Review

The future of FMRI connectivity

Stephen M. Smith
OXFORD UNIVERSITY CENTRE FOR FUNCTIONAL MRI OF THE BRAIN (FMRIB), UK

ABSTRACT

"FMRI connectivity" encompasses many areas of research, including resting-state networks, biophysical modelling of task-FMRI data and bottom-up simulation of multiple individual neurons interacting with each other. In this brief paper I discuss several outstanding areas that I believe will see exciting developments in the next few years, in particular concentrating on how I think the currently separate approaches will increasingly need to take advantage of each others' respective complementarities.

© 2012 Elsevier Inc. All rights reserved.

Contents

Introduction - brief review of concepts ................................................... 1257
Network modelling via nodes and edges; functional vs. effective connectivity. .......................................................... 1258
Spatial patterns of connectivity ............................................................ 1258
Connectivity modelling from multiple subjects ...................................... 1259
Model complexity ............................................................................ 1259
Bottom-up modelling ...................................................................... 1259
Graph theory ................................................................................. 1259
FMRI network modelling methods .................................................... 1260
Causality .......................................................................................... 1261
Patterns of conditional independence; observational vs. interventional studies ......................................................... 1261
Dynamic biological Bayesian models .................................................. 1262
Future ............................................................................................. 1263
Nonlinearities and temporal nonstationarities ...................................... 1263
Other issues... and conclusions .......................................................... 1264
Acknowledgments .......................................................................... 1265
References ..................................................................................... 1265

Introduction - brief review of concepts

Much of the neuroimaging community is shifting emphasis from blobology (functional specialisation/segregation) towards connectology (functional integration). While non-MRI modalities such as MEG are showing increasing promise in aiding this, FMRI continues to be a major tool in the quest for inferring "the" brain network model. A large amount of FMRI connectivity research centres around the development, validation and interpretation of new analysis methods, with a plethora of different approaches appearing, each with its own limitations and promise, and each potentially asking a different question regarding neural connectivity.

A general meaning ascribed to the phrase FMRI connectivity is the study of interactions between distinct brain areas using FMRI. As soon as this is made explicit, it becomes obvious that FMRI connectivity covers many areas, including resting-state networks, DCM, graph theory and the connectome — and yet there is sufficient (and increasing) interplay...
between such areas, that hopefully it can make sense to discuss several of them together in the general context of “connectivity”. This brief paper does not give a detailed retrospective review, as it is supposed to be a future-looking paper, and there are several relevant reviews by other authors in this special issue, in addition to excellent recent reviews elsewhere (e.g., (Friston, 2011)). In the remainder of this section I recapitulate some basic network modelling concepts, and in the rest of the paper discuss several outstanding areas for future work that I personally think are important and exciting.

**Network modelling via nodes and edges: functional vs. effective connectivity**

The mapping of the brain's networks often starts by identifying a set of “nodes”, and then attempts to estimate the set of connections or “edges” between these nodes, based on an analysis of the FMRI timeseries associated with the nodes. In some cases, the directionality of these connections is estimated, in an attempt to show how information flows through the network.

There are many ways to define network nodes from FMRI; nodes are often defined as spatial regions of interest, for example, as obtained from task-FMRI activation or from brain atlases. Alternatively, parcellation via a data-driven clustering of the FMRI data itself (e.g., hierarchical clustering or independent component analysis) can be run to define clusters or components (spatial maps with associated timecourses), which can be considered network nodes, although the extent to which this makes sense depends on the number of components extracted (e.g., the ICA dimensionality). If a low number of components is estimated (Kiviniemi et al., 2003), then it makes more sense to think of each component itself as a “network”. This will often include several non-contiguous regions, all having the same timescourse (according to the model), and hence with-in-component network analysis is not possible without further processing, such as splitting the components and re-estimating each resulting node's timeseries. Furthermore, between-component network analysis is quite possibly not very meaningful, as each component will in itself constitute a gross, complex functional system. However, if a higher number of components is estimated (Kiviniemi et al., 2009), these are more likely to be smaller, isolated regions (parcels), which can more sensibly be then considered as nodes for use in network analysis.

Once the nodes are defined, each has its own associated timecourse (e.g., the average timeseries from all voxels within the node). These are then used to estimate the connections (edges) between nodes — in general, the more similar the timescourses are between any given pair of nodes, the more likely it is that there is a functional connection between those nodes. Of course, correlation (between two nodes' timeseries) does not imply either causality (in itself it tells one nothing about the direction of information flow), or even whether the functional connection between two nodes is direct (there may be a third node “in-between” the two under consideration, or a third node may be feeding into the two, without a direct, or even causally-indirect, connection existing between them). This distinction between simple correlation and trying to estimate the underlying, direct, causal connections (sometimes referred to as the distinction between functional and effective connectivity respectively (Friston, 1994)) is very important if one cares about the underlying biological network. For example, in a 3-node network $A \rightarrow B \rightarrow C$, with distinct external (e.g., sensory) inputs feeding into all nodes, all three nodes' timeseries will be correlated with each other, so the “network estimation method” of simple correlation will incorrectly estimate a triangular network. However, another simple estimation method, partial correlation, can correctly estimate the direct connections (though not their directionals); this works by taking each pair of timeseries in turn, and regressing out the third from each of the two timeseries in question, before estimating the correlation between the two. If $B$ is regressed out of $A$ and $C$, there will no longer be any correlation between $A$ and $C$, and hence the spurious third edge of the network ($A \rightarrow C$) is correctly eliminated.

The question of directionality is also often of interest, but in general is harder to estimate than whether a connection exists or not (Smith et al., 2011). For example, many methods, such as the two mentioned earlier (full correlation and partial correlation) give no directional information at all. The methods that do attempt to estimate directionality fall into a few general classes. One class is lag-based (more generally, multivariate autoregressive modelling (Valdes-Sosa et al., 2011)), the most common example being Granger causality (Granger, 1969). Here it is assumed that if one timeseries looks like a time-shifted version of the other, then the one with temporal precedence caused the other, giving an estimation of connection directionality. A second class (e.g., Bayes nets and structural equation modelling) is based on the idea of conditional independence, and (for FMRI) often starts just by estimating the (zero-lag) covariance matrix between all nodes' timeseries (hence such methods are based on the same raw measure of connectivity as correlation-based approaches — but attempt to go further in utilising this matrix to draw more complex inferences about the network). Such methods may look at the probability of pairs of variables conditional on sets of other variables; for example, Bayes net methods (Ramsey et al., 2010) in general estimate directionality by first orienting "unshielded colliders" (paths of the form $A \rightarrow B \leftarrow C$, where a node is fed into by the others) and then drawing inferences based on algorithm-specific assumptions regarding what further orientations are implied by these colliders. A third class of methods utilises higher order statistics than just the covariance; for example, Patel's pairwise conditional probability approach (Patel et al., 2006) looks at the probability of $A$ given $B$, and $B$ given $A$ (under a non-Gaussian data distribution model), with asymmetry in these probabilities being interpreted as indicating causality.

**Spatial patterns of connectivity**

There is also a significant amount of FMRI connectivity research that is not working within the nodes+edges network framework. The most obvious area is the seed-based analysis of resting-FMRI data, where one might take a single voxel's resting timescourse and regress all other voxels' timescourses against this, resulting in a spatial map of correlation scores (Biswal et al., 1995). Such spatial (voxelwise) investigations of functional connectivity can have some advantages when compared with a nodes+edges analysis, for example, if a connectivity difference between groups of subjects is one of shape rather than correlation strength. A similar point can be made with respect to the dual-regression approach (Beckmann et al., 2009); here, a set of group spatial maps (e.g., from a low-dimensional group-
ICA) is effectively treated as a set of multiple extended seed regions, and hence regressed into individual datasets to obtain the “seed” time-courses, which are then regressed into the same datasets to obtain subject-specific maps correlating with those seeds. Although this has some important differences to standard seed-based correlation,\(^7\) the final voxelwise cross-subject comparisons of the spatial maps is related; because voxelwise (spatial) tests are being carried out, such analysis can show changes in functional connectivity of spatial shape, and not just strength.

Such investigations of spatial patterns of connectivity are (relatively) free from the spatial over-simplifications imposed by a strict parcellation model of the brain. For example, (van den Heuvel and Hulshoff Pol, 2010) shows striking results from seed-based analysis of the primary motor regions; as the seed point moves continuously up the motor strip, the corresponding point of maximum correlation on the contralateral hemisphere also moves up in a continuous (as opposed to parcellated) manner. In some parts of the brain, the connectivity “gradient” (how much the connectivity pattern varies from one seed-point in the brain to a neighbouring seed-point) is much higher than in others. Given this, it must be the case that nodes + edges is an inaccurate model (implying some loss of “correctness” in the derived network connectivity), and indeed that some of the parcellation “boundaries” may be more arbitrarily placed than others. Hence, one might choose to define parcels (nodes) in terms of their boundaries (Cohen et al., 2008) rather than their centroids (as is effectively the case with methods such as ICA); this may ameliorate the effect of a continuum of connectivity gradients, but one will still need to think hard about how to “average” the connectivity patterns found when seeding from all voxels within such parcels.

**Connectivity modelling from multiple subjects**

Finally, there is still much connectivity-related work remaining to be done with respect to the analysis of multiple subjects’ data. A major challenge here is to achieve the best within-subject modelling while being able to robustly achieve correspondence across subjects. For example, if one doesn’t have an equivalent functional parcellation in all subjects, the nodes and edges don’t mean the same thing in all subjects, and hence, how can one combine any network modelling across them? Likewise, in an ICA decomposition, if one carries out subject-specific ICA (which is good from the point of view of modelling-out session-specific artefacts), how can one robustly guarantee that the components found are compatible across subjects? There is ongoing disagreement on this (hugely important) question, but my personal feeling is that a “core” parcellation will have to be carried out first at the group-level, to enforce correspondence from the beginning, and then, using that as a constraint, the within-subject parcellation can be refined/revisited (including potentially the modelling of “outlier” parcels/components which are missing/ additional compared with the group). There is also the question of how to carry out group-wise network modelling (assuming that the parcellation correspondence question can be resolved); I don’t think this is as difficult/fundamental a short-term problem as that of multiple-subject parcellation, but this will definitely be an exciting area. For examples of nice initial work, see (Varoquaux et al., 2010; Ramsey et al., 2011).

**Model complexity**

The scope of network-related research almost falls onto a one-dimensional continuum that starts with neural-level simulations at one end, passes through network modelling methods that are applied to real FMRI data, and ends with the most abstract of the graph-theoretic summaries of a network matrix (Fig. 1). The various distinctions between the different levels are worth noting, as they relate to many of the respective strengths and weaknesses of different approaches, and also inform some thoughts about valuable future directions.

**Bottom-up modelling**

At the lowest modelling level, there is an increasing amount of exciting neural-network simulation work, some of which simulates networks of individual neurons, but most of which shows/assumes that groups of neurons can be treated as single units (Deco et al., 2008). While the majority of this work has to date been well-informed by empirical data regarding neuronal dynamics (e.g., via single-cell recordings), it has been so distant from FMRI data that it can be hard to be sure how well the simulated large-scale network behaviour relates to real FMRI data. Indeed, while such bottom-up modelling has been used to generate simulated FMRI data (and in some cases show that some characteristics of the resulting data match what is seen in reality (Honey et al., 2007)), this work is more likely to interact richly with MEG data before it can with FMRI data. This is reflected in the fact that such models (e.g., “neural mass” modelling) are used in the DCM-MEG forward model, but deemed too distant/detailed to be worth feeding into the DCM-FMRI forward model. One might go as far as to suggest that the limitations of FMRI (including in particular the limit on temporal resolution imposed by the haemodynamic blurring) may prove an insuperable barrier to ever relating FMRI data usefully to the most detailed bottom-up models; however, the potential benefits are too great to not pursue such approaches, and at the very least, we can look forward to combinations of FMRI and electrophysiological methods together being related to the most detailed neural models.

**Graph theory**

At the highest modelling level, defined here as being post-network-matrix-estimation, there is much (often relatively abstract) work generally termed graph theory (Rubinov and Sporns, 2010). This includes the study of network clustering and hierarchies, the study of network hubs (nodes or clusters that are particularly highly connected to other parts of the network), and deriving network summary statistical measures such as small-worldness (looking at how the clustering acts over multiple scales), and measures of general network efficiency. A lot of this work has utilised impressive mathematics, but has often seemed to me to be a little too distant from real data; for example, it has sometimes seemed that the focus on advanced graph-theoretic modelling has come at the expense of doing a thorough job of estimating an accurate network matrix in the first place (to feed into the graph theory). One practical danger is the use of inappropriate node definition (Smith et al., 2011; Craddock et al., in press), where a gross structural atlas-based parcellation may not correspond at all well to real functional boundaries in the data, resulting in network matrices that are probably not very meaningful (and hence neither is the further graph theory applied to the network matrix). A second problematic aspect of such work is that it is generally the (often thresholded \(^7\)) correlation matrix that is fed into the graph theory, and so no attempt has been made to estimate only the direct network connections for analysis by the graph theory.\(^6\)

---

\(^7\) The choice of threshold itself being a practical and interpretive problem.

\(^6\) However, I think this criticism is possibly ameliorated in the cases where graph theory is primarily being used to identify modules (functional clusters of nodes); this is because the full correlation (functional connectivity), by definition, is describing which nodes are functionally linked (whether directly connected or not), so this use of the correlation matrix seems quite reasonable. It is also the case that the use of graph theory when applied to structural (e.g., diffusion MRI derived) connectivity matrices does not suffer from this problem to nearly the same extent.
This is particularly problematic where the measure (derived via graph theory) is supposed to relate to "path length", or where simulated "lesioning" is supposed to relate to real connections!

A more general danger is that graph theory is able to abstract the network matrix to such a high degree (e.g., summarising an entire study down to a single number representing overall network efficiency, or small-worldness), that one is very hard pressed to be confident that any change in this (e.g., between patients and controls) really reflects a change in the brain connectivity, as opposed to being driven by any one of a myriad of potential confounds (e.g., factors as basic as systematic group differences in head motion or heart rate). However, I don’t want to sound over-pessimistic; I do believe that the future holds great things for this area of work, particularly as more accurate and meaningful network matrices are fed into it. One area that I think will be particularly exciting will be to see how graph theory can help us decide what clusters of nodes are functionally meaningful, including addressing questions of how to deal with overlap between clusters, and how best to define hierarchies of clusters that are useful functional descriptions. Additionally, I would hope that such work will feed back down, helping direct us in our search for better methods for estimating the network matrix in the first place.

**FMRI network modelling methods**

Finally there is the middle-ground, containing the majority of the brain connectivity work that has most closely related to practical network modelling from real FMRI data. Again, here we have a continuum with respect to many distinct factors. At one extreme we have highly-complex models of effective connectivity with many free parameters, each representing a biological or physical concept, such as neuronal activity and (separately) the haemodynamic response to neural activity; this model is “fit” to data ideally using probabilistic (e.g., Bayesian) methods. The obvious example of such a method is DCM (Friston et al., 2003). Not only is the model complex, but so is the inference method (for all its advantages, Bayesian inference is considerably more complex than simple, e.g., “point estimate”, model fitting). At the other extreme we have mathematically very simple methods, such as correlation (between node timeseries). The simpler methods are in general more “robust” (in fitting the model to the data), and faster to compute, than the complex methods. Related to this, and the fact that they have many fewer parameters to estimate, the simpler methods can handle a much larger number of network nodes than the more complex methods. Additionally, the simpler methods do not require the scope of possible network models to be pre-specified or constrained, i.e., they are computationally practical for attempting network search or discovery, which is more difficult for the most complex methods, that have traditionally not been able to search over all possible network matrices.

However there is a serious downside to the simpler methods (at the functional connectivity end of the spectrum); they are really just descriptions of the data, rather than relating to underlying, interpretable network parameters. For example, as mentioned earlier, correlation tells one nothing quantitative about causality or network connection strengths, and as a result is more vulnerable to being affected by confounds in the data. Correlation is affected by factors such as noise level, neural input amplitude and does not just reflect local connection strength, but is also affected by distant changes in brain function (Friston, 2011). Moving towards the more complex end of the modelling spectrum, with methods such as SEM (structural

---

9 For example, one major advantage of biophysically-based Bayesian modelling (such as DCM) over non-biological point-estimate approaches is that the modelling ends up “knowing” which of the biological parameters are (relatively) unambiguously identifiable from the data.

10 To be fair to the more complex methods, these have often not claimed to be able to carry out robust network “discovery”. Indeed, strongly hypothesis-driven carefully thought-out connectivity experimentation often compares rather favourably against some of the more “fishing-trip” resting-FMRI experiments that end up with sometimes rather peculiar/non-refutable results!
Causality

According to the clear and brief overview of causality given on Wikipedia,13 “Philosophers ... have defined causation in terms of a cause preceding and increasing the probability of the effect”. The italics are mine, to emphasise that these two (both quite sensible) measures of causality can potentially be found quite independently of each other, from a given dataset. A natural urge for many neuroscientists is to look for causal structures in the brain; for example, a flashing light causes the eyes to send a signal to V1, which activates as a result, and this activation then causes V2 to activate. However, some may then point out that few (if any) brain connections are unidirectional, with feed-back/top-down connections generally sitting in parallel to feed-forward connections. Nevertheless, many are still interested in estimating at least the dominant direction of information flow for a given connection (or, in the case of models such as DCM, even attempting to estimate the forwards and backwards connection strengths separately).

Thus, with respect to the idea of the cause increasing the probability of the effect, we are most likely in practice aiming to find the “relative causality”; if the dominant flow of information is from A to B, then [the probability of B given A] is greater than the [probability of A given B], or \( P(B|A) - P(A|B) > 0 \). Interestingly, it was exactly this measure that was used in the pairwise directionality estimation approach “Patel’s r”, which was the most successful measure of directionality of all those tested in our recent simulation-based evaluation of network modelling methods (Patel et al., 2006; Smith et al., 2011).14 Conditional probabilities do not just have to be considered with respect to a pair of nodes at a time; the more sophisticated Bayes nets methods discussed in the following section can utilise the full (nodes \( \times \) nodes) dependence structure to try to do an even better job of estimating causality. For example, recent simulations15 suggest that the full covariance structure can be used, at least with multiple subjects’ datasets, to correctly infer directionality from FMRI data. Other recent work16 utilised non-Gaussianities (Shimizu et al., 2006) in the data to show even more robust estimation of (dominant) causality. A simple explanation of the use of non-Gaussianities invokes the central-limit theorem; if A and B each have their own inputs, and also A feeds into B, then B will be more Gaussian than A. Although such a simplistic approach would be biased by differential measurement noise, there are more sophisticated measures which reduce such sensitivity (Hyvärinen, 2010). However, there are still some major limitations to such methods, some of which are covered subsequently.

Finally, as mentioned at the start of this section, we can also look to temporal precedence to tell us about causality: if B happened after A, then A caused B. This is only useful if the relevant temporal information is available. This is a crucial point in the case of FMRI, where, unfortunately, the (generally unknown amounts of) haemodynamic blurring and delaying renders estimation of the neural temporal precedence unknown. Thus temporal lag is unlikely to be the best way to infer causality from FMRI data, and it is important to remember that temporal lag is not the only way by which we can infer causality (as discussed earlier and later).

*Patterns of conditional independence; observational vs. interventional studies*

Graphical causal models (Bayes nets) try to do the best they can to find the causal network structure, given the apparent probabilistic dependencies of the different nodes’ timeseries on each other; this might be done just via the covariances (i.e., assuming the data is Gaussian), or might try to utilise non-Gaussianities in the data to gain further information about the dependencies. Specific patterns of probabilistic dependencies can be used to infer aspects of the causal structure; however, in addition to the caveats listed below (e.g., with respect to hidden external inputs), it is generally the case that

---

14 It is possible that causality estimation when the brain is “at rest” (containing potentially many different functional processes mixed together, and hence estimated as the average “relative causality” over all possible spontaneous fluctuations, quite likely via a temporally stationary model) may be so different from what is seen during individual focussed tasks that it might be too much to hope that dominant causalities found from resting-FMRI will relate meaningfully to the route by which information flows around the “brain network” when triggered by external events; however I do hope that they will!
15 testing the IMAGES multi-subject Bayes net method (Ramsey et al., 2011)
16 (Hyvärinen, 2010) and the use of the LOFS method in (Ramsey et al., 2011), both, in fact, (directly and indirectly respectively) inspired by the Patel’s r results
for many patterns of dependencies, there is more than one possible causal network structure that could give rise to the data. The set of possible networks, which it is not possible to disambiguate between (given the data), is known as the Markov equivalence class. Related to this limitation, Bayes nets are in general only able to correctly identify networks that are directed acyclic graphs (DAGs), meaning that no networks with closed cycles of causality can be robustly estimated. Such limitations will provide ample scope for valuable research in coming years, given that the brain certainly contains causal cycles!

Experts in causal inference often draw a strong distinction between observational vs. interventional experimentation (for example, see the first half of Pearl (2009) by Judea Pearl, an accessible summary of concepts in causality). There is clearly a (not perfect) parallel between this distinction, and the distinction of resting-FMRI vs. task-FMRI. Resting-FMRI data is by definition purely observational (no external interventions) and task-FMRI must contain some external interventions. According to many causality experts, the conclusion that can be drawn from purely observational studies are, in theory, seriously limited, in terms of the confidence/robustness with which one can draw conclusions about causalities within the system being studied (in our case, the brain). The distinction largely comes down to the fact that observational data requires a greater number of assumptions to be made in order for causal inference to take place, and there is a good chance that these assumptions could be violated. One example confound is the presence of unseen external factors (e.g., hidden common causes) that generate apparent causalities between two viewed nodes; this is a greater problem for resting-FMRI than for task-FMRI. An example (which we have recently seen in real resting data) is where we have found spontaneous activity in the frontal eye fields “causing” activity in early visual areas, but of course we could not tell if this was due to FEF activity causing the eyes to move, which caused an external (hidden) change in what is then seen by the early visual areas, or whether we were seeing part of a “top-down” modulatory effect. A careful vision (task-FMRI) experiment would explicitly control for various confounding factors and aim to lead to more meaningful specific interpretations of the data. A related, more mundane, issue is that artefactual correlations in the data are more likely to interfere with resting-state connectivity estimations than they are to confound a task-based connectivity experiment (because, in resting-FMRI, there is no known timing information that can help to reject the artefactual effects).

A scenario that helps illustrate the difference between observational and interventional data, and also between the raw covariance pattern and the use that Bayes nets can make of it, is that of “Berkson’s paradox” (or “explaining away”, a special case of “selection bias”). Here we have the causal structure $A \rightarrow B \rightarrow C$, where $A$ and $C$ are not directly connected. If we actively hold $B$ constant, via experimental intervention, the lack of a causal link between $A$ and $C$ will be correctly observed in the data. However, things are different if we do not actively intervene, but try to account for (“condition on”) $B$ in simple analysis; in a partial correlation analysis (i.e., testing for conditional independence), by regressing $B$ out of $A$ and $C$, we induce a negative correlation between $A$ and $C$22 This is clearly problematic, but we can maybe hope to recognise this scenario by noting that the full correlation (testing for “marginal independence”) shows that $A$ and $C$ are uncorrelated. Finally, however, if we can correctly assume that there are no hidden causes in play,23 A Bayes net analysis of the full set of (marginal and conditional) probabilistic dependencies will be able to fully identify the correct causalities.24 Even if one really cannot be sure about the “causal” conclusions being derived from resting-FMRI experimentation, it still has a role in “discovery science”. For example, in (Steyvers et al., 2003), we are told that “In laying out his approach to scientific discovery, Mill [in 1874] noted that while only experiments can prove causation, pure observation still plays an important role as the natural guide for experimentation. Entering a new domain, scientists often do not know what questions are worth asking, and which experiments worth doing, until they observe a surprising phenomenon along with some other correlated events that might be potential causes. Like scientists, people might use observations primarily to form hypotheses and interventions primarily to test those hypotheses.” For example, resting-FMRI can suggest observations primarily to form hypotheses and interventions primarily to test those hypotheses. Furthermore, despite the concerns described earlier, it is generally the case that in practice people are more ambitious in the number of nodes that they extract and use in network modelling from resting-FMRI data than from task-FMRI experiments, adding (hopefully useful) further richness to the “discoveries” that can be made from resting-FMRI network analysis.

**Dynamic biological Bayesian models**

“Standard” Bayes nets are static (aka “instantaneous” or “zero-lag”) models, meaning that the data is reduced to summary-statistics (e.g., covariances) between timeseries, collapsing over all time, as opposed to dynamic methods which fit a temporal model directly to the entire (set of nodes’) timeseries. Dynamic models (e.g., dynamic Bayes nets or DCM) can in theory utilise both temporal lag information and the conditional dependencies, in order to infer causality. This might be expected to give such methods an advantage (making use of more,22 For example, PhD supervisors might only allow their students to attend the OHBM conference if they can pay their own costs, or if their work is good enough to be given an oral presentation (or both). Because the students who fail on both counts do not attend, the average correlation (conditional on going to the conference) between rich and clever students appears to be negative!

23 This is a very big “if!” However, there are variants of Bayes net approaches that do explicitly try to deal with even this issue (Zhang, 2008).

24 A further, related, example is given in Pearl (2009), which we include for the interest of the brave reader who is comfortable with this description of Berkson’s paradox, and wants more: “...the requirement of holding the mediating variables fixed [where mediating or intermediate refers to network nodes that sit between other nodes of interest in the network’s chain of causality] must be interpreted as (hypothetically) setting the intermediate variables to constants by physical intervention, not by analytical means such as selection, conditioning, or adjustment. [For example, by regressing an intermediate’s timeseries out of the others under consideration.] For example, it will not be sufficient to measure the association between gender (X) and hiring (Y) for a given level of qualification Z, because, by conditioning on the mediator Z, we may create spurious associations between X and Y even when there is no direct effect of X on Y... This can easily be illustrated in the model $X \rightarrow Z \rightarrow Y$; where X has no direct effect on Y. Physically holding Z constant would sustain the independence between X and Y, as can be seen by deleting all arrows entering Z. But if we were to condition on Z [e.g., carry out a partial correlation analysis on resting-FMRI data] a spurious association would be created through U (unobserved) that might be construed as a direct effect of X on Y.”
richer, aspects of the data). However, this is not necessarily the case. For example, temporal lag between two FMRI timeseries is a summation of neural lag and differential haemodynamic delays, and a fitted biological generative model is likely to report simply that these two effects cannot be disambiguated from each other given the data\textsuperscript{25}, hence little or no use will be made of the temporal lag in terms of inferring (neural) causality. In this case, all we are left with is the set of conditional probabilities (e.g., just the covariance structure of the data, in the case of Gaussian data/modelling), and any inference that can be drawn faces all of the limitations and dangers discussed earlier. It is possibly considerations such as this that have led some, rather harshly, to describe DCM as being little more than Bayesian SEM with HRF deconvolution!

Of course, this is not the whole story. The fact that the Bayesian inference in DCM will be able to work out for itself that temporal lag is probably not informative about neural causality is indeed valuable! If we are able to generate data where the lag could be usable (e.g., if we can pre-characterise the haemodynamics accurately in all regions being studied, and use that knowledge to feed into the DCM haemodynamic priors, and if we can generate high-quality short-TR FMRI data, and if we include neural lag in the DCM model), then DCM will immediately “know” that it can start making use of this aspect of the data to infer causality. This (model-fitting utilising knowledge about its own strengths and weaknesses) is something which methods based purely on data-descriptive auto-regressive modelling (e.g., Granger causality) simply cannot do. More generally, the fact that the parameters being fit in DCM are biophysical means that we learn about the quantities that we really care about, even if all that we are told is that a given parameter cannot be estimated with high precision from the data. Finally, to the extent that the biophysical model of DCM predicts nonlinearities/non-Gaussianities in the data, these aspects of the data can be exploited, as long as the biophysical model is sufficiently accurate; this may help get around some of the (causality inference) limitations described at the start of this section.

**Future**

So, what of the future of causality estimation? The initial results reported in (Ramsey et al., 2011) are indeed encouraging, suggesting that both Bayes nets and non-Gaussianity-based methods can be used to estimate causality. Although these approaches do not escape from all of the limitations discussed earlier, they do show that directionality is in principle estimable from FMRI data. Although it is generally assumed that in theory methods such as Bayes nets and approaches based on non-Gaussianities can only estimate acyclic networks, these results suggest that in practice this is not always the case.

However, concerns regarding observational studies are worth taking seriously. I believe. To be more confident of the directionality (or, even better, the parameters of the forwards and backwards connections between any two regions), we will likely require task studies (or even more “active” interventions such as TMS or pharmacological manipulation). Hence I would hope that in the near future the complementarities between resting-FMRI and task-FMRI (and between pragmatic, data-descriptive analyses and sophisticated, highly parameterised, biophysical modelling) will be more explicitly appreciated. Hopefully, this will cause greater use of the combination of rest and task, and of different analysis methods, taking the respective strengths from each; hopefully the rigour of the “scientific method” of hypothesis-based experimentation can merge with the larger-scale discovery methods.

**Nonlinearities and temporal nonstationarities**

The majority of FMRI connectivity research to date has worked with the (explicit or implicit) assumptions of stationarity and linearity. Stationarity in general means that some statistic or model parameter of interest is non-changing, and in this context is generally used to mean that some measure of connectivity (e.g., correlation between two regions, or connection strength parameter in a DCM model) is not changing over time. Linearity, in this context, might be referring to the output of a node being a linear combination of its inputs, or to the haemodynamic response being a linear function of the neural activity (e.g., if the activity is doubled, then so is the haemodynamic response). As data quality and analysis methods improve, we become more able to see and model nonstationarities and nonlinearities, and I expect such improvements in modelling sophistication to be a major growth area in the next few years.

It can be somewhat confusing that these terms are often used with some vagueness, and this is not helped by the fact that a given dataset may appear linear and stationary from the point of view of one modelling approach, while appearing to be nonlinear and nonstationary from another. For example, DCM may happily be able to model a 3-node task-FMRI dataset where the activity at node C modulates the strength of the (otherwise linear and unchanging) connection between nodes A and B (Stephan et al., 2008). Thus we have a system which DCM considers well-modelled, and linear/stationary in all its components, but which a simple correlation analysis between nodes A and B will see as both nonlinear (B is not apparently a linear function of A) and nonstationary (the connection strength appears to vary over time). One can justifiably see this as a weakness of the less quantitative/complete analysis methods such as correlation; however, even with respect to the most sophisticated of methods, one can easily imagine scenarios where their assumptions (of linearity/stationarity) are unjustified—for example, if the strength of the modulatory influence is varying over time, controlled by some unseen factor.

Note that although some modelling methods (e.g., DCM and dynamic Bayes nets) call themselves “dynamic”, that generally refers to modelling timeseries dynamics, and does not imply that their underlying parameters (in particular strengths of connections) are allowed to vary freely over time.\textsuperscript{26} However, there is recent related work which attempts to expand the range of modellable scenarios that under previous models would appear to be nonstationary. For example, see Jason Smith’s work on a DCM-like approach that models the data with a temporally-varying alternation between multiple different network models (Smith et al., 2010).

Unfortunately, it is still the case that such sophisticated, highly parameterised models are not yet applicable to carrying out network discovery on large numbers of nodes, bringing us back to the practical tradeoff between the complexity/interpretability of the modelling, and the number of nodes that can be handled. This is one reason why the simplest methods (such as correlation) are still widely used, particularly for resting-FMRI datasets. As a result, some researchers who are interested in nonstationarities are starting to look into adaptation of these simplest methods, e.g., using sliding-window correlation rather than just estimating one correlation value across the whole timeseries.

In just the last couple of years the resting-FMRI community has started looking into nonstationarities, so far, in order to better understand the dynamics of resting-state “anticorrelations” (such as seen between the “default mode network” and the “task positive network”), but there is no reason to restrict the study of nonstationarities to

\textsuperscript{25} The two parameters, which are very similar to each other in terms of their effect on the (apparent lag in the) data, will each have a very wide (marginal) posterior distribution; the uncertainty on each is very high, unless one parameter is already known.

\textsuperscript{26} A nice introduction to (static and dynamic) Bayes net models http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html puts this well: “Note that ‘temporal Bayesian network’ would be a better name than ‘dynamic Bayesian network’, since it is assumed that the model structure does not change, but the term DBN has become entrenched.”
negative correlations. One of the first groups to inspire this was the experimental work of Daniela Popa and Denis Paré, where LFP and unit recordings were acquired in functionally anticorrelated regions (Popa et al., 2009). They showed that there were distinct periods when two regions were positively correlated, and others when they were anticorrelated—a simple example of nonstationarity. One important future issue for such work will be to try to distinguish between such changes in correlation being due to different nodes being part of multiple overlapping functional networks vs. the internal connections within any given network being non-constant, and it is not apparent which of these factors dominates in practice in (apparent nonstationarities seen in) typical resting-FMRI data. Indeed, there are many other sources of apparent variability in correlation, e.g., as seen by sliding window correlation. Most of the criticisms (discussed earlier) of the non-quantitative nature of correlation (as a connectivity measure) by Friston are equally valid criticisms of sliding-window correlation; for example, changing noise level or changing level of neural activity will cause an apparent nonstationarity in correlation. In addition, it is easy to end up with a suboptimal implementation, such as applying a sliding-window length that does not encompass (at least) several cycles of the resting fluctuations, a particular problem, given that non-highpass-filtered data will have the dominant fluctuation power at quite low frequencies ($\approx 0.015$ Hz).27 If only a fraction of a cycle is seen within the sliding window, then the apparent correlation is expected to appear to fluctuate wildly over time, even if the network structure is stationary.

A richer, related approach is to look at coherence (and relative phase) between two timeseries over a range of frequencies and associated window lengths (matched to the frequencies to avoid the problem just discussed28), typically achieved via wavelet transforms. Each node’s timeseries is transformed into a set of new timeseries, each one estimating a bandpassed version of the data with a different centre frequency. Then two nodes’ wavelet decompositions can be compared (as a 2D function of time and frequency), with a range of relative measures estimable, such as relative phase (are the nodes temporally correlated or anticorrelated?) and coherence (do they both have significant power at this frequency at this point in time, with locally constant relative phase?). Early work on this was carried out by Catie Chang and Gary Glover, showing from FMRI data that the default-mode anticorrelations were varying over time (Chang and Glover, 2010).

As mentioned earlier, one potential source of apparent nonstationarity in correlation is where different nodes are part of multiple overlapping functional networks. This is a subject we are currently investigating, by seeking to identify independent networks, as opposed to different networks being spatially independent (in the latter case the working assumption is that the networks should be spatially largely non-overlapping). However, the temporal richness in the data required for such analyses is very challenging given typical FMRI acquisitions, and hence we are working with short-TR accelerated EPI data (Feinberg et al., 2010) to identify temporally independent modes of brain activity. We are hopeful that in the near future we will be able to report results showing some quite distinct functional architecture from that currently seen (e.g., with low-dimensional spatial ICA or seed-based correlations).

Finally, outside the general network model of nodes and edges, and not estimable by simple “outer product” models of resting-state networks (such as ICA), there is some initial evidence of nonstationarities appearing in the form of spatiotemporal patterns of spontaneous activity that propagate across the brain and may explain at least some of what we see as nonstationarity in resting-state networks.

The group of Shella Keilholz has a good deal of data from animals and humans that suggest that such patterns are repeatedly found, i.e., are not just random fluctuations across space and time (Majee et al., 2010). If this work is indeed showing spontaneous neural spatiotemporal processes, and if these are a significant component of resting-state data, we will need to think hard about even the most common, basic current methods of analysis!

**Other issues... and conclusions**

This paper was supposed to concentrate on FMRI connectivity, and so I have barely mentioned other modalities, but it is clear that combining different experimental techniques will be crucial in the future of brain connectivity research—quite possibly a lot more important than the primarily-FMRI areas that I have discussed. The many modalities that are able to probe neural processes more directly than can FMRI include non-invasive modalities such as MEG and EEG, and many different invasive electrophysiological techniques such as the study of local field potentials, and single-unit recordings. The improved temporal resolution and neural interpretability of many of these other modalities will have a huge impact on our ability to look at causalities, nonstationarities and modulatory effects. It seems that experiments using simultaneous-FMRI + EEG and non-simultaneous–FMRI + MEG have been somewhat slow to take off, partly because of the difficulty of determining which (of the many) sources of signal in the EEG/MEG data really relate to what we know of as resting-state networks in the FMRI data. However such work is moving forwards, and in particular, I think there will be quite an explosion of resting-MEG experimentation in the next few years; for example, (de Pasquale et al., 2010) identified the default mode and task positive networks via resting-MEG data, and, more recently, (Brookes et al., 2011) found 8 different resting-state networks in resting-MEG data that were spatially an excellent match to networks previously shown in resting-FMRI. Other related areas of research29 are studies of (mostly) low-frequency oscillations in rest vs. sleep vs. anaesthesia, in humans and animals, using a range of electrophysiological modalities, such as “DC-EEG” (meaning very low frequency, $<1$ Hz EEG). There are suggestions that there could be overlap between such work (including travelling electrical wave studies/slow cortical potentials (Ridner et al., 2011; He et al., 2008)) and what we know of as resting-state FMRI networks, but I’m not aware of any work yet that has unambiguously linked such things together.

There is also a wide range of interventional techniques, including those that are “external” (TMS, TDCS, etc.) and others that are “internal” (optogenetics, direct electrical stimulation, etc.); these further expand the range of experiments and questions that we can ask, and may help ameliorate the limitations of purely “observational” data. While it is clear that FMRI will not be the only methodology used to study brain connectivity in the future, it remains the case that there is no other modality that is close to what FMRI can give us with respect to its specific combination of non-invasiveness, whole-brain coverage and spatial resolution.

Of course the future also holds much promise in terms of the other MRI modalities. An obvious complement to FMRI connectivity is diffusion-based connectivity, which can give robust and detailed estimation of structural connectivity. One might over-simplify the complementarities between diffusion-based structural connectivity and FMRI-based functional/effective connectivity by saying that the former suffers primarily from false negatives while the latter suffers from false positives (in estimating the “connectome”); with respect to the problems of distinguishing direct from indirect connections

27 There is some recent evidence (Niazy et al., 2011) that good quality resting-FMRI data can allow resting-state networks to be estimated up to at least 0.2 Hz, hence aggressive highpass filtering could hopefully ameliorate this problem.

28 and also, ideally using methods such as Monte Carlo simulations to identify which correlations/coherences are stronger than chance

29 of which I am embarrassingly ignorant, and hence can only refer to vaguely, but direct the interested reader to a recent book put together following an excellent low-frequency oscillations workshop in Amsterdam (Van Someren, 2011).
from FMRI data, we can hope that the structural connections can serve as valuable priors (or aim for a more integrated analysis of true connectivity). Indeed, resting-FMRI and diffusion MRI are the two primary modalities\(^{30}\) mandated to be used in the NIH Human Connectome Project; this began in 2010, and should produce the most detailed, large-scale \textit{in vivo} whole-brain connectivity mapping achieved to date. The HCP will generate leading-edge quality data from long imaging sessions with over a thousand subjects. It will be interesting to see how other ventures,\(^{31}\) that seek to bring together even larger numbers of subjects (and covering a wider range of subject groups such as different pathologies, but subject to the caveat of having greater heterogeneity of scanning parameters and lower overall data quality), will complement studies with smaller numbers of subjects and higher-quality data. Hopefully the larger, heterogeneous databases of connectivity datasets, while not supporting the most sophisticated analysis techniques, will complement studies such as HCP by being able to find gross imaging phenotypes and carrying out very-large-N subject-pathology correlations. In addition to FMRI and diffusion MRI, there is increasing evidence that even structural MRI (e.g., acquiring a single T1-weighted image per subject) can tell us about functional networks, through covariance analysis (Seeley et al., 2009), the idea being that functionally-connected regions covary with each other across subjects in terms of some structural characteristics (such as cortical thickness).

FMRI itself is by no means at the limits of its technical abilities. Raising the field strength continues to be of increasing value — possibly more so than expected, for resting-FMRI, given that the occasional mantra of “you don’t get what you hope for when you increase field strength” does not take into account that a lot of the increased “physiological noise” is the very resting-FMRI spontaneous fluctuations that we are now using as signal! Indeed, resting and task-FMRI at 7 T is already seeing spectacular gains in spatial resolution and/or effective SNR. Additionally, several groups are producing exciting new work on accelerated FMRI, with the ability to achieve sub-second whole-brain imaging (Feinberg et al., 2010). Such short-TR FMRI can, dependent on the analysis being carried out, give much greater gains in effective SNR than might initially be predicted — for example, when using methods\(^{32}\) which are dependent on high temporal degrees-of-freedom, or applying artefact modelling approaches that benefit from dense temporal sampling. Additionally, we should be able to do much better with analyses looking at non-Gaussianians, non-linearities and nonstationarities, when moving to faster-sampled FMRI data.

Finally, I have not yet mentioned \textit{clinical applications of FMRI connectivity}, but this is clearly going to be a huge growth area. There are already many papers linking \textit{changes} in connectivity to different diseases (Filippini et al., 2009), and resting-FMRI (and FMRI connectivity in general) should hopefully become a powerful clinical marker, albeit with the interpretive caveats listed earlier regarding the quantitative and biological issues associated with the simpler functional connectivity measures (which can become even more problematic when disease is associated with changes to neurovascular coupling, either as a direct result of the disease, or due to pharmacological interventions). Ultimately, in order to maximise the \textit{interpretability} of changes seen (across different clinical or cognitive conditions), we will need to move from \textit{functional} towards \textit{effective} connectivity modelling.\(^{33}\) One important advance that will help bring even greater benefit to the use of connectivity measures in clinical applications will be the use of \textit{multivariate} classifiers, rather than just simple univariate tests (Craddock et al., 2009). In addition to the growth of the more sophisticated, quantitative and biologically interpretable modelling methods (Brodersen et al., 2011), we will see FMRI connectivity become not just a powerful clinical marker, but a tool for investigating disease \textit{mechanism}. For clinical and non-clinical investigation of brain structure, function, development and pathologies, \textit{FMRI connectivity} will remain a powerful, sensitive non-invasive tool, and over the coming years I see huge potential for further growth, in terms of both the underlying technical and modelling challenges, and in its applications.

\textbf{Acknowledgments}

I am extremely grateful to Christian Beckmann and Mark Woolrich for (patiently!) teaching me about FMRI connectivity over the years. I am also very grateful, for useful discussions and/or helpful comments on this paper, to: Faysal Ahmad, Christian Beckmann, Karl Friston, Clark Glymour, Aapo McIntosh, Karla Miller, Tom Nichols, Olaf Sporns and Mark Woolrich.

\textbf{References}


\(^{30}\) with other modalities to include task-FMRI and MEG.

\(^{31}\) For example the “1000 functional connectomes” project (aka “KFC”) (Biswal et al., 2010), created by Mike Milham.

\(^{32}\) such as parcellation through high-dimensional ICA, or when using partial correlation or SEM applied to large numbers of nodes

\(^{33}\) Note that the study of changes in cognitive state has been an active and successful application of methods such as SEM for many years.


