Report for diagnosing problems with HMC

February 13, 2020

Theoretical Neurosciences Group
Institut de Neurosciences des Systèmes
Aix-Marseille Université

Introduction

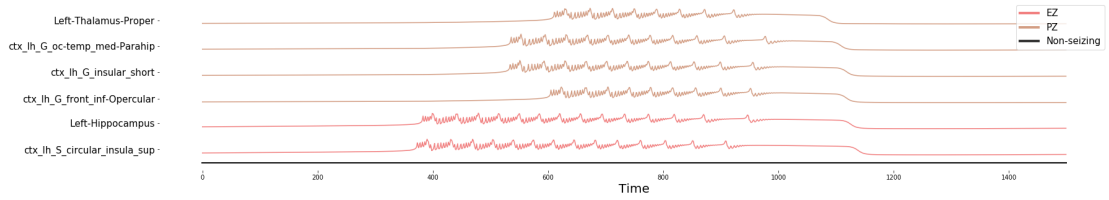
The objective of this work is to identify spatio-temporal seizure propagation patterns i.e to find the regions of brain where seizure starts and propagates towards by inverting a dynamical model of epileptic seizures namely Epileptor [2]. Coupled Epileptor model [3] is shown to reproduce patient specific seizures by adjusting the parameters of the model. So we aim to build virtual epileptic patient models using STAN to infer the parameter subspaces that would best explain log. power profiles of each patients intracranial recordings. More details about the approach are given in [1]

Notations

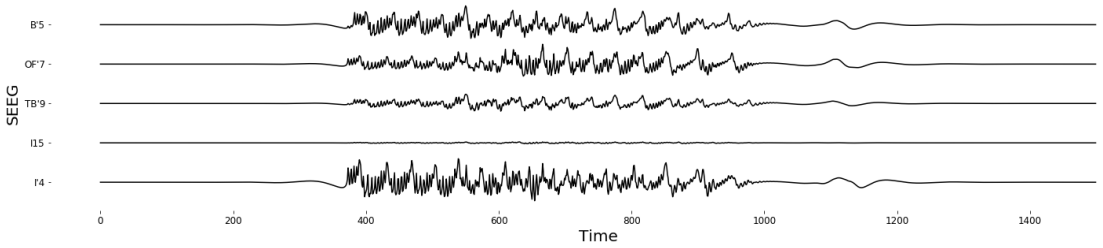
The network of regions which initiate the seizure is referred to as Epileptogenic Zone(EZ) and the network of regions where seizure propagates is referred to as Propagation Zone(PZ)

Dataset

In order test the model, a synthetic dataset is generated using 5D Epileptor



(a) Source activity



(b) SEEG activity

Figure 1: Simulated seizure propagation pattern

Modeled data features

Given appropriate parameter values 2D Epileptor can capture some of the important characteristics of a seizure namely seizure onset and seizure length, which are sufficient for the purposes of identifying spatio-temporal seizure propagation patterns. Log. SEEG power encompasses both these features and hence forms a good candidate data feature that can be modeled using 2D Epileptor. Figure 2 shows the features extracted from synthetic dataset shown in figure 1

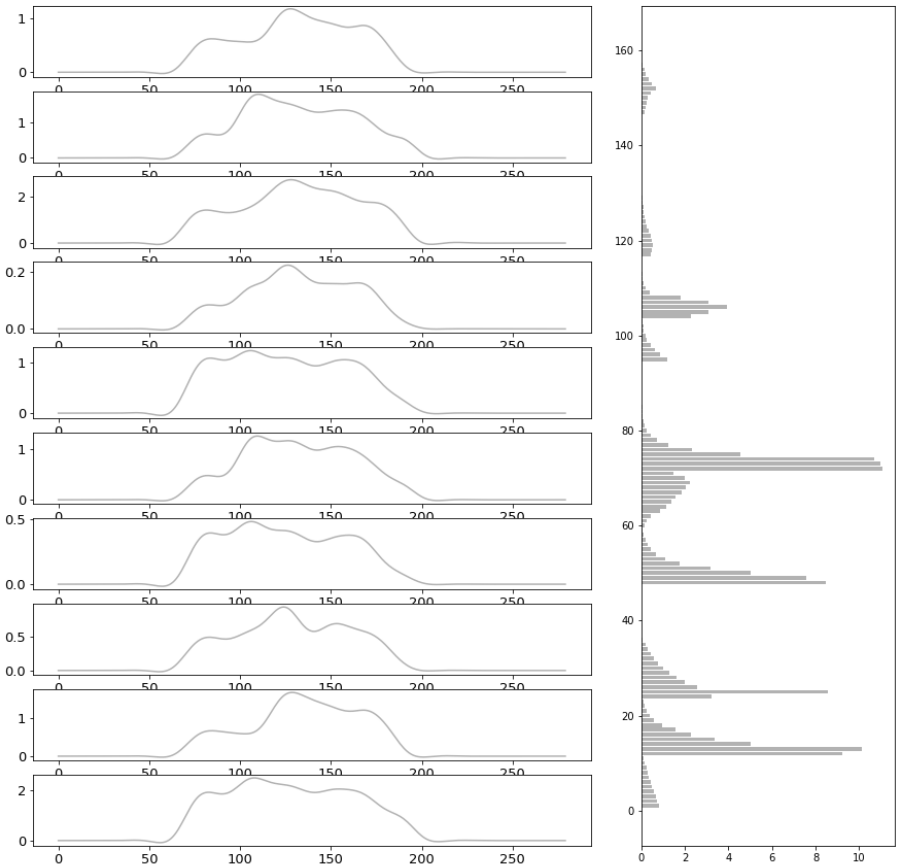


Figure 2: Modeled data features. SEEG log power of 10 sensors(left) and total power per sensor(right)

Model

The dependency structure between parameters of virtual epileptic patient(VEP) model is shown in figure 3

$$P(\theta, \mathbf{X}|\mathbf{D}) \propto P(\mathbf{D}|\theta, \mathbf{X})P(\mathbf{X}, \theta)$$

where,

$$\mathbf{D} = (\mathbf{S}, \vec{\rho}), \mathbf{S} = \text{Log SEEG power}, \vec{\rho} = \text{Total power in each sensor}$$

$$\mathbf{S} = \begin{pmatrix} s^1_{t_1} & s^2_{t_1} & \cdots & s^M_{t_1} \\ s^1_{t_2} & s^2_{t_2} & \cdots & s^M_{t_2} \\ \vdots & \vdots & \ddots & \vdots \\ s^1_{t_T} & s^2_{t_T} & \cdots & s^M_{t_T} \end{pmatrix}_{T \times M} \quad \mathbf{X} = \begin{pmatrix} x^1_{t_1} & x^1_{t_2} & \cdots & x^1_{t_T} \\ x^2_{t_1} & x^2_{t_2} & \cdots & x^2_{t_T} \\ \vdots & \vdots & \ddots & \vdots \\ x^N_{t_1} & x^N_{t_2} & \cdots & x^N_{t_T} \end{pmatrix}_{N \times T}$$

$$\theta = (\vec{x}_0, k, \tau_0, \alpha, \beta)$$

Likelihood

$$P(\mathbf{S}, \rho|\mathbf{X}, \theta) = P(\mathbf{S}|\mathbf{X}, \theta)P(\rho|\mathbf{S}, \theta)$$

$$P(\mathbf{S}|\mathbf{X}, \theta) = \prod_{i=1}^{M} \prod_{j=1}^{T} P(s^i_{t_j}|\vec{x}_{t_j}, \theta)$$

$$P(s^i_t|\vec{x}_t, \theta) \sim \mathcal{N}(\alpha(\log\langle G_i\,, e^{\vec{x}_t}\rangle + \beta), \epsilon_1), \text{where,}$$

G is a projection matrix from source space to sensor space

$$P(\vec{\rho}^i|\mathbf{S}, \theta) \sim \prod_{i=1}^{M} \mathcal{N}(\frac{1}{T}\sum_{j=1}^{T}(s^i_{t_j})^2, \epsilon_2)$$

Priors

Source dynamics are assumed to follow trajectories given by 2D Epileptor

$$\dot{x}_i = 1 - x_i^3 - 2x_i^2 - z_i + I_1$$

$$\dot{z}_i = \frac{1}{\tau_0}\left(4(x_i - x_0) - z_i - \sum_{j=1}^{N} K_{ij}(x_j - x_i)\right)$$

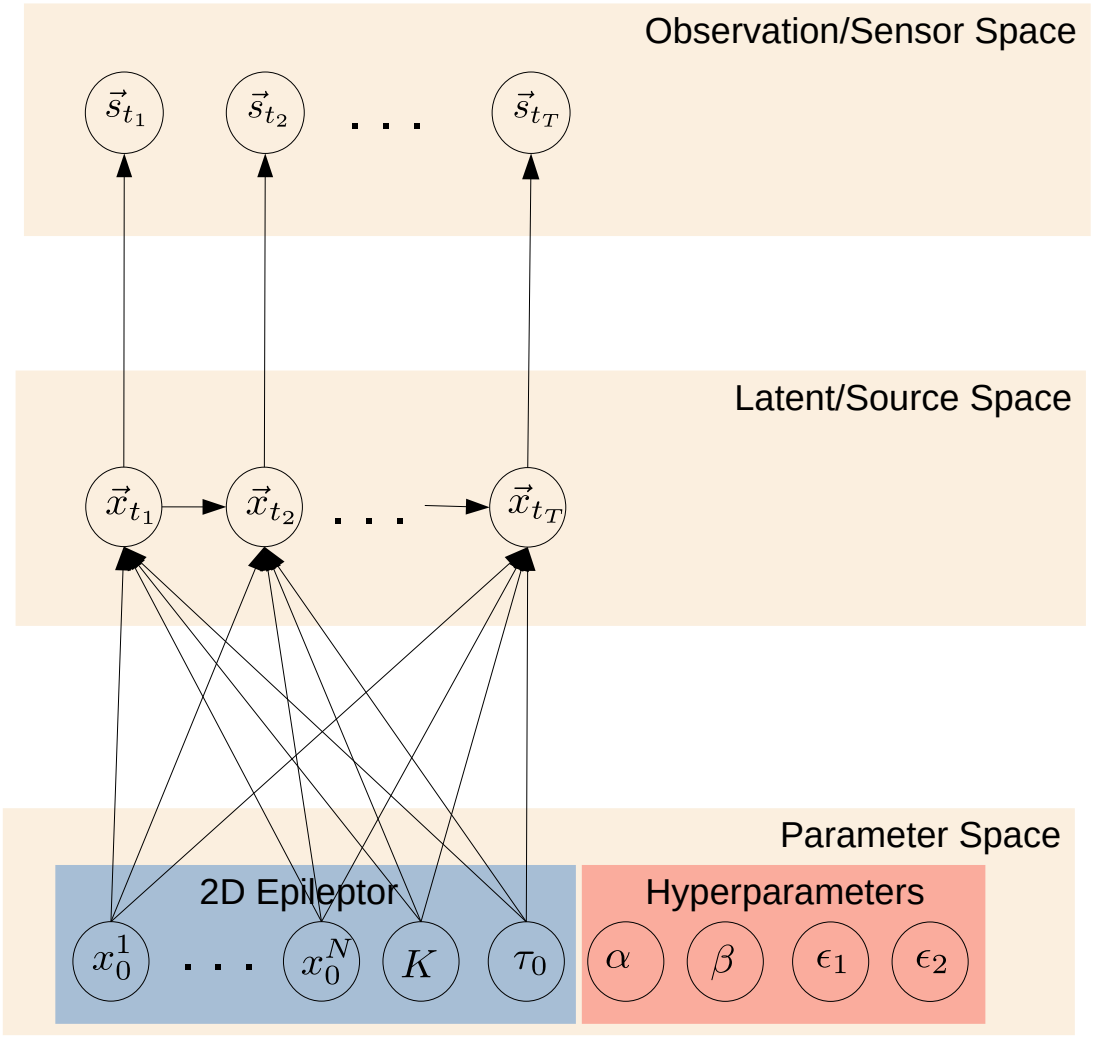


Figure 3: Probabilistic graphical model of virtual epileptic patient

Parameter	*Prior*
x_0^i	$\mathcal{N}(\mu_{x_0}^i, 1)$
K	$\mathcal{N}(1, 10)$
τ_0	$\mathcal{N}(20, 10)$
α	$\mathcal{N}(1, 10)$
β	$\mathcal{N}(0, 10)$
ϵ_1	$\mathcal{N}(1, 10)$
ϵ_2	$\mathcal{N}(1, 10)$

Table 1: Prior distributions of all inferred parameters

$$
\begin{aligned}
P(\theta, \mathbf{X}) &= P(\mathbf{X}|\theta)P(\theta) \\
P(\mathbf{X}|\theta) &= P(\vec{x}_{t_1}|\theta) \prod_{j=2}^{T} P(\vec{x}_{t_j}|\vec{x}_{t_{j-1}}, \theta) \\
P(\vec{x}_{t_j}|\vec{x}_{t_{j-1}}, \theta) &= \delta(x_{t_j} - f(x_{t_{j-1}}, \theta)) \\
f(x_{t_{j-1}}, \theta)) &= x_{t_{j-1}} + dt\dot{x} \\
P(\theta) &= P(x_0, K, \tau_0, \alpha, \beta) \\
&= P(x_0)P(K)P(\tau_0)P(\alpha)P(\beta)
\end{aligned}
$$

Prior distributions of all inferred parameters are given in table 1

Results

NUTS parameters across iterations and pair plots for inferred scalar parameters from 4 chains are shown in figures 4 - 7.

sampler parameters:

No. of warmup iterations	:	200
No. of sampling iterations	:	200
Target acceptance statistic (δ)	:	0.9
Max. tree depth	:	10
Metric	:	diagonal

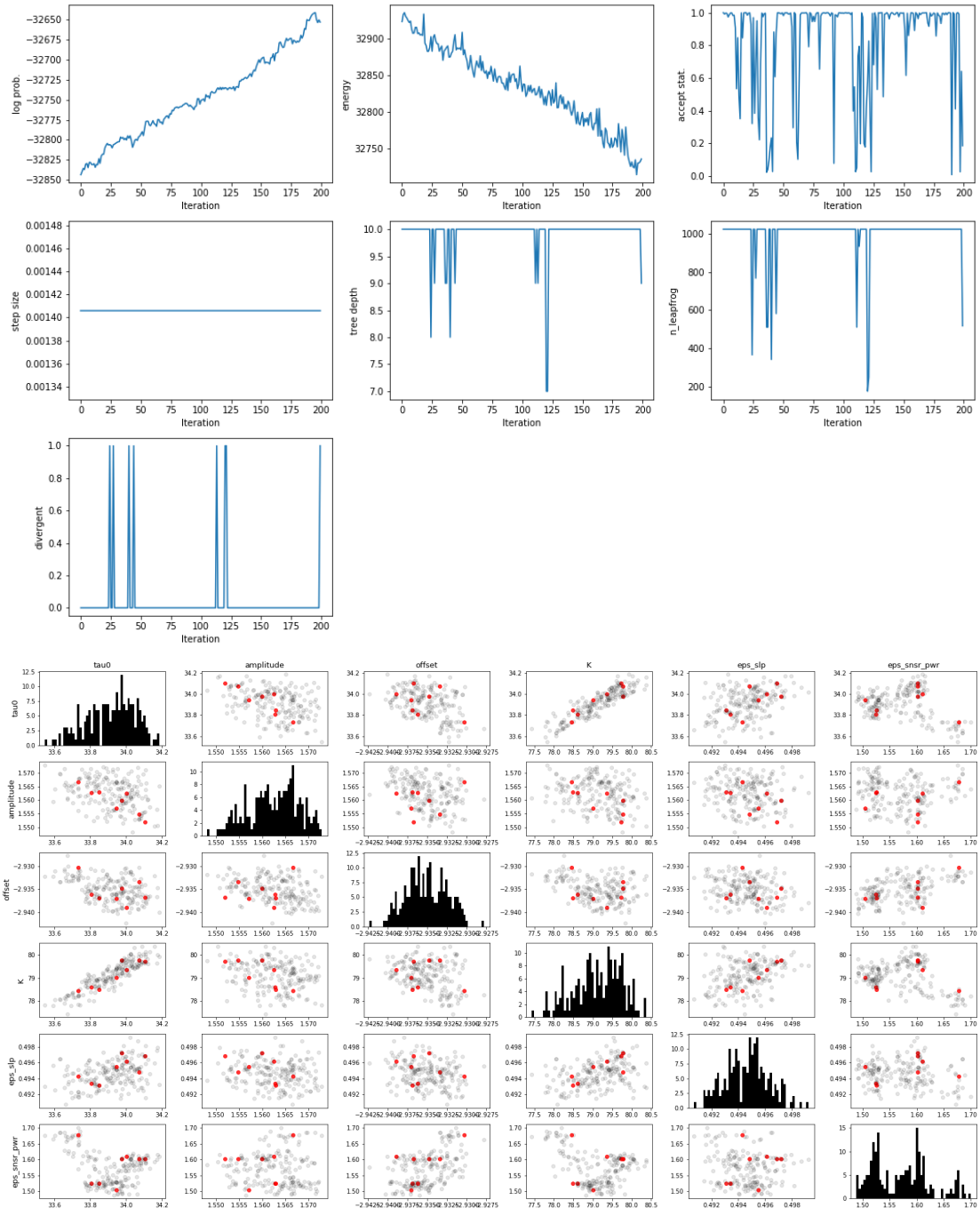


Figure 4: NUTS diagnostics and pair plots of scalar parameters from Chain 1

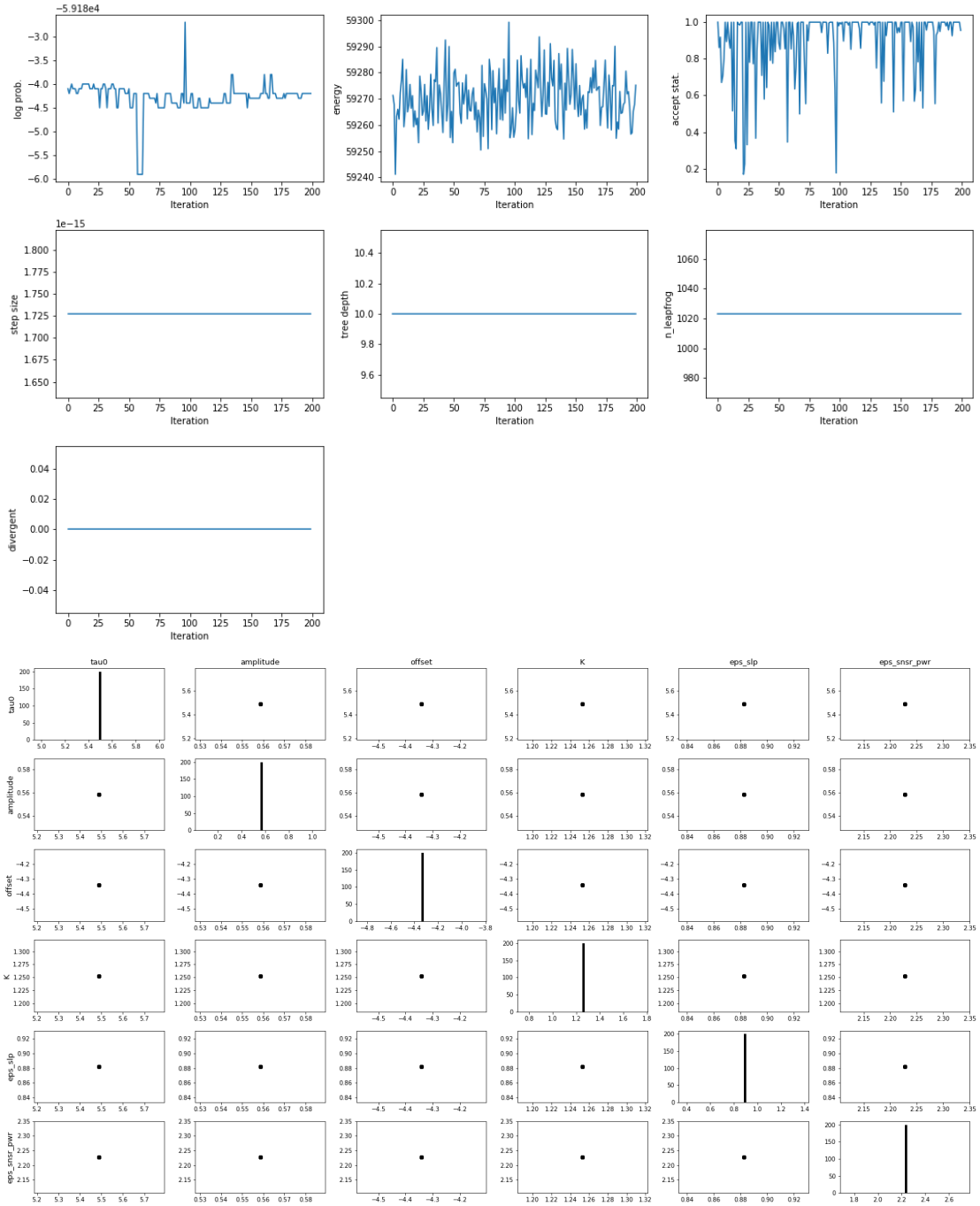


Figure 5: NUTS diagnostics and pair plots of scalar parameters from Chain 2

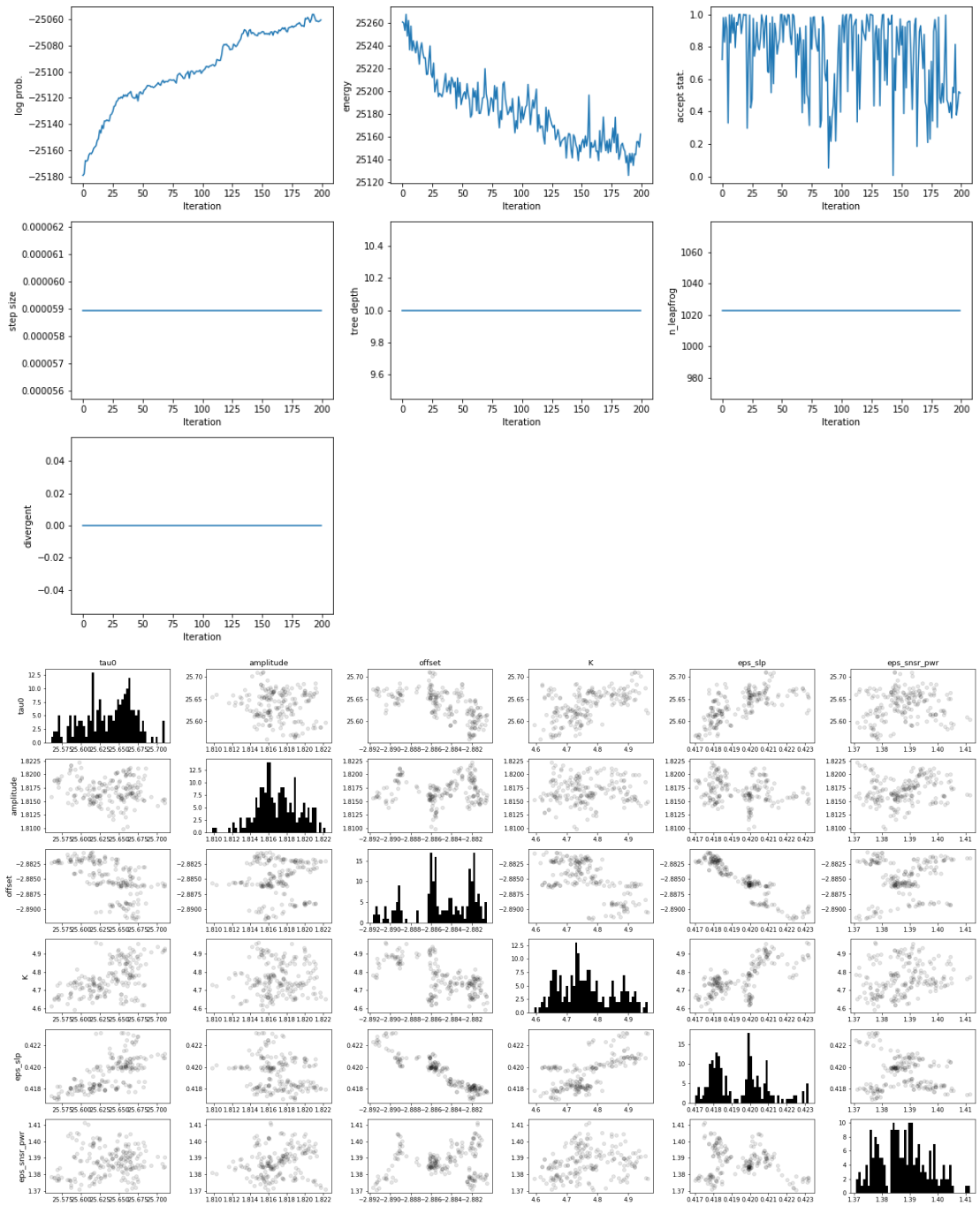


Figure 6: NUTS diagnostics and pair plots of scalar parameters from Chain 3

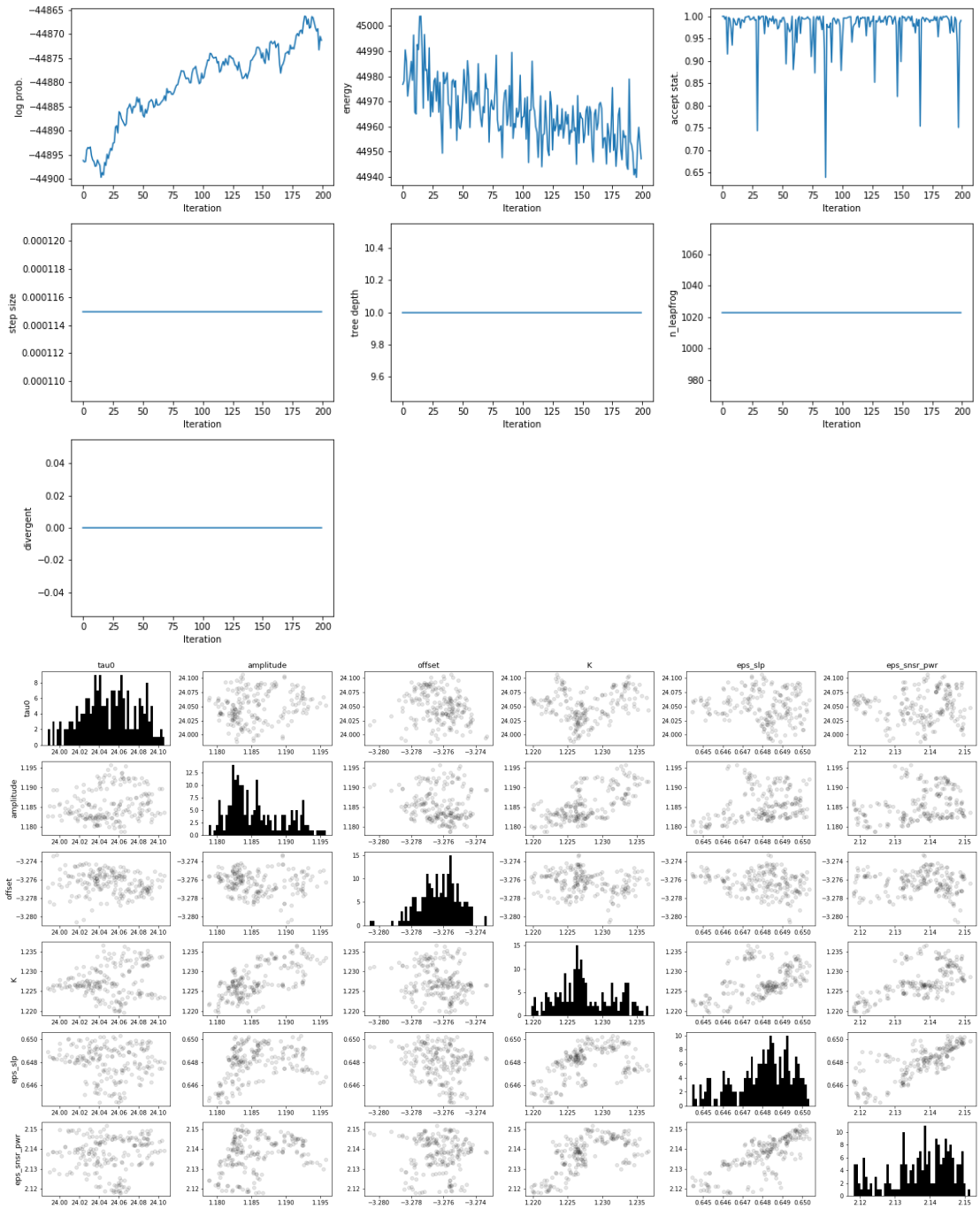


Figure 7: NUTS diagnostics and pair plots of scalar parameters from Chain 4

Problems/Diagnostics

- All chains hit max. tree depth
- Sampling is very slow, it takes 3-4 days to finish 200 warmup + 200 sampling itearations
- Some chains report divergent transitions and fixing observation noise($\epsilon_1 = 1, \epsilon_2 = 1$) causes many divergent transitions in all chains.

Some diagnostics from cmdstan *diagnose* utility:

```
789 of 1600 (49%) transitions hit the maximum treedepth
  limit of 10, or 2^10 leapfrog steps. Trajectories that
  are prematurely terminated due to this limit will
  result in slow exploration and you should increase the
  limit to ensure optimal performance.

8 of 1600 (0.5%) transitions ended with a divergence.
  These divergent transitions indicate that HMC is not
  fully able to explore the posterior distribution. Try
  rerunning with adapt delta set to a larger value and
  see if the divergences vanish. If increasing adapt
  delta towards 1 does not remove the divergences then
  you will likely need to reparameterize your model.

The E-BFMI, 0.0049, is below the nominal threshold of 0.3
  which suggests that HMC may have trouble exploring the
  target distribution. You should consider any
  reparameterizations if possible.

The following parameters had split R-hat greater than 1.1:
Almost all the parameters
```

Questions

- Inference works well while using the source activity as observations but performs poorly while using the sensor activity as observations. Does this mean nonlinear correlations between parameters are induced by the transformation from source space to sensor space? If so how could we address this?

- Some chains for example chain 2 (see fig. 5) get stuck at a point and lead to samples with almost zero variance. Is this due to pathological geometries like Neal's funnel in the posterior? Are there any other reasons for such a behavior?

References

[1] V. K. Jirsa, T. Proix, D. Perdikis, M. M. Woodman, H. Wang, J. Gonzalez-Martinez, C. Bernard, C. Bénar, M. Guye, P. Chauvel, et al. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. *Neuroimage*, 145:377–388, 2017.

[2] V. K. Jirsa, W. C. Stacey, P. P. Quilichini, A. I. Ivanov, and C. Bernard. On the nature of seizure dynamics. *Brain*, 137(8):2210–2230, 2014.

[3] T. Proix, F. Bartolomei, P. Chauvel, C. Bernard, and V. K. Jirsa. Permittivity coupling across brain regions determines seizure recruitment in partial epilepsy. *Journal of Neuroscience*, 34(45):15009–15021, 2014.