Basic mechanisms of DBS for Parkinson’s disease: computational and experimental studies on neural dynamics
Çanan, H.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

Dynamic Causal Modeling: Inferring sub-cortical changes in connectivity strengths due to dopaminergic therapy

Authors

Hayriye Cagnan, Vladimir Litvak, Rosalyn Moran, Alexandre Eusebio, Alek Pogosyan, Peter Brown, Wytse J. Wadman, Hubert C.F. Martens, Karl Friston

Manuscript in preparation
6.1 Introduction

The basal ganglia-thalamocortical loop exhibits continuous synchronized oscillations in the theta (3.5-7 Hz) and beta (13-30 Hz) frequency bands during Parkinson's disease (PD) (Boraud et al. 2005; Brown 2003; Brown et al. 2001; Cassidy et al. 2002; Chen et al. 2007; Hammond et al. 2007; Kuhn et al. 2008; Kuhn et al. 2006; Kuhn et al. 2005; Kuhn et al. 2009; Kuhn et al. 2004; Trottenberg et al. 2006). Several studies used local field potential or single/multi unit micro electrode recordings obtained from sub-cortical regions to study changes which occur in neuronal activity patterns under different experimental conditions, such as during movement or following LevoDopa administration (i.e. dopamine enhancement therapy). LevoDopa administration, used in the management of PD motor symptoms, suppresses continuous synchronized beta band oscillations in the Basal Ganglia (Kuhn et al. 2006). Additionally, experimental studies indicate that stimulation at frequencies within the beta band exacerbates motor impairments in 6-OHDA lesioned rodents and in PD patients (Chen et al. 2007; Eusebio et al. 2008; Gradinaru et al. 2009). These studies have led to the hypothesis that continuous synchronized oscillations in the beta band observed in the basal ganglia-thalamocortical loop are correlated with PD motor symptoms such as tremor, rigidity and bradykinesia (Brown 2003; Brown et al. 2001; Kuhn et al. 2006; Kuhn et al. 2009).

In the classical model of the basal ganglia-thalamocortical loop, dopaminergic projections from the substantia nigra pars compacta predominantly target the striatal neurons (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990). Moreover, in this model, changes in the basal ganglia-thalamocortical loop, attributed to the loss of dopaminergic neurons during PD, are reflected as an imbalance of the indirect and direct pathways and hyperactivity of the subthalamic nucleus (STN) and the globus pallidus internum (GPI) neurons, as a direct consequence of this imbalance (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990). Recent experimental evidence on the other hand indicates that the substantia nigra pars compacta dopaminergic neurons project onto several other nuclei within the basal ganglia-thalamocortical loop (Joel and Weiner 2000; Smith and Kieval 2000).
Eusebio et al. (2009) investigated dopamine-level dependent changes that occur in the basal ganglia-thalamocortical network activity. It has been hypothesized that the decrease observed in dopamine levels during PD facilitates resonance in the basal ganglia-thalamocortical loop at frequencies around 20 Hz. This hypothesis was tested by applying short pulses at the STN at various frequencies (i.e. 5-30 Hz), while recording the cortical evoked potentials (cEPs) (Eusebio et al. 2009). Results indicated that the basal ganglia-thalamocortical network has a tendency to resonate at around 20 Hz during PD and dopaminergic treatment alters this resonant behavior (Eusebio et al. 2009).

The modulatory effect of dopaminergic treatment during PD on the network dynamics and properties of the basal ganglia-thalamocortical loop remains unknown. Here, we employ dynamic causal modeling (DCM) to study the physiological causes underlying the changes observed in the cEPs arising from administration of Levodopa (Eusebio et al. 2009). DCM has been previously used to study interactions between different cortical regions and to infer changes in system parameters that occur as a result of different experimental paradigms (Garrido et al. 2007; Stephan et al. 2007; Stephan et al. 2010). In this chapter, we use DCM to study interactions between a cortical region and underlying sub-cortical regions. The model describing the connectivity within the basal ganglia-thalamocortical loop is based on the well established model of Alexander et al. (1986) and DeLong et al. (1990) (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990). We use neural mass models to characterize nuclei making up the basal ganglia-thalamocortical loop, preserving the type of projection neurons and projection types (i.e. Glutamatergic, GABAergic) and estimate the parameters that determine the underlying neural dynamics using the cEPs and the network architecture of the basal ganglia-thalamocortical loop. Three variants of the model, which differ with respect to connections that are allowed to exhibit off-on dopamine-specific changes, are compared using Bayesian Model Comparison (Garrido et al. 2007; Penny et al. 2004; Stephan et al. 2009; Stephan et al. 2010). This comparison was used to establish which connectivity changes within the basal ganglia-thalamocortical loop best explain the differences in the cEPs arising from administration of Levodopa. Our focus was on models that limit these dopamine-specific changes to
connections within the Basal Ganglia, pathways involving the cortex or both.
6.2 Materials and Methods

cEPs analyzed in this chapter were obtained from PD patients undergoing Deep Brain Stimulation (DBS) surgery, as described previously by Eusebio et al. (2009). Sections on patients and protocol (6.2.1) and data analysis (6.2.2) are abridged accounts of the previously published data acquisition and analysis techniques (Eusebio et al. 2009).

6.2.1 Patients and Protocol

Implantation of the DBS electrode and acquisition of the data took place at two centers; Institute of Neurology in London and Timone University Hospital in Marseille. Patients participating in this study participated with written informed consent and the permission of the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology and the local Ethics Committee (CPP Marseille 2). cEPS were elicited with an implanted Kinetra pulse generator and a DBS electrode (model 3389: Medtronic, Minneapolis, USA). DBS electrodes were implanted at least six months prior to acquisition of the EEG recordings. Further information regarding the clinical history of the patients and the surgical procedure can be found in Eusebio et al. (2009).

Several studies have reported induction of cEPs as a result of STN-DBS (Baker et al. 2002; Eusebio et al. 2009; MacKinnon et al. 2005). The core mechanism behind generation of cEPs has been hypothesized to be orthodromic activation of fibers projecting from the Basal Ganglia onto the cortex (Baker et al. 2002; Eusebio et al. 2009; MacKinnon et al. 2005). In Eusebio et al. (2009) cEPs were induced by stimulating the STN using the DBS contacts, giving rise to the best clinical outcome to ensure activation of the fibers contributing to suppression of PD motor symptoms during STN-DBS. The recordings were made under two distinct experimental conditions: off-dopamine (i.e. following withdrawal from Levodopa) and on-dopamine (i.e. while the patient was on Levodopa). In this chapter, we analyze the recordings obtained during 5 Hz stimulation of the STN for both off-dopamine and on-dopamine conditions (Eusebio et al. 2009). Recordings made during other stimulation frequencies (i.e. 10-30 Hz) can also be used in this analysis. During acquisition of the data, bipolar stimulation was used to minimize stimulus artifact. The
Kineta stimulator stimulates the two hemispheres of the brain in an alternating fashion therefore, to avoid recording the stimulus artifact arising from stimulation of the other hemisphere, only one side, which gave rise to the best clinical efficacy, was stimulated. Off-dopamine and on-dopamine recordings were made from the same hemisphere. During data acquisition, the STN was stimulated at 5 Hz and no additional high frequency stimulation was applied during the protocol.

6.2.2 Data analysis

Figure 1: Cortical evoked potentials from 9 patients. Black lines indicate recordings made following withdrawal from Levodopa (off-dopamine) and grey lines indicate recordings made while the patient was on Levodopa (on-dopamine).

EEG recordings were obtained while the patients were at rest (Eusebio et al. 2009). Scalp EEG recordings at the Institute of Neurology in London were obtained using 19 Ag/AgCl electrodes and referenced to linked ear electrodes. Recordings made at the Timone University Hospital in Marseille were obtained using 11 Ag/AgCl electrodes and
again referenced to linked ear electrodes. Sampling rate was set at 1500 Hz for two recordings (Patients 2 and 3) and at 2048 Hz for the other seven recordings (Patients 1, 4-9). Further particulars regarding the EEG recordings can be found in Eusebio et al. (2009).

EEG recordings were analyzed using Spike® version 2.06. Artifacts observed due to eye movement or scalp muscle activation were deleted manually and excluded from further analysis. We studied recordings made from EEG channels F3, F4, C3, C4, Cz and Fz based on a previous report indicating that the amplitudes of the cEPs recorded from these channels were significantly higher than those recorded from other channels (Eusebio et al. 2009). Monopolar signals were generated by taking the difference between each channel and the average of all the channels (i.e. common mode subtraction). cEPs were obtained by averaging the signals around the stimulus artifact and low-pass filtered using a fourth order Butterworth filter with cut-off frequency of 100 Hz. Cut-off frequency of the low-pass filter was chosen well above the frequency range of interest. Mean of the filtered cEPs were also removed (i.e. mean equal to zero). Per patient, the cEP with the highest amplitude was chosen for further analysis and modeling (Fig. 1).

6.2.3 Dynamic Causal Modeling

a) The Generative Model

DCM makes use of a generative model to account for biophysical mechanisms underlying various experimental conditions (David et al. 2006; Friston et al. 2003; Garrido et al. 2007; Stephan et al. 2007; Stephan et al. 2010). Three different architectures were tested in this study: 1) architecture with a single cortical source, 2) a full basal ganglia-thalamocortical network model, and 3) a reduced basal ganglia-thalamocortical network model. Source activity was described via neural mass models (David and Friston 2003; David et al. 2006; Friston et al. 2003; Garrido et al. 2007; Marreiros et al.; Stephan et al. 2007). The single cortical source was modeled using three interacting sub-populations (two excitatory and one inhibitory) (Fig. 2B). In the full model, each source was modeled with three interacting sub-populations, two excitatory and one inhibitory, i.e. a realistic representation of a generic local neuronal network.
architecture (Fig. 2) (Alexander and Crutcher 1990; Delong 1990). In the reduced model, only the cortex was modeled with three interacting sub-populations (i.e. two excitatory and one inhibitory), while the STN and thalamus were modeled using one excitatory sub-population and the striatum, globus pallidus externum (GPe) and GPi were modeled using an inhibitory sub-population (Fig. 3A) (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990). The reduced model is based on the main projection cell types at each source (i.e. Glutamatergic versus GABAergic) (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990).

Neural mass models have been used in theoretical studies to investigate spontaneous rhythms generated by cortical or thalamo-cortical networks (David and Friston 2003; Jansen and Rit 1995; Lopes da Silva et al. 1974). Neural mass models have also been used to study event related responses of large neuronal populations (David et al. 2006; Garrido et al. 2007). Most commonly, the neuronal dynamics of each population

Figure 2: A: The full model where each source is described in terms of B: three interacting sub-populations, two excitatory and one inhibitory. Solid lines indicate inhibitory projections and dashed lines indicate excitatory projections (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990).
are described through coupled first order differential equations, where changes in the membrane potential are a function of excitatory and inhibitory currents and changes in the excitatory/inhibitory current are a function of the excitatory/inhibitory current and the membrane potential (David et al. 2006; Garrido et al. 2007; Stephan et al. 2007). The nonlinear nature of neuronal dynamics is captured with a sigmoid function that translates changes in membrane potential into firing rate, which then form synaptic input to other subpopulations (David and Friston 2003; Garrido et al. 2007; Stephan et al. 2007). Synaptic impulse response function translates firing rate into membrane potential changes (David and Friston 2003; Garrido et al. 2007; Stephan et al. 2007).

Figure 3: A: The reduced model where only the cortex is described in terms of B: three interacting sub-populations (i.e. two excitatory and one inhibitory). The STN and thalamus are described using one excitatory sub-population while the striatum, GPe and Gpi are described using an inhibitory sub-population. Solid lines indicate inhibitory projections and dashed lines indicate excitatory projections (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990).
Dynamic Causal Modeling

Neural mass models have been used to describe the mean response of an interacting population of excitatory and inhibitory neurons both in the time domain and the frequency domain (i.e. steady state response). Marreiros et al. (2009) has described a simplified mean-field model of neuronal dynamics by using the Laplacian approximation and expressing the neuronal dynamics as a function of the mean and the variance of the underlying neuronal states. Neural mass models are obtained as a special case of the mean-field model, where the variance of the neuronal states is assumed to be constant (Marreiros et al. 2009). This is the assumption we make here.

Each sub-population \( (i) \) is described through a reduced Morris and Lecar model, where membrane dynamics are described through the interaction of various ion channels lumped into excitatory and inhibitory synaptic conductances (Morris and Lecar 1981). Changes in the membrane potential (Eq. 1) and excitatory/inhibitory conductance (Eqs. 2 and 3) are described using first order differential equations (Marreiros et al. 2009; Morris and Lecar 1981).

\[
CV^i = g_L(V_L - V^i) + g_E^i(V_E^i - V^i) + g_I^i(V_I - V^i) + \exp(C) U \tag{1}
\]

\[
\begin{align*}
\dot{g}_E^i &= \left(1/(\exp(T_s) \tau_E)\right)(\varsigma_E^i - g_E^i) + (\exp(X))M \\
\dot{g}_I^i &= \left(1/\tau_i\right)(\varsigma_I^i - g_I^i) + (\exp(X))M \tag{3}
\end{align*}
\]

\[
\varsigma_k^i = \sum_j \exp \left(G_s^{ij} \nu_{ij}^k \right) S(\mu_V^i - V_R, \Sigma^i) \tag{4}
\]

\[
U = 32 \exp\left(-t^2/(64 \exp(R))\right) \tag{5}
\]

Equations 1-5 describe the dynamics of sub-population \( i \) of source \( s \) (Marreiros et al. 2009). Equation 1 describes the membrane dynamics of sub-population \( i \) as a function of excitatory and inhibitory synaptic conductances and exogenous input \( U \). Exogenous input \( U \) is used to model membrane currents induced by DBS. Equations 2 and 3 describe changes in the excitatory/inhibitory synaptic conductance as a function of the excitatory/inhibitory
synaptic conductance and the pre-synaptic membrane potential. Equation 4 describes the increase in the excitatory/inhibitory synaptic conductance due to the pre-synaptic membrane potential and equation 5 indicates the exogenous input applied at the STN (full and reduced models) or at the cortex (cortical model).

The model for the cEPs (i.e. predictions of experimental data) is derived from the average membrane potential of the glutamatergic projection cells of the cortex, following a linear transformation. The free parameters, which are estimated, enter as log-scaling factors in equations 1-5. These control excitatory synaptic time constants ($\tau_E$), excitatory synaptic strengths between sub-populations ($\gamma_{ij}^E$, $\gamma_{ji}^E$), background synaptic activity ($M$), delays in synaptic connections between different sub-populations and sources, exogenous input amplitude and dispersion, and excitatory and inhibitory synaptic strengths between sources ($L_E$, $L_I$). Excitatory synaptic time constants ($\tau_E$) and excitatory synaptic strengths between sub-populations ($\gamma_{ij}^E$, $\gamma_{ji}^E$) are optimized individually per source. Other parameters have the same value for all sources. Parameter values scaled by the scale-parameters (i.e. prior expectations) are indicated in Table 1, and prior densities of the (log) scale-parameters are shown in Table 2 (Marreiros et al. 2009).

We account for two different types of synaptic connections ($k$) between different sub-populations and sources: excitatory ($E$) and inhibitory ($I$) (Marreiros et al. 2009). Delays in synaptic connections between different sub-populations and sources are set to 2 ms and 16 ms, respectively (David and Friston 2003; David et al. 2006; Lumer et al. 1997). $\gamma_{ij}^k$ indicates intrinsic connectivity strengths (i.e. between different sub-populations (from sub-population $j$ to sub-population $i$) within the same source ($s$)) and $A_{ss}^{ij}$ indicates extrinsic connectivity strengths (i.e. from source $s$, population $j$ to source $s'$, population $i$). Sub-population 1 corresponds to GABAergic cells, sub-population 2 corresponds to glutamatergic interneurons and sub-population 3 corresponds to glutamatergic projection cells (Figs. 2 and 3). Crucially, $B_{ij}^{ss}$ accounts for the changes in extrinsic connectivity strengths that model the differences in the cEPs due to administration of Levodopa (off-dopamine to on-dopamine differences). $S(\cdot)$ is the sigmoid function which is used to map mean depolarization in a population to
expected firing rate (Marreiros et al. 2008). The exogenous input \((U)\) was applied to the glutamatergic projection cells (sub-population 3) of the STN. In the generative model using a single cortical source, the exogenous input was applied to the glutamatergic interneurons (sub-population 2), in line with previous DCM studies (Garrido et al. 2007; Stephan et al. 2007).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Physiological connotation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g_L)</td>
<td>Leak conductance</td>
<td>1 mS</td>
</tr>
<tr>
<td>(\tau_E, \tau_I)</td>
<td>Postsynaptic time constants</td>
<td>4, 16ms</td>
</tr>
<tr>
<td>(\gamma_{32}E, \gamma_{23}E, \gamma_{13}E, \gamma_{31}I, \gamma_{21}I)</td>
<td>Intrinsic connectivity</td>
<td>0.5, 0.5, 1, 1, 0.25</td>
</tr>
<tr>
<td>(V_L, V_E, V_I)</td>
<td>Reversal potential</td>
<td>-70, 60, -90 mV</td>
</tr>
<tr>
<td>(V_R)</td>
<td>Threshold potential</td>
<td>-40 mV</td>
</tr>
<tr>
<td>(L_E, L_I)</td>
<td>Extrinsic connectivity</td>
<td>0.5, 0.25</td>
</tr>
<tr>
<td>(M)</td>
<td>Synaptic background activity</td>
<td>1/8</td>
</tr>
</tbody>
</table>

Table 1: Parameter values used for generating population dynamics (Marreiros et al. 2009)

The complete set of differential equations describing the dynamics of each sub-population, together with differential equations describing the coupling between different sub-populations, can be found in Marreiros et al. (2009).

b) Prior Assumptions

Under the assumption that the prior distributions of the (log) scale-parameters (Table 1) are Gaussian, the prior distribution of each parameter is described by its mean and variance (Marreiros et al. 2009). These prior densities are shown in Table 2 (Marreiros et al. 2009).

Priors on \(B_{ij}^{sl}\) determine where dopamine-level dependent differences are expressed. These priors define the key variants of the models we were interested in comparing. We examined three variants (A, B and C) corresponding to models in which dopamine exerted its effects on connections within the Basal Ganglia (model B), connections involving the
cortex (including the hyper-direct pathway; model C) and a model in which dopaminergic effects can be expressed in all extrinsic connections (model A). Heuristically, model A subsumes models B and C (Fig. 4).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{ij}^{ss}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$B_{ij}^{ss}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$C$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$R$</td>
<td>0</td>
<td>1/16</td>
</tr>
<tr>
<td>$D$</td>
<td>0</td>
<td>1/128</td>
</tr>
<tr>
<td>$X$</td>
<td>0</td>
<td>1/128</td>
</tr>
<tr>
<td>$G^S$</td>
<td>0</td>
<td>1/32</td>
</tr>
<tr>
<td>$T^S$</td>
<td>0</td>
<td>1/32</td>
</tr>
</tbody>
</table>

Table 2: Prior densities of parameters under Gaussian assumption

Experimental studies highlighted the modulatory effect dopaminergic receptors have on Ca$^{2+}$ and K$^+$ channels and on Na$^+$-K$^+$ pumps. In DCM, changes in connection strengths between different brain regions are used in order to account for differences in experimental conditions (i.e. altered dopamine levels). Synaptic connections between different regions are inputted directly to the equations defining dynamics of a specific brain region (Eq. 1). As a result, modulatory effect of dopamine on excitability of a nucleus is captured through altered synaptic connection strengths impinging on that nucleus.

Models A, B and C determine where dopamine can change sensitivity to extrinsic afferents. Based on data obtained from MPTP treated monkeys it was hypothesized that degeneration of dopaminergic neurons in the substantia nigra pars compacta gives rise to increased inhibition of the GPe and decreased inhibition of the GPi via the indirect and direct pathways, respectively. A model was put forward based on these connection changes (i.e. classical model of the Basal Ganglia) (Alexander et al. 1986) accounting for alterations in firing rates in the basal ganglia-thalamocortical network during PD. In this model, imbalance of activity in the direct and indirect pathways was used to account for hyperactivity of the STN and GPi, observed following reduction in dopamine levels with respect to the normal state. In model B, it is
assumed that the same connections hypothesized to be altered following degeneration of dopaminergic neurons in the classical model of the Basal Ganglia, change as a result of dopamine replacement therapy (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990).

In the classical model, hyperactivity of the STN during PD is associated with increased inhibition of the GPe via indirect pathway. In addition to receiving inhibitory input from the GPe, the STN receives excitatory input from the cortex via the hyper-direct pathway. Experimental studies highlighted that activity levels in the STN increase prior to depletion of the striatal dopamine and that hyper-activity of the STN is not necessarily accompanied by a reduction in the GPe activity levels (Bezard et al. 1999; Hartmannvonmonakow et al. 1978; Nambu et al. 1996; Nambu et al. 2002; Obeso et al. 2000; Vila et al. 2000). STN-GPe network is capable of generating sustained oscillatory activity patterns in the theta frequency band (Terman et al. 2002). The hyper-direct pathway on the other hand has been shown to play a crucial role in the generation of sustained oscillations in the BG nuclei in the beta frequency band during PD (Baufreton et al. 2005; Beurrier et al. 2001; Wilson et al. 2006). Taking these observations into account, we test the hypothesis that changes in the basal ganglia-thalamocortical loop are dependent on the hyper-direct pathway. Therefore, in model C, we allow the hyper-direct pathway and the down-stream connections from the STN to GPi, GPi to thalamus and thalamus to cortex to change.

Last but not least, in model A which is a combination of models B and C, all connections between different sources are allowed to change to test the hypothesis that altering dopamine levels have wide-spread effects on all the connections and alters the indirect, direct and hyper-direct pathways.

c) Bayesian Model Comparison

Bayesian inference makes use of the posterior probability on the parameters \( p(\theta | y, m) \) to summarize our current knowledge on model parameters after incorporating the likelihood of observing the data \( y \) given a model, \( m \), and its parameters, \( \theta \), with our a priori knowledge on the model parameters, \( \theta \), expressed through the prior probability
Additionally, Bayesian inference provides a measure for how well a model explains a data set (i.e. model evidence, $p(y|m)$). Model evidence can also be used to infer among a group of models which one best describes the given data set. This property is used in this study to infer among three variants of the generative model (Figs. 2 and 3) which one best captures off-on dopamine specific changes in the cEPs. Classical statistics on the other hand allows for testing a hypothesis against an alternative hypothesis and does not provide a measure for how well a given hypothesis explains the data. One potential constraint of Bayesian analysis is the strong dependence of the inference made on the prior probability, $p(\theta, m)$. Dependence on the prior probability can be minimized by making use of uninformative priors. For instance, in this study, priors on extrinsic coupling ($A_{ij}^{ss}, B_{ij}^{ss}$) and exogenous input amplitude ($C$) are relatively uninformative priors (i.e. variance of the prior densities is one).

According to the Bayes' rule, the posterior probability of the parameters ($\theta$) conditioned upon the data ($y$) and the model ($m$) is (Friston et al. 2007; Garrido et al. 2007; Penny et al. 2004; Stephan et al. 2009):

$$p(\theta|y, m) = \frac{p(y|\theta, m)p(\theta, m)}{p(y|m)}$$

(6)

The model evidence $p(y|m)$, which is the probability of obtaining the observed data given model, $m$, is defined as (Friston et al. 2007; Garrido et al. 2007; Stephan et al. 2009):

$$p(y|m) = \int p(y|\theta, m)p(\theta, m)d\theta$$

(7)

In DCM, the model evidence is approximated using variational Bayes, where an approximating posterior density $q(\theta)$ is proposed and is optimized with respect to a free energy bound ($F$) on the log-evidence (i.e. $\ln(p(y|m))$). More information on approximating the model evidence using variational Bayes can be found in Stephan et al. (2009) and Friston et al. (2007).

Free energy bound ($F$) is also referred to as negative-free energy and provides a lower bound for the log-evidence (Friston et al. 2007; Stephan et al. 2009).
\[ F = \ln(p(y|m)) - \text{KL}[q(\theta), p(\theta|y, m)] \]  

(8)

The second term in equation 8 is the Kullback-Leibler divergence between the approximating posterior density \( q(\theta) \) and the posterior probability of the parameters conditioned upon the data and the model, \( p(\theta|y, m) \) (Kullback and Leibler 1951; Penny et al. 2004; Stephan et al. 2009). When the approximating posterior density, \( q(\theta) \), is equivalent to the posterior probability, \( p(\theta|y, m) \), the negative-free energy is equivalent to the log evidence (Friston et al. 2007; Stephan et al. 2009). The definition of negative-free energy can be rearranged such that it is partitioned into accuracy (i.e. \( \langle \ln(p(y|\theta, m)\rangle_q \)) and complexity terms (Stephan et al. 2009).

\[ F = \langle \ln(p(y|\theta, m))\rangle_q - \text{KL}[q(\theta), p(\theta|m)] \]  

(9)

The complexity term is the Kullback-Leibler divergence between the approximating posterior density \( q(\theta) \) and the prior probability on the parameters \( p(\theta|m) \), which is equivalent to (Stephan et al. 2009):

\[
\text{KL}[q(\theta), p(\theta|m)] = \frac{1}{2} \ln |C_\theta| - \frac{1}{2} \ln |C_{\theta|y}| \\
+ \frac{1}{2} (\mu_{\theta|y} - \mu_\theta)^T C_\theta^{-1}(\mu_{\theta|y} - \mu_\theta)
\]  

(10)

where \( C_\theta \) is the prior covariance, \( C_{\theta|y} \) is the posterior covariance, \( \mu_{\theta|y} \) is the posterior mean and \( \mu_\theta \) is the prior mean (Stephan et al. 2009). The first term defines the level of dependence of the parameters a priori; the second term defines the level of dependence of the parameters a posteriori, while the third term describes the distance between the posterior and prior means (Stephan et al. 2009). Hence, the free-energy is sensitive to the accuracy of the model but also penalizes complexity (e.g., number of free parameters) in an optimum way.

For a given data set \( (y) \), the model describing the data best is the model with the highest log-evidence \( \ln p(y|m) \) (Friston et al. 2007; Garrido et al. 2007; Penny et al. 2004; Stephan et al. 2009). In this chapter, we compare three variants (i.e. Models A, B and C) of the generative model with full and reduced architectures to null models, where none of the connections among sources were allowed to change (Fig. 4).
\[ \ln B_{i,\text{null}} = \ln p(y|m_i) - \ln p(y|m_{\text{null}}) \]  

A positive value for \( \ln B_{i,\text{null}} \) indicates that model \( i \) describes the data better than the null model (i.e. higher log-evidence) and a negative value for \( \ln B_{i,\text{null}} \) indicates that the null model describes the data better than model \( i \) (i.e. the null model has higher log-evidence). A difference in free-energy (log-evidence) of three or more which corresponds to a p-value less than or equal to 0.05 is usually considered significant (i.e. a log-odds ratio of 20:1) (Penny et al. 2004). In what follows, we will present the results of model comparison for each subject separately and for the entire data over all subjects. This is achieved simply by summing the log-evidences for each model over subjects, under the assumption that each subject’s data are conditionally independent and the optimal model is the same for each subject in the population. The latter assumption holds when studying a physiological mechanism that is improbable to vary across subjects (Stephan et al. 2010) which in this study is the modulatory effect of altering dopamine levels. The subject-specific profiles of model log-evidence are used to indicate the between-subject consistency of inference on models; while definitive conclusions are based on the evidence for the different models at the group level. Here, we were primarily interested in selecting among models A, B and C; and, secondarily, in whether the use of reduced models provided better models.

Furthermore, Bayesian parameter averaging is used to infer average changes in model parameters across all subjects using \( \mu = \Lambda^{-1} \sum_{l=1}^{9} A_l \mu_l \) and \( \Lambda = \sum_{l=1}^{9} A_l \), where \( A_l = \sigma_l^{-1} \), under the assumption that posterior distribution on the model parameters are Gaussian for each subject \( l \) (i.e. \( p(\theta_l|y_l,m) = N(\mu_l,\sigma_l) \)). Comparable to the assumptions made for performing model comparison over all subjects, Bayesian parameter averaging can be used under the assumption that the optimal model is identical for all subjects (Stephan et al. 2010). Bayesian parameter averaging takes into account variances of the model parameters together with interdependencies of model parameters through the covariance matrix \( \sigma_l \).
d) Software note

The software used in this study is available as part of the academic freeware SPM version 8b (downloaded from www.filion.ucl.ac.uk/spm).

Figure 4: Variations of the generative model tested to account for off-dopamine to on-dopamine changes observed in the cEPs: Model A: All connections between sources are allowed to change; Model B: Connections used in the classical model of Basal Ganglia put forward by DeLong et al. (1990) to account for alterations in the network arising from dopamine depletion are allowed to change; Model C: Hyper-direct pathway and the connections down-stream the hyper-direct pathway are allowed to change. (Connections which are allowed to change are indicated with dashed lines and connections which are not allowed to change are indicated with solid lines)
6.3 Results

6.3.1 Cortex

First, we considered the generative model with a single cortical source. The exogenous input was applied at the second sub-population of the cortical source. The total number of free parameters was six (i.e. excitatory synaptic time constant ($\tau_E$), excitatory synaptic strength between sub-populations ($\gamma_{32}^E, \gamma_{23}^E, \gamma_{13}^E$), background synaptic activity ($M$), delay in synaptic connections between different sub-populations, and exogenous input amplitude and dispersion). Figure 5 shows the cEP and the fit for one patient data set (i.e. cEP obtained off-dopamine and cEP obtained on-dopamine). This generative model can only capture the initial peak in the cEPs and not the oscillatory nature of the cEP that follows. Similar results were observed for other patient data sets (data not shown). This model inversion (i.e. fit) is presented to show the need to use more realistic models, as outlined below.

Figure 5: cEP and the fit obtained using single cortical source as the generative model. cEPs obtained A: off-dopamine and B: on-dopamine are indicated with solid lines and the fits are indicated with dashed lines. Solid grey lines indicate one standard error of the mean (i.e. cEP). None of the generative model parameters were optimized to account for off-on dopamine-specific changes (i.e. fits for off-dopamine and on-dopamine are equivalent).
6.3.2 Full Model

We next inverted the full model where all six sources (i.e. the cortex, STN, GPe, GPi and thalamus) were each represented with three sub-populations (Fig. 2). Fifteen parameters describing source dynamics, two parameters defining exogenous input characteristics, and eleven parameters describing connection strengths between sources were estimated. Depending on the model used to describe the off-on dopamine specific changes in the cEPs, parameters, capturing the increase/decrease in the connection strengths, were also estimated. For model A eleven, for model B six and for model C four additional parameters were estimated (Fig. 4).

Figure 6: cEP and the fit obtained using the full model as the generative model. cEPs obtained A: off-dopamine and B: on-dopamine are indicated with solid lines and the fits are indicated with dashed lines. Solid grey lines indicate one standard error of the mean (i.e. cEP). All the extrinsic connections between sources were allowed to change to account for off-on dopamine specific changes (Fig. 4A).

Figure 6 shows the cEP and the fit for one patient data set. The generative model can now model both the initial peak in the cEPs and the ensuing oscillatory nature of the cEP, following the initial peak. Similar results were observed for other patient data sets (data not shown).

Using Bayesian model comparison, we compared three variants of this full model (Fig. 4 A, B and C), which differed
with respect to connections that were allowed to exhibit off-on dopamine-specific changes, with respect to a null model.

For six patient data sets out of nine, the best model was the model allowing all the extrinsic connection strengths between sources to change to account for off-on dopamine-specific changes (Patients 2-7) (Fig. 7A). In one patient data set the dopamine specific changes were best described by model C (Patient 9) (Fig. 7A). For two patient data sets, the null model was favored over the three variants (Patients 1 and 8) (Fig. 7A). The best model per patient was determined based on the differences in log-evidences (i.e. a difference in log-evidence of three or more which corresponds to a p-value less than or equal to 0.05 (Penny et al. 2004)).

6.3.3 Reduced Model

We then repeated the above procedure using the reduced model where the STN, striatum, GPe, GPi and thalamus were described using one population (Fig. 3). Since the STN, striatum, GPe, GPi and thalamus are each described by one population; excitatory synaptic strengths between sub-populations are not optimized for these populations, resulting in a difference of five parameters in the total number of parameters estimated with respect to the full model.

A

Full Model
Figure 7: Log evidence of the variations of the generative model with respect to null models: Model A: All connections between sources are allowed to change (Fig. 4A); Model B: Connections utilized in the classical model of the basal ganglia-thalamocortical loop put forward by DeLong et al. (1990) are allowed to change (Fig. 4B); Model C: Hyper-direct pathway and the connections downstream the hyper-direct pathway are allowed to change (Fig. 4B). The generative models were tested using full (Fig. 7A) and reduced architectures (Fig. 7B).

For five patient data sets out of nine, the best model was model A, describing the off-on dopamine-specific changes based on changes in all the extrinsic connection strengths between sources (Patients 1, 2, 3, 6, 8) (Fig. 7B). In two patient data sets dopamine specific changes were best captured through the classical changes outlined by DeLong et al. (1990) (Patients 4 and 7) while in one patient data set the dopamine specific changes were described through the hyper-direct pathway and the projections downstream the hyper-direct pathway (Patient 5). For one patient data set, the null model was the best model (Patient 9) (Fig. 7B).
6.3.4 Synthesis of model comparison

Figure 8: Log-evidence (free-energy) over all the models tested, after pooling (summing) over subjects. These models are based on models A, B and C (which determine where dopamine can change sensitivity to extrinsic afferents). For each of these models, we considered full and reduced architectures.

To make final inferences about the different models, we pooled the model evidence over all subjects. Figure 8 shows the log-evidence (free-energy) over all the models tested, after summing over subjects. These models are based on models A, B and C (i.e. which determine connections that are allowed to vary with changes in dopamine levels). For each of these models, we considered full and reduced architectures. One can see that the best class of model was of variant A for both full and reduced architectures. Interestingly, the reduced architecture had a greater log-evidence for every variant of the model. This is important because the reduced models had fewer parameters, suggesting that the reduction in complexity more than compensated for the decrease in accuracy, when using reduced models informed by main projection cell types at each source. It should be noted that inferences made only holds for the nine subjects used in this study.

In summary these results are clear evidence that dopamine affects both the connections within the Basal Ganglia and the hyper-direct pathway (and downstream connections). This is because a model that allowed for changes in both subsets of connections was better than the models that limited changes to one or the other. In this context, a change in connection strength means that the sub-populations become more or less sensitive to extrinsic afferents. Quantitatively, this change in
gain is encoded by the $B_{ij}^{ss}$ parameters. Figure 9 shows these average changes over all subjects. Projections between the cortex and the thalamus were altered the least (4 %) while the connection strength from the GPi to the thalamus shows the largest reduction following Levodopa administration. Additionally, the sensitivity to the hyper-direct pathway and the excitatory projection from the STN to GPi decreases. According to the classical model of the Basal Ganglia, reduction in dopamine levels is accompanied with increased inhibition of the GPe via the indirect pathway and decreased inhibition of the GPi via the direct pathway. Changes in the striatum-GPe and striatum-GPi connections estimated over all subjects show increased inhibition of the GPe and decreased inhibition of the GPi following Levodopa administration.

Figure 9: Changes in connection strengths estimated over all subjects using Bayesian parameter averaging. Values indicate percentage increase (+) or decrease (-) in connection strengths for on-dopamine condition with respect to connection strengths obtained for off-dopamine condition (i.e. $B_{ij}^{ss}/A_{ij}^{ss}$). The estimate applies only for the nine subjects analyzed and inferences cannot be made regarding the entire population the subjects are chosen from.
6.4 Discussion

Continuous synchronized oscillations in the 3-30 Hz frequency range are observed in the basal ganglia-thalamocortical loop during PD (Boraud et al. 2005; Brown 2003; Brown et al. 2001; Cassidy et al. 2002; Chen et al. 2007; Hammond et al. 2007; Kuhn et al. 2008; Kuhn et al. 2006; Kuhn et al. 2005; Kuhn et al. 2009; Kuhn et al. 2004; Trottenberg et al. 2006). Administration of anti-parkinsonian drugs or surgical interventions such as lesioning or DBS suppresses these synchronized oscillations and alleviates PD motor symptoms (Kuhn et al. 2006). This observation has lead to the hypothesis that continuous synchronized oscillations in the 3-30 Hz frequency range are linked to the PD motor symptoms (Brown 2003).

It has been suggested that loss of dopaminergic neurons in the substantia nigra pars compacta during PD exposes the basal ganglia-thalamocortical loop to resonance around 20 Hz (Eusebio et al. 2009). To test this hypothesis, Eusebio et al. (2009) recorded EEGs during STN-DBS both when the patients were off anti-parkinsonian drugs and when the patients were on anti-parkinsonian drugs. It was observed that amplitude of the cEPs during 20 Hz STN-DBS were larger than those recorded at other frequencies (Eusebio et al. 2009). Moreover, administration of anti-parkinsonian drugs modulated the cEPs (Eusebio et al. 2009). The cEPs were successfully fitted using a damped oscillator model, with natural frequency around 20 Hz (Eusebio et al. 2009). The damped oscillator model revealed that changes arising in cEPs due to administration of Levodopa could be captured as an increase in the damping factor of the damped oscillator model (Eusebio et al. 2009).

We have studied the potential biophysical mechanisms underlying these phenomena using DCM. The generative model employed in DCM is based on the well established model of the basal ganglia-thalamocortical network (Alexander and Crutcher 1990; Alexander et al. 1986; Delong 1990). DCM allows one to estimate parameters describing the underlying neuronal dynamics given a data set and a generative model (Garrido et al. 2007; Stephan et al. 2007). The a priori assumptions made by DCM are that changes in dopamine levels are captured solely as changes in connection strengths between sources (i.e. a selective change in post-
Dynamic Causal Modeling

synaptic sensitivity to afferent extrinsic connections). In this chapter, we tested three models to account for dopamine specific changes: model A: all connections between sources can change; model B: connections within the Basal Ganglia can change; model C: the hyper-direct pathway and connections making up the outer loop of the basal ganglia-thalamocortical loop can change. Model A is based on the hypothesis that dopaminergic treatment does not only modify the connection strengths within the Basal Ganglia and has an effect on the connection strengths between all the sources within the basal ganglia-thalamocortical network. Model B is based on the classical model of the Basal Ganglia and captures the hypothesis that dopaminergic treatment solely modifies the connection strengths within the Basal Ganglia (Alexander and Crutcher 1990; Delong 1990). Model C is derived from the hypothesis that administration of Levodopa has an effect on the hyper-direct pathway and the connections making up the outer loop of the basal ganglia-thalamocortical loop; i.e. the smallest loop connecting the hyper-direct pathway back to the cortex. One can regard Model A as a combination of the two more restricted hypotheses about the site of action of dopamine.

Bayesian model comparison was used to compare three variants (Fig. 4) of the generative models (i.e. full and reduced architectures (Figs. 2-3)) to evaluate the evidence in favor of one model with respect to the other one. Bayesian model comparison exploits the negative free energy in model evaluation. The negative free energy is made up of a complexity and an accuracy term (Eqs. 9 and 10). In DCM, the more accurate model is favored while the more complex model is penalized. The outcome of model comparison is solely dependent on the models evaluated and does not eliminate the possibility that there can be other equally or more plausible models.

Model comparison provided clear evidence, at the group level, that dopamine affects both the connections within the Basal Ganglia and the hyper-direct pathway (and downstream connections). This is because a model (i.e. model A) that allowed for changes in both subsets of connections was better than models that limited changes to one or the other (models B and C) (Fig. 8).

We also tested for the effect of source modeling. In the reduced basal ganglia-thalamocortical network model,
sources STN, striatum, GPe, GPi and thalamus, were modeled with a single population instead of three interconnected populations. As a result, in the reduced model, neuronal dynamics were described with fewer parameters with respect to the full model (i.e. five parameters). With the reduced model, the range of dynamics each source is able to exhibit is inherently limited, which means the complexity of these models is smaller; however, as we have seen, their evidence can surpass more complex models. The relative difference between the models A, B and C was similar under full and reduced architectures (Fig. 8).

Based on the description of the generative model and the set of priors used, log-evidences of the models varied per patient (Fig. 7). More than 50% of the patient data sets were best described by model A (for both full and reduced models), where all extrinsic connection strengths between sources change to take into account off-on dopamine-specific changes. It is interesting to note that, for all the generative models tested in this chapter, only a small subset of the patient data was best described by model B, where off-on dopamine-specific changes are captured through the changes in the connection strengths in the Basal Ganglia and model C, where dopamine modulates sensitivity to the hyper-direct pathway and down-stream connections. Our results indicate Levodopa administration leads to wide-spread effects or changes in the connectivity strengths that include both the Basal Ganglia and the hyper-direct cortex-STN-thalamocortical loop.

One of the most important challenges involved with using DCM to model data from PD patients is the prior densities of parameters used. Standard prior densities used in DCM are based on literature and are aimed at capturing the characteristics of a “healthy” brain. In the future, making use of a larger patient data set could enable identification of different prior densities derived from posterior densities averaged across patients. In this study, the striatum, STN, GPe, GPi and thalamus contribute to the parameter estimates and the fit for the experimental data through the network topology and specifically via the projections onto the cortex. Lack of (invasive) source data, which could be used to constraint the dynamics of the striatum, STN, GPe, GPi and thalamus puts even further emphasis on the prior densities. Incorporation of experimental observations highlighting the
effects of dopamine depletion on the dynamics may be a crucial step in the future applications of DCM.

In conclusion, DCM is a powerful and unique tool which provides a new platform for theoretical studies involving PD patients, where one can make use of data in the form of local field potentials and cEPs to make inferences about changes in network dynamics due to different experimental conditions.
References


Lumer ED, Edelman GM, and Tononi G. Neural Dynamics in a Model of the Thalamocortical System. 1. Layers, Loops and


