eLife

Training deep neural density estimators to identify mechanistic models of neural dynamics

Pedro J Gonçalves[1,2†*], Jan-Matthis Lueckmann[1,2†*], Michael Deistler[1,3†*], Marcel Nonnenmacher[1,2,4], Kaan Öcal[2,5], Giacomo Bassetto[1,2], Chaitanya Chintaluri[6,7], William F Podlaski[6], Sara A Haddad[8], Tim P Vogels[6,7], David S Greenberg[1,4], Jakob H Macke[1,2,3,9*]

Computational and Systems Neuroscience Journal Club

11/3/2022

Abuzar Mahmood - Katz Lab

Layout

- Motivation
- Caveats (what this talk will tell you, and what it won't)
- Background
    - Conditional Density Estimation
    - Mixing Neural Networks with Probabilistic Models
    - Prior/other methods (epsilon / likelihood-free methods)
        - ABC and SMB-ABC
- Results

1. Recap of Methodology
2. Comparison with established methodologies
3. Inferring visual receptive field
4. ~~Inferring ion channel parameters~~
5. Inferring parameters for Hodgkin-Huxley Model
6. Inferring parameters for STG and analyzing posterior distribution

Trade

Increasing
Model
Complexity

Motivation - We need mechanistic models

- Does it work how we think it works?
- Model vs parameters?

Systems Biology

Cosmology and Particle Physics

Neuroscience

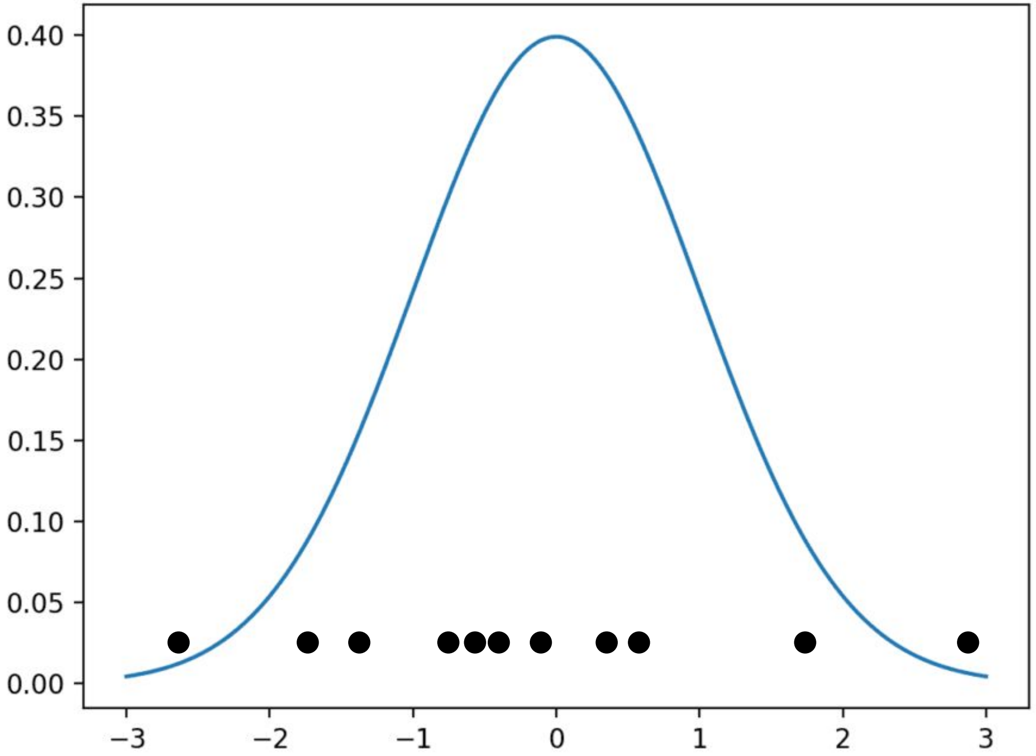
https://dark.nbi.ku.dk/news/dark-news-21/simulation-based-inference-of-dynamical-galaxy-cluster-masses-with-3d-convolutional-neural-networks/
https://sethna.lassp.cornell.edu/Sloppy/SloppySignalTransduction.html

Motivation - We need the distribution of relevant parameters

- Adjusting free parameters to be consistent with experimental observations
  - Essential for investigating whether the model agrees with reality
  - For **gaining insight into processes which cannot be measured experimentally (latent variables)**
- To understand how a model quantitatively explains data, necessary to find not only the best, but all parameter settings consistent with experimental observations
  - Something not possible/very difficult with point optimization techniques

Motivation – Likelihood-free inference

- What is a likelihood?
  - How likely are data to come from a parametric distribution?



$$p(x_n) = \mathcal{N}(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$p(\{x_1, ..., x_N\}|\pi_k, \mu_k, \sigma_k) = \prod_{n=1}^{N} p(x_n),$$

Likelihood allows us to use gradient-based methods to fit parameters

No likelihood □ Not discussing likelihood-based methods

Motivation – Current State of Affairs

- Gold standard for automated parameter identification is statistical inference, which uses the likelihood to quantify the match between parameters and data
- Likelihoods can be efficiently computed for purely statistical models commonly used in neuroscience, but are computationally intractable for most mechanistic models

Likelihood-based Inference (on parametric distributions)

Likelihood-free inference

Layout

- Motivation
- **Caveats (what this talk will tell you, and what it won't)**
- Background
  - Conditional Density Estimation
  - Mixing Neural Networks with Probabilistic Models
  - Related work (epsilon / likelihood-free methods)
    - ABC and SMB-ABC
- Results
  1. Recap of Methodology
  2. Comparison with established methodologies
  3. Inferring visual receptive field
  4. ~~Inferring ion channel parameters~~
  5. Inferring parameters for Hodgkin-Huxley Model
  6. Inferring parameters for STG and analyzing posterior distribution

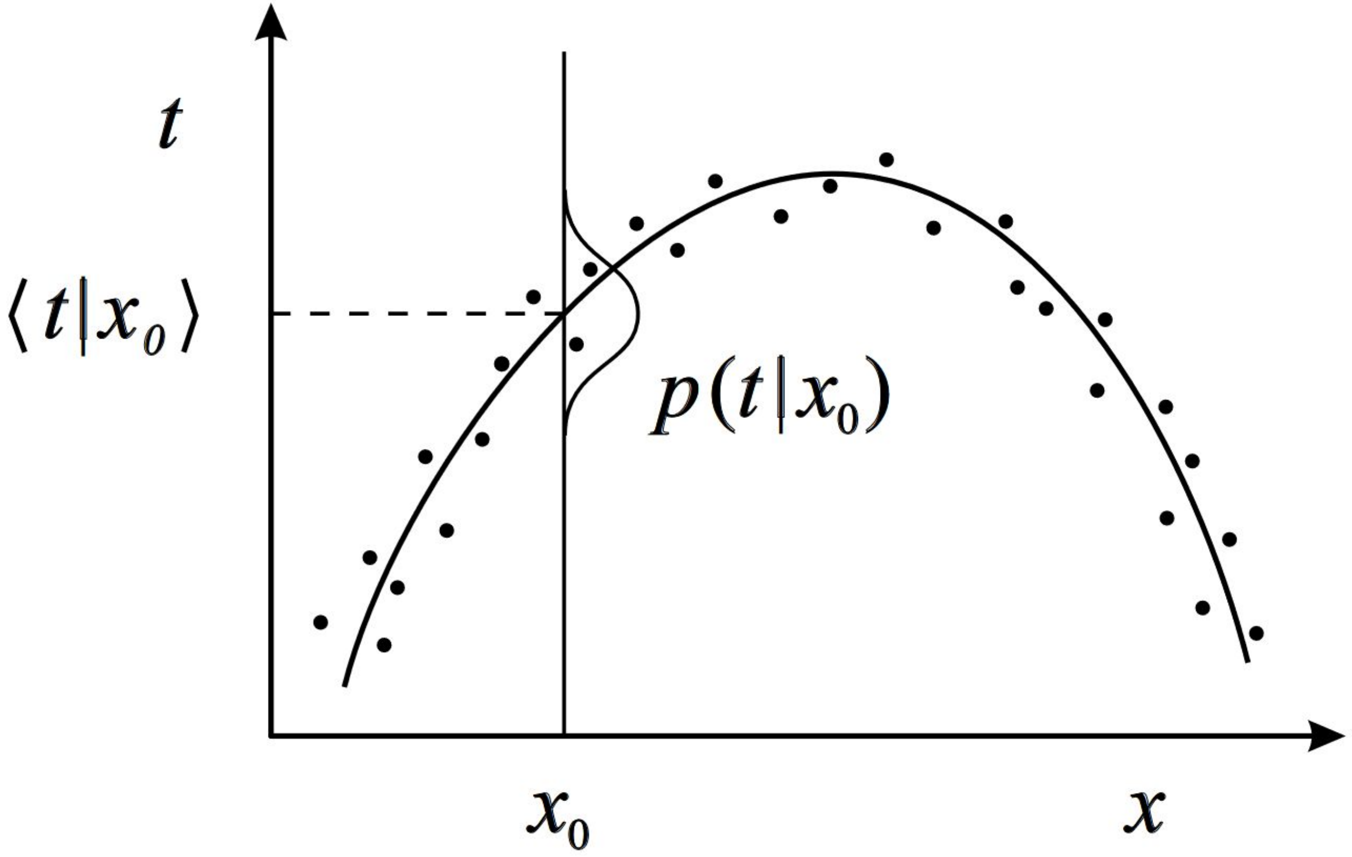
Caveats

- Will not be diving into the math
- High-level concepts and results
- No new science :/

Layout

- Motivation
- Caveats (what this talk will tell you, and what it won't)
- **Background**
  - **Conditional Density Estimation**
  - **Mixing Neural Networks with Probabilistic Models**
  - Related work (epsilon / likelihood-free methods)
    - ABC and SMB-ABC
- Results
  1. Recap of Methodology
  2. Comparison with established methodologies
  3. Inferring visual receptive field
  4. ~~Inferring ion channel parameters~~
  5. Inferring parameters for Hodgkin-Huxley Model
  6. Inferring parameters for STG and analyzing posterior distribution

Background – Conditional Density Estimation

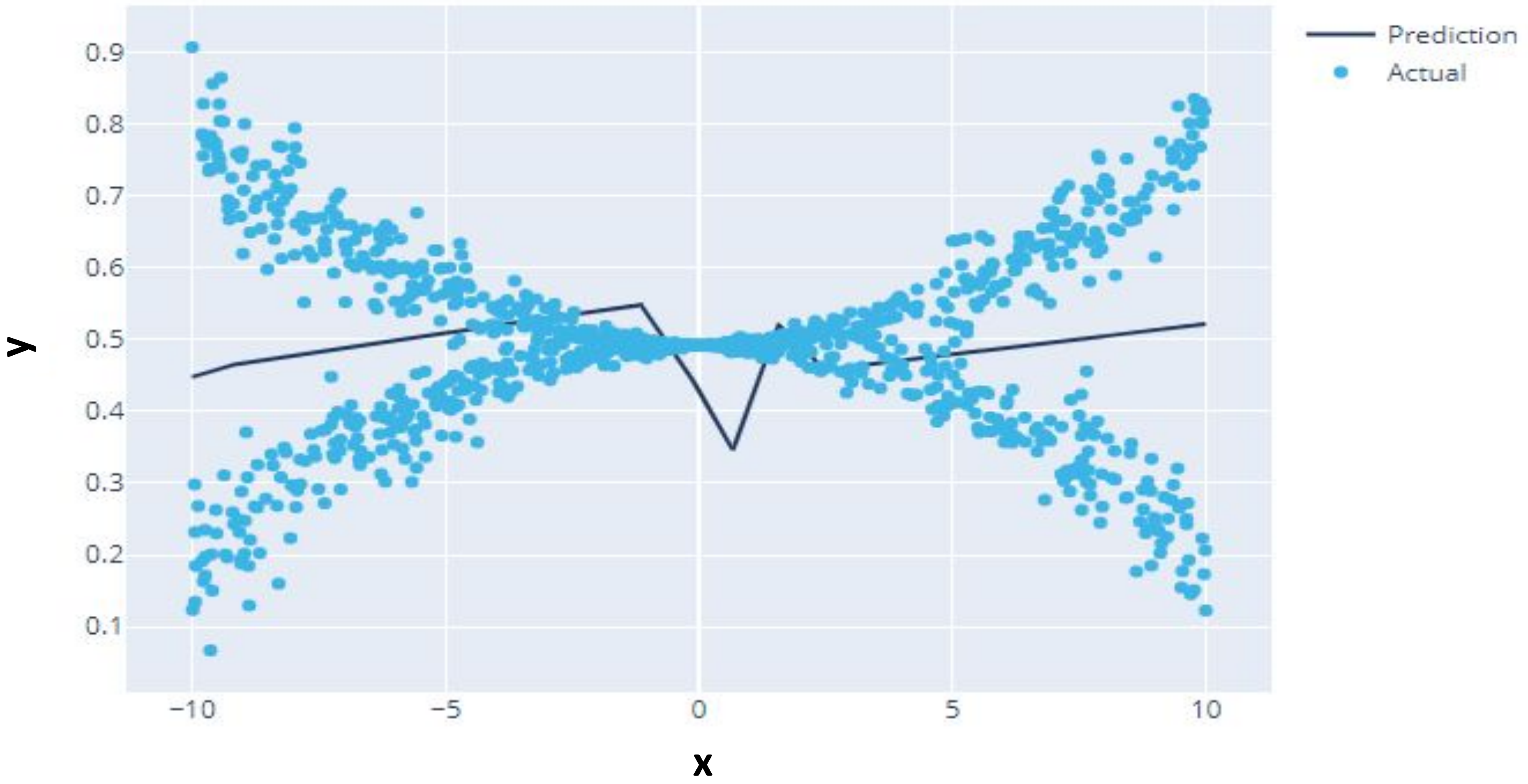


1. Why do we want to learn a conditional density?
   - Posterior → $p(\theta|x)$

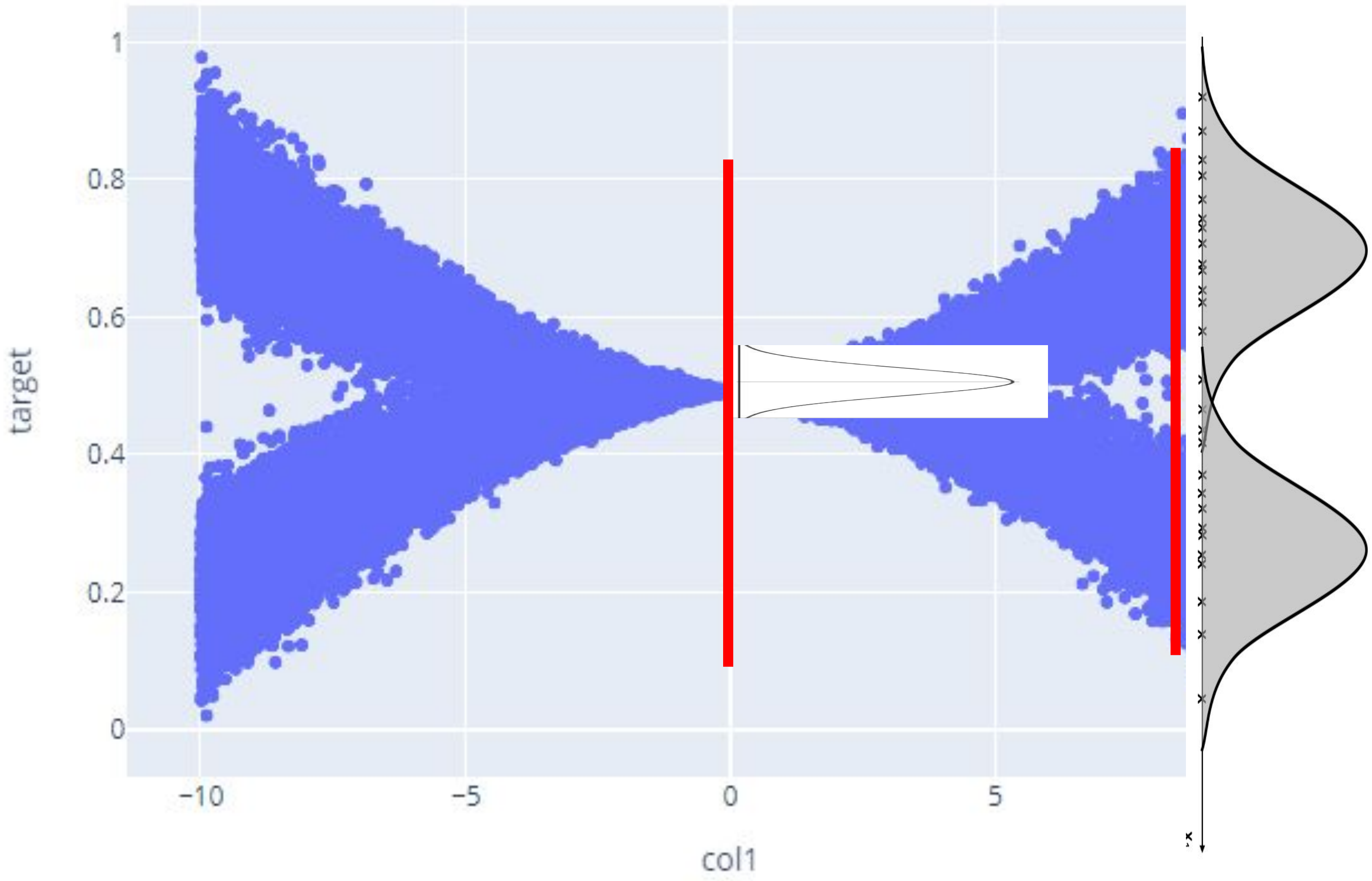
2. Why do we need to learn a density rather than the mean value here?

Background – Conditional Density Estimation

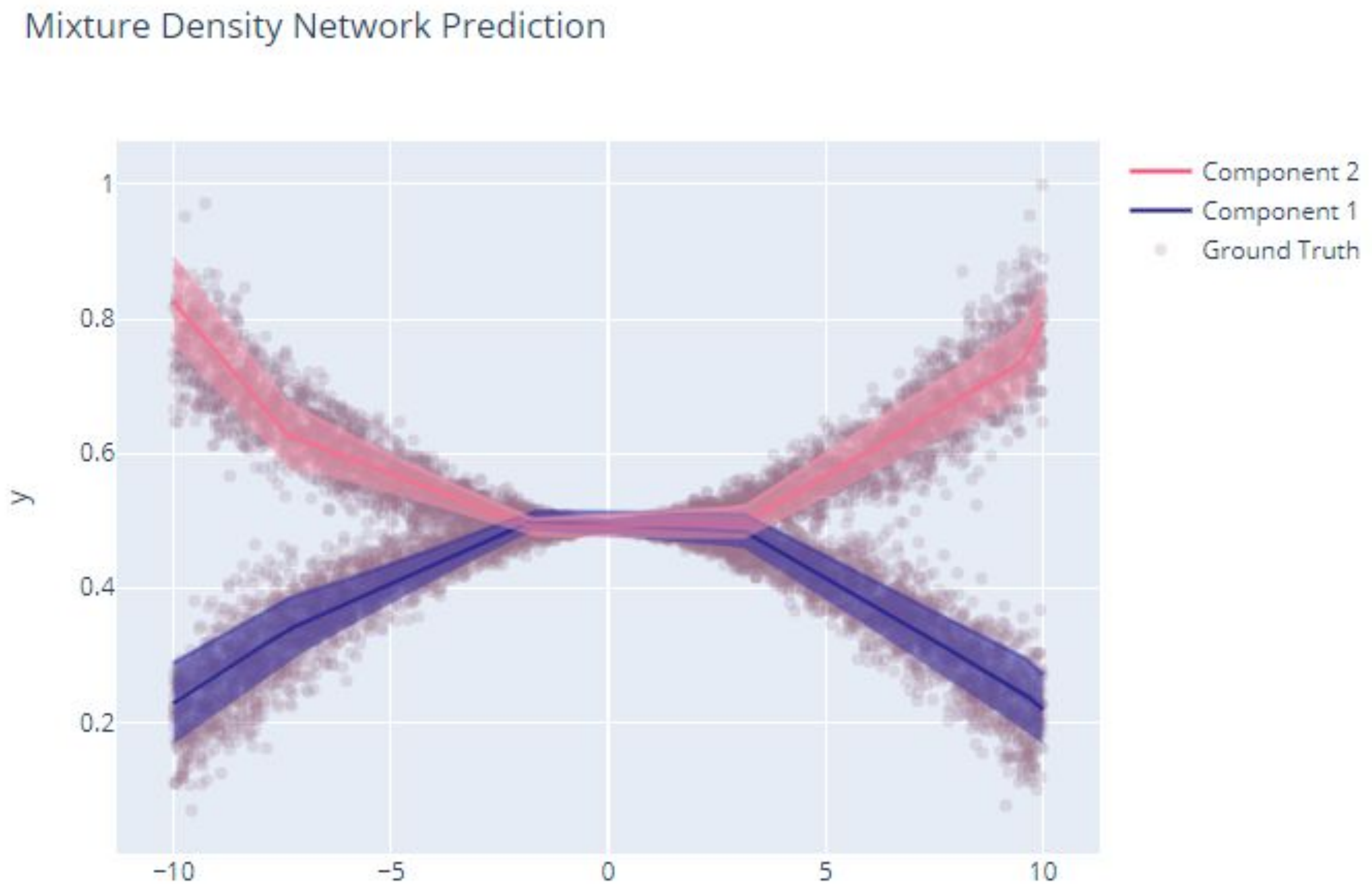
Vanilla Neural Network



Background – Conditional Density Estimation



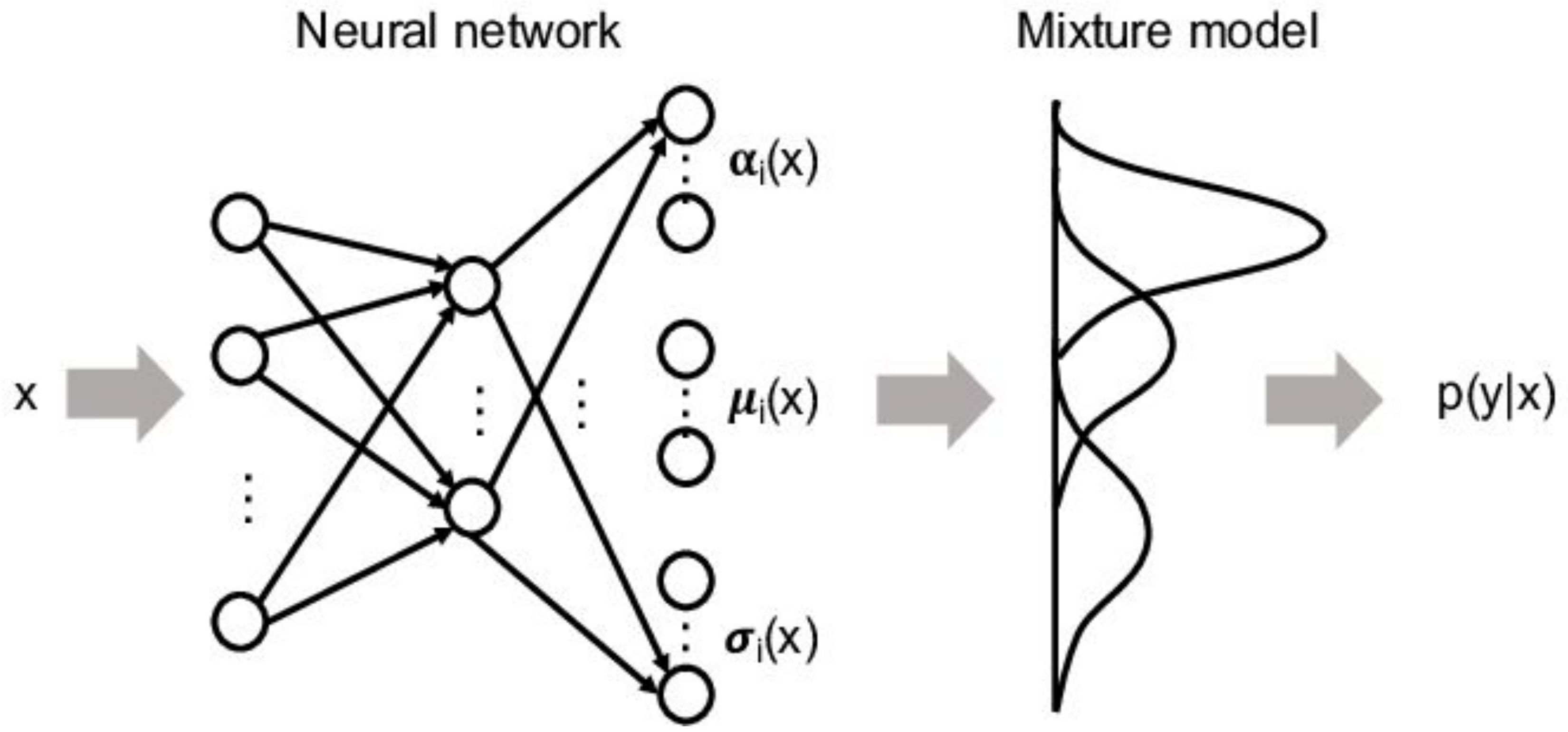
Background – Conditional Density Estimation



1. How does one learn the conditional density?

Background – Mixing Neural Networks with Probabilistic Models

- Example – Mixture Density Network

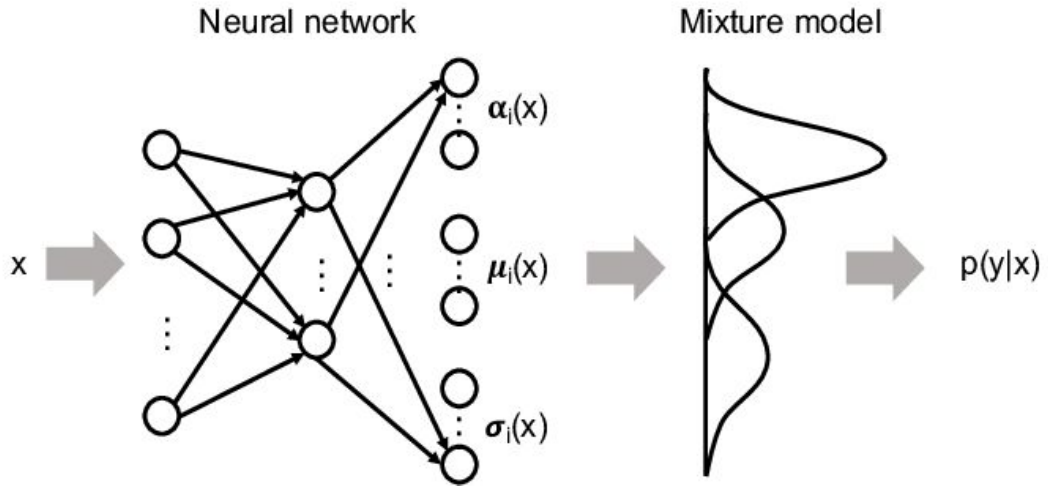


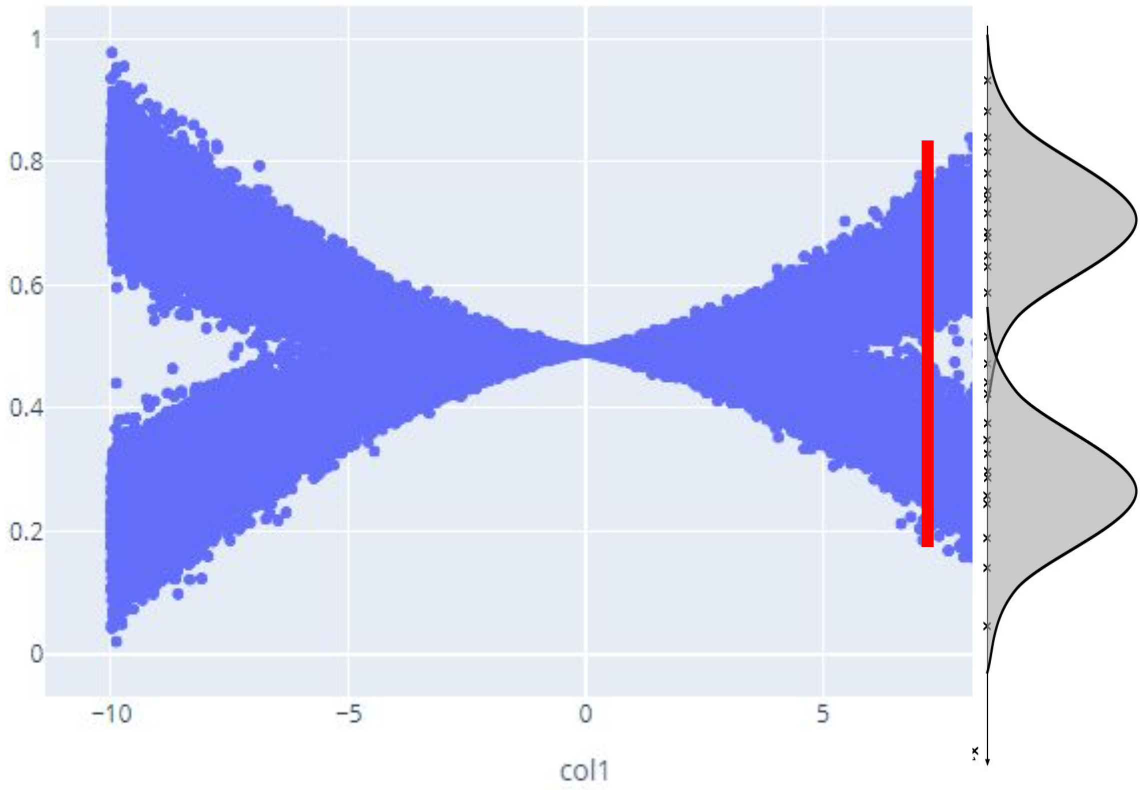
- Are gaussian mixtures enough to learn complex distributions?
  - YES, given enough components
  - But authors also use normalizing flows

Cost function □ Gaussian Mixture Log-likelihood

Background – Mixing Neural Networks with Probabilistic Models

- Example – Mixture Density Network
  - **Differentiable representation of the density – Important for later**





Layout

- Motivation
- Caveats (what this talk will tell you, and what it won't)
- **Background**
  - Conditional Density Estimation
  - Mixing Neural Networks with Probabilistic Models
  - **Related work (epsilon / likelihood-free methods)**
    - **ABC and SMB-ABC**
- Results
  1. Recap of Methodology
  2. Comparison with established methodologies
  3. Inferring visual receptive field
  4. ~~Inferring ion channel parameters~~
  5. Inferring parameters for Hodgkin-Huxley Model
  6. Inferring parameters for STG and analyzing posterior distribution

Related work – Likelihood-free inference

- Approximate Bayesian Inference (ABC)
- Rejection sampling ABC (ϵ)
  1. Given model
  2. Initialize random hyperparameters/parental variables from prior distributions
  3. Generate sample
  4. If | target – sample | < ϵ (according to some distance metric), keep sample, else reject
- Problems
  - Rejection rate too high for any interesting problem
  - Sampling in areas of low likelihood

- Introduction to Approximate Bayesian Computation (ABC)
  - https://darrenjw.wordpress.com/2013/03/31/introduction-to-approximate-bayesian-computation-abc/
- Tiny Data, Approximate Bayesian Computation and the Socks of Karl Broman
  - https://www.sumsar.net/blog/2014/10/tiny-data-and-the-socks-of-karl-broman/

Related work – Likelihood-free inference

- Sequential Monte-Carlo ABC
  - Similar to regular SMC/Particle Filter, but with likelihood replaced with a cost/distance function
  1. Given model
  2. Initialize random hyperparameters/parental variables from prior distributions
  3. Generate sample
  4. If | target – sample | < ϵ (according to some distance metric), keep sample, else reject
  5. Kept samples + noise (from pre-specified kernel) become new prior distributions
- Like a genetic algorithm, but with distributions

Training deep neural density estimators to identify mechanistic models of neural dynamics

Pedro J Gonçalves[1,2†*], Jan-Matthis Lueckmann[1,2†*], Michael Deistler[1,3†*], Marcel Nonnenmacher[1,2,4], Kaan Öcal[2,5], Giacomo Bassetto[1,2], Chaitanya Chintaluri[6,7], William F Podlaski[6], Sara A Haddad[8], Tim P Vogels[6,7], David S Greenberg[1,4], Jakob H Macke[1,2,3,9*]

Part Deux

Computational and Systems Neuroscience Journal Club

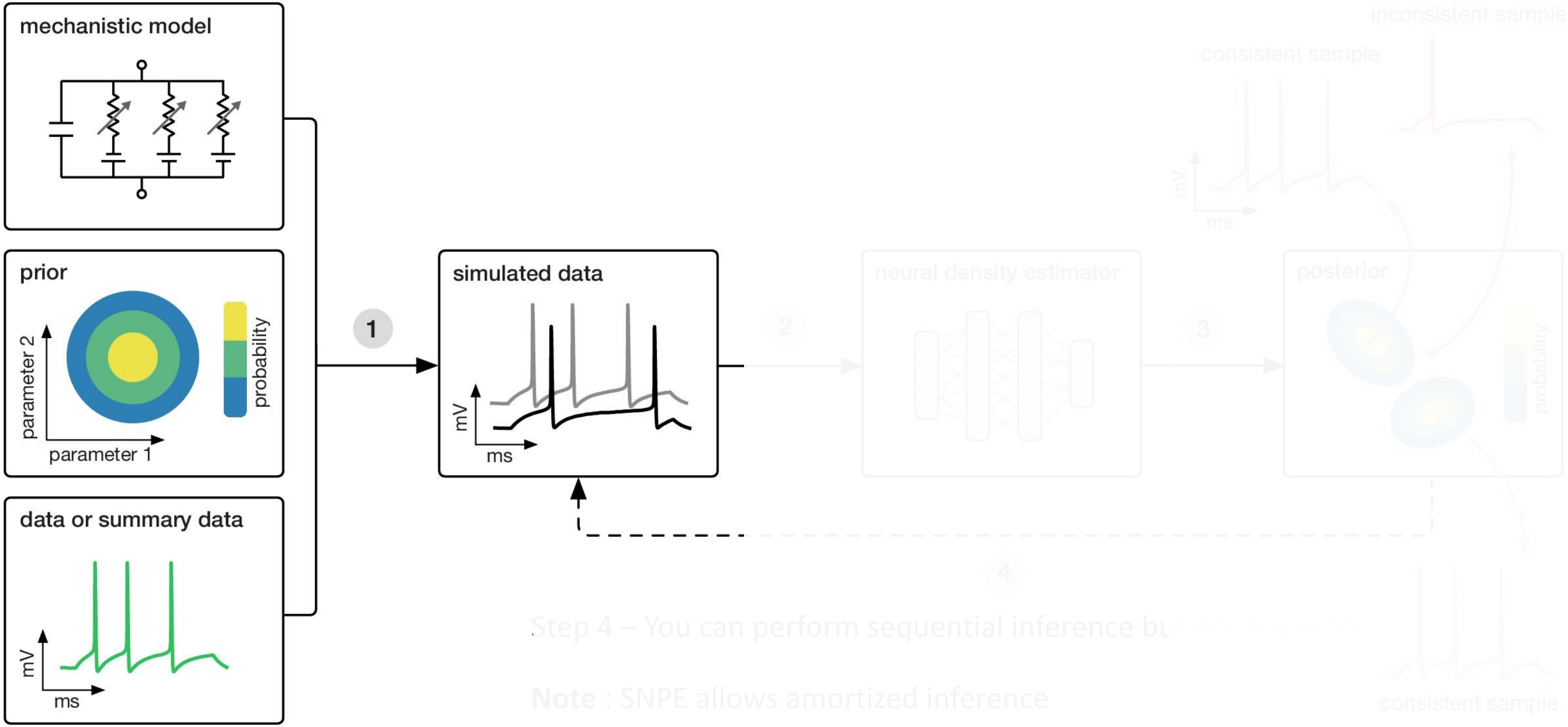
11/10/2022

Abuzar Mahmood - Katz Lab

Layout

- Motivation
- Caveats (what this talk will tell you, and what it won’t)
- Background
  - Conditional Density Estimation
  - Mixing Neural Networks with Probabilistic Models
  - Related work (epsilon / likelihood-free methods)
    - ABC and SMB-ABC
- **Results**
  1. **Recap of Methodology**
  2. **Comparison with established methodologies**
  3. **Inferring visual receptive field**
  4. ~~Inferring ion channel parameters~~
  5. Inferring parameters for Hodgkin-Huxley Model
  6. Inferring parameters for STG and analyzing posterior distribution

Results – Recap of Methodology
SNPE – Sequential Neural Posterior Estimation

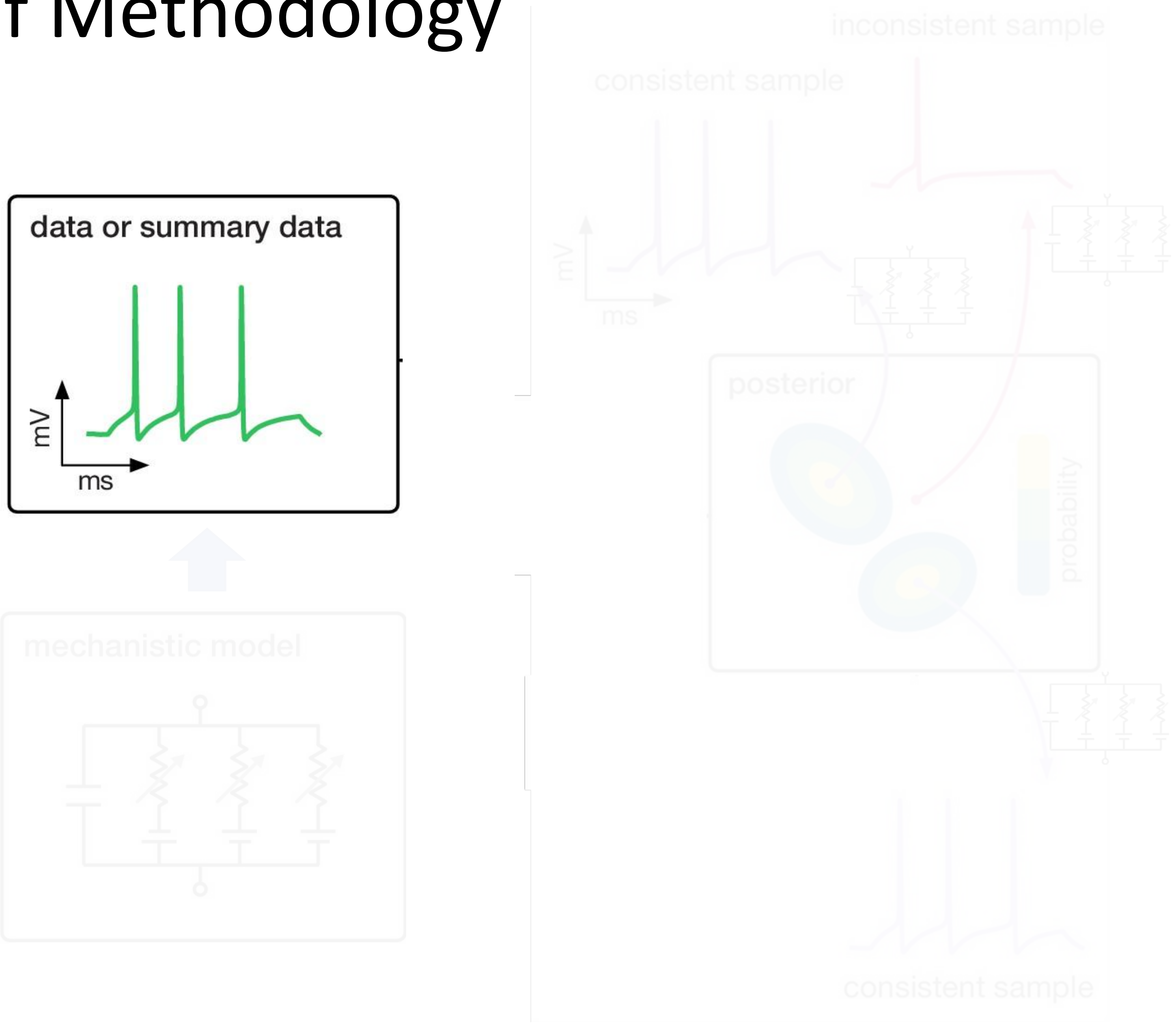


Step 4 – You can perform sequential inference bu

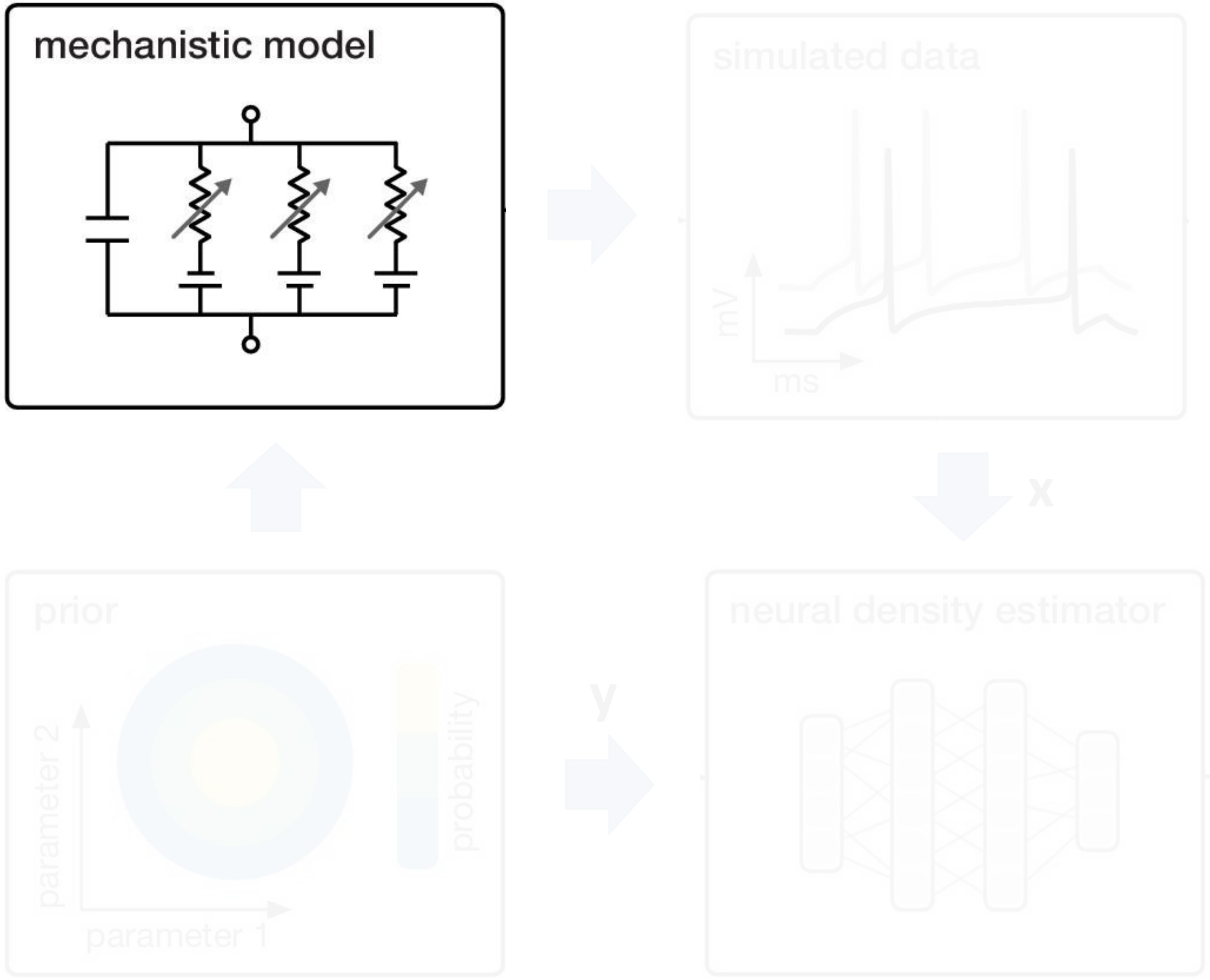
Note : SNPE allows amortized inference

Results – Recap of Methodology The Goal

- Not prescribing utility of having the posterior distribution
- Distribution of parameters can provide similar outputs
- Having full distribution provides understanding of variance (uncertainty) and covariances



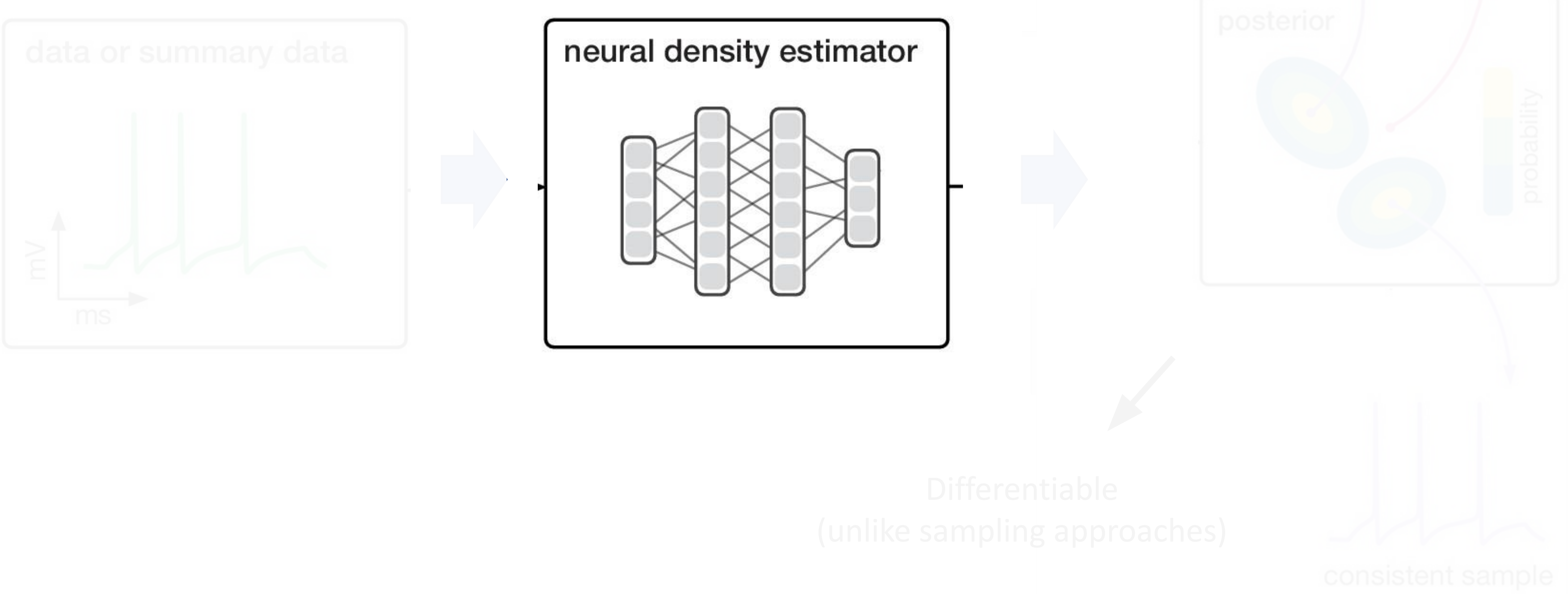
Results – Recap of Methodology
SNPE – Sequential Neural Posterior Estimation



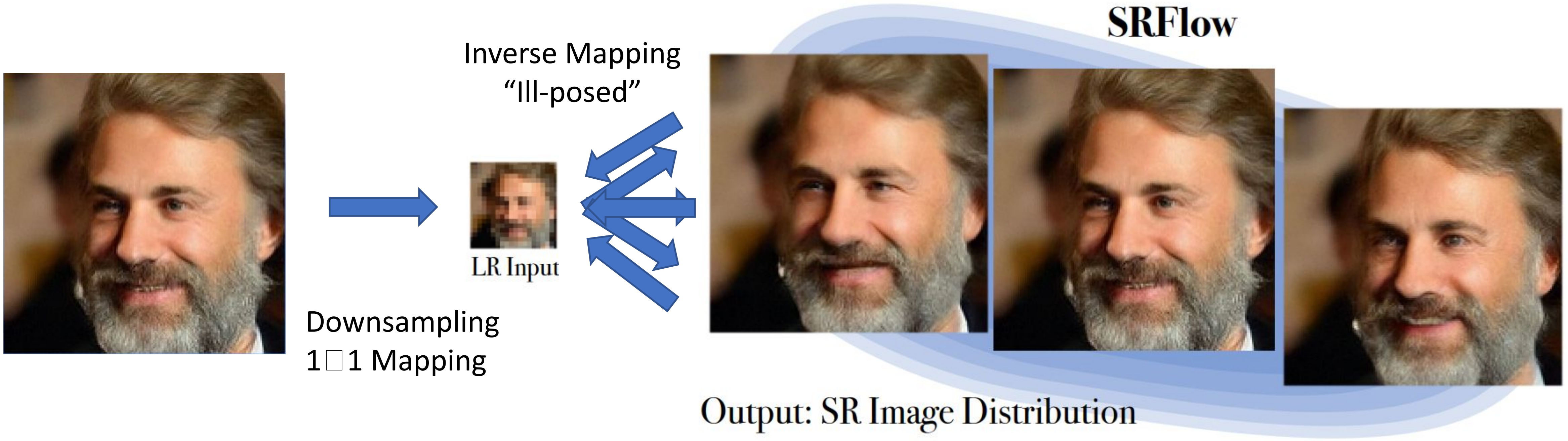
Learning distribution p(y|x)
- Enter data (x)
- Returned P(y|x)

Results – Recap of Methodology SNPE – Sequential Neural Posterior Estimation

Trained

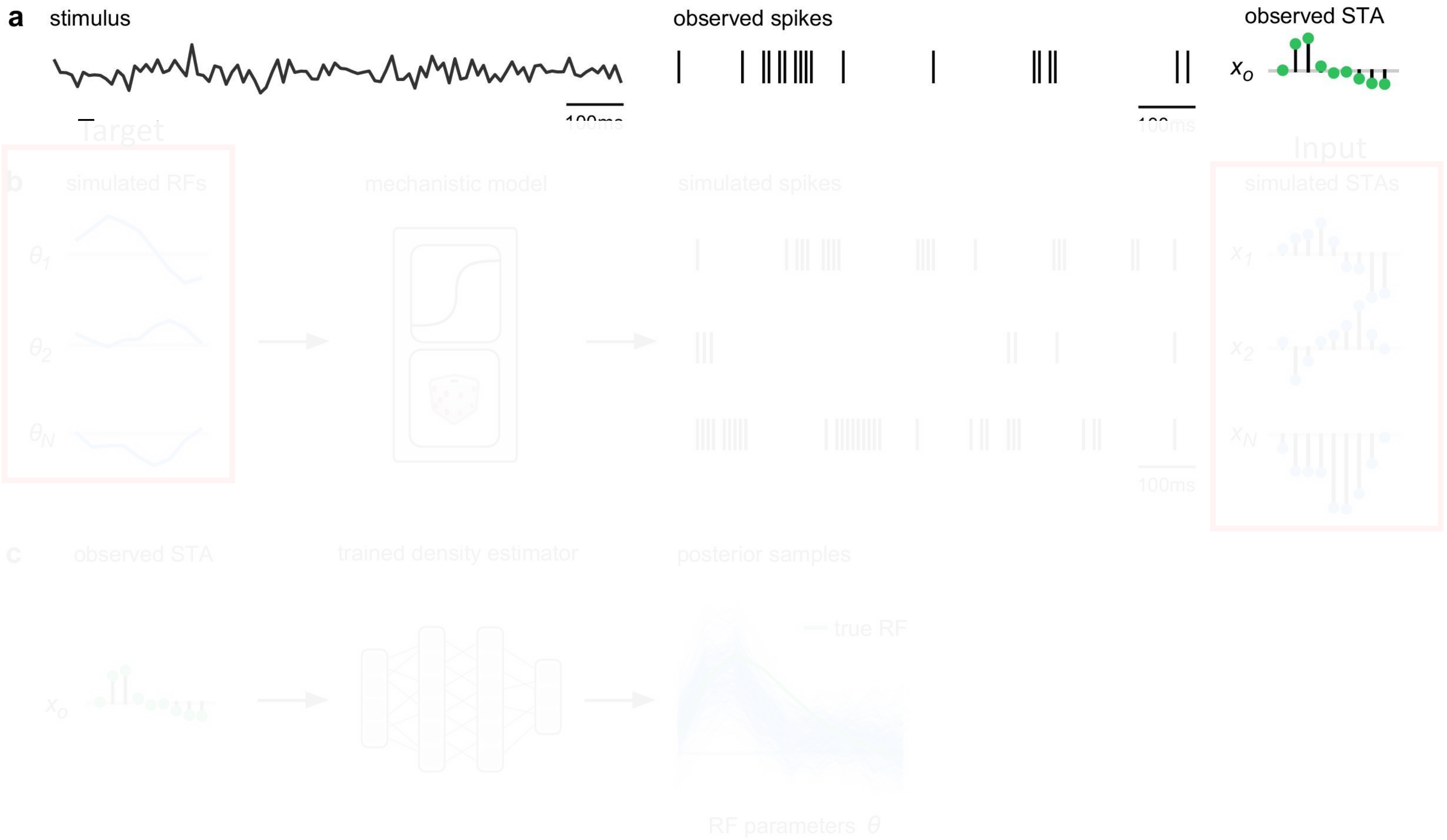


Similar “Interesting” Example

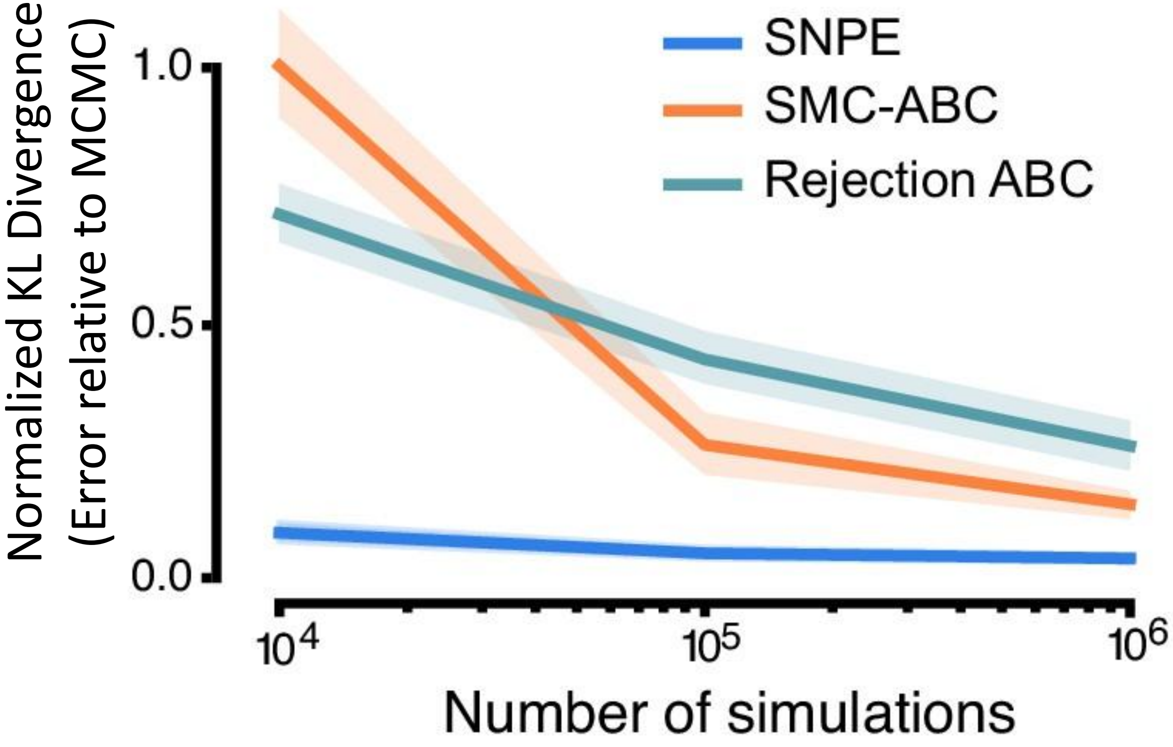


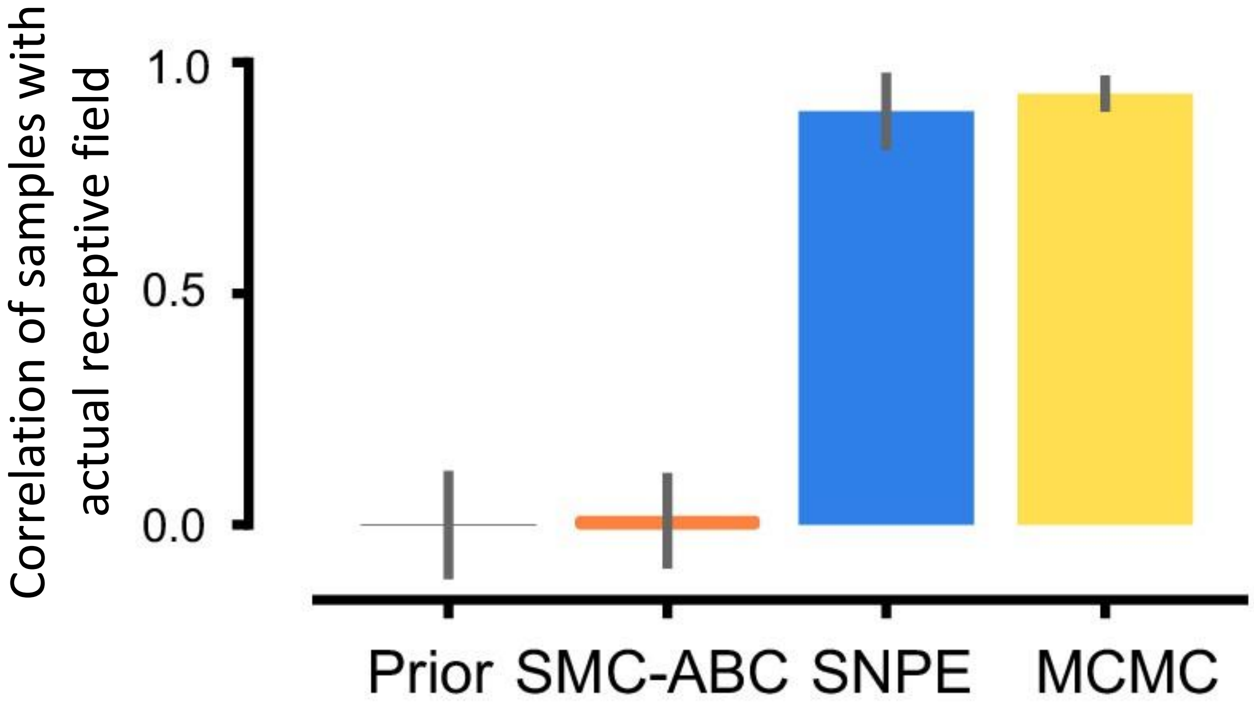
SRFlow: Learning the Super-Resolution Space with Normalizing Flow | https://arxiv.org/pdf/2006.14200.pdf

Results - Inferring receptive field (Linear-Nonlinear encoding model)



Results - Comparison with established methodologies (Inferring receptive field)



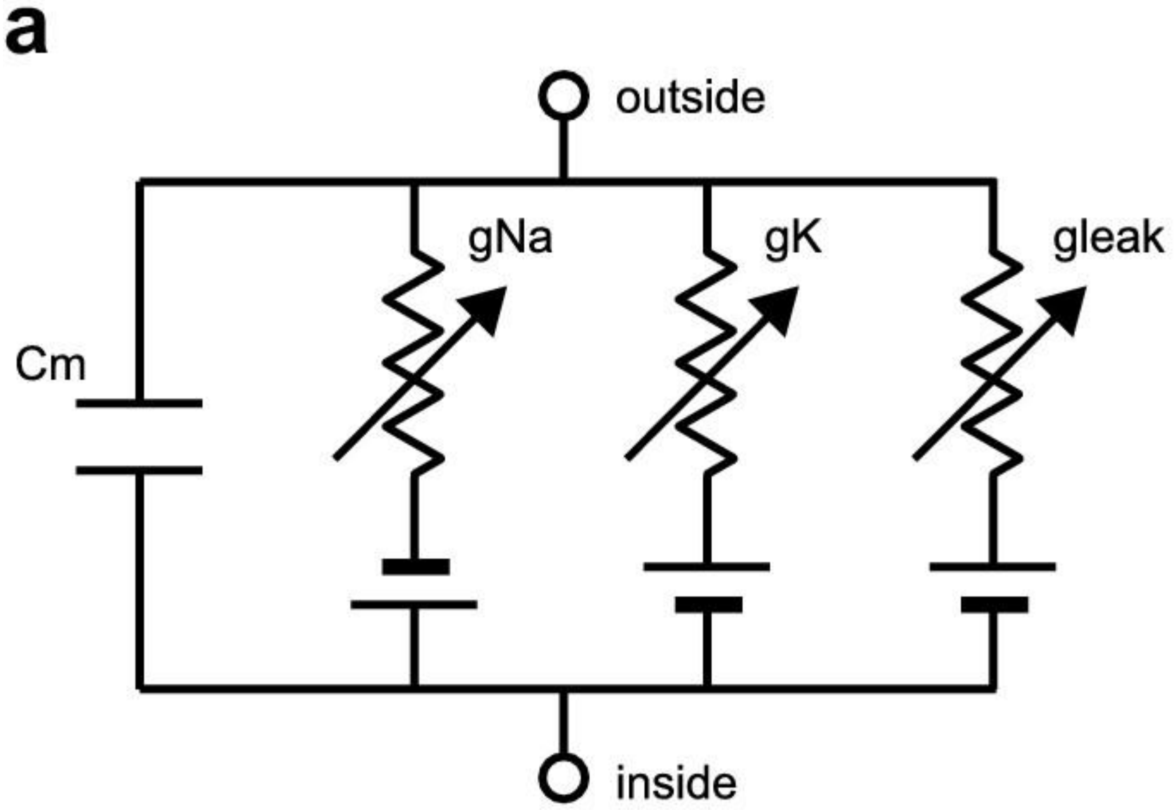


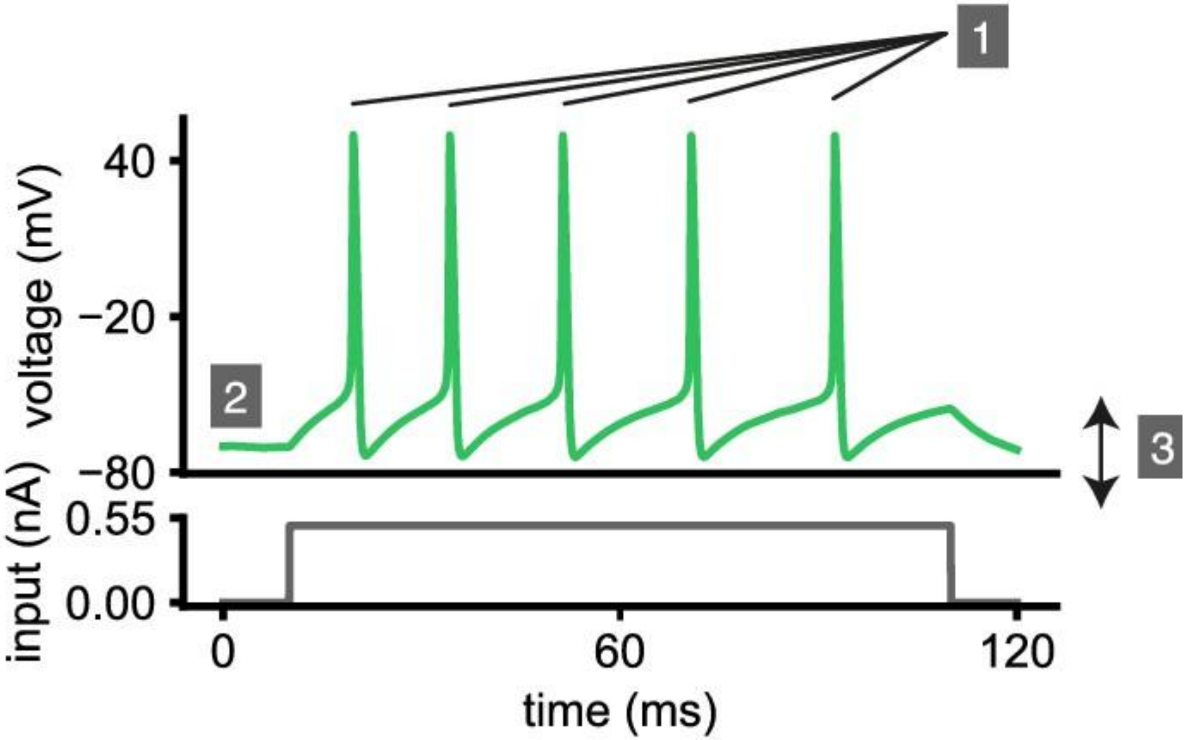
Layout

- Motivation
- Caveats (what this talk will tell you, and what it won't)
- Background
  - Conditional Density Estimation
  - Mixing Neural Networks with Probabilistic Models
  - Related work (epsilon / likelihood-free methods)
    - ABC and SMB-ABC
- **Results**
  1. Recap of Methodology
  2. Comparison with established methodologies
  3. Inferring visual receptive field
  4. ~~Inferring ion channel parameters~~
  5. **Inferring parameters for Hodgkin-Huxley Model**
  6. Inferring parameters for STG and analyzing posterior distribution

Results - Inferring parameters for Hodgkin-Huxley Model

(1) number of spikes,
(2) mean resting potential
(3) standard deviation of the pre-stimulus resting potential.

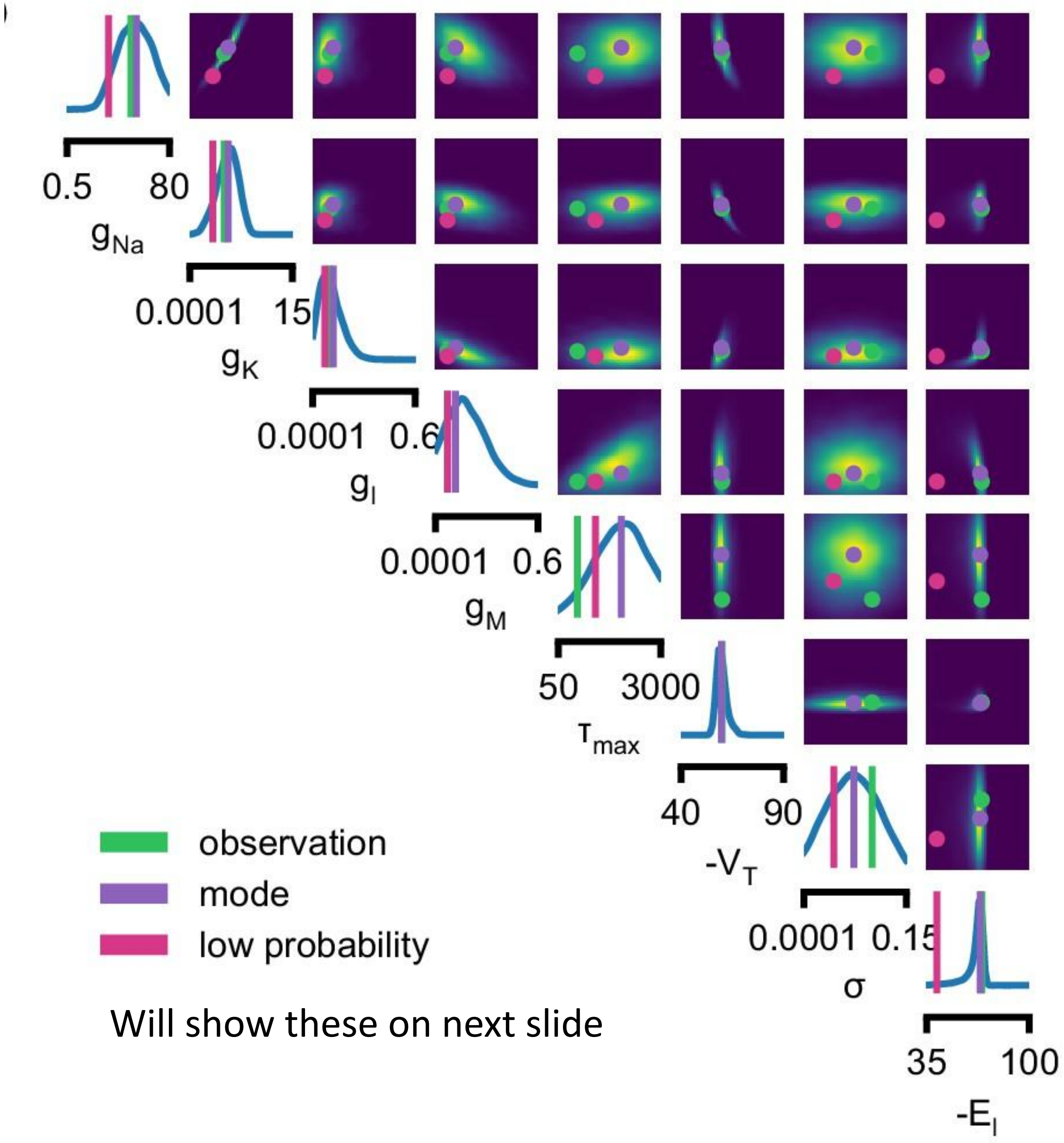




100,000 samples for 8 parameters
5^8 ~= 400,000

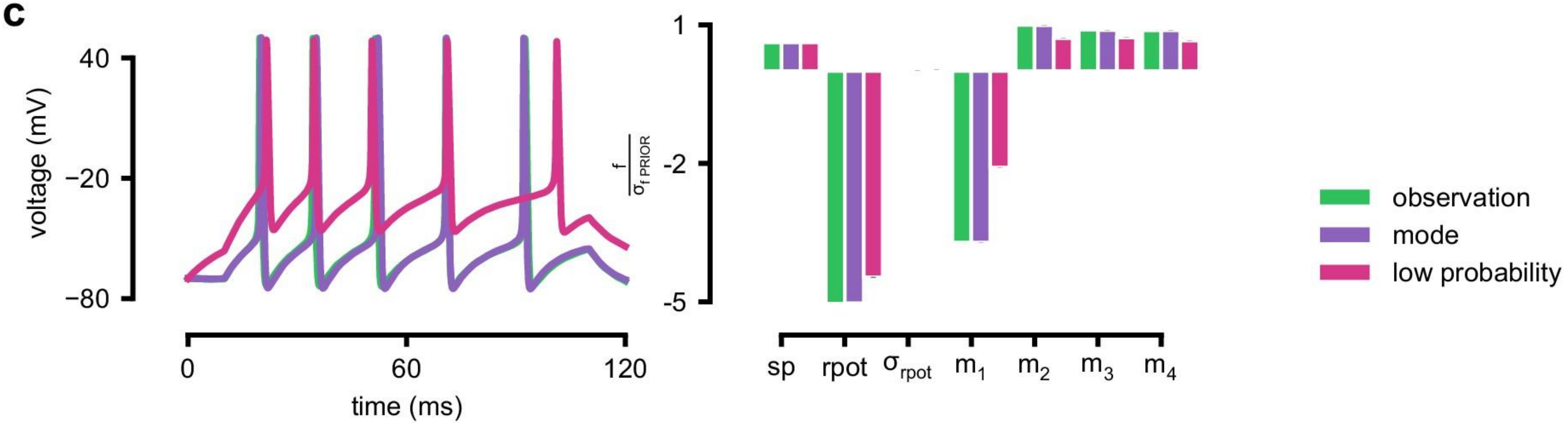
Results - Inferring parameters for Hodgkin-Huxley Model

- Interesting to note parameters which need to be tightly vs loosely tuned
- **Importantly :** Parameters for observation are NOT mode of posterior



Will show these on next slide

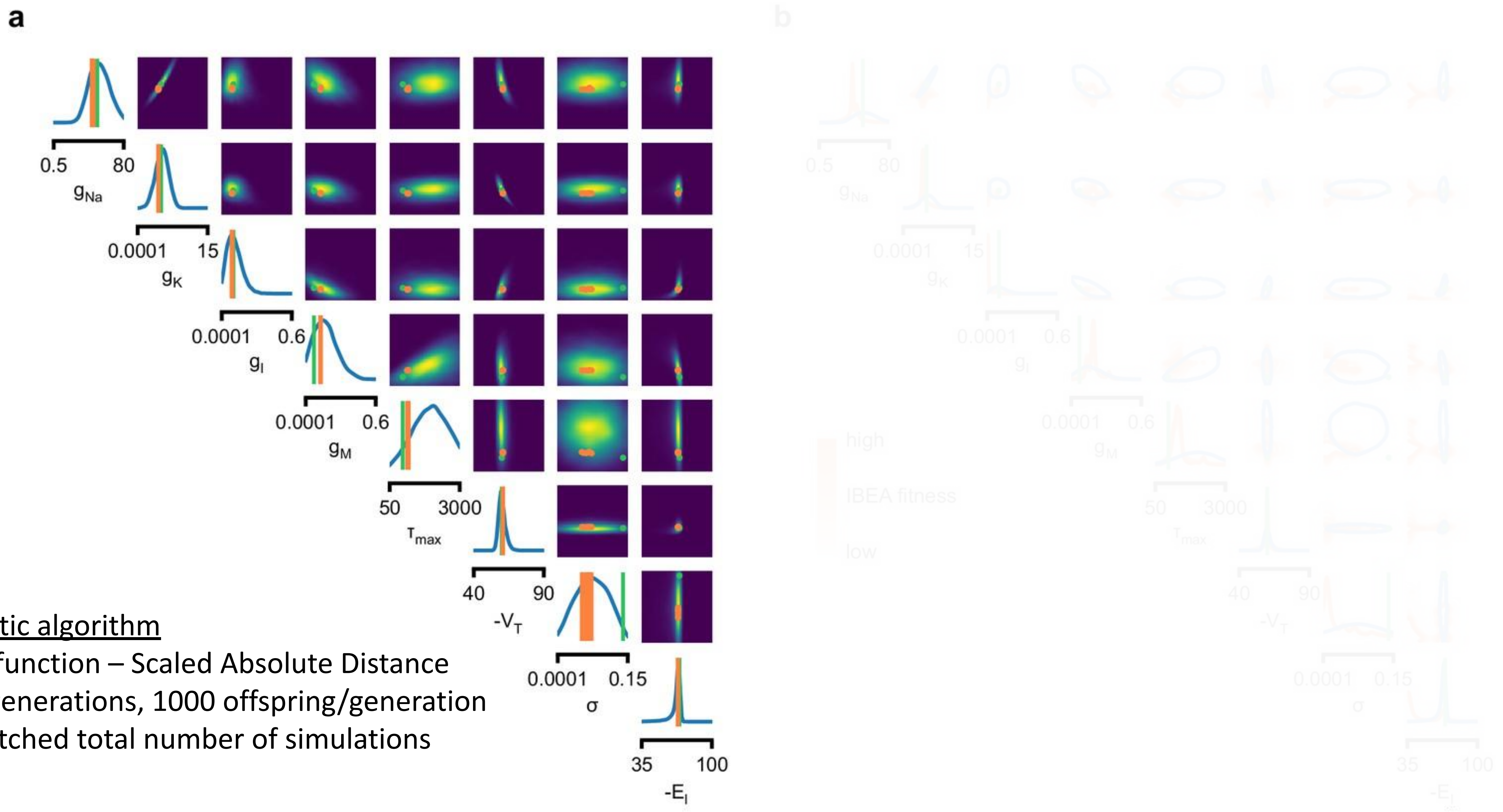
Results - Inferring parameters for Hodgkin-Huxley Model



-

sp : number of spikes,
rpot : mean resting potential,
srpot : standard deviation of the resting potential,
first four voltage moments: mean (m1), standard deviation (m2), skewness (m3) and kurtosis (m4)

Results - Comparison with genetic algorithm



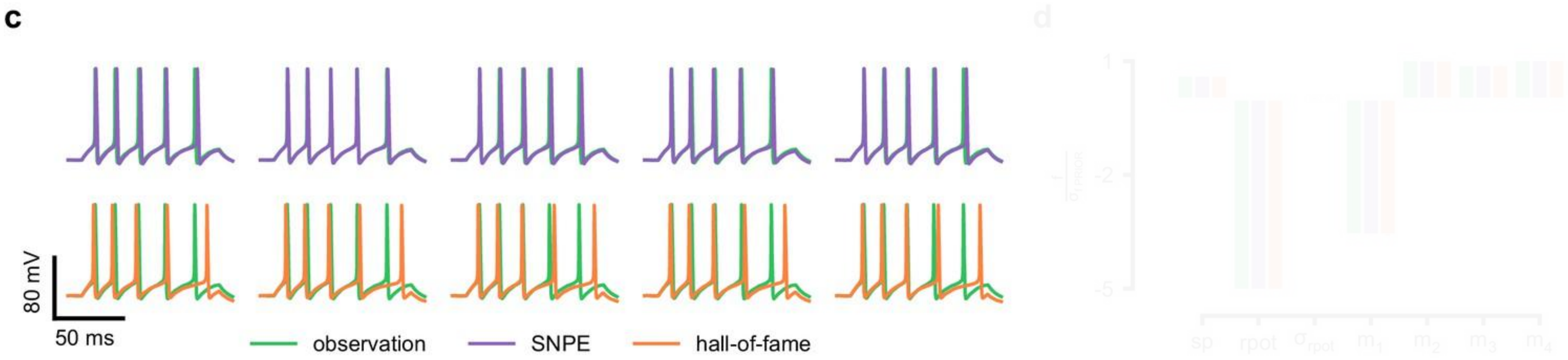
<u>Genetic algorithm</u>

Cost function – Scaled Absolute Distance

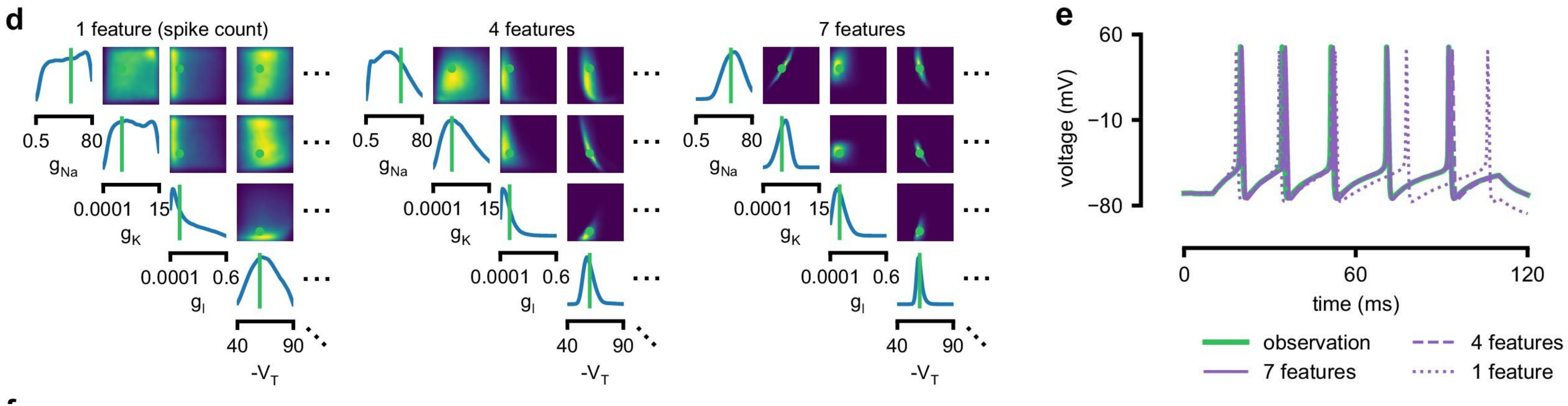
100 generations, 1000 offspring/generation

□Matched total number of simulations

Results - Comparison with genetic algorithm



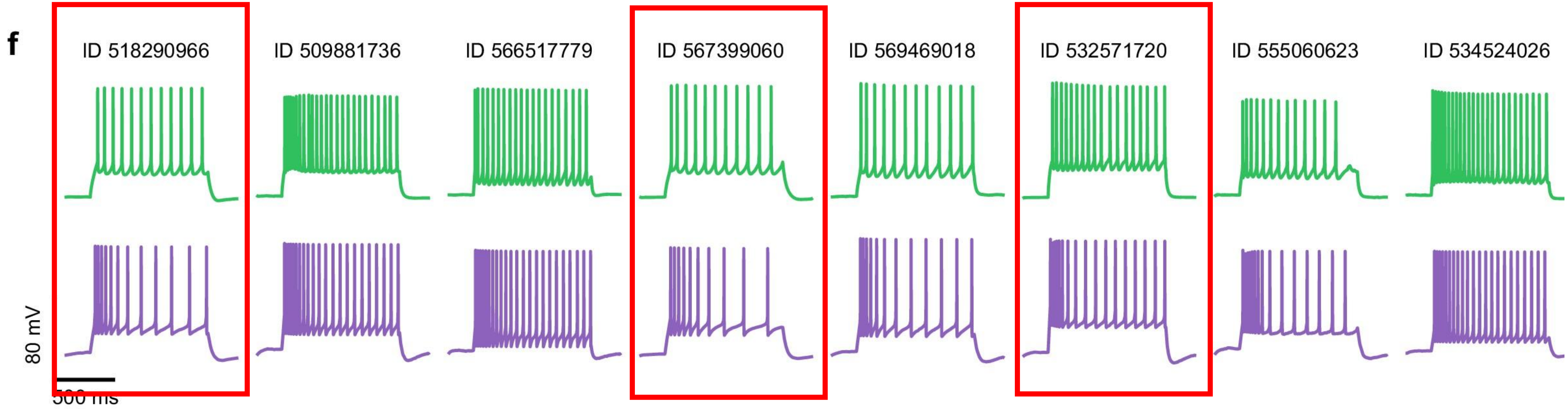
Results - Inferring parameters for Hodgkin-Huxley Model



The more features you use, the tighter the distribution becomes

Results - Inferring parameters for Hodgkin-Huxley Model

Observations from Allen Cell Types Database (green) and corresponding mode samples (purple)

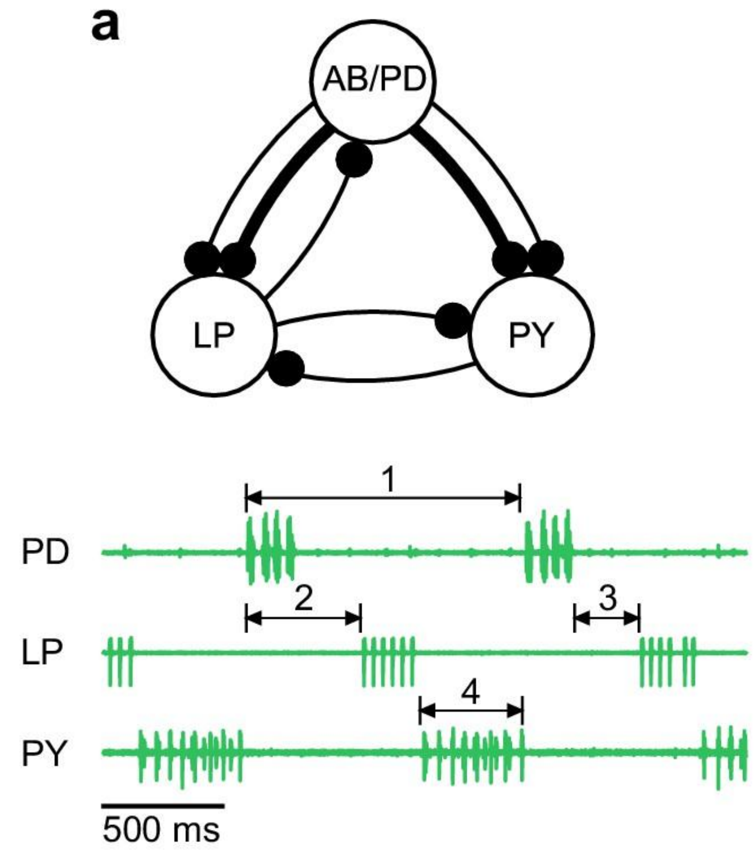


- Poor fits suggest amortized inference claim likely needs to be tempered
- The authors note that some parameters match better than others

Layout

- Motivation
- Caveats (what this talk will tell you, and what it won't)
- Background
  - Conditional Density Estimation
  - Mixing Neural Networks with Probabilistic Models
  - Related work (epsilon / likelihood-free methods)
    - ABC and SMB-ABC
- **Results**
  1. Recap of Methodology
  2. Comparison with established methodologies
  3. Inferring visual receptive field
  4. ~~Inferring ion channel parameters~~
  5. Inferring parameters for Hodgkin-Huxley Model
  6. **Inferring parameters for STG and analyzing posterior distribution**

Results - Inferring parameters for STG



1. cycle period
2. phase delays
3. phase gaps
4. burst durations

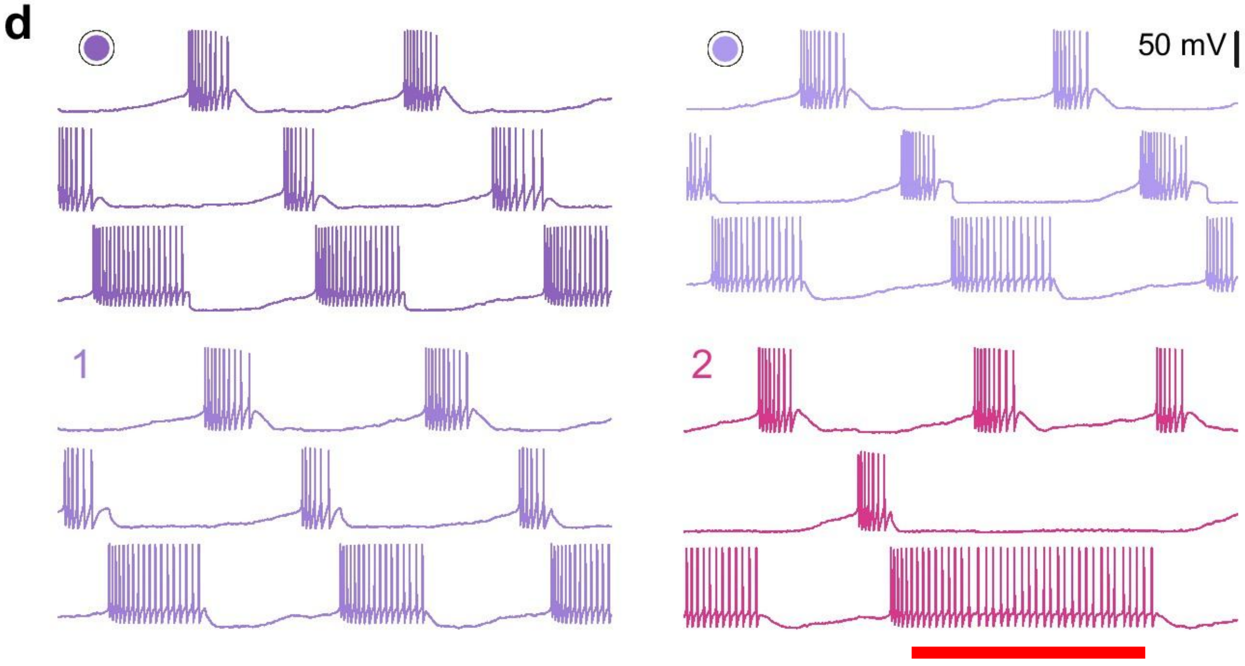
Claim - Possible due to having analytical/differentiable posterior

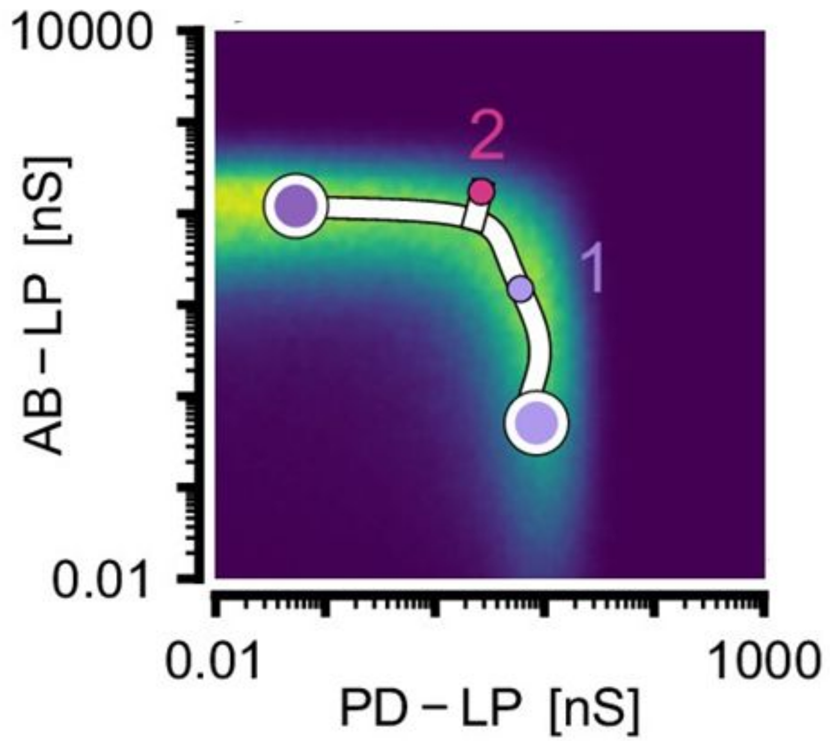
24 membrane parameters and 7 synaptic parameters

31 parameters in total

18 million samples used here vs. 2^31 (~2 billion) for grid with 2 values per parameter

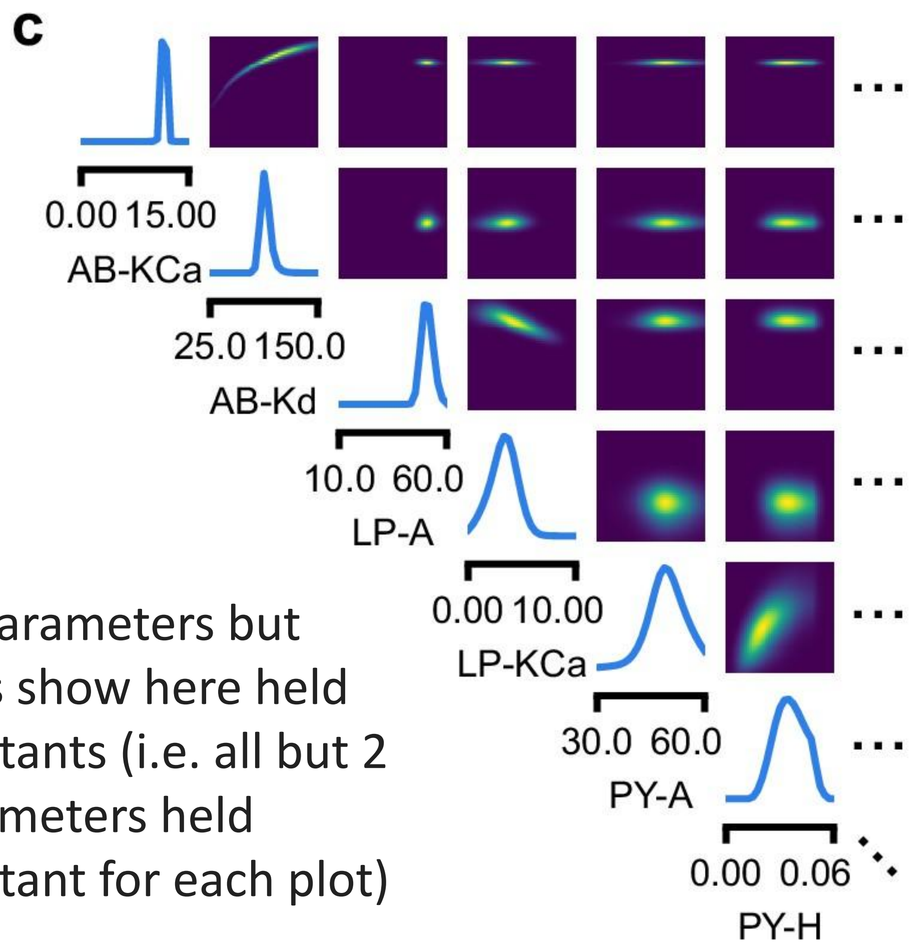
Results - Inferring parameters for STG





Results – Analysis of STG Parameter Posterior

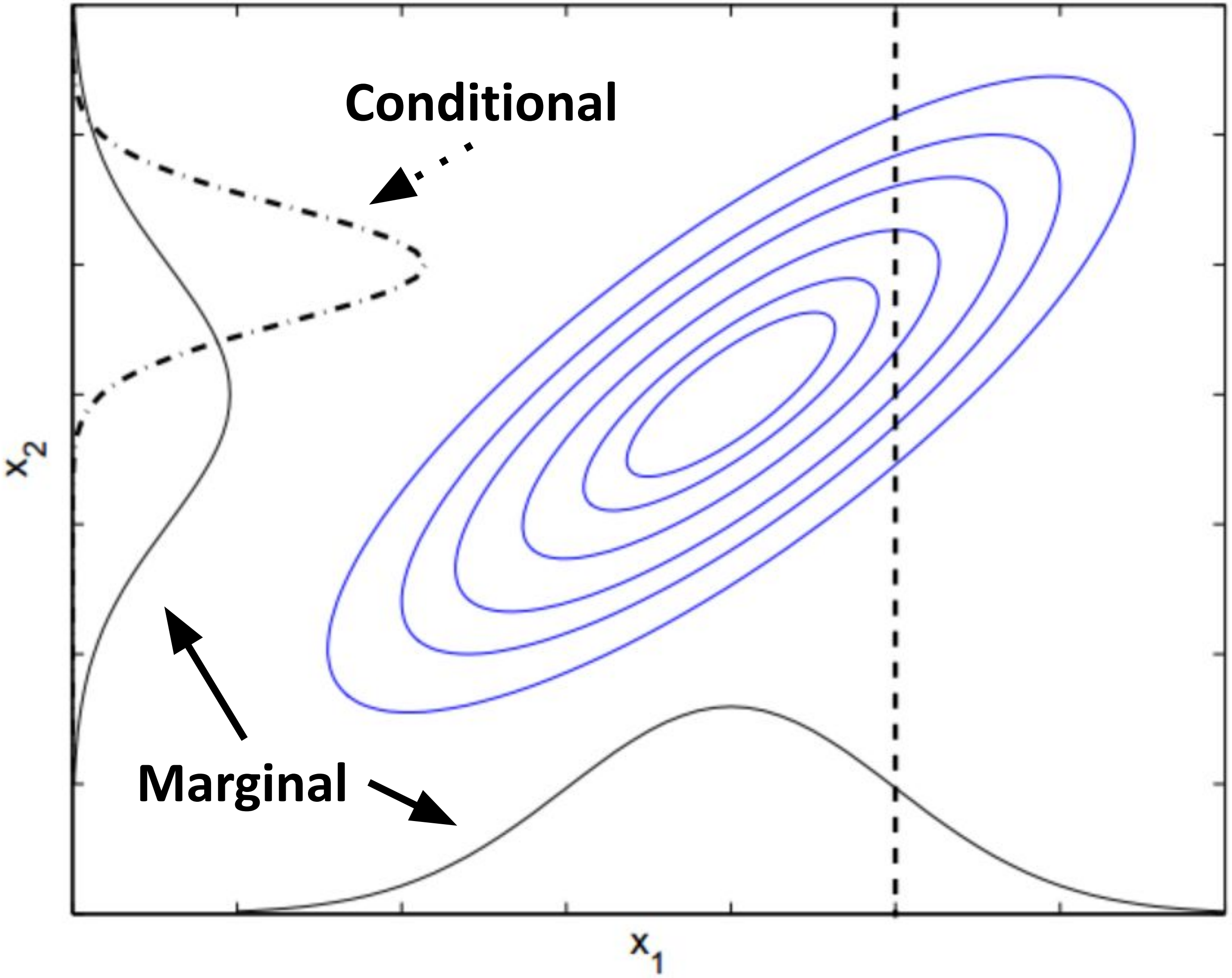
Strong correlations between parameters makes setting parameters individually problematic



All parameters but pairs show here held constants (i.e. all but 2 parameters held constant for each plot)

Conditional vs Marginal Densities

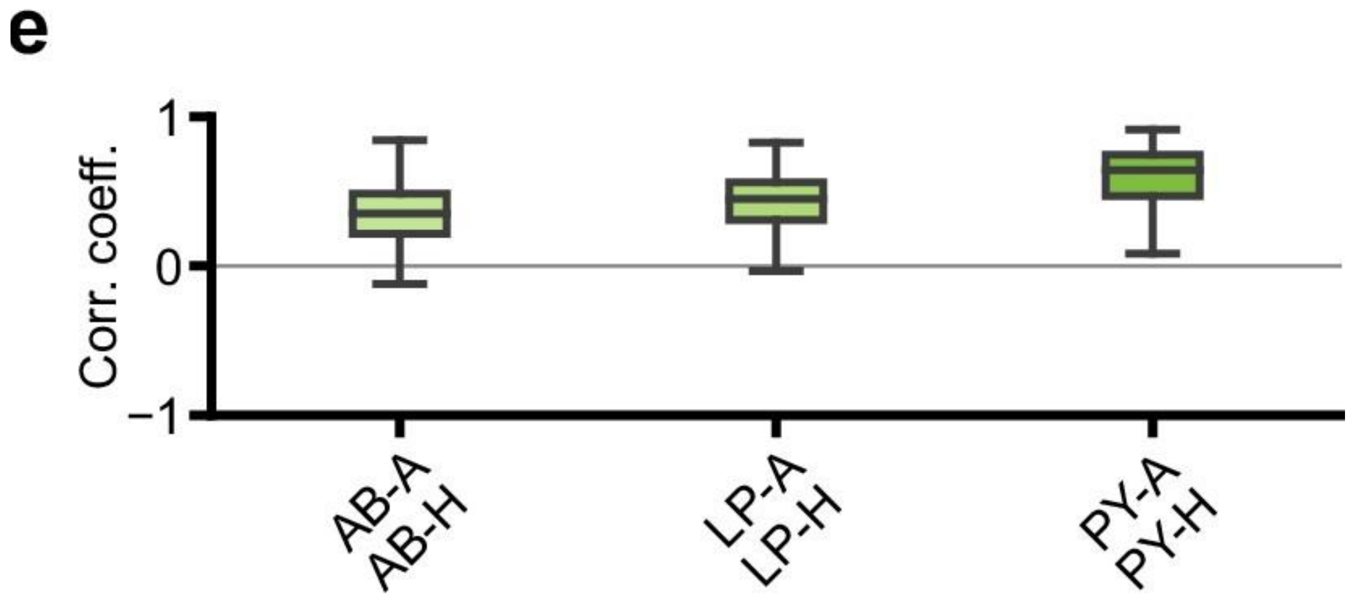
- Marginal
  - Summing over (or ignoring) other variables
- Conditional
  - Holding other variables at constant value



https://stackoverflow.com/questions/21867156/bivariate-normal-with-marginal-and-conditional-densities

Results – Analysis of STG Parameter Posterior

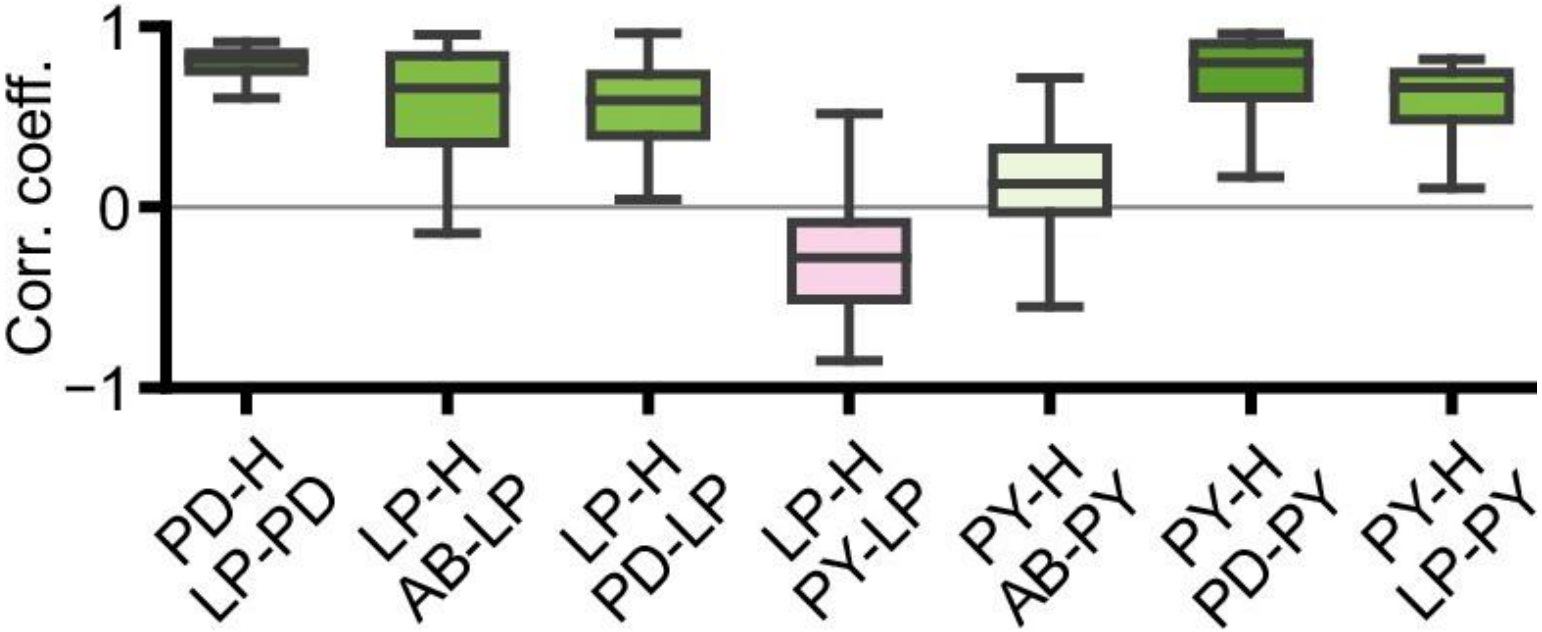
- For the PD and the LP neuron
- overexpression of the fast transient potassium current (IA) leads to a compensating increase of the hyperpolarization current (IH)
- (MacLean et al., 2003; MacLean et al., 2005).



Significantly different from zero and positive

Results – Analysis of STG Parameter Posterior

- Using the dynamic clamp,
- Diverse combinations of the synaptic input strength and the maximal conductance of IH lead to similar activity in the LP and the PD neuron
- (Grashow et al., 2010; Marder, 2011).



Some relationship

Conclusion

- Novel methodology to estimate posterior distribution of parameters for given data
- Appears to be more efficient than current standard methods
- Having full posterior allows exploring relationships between parameters

- Criticism:
  - Did not compare with more advanced ABC samplers (ABC-MCMC) which might be more efficient
- Advantage:
  - Amortization of inference (do not need to rerun inference for new sample once neural network has been trained)
  - Sequential inference (used, but not explored here) allows faster training
  - No need to define distance function or distance threshold
  - Uses all samples, unlike grid sampling

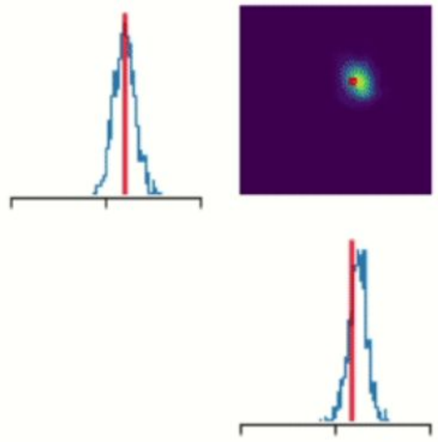
Simulation based inference in Python

sbi: simulation-based inference
https://www.mackelab.org/sbi/

```python
prior = BoxUniform(low=zeros(2), high=2*ones(2))   # Box prior [0,2]x[0,2]
def simulator(θ): return θ + 0.1*randn_like(θ)     # Gaussian in 2D
posterior = infer(simulator, prior, method='SNPE', num_simulations=500)
```

```
Running 500 simulations.: 100%|██████████| 500/500 [00:00<00:00, 57141.55it/s]
Neural network successfully converged after 109 epochs.
```

```python
samples = posterior.sample((1000,), x=observed)
pairplot(samples, points=ground_truth, **plot_style);
```



Simulation based inference in Python

ELFI - Engine for Likelihood-Free Inference
https://elfi.readthedocs.io/en/latest/

Currently implemented LFI methods:

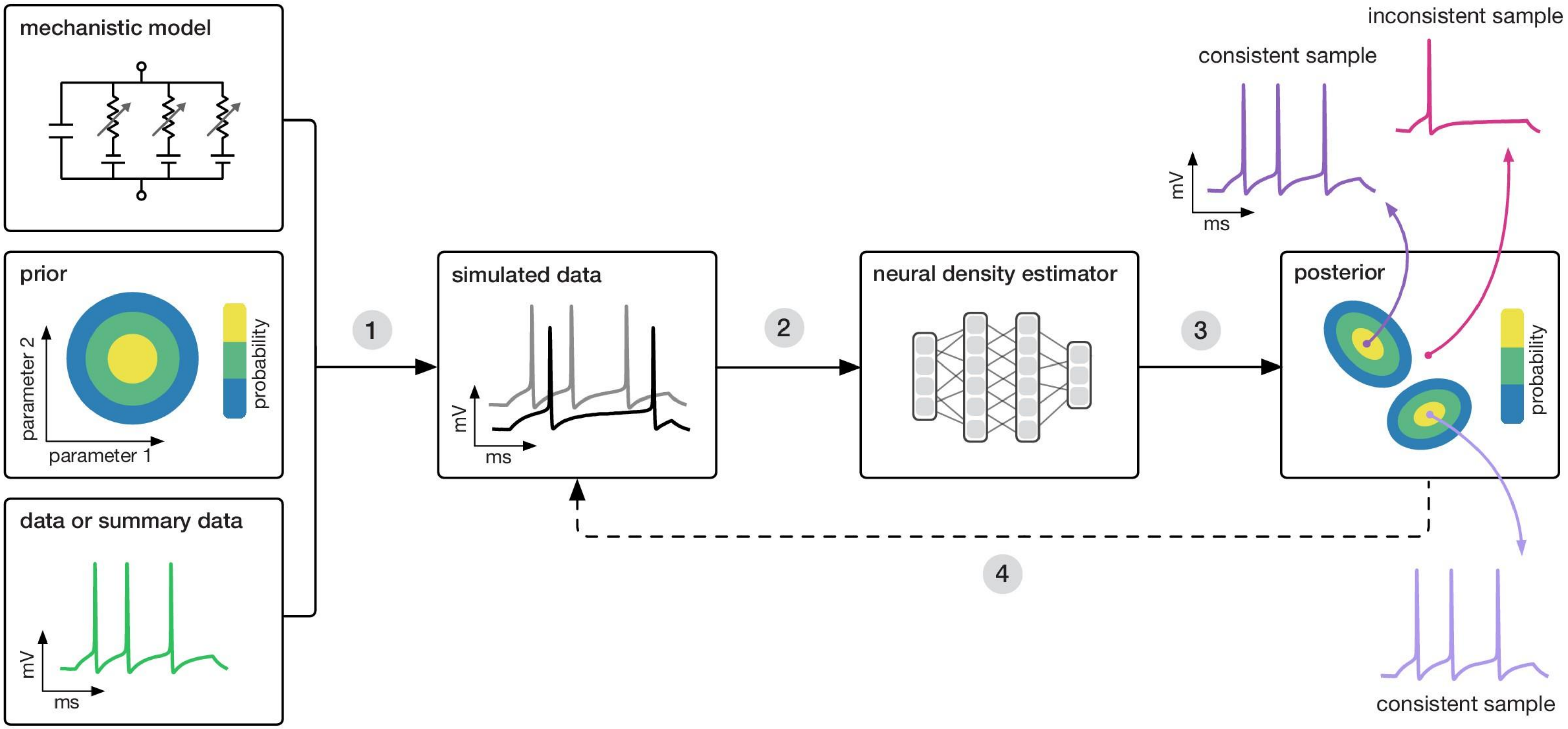
- ABC rejection sampler
- Sequential Monte Carlo ABC sampler
- ABC-SMC sampler with adaptive distance
- ABC-SMC sampler with adaptive threshold selection
- Bayesian Optimization for Likelihood-Free Inference (BOLFI) framework
- Robust Optimization Monte Carlo (ROMC) framework

ELFI also has the following non LFI methods:

- Bayesian Optimization
- No-U-Turn-Sampler, a Hamiltonian Monte Carlo MCMC sampler

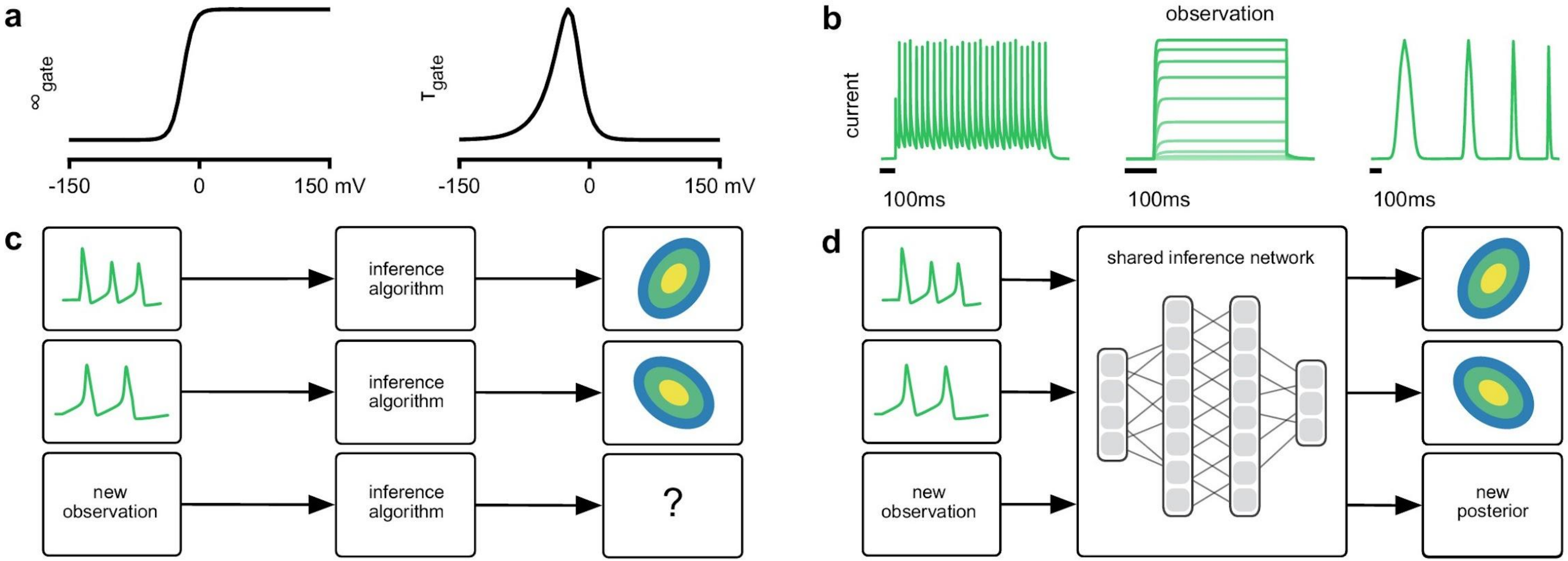
Additionally, ELFI integrates tools for visualization, model comparison, diagnostics and post-processing.

Questions?

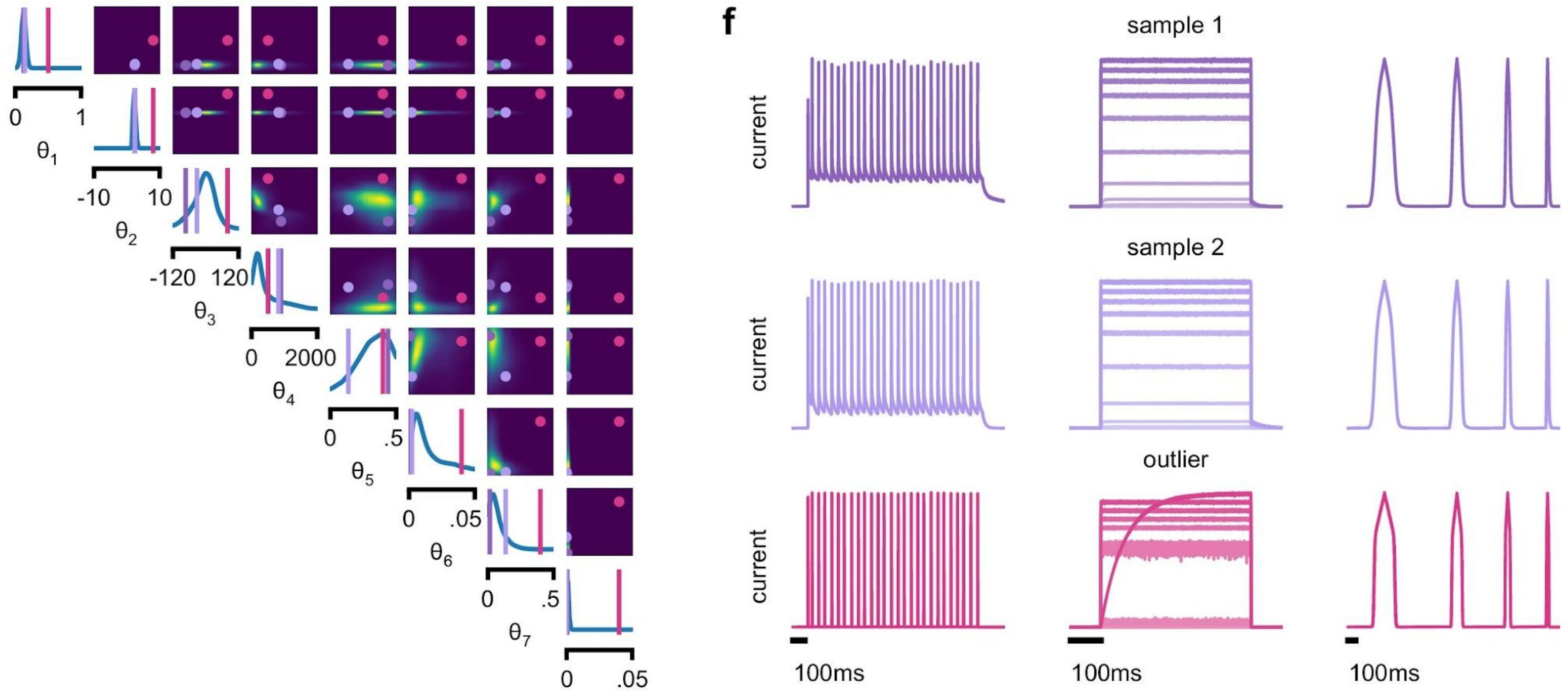


mechanistic model
prior
parameter 2
parameter 1
probability
data or summary data
mV
ms
1
simulated data
mV
ms
2
neural density estimator
3
posterior
probability
4
consistent sample
inconsistent sample
mV
ms
consistent sample

Results - Inferring ion channel parameters



Results - Inferring ion channel parameters



Results - Inferring ion channel parameters