Modeling Covariance in Multilevel ROI Analyses of Neuroimaging Data

Sean Spinney

April 23, 2025

Abstract

Contents

1 Introduction

In neuroimaging research, analysts often extract measures from multiple brain regions of interest (ROIs) — for example, functional MRI activation in predefined regions, diffusion tensor imaging metrics in fiber tracts, or cortical thickness in anatomical areas. A common analytical pipeline is to perform separate statistical tests for each ROI (or each voxel) and then apply a correction for multiple comparisons. This mass-univariate approach treats each ROI as an independent unit of analysis Chen et al., 2019. While convenient, it ignores the rich covariance structure among brain regions. In reality, ROI measures are inter-correlated due to anatomical, developmental, and functional relationships in the brain. Ignoring these dependencies can lead to suboptimal inference, reduced power, and an excessive focus on controlling false positives at the expense of sensitivity. This document provides a comprehensive overview of why explicitly modeling the covariance structure across ROIs is statistically superior and how to do so using multilevel (hierarchical) models. We focus on general predictive modeling scenarios (e.g. relating brain measures to behavioral or clinical outcomes) rather than disease-specific findings. We will review recent methodological research that leverages ROI structural information within unified models, and illustrate practical implementation using R packages like lme4, brms, and rstanarm. Key sections include: background on current practices and their limitations, advantages of modeling ROI covariance, alternative modeling approaches (multivariate and multilevel models), highlights of recent research, example code, and conclusions.

2 Background

2.1 ROI Analyses in Neuroimaging

ROI-based analysis is a popular alternative to voxel-wise analysis, offering dimensionality reduction and interpretability by averaging signals within predefined regions. Typically, researchers test each ROI for associations with an outcome (e.g. a behavioral score or group difference) using separate statistical models, then apply a multiple-comparison correction (such as Bonferroni or False Discovery Rate). This massively univariate strategy indeed controls false positive rate across many tests, but treats ROIs as if they were statistically independent. In practice, ROIs often show substantial covariance — for example, structural MRI volumes of anatomically proximal regions tend to correlate across individuals, and functionally connected regions exhibit coordinated activation. Standard approaches that ignore these within-subject dependencies are statistically inefficient and can miss subtle distributed effects.

2.2 Multiplicity and Independence Assumptions

The default ROI analysis approach (separate tests + multiple comparison correction) has several drawbacks. First, conducting many independent tests wastes information about shared variance. Second, corrections like Bonferroni assume independence or use only the marginal distribution of test statistics, which can be overly conservative when predictors are correlated. Third, some designs (e.g. repeated-measures or multimodal imaging within the same subjects) cannot be fully analyzed by simple ROI-by-ROI tests without making unrealistic assumptions about the covariance across measures McFarquhar et al., 2016. In fact, even software pipelines that allow repeated-measures ROI analysis often impose simplistic covariance structures (such as assuming equal correlations or compound symmetry among measurements). These simplifications may be violated in practice, potentially affecting validity.

2.3 Example Scenario

Consider a study measuring gray matter volumes of 20 cortical regions in each subject, with a goal to predict a cognitive performance score. A naive approach might run 20 separate regressions (cognitive score vs. volume in ROI i for i=1...20) and then apply an FDR correction to identify significant regions. This approach ignores that volumes of, say, adjacent frontal regions are correlated across subjects. By not modeling this covariance, one may lose power to detect a distributed effect where combined information from several correlated regions predicts the outcome, even if each region's marginal effect is modest. Additionally, the ROI-by-ROI approach cannot easily capture patterns (like a global atrophy factor) that influence all regions simultaneously.

3 Methodological Limitations of Independent ROI Analysis

Treating each ROI as independent in statistical analysis is a limitation because it fails to account for the brain's structured covariance, leading to inefficient and possibly biased inference. Key issues include:

3.1 Multiple Comparisons and Reduced Power

Splitting the analysis into many separate tests incurs a severe multiple comparisons penalty. The more ROIs tested, the higher the threshold for significance after correction, which reduces statistical power. Yet many of those ROIs carry redundant information due to intercorrelation. Modeling them jointly can borrow strength from shared variance, effectively increasing power. In fact, accounting for the actual correlation structure via random effects or multivariate models can increase power and reduce inferential bias Parekh et al., 2024. When ROI effects are analyzed in isolation, we may be underpowered to detect effects that are moderate in each ROI but consistent across many ROIs.

3.2 Ignoring Shared Variance and Structure

By not modeling covariance, the standard approach ignores shared information among ROIs. This can lead to unstable or misleading results. For example, if two regions are highly correlated, separate models might each show a non-significant trend, whereas a joint model could reveal that a linear combination (e.g. average or difference) of the two carries significant predictive signal. Conventional corrections do not capitalize on such shared variance. A recent methods study noted that conventional group analyses often favor multiple independent models over a unified repeated-measures model, and as a result certain comparisons of interest become impossible or complex. In other words, important multivariate patterns can be missed when using univariate models.

3.3 Unrealistic Statistical Assumptions

Even when researchers attempt to account for within-subject ROI dependence (e.g., using a repeated-measures ANOVA across ROIs or time points), they often must assume a specific covariance structure (such as compound symmetry or sphericity). These assumptions may be violated if, for instance, some ROI pairs are more correlated than others. McFarquhar et al., 2016 highlighted that popular neuroimaging software impose potentially unrealistic assumptions about the covariance structure across the brain. Violations of these assumptions can lead to incorrect degrees of freedom or inefficient estimates. A more flexible approach would let the data inform the covariance structure instead of assuming independence or simple forms.

3.4 Multiplicity as Modeling Inefficiency

From a statistical perspective, the multiple comparisons problem in ROI analysis is not just about quantity of tests, but about modeling efficiency. If we had a single model that explains all ROIs together, we wouldn't be "multiplying" tests in the first place Chen et al., 2019. The common practice of thresholding each ROI's p-value and then applying a sharp cutoff (e.g. $p < 0.05$ after correction) can be seen as symptomatic of an inefficient modeling strategy, where we reduce multidimensional data to a set of thresholded univariate results. This not only risks false negatives but also complicates interpretation (e.g. borderline ROIs might be ignored despite collectively indicating a pattern).

In summary, treating ROI measures as independent features and correcting post hoc is a suboptimal strategy that can obscure true brain-behavior relationships. It fails to make use of known structure (anatomical or functional) in the data and often leads to an overly conservative analysis. The next sections discuss why explicitly modeling the covariance among ROIs can overcome these limitations and how to implement such models.

4 Advantages of Explicitly Modeling ROI Covariance

Incorporating the covariance structure of ROIs directly into a unified model has several clear advantages:

4.1 Improved Statistical Power

By modeling correlations, a joint model can distinguish signal from noise more effectively than many separate tests. As one study noted, accounting for the actual correlations in the data by means of random effects can lead to an increase in statistical power and reduction of inferential biases Parekh et al., 2024. Intuitively, a multivariate model “knows” that ROIs move together, so it doesn't treat correlated fluctuations as independent evidence against an effect. Instead, it pools evidence across ROIs appropriately. This often allows detection of effects that are diffuse or spread across regions, which a univariate approach might miss.

4.2 Partial Pooling and Regularization

Multilevel (hierarchical) models perform partial pooling of information across units (here, across ROIs). Rather than estimating a completely separate effect for each ROI (no pooling) or forcing one identical effect for all ROIs (complete pooling), a hierarchical model finds a middle ground. It estimates an overall effect and ROI-specific deviations, shrinking individual ROI estimates toward the group mean based on the data. This adaptive regularization reduces the risk of false positives (extreme values are pulled toward the center) while improving the precision of each ROI's estimate by borrowing strength from the others. Essentially, the model “learns” the similarity among ROIs – if many ROIs show a similar effect, the model is confident in that shared effect; if one ROI is an outlier with little support from others, it is shrunk more toward zero.

4.3 No Need for Post-hoc Multiplicity Correction

When the covariance structure is modeled within a single framework, the notion of performing N separate tests largely disappears – the model yields joint inferences. A Bayesian multilevel approach, for example, inherently handles the multiplicity by treating ROI effects as random variables with a distribution, rather than fixed unknowns each needing a separate test. As Chen et al. (2019) put it, instead of separately correcting for multiple testing, Bayesian multilevel modeling (BML) incorporates multiple testing as part of the model by assigning a prior distribution among the ROIs (treating ROIs as random effects). The result is that the “multiple comparisons problem” is largely mitigated – credible effects emerge from the model with appropriate shrinkage, and one does not apply arbitrary p-value thresholds that depend on the number of tests. Even in frequentist multivariate models (e.g. MANOVA or multivariate linear models), the overall test of an effect uses the combined covariance of outcomes, yielding one significance test that inherently accounts for multiple degrees of freedom without requiring Bonferroni-type adjustments.

4.4 Greater Inference Efficiency and Reliability

A single integrative model makes use of all available data and constraints simultaneously, often leading to more stable and reliable estimates. For instance, a recent Bayesian ROI analysis framework demonstrated that using one multilevel model for all regions improves inference efficiency and places all the regions on an equal footing, so that small ROIs are not disadvantaged by having lower signal or volume. In classical analyses, a small ROI might fail to reach significance simply because its effect estimate is noisier (higher variance), whereas a multilevel model will appropriately propagate that uncertainty and possibly share information from larger, more reliable ROIs. This contributes to fairer and more nuanced inference, where we can report the full distribution of effects across ROIs rather than a dichotomous significant/not-significant outcome for each region. Modeling covariance can also increase spatial specificity – by accounting for broad patterns, the model can more sharply identify which regions deviate from the pattern in relation to the outcome.

4.5 Capturing Complex Biological Structure

The brain's organization is hierarchical and network-based. By leveraging known ROI relationships (e.g., grouping ROIs into larger networks or using anatomical proximity), one can specify structured covariance in a multilevel model. This goes beyond simply allowing covariance – it uses prior knowledge to inform the covariance model. For example, one could allow ROI random effects to have a covariance that is higher for ROIs within the same lobe or network, or impose a spatial correlation decay with distance. This way, the model formalizes the idea that certain ROIs should exhibit similar effects. The result can be more biologically interpretable models that mirror the brain's structure. Even without explicit anatomical priors, multivariate methods naturally capture common variations: e.g., Mahalanobis distance (Hotelling's T^2) combines multiple ROI metrics into a single multivariate deviation score, inherently accounting for their covariance. Such approaches have been shown to better characterize complex brain-behavior relationships than multiple univariate tests Tremblay et al., 2024.

In essence, explicitly modeling ROI covariance yields more informative and powerful analyses by using the data's covariance structure as an asset rather than a nuisance. It shifts the focus from controlling false positives via harsh penalties to modeling the data more faithfully, thus achieving both control of false positives and improved sensitivity through shrinkage and information-sharing. The following section will outline the methodological approaches to implement these ideas.

5 Alternative Modeling Approaches

There are two broad statistical strategies to account for covariance among ROI measures in a unified analysis: multivariate outcome models and multilevel (mixed-effects) models. In practice, these approaches overlap and can be combined (for example, a multilevel model can have multivariate outcomes). We will discuss each in turn, along with how they leverage ROI structural information. The choice of approach may depend on the research question and data structure (e.g., whether we have multiple ROI measurements as outcomes, or ROIs as predictors of an outcome, etc.), but both aim to move beyond treating ROIs independently.

5.1 Multivariate GLM and MANOVA Approaches

A multivariate General Linear Model (GLM) treats multiple ROI measures as simultaneous dependent variables in one model. For example, instead of running separate univariate regressions for each of 10 ROIs, one could run a single multivariate regression with a 10-dimensional outcome vector. Classical statistics provides tests like Hotelling's T^2 (for comparing multivariate means between two groups) or MANOVA (multivariate ANOVA) for factor effects on multiple outcomes. In neuroimaging, this approach has been implemented for group analysis where each ROI or condition is considered an outcome in a multivariate test. Software like AFNI's 3dMVM Chen et al., 2014 and the MATLAB toolbox MRM (Multivariate and Repeated Measures) McFarquhar et al., 2016 are designed for such analyses at the group level.

The multivariate model estimates a covariance matrix for the ROI outcomes and uses it to evaluate hypotheses. For example, MANOVA might test if the vector of mean ROI values differs by group, taking into account the ROI covariance. If an overall effect is found, one can probe which ROIs contribute via post-hoc tests or dimension-reduction (e.g., canonical variates). Crucially, the test inherently accounts for ROI correlations, offering a single omnibus assessment of an effect across all regions. This bypasses the need for multiple univariate tests and corrections. This approach directly addresses whether there is any multivariate effect of interest (e.g., does the brain ROI profile as a whole predict the outcome?). It capitalizes on covariance: a combination of small effects across ROIs can yield a significant multivariate effect even if each ROI alone was not significant. This study (McFarquhar et al., 2016) shows that a multivariate framework can handle repeated measures and multimodal data without the restrictive assumptions of univariate repeated-measures GLM. They describe how conventional methods either ignore repeated measures or assume constant covariance (sphericity), and show that a multivariate GLM can address these issues in a simple and efficient manner.

Traditional MANOVA/MVM can be less familiar to researchers and often has higher computational cost. Indeed, multivariate tests like Wilks' Lambda or Pillai's Trace involve matrix operations that scale with the number of ROIs. For large numbers of ROIs or voxels, these tests become computationally heavy. Additionally, while an omnibus test is useful, researchers often still want to know which ROIs are driving an effect. This requires follow-up analysis (such as examining the estimated coefficients for each ROI or performing stepdown tests). New tools like the MRM toolbox provide permutation-based inference to alleviate distributional concerns and even incorporate discriminant analysis to interpret the multivariate effects.

Example: Suppose we have 5 ROI connectivity metrics as outcomes and a binary group (patients vs. controls) as predictor. A MANOVA can test the hypothesis of any group difference in the 5-dimensional ROI profile. If significant, one might then look at the group difference vector or use canonical correlation analysis to see which linear combination of ROIs best separates the groups. In fact, one study combined mixed-effects multilevel analysis with canonical correlation analysis (CCA) to relate individual variability in fMRI ROI activations to behavior Jo et al., 2021 – the mixed-effects model captured subject-specific ROI activation patterns, and CCA found a multivariate brain-behavior relationship, illustrating a powerful synergy of multilevel and multivariate methods.

Multivariate GLM approaches are a rigorous way to handle multiple ROI outcomes together. They are particularly useful when one wants an overall test of brain-behavior association or when the number of ROIs is moderate. For very high-dimensional ROI sets (e.g., whole-brain parcellations with >100 regions), purely multivariate methods may become unwieldy, and that is where multilevel models and Bayesian methods (discussed next) become attractive for incorporating regularization.

5.2 Multilevel Mixed-Effects Models

What it is. A multilevel model (also known as a mixed-effects or hierarchical model) introduces random effects to represent the hierarchy or clustering in data – here, the clustering of measurements within subjects (and possibly within ROIs). In ROI analysis, we commonly have subjects as one level and ROI measurements as a repeated measure within subjects. A multilevel model can explicitly include effects that vary by ROI and/or by subject. For example, one can model an outcome (behavior) as a function of brain ROI measures with subject-level random intercepts (to account for overall subject differences) and ROI-level random effects (to account for variability across regions, or correlations within a region).

Two perspectives. There are two equivalent ways to apply multilevel modeling in this context, depending on whether the brain measures are predictors or outcomes in the model:

- **(a) ROI-as-outcome approach:** Here, the ROI measure is the dependent variable, and subject-level predictors (or group labels) are independent variables. For example, Y_{ij} is the brain measure in ROI j for subject i. We can fit a mixed model like:

  $$Y_{ij} = \beta_0 + \beta_1 X_i + u_j + v_j X_i + a_i + \epsilon_{ij},$$

  where X_i is a predictor (e.g., a clinical score or a group indicator). In this setup, u_j is a random intercept for ROI j (accounting for differing overall means by region), v_j is a random slope for the effect of X in ROI j (so each ROI can have a different association with the predictor), and a_i is a random intercept for subject i (accounting for subject-specific offsets affecting all their ROIs). This is a cross-classified multilevel model: ROI and subject are two grouping factors. The model yields an overall effect β_1 (the average relationship between X and the ROI measure across regions) and ROI-specific effects $(\beta_1 + v_j)$ that deviate for particular regions. Crucially, we can model a covariance structure for the random effects v_j (and u_j), meaning the ROI-specific deviations are assumed to come from a distribution, often multivariate normal. If we allow u_j and v_j to be correlated random effects, the model learns the covariance between ROI baselines and their sensitivity to X. We also inherently model the within-subject covariance: because each subject has a random intercept a_i, any two measures from the same subject are correlated through a_i. This structure is analogous to a repeated-measures ANOVA but far more flexible (not requiring equal correlations or sphericity).

- **(b) ROI-as-predictors approach:** Alternatively, one may be interested in predicting a single outcome per subject (say a cognitive score Z_i) using multiple ROI measures as predictors. A direct approach is a multiple regression:

  $$Z_i = \gamma_0 + \sum_{j=1}^{p} \gamma_j Y_{ij} + \epsilon_i.$$

However, this is prone to overfitting and multicollinearity if p (number of ROIs) is large relative to n (subjects). A multilevel Bayesian approach can help here by treating the coefficients γ_j as coming from a distribution – effectively a random effects prior on the ROI regression coefficients. This imposes shrinkage and can incorporate knowledge about ROI structure in the prior covariance of γ_j. In practice, one might perform dimension reduction (PCA or partial least squares) as a pre-processing step to handle many collinear predictors. But an appealing fully Bayesian solution is to use hierarchical priors: for example, assume $\gamma_j \sim \mathcal{N}(\mu, \tau^2)$ for all ROIs j, which couples the strength of all ROI effects. This shrinks all γ_j toward a common mean μ, essentially partial pooling in the predictor space. More sophisticated models could let μ vary by ROI group (network) or specify a prior covariance matrix between the γ_j's based on ROI distances or connectivity. While this approach is conceptually slightly different (outcome as response rather than ROI measures), it aligns with the same philosophy: exploit ROI covariance (through prior or random effects) to stabilize estimates. It is often implemented via Bayesian regression (e.g., using regularized horseshoe priors or Gaussian processes over ROI indices).

Why it works. Multilevel models shine in sharing information across units. In our context, that means sharing information across ROIs and/or across subjects. By structuring the model in levels, we preserve the rich dependency structure in the data instead of averaging it away or testing it piecewise. The strength of this approach is summarized by Chen et al., 2019 the strength of multilevel modeling lies in its capability of stratifying the data so that complex dependency or correlation structures can be properly accounted for coherently within a single model. In a crossed ROI-by-subject design, a single multilevel model can replace what would otherwise be a "laundry list" of separate tests (one per ROI) and corrections. The model's estimates of ROI-specific effects are more reliable due to partial pooling, and the overall evidence for an association is integrated. In essence, the multilevel model is doing an adaptive pooling of ROIs: if ROIs behave similarly with respect to the outcome, it will treat them almost like a group (increasing effective sample size for that common effect); if a particular ROI stands out, it can still be detected as a deviation, but with acknowledgment of the uncertainty.

Incorporating Structural Information. One can enhance multilevel models by informing the random effect covariance structure. For example, in a Bayesian multilevel model, one might specify a covariance matrix for the ROI random slopes v_j such that $\mathrm{Cov}(v_j, v_{j'})$ is high if ROIs j and j' are in the same network or anatomical structure, and lower if they are distant. This is akin to a random effect correlation matrix informed by prior knowledge. In practice, a Conditional Autoregressive (CAR) or Gaussian Process prior can be placed on ROI random effects to encode spatial or network smoothness. Some software (like `brms` in R) supports this via custom covariance structures (e.g., `cor_car()` for spatial correlation given an adjacency matrix of ROIs). By leveraging such structural info, we prevent the model from overfitting noise in any single ROI and encourage it to learn coherent patterns. Even without explicit priors, the data itself can reveal structure: for instance, if all ROIs within the temporal lobe show similar random slopes, the posterior of their effects will be more tightly clustered, effectively discovering that as a grouping.

Bayesian vs. Frequentist Implementation. While one can fit multilevel models with frequentist methods (REML via packages like `lme4` or `nlme` in R), complex covariance structures and many random effects can lead to convergence issues or high computational burden. Bayesian methods (via MCMC) are more computationally intensive per iteration, but they offer more flexibility (any prior covariance structure can be implemented) and regularization that can actually aid convergence by ruling out degenerate parameter combinations. Modern Bayesian tools (`Stan`, `rstanarm`, `brms`) have made it practical to do full Bayesian multilevel modeling even with moderately large ROI sets. For example, the No-U-Turn Sampler (NUTS) in Stan enables efficient exploration of high-dimensional posteriors, making it feasible to perform full Bayesian inferences for BML with datasets of moderate size. If the number of ROIs is very large (hundreds), one may still need to be careful; methods like variational Bayes or reducing dimensionality might be used as approximations. Recent developments like the FEMA (Fast and Efficient Mixed-Effects Algorithm) aim to fit massive mixed models (e.g., whole-brain voxel/vertex-wise analyses) faster, indicating the field's push towards incorporating multilevel models at scale.

Example. Using the earlier scenario of 20 ROIs predicting a cognitive score, a Bayesian multilevel regression could treat the 20 ROI coefficients as random draws from a common distribution. If the model finds that most ROIs have near-zero association, but a subset (say hippocampal ROIs) have stronger positive coefficients, those hippocampal effects will stand out but also be estimated with partial pooling (less uncertainty than separate regressions). Conversely, if all ROIs truly have some similar association with the score, the model will capture a common effect (in the group-level μ) and shrink individual ROI differences, yielding a more precise estimate of the general brain-behavior effect. Instead of multiple comparison corrected p-values per ROI, one would directly obtain posterior credible intervals for each ROI's effect and for the global effect, with the model inherently controlling false positives through shrinkage.

6 Current Research and Developments

There is a growing body of research proposing and utilizing models that account for ROI covariance in neuroimaging analysis. We highlight a few notable developments and studies:

Bayesian Multilevel Modeling for Neuroimaging (Chen et al., 2019). Chen and colleagues introduced a Bayesian multilevel (BML) framework (implemented in AFNI's `3dLMEr` and related tools) specifically to tackle the multiplicity problem at the ROI or voxel level. Their approach pools information across ROIs, controlling type I error by shrinkage rather than after-the-fact correction. They demonstrated that BML can "dissolve" the multiple testing issue by having one integrative model, and showed in examples that small ROIs with weaker signals are no longer unfairly penalized – instead, the model yields results for all ROIs with associated uncertainty, often increasing the overall discovery of meaningful effects. This work is part of a broader shift towards Bayesian inference in neuroimaging, emphasizing estimation over strict hypothesis testing. The authors also highlight quality control measures like posterior predictive checks to ensure the model fits the data well, which is an advantage of Bayesian implementation. Earlier, Chen et al. developed `3dMVM` (for multivariate modeling) and `3dLME` (for linear mixed-effects) in AFNI, enabling group-level fMRI analysis that can handle complex designs (e.g., within-subject factors, multiple conditions or ROIs) in a single model. These tools brought multivariate ANOVA and mixed modeling into practical use for neuroimagers not comfortable coding their own models. For example, `3dMVM` can test an interaction effect across multiple ROIs jointly, and `3dLME` can accommodate longitudinal and repeated measures without needing to flatten data into one time point per subject. The Multivariate and Repeated Measures (MRM) toolbox was introduced to fill the gap in general neuroimaging software for true multivariate modeling. It provides a user-friendly GUI to specify design matrices with multiple outcomes (ROIs, time points, modalities) and runs permutation tests for significance. The authors compared their approach to standard ones and found that MRM could detect effects that a mass-univariate approach would miss, especially in designs with within-subject factors. MRM's use of permutations also offers a robust way to obtain p-values without relying on large-sample assumptions. Recognizing the computational challenge of fitting multilevel models for high-dimensional neuroimaging data, Parekh et al. introduced FEMA, which streamlines LME model fitting for each voxel or ROI in large datasets (e.g., UK Biobank). While FEMA is aimed at mass-univariate analysis (voxel-wise mixed models), the technology is relevant for ROI-based analysis too: it shows that even with thousands of vertices/regions, one can computationally manage a mixed model by clever factorization and parallelization. This means modeling covariance at the whole-brain level is becoming more feasible, narrowing the gap between ideal modeling and practical execution. An example of cutting-edge application is the combination of hierarchical Bayesian modeling with multivariate fMRI decoding to assess individual differences Freund et al., 2023. Freund et al. showed that by using a multilevel model, they could improve the reliability of individual-specific brain activation measures over time, and by using multivariate (pattern-based) measures, they increased precision in correlations. Although this study was not about ROI per se, it exemplifies the benefits of integrating multilevel and multivariate thinking: hierarchical models for subjects/trials and multivariate models for brain response patterns. It underscores a general point that echoes in ROI analyses: integrating across levels and regions yields more reliable individual effect estimates than analyzing each piece separately. Outside of explicit ROI modeling, there are multivariate pattern analysis (MVPA) techniques and machine learning methods that inherently account for multiple features (voxels or ROIs) together (e.g., support vector machines, regularized regression, deep nets). While often used for prediction rather than inference, these methods demonstrate the power of combining features. Recent trends aim to marry these techniques with statistical inference – for example, performing permutation importance testing in a random forest to see which ROI features are important, or using graphical models to identify network connections predictive of outcome Lukemire et al., 2021. Some approaches, like multivariate distance matrix regression (MDMR), specifically test for associations between a set of variables (like all ROI measures) and an outcome by considering the multivariate distance among subjects' ROI profiles. These are beyond the scope of classic multilevel models but are part of the ecosystem of solutions to the same core problem: capturing the joint effect of correlated brain measures. The methods community is active. One can find recent preprints on arXiv/bioRxiv dealing with topics like hierarchical Bayesian models for brain connectivity Colclough et al., 2018, or normative modeling with hierarchical regression Kia et al., 2021. Many of these use ROI-based representations. For instance, a preprint by Mejia et al., 2020 introduced Bayesian joint modeling of multiple brain networks, treating ROI connectivities as multivariate outcomes with sparsity priors (to identify which connections matter). The development of methods such as Bayesian covariance regression (which allows covariates to influence the covariance between ROIs) and hierarchical graphical models (which infer networks at group and subject levels) are particularly relevant for DTI and functional connectivity data, where the interest lies in covariance (connectivity) itself. These advanced models go a step further: not only do they account for covariance, they make the covariance the object of study.

In summary, the field is clearly moving toward more integrated models. Both software tools and statistical methodologies are being refined to handle the complexity of neuroimaging data. Researchers today have access to packages in R (as we'll see next) and other platforms that make these approaches accessible without needing to reinvent the wheel. The consensus from recent literature is that modeling the covariance structure is not just an academic exercise – it has tangible benefits for discovery and reproducibility in brain-behavior studies.

7 Conclusion

Modern neuroimaging studies increasingly recognize that the brain does not operate in isolated pieces — our analysis methods should reflect this reality. Modeling the covariance structure among ROI measures using multilevel models (and related multivariate techniques) provides a statistically principled way to analyze brain data holistically. By moving away from treating each ROI as an independent test, we gain in multiple ways: higher power to detect effects that are distributed across regions, greater protection against false positives through shrinkage, and more nuanced insight into how different regions jointly relate to behaviors or clinical outcomes.

We have discussed how the traditional ROI-by-ROI analysis can be improved by explicitly accounting for inter-ROI correlations. Multivariate GLM approaches (like MANOVA or MVM) offer omnibus tests that incorporate covariance, while hierarchical models offer flexibility in partial pooling and the ability to model complex structures (crossed random effects, spatial priors, etc.). The consensus from recent research is that such models are statistically superior because they make use of all available information and reflect the data-generating process more closely. As a result, they can reveal effects that a mass-univariate approach would overlook, and they report results in a way that is more quantitative (effect size distributions) rather than qualitative (significance yes/no per ROI).

For researchers, the practical barriers to adopting these models are coming down. User-friendly tools exist for both frequentist and Bayesian paradigms, and computational advances (along with increasing computing power) allow even whole-brain multilevel models to be fitted in reasonable time. It is now possible to include hundreds of ROIs or even voxel-wise data in mixed models with approaches like FEMA, or to use Bayesian inference to enrich group analyses with full posterior uncertainty on effects. With these tools, one can leverage known ROI structure (anatomical hierarchies, networks, etc.) to inform the analysis, leading to findings that are not only statistically sound but also neuroscientifically interpretable.

In conclusion, modeling ROI covariance within a unified multilevel model represents a significant advance in neuroimaging statistics. It aligns with the move toward multivariate, integrated analysis in many fields of science. By embracing these methods, researchers can improve the sensitivity of their studies (finding true effects with fewer subjects or scans), reduce spurious results (through proper error control and regularization), and ultimately paint a more coherent picture of how brain structure and function relate to the variables of interest. As datasets grow in size and complexity (think multimodal, longitudinal, multi-center studies), multilevel and multivariate modeling will be not just advantageous but essential to fully exploit the data. The methodologies and examples outlined here provide a roadmap for incorporating these techniques into ROI analyses today, enabling more robust and insightful neuroscience

References

Chen, G., Taylor, P. A., & Cox, R. W. (2019). Handling multiplicity in neuroimaging through bayesian lenses with multilevel modeling. *Frontiers in Neuroscience*, *13*, 300.

McFarquhar, M., McNab, F., Jbabdi, S., Byatt, G., & Henson, R. N. (2016). Multivariate and repeated measures (mrm): A new toolbox for dependent and multimodal group-level neuroimaging data. *NeuroImage*, *132*, 373–389.

Parekh, R. et al. (2024). Fema: Fast and efficient mixed-effects algorithm for large sample whole-brain imaging data [In press]. *Human Brain Mapping*.

Tremblay, et al. (2024). Mvcomp toolbox: Multivariate comparisons of brain mri features accounting for common information across measures. *Aperture Neuro*, *4*, 118427.

Chen, G., Adleman, N. E., Leibenluft, E., & Cox, R. W. (2014). Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate glm. *NeuroImage*, *99*, 571–588.

Jo, S., Kim, H.-C., Lustig, N., Chen, G., & Lee, J.-H. (2021). Mixed-effects multilevel analysis followed by canonical correlation analysis is an effective fmri tool for the investigation of idiosyncrasies [Epub 2021 Aug 20]. *Human Brain Mapping*, *42*(16), 5374–5396. https://doi.org/10.1002/hbm.25627

Freund, M. C. et al. (2023). Complementary benefits of multivariate and hierarchical models for identifying individual differences in cognitive control. *Imaging Neuroscience*, *2*(1), e10.

Lukemire, J., Kundu, S., Pagnoni, G., & Guo, Y. (2021). Bayesian joint modeling of multiple brain functional networks [Epub 2020 Sep 1]. *Journal of the American Statistical Association*, *116*(534), 518–530. https://doi.org/10.1080/01621459.2020.1796357

Colclough, G. L., Woolrich, M. W., Harrison, S. J., Rojas López, P. A., Valdes-Sosa, P. A., & Smith, S. M. (2018). Multi-subject hierarchical inverse covariance modelling improves estimation of functional brain networks. *NeuroImage*, *178*, 370–384. https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.04.077

Kia, S. M., Huijsdens, H., Rutherford, S., Dinga, R., Wolfers, T., Mennes, M., Andreassen, O. A., Westlye, L. T., Beckmann, C. F., & Marquand, A. F. (2021). Federated multi-site normative modeling using hierarchical bayesian regression. *bioRxiv*. https://doi.org/10.1101/2021.05.28.446120

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 [Epub 2019 Jun 12]. *Journal of the American Statistical Association*, *115*(530), 501–520. https://doi.org/10.1080/01621459.2019.1611582