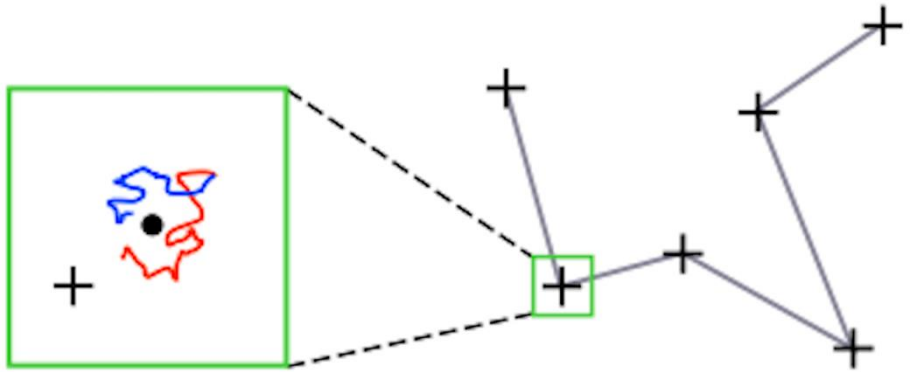
HAMLET-Physics, Copenhagen Aug 22, 2025.

Achieving subsampling time resolution in the analysis of two-state single molecule trajectories

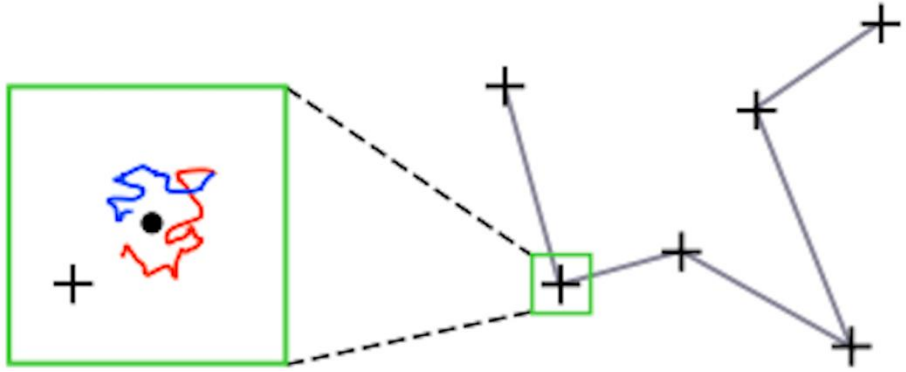


By Tobias Ambjörnsson

Computational Science for Health and Environment (COSHE), Lund University

Super-(spatial)-resolution microscopy is done(!?) – we reached nm scales.

Can we increase time-resolution beyond milliseconds?



Purpose:

Using computations (+ experiments) to quantify biomolecule dynamics at time scales faster than the sampling (exposure) times (milliseconds).

Erik Clarkson, Lund Uni.

Bayesian data analysis - a primer (for ANNers and LLMers)

"Machine learning when we do not have massive amounts of data".

$$P(\theta|O) = \frac{L(O|\theta)\pi(\theta)}{Z}$$

The Likelihood is our model

$$Z = \int L(O|\theta)\pi(\theta)\, d\theta$$

O – data
L – likelihood
θ - parameters
π – prior
Z – evidence
P - posterior

How to use:

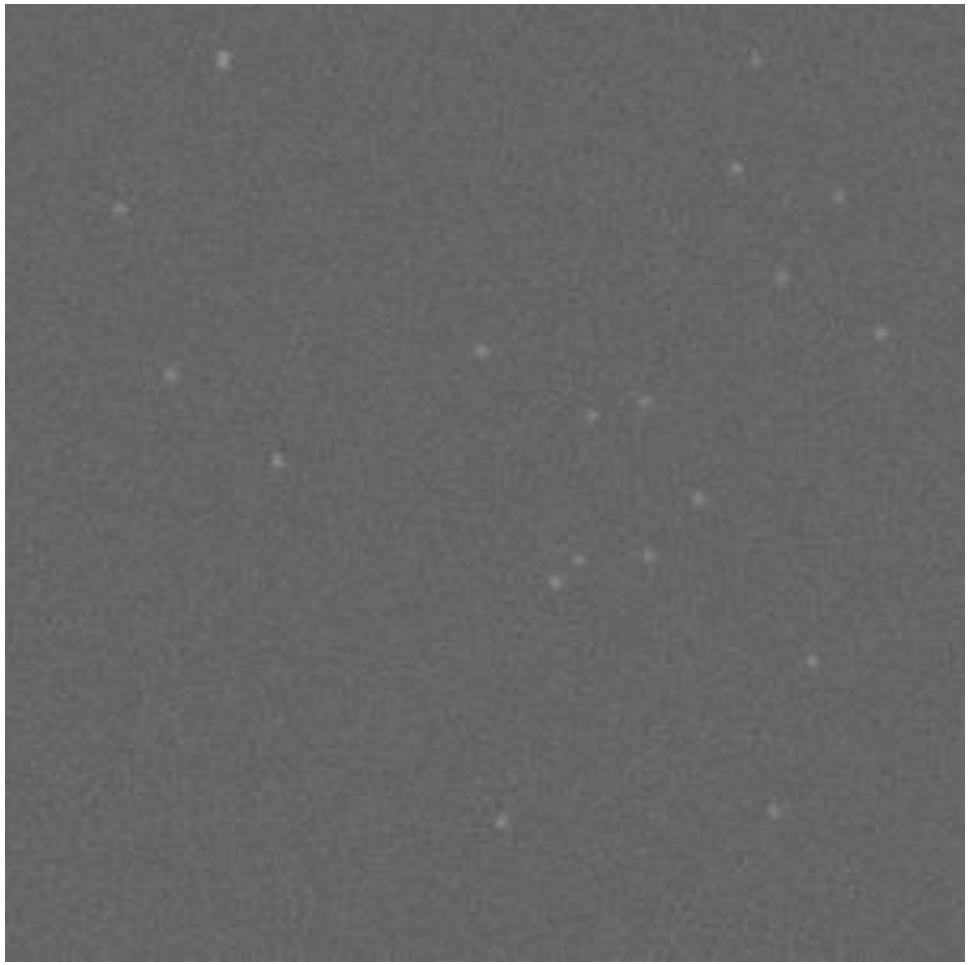
1. **Choose best model**: maximize Z [Occam's razor built-in, no overfitting).
2. **"Learning"**: Posterior gives model parameters ["loss function" = - log(P)]
3. **Generate new data:** Posterior predictives [given the "old" data, what is the probability for "new data"]

NOTE. Automatic "interpretability".

Reverend Thomas Bayes, 1740s
Pierre-Simon Laplace, 1774

The DATA – two-state diffusion single-molecule trajectories

- Wide-field fluorescence microscopy
- Fluorescent tags on T-cell receptors (TCRs)
- Two (hidden) states due to binding events of TCRs to pMHC molecules (“slow” and “fast” diffusion).



Experiments: Peter Jönsson’s group, Dept. of Chemistry, Lund University

Outline for the rest of the talk

Bayesian data analysis for:

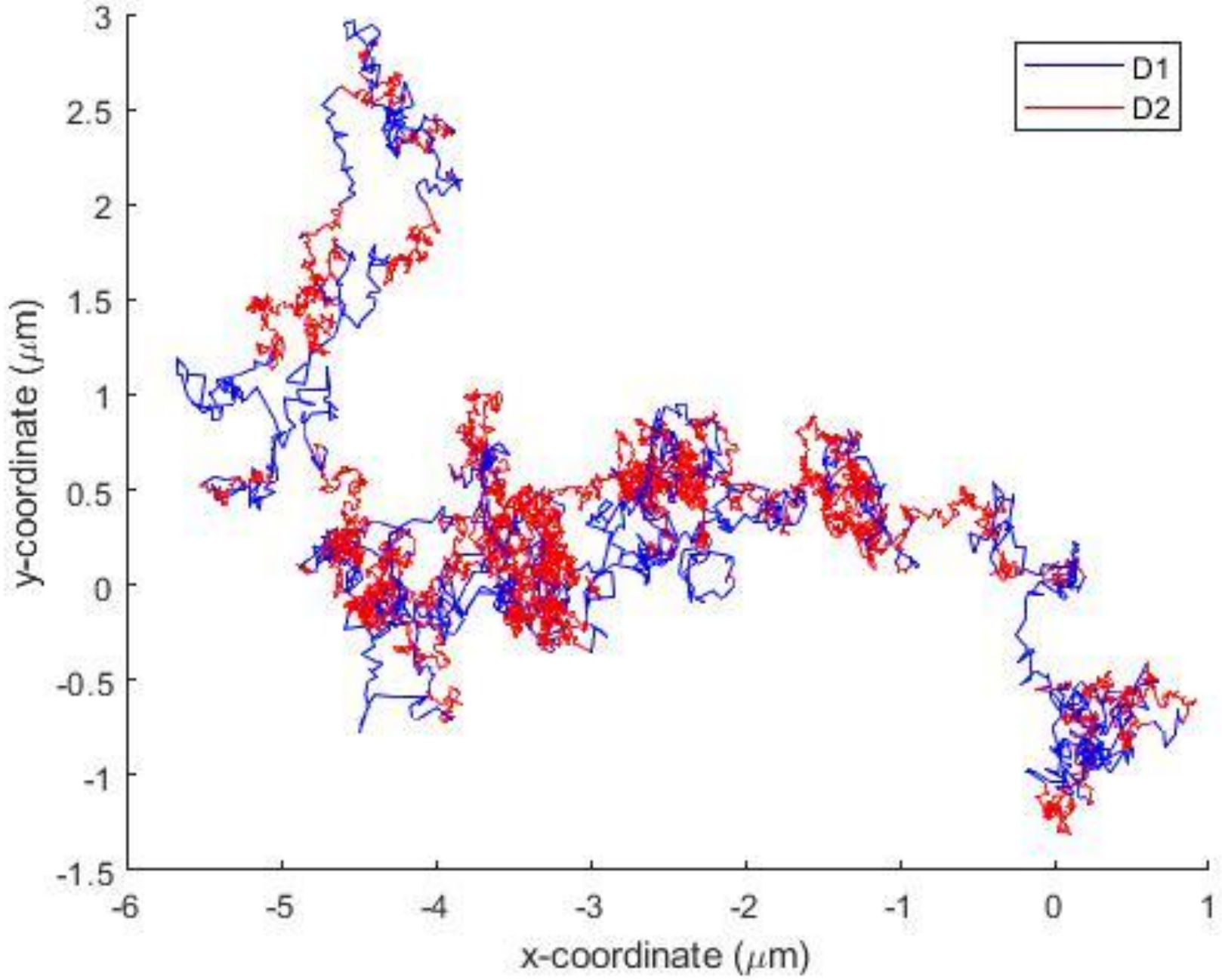
- Warmup: Two-state **slow switching** diffusion and Bayesian inference.
- Increasing time-resolution for two-state diffusion – **stHMM (subtime Hidden Markov Model).**

  [Discretely sampled (over a time window) continuous-time processes]

Two-state diffusion with **slow switching**

The DATA (synthetic)

NOTE: colors (ground truth state) are not seen in experiments



Now, Bayes!

The model (likelihood)

Displacements

Transition probabilities

Data/Observations $O = \Delta_1, \dots, \Delta_N$

Model parameters $\theta = D_1, D_2, p_{12}, p_{21}$

Hidden ("latent") states $s_1, \dots, s_N$

Sampling time

L = likelihood

$$P(O|\theta) = \sum_{s_0,\dots,s_N} P(O|s_1, \dots, s_N, \theta)\, P(s_1, \dots, s_N)$$

$$\sigma(s)^2 = 2\tau D(s)$$

Hidden Markov Model (HMM)

$$= \sum_{s_0,\dots,s_N} \prod_{j=1}^{N} \frac{1}{2\pi\sigma(s_j)^2} \exp\left(-\frac{\Delta_j^2}{2\sigma(s_j)^2}\right) P(s_1)P(s_2|s_1)\cdots P(s_{N-1}|s_{N-2})P(s_N|s_{N-1})$$

EMISSION PROBABILITIES = Path probability (fixed state sequence)

Likelihood computation in practice

- Forward algorithm avoids "2^N-problem"

$$\alpha_n(i) = \sum_{k=1}^{2} \alpha_{n-1}(k) P(s_n = i \,|\, s_{n-1} = k) P(\Delta_n | s_n = i)$$

$$L(O|\theta) = \sum_{i=1}^{2} \alpha_N(i)$$

R. Das et al. "A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton". PLoS Comput Biol 5.11 (2009).
E. Clarkson and T.A., Bayesian and frequentist analyses of two-state single-molecule diffusion trajectories, J. Phys. A 58 (2025)

Sampling model parameters – Nested sampling

- Evolve a number of ‘live points’
- Compress parameter space around high-likelihood regions
- Output multivariate posterior, $P(\theta|O)$, and evidence, Z.
- Project posterior distribution onto one parameter axis

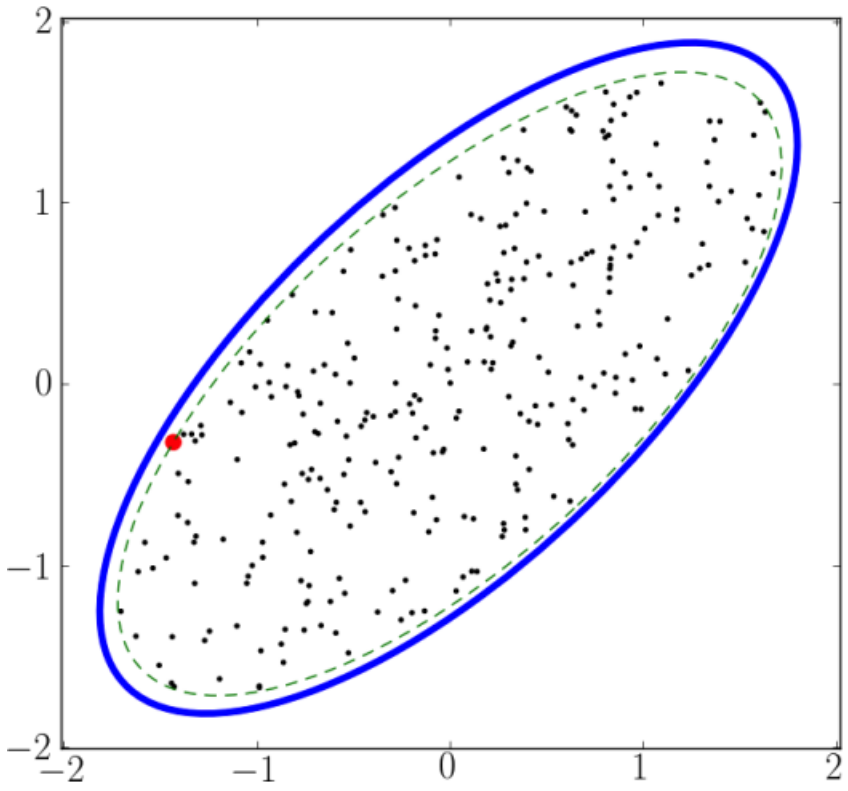
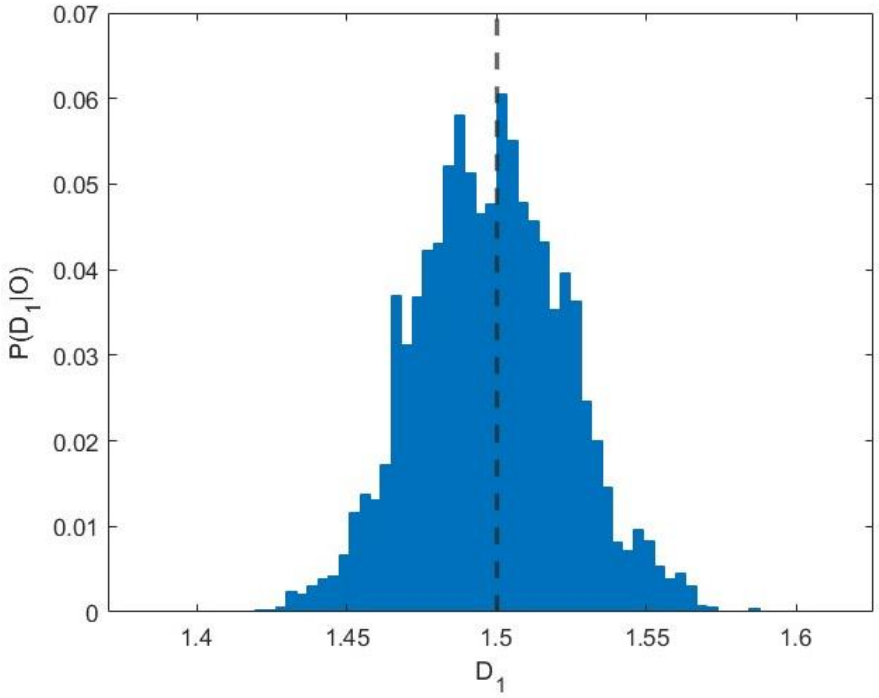
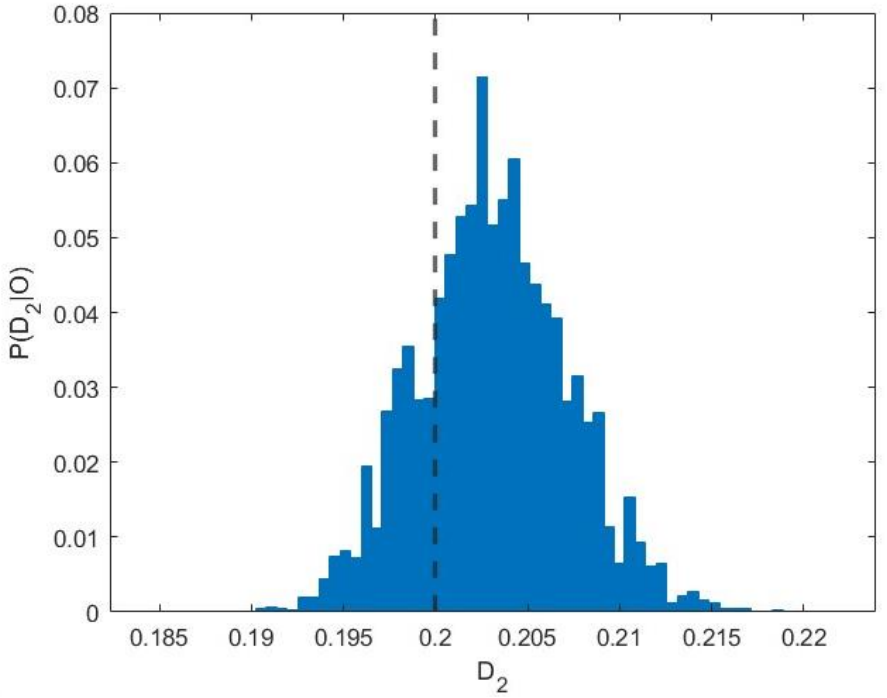
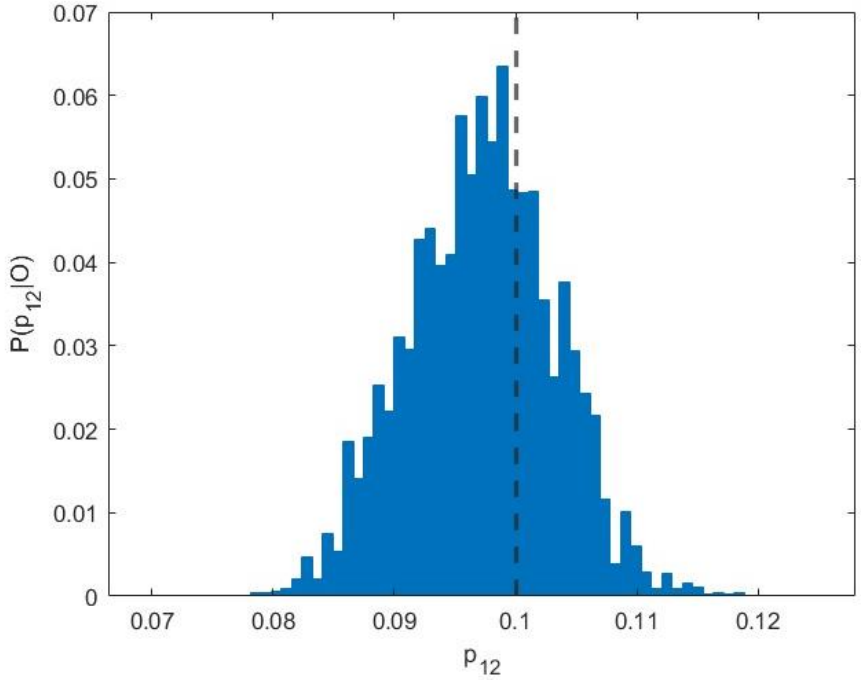


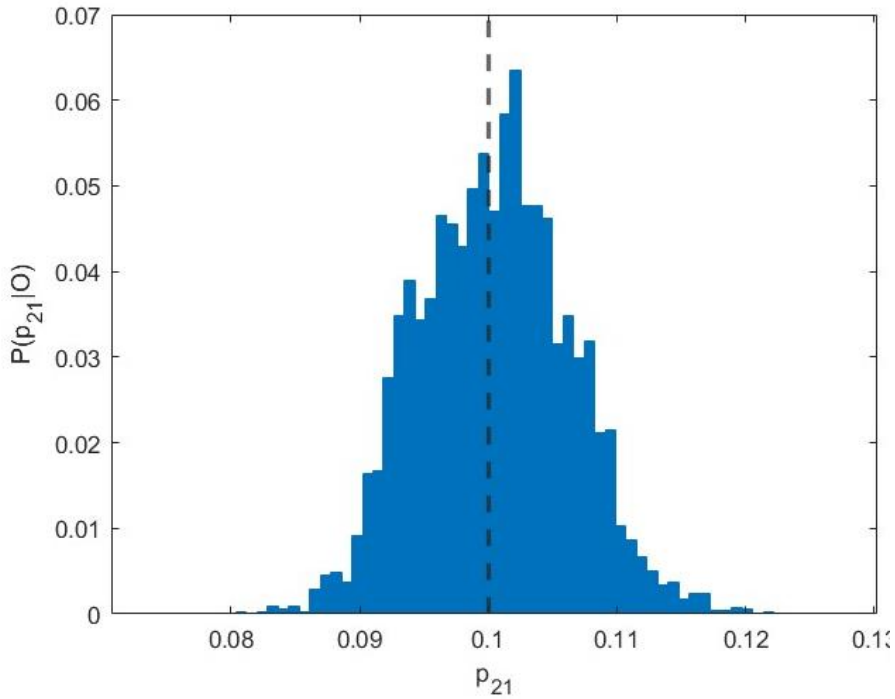
Figure 4. 2D-slice of parameter space with live points.

Results


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
P(D1|O)
1.4
1.45
1.5
1.55
1.6
D1


0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
P(D2|O)
0.185
0.19
0.195
0.2
0.205
0.21
0.215
0.22
D2


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
P(p12|O)
0.07
0.08
0.09
0.1
0.11
0.12
p12

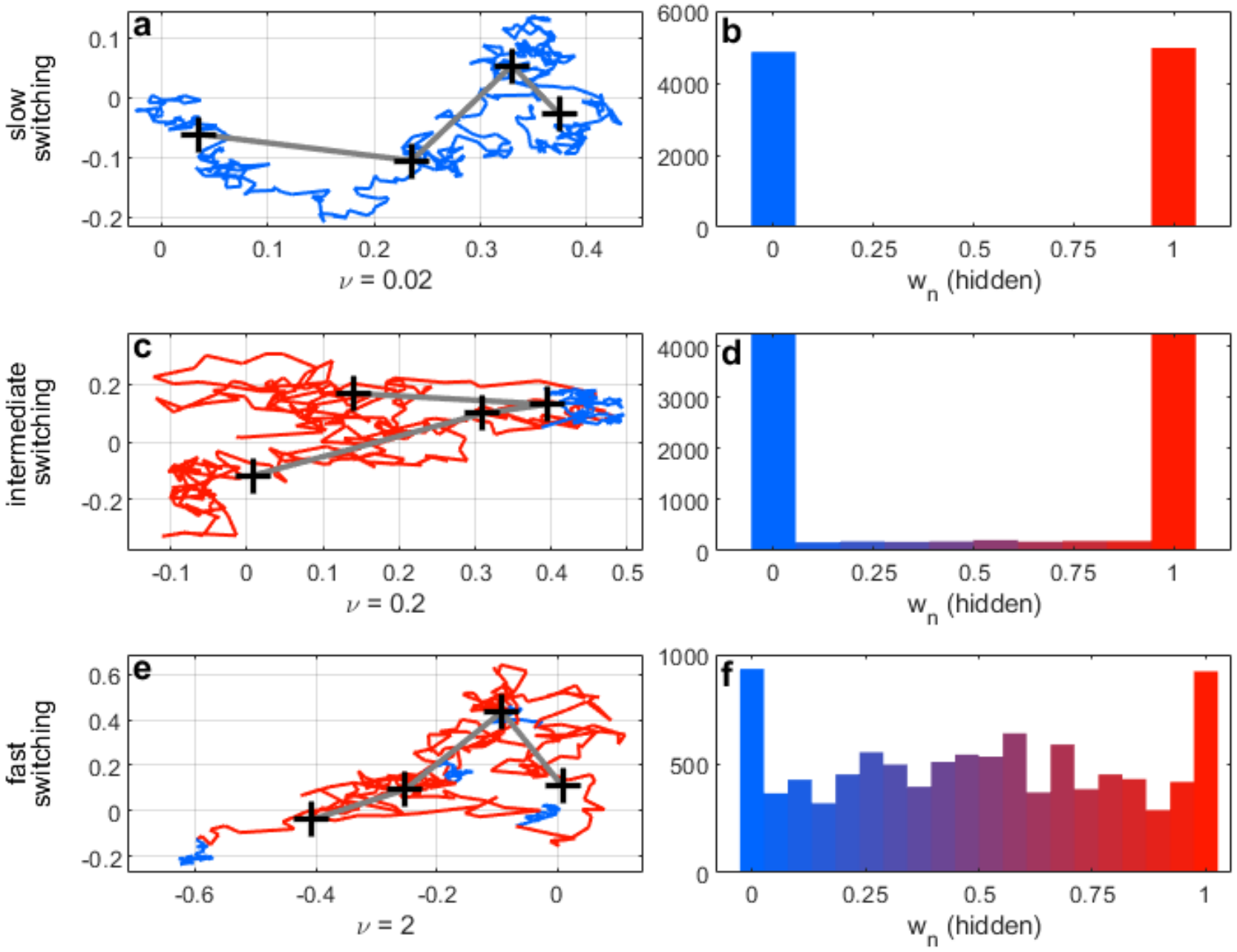

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
P(p21|O)
0.08
0.09
0.1
0.11
0.12
p21

How to increase time resolution using computations?

subtime-HMM (stHMM)

Applications: Discretely sampled continuous-time processes.

The DATA



NOTE: colors (ground truth state) are not seen in experiments

w_n = fraction of time spent in state 1 during time interval n.

Now, Bayes!

Emission probabilities

depends only on w_n

Emission probabilities have mean = 0 and variance:

$$\sigma(w_n)^2 = 2\tau D(w_n) = 2\tau(w_n D_1 + (1 - w_n)D_2)$$

w_n = fraction of time spent in state 1 during time interval n.

Likelihood for continuous time

stHMM likelihood

Integrate out unobserved (hidden/latent) transitions within each time interval

$$P(\Delta_1, \ldots, \Delta_N | \theta) =$$

$$\sum_{s_0=1}^{2} \cdots \sum_{s_N=1}^{2} \int_0^1 \mathrm{d}w_1 \cdots \int_0^1 \mathrm{d}w_N$$

$$P(\Delta_1 | w_1, \theta) P(s_0, \theta) P(w_1, s_1 | s_0, \theta) \cdots$$

$$P(\Delta_N | w_N, \theta) P(w_N, s_N | s_{N-1}, \theta)$$

Emission probability

Transition-accretion probability

Forward algorithm can be extended. The one-variable integral is computed numerically.

Many weeks later …

- Moment generating function
- Fixed start and end states
- Four cases need to be separately dealt with

Example:

$$P(w_n, s_n = 1 | s_{n-1} = 2, \theta)$$

$$= \tau k_{21} e^{-\tau(k_{12} w_n + k_{21}(1-w_n))}$$

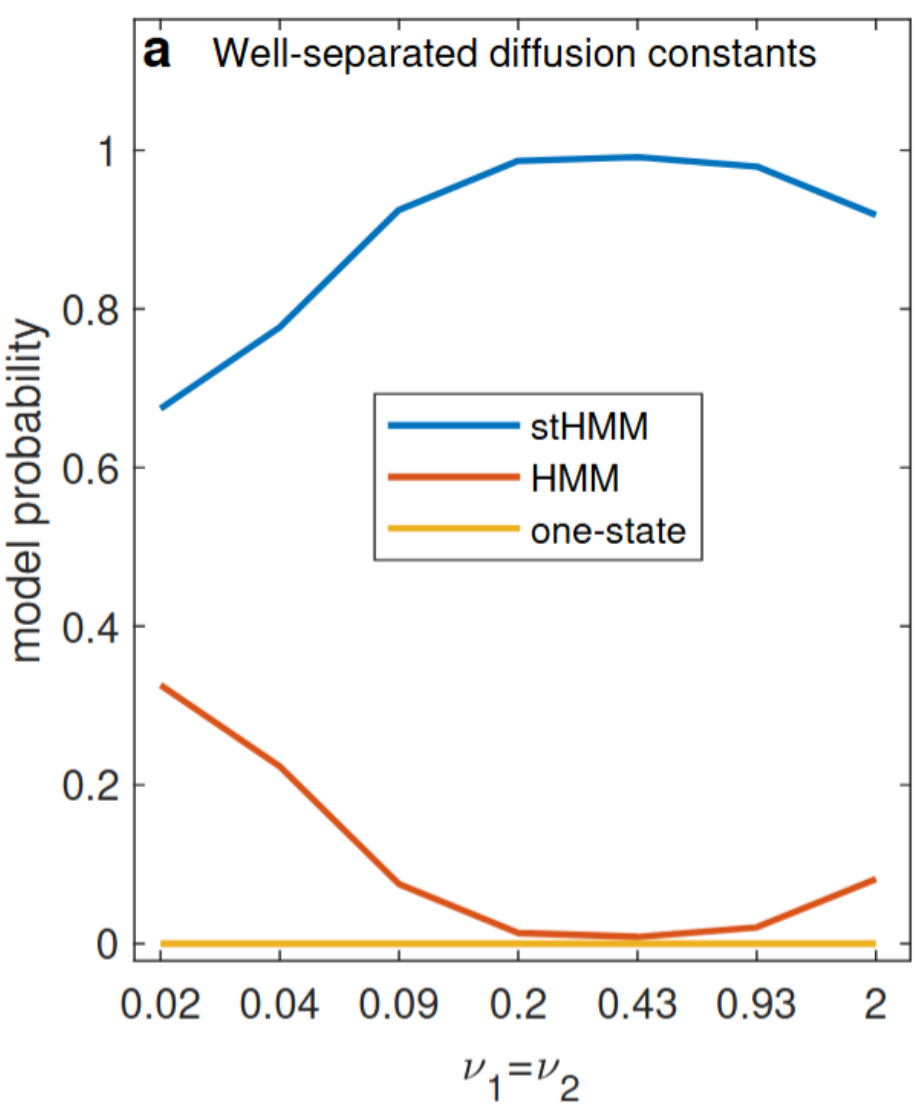
$$I_0\left(2\tau\sqrt{k_{12}k_{21}w_n(1-w_n)}\right),$$

2 → 1;
#switches > 0

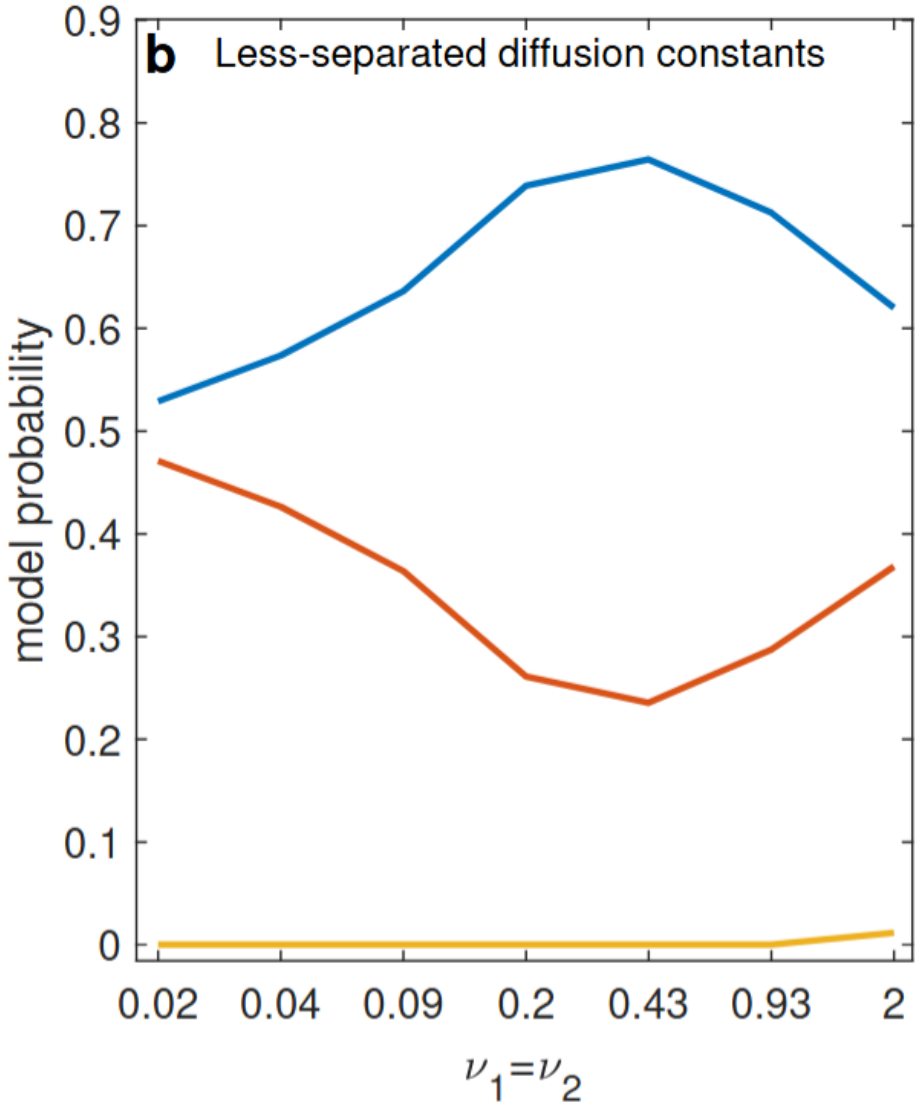
Modified Bessel function

Results, Model selection

ν = k τ = average number of switches per sampling time



D_1/D_2 =10



D_1/D_2 =3

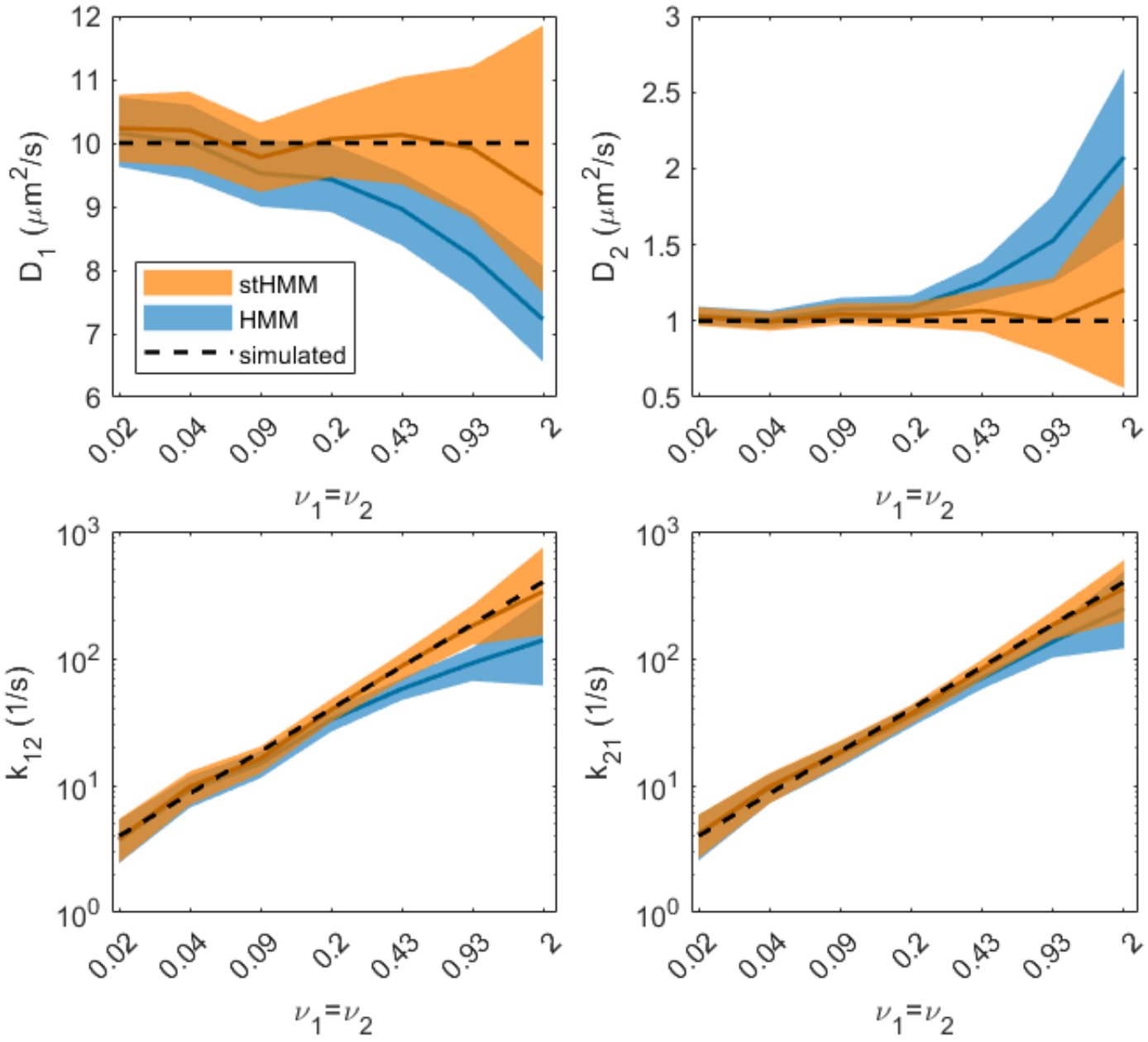
Results, Parameter estimation

Current state of the art:

HMM - neglects subtime events.

stHMM helps to solve the "photon budget dilemma"

E. Clarkson and T. Ambjörnsson, Achieving subsampling time resolution in the analysis of two-state single molecule trajectories, submitted



$\nu = k\,\tau$

Summary

- Bayesian data analysis for model selection and parameter estimation from single-molecule trajectories.
- Through stHMM we reach sub-sampling-time resolution. Main novelty: transition-accretion probabilities.
- stHMM outperforms previous state-of-the-art method (HMM).
- General framework for discretely sampled continuous time processes(?).

Outlook

- Extend to N states instead of 2. Use Bayesian model selection to determine optimal value of N.
- Apply to actual experimental data ...

Funding: Swedish Research Council, 2023-2027.