UNIVERSITEIT GENT

Dynamic Causal Modeling

Hannes Almgren, Frederik Van De Steen, Daniele Marinazzo

Department of Data Analysis
University of Ghent, Belgium

hannes.almgren@ugent.be; frederik.vandesteen@ugent.be;
daniele.marinazzo@ugent.be

www.ugent.be/~dmarinaz

@Hannes_Almgren @Frederikvds_ @dan_marinazzo

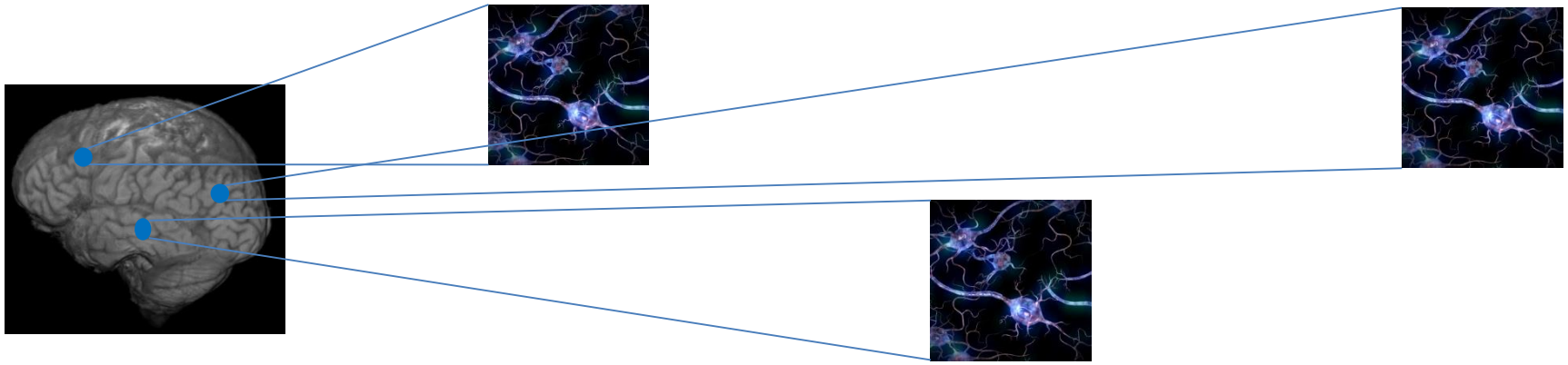
UNIVERSITEIT GENT

Model of brain mechanisms

UNIVERSITEIT GENT

Neural model

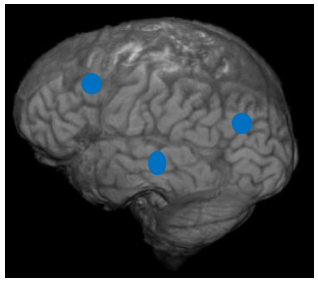
Neural populations

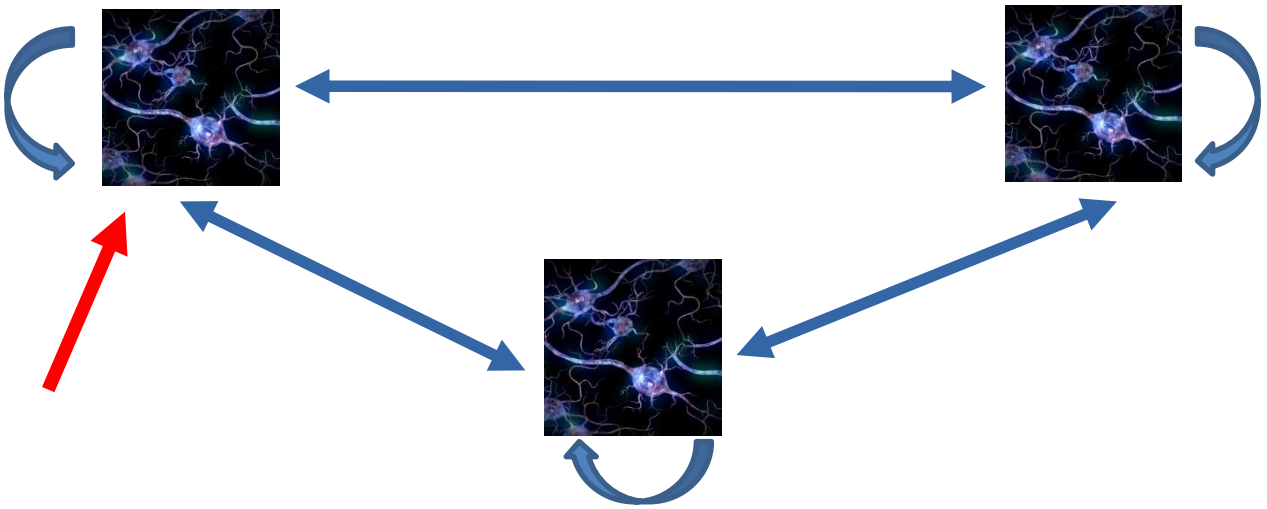


Neural model

Interactions between and within neural populations

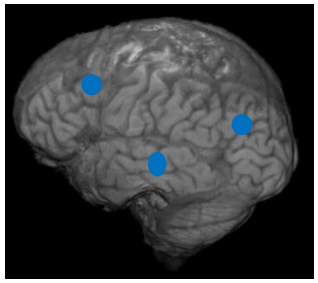
$$\dot{x} = f(x, u, \theta^e)$$

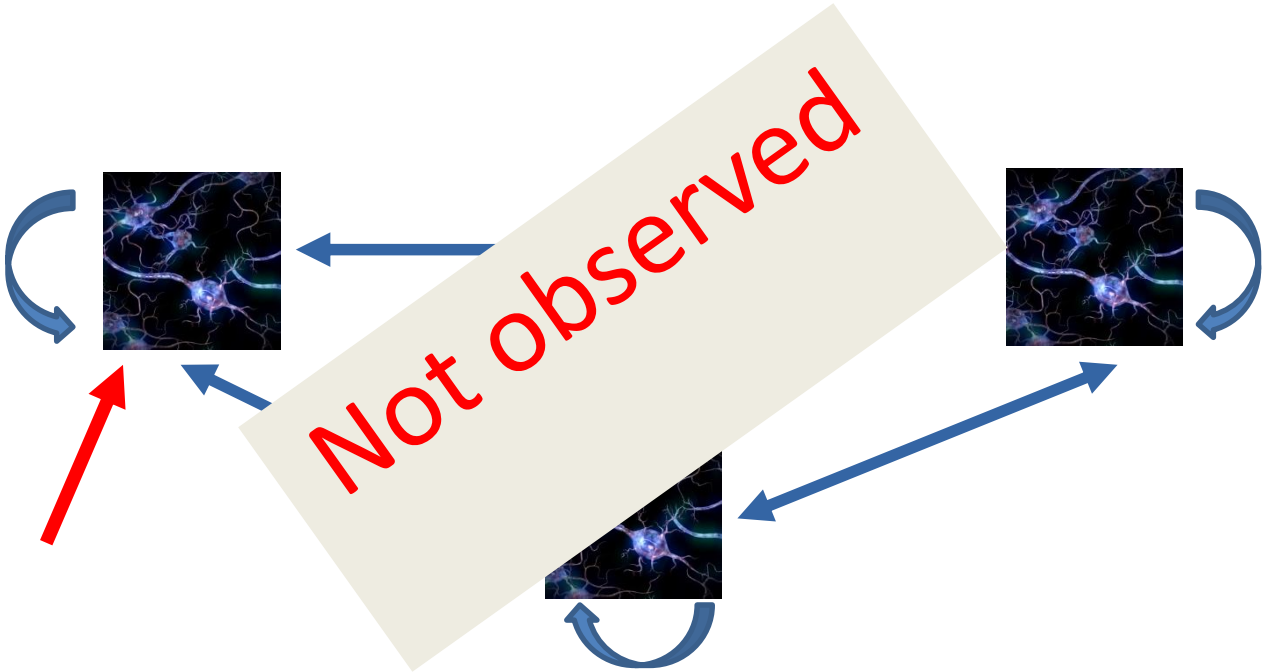




Neural model

Interactions between and within neural populations



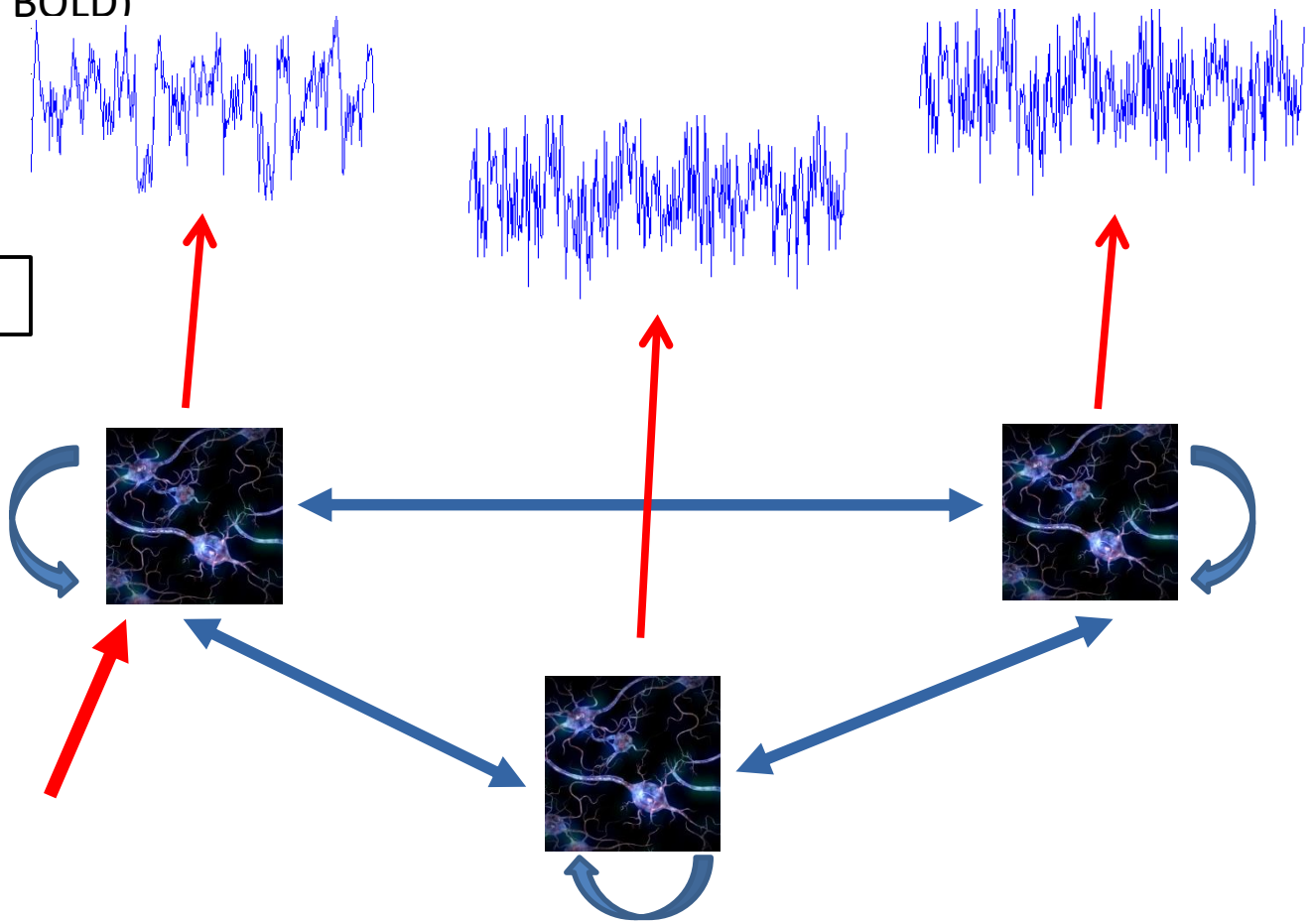


Forward model

OBSERVED signals (e.g., BOLD)

$$y = g(x, \theta) + \varepsilon$$

UNOBSERVED neural states & interactions

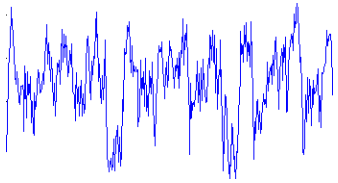


UNIVERSITEIT GENT

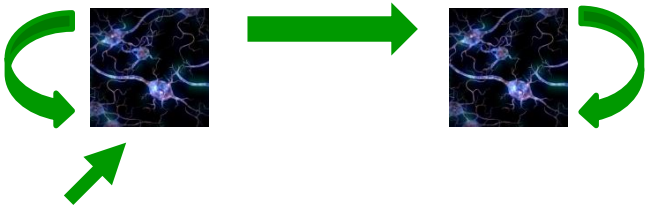
Generative model

$$P(\theta \mid y, m) = \frac{P(y \mid \theta, m) * P(\theta \mid m)}{P(y \mid m)}$$

y = data



m = model

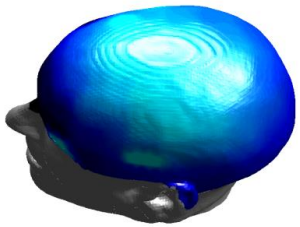


θ = parameter

θ^c = neural parameters

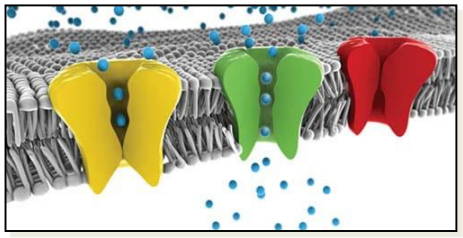
θ^h = (hemodynamic) parameters

Generative model



$$p(y \mid \theta, m) \cdot p(\theta \mid m)$$

$$p(\theta \mid y, m)$$



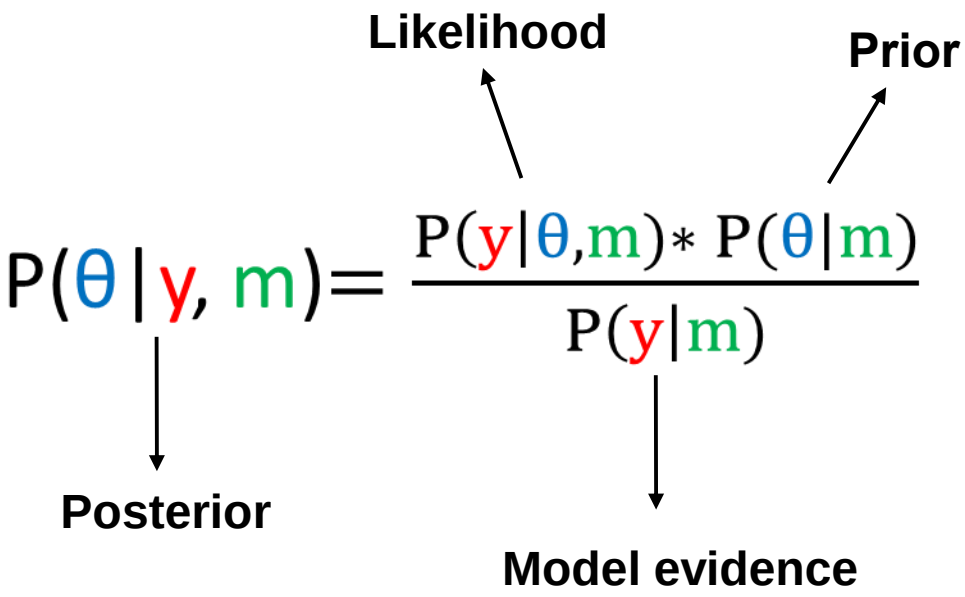
1. enforces mechanistic thinking: how could the data have been caused?
2. generate synthetic data (observations) by sampling from the prior – can model explain certain phenomena at all?
3. inference about model structure: formal approach to disambiguate mechanisms → p(m|y)
4. inference about parameters → p(θ|y)

Adapted from K.E.Stephan

UNIVERSITEIT GENT

Bayesian model inversion

What **parameter estimates** (θ) have highest probability given the **data** (y) and the **model** (m)?



UNIVERSITEIT GENT

Bayesian model inversion

- **Prior:** Specifies what connections are included in the model

- **Likelihood:** Incorporates the generative model and prediction errors

- **Model evidence:** Quantifies the 'goodness' of a model (i.e., accuracy minus complexity). Used to draw inference on model structure.

- **Posterior:** Probability density function of the parameters given the data and model. Used to draw inference on model parameters.

$$P(\theta \mid y, m) = \frac{P(y \mid \theta, m) * P(\theta \mid m)}{P(y \mid m)}$$

Bayesian system identification workflow

$u(t)$

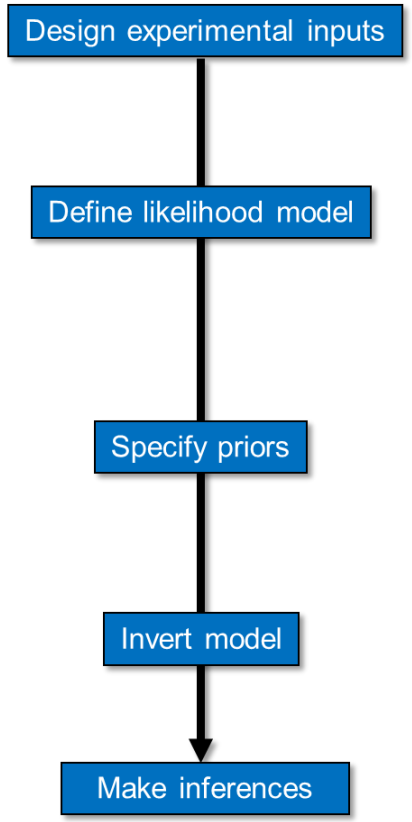
Neural dynamics: $dx/dt = f(x,u,\theta)$

Observer function: $y = g(x,\theta) + \varepsilon$

$$p(y \mid \theta, m) = N(g(\theta), \Sigma(\theta))$$
$$p(\theta, m) = N(\mu_\theta, \Sigma_\theta)$$

Inference on model structure: $p(y \mid m) = \int p(y \mid \theta, m)\, p(\theta) d\theta$

Inference on parameters: $p(\theta \mid y, m) = \dfrac{p(y \mid \theta, m)\, p(\theta, m)}{p(y \mid m)}$

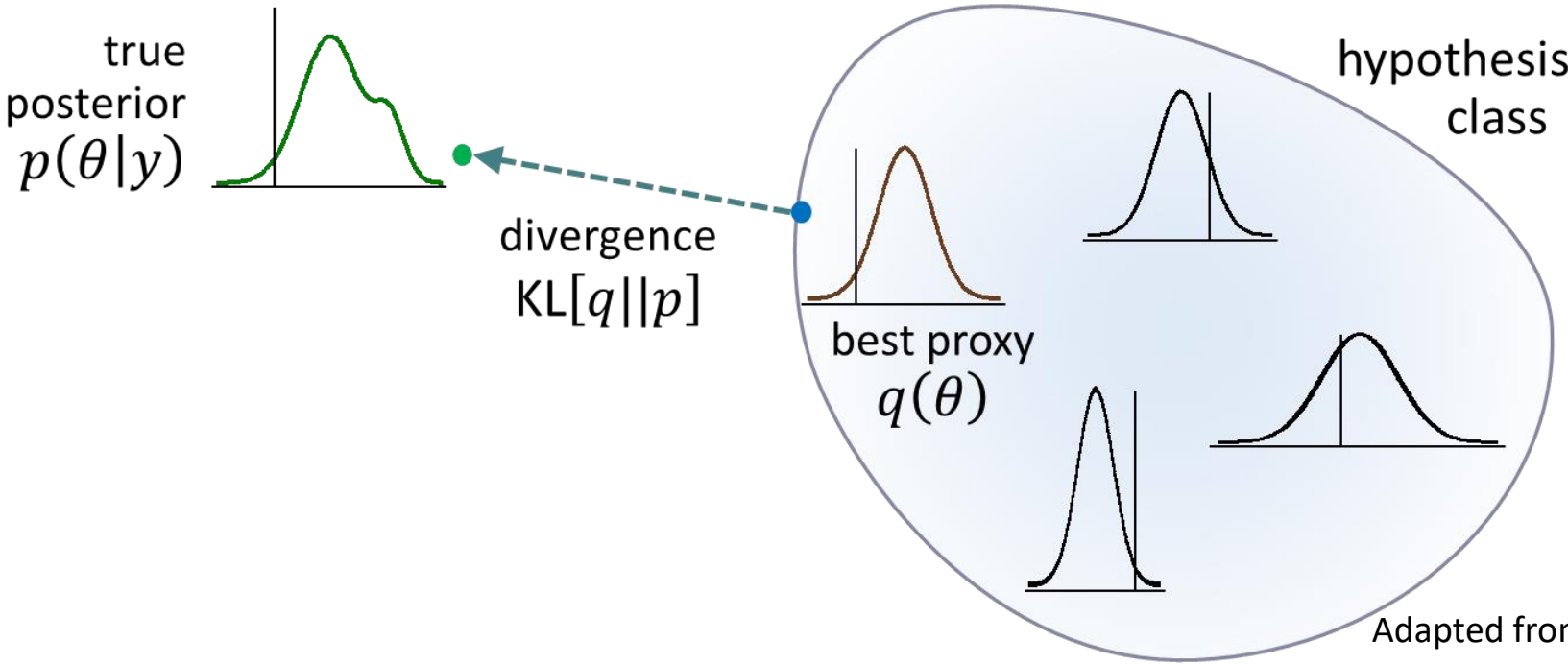


Adapted from K.E.Stephan

Variational Bayes

Idea: find an approximate density $q(\theta)$ that is maximally similar to the true posterior $p(\theta|y)$.

This is often done by assuming a particular form for q (fixed form VB) and then optimizing its sufficient statistics.



Adapted from K.E.Stephan

Variational Bayes

$$\ln p(y) = \underbrace{\mathrm{KL}[q||p]}_{\text{divergence}} + \underbrace{F(q,y)}_{\text{neg. free energy}}$$

divergence ≥ 0 (unknown)

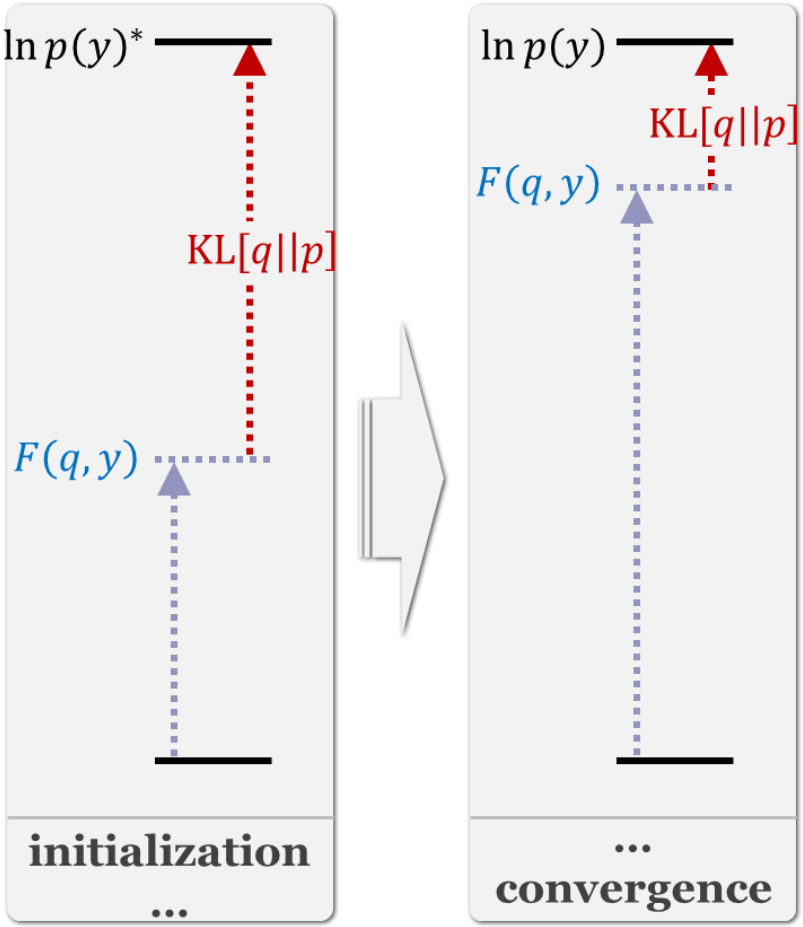
neg. free energy (easy to evaluate for a given q)

$F(q,y)$ is a functional wrt. the approximate posterior $q(\theta)$.

Maximizing $F(q,y)$ is equivalent to:

- minimizing $\mathrm{KL}[q||p]$
- tightening $F(q,y)$ as a lower bound to the log model evidence

When $F(q,y)$ is maximized, $q(\theta)$ is our best estimate of the posterior.

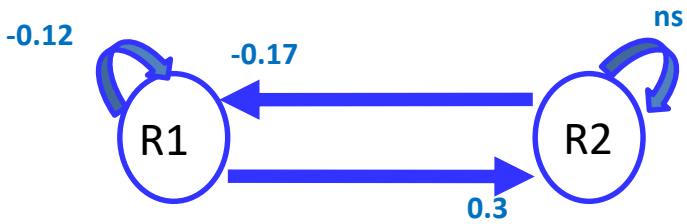


Adapted from K.E.Stephan

Inference

- On the level of **model structure**: Which model (or family of models) has highest evidence?

- On the level of **model parameters**: What parameters are statistically significant, and what is their size/sign?



UNIVERSITEIT GENT

Inference on model structure

- Inference on **model structure** is a necessary step in DCM studies

→ Unless strong prior knowledge about model structure

- **Bayesian model selection (BMS)** compares the (log) model evidence of different models (i.e., probability of the data given model)

→ log model evidence is approximated by free energy

$$\ln p(y|m) = F(y,q) + D_{KL}[q(\Theta)||p(\Theta|y,m)]$$

UNIVERSITEIT GENT

Inference on model parameters

- Inference on **model parameters** is often a second step in DCM studies

- If a clear **'winning' model**:

→ Inference on parameters of this optimal model

Inference on model structure

- If no clear ‘winning’ model (or if optimal model structure differs between groups) then **Bayesian model averaging (BMA)** is an option

→ Final parameters are weighted average of individual model parameters and posterior probabilities

Group-level inference

- Different DCM's are fitted to the data for every subject.
- Group inference on the models: themselves or groups of models (in DCM terminology families of models e.g. all models with input to DLPFC vs. input to FFA vs. both three families):
  ➔**Bayesian model selection**
- Winning model/family is the one with highest exceedance probability

UNIVERSITEIT GENT

Group-level inference

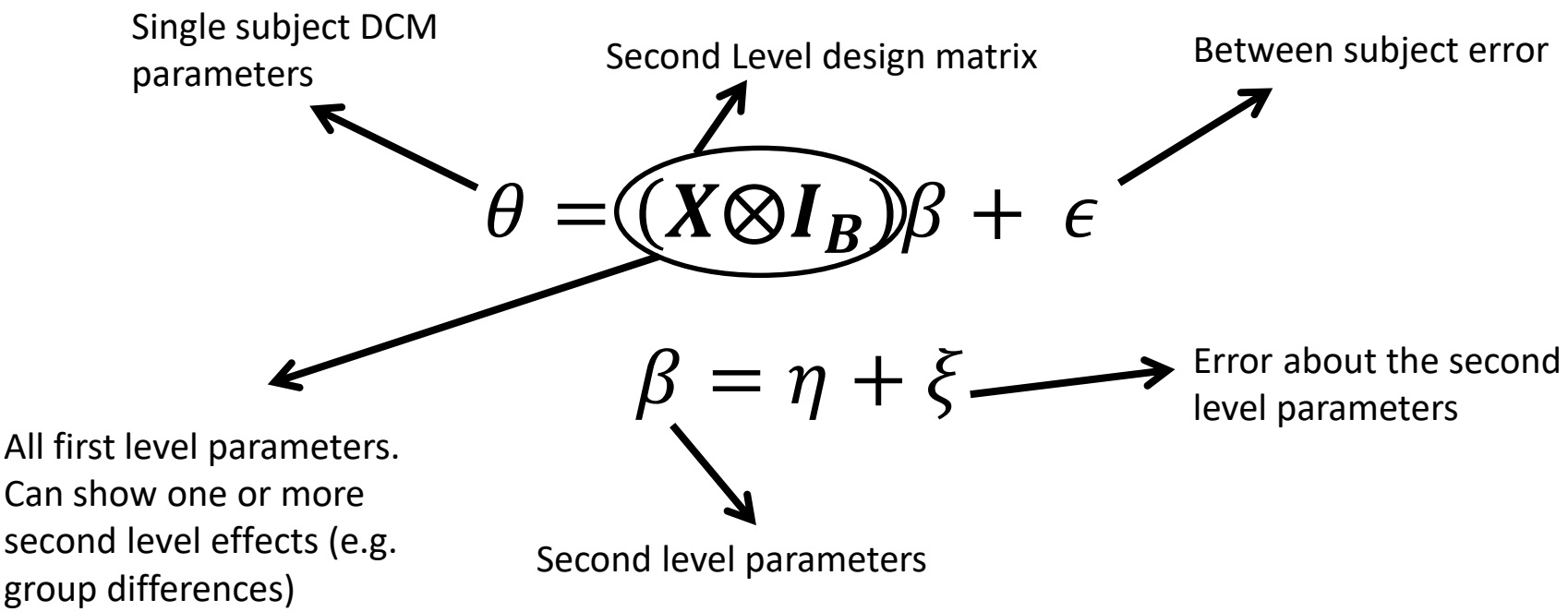
- Group inference on model parameter: Either on the winning model or Bayesian model averaging (BMA) across models (within a winning family or all models when BMS reveal no clear winner)

- (BMA) Parameter(s) of interest are harvested for every subject and subjected to frequentist inference (e.g. t-test)

- Alternatively, one could use the newly developed **PEB**-framework on a full-model. Using Bayesian model reduction all nested models can be efficiently scored and compared at the group-level (Friston et al, 2016).

Parametric empirical Bayes (PEB)

PEB: Hierarchical general linear model



UNIVERSITEIT GENT

Recap: generative models & model selection

any DCM = a particular generative model of how the data (may) have been caused

generative modelling: comparing competing hypotheses about the mechanisms underlying observed data

→ a priori definition of hypothesis set (model space) is crucial

→ determine the most plausible hypothesis (model), given the data

model selection ≠ model validation!

→ model validation requires external criteria (external to the measured data)

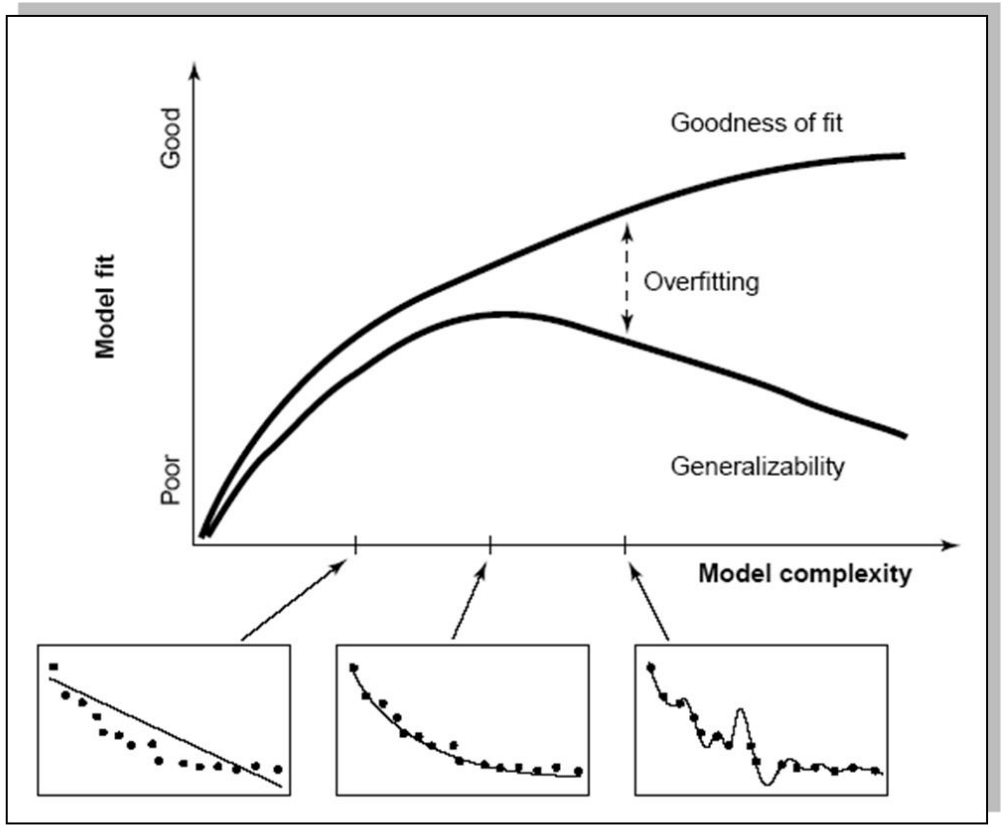
Slide from K.E.Stephan

Model comparison and selection

Given competing hypotheses on structure & functional mechanisms of a system, which model is the best?

Which model represents the best balance between model fit and model complexity?

For which model m does $p(y|m)$ become maximal?



Pitt & Miyung (2002) *TICS*

Adapted from K.E.Stephan

Bayesian model selection

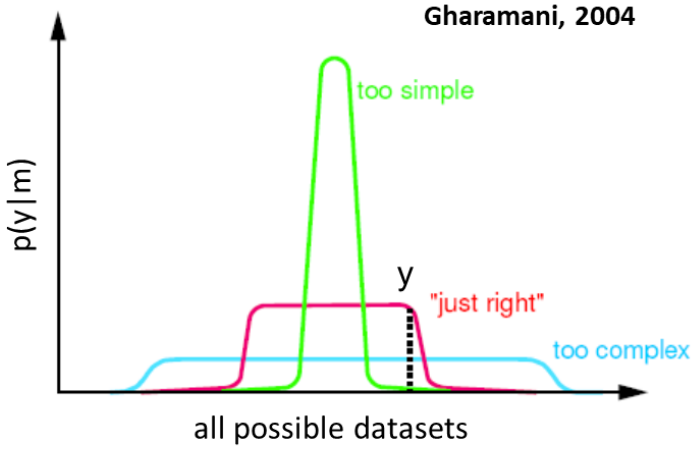
Model evidence:

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

accounts for both accuracy and complexity of the model

allows for inference about model structure

Gharamani, 2004



Various approximations, e.g.:

- negative free energy, AIC, BIC

McKay 1992, *Neural Comput.*
Penny et al. 2004a, *NeuroImage*

Adapted from K.E.Stephan

Approximation to model evidence in DCM

Logarithm is a monotonic function → Maximizing log model evidence = Maximizing model evidence

Log model evidence = balance between fit and complexity

$$\log p(y \mid m) = accuracy\ (m) - complexity\ (m)$$
$$= \log p(y \mid \theta, m) - complexity\ (m)$$

Akaike Information Criterion: $AIC = \log p(y \mid \theta, m) - p$

Bayesian Information Criterion: $BIC = \log p(y \mid \theta, m) - \frac{p}{2} \log N$

No. of parameters

No. of data points

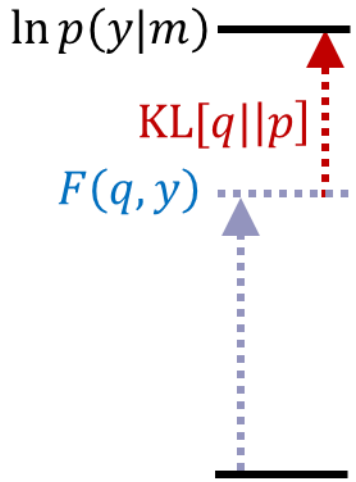
Penny et al. 2004a, *NeuroImage*

Adapted from K.E.Stephan

The (negative) free energy approximation F

Negative free energy is a lower bound on log model evidence:

$$\log p(y \mid m) = F + KL\left[q(\theta), p(\theta \mid y, m)\right]$$



Like AIC/BIC, F is an accuracy/complexity tradeoff:

$$F = \underbrace{\left\langle \log p(y \mid \theta, m) \right\rangle}_{accuracy} - \underbrace{KL\left[q(\theta), p(\theta \mid m)\right]}_{complexity}$$

Adapted from K.E.Stephan

The complexity term in F

- In contrast to AIC & BIC, the complexity term of the negative free energy *F* accounts for parameter interdependencies.

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

- The complexity term of *F* is higher
  - the more independent the prior parameters (↑ effective DFs)
  - the more dependent the posterior parameters
  - the more the posterior mean deviates from the prior mean

Adapted from K.E.Stephan

UNIVERSITEIT GENT

With so many parameters, does DCM overfit?

this is "overfitting": an increasingly complex model will, at some point, start to fit noise that is specific to one data set and thus become less generalizable across multiple realizations of the same underlying generative process.

Therefore, the question "What is the optimal model?" can be reformulated more precisely as "What is the model that represents the best balance between fit and complexity?". In a Bayesian context, the latter question can be addressed by comparing the evidence, $p(y|m)$, of different models. According to Bayes theorem

$$p(\theta|y,m) = \frac{p(y|\theta,m)p(\theta|m)}{p(y|m)} \tag{35.7}$$

the model evidence can be considered as a normalization constant for the product of the likelihood of the data and the prior probability of the parameters, therefore

$$p(y|m) = \int p(\theta|y,m)p(\theta|m)d\theta \tag{35.8}$$

Here, the number of free parameters (as well as the functional form) are considered by the integration. Unfortunately, this integral cannot usually be solved analytically, therefore an approximation to the model evidence is needed. One such approximation used by DCM, and many other models in SPM, is to make use of the Laplace approximation [1].

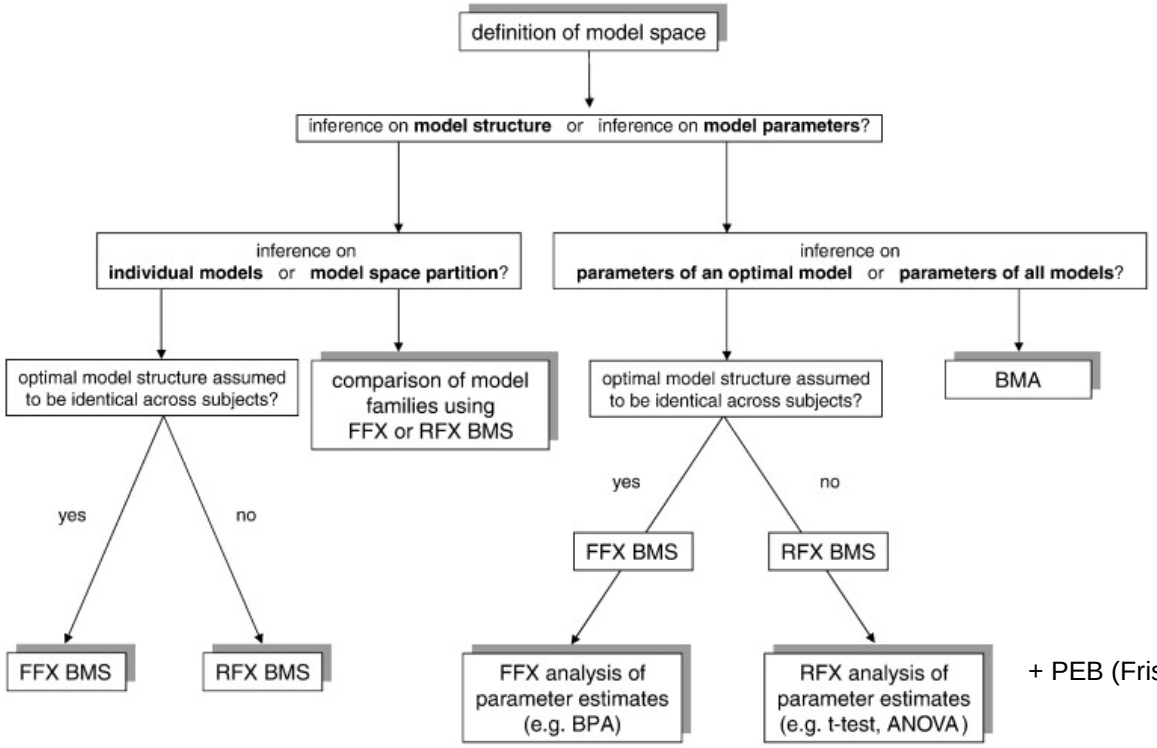
As shown in [95], this yields the following expression for the natural logarithm (ln) of the model evidence ($\eta_{\theta|y}$ denotes the posterior mean, $C_{\theta|y}$ is the posterior covariance of the parameters, C_e is the error covariance, θ_p is the prior mean of the parameters, and C_p is the prior covariance):

$$\begin{aligned} lnp(y|m) &= accuracy(m) - complexity(m) \\ &= \left[-\frac{1}{2}ln|C_e| - \frac{1}{2}(y - h(u,\eta_{\theta|y}))^T C_e^{-1}(y - h(u,\eta_{\theta|y}))\right] \\ &- \left[\frac{1}{2}ln|C_p| - \frac{1}{2}ln|C_{\theta|y}| + \frac{1}{2}(\eta_{\theta|y} - \theta_p)^T C_p^{-1}(\eta_{\theta|y} - \theta_p)\right] \end{aligned} \tag{35.9}$$

This expression properly reflects the requirement, as discussed above, that the optimal model should represent the best compromise between model fit (accuracy) and model complexity. The complexity term depends on the prior density, for example, the prior covariance of the intrinsic connections.

Generally speaking, a model with a poor fit can have more evidence than a model with a good fit. For example, if you give a model pure measurement noise, a good model should properly identify that there is noise and no signal. This model will be better than a model that tries to (over) fit noisy fluctuations. It is important to note that in most cases of model inversion it is not just model parameters that are optimised but also the estimate of the amplitude of observation noise.

Inference: summary



+ PEB (Friston et al., 2016):
~Bayesian GLM

Stephan et al., 2010

UNIVERSITEIT GENT

Different variants of DCM

Different variants of DCM

- DCM has been developed for specific contexts (e.g., fMRI and EEG data, time and frequency domain, task-induced and resting state paradigms,...)

- The following types of DCM are often used:

  - DCM for task-fMRI (Friston et al., 2003)*

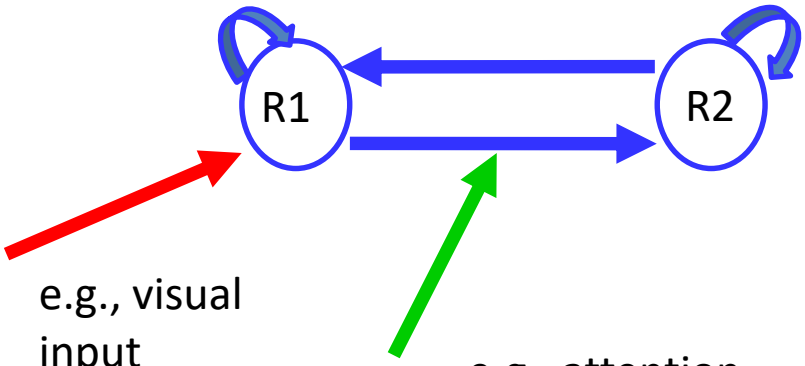
  - DCM for resting state fMRI (Friston et al., 2014)*

  - DCM for ERP/ERF (David et al., 2006)

*stochastic DCM (Friston et al., 2010, Li et al., 2011) is also applicable to both task- and resting state fMRI

DCM for task-fMRI

- Neural model



- Structure (effective connections)

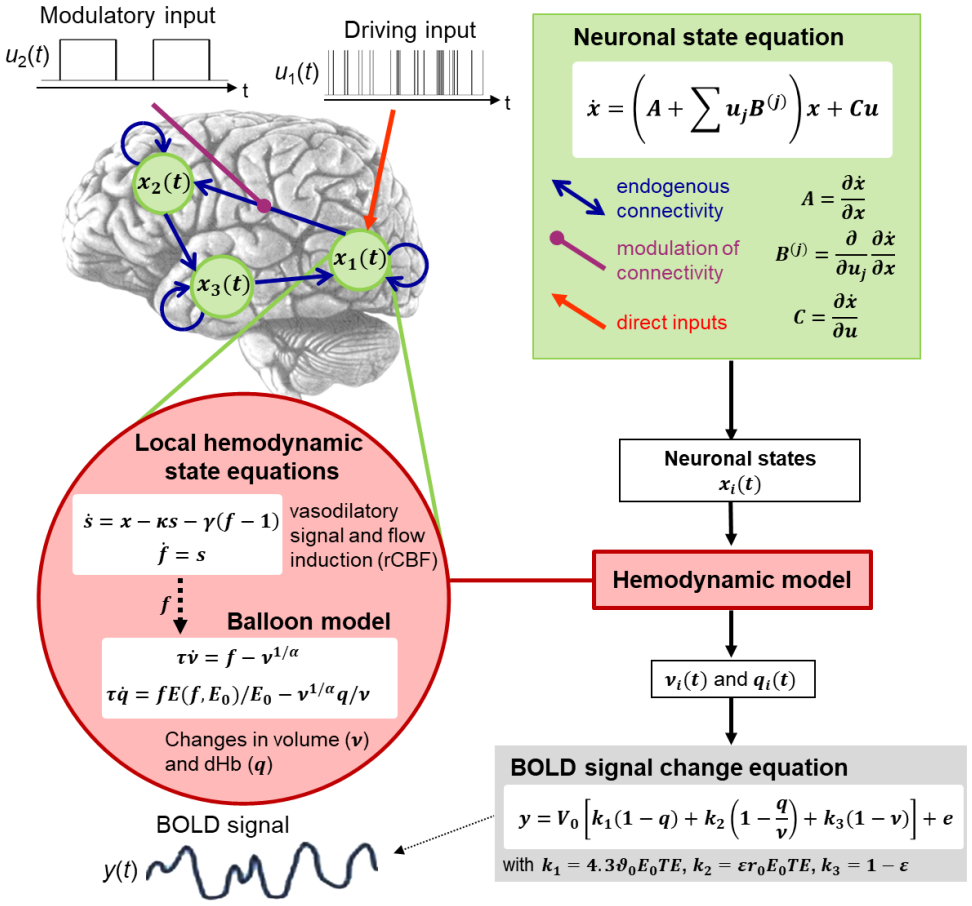
- Modulation of connections

- Driving inputs

$$\dot{x} = (A + \Sigma_j u_j B^j) x + Cu$$

UNIVERSITEIT GENT

DCM for task-fMRI



Stephan et al. 2015, *Neuron*

UNIVERSITEIT GENT

Connectivity is measured in Hz. Why?

Integration of a first-order linear differential equation gives an exponential function:

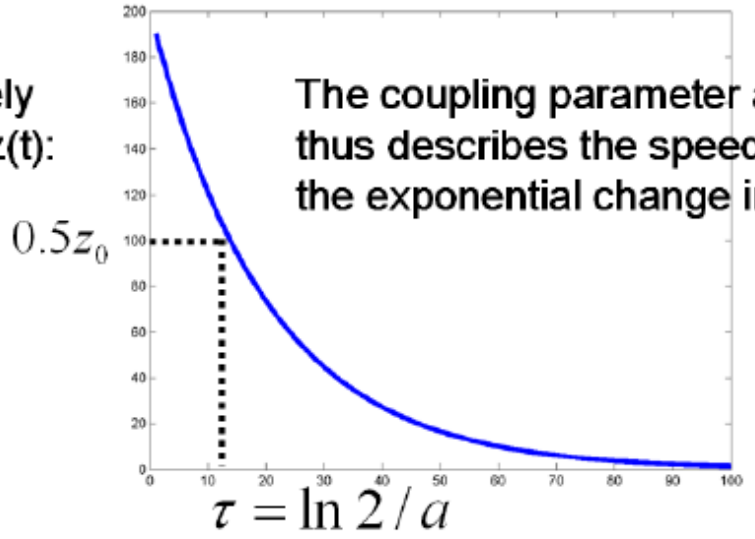
$$\frac{dz}{dt} = az \quad \Rightarrow \quad z(t) = z_0 \exp(at)$$

Coupling parameter a is inversely proportional to the half life τ of z(t):

$$z(\tau) = 0.5 z_0$$

$$= z_0 \exp(a\tau)$$

$$\Rightarrow \boxed{a = \ln 2 / \tau}$$



The coupling parameter a thus describes the speed of the exponential change in z(t)

UNIVERSITEIT GENT

DCM for task-fMRI

- Bayesian Model Inversion

  → Variational Expectation Maximization

  Assumes (approximate) posterior is **Gaussian**

  Maximizes free energy by updating (hyper)parameters

DCM for resting state fMRI

- Neural model

- A → Structure (effective connections)
-
- v → neuronal fluctuations (drive the system)

$$\dot{x} = Ax + v$$

UNIVERSITEIT GENT

Spectral DCM

- Parametrization of spectral densities

$g_v(\omega,\theta) = \alpha_v \omega^{-\beta v}$ → neuronal fluctuations

$g_e(\omega,\theta) = \alpha_e \omega^{-\beta e}$ → observation noise

UNIVERSITEIT GENT

Spectral DCM

- Forward model:

  - Modeled with Volterra kernels [κ(t)]

  - Is a function of effective connectivity

UNIVERSITEIT GENT

Spectral DCM

- Generative model (in frequency domain)

$$g_y(\omega,\theta) = |K(\omega)|^2 \, g_v(\omega,\theta) + g_e(\omega,\theta)$$

Predicted cross spectra

Fourier transform Volterra kernels

UNIVERSITEIT GENT

Spectral DCM

- Bayesian Model Inversion

→ Variational Expectation Maximization

UNIVERSITEIT GENT

Spectral DCM

- deterministic model that generates predicted cross-spectra in a distributed neuronal network or graph

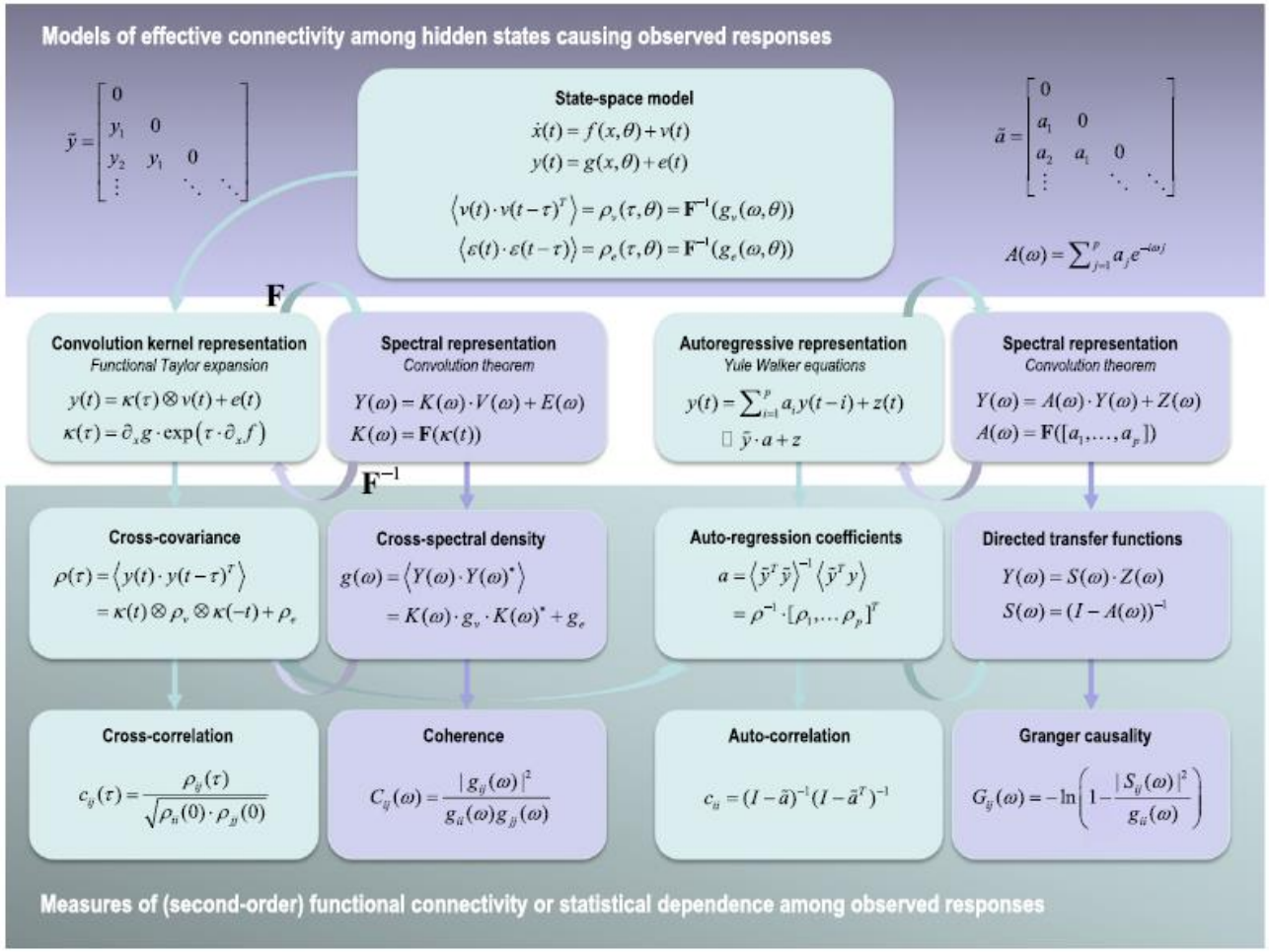
- finds the effective connectivity among hidden neuronal states that best explains the observed functional connectivity among hemodynamic responses

- advantage:
  - replaces an optimisation problem wrt. stochastic differential equations with a deterministic approach from linear systems theory
    → computationally very efficient

- disadvantages:
  - assumes stationarity

UNIVERSITEIT GENT

Families of (spectral) dependencies



Quality check

- After estimating DCMs, diagnostics should be consulted

- Code for fMRI: spm_dcm_fmri_check(DCM):

    - Proportion explained variance should be sufficient (at least 10% for task-fMRI)

    - Largest connection's strength should be above 0.125Hz

    - Estimable parameters should be greater than 1

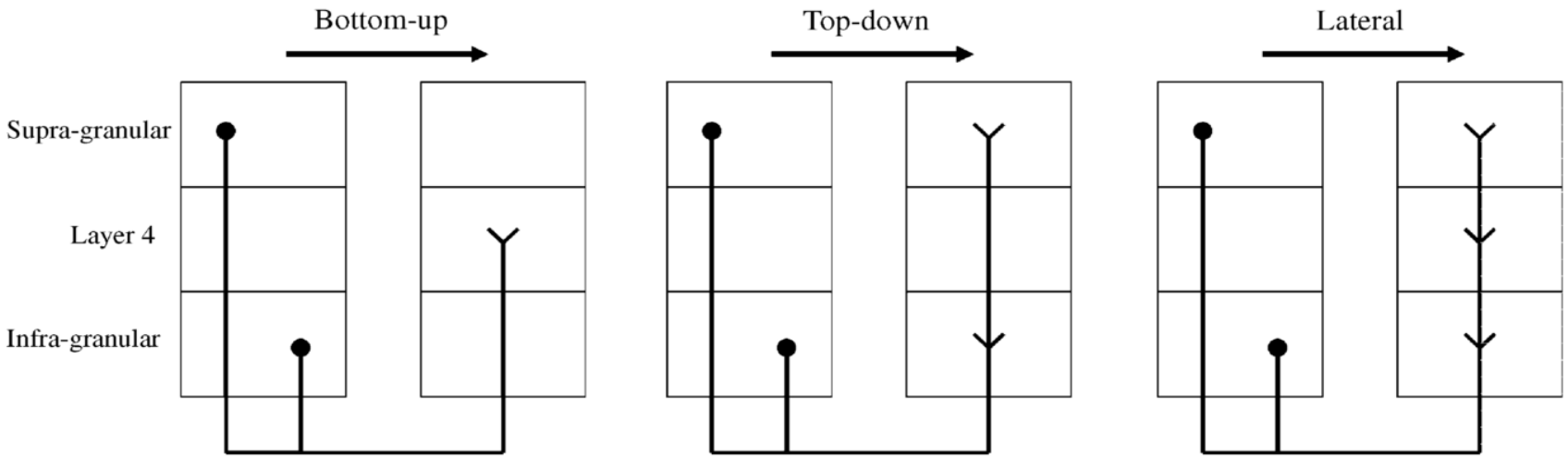
→ If one of the above is not satisfied, respective subtitle will be shown in red

UNIVERSITEIT GENT

DCM for ERP/ERF

- Neural model: much more complex compared to DCM for fMRI
- Each region (‘node’) is modelled with neural mass (or field) models
- 3 layer per node: supra, infra-granular layer and layer IV
- Nodes are connected by either forward, backward or lateral connection (the extrinsic connections)

DCM for ERP/ERF



- Bottom-up: connection from low to high hierarchical areas (Felleman 1991)
- top-up: connection from high to low hierarchical areas (Felleman 1991)
- Lateral: same level in hierarchical organization (e.g. interhemispheric connection)
- Prior on connection: forward>backward>lateral

UNIVERSITEIT GENT

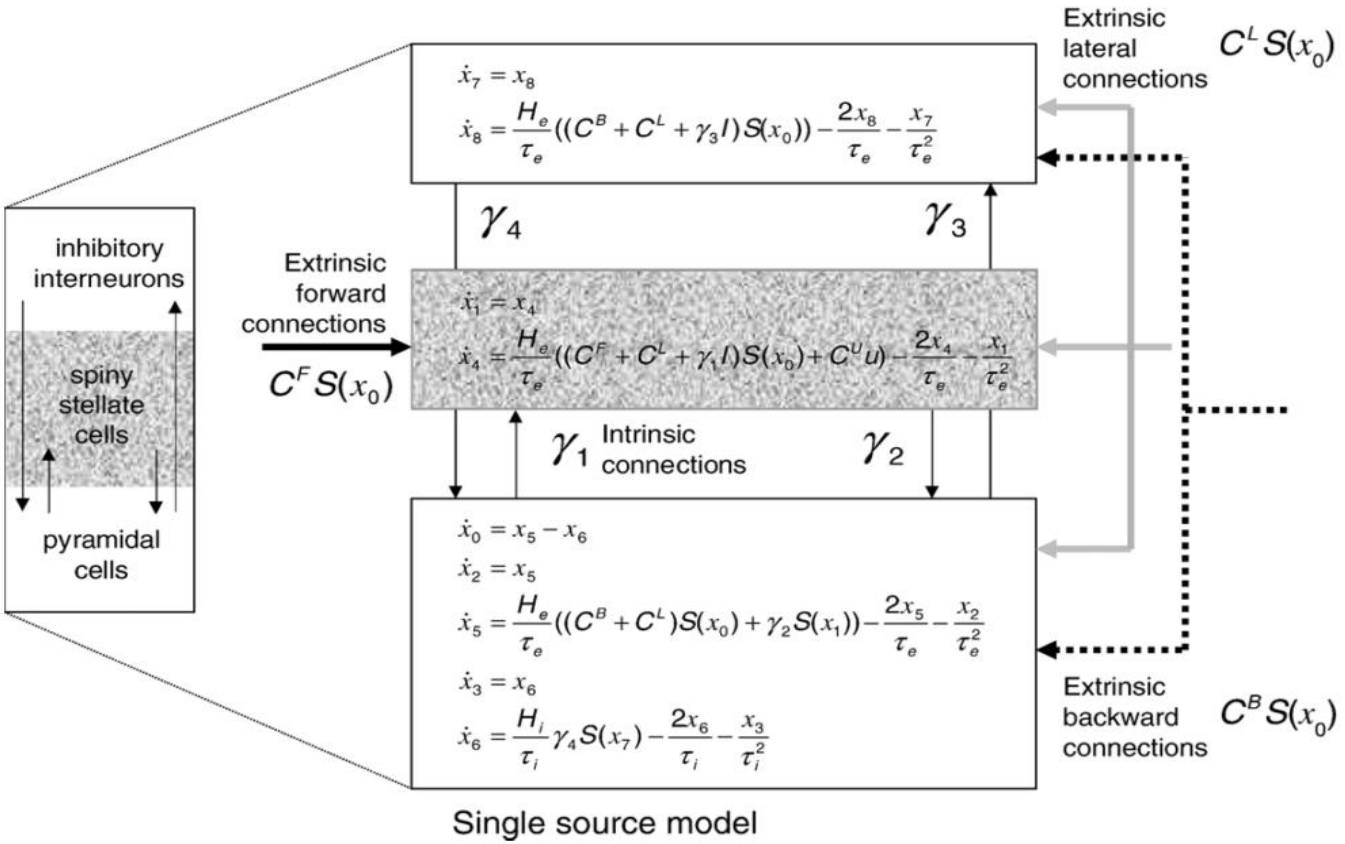
DCM for ERP/ERF

- Layers within regions interact via intrinsic connection

- What is measured with the EEG/MEG sensor are the potentials generated by pyramidal cells

UNIVERSITEIT GENT

DCM for ERP/ERF



Single source model

UNIVERSITEIT GENT

DCM for ERP/ERF

- The forward model in EEG/MEG is much simpler compared to fMRI:

- $Y(t) = LX_0(t) + \varepsilon(t)$

- Y(t) are the channel time series, L is the leadfield (conduction of electromagnetic fields), X_0 are the pyramidal potentials of all sources and ε is measurement noise.
- In words: each channel is a weighted sum of source activities where the weights depend on position and orientation of the sources and channels

UNIVERSITEIT GENT

DCM for ERP/ERF

- ROIs need to be specified based on prior knowledge/assumptions regarding the location of the sources or based on data itself via source reconstruction
- Models with different ROIs can be compared (not the case with fMRI)
- Search in literature for determining type of connection between ROIs (e.g. forward connection from low to higher cortical areas)

Recommended articles

DCM for task-fMRI:

→ Friston et al., 2003: Dynamic causal modeling (NI)

DCM for resting state fMRI:

→ Friston et al., 2014: A DCM for resting state fMRI (NI)

→ Razi et al., 2015: Construct validation of a DCM for resting state fMRI (NI)

DCM for ERP/ERF:

→ David et al., 2006: Dynamical causal modelling of evoked responses in EEG and MEG (NI)

PEB:

-> Friston et al., 2016: Bayesian model reduction and empirical Bayes for group DCM studies (NI)

-> Zeidman et al. 2019: A tutorial on group effective connectivity analysis, part 1: first level analysis with DCM for fMRI

-> Zeidman et al. 2019: A tutorial on group effective connectivity analysis, part 2: second level analysis with PEB

Practical recommendations:

→ Stephan et al., 2010: Ten simple rules for dynamic causal modeling (NI)

→ Penny et al., 2004: Comparing Dynamic causal models (NI)