

RESEARCH ARTICLE

# Experimental design modulates variance in BOLD activation: The variance design general linear model

Garren Gaut<sup>1</sup>  | Xiangrui Li<sup>2,3</sup> | Zhong-Lin Lu<sup>2,3</sup> | Mark Steyvers<sup>1</sup> 

<sup>1</sup>Department of Cognitive Science, University of California Irvine, Irvine, California

<sup>2</sup>Center for Cognitive and Behavioral Brain Imaging, The Ohio State University, Columbus, Ohio

<sup>3</sup>Department of Psychology, The Ohio State University, Columbus, Ohio

## Correspondence

Mark Steyvers, Department of Cognitive Science, University of California Irvine, Irvine, CA.

Email: mark.steyvers@uci.edu

## Funding information

National Science Foundation, Grant/Award Numbers: 1533500, 1533661

## Abstract

Typical fMRI studies have focused on either the mean trend in the blood-oxygen-level-dependent (BOLD) time course or functional connectivity (FC). However, other statistics of the neuroimaging data may contain important information. Despite studies showing links between the variance in the BOLD time series (BV) and age and cognitive performance, a formal framework for testing these effects has not yet been developed. We introduce the variance design general linear model (VDGLM), a novel framework that facilitates the detection of variance effects. We designed the framework for general use in any fMRI study by modeling both mean and variance in BOLD activation as a function of experimental design. The flexibility of this approach allows the VDGLM to (a) simultaneously make inferences about a mean or variance effect while controlling for the other and (b) test for variance effects that could be associated with multiple conditions and/or noise regressors. We demonstrate the use of the VDGLM in a working memory application and show that engagement in a working memory task is associated with whole-brain decreases in BOLD variance.

## KEYWORDS

brain mapping, functional magnetic resonance imaging, image processing, linear models

## 1 | INTRODUCTION

At their core, neuroimaging analyses consist of relating a summary statistic of the blood-oxygen-level-dependent (BOLD) time course to experimental condition, behavior, or individual characteristics. The primary method for fMRI analysis, the general linear model (GLM; Bullmore et al., 1996; Friston, Jezzard, & Turner, 1994), focuses on the mean trend in the BOLD activation. Recently, researchers have moved beyond mean BOLD trend by studying functional connectivity (FC), which is calculated as the Pearson correlation between regions. However, it is possible that other statistics of the neuroimaging data may contain important information. A natural candidate is BOLD variability (BV) defined as the variance in the BOLD time series. BV can be thought of as intermediate to mean BOLD trend and FC in terms of computational complexity; BV is based on a locally independent computation whereas FC incorporates between-region dependencies. Importantly, as FC and mean BOLD trend have led to distinct avenues

of research, BV could be a fundamentally different channel for studying brain function.

This article introduces the variance design general linear model (VDGLM), a novel framework that allows researchers to simultaneously test for effects on BV and on mean BOLD activation. The VDGLM can be conceptualized as a GLM that explicitly incorporates the experimental design into the model of the variance. Direct incorporation of the experimental design allows the VDGLM to be flexible enough to be used in any fMRI experiment. This new framework facilitates the analysis of BV effects and enables new discoveries that relate BV to disease, individual characteristics, and human behavior.

The development of the VDGLM was motivated in part by studies in EEG and fMRI that have demonstrated relationships between brain fluctuations and cognitive processes, behavior, and age. EEG studies have shown variance effects in the form of suppression of alpha and theta oscillatory waves, that is, reduction of the amplitude of the oscillations (see (Klimesch, 1999) for a review). Alpha suppression is

related to task engagement (Williamson, Kaufman, Lu, Wang, & Karron, 1997), opening the eyes (Berger, 1929), sleep (Dement & Kleitman, 1957), and cognitive performance (Klimesch, 1999). Aging studies have found that alpha, delta, and theta suppression increases with age during youth (Somsen, van't Klooster, van der Molen, van Leeuwen, & Licht, 1997) and that alpha suppression decreases with age in older populations (Duffy, Albert, McNulty, & Garvey, 1984). Alpha suppression's relationship to cognitive and individual differences is mediated by inhibitory-control processes involved in attention (Klimesch, 2012; Klimesch, Schack, & Sauseng, 2005).

In fMRI, brain fluctuations measured by BV have been shown to vary with behavior and age. BV varies between task and fixation, particularly in younger adults. Differences in BV between task and fixation are associated with higher visual discrimination performance (Wutte, Smith, Flanagan, & Wolbers, 2011) and track task difficulty (Garrett, McIntosh, et al., 2013). Age has also been related to BV; BV was shown to predict age with five times the explanatory power of mean BOLD (Garrett, Kovacevic, McIntosh, & Grady, 2010) and was indicative of younger, faster, and more consistently performing subjects (Garrett, Kovacevic, McIntosh, & Grady, 2011). In both studies, the spatial distribution of BV effects was orthogonal to the distribution of mean effects. Furthermore, a follow-up study found that age related BV affects were robust to vascular controls (Garrett, Lindenberger, Hoge, & Gauthier, 2017). Gaut et al. (2018) found that BV could be used to accurately predict the task a subject was performing and subject identity. Despite these links to behavior and age, the study of BOLD variability has not been widely pursued in fMRI.

While the aforementioned studies have led to the development of general pipelines for BV analysis (e.g., Garrett et al., 2010; Garrett, McIntosh, et al., 2013), the studies did not use a general statistical framework for simultaneously studying BV and mean effects. Therefore, a key motivation for the VDGLM framework is to introduce a parametric approach that *jointly* models the mean and variance by explicitly incorporating the experimental design into the variance formulation. The inclusion of the design in the variance allows us to: (a) jointly model mean and variance effects, (b) explicitly model the temporal dynamics between BV and experimental condition, and (c) include multiple experimental conditions in our variance analyses. The explicit structure of mean and variance design is supplied by the researcher, which allows for easy generalization to any experiment, and model fitting is computationally efficient enough to run region of interest (ROI) analyses on large brain imaging studies. By developing this framework, we are providing an important tool to test for variance effects that has the potential to spur new research developments in various fields.

The plan for the rest of the article is as follows. We first provide an overview of the GLM, and establish the theory behind the VDGLM. Next we provide an application of the VDGLM to Working Memory data from the Human Connectome Project Healthy Adult dataset. We finish with a discussion of the choices made when using the VDGLM and how variance measured by the VDGLM compares to other measures of variance.

## 2 | A NOVEL FRAMEWORK FOR STUDYING BV

To motivate the VDGLM, consider a hypothetical BOLD activation time series where BV is affected by an experimental condition that indicates fixation versus task (see Figure 1). The single experimental condition is plotted in black and the BOLD time series from a single voxel is plotted in blue. For conceptual simplicity, the experimental condition time series is not convolved with a hemodynamic response function (HRF) model. The voxel time series varies as a function of task condition; the variance is higher during fixation compared to task, consistent with reduced default mode network activations during cognitive tasks. We can describe these effects on the mean and variance using a simple VDGLM:

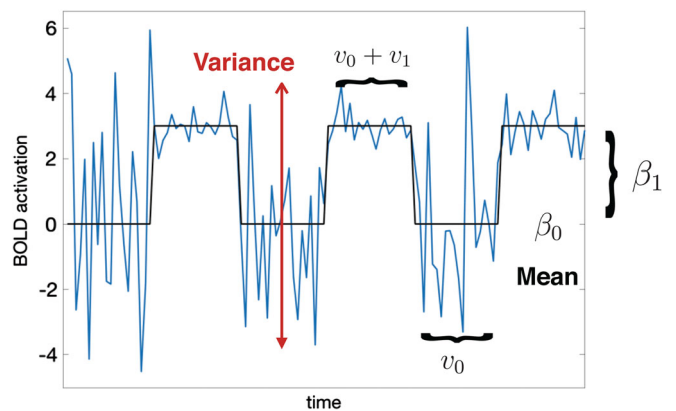
$$y \sim N(\beta_0 + x^T \beta_1, (v_0 + x^T v_1)I) \quad (1)$$

where,  $y$  is the BOLD time series,  $x$  represents the condition indicator,  $I$  is the identity matrix,  $\beta_0$  captures the mean activation, and  $v_0$  captures the measurement variance, that is, the out-of-task variation. Then  $\beta_1$  and  $v_1$  capture the degree of change in mean and variance due to task engagement, respectively. If the model is applied to the data from Figure 1, we expect  $\beta_0 \approx 0$ ,  $\beta_1 \approx 3$ ,  $v_0$  to be a large positive value, and  $v_1$  to be negative, but with the constraint that  $|x_t v_1| < |v_0| \quad \forall t$ . Here, the parameter  $v_1 < 0$  reflects the fact that BOLD variation is lower within task compared to fixation.

For comparison, the GLM estimates a single variance parameter over the entire time series and ignores the change in variance due to the experimental manipulation:

$$y \sim N(\beta_0 + x^T \beta_1, v_0 I) \quad (2)$$

Note that Equation (2) is equivalent to Equation (1) when  $v_1 = 0$ , that is, the GLM is a nested model of the VDGLM where there is no



**FIGURE 1** Illustration of artificial data where the presence of a single experimental condition (black) increases the mean but lowers the variance of the BOLD time course (blue). The values used to create this visualization are based on Equation (1) with  $\beta_0 = 0$ ,  $\beta_1 = 3$ ,  $v_0 = 2$ , and  $v_1 = -1.5$ . BOLD, blood-oxygen-level-dependent [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

experimental modification of the variance. The GLM is a null model for no effect of the variance that can be compared to the VDGLM (a) to explicitly test for variance inclusion and (b) test whether the mean effects found by the VDGLM are similar to the mean effects found by the GLM.

## 2.1 | VDGLM analysis pipeline

One goal of the VDGLM framework is to allow the VDGLM to be inserted into any standard fMRI analysis pipeline with minimal modifications (see Figure 2). The VDGLM does not change data acquisition, preprocessing, model comparison, or results dissemination. The main step that must change is model formulation and estimation. In some cases, the inference step is not affected (e.g., computing effect sizes using Cohen's  $d$ ). However, more sophisticated inference such as parameter significance testing will require modification of the inference step to include statistics for testing variance parameters (see Section 2.5). In a BV focused analysis, we also recommend additional preprocessing steps to remove variance confounds such as censoring and head motion correction, but these additional steps are not necessary to use the VDGLM.

## 2.2 | Matrix notation

We can write the GLM and the VDGLM in matrix notation to highlight the concept of inserting the design matrix into the variance formulation. The GLM models the BOLD time series  $y_{[T \times 1]}$  from a single voxel as a linear function of the experimental design (Beckmann, Jenkinson, & Smith, 2003; Bullmore et al., 1996; Friston et al., 1994; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004). Formally, the GLM is defined:

$$\begin{aligned} y &= X\beta + \epsilon \\ \epsilon &\sim N(0, I\sigma^2) \end{aligned} \quad (3)$$

where,  $X$  is a  $T \times p$  design matrix,  $\beta$  is a  $p \times 1$  vector of mean parameters,  $\sigma^2$  is a variance parameter, and  $I$  is the identity matrix. The columns of the design matrix  $X$  include experimental events,

experimental blocks, stimulus presentation, or mean activation. The VDGLM has the same formulation, but extends the variance model:

$$\begin{aligned} y &= X_m\beta + \eta \\ \eta &\sim N(0, \text{diag}(X_v v)I) \\ \text{diag}(X_v v)I &> 0 \end{aligned} \quad (4)$$

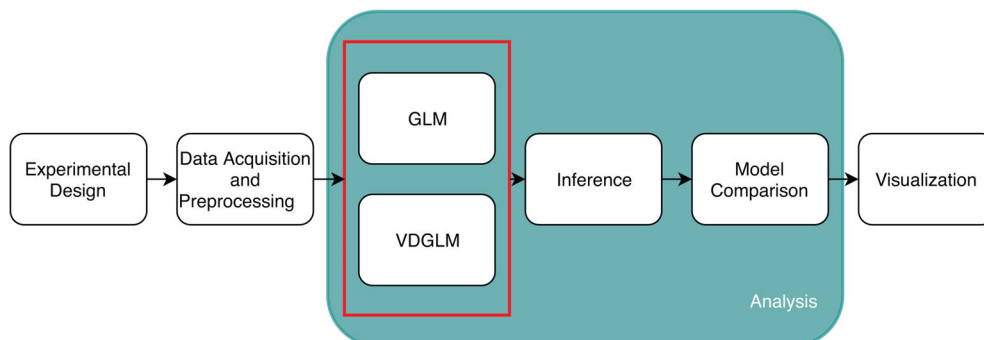
where,  $\text{diag}(x)$  is the matrix with the entries of the vector  $x$  along the diagonal. To emphasize that the mean and variance design matrices can be distinct, we use the notation  $X_m$  and  $X_v$  to denote the mean and variance designs, respectively. The parameters,  $\beta$  and  $v$ , capture mean and variance effects, respectively. It is clear that the GLM (Equation (3)) is a special case of the VDGLM for which the variance design matrix  $X_v$  is a single column of ones and  $v = \sigma^2$ .

## 2.3 | Prewhitening and noise Regressors

In GLM analyses, BOLD time series are often “prewhitened” to account for BOLD autocorrelation (Bullmore et al., 1996; Woolrich, Ripley, Brady, & Smith, 2001). The results in this article are based on VDGLM without a prewhitening step to avoid interfering with the variance estimates. We have also investigated the results that includes a prewhitening step and verified that the variance effects found by the VDGLM are not artifacts caused by autocorrelation. The results with the prewhitening step are very similar to the results without prewhitening reported in this article. Other techniques for controlling noise (e.g., coloring or head motion correction) involve the addition of noise regressors. As with GLM analyses, the VDGLM can incorporate these techniques by including the appropriate regressors in the design matrices.

## 2.4 | Estimation

Univariate GLM mean parameter estimation can proceed in one of two ways: (a) ordinary or general least squares (Beckmann et al., 2003; Friston, Holmes, et al., 1994) or (b) fully Bayesian inference



**FIGURE 2** An illustration of a typical fMRI pipeline that uses either the GLM or the VDGLM for analysis. To use the VDGLM, the only steps from a traditional pipeline that must change are model formulation and estimation. Data acquisition, preprocessing, model comparison, and methods for result dissemination can remain the same. Some inference steps, such as effect size estimation, could remain the same. However, other inference procedures, such as significance testing, require statistics developed expressly for the VDGLM. GLM, general linear model; VDGLM, variance design general linear model [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(Woolrich et al., 2004). Approximate Bayesian inference has also been used, but in a single group-level analysis that combines first and second-level models (Friston et al., 2002). Variance estimation is usually done by iteratively computing OLS estimates (Woolrich et al., 2001; Worsley et al., 2002), but can require more advanced methods depending on the structure of group-level models (see Section 2.5).

There are potentially many approaches that could be used to estimate the VDGLM, including Bayesian and maximum likelihood approaches. For simplicity, we use a maximum likelihood approach (see (Eklund, Lindquist, & Villani, 2017) for a Bayesian approach that has the goal of removing motion artifacts). Estimation approaches must be computationally efficient enough to handle the high dimensionality of fMRI data and in practice, we found that sampling techniques were too slow to be practical for large data sets. Maximum likelihood (or maximum a posteriori) estimation using mode-finding algorithms is efficient enough to estimate parameters for an ROI analysis from a large fMRI dataset in about half a day using parallel computing techniques.

There is known bias in variance estimates when computing maximum likelihood estimate solutions (Harville, 1977). However, in fMRI studies, the large number of time measurements will lead to small biases that can be considered negligible (e.g., a bias of  $1/405 \approx 0.0025$  for the Working Memory Task of the HCP Healthy Adults data).

## 2.5 | Group level analysis

Group-level GLM analyses typically incorporate two-stages, in which second stage analysis is based on summary statistics from the first (Beckmann et al., 2003; Holmes & Friston, 1998; Woolrich et al., 2004). The methodology for group-level significance testing depends on the experimental design. *T*-tests can be used provided that the experiment is balanced (Holmes & Friston, 1998). For unbalanced data, if the variance components of the data are known, then principled group-level inference can be done using univariate parameter estimates and their covariance estimates (Beckmann et al., 2003). In most cases, these variance components are not known. Second-level variance parameter estimates have been found using the EM algorithm (Worsley et al., 2002), approximate Bayesian inference (Friston et al., 2002), fully Bayesian inference (Woolrich et al., 2004) and restricted maximum likelihood approaches (Chen, Saad, Nath, Beauchamp, & Cox, 2012). These same ideas extend to the VDGLM. For the simple balanced-experiment setting, group-level inference can be computed using *t*-tests. In the unbalanced case, more work is needed due to the difficulty in computing the covariance of variance parameters. We initially tried to develop group-level inference procedures using asymptotic statistics (Wald test), but these tests were ill-behaved for several subjects due to high-condition number matrix inversions. We leave development of alternative statistics and a fully Bayesian framework to future work.

Group effect sizes can be estimated using the set of parameter estimates from all subjects. In our application, we compute Cohen's *d*, which measures the standardized mean between two populations,

and is popular in fMRI for avoiding the multiple comparison problem inherent in significance testing.

## 2.6 | Model comparison

Model comparison also proceeds as in a traditional fMRI pipeline. Model comparison can be done using AIC (Akaike, 1974), BIC (Schwarz, 1978), or any other log-likelihood-based metric that is a function of a point estimate. Model comparisons can consist of traditional in-sample comparisons or can be generalized to new data using out-of-sample comparisons (Mosteller & Tukey, 1968). The outcomes of univariate comparisons can be aggregated into group level results that test whether a subject tends to prefer a certain model across the brain or whether a particular region tends to prefer a certain model across subjects.

## 3 | EXAMPLE APPLICATION: BV IN WORKING MEMORY

In this example application, we used the VDGLM to find brain regions that are involved in working memory via changes in BV. We examined whether these regions differ from regions involved via changes in mean BOLD activation, and tested whether the VDGLM better describes working memory data than the GLM. The goal is to illustrate how to use the VDGLM and to showcase its utility.

We used data from the Human Connectome Project (HCP) Working Memory Experiment. In the experiment, subjects alternated between fixation blocks and two different task blocks during which they were presented with sequences of visual stimuli. In a 2-back task block, subjects indicated whether the current stimulus was the same as the one two presentations ago. In a 0-back task block, subjects indicated when a target stimulus was presented.

GLM analyses of the HCP data have found that engagement in the 2-back working memory task invokes regions thought to be involved in a cognitive control network, including bilateral dorsal and ventral prefrontal cortex, dorsal parietal cortex, and dorsal anterior cingulate. Task engagement leads to a deactivation in the default mode network, namely in the medial prefrontal cortex, posterior cingulate, and the occipital parietal junction (Barch et al., 2013). Similar activation patterns are found even when comparing 2-back versus 0-back. A 24 study meta-analysis of *N*-back studies found consistent activation in frontal and parietal areas, namely bilateral and medial posterior parietal cortex, bilateral premotor cortex, dorsal cingulate/medial premotor cortex, bilateral rostral prefrontal cortex or frontal poles, bilateral dorsolateral prefrontal cortex, and bilateral mid-ventrolateral prefrontal cortex or frontal operculum (Owen, McMillan, Laird, & Bullmore, 2005).

In our VDGLM analysis, the goal was to find both mean effects that overlap with known mean effects and also variance effects that could be spatially orthogonal to known mean effects.

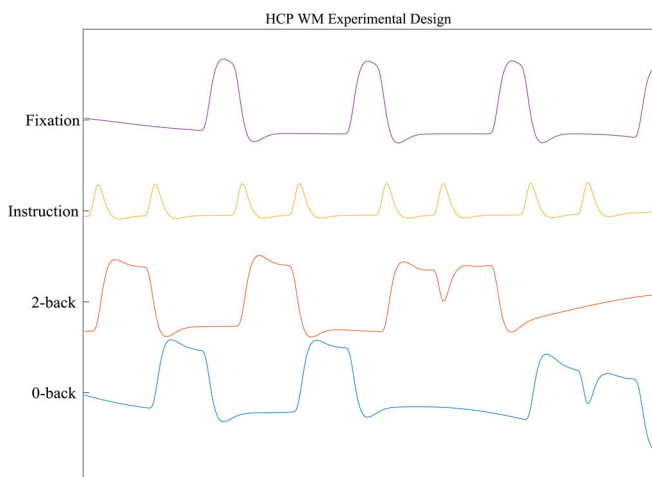
### 3.1 | Methods

#### 3.1.1 | Data acquisition and preprocessing

The data was collected by the Washington University—University of Minnesota Consortium Human Connectome Project (HCP, Van Essen et al., 2013). We used the Working Memory task data from the 1,200 Subjects release using the minimal preprocessing pipeline (Glasser et al., 2013). Details of task fMRI processing can be found in (Barch et al., 2013). We included subjects that performed the left-to-right phase encoded Working Memory task, resulting in 875/1200 total subjects for analysis. The downloaded data were in grayordinate system (Glasser et al., 2013), and the time series for 333 surface regions of interest (ROIs) based on Gordon et al. (2014) were extracted for further analysis. We perform additional preprocessing including scrubbing (relative RMS greater than 0.5 mm) and regression of motion estimates to minimize motion artifacts (Burgess et al., 2016).

#### 3.1.2 | Task design

During the Working Memory experiment, subjects alternatively engaged in a 0-back and 2-back tasks that use faces, places, tools, and body parts as the four categories of stimuli. Within each run, subjects were presented with blocks of stimuli, where all stimuli within a block were from the same category. For half of the blocks, subjects were given a “target” stimulus and were instructed to press a button whenever that stimulus was presented (0-back task). For the other half of blocks, subjects were instructed to respond when the stimulus was the same as the one presented two presentations ago (2-back task). Task blocks were interwoven with 15 s fixation blocks and instruction cues indicating the task type and “target” stimulus if the task was the 0-back task. Each run contained eight task blocks. We combined blocks from each stimuli type to create two task indicators (one for 0-back and one for 2-back). In total, the experimental design contained four conditions: the 0-back task, the 2-back task, fixation, and instruction (see Figure 3).



**FIGURE 3** HCP working memory experimental design convolved with the canonical HRF. Condition is indicated on the y-axis. HCP, Human Connectome Project; HRF, hemodynamic response function [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

#### 3.1.3 | Modeling

We applied the VDGLM model (Equation 4) to the data. Building the VDGLM required specifying both the mean and variance design matrix. In the mean design, we included the 0-back, 2-back, Fixation, and instruction conditions. In the variance design, we included the same regressors as in the mean design, but with an additional intercept regressor to reflect the assumption that there exists some measurement noise not captured by the other variance regressors. Additionally, we added a mean frame displacement noise regressor to the mean design.

We also fit a GLM (Equation 3) model to the data using a design matrix that is equivalent to the mean design matrix used in the VDGLM.

#### 3.1.4 | Estimation

We estimated GLM parameters using ordinary least squares. Since the VDGLM is analytically intractable, we estimated parameters using constrained trust-region optimization (Moré & Sorensen, 1983; see Appendix A for optimization details, and (Yuan, 2000) for a review of trust-region optimization). We performed mass univariate estimation, that is, we fit the VDGLM and the GLM for each ROI and subject. From parameter estimates, we created parameter contrasts for the 2-back minus Fixation, 0-back minus Fixation, and 2-back minus 0-back conditions.

#### 3.1.5 | Group level effect sizes

We estimated group-level effect sizes for each contrast using Cohen's *d* (the difference in standardized means) computed over subjects (Cohen, 2013). We visualized these effect sizes for each ROI using the HCP workbench software (Marcus et al., 2011). For a single region there exist three possible group-level effect patterns on BOLD: (a) both mean and variance effects are shown, (b) either mean or variance effects are shown, or (c) neither type of effect is shown. We plotted the whole-brain spatial distribution of each type of effect at small and medium effect sizes (Cohen's *d* of 0.2, 0.5, respectively). Additionally, we compare our VDGLM mean estimates to GLM mean estimates, to see whether modeling the variance changes known mean inferences.

#### 3.1.6 | Model comparison

Because the VDGLM has more parameters than the GLM, it has the potential to explain more variability of the observed data, thus any model comparison metric should take complexity into account. We achieved this using out-of-sample log likelihood (OOSLL), which penalizes overfitting by testing how well a model generalizes to new unseen data. We used an out-of-sample metric, rather than traditional metrics of model fit (such as goodness of fit tests, or information criteria) to balance the goals of our analysis between prediction and explanation (Yarkoni & Westfall, 2016); the model was compared to



other models using predictive performance, but also evaluated on its explanatory power. We used 10-fold cross validation to compute the out-of-sample log likelihood. For a single subject and region, we split the time series into 10 folds that each contains a test and training set. For each fold, we fit our model on the training set and computed the out-of-sample log likelihood of the test set given the parameters computed during training. To understand the level of general preference for the VDGLM, we compute the percent of subject/ROI time series with higher OOSLL for the VDGLM compared to the GLM. To understand subject VDGLM preference, we compute the percent of regions that prefer the VDGLM model for each subject.

To test that our model comparison results did not occur by chance, we compared the prevalence of VDGLM preference found in the real data to that found for a dataset simulated from the GLM (i.e., data without variance effects). In this comparison, we wanted to make explicit assurances that the VDGLM was preferred because of real effects in the data, that is, that preference was not due to autocorrelation artifacts. We did this by adding autocorrelation to the dataset simulated from the GLM, where we estimated the autocorrelation from the real data. We generated a time series for each subject and ROI independently. The generation of a sample time series from a real time series  $y_k$  proceeded as follows:

1. Compute the GLM OLS solution  $\hat{\beta}$  and variance solution  $\hat{\sigma}^2$ . We use GLM parameters estimated from the real data to better account for subject heterogeneity than if we simulated the underlying GLM parameters.
2. Estimate the autocorrelation of the residuals  $y_k - X\hat{\beta}$  using an AR(2) process and generate an estimated autocovariance matrix  $A$ .
3. Generate a sample time series  $y_{\text{samp}} = X\hat{\beta} + \epsilon_{\text{samp}}$ , where  $\epsilon_{\text{samp}} \sim \text{MVN}(0, \hat{\sigma}^2 A)$ .

For each subject and region, we generate a single sample time series (we generate only one sample to reduce the computational complexity of this test). Using the generated dataset, we fit the VDGLM to each subject and region. Due to having just one sample from each subject and ROI, we cannot make statements about whether model comparisons for a single subject and ROI are due to chance. However, we can analyze the percent of ROIs for a given subject that prefer the VDGLM to assess whether the amount of subject-level preference is due to chance. We compute whether the subject-level VDGLM preference is greater in the actual data compared to the simulated data. This test shows whether VDGLM preference is caused by overfitting to autocorrelation or whether there is true variance-related signal in the data. Model fitting and model comparison for 875 subjects took  $\sim 1/2$  a day to fit on the UCI High Performance Cluster.

## 4 | RESULTS

The analyses we ran on the VDGLM were designed with three goals in mind. First, we wanted to test for the existence of effects on BOLD variation during working memory engagement. We did this by

computing effect sizes of variance parameter estimates. Second, we wanted to see whether these effects occur in regions that are spatially orthogonal to regions that exhibit mean effects. To do this, we visually examined whether mean and variance effect sizes are correlated and plotted whole-brain visualizations of where mean and variance effects occur. Finally, we wanted to verify that the VDGLM provides a better account of the data than the GLM using model comparison metrics based on out-of-sample log-likelihood.

### 4.1 | Group level effect sizes

The first goal in our analysis was to test for existence of variance effects caused by working memory engagement. We measured effects by computing Cohen's  $d$  over subjects. For each parameter contrast (2-back minus Fixation, 0-back minus Fixation, and 2-back minus 0-back) we plotted whole-brain Cohen's  $d$  (see Figure 4, bottom row). We also plotted Cohen's  $d$  for mean parameter contrasts (top row) to verify that the VDGLM preserves known mean effects.

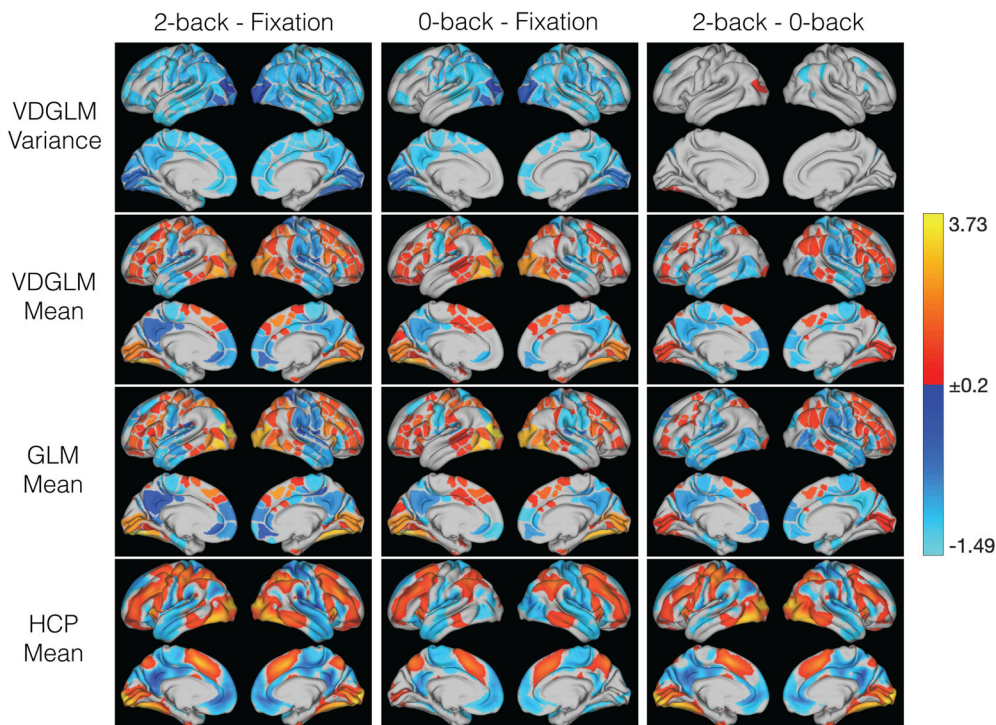
We found small and medium sized variance effects during the 2-back and 0-back tasks compared to fixation across much of the entire brain. Both 2-back and 0-back engagement evoked less BOLD variation (i.e., negative Cohen's  $d$ ) compared to Fixation across the whole brain. For the 2-back minus 0-back contrast, variance Cohen's  $d$  was low for some areas in the default mode network, some areas of the dorsal attention network, some parts of visual cortex, and some parts of the fronto-parietal network.

The VDGLM found mean effects that overlap existing HCP GLM results (Barch et al., 2013; see Figure 4). The 2-back minus Fixation and 0-back minus Fixation contrasts showed activation in the bilateral frontal-parietal network, bilateral visual cortex, and deactivation in the default mode network, including medial prefrontal cortex, posterior cingulate, and the occipital parietal junction. These same regions were activated, but less intensely for the 2-back minus 0-back contrast.

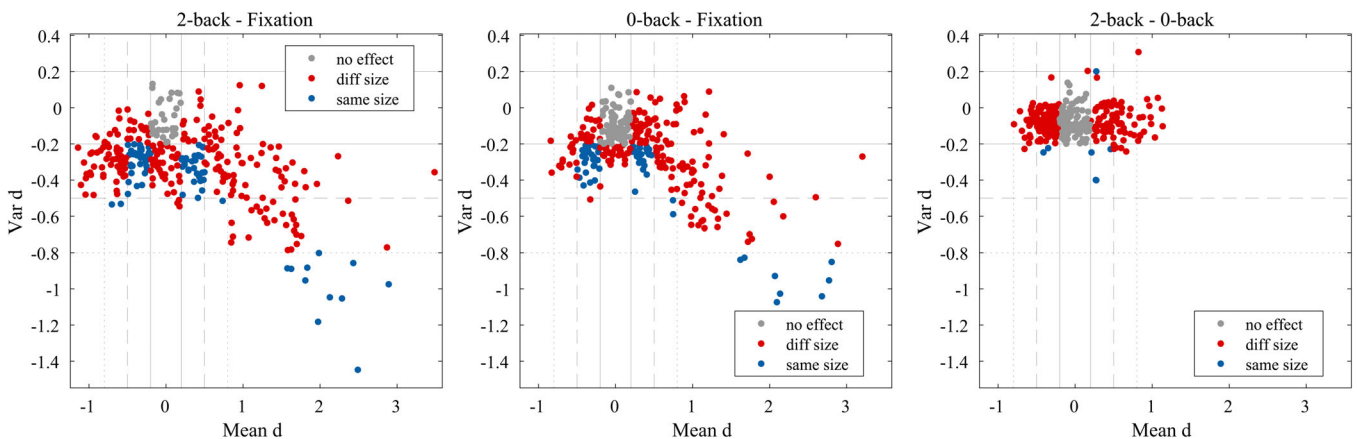
In general, task engagement leads to both positive and negative mean effects sizes, but predominantly negative variance effects sizes.

### 4.2 | Orthogonality of mean and variance effects

The second goal in our analysis was to examine whether the VDGLM finds variance effects that are orthogonal to known mean effects. To analyze the degree of orthogonality between mean versus variance effects, we plotted regional Cohen's  $d$  for the mean contrasts versus variance contrasts (see Figure 5). We grouped ROIs by effect size. ROIs with the same effect size in the mean and variance are plotted in red, those with difference effect sizes in the mean and variance are plotted in blue, ROIs with no mean nor variance effects are plotted in gray. The small, medium, and large, effect thresholds are plotted by the solid, dashed, and dotted black lines, respectively. ROIs exhibited mean and variance effects that span all possible combinations of effect sizes, although there were no large variance effects for the 0-back minus Fixation nor 2-back minus 0-back contrasts (see Figure 5). In general, mean effects were larger than variance effects. Effects were also much larger for the 2-back minus Fixation and



**FIGURE 4** Group-wise Cohen's  $d$  for the 2-back minus fixation, 0-back minus fixation, and 2-back minus 0-back contrasts. The top row shows VDGLM variance effects, the middle row shows VDGLM mean effects, and the bottom row shows voxel-wise GLM results from previous analysis. Maps are thresholded at  $(-0.2, 0.2)$ . VDGLM, variance design general linear model [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



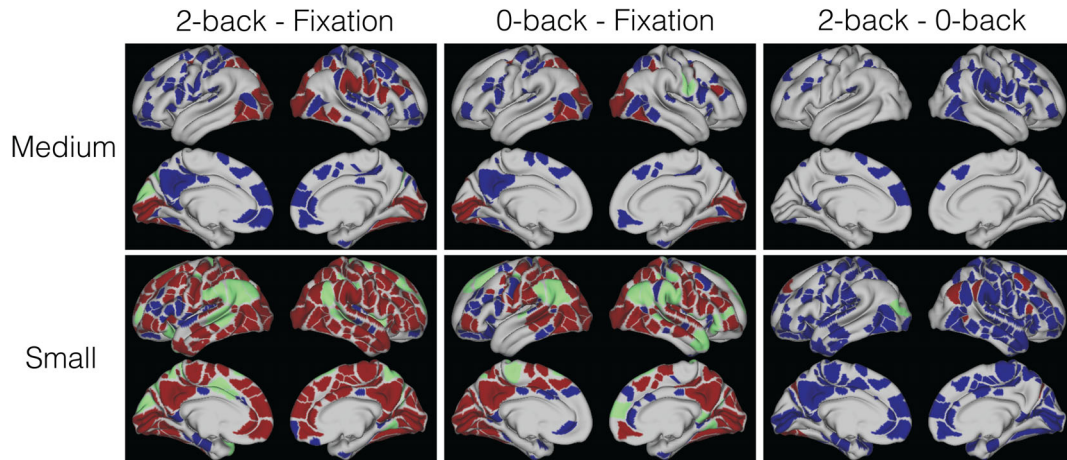
**FIGURE 5** The magnitude of mean and variance effects sizes for the 2-back minus fixation, 0-back minus fixation, and 2-back minus 0-back contrasts. Each circle represents an ROI. ROIs are grouped by whether they exhibit the same size effect (blue) in the mean and variance or different sized effects (red). The black lines indicate the small (solid), medium (dashed), and large (dotted) effect sizes. ROI, region of interest [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

0-back minus Fixation contrasts compared to the 2-back minus 0-back contrast. While there was a negative correlation between mean Cohen's  $d$  and variance Cohen's  $d$  for the 2-back minus Fixation and 0-back minus Fixation contrasts ( $R^2 = .265$  and  $R^2 = .357$ , respectively), the other contrast was uncorrelated ( $R^2 = .000$ ). Hence, mean and variance effects are orthogonal for the 2-back minus 0-back contrast.

### 4.3 | Spatial distribution of effects

In order to understand where the mean and/or variance effects were occurring, we plotted the type of small and medium effects that occur

in each region (see Figure 6). A region with mean effect only is plotted in blue, variance effect only in green, and both effects in red. For the 2-back minus Fixation contrast, there are regions that exhibited all types of effects. Variance only effects occurred primarily in the default mode network, but also in sensorimotor mouth regions, and regions in the visual, dorsolateral attention, and cingulo-opercular networks. Mean only effects occurred in the frontoparietal, the auditory, cingulo-opercular, visual, and dorsolateral attention networks. Effects overlapped in some regions of the default mode, frontoparietal, visual, dorsal attention, and cingulo-opercular networks and in some sensorimotor regions. For the 0-back minus Fixation contrast, less regions exhibited medium variance effects than the 2-back minus Fixation



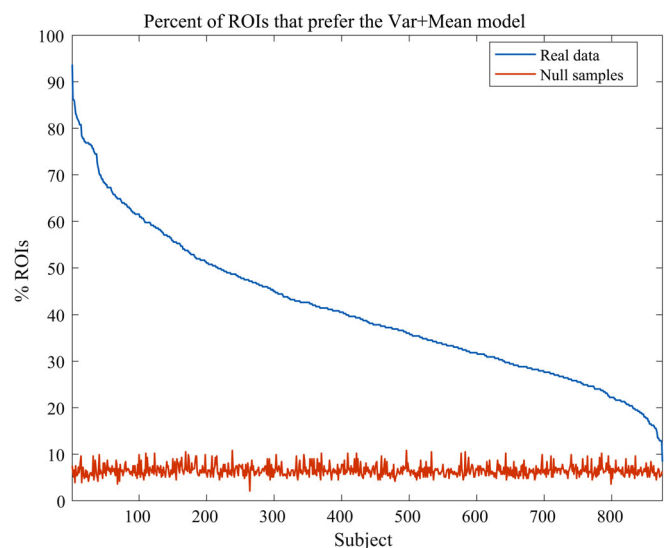
**FIGURE 6** The figure shows which types of effects occur in which regions. Regions can have only a mean effect (blue), only a variance effect (green), or both effects (red). We plot effects for the 2-back minus fixation, 0-back minus fixation, and 2-back minus 0-back contrasts. We plot small (Cohen's  $d \in [-0.2, 0.2]$ ) and medium (Cohen's  $d \in [-0.5, 0.5]$ ) effects [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

contrast, suggesting that cognitive demand plays a role in the size of variance effects. There were, however, effects in sensorimotor areas used in control of the mouth and hand, a region in visual cortex, and some regions in the dorsolateral attention network. These effects were all present in the 2-back minus Fixation contrast as well. Mean 0-back minus Fixation effects also largely mirror the mean effects in the 2-back minus Fixation tasks. There were no medium sized variance 2-back minus 0-back contrasts. This suggests that while there were different sized effects between 2-back minus Fixation and 0-back minus Fixation, these differences in effect sizes were small. The main differences in variance effects tended to be caused by task engagement rather than cognitive load.

#### 4.4 | Model comparison

The VDGLM inferred that engagement in a working memory task leads to less BOLD variation. In this section, we pursue a different question. Does a model with these additional parameters give a significantly better account of the BOLD time series than a model without them? To test this, we perform model comparisons between the VDGLM and the GLM. Since, the VDGLM has more parameters and will trivially better fit the data, we use out-of-sample log-likelihood to check the VDGLM's ability to describe new unseen data. We perform a model comparison for each subject and ROI time series in a mass univariate approach.

We found significant preference for the VDGLM model; 41% of subjects/ROIs had higher OOSLL for the VDGLM model over the GLM in the real data (7% in the simulated data). Model preference varies by subject and region (see Figure 7). The figure plots the percentage of ROIs that prefer each model in the real data (blue) and the simulated data (red). The figure is ordered by a subject's proportion of ROIs that prefer the VDGLM model in the real data. A subject's percentage of regions that favored the VDGLM ranged from 14% for GLM-leaning subjects to 73% for VDGLM-leaning subjects. For all 875 subjects, some number of regions (but not all) preferred the



**FIGURE 7** The percent of regions for each subject for which the VDGLM model better describes the HCP data (blue) and data simulated from the mean model (red). Subjects are ordered by percent of regions for which the VDGLM model has higher OOSLL. OOSLL, out-of-sample log likelihood (OOSLL); VDGLM, variance design general linear model [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

VDGLM model. Similarly, for all 333 ROIs, some number of subjects (but not all) preferred the VDGLM model.

To check that the VDGLM was not just fitting autocorrelation, we performed model comparisons for models fitted to data generated from the GLM with added autocorrelation. For the simulated data, only 7% of subjects/ROIs preferred the VDGLM model. For every subject, the percent of ROIs that preferred the VDGLM model was larger in the real data than in the simulated data, indicating a significant preference for the VDGLM that was not just due to fitting to autocorrelation.



## 4.5 | Application summary

The working memory application highlights the usage of the VDGLM. We used the model to show that working memory task engagement was related to a decrease in variance compared to fixation. Variance effects and mean effects were not spatially correlated, suggesting that the VDGLM reveals distinct brain patterns not captured by mean-based approaches. We also found that the spatial distribution of mean effects was similar to the spatial distribution of voxel-wise mean effects from previous analyses (Barch et al., 2013). Variance tended to be reduced across the whole brain compared to baseline. We want to highlight that while many of the variance effect sizes in this application were fairly small, this will not necessarily be true in future applications. Importantly, if there exists some quantity of interest that consistently relates to small variance effect, then these small effects are worth studying. This is especially true in disease studies where discoveries have the potential to impact human lives (D'Esposito, Deouell, & Gazzaley, 2003; Rombouts, Barkhof, Goekoop, Stam, & Scheltens, 2005).

The application focused on testing for effects of BV, so we designed our application to reduce potential confounds. By using HCP data we minimized the effects of noise from CSF, large veins, and white matter; high resolution data collection (2 mm) and registration to the cortical surface leads to less voxel-by-voxel overlap with these noise sources than compared to other data sources (Glasser et al., 2013). We corrected for head motion by including nuisance motion regressors and scrubbing particularly noisy volumes. We did not account for heart beat nor respiration, which are known to affect resting state BOLD variability (Biswal, Deyoe, & Hyde, 1996; Kannurpatti & Biswal, 2008; Kannurpatti, Motes, Rypma, & Biswal, 2010; Kannurpatti, Motes, Rypma, & Biswal, 2011). However, we tested at the group-level whether mean within-block heart rate and respiration were correlated with mean within-block activation. We found some significant correlations although the mean size of correlations were small (between  $-0.088$  and  $0.112$  for heart rate,  $-0.061$  and  $0.25$  for respiration), which suggests that physiological noise may be a driver of the variance effects we demonstrated.

Since neither heart rate nor respiration are highly correlated with the task design, presence of these sources of noise increases variance during task. Thus, we suspect that correcting for physiological noise in future studies would lead to larger variance effect sizes. We performed several posthoc analyses to check that VDGLM preference was not related to mean frame displacement, nor grand mean intensity scaling factor (Turner, Lopez, Santander, & Miller, 2015) ( $\text{adj}R^2 = -0.00114, -0.0011$ , respectively).

In this application, we fit a single VDGLM model and a single corresponding GLM model. In practice, we could fit several VDGLM models to test hypotheses of the form: "should condition C be included in our model and does it affect the mean or the variance in BOLD activation." In this set-up, each model takes the form of Equation (4) and the conditions to be tested are defined by the entries of the mean and variance design matrices. For example, we could test a model with only intercept effects versus a model that allows each

condition to affect the variance, but not the mean. Then model comparison indicates which experimental conditions are necessary in the model and whether those conditions are necessary as mean or variance regressors. This formulation allows us to define several nested models in the classical sense—that is, that the set of mean regressors in the nested model is a subset of the regressors of the full model—or in a novel way where a combination mean and variance regressors are nested.

To effectively develop and test the VDGLM, we chose an ROI approach to have more reliable BOLD signal and a lower computational load. Application of the VDGLM to voxel-wise analyses is left to future work.

## 5 | DISCUSSION

Traditional fMRI analyses treat BOLD variation as a "nuisance parameter" despite results linking BOLD variation to age, behavioral performance, and task engagement. The VDGLM fills this gap by providing a flexible framework for linking variance effects to experimental design. By directly incorporating the design matrix, the VDGLM can assess the independent contributions to BOLD variance from multiple experimental conditions while controlling for confounding factors. The VDGLM also controls for confounding between mean and variance effects; since both effects are modeled simultaneously, we can make inferences about one effect while controlling for the other. The VDGLM is fit in a mass univariate approach, which allows analysis at a more fine-grained resolution than previous empirical studies that analyzed latent structures of large spatial patterns in BV (Garrett et al., 2010, 2011; Garrett, Samanez-Larkin, et al., 2013). Under the VDGLM framework hypothesis generation and comparison is easy; each hypothesis corresponds to an instantiation of the model and can be tested using model comparison.

In our application, we showed that the VDGLM can be used to find variance effects caused by working memory engagement (Figure 4). We showed that the working memory effects (2-back vs. 0-back) are spatially orthogonal to mean effects (Figures 5 and 6) and finally, we compared the GLM and VDGLM and showed that VDGLM provides a better description of the data even while accounting for model complexity (Figure 7).

An important feature the VDGLM is the facility for modeling mean and variance simultaneously while allowing for orthogonal spatial inferences. In the BV-age fMRI literature, variance effects were orthogonal to mean effects (Garrett et al., 2010, 2011). This trend generalized to our working memory application, where task engagement resulted in predominantly negative BV effects across the brain, but a mix of positive and negative mean effects. The negative BV effects we found in the working memory application of the VDGLM are different from (Garrett et al., 2010, 2013; Garrett, Samanez-Larkin, et al., 2013) in which the authors found increased BOLD variability associated with cognitive demand. One possible explanation is that different cognitive tasks were used in these studies, and it is possible that different tasks may lead to BV effects with different signs.

Prior to the development of the VDGLM, we analyzed the Working Memory data in the HCP dataset by simply computing the mean and variance in each task blocks. We found reduced BOLD variability associated with both the 2-back and 0-back tasks relative to fixation. The results were consistent with the VDGLM analysis. Although the sign of the BV effect may change depending on the task, these results constitute a growing body of evidence that BV is a novel dimension for studying brain function.

We want to highlight that there are alternatives to the methodological choices we made in our application. Alternative choices can be made regarding (a) the inference statistic or effect size estimate, (b) the model comparison metric, and (c) the method for assessing model comparison significance (see Figure 2). The comparison metric can easily be substituted for another model and metric, and the additional model comparison significance test, while powerful, is not necessary in most standard analyses. Using an alternative choice of inference statistic may require further work. Our choice to use effects size was motivated by the use of effect size in previous analyses (Van Essen et al., 2013) and the ease of using a statistic that depends solely on parameter estimates. In an effort to provide alternative inference statistics, we developed approximate *t*-tests for variance effects. However, we found that the estimates were sensitive to the condition number of the Hessian matrix specified by the VDGLM (which is required for computation of approximate *t*-tests). We tested the accuracy of the approximate *t*-tests for mean effects by comparing them to standard *t*-tests made by the GLM. While we found that while they were close for most subjects and ROIs, for other subjects with poorly conditioned Hessian matrices the *t*-tests tended to be unrealistically large. While any individual data point could be excluded from group-level analysis using condition number threshold, we found this approach too cumbersome for a framework aimed at general public use. Development of well-behaved statistics for inference is ongoing work.

Many different measures of BV have been used in past fMRI studies: empirical variance (He, 2011), parametric variance (Wutte et al., 2011), block-normalized standard deviation (Garrett et al., 2010; Garrett, Kovacevic, McIntosh, & Grady, 2013), and mean squared successive difference (Leo et al., 2012; Samanez-Larkin, Kohnen, Yoo, & Knutson, 2010). The goals of mean squared successive difference and block-normalized standard deviation are to measure the variance not accounted for by mean trends in the data. Since the VDGLM models the variance/standard deviation in BOLD activation after accounting for the mean trend, its variance parameters can be conceptualized to measure a construct similar to mean squared successive difference or block-normalized variance (but where blocks are convolved with the canonical HRF). The parameterized model in Wutte et al., 2011 is similar to the VDGLM, but uses a mixing parameter to capture shared variance between task and fixation blocks rather than modeling the variance as a function of convolved experimental design. We expect that this approach leads to similar results, but with the caveat that it only incorporates a single experimental condition. Lastly, we consider the interquartile range, which is not used in the VDGLM and to our knowledge has not been used in fMRI analysis to date. The goal of the interquartile-range is to summarize the dispersion while limiting the effects of any highly outlying time points. In our application, we used

scrubbing to a similar effect by manually removing any outlying time points and recommend this approach if large outliers are present.

The VDGLM could be improved by implementing it in a Bayesian framework. Bayesian frameworks would allow us to make more robust inferences, incorporate prior beliefs about regions likely (or unlikely) to exhibit BV effects, and to better quantify model comparisons. The main disadvantage of Bayesian methods, and the reason we did not develop a Bayesian VDGLM, is the computational complexity of inference.

The VDGLM could also be improved by transforming the variance so that we did not need to enforce positivity. Log transformations have been widely used for covariance and variance estimation (Pourahmadi, 2011), however in the case of the VDGLM lead to a drastic conceptual change in the model. Since the VDGLM incorporates the design matrix into its variance formulation, the exponential transformation results in a variance parameters that are raised to the power of elements of the design matrix. We opted to keep the VDGLM as an additive variance model that requires constraints rather than as a multiplicative variance model so that parameters were more interpretable. However, we expect that the study of variance transformations could lead to stronger inferences in future work.

## 6 | CONCLUSION

Studies have demonstrably shown that variance in BOLD activation is a functional construct orthogonal to mean BOLD that should be taken into account in future imaging analyses.

This work developed the VDGLM, a coherent statistical framework for incorporating BV into standard fMRI analyses. The VDGLM was motivated by strong evidence that variance in BOLD activation is linked to individuals and behavior. The VDGLM can be easily applied in any experimental setting and will allow for increased ease and flexibility in research on BOLD variability. We expect that it will lead to exciting new discoveries relating BOLD variability to human characteristics and behavior.

## ACKNOWLEDGMENTS

This work was supported by a National Sciences Foundation Integrative Strategies for Understanding Neural and Cognitive Systems Collaborative Research Grant (1533500 and 1533661).

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

## Data Availability Statement

The preprocessed version of our data is available through the Open Science Framework website: (<https://osf.io/4rvbz/>). MATLAB code to run the VDGLM model can be found on GitHub (<https://github.com/geebioso/VDGLM>).

## ORCID

Garren Gaut  <https://orcid.org/0000-0001-5661-6478>

Mark Steyvers  <https://orcid.org/0000-0003-1466-5647>

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., ... WU-Minn HCP Consortium. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 80, 169–189.
- Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in fMRI. *NeuroImage*, 20(2), 1052–1063.
- Berger, H. (1929). Über das elektroencephalogramm des menschen. *Archiv für Psychiatrie Und Nervenkrankheiten*, 87(1), 527–570.
- Biswal, B., Deyoe, E. A., & Hyde, J. S. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, 35(1), 107–113.
- Bullmore, E., Brammer, M., Williams, S. C., Rabe-Hesketh, S., Janot, N., David, A., ... Sham, P. (1996). Statistical methods of estimation and inference for functional MRI image analysis. *Magnetic Resonance in Medicine*, 35(2), 261–277.
- Burgess, G. C., Kandala, S., Nolan, D., Laumann, T. O., Power, J. D., Adeyemo, B., ... Barch, D. M. (2016). Evaluation of denoising strategies to address motion-correlated artifacts in resting-state functional magnetic resonance imaging data from the human connectome project. *Brain Connectivity*, 6(9), 669–680.
- Byrd, R. H., Schnabel, R. B., & Shultz, G. A. (1988). Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming*, 40(1), 247–263.
- Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., & Cox, R. W. (2012). fMRI group analysis combining effect estimates and their variances. *NeuroImage*, 60(1), 747–765.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 9(4), 673–690.
- D'Esposito, M., Deouell, L. Y., & Gazzaley, A. (2003). Alterations in the bold fMRI signal with ageing and disease: A challenge for neuroimaging. *Nature Reviews Neuroscience*, 4(11), 863–872.
- Duffy, F. H., Albert, M. S., McAnulty, G., & Garvey, A. J. (1984). Age-related differences in brain electrical activity of healthy subjects. *Annals of Neurology*, 16(4), 430–438.
- Eklund, A., Lindquist, M. A., & Villani, M. (2017). A bayesian heteroscedastic GLM with application to fMRI data with motion spikes. *NeuroImage*, 155, 354–369.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.
- Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, 1(2), 153–171.
- Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. (2002). Classical and bayesian inference in neuroimaging: Theory. *NeuroImage*, 16(2), 465–483.
- Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2010). Blood oxygen level-dependent signal variability is more than just noise. *Journal of Neuroscience*, 30(14), 4914–4921.
- Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2011). The importance of being variable. *Journal of Neuroscience*, 31(12), 4496–4503.
- Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2013). The modulation of bold variability between cognitive states varies by age and processing speed. *Cerebral Cortex*, 23(3), 684–693.
- Garrett, D. D., Lindenberger, U., Hoge, R. D., & Gauthier, C. J. (2017). Age differences in brain signal variability are robust to multiple vascular controls. *Scientific Reports*, 7(1), 10149.
- Garrett, D. D., McIntosh, A. R., & Grady, C. L. (2013). Brain signal variability is parametrically modifiable. *Cerebral Cortex*, 24(11), 2931–2940.
- Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W., Lindenberger, U., McIntosh, A. R., & Grady, C. L. (2013). Moment-to-moment brain signal variability: A next frontier in human brain mapping? *Neuroscience & Biobehavioral Reviews*, 37(4), 610–624.
- Gaut, G., Li, X., Turner, B., Cunningham, W. A., Lu, Z.-L., & Steyvers, M. (2018). Predicting task and subject differences with functional connectivity and bold variability. *Brain Connectivity*.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... WU-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80, 105–124.
- Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2014). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*, 26(1), 288–303.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358), 320–338.
- He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *The Journal of Neuroscience*, 31(39), 13786–13795.
- Holmes, A., & Friston, K. (1998). Generalisability, random effects & population inference. *NeuroImage*, 7, S754.
- Kannurpatti, S. S., & Biswal, B. B. (2008). Detection and scaling of task-induced fMRI-bold response using resting state fluctuations. *NeuroImage*, 40, 1567–1574.
- Kannurpatti, S. S., Motes, M. A., Rypma, B., & Biswal, B. B. (2010). Neural and vascular variability and the fMRI-bold response in normal aging. *Magnetic Resonance Imaging*, 28(4), 466–476.
- Kannurpatti, S. S., Motes, M. A., Rypma, B., & Biswal, B. B. (2011). Increasing measurement accuracy of age-related bold signal change: Minimizing vascular contributions by resting-state-fluctuation-of-amplitude scaling. *Human Brain Mapping*, 32(7), 1125–1140.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2–3), 169–195.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12), 606–617.
- Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations. *Experimental Psychology*, 52(2), 99–108.
- Leo, A., Bernardi, G., Handjaras, G., Bonino, D., Ricciardi, E., & Pietrini, P. (2012). Increased bold variability in the parietal cortex and enhanced parieto-occipital connectivity during tactile perception in congenitally blind individuals. *Neural Plasticity*, 2012, 1–8.
- Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., ... Van Essen, D. C. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics*, 5, 4.
- Moré, J. J., & Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3), 553–572.
- Mosteller, F. and Tukey, J. W. (1968). Data analysis, including statistics. *Handbook of Social Psychology* (Vol. 2, pp. 80–203).
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46–59.

- Pourahmadi, M. (2011). Covariance estimation: The glm and regularization perspectives. *Statistical Science*, 26, 369–387.
- Rombouts, S. A., Barkhof, F., Goekoop, R., Stam, C. J., & Scheltens, P. (2005). Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: An fmri study. *Human Brain Mapping*, 26(4), 231–239.
- Samanez-Larkin, G. R., Kuhnen, C. M., Yoo, D. J., & Knutson, B. (2010). Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. *Journal of Neuroscience*, 30(4), 1426–1434.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Somsen, R. J., van't Klooster, B. J., van der Molen, M. W., van Leeuwen, H. M., & Licht, R. (1997). Growth spurts in brain maturation during middle childhood as indexed by eeg power spectra. *Biological Psychology*, 44(3), 187–209.
- Turner, B. O., Lopez, B., Santander, T., & Miller, M. B. (2015). One dataset, many conclusions: Bold variability's complicated relationships with age and motion artifacts. *Brain Imaging and Behavior*, 9(1), 115–127.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The wu-minn human connectome project: An overview. *NeuroImage*, 80, 62–79.
- Williamson, S., Kaufman, L., Lu, Z.-L., Wang, J.-Z., & Karron, D. (1997). Study of human occipital alpha rhythm: The alphon hypothesis and alpha suppression. *International Journal of Psychophysiology*, 26(1–3), 63–76.
- Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for fmri group analysis using Bayesian inference. *NeuroImage*, 21(4), 1732–1747.
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, 14(6), 1370–1386.
- Worsley, K. J., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., & Evans, A. (2002). A general statistical analysis for fmri data. *NeuroImage*, 15(1), 1–15.
- Wutte, M. G., Smith, M. T., Flanagan, V. L., & Wolbers, T. (2011). Physiological signal variability in hmt+ reflects performance on a direction discrimination task. *Frontiers in Psychology*, 2, 185.
- Yarkoni, T. and Westfall, J. (2016). Choosing prediction over explanation in psychology: Lessons from machine learning. Unpublished manuscript. Retrieved from [http://jakewestfall.org/publications/Yarkoni\\_Westfall\\_choosing\\_prediction.pdf](http://jakewestfall.org/publications/Yarkoni_Westfall_choosing_prediction.pdf).
- Yuan, Y. (2000). A review of trust region algorithms for optimization. *ICIAM*, 99, 271–282.

**How to cite this article:** Gaut G, Li X, Lu Z, Steyvers M. Experimental design modulates variance in BOLD activation: The variance design general linear model. *Hum Brain Mapp*. 2019;40:3918–3929. <https://doi.org/10.1002/hbm.24677>

## APPENDIX A: VDGLM OPTIMIZATION

We perform maximum likelihood estimation using trust-region optimization (TRO). TRO is an iterative procedure for minimization. At each iteration, TRO locally approximates the negative log likelihood function using a Taylor expansion and finds a minimum within that step's

trust-region, that is, the region for which the local approximation accurately approximates the objective function. We used a built-in function in MATLAB that restricts the local approximation to a two-dimensional subspace to allow for faster convergence. The algorithm locally minimizes along the two-dimensional subspace spanned by the direction of steepest descent and one of either (a) the approximate newton direction, if it exists, or (b) the direction of negative curvature (Byrd, Schnabel, & Shultz, 1988). For a time series  $y$  from a single subject and region, we minimize the negative log likelihood  $-\log p(y|\theta)$  where the likelihood is defined:

$$p(y, \theta) = \left[ \prod_{t=1}^T p(y_t | \beta, v, X_t^M, X_t^V) \right]$$

where  $X_t^M$  is a  $[1 \times p_M]$  vector: the single row of the mean design matrix at time  $t$ .  $X_t^V$  is a  $[1 \times p_V]$  vector: the single row of the variance design matrix at time  $t$ . For single point in the time series, the likelihood is

$$p(y_t | \beta, v, X_t^M, X_t^V) = \frac{1}{\sqrt{2\pi X_t^V v}} \exp\left(-\frac{1}{2X_t^V v} (y_t - X_t^M \beta)^2\right).$$

and we can compute the joint log-likelihood as the product of the log-likelihoods for each point:

$$\log p(y | \theta) = -\frac{T}{2} \log 2\pi + \left[ \sum_{t=1}^T -\frac{1}{2} \log(X_t^V v) - \frac{1}{2X_t^V v} (y_t - (X_t^M \beta))^2 \right]$$

subject to the inequality constraint that the variance is nonzero, that is:

$$X_t^V v > 0 \quad \forall t \in \{1, \dots, T\}$$

This constraint can be conceptualized in a Bayesian setting as a uniform prior over the constrained area. Each iteration of the trust-region algorithm uses a Newton–Raphson step to update. We supply the analytical gradients:

$$\begin{aligned} \frac{\partial \log p(\beta | \bullet)}{\partial \beta} &= (X_t^M)^T \left[ \sum_{t=1}^T \frac{(y_t - X_t^M \beta)}{X_t^V v} \right] \\ \frac{\partial \log p(v | \bullet)}{\partial v} &= (X_t^V)^T \left[ \sum_{t=1}^T \frac{-1}{2X_t^V v} + \frac{(y_t - X_t^M \beta)^2}{2(X_t^V v)^2} \right]. \end{aligned}$$

We stop the optimization routine when the magnitude of the gradient is smaller than  $1e-6$ , the change in objective value is smaller than  $1e-6$ , the size of the trust region is below  $1e-6$ , or the optimization routine reaches 1,000 iterations.