# SCIENTIFIC REP♦RTS

# Multiple Kernel Learning Model for Relating Structural and Functional Connectivity in the Brain

Sriniwas Govinda Surampudi[1], Shruti Naik[2], Raju Bapi Surampudi[2,3], Viktor K. Jirsa[4], Avinash Sharma[1] & Dipanjan Roy[5]

A challenging problem in cognitive neuroscience is to relate the structural connectivity (SC) to the functional connectivity (FC) to better understand how large-scale network dynamics underlying human cognition emerges from the relatively fixed SC architecture. Recent modeling attempts point to the possibility of a single diffusion kernel giving a good estimate of the FC. We highlight the shortcomings of the single-diffusion-kernel model (SDK) and propose a multi-scale diffusion scheme. Our multi-scale model is formulated as a reaction-diffusion system giving rise to spatio-temporal patterns on a fixed topology. We hypothesize the presence of inter-regional co-activations (latent parameters) that combine diffusion kernels at multiple scales to characterize how FC could arise from SC. We formulated a multiple kernel learning (MKL) scheme to estimate the latent parameters from training data. Our model is analytically tractable and complex enough to capture the details of the underlying biological phenomena. The parameters learned by the MKL model lead to highly accurate predictions of subject-specific FCs from test datasets at a rate of 71%, surpassing the performance of the existing linear and non-linear models. We provide an example of how these latent parameters could be used to characterize age-specific reorganization in the brain structure and function.

Over the last two decades, investigation of the slow correlated fluctuations in the resting state functional magnetic resonance images (rs-fMRI) of the brain have yielded valuable insights about the spontaneous functional organization of the brain. The structural connectivity (SC) derived from diffusion tensor imaging (DTI) reveals the white-matter fiber connections between regions of interest (ROIs)[1,2]. On the other hand, functional connectivity (FC) estimates the linear statistical dependency between ROIs based on blood oxygen level dependent (BOLD) activation[3]. It has been observed that BOLD signals are non-stationary in nature[4–6]. This non-stationarity is understood in terms of states (measured in small time windows) and their transitions and are expressed as functional connectivity dynamics (FCD)[7]. On the other hand, static FC enables investigation of the analytical properties of network behavior[8].

Even in resting conditions BOLD activations of ROIs have been observed to form self-organizing patterns called resting state networks (RSNs). FC can be interpreted as a superposition of all these RSNs. Relating structural connectivity (SC) of the brain to its functional connectivity (FC) is a fundamental goal in neuroscience because it enables our understanding of how the relatively fixed SC architecture underlies human cognition and diverse behaviors[9]. It has been observed that regions with direct structural link are also functionally connected. However the converse is not necessarily true[10]. The question of how SC shapes FC has been the object of computational modeling but remains an open question[11]. In the recent years, connectivity analysis using whole brain computational models and graph theoretic techniques have given unprecedented insights about brain-wide correlations during rest and task conditions[12–14]. Computational models are designed to expand our understanding and explaining the functioning of the brain. The more biologically real the model is, the more computationally expensive it is. Hence, gaining analytical insights becomes increasingly difficult with complex models.

In the realm of noise-induced correlated deviations, there are linear and non-linear mean field models that attempt to answer this open question incorporating various kinds of dynamics and biological details[15–18].

[1]CVIT, IIIT-Hyderabad, Hyderabad, 500032, India. [2]Cognitive Science Lab, IIIT-Hyderabad, Hyderabad, 500032, India. [3]School of Computer and Information Sciences, University of Hyderabad, Hyderabad, 500046, India. [4]Aix Marseille Univ, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France. [5]Cognitive Brain Dynamics Lab, National Brain Research Centre, Manesar, Gurgaon, Haryana, 122 051, India. Correspondence and requests for materials should be addressed to A.S. (email: asharma@iiit.ac.in) or D.R. (email: dipanjan.nbrc@gov.in)

A biophysical attempt to relate SC to FC is a linear model based on graph diffusion of brain dynamics is outlined in[18]. This linear diffusion model considers brain dynamics, the diffusing quantity, firing rate of the neuronal population, undergoing random walk on the SC graph. This linear diffusion model considers that the mean regional activity diffuses over the anatomical fibers governed by a deterministic linear differential equation[18]. The analytically tractable solution becomes the graph diffusion kernel which is hypothesized to resemble the FC. This model fixes one global parameter across all subjects. Another model proposed by Saggio et al.[19] considers a linear auto-regressive model with additive Gaussian white noise, coupling matrix being SC. This model becomes a linear system of coupled stochastic first order differential equations in which the BOLD activities diffuse on the anatomical constraints, i.e. SC. This model computes covariance between regional activities whose analytical expression works out to be a function of SC. Such a model would find it difficult to account for inter-subject variability in the functional expression. Extending the idea of linearity to super-critical bifurcations and multi-stability, a series of non-linear stochastic models have been proposed that explain the underlying biological behavior efficiently[20–22]. These models differ in their representation scheme for the ROIs[23]. Whereas Kuramoto oscillator model[24] abstracts out the biophysical details, Deco and Jirsa's mean-field-models[20,21] consider dynamics of specific biological analogues such as mean firing rate and mean activity of the regions. These neural and meso-scopic models can be seen as variants of reaction-diffusion system at the heart of which lie the Wilson-Cowan equations[25,26]. Wilson-Cowan equations, a variant of reaction-diffusion systems, provide a coarse-grained description of the large-scale neuronal network in terms of oscillatory self-organizing patterns. New experimental evidence supports these equations[27].

Recently, a new paradigm of understanding the oscillatory patterns of cortico-cortical activity is proposed that utilizes spectral analysis of the connectome or structural connectivity (SC)[28]. It has been observed that these connectome-specific harmonics predict oscillatory functional networks of the human brain possibly through interplay of excitation; for instance mediated by the glutamatergic principal cells, and inhibition; for instance mediated by the GABAergic interneurons. The push-and-pull between diffusing excitatory cells and suppressing inhibitory cells can result in self-organizing pattern formation. The emergent harmonics or the standing waves are the allowed spatial frequencies, or the eigenfunctions of the graph Laplacian operator on the anatomically constrained SC largely determined by the selection of the diffusion parameters of excitation and inhibition. Surampudi et al.[29] observed that physical diffusion on large-scale graphs, i.e. SC, at multiple *diffusion scales* exhibits scale-dependent relationships among various regions of interest (ROIs). These multi-scale diffusion kernels are similarly motivated to capture reaction-diffusion systems operating on a fixed underlying connectome (SC) and hence can be interpreted as components of FC at different diffusion scales (see Supplementary Fig. S1 for the motivation of multi-scale approach). However, our investigations revealed that a combination of multiple diffusion kernels was not sufficient to explain the self-organizing resting state patterns found in FC and hence necessitates the need of additional explanatory parameters.

The extant whole brain computational models can be characterized along two dimensions–interpretability and complexity, where the linear and non-linear models lie at the opposite ends of the spectrum. The former are analytical models with few parameters that can be interpreted and tuned easily, whereas the latter are fairly complex models with richer dynamics but tend to be analytically intractable. The proposed model possesses the analytical beauty of linear models and yet is complex enough to capture the biological details. We hypothesize that the presence of regional multi-scale co-activations that initiate diffusion would be necessary to bridge the gap between structurally confined diffusion phenomenon and empirically observed FC and that these co-activations would be common across the cohort. We further provide a plausible mathematical reasoning for the existence of these co-activations along with diffusion kernels by linearizing a variant of reaction-diffusion model and extending it to generate FC. Moreover, we also describe a succinct multiple kernel learning (MKL) procedure to retrieve these co-activations by formulating it as an optimization formulation. MKL techniques are well explored in the machine learning community[30,31]. Our proposed model while retaining the parsimony of a simple linear approach, proposes a novel learning scheme for optimizing the best-fitting kernels for SC-to-FC mapping. Our detailed empirical results demonstrate the validity of the proposed model on a larger dataset.

## Results

**Model description.**     Our model is based on the hypothesis that there are inter-regional multi-scale co-activations (among ROIs) for initiating diffusion. These co-activations that are shared across all subjects enable our model to generate the observed FC patterns from the structurally confined diffusion kernels. This is achieved by first post-multiplying the co-activation parameters (denoted henceforth as $\pi_i'$s) with the respective diffusion kernel matrices derived from subject-specific structural matrix (SC) and then taking a linear combination of these for approximating/predicting the functional connectivity (FC). We propose to learn these co-activation parameters in a multiple kernel learning (MKL) framework, thus retaining the parsimony of a simple linear approach. In the experiments reported in this paper, 16 diffusion kernels were used and their respective inter-regional co-activation parameters (16 corresponding $\pi_i'$s) were learned (see Supplementary Figs S2–S4 along with their corresponding sections for scale selection methodology). To evaluate our model, we mainly used the data scanned at the Berlin Centre for Advanced Imaging, Charité University, Berlin, Germany that has SC-FC pairs of total 47 subjects (see Supplementary section 8 for data pre-processing methodology).

**Model performance.**     We compared performance of the proposed model (henceforth designated as MKL model) with two previously proposed approaches: single-diffusion-kernel (SDK) model of Abdelnour et al.[18] and the non-linear dynamic-mean-field (DMF) model described in Deco et al.[32] (see *Methods* for implementation details of these models). To remain consistent with the previous studies, we used Pearson correlation coefficient between empirical and predicted functional connectivities (FC) as the measure of model performance. To obtain

(**a**)   Performance of models.



(**b**)   Leave-one-out validation.



(**c**)   5-fold cross-validation
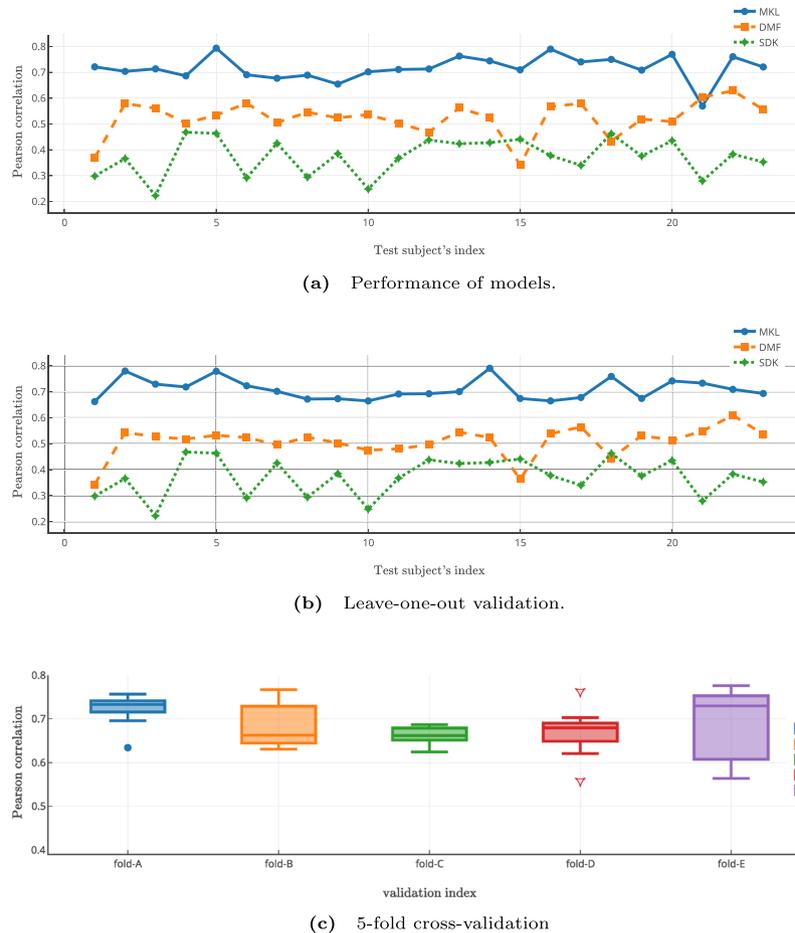
**Figure 1.** Comparison of Model Performance on Individual Test subjects. (**a**) Pearson correlation between empirical and predicted FCs of all the test subjects by multiple kernel learning (MKL) model and performance comparison with the predictions by the other two models. While MKL model has superior performance compared to that of dynamic mean field (DMF) and single diffusion kernel (SDK), DMF model performs slightly better than the SDK model. (**b**) Results of leave-one-out cross-validation on the test subjects also yield similar comparative performance. Note that the subject indices are kept identical between sub-figures (a) and (b). This plot suggests that MKL model can handle an increase in the number of training subjects without necessarily any over-fitting. (**c**) Box-plots of Pearson correlation measure on 9 randomly chosen validation subjects for each of the 5 folds for the MKL model. Points lying outside the quartiles are the suspected outliers. Compactness of boxes suggests inter-subject consistency of model's performance. Further, these 5-fold cross-validation results suggest that MKL model performs consistently well on unseen subjects.

a benchmark, we computed the Pearson correlation between empirical SC-FC pairs for all subjects and found mean value for these correlations to be 0.3 with a standard deviation of 0.02. These values are taken as baseline correlation values henceforth.

Figure 1 shows the performance comparison of the proposed method with other two models in three different setups. In the first setup, a randomly chosen set of half of the subjects was used for training (23 pairs) and the other half (23 pairs) for testing. Figure 1(a) shows the model performance for all the test participants for the three models. Since SDK and DMF models do not incorporate learning in their formulation, we gleaned the optimal values based on the training subjects. The optimal parameter settings were taken as the values at the mode of the performance distribution histogram for the training set and the same were used for estimating FC for test subjects. We took the best fitting scale on the training subjects for the SDK model and similarly selected the optimal global coupling parameter, $G$, for the DMF model. Optimal scale for SDK model worked out to be 0.8 and similarly optimal value of $G$ for DMF was 2.85. As can be seen from Fig. 1(a) and (b), the MKL model performs consistently better for each test subject when compared to the other two models. In the remaining two setups, in order to crosscheck whether MKL model suffers from over-fitting, we computed leave-one-out (Fig. 1b) and 5-fold cross-validation (Fig. 1(c)) results. These results clearly show the consistency in the performance of MKL model and indicate that the performance is not due to over-fitting nor it is due to any particular optimistic train-test split.

In all the experiments and for all the three models in order to compare group statistics, we compute the predicted FC for each test subject and then find the Pearson correlation coefficient with the corresponding empirical FC, followed by taking the mean of all these correlation coefficients. We designate the resulting mean correlation as mean FC in the rest of the paper.

**(a)** Emp        **(b)** MKL        **(c)** DMF        **(d)** SDK
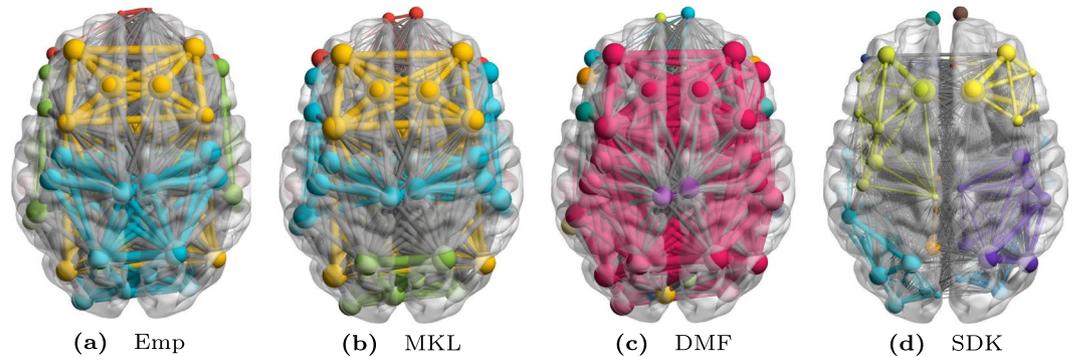
**Figure 2.** Functional Connectivity (FC) Networks Derived from Group-Mean FCs of the Test Dataset. The mean FC networks depict edge-connectivity patterns for (**a**) Empirical FC and FCs predicted by (**b**) MKL model; (**c**) DMF model[32] and (**d**) SDK model[18], respectively. Note the similarity of the MKL model and the empirical FC in terms of community assignment and inter-hemispheric connections. DMF model predicts a denser network while the single scale model predicts coarser network than the empirically observed FC. Brain-net-viewer[47] was used for visualization of the four communities detected from the Louvian algorithm available in brain-connectivity-toolbox[48]. Colors of the edges and nodes are only to demarcate the communities observed and do not have any correspondence across brain surfaces for different models. Thus the comparison of community structure across models is qualitative in nature.

**Edge-Connectivity Analysis.** Mean FCs are visualized primarily in two modes, via the edge-connectivity pattern analysis and using the seed-based connectivity analysis. To understand the edge and node distribution across the communities, we rendered the mean predicted FCs on brain surface. The visualization of edge-connectivity patterns of four mean FCs is shown in Fig. 2 (see Supplementary Fig. S6 for the estimated FCs). In the figure, the colors demarcate the communities for a particular model on the corresponding brain surface. It can be seen that the community structure of the mean FC predicted by the MKL model (shown in Fig. 2(b)) best resembles that of the mean empirical FC (shown in Fig. 2(a)). The other two models predict either a dense FC network (as shown in Fig. 2(c) for the DMF model) or a sparse FC network (as shown in Fig. 2(d) for the SDK model), where both the scenarios are far from the empirically observed network. Additionally, the predicted mean FC by MKL model and the empirical mean FC seem similar in terms of community assignment and inter-hemispheric connections.

Further, to see element-wise variance in the mean predicted FCs, we also drew scatter plots between the predicted and empirical FCs in Supplementary Fig. S7, where only the non-diagonal lower triangular matrix entries were displayed. These plots suggest that MKL model preserves the global structure of the empirical FC as well as the element-wise connectivity patterns significantly better than the other models.

**Seed-based Connectivity Analysis.** To further validate the nature of reconstruction of the connectivity patterns for various ROIs, we performed a seed-based correlation analysis using the mean FC matrices predicted from the three models. We chose the left posterior cingulate cortex (PCC) as a seed region since it has been known to have an important functional role as a hub region of the default mode network[33]. Figure 3 plots the correlation values between left PCC and all other regions on the brain surface reconstructed from the Desikan-Killiany atlas[34]. Cool (hot) colors suggest low (high) connectivity (correlation) of that particular region with the left PCC (shown as dark red color in the mean empirical FC). The MKL model could reconstruct the connectivity pattern with higher precision than the other two models. It appears that due to very high correlation of left PCC with all other regions, DMF model could not as clearly distinguish the boundary between regions. SDK model could not possibly distinguish them due to very sparse connectivity between left PCC and all other regions.

**Effect of Thresholding.** The rich-club organization of the structural connectivity (SC) matrix is previously demonstrated to be the backbone for generating the functional connectivity patterns[35–37]. Therefore we set out to investigate the impact of using thresholded SCs for predicting FCs. We pruned the SCs of all subjects by keeping only top T% of the connections (see Supplementary Section 1.4 for details). Each of these sparse matrices was passed as input to the learned MKL model. For each sparse SC, corresponding FC was predicted. Pre-learned $\pi_i$'s were used for predicting FCs in the MKL model. Similarly, fixed diffusion scale and $G$ parameters were used for SDK and DMF models for comparative evaluation. Figure 4 shows the mean correlation between empirical FC and predicted FCs for each of the sparse SC matrices. As can be seen, DMF and SDK models attain their respective optimum performance even when only few (as low as 10%) strongest edges in SC remained. On the other hand, MKL model requires both strong edges and few local edges, and hence its performance starts increasing from $T = 15\%$ and is significantly superior at all thresholds above this value. This result suggests that functional patterns may be primarily decided by the initial co-activations captured by $\pi_i$'s and that the structural constraints of individual SCs provide paths for these activities to diffuse, giving support to our hypothesis. Nevertheless, stable performance with sparsification as high as with $T = 20\%$ indicates that all the models obey the basic rich-club principle.
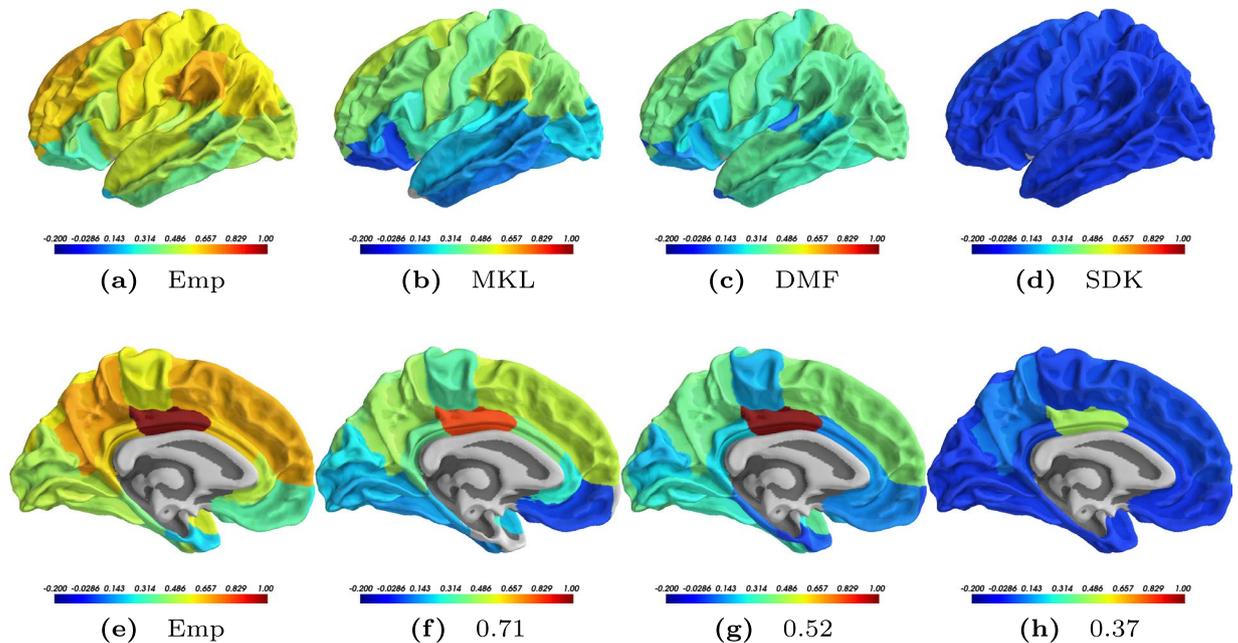
**Figure 3.** Results of Seed-based Correlation. Mean correlation maps resulting from considering the left Posterior Cingulate Cortex as a seed region and then calculating the seed-based correlations of all other regions. These maps are rendered on the left lateral sagittal view in the top sub-figures (**a**–**d**) and on the medial sagittal surface in the bottom sub-figures (**e**–**h**). While sub-figures (**a**) and (**e**) depict the maps for Empirical FC, the maps from the predicted FCs of MKL model are in (**b**) and (**f**); those of DMF in (**c**) and (**g**); and those of SDK in (**d**) and (**h**), respectively. Captions in the top row mention the model name and those in the bottom row indicate the mean correlation value on the test subjects. As can be observed, the correlation maps of MKL model seem to have greater correspondence with those of the mean empirical FC. Moreover, as depicted by the contrasts in the colors, MKL model is able to distinguish between the correlations at a better resolution than the other two models.
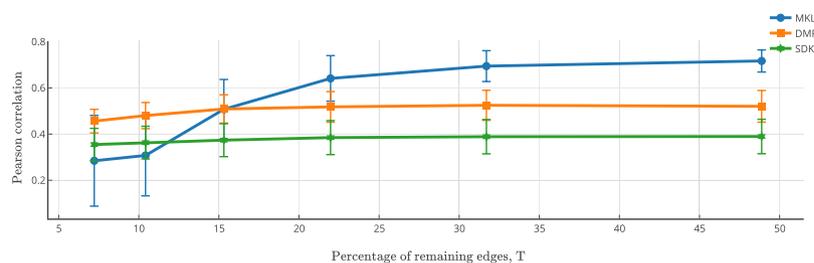


**Figure 4.** Model Performance on Sparse/Thresholded SCs. Structural Connectivity (SC) matrices for each subject were sparsified using 9 sparsification values. The three models were tested using these sparsified SCs. The plot shows the mean correlation values along with the standard deviation across subjects at each sparsification level. While MKL model was pre-trained with the un-sparsified SCs, a single optimal parameter was derived from the training data for each of the DMF and SDK models and used for estimating FC for the test subjects. The performance of MKL model starts to be superior after the sparsification level of 15% of the remaining edges. It appears that since MKL model selects the scales closely based on the SCs, the model performance degrades at very high sparsification levels.

Interestingly, MKL model captures the differences in sparsity levels better when compared to the other two models, especially when SC was pruned to keep the strongest edges between 10–20%. This behavior suggest that pre-trained $\pi_i$'s in MKL model do not compensate for major loss of information in sparsified SCs, thereby indicating avoidance of overfitting. Please see Supplementary Section 1.4 for the exact threshold values for this procedure.

**Robustness of the MKL Model.** The proposed model learns a latent representation, $\Pi$ that maps the relationship between SC and FC. This being the crucial difference between MKL and other models, we performed extensive robustness tests to verify the usefulness of learning the $\pi_i$'s. To ascertain that the model's representation learns important features and does not capture the SC-FC mapping by chance, we conducted the following four

**(a)** MKL      **(b)** DMF      **(c)** SDK
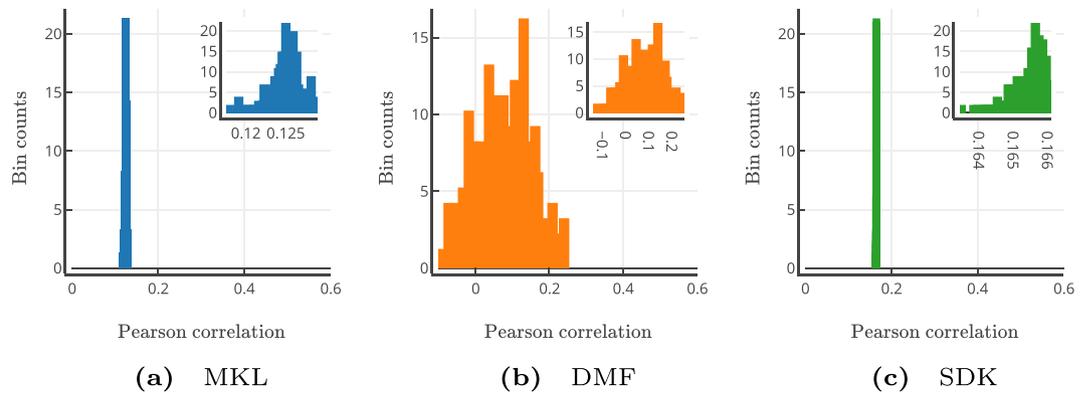
**Figure 5.** Model Performance with Perturbed Structural Connectivity (SC) Matrices. Randomly perturbed SCs ($N = 250$ sets) of the test subjects were used for estimating the FCs with the trained models. Sub-figures depict histograms of model's mean performance (Pearson correlation between empirical and predicted FCs): (**a**) MKL; (**b**) DMF; and (**c**) SDK model, respectively. The sub-plots (top right corners) within these sub-figures zoom in on the histograms for clarity. As expected, all the models depict degradation in performance with perturbed SCs.
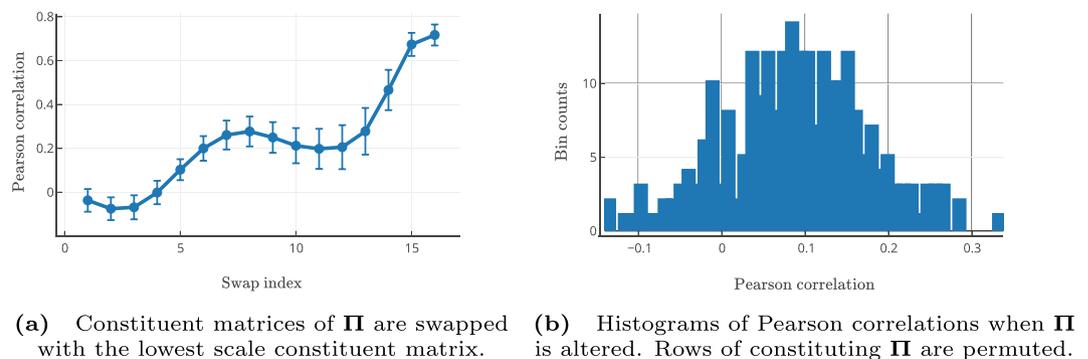


**(a)** Constituent matrices of **Π** are swapped with the lowest scale constituent matrix.      **(b)** Histograms of Pearson correlations when **Π** is altered. Rows of constituting **Π** are permuted.

**Figure 6.** Investigation of the Impact of Altering the Scale-specificity of the Parameters $\pi$'s. Two studies are conducted where the first study (**a**) looks at the impact of changing the scale-specificity of the $\pi$'s and the second study (**b**) looks at the impact of a larger-scale alteration when the components of individual $\pi$ matrices are themselves altered. (**a**) This sub-figure depicts the result of swapping each of the $\pi_i$ matrices with the last matrix, i.e., with $\pi_{16}$. For example, the first data point shows the mean performance when $\pi_1$ is swapped with $\pi_{16}$, the second data point corresponds to the case when $\pi_2$ is swapped with $\pi_{16}$ and so on for each of the 16 $\pi_i$ matrices being swapped with the last matrix $\pi_{16}$. Thus the last data point corresponds to the case when the original order was retained. The error bars represent the standard deviation. The results suggest scale-specificity of the learned parameters, i.e., in the sense that the performance degrades drastically if the $\pi$ matrices of one scale are swapped with a $\pi$ matrix of a distant scale. (**b**) The histogram of Pearson correlations depicts the performance when all the $\pi_i$ matrices are stacked together and the rows of the resulting stacked Π-matrix are swapped randomly 250 times. Such global alteration drastically degrades the performance. Together, these results indicate that the learned parameters do not predict FCs by chance but play a crucial role in the MKL model.

randomization experiments. In the first one we randomize the input to the model (i.e., SCs) (see Fig. 5) and in the second the learning itself is conducted based on perturbed SCs (see Supplementary Fig. S5). In the third experiment we disturb the scale-specific relation between the learned $\pi_i$'s and $H_i$'s (see Eq. 21) and finally in the fourth experiment the constituent rows of $\pi_i$'s are randomly permuted (see Fig. 6).

*Perturbing the model input.* To verify whether the model learns the SC-FC relationship correctly or predicts the average FC independent of SC, we provided the MKL model with perturbed SCs in two possible scenarios: first, while testing, and second, while training.

Each subject-wise SC was perturbed $N = 250$ times, hence forming 250 sets of subject-wise perturbed SC-empirical FC pairs (please see Supplementary section 1.5 for perturbation procedure). In the first perturbation analysis, we trained the MKL model with the original subject-wise training SC-FC pairs, and tested the model with these 250 perturbation sets. These same sets were used for evaluating the other two models. We calculated the mean correlation values between predicted and empirical FCs, thus obtaining 250 mean correlation values for every model. Figure 5(a)–(c) show the histograms of these mean correlation values for MKL, DMF and

SDK models, respectively. As expected, all the three models have significant drop in their performance indicating their sensitivity towards meaningful SC matrices while arriving at predictions.

In the second perturbation analysis, we trained $N = 250$ MKL models using the 250 purtubed sets and evaluated them using the subject-wise empirical SC-FC pairs. We did not have to perform this analysis for the other two models as this analysis is the same as that of the above for these models. Supplementary Fig. S5 shows the histogram of the 250 mean correlation values that is distributed across a wide range of correlation values instead of peaking at a particular value, thus indicating a *garbage-in, garbage-out* phenomenon from a machine learning perspective! This result, along with the results in Fig. 4, demonstrate that MKL model is not learning just a transformation from a subgraph of SC to an average FC but that the learning is holistic.

*Altering the model parameters.* After confirming that the model does not learn a random mapping between SC-FC pairs, we alter the learned mapping to further confirm model's robustness. We considered two ways of altering the model parameters ($\pi_i$'s). These parameters are mathematically represented as a set of $m$ matrices ($\pi_i$'s) corresponding to $m$ diffusion scales ($m$ here is set to 16, also see Eq. 21). We sought to experimentally verify that $\Pi$ can be interpreted as holding complementary information of a cohort of SCs. Hence it is likely that any perturbation of $\Pi$ would disturb the synergistic correspondence to empirical SCs and cause performance degradation. In order to experimentally validate this intuition, we ran two types of permutation tests.

Firstly, we sought to estimate the importance of the arrangement of $\pi_i$'s, i.e., the ordering of the scale-specific matrices constituting $\Pi$. For this we swapped every matrix $\pi_i$ ($1 \leq i \leq m$) one at a time with $\pi_m$ (corresponding to the lowest scale, i.e., $\pi_{16}$). Figure 6(a) shows the mean correlation while performing swapping. Pearson correlations are plotted against the swapped indices. Because of no-swap the last correlation (corresponding to $i = 16$) depicts optimal performance. This plot suggests that indeed matrices have positional significance (in other words, scale-specificity), so they cannot be reorganized to predict FC. This is a property that is also subtly captured in Eq. (21) in the sense that these matrices have a strict correspondence to their scales, consequently they embed scale-specific diffusion kernels to enable correct prediction of FC.

Secondly, we sought to estimate regional importance of the entries of $\pi_i$ matrices across scales. We concatenate all $m$ $\pi_i$'s into a single matrix ($\Pi$) of size $mn \times n$. We permute the rows of this matrix and test the model performance. A row of $\Pi$ captures regional co-activations at that scale between the region corresponding to that row and all other regions. We permute the rows of this large matrix 250 ($N$). Each newly generated $\Pi$ is used for testing the model performance. Figure 6(b) shows the histogram of the mean correlations of all the $N$ permutations. Clearly the plot shows that permuted $\Pi$ significantly deteriorates the model performance. This figure underlines the importance of maintaining the structure of co-activation between pairs of regions.

## Discussion

The holy grail in cognitive neuroscience is understanding how the static brain structure gives rise to dynamic function both during rest and task conditions. Several models have been proposed to characterize the structure-function relationship[38]. Simple linear diffusion models[18,19] as well as complex non-linear, whole-brain computational models[32] have been proposed. Linear graph models[18] admit closed form deterministic and testable solution to macroscopic interactions of brain activity without requiring any details of neural coding or their biophysical substrate. On the other hand nonlinear complex drift-diffusion models based on excitatory and inhibitory neuronal populations, though not analytically tractable, give rise to rich dynamics[32].

Abdelnour *et al.*[18] conceived a model of functional connectivity (FC) with only one diffusion kernel defined at an optimal scale. This optimal kernel operates on an identity matrix, meaning that the amount of activity reaching other regions from a single source is representative of the statistical dependence between those regions. This statistical dependence resembles activity heat maps which exhibit inter-individual variations. However, Surampudi *et al.*[29] showed that single kernel models do not generalize to a larger cohort and demonstrated that FC can be decomposed into multiple diffusion kernels with subject non-specific combination coefficients.

In this work, we proposed a *multiple kernel learning* (MKL) method that learns inter-regional co-activations (denoted as $\pi_i$'s) and reshapes the structurally confined diffusion kernels to give rise to functional connectivity estimates. MKL model is a generalization of the SDK diffusion model (see Supplementary and Fig. S1 for the strengths and limitations of SDK models). Resting state functional connectivity could be considered as a signal on a brain graph expressed at multiple different spatio-temporal scales. Our approach essentially finds a way to unfold these solutions on the brain graph combining multiple scales to accurately estimate the empirical FC. One way to interpret the proposed multi-scale diffusion model is to treat it as a variant of a reaction-diffusion system on the graph determined by the underlying structural connectivity (SC) matrix.

In this work, we adopt the representation of the graph signal in terms of eigenvectors of the graph Laplacian similar to what has been recently proposed[28]. The proposed MKL framework devises a scheme for learning the hidden parameters ($\pi_i$'s) to estimate FC. The initial regional activity $u_0$ in the reaction-diffusion type model is a vector, hence the matrix $u_0 u_0^T$ is a rank 1 matrix. As it is a positive semi-definite (PSD) matrix, it will only have one non-zero eigenvalue. Eigen-decomposition in Eq. (11) suggests a possible physical interpretation, that the initial mean activity distribution, an eigenvector of the graph Laplacian, resembles standing wave patterns on the graph. Total number of such standing waves is equal to the number of nodes of the graph. Hence our hypothesis is that the initial regional co-activations ($\pi_i$'s) correspond to one of the standing waves present at some time $k_i \tau$ significantly changing the pattern at that reaction instance (please refer to section titled *Methods/Proposed MKL model* for notations). Functional connectivity can then be articulated as a superposition of such standing wave patterns and their regional co-activations.

In order to predict FC from the proposed diffusion model, we estimated $\Pi$ by solving a LASSO optimization formulation. We hypothesized that these hidden parameters are learnable from the training data and remain fixed at the time of testing. Consequently different FC matrices for the test subjects would be arising by virtue

of the underlying differences in the respective structural connectivity matrices (SCs). This would mean that the parameters $\Pi$ are not merely a derivative of SCs but instead they complement the missing aspects by capturing the statistical dependence between two regions that are modulated by some intermediate region that may not be in physical proximity and that too operating at multiple resolutions or scales. Thus by incorporating the inter- and intra-hemispheric functional connectivity terms for a brain region, the learned optimal $\Pi$ parameters enable more accurate matching of the structure-function correlation. All the computational models can be visualized to lie on the spectrum spanned by biological interpretability and analytical ease. Whereas linear models enjoy simplicity of solution of their models, non-linear models tend to explain the complex biological reality. MKL model seems to find a sweet spot and enjoys best of both by analytically providing the solution and explaining the patterns in terms of large-scale excitatory-inhibitory interactions. Since LASSO optimization is the most expensive computational step, the computational complexity of the proposed MKL model would be dominated by the cost of LASSO optimization.

In summary, on the model continuum, the proposed MKL model lies somewhere between simple linear diffusion models[18,19] and complex non-linear drift diffusion models[32]. Consequently, we compared our simulation results predicting BOLD functional connectivity using the proposed model with models at either end of the complexity spectrum. The experimental results showed that the correlation structure of BOLD functional resting state brain networks is significantly well captured by our model. Prediction accuracy of the MKL model for the 23 test subjects is close to 0.70 whereas that of the non-linear model comes second best at 0.52 and that of the SDK model around 0.37. We conducted a series of tests that perturbed the inputs to the model as well as permuted the learned parameters $\Pi$. The test results attest to the robustness of the proposed model. Interestingly the model not only captures the variability of scales across participants but also demonstrates a possible application in characterizing age-related differences in learning optimal parameters for the accurate estimation of FC (refer to Supplementary section 4 and Figs S8–S10). Even in the face of considerable amount of variability present in the data, the proposed MKL model is still able to predict subject-specific FCs with high accuracy. Beyond this, functional connectivity subsumes the influence of different regions across scales and age groups providing a viability of $\Pi$ being a useful parameter for classification purposes for other domains of application in health and disease. Overall, our method might be considered the missing link in the estimation and improvement of predicting subject-specific resting-state functional connectivity that remained elusive so far for complex non-linear and linear models. Given the strength of the analytical approach and tractability, the proposed model could be a suitable method for predicting task-based functional connectivity across different age groups.

One major limitation of our work is that it is not so straightforward like the linear diffusion model to invert the FC to recover the SC matrix. Currently, in the MKL model the procedure to predict SC from FC would rely on a given $\Pi$. One way of finding SC is by estimating the diffusion kernels for individual subjects by solving the same system of linear equations used to find FC. Laplacian of a graph could then be estimated. Carefully recovering multiple diffusion kernels might turn out to cause numerical instability to the Laplacian (see Supplementary section 5 for details of the proposed inversion) and this issue needs to be resolved in the future. While in the current formulation we are empirically determining the number of scales ($m$) and their spacing, optimization formulation could be modified to estimate these automatically. Moreover, the current model does not consider the non-stationary nature of functional connectivity, the so called functional connectivity dynamics (FCD). Future studies can look at optimization procedures of MKL for modeling the dynamic functional connectivity which is more realistic than modeling stationary FC.

## Methods

**Notations.** This section introduces the notations used here as well as in the Supplementary material. Please refer to Table 1 for all the notations.

**Proposed Multiple Kernel Learning (MKL) Model.** In this section we describe the proposed model as a variant of Reaction-diffusion (RD) systems[39] wherein the regional mean activities diffuse on the graph determined by anatomical pathways (SC). RD systems explain the formation of complex self-organizing patterns naturally occurring in nature[40]. RD systems have been employed to model interaction among populations of neurons and the emerging patterns of functional connectivity among neural ensembles[41–44]. Just as statistical thermodynamics relates brownian motion of fluid particles to mean motion of a whole fluid, Wilson-Cowan equations characterize the macro-scopic statistical behavior of mean fields of the resulting neural activities[26,45]. Atasoy *et al.*[28] embed anatomical constraints in terms of the graph Laplacian matrix of the SC matrix in the Wilson-Cowan equations to explain the macro-scale excitatory and/or inhibitory interactions of the regional activities. These excitatory and/or inhibitory interactions result in the formation of complex functional patterns such as RSNs. We extend our model from Atasoy *et al.*[28] and explain the formation of FC through RSNs. We hypothesize that the cumulative mean activities of all the regions is generated by intra-regional micro-scale dynamics which diffuses inter-regionally on the structural connectome. We propose a physical model that implicitly captures the pairwise functional interactions between ROIs by explicitly associating them with their extent of influence through the diffusion kernels on the SC (see Supplementary section 3 for details on graph Laplacian and diffusion kernels).

The derivation of the expression for FC consists of five major stages. We consider that FC matrix encompasses effects of diffusion from multiple reactions. In the first stage, we formulate the differential equation for the time evolution of regional activities (Eqs 1–3). In second stage, we characterize the time evolution of the regional activities in an arbitrarily small time interval (Eqs 5 and 6). In the third stage, we integrate the diffusion process over all the connectome harmonics (Eqs 7 and 8). In the fourth stage we accumulate the diffusions happening in various time intervals to generate the complete expression for FC (Eqs 9–21). This FC assumes the form of a combination

| Object | Description |
|---|---|
| n | Number of ROIs or the number of nodes in the brain graph. |
| p | Number of training subjects. |
| SC | Structural connectivity matrix. |
| $SC^s$ | SC matrix for subject $s$. |
| $D^s$ | Degree matrix for subject $s$; sum of edge weights for every region. |
| FC | Functional connectivity matrix. |
| $FC^s$ | FC matrix for subject $s$. |
|  | $[f_1^s, \cdots, f_n^s]$ |
| $\mathbf{W}_{n\times n}$ | Weighted adjacency matrix of a graph. |
| $\mathbf{D}_{n\times n}$ | Degree matrix of a graph, computed by taking the sum of all weights on every node and diagonalizing the vector. |
| $\mathbf{L}_{n\times n}^s$ | Laplacian matrix of subject $s$. |
| $\Psi_{n\times n}^s$ | Eigenvector matrix of the graph Laplacian of subject $s$. |
| $\Lambda_{n\times n}^s$ | Eigenvalue matrix, diagonal matrix with increasing order of eigenvalues, of the graph Laplacian of subject $s$. |
| $\gamma_i$ | A scale at which diffusion kernel is defined. |
| $\mathbf{H}_{i\,n\times n}^s$ | Diffusion kernel at scale $\gamma_i$ for subject $s$. |
| $m$ | Number of scales |
| $\mathbf{H}_{n\times mn}^s$ | Collection of all $m$ diffusion kernels of a subject $s$. $\left[\mathbf{H}_{1\,n\times n}^s \quad \cdots \quad \mathbf{H}_{m\,n\times n}^s\right]$ |
| $\pi_{i n\times n}$ | Interregional co-activations corresponding to scale $\gamma_i$. |
| $\mathbf{\Pi}_{mn\times n}$ | Interregional co-activations collectively represented at all scales. $\begin{bmatrix} \pi_{1\,n\times n} \\ \vdots \\ \pi_{mn\times n} \end{bmatrix} = [\Pi_{mn\times 1}^1 \quad \cdots \quad \Pi_{mn\times 1}^n]$ |
| $X_{pn\times mn}$ | $\begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^p \end{bmatrix}$ |
| $Y_{pn\times n}$ | $\begin{bmatrix} FC_{n\times n}^1 \\ \vdots \\ FC_{n\times n}^p \end{bmatrix} = \begin{bmatrix} f_1^1 & \cdots & f_n^1 \\ & \vdots & \\ f_1^p & \cdots & f_n^p \end{bmatrix} = \begin{bmatrix} Y_{1\,pn\times 1} & \cdots & Y_{n\,pn\times 1} \end{bmatrix}$ |
| $\mathbf{C}_f^s$ | Predicted FC $\sum_{i=1}^m \mathbf{H}_i^s \pi_i$ |
| $\mathbf{C}_f\big|_{k_0}$ | Functional connectivity FC when reaction only happens at $k_0\tau$. |

**Table 1.** Notations used in the models and optimization formulation.

of diffusion kernels weighted by scale-specific parameters (Eq. 21). In the final stage, we propose an optimization framework for estimating these global parameters (Eqs 22–27).

Let the cumulative mean activities for all regions be denoted by $\mathbf{u}(t)_{n\times 1}$ at time $t$. We assume that these activities belong to either excitatory and/or inhibitory interactions. The temporal evolution of regional activities are modeled as the following linear variant of Wilson-Cowan equations:

$$\tau\frac{\partial}{\partial t}\mathbf{u}(t) = -\mathbf{u}(t) + \mathcal{D}[\mathbf{u}(t)]. \tag{1}$$

where $\mathcal{D}$ is the spatial propagation operator. $\tau$ is a characteristic time scale that speeds up or slows down the evolution of the system. Mean activity of a region $i$, $u_i(t)$, can be abstracted out from biological details as a one-dimensional (1-D) time varying signal. A vector of these 1-D signals indexed by the nodes of the graph represents a *graph signal*. We represent the graph signal in terms of its Fourier components using graph Fourier transform[46]:

$$\mathbf{u}(t) = \Psi\beta(t), \tag{2}$$

where $\Psi$ is the eigenvector matrix of graph Laplacian (see Supplementary section 3) and $\beta(t)$ is its Fourier representation at time $t$. With this decomposition temporal dynamics is explicitly represented using spatial basis functions. Further we conceptualize the spatial operator $\mathcal{D}$ in the form of a diffusion kernel defined at scale $\sigma^2/2$ on the structural brain graph Laplacian ($\Lambda$) corresponding to the time interval $\tau$ between two consecutive reaction instances.

$$\mathcal{D}[\mathbf{u}(t)] = \Psi e^{-\Lambda\,\sigma^2/2}\Psi^T\mathbf{u}(t) = \Psi e^{-\Lambda\,\sigma^2/2}\beta(t). \tag{3}$$

Substituting Equations (2) and (3) and combining the fact that $\Psi$ is invertible, differential Equation (1) can be solved for $\beta(t)$ which represents the signal evolution in the time interval between two reaction instances, as follows:

$$\tau\Psi\frac{d}{dt}\beta(t) = -\Psi\beta(t) + \Psi e^{-\Lambda\sigma^2/2}\beta(t) \tag{4}$$

$$\frac{d}{dt}\beta(t) = -\frac{1}{\tau}(\mathbf{I}_n - e^{-\Lambda\sigma^2/2})\beta(t) \tag{5}$$

$$\beta(t) = e^{-1\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\beta_0, \tag{6}$$

where $\beta_0$ represents the initial mean activity. Equation (6) depicts how the mean activity ($\beta_0$) of every region diffuses on the graph. Finally the graph signal between two reaction times can be expressed in a closed form (substituting Equation (6) in (2)) as:

$$\mathbf{u}(t) = \Psi e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\beta_0 \tag{7}$$

$$= \Psi e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Psi^T\mathbf{u}_0. \tag{8}$$

where, $\mathbf{u}_0 = \Psi\beta_0$ captures initial activity just after reaction, or at the start of diffusion. $\mathbf{u}_0$ depends on the magnitude of reaction phenomenon, hence may change after every reaction instance. Given the temporal evolution of graph signal, we will next derive how this leads to the evolution of functional connectivity and RSNs. RSNs have unique correspondence with graph-harmonics/eigenvectors of the structural graph Laplacian[28]. We develop the model for a single graph-harmonic, i.e., for all RSNs corresponding to that graph-harmonic. Finally, we superpose all the patterns of the resting state networks and explain the formation of FC.

The graph signal $\mathbf{u}_0$ may not change significantly in every reaction. Equation (8) represents the diffusive phenomenon of the graph signal over the characteristic time $\tau$. Let $\mathbf{u}_0$ change significantly at scalar multiples of $\tau$, i.e., $t + k_0\tau$, $t + k_0\tau + k_1\tau$, … with corresponding amplitudes $a_0$, $a_1$, …, respectively. For now we consider generating the functional connectivity $\left(\mathbf{C}_f|_{k_0}\right)$ for the time interval between two consecutive reactions; at times $t + k_0\tau$ and $t + k_0\tau + k_1\tau$.

$$\mathbf{C}_f|_{k_0} = \int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}\mathbf{u}(t)\mathbf{u}(t)^{\mathrm{T}}dt \tag{9}$$

$$= \int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}\Psi e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Psi^{\mathrm{T}}(a_0\mathbf{u}_0)(a_0\mathbf{u}_0^{\mathrm{T}})\Psi e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Psi^{\mathrm{T}}dt. \tag{10}$$

As $\mathbf{u}_0$ is also a signal on graph, we can express the positive semi-definite (PSD) matrix $\mathbf{u}_0\mathbf{u}_0^{\mathrm{T}}$ in terms of its eigen-decomposition. And as it is only a single harmonic, $\Delta$ is a diagonal matrix with only one non-zero entry.

$$\mathbf{u}_0\mathbf{u}_0^{\mathrm{T}} = \Psi\Delta\Psi^{\mathrm{T}}. \tag{11}$$

Hence, $\mathbf{C}_f|_{k_0}$ takes the following form:

$$\mathbf{C}_f|_{k_0} = a_0^2\int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}\Psi e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Delta e^{-1/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Psi^{\mathrm{T}}dt \tag{12}$$

$$= a_0^2\int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}\Psi e^{-2/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Delta\Psi^{\mathrm{T}}dt \tag{13}$$

$$= a_0^2\int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}\left\{\Psi e^{-2/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}\Psi^{\mathrm{T}}\right\}\{\Psi\Delta\Psi^{\mathrm{T}}\}dt \tag{14}$$

$$= \Psi\int_{t+k_0\tau}^{t+k_0\tau+k_1\tau}e^{-2/\tau\left(\mathbf{I}_n - e^{-\Lambda\sigma^2/2}\right)t}dt\Psi^{\mathrm{T}}\{a_0^2\theta_{k_0}\}. \tag{15}$$

We can denote the initial activity matrix as $a_0^2\theta_{k_0}$. As reaction instances are not usually far apart in time, instead of double exponentiation we utilize the first order Taylor approximation for the exponent of the integrand; i.e. $\mathbf{I}_n - e^{-\Lambda\sigma^2/2} \approx \Lambda\sigma^2/2$. Hence, $\mathbf{C}_f|_{k_0}$ becomes

$$\mathbf{C}_f|_{k_0} = \left(\Psi e^{-\sigma^2/\tau\Lambda(t+k_0\tau)}\Psi^{\mathrm{T}} - \Psi e^{-\sigma^2/\tau\Lambda(t+k_0\tau+k_1\tau)}\Psi^{\mathrm{T}}\right)\left(a_0^2\tau/\sigma^2\Lambda^{-1}\theta_{k_0}\right). \tag{16}$$

We call the matrix independent of time as

$$\pi_0 = a_0^2 \tau / \sigma^2 \Lambda^{-1} \theta_{k_0} \tag{17}$$

Now with multiple reactions happening at multiples of $\tau$, we sum over all the reaction instances to get the functional connectivity matrix:

$$\mathbf{C}_f = \sum_i \Psi e^{-\sigma^2/\tau \Lambda k_i} \Psi^{\mathrm{T}} \pi_i, \tag{18}$$

where,

$$\pi_i = (a_i^2 - a_{i-1}^2) \tau / \sigma^2 \Lambda^{-1} \theta_{k_i}. \tag{19}$$

Observing the structure of the FC matrix, FC is conceptualized as being represented by diffusion kernels and their corresponding inter-regional mean activities. So, the larger the value of $k_i$, the lesser is its contribution to FC. This means that summation on a finite number of diffusion scales is sufficient for reproducing FC (in this work we considered 16 diffusion scales based on pilot simulations). Now after combining the functional patterns of all the graph-harmonics, we approximate empirical FC with $m$ number of diffusion scales $\gamma_i$'s. The model thus takes the form as follows:

$$\mathbf{C}_f = \sum_{i=1}^{m} \Psi e^{-\Lambda \gamma_i} \Psi^{\mathrm{T}} \pi_i \tag{20}$$

$$= \sum_{i=1}^{m} \mathbf{H}_i \pi_i. \tag{21}$$

where, $\mathbf{H}_i$ denotes the diffusion kernel at scale $\gamma_i$. Further the model in Equation (21) suggests that the scale of diffusion is determined by a characteristic time constant ($\tau$), spatial diffusion variance ($\sigma^2$) and the time interval between consecutive reaction instances. Matrix $\pi_i$ represents the scale-specific initial relationships in the mean regional activities.

Proposed model represents the functional connectivity in terms of diffusion kernels operating on scale-specific matrices. It can be inferred that Adelnour et al.[18] envisage FC comprising only one diffusion kernel defined at an optimal scale. The optimal kernel operates on an identity matrix; meaning only the concerned region has non-zero mean activity independent of other regions, i.e., the amount of activity reaching other regions from the single source is representative of the statistical dependence between those two regions. Surampudi et al.[29] demonstrated that FC can be decomposed into multiple diffusion kernels whose combination coefficients are unique to the cohort. In addition to the multiple scales, proposed model provides inter-regional relationships instead of individually active regions. The proposed model generalizes both the aforementioned models as statistical dependence between two regions may be modulated by some intermediate regions without physical proximity that too at multiple resolutions or scales. Moreover, the model provides a biological interpretation of the diffusion scales and has an organic relationship to the reaction-diffusion system.

**Optimization formulation.** We hypothesize that the global parameters $\pi_i$'s are estimated from the training subjects (indexed by $s$ and varies from 1 to $p$) and remain fixed for all the test subjects. In order to estimate $\pi_i$'s we utilize an optimization formulation that minimizes an objective function $J$ comprising the mean squared error between empirical and predicted FCs.

$$\mathbf{J} = \sum_{s=1}^{p} \left\| \mathbf{C}_f^s - \mathrm{FC}^s \right\|_F^2 \tag{22}$$

$$= \sum_{s=1}^{p} \left\| \sum_{i=1}^{m} \mathbf{H}_i^s \pi_i - \mathrm{FC}^s \right\|_F^2 \tag{23}$$

$$= \left\| \sum_{s=1}^{p} \sum_{j=1}^{n} \mathbf{H}^s \Pi^j - f_j^s \right\|_F^2 \tag{24}$$

$$\leq \sum_{j=1}^{n} \left\| \sum_{s=1}^{p} \mathbf{H}^s \Pi^j - f_j^s \right\|_F^2. \tag{25}$$

To keep the number of reacting regions less, we employ $L_1$ norm on $\mathbf{\Pi}^j$'s.

$$\mathbf{J} \leq \sum_{j=1}^{n} \left\| \sum_{s=1}^{p} \mathbf{H}^s \Pi^j - f_j^s \right\|_F^2 + \lambda \|\Pi^j\|_1 \tag{26}$$

$$= \sum_{j=1}^{n} \left\| X\Pi^j - Y_j \right\|_F^2 + \lambda \|\Pi^j\|_1 \tag{27}$$

We apply sparsity on each column and compute every column of $\Pi^j$ separately. The objective function takes the form well known in regression analysis as *least absolute shrinkage and selection operator* (LASSO) that performs both variable selection and regularization. We passed the respective matrices into *lasso solver* to get the solution. We arrived at the model parameters experimentally, for example, the number of scales $m$ is empirically chosen.

**DMF Model.** We used the reduced dynamic mean field model as the non-linear model for comparative analysis[32]. This approach considers models with synaptic gating variable with passive decay differential equation along with Gaussian fluctuations. Firing rate was approximated based on input-output sigmoid function of the synaptic gating variable. The whole dynamics of each local network of excitatory and inhibitory populations of spiking neurons interconnected via excitatory synapses can be expressed by a single one-dimensional equation. The global brain dynamics of the network of interconnected local networks can be described by the following set of coupled non-linear stochastic differential equations[32]:

$$\frac{dS_i}{dt} = -\frac{S_i}{\tau_S} + (1 - S_i)\gamma H(x_i) + \sigma \nu_i(t) \tag{28}$$

$$H(x_i) = \frac{ax_i - b}{1 - exp(-d(ax_i - b))} \tag{29}$$

$$x_i = wJ_N S_i + GJ_N \sum_j C_{ij} S_j + I_0 \tag{30}$$

Here $S_i$ is synaptic gating variable of area $i$. $x_i$ is population mean firing rate for region $i$. $J_N$ is the excitatory synaptic coupling. $\nu_i$ in (28) is uncorrelated standard Gaussian noise with noise amplitude $\sigma = 0.001$ nA. $I_0$ is the external input current. $C_{ij}$ represents entries of the SC matrix which captures the structural connectivity between regions $i$ and $j$. Parameter values were selected as in Deco *et al.*[32]. A forward BOLD model was used that converts the local synaptic activity of a given cortical area into an observable BOLD signal. The simulated BOLD signal was down-sampled at 2 secs to have the same temporal resolution as in the empirically measured BOLD signal. Simulation length for computing the model FC was equivalent to 8 minutes. The coupling parameter $G$ (see Equation (30)) is varied between 0 to 3. We use individual empirical SC - FC matrices for exploration of subject-wise parameters for optimal fit. The optimal $G$ value varied among the subjects from 0.5 to 3. The mode of the distribution of the parameters obtained for training subjects was taken as the optimal $G$s for the training cohort and was found to be 2.85. The same value was used to estimate predicted FCs for all the test subjects.

**SDK model.** We used a linear diffusion model described in Abdelnour *et al.*[18]. This model considers SC matrix of a participant as the weighted adjacency matrix and computes graph Laplacian (described in Supplementary section 3). Then it assumes a set of scales and defines diffusion kernels at each scale. Iteratively for each diffusion kernel, the estimated FC is compared with empirical FC of that subject in terms of Pearson correlation. The scale at which Pearson correlation was found to be maximum is considered subject-specific optimum scale and its corresponding diffusion kernel is hypothesized as FC. Amongst the training subjects, we found that the mode of the optimal diffusion scale was 0.8, and this was maintained as a fixed parameter for all the test subjects.

**Data availability statement.** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. The codes for the MKL model are available at https://github.com/govindasurampudi/MKL.

## References

1. Hagmann, P. *From diffusion MRI to brain connectomics*. Ph.D. thesis, Université de Lausanne (2005).
2. Sporns, O., Tononi, G. & Kötter, R. The human connectome: a structural description of the human brain. *PLoS Comput Biol* **1**, e42 (2005).
3. David, O., Cosmelli, D. & Friston, K. J. Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage* **21**, 659–673 (2004).
4. Ioannides, A. A. Dynamic functional connectivity. *Current opinion in neurobiology* **17**, 161–170 (2007).
5. Hutchison, R. M. *et al.* Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* **80**, 360–378 (2013).
6. Hutchison, R. M., Womelsdorf, T., Gati, J. S., Everling, S. & Menon, R. S. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Human brain mapping* **34**, 2154–2177 (2013).
7. Ryali, S. *et al.* Temporal dynamics and developmental maturation of salience, default and central-executive network interactions revealed by variational bayes hidden markov modeling. *PLOS Computational Biology* **12**, e1005138 (2016).
8. Cabral, J., Kringelbach, M. L. & Deco, G. Functional graph alterations in schizophrenia: a result from a global anatomic decoupling? *Pharmacopsychiatry* **45**, S57–S64 (2012).

9. Bettinardi, R. G. *et al.* How structure sculpts function: Unveiling the contribution of anatomical connectivity to the brain's spontaneous correlation structure. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 047409 (2017).

10. Honey, C. *et al.* Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* **106**, 2035–2040 (2009).

11. Bressler, S. L. & Menon, V. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences* **14**, 277–290 (2010).

12. Deco, G. & Kringelbach, M. L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron* **84**, 892–905 (2014).

13. Deco, G., Rolls, E. T. & Horwitz, B. "what" and "where" in visual working memory: a computational neurodynamical perspective for integrating fmri and single-neuron data. *Journal of Cognitive Neuroscience* **16**, 683–701 (2004).

14. Newman, S. D., Carpenter, P. A., Varma, S. & Just, M. A. Frontal and parietal participation in problem solving in the tower of london: fmri and computational modeling of planning and high-level perception. *Neuropsychologia* **41**, 1668–1682 (2003).

15. Galán, R. F. On how network architecture determines the dominant patterns of spontaneous neural activity. *PloS one* **3**, e2148 (2008).

16. Barnett, L., Buckley, C. L. & Bullock, S. Neural complexity and structural connectivity. *Physical Review E* **79**, 051914 (2009).

17. Hlinka, J. & Coombes, S. Using computational models to relate structural and functional brain connectivity. *European Journal of Neuroscience* **36**, 2137–2145 (2012).

18. Abdelnour, F., Voss, H. U. & Raj, A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *Neuroimage* **90**, 335–347 (2014).

19. Saggio, M. L., Ritter, P. & Jirsa, V. K. Analytical operations relate structural and functional connectivity in the brain. *PloS one* **11**, e0157292 (2016).

20. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews. Neuroscience* **12**, 43–56 (2011).

21. Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O. & Kötter, R. Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences* **106**, 10302–10307 (2009).

22. Hahn, G. *et al.* Spontaneous cortical activity is transiently poised close to criticality. *PLOS Computational Biology* **13**, e1005543 (2017).

23. Nakagawa, T. T., Adhikari, M. H. & Deco, G. Large-scale computational models of ongoing brain activity. *Computational Models of Brain and Behavior* 425–437.

24. Kuramoto, Y. *Chemical oscillations, waves, and turbulence*, vol. 19 (Springer Science & Business Media 2012).

25. Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12**, 1–24 (1972).

26. Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics* **13**, 55–80 (1973).

27. Destexhe, A. & Sejnowski, T. J. The wilson–cowan model, 36 years later. *Biological cybernetics* **101**, 1–2 (2009).

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

29. Combining multiscale diffusion kernels for learning the structural and functional brain connectivity. *bioRxiv* 078766 (2016).

30. Lanckriet, G. R., Cristianini, N., Bartlett, P., Ghaoui, L. E. & Jordan, M. I. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research* **5**, 27–72 (2004).

31. Gonen, M. & Alpaydin, E. Multiple kernel learning algorithms. *Journal of Machine Learning Research* **12**, 2211–2268 (2011).

32. Deco, G. *et al.* Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience* **33**, 11239–11252 (2013).

33. Damoiseaux, J. *et al.* Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences* **103**, 13848–13853 (2006).

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

35. Ven Den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *Journal of Neuroscience* **31**, 15775–15786 (2011).

36. Collin, G., Sporns, O., Mandl, R. C. & van den Heuvel, M. P. Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cerebral cortex* **24**, 2258–2267 (2014).

37. Senden, M., Deco, G., de Reus, M. A., Goebel, R. & van den Heuvel, M. P. Rich club organization supports a diverse set of functional network configurations. *Neuroimage* **96**, 174–182 (2014).

38. Pillai, A. S. & Jirsa, V. K. Symmetry breaking in space-time hierarchies shapes brain dynamics and behavior. *Neuron* **94**, 1010–1026 (2017).

39. Morrison, P. The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions (1965).

40. Camazine, S. *Self-organization in biological systems* (Princeton University Press, 2003).

41. Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).

42. Turing, A. M. The chemical basis of morphogenesis. *Bulletin of mathematical biology* **52**, 153–197 (1990).

43. Kondo, S. & Miura, T. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* **329**, 1616–1620 (2010).

44. Angstmann, C. N., Donnelly, I. C. & Henry, B. I. Pattern formation on networks with reactions: A continuous-time random-walk approach. *Physical Review E* **87**, 032804 (2013).

45. Kilpatrick, Z. P. Wilson-cowan model. *Encyclopedia of Computational Neuroscience* 3159–3163 (2015).

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

47. Xia, M., Wang, J. & He, Y. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one* **8**, e68910 (2013).

48. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).

## Acknowledgements

## Author Contributions

S.G.S., A.S. and D.R. designed the study. S.G.S. and S.N. collected data. S.G.S. and S.N. ran the analyses and generated figures. S.G.S., A.S., R.B.S., D.R. and V.J. analyzed the results as well as the mathematical formulation. All authors have contributed in writing the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-018-21456-0.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.