortical atrophy patterns, revealing diverse, pathology-driven deviation profiles. While group-level studies provide insights into common atrophy signatures, they may obscure individual-specific deviations[71]. This heterogeneity may reflect comorbidities[65,72–75], which influence each person's atrophy pattern across a varied landscape. Studies typically address this with clustering approaches to categorize pathological cohorts into common atrophy subtypes[64,66,76–80]; however, our findings reiterate that AD-related cortical deviations remain notably heterogeneous[81].

Clustering methods, which group individuals into subtype patterns, may also obscure unique deviations when averaging patterns within subtypes, limiting the precision of subtype assignment for individuals[82]. We observed that some AD-diagnosed individuals exhibited negligible signs of atrophy, aligning with previous findings of subtypes characterized by minimal or no atrophy[56,64,66,79]. Conversely, some HC/MCI individuals showed significant atrophy without severe cognitive decline, possibly indicating markers of brain resilience[83,84]. These findings suggest that analyzing individual atrophy patterns along continuous principal or latent dimensions[85] offers deeper insights into neurodegenerative variability than traditional subtype classifications, while also facilitating the study of resilience mechanisms in brain aging.

While these clinical findings underscore the potential of SNM for early detection and characteriza-

tion of atrophy patterns, they remain preliminary. Future studies focused on specific clinical applications are necessary to fully assess SNM's role as a diagnostic and prognostic aid in clinical settings and its broader potential in personalized medicine. For example, SNM could enhance diagnostic accuracy and aid in differential diagnosis, with deviation maps serving to distinguish between various types of dementia, such as Alzheimer's disease, vascular dementia, Lewy body dementia, and frontotemporal dementia [86]. Additionally, SNM shows promise as a prognostic tool; normative deviation maps in stroke or TBI patients could help evaluate the extent of damage to functional brain regions, allowing for more personalized rehabilitation plans and improved treatment monitoring. Finally, SNM's heightened sensitivity to subtle, localized atrophy patterns may facilitate the detection of early biomarkers, such as microinfarcts, which often remain undetected but could identify individuals at risk, offering opportunities for preventive interventions before more severe events occur [87].

3.5 Future Directions and Limitations

The present evaluations demonstrate the feasibility of SNM as a flexible normative modeling framework with the potential for numerous applications in future research. An immediate next step could involve training SNM on a large sample of brain imaging scans, comprising hundreds of thousands of scans from multiple biobanks. This would support the development of a comprehensive, open-access database of spectral normative brain charts.

While the current study employs eigenmodes defined over the cortical surface mesh, future research could extend SNM to volumetric template spaces. By extracting voxel-wise brain eigenmodes[88], SNM can also capture volumetric norms. Additionally, this framework could be applied to explore normative trajectories of high-resolution brain phenotypes across different modalities. For example, quantitative measures of high-resolution structural volume[13], white matter microstructure[89–91], functional organization[92,93], or metabolic processes quantified by Positron Emission Tomography[94] could serve as alternative normative phenotypes of interest. Such explorations would contribute to a deeper, multifaceted understanding of healthy brain development and organization.

Several limitations warrant consideration. First, the current spectral model assumes Gaussianity, modeling phenotypes as normally distributed across the population. This model does not extend to phenotypes with substantial skewness or kurtosis. Although handling non-Gaussian distributions is beyond the scope of this paper, this limitation could be addressed in future work by integrating SNM with recent advancements in likelihood warping techniques[16,91] to model non-Gaussian distributions.

Second, SNM assumes that, with an adequate number of low-frequency eigenmodes, the phenomenon under evaluation (e.g., cortical thickness) is well-captured by the spectral basis set. While our study has demonstrated this for cortical thickness (Figure 2), the utility of this approach for other modalities and phenotypes depends on whether this assumption holds true. Similarly, the spatial query under study should be restricted to a low-pass graph spectral regime. Future applications of this method for different phenotypes or novel queries should include tests of reconstruction accuracy to ensure the validity of the spectral basis set for those specific contexts.

Lastly, this study utilized a spatial basis set constructed from connectome eigenmodes via singular value decomposition of the random walk Laplacian shift operator. However, spectral normative modeling can be applied using various orthonormal basis sets. For example, eigenmodes can be derived from different brain features (e.g., geometry[95], diffusion[25,27], or function[96,97]) or other shift operators (e.g. combinatorial Laplacian, symmetric normalized Laplacian, or the fractional Laplacian)[98,99]. We chose to use connectome eigenmodes due to their ease of extension across both cortical hemispheres while aligning with anatomical and geometrical brain landmarks, and the random walk Laplacian shift operator for its smoothness and robustness to degree distribution [100]. Nevertheless, it should be noted that SNM is theoretically generalizable to any orthogonal basis set for information reconstruction and is not limited to brain eigenmodes.

4 Methods

In the ensuing sections, we first describe the brain imaging datasets utilized in this work. We provide a formal description of the high-resolution cortical thickness phenotypes used as input to the normative modeling framework. Next, we introduce the mathematical foundations of SNM, covering core concepts

from graph signal processing and normative modeling. Finally, we describe experiments undertaken to evaluate SNM's performance and suitability.

4.1 Brain Imaging Data

This study uses three publicly available healthy brain imaging datasets that were provided and curated by the Human Connectome Project (HCP)[101]; namely, HCP's Young Adult[102], Aging[103], and Development[104] cohorts. These cohorts collectively comprised 2473 healthy individuals (54.7% female) spanning a wide age range, from 5 to 100 years old. The imaging acquisition and preprocessing followed previously established pipelines and were broadly consistent across the three cohorts[105,106]. In addition to the healthy datasets used to train the normative model, we evaluated the model's translational capabilities to real-world data from memory clinics (Memory, Ageing & Cognition Centre at the National University of Singapore: MACC)[107,108]. This data comprised 542 samples (ages 50–91, 61.4% female), including 132 cognitively healthy individuals (24.3%), 202 individuals with mild cognitive impairment (MCI; 37.3%), and 208 individuals diagnosed with Alzheimer's Disease (38.4%) based on DSM-IV criteria[108]. Across all datasets, FreeSurfer processing of the T1-weighted (T1w) structural brain imaging data was used to extract high-resolution maps of cortical thickness. Additional details regarding acquisition and preprocessing procedures for each dataset are provided in the Supplementary Information (see Section A.1 Image Acquisition and Preprocessing).

4.2 Train Test Split

Healthy individuals were stratified into train and test splits to enable validation of the fitted models on unseen data. Namely, 80% (N_{train}=1978 individuals) of the healthy data were used to train the model, and the remaining 20% (N_{test}=495) were used for out-of-sample validation ($N_p = N_{train} + N_{test}$). A single split fold was used, with a randomized approach that controlled for covariate distributions of age, sex, and dataset by stratification; this ensured that the covariates had similar distributions across splits[109]. Supplementary Figure S.1 provides a visual summary of covariate distributions before and after splitting.

4.3 Cortical Thickness

For all cohorts, cortical thickness estimates for each vertex comprising the cortical surface mesh was sourced from preprocessed data. HCP's minimal processing pipeline[105] contains procedures utilizing FreeSurfer outputs[110] to provide high-resolution thickness estimates for the fs-LR 32k surface mesh template. This template surface uses 32,492 vertices to model each hemisphere, resulting in an average inter-vertex distance of approximately 2 mm; it includes a total of 59,412 vertices after the exclusion of the medial wall. Vertices are aligned between the left and right hemispheres, enabling inter-hemispheric comparison)[102]. The same transformation was used to project MACC sample's FreeSurer cortical thickness estimates to the fs-LR surface space.

In this space, cortical thickness is represented as a high-resolution vector $T_i \in \mathbb{R}^{1\times N_v}$ ($N_v = 59,412$), providing a thickness estimate for each vertex of the ith individual's cortical surface. The collection of these high-resolution thickness features across all participants forms the complete high-resolution thickness data sample $T \in \mathbb{R}^{N_p\times N_v}$ which is used to estimate and evaluate normative ranges of thickness variation across the human lifespan. These data are divided into a training sample $T_{train} \in \mathbb{R}^{N_{train}\times N_v}$ used to construct the normative model and an independent test sample $T_{test} \in \mathbb{R}^{N_{test}\times N_v}$ used to evaluate the normative model's goodness of fit.

4.4 Normative Covariates

In normative modeling, the aim is to provide a statistical model that describes the distribution of a variable of interest y with respect to a set of covariates C (fixed effects), while potentially accounting for batch effect Z (random effects)[111]. For simplicity, we consider the covariates of age and sex, although other covariates can be straightforwardly incorporated. Moreover, batch effects were included in the model to facilitate harmonization among the three imaging cohorts (development, young adult, and aging).

4.5 Clinical Fine-tuning

The normative model, pre-trained on N_{train} healthy individuals, was subsequently fine-tuned for the clinical sample of $N_{MACC} = 674$ individuals using transfer learning. During fine-tuning, the model parameters related to age and sex were kept frozen to maintain consistency with the initial healthy cohort model. However, site parameters for dataset harmonization were fine-tuned on a subsample of 66 healthy individuals (50% of the HC cohort) from the MACC clinical cohort. The fine-tuned model was then applied to the rest of the clinical sample to generate individualized, high-resolution normative deviation maps. These maps were utilized in normative evaluations of the clinical cohort.

4.6 High-resolution Connectomes

Our proposed approach uses brain connectivity eigenmodes computed from a reference brain connectivity structure[23]. Connectivity eigenmodes were selected for their ability to span both cortical hemispheres, enabling a cohesive representation of interhemispheric relationships. In contrast, geometric modes analyze hemispheres separately, overlooking interhemispheric interactions. Additionally, as many brain disorders with structural deviations from normative patterns are hypothesized to express spatial signatures aligned with anatomical connectivity, connectome eigenmodes are optimally suited to capture such abnormalities[28–33]. High-resolution structural connectivity data were utilized for this purpose[35].

To this end, we sourced connectomes computed from a tractography pipeline detailed elsewhere[112]. In brief, diffusion-weighted imaging data were used to estimate structural connectivity. Probabilistic tractography was conducted using MRtrix3 to reconstruct whole-brain tractograms (5 million streamlines, with anatomically constrained tractography). Tractography endpoints were used to construct connectomes encoding streamline count at the resolution of the fs-LR 32k template (same space as the thickness information, with $N_v = 59,412$ vertices representing network nodes). Individual connectomes were combined to form an estimate of group-level connectivity. Connectome spatial smoothing was performed to account for endpoint inaccuracies and improve connectome reliability[113,114]. This yielded a high-resolution weighted adjacency matrix $\mathbf{A} \in \mathbb{R}_{+}^{N_v \times N_v}$ in which element $A_{i,j}$ denotes the group average strength of structural connectivity between nodes i and j. The eigenmodes resulting from this high-resolution connectome mapping pipeline were previously tested to assess their accuracy in encoding brain signals[23].

4.7 Brain Signal Reconstruction

Graph signal processing enables the study of data encoded on a graph/network structure[115]. As such, considering that the human brain is fundamentally a network structure, graph signal processing can be utilized to design new ways to analyze and study the brain[116]. Here, we utilize dimensionality reduction and signal reconstruction techniques that encode cortical information in a lower-dimensional latent space[117,118]. To this end, we use high-resolution connectomes as a weighted graph $\mathcal{G} : (\mathcal{V}, \mathbf{A})$ where $\mathcal{V} = \{1, 2, \cdots, N_v\}$ is the set of high-resolution cortical nodes (vertices) and $\mathbf{A} \in \mathbb{R}_{+}^{N_v \times N_v}$ is a weighted (non-negative) adjacency matrix. We utilize GSP to model any brain signal $x \in \mathbb{R}^{N_v \times 1}$ that is defined on the set of vertices $\mathcal{V}$. In this context, we model vertex-wise cortical thickness estimates as a brain signal.

The random-walk Laplacian matrix $L_{rw} = I - D^{-1}\mathbf{A}$ is used as the graph shift operator (where $D \in \mathbb{R}^{N_v \times N_v}$ denotes the diagonal strength matrix). This shift operator is diagonalized via singular value decomposition $L_{rw} = \Psi \Lambda \Psi^{-1}$ to compute an orthogonal basis for information reconstruction. Here, $\Psi \in \mathbb{R}^{N_v \times N_v}$ denotes the full set of eigenvectors (left singular vectors, also referred to as eigenmodes), and $\Lambda \in \mathbb{R}^{N_v \times N_v}$ denotes the diagonal matrix of eigenvalues (singular values) associated with the eigenmodes such that $\Lambda_{i,i} = \lambda_i$ is the ith smallest eigenvalue and $\Psi_i \in \mathbb{R}^{N_v \times 1}$ (ith column of Ψ) refers to its associated eigenmode. For any brain signal $x \in \mathbb{R}^{N_v \times 1}$, a graph Fourier transform (GFT) can encode the signal in graph spectral domain $\tilde{x} = \Psi^T x$, where $\tilde{x} \in \mathbb{R}^{N_v \times 1}$ denotes the encoded signal, where $\tilde{x}_i$ is the ith element of $\tilde{x}$ which quantifies the loading of the brain signal on the ith eigenmode ($\tilde{x}_i = \Psi_i^T x$). Conceptually, this encoding transforms the brain signal x defined in the spatial domain to a latent embedding $\tilde{x}$ in graph spectral domain. The encoded signal in the graph spectral domain can be mapped back to the spatial domain via the inverse GFT: $x = \Psi \tilde{x} = \Psi \Psi^T x = Ix$. Notably, this

transformation, utilizing the full set of eigenmodes, yields an exact reconstruction of any brain signal. Our goal is to use only a fraction of Ψ to derive a low-dimensional representation of brain signals. We selected the random-walk Laplacian shift operator as it yields eigenmodes encoding topological frequency such that eigenmodes associated with smaller eigenvalues capture signals that vary more smoothly along network edges[100].

Deriving the full set of eigenmodes (Ψ_1 to Ψ_{N_v}) is computationally demanding; however, graph Fourier filtering enables efficient approximation of brain signals using a filter that only requires a limited set of eigenmodes. Specifically, we can define a diagonal filtering matrix $H \in \mathbb{R}^{N_v \times N_v}$, where $H_{i,i} = h(\lambda_i)$ denotes the filter's frequency response to the eigenmode Ψ_i associated with the frequency λ_i. An ideal low-pass filter $H_{(k)}$ with frequency response h_k that discards the information of eigenmodes with graph frequencies higher than λ_k is defined using the step function:

$$H_{(k)_{i,i}} = h_k(\lambda_i) \tag{1}$$

$$h_k(\lambda) = \begin{cases} 1 & \text{if } \lambda \leq \lambda_k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Using this low-pass filter, brain signal x is approximated by $x \approx \hat{x}_{(k)} = \Psi H_{(k)} \Psi^T x = \Psi H_{(k)} \tilde{x}$, where $\hat{x}_{(k)}$ is the low-pass filtered approximation of x reconstructed from the first k eigenmodes. It should be noted that all elements beyond the kth row and column of $H_{(k)}$ are zeros. As a result, this approximation does not require knowledge about higher frequency eigenmodes $\Psi_i : i > k$. Specifically, let $\hat{\Psi}_{(k)} \in \mathbb{R}^{N_v \times k}$ denote the set of first k eigenmodes (i.e. first k columns of Ψ); the low-pass approximation of brain signal x can be achieved as follows:

$$\hat{x}_{(k)} = \hat{\Psi}_{(k)} \hat{\Psi}_{(k)}^T x = \hat{\Psi}_{(k)} \tilde{x}_{(k)} \tag{3}$$

Where $\tilde{x}_{(k)} \in \mathbb{R}^{k \times 1}$ is the k-dimensional encoding of the signal based on the first k eigenmodes in the graph spectral domain. As such, low-pass graph frequency filtering effectively reduces an N_v-dimensional brain signal x to a k-dimensional latent graph frequency representation. We anticipate that the majority of spatial variation in a brain signal such as cortical thickness is captured by a low-pass approximation of adequate bandwidth (i.e. an appropriate choice of k). This provides a tool to control the trade-off between computational complexity (dimensionality) and spatial specificity. Choosing a lower value for k reduces the computational demand but increases simplification and lowers spatial specificity. Selecting a k that gives an optimal balance between computational tractability and spatial specificity is thus important.

4.8 Normative Modeling Framework

In the context of our paper, conventional normative models encompass all univariate approaches in which the normative ranges of a fixed spatial query (e.g., the thickness of a region of interest) are modeled as a function of demographic covariates. SNM is a general framework that is compatible with alternative normative models, such as GAMLSS[2,119], BLR[17,120], and HBR[15,111], etc.). For clarity and without loss of generality, we focus on an exemplary instantiation using Hierarchical Bayesian Regression (HBR)[15]. We begin by formally defining a spatial normative query to specify the region of interest, followed by an overview of HBR (referred to as the direct method, see Figure 1A). We then describe how SNM extends the direct approach by using eigenmodes, facilitating the estimation of arbitrary, a posteriori-defined spatial normative queries.

4.8.1 Spatial Normative Query

The concept of a spatial normative query is new to this work. A researcher may wish to determine, or *query*, a normative range over a specific spatial extent, such as an individual vertex on the cortical mesh, a cortical region or hemisphere, or the entire brain. A spatial normative query denotes a brain signal $x \in \mathbb{R}^{N_v \times 1}$ that is the result of such a query. Typically, this query characterizes averaging the phenotype over the spatial region of interest, although other descriptive statistics could be applied. For instance, if we want to determine the normative range for cortical thickness averaged over the entire

cortex, the spatial query would store the associated weights for the cortex-wide averaging operation, i.e. $x = \mathbf{1}_{N_v} \cdot \frac{1}{N_v}$ where $\mathbf{1}_n$ denotes a unit vector of length n ($\mathbf{1}_n = (1, 1, \cdots, 1)^T \in \mathbb{R}^{n \times 1}$). Similarly, if we are interested in modeling the average thickness of an arbitrary brain region with N_r vertices, the respective spatial query x can be defined as follows (x_i denotes the ith element/row of $x = (x_1, x_2, \cdots, x_{N_v})^T$):

$$x_i = \begin{cases} \frac{1}{N_r} & \text{if vertex } v_i \text{ belongs to the region of interest} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Our definition of a spatial query provides significant flexibility to establish reference ranges across arbitrarily defined regions and networks. For instance, the normative thickness of a functional network can be evaluated using a spatial query with the weights of x proportional to the probability of a vertex belonging to that network (such that $\sum_{i \in \{1,2,\cdots,N_v\}} x_i = 1$). Moreover, this notion enables the assessment of comparative normative questions. For instance, we can assess the difference in average thickness of the left and right hemispheres to assess cortical asymmetry/lateralization:

$$x_i = \begin{cases} \frac{1}{N_{left}} & \text{if } v_i \text{ belongs to the left hemisphere} \\ \frac{-1}{N_{right}} & \text{if } v_i \text{ belongs to the right hemisphere} \end{cases} \tag{5}$$

where N_{left}, N_{right} respectively denote the number of vertices in each hemisphere.

4.8.2 Direct Method

The direct method refers to the conventional normative modeling approach in which a fixed a-priori-defined spatial query is used (e.g. mean thickness). In this manuscript, we utilize HBR as the normative approach used for the direct method[15,111]. Using a fixed spatial query x, the direct method models the expression of the phenotype (e.g. cortical thickness) on the query in the training set, $y \in \mathbb{R}^{N_{train} \times 1}$; which is computed as the inner product between the spatial query x and the high-resolution thickness data T, $y = T_{train} \,.\, x$. HBR assumes that these observations follow a normal distribution $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and aims to model its moments (mean and variance) as a function of covariates (age and sex) while controlling for batch effects:

$$\mu_y = f_\mu(age,\ sex,\ batch) = w_\mu \Phi(age) + \alpha_\mu(sex) + \beta_\mu(batch) + \epsilon_\mu \tag{6}$$

$$\sigma_y = f_\sigma^+(age,\ sex,\ batch) = w_\sigma^+ \Phi(age) + \alpha_\sigma^+(sex) + \beta_\sigma^+(batch) + \epsilon_\sigma \tag{7}$$

In particular, w_μ and w_σ^+ are weights that model the fixed effect of age by a basis expansion Φ. A B-spline basis expansion (cubic spline with three evenly spaced knots; df=5) was used to capture the non-linear effects of age. The fixed effect of sex is modeled by intercepts for mean (α_μ) and deviation (α_σ^+). Batch effects (for dataset harmonization) are modeled as intercepts for the mean ($\beta_\mu \sim \mathcal{N}(\mu_{\beta_\mu}, \sigma_{\beta_\mu})$) and deviation ($\beta_\sigma^+ \sim \mathcal{N}(\mu_{\beta_\sigma^+}, \sigma^2_{\beta_\sigma^+})$) that are randomly drawn from another prior distribution; this results in hierarchically modeling batch as a random effect. The prior assumptions of the Bayesian model can be summarized as follows:

$$w_\mu(i) \sim \mathcal{N}(0, \sigma^2_{w_\mu(i)}) \ \ , \ \ \alpha_\mu \sim \mathcal{N}(0, \sigma^2_{\alpha_\mu}) \ \ , \ \ \beta_\mu \sim \mathcal{N}(\mu_{\beta_\mu}, \sigma^2_{\beta_\mu}) \tag{8}$$

$$w_\sigma^+(i) \sim \mathcal{N}^+(0, \sigma^2_{w_\sigma^+(i)}) \ \ , \ \ \alpha_\sigma^+ \sim \mathcal{N}^+(0, \sigma^2_{\alpha_\sigma^+}) \ \ , \ \ \beta_\sigma \sim \mathcal{N}^+(\mu_{\beta_\sigma}, \sigma^2_{\beta_\sigma}) \tag{9}$$

Where $\mathcal{N}^+$ denotes a positive half-normal distribution. The random effect parameters are hierarchically sampled from a set of batch hyperparameters:

$$\mu_{\beta_\mu} \sim \mathcal{N}(0, \sigma^2_{\mu_{\beta_\mu}}) \ \ , \ \ \sigma_{\beta_\mu} \sim \mathcal{N}^+(0, \sigma^2_{\sigma_{\beta_\mu}}) \ \ , \ \ \mu_{\beta_\sigma} \sim \mathcal{N}^+(0, \sigma^2_{\mu_{\beta_\sigma}}) \ \ , \ \ \sigma_{\beta_\sigma} \sim \mathcal{N}^+(0, \sigma^2_{\sigma_{\beta_\sigma}}) \tag{10}$$

Notably, the model can quantify the normative deviation of an individual subject y_i using a Z-score:

$$z = \frac{y_i - \mu_y}{\sigma_y} \tag{11}$$

This Z-score can be used to compute normative centiles using the cumulative distribution function (CDF) of the normal distribution.

4.8.3 Spectral Normative Model (SNM)

Next, we explain the theory behind the proposed SNM framework. In the training phase, k direct normative models are independently fitted to the projection of phenotypes on the first k brain eigenmodes with the lowest graph frequencies. This forms the low-dimensional latent approximation containing information about the low-frequency characteristics of the cortical phenotype. For example, in the case of cortical thickness, for each $i \leq k$, the spectral coefficient $s_i \in \mathbb{R}^{N_{train} \times 1}$ associated with Ψ_i is first computed:

$$s_i = T_{train} \cdot \Psi_i \tag{12}$$

Similar to the direct model, normative estimators of the mean (f_{μ_i}) and deviation ($f^+_{\sigma\ i}$) are trained to model normative ranges of all low-frequency spectral coefficients:

$$\forall i \leq k : s_i \sim \mathcal{N}(\mu_{s_i}, \sigma^2_{s_i}) \Rightarrow \begin{cases} \mu_{s_i} = f_{\mu_i}(age,\ sex,\ batch) \\ \sigma_{s_i} = f^+_{\sigma\ i}(age,\ sex,\ batch) \end{cases} \tag{13}$$

The normative model for each spectral coefficient is the same as the direct case (see equations 6, 7, 8, 9, 10). Next, the normative ranges of any arbitrary spatial query is approximated using this set of k pre-trained normative models. For any spatial query x, its observed cortical phenotype $y = T_{train} \cdot x$ is approximated by a low-pass graph filter (see equation 3):

$$\begin{aligned} y = T_{train} \cdot x \approx\ & \hat{y}_{(k)} = T_{train} \cdot \hat{x}_{(k)} \\ =\ & T_{train} \hat{\Psi}_{(k)} \hat{\Psi}^T_{(k)} x = T_{train} \hat{\Psi}_{(k)} \tilde{x}_{(k)} \\ =\ & \sum_{i \leq k} T_{train} \Psi_i \tilde{x}_i = \sum_{i \leq k} s_i \tilde{x}_i \end{aligned} \tag{14}$$

The variable of interest y can hence be approximated by a linear combination of the spectral coefficients s_i. Notably, if the spatial query is frequency-bounded this approximation will be exact (i.e. if the query is a linear combination of low-frequency eigenmodes such that $\tilde{x}_i = 0$ for $i > k$). Equation 14 suggests that for any spatial query x, the associated phenotype y can be modeled as a multivariate normal distribution such that its normative ranges are estimated by the linear combination $y = \sum_{i \leq k} s_i \tilde{x}_i$. This is formalized by the following equation:

$$\forall x \in \mathbb{R}^{N_v \times 1},\ y = T_{train} \cdot x :\ y \sim \mathcal{N}(\mu_y, \sigma^2_y) \Rightarrow \begin{cases} \mu_y \approx g_\mu(F_{\mu(k)}, \tilde{x}_{(k)}) \\ \sigma_y \approx g^+_\sigma(\mathbf{\Sigma}, \tilde{x}_{(k)}) \end{cases} \tag{15}$$

Where g_μ and g^+_σ are functions that respectively approximate the mean and standard deviation of the multivariate normal distribution based on the trained low-pass spectral moments ($F_{\mu(k)}, \mathbf{\Sigma}$). Explicitly, g_μ defines the mean of the normal distribution y as a linear combination of the means of each spectral coefficient and is based on the graph spectral encoding $\tilde{x}_{(k)}$:

$$g_\mu(F_{\mu(k)}, \tilde{x}_{(k)}) = \sum_{i \leq k} f_{\mu_i} \tilde{x}_i = \begin{pmatrix} f_{\mu_1}, & f_{\mu_2}, & \cdots, & f_{\mu_k} \end{pmatrix}_{1 \times k} \cdot \tilde{x}_{(k)} = F_{\mu(k)} \cdot \tilde{x}_{(k)} \tag{16}$$

where $F_{\mu(k)}$ is used to summarize the set of estimated means for the first k spectral coefficients $(y_1, \cdots, y_k)$ into a $1 \times k$ vector. Note that equation 16 arises because the expected value of a linear combination of a set of random variables is equal to the linear combination of their respective expected values:

$$E[a_1 X_1 + a_2 X_2 + ... + a_k X_k] = a_1 E[X_1] + a_2 E[X_2] + ... + a_k E[X_k] \tag{17}$$

Moreover, g^+_σ defines the standard deviation of the normal distribution y as a function of the covariance matrix describing cross-basis dependencies of spectral coefficients $\mathbf{\Sigma} \in \mathbb{R}^{k \times k}_+$:

$$g^+_\sigma = \sqrt{\sum_{i,j \leq k} \tilde{x}_i \tilde{x}_j \mathbf{\Sigma}_{i,j}} = \sqrt{\tilde{x}^T_{(k)} \mathbf{\Sigma} \tilde{x}_{(k)}} \tag{18}$$

Equation 18 describes how the standard deviation of the variable of interest y is related to the covariance matrix $\mathbf{\Sigma}$ describing dependencies across spectral coefficients s_i. This equation is based on the following notion that relates the variance of a linear combination of a set of normally distributed variables to their covariance structure:

$$\mathrm{Var}\left(\sum_{i=1}^{k} a_i X_i\right) = \left(\sum_{i=1}^{k} a_i^2 \mathrm{Var}(X_i)\right) + \left(\sum_{i=1}^{k} \sum_{j=1(j\neq i)}^{k} a_i a_j \mathrm{Cov}(X_i, X_j)\right) \tag{19}$$

Elements of $\mathbf{\Sigma}$ can also change as a function of covariates and batch effects. We note that for a pair of random variables (X and Y), covariance ($\mathrm{Cov}(X,Y)$) and correlation ($\mathrm{Corr}(X,Y)$) are related by $\mathrm{Cov}(X,Y) = \sigma_X \sigma_Y \mathrm{Corr}(X,Y)$, where σ denotes the standard deviation of each variable. We use this notion to separate the effect of within-mode variance from the effect of cross-mode correlation on $\mathbf{\Sigma}$. In particular, the following equation decomposes the cross-basis covariance matrix $\mathbf{\Sigma}$ into a correlation matrix $\mathbf{P}$ and an independent set of eigenmode standard deviations:

$$\mathbf{\Sigma} = \mathrm{diag}\left(F^{+}_{\sigma\,(k)}\right) \,.\, \mathbf{P} \,.\, \mathrm{diag}\left(F^{+}_{\sigma\,(k)}\right) \tag{20}$$

where $F^{+}_{\sigma\,(k)}$ summarizes the set of estimated standard deviations of the first k spectral coefficients $(y_1, \cdots, y_k)$ to a $1 \times k$ vector, and the *diag()* operator is used to map a $1 \times k$ vector to a $k \times k$ diagonal matrix:

$$\mathrm{diag}\left(F^{+}_{\sigma\,(k)}\right) = \begin{pmatrix} f^{+}_{\sigma\,1}, & 0, & \cdots, & 0 \\ 0, & f^{+}_{\sigma\,2}, & \cdots, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots \\ 0, & 0, & \cdots, & f^{+}_{\sigma\,k} \end{pmatrix} \tag{21}$$

Thus far, we show that any spatial query can be represented as a multivariate normal distribution across k spectral components, the normative ranges of which can be approximated from the mean ($F_{\mu(k)}$), standard deviation ($F^{+}_{\sigma\,(k)}$), and cross-correlation ($\mathbf{P}$) of spectral coefficients. Mean and standard deviation estimates are computed similarly to a direct HBR model. However, approximations may be needed to estimate the cross-correlation structure. This is because the complexity of the correlation structure increases quadratically with the number of eigenmodes included in the low-pass approximation (k). We hence impose a sparsity constraint on the correlation structure using a threshold ($\rho = 0.25$) on the observed cross-correlation of spectral phenotypes within the training sample. Our evaluations indicate that the cross-correlation matrix is naturally sparse (see Figure 1D). As a result, the majority of elements $\mathbf{P}_{i,j} : \mathrm{Corr}(s_i, s_j) \leq \rho$ are small and will hence be replaced with zero through this sparsification step. The remaining pairs of suprathreshold correlations form a sparse matrix representation as summarized by the following derivation of $\mathbf{P}$:

$$\mathbf{P}_{i,j} = \begin{cases} 0 & \text{if } \mathrm{Corr}(s_i, s_j) \leq \rho \\ \rho_{s_i,s_j} & \text{if } \mathrm{Corr}(s_i, s_j) > \rho \end{cases} \tag{22}$$

The suprathreshold elements ρ_{s_i,s_j} are affected by covariates (age and sex) and batch effects. As such, for every suprathreshold element, the spectral coefficient pair was modeled as a bivariate normal distribution:

$$\begin{bmatrix} s_i \\ s_j \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{s_i} \\ \mu_{s_j} \end{bmatrix}, \begin{bmatrix} \sigma^2_{s_i} & \rho_{s_i,s_j}\sigma_{s_i}\sigma_{s_j} \\ \rho_{s_i,s_j}\sigma_{s_i}\sigma_{s_j} & \sigma^2_{s_j} \end{bmatrix}\right) \Rightarrow \begin{cases} \text{constants} \begin{cases} \mu_{s_i} = f_{\mu_i};\ \mu_{s_j} = f_{\mu_j} \\ \sigma_{s_i} = f^{+}_{\sigma\,i};\ \sigma_{s_j} = f^{+}_{\sigma\,j} \end{cases} \\ \\ \rho_{s_i,s_j} = tanh\left(f_{\rho_{i,j}}(age,\ sex,\ batch)\right) \end{cases} \tag{23}$$

where spectral coefficients' estimates of the mean (μ_{s_i}, μ_{s_j}) and deviation ($\sigma_{s_i}, \sigma_{s_j}$) were inputted as constants (yielded from the solution of Equation 13). The cross-correlation estimate (ρ_{s_i,s_j}) was

adjusted by a hyperbolic tangent function to limit the range of possible correlation values to $[-1, 1]$. Effects of covariates and batch were modeled as follows:

$$f_{\rho_{i,j}}(age,\ sex,\ batch) = w_\rho \Phi(age) + \alpha_\rho(sex) + \beta_\rho(batch) + \epsilon_\rho \tag{24}$$

with the following prior assumptions:

$$w_\rho(i) \sim \mathcal{N}(0, \sigma^2_{w_\rho(i)}) \ \ , \ \ \alpha_\rho \sim \mathcal{N}(0, \sigma^2_{\alpha_\rho}) \ \ , \ \ \beta_\rho \sim \mathcal{N}(\mu_{\beta_\rho}, \sigma^2_{\beta_\rho}) \tag{25}$$

and the following hierarchical batch priors:

$$\mu_{\beta_\rho} \sim \mathcal{N}(0, \sigma^2_{\mu_{\beta_\rho}}) \ \ , \ \ \sigma_{\beta_\rho} \sim \mathcal{N}^+(0, \sigma^2_{\sigma_{\beta_\rho}}) \tag{26}$$

4.9 Model Evaluation

We use three different families of spatial normative queries (x) to evaluate SNM's performance. These families include (I) *brain-wide*, (II) *regional*, and (III) *high-resolution* signals, each of which groups queries according to their spatial scale. Brain-wide queries are regional masks that describe coarse and large-scale characteristics along the cortex. Namely, this includes a total of 25 queries that represent the average of total cortical thickness, as well as the average thickness over predefined brain networks that divide the brain into 7/17 functionally distinct segments (Yeo functional networks)[39]. Regional queries included 200 signals, each describing the average cortical thickness within a region of interest as defined by a brain parcellation. For regional queries, we used a homotopic parcellation of the cortex comprising 200 cortical regions[121] (Yan200 atlas). In the supplementary analysis, we evaluate the sensitivity of our findings to atlas granularity (see Supplementary Figures S.5, S.6). Finally, a total of 400 high-resolution signals were constructed, each of which describes the average cortical thickness in the vicinity of a specific cortical vertex. In this case, we first randomly select one vertex from every region of the Yan400 cortical parcellation[121] and compute an 8mm FWHM smoothing kernel based on the geodesic distance to that vertex[114]. The smoothing kernel serves two key purposes: (1) to mitigate the impact of local intersubject registration misalignments and (2) to ensure vertex-resolution queries are appropriately reconstructed by the low-pass graph filter, thereby reducing ringing effects in reconstructions. In a supplementary analysis, we evaluate the sensitivity of our evaluations to the smoothing kernel choice (see Supplementary Figures S.5, S.7). The signals evaluated here quantify the average thickness of the cortex with different spatial granularities (ranging from tens of centimeters to a few millimeters). While such average queries are the only type of normative queries that have been previously studied for brain charting, in supplementary analyses we evaluate the applicability of our proposed model to other possible cases, e.g. when a comparative asymmetric norm is being studied (see Supplementary Figures S.8, S.9, S.10, and S.11).

For all signal families (whether presented in the main text or supplementary information), we evaluated the performance of SNM (using different values of k) based on (i) reconstruction accuracy, normative model's (ii) goodness of fit of the central tendency, and (iii) total goodness of fit of the estimated distribution. These evaluation criteria are detailed in the ensuing sections.

4.9.1 Reconstruction Accuracy

For the spectral approach to provide accurate estimates of normative ranges, both the cortical phenotype under study (i.e. cortical thickness) and the spatial normative query must be adequately reconstructed by the low-dimensional latent space defined by the brain eigenmodes. We hence use the following two metrics to evaluate the validity of the brain eigenmodes for reconstructing brain signals:

Energy Proportion: For any arbitrary brain signal $x \in \mathbb{R}^{N_v \times 1}$ with graph spectral encoding $\tilde{x} = \Psi^T x$, the total energy of the signal can be quantified by $E(x) = \sum_{i \in \{0 \cdots N_v\}} x_i^2 = \sum_{i \in \{0 \cdots N_v\}} \tilde{x}_i^2 = E(\tilde{x})$. In other words, the signal energy in vertex domain is equal to the signal energy in the spectral domain (Parseval's theorem[122]). As such, the proportional contribution of every single eigenmode Ψ_i to the total signal energy can be quantified by $\tilde{x}_i^2 / E(x)$. We provide a cumulative line plot of observed ranges for this proportional energy for different brain signals. This indicates the number of eigenmodes that capture major energy proportions for different brain signals.

Standardized Mean Squared Error (SMSE): This measure normalizes the mean squared error (MSE) by signal variance to provide an error metric that does not depend on signal variance[1,2,3]. Using this metric, a trivial reconstruction that replaces values of all vertices x_i with the signal mean will have an SMSE of one, and a perfect reconstruction will result in an SMSE of zero. By considering $\hat{x}_{(k)}$ to be a reconstruction of the original brain signal x, we compute SMSE for different ks to quantify reconstruction error at different low-pass frequency bandwidths. A high SMSE (close to one) indicates poor reconstruction performance and a low SMSE (close to zero) marks accurate reconstruction.

4.9.2 Model's Central Tendency

Independent of how accurately the eigenmode basis reconstructs various brain signals, we also need to validate that SNM yields accurate normative fits for spatial queries. This is quantified by measures of goodness of fit of the central tendency[2]. These measures can particularly indicate the accuracy of the normative model in explaining trends in the data. To this end, we utilize the mean absolute error (MAE) of model predictions. An out-of-sample MAE score was evaluated using 20% of the data in the test set. Specifically, for any spatial query x, the thickness of that query in the test set is measured $y \in \mathbb{R}^{N_{test} \times 1} = T_{test} \cdot x$ and MAE quantifies the extent to which these thickness values deviate from model predictions $\text{MAE} = \mathbb{E}[|y - \mu_y|] = \sum |y_i - \mu_{y_i}|/N_{test}$. MAE was computed for each of the three families of spatial queries, comparing the performance of a direct HBR model against four alternative SNMs, which utilized the first 10, 100, 1,000, and 10,000 eigenmodes, respectively. Since all models provided mean thickness estimates in millimeters, MAE offers a practical measure of the average error in millimeters when inferring the thickness of a given spatial query.

4.9.3 Model's Total Goodness of Fit

Measures of central tendency, such as mean or median, assess only the accuracy of modeling normative trends but do not capture the validity of the estimated distribution's deviations. To evaluate the model's overall goodness of fit—including both mean and variance estimates—we utilize the mean standardized log loss (MSLL)[1,2,3]. Mean log loss quantifies the mean negative log probability of the observed data (in the test sample) under the hypothesis of the fitted normative model:

$$\text{log-loss(model)} = -\frac{1}{N_{test}} \log\left(p(y|\mu_y, \sigma_y)\right) = -\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left(\frac{1}{2} \log(2\pi\sigma_{y_i}^2) + \frac{(y_i - \mu_{y_i})^2}{2\sigma_{y_i}^2} \right) \tag{27}$$

MSLL is computed by comparing the log loss of competing models against the log loss of a trivial reference model that is based on the mean and standard deviation of the training sample, MSLL = log-loss(model) − log-loss(reference). By definition, the trivial model achieves an MSLL of zero, and models providing better than trivial normative ranges will yield negative MSLL values. Here, we estimate log loss using a censored log-likelihood measure to reduce the sensitivity to outliers[2,124]. As such, for observations falling outside of the range of 1st to 99th percentile, the value of log loss was replaced by $-\log(0.02)$. This censored MSLL score favors models with accurate distribution estimates within this percentile range and is thus robust against extreme outliers.

4.10 Clinical Evaluations

In Section 2.5 SNM Uncovers Cortical Signatures of Atrophy in Alzheimer's Disease and Section 2.6 SNM Highlights Heterogeneity in Individual Cortical Atrophy, we examined the clinical insights provided by SNM when applied to the clinical MACC dataset. A subset of 66 healthy individuals from the MACC cohort was used to fine-tune the pretrained SNM for the clinical sample. During fine-tuning, parameters linking cortical phenotypes to age and sex $(w_\mu, w_\sigma^+, w_\rho, \alpha_\mu, \alpha_\sigma^+, \alpha_\rho)$ were held constant, while only site-specific parameters for batch effect harmonization $(\beta_\mu, \beta_\sigma^+, \beta_\rho)$ were updated based on the MACC data. The fine-tuned model was then applied to generate normative deviation maps for smoothed, vertex-resolution cortical thickness phenotypes.

4.10.1 HC vs. AD Comparison

A two-sample t-test was conducted to compare high-resolution deviation maps between the healthy and AD groups within the MACC dataset. To identify vertices with statistically significant differences, a nonparametric false discovery rate (FDR) correction[125] was applied across all vertices (FDR-$\alpha = 5\%$, 1000 permutations), controlling for multiple comparisons. Vertices surpassing the corrected threshold were deemed to exhibit significantly different deviation patterns between groups.

4.10.2 Linking Atrophy to Cognitive Impairment

We next evaluated the potential of normative deviation maps to serve as biomarkers of cognitive impairment. To begin, we performed a univariate vertex-wise assessment, measuring the Pearson correlation coefficient between normative z-scores and cognitive performance scores (MMSE). The significance of associations was evaluated with a nonparametric FDR correction via permutation testing. Specifically, we shuffled MMSE scores across 1000 permutations, calculating correlation coefficients for each permutation to produce a null distribution under the hypothesis of no association between normative deviation and cognitive performance. Using this null distribution, we derived nonparametric p-values for each vertex, which were subsequently FDR-corrected at $\alpha = 5\%$.

In addition, we derived an extreme value statistic[89] for atrophy by counting vertices with a z-score below a threshold of -1.96 for each individual. This measure, termed Extremely Thin Vertex Count (ETVC), was then tested for its predictive power on MMSE scores via Pearson correlation. To assess the sensitivity of ETVC to the threshold choice, we repeated this analysis for a range of z-score thresholds. We performed these assessments within each cohort—HC, MCI, and AD—to evaluate the specificity of ETVC as a biomarker for AD-related cognitive impairment (see Supplementary Section A.6 Within-group Cognitive Associations).

4.10.3 Examining Interindividual Heterogeneity

Next, we aimed to demonstrate the capabilities of the SNM in revealing the heterogeneity landscape of brain atrophy in AD. We quantified interindividual differences in atrophy by computing the Euclidean distance between vectors of vertex-level z-score deviation maps, yielding a $N_{MACC} \times N_{MACC}$ distance matrix that captures interindividual differences. This matrix was visualized as a heatmap to provide an overview of the variation across individuals. Additionally, we applied multidimensional scaling (MDS) [126] to reduce the distance matrix to a 2-dimensional embedding, providing a spatial representation of interindividual differences. Average deviation maps for subgroups within local vicinities in this 2-dimensional space were then visualized as cortical projections (see Figure 6 and Supplementary Section A.8 Heterogeneity Landscape Subgroups). These visualizations, along with the original heatmap, illustrate that AD is associated with heterogeneous atrophy patterns: while individuals with AD deviate from healthy norms, they do not conform to a single or even multiple stereotypical patterns of deviation.

Code and Data Availability

The code for implementing, evaluating, and generalizing the SNM framework to independent cohorts will be made openly available in the following Git repository: **[Link to the repository will be added here.]**

This repository will also host the fitted parameters of the lifespan SNM model. Analyses involving the healthy lifespan sample were conducted using publicly available data from the Human Connectome Project, accessible through the Connectome Coordination Facility (`www.humanconnectome.org`). The clinical AD dataset can be obtained via a data-transfer agreement with the MACC (`http://www.macc.sg/`).

Acknowledgments

The analysis for this study was supported by the Spartan High-Performance Computing infrastructure[127,128] and dedicated computing and storage solutions provided by Research Computing Services at the University of Melbourne. We gratefully acknowledge the invaluable contribution of several open-source software packages that significantly facilitated our data analysis and interpretation. The analytical pipeline made use of several open-source Python packages, including Numpy[129], Scipy[130], Matplotlib[131], Pandas[132], scikit-learn[133], PyMC[134], xarray[135], and Nibabel[136]. The connectome spatial smoothing package[113] was utilized to compute smoothing kernels. Brain visualizations were generated by the Cerebro brain viewer[137]. This study was supported by the NHMRC Ideas Grant (GNT2011592) titled Normative Reference Ranges for Brain Phenotypes, awarded to A. Zalesky, C. Bousman, C. Pantelis, and M. Di Biase, as well as the NUS Yong Loo Lin School of Medicine (NUHSRO/2020/124/TMR/LOA), the Singapore National Medical Research Council (NMRC) LCG (OFLCG19May-0035), NMRC CTG-IIT (CTGIIT23jan-0001), NMRC OF-IRG (OFIRG24jan-0030), NMRC STaR (STaR20nov-0003), Singapore Ministry of Health (MOH) Centre Grant (CG21APR1009), the Temasek Foundation (TF2223-IMH-01), and the United States National Institutes of Health (R01MH133334). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the funders.

In addition, this study was made possible by data obtained from the following studies:

HCP-Development: Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number U01MH109589 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The HCP-Development 2.0 Release data used in this report came from DOI: 10.15154/1520708.

HCP-Young Adult: Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

HCP-Aging: Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number U01AG052564 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The HCP-Aging 2.0 Release data used in this report came from DOI: 10.15154/1520707.

MACC: The translational evaluation of our framework utilized clinical data from the Memory, Aging, and Cognition Center (MACC) at the National University of Singapore, acquired through psychiatry-led memory services.

References

1. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biological Psychiatry* **80,** 552–561. (2016).
2. Dinga, R. *et al.* Normative modeling of neuroimaging data using generalized additive models of location scale and shape. *bioRxiv.* (2021).
3. Rutherford, S. *et al.* The normative modeling framework for computational psychiatry. *Nature Protocols* **17,** 1711–1734. (July 2022).
4. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry* **24,** 1415–1424. (2019).
5. Bethlehem, R. A. *et al.* A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. *Communications Biology* **3.** (2020).

6. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604,** 525–533. (Apr. 2022).
7. Mirnezami, R., Nicholson, J. & Darzi, A. Preparing for Precision Medicine. *New England Journal of Medicine* **366,** 489–491. (Feb. 2012).
8. Zabihi, M. *et al.* Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **4,** 567–578. (2019).
9. Parkes, L. *et al.* Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure. *Translational Psychiatry* **11.** (2021).
10. Lv, J. *et al.* Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Molecular Psychiatry* **26,** 3512–3523. (July 2021).
11. Habes, M. *et al.* The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimer's and Dementia* **17,** 89–102. (2021).
12. Alexander-Bloch, A. *et al.* Copy Number Variant Risk Scores Associated with Cognition, Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. *JAMA Psychiatry* **79,** 699–709. (2022).
13. Holz, N. E. *et al.* A stable and replicable neural signature of lifespan adversity in the adult brain. *Nature Neuroscience* **26,** 1603–1612. (Sept. 2023).
14. Verdi, S. *et al.* Personalising Alzheimer's Disease progression using brain atrophy markers, 1–29 (2023).
15. Kia, S. M. *et al. Hierarchical bayesian regression for multi-site normative modeling of neuroimaging data* in *Medical Image Computing and Computer Assisted Intervention–MICCAI* (2020), 699–709.
16. Fraza, C. J., Dinga, R., Beckmann, C. F. & Marquand, A. F. Warped Bayesian linear regression for normative modelling of big data. *NeuroImage* **245,** 118715. (Dec. 2021).
17. Rutherford, S. *et al.* Charting brain growth and aging at high spatial precision. *eLife* **11,** 1–15. (Feb. 2022).
18. Wolfers, T. *et al.* Individual differences v. the average patient: Mapping the heterogeneity in ADHD using normative models. *Psychological Medicine* **50,** 314–323. (2019).
19. Petersen, M. *et al.* Brain network architecture constrains age-related cortical thinning. *NeuroImage* **264,** 119721. (2022).
20. Kalantar-Hormozi, H. *et al.* A cross-sectional and longitudinal study of human brain development: The integration of cortical thickness, surface area, gyrification index, and cortical curvature into a unified analytical framework. *NeuroImage* **268,** 119885. (2023).
21. Zhang, R., Chen, L., Oliver, L. D., Voineskos, A. N. & Park, J. Y. SAN: Mitigating spatial covariance heterogeneity in cortical thickness data collected from multiple scanners or sites. *Human Brain Mapping* **45,** 1–14. (2024).
22. Pang, J. C. *et al.* Geometric constraints on human brain function. *Nature* **618,** 566–574. (June 2023).
23. Mansour L., S. *et al. Eigenmodes of the brain: revisiting connectomics and geometry* Apr. 2024.
24. Olsen, A. S. *et al. On reconstruction of cortical functional maps using subject-specific geometric and connectome eigenmodes* Oct. 2024.
25. Robinson, P. *et al.* Eigenmodes of brain activity: Neural field theory predictions and comparison with experiment. *NeuroImage* **142,** 79–98. (Nov. 2016).

26. Deslauriers-Gauthier, S., Zucchelli, M., Frigo, M. & Deriche, R. A unified framework for multimodal structure–function mapping based on eigenmodes. *Medical Image Analysis* **66,** 101799. (2020).

27. Atasoy, S., Donnelly, I. & Pearson, J. Human brain networks function in connectome-specific harmonic waves. *Nature Communications* **7,** 10340. (Apr. 2016).

28. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nature Reviews Neuroscience* **16,** 159–172. (Mar. 2015).

29. Pievani, M., Filippini, N., Van Den Heuvel, M. P., Cappa, S. F. & Frisoni, G. B. *Brain connectivity in neurodegenerative diseases - From phenotype to proteinopathy* 2014. .

30. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron* **62,** 42–52. (2009).

31. Wannan, C. M. *et al.* Evidence for network-based cortical thickness reductions in schizophrenia. *American Journal of Psychiatry* **176,** 552–563. (2019).

32. Vogel, J. W. *et al.* Connectome-based modelling of neurodegenerative diseases: towards precision medicine and mechanistic insight. *Nature Reviews Neuroscience* **24,** 620–639. (2023).

33. Chopra, S. *et al.* Network-Based Spreading of Gray Matter Changes Across Different Stages of Psychosis. *JAMA Psychiatry* **80,** 1246–1257. (2023).

34. Sebenius, I. *et al.* Structural MRI of brain similarity networks. *Nature Reviews Neuroscience* **26.** (2024).

35. Mansour L, S., Tian, Y., Yeo, B. T., Cropley, V. & Zalesky, A. High-resolution connectomic fingerprints: Mapping neural identity and behavior. *NeuroImage* **229,** 117695. (Apr. 2021).

36. Fischl, B. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex* **14,** 11–22. (Jan. 2004).

37. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31,** 968–980. (July 2006).

38. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53,** 1–15. (Oct. 2010).

39. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106,** 1125–1165. (2011).

40. Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72,** 665–678. (Nov. 2011).

41. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex* **26,** 3508–3526. (2016).

42. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex* **26,** 288–303. (2016).

43. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536,** 171–178. (Aug. 2016).

44. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex* **28,** 3095–3114. (2018).

45. Eickhoff, S. B., Yeo, B. T. & Genon, S. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience* **19,** 672–686. (2018).

46. Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex* **29,** 2533–2551. (2019).

47. Poline, J. B. *et al.* Data sharing in neuroimaging research. *Frontiers in Neuroinformatics* **6,** 1–13. (2012).

48. Eickhoff, S., Nichols, T. E., Van Horn, J. D. & Turner, J. A. Sharing the wealth: Neuroimaging data repositories. *NeuroImage* **124,** 1065–1068. (2016).

49. Jwa, A. S. & Poldrack, R. A. The spectrum of data sharing policies in neuroimaging data repositories. *Human Brain Mapping* **43,** 2707–2721. (2022).

50. Giehl, K. *et al.* Sharing brain imaging data in the Open Science era: how and why? *The Lancet Digital Health* **6,** e526–e535. (2024).

51. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22,** 1345–1359. (2010).

52. Dufumier, B. *et al.* Exploring the potential of representation and transfer learning for anatomical neuroimaging: Application to psychiatry. *NeuroImage* **296,** 120665. (2024).

53. Kia, S. M. *et al.* Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression. *PLoS ONE* **17,** 1–29. (2022).

54. Busche, M. A. & Hyman, B. T. Synergy between amyloid-β and tau in Alzheimer's disease. *Nature Neuroscience* **23,** 1183–1193. (2020).

55. Hampel, H. *et al.* The Amyloid-β Pathway in Alzheimer's Disease. *Molecular Psychiatry* **26,** 5481–5503. (2021).

56. Das, S. R. *et al.* Tau-Atrophy Variability Reveals Phenotypic Heterogeneity in Alzheimer's Disease. *Annals of Neurology* **90,** 751–762. (2021).

57. Frigerio, I. *et al.* Amyloid-β, p-tau and reactive microglia are pathological correlates of MRI cortical atrophy in Alzheimer's disease. *Brain Communications* **3,** 1–13. (2021).

58. Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N. & Fox, N. C. Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 4703–4707. (2002).

59. Dickerson, B. C. *et al.* Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology* **76,** 1395–1402. (2011).

60. Zetterberg, H. & Bendlin, B. B. Biomarkers for Alzheimer's disease—preparing for a new era of disease-modifying therapies. *Molecular Psychiatry* **26,** 296–308. (2021).

61. Bakkour, A., Morris, J. C. & Dickerson, B. C. The cortical signature of prodromal AD: Regional thinning predicts mild AD dementia. *Neurology* **72,** 1048–1055. (2009).

62. Dickerson, B. C. *et al.* The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex* **19,** 497–510. (2009).

63. Whitwell, J. L. Progression of atrophy in Alzheimer's disease and related disorders. *Neurotoxicity Research* **18,** 339–346. (2010).

64. Poulakis, K. *et al.* Heterogeneous patterns of brain atrophy in Alzheimer's disease. *Neurobiology of Aging* **65,** 98–108. (2018).

65. Sun, N., Mormino, E. C., Chen, J., Sabuncu, M. R. & Yeo, B. T. Multi-modal latent factor exploration of atrophy, cognitive and tau heterogeneity in Alzheimer's disease. *NeuroImage* **201.** (2019).

66. Poulakis, K. *et al.* Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. *Nature Communications* **13.** (2022).

67. Jones, B. F. *et al.* Differential regional atrophy of the cingulate gyrus in Alzheimer disease: A volumetric MRI study. *Cerebral Cortex* **16,** 1701–1708. (2006).

68. Benzinger, T. L. *et al.* Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **110.** (2013).

69. Phan, T. X. *et al.* Increased Cortical Thickness in Alzheimer's Disease. *Annals of Neurology* **95,** 929–940. (2024).

70. Iacono, D. *et al.* Neuronal hypertrophy in asymptomatic Alzheimer disease. *Journal of Neuropathology and Experimental Neurology* **67,** 578–589. (2008).

71. Verdi, S., Marquand, A. F., Schott, J. M. & Cole, J. H. Beyond the average patient: How neuroimaging models can address heterogeneity in dementia. *Brain* **144,** 2946–2953. (2021).

72. Ownby, R. L., Crocco, E., Acevedo, A., John, V. & Loewenstein, D. Depression and Risk for Alzheimer Disease. *Archives of General Psychiatry* **63,** 530. (2006).

73. Körner, S. *et al.* Prevalence and prognostic impact of comorbidities in amyotrophic lateral sclerosis. *European Journal of Neurology* **20,** 647–654. (2013).

74. Wang, J. H., Wu, Y. J., Tee, B. L. & Lo, R. Y. Medical Comorbidity in Alzheimer's Disease: A Nested Case-Control Study. *Journal of Alzheimer's Disease* **63,** 773–781. (2018).

75. Avitan, I. *et al.* Towards a consensus on alzheimer's disease comorbidity? *Journal of Clinical Medicine* **10,** 1–13. (2021).

76. Whitwell, J. L. *et al.* Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *The Lancet Neurology* **11,** 868–877. (2012).

77. Noh, Y. *et al.* Anatomical heterogeneity of Alzheimer disease. *Neurology* **83,** 1936–1944. (Nov. 2014).

78. Zhang, X. *et al.* Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **113,** E6535–E6544. (2016).

79. Ferreira, D. *et al.* Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: Longitudinal trajectories and clinical applications. *Scientific Reports* **7,** 1–13. (2017).

80. Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nature Medicine* **27,** 871–881. (2021).

81. Verdi, S. *et al.* Revealing Individual Neuroanatomical Heterogeneity in Alzheimer Disease Using Neuroanatomical Normative Modeling. *Neurology* **100,** E2442–E2453. (2023).

82. Mohanty, R. *et al.* Comparison of subtyping methods for neuroimaging studies in Alzheimer's disease: A call for harmonization. *Brain Communications* **2,** 1–16. (2020).

83. Hohman, T. J. *et al.* Asymptomatic Alzheimer disease. *Neurology* **87,** 2443–2450. (2016).

84. Gómez-Isla, T. & Frosch, M. P. Lesions without symptoms: understanding resilience to Alzheimer disease neuropathological changes. *Nature Reviews Neurology* **18,** 323–332. (2022).

85. Yang, Z. *et al.* Brain aging patterns in a large and diverse cohort of 49,482 individuals. *Nature Medicine.* (2024).

86. Chouliaras, L. & O'Brien, J. T. The use of neuroimaging techniques in the early and differential diagnosis of dementia. *Molecular Psychiatry* **28,** 4084–4097. (2023).

87. Huang, J. *et al.* Cerebral microinfarcts revisited: Detection, causes, and clinical relevance. *International Journal of Stroke* **19,** 7–15. (2024).

88. Tarun, A., Abramian, D., Behjat, H. & De Ville, D. V. *Graph Spectral Analysis of Voxel-Wise Brain Graphs from Diffusion-Weighted Mri* in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (IEEE, Apr. 2019), 159–163.

89. Fraza, C., Zabihi, M., Beckmann, C. F. & Marquand, A. F. *Reconceptualizing psychopathology as extreme deviations from a normative reference model* Aug. 2022.

90. Villalón-Reina, J. E. *et al. Large-scale Normative Modeling of Brain Microstructure* in *2023 19th International Symposium on Medical Information Processing and Analysis (SIPAIM)* (IEEE, Nov. 2023), 1–5.

91. De Boer, A. A. A. *et al.* Non-Gaussian normative modelling with hierarchical Bayesian regression. *Imaging Neuroscience* **2,** 1–36. (Apr. 2024).

92. Rutherford, S. *et al.* Evidence for embracing normative modeling. *eLife* **12,** 1–24. (2023).

93. Kasper, J. *et al.* Resting state changes in aging and Parkinson's disease are shaped by underlying neurotransmission – a normative modeling study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,* 1–12. (2024).

94. Therriault, J. *et al.* Biomarker-based staging of Alzheimer disease: rationale and clinical applications. *Nature Reviews Neurology* **20,** 232–244. (2024).

95. Pang, J. C. *et al.* Reply to: Commentary on Pang et al. (2023). *Nature,* 1–7 (2023).

96. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* **113,** 12574–12579. (Nov. 2016).

97. Becker, C. O. *et al.* Spectral mapping of brain functional connectivity from diffusion imaging. *Scientific Reports* **8,** 1–15. (2018).

98. Gavili, A. & Zhang, X.-P. On the Shift Operator, Graph Frequency, and Optimal Filtering in Graph Signal Processing. *IEEE Transactions on Signal Processing* **65,** 6303–6318. (Dec. 2017).

99. Benzi, M., Bertaccini, D., Durastante, F. & Simunec, I. Non-local network dynamics via fractional graph Laplacians. *Journal of Complex Networks* **8,** 1–29. (2020).

100. Boukrab, R. & Pagès-Zamora, A. Random-walk Laplacian for frequency analysis in periodic graphs. *Sensors (Switzerland)* **21,** 1–13. (2021).

101. Van Essen, D. C. *et al.* The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62,** 2222–2231. (2012).

102. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80,** 62–79. (2013).

103. Bookheimer, S. Y. *et al.* The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage* **185,** 335–348. (2019).

104. Somerville, L. H. *et al.* The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* **183,** 456–468. (2018).

105. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80,** 105–124. (Oct. 2013).

106. Harms, M. P. *et al.* Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage* **183,** 972–984. (2018).

107. Hilal, S. *et al.* Markers of cardiac dysfunction in cognitive impairment and dementia. *Medicine (United States)* **94,** e297. (2015).

108. Hilal, S. *et al.* Cortical cerebral microinfarcts predict cognitive decline in memory clinic patients. *Journal of Cerebral Blood Flow and Metabolism* **40,** 44–53. (2020).

109. Sechidis, K., Tsoumakas, G. & Vlahavas, I. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* PART 3, 145–158 (2011).

110. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical Surface-Based Analysis. **194,** 179–194 (1999).

111. Bayer, J. M. *et al.* Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. *NeuroImage* **264,** 119699. (Dec. 2022).

112. Mansour L., S., Di Biase, M. A., Smith, R. E., Zalesky, A. & Seguin, C. Connectomes for 40,000 UK Biobank participants: A multi-modal, multi-scale brain network resource. *NeuroImage* **283,** 120407. (Dec. 2023).

113. Mansour L., S., Seguin, C., Smith, R. E. & Zalesky, A. Connectome Spatial Smoothing v.0.1.1 (v.0.1.1). *Zenodo.* (2021).

114. Mansour L, S., Seguin, C., Smith, R. E. & Zalesky, A. Connectome spatial smoothing (CSS): Concepts, methods, and evaluation. *NeuroImage* **250,** 118930. (Apr. 2022).

115. Ortega, A., Frossard, P., Kovacevic, J., Moura, J. M. F. & Vandergheynst, P. Graph Signal Processing: Overview, Challenges, and Applications. *Proceedings of the IEEE* **106,** 808–828. (May 2018).

116. Huang, W. *et al.* A Graph Signal Processing Perspective on Functional Brain Imaging. *Proceedings of the IEEE* **106,** 868–885. (May 2018).

117. Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A. & Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine* **30,** 83–98. (May 2013).

118. Sandryhaila, A. & Moura, J. M. Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine* **31,** 80–90. (Sept. 2014).

119. Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V. & Bastiani, F. D. *Flexible Regression and Smoothing* (Chapman and Hall/CRC, Apr. 2017).

120. Huertas, I. *et al.* A Bayesian spatial model for neuroimaging data based on biologically informed basis functions. *NeuroImage* **161,** 134–148. (2017).

121. Yan, X. *et al.* Homotopic local-global parcellation of the human cerebral cortex from resting-state functional connectivity. *NeuroImage* **273,** 120010. (2023).

122. Hammond, D. K., Vandergheynst, P. & Gribonval, R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis* **30,** 129–150. (2011).

123. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (2006).

124. Diks, C., Panchenko, V. & Van Dijk, D. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics* **163,** 215–230. (2011).

125. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92,** 381–397. (2014).

126. Borg, I. & Groenen, P. *Modern Multidimensional Scaling: Theory and Applications* (2007).

127. Meade, B., Lafayette, L., Sauter, G. & Tosello, D. *Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility* 2017.

128. Lafayette, L. & Wiebelt, B. Spartan and NEMO: Two HPC-cloud hybrid implementations. *Proceedings - 13th IEEE International Conference on eScience, eScience 2017,* 458–459. (2017).

129. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585,** 357–362. (2020).

130. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17,** 261–272. (2020).

131. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science \& Engineering* **9,** 90–95. (2007).

132. Pandas development team, T. *pandas-dev/pandas: Pandas* Apr. 2024.

133. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

134. Abril-Pla, O. *et al.* PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science* **9,** e1516. (Sept. 2023).

135. Hoyer, S. & Hamman, J. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software* **5,** 10. (2017).

136. Brett, M. *et al. nipy/nibabel: 5.2.1* Feb. 2024.

137. Mansour L., S. *et al. sina-mansour/Cerebro_ Viewer: v0.0.10.4* Aug. 2023.

138. Robinson, E. C. *et al.* MSM: A new flexible framework for multimodal surface matching. *NeuroImage* **100,** 414–426. (2014).

139. Robinson, E. C. *et al.* Multimodal surface matching with higher-order smoothness constraints. *NeuroImage* **167,** 453–465. (2018).

140. Chong, J. S. X. *et al.* Influence of cerebrovascular disease on brain networks in prodromal and clinical Alzheimer's disease. *Brain* **140,** 3012–3022. (Nov. 2017).

141. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33,** 341–355. (2002).

142. Fischl, B. FreeSurfer. *NeuroImage* **62,** 774–781. (Aug. 2012).

Supplementary material for:

Spectral normative modeling of brain structure

Mansour L., S. et al. (2025)

A Supplementary Information

A.1 Image Acquisition and Preprocessing

A.1.1 HCP Lifespan Data

Healthy human brain imaging data from three cohorts of the Human Connectome Project (HCP) was aggregated to encompass a broad age range spanning the human lifespan. Table S.1 summarizes the population demographics across datasets and their respective train-test splits. The following paragraphs detail each cohort's image acquisition protocols and preprocessing steps.

Cohort	HCP-D		HCP-YA		HCP-A	
Split	Train	Test	Train	Test	Train	Test
Sample Size	521	131	877	219	580	145
Age	14.41 ± 4.06	14.53 ± 4.07	28.79 ± 3.69	28.78 ± 3.70	60.29 ± 15.70	60.61 ± 15.89
Sex	53.7% female	54.2% female	54.4% female	54.3% female	55.9% female	56.6% female
Average Thickness (mm)	3.01 ± 0.14	3.03 ± 0.14	2.64 ± 0.08	2.65 ± 0.08	2.57 ± 0.15	2.56 ± 0.17

Table S.1. Sample Characteristic of HCP lifespan dataset. Note: Values presented as $\mu \pm \sigma$ indicate the mean (μ) and standard deviation (σ).

HCP Young Adult Data Young adult imaging data was sourced from the HCP S1200 release[102] (HCP-YA). This contained structural MRI images obtained from 1096 healthy participants (54.4% female) aged 22 to 37 years old. A comprehensive report on imaging acquisition and preprocessing is available elsewhere[105]. In brief, the images were acquired with a Siemens 3T Skyra scanner with a 32-channel head coil (housed at Washington University). T1w images were acquired using a single-echo MPRAGE sequence (0.7mm isotropic resolution, TR = 2400ms, TE = 2.14ms, TI = 1000ms, flip angle = 8°, no partial Fourier)[105]. Acquired images underwent preprocessing using the HCP's minimal preprocessing pipeline. Notably, structural images underwent gradient distortion correction, skull stripping, readout distortion correction, and bias field correction, followed by a nonlinear registration to a standard template volumetric space. Cortical surface reconstruction was performed using FreeSurfer (version 5.2) which produced 3-dimensional meshes for the white and pial cortical surface along with segmentation and folding-based registration to the *fsaverage* surface. Finally, the outputs

were registered to the *Conte69* surface template with a multimodal surface matching algorithm (MSMAll)[138] and downsampled to triangulated meshes with ~32k vertices per cortical hemisphere, i.e. the *fs-LR* 32k space.

HCP Aging Data The aging imaging data was sourced from the HCP-Aging Lifespan 2.0 release[103] (HCP-A) which included structural MRI images from 725 healthy adults (56% female) aged 36 to 100+ years old. For all participants aged above 90 years old, age was reported as 100 to avoid releasing protected health information. The acquisition and preprocessing protocols for this cohort were generally consistent with the earlier acquired HCP-YA protocols, with particular necessary changes due to hardware differences that are detailed elsewhere[106]. In summary, data were collected on Siemens 3T Prisma scanners (with 32-channel head coils) located at 4 different acquisition sites in the United States. Particularly, T1w images were collected by a multi-echo MPRAGE sequence with real-time motion correction (0.8mm isotropic voxels, TR = 2500ms, TE = 1.8/3.6/5.4/7.2ms, TI = 1000ms, flip angle = 8°, 6/8 slice partial Fourier)[106]. The preprocessing pipeline for structural images was similar to that of HCP-YA following the minimal preprocessing pipeline[105] with minor variations including a different version of FreeSurfer (version 6.0) and an update to MSMAll using higher order smoothness constraints[139].

HCP Development Data The development imaging data was sourced from HCP-Development Lifespan 2.0 release[104] (HCP-D) which contained structural MRI images of 652 healthy participants (53.8% female) aged 5 to 22 years old. This cohort aimed to adapt existing HCP protocols to the practical challenges of studying developmental populations. The acquisition and processing of the HCP-D dataset[104] were generally consistent with the HCP-A dataset[103]. Namely, similar Prisma scanners (with 32-channel head coils) were used to collect imaging data across four acquisition sites, while a tailored 32-channel head coil was used for 5-7 year-old participants[104,106]. Identical to HCP-A, a multi-echo MPRAGE sequence was used to collect T1w images followed by the same preprocessing procedures[106].

A.1.2 MACC Clinical Data

The clinical data (part of MACC harmonization study) was acquired by the Memory, Ageing & Cognition Centre at the National University of Singapore[107,108]. The dataset includes 542 participants aged 50 to 91 years (61.4% female, see Table S.2 for detail), grouped into three cohorts: i) a healthy cohort (HC) of 132 cognitively healthy individuals, ii) a mild cognitive impairment cohort (MCI) of 202 individuals with cognitive impairment but no AD, and iii) an AD cohort of 208 individuals with an AD diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria. AD diagnosis followed internationally accepted criteria using the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) guidelines.

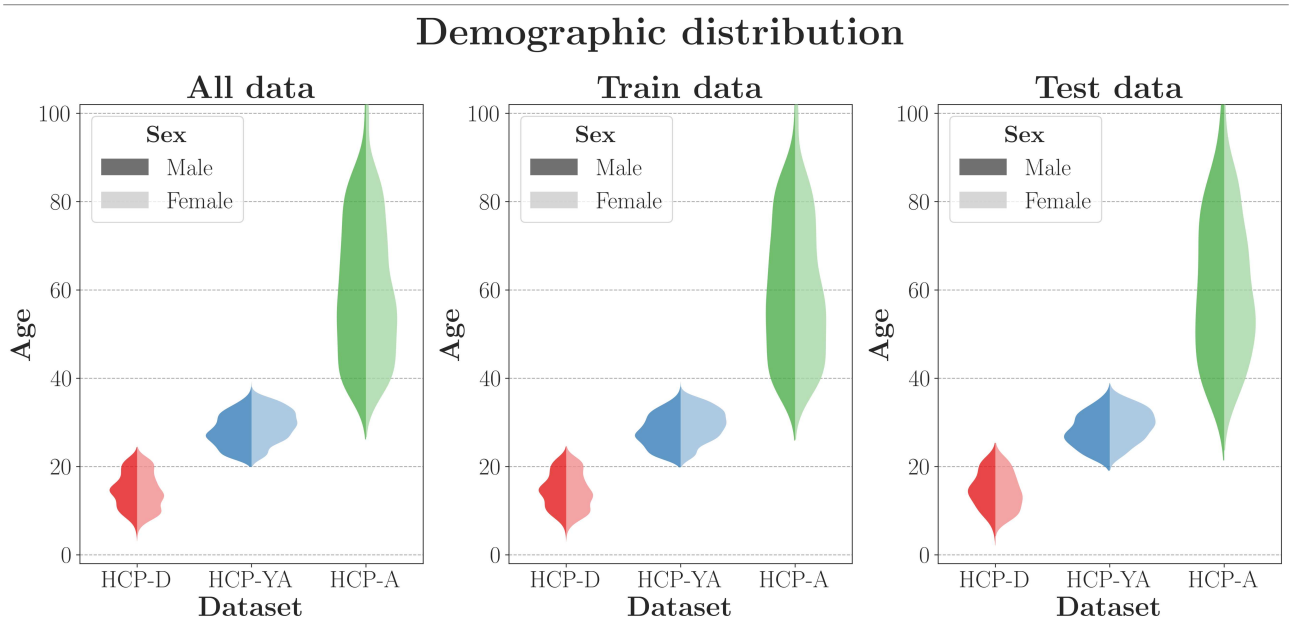
Each participant underwent neuroimaging and comprehensive clinical and neuropsychological evaluation, including the Mini-Mental State Examination (MMSE). MRI was performed on a 3T Siemens Magnetom Trio Tim scanner with a 32-channel head coil at the Clinical Imaging Research Centre of the National University of Singapore. The imaging protocol included a T1-weighted MPRAGE sequence (1 mm isotropic resolution, TR = 2300 ms, TE = 1.9 ms, TI = 900 ms, flip angle = 9°, 192 sagittal slices, matrix size = 256 × 256)[140]. T1-weighted scans were preprocessed using FreeSurfer 6.0 `recon-all` pipeline to generate cortical surfaces and thickness estimates in the fsnative surface space [141,142]. Matlab scripts and commands from Connectome Workbench and FreeSurfer were then used to project thickness data to the fs-LR surface space, aligning it with HCP thickness data to enable transfer learning from HCP datasets to MACC.

Cohort	HC	MCI	AD
Description	No cognitive impairment	Cognitive impairment no dementia	Diagnosis of AD
Sample Size	132	202	208
Age	68.49 ± 7.58	74.29 ± 6.70	75.49 ± 7.19
Sex	54.5% female	58.4% female	68.8% female
MMSE	27.39 ± 1.90	24.04 ± 3.88	15.60 ± 4.96

Table S.2. Sample Characteristic of MACC-H dataset. Note: Values presented as $\mu \pm \sigma$ indicate the mean (μ) and standard deviation (σ).

A.2 Stratified Sample Splitting

As described in the Methods section, train test splitting was performed by a randomized approach that controls for the distribution of sample covariates (age, sec, and dataset/batch). In particular, this was implemented by `scikit-learn`'s `train_test_split` function by utilizing the `stratify` parameter. First, continuous values for individual age were grouped into 20 bins by Quantile-based discretization. Next, discrete values encoding age, sex, and dataset information were utilized for stratified splitting. Supplementary Figure S.1 demonstrates the stability of covariate distribution after train test splitting.



Supplementary Figure. S.1. Distribution of demographic information before and after splitting. The distribution of demographic information is displayed for the entire sample before splitting (left) and after splitting into training (center) and test (right) sets. The violin plots illustrate the age distribution (y-axis) across different datasets (x-axis). Each plot differentiates between male (dark shades) and female (bright shades) participants. The stratified group splitting ensures that the demographic distributions of covariates are matched between the training and test samples.

A.3 Cortical projections of normative performance metrics

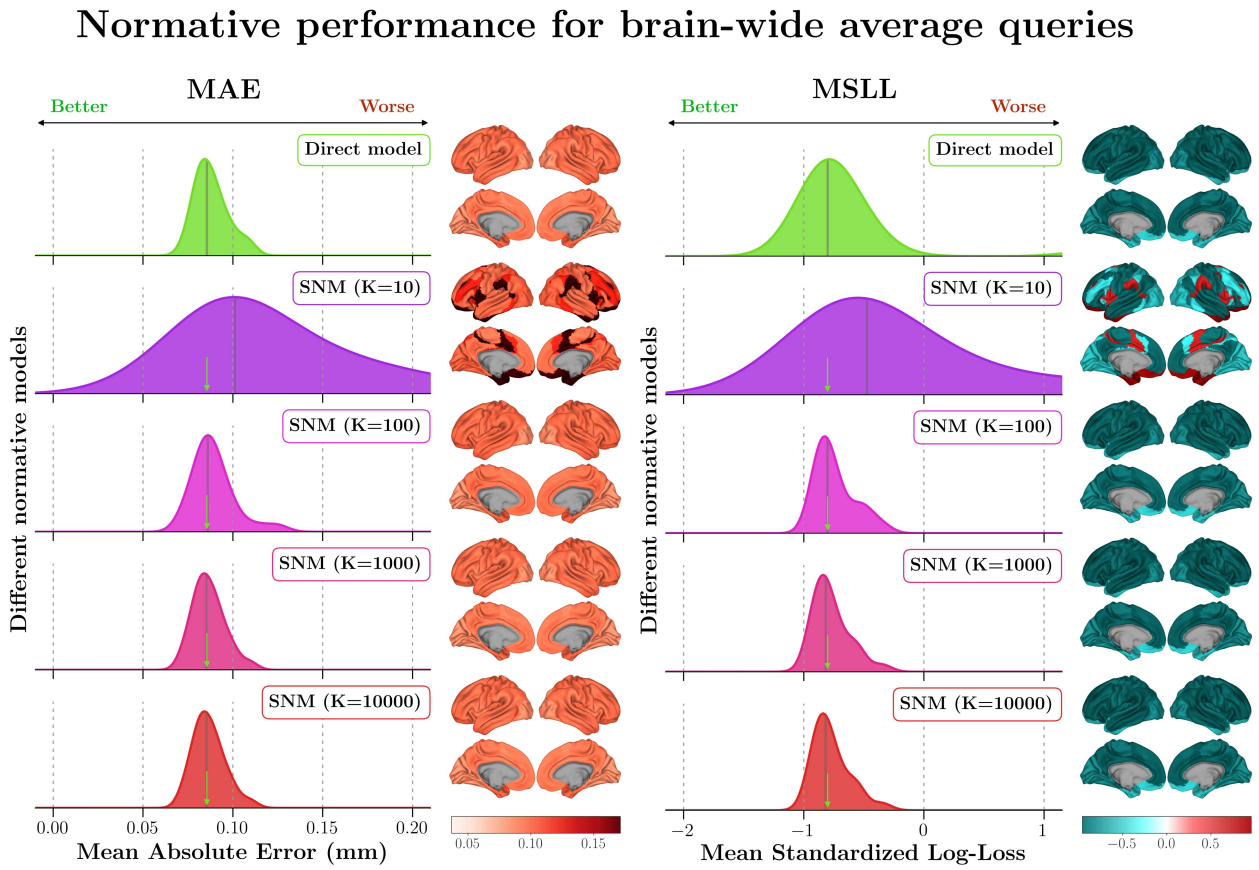
Figure 3 presents a summary report comparing SNM performance with that of the direct model. In the following sections, we expand upon these findings, detailing the regional distribution of performance metrics to offer deeper insights into comparative SNM evaluations.

A.3.1 Brain-wide normative estimates

Figure S.2 presents the performance of various SNMs in reconstructing brain-wide thickness maps, compared to a direct model. The direct model is trained on observed cortical phenotypes for each

spatial query ($y = T_{train} \cdot x$), while SNMs rely solely on spectral approximations ($y \approx \sum_{i \leq k} s_i \tilde{x}_i$). Four SNMs are evaluated, which respectively utilize the first $k = \{10, 10^2, 10^3, 10^4\}$ eigenmodes.

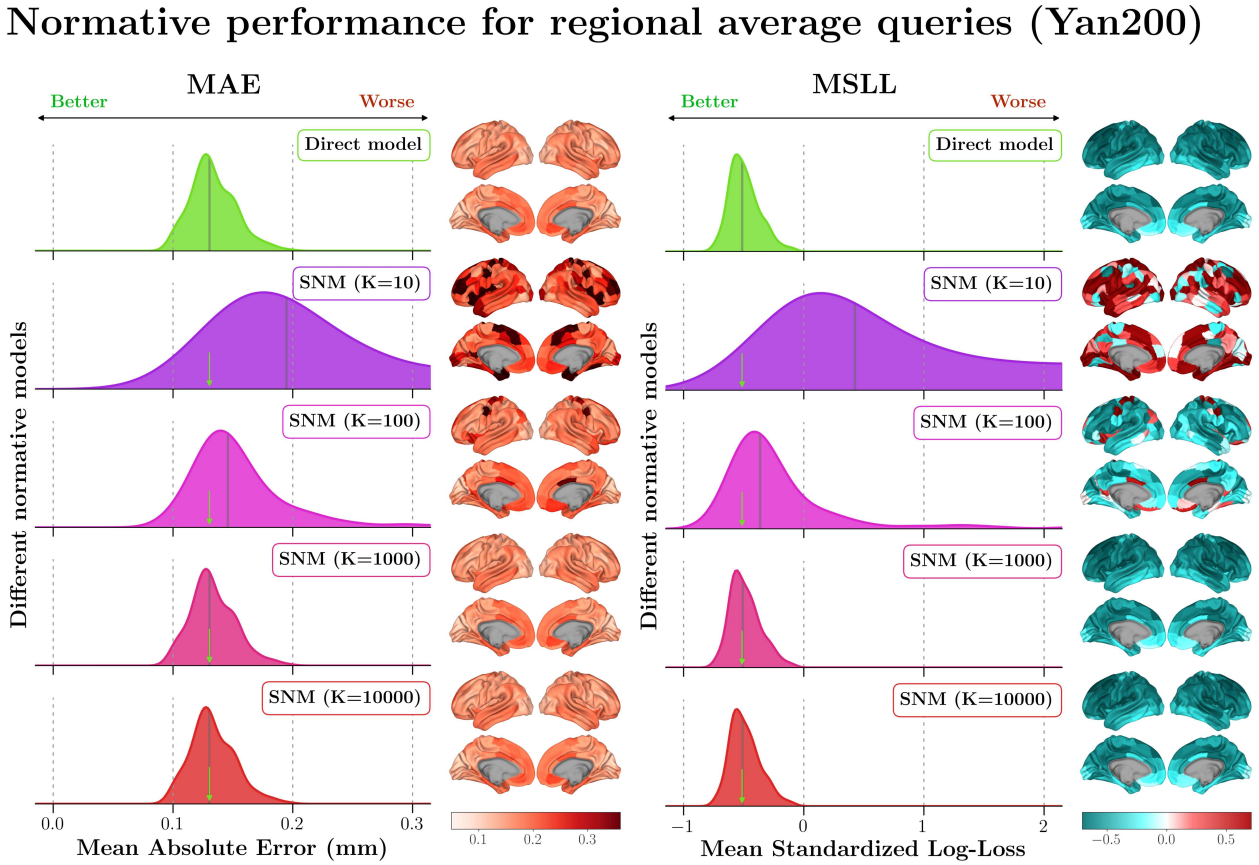
With as few as 10 modes, spectral normative models can produce acceptable normative estimates for some brain-wide signals, indicated by negative MSLL values. However, specific brain-wide signals, such as the average thickness of the ventral attention network and the limbic network (from Yeo's 7 functional networks[39]), exhibit poor normative estimates at this level, showing noticeably inferior MAE and MSLL scores compared to the direct model. However, including at least 100 modes in the spectral normative model yields performance comparable to the direct model. These findings were also extended in sensitivity analysis for brain-wide asymmetric queries (see Supplementary Figure S.9).



Supplementary Figure. S.2. Comparison of performance for different normative models assessing brain-wide spatial queries. Normative performance is evaluated using out-of-sample assessments of direct and spectral normative models. The top row displays the performance of the direct model, while the next four rows show the performance of spectral models with 10, 100, 1,000, and 10,000 eigenmodes, respectively. Mean absolute error (MAE, left) and mean standardized log-loss (MSLL, right) are used to quantify model performance. Violin plots illustrate the distribution of performance measures across 25 different brain-wide signals, including cortex-wide average thickness and thickness averaged over 7 and 17 functional networks. Median values are indicated by gray lines on the violin plots. A green arrow marks the median performance of the direct model over spectral performance distributions for visual comparison. Cortical surface projections show normative performance values for 7 functional networks, with shared colorbars across all five rows to facilitate visual comparison.

A.3.2 Regional normative estimates

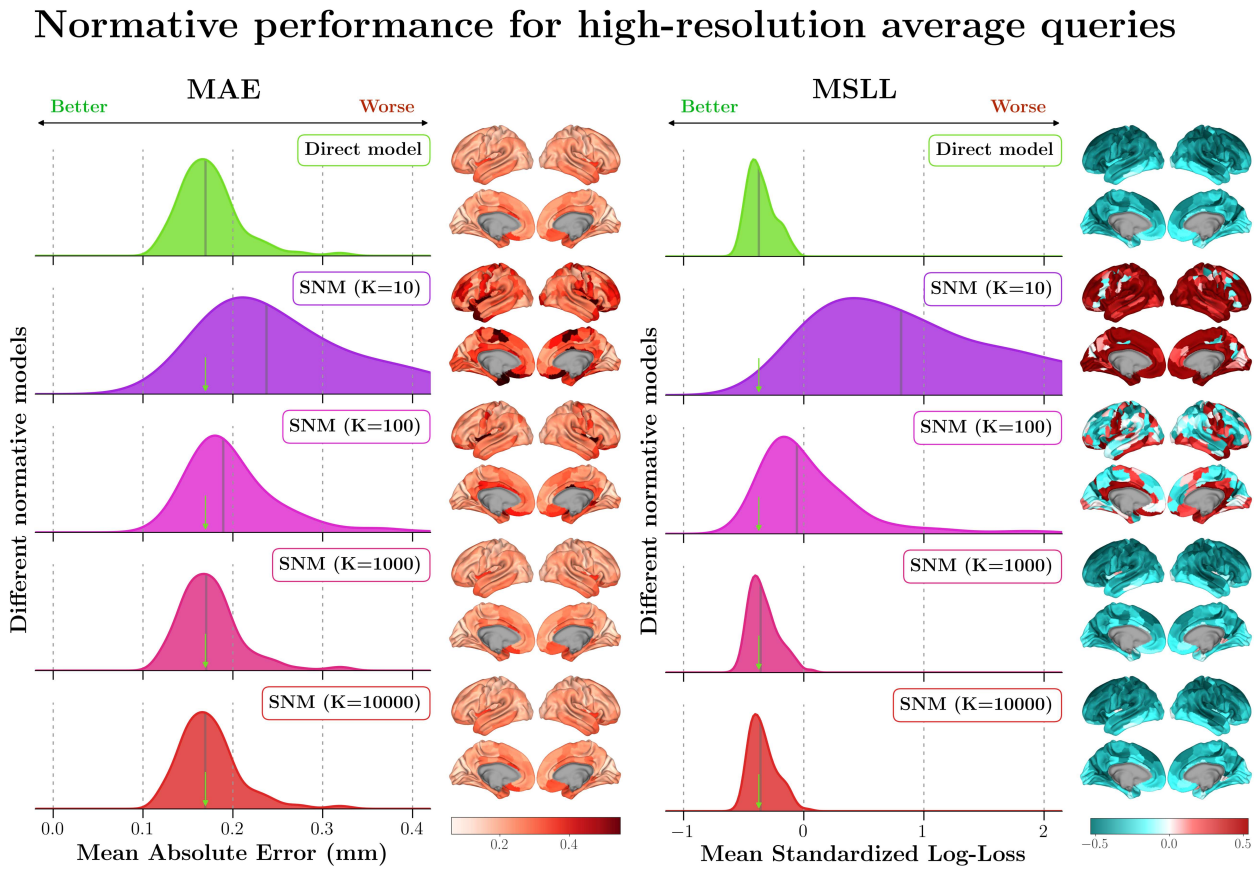
Figure S.3 extends the previous comparison to regional signals. At this level of spatial specificity, SNM with only 10 modes shows substantially inferior performance across several brain regions. Including the first 100 modes results in acceptable performance for most brain regions, except for regions on the post-central gyrus, insular cortex, and orbitofrontal cortex. Incorporating up to 1000 modes in the spectral model achieves performance matching that of the direct model across all brain regions. Further inclusion of modes (up to 10,000) provides minimal additional benefits in goodness of fit, as measured by MAE or MSLL. Sensitivity analyses evaluate the robustness of these findings to regional granularity and signal asymmetry (see Supplementary Figures S.6, S.10).



Supplementary Figure. S.3. Comparison of performance for different normative models assessing regional spatial queries. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. The top row shows the performance of the direct model, while the following four rows present the performance of spectral models with 10, 100, 1,000, and 10,000 eigenmodes, respectively. Mean absolute error (MAE, left) and mean standardized log-loss (MSLL, right) are used to quantify model performance. Violin plots display the distribution of performance measures across 200 different spatial queries, representing regional average thickness based on the Yan200 atlas. Median values are marked by gray lines on the violin plots. A green arrow marks the median performance of the direct model over spectral performance distributions for visual comparison. Cortical surface projections show normative performance values for each region, with shared colorbars across all five rows to aid visual comparison.

A.3.3 High-resolution normative estimates

Finally, Figure S.4 assesses SNM's performance in estimating high-resolution normative charts for spatial queries centered on various cortical surface voxels. Given the computational infeasibility of applying the direct model to every vertex, we evaluated a randomly selected subset of 400 vertices across the cortical surface. At this level of spatial specificity, spectral normative models with 10 and 100 modes fail to estimate normative ranges effectively compared to the direct model. However, the spectral model with 1000 modes achieves performance comparable to the direct model. Further inclusion of 10,000 modes yields marginal performance improvements. We also evaluated the sensitivity of these results to spatial granularity and asymmetry in normative queries (see Supplementary Figures S.7, S.11).



Supplementary Figure. S.4. Comparison of performance for different normative models assessing high-resolution spatial queries. Normative performance is evaluated through out-of-sample assessments of both direct and spectral normative models. The top row displays the performance of the direct model, while the subsequent four rows show the performance of spectral models with 10, 100, 1,000, and 10,000 eigenmodes, respectively. Mean absolute error (MAE, left) and mean standardized log-loss (MSLL, right) quantify model performance. Violin plots illustrate the distribution of performance measures across 400 different spatial queries, each representing an 8mm FWHM Gaussian kernel centered at a high-resolution cortical vertex (one vertex was randomly selected from each region of the Yan400 atlas). Median values are indicated by gray lines on the violin plots. A green arrow highlights the median performance of the direct model in comparison to SNMs. Cortical surface projections show normative performance values for each vertex, depicted within its associated Yan400 region, with shared colorbars across all five rows to facilitate visual comparison.

A.4 Sensitivity Analyses

Several sensitivity analyses were conducted to evaluate the robustness of our findings to various parameters. These analyses can be categorized into two main groups: those assessing the robustness of results against the granularity of the spatial queries, and those evaluating the impact of asymmetry in the spatial queries. These analyses will be respectively detailed in the ensuing sections.

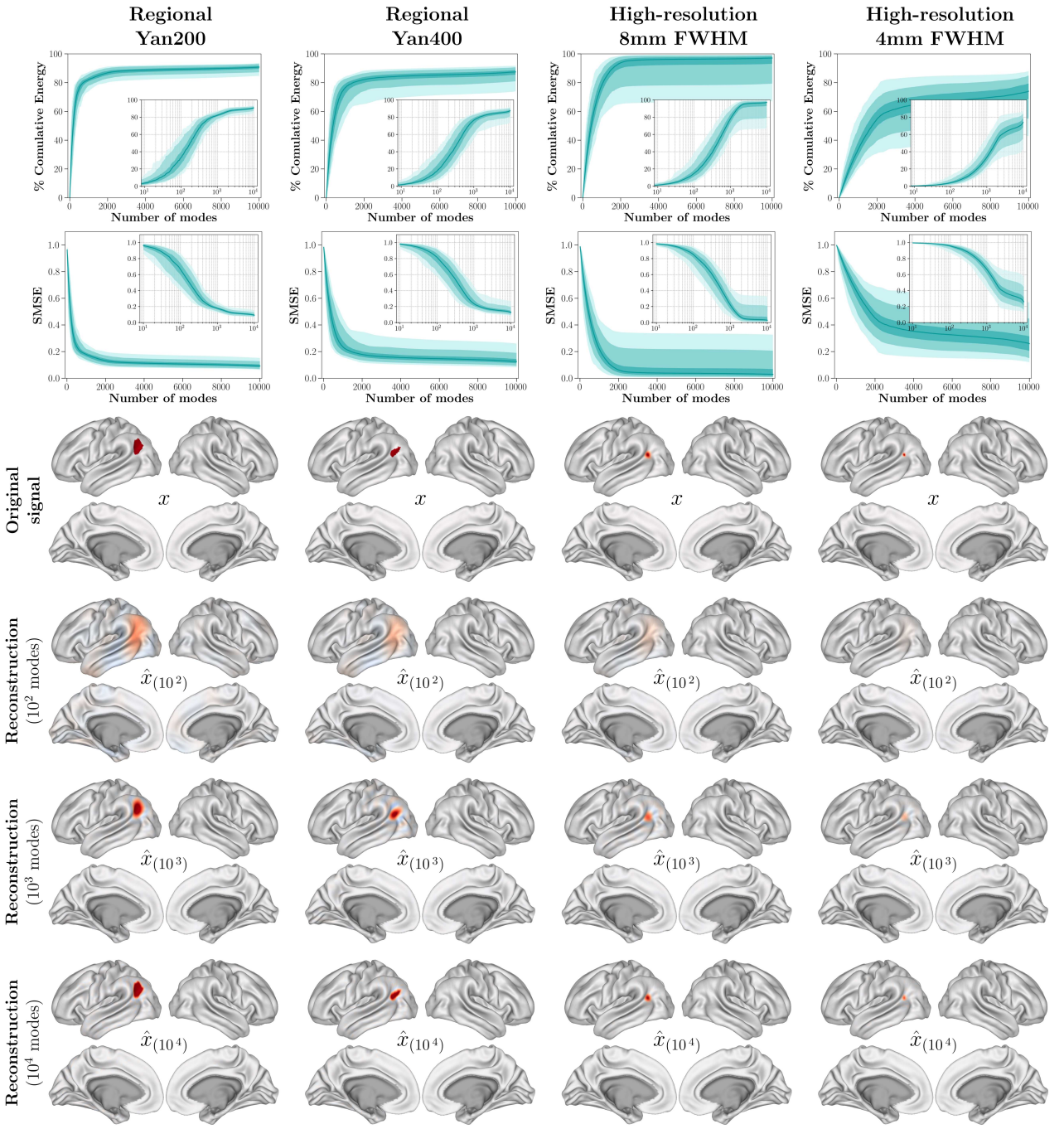
A.4.1 Effect of Spatial Granularity

To evaluate the sensitivity of our findings to the spatial granularity of normative queries, we repeat the evaluations for both regional and high-resolution queries with altered parameters. For regional queries, we assess the impact of increased spatial granularity by using the Yan400 parcellation template, which contains twice as many brain regions as the Yan200 atlas, thereby doubling the spatial granularity of the regional signals. Similarly, the spatial granularity of the high-resolution signals are altered by adjusting the spatial smoothing kernel strength. Specifically, high-resolution evaluations are repeated using a 4mm FWHM Gaussian smoothing kernel, which effectively doubles the spatial granularity of the high-resolution signals.

Supplementary Figure S.5 provides a comparative view of reconstruction accuracy performance as affected by signal granularity. As anticipated, increasing spatial granularity introduces higher spatial frequencies that are captured by higher graph frequencies. This is evident from the slight shift in proportional energy plots, which indicates an increased contribution of higher frequency modes to

signal approximation. Additionally, the results show that the signal reconstruction accuracy for both regional and high-resolution spatial queries reaches a plateau after including the first 1,000 and 2,000 eigenmodes, respectively. Finally, the relatively higher values of SMSE for high-resolution signals with a 4mm FWHM smoothing kernel suggest that high-resolution spectral evaluations are better suited for signals smoothed to 8mm FWHM.

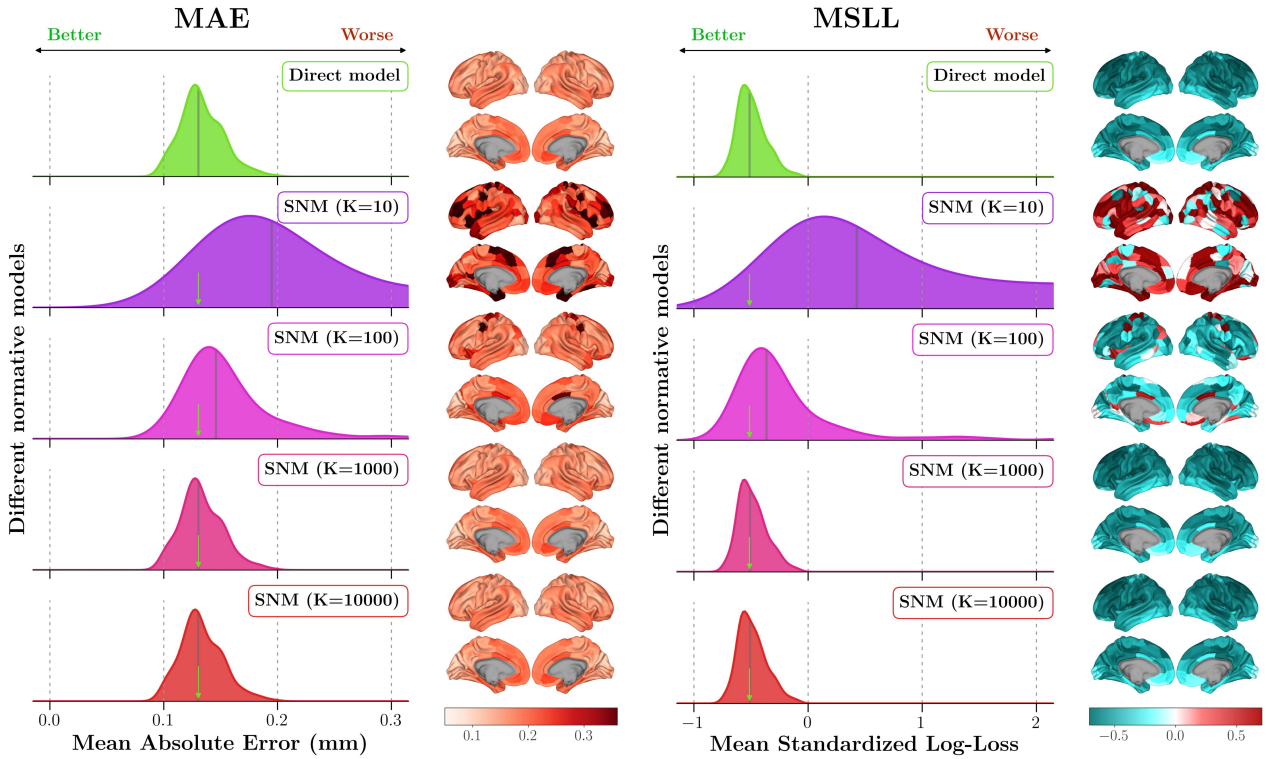
Supplementary Figure S.6 illustrates the impact of regional signal granularity on the goodness of fit of normative models. Repeating the evaluations shown in Figure S.3, we demonstrate that, irrespective of the regional signal granularity, the inclusion of 1,000 eigenmodes in the spectral model achieves performance comparable to that of a direct model. Similarly, Supplementary Figure S.7 repeats the evaluations presented in Figure S.4 and shows that a spectral model with 1,000 modes yields normative estimates comparable to those produced by the direct model. While these results indicate the robustness of the spectral normative framework to changes in spatial granularity, any query used with the spectral model should nevertheless be tested to assess adherence to the low-pass spectral regime (e.g., by evaluating reconstruction accuracy as quantified by metrics such as SMSE).



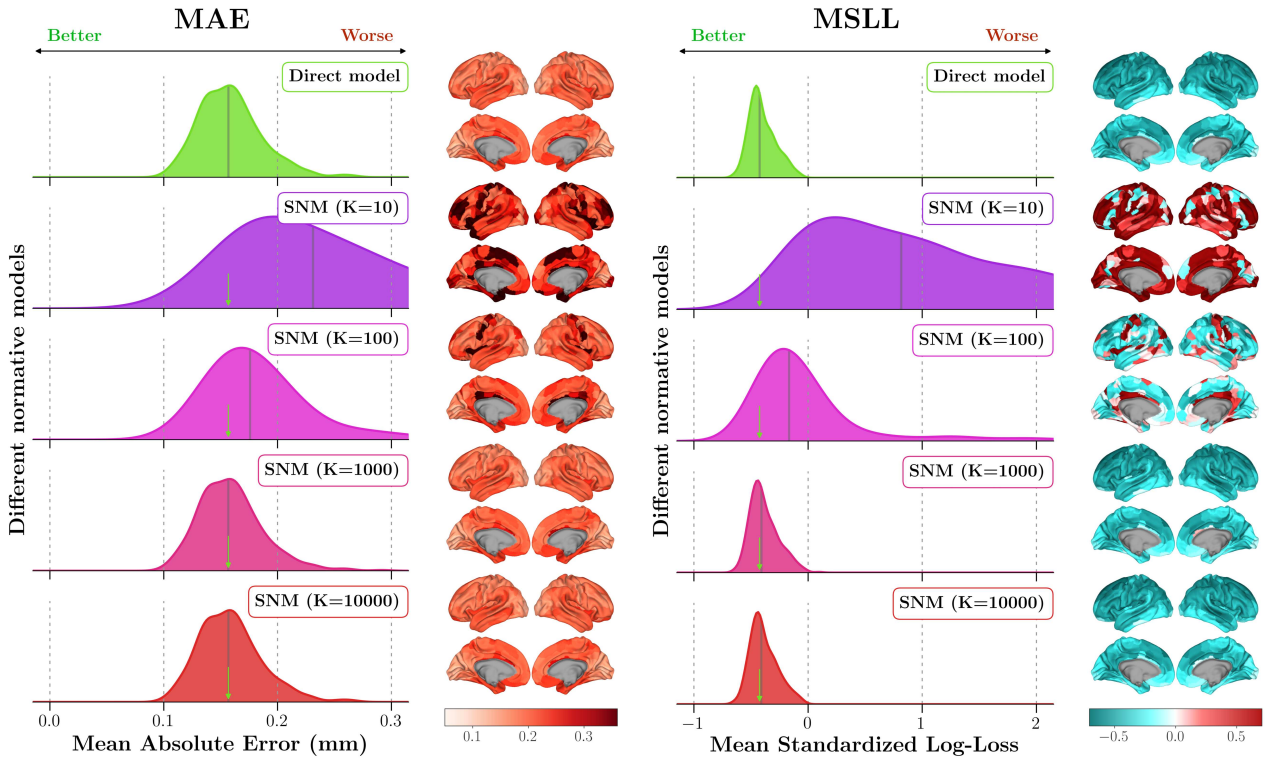
Supplementary Figure. S.5. Signal reconstruction accuracies across different granularity scales. The evaluations presented in Figure 2 are repeated to examine the effect of doubling the spatial granularity of the brain signals. The columns (from left to right) display spectral reconstruction accuracies for regional signals using the Yan200 and Yan400 atlases, as well as for high-resolution signals defined by 8mm FWHM and 4mm FWHM smoothing kernels. In the shaded line plots, the lines represent the median across all observations, while the shades indicate the [25, 75], [5, 95], and [1, 99] percentiles. The first row shows the proportional energy independently contributed by each eigenmode (logarithmic x-axis). The second row presents the standardized mean square error (SMSE) as a function of the number of low-frequency eigenmodes used for reconstruction (logarithmic x-axis for the insets). The third row illustrates one exemplary brain signal from each category, while the last three rows show the same signal reconstructed using 100, 1,000, and 10,000 eigenmodes, respectively.

Normative performance for regional queries: granularity

A | Performance on Yan200 atlas (average queries)



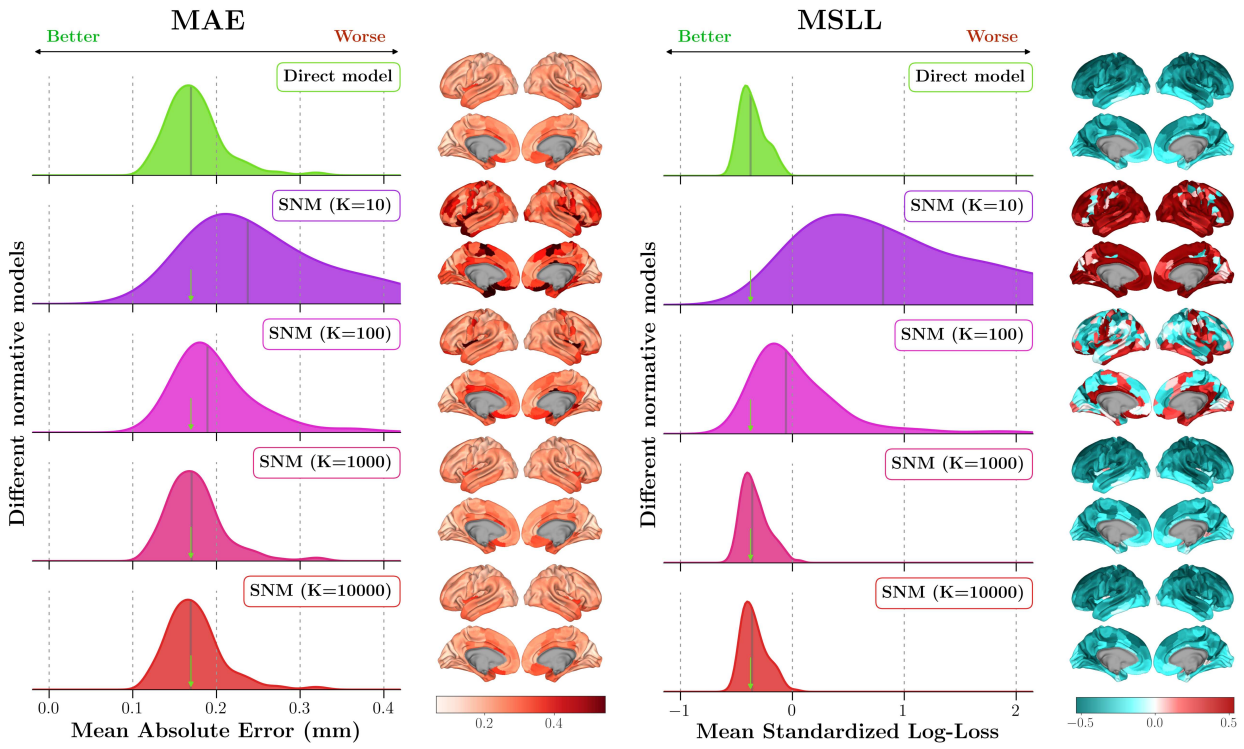
B | Performance on Yan400 atlas (average queries)



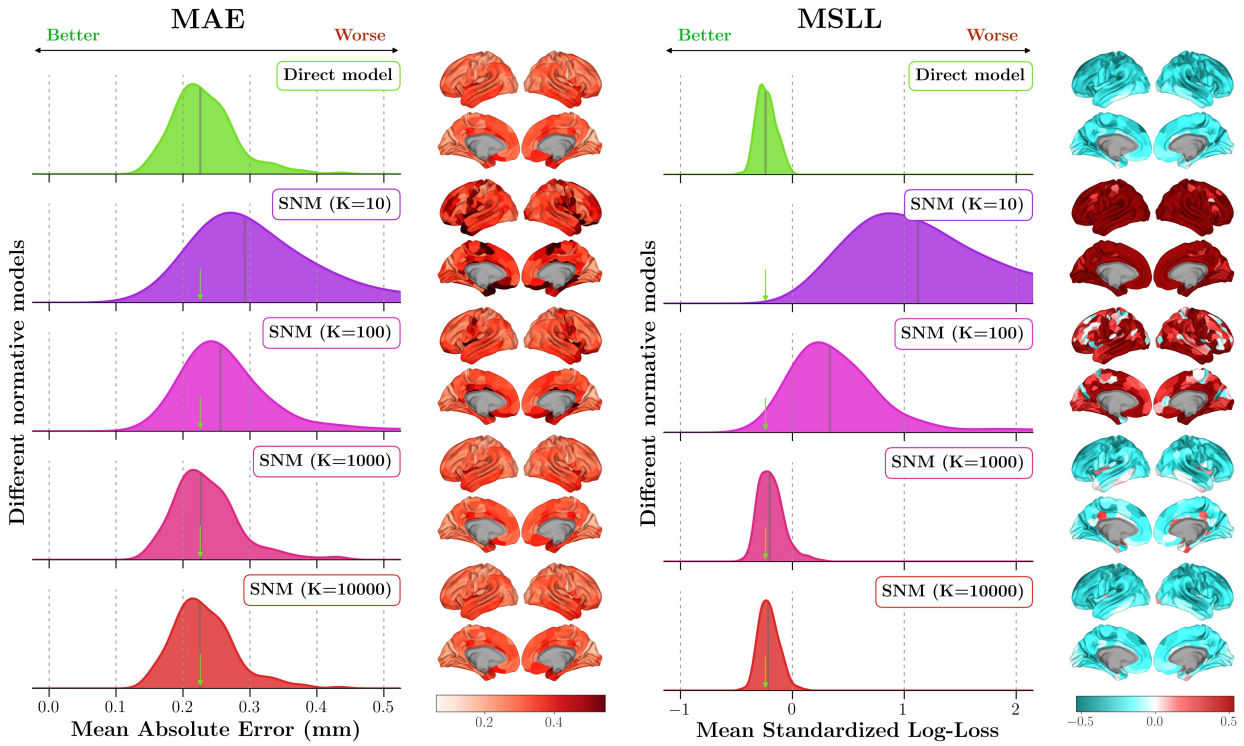
Supplementary Figure. S.6. Sensitivity of normative models assessing regional spatial queries at different granularity scales. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. **(A)** same as Figure S.3, repeated to aid visual comparison. **(B)** Same evaluations as panel A, but on regional queries from Yan400 atlas (increased spatial granularity).

Normative performance for high-resolution queries: granularity

A | Performance on 8mm FWHM smoothing (average queries)



B | Performance on 4mm FWHM smoothing (average queries)



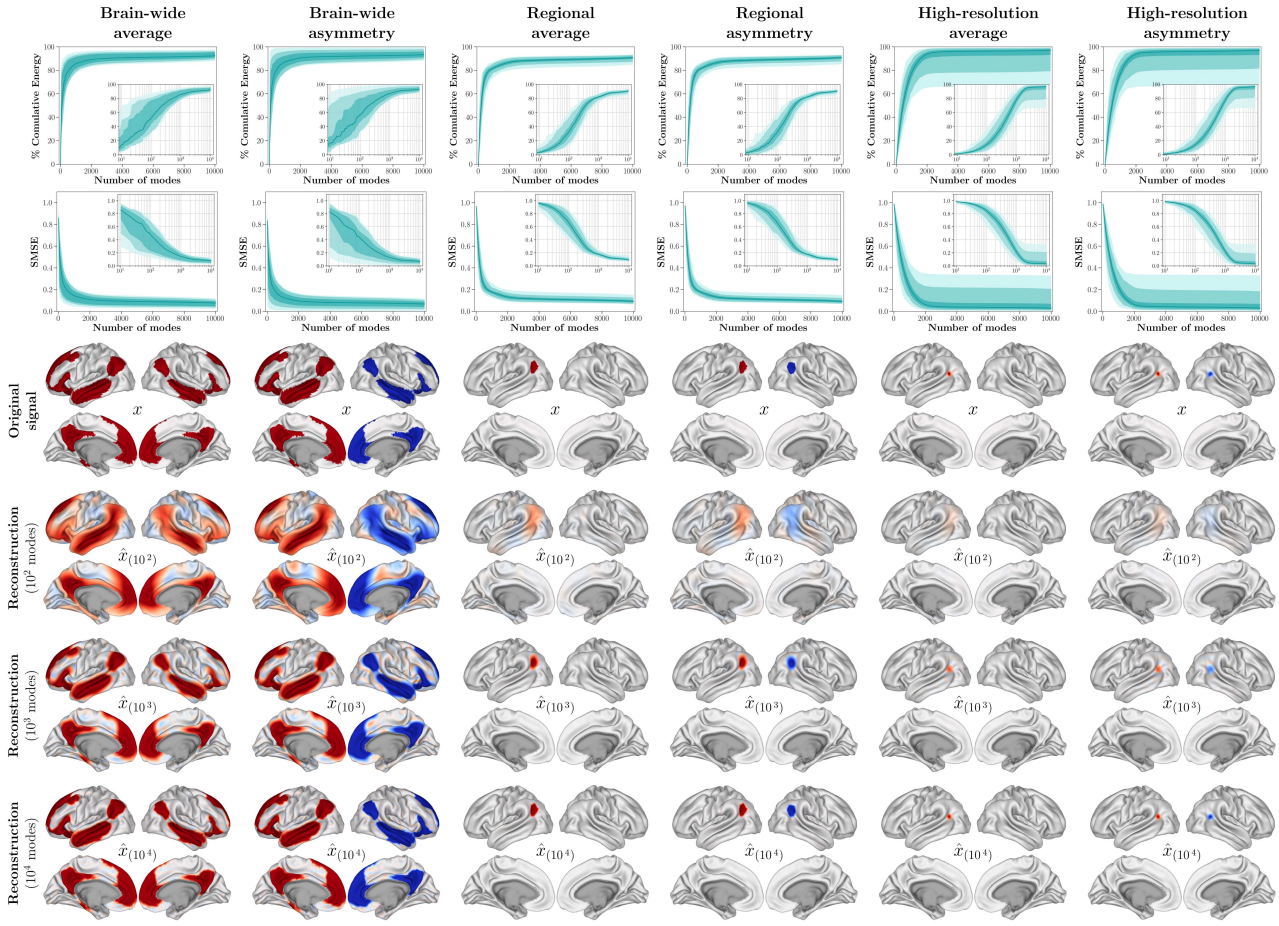
Supplementary Figure. S.7. Sensitivity of normative models assessing high-resolution spatial queries at different granularity scales. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. **(A)** same as Figure S.4, repeated to aid visual comparison. High-resolution spatial queries are generated using an 8mm FWHM Gaussian smoothing kernel. **(B)** Same evaluations as panel A, but on high-resolution queries generated using a 4mm FWHM Gaussian smoothing kernel (increased spatial granularity).

A.4.2 Effect of Spatial Asymmetry

A notable improvement of the spectral normative framework is its spatial versatility. This enables spectral models to extend beyond conventional average-based normative queries and gives them the ability to model a wider range of possible queries. As long as the query of interest adheres to the low-pass spectral regime, it can include various spatial patterns of interest. Asymmetric spatial queries are an interesting example of such patterns. Specifically, spectral normative models can be used to infer abnormal deviations in thickness lateralization by constructing a query that assesses the difference between the average thickness of a region/locus on the left hemisphere and its contralateral counterpart on the right hemisphere. This added benefit requires no extra training time and works out of the box. Hence, we evaluate the sensitivity of our main findings to the spatial asymmetry of normative queries.

We repeat the evaluations for brain-wide, regional, and high-resolution queries after introducing lateralization asymmetry. For brain-wide queries, we use a query that computes the average thickness of the query on the left cortex and subtracts it from the average of the part of the query that lies on the right cortex. For regional queries, given that the Yan atlas provides a homotopic parcellation of the cerebral cortex (i.e., parcels on the left and right cortices are paired), we evaluate 100 spatial normative queries, each assessing the difference in the average thickness of one homotopic parcel pair from Yan200 parcellation. Finally, for high-resolution queries, given that the fs-LR template space is aligned across the left and right cortices, we select pairs of homotopic vertices and construct spatial queries that compare their average thicknesses as described by 8mm FWHM smoothing kernels centered on respective vertex pairs.

Supplementary Figure S.8 provides a comparative view of reconstruction accuracy performance as affected by signal asymmetry. Results indicate that signal asymmetry has negligible impact on eigenmode energy proportions and reconstruction accuracy (quantified by SMSE). Comparing asymmetric queries to average queries (presented in Figure 2), we observe that asymmetric signals require a comparable number of modes for accurate reconstruction as average signals within each spatial query family (brain-wide, regional, and high-resolution). Regardless of signal symmetry, SMSE for the reconstruction of brain-wide, regional, and high-resolution signals respectively reaches below 0.2 after the inclusion of the first 400, 600, and 1,000 eigenmodes. Thus, the reported findings on the number of eigenmodes required to reconstruct brain signals remain consistent regardless of signal asymmetry.



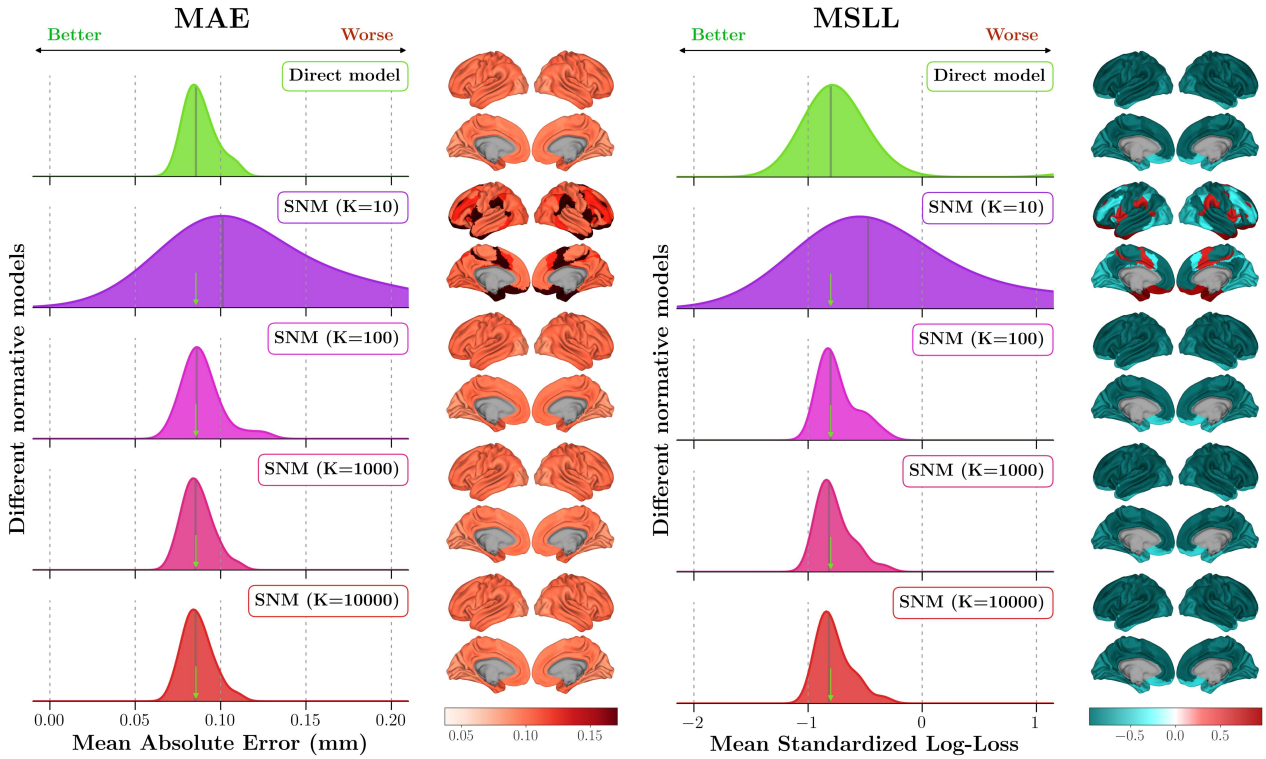
Supplementary Figure. S.8. Signal reconstruction accuracies for average vs. asymmetric signals. The evaluations presented in Figure 2 are repeated to examine the effect of introducing asymmetry to the brain signals. The columns display spectral reconstruction accuracies for pairs of (average, vs. lateralized/asymmetric) brain-wide, regional, and high-resolution signals. In the shaded line plots, the lines represent the median across all observations, while the shades indicate the [25, 75], [5, 95], and [1, 99] percentiles. The first row shows the proportional energy independently contributed by each eigenmode (logarithmic x-axis). The second row presents the standardized mean square error (SMSE) as a function of the number of low-frequency eigenmodes used for reconstruction (logarithmic x-axis for the insets). The third row illustrates one exemplary brain signal from each category, while the last three rows show the same signal reconstructed using 100, 1,000, and 10,000 eigenmodes, respectively.

Supplementary Figures S.9, S.10, and S.11 illustrate the impact of signal asymmetry on the goodness of fit of the associated normative models. Repeating the evaluations shown in Figures S.2, S.3, and S.4 shows that irrespective of the normative modeling framework (direct vs. spectral), goodness of fit, particularly when assessed by MSLL, is lower for lateralized norms. This effect is more pronounced at higher resolutions, suggesting that healthy norms of thickness lateralization are better studied at lower spatial specificity (functional networks or parcels). As the goodness of fit in predicting the central tendency (quantified via MAE) is less affected, we speculate that this reduction in the accuracy of lateralized queries is due to the misaligned gyrification patterns between contralaterally aligned high-resolution vertices, rendering lateralization evaluations less meaningful at higher resolutions.

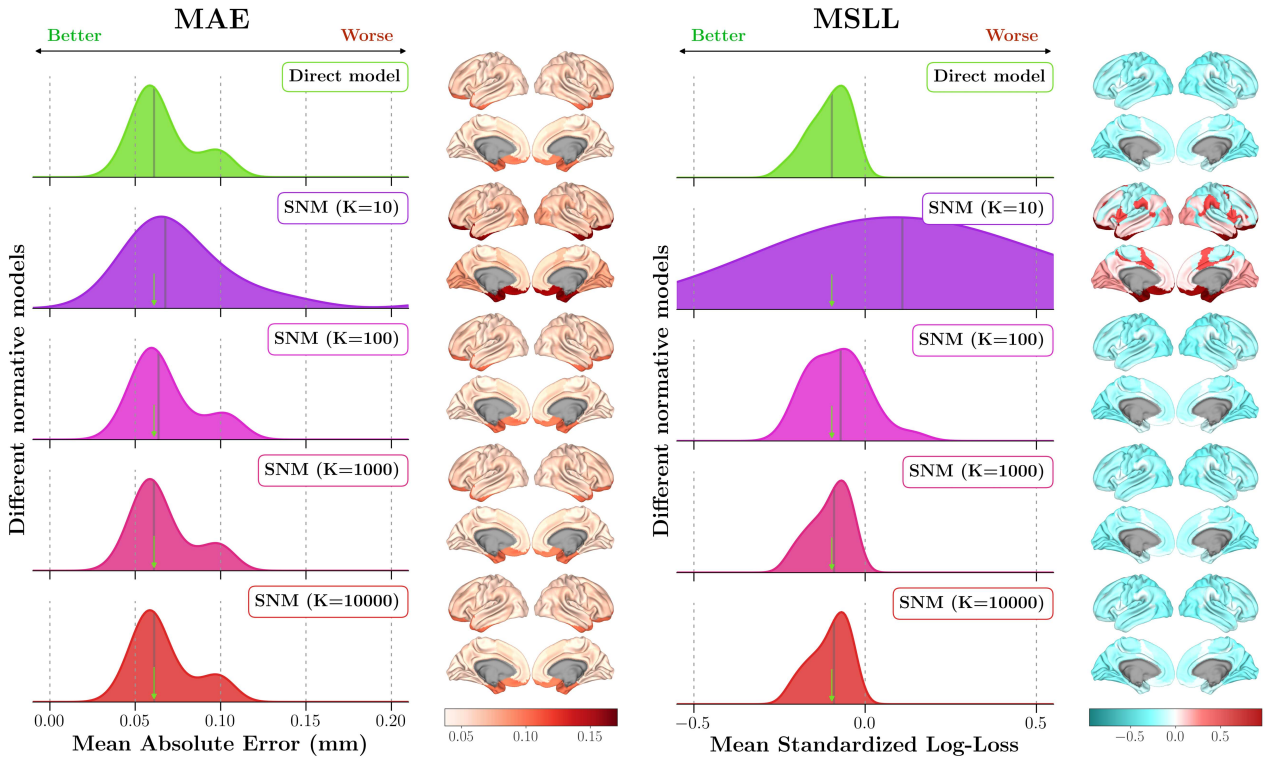
Despite this effect, which was not specific to the spectral model, we observe that, similar to the main findings, the inclusion of 100 eigenmodes is sufficient to model brain-wide normative thickness ranges with the spectral model, achieving performance on par with the direct model. As with the main findings, the number of modes required to achieve comparable performance to that of a direct model at the resolution of regional or vertex-wise queries increases to 1,000 eigenmodes. These sensitivity evaluations demonstrate that our findings are robustly replicable for signals with asymmetry, and that the inclusion of 1,000 eigenmodes can provide comparable performance to that of a conventional model across a wide range of spatial queries.

Normative performance for brain-wide queries: symmetry

A | Performance on average queries (symmetric)



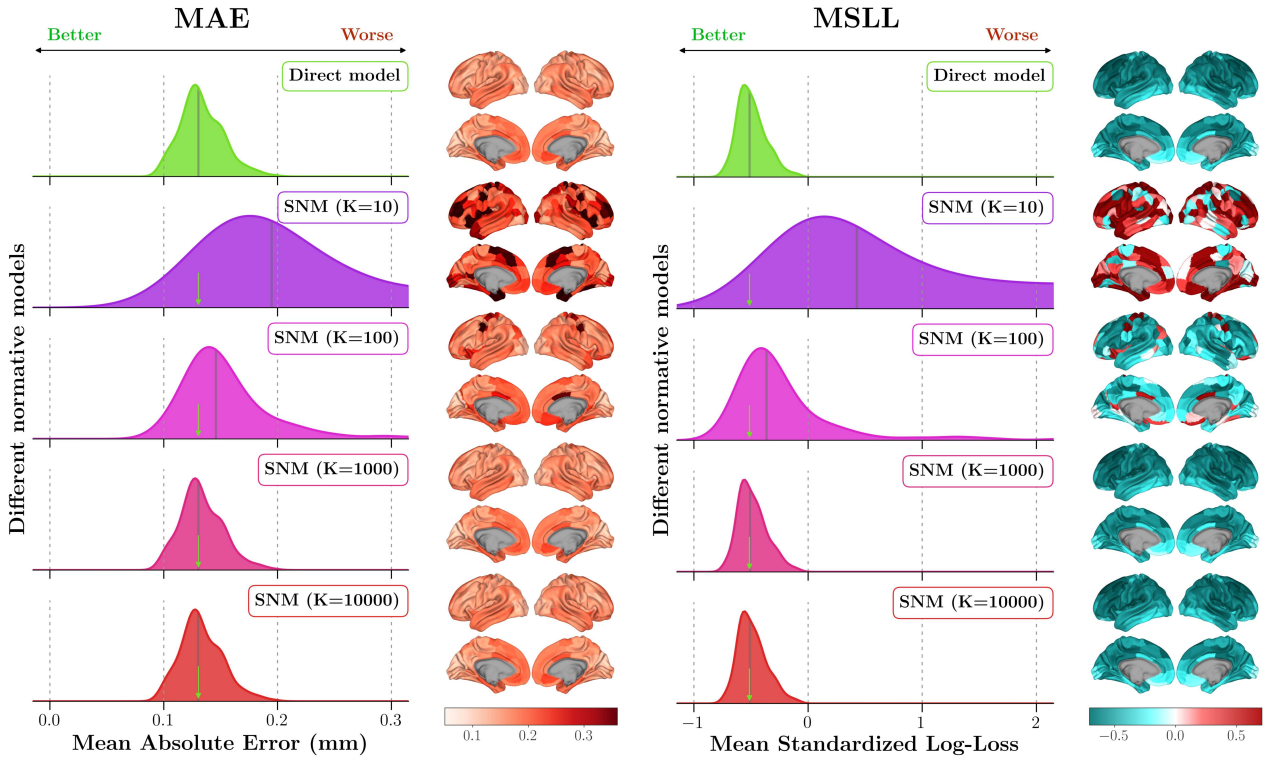
B | Performance on asymmetry queries



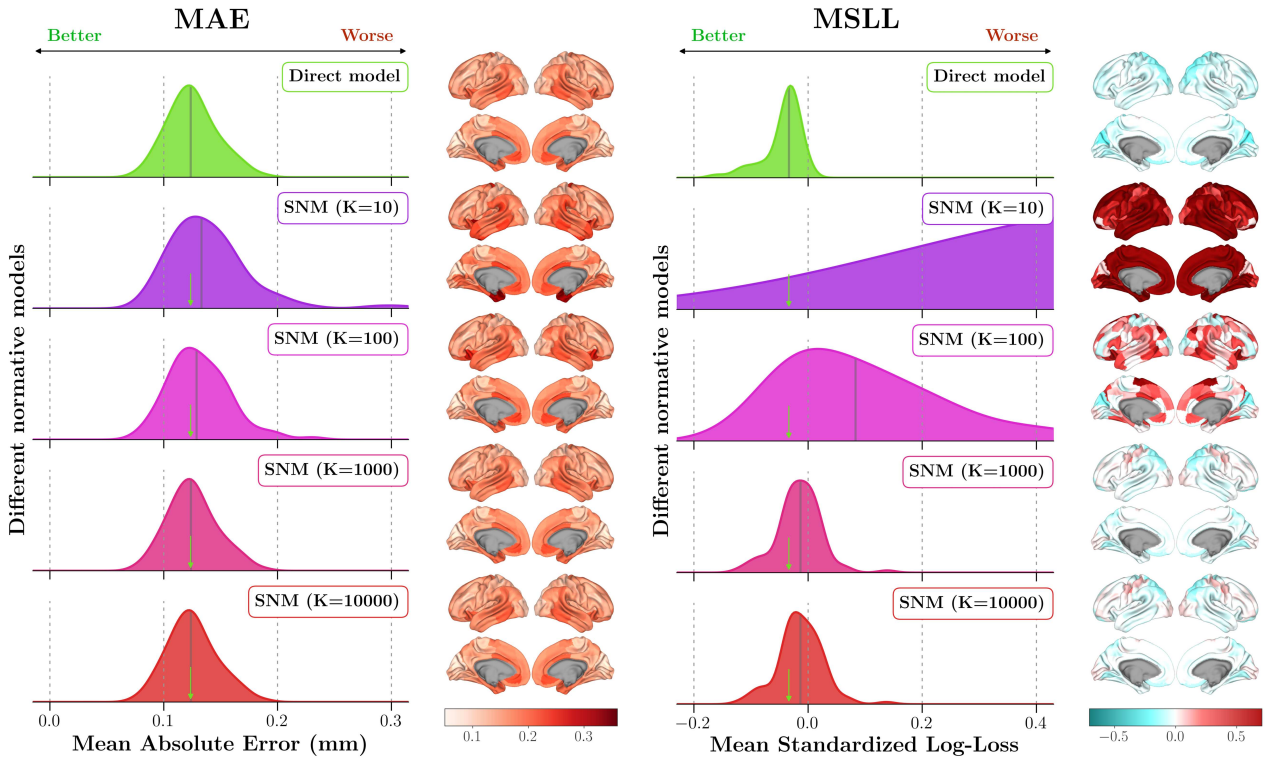
Supplementary Figure. S.9. Sensitivity of normative models to assessing lateralized brain-wide spatial queries. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. **(A)** same as Figure S.2, repeated to aid visual comparison. **(B)** Same evaluations as panel A, but on lateralized counterparts of the brain-wide queries.

Normative performance for regional queries: symmetry

A | Performance on average queries (symmetric)



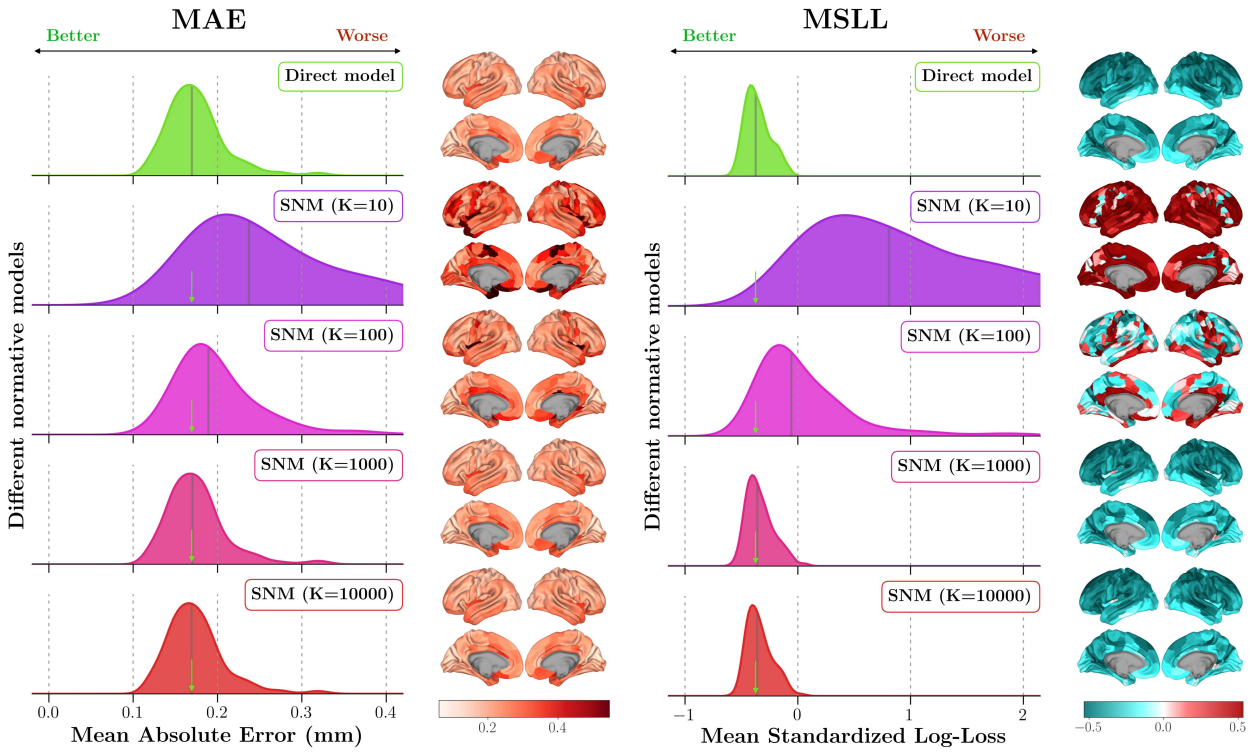
B | Performance on asymmetry queries



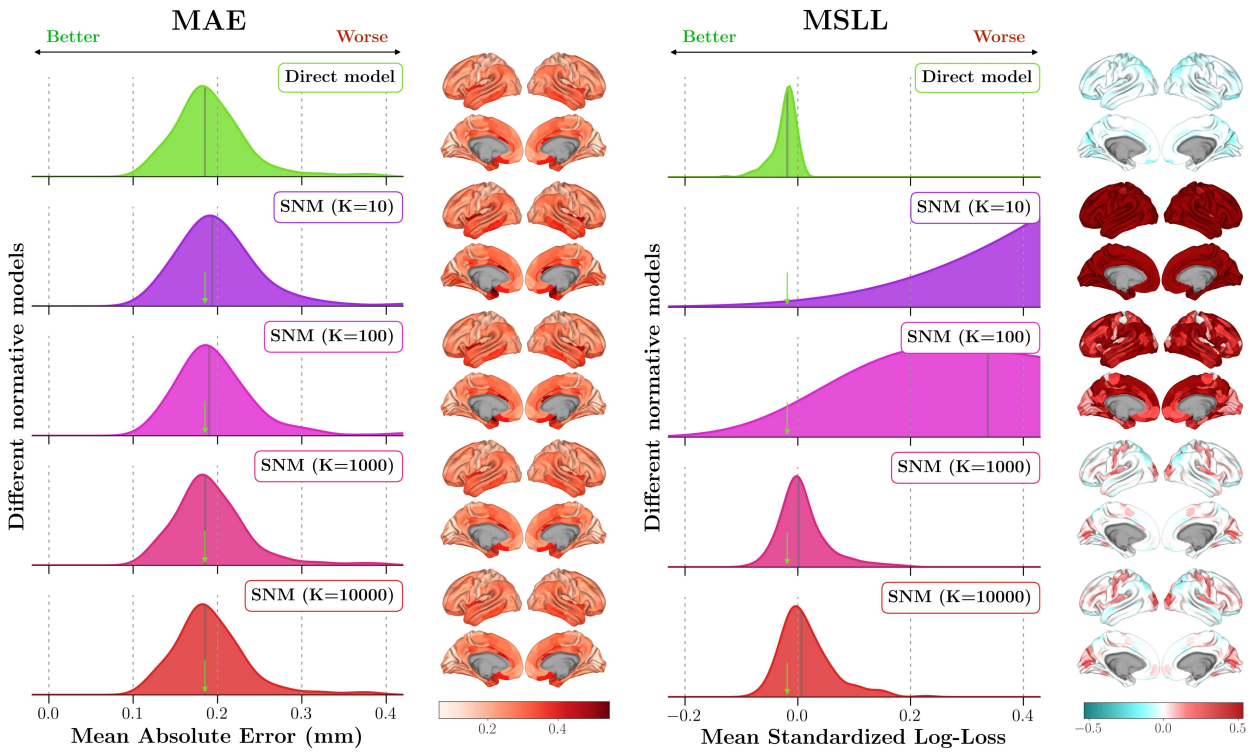
Supplementary Figure. S.10. Sensitivity of normative models to assessing lateralized regional spatial queries. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. **(A)** same as Figure S.3, repeated to aid visual comparison. **(B)** Same evaluations as panel A, but on lateralized counterparts of the regional queries.

Normative performance for high-resolution queries: symmetry

A | Performance on average queries (symmetric)



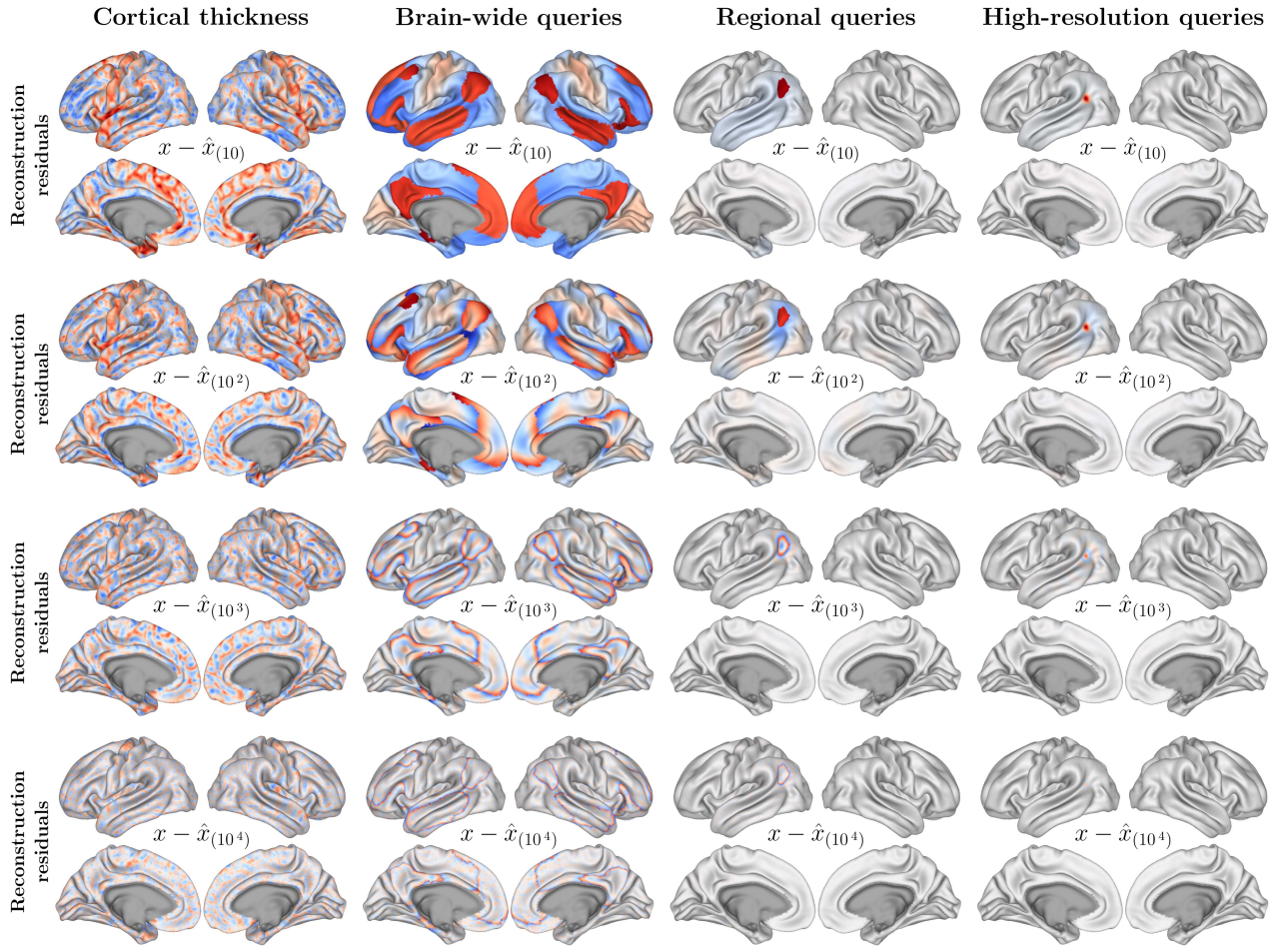
B | Performance on asymmetry queries



Supplementary Figure. S.11. Sensitivity of normative models to assessing lateralized high-resolution spatial queries. Normative performance is evaluated using out-of-sample assessments of both direct and spectral normative models. **(A)** same as Figure S.4, repeated to aid visual comparison. **(B)** Same evaluations as panel A, but on lateralized counterparts of the high-resolution queries.

A.5 Reconstruction Residuals

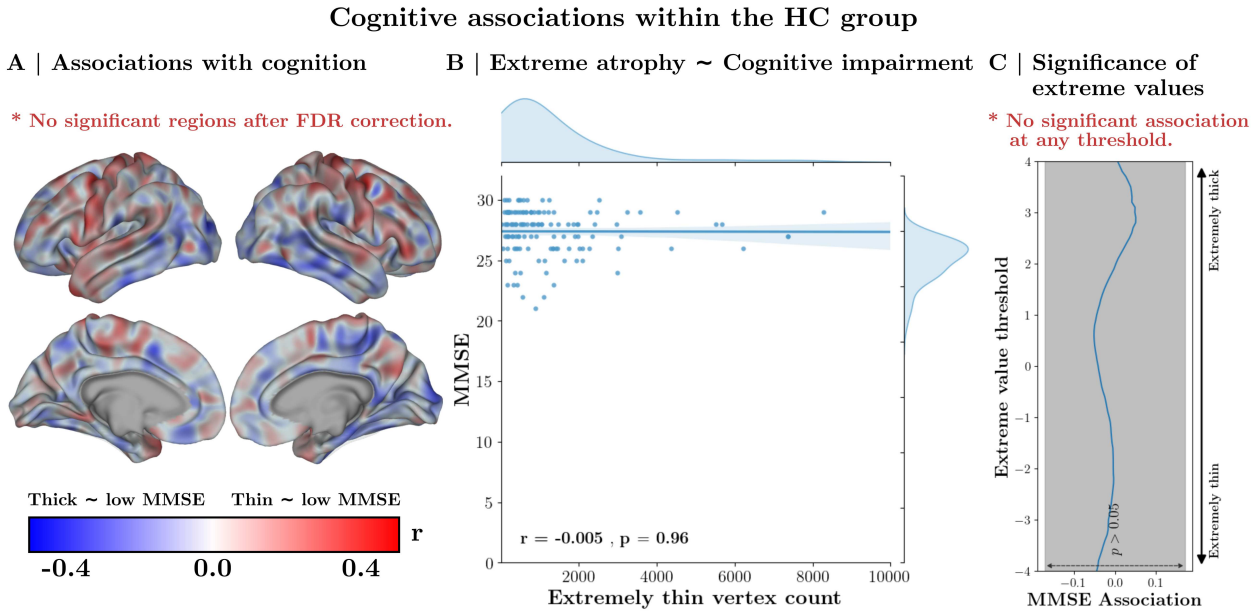
In Figure 2, cortical projections of exemplary brain signals were presented alongside the original signal (last four rows). Supplementary Figure S.12 shows the residuals of low-pass filtered graph approximations of the same signals. For four different low-pass approximations ($k = 10, 10^2, 10^3, 10^4$), the approximation residuals ($x - \hat{x}_{(k)}$) are projected onto the cortical surface. As anticipated, the magnitude of the approximation error (indicated by color intensity in the cortical projections of residuals) decreases as the number of incorporated eigenmodes increases. Furthermore, with the inclusion of more modes, only regions containing high spatial frequency information (such as the transition loci along parcellation borders) have higher approximation errors. This verifies our expectations that low-pass filtering by eigenmodes can accurately approximate smooth signals (i.e. lower graph frequencies).



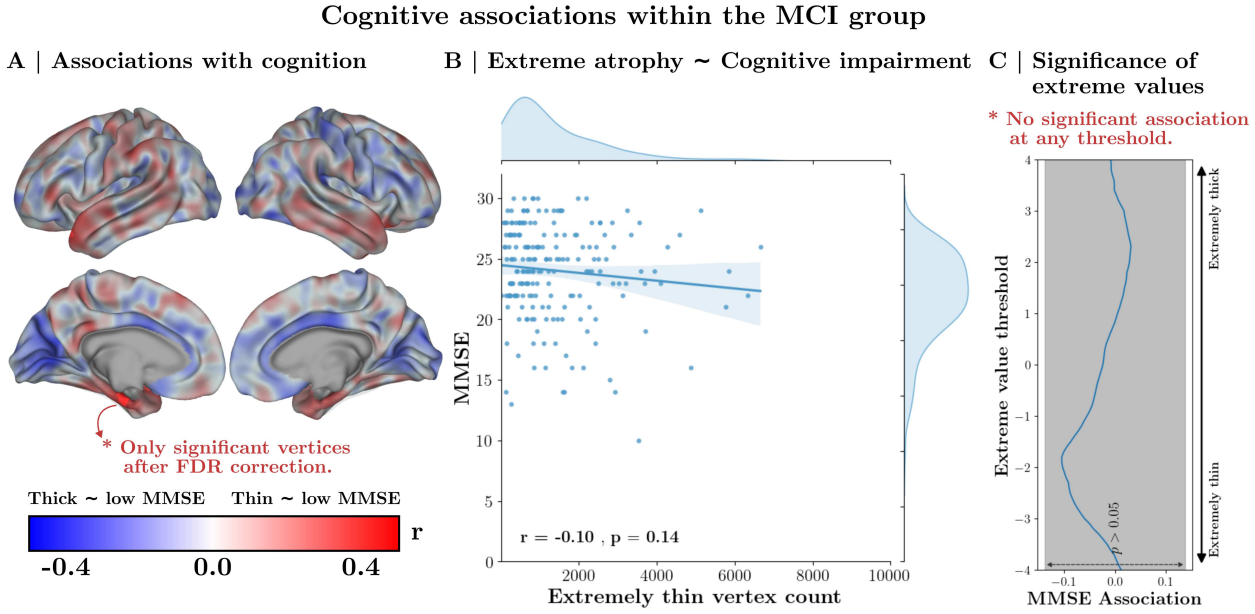
Supplementary Figure. S.12. Signal reconstruction residuals. The columns (from left to right) display the cortical projections of approximation residuals for individual cortical thickness signals, as well as brain-wide, regional, and high-resolution spatial queries (similar to Figure 2). The rows show the residuals for the same signal reconstructed using 10, 100, 1,000, and 10,000 eigenmodes, respectively.

A.6 Within-group Cognitive Associations

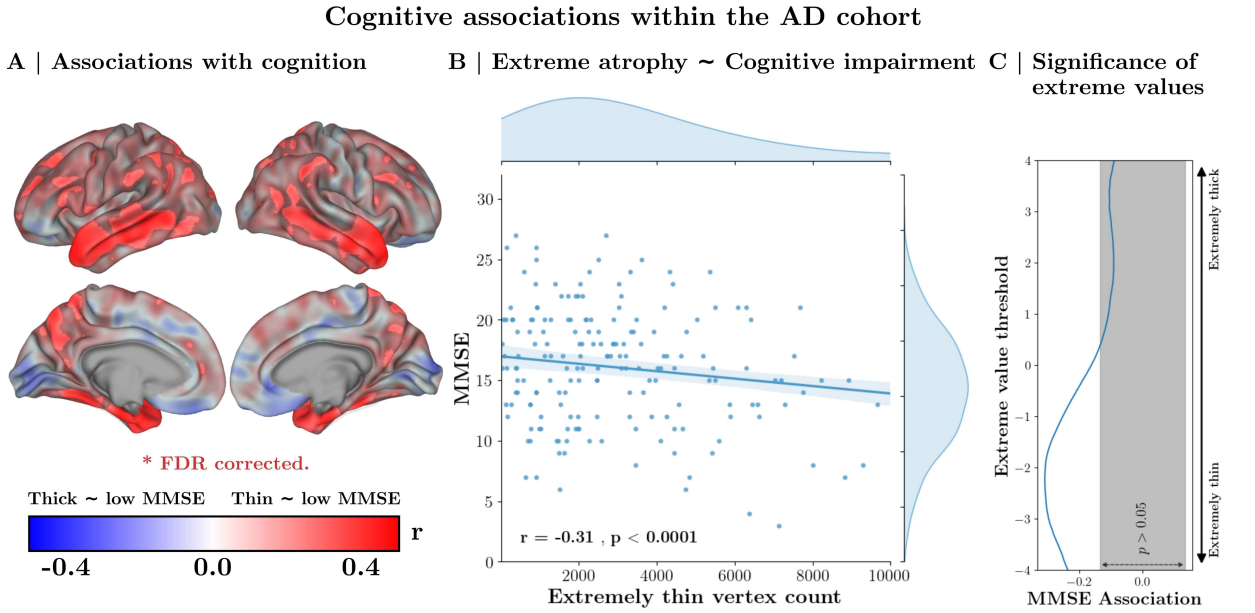
To assess the significance of cognitive associations within each cohort, we repeated the analyses from Section 2.5 SNM Uncovers Cortical Signatures of Atrophy in Alzheimer's Disease on the HC, MCI, and AD subsets. For healthy controls (HC), high-resolution cortical associations did not reach significance in any region after FDR correction (Supplementary Figure S.13), and ETVC showed no association with cognitive performance, irrespective of the z-threshold. In the MCI cohort, ETVC was similarly unrelated to cognitive impairment (Supplementary Figure S.14), but regional analyses identified significant atrophy in the left entorhinal cortex and left frontal pole to be associated with cognitive impairment. In contrast, the AD cohort displayed widespread neocortical atrophy patterns significantly linked to cognitive impairment (Supplementary Figure S.15), with reductions in the temporal pole, superior and middle temporal gyri, entorhinal cortex, parahippocampal cortex, precuneus, supramarginal gyrus, and superior, middle, and inferior frontal gyri. Additionally, ETVC showed a strong association with cognitive impairment in the AD cohort, specific to extremely thin vertices. Interestingly, this predictive capacity diminished when the z-score threshold was raised to zero or any positive value, with optimal predictions achieved between z-score thresholds of -2 and -3. Together, these findings suggest that the derived normative atrophy marker of cognitive impairment is specific to AD and most sensitive to extreme signs of atrophy. However, it should be noted that limited variability in cognitive performance within HC and MCI cohorts has potentially reduced the sensitivity for detecting smaller effects in these groups.



Supplementary Figure. S.13. Cognitive Association Tests within the Healthy Cohort (HC). (A) Vertex-level normative assessments were tested for linear associations with cognitive performance (MMSE), but no vertices reached significance after nonparametric FDR correction. **(B)** Extreme value statistics for severe atrophy (ETVC, z-score threshold of -1.96) were assessed for their predictive capability on cognitive performance, revealing no sensitivity of ETVC to cognitive variations in HC. **(C)** A range of alternative z-score thresholds was tested, demonstrating that ETVC does not predict cognitive performance in HC, regardless of threshold choice.



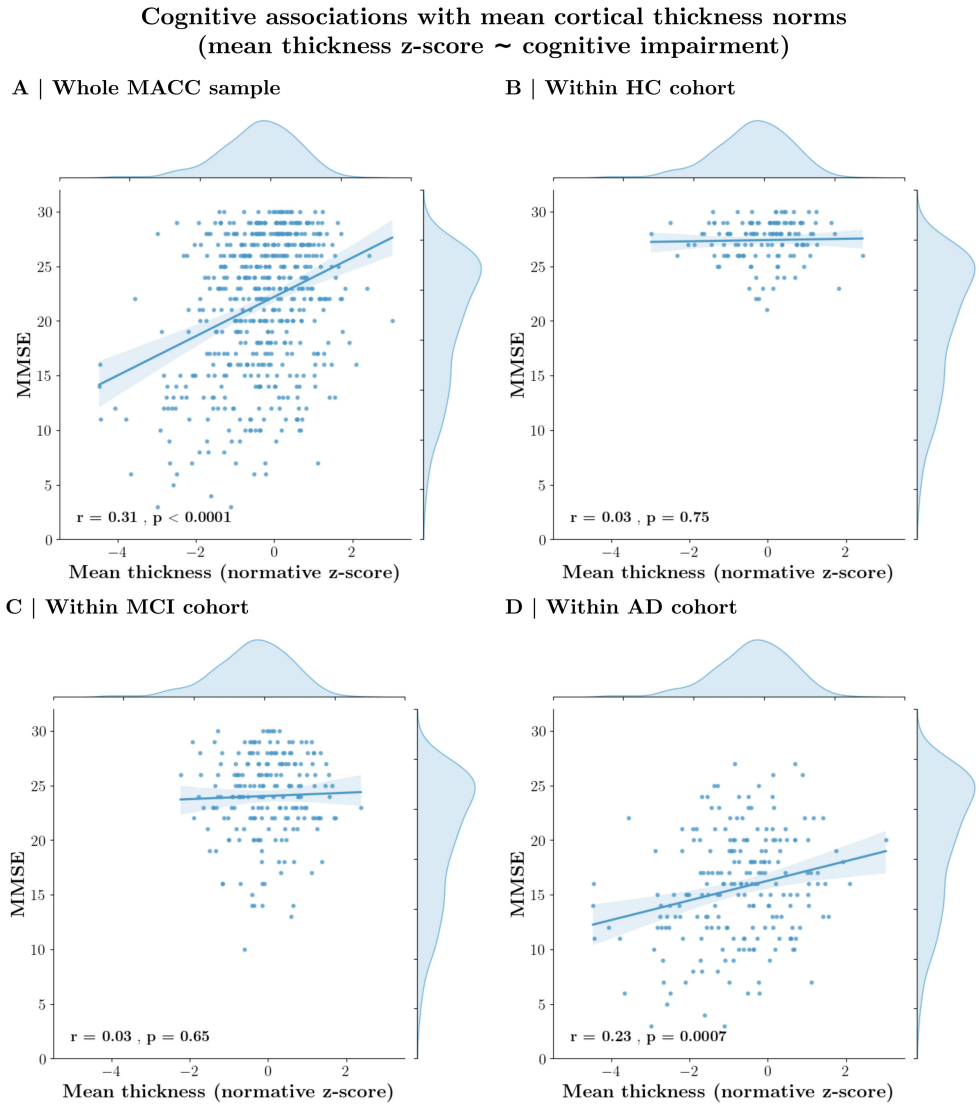
Supplementary Figure. S.14. Cognitive Association Tests within the MCI Cohort. (A) Vertex-level normative assessments were tested for linear associations with cognitive performance (MMSE) within the MCI cohort. A limited set of vertices within the entorhinal cortex reached statistical significance, indicating that extreme atrophy in this region may serve as a marker of cognitive impairment. **(B)** Extreme value statistics for severe atrophy (ETVC, z-score threshold of -1.96) were evaluated for their predictive capability on cognitive performance, showing no sensitivity of ETVC to cognitive variations within MCI. **(C)** Testing a range of alternative z-score thresholds confirmed that ETVC does not predict cognitive performance in MCI, regardless of the chosen threshold.



Supplementary Figure. S.15. Cognitive Association Tests within the AD Cohort. (A) Vertex-level normative assessments were tested for linear associations with cognitive performance (MMSE) within the AD cohort. A widespread set of cortical regions reached statistical significance, highlighting extensive brain-wide associations as markers of AD-related cognitive impairment. **(B)** Extreme value statistics for severe atrophy (ETVC, z-score threshold of -1.96) were evaluated for predictive capability regarding cognitive performance, revealing significant predictive power of ETVC to capture cognitive variations within the AD cohort. **(C)** Testing across a range of alternative z-score thresholds indicated that only ETVCs with a negative z-threshold significantly predicted cognitive performance in the AD cohort. Notably, extreme thinning (z-thresholds from -2 to -3) proved to be the strongest predictor of cognitive impairment.

A.7 Brain-wide Cognitive Associations

To highlight the benefits of inferring high-resolution spatial norms, we compare the predictive performance of the ETVC normative measure derived from high-resolution brain charting with that of a brain-wide normative z-score computed from a spatially coarse brain chart of mean cortical thickness (averaged across the entire cortex). Using the fine-tuned SNM, we calculate individual z-scores for deviations in average cortical thickness and replicate the evaluations from Figures 5C, S.13B, S.14B, and S.15B. As shown in Figure S.16, the results consistently indicate that while brain-wide normative measures can predict cognitive impairments, high-resolution ETVC measures outperform them. Specifically, ETVC achieves higher prediction accuracy in the entire sample (ETVC: $|r| = 0.45$, brain-wide: $|r| = 0.31$) and in the AD subsample (ETVC: $|r| = 0.31$, brain-wide: $|r| = 0.23$). Notably, both normative metrics were specifically sensitive to AD-related cognitive differences as neither predicted cognition in the HC or MCI subsets.



Supplementary Figure. S.16. Cognitive Association Tests with a Coarse Brain-Wide Normative Metric. Normative deviations (z-scores) in mean cortical thickness were evaluated for their predictive capability regarding cognitive performance (quantified by MMSE). Predictive performance was assessed in **(A)** the entire MACC dataset, as well as within the **(B)** HC, **(C)** MCI, and **(D)** AD cohorts individually.

A.8 Heterogeneity Landscape Subgroups

Figure 6D presents snapshots of cortical deviation patterns across various regions within the 2-dimensional atrophy landscape. To more comprehensively illustrate the overall cortical signature in this landscape, we have provided four supplementary videos (one each for the whole sample, HC, MCI, and AD subsets). These videos display changes in the cortical deviation patterns as the local region is gradually shifted over a circular area within the 2-dimensional space.