



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Gao, X;Robinson, PA

Title:

Importance of self-connections for brain connectivity and spectral connectomics

Date:

2020-12

Citation:

Gao, X. & Robinson, P. A. (2020). Importance of self-connections for brain connectivity and spectral connectomics. *Biological Cybernetics*, 114 (6), pp.643-651. <https://doi.org/10.1007/s00422-020-00847-5>.

Persistent Link:

<https://hdl.handle.net/11343/272200>

License:

[CC BY](#)



ORIGINAL ARTICLE

# Importance of self-connections for brain connectivity and spectral connectomics

Xiao Gao<sup>1,2,3</sup> · P. A. Robinson<sup>2,3</sup> 

Received: 30 June 2020 / Accepted: 2 November 2020 / Published online: 26 November 2020  
© The Author(s) 2020

## Abstract

Spectral analysis and neural field theory are used to investigate the role of local connections in brain connectivity matrices (CMs) that quantify connectivity between pairs of discretized brain regions. This work investigates how the common procedure of omitting such self-connections (i.e., the diagonal elements of CMs) in published studies of brain connectivity affects the properties of functional CMs (fCMs) and the mutually consistent effective CMs (eCMs) that correspond to them. It is shown that retention of self-connections in the fCM calculated from two-point activity covariances is essential for the fCM to be a true covariance matrix, to enable correct inference of the direct total eCMs from the fCM, and to ensure their compatibility with it; the deCM and teCM represent the strengths of direct connections and all connections between points, respectively. When self-connections are retained, inferred eCMs are found to have net inhibitory self-connections that represent the local inhibition needed to balance excitation via white matter fibers at longer ranges. This inference of spatially unresolved connectivity exemplifies the power of spectral connectivity methods, which also enable transformation of CMs to compact diagonal forms that allow accurate approximation of the fCM and total eCM in terms of just a few modes, rather than the full  $N^2$  CM entries for connections between  $N$  brain regions. It is found that omission of fCM self-connections affects both local and long-range connections in eCMs, so they cannot be omitted even when studying the large-scale. Moreover, retention of local connections enables inference of subgrid short-range inhibitory connectivity. The results are verified and illustrated using the NKI-Rockland dataset from the University of Southern California Multimodal Connectivity Database. Deletion of self-connections is common in the field; this does not affect case-control studies but the present results imply that such fCMs must have self-connections restored before eCMs can be inferred from them.

**Keywords** Brain connectivity · Eigenmodes analysis · FMRI · Neural field theory · Self-connections

---

Communicated by Karl Friston.

---

This work was supported by the Australian Research Council Center of Excellence Grant CE140100007, the Australian Research Council Laureate Fellowship Grant FL1401000225, and a McKenzie Fellowship from The University of Melbourne.

---

✉ P. A. Robinson  
robinson@physics.usyd.edu.au

Xiao Gao  
xiao.gao@unimelb.edu.au

- <sup>1</sup> Department of Biomedical Engineering, University of Melbourne, Parkville, VIC 3052, Australia
- <sup>2</sup> School of Physics, The University of Sydney, Sydney, NSW 2006, Australia
- <sup>3</sup> Center of Excellence for Integrative Brain Function, The University of Sydney, Sydney, NSW 2006, Australia

## 1 Introduction

A typical method to analyze brain connectivity is through connectivity matrices (CMs), which contain the strengths of connections between pairs of discretized regions of interest (RoIs), which are usually chosen to be functionally homogeneous and spatially contiguous; for example, RoIs can be chosen to be voxels in functional magnetic resonance imaging (fMRI), subdivisions of specific Brodman areas, or based on the subject's own anatomy (Friston et al. 2006; Poldrack 2007). In CMs, the rows and columns in the matrices represent brain regions and entries represent the connections between brain regions (Bullmore and Sporns 2009; Sporns 2010; Friston 2011), so a CM is really a four-tensor that maps the 2D cortex to itself (Robinson 2019). The strengths of connection are defined in a variety of ways. Structural connectivity is often measured using anatomical CMs (aCMs),

which summarize the density of axonal bundles between RoIs as measured by methods such as diffusion tensor imaging (DTI) and related diffusion-weighted magnetic resonance imaging (dwMRI) (Basser et al. 2000; Hofer and Frahm 2006). Functional connectivity is commonly measured via functional CMs (fCMs), which are commonly determined from the two-point covariance (equal-time correlations) of the activity in brain regions using fMRI, on the assumption that regions that are correlated are likely to be functionally related. Effective connectivity matrices (eCMs), sometimes termed gain matrices, quantify the actual effect of one brain region to another, including the strength of connections. The CMs that embody the strengths of direct connections between points in a given brain state are termed direct effective CMs (deCMs), whereas total effective CMs (teCMs) describe the total connectivity between points via both direct and indirect paths (Robinson 2012).

Analyses of CMs are highly topical. Previously, (Robinson 2012) showed how eCMs correspond to propagators, using neural field theory (NFT) (Galán 2008), and used the deCM to compute teCMs and fCMs. Later, the inverse problem of inferring the eCMs from fCMs was studied (Robinson et al. 2014), and eCMs were determined from the experimental fCMs of resting state brain activity using NFT and eigenfunction analysis. On the widely used assumption that the deCM is proportional to the aCM, the theory interrelated structural, functional, and effective CMs in terms of propagators (Robinson 2012; Robinson et al. 2014), which enabled the propagator theory to be applied to analysis of connectivity. The existence and strength of connections that were not detected using dwMRI were also inferred in this work, particularly interhemispheric links via the corpus callosum. Later, Robinson et al. (2016) applied NFT to predict and analyze the activity eigenmodes of the bihemispheric brain, focusing particularly on their spatial structure. The eigenmodes of a single brain hemisphere were found to be close analogs of spherical harmonics, which are the natural modes of a sphere. The results showed a close match to experimental brain connectivity data (Robinson et al. 2016).

Self-connections are commonly removed from experimental fCMs because they are thought to be trivial or because correlations are only computed between time series from different locations, i.e., the diagonal elements in the experimental fCMs are omitted or set to zero, as in many central papers in the field, such as Hagmann et al. (2008); Brown et al. (2012), and in the key examples cited from the literature in foundational texts, such as those by Sporns (2010) and Fornito et al. (2016). This does not affect case-control comparisons where the difference between CMs is examined directly and terms involving the self-connections cancel out. However, it is known that the diagonal entries in any normalized covariance matrix are all 1, by definition. In this paper, NFT is used to investigate how the deleted self-

connections impact the properties and structure of CMs. In particular, by interrelating the fCM, teCM, and deCM using spectral analysis and NFT (Robinson 2012; Robinson et al. 2014), we investigate to what extent the missing self-connections in the fCM change the mutually consistent inferred eCM entries. First, an investigation of the importance of the self-connections in CMs is made. The role of the diagonal elements in the experimental fCMs is then clarified, the eCMs are calculated using NFT, and the effects of removing fCM self-connections on the inferred eCMs are demonstrated, showing that serious errors flow from failing to preserve the positive definiteness of the fCM, even though the proportion of connections omitted is tiny. The connectivity is also decomposed using eigenmode analysis, which enables representation of this modal brain connectivity in a compact diagonal matrix form and the rapid convergence is demonstrated (Robinson 2019).

It is worth stressing that the central aim of the present paper is *not* simply to prove that omission of diagonal entries changes the properties of the fCM in some way—that is entirely obvious (if you change any matrix you will change its properties). Rather, it is concerned with *how* the properties of the functional CM are changed and how these changes affect effective CMs inferred from the fCM via NFT. Second, it focuses on new results that can be obtained when self-connections are correctly retained.

This paper is organized as follows. Section 2 presents the theoretical background to NFT in matrix form, the calculation of eCMs from experimental fCMs, and the experimental dataset used in this paper to test the predictions. The results are presented in Sect. 3, where we show how the deleted self-connections in experimental fCM affect the structure of the fCM and the inferred eCMs. Methods that avoid these errors are discussed and the CMs are further analyzed using spectral methods that isolate individual eigenmode contributions to the total connectivity. The discussion and conclusions are presented in Sect. 4, where the broad applicability of the results to brain connectivity studies is emphasized.

## 2 Methods

A key message of the present paper is that it is necessary to define all mathematical and physical quantities carefully and to respect their basic properties. Hence, in this section we briefly recapitulate the relevant aspects of linear neural field theory and how its propagators (Green functions) relate to connection matrices.

### 2.1 NFT in matrix representation

Following the approach in (Robinson 2012), which pointed out that normal brain activity comprises mainly perturba-

tions from a mean level that corresponds to a fixed point of the dynamics. Thus, we write the perturbation field of the synaptic activity of neuron population  $a$  at location  $\mathbf{r}$  and time  $t$  as  $\Phi_a(\mathbf{r}, t)$ .

Since activity in a neuron population  $a = 1, \dots, P$  (where  $P$  is the total number of populations) is caused by neural input from populations  $b = 1, \dots, P$  (including  $a$ ) and external input  $N_a$ , we can write

$$\Phi_a(\mathbf{r}, t) = \sum_b \int \int \Lambda_{ab}(\mathbf{r}, t, \mathbf{r}', t') \Phi_b(\mathbf{r}', t') d\mathbf{r}' dt' + N_a(\mathbf{r}, t), \tag{1}$$

where the propagator  $\Lambda_{ab}$  quantifies the activity evoked in neuron population  $a$  at location  $\mathbf{r}$  and time  $t$  by activity afferent from neuron population  $b$  at  $\mathbf{r}', t'$ . To preserve causality,  $\Lambda_{ab}(\mathbf{r}, t, \mathbf{r}', t') = 0$  for  $t < t'$ .

In CM analysis, the locations and time are usually discretized, so in Eq. (1) the integrals are replaced by sums over discrete values of  $\mathbf{r}$  and  $t$ . Here, we discretize each of the  $P$  populations into  $M$  spatial regions and view synaptic activity  $\Phi_a(\mathbf{r}, t)$  and external input  $N_a(\mathbf{r}, t)$  as  $MP$ -element column vectors, in which  $P$  groups of  $M$  elements each represents one population’s activity on the  $M$  chosen regions (Robinson 2019). Then in matrix format Eq. (1) becomes

$$\Phi(t) = \int \Lambda(t, t') \Phi(t') dt' + N(t), \tag{2}$$

where  $\Phi$  and  $N$  are  $MP$ -element column vectors that represent activities through all spatial points for each neural population in turn, and  $\Lambda$  is an  $MP \times MP$  matrix. We note that each element in the matrices  $\Phi$ ,  $N$ , and  $\Lambda$  includes an implicit factor that corresponds to the volume element represented by that point, and this highlights the need to use a fine discretization if the integral is to be done accurately, and also possible to determine when experimental data suffice to yield convergent results that reflect properties of the brain, rather than of discretization and thresholding (Robinson 2019). In fCM measurements, only large-scale connections of excitatory neurons via white matter are measured; hence, we only consider connectivity of the excitatory population explicitly. This means that subscripts  $a$  and  $b$  are omitted henceforth,  $\Phi$  and  $N$  reduce to  $M$ -element vectors, and  $\Lambda$  is of size  $M \times M$ .

### 2.2 Inferring eCMs from fCMs

The method of calculation of eCMs from experimental fCMs was introduced in Robinson et al. (2014). We emphasize that we only consider the symmetric case because fCMs are determined from covariances, and are symmetric by definition; they do not include information on the directionality of links.

If  $\Lambda$  can be approximated as static on the timescale of cortical activity, it depends only on  $t - t'$ . Following Eq. (2), we then have

$$\Phi(t) = \int \Lambda(t - t') \Phi(t') dt' + N(t), \tag{3}$$

where the propagator  $\Lambda$  is now identified as being the spatiotemporal deCM (Robinson 2012; Robinson et al. 2014). By Fourier transforming Eq. (3) versus time, we have

$$\begin{aligned} \Phi(\omega) &= [I - \Lambda(\omega)]^{-1} N(\omega), \tag{4} \\ &= T(\omega) N(\omega), \tag{5} \end{aligned}$$

where  $\omega$  is the angular frequency (this argument distinguishes Fourier transformed quantities from temporal ones, which have time as their argument),  $I$  is the unit matrix, and  $T$  is the transfer matrix that links activity  $\Phi$  to input  $N$ , which is approximated as white noise in the resting state, i.e.,  $T$  is the teCM.

The teCM  $T$  can be expanded in a Taylor series to give

$$T(\omega) = \sum_{m=0}^{\infty} [\Lambda(\omega)]^m, \tag{6}$$

where the powers represent successively higher-order polysynaptic paths from input locations to the cortex (Robinson 2012; Robinson et al. 2014; Mehta-Pandjee et al. 2017)

We define the fCM  $\tilde{C}$  to be the covariance matrix of the activities (Robinson et al. 2014),

$$\tilde{C}(\tau) = \langle \Phi(t + \tau) \Phi^T(t) \rangle, \tag{7}$$

where the angle brackets indicate an average over  $t$ . With  $\tau = 0$ , we write Eq. (7) in terms of transfer function as (Robinson et al. 2014)

$$\tilde{C}(\omega) = T(\omega) T^\dagger(\omega), \tag{8}$$

in the resting state where the dagger indicates a Hermitian conjugate and is just the transpose (denoted by the superscript  $T$  below) at  $\omega = 0$  where  $T$  is real. For very low frequencies of fMRI, i.e.,  $\omega \lesssim 1 \text{ s}^{-1}$  (Jezzard et al. 2001; Aquino et al. 2012), we can approximate Eq. (8) as

$$\tilde{C} \approx \tilde{C}(\omega = 0). \tag{9}$$

We note that all matrices relevant to fMRI involve very low frequencies, so  $\omega \approx 0$ , and we omit this argument henceforth.

Most experimental studies use the normalized covariance matrix  $\tilde{C}$  to define the fCM; this is obtained from  $\tilde{C}$  by dividing all its elements  $\tilde{c}_{ij}$  by the geometric mean of  $\tilde{c}_{ii}$  and  $\tilde{c}_{jj}$ , with

$i = 1, \dots, p$  and  $j = 1, \dots, M$ . If we write the elements of  $\mathcal{C}$  as  $\tilde{c}_{ij}$ , we thus have

$$c_{ij} = \frac{\tilde{c}_{ij}}{\sqrt{\tilde{c}_{ii}}\sqrt{\tilde{c}_{jj}}}. \quad (10)$$

If we approximate the diagonal elements by their average  $\langle \tilde{c}_{ii} \rangle$ , Eq. (10) can be simplified to

$$c_{ij} = \frac{\tilde{c}_{ij}}{\langle \tilde{c}_{ii} \rangle}, \quad (11)$$

whence

$$C = \frac{1}{\langle \tilde{c}_{ii} \rangle} \tilde{C}. \quad (12)$$

We emphasize that we use the normalized covariance matrix  $\mathcal{C}$  to define the fCM. Equations (4)–(6), and (8) show that  $\Lambda$ ,  $T$ ,  $C$ , and  $\tilde{C}$  commute. Hence using standard matrix theory for symmetric  $\Lambda$ , we can write

$$\Lambda = ULU^\dagger, \quad (13)$$

where  $U$  is a unitary matrix whose columns are the eigenvectors of  $\Lambda$ ,  $U^\dagger = U^{-1}$ , and  $L$  is a diagonal matrix of the eigenvalues  $\lambda_j$  of  $\Lambda$ , written

$$L = \text{diag}(\lambda_j). \quad (14)$$

For  $\omega = 0$ , the unitary matrices become real orthonormal matrices and the Hermitian conjugates above are equivalent to transposes.

We can also write  $T$  and  $C$  in diagonal form, analogous to Eq. (13):

$$T = U\Theta U^\dagger, \quad (15)$$

$$C = UKU^\dagger, \quad (16)$$

where  $\Theta$  as a diagonal matrix of the eigenvalues of  $T$  and  $K$  as a diagonal matrix of the eigenvalues of  $C$ . Following (Robinson et al. 2014; Pinotsis et al. 2014),

$$\Theta = \text{diag}(\theta_j) = \text{diag}([1 - \lambda_j]^{-1}), \quad (17)$$

$$K = \text{diag}(\kappa_j) = \text{diag}(|1 - \lambda_j|^{-2}), \quad (18)$$

where Eqs (17) and (18) express  $\theta_j$  and  $\kappa_j$  in terms of  $\lambda_j$ . The matrices  $L$ ,  $\Theta$ , and  $K$  are the deCM, teCM, and fCM expressed in terms of eigenfunctions (i.e., in the eigenfunction basis, rather than the coordinate basis defined by RoIs).

### 3 Results

In this section, we first point out some theoretical and practical errors that result from removing self-connections in fCMs and stress that, although the mathematical cause is quite simple, the result is that serious errors have affected a whole research field, and continue to do so. We then demonstrate and illustrate the results, especially for non-mathematical audiences, and investigate how the removal of self-connections impacts the structure of fCMs and eCMs inferred from them using the publicly available NKI-Rockland experimental dataset, obtained from the USC Multimodal Connectivity Database (Nooner et al. 2012). The fCM (of size  $165 \times 165$ ) used in this paper is based on group-average data and is a normalized covariance matrix. In each connectivity matrix, the elements are ordered so that the first 81 elements in each row and column are from the left hemisphere and the next 84 are from the right hemisphere.

A detailed description of the NKI-Rockland dataset can be found in Brown et al. (2012), so we do not reproduce this material in full here. In brief, fMRI time series were collected over several minutes per scan in voxels of  $(1\text{--}2 \text{ mm})^3$ . Preprocessing to remove motion artifact, reduce noise, and correct for a range of systematic effects in the experiment was then carried out. The resulting time series were clustered into 188 regions of interest (RoIs), of which 165 were cortical. The mean time series for each RoI was then calculated and correlated with the others to obtain the fCM used here, which is restricted to the cortical RoIs.

#### 3.1 Effect of deleting self-connections on CMs

In the NKI-Rockland fCMs, as in many others, the self-connections in the experimental fCM have been removed by setting the diagonal elements to 0. However, we know that the diagonal entries in any normalized correlation matrix are 1 by definition. In this section, we investigate how the properties of the experimental fCM are affected by deletion of the self-connections, as well as the effect of this step on the eCMs that are inferred via Eqs (13)–(18).

We write the normalized experimental fCMs with diagonal entries as  $C$ , and the fCM with diagonal entries deleted as

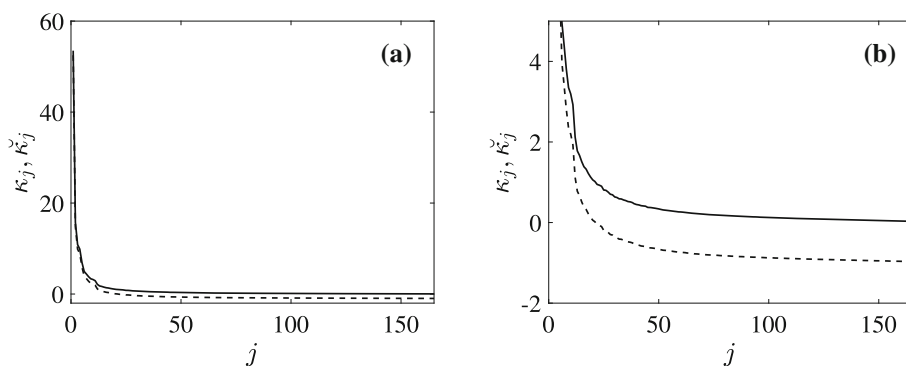
$$\check{C} = C - I, \quad (19)$$

where  $I$  is the unit matrix.

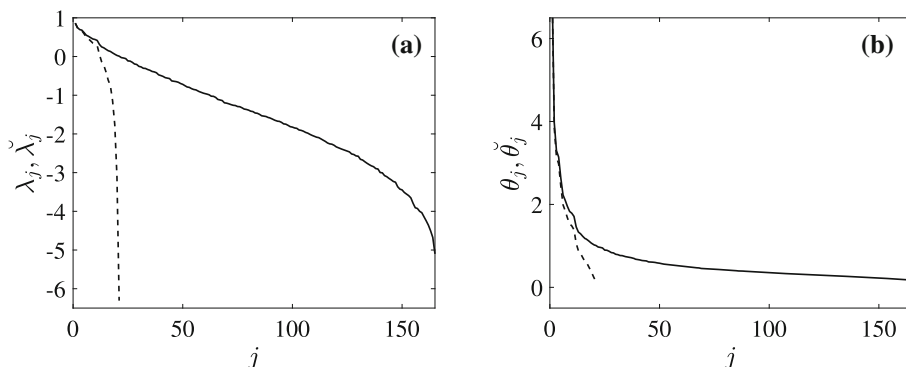
Now  $C$  and  $\check{C}$  commute because  $I$  commutes with any matrix. Hence, the three matrices share the same eigenvectors, and we can write

$$\check{C} = U\check{K}U^\dagger, \quad (20)$$

**Fig. 1** Comparison between the eigenvalues  $\kappa_j$  of  $C$  (solid) and  $\check{\kappa}_j$  of  $\check{C}$  (dashed). **a** All eigenvalues. **b** Expanded view of eigenvalues with  $\kappa_j < 5$



**Fig. 2** Comparison between the eigenvalues  $\lambda_j$  and  $\theta_j$  of  $\Lambda$  and  $T$  calculated from  $C$  (solid) and  $\check{\lambda}_j$  and  $\check{\theta}_j$  of  $\check{\Lambda}$  and  $\check{T}$  of  $\check{C}$  (dashed). **a** Eigenvalues of the deCMs. **b** Eigenvalues of the teCMs



where  $\check{K}$  is the diagonal matrix of the eigenvalues  $\check{\kappa}_j$  of  $\check{C}$ . We know that the eigenvalues of a unit matrix are all 1, so

$$\check{\kappa}_j = \kappa_j - 1. \tag{21}$$

We use the notation  $\check{\Lambda}$ ,  $\check{T}$  for the deCM and teCM calculated from  $\check{C}$ , respectively, and  $\check{\lambda}_j$  and  $\check{\theta}_j$  for the eigenvalues of  $\check{\Lambda}$  and  $\check{T}$ , respectively.

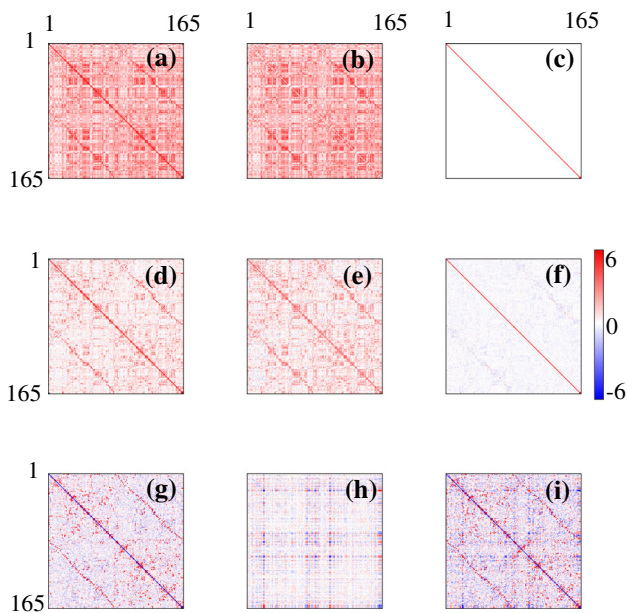
Figure 1 compares the eigenvalues of experimental  $C$  and  $\check{C}$ , arranged in decreasing order. The results agree with Eq. (21) to within numerical round-off errors. We notice that  $\check{\kappa}_j < 0$  for  $j \geq 21$ , which demonstrates a fundamental error in the use of  $\check{C}$  because covariance matrices are positive definite by definition and thus must have positive eigenvalues. If we were to use  $\check{\kappa}_j$  to calculate the eigenvalues of the deCM and teCM, only 21 eigenvalues of  $C$  would potentially be available to calculate  $\check{\lambda}_j$  and  $\check{\theta}_j$  and even these are not accurate. Hence, diagonal elements of the fCM must be retained to obtain valid results.

Figures 2a, b, respectively, show the eigenvalues of the deCMs and teCMs calculated from  $C$  (solid) and  $\check{C}$  (dashed). We observe that use of  $\check{\kappa}_j$  leads to incorrect estimation of the eigenvalues of deCM and teCM, although the first few are approximately correct. We also note that the largest eigenvalues of the deCMs are both approximately 0.87, which is also in close agreement with prior EEG- and fMRI-based results (Robinson 2017).

Figure 3 compares the effect of self-connectivity on fCMs and the corresponding inferred eCMs. Comparing Figs 3a,

b we notice the absence of diagonal connections in  $\check{C}$ . Figure 3c shows the teCM  $T$  calculated from  $C$ , we observe structure in Fig. 3c that is similar to that in Fig. 3a, except that there is a smaller fraction of strong connections, and these are concentrated more tightly around the main diagonal and the secondary diagonals that represent interhemispheric connections between homologous regions. Some block-like structure is seen, although this is mostly an illusion that is caused by mapping the 2D cortex onto a 1D list of RoI label, it does not represent discrete modularity (Henderson and Robinson 2011)

Figure 3e shows the deCM  $\Lambda$  calculated from  $C$ . Again, we observe that the strongest entries are on the main diagonal and secondary (interhemispheric, between homologous regions) diagonals. However, entries near the main diagonal are negative. These negative elements reflect the presence of net local inhibition that approximately balances the net excitation at longer ranges, as is required for the brain to have its observed overall marginal stability (Robinson 2017). Notably, they also demonstrate the ability of the inversion method to infer the effects of spatially unresolved structure with scales ( $< 1$  mm) well below those of the cortical discretization ( $\sim 4$  cm). In contrast to Fig. 3a, c, many entries in  $\Lambda$  are negative, this is because  $\Lambda$  only quantifies the strength of direct connections between RoIs. One important point to stress here is that the net local inhibition at scales of one grid point (representing several square cm of cortex) with longer range excitation via white matter fibers must not be confused

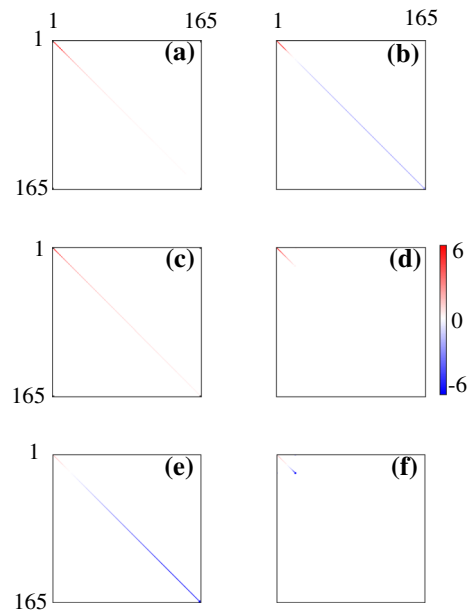


**Fig. 3** Functional CM and the corresponding inferred eCMs with the strengths of entries given by the color bar at right. **a**  $C$ , **b**  $\check{C}$ , **c**  $T$ , **d**  $\check{T}$ , **e**  $\Lambda$ , and **f**  $\check{\Lambda}$

with the well-known sub-mm central excitation and surround inhibition of individual neurons. The implication is that the latter is the stronger, when integrated over distances of 1 mm or more, and that it must then be balanced by even longer-range excitation to achieve the observed near-criticality of the cortex.

We next examine the CMs calculated from  $\check{C}$  in Fig. 3b. We see that the incorrect estimation of the eigenvalues  $\check{\theta}_j$  leads to incorrect estimation of  $\check{T}$ . Although the diagonals and block-like structures appear in  $\check{T}$ , the connectivity strength, especially the entries in the main diagonal are significantly reduced in  $\check{T}$ . Figure 3f shows  $\check{\Lambda}$ , which shows almost no sign of the main structures seen in  $\Lambda$ . This is because most of the negative eigenvalues in  $\Lambda$  are not captured in  $\check{\Lambda}$ . Thus, our results show that the missing self-connections in the fCM greatly affect the estimated connections within and between hemispheres in the inferred eCMs and thus must be retained to obtain the accurate results. At one level this is to be expected, but the extent of the changes is surprising, given how small a proportion of the connections are deleted in the fCM—only 0.6% in the present case.

From the theoretical analysis, we know that the eigenmodes of the fCM and the inferred eCMs are the same. By using eigenmode analysis and spectral method we can write the connectivity matrix in diagonal form. Figure 4a, b highlights the dramatic simplification relative to Fig. 3a, b by showing the diagonalized fCMs  $K$  and  $\check{K}$ , respectively. The entries in Fig. 4a are all positive because  $C$  is a covariance matrix and thus can have only positive eigenvalues. However,



**Fig. 4** Diagonalized fCMs and the corresponding inferred eCMs with the strengths of entries given by the color bar at right. **a**  $C$ , **b**  $\check{C}$ , **c**  $T$ , **d**  $\check{T}$ , **e**  $\Lambda$ , and **f**  $\check{\Lambda}$

in Fig. 4b, only 21 eigenvalues are positive, as was discussed in Sect. 3.1. Figure 4c shows the diagonalized  $T$  (i.e.,  $\Theta$ ) calculated via Eq. (17), again showing an enormous simplification with all eigenvalues positive. Figure 4e shows the diagonalized  $\Lambda$  (i.e.,  $L$ ) calculated using Eq. (13). In this case, we observe negative  $\lambda_j$  at large  $j$  in Fig. 4e. Figure 4d, f shows the diagonal matrices  $\check{\Theta}$  and  $\check{L}$ , respectively. Since only 21 eigenvalues in  $\check{C}$  can be used in calculating  $\check{T}$  and  $\check{\Lambda}$ , we only observe limited eigenvalues in both figures, which leads to the incorrect estimation of eCMs shown in Fig. 3. We note that NFT in a spherical geometry (Robinson et al. 2016; Robinson 2019) implies

$$\lambda_j \sim -j, \quad (22)$$

$$\theta_j \sim j^{-1}, \quad (23)$$

and

$$\kappa_j \sim j^{-2}, \quad (24)$$

at large  $j$ , shown in Figs. 1 and 2, all of which are consistent with the results shown in Fig. 4. Therefore,  $\theta_j$  and  $\kappa_j$  converge to 0 as  $j$  increases, whereas  $\lambda_j$  grows. Thus, only  $C$  and  $T$  are expected to give diagonal forms that are dominated by just a few modes, in accord with the results in Fig. 4. The teCM  $T$  is far more important dynamically than  $\Lambda$  because it includes all connections, not just direct ones (Robinson 2012; Robinson et al. 2014, 2016).

### 3.2 Modal analysis of CMs

Using the above data, we now investigate the contributions of individual modes to the CMs  $C$  and  $T$ , which have the most compact forms. Previous work has shown the lowest terms to be dominant (Robinson et al. 2016), but the form of the individual contributions to the CMs has not been previously studied. The contributions of the  $j$ th eigenmode to  $C$  and  $T$  are

$$C_j = UK_jU^\dagger, \tag{25}$$

and

$$T_j = U\Theta_jU^\dagger, \tag{26}$$

respectively, where

$$K_j = \text{diag}(0, \dots, 0, \kappa_j, 0 \dots, 0), \tag{27}$$

and

$$\Theta_j = \text{diag}(0, \dots, 0, \theta_j, 0 \dots, 0). \tag{28}$$

We also define the partial sums of the first  $m$  eigenmode contributions to  $C$  and  $T$  to be

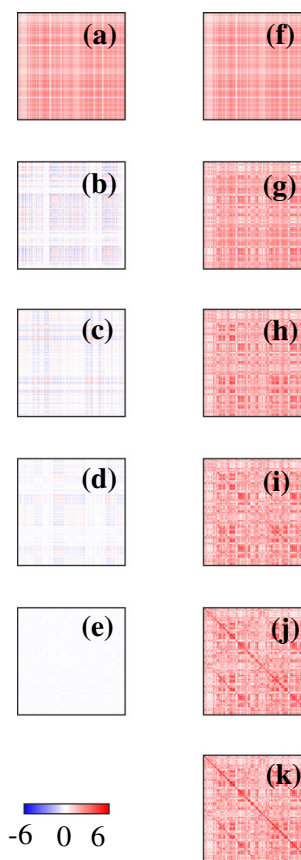
$$S_m = \sum_{j=1}^m C_j, \tag{29}$$

and

$$V_m = \sum_{j=1}^m T_j, \tag{30}$$

respectively.

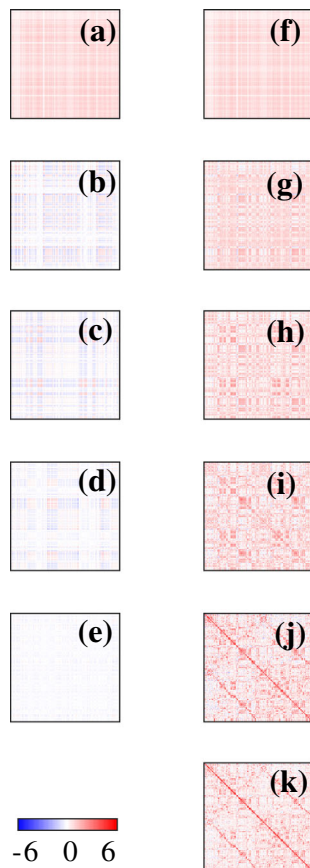
Figure 5 shows modal contributions to  $C$  and their partial sums. Figure 5a–e shows the contributions  $C_1, C_2, C_3, C_5$  and  $C_{20}$ , respectively, while Fig. 5f–j shows the corresponding partial sums  $S_1, S_2, S_3, S_5$ , and Fig. 5(k) shows  $C$  for comparison. The first mode, shown in Fig. 5a, is approximately spatially uniform and has no negative entries, which is consistent with the previous results that this corresponds to a uniform mode (Robinson et al. 2016; Gabay and Robinson 2017). The next modes have approximately equal numbers of positive and negative entries, consistent with their mean values being zero because of orthogonality to the lowest, uniform mode. These contributions are spatially nonuniform because their spatial eigenmodes have nodal lines that divide positive from negative regions (Robinson et al. 2016; Gabay and Robinson 2017). The size of these contributions drops rapidly from  $C_1$  to  $C_{20}$  as the eigenvalues decrease. As shown in Fig. 5g–i, the block-like structures start to appear in  $S_2$ ,



**Fig. 5** Functional connectivity matrix in modal analysis with the strengths of entries given by the color bar at bottom left. **a–e** show modal contributions  $C_1, C_2, C_3, C_5$  and  $C_{20}$ , respectively; **f–j** show partial sums  $S_1, S_2, S_3, S_5$  and  $S_{20}$ , respectively; and **k** shows  $C$

the diagonals start to appear in  $S_3$  and most of the strongest connectivity is already present in  $S_5$ , in accord with just the first few eigenmodes being dominant (Robinson et al. 2016). Comparing Fig. 5j, k, we observe that the connectivity strength is almost the same for  $S_{20}$  and  $C$ . This shows that to obtain the full sharpness and strength of the diagonals, requires superposition of around 20 or more modes to obtain sufficient spatial localization.

Figure 6 shows modal contributions to  $T$  and their partial sums. Figure 6a–e shows modal contributions  $T_1, T_2, T_3, T_5$  and  $T_{20}$ , respectively, Fig. 6f–j shows partial sums  $V_1, V_2, V_3, V_5$  and  $V_{20}$ , respectively, and Fig. 6k shows  $T$ . As for  $C_1$ , we find that  $T_1$  in Fig. 6a is approximately uniform. Indeed, the structure of each eigenmode contribution is the same as for  $C$  because  $C$  and  $T$  have the same eigenvectors, but the weights decrease more slowly in  $T$  because of the slower decrease of its eigenvalues. Comparing Fig. 6j–k, we again observe that accurate representation of the main diagonal requires retention of more modes than other parts of the matrix.



**Fig. 6** Total effective connectivity matrix in modal analysis with the strengths of entries given by the color bar at bottom left. **a–e** show  $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_5$  and  $T_{20}$ , respectively; **f–j** show  $V_1$ ,  $V_2$ ,  $V_3$ ,  $V_5$  and  $V_{20}$ , respectively; and **k** shows  $T$

## 4 Conclusion

The importance of retaining self-connections in experimental fCMs and the eCMs inferred from them has been investigated using methods from spectral analysis and neural field theory. We have introduced new methods to explore the connectivities and their effects on the relationships between functional and effective CMs using eigenmode analysis and spectral methods through which we also represent the brain connectivity in a compact diagonal matrix form. The results have been illustrated and verified using NKI-Rockland data, which underline the need to adopt these new methods to ensure the accurate results. The main findings are:

- (i) The near-universal step of in published studies, deleting diagonal entries from functional CMs defined by activity covariances, is invalid. It violates fundamental mathematical requirements on covariance matrices and invalidates the physical relationships between functional and effective CMs, as detailed in the following points. Again, we stress that the fact that there is a

change of some kind is mathematically obvious and the central purpose of the paper is thus not merely to establish its existence, but to explore its nature and the resulting effects on both the fCM and, more importantly, on the effective CMs that are inferred from it via NFT.

- (ii) The analytical and numerical results show that all eigenvalues of  $C$  decrease by 1 after deleting the self-connections. Therefore, most of the eigenvalues of the fCM without self-connections become negative, which is impossible for a covariance matrix, and indicates a fundamental error. These negative values cannot be used to validly calculate eCMs (Robinson et al. 2014). Thus, it is mathematically essential for self-connections to be retained.
- (iii) Because the fCM, deCM, and teCM are symmetric and commute, they share the same eigenvectors and can be represented in closely related diagonalized forms. However, unless self-connections are retained, the inferred eCM structures are severely affected, especially for the deCM, whose diagonals and block-like structures are not captured. These changes are fundamental and disproportionate to the small number of connections deleted. Deletion of fCM self-connections also implies widespread long-range differences in the corresponding effective connectivities and removes the ability to infer short-range net inhibitory connections at subgrid scales. (Note that these include the sub-mm Mexican-hat structure of very short range excitatory connections with inhibitory surround, plus longer excitatory connections out to  $\sim 2$  cm but within the same region of interest.)
- (iv) We have decomposed the fCM and teCM via eigenmode analysis, retaining self-connections and confirming that the first few eigenmodes suffice to reproduce the main features of  $C$  (Robinson et al. 2014, 2016) and  $T$ . We also showed that to accurately represent diagonal entries of the fCM, only around 20 eigenmodes are needed for NKI-Rockland data with  $N = 165$  regions of interest. This contrasts with the  $N(N + 1)/2 = 13695$  entries that are required in a conventional representation.
- (v) The deCM  $A$  has more large entries in its diagonal representation than do  $C$  and  $T$ , because  $|\lambda_j|$  increases at large  $j$ , but  $A$  only summarizes direct connections between RoIs, whereas  $T$  is the quantity of direct dynamical relevance.

In summary, we have investigated the importance of retaining self-connections when analyzing the experimental fCM and the corresponding inferred eCMs. We have shown that the self-connections in the fCM play essential mathematical and physical roles in the correspondence between fCMs and eCMs and thus cannot be deleted from experimental data if one wishes to have mutual consistency between

these quantities; this is contrary to their usual removal. We note that this does not make a difference in case-control comparisons, where the self-connection terms cancel, but inferred relationships between fCMs and eCMs in existing studies in which diagonal fCM elements have been deleted should be reviewed. This underlines the need to respect and preserve the formal properties of quantities being measured (in this case, the covariance matrix) to avoid the potential for mathematically invalid analysis steps. We have also shown that the methods discussed here deliver compact spectral representations of CMs and verify that these can greatly simplify treatment of brain connectivity, thus promising new and more tractable analyses and insights. These latter results have been illustrated in familiar CM form to emphasize their utility.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aquino KM, Schira MM, Robinson PA, Drysdale RM, Breakspear M (2012) Hemodynamic traveling waves in human visual cortex. *PLoS Comput Biol* 8:e1002435
- Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A (2000) In vivo fiber tractography using DT-MRI data. *Magn Reson Med* 44:625–632
- Brown JA, Rudie JD, Bandrowski A, Van Horn JD, Bookheimer SY (2012) The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. *Front Neuroinform* 6:28
- Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:189–198
- Fornito A, Zalesky A, Bullmore ET (2016) Fundamentals of brain network analysis. Academic Press, London
- Friston KJ (2011) Functional and effective connectivity: a review. *Brain Connect* 1:13–36
- Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. *NeuroImage* 30:1077–1087
- Gabay NC, Robinson PA (2017) Cortical geometry as a determinant of brain activity eigenmodes: neural field analysis. *Phys Rev E* 96:032413
- Galán RF (2008) On how network architecture determines the dominant patterns of spontaneous neural activity. *PLoS ONE* 3:e2148
- Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ et al (2008) Mapping the structural core of human cerebral cortex. *PLoS Biol* 6:e159
- Henderson JA, Robinson PA (2011) Geometric effects on complex network structure in the cortex. *Phys Rev Lett* 107:018102
- Hofer S, Frahm J (2006) Topography of the human corpus callosum revisited—comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage* 32:989–994
- Jezzard P, Matthews PM, Smith SM (2001) Functional magnetic resonance imaging: an introduction to methods. Oxford University Press, Oxford
- Mehta-Pandey G, Robinson PA, Henderson JA, Aquino KM, Sarkar S (2017) Inference of direct and multistep effective connectivities from functional connectivity of the brain and of relationships to cortical geometry. *J Neurosci Methods* 283:42–54
- Nooner KB, Colcombe SJ, Tobe RH, Mennes M, Benedict MM, Moreno AL et al (2012) The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front Neurosci* 6:152
- Pinotsis D, Robinson PA, Friston K (2014) Neural masses and fields: modeling the dynamics of brain activity. *Front Comput Neurosci* 8:149
- Poldrack RA (2007) Region of interest analysis for fMRI. *Soc Cognit Affect Neurosci* 2:67–70
- Robinson PA (2012) Interrelating anatomical, effective, and functional brain connectivity using propagators and neural field theory. *Phys Rev E* 85:011912
- Robinson PA (2017) The balanced and introspective brain. *J R Soc Interface* 14:20160994
- Robinson PA (2019) Physical brain connectomics. *Phys Rev E* 99:012421
- Robinson PA, Sarkar S, Pandey GM, Henderson JA (2014) Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. *Phys Rev E* 90:012707
- Robinson PA, Zhao X, Aquino KM, Griffiths JD, Sarkar S, Pandey GM (2016) Eigenmodes of brain activity: neural field theory predictions and comparison with experiment. *NeuroImage* 42:79–98
- Sporns O (2010) Networks of the brain. MIT Press, Cambridge

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