Chapter 10

# Multi-Scale Information, Network, Causality, and Dynamics: Mathematical Computation and Bayesian Inference to Cognitive Neuroscience and Aging

Michelle Yongmei Wang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55262

## 1. Introduction

The human brain is estimated to contain 100 billion or so neurons and 10 thousand times as many connections. Neurons never function in isolation: each of them is connected to 10, 000 others and they interact extensively every millisecond. Brain cells are organized into neural circuits often in a dynamic way, processing specific types of information and providing the foundation of perception, cognition, and behavior. Brain anatomy and activity can be described at various levels of resolution and are organized on a hierarchy of scales, ranging from molecules to organisms and spanning 10 and 15 orders of magnitude in space and time, respectively. Different dynamic processes on local and global scales generate multiple levels of segregation and integration, and lead to spatially distinct patterns of coherence. At each scale, neural dynamics is determined by processes at the same scale, as well as smaller and larger scales, with no scale being privileged over others. These scales interact with each other and are mutually dependent; the coordinated action yields overall functional properties of cells and organisms.

An ultimate goal of neuroscience is to understand the brain's driving forces and organizational principles, and how the nervous systems function together to generate behavior. This raises a challenge issue for researchers in the neuroscience community: integrate the diverse knowledge derived from multiple levels of analyses into a coherent understanding of brain structure and function. The accelerating availability of neuroscience data is placing a huge need on mining and modeling methods. These data are generated at different description resolutions, for example, from neuron spike trains to electroencephalogram (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). A key theme in modern

neuroscience is to move from localization of function to characterization of brain networks; mathematical approaches aiming at extracting directed causal connectivity from neural or neuroimaging signals are increasingly in demand. Despite differences in spatiotemporal scales of the brain signals, the data analysis and modeling share some fundamental computation strategies.

Among the diverse computational methods, probabilistic modeling and Bayesian inference play a significant role, and can contribute to neuroscience from different perspectives. Bayesian approaches can be used to analyze or decode brain signals such as spike trains and structural and functional neuroimaging data. Normative predictions can be made regarding how an ideal perceptual system integrate prior knowledge with sensory observations, and thus enable principled interpretations of data from behavioral and psychological experiments. Moreover, algorithms for Bayesian estimation could provide mechanistic interpretations of neural circuits and cognition in the brain. In addition, better understanding of the brain's computational mechanisms would have a synergistic impact on developing novel algorithms in Bayesian computation, resulting in new technologies and applications.

This chapter reviews and categorizes varieties of mathematical and statistical approaches for measuring and estimating information, networks, causality and dynamics in the multi-scale brain. Specifically, in Section 3, we introduce the fundamentals in information theory and the extended concepts and metrics for describing information processing in the brain, with validity and applications demonstrated on neural signals from multiple scales and aging research. Bayesian inference for neuroimaging data analysis, and cognition modeling of observations from psychological and behavioral experiments as well as the corresponding neural/neuronal underpinnings are provided in Section 4. Graphical models, Bayesian and dynamic Bayesian networks, and some new development, together with their applications in detecting causal connectivity and longitudinal morphological changes are presented in Section 5. We illustrate the attractor dynamics and the associated interpretations for aging brain in Section 6. Conclusions and future directions are given in Section 7.

## 2. Neuroscience data/signals and brain connectivity

### 2.1. Recording and imaging techniques at multiple scales

An important breakthrough regarding neuronal activity and neurotransmission is that electrophysiological recordings of single neurons were carried out in the intact brain of an awake or anesthetized animal, or in an explanted piece of tissue [1]. Such recordings have extremely high spatial (micrometer) and temporal (millisecond) resolution and allow direct observation of electrical currents and potentials generated by single nerve cells, which, however, at considerable cost since all cellular recording techniques are highly invasive, requiring surgical intervention and placement of recording electrodes within brain tissue. Neurons communicate via action potentials or spikes; neural recordings are usually transformed into series of discrete spiking events that can be characterized in terms of rate and timing. Less direct observations of electrical brain activity are electromagnetic potentials

http://dx.doi.org/10.5772/55262

generated by combined electrical currents of large neuronal populations, i.e. electroencephalography (EEG) and magnetoencephalography (MEG). They are non-invasive as recordings are made through sensors placed on, or near, the surface of the head. EEG and MEG directly record signals of neuronal activity and thus have a high temporal resolution. But the spatial resolution is relatively poor as neither technique allows an unambiguous reconstruction of the electrical sources responsible for the recorded signal. EEG and MEG signals are often processed in sensor space as sources are difficult to localize in anatomical space.

With the development of magnetic resonance imaging (MRI) in 1980s [2], brain imaging took a huge step forward. The strong magnetic field and radiofrequency pulse used in MRI scanning are harmless, making this technique completely noninvasive. MRI is also extremely versatile: by changing the scanning parameters, we can acquire images based on a wide variety of different contrast mechanisms. For example, diffusion MRI is a MRI method allows the mapping of diffusion process of molecules, mainly water, in biological tissues, in vivo and non-invasively. Water molecule diffusion patterns can consequently reveal microscopic details about tissue architecture in the brain. Functional magnetic resonance imaging (fMRI) measures hemodynamic signals, only indirectly related to neural activity. These techniques allow the reconstruction of spatially localized signals at millimeter-scale resolution across the imaged brain volume. In fMRI, the primary measure of activity is the contrast between the magnetic susceptibility of oxygenated and deoxygenated hemoglobin within each voxel; so it is called the *blood oxygen level-dependent* (BOLD) signal. BOLD signal can only be viewed as an indirect measure of neural activity, In addition, the slow time constants of the BOLD response result in poor temporal resolution on the order of seconds. A critical objective of neuroimaging data analysis is the inference of neural processes responsible for the observed data, that is, the estimation of the hemodynamic response functions.

Neural signals recorded via the above techniques differ significantly in both spatial and temporal resolutions and in the directness with which neuronal activity is detected. Simultaneously using two more recording methods within the same experiment can reveal how different neural or metabolic signals are interrelated [3]. Each technique measures a different aspect of neural dynamics and organization, and interpreting neural data sets shall take these differences into account. All methods for observing brain structure and function have advantages but also disadvantages: some methods provide great structural detail but are invasive or cover only a small part of the brain, while others may be noninvasive but have poor spatial or temporal resolution. Nervous systems are organized at multiple scales, from synaptic connections between single cells, to the organization of cell populations within individual anatomical regions, and finally to the large-scale architecture of brain regions and their interconnections or network connectivity. Different techniques are sensitive to different levels of organization. The multi-scale aspect of the nervous system is an essential feature of its organization and network architecture [4].

### 2.2. Categorization of brain network connectivity

Given the diverse techniques for observing the brain, there are many different ways to describe and measure brain connectivity [5, 6]. Brain connectivity can be derived from histological

sections revealing anatomical connections, from electrical recordings of single nerve cells, or from functional imaging of the entire brain. Even with a single recording technique, different ways of processing and analyzing neural data may result in different descriptions of the underlying network. Structural connectivity is a wiring diagram if physical links while functional connectivity describes dynamic interactions. A third class of brain networks is effective connectivity, which encompasses the network of directed interactions between neural elements. Effective connectivity goes beyond structural and functional connectivity by detecting patterns of causal influence among neural elements. These three main types of brain connectivity are defined more precisely as below.

*Structural connectivity* refers to a set of physical or structural (anatomical) connections that links neural elements. These anatomical connections range in scale from those of local circuits of single cells to large-scale networks of interregional pathways. Their physical pattern can be treated as relatively static at shorter time scales (seconds to minutes) but may be dynamic at longer time scales (hours to days). *Functional connectivity* describes patterns of deviations from statistical independence between distributed and often spatially remote neuronal units. The basis of functional connectivity is time series data from neural recordings such as cellular recording, EEG, MEG, and fMRI. Deviations from statistical independence typically indicates dynamic coupling and can be measured by estimating the correlation or covariance, spectral coherence, or other metrics. Functional connectivity is very time dependent, and can be statistically nonstationary. It is also modulated by external task demands and sensory stimulation, as well as internal state of the organism. But functional connectivity does not make any explicit reference to causal effects among neural elements. *Effective connectivity* captures the network causal effects between neural elements, and can be inferred through time series analysis, statistical modeling, or experimental perturbation. Same as functional connectivity, effective connectivity is also time dependent and can be rapidly modulated by external stimuli or tasks, and internal state. Some methods for effective connectivity inference are model-free without assuming anatomical pathways, while others require the specification of an explicit causal model including structural parameters. In general, the estimation of effective connectivity needs complex data processing and modeling techniques. Thus, in this chapter, regarding the networks, I mainly review strategies for estimation of effective connectivity or causal inference.

## 3. Information theory and processing

### 3.1. Fundamentals and definitions: Entropy, Kullback-Leibler divergence, and mutual information

A major objective of neuroscience is to understand how the brain processes information. Here we provide probabilistic notations and information-theoretic definitions that will be used in this section (definitions denoted with $\triangleq$ ). We define $x^n \triangleq x_1^n = (x_1, \ldots, x_n)$. More generally, for integers $i \leq j$, $x_i^j \triangleq (x_i, \ldots, x_j)$. For a random variable $X$, $\mathcal{X}$ corresponds to a measurable space that $X$ takes values in, and $x \in \mathcal{X}$ are specific realizations. The probability mass function (PMF)

http://dx.doi.org/10.5772/55262

of a discrete random variable $X$ is defined as $\mathrm{P}_X(x) \triangleq P(X=x)$, and the probability density function (PDF) of a continuous random variable is denoted as $p_X(x)$.

The *information* or *surprise* [7] of a discrete random variable is defined as:

$$\log \frac{1}{\mathrm{P}_X(x)} = -\log P(X=x) \quad .$$

The choice of logarithmic base determines the unit. The most common unit of information is the *bit*, based on the binary logarithm. The information is zero for a fully predicted outcome $x$ with $P(X=x)=1$, and it increases as $P(X=x)$ decreases.

The *entropy* of a discrete random variable $X$ is defined to be the average information from observing this variable:

$$\mathrm{H}(X) = \sum_{x \in \mathrm{X}} -\mathrm{P}_X(x) \log \mathrm{P}_X(x) \quad .$$

Entropy is a measure of randomness or uncertainty of the distribution: the more random the distribution, the more information is gathered by observing its value. Specifically, entropy is zero for a deterministic variable and is maximized for a uniform distribution. The conditional entropy is given as below:

$$\mathrm{H}(Y \mid X) = \sum_{x \in \mathrm{X}} \sum_{y \in \mathrm{Y}} -\mathrm{P}_{X,Y}(x, y) \log \mathrm{P}_{Y \mid X}(y \mid x) \quad .$$

The chain rule for entropy is

$$\mathrm{H}(X^n) = \sum_{i=1}^{n} \mathrm{H}(X_i \mid X^{i-1}) \quad .$$

The *Kullback-Leibler* (KL) *divergence* (also called the relative entropy) between two probability distributions $P$ and $Q$ on X is defined as their average difference:

$$\mathrm{D}(\mathrm{P} \| \mathrm{Q}) \triangleq \mathrm{E}_{\mathrm{P}}\left[\log \frac{\mathrm{P}(X)}{\mathrm{Q}(X)}\right] = \sum_{x \in \mathrm{X}} \mathrm{P}(x) \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)} \geq 0 \quad .$$

It is a measure of the difference of two distributions, but does not usually satisfy the symmetry condition, that is, $\mathrm{D}(\mathrm{P} \| \mathrm{Q}) \neq \mathrm{D}(\mathrm{Q} \| \mathrm{P})$. So, it cannot be called "distance".

The *Mutual information* of two discrete random variables $X$ and $Y$ is defined as:

$$\mathrm{I}(X;Y) \triangleq \mathrm{D}(\mathrm{P}_{XY}(\bullet, \bullet) \| \mathrm{P}_X(\bullet)\mathrm{P}_Y(\bullet)) = \mathrm{E}_{\mathrm{P}_{XY}}\left[\log \frac{\mathrm{P}_{Y \mid X}(Y \mid X)}{\mathrm{P}_Y(Y)}\right]$$

$$= \sum_{x \in \mathrm{X}} \sum_{y \in \mathrm{Y}} \mathrm{P}_{X,Y}(x, y) \frac{\mathrm{P}_{Y \mid X}(Y \mid X)}{\mathrm{P}_Y(Y)} = \mathrm{H}(Y) - \mathrm{H}(Y \mid X) \quad .$$

Intuitively, mutual information measures the information that $X$ and $Y$ share: it measures how much knowing one of the variables reduces uncertainty about the other. The mutual information is known to be symmetric: $\mathrm{I}(X;Y) = \mathrm{I}(Y;X)$. The chain rule for mutual information is

$$\mathrm{I}(X^n; Y^n) = \sum_{i=1}^{n} \mathrm{I}(Y_i; X^n \mid Y^{i-1}) \quad ,$$

with the conditional mutual information given as following:

$$I(X;Y \mid Z) = E_{P_{XYZ}}\left[\log \frac{P_{Y \mid X,Z}(Y \mid X, Z)}{P_{Y \mid Z}(Y \mid Z)}\right] \quad .$$

### 3.2. Causal inference: Granger causality, transfer entropy, and directed information

*Granger Causality:* A widely-established technique for extracting causal relations or effective connectivity from data is *Granger causality* [8-11]. The principle of Granger causality is based on the concept of cross prediction. Accordingly, if incorporating the past values of times series X improves the future prediction of time series Y, the X is said to have a causal influence on Y [8]. Exploring Granger causality is closely related to analysis of vector autoregressive (VAR) models, by calculating the variances to correlation terms for autoregressive models. Using terminology introduced in [10], let $X=(X_i : i \geq 1)$ and $Y=(Y_i : i \geq 1)$ be the two time series for determining whether X causally influences Y. Y is first modeled as an univariate autoregressive series with error term $V_i$, and then modeled again using the X series as causal information. That is:

$$Y_i = \sum_{j=1}^{p} a_j Y_{i-j} + V_i \quad ,$$

$$Y_i = \sum_{j=1}^{p} b_j Y_{i-j} + c_j X_{i-j} + W_i \quad , \tag{1}$$

where $W_i$ in Eq. (1) is the new error term. The number of time-lags or model order $p$ can be a fixed prior or specified by minimizing a criterion (for example, Akaike information criterion [12] or Bayesian information criterion [13]) that balances the variance accounted for by the model, against the number of coefficients to be estimated. The Granger causality is defined as below, examining the ratio of the variances of the error terms:

$$G_{X \to Y} \triangleq \log \frac{\mathrm{var}(V)}{\mathrm{var}(W)} \quad .$$

If including X in the modeling decreases the variance of the error term, $G_{X \to Y} > 0$. Typically by comparing $G_{X \to Y}$ and $G_{Y \to X}$, we determine the causal direction as the larger one. The directed transfer function transforms the autoregressive model into the spectral domain [14], and also uses multivariate models rather than univariate and bivariate models for each time series to consider the full covariance matrix for improved modeling. Granger causality, the directed transfer function, and their derivative methods are usually fast to calculate and easy to interpret. Despite the advantages, they may not be statistically suitable for inference questions associated with neural spike train data that are often modeled as point processes due to the sample-variance computation.

*Transfer entropy* is a measure of effective connectivity based on information theory [15, 16]. It does not require a model of interaction, is inherently non-linear, and thus provides a reasonable basis to precisely formulate causal hypotheses. Assume that the two time series $X=(X_i : i \geq 1)$ and $Y=(Y_i : i \geq 1)$ can be approximated by Markov processes:

http://dx.doi.org/10.5772/55262

$$\mathrm{P}_{Y_{n+1}|Y^{n},X^{n}}\left(y_{n+1} \mid y^{n},\, x^{n}\right)=\mathrm{P}_{Y_{n+1}|Y^{n}_{n-J+1},X^{n}_{n-K+1}}\left(y_{n+1} \mid y^{n}_{n-J+1},\, x^{n}_{n-K+1}\right) \ ,$$

where $J$ and $K$ are respectively the orders (memory) of the Markov processes for X and Y. The transfer entropy is defined as conditional mutual information [15]:

$$\mathrm{T}_{X\to Y}(i)=\mathrm{I}\left(Y_{i+1};\, X^{i}_{i-K+1} \mid Y^{i}_{i-J+1}\right) \ . \tag{2}$$

Transfer entropy is asymmetric and based on transition probabilities; it thus provides directional and dynamic information. The key feature of this information theoretic functional for identifying causality is that, theoretically, it does not assume any particular model for the interaction between the two time series. So, transfer entropy is sensitive to all order correlations, which makes it suitable for exploratory analyses over Granger causality or other model based approaches. This is especially advantages if some unknown non-linear interactions are embedded in the systems to be discovered. It is shown in [17] that for Gaussian variables, Granger causality and transfer entropy are equivalent, which bridges autoregressive and information-theoretic methods in causal inference. Another issue with transfer entropy is that its performance depends on the estimation of transitional probabilities; this requires the order selection for both the driven and driving systems.

*Directed information*, proposed by Marko [18] and re-formalized by others [19, 20], is more general for quantifying directional dependencies, and has recently attracted attention [10, 21]. It is modified from the mutual information to capture causal influences, denoted as $\mathrm{I}(\mathrm{X}\to\mathrm{Y})$ for two stochastic processes X and Y. For vectors $X^{n}$ and $Y^{n}$, the mutual information can be shown to be:

$$\begin{aligned} I\left(X^{n};Y^{n}\right) &= \sum_{i=1}^{n} I\left(X^{n};Y_{i} \mid Y^{i-1}\right) \\ &= E\left[\sum_{i=1}^{n} \log \frac{P_{Y_{i}|Y^{i-1},X^{n}}\left(Y_{i} \mid Y^{i-1},X^{n}\right)}{P_{Y_{i}|Y^{i-1}}\left(Y_{i} \mid Y^{i-1}\right)}\right] \\ &= \sum_{i=1}^{n} D(P_{Y_{i}|Y^{i-1},X^{n}} \,\|\, P_{Y_{i}|Y^{i-1}}) \ . \end{aligned} \tag{3}$$

The mutual information is symmetric and only measures the correlation or statistical dependence between random processes, but cannot identify causal directionality. The directed information is defined as:

$$\mathrm{I}\left(X^{n}\to Y^{n}\right) \triangleq \sum_{i=1}^{n} \mathrm{I}\left(X^{i};Y_{i} \mid Y^{i-1}\right) \tag{4}$$

$$= \mathrm{E}\left[\sum_{i=1}^{n} \log \frac{\mathrm{P}_{Y_i|Y^{i-1},X^i}(Y_i \mid Y^{i-1},\ X^i)}{\mathrm{P}_{Y_i|Y^{i-1}}(Y_i \mid Y^{i-1})}\right] \tag{5}$$

$$= \sum_{i=1}^{n} \mathrm{D}\left(\mathrm{P}_{Y_i|Y^{i-1},X^i} \,\|\, \mathrm{P}_{Y_i|Y^{i-1}}\right) \quad . \tag{6}$$

It can also be written as following with the chain rule for entropy:

$\mathrm{I}(X^n \to Y^n) = \mathrm{H}(Y^n) - \mathrm{H}(Y^n \,\|\, X^n)$ ,

where the $\mathrm{H}(Y^n \,\|\, X^n)$ is the causally conditioned entropy given by [22]:

$$\mathrm{H}\left(Y^n \,\|\, X^n\right) \triangleq \sum_{i=1}^{n} \mathrm{H}\left(Y_i \mid Y^{i-1},\ X^i\right) \quad .$$

The difference between mutual information in Eq. (3) and directed information in Eq. (5) is that $X^n$ is changed to $X^i$; so the causal influence of X on the current $Y_i$ at each time $i$ can be captured by directed information. Compared with Granger causality, directed information is a sum of divergences (Eq. (6)), and well-defined for any joint probability distributions including point processes. In addition, directed information is not tied to any particular statistical model; it operates on log likelihood ratios, and thus is more flexible and can be directly applied to varieties of modalities such as neural spike trains. By calculating the mutual information in bits, a degree of correlation (or statistical interdependence) is determined. Similarly, we can also quantify a degree of causation in bits through calculating the directed information. It is demonstrated by Amblard et al. [20]: for linear Gaussian processes, directed information and Granger causality are equivalent. Note that the transfer entropy defined in Eq. (2) is part of the sum terms in Eq. (4) for directed information. Amblard et al. also proved that for a stationary process, directed information rate can be decomposed into two parts: one is equivalent to a particular instance of the transfer entropy, and the other to the instantaneous information change rate. In fact, it has recently shown in [23] that transfer entropy is equal to the upper bound of directed information rate.

### 3.3. Applications and validity in neuroscience and aging research

*Granger Causality:* Li et al. [24] performed a longitudinal MRI study to examine the gray matter changes due to Alzheimer's disease (AD) progression. A standard voxel-based morphometry method was used to localize the abnormal brain regions, and the absolute atrophy rate in these regions was calculated with a robust regression method. The hippocampus and middle temporal gyrus (MTG) were identified as the primary foci of atrophy. A model based Granger causality approach was developed to examine the cause–effect relationship over time between these regions based on gray matter concentration. It is shown that primary pathological foci are in the hippocampus and entorhinal cortex in the earlier stages of AD, and appears to subsume the MTG subsequently. The causality results indicate that there are larger differences in MTG between AD and age-matched healthy control but little in hippocampus, which implies local pathology in MTG being the predominant progressive abnormality during intermediate

http://dx.doi.org/10.5772/55262

stages of AD development. In [25], the authors would like to address ongoing issues regarding how the default-mode network (DMN) hubs, including posterior cingulate cortex (PCC), medial prefrontal cortex (MPFC) and inferior parietal cortex (IPC), interact to each other, and the altered pattern of hubs in AD. Causal influences were examined between any pair of nodes within the DMN using Granger causality analysis and graph-theoretic methods on resting-state fMRI of 12 young subjects, 16 old normal controls and 15 AD patients. Results support the hub configuration of the DMN from the perspective of causal relationship, and reveal abnormal pattern of the DMN hubs in AD. Findings from young subjects give additional evidence for the role of PCC/MPFC/IPC acting as hubs in the DMN. Compared to old control, MPFC and IPC lost their roles as hubs due to the obvious causal interaction disruption, and PCC was preserved as the only hub with significant causal relations with all other nodes. Deshpande et al. [11] proposed a combination of multivariate Ganger causality analysis through temporal down-sampling of fMRI time series, to investigate causal brain networks and their dynamics. The method was applied to study epoch-to-epoch changes in a hand-gripping, muscle fatigue experiment. Causal influences between the activated regions were analyzed by applying the directed transfer function analysis of multivariate Granger causality with the integrated epoch response as the input, to account for the effects of several relevant regions simultaneously. The authors separately modeled the early, middle, and late periods in the fatigue. The results demonstrate the temporal evolution of the network and reveal that motor fatigue leads to a disconnection in the associated neural network.

*Transfer Entropy and Directed Information:* Vicente et al. [16] investigated the applicability of transfer entropy as a measure to electrophysiological data from simulations and MEG recordings in a motor task. Specifically, they demonstrated that transfer entropy improved the effective connectivity identification for non-linear interactions, and for sensor level MEG signals where linear approaches are hampered by signal-cross-talk due to volume conduction. Utilizing transfer entropy at the source-level, Wibral et al. [26] analyzed MEG data from an auditory short-term memory experiments and found that changes in the network between different task types can be detected. Prominently involved areas for the changes include left temporal pole and cerebellum, which have previously been implied to be involved in auditory short-term or working memory. Amblard and Michel [20] extracted Granger causality graphs using directed information, and such techniques were shown to be necessary to analyze the structure of systems with feedback in general, and neural systems specifically. Quinn et al. [10] proposed a nonlinear robust extension of the linear Granger tools also based directed information. They used point process models of neural spike trains, performed parameter and model order selection with minimal description length, and applied the analysis to infer the interactions and dynamics of neural ensembles in the primary motor cortex (MI) of macaque monkeys.

*Multi-Scale Information and Multi-Scale Entropy:* There is increasing evidence that brain signals are expressed with variability of the neural network dynamics [27]. Effective characterization of this variability in the complex systems can bring new insight to empirical studies. A number of tools have recently been developed, integrating information theory, nonlinear dynamics, and complex systems, to support the empirical research and unravel

the principles of brain dynamics [28]. In particular, approximate entropy and sample entropy were proposed to quantify the complexity of short and noisy time series, and with later correcting the bias effect in approximate entropy. Higher values of sample entropy are associated with the signals having more complexity and less regular patterns, while smaller values indicate less irregularity in their representation. Note that signaling in the brain is not instantaneous, and neural activity propagation takes time. Utilizing multi-scale entropy (MSE) is a reasonable strategy to control for the embedding delay of the brain system. This can be achieved through down-sampling the original time series by factors 2, 4, 8, etc., which, would alleviate the effects of linear correlations between consecutive samples. A similar idea was previously introduced in [29], using a complexity measure based on the Shannon entropy at various scales. Some studies used the approximate and sample entropy statistics to quantify the brain signal variability for both the electrode measurements [30] and source dynamics [31]. In [32], in order to test the hypothesis that complexity of BOLD activity is reduced with aging and is correlated with cognitive performance in the elderly, the authors employed the MSE analysis, and investigated appropriate parameters for MSE calculation. Compared with younger subjects, the older group had the most significant reductions in MSE of BOLD signals in posterior cingulate gyrus and hippocampal cortex. MSE of BOLD signals from DMN areas were found to be positively correlated with major cognitive functions including attention, short-term memory and language, etc. The MSE approach was also applied to reveal the differences in the EEG signals, between normal subjects and patients with AD. The resting-state EEG was utilized in [33] with MSE curves (scales 1-16) averaged over channels and individuals for three groups: normal population, subjects with mild cognitive impairment (MCI), and AD patients. The three groups have some common features for the MSE curves, i.e. the sample entropy reached its maximum at scales 5-7 and then gradually decreased. Severe AD patients had a significantly lower level of sample entropy values than that of the normal group at scale 2-16. The maximal difference in the complexity was observed at scales 6-8. Between MCI and normal subjects, the main difference in the MSE curve was the shift of the peak in sample entropy toward coarse timescales for the MCI group.

## 4. Probabilistic modeling and Bayesian inference for neural computation, cognition, and behavior

### 4.1. Bayes' theorem and approximate inference

A generic problem in science is: given the observed data $\mathsf{D}$ and some knowledge of the underlying data generating mechanism, can you tell something about the variable $\theta$? Based on Bayes' theorem, our interest is the quantity:

$$p(\theta \mid \mathsf{D}) = \frac{p(\mathsf{D} \mid \theta)p(\theta)}{p(\mathsf{D})} = \frac{p(\mathsf{D} \mid \theta)p(\theta)}{\int_\theta p(\mathsf{D} \mid \theta)p(\theta)d\theta} \quad .$$

That is: from a *generative model* $p(\mathsf{D} \mid \theta)$ of dataset and a *prior* belief $p(\theta)$ about which variable values are appropriate, we can infer the *posterior* distribution $p(\theta \mid \mathsf{D})$ of the variable in light

http://dx.doi.org/10.5772/55262

of the observed data. When a particular observation is made, $p(\mathsf{D} \mid \theta)$ is called the likelihood. The *maximum a posteriori* (MAP) estimate maximizes the posterior, $\theta_* = \text{argmax}_\theta p(\theta \mid \mathsf{D})$. For a flat prior, i.e. for $p(\theta)$ being a constant, the MAP solution is equivalent to the *maximum likelihood*, with $\theta$ maximizing the likelihood $p(\mathsf{D} \mid \theta)$ of the model generating the observed data. The MAP can incorporate our prior knowledge about the variable, but it is still a point estimate. Bayesian estimate gives the full probability distribution or density of the posterior $p(\theta \mid \mathsf{D})$. For example, when the distribution is wide or even has multiple peaks, the corresponding outputs can be averaged to make a more conservative estimate instead of just using a single point estimate.

A key algorithm challenge for Bayesian inference is for many models of interest, analytical tractability of the above posterior is elusive due to the integral in the denominator. We therefore resort to approximation inference, where the approaches tend to fall into one of following two classes: 1) *Monte Carlo methods* [34] provide approximate answers with accuracy depending on the number of generated samples. Importance sampling is a simple Monte Carlo approximation while Markov chain Monte Carlo (MCMC) is more efficient and popular. MCMC generates each sample by making a random change to the preceding sample. So we can think of an MCMC algorithm as being in a particular current state specifying a value for every variable and generating a next state by making random changes to the current state. Special cases of MCMC include Gibbs sampling and the Metropolis-Hasting algorithm. 2) *Variational approximations* [35, 36] are a series of deterministic techniques that make approximate inference for the parameters in complex statistical models. Compared with MCMC, they are much faster, especially for large models, but limited in their approximation accuracy. The mean-field approximation is a simplest example, which exploits the law of large numbers to approximate large sums of random variable by their means. Variational parameters are introduced and iteratively updated so as to minimize the KL divergence between the approximate and true probability distributions. Updating the variational parameters becomes a proxy for inference. The mean-field approximation produces a lower bound on the likelihood. More sophisticated methods are possible, which give tighter lower (and upper) bounds.

### 4.2. Neuroimaging data analyses using Bayesian approaches

Here I focus on Bayesian inference in fMRI data analysis, mainly for activation detection and hemodynamic response function (HRF) estimation, although the key concepts of Bayesian methods have been applied to structural MRI images as well [37-39]. Graphical model based Bayesian and dynamic Bayesian networks and their applications will be discussed in Section 5.

Bayesian inference has taken fMRI analysis research into an area that classical frequentist statistics have difficulty to address because of some challenging issues associated with the data. For example, fMRI response to stimuli is not instantaneous, but lagged and damped by the hemodynamic response. Estimating HRFs has gained increasing interests, since it provides not only a deep insight into the underlying dynamics of human brain, but also a basis for making inference of brain activation regions. How do we account for the HRF

properties such as the nonlinearities and variability over different brain regions? fMRI is a 4-dimensional signal though with spatial and temporal noise correlations [40, 41]. How to incorporate the modeling of the presence of these correlations into the data analysis, alongside considering the clustered pattern of activation? Moreover, group level statistical inference of fMRI time series is usually needed to answer imaging-based scientific questions. How to make valid, sensitive and robust estimation of activation effects in populations of subjects? In fMRI analysis, what we often do is taking acquired data plus a generative model and extracting pertinent information about the brain, i.e. making inference on the model and its parameters. Bayesian statistics requires a prior probabilistic belief about the model parameters to be specified. Such models are typically HRF models, spatial models, and hierarchical multi-subject models, to respectively address the challenges listed above.

*HRF models* can incorporate biophysical or regularization priors for flexible HRF modeling across brain voxels and over subjects. Several similar Bayesian approaches in the literature use parametric HRFs with parameters describing features such as time-to-peak and undershoot size [42, 43]. Priors placed on these HRF parameters can ensure biological plausibility and result in increased sensitivity. An early example of more advanced HRF modeling is in [44], which uses Bayes to infer on a fully Bayesian biologically informed generative model. The reason of introducing regularization priors is the models have too many parameters to infer stably without regularization. Bayesian regularization places priors on HRF parameters to encode the prior belief that HRF is smooth temporally without strong assumptions about the shape of the response function. Thus such priors are suitable for exploratory approaches or possibly abnormal HRFs. Regularization priors can also be achieved through semi-parametric Bayesian for HRF modeling [45-47]. In semi-parametric approaches, HRF does not have a fixed parametric format but can take any form with a parameter describing the HRF size at each time point.

*Spatial models* for regularization using spatial Markov random field (MRF) priors to tackle spatial correlation in fMRI were proposed in [38, 48, 49], followed by MCMC numerical integration for inference. To overcome the large computation cost for spatial model inference in MCMC, Variational Bayesian approaches were developed [50, 51] without time-consuming numerical integration. Variational Bayes approximate the true posterior distribution through estimation using a posterior factorized over subsets of the model parameters, which results in update equations with the desired approximate posterior distributions in a much more efficient way than techniques such as MCMC. MRF-based work has recently been extended to using more flexible spatial Gaussian Process priors, to allow for the modeling of spatial non-stationarities [52] and the combining of spatial and non-spatial prior information [53]. The hyperparameters of the spatial priors can be estimated via Bayesian inference together with the rest of the model, which is a key advantage of fully Bayesian methods. Some other spatial models include mixture models representing the active and non-active voxels [54-56] and a Bayesian wavelets approach [57]. The popular mixture modeling, however, can be hampered by the presence of structured noise artifacts (e.g. stimulus correlated motion, spontaneous networks of

http://dx.doi.org/10.5772/55262

activity) violating the distributional assumptions. More sophisticated modeling of structured noise could be needed to render the distributional assumptions valid. Recent development of nonparametric Bayes can also be used to handle the mixture modeling, though a massive number of model parameters need to be estimated. Infinite mixture models based on Dirichlet process priors [58] involve effectively an infinite number of distributions. An application of such methods in fMRI for activation regions is in [59] using a spatial mixture model.

*Hierarchical models for group inference* was first proposed in [60], which fit naturally into the Bayesian framework via a cascade of conditional probabilities to handle activation effects over multiple subjects. In classical fMRI analysis, group-level inferences are usually made using the results of separate first-level analyses to decrease computation cost. This is the so-called summary statistics approach. The widely-used frequentist group analysis in [61] employed parameter estimates from the general linear model regression as summary statistics, which however, was only optimal under certain conditions due to the required balanced designs. On the contrary, Woolrich et al. [55] utilized Bayes to incorporate the summary statistics without restrictions, with information regarding both the effect sizes from the lower levels and their variances passed up.

### 4.3. Bayesian brain: Cognition, perception, uncertainty, behavior and neural representations

The neuroscience principle that the nervous system of animals and humans is adapted to the statistical properties of the environment is reflected across all organizational levels, from the activity of single neurons to networks and behavior [62]. A critical aim of the nervous system is to estimate the world state from incomplete and noisy data. During such process, a challenge issue that brains must handle is uncertainty. For example, when we perceive the physical world, make a decision, and take an action, there is uncertainty associated with the sensory system, the motor apparatus, one's own knowledge, and the world itself. Probability has played a central role in perception and cognition modeling. Specifically, the Bayesian framework of statistical estimation provides a systematic way of dealing with these uncertainties for optimal estimation. Comparison between the optimal and actual behavior gives rise to better understanding about how the nervous system works. Bayesian models have been used to explain results in perception, cognition, behavior, and neural coding in diverse forms [63-67], with differences in distinct assumptions about the world variables and how they relate to each other. However, the same key idea shared by all these Bayesian models is that different sources of information can be integrated for estimation of the relevant variables. Thus the Bayesian approach unifies an enormous range of otherwise apparently disparate behavior within one coherent framework.

A key aim of cognitive science is to reverse-engineer the mind. Cognition modeling based on the probabilistic method begins by identifying ideal solutions to these inductive problems, and then uses algorithms to model the mental processes for approximating these solutions. Neural processes are viewed as mechanisms for implementing these algorithms. Probabilistic models of cognition pursue a top-down strategy, which begins with abstract principles allowing agents to solve problems posed by the world (i.e. the func-

tions minds performing) and then aims to reduce these principles to psychological and neural processes. This analysis results in better flexibility in exploration of the representations and inductive biases underlying human cognition. On the contrary, connectionist models usually follow a bottom-up approach that starts with a neural mechanism characterization and explores what macro-level functional phenomena might emerge. With a formal characterization of an inductive problem, a probabilistic model specifies the hypotheses under investigation, the relation between these hypotheses and observable data, and the prior probability of each hypothesis. By assuming different prior distribution for the hypotheses, different inductive biases can be captured. Although the link between probabilistic inference and neural computation/function is drawing attention of modelers from different backgrounds, little is known concerning how these structured representations can be implemented in neural systems for high-level cognition.

Sufficient results in perception have shown that the nervous system represents its uncertainty about the true state of the world probabilistically and such representations are utilized in two related cognitive areas: information fusion and perceptual decision-making. To fuse information from different sources about the same object, inferences about the object should rely on these sources commensurate with their corresponding uncertainty, as demonstrated in multisensory integration [68, 69] with the sources of different sensory modalities, or between information coming from the senses and being stored in memory [70, 71]. With the Bayesian framework, the organism calculates probability distributions over parameters describing the state of the world, with computation based on sensory information and knowledge accrued from experience. Although the particular sensory information and prior knowledge are specific to the task, the computation follows the same probability rules. Psychological evidence at the behavior level that animals and humans represent uncertainty during perceptual processes caused research into the neural underpinnings of such probabilistic representations. That is: how neurons compute with sensory uncertainty information or even full probability distributions? One scheme is the probabilistic population coding [72] that involves making use of the likelihood function encoded in neural population activity (as described below). Beyond perception, the neural implementation of cognitive probabilistic models has basically not been explored yet [64, 73].

*Neural/Neuronal Models of Probabilistic Computation (Probabilistic Population Coding)*: Perception modeling has the potential to constrain neural implementation of perceptual computation. In order to form a neural model from a behavioral model, one needs to first define the relevant level of neural variables. A common candidate is the level of spike counts in sensory and decision-making neurons. For example, an orientated stimulus s might elicit a set of spike counts $\mathbf{r}=(r_1, \ldots, r_n)$ in a population of orientation-tuned cells in primary visual cortex. There is trial-to-trial variability in the population activity, which can be described by a distribution $p(\mathbf{r}|s)$. The connection between $\mathbf{r}$ and s, is that the latter (the scalar stimulus in a behavioral model) is the value maximizing the neural likelihood function, $\mathrm{L}(s)=p(\mathbf{r}|s)$ [74]. The likelihood function $\mathrm{L}(s)$ has a width, $\sigma$, reflecting the observer's uncertainty about the stimulus. The variable $\mathbf{r}$, is high-dimensional with sufficient degrees of freedom to encode $\sigma$ on a trial-by-trial basis. With neural likelihood functions, Bayesian models of

http://dx.doi.org/10.5772/55262

behavior can be mapped to neural operations. This scheme has been successfully applied to cue combination [72], decision-making [75], etc. Some alternative approaches for encoding likelihood functions or probability distributions using neurons have also been proposed in the literature [65, 66, 76, 77].

## 5. Graphical models, Bayesian and dynamic Bayesian networks

### 5.1. Mathematical description and solution

Graphical models, intersecting probability and graph theories, provide a natural tool for handling uncertainty and complexity that frequently occur in applied mathematics and engineering, and scientific domains involving computation. Many of the classical multivariate probabilistic techniques are special cases of the general graphical models, such as mixture models, factor analysis, hidden Markov models, Kalman filters and Ising models [35, 78, 79]. A graph consists of *nodes* connected by *links* (also called *arcs* or *edges*). The nodes in probabilistic graphical models represent random variables, and the links or arcs express probabilistic relationships between these variables. The lack-of-arcs represent conditional independence assumptions. This provides a compact representation of joint probability distributions over all of the random variables, which can be decomposed into a product of factors each depending on a subset of variables. One category of graphical models is *Markov Random Fields* (MRFs), also known as *undirected graphical models*, in which the links do not have arrows and thus do not provide directional significance. For example, two sets of nodes $A$ and $B$ are conditionally independent given a third set, $C$, if all paths between the nodes in $A$ and $B$ are separated by a node in $C$. The other major class is *Bayesian Networks* or *Belief Networks* (BNs), also known as *directed graphical models*, in which the links carry arrows indicating a particular directionality in the notion of independence. Despite the complexity, directed models do have several advantages compared to undirected models; and the most important is that they can express causal relationships between random variables, whereas undirected graphics are more suitable for soft constraints between random variables.

In Bayesian Networks, if there is an arrow from node $X$ to node $Y$, $X$ is said to be a *parent* of $Y$. Each node $X_i$ is associated with a conditional probability distribution (CPD) $\mathrm{P}(X_i \mid Parents(X_i))$, quantifying the effect of the parents on the node. If the variables are discrete, it is represented as a table (CPT), listing the probability that the child node takes on each of its different values for each combination of its parents' values. The network in BNs can be viewed as a representation of the joint probability distribution (JPD), or as an encoding of a collection of conditional independence statements. Let the joint distribution be $\mathrm{P}(x_1, \ldots, x_n)$; and we have

$$\mathrm{P}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathrm{P}(x_i \mid parents(X_i)) \ , \tag{7}$$

where $parents(X_i)$ denotes the values of $Parents(X_i)$ appearing in $x_1, \ldots, x_n$. The CPTs are essentially conditional probability tables based on Eq. (7). In general, given $n$ binary nodes, the full joint would require $O^{2^n}$ space to represent, but due to the presence of independence in the graphical modeling, the factored form would require $O^{n2^k}$ space, where $k$ is the maximum fan-in of a node. Fewer parameters make learning easier.

Note that Bayesian networks do not necessarily imply Bayesian statistics. In fact, it is common to use frequentists methods to estimate the parameters of the CPDs. They are so called because they use Bayes' rule for probabilistic inference. Nevertheless, Bayes net are a useful representation for hierarchical Bayesian models, which form the foundation of applied Bayesian statistics. Bayesian statistical methods in conjunction with Bayesian networks offer an efficient and principled approach for avoiding the data overfitting. Dynamic Bayesian Networks (DBNs) are directed graphical models of stochastic processes, and generalization of hidden Markov models (HMMs) and linear dynamical systems (LDSs). DBN represent the hidden (and observed) state in terms of state variables, which can have complex interdependencies. The simplest DBN is a HMM, with one discrete hidden node and one discrete or continuous observed node per slice. A LDS has the same topology as an HMM, but all the nodes are assumed to have linear-Gaussian distributions. Kalman filter is an online filtering of this model.

A graphical model specifies a complete JPD over all the variables; and all possible inference queries can be answered by marginalization, i.e. summing out over irrelevant variables. However, the JPD has size $O^{2^n}$, with $n$ the number of nodes, and each node is assumed to have 2 states. So, summing over the JPD takes exponential time. More efficient methods are thus desirable, including variable elimination [80], dynamic programming [81], approximation algorithms [34, 35] (Monte Carlo methods, variational methods), etc. For the learning part, a BN has two components that need to be specified, i.e. the graph topology (structure) and the parameters (CPD of each node). It is possible to learn both of these from data, though learning structure is much harder than learning parameters. Also, learning when some of the nodes are hidden, or we have missing data, is much harder than when everything is observed. This gives rise to 4 cases and the respective algorithms: 1) known structure and full observability: Maximum Likelihood Estimation; 2) known structure and partial observability: Expectation Maximization (EM) algorithm; 3) unknown structure, full observability: search through model space; 4) unknown structure, partial observability: EM and search through model space.

### 5.2. Applications and validity in neuroimaging and aging research

*Functional MRI:* Bayesian networks (BNs) were used in [82] to learn the structure of effective connectivity involved in a fMRI experiment. The approach is exploratory, does not require a priori hypothesized model, and was validated using synthetic data and fMRI data collected in silent word reading and counting Stroop tasks. However, BNs provide a single snapshot of effective connectivity of the entire experiment and thus are not suitable for accurately inferring the temporal characteristics of connectivity. Dynamic Bayesian networks (DBNs) were then proposed [83] to learn the structure of effective brain connectivity in an exploratory way. A

http://dx.doi.org/10.5772/55262

Markov chain was employed to model fMRI time-series for discovery of temporal interactions among brain regions. DBNs yield more accurate and informative brain connectivity than earlier methods since temporal characteristics of time-series are explicitly accounted. The functional structures captured on two fMRI datasets are consistent with the previous literature findings and more accurate than those identified by BN. Li et al. [84] aimed to extrapolate BN results from one subject to an entire population while addressing inter-subject, within-group variability. The authors explored two group analysis approaches in fMRI using DBNs: constructing a group network based on a common structure assumption across individuals, and identifying significant structure features by examining DBNs individually-trained. The methods were validated on subjects performing a motor task at three progressive levels of difficulty, and statistically significant, biologically plausible connectivity was detected.

*Structural MRI:* Detecting interactions among brain regions from structural MRI presents a major challenge in computational neuroanatomy. Instead of traditional univariate analysis for brain morphometry, a network analysis based on a BN representation of variables was investigated in [85] to take into account interactions among brain structures in explaining a clinical outcome. Results on a cross-sectional study of mild cognitive impairment (MCI) demonstrated nonlinear and complex multivariate associations among morphological changes in the left hippocampus, the right thalamus, and the presence of MCI. This indicates that the BN has the potential to predict the presence of MCI from structural MRI. Chen et al. [86] proposed to use DBN to represent evolving inter-regional dependencies and identify longitudinal morphological changes in the human brain. The main advantage of DBN modeling is that it can represent complicated interactions among temporal processes. The approach wad validated by analyzing a simulated atrophy study: only a small number of samples were needed to detect the ground-truth temporal model. The method was also applied to a longitudinal study of normal aging and MCI — the Baltimore Longitudinal Study of Aging. It was shown that interactions among regional volume-change rates for the MCI group were different from those for the normal aging group.

*Further Development of Sparse BNs and Time-Varying DBNs:* There are some recent new development in the area of BNs and DBNs. Sparse BN for effective connectivity modeling was investigated in [87], with a novel formulation for the structure learning of BNs. A L1-norm penalty term imposes sparsity and another penalty ensures the learned networks to satisfy the required property of BNs (i.e. directed acyclic graph). Both theoretical analysis and experiments on moderate and large benchmark networks demonstrate that the approach has enhanced learning accuracy and scalability compared with existing algorithms. The authors also applied the proposed method to brain images of 42 Alzheimer's disease (AD) and 67 normal controls (NC); the revealed effective connectivity of AD was shown to be different from that of NC, for example, in the global-scale effective connectivity, intra-lobe, inter-lobe, and inter-hemispheric effective connectivity distributions, and the effective connectivity corresponding to specific brain regions. Graphical model results are often based on static networks, assuming networks with invariant topology. For certain situations, it is desirable to understand and quantitatively model the dynamic topological and functional properties of biological or brain networks. This yields time or condition specific time-varying or non-stationary net-

works. In order to capture the dynamic causal influences between covariates, time-varying dynamic Bayesian networks (TV-DBNs) was proposed [88]. It models the varying directed dependency structures underlying non-stationary biological/neural time series. A kernel reweighted L1-regularized auto-regressive procedure was employed, with desirable properties including computational efficiency and asymptotic consistency. Application of the TV-DBNs to simulated data and brain EEG signals to visual stimuli show that the technique can identify temporally rewiring networks due to system dynamic transformation.

## 6. Dynamical brain system

### 6.1. Attractors and brain dynamics

Computational neuroscience illustrates the network dynamics of neurons and synapses with models to reproduce emergent properties or predict observed neurophysiology (e.g. single- and multiple-cell recordings, EEG, MEG, fMRI) and associated behavior [27]. Attractor theory [89] is a powerful theoretical framework that can capture the neural computations inherence in cognitive functions such as attention, memory, and decision making. It is based on mathematical models formulated at the level of neuronal spiking and synaptic activity. An attractor of a dynamical system is a subset of the state space to which orbits originating from typical initial conditions evolve over time. It is common for dynamical system to have more than one attractor. For each such attractor, its *basin of attraction* is the set of initial conditions that give rise to long-time behavior approaching that attractor. Reduced depths in the basins of attraction of prefrontal cortical networks and the noise effects could result in some cognitive symptoms like poor short-term memory and attention. The hypothesis is that reduced depth in the basins of attraction would make short-term memory unstable. Hence the continuing firing of neurons implementing short-term memory sometimes would cease, and the system under noise influence would fall back out of the short-term memory state into spontaneous firing. Top-down attention requires a short-term memory to hold the object of attention in mind. This is the source of the top-down attentional bias that influences competition in other networks receiving incoming signals. Therefore, disruption of short-term memory is also predicted to impair the attention stability.

### 6.2. Attractors dynamics in aging

The stochastic dynamical theory to brain function given above has implications in aging research. In the following, we describe effects of these factors and the associated hypotheses to aging [90]. The stochastic dynamic approach to aging can provide a way to test combinations of pharmacological treatments, which may together help to minimize the cognitive symptoms of aging.

*NMDA Receptor Hypofunction*: NMDA receptor functionality tends to decrease with aging [91]. This would act to reduce the depth of the basins of attraction, by reducing firing rate of the neurons in the active attractor, and by decreasing the strength of the potentiated synaptic connections that support each attractor. The reduced depth in the basins of attraction could

http://dx.doi.org/10.5772/55262

have several effects to cognitive changes in aging. First, the stability of short-term memory networks would be impaired, which may cause difficulty in hold items in short-term memory for long. Second, top-down attention would be impaired. Third, the recall of information from episodic memory systems in the temporal lobe would be impaired [92]. Lastly, any reduction of the firing rate of the pyramidal cells caused by NMDA receptor hypofunction would itself be likely to impair new learning involving long-term potentiation (LTP).

*Dopamine*: D1 receptor blockade in the prefrontal cortex can impair short-term memory [93]. Partial reason for this may be that D1 receptor blockade can decrease NMDA receptor activated ion channel conductances. Hence part of the role of dopamine in prefrontal cortex in short-term memory can be accounted for by a decreased depth in the basins of attraction of prefrontal attractor networks [94]. The decreased depth would be caused by both the decreased firing rate of the neurons, and the reduced efficacy of the modified synapse since their ion channels would be less conductive. Dopaminergic function in the prefrontal cortex may decline with aging [95], which could contribute to the reduced short-term memory and attention in aging.

*Impaired Synaptic Modification*: Long-lasting associative synaptic modification may also contribute to the cognitive changes in aging, as LTP is more difficult to achieve in older animals and decays more quickly [91, 96]. This would tend to make the synaptic strengths support an attractor weaker and weaken further over time, and thus directly reduces the depth of the attractor basins. This would impact episodic memory, the memory for particular past episodes. The reduction of synaptic strength over time could also affect short-term memory, which requires the synapses supporting a short-term memory attractor be modified in the first place using LTP, before the attractor is used [97].

*Cholinergic Function*: Acetylcholine in the neocortex has its origin largely in the cholinergic neurons in the basal magnocellular forebrain nuclei of Meynert. The correlation of clinical dementia ratings with the reductions in a number of cortical cholinergic markers such as choline acetyltransferase, muscarinic and nicotinic acetylcholine receptor binding, as well as levels of acetylcholine, implied an association of cholinergic hypothesis of memory dysfunction in senescence and AD [98]. Cholinergic system could also alter the cerebral cortex function in ways that can be illuminated by stochastic neurodynamics [99]. Enhancing cholinergic function will likely help to reduce the instability of attractor networks involved in short-term memory and attention that may occur in aging.

## 7. Conclusions

Brain structure and activity can be described at various levels of resolution. Recent developments in biotechnology have provided us the ability to measure and record population neuronal activity with more precision and accuracy than ever before, allowing researchers to study and perform detailed analyses which may have been impossible just a few years ago. Brain imaging techniques, such as EEG, MEG, and structural/functional MRI, open macroscopic windows on processes in the working brain. These methods yield high dimensional data sets that are organized in space and time [100]. This creates a huge analysis need to extract

interpretable signals and information from the big data, harvesting the full richness of the multi-modality measurements of the multi-scale brain. One of the future directions on the computation side is to develop high-dimensional analysis methods for mining and modeling of the neuroscience data, and thus to assess and interpret properties in the joint data set combining imaging and behavior/stimulus measurements. The objective is to further our understanding about how neural structures of humans and other animals develop, are aged, and create systems able to accomplish basic and complex behavioral tasks.

## Acknowledgements

Preparation of this chapter is supported in part by a grant from the National Institute of Aging, K25AG033725.

## Author details

Michelle Yongmei Wang*

Address all correspondence to: ymw@illinois.edu

Departments of Statistics, Psychology, and Bioengineering, Beckman Institute, University of Illinois at Urbana-Champaign, U.S.A.

## References

[1] Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., & White, L. E. *Neuroscience*, Sinauer Associates, Inc., (2008).

[2] Ramachandran, V. S. *Encyclopedia of the Human Brain*, (2002), 3.

[3] Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal, *Nature,* (2001), 412: 150-157.

[4] Sporns, O. *Networks of the Brain*, Massachusetts Institute of Technology, (2011).

[5] Friston, K. J. Functional and effective connectivity: a review, *Brain Connectivity,* (2011), 1: 13-36.

[6] Sporns, O. *Discovering the Human Connectome*, Massachusetts Institute of Technology, (2012).

[7] Shannon, C. E. A mathematical theory of communication, *Bell System Technical Journal,* (1948), 27: 379-423.

http://dx.doi.org/10.5772/55262

[8] Granger, C. Investigating causal relations by econoetric models and cross-spectral methods, *Econometrica,* (1969), 37: 424-438.

[9] Seth, A. K. A MATLAB toolbox for Granger causal connectivity analysis, *J Neurosci Methods,* (2010), 186: 262-273.

[10] Quinn, C. J., Coleman, T. P., Kiyavash, N., & Hatsopoulos, N. G. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings, *J Comput Neurosci,* (2011), 30: 17-44.

[11] Deshpande, G., LaConte, S., James, G. A., Peltier, S., & Hu, X. Multivariate Granger causality analysis of fMRI data, *Hum Brain Mapp, (2009),* 30: 1361-1373.

[12] Akaike, H. A new look at the statistical model identification, *IEEE Transactions on Automatic Control,* (1974), 19: 716-723.

[13] Schwartz, G. Estimating the dimension of a model, *Annals of Statistics,* (1978), 5: 461-464.

[14] Kaminski, M., Ding, M., Truccolo, W. A., & Bressker, S. L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance, *Biological Cybernetics,* (2001), 85: 145-157.

[15] Schreiber, T. Measuring information transfer, *Phys Rev Lett,* (2000), 85: 461-464.

[16] Vicente, R., Wibral, M., Lindner, M., & Pipa, G. Transfer entropy--a model-free measure of effective connectivity for the neurosciences, *J Comput Neurosci,* (2011), 30: 45-67.

[17] Barnett, L., Barrett, A. B., & Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables, *Phys Rev Lett,* (2009), 103: 238701.

[18] Marko, H. The bidirectional communication theory- a generalization of information theory, *IEEE Transactions on Communications,* (1973), 21: 1345-1351.

[19] Massey, G. Causality, feedback and directed information, in *Proceedings of International Symposium on Information Theory and Its Applications*, (1990), pp. 27-30.

[20] Amblard, P. O., & Michel, O. J. On directed information theory and Granger causality graphs, *J Comput Neurosci,* (2011), 30: 7-16.

[21] Seghouane, A. K., & Amari, S. Identification of directed influence: Granger causality, Kullback-Leibler divergence, and complexity, *Neural Comput,* (2012), 24: 1722-1739.

[22] Kramer, G. *Directed Information for Channels with Feedback*, Ph.D. Thesis, University of Manitoba, Canada, (1998).

[23] Liu, Y., & Aviyente, S. The relationship between transfer entropy and directed information, in *IEEE Statistical Signal Processing Workshop*, (2012), pp. 73-76.

[24] Li, X., Coyle, D., Maguire, L., Watson, D. R., & Mcginnity, T. M. Gray matter concentration and effective connectivity changes in Alzheimber's disease: a longitudinal structural MRI study, *Diagnostic Neuroradiology,* (2011), 53: 773-748.

[25] Miao, X., Wu, X., Li, R., Chen, K., & Yao, L. Altered connectivity pattern of hubs in default-mode network with Alzheimer's disease: an Granger causality modeling approach, *PLoS One,* (2011), 6: e25546.

[26] Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., & Kaiser, J. Transfer entropy in magnetoencephalographic data: quantifying information flow in cortical and cerebellar networks, *Prog Biophys Mol Biol,* (2011), 105: 80-97.

[27] Rabinovich, M. I., Friston, K. J., & Varona, P. *Principles of Brain Dynamics*, Massachusetts Institute of Technology, (2012).

[28] Vakorin, V. A., Ross, B., Krakovska, O., Bardouille, T., Cheyne, D., & Mcintosh, A. R. Complexity analysis of source activity underlying the neuromagnetic somatosensory steady-state response, *Neuroimage,* (2010), 51: 83-90.

[29] Zhang, Y.-C. Complexity and 1/f noise. A phase space approach, *Journal of Physique I France,* (1991), 1: 971-977.

[30] Abasolo, D., Hornero, R., Espino, P., Alvarez, D., & Poza, J. Entropy analysis of the EEG background activity in Alzheimer's disease patients, *Physiological Mesurement*, (2006), 27: 241-253.

[31] Misic, B., Mills, T., Taylor, M. J., & Mcintosh, A. R. Brain noise is task-dependent and region-specific, *Journal of Neurophysiology,* (2010), 104: 2667-2676.

[32] Yang, A. C., Huang, C.-C., Yeh, H.-L., Liu, M.-E., Hong, C.-J., & Tu, P.-C. Complexity of spontaneous BOLD activity in default mode network is correlated with cognitive function in normal male elderly: a multiscale entropy analysis, *Neurobiology of Aging,* (2013), 34: 428-438.

[33] Park, J. H., Kim, S., Kim, C. H., Cichocki, A., & Kim, K. Multiscale entropy analysis of EEG from patients under different pathological conditions, *Fractals,* (2007), 15: 399-404.

[34] Robert, C., & Casella, G. *Monte Carlo Statistical Methods*, Berlin: Springer-Verlag, (2004).

[35] Jordan, M., Ghahramani, Z., Jaakkola, T., & Saul, L. An introduction to variational methods for graphical models, *Machine Learning,* (1999), 37: 183-233.

[36] Ormerod, J. T., & Wand, M. P. Explaining variational approximations, *The American Statistician,* (2010), 64: 140-153.

[37] Wang, Y. *Statistical Shape Analysis for Image Segmentation and Physical Model-Based Non-Rigid Registration*, Ph.D. Thesis, Department of Electrical Engineering, Yale University, (1999).

http://dx.doi.org/10.5772/55262

[38] Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., & Behrens, T. Bayesian analysis of neuroimaging data in FSL, *Neuroimage,* (2009), 45: S173-S186.

[39] Wang, Y, & Staib, L. H. Boundary finding with prior shape and smoothness models, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* (2000), 22: 738-743.

[40] Wang, Y. M., & Xia, J. Unified framework for robust estimation of brain networks from fMRI using temporal and spatial correlation analyses, *IEEE Transactions on Medical Imaging,* (2009), 28: 1296-1307.

[41] Wang, Y. M. Modeling and nonlinear analysis in fMRI via statistical learning, in *Advanced Image Processing in Magnetic Resonance Imaging*, Landini, L. Positano, V., & Santarelli, M. F., Eds., Marcel Dekker International Publisher, (2005), pp. 565-586.

[42] Genovese, C. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion), *Journal of the American Statistical Association,* (2000), 95: 691-703.

[43] Gossl, C., Fahrmeir, I., & Auer, D. P. Bayesian modeling of the hemodynamic response function in bold fmri, *Neuroimage,* (2001), 14: 140-148.

[44] Friston, K. J. Bayesian estimation of dynamical systems: an application to fmri, *Neuroimage,* (2002), 16: 513-530.

[45] Ciuciu, P., Poline, J. B., Marrelec, G., Idier, J., Pallier, C., & Benali, H. Unsupervised robust nonparametric estimation of the hemodynamic response function for any fmri experiment, *IEEE Transactions on Medical Imaging,* (2003), 22: 1235-1251.

[46] Goutte, C., Nielsen, F. A, & Hansen, L. K. Modeling the haemodynamic response in fmri using smooth fir filters, *IEEE Transactions on Medical Imaging,* (2000), 19: 1188-1201.

[47] Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., & Poline, J. B. Robust Bayesian estimation of the hemodynamic response function in event-related bold fmri using basic physiological information, *Human Brain Mapping,* (2003), 15: 1-25.

[48] Gossl, C., Auer, D. P., & Fahrmeir, L. Bayeisan spatiotemporal inference in functional magnetic resonance imaging, *Biometrics,* (2001), 57: 554-562.

[49] Xia, J., Liang, F., & Wang, Y. M. FMRI analysis through Bayesian variable selection with a spatial prior, in *IEEE International Symposium on Biomedical Imaging*, (2009), pp. 714-717.

[50] Penny, W. D., Trujillo-Barreto, N. J, & Friston, K. J. Bayesian fmri time series analysis with spatial priors, *Neuroimage,* (2005), 24: 350-362.

[51] Woolrich, M., Behrens, T., & Smith, S. Constrained linear basis sets for HRF modeling using variational Bayes, *Neuroimage,* (2004), 21: 1748-1761.

[52] Harrison, L. M., Penny, W. D., Ashburner, J., Trujillo-Barreto, N., & Friston, K. J. Diffusion-based spatial priors for imaging, *Neuroimage,* (2007), 38: 677-695.

[53] Groves, A. R., Chappell, M. A., & Woolrich, M. W. Combined spatial and non-spatial prior for inference on MRI time-series, *Neuroimage,* (2009), 45: 795-809.

[54] Hartvig, N. V., & Jensen, J. L. Spatial mixture modeling of fMRI data, *Hum Brain Mapp,* (2000), 11: 233-248.

[55] Woolrich, M., Behrens, T, Beckmann, C, & Smith, S. Mixture models with adaptive spatial regularization for segmentation with an application to fmri data, *IEEE Transactions on Medical Imaging,* (2005), 24: 1-11.

[56] Xia, J., Liang, F., & Wang, Y. M. On clustering fMRI using Potts and mixture regression models, in *IEEE Engineering in Medicine and Biology Society Conference*, (2009), pp. 4795-4798.

[57] Flandin, G., & Penny, W. D. Bayesian fMRI data analysis with sparse spatial basis function priors, *Neuroimage,* (2007), 34: 1108-1125.

[58] Fergusson, T. A Bayesian analysis of some nonparametric problems, *Annals of Statistics,* (1973), 1: 209-230.

[59] Kim, S., & Smyth, P. Hierarchical Dirichlet Processes with random effects, in *Neural Information Processing Systems*, (2006), pp. 697-704.

[60] Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. Classical and Bayesian inference in neuroimaging: theory, *Neuroimage,* (2002), 16: 465-483.

[61] Holmes, A., & Friston, K. Generalisability, random effects & population inference, in *Fourth International Conference on Functional Mapping of the Human Brain: Neuroimage*, (1998), pp. S754.

[62] Geisler, W. S., & Diehl, R. L. Bayesian natural selection and the evolution of perceptual systems, *Philos Trans R Soc Lond B Biol Sci,* (2002), 357: 419-448.

[63] Ma, W. J. Organizing probabilistic models of perception, *Trends Cogn Sci,* (2012), 16: 511-518.

[64] Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. Probabilistic models of cognition: exploring representations and inductive biases, *Trends Cogn Sci,* (2010), 14: 357-364.

[65] Fiser, J., Berkes, P., Orban, G., & Lengyel, M. Statistically optimal perception and learning: from behavior to neural representations, *Trends Cogn Sci,* (2010), 14: 119-130.

[66] Vilares, I., & Kording, K. Bayesian models: the structure of the world, uncertainty, behavior, and the brain, *Ann N Y Acad Sci,* (2011), 1224: 22-39.

http://dx.doi.org/10.5772/55262

[67] Knill, D. C., & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation, *Trends Neurosci,* (2004), 27: 712-719.

[68] Atkins, J. E., Fiser, J., & Jacobs, R. A. Experience-dependent visual cue integration based on consistencies between visual and haptic percepts, *Vision Res,* (2001), 41: 449-461.

[69] Ernst, M. O., & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion, *Nature,* (2002), 415: 429-433.

[70] Weiss, Y., Simoncelli, E. P., & Adelson, E. H. Motion illusions as optimal percepts, *Nat Neurosci,* (2002), 5: 598-604.

[71] Kording, K. P., & Wolpert, D. M. Bayesian integration in sensorimotor learning, *Nature,* (2004), 427: 244-247.

[72] Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. Bayesian inference with probabilistic population codes, *Nat Neurosci,* (2006), 9: 1432-1438.

[73] Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. Exemplar models as a mechanism for performing Bayesian inference, *Psychon Bull Rev,* (2010), 17: 443-464.

[74] Sanger, T. D. Probability density estimation for the interpretation of neural population codes, *J Neurophysiol,* (1996), 76: 2790-2793.

[75] Huys, Q. J. M., Zemel, R. S., Natarajan, R., & Dayan, P. Fast population coding, *Neural Computation,* (2007), 19: 404-441.

[76] Deneve, S. Bayesian spiking neurons I: inference, *Neural Comput,* Jan (2008). , 20, 91-117.

[77] Jazayeri, M., & Movshon, J. A. Optimal representation of sensory information by neural populations, *Nat Neurosci,* (2006), 9: 690-696.

[78] Murphy, K. P. *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. Thesis, Department of Computer Science, University of California, Berkeley, (2002).

[79] Bishop, C. M. *Pattern Recognition and Machine Learning*, Springer Science + Business Media, LLC, (2006).

[80] Kschischang, F. R., Frey, B. J., & Loeliger, H.-A. Factor graphs and the sum-product algorithm, *IEEE Transactions on Information Theory,* (2001), 47: 498-519.

[81] Peot, M. A., & Shachter, R. D. Fusion and propagation with multiple observations in belief networks, *Artificial Intelligence,* (1991), 48: 299-318.

[82] Zheng, X., & Rajapakse, J. C. Learning functional structure from fMR images, *Neuroimage,* (2006), 31: 1601-1613.

[83] Rajapakse, J. C., & Zhou, J. Learning effective brain connectivity with dynamic Bayesian networks, *Neuroimage,* (2007), 37: 749-760.

[84] Li, J., Wang, Z. J., & Mckeown, M. J. Multi-subject, A. dynamic Bayesian networks (DBNs) framework for brain effective connectivity, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2007). pp. I-429 – I-432.

[85] Chen, R., & Herskovits, E. H. Network analysis of mild cognitive impairment, *Neuroimage,* (2006), 29: 1252-1259.

[86] Chen, R., Resnick, S. M., Davatzikos, C., & Herskovits, E. H. Dynamic Bayesian network modeling for longitudinal brain morphometry, *Neuroimage,* Feb 1 (2012). , 59, 2330-2338.

[87] Huang, S., Li, J., Ye, J., Fleisher, A., Chen, K., & Wu, T. Brain effective connectivity modeling for Alzheimer's disease study by sparse Bayesian network, in *The Seventeenth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining* (2011), pp. 931-939.

[88] Song, L., Kolar, M., & Xing, E. P. Time-varying dynamic Bayesian networks, in *Proceeding of the 23rd Neural Information Processing Systems*, (2009), pp. 1732-1740.

[89] Brunel, N., & Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition, *J Comput Neurosci,* (2001), 11: 63-85.

[90] Rolls, E. T., Deco, G., & Loh, M. *A Stochastic Neurodynamics Approach to the Changes in Cognition and Memory in Aging*, (2010).

[91] Kelly, K. M., Nadon, N. L., Morrison, J. H., Thibault, O., Barnes, C. A., & Blalock, E. M. The neurobiology of aging, *Epilepsy Res,* Suppl 1, 2006), 68: S5-20.

[92] Dere, E., Easton, A., Nadel, I., & Huston, J. P. *Handbook of Episodic Memory*, Elsevier, Amsterdam, (2008).

[93] Goldman-Rakic, P. S. The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia, *Biol Psychiatry,* (1999), 46: 650-661.

[94] Loh, M., Rolls, E. T., & Deco, G. Statistical fluctuations in attractor networks related to schizophrenia, *Pharmacopsychiatry,* (2007), 40: S78-S84.

[95] Sikstrom, S. Computational perspectives on neuromodulation of aging, *Acta Neurochir Suppl,* (2007), 97: 513-518.

[96] Burke, S. N., & Barnes, C. A. Neural plasticity in the ageing brain, *Nat Rev Neurosci,* (2006), 7: 30-40.

[97] Kesner, R. P., & Rolls, E. T. Role of long-term synaptic modification in short-term memory, *Hippocampus,* (2001), 11: 240-250.

[98] Schliebs, R., & Arendt, T. The significance of the cholinergic system in the brain during aging and in Alzheimer's disease, *J Neural Transm,* (2006), 113: 1625-1644.

http://dx.doi.org/10.5772/55262

[99] Rolls, E. T., & Deco, G. *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*, Oxford University Press, (2012).

[100] Wang, M. Y., Zhou, C., & Xia, J. Statistical analysis for recovery of structure and function from brain images, in *Biomedical Engineering, Trends, Researches and Technologies*, Komorowska, M. A. & Olsztynska-Janus, S., Eds., (2011), pp. 169-196.