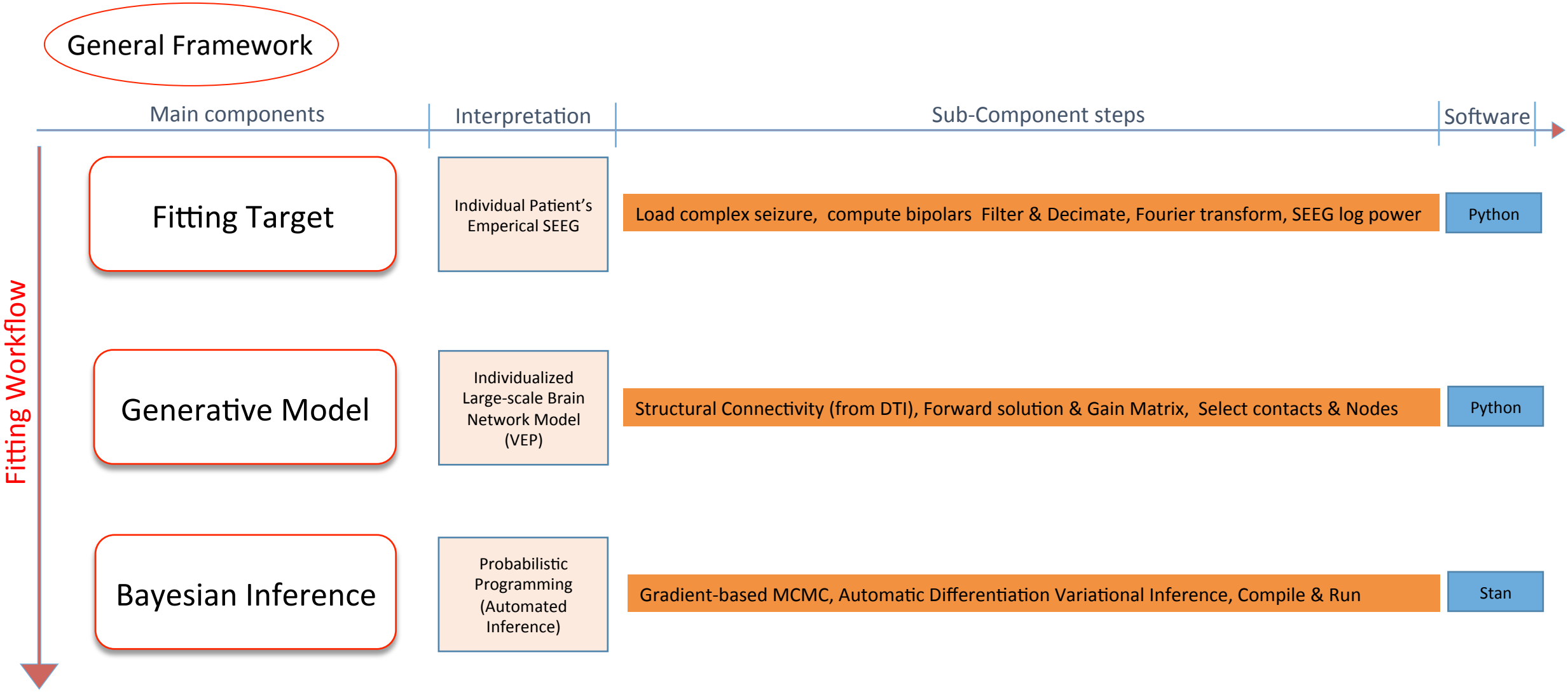
A pipeline for Bayesian inference on seizure propagation seen in SEEG data

First, the general framework for the current workflow aimed to fit brain model parameters against the brain functional data such as stereotactic EEG (EEG) is illustrated. The main components namely fitting target, generative model, and Bayesian inference framework and the sub-components are explained in the following section.



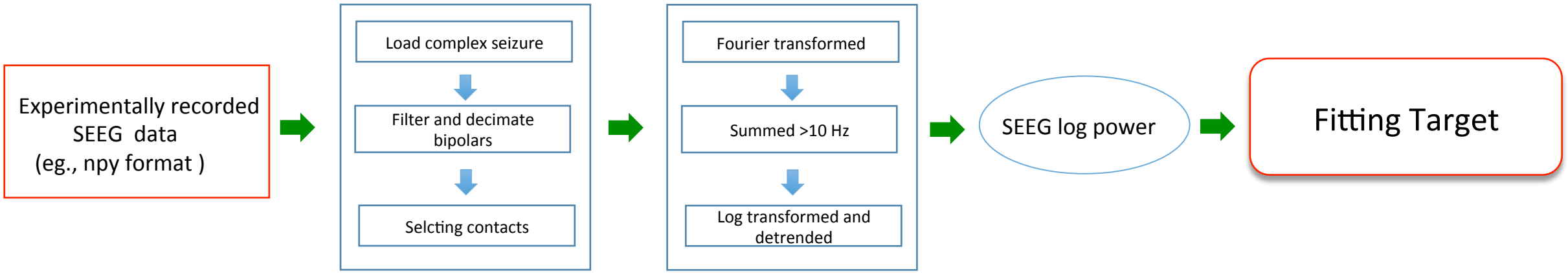
In this systematic workflow, we aim to fit large-scale brain model parameters to the functional data such as stereotactic EEG (SEEG). To achieve this goal, we need the following components:

a) Fitting target: taking this step, from empirical patient data such as SEEG signals comprising patient seizure, we obtain sensor log power as the target for fitting (known as observation in Bayesian framework).

Having recorded patient data such as SEEG signals, it is necessary to preprocess and select a frequency band of interest and reduce the size of the data. To this end, the raw time series of electrodes are required to be processed.

- First, the complex seizure data are loaded in python (eg., npy format).
- Then, the bipolars were filtered and decimated from complex seizure data. This reduces the computationally costs of the data fitting task.
- We select some of the contacts which more contributed in the bipolar data. According to mean energy of differences in bipolars, we select some of the contacts as the fitting target.
- Finally, the log power of high frequency activity, predicting sensor log power, is computed as the target for fitting task. More precisely, the SEEG data are windowed and Fourier transformed to obtain estimates of their spectral density over time. Then SEEG power above 10 Hz is summed to capture the temporal variation of the fast activity. These time series are corrected to a preictal baseline, log-transformed and linearly detrended over the time window encompassing the seizure.

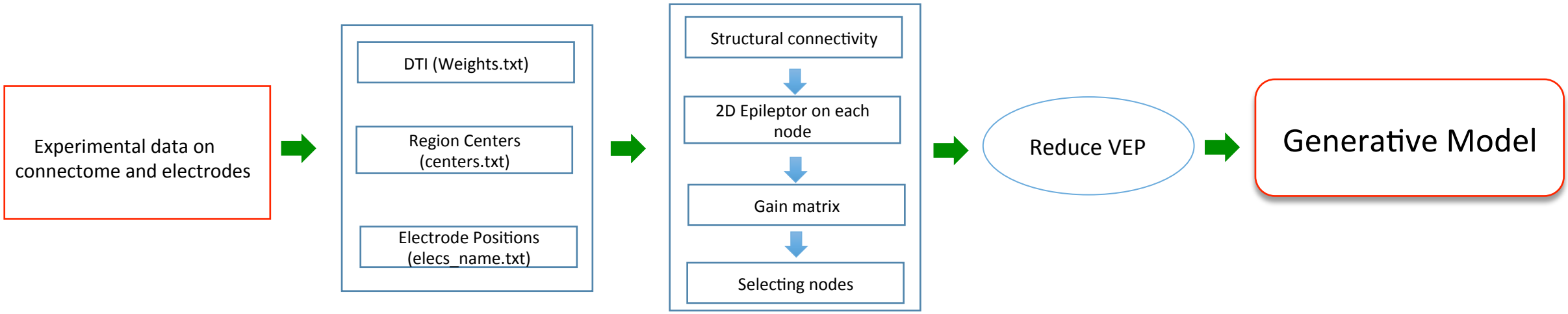
All these steps leading to fitting target have been performed in Python.



b) Generative Model: taking this step, we construct the brain network model including the parameters of interest for fitting (known as the mathematical generative model in Bayesian framework).

Having the mathematical model, it is necessary to preprocess and select some of the network nodes of interest and reduce the complexity of the generative model while the model is still able to reproduce the characteristic features of seizures as observed experimentally. The Virtual Epileptic Patient (VEP) based on personalized brain network models derived from non-invasive structural data of individual patients gives us the generative model for fitting task (Jirsa et al 2016). Using patient-specific connectomes in large-scale brain networks as generative models of neuroimaging signals, we fit and validate the brain model against the patient's empirical SEEG data. The approach to build VEP brain model comprises:

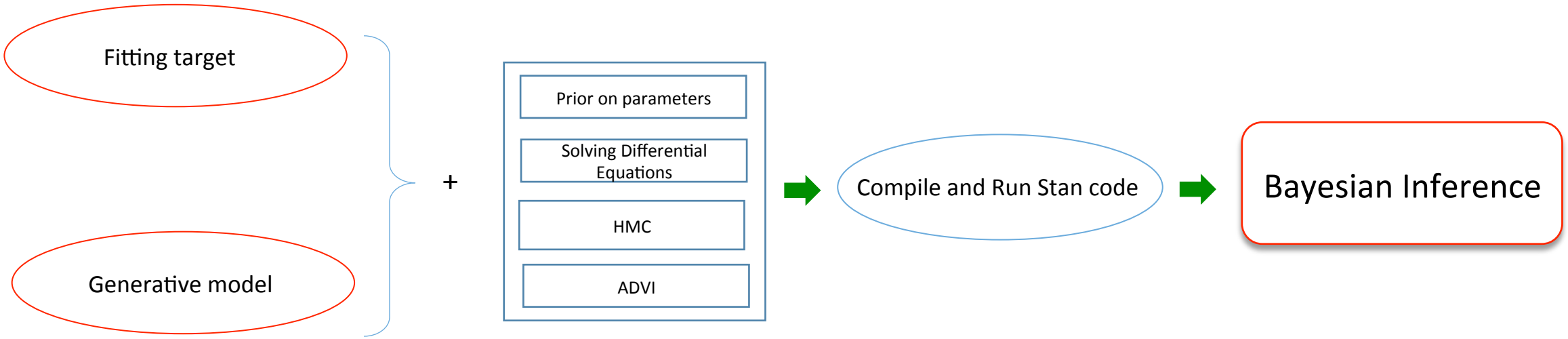
- Structural connectivity: reconstruction of the patient's individual brain network topography DTI to obtain connectivity between nodes.
- Functional network modeling: 2D reduction of coupled Epileptors are defined on each network node (Prix et al 2014).
- Gain matrix (Forward solution): Like other modalities, the SEEG measurements can be modeled using a forward solution that describes the contribution of each source dipole to each contact's measurement. Computing gain matrix follows the standard definition of the potential due to a point dipole in homogeneous space. We need region centers and electrode positions to compute gain matrix.
- Selecting nodes: As described above, we reduced the observations by selecting some of the contacts according to the mean energy of bipolars. Moreover, we assume that few nodes in the whole brain network play a critical role in the seizure and thus contribute to the observed data. This allows us to reduce number of nodes in the network on which the inference is performed. In this step, we select the nodes according to gain matrix information.



c) Bayesian Inference: a parameter estimation framework to fit large-scale brain network models against signals from the imaging modality such as SEEG. Note that taking the aforementioned steps i.e., defining fitting target and having generative model are critical to perform Bayesian inference in order to obtain the posterior distribution of excitability parameter of the described large-scale brain network model from SEEG signals. The excitability parameters give us valuable information about Epileptogenic Zone (EZ) and Propagation Zone (PZ), which are influential for clinician to perform more accurate surgery.

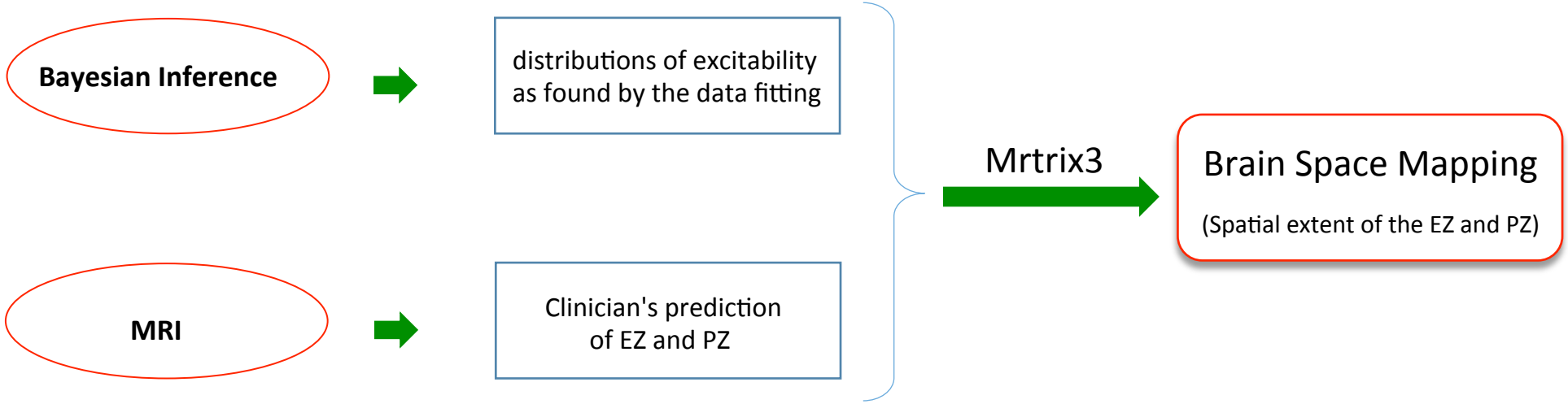
We have a large set of empirical data, and a complex mathematical model comprising neural mass models on each network node connected via patient's connectome. Consequently, we need a complex inference framework to achieve reliable estimates. It is important to note that deriving Bayesian inference algorithms requires tedious model-specific calculations. Satn is an Open-source framework for Bayesian inference without the hand-tuning the algorithm's parameters for different case studies. This software is a flexible probabilistic programming system allowing us to automatically implement both Hamiltonian Monte Carlo (HMC) and Variatinal Inference (VI) in pure Stan code, which then translated to C++ by Stanc compiler. Stan has different interfaces such as Python and shell, running on all major platforms (Linux, Mac,Windows). Stan lisence: New BSD Lisence. Stan Homepage: mc-stan.org

It is important to note that performing each Bayesian inference requires to define the fitting target as well as the generative model. Furthermore, we need to define our prior knowledge on the model parameters. Uninformative priors are placed on the hidden states' initial conditions, while their evolution follows a Euler-Maruyama discretization of the corresponding stochastic differential equations. Finally, we compile and run HMC and ADVI using Stan code to obtain posterior distribution of excitability parameters.



Fitting evaluation: At the end, we evaluate the fitting results using two defined metric namely Predictive Score and Least Square Error (LSE). Predictive Score indicates how many of the nodes are correctly predicted according to the clinical expertise. LSE indicate the error between empirical and fitted signal. Higher value of Predictive Score and lower value of LSE indicate a better fitting.

Furthermore, we virtualize and compare distributions of excitability as found by the data fitting to the clinical observations. Spatial extent of the excitability zone expressed through the parameter distributions and those according to clinician expertise can be plotted over an MRI view. This leads to comparison of the distribution of excitabilities found by fitting the model to the SEEG data in order to evaluate data fitting identifies. To achieve this goal, we use *mrconvert* and *overlay* from Mrtrix3. The software MRtrix3 is freely available under an open-source license which provides a set of tools to perform various types of diffusion MRI analyses.



In sum, we load and preprocess SEEG data, sensors and connectivity (in python), then build a dynamical model (Reduced Epileptor) linking the data, and run HMC and VI algorithms as implemented by the Stan software. Finally, the spatial distributions of excitability as found by the data fitting is compared against clinical prediction using Mrtrix3.