The impact of heterogeneous spatial autocorrelation on comparisons of brain maps

Leech R[1], Smallwood JS[2], Moran R[1], Jones EJH[3], Vowles N[1], Leech D[4], Viegas EM[1], Turkheimer FE[1], Alberti F[5], Margulies, D[5], Jefferies E[6], Bernhardt B[7], Váša F[1]

1 Institute of Psychiatry, Psychology & Neuroscience, King's College London
2 Department of Psychology, Queens University
3 Centre for Brain & Cognitive Development Birkbeck, University of London
4 Department of Economics, University of Warwick
5 Integrative Neuroscience and Cognition Center, University of Paris.
6 Department of Psychology, University of York
7 Montreal Neurological Institute-Hospital, McGill University

Abstract
It is increasingly common to statistically compare macroscopic brain maps to assess how spatially similar they are. Due to the presence of spatial autocorrelation, statistical inference can be challenging; to address this, random permutation approaches based on null models are widely used. Here, we show how that the presence of heterogeneity in the spatial autocorrelation across brain maps may affect statistical inference in common approaches for spatially correlated maps. In response, we highlight the need to explicitly model spatial processes, including non-stationarity , for more accurate statistical inference. We illustrate how a Bayesian spatial regression approach can be applied to compare functional and structural cortical brain maps, even in the presence of heterogeneity. By explicitly modelling spatial processes underlying the data, a much wider and more sophisticated range of neurobiological questions can be answered about the relationship between brain maps than with current approaches.

Introduction
With the increasing availability of openly shared datasets and results (Milham et al., 2018; Poldrack and Gorgolewski, 2014), a growing number of studies have been comparing newly-generated whole-brain maps to existing reference patterns of spatial variability generated using a range of imaging and other modalities. For example, recent studies have compared whole-brain spatial maps measured with fMRI at rest with a range of morphological, molecular and genetic whole-brain spatial patterns (Hansen et al., 2022; Larivière et al., 2023; Paquola et al., 2019; Váša et al., 2020; Wagstyl et al., 2024).

Statistical comparisons between whole-brain maps are challenging because of the marked and ubiquitous presence of spatial autocorrelation. This phenomenon refers to spatially proximal measurements being more likely similar, regardless of how the whole-brain maps were derived (Burt et al., 2020; Leech et al., 2023; Shinn et al., 2022). Spatial autocorrelation results from a mix of biological processes underlying spatial neural organisation as well as measurement problems, interpolation induced by data processing and the somewhat arbitrary spatial resolution of the imaging data. In the presence of spatial autocorrelation, the significance of similarity between pairs of brain maps is likely to be inflated. In fact, p-values resulting from measures of statistical association such as the correlation coefficient naïvely assume that underlying datapoints are independent (and the number of datapoints directly translates into degrees of freedom for the correlation). This assumption is violated by spatial autocorrelation (Alexander-Bloch et al., 2018; Váša and Mišić, 2022; Worsley et al., 1999).

Approaches for addressing the consequences of spatial autocorrelation on comparisons of brain maps have predominantly involved the use of null models with preserved spatial autocorrelation (Alexander-Bloch et al., 2018; Burt et al., 2020; Markello and Misic, 2021; Vos de Wael et al., 2020). One approach involves iterative randomisation and smoothing of brain maps to generate surrogate

maps with randomized topography but preserved spatial autocorrelation (Burt et al., 2020, 2018). Another popular approach applies a spatial permutation or "spin" to a spherical projection of the data (Alexander-Bloch et al., 2018; Váša et al., 2018). Alternative approaches propose to permute spectral decompositions of the geodesic distance matrix (Vos de Wael et al., 2020), or to randomly rotate a spectral decomposition of cortical and subcortical surfaces into geometric eigenmodes (Koussis et al. 2024). These spatial null models can be used to benchmark the magnitude of the similarity between two brain maps against a null distribution corresponding to the similarity between one of the maps, and numerous surrogate instances of the other map (Váša and Mišić, 2022).

These permutation approaches implicitly assume that autocorrelation across the underlying brain maps is spatially stationary, or homogeneous (see left panel in Figure 1). In other words, the data is assumed to come from a random distribution with statistical properties that do not change over space. However, evidence suggests that there is considerable spatial heterogeneity in autocorrelation across the cortex for functional MRI (Hayasaka et al., 2004; Leech et al., 2023; Leech and Leech, 2011; Li et al., 2015), and such findings are likely similar for other modalities (Paquola et al., 2023). As such, the above permutation approaches may be inappropriate in many application scenarios.

There are many possible potential spatial processes that could create this heterogeneity in the autocorrelation (some examples are illustrated in **Figure 1**). For example, this could result from non-stationarity in the dependency structure of the random noise process across space or a spatially varying correlation structure across the maps. We can have different random noise processes within the target map (**Figure 1A**), including: (i) white noise with no spatial structure; (ii) homogeneous spatial autocorrelation; (iii) or a spatially varying noise structure (e.g., changing length scale of the noise over the brain). When we want to compare one spatial map (e.g., **Figure 1B**) to another, there are a range of possible statistical scenarios (illustrated in **Figure 1C**). The simplest is the case of white noise (*i.e.,* no spatial autocorrelation) and a homogeneous correlation between the two maps across space (**Figure 1C i**); in this situation simple correlation or ordinary least squares regression is appropriate for comparing the two maps. The second situation (**Figure 1C ii**) is one where there is homogeneous spatial autocorrelation and a spatially homogeneous correlation across maps; here, existing null permutation approaches are appropriate. The third illustration (**Figure 1C iii**) involves heterogeneous spatial noise and a consistent spatial correlation across the two maps. The fourth example (**Figure 1C iv)** has homogeneous spatial autocorrelation of the noise, but the correlation between the maps varies over space (with a positive correlation coefficient in the posterior parts of the cortex and a negative correlation coefficient in the inferior parts). The final illustration (**Figure 1C v**) is where there is no spatial correlation between the two maps; however, the spatial autocorrelation in the second map varies depending on values of the first map. These final three hypothetical examples pose challenges for existing null model-based approaches for comparing spatial maps.

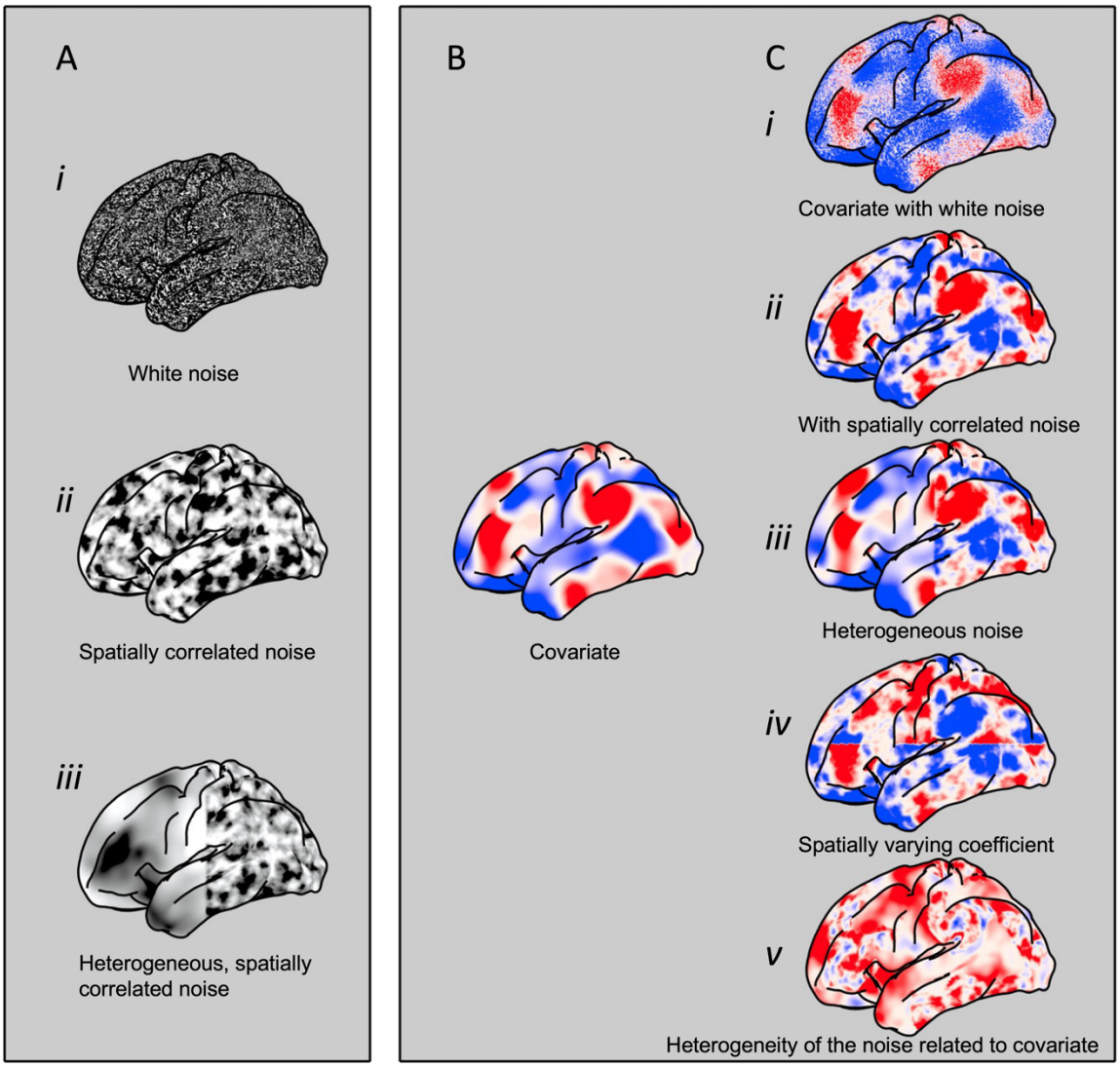


Figure 1. Heterogeneous spatial structure in brain maps. *Left panel: A, Illustration of different types of spatial noise across the cortex:* ***(i)*** *independently and identically distributed white noise, (ii) spatially correlated noise generated with a spatially homogeneous covariance structure; (iii) spatially heterogeneous noise, with a greater spatial noise scale anteriorly than posteriorly. Right panel: some of the potential relationships between two spatial maps: B is a randomly generated map with spatially-correlated structure; C, are example illustrations of other maps which are related to the first map in different ways: (i) white noise and a consistent spatial relationship between the two maps; (ii) homogeneous spatially correlated noise and a homogeneous spatial relationship between the two maps; (iii) spatially heterogeneous spatial noise and a consistent spatial relationship between maps; (iv) homogeneous spatial noise and a spatially varying coefficient between the two maps (superior half has a positive relationship, inferior half has a negative relationship). (v) no spatially varying correlation, but spatially varying autocorrelated noise between the two maps.*

We first illustrate that there is a likely problem with spatial homogeneity assumptions when comparing spatial maps commonly used in the neuroimaging field. We quantify the extent of heterogeneity in spatial autocorrelation exhibited by multiple brain maps generated using different imaging modalities. We then use synthetic data to illustrate why such heterogeneity can lead to problems with statistical inference when comparing spatial maps using existing null model approaches.

The issues of heterogeneity in spatial correlation have long been recognized in other fields, notably ecology, climatology and epidemiology (Daly et al., 1994) and this body of work provides a more comprehensive framework for assessing statistical similarity in neuroimaging. Recent theoretical and computational approaches have allowed for sophisticated modelling of spatial structure, including different forms of heterogeneity (as illustrated in **Figure 1**) even for very large datasets. We use one such approach based on stochastic partial differential equations to explicitly model and compare potential spatial processes underlying brain maps (Lindgren and Rue, 2015). Using this approach, we show how statistical assessments between brain maps can be made while modelling different forms of spatial heterogeneity, using both synthetic and real neuroimaging data.

Results

Heterogeneity in spatial autocorrelation across the cortex

We first consider several canonical maps of cortical organization from different imaging modalities obtained from *neuromaps* (Markello et al., 2022). For this example, we consider the T1w/T2w ratio (a proxy for intracortical myelin) and cortical thickness (Glasser et al., 2013), the principal functional connectivity gradient derived from resting-state fMRI (Margulies et al., 2016), and a map of evolutionary expansion from non-human primates and humans (Hill et al., 2010) (**Figure 2A**). For each map, we calculated the local Moran's I statistic (Anselin, 1995; Moran, 1950) for each vertex using a geodesic distance matrix along the cortical surface. Moran's I is calculated based on the pairwise distances and dissimilarities between all data points, and primarily reflects spatial autocorrelation, producing a statistic with values between -1 and 1 which indexes the magnitude and sign of spatial autocorrelation, while 0 indicates the absence of autocorrelation. Local Moran's I calculates a similar metric but separately for each point compared to all other points (See **Figure 2B** for an illustration).

In **Figure 2A**, we observe positive spatial autocorrelation (and a near absence of negative spatial autocorrelation) across all maps. Further, the distribution of local Moran's I displays considerable heterogeneity in the spatial autocorrelation within each map. For example, in the T1w/T2w map we see widespread increases in spatial dependencies in early somatosensory and visual regions as well as a range of anterior temporal and frontal regions. In **Figure 2C**, we observe the cross-correlations between local Moran's I maps. Spearman's rank correlation values vary in amplitude and range from negative to positive for the different map pairs, reflective of varying alignment in spatial heterogeneity across different pairs of maps.

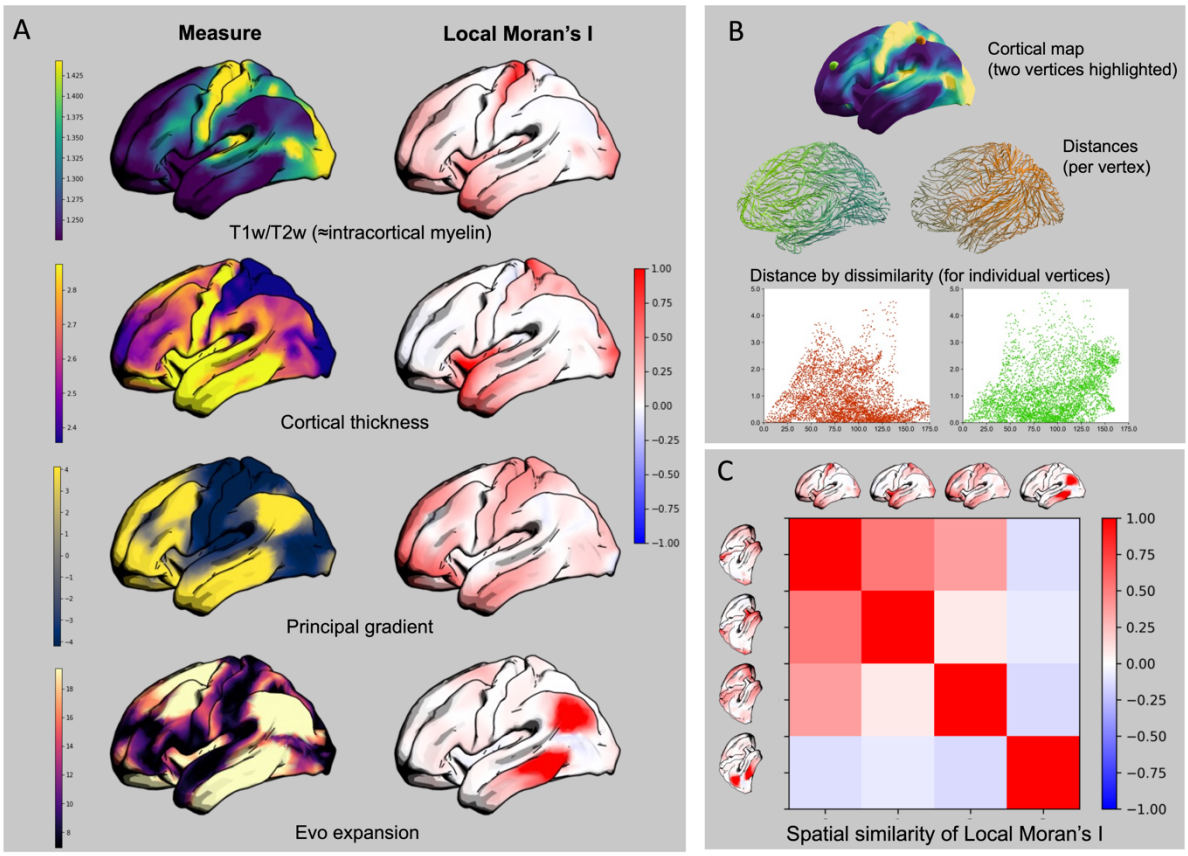


Figure 2. Regionally heterogeneous spatial autocorrelation. *A, left: Four exemplar cortical maps: T1w/T2w ratio, cortical thickness, the first functional connectivity gradient, cortical expansion between macaque monkey and human. Right, the distribution of local Moran's I across the cortex (measuring regional variability in spatial autocorrelation). B. Illustration of calculating local Moran's I for two illustrative vertices from the T1w/T2w cortical map, combining dissimilarity to all other vertices with the vector of distances to other vertices. C, the spatial similarity in the distribution of local Moran's I across maps, quantified using Spearman's ρ.*

A 1-dimensional illustration of the problem

To understand the effects of these types of heterogeneity on spatial permutation approaches such as the spin test (Alexander-Bloch et al., 2018; Váša et al., 2018), we next consider some simulated datasets. Note that the objective of these simulations is to highlight the problem in a transparent way which is easy to follow, rather than to replicate the type of stochasticity and signals corresponding to empirical cortical maps.

We first simulate a pair of 1-dimensional spatial signals with basic spatial autocorrelation and externally imposed heterogeneity. Two 1D vectors of different length Gaussian random noise are generated and resampled to create vectors with different amounts of spatial smoothness. We subsequently concatenate these vectors, to create a spatially heterogeneous signal, with two regimes.

Figure 3 illustrates two scenarios with pairs of heterogeneous signals (in dotted and solid). These signals were both generated from random Gaussian noise independent from each other; therefore, a statistical test quantifying the similarity between each pair of signals should fail to reject the null hypothesis that the signals are independent. In one of the two scenarios **(Figure 3A)**, the heterogeneity is spatially aligned *i.e.,* high or low spatial autocorrelation is present in the same locations in both signals. In the second scenario **(Figure 3B)**, the heterogeneity is inversely aligned. These scenarios are directly comparable to relationships between the autocorrelation of pairs of empirical maps, quantified by correlations between their local Moran's I. The first simulated scenario

corresponds to positive correlations between local Moran's I, as in the comparison of the T1w/T2w ratio and cortical thickness. The second simulated scenario corresponds to negative correlations between local Moran's I, as in the comparison of the T1w/T2w ratio with the evolutionary expansion map (see **Figure 2C**).

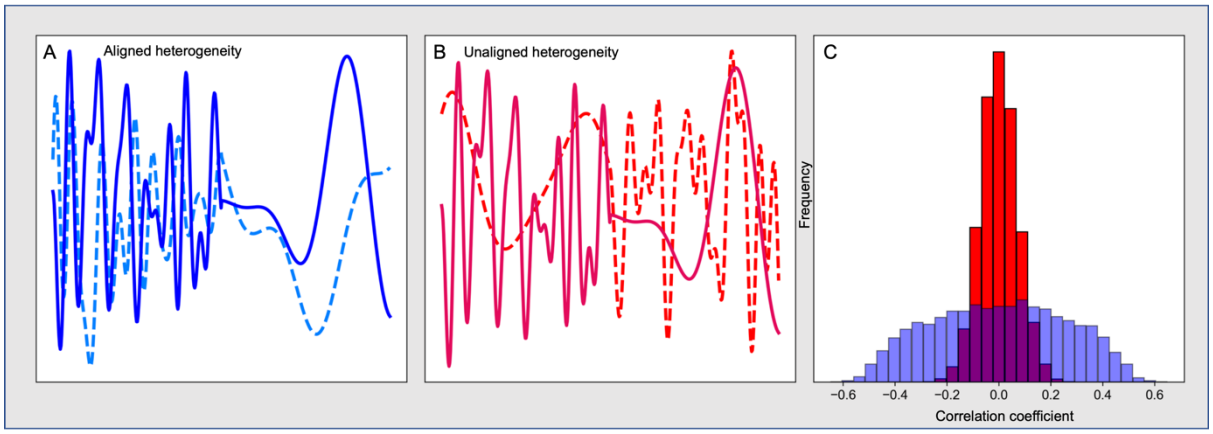


Figure 3**. An illustration of different forms of heterogeneity in 1D spatial autocorrelation.* *Two random signals created by spatially concatenating vectors with high and low levels of smoothing are compared (dotted and solid lines).* ***A*) A case in which the heterogeneity of the two signals is perfectly aligned. B) A case where the heterogeneity of the two signals is inversely aligned. C: the distribution of Spearman's correlation coefficients comparing the two signals that arise in the two scenarios, blue are aligned, red are unaligned (from 5000 repetitions with different random seeds).*

In **Figure 3C** we observe the distribution of Spearman's Rank correlation coefficients between both variables, for 5000 repetitions of the data generation process. Both for the aligned and non-aligned scenarios, we observe symmetric distributions of correlation coefficients (as expected because there is no correlation between the variables. However, the non-aligned scenario shows a relatively narrow range of correlations (high smoothness is paired with low smoothness), while the aligned variant shows a much wider range of correlation values and a much flatter distribution. The difference between these two scenarios results from the spatial translation of one of the variables in relation to the other. Therefore, spatial translation applied to one of the signals (as is done with spin permutations to build a null distribution) may result in null distributions with different statistical properties to the empirical result (*e.g.,* more extreme values). This can potentially lead to inappropriate statistical inference; whether it will inflate or deflate the probability of a false positive depends on the nature of the alignment of the heterogeneity across the true signals compared to the translated signals.

A surface-based illustration of the problem

Above we have observed that the alignment of heterogeneity in the smoothness of a 1D signal can profoundly affect the distribution of correlations. To illustrate how this may affect statistical inference following spin permutation tests (commonly used for surface-based neuroimaging analyses; Alexander-Bloch et al., 2018; Váša et al., 2018), we created two more simulated ground-truth datasets on a spherical surface. In one scenario, the data was generated by a spatially homogeneous process, where we anticipated that the spin permutation test would lead to appropriate statistical inference. In the other scenario, the data was generated by a spatially heterogeneous process that is a 2D spherical analogue of the scenario in **Figure 3A**, which would be challenging for the spin permutation test. For the heterogeneous situation, we used a previously proposed method (Ingebrigtsen et al., 2014) to adjust the variance and length scale of the autocorrelation depending on the spatial location (**Figure 4A**). For each scenario, we created 5000 pairs of spheres using different random seeds. We then used the spin permutation procedure (with 100 iterations) (Alexander-Bloch et al., 2018; Markello and Misic, 2021) to generate a null distribution of correlations and obtain a measure of

statistical significance as the proportion of null correlations which exceeds the true correlation coefficient. We repeated this process 5000 times for both the homogeneous and heterogeneous scenarios, resulting in a distribution of "true" correlation coefficients (between each pair of unpermuted maps) and a distribution of associated p-values derived from the spin permutations, for each scenario.

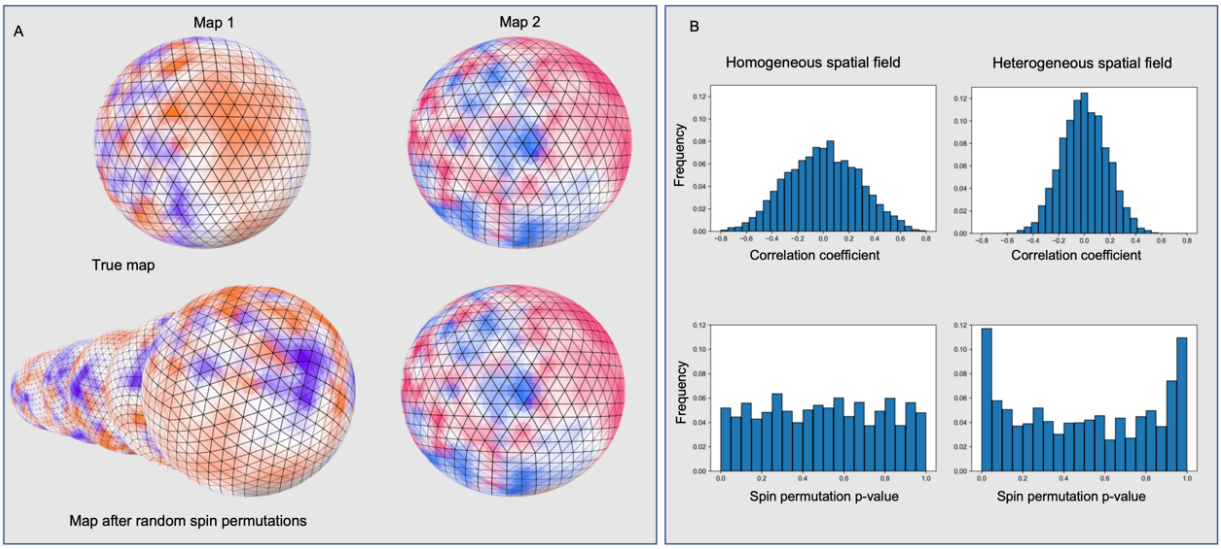


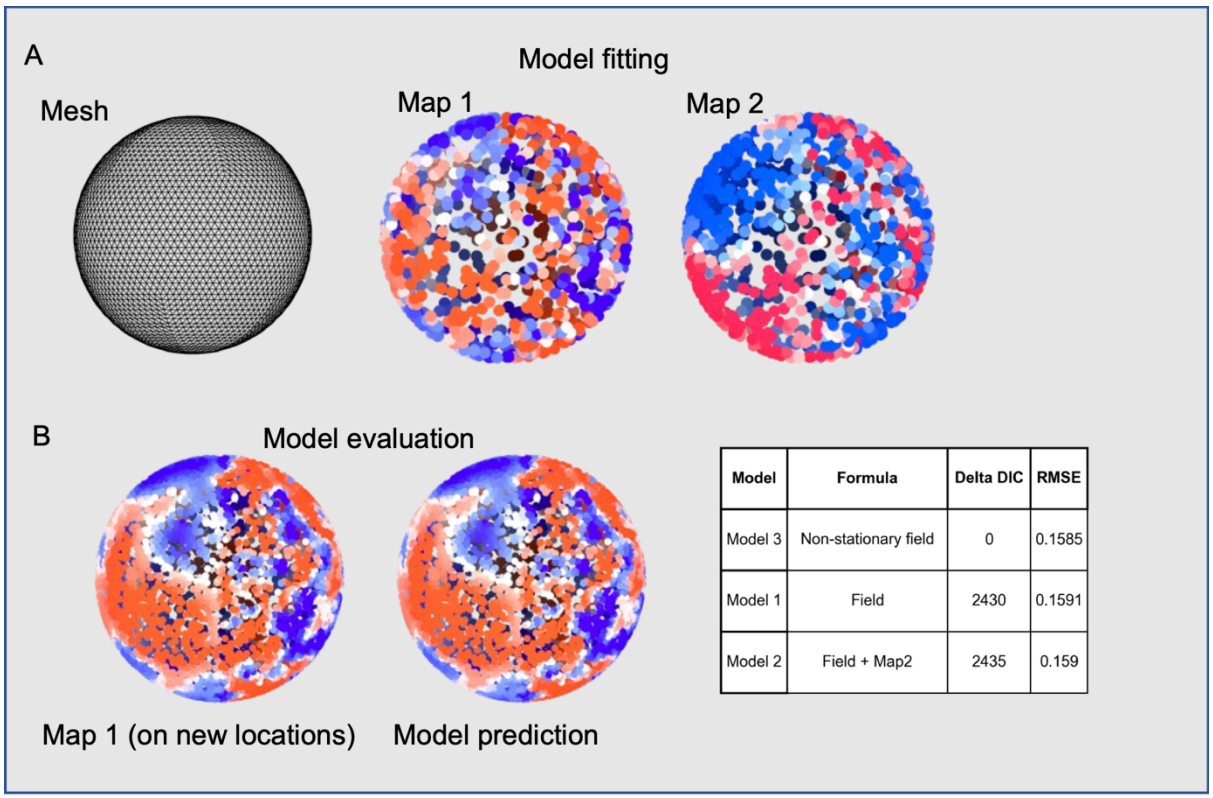
Figure 4. The impact of spatially heterogeneous autocorrelation on the spin permutation test. A*) Top: an illustration of two randomly generated maps with spatially-varying autocorrelation (created from a realisation of a stochastic partial differential equation on an icosphere surface). Bottom: one of these maps was randomly rotated 100 times to build a permutation distribution. In the heterogeneous scenario, the spatial autocorrelation varies based on the spatial location on the sphere (0 degree meridian). Pairs of these maps were generated with different random values and only aligned in terms of the heterogeneity of the spatial autocorrelation. Two scenarios were considered, homogeneous and heterogeneous spatial autocorrelation across the surface.* ***B****) Top: A distribution of 5000 pairs of maps were generated and then spatially correlated (i.e., correlating the values across Maps 1 and 2), to generate "true" Spearman Rank correlation values. Bottom : Subsequently one map was randomly rotated 100 times to generate null distributions of correlations and associated p-values (defined as the proportion of null correlations exceeding the true correlation coefficient).*

Figure 4B illustrates the distribution of correlation coefficients observed for spatially correlating the two maps in both the homogeneous and heterogeneous scenarios and the associated distribution of spin permutation p-values. We observe that the homogeneous scenario results in a wider distribution of correlation coefficients; this is because it was generated only with a high level of spatial smoothing, while the heterogeneous scenario was generated with mix of high and low spatial smoothing. For the homogeneous scenario, although the amount of spatial autocorrelation is higher resulting in more extreme correlation coefficients, the spin permutation test works appropriately at controlling the error rate, with an approximately flat distribution of permutation p-values. In contrast, the heterogeneous scenario has a narrower range of correlation coefficients, but the spin permutation test fails to adequately control the type 1 error, with a disproportionate number of extreme correlation values compared to the null. In this scenario, approximately 1/4 of correlation coefficients are within bins at the extremes of the histogram, compared to the expected 1/10 in the homogeneous situation. As such, any analysis involving comparisons between brain maps with heterogeneous spatial autocorrelation is likely to result in substantial inflation of p-values; the precise amount of inflation will depend on the degree of heterogeneity, the amount of spatial autocorrelation and the spatial resolution of the data.

Explicitly modelling spatial dependencies in the data

An alternative framework for statistically assessing the relationship between spatial maps in neuroimaging involves attempting to explicitly quantify the spatial autocorrelation while modelling

the relationship between these maps. Several approaches have been developed in other scientific fields to this end *e.g.,* R-INLA (Integrated Nested Laplace Approximation in R) (Bachl et al., 2019; Lindgren et al., 2011; Lindgren and Rue, 2015) and sdmTMB (Anderson et al. 2024). R-INLA uses approximate Bayesian methods to estimate a model of a continuous spatial process with stochastic partial differential equations realised on a discrete mesh. It allows creation of a generative model of spatial data, including various types of non-stationary spatial phenomena, as well as spatial covariates. Model comparison, preferably using information criteria and/or out-of-sample cross-validation, enables differentiating between different hypotheses about the role of spatial covariates and assessing the spatial autocorrelation of the signal.



Model	Formula	Delta DIC	RMSE
Model 3	Non-stationary field	0	0.1585
Model 1	Field	2430	0.1591
Model 2	Field + Map2	2435	0.159

Figure 5**. Explicitly modelling autocorrelation in a simulated dataset on a sphere, generated by a non-stationary process. A)* *Model fitting: a spherical mesh is approximated, and a pair of maps is generated at random locations on the sphere; then, the mesh is used to fit three different models to predict Map 1: (i) a stationary spatial field; (ii) a stationary spatial field plus Map 2; (iii) a non-stationary spatial field.* ***B) *The best model is selected by evaluating model performance on either the deviance information criteria (DIC) or the root mean squared error between predicted and actual values of Map 1 at out-of-sample spatial locations.*

We first use R-INLA with a ground-truth model generated by a non-stationary process on a mesh (see **Figure 5**) generated with the same procedure as used in **Figure 4** above. Two maps (Map 1 and Map 2) were generated at 1000 random spatial locations on the sphere. Maps have different random seeds but the same non-stationary autocorrelation structure, with a binary variable based on the map's *x*-coordinate included in the model's dependency structure (see Methods and Code for full details). Three different models were then fit to the training data, predicting Map 1 with: *(i)* a stationary Gaussian random field as the intercept; *(ii)* a stationary field plus Map 2 as a covariate; *(iii)* a non-stationary Gaussian field as intercept (similar to the one used to generate the data). Model comparison was performed using the Deviance Information Criterion (Spiegelhalter et al., 2002) as well as root-mean-squared-error (RMSE) comparing predicted and actual values of Map 1 at 5000 out-of-sample coordinates on the sphere. For both DIC and RMSE, R-INLA correctly identifies the non-stationary model without a covariate for Map 2, as the best model (which is consistent with how the

data was generated). Note that the models with stationary fields (with and without an intercept) are much more similar (in terms of both RMSE and DIC).

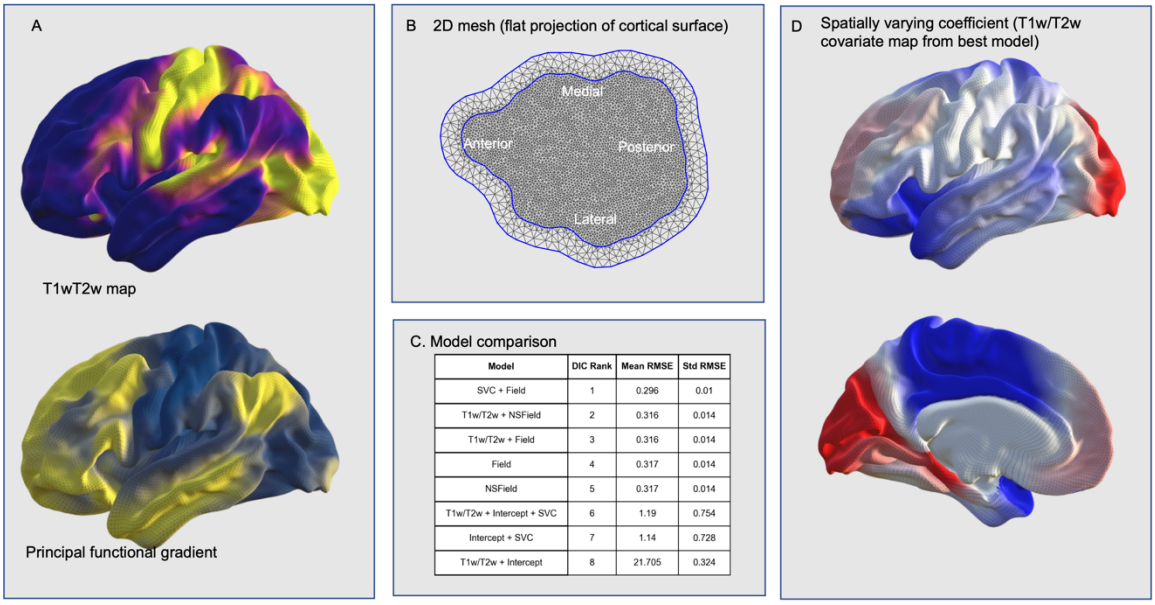
The above simple simulated example illustrates a general alternative to existing approaches typically used in neuroimaging for studying the relationship between different maps. We next apply this alternative approach to compare two empirical cortical maps: the principal functional gradient (Margulies et al. 2016) and a T1w/T2w based proxy for intracortical myelin (Glasser et al. 2011) (**Figure 6A**). We also illustrate how explicit modelling of the spatial process using a range of different models can account for different types of spatial relationships between the maps (see **Figure 1**). In particular, we can evaluate different non-stationary contributions to spatial autocorrelation, including maps with spatially-varying coefficients, where instead of a single coefficient being used to predict the whole map (e.g., in spatial ordinary least squares regression), the coefficient can vary across the cortex.

We first split the data into 10 folds. Within each fold a random proportion of the data (10% of all vertices excluding the medial wall: approximately 3000 vertices) is used to make a mesh using the *fmesher* tool (Bachl et al., 2019). For simplicity, we use a flat projection of the cortex, although it is also possible to use other projections to make the mesh. Subsequently, a range of different models were fit to a random 30% of the remaining data. This is a non-exhaustive but comprehensive range of models accounting for multiple spatial dependency scenarios. In our simulation, the principal gradient map (Margulies et al. 2016) was predicted by each of the following models:

- *(i)* A basic linear relationship with T1w/T2w and constant intercept with no Gaussian random fields;
- *(ii)* A Gaussian random field and no covariate;
- *(iii)* A Gaussian random field plus the T1w/T2w map;
- *(iv)* A non-stationary field with the T1w/T2w map as the explanatory covariate in the dependency structure;
- *(v)* The same non-stationary field as *(iv)* with an additional constant T1w/T2w covariate;
- *(vi)* A spatially varying coefficient (T1w/T2w) plus a Gaussian random field;
- *(vii)* A spatially varying coefficient (T1w/T2w) with a constant intercept;
- *(viii)* A spatially varying coefficient (T1w/T2w) with a constant intercept and a constant T1w/T2w covariate.

Model selection was performed both with Deviance information criteria on the fitted model and out-of-sample prediction on 10% of the data that was held out from mesh creation and model training for each fold. (Note that the whole dataset was not used in the interest of computational efficiency.)

Based on both the Deviance Information Criteria and out-of-sample RMSE, the spatially-varying-coefficient and Gaussian field model performed best for every fold (See **Figure 6C**). This suggests that there is a relationship between the principal functional gradient derived from rs-fMRI and the T1w/T2w map, but that it is spatially varying. The spatially-varying coefficients can be projected onto the cortical surface (**Figure 6D**). A range of qualitatively different relationships can be seen across cortex, ranging from predominantly negative relationships centered around the insula and paracentral lobule, through a lack of relationship in in a subset of regions, to positive relationships centered around the occipital pole (see **Figure 6D;** this can also be observed in the raw data in **Figure 6A**).



C. Model comparison

Model	DIC Rank	Mean RMSE	Std RMSE
SVC + Field	1	0.296	0.01
T1w/T2w + NSField	2	0.316	0.014
T1w/T2w + Field	3	0.316	0.014
Field	4	0.317	0.014
NSField	5	0.317	0.014
T1w/T2w + Intercept + SVC	6	1.19	0.754
Intercept + SVC	7	1.14	0.728
T1w/T2w + Intercept	8	21.705	0.324

Figure 6. Comparing two empirical cortical maps by explicitly modelling spatial processes. A) *The maps being compared: the T1w/T2w ratio (an approximate measure of intracortical myelin) and the principal functional gradient.* ***B)*** *A mesh made from a subset of the data projected onto a plane, including an outer layer to control for potential boundary effects. This mesh was used to fit models on a different subset of the data from both T1w/T2w and the functional gradient.* ***C)*** *Model comparison reveals that a spatially varying coefficient (T1w/T2w) model with a stationary field has the lowest Deviance Information Criterion and mean out of sample error (across 10-fold cross-validation).* ***D)*** *The spatially varying coefficient model reprojected back onto the cortical surface (cold/warm colors indicate a negative/positive relationship between T1w/T2w and the principal functional gradient).*

All code to repeat the analyses is openly available at:
https://github.com/ActiveNeuroImaging/SpatialNonStationarity

Discussion

Funnelled by advances in multimodal imaging and transcriptomics, as well as increased efforts to make neural data openly available in a common reference frame (Amunts et al., 2013; DeKraker et al., 2024; Larivière et al., 2023, 2021; Markello et al., 2022), recent years have witnessed a surge in studies assessing the spatial correspondence of different maps of brain organization (e.g., DeKraker et al., 2024; Larivière et al., 2023, 2021; Markello et al., 2022). Here, we have illustrated that in different cortical maps frequently used in the literature, there is considerable heterogeneity in the spatial similarity of the signal *within* individual maps. We also show why the presence of this heterogeneity may lead to problems for common techniques for comparing spatial maps. In other scientific fields, a range of techniques have been developed for modelling spatial processes to characterize spatial maps and their relationships. Here, we illustrate one such approach based on modelling spatial processes as stochastic partial differential equations (Lindgren et al., 2011); this provides a more comprehensive account of the relationship between spatial maps, which can also explicitly model and compare different forms of heterogeneity.

Incorporating explicit modelling of spatial processes also allows opportunities for studying more complex spatial phenomena. We have already seen that we can observe spatially varying relationships that disentangle the sign and amplitude of regionally-specific relationships, rather than a single spatial correlation value. Using this approach, it is also possible to include multiple replicates within a single spatial model (e.g., investigating the relationship between maps of many individuals including group

differences), which has the advantage of being able to disambiguate more complicated models. For example, non-stationarity in a spatial process can reflect variation in the mean or the covariance structure, and this difference can only be differentiated with multiple replicates (e.g., with different maps of different individuals). In a related vein, spatio-temporal models can also be realised, modelling a temporal autoregressive process in additional to the spatial one, allowing for a more comprehensive perspective on neural dynamics (Mejia et al., 2020), including task-evoked signal change and functional connectivity. Finally, we can also extend the model to considering multiple covariates simultaneously (e.g., Tavor et al., 2016), for example to assess how a range of neurotransmitter maps or microstructural features reflect functional properties such as the principal functional gradients, with different random Gaussian fields with spatial structures for the different covariates. Such powerful statistical approaches can help model and disentangle the likely complex spatial relationships between these maps, and between the underlying genetic, developmental, anatomical and physiological processes.

While approaches to explicitly modelling the underlying spatial processes can better characterize the statistical relationships between spatial maps, the added flexibility carries a cost. Whereas approaches such as spin permutation tests (Alexander-Bloch et al., 2018; Váša et al., 2018) are easy to apply, using R-INLA or other approaches to model spatial processes within brain maps requires domain knowledge to inform the selection of model priors and choice of covariates. Also, results may be more ambiguous and harder to interpret than the resulting *r* and *p*-values from null-modeling approaches. It is also worth emphasizing that the existing null model approaches are still likely to be appropriate in a wide range of scenarios, for example, if the compared maps exhibit relatively limited spatial autocorrelation, and this is relatively spatially homogeneous. This may be the case when the analysis is restricted to local regions. Finally, approaches such as R-INLA have been developed primarily for 2D or spherical maps; in future, it would be beneficial to build 3D meshes to more accurately model the cortical surface, and to consider approaches which can be applied to 3D volumes.

In summary, variability within cortical maps poses challenges for statistical comparisons between them using the existing null model framework. Alternative methods such as stochastic partial differential equations offer a more comprehensive approach to modeling and comparing spatial maps, accommodating different forms of spatial heterogeneity.

Methods

Neuroimaging data

All neuroimaging data were obtained from *neuromaps* (Markello et al., 2022) in the FSLR surface projection. We used four different maps from three different studies. The T1w/T2w maps and the cortical thickness maps are the group averages from the Human Connectome Project's Young Adult 1000 participant release (Glasser et al. 2013). The principal gradient of functional connectivity was calculated on the resting state fMRI data from the HCP Young Adult dataset using diffusion embedding (Margulies et al., 2016). The evolutionary expansion map was created by Hill et al. (2010). For full details of the pre-processing pipelines and subsequent creation of each map, see the corresponding publications.

Local Moran's I

Code to calculate local Moran's I was adapted from Markello and Mišić (2021). For computational reasons, the four cortical maps were downsampled to 10,000 vertices; the distance matrix was calculated based on the down-sampled version of the FSLR atlas (with 5000 vertices in each hemisphere). Local Moran's I was calculated for each vertex in each of the four maps.

One-dimensional simulation

Code to reproduce the 1D scenario is available on Github (link in Results). Two independently and identically distributed arrays were created with: (i) 5 samples; (ii) 100 samples. Both were then interpolated with the *scipy* resample function to 1000 samples. These two 1000-sample arrays were normalized (demeaned and then divided by the standard deviation) and then concatenated to form a 2000-sample 1D array with both a high and low smoothness regime. Pairs of 1D arrays were created with different random seeds. Differences in the order of the concatenation (high-low or low-high) across each of the pairs determined whether the heterogeneity of smoothness was aligned or not. Finally, the pairs of arrays were correlated. This was repeated 10,000 times to build distributions of correlation coefficients for aligned and inversely aligned smoothness.

Surface-based simulation on a sphere

Spherical spatially heterogeneous and homogeneous spatial autocorrelation was generated using R-INLA (Lindgren and Rue, 2015) and the INLABRU (Bachl et al., 2019) wrapper package. The full code is available on Github. A spherical mesh was generated from an icosphere with 30 subdivisions (using the *fmesher* library). A binary variable was created for each point on the mesh, based on its *x*-coordinate; this was then demeaned. A model of a Gaussian random field with a Matern covariance kernel was instantiated on the mesh; either a standard homogeneous Guassian random field or a spatially varying Gaussian random field (based on the binary variable) were used. Five thousand pairs of sample maps were created with different random seeds. Full details of the model parameterisation are in the associated R code. To assess spin test performance, these pairs of random maps with either homogeneous or aligned spatially varying autocorrelation structure were compared using code from Markello and Mišić (2021).

In addition, a separate simulation was performed in Python (see Github) by explicitly smoothing points on the sphere based on two smoothing kernels, cutting them in half, normalizing them, and then adding them together to create aligned heterogeneity (similar to the 1D simulated data above). This had qualitatively similar results to the simulated data generated from the R-INLA model.

R-INLA modelling of spatial processes

(i) Simulated data

A fine spherical mesh (icosphere with 30 subdivisions) was created with the same model used to generate heterogeneous spatial maps for evaluating the spin test. Two randomly-seeded maps were then generated (with an identical spatial process) at 1000 random locations on the sphere (not on the mesh). A new lower-resolution mesh was created (again an icosphere, but with 20 subdivisions). Three models were fit to the data, predicting Map 1, with either: *(i)* A stationary random Gaussian field; *(ii)* A stationary random field and Map 2 as a covariate; *(iii)* Or a spatially varying random Gaussian field (with the hemisphere controlling the spatial variability in the field). Model performance was assessed with two metrics: (a) the Deviance Information Criterion (Spiegelhalter et al., 2002) on the trained data; and (b) out of sample root-mean-squared error on 5000 novel randomly selected locations on the sphere.

(ii) Cortical maps

The group average principal functional gradient (Margulies et al., 2016) and the T1w/T2w map (Glasser et al. 2013) were used. The 2D projection was the default flat map from the HCP Young Adult group average data release. The model was fit with 10-fold cross validation: different subsets of the training data (i.e., different spatial locations) were used for mesh creation, and a separate set was used for model fitting for each fold. Model evaluation again used both the Deviance Information Criteria and root mean squared error on held out 10% of the data used for evaluation at each fold. Within each fold the 2D mesh was generated from 10% of the data using the *fmesher* library. The same 8 models (listed in the results) were fit to each fold assessing combinations of linear effects, stationary and non-

stationary Gaussian random fields, and spatially varying coefficients. For non-stationary Gaussian random fields, the demeaned T1w/T2w ratio was used as a covariate in the formula for the field.

References

Alexander-Bloch, A.F., Shou, H., Liu, S., Satterthwaite, T.D., Glahn, D.C., Shinohara, R.T., Vandekar, S.N., Raznahan, A., 2018. On testing for spatial correspondence between maps of human brain structure and function. NeuroImage 178, 540–551. https://doi.org/10.1016/j.neuroimage.2018.05.070

Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.-É., Bludau, S., Bazin, P.-L., Lewis, L.B., Oros-Peusquens, A.-M., Shah, N.J., Lippert, T., Zilles, K., Evans, A.C., 2013. BigBrain: An Ultrahigh-Resolution 3D Human Brain Model. Science 340, 1472–1475. https://doi.org/10.1126/science.1235381

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geographical Analysis 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods in Ecology and Evolution 10, 760–766. https://doi.org/10.1111/2041-210X.13168

Burt, J.B., Demirtaş, M., Eckner, W.J., Navejar, N.M., Ji, J.L., Martin, W.J., Bernacchia, A., Anticevic, A., Murray, J.D., 2018. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat Neurosci 21, 1251–1259. https://doi.org/10.1038/s41593-018-0195-0

Burt, J.B., Helmer, M., Shinn, M., Anticevic, A., Murray, J.D., 2020. Generative modeling of brain maps with spatial autocorrelation. NeuroImage 220, 117038. https://doi.org/10.1016/j.neuroimage.2020.117038

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. Journal of Applied Meteorology and Climatology 33, 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

DeKraker, J., Cabalo, D.G., Royer, J., Khan, A.R., Karat, B., Benkarim, O., Rodriguez-Cruces, R., Frauscher, B., Pana, R., Hansen, J.Y., Misic, B., Valk, S.L., Lau, J.C., Kirschner, M., Bernasconi, A., Bernasconi, N., Muenzing, S., Axer, M., Amunts, K., Evans, A.C., Bernhardt, B.C., 2024. HippoMaps: Multiscale cartography of human hippocampal organization. https://doi.org/10.1101/2024.02.23.581734

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage, Mapping the Connectome 80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Hansen, J.Y., Shafiei, G., Markello, R.D., Smart, K., Cox, S.M.L., Nørgaard, M., Beliveau, V., Wu, Y., Gallezot, J.-D., Aumont, É., Servaes, S., Scala, S.G., DuBois, J.M., Wainstein, G., Bezgin, G., Funck, T., Schmitz, T.W., Spreng, R.N., Galovic, M., Koepp, M.J., Duncan, J.S., Coles, J.P., Fryer, T.D., Aigbirhio, F.I., McGinnity, C.J., Hammers, A., Soucy, J.-P., Baillet, S., Guimond, S., Hietala, J., Bedard, M.-A., Leyton, M., Kobayashi, E., Rosa-Neto, P., Ganz, M., Knudsen, G.M., Palomero-Gallagher, N., Shine, J.M., Carson, R.E., Tuominen, L., Dagher, A., Misic, B., 2022. Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. Nat Neurosci 25, 1569–1581. https://doi.org/10.1038/s41593-022-01186-3

Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 22, 676–687. https://doi.org/10.1016/j.neuroimage.2004.01.041

Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., Van Essen, D., 2010. Similar patterns of cortical expansion during human development and evolution. Proceedings of the National Academy of Sciences 107, 13135–13140. https://doi.org/10.1073/pnas.1001229107

Ingebrigtsen, R., Lindgren, F., Steinsland, I., 2014. Spatial models with explanatory variables in the dependence structure. Spatial Statistics, Spatial Statistics Miami 8, 20–38. https://doi.org/10.1016/j.spasta.2013.06.002

Larivière, S., Bayrak, Ş., Vos de Wael, R., Benkarim, O., Herholz, P., Rodriguez-Cruces, R., Paquola, C., Hong, S.-J., Misic, B., Evans, A.C., Valk, S.L., Bernhardt, B.C., 2023. BrainStat: A toolbox for brain-wide statistics and multimodal feature associations. NeuroImage 266, 119807. https://doi.org/10.1016/j.neuroimage.2022.119807

Larivière, S., Paquola, C., Park, B., Royer, J., Wang, Y., Benkarim, O., Vos de Wael, R., Valk, S.L., Thomopoulos, S.I., Kirschner, M., Lewis, L.B., Evans, A.C., Sisodiya, S.M., McDonald, C.R., Thompson, P.M., Bernhardt, B.C., 2021. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. Nat Methods 18, 698–700. https://doi.org/10.1038/s41592-021-01186-4

Leech, R., Leech, D., 2011. Testing for spatial heterogeneity in functional MRI using the multivariate general linear model. IEEE Transactions on Medical Imaging 30, 1293–1302.

Leech, R., Vos De Wael, R., Váša, F., Xu, T., Austin Benn, R., Scholz, R., Braga, R.M., Milham, M.P., Royer, J., Bernhardt, B.C., Jones, E.J.H., Jefferies, E., Margulies, D.S., Smallwood, J., 2023. Variation in spatial dependencies across the cortical mantle discriminates the functional behaviour of primary and association cortex. Nat Commun 14, 5656. https://doi.org/10.1038/s41467-023-41334-2

Li, H., Nickerson, L.D., Zhao, X., Nichols, T.E., Gao, J.-H., 2015. A voxelation-corrected non-stationary 3D cluster-size test based on random field theory. NeuroImage 118, 676–682. https://doi.org/10.1016/j.neuroimage.2015.05.094

Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with R-INLA. Journal of Statistical Software 63.

Lindgren, F., Rue, H., Lindström, J., 2011. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. Journal of the Royal Statistical Society Series B: Statistical Methodology 73, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences 113, 12574–12579. https://doi.org/10.1073/pnas.1608282113

Markello, R.D., Hansen, J.Y., Liu, Z.-Q., Bazinet, V., Shafiei, G., Suárez, L.E., Blostein, N., Seidlitz, J., Baillet, S., Satterthwaite, T.D., Chakravarty, M.M., Raznahan, A., Misic, B., 2022. neuromaps: structural and functional interpretation of brain maps. Nat Methods 19, 1472–1479. https://doi.org/10.1038/s41592-022-01625-w

Markello, R.D., Misic, B., 2021. Comparing spatial null models for brain maps. NeuroImage 236, 118052. https://doi.org/10.1016/j.neuroimage.2021.118052

Markello, R.D., Mišić, B., 2021. netneurolab/markello_spatialnulls: 0.2. https://doi.org/10.5281/zenodo.4560668

Mejia, A.F., Yue, Y.R., Bolin, D., Lindgren, F., Lindquist, M.A., 2020. A Bayesian General Linear Modeling Approach to Cortical Surface fMRI Data Analysis. J Am Stat Assoc 115, 501–520. https://doi.org/10.1080/01621459.2019.1611582

Milham, M.P., Craddock, R.C., Son, J.J., Fleischmann, M., Clucas, J., Xu, H., Koo, B., Krishnakumar, A., Biswal, B.B., Castellanos, F.X., Colcombe, S., Di Martino, A., Zuo, X.-N., Klein, A., 2018. Assessment of the impact of shared brain imaging data on the scientific literature. Nat Commun 9, 2818. https://doi.org/10.1038/s41467-018-04976-1

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Paquola, C., Garber, M., Frässle, S., Royer, J., Tavakol, S., Rodriguez-Cruces, R., Cabalo, D.G., Valk, S., Eickhoff, S., Margulies, D.S., Evans, A., Amunts, K., Jefferies, E., Smallwood, J., Bernhardt, B.C., 2023. The unique cytoarchitecture and wiring of the human default mode network. https://doi.org/10.1101/2021.11.22.469533

Paquola, C., Wael, R.V.D., Wagstyl, K., Bethlehem, R.A.I., Hong, S.-J., Seidlitz, J., Bullmore, E.T., Evans, A.C., Misic, B., Margulies, D.S., Smallwood, J., Bernhardt, B.C., 2019. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. PLOS Biology 17, e3000284. https://doi.org/10.1371/journal.pbio.3000284

Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat Neurosci 17, 1510–1517. https://doi.org/10.1038/nn.3818

Shinn, M., Hu, A., Turner, L., Noble, S., Preller, K.H., Ji, J.L., Moujaes, F., Achard, S., Scheinost, D., Constable, R.T., Krystal, J.H., Vollenweider, F.X., Lee, D., Anticevic, A., Bullmore, E.T., Murray, J.D., 2022. Spatial and temporal autocorrelation weave complexity in brain networks. https://doi.org/10.1101/2021.06.01.446561

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society Series B: Statistical Methodology 64, 583–639. https://doi.org/10.1111/1467-9868.00353

Tavor, I., Jones, O.P., Mars, R.B., Smith, S.M., Behrens, T.E., Jbabdi, S., 2016. Task-free MRI predicts individual differences in brain activity during task performance. Science 352, 216–220. https://doi.org/10.1126/science.aad8127

Váša, F., Mišić, B., 2022. Null models in network neuroscience. Nat Rev Neurosci 23, 493–504. https://doi.org/10.1038/s41583-022-00601-9

Váša, F., Romero-Garcia, R., Kitzbichler, M.G., Seidlitz, J., Whitaker, K.J., Vaghi, M.M., Kundu, P., Patel, A.X., Fonagy, P., Dolan, R.J., Jones, P.B., Goodyer, I.M., the NSPN Consortium, Vértes, P.E., Bullmore, E.T., 2020. Conservative and disruptive modes of adolescent change in human brain functional connectivity. Proceedings of the National Academy of Sciences 117, 3248–3253. https://doi.org/10.1073/pnas.1906144117

Váša, F., Seidlitz, J., Romero-Garcia, R., Whitaker, K.J., Rosenthal, G., Vértes, P.E., Shinn, M., Alexander-Bloch, A., Fonagy, P., Dolan, R.J., Jones, P.B., Goodyer, I.M., the NSPN consortium, Sporns, O., Bullmore, E.T., 2018. Adolescent Tuning of Association Cortex in Human Structural Brain Networks. Cerebral Cortex 28, 281–294. https://doi.org/10.1093/cercor/bhx249

Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.-J., Langs, G., Valk, S., 2020. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Communications biology 3, 1–10.

Wagstyl, K., Adler, S., Seidlitz, J., Vandekar, S., Mallard, T.T., Dear, R., DeCasien, A.R., Satterthwaite, T.D., Liu, S., Vértes, P.E., Shinohara, R.T., Alexander-Bloch, A., Geschwind, D.H., Raznahan, A., 2024. Transcriptional Cartography Integrates Multiscale Biology of the Human Cortex. eLife 12. https://doi.org/10.7554/eLife.86933.2

Worsley, K. j., Andermann, M., Koulis, T., MacDonald, D., Evans, A. c., 1999. Detecting changes in nonisotropic images. Human Brain Mapping 8, 98–101. https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<98::AID-HBM5>3.0.CO;2-F