Computing transition-path statistics from short-trajectory data

Aaron Dinner,

John Strahan, Adam Antoszewski, Justin Finkel,

Spencer Guo, Chatipat Lorpaiboon, Erik Thiede, Bodhi Vani

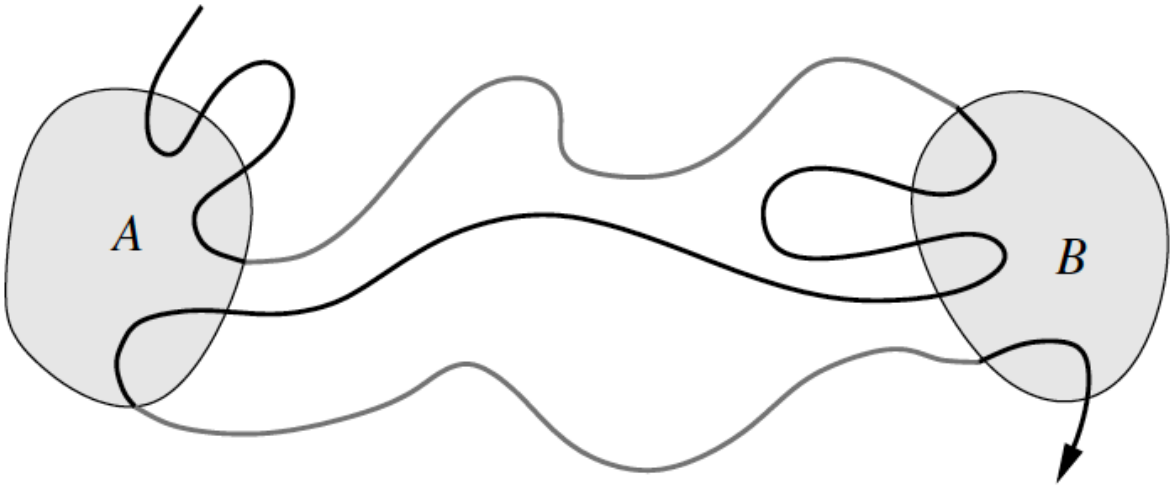
University of Chicago

Jonathan Weare

New York University

This work was supported by NIH and NSF.

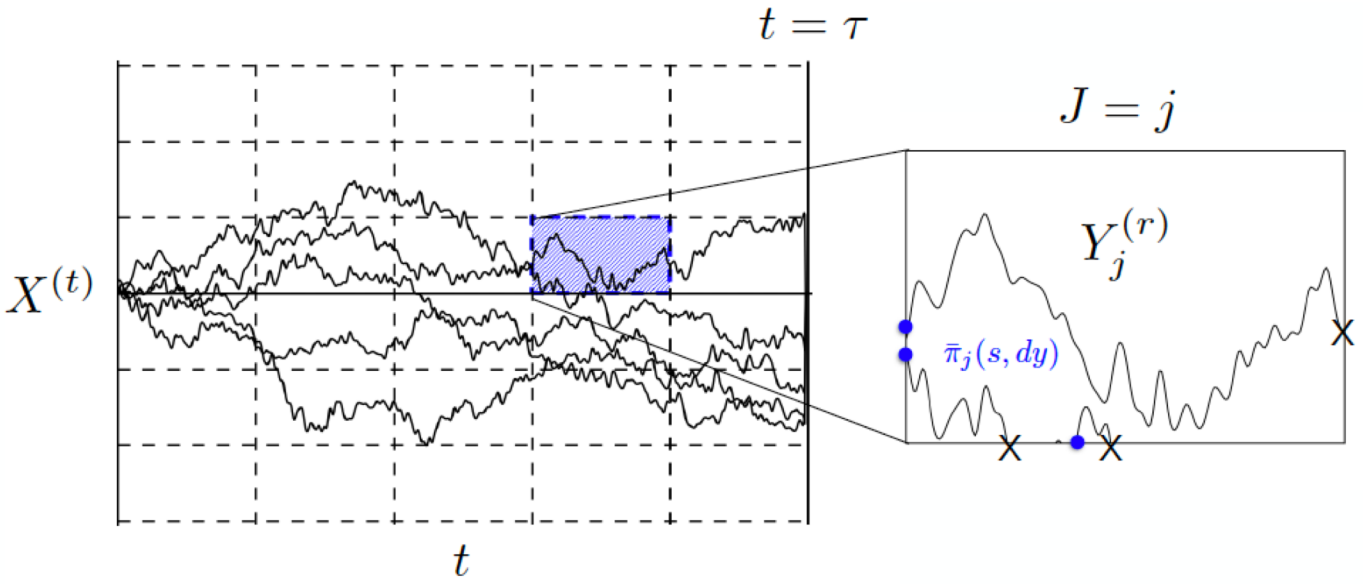
For reactions between two states (*A* and *B*), we want to compute committors, mean first passage times, reactive currents, rates, and potentials of mean force.



This is challenging by direct simulation.

picture from Vanden-Eijnden (2006)

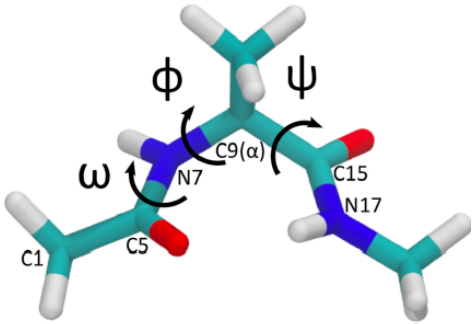
One solution is to use enhanced sampling.



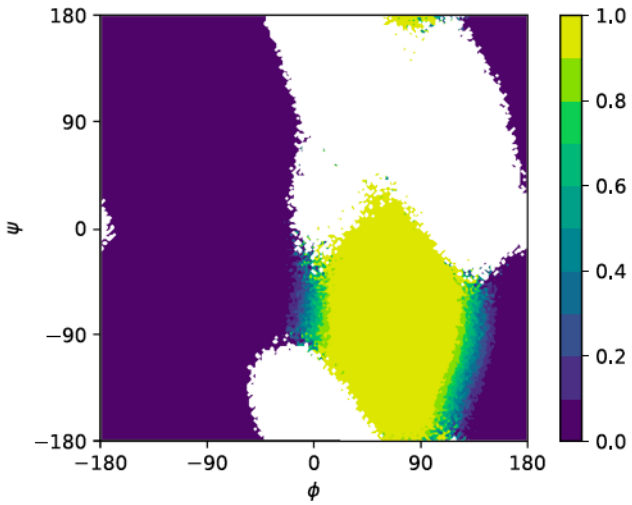
See for example Vani, Weare, and Dinner, J Chem Phys (2022), where we split the ensemble of paths based on their origins and use trajectory stratification (nonequilibrium umbrella sampling) to improve statistics.

This builds on work from our group going back to 2007, and Vanden-Eijnden and Venturoli (2009); see also exact milestoning from Elber.

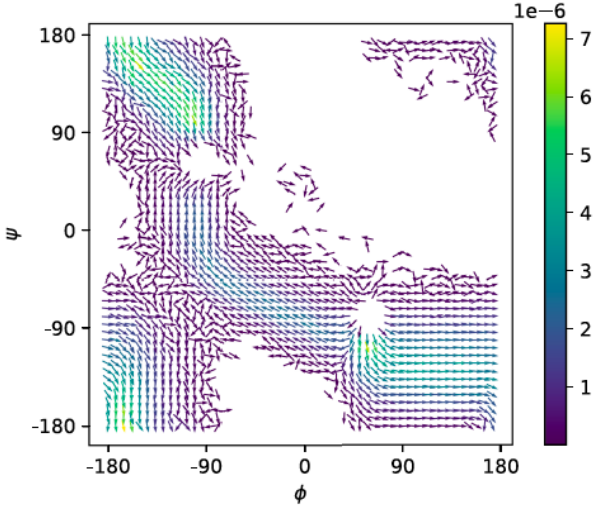
One solution is to use enhanced sampling.



committor



reactive currents



See for example Vani, Weare, and Dinner, J Chem Phys (2022), where we split the ensemble of paths based on their origins and use trajectory stratification (nonequilibrium umbrella sampling) to improve statistics.

This builds on work from our group going back to 2007, and Vanden-Eijnden and Venturoli (2009); see also exact milestoning from Elber.

In this talk, we will instead make a Markov assumption and solve operator equations of the form

$$(\mathcal{S}^\tau - \mathcal{I})u(x) = -\mathbb{E}_x\left[\int_0^{\tau\wedge T} h(X_s)ds\right] \text{ for } x \in D = (A \cup B)^c$$

$$u(x) = g(x) \text{ for } x \notin D$$

where $\mathcal{S}^\tau$ is the "stopped" transition operator:

$$\mathcal{S}^\tau f(x) = \mathbb{E}_x\left[f(X_{\tau\wedge T})\right] = \mathbb{E}\left[f(X_{\tau\wedge T})|X_0 = x\right]$$

for $T = \inf\{t > 0 : X_t \in A \cup B\}$.

For example, the committor:

$$q_+(x) = \mathbb{P}_x[X_T \in B] = \mathbb{E}_x[\mathbb{1}_B(X_T)]$$

satisfies

$$\mathcal{S}^\tau q_+(x) = q_+(x) \text{ for } x \in D$$

$$q_+(x) = \mathbb{1}_B \text{ for} x \notin D$$

The lead time (mean first passage time conditioned on entering B before A):

$$m_{AB}(x) = \frac{\mathbb{E}_x[T\mathbb{1}_B(X_T)]}{\mathbb{E}_x[\mathbb{1}_B(X_T)]} \tag{1}$$

satisfies

$$\mathcal{S}^\tau[m_{AB}q](x) = -\mathbb{E}_x\left[\int_0^{\tau\wedge T} q(X_s)ds\right] \tag{2}$$

$$m_{AB} = 0 \text{ for } x \in A \cup B.$$

For the special case of Langevin dynamics

$$\dot{X}(t) = b(X(t)) + \sqrt{2}\sigma(X(t))\eta(t),$$

the Itô formula of stochastic calculus leads to an explicit form for the infinitesimal generator of the transition operator:

$$\mathcal{L}f(x) = \sum_{i=1}^{n} b_i(x)\frac{\partial f}{\partial x_i} + \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

However, in general, we do not know the form of $\mathcal{S}^\tau$ and $\mathcal{L}$.

This talk is about how we can still solve equations of $\mathcal{S}^\tau$ using partial observations sampled at finite times.

Outline

- A Galerkin approach
  - Applications to protein dynamics
- A neural-network approach
  - Numerical experiments illustrating key features
  - Application to a model of sudden stratospheric warming
- Addressing the two-trajectory requirement

We want to solve equations of the form

$$(\mathcal{S}^{\tau} - \mathcal{I})u(x) = -\mathbb{E}_x \left[\int_0^{\tau \wedge T} h(X_s) ds \right] \text{ for } x \in D = (A \cup B)^c$$

$$u(x) = g(x) \text{ for } x \notin D.$$

Expand $u(x)$ in terms of basis functions $\phi_j(x)$:

$$u(x) \approx \psi(x) + \sum_{j=1}^{n} v_j \phi_j(x),$$

where $\psi(x)$ is a guess that homogenizes the boundary conditions.

By standard means, we can convert the operator equation into a linear algebra problem:

$$(C^\tau - C^0)v = r^\tau,$$

where

$$C_{ij}^s = \langle \phi_i, \mathcal{S}^s \phi_j \rangle_\mu \quad \text{for} \quad s = 0, \tau$$

$$r_i^\tau = \langle \phi_i, \psi(x) - \mathcal{S}^\tau \psi(x) \rangle_\mu ,$$

and the inner product is over an *arbitrary* initial density $\mu(x)$:

$$\langle \alpha(x), \beta(x) \rangle = \int \alpha(x)\beta(x)\mu(x)dx.$$

In practice, cover the space with initial points and run (short) simulations.

The fact that μ is arbitrary enables us to choose the initial points so as to maximize the sampling efficiency.

Compute

$$\overline{\langle \phi_i, \mathcal{S}^\tau \phi_j \rangle} = \frac{1}{N} \sum_{n=1}^{N} \phi_i(X_0^{(n)}) \phi_j(X_{\tau \wedge T}^{(n)})$$

and similarly for the inner products involving ψ.

Solve $(C^\tau - C^0)v = r^\tau$ by standard linear algebra methods, and use v to construct u.

Benchmark system: Trp-cage folding

Fast-folding miniprotein extensively studied both experimentally and computationally (Zhou, Berne, Roitberg, Hagen, Bohuis, Garcia, Laio, Piana, Shaw, Dror, Mukamel, Wang, Dickson, Brooks, Zuckerman, Levy, Kevrekidis, Debenedetti, Ferguson, others).

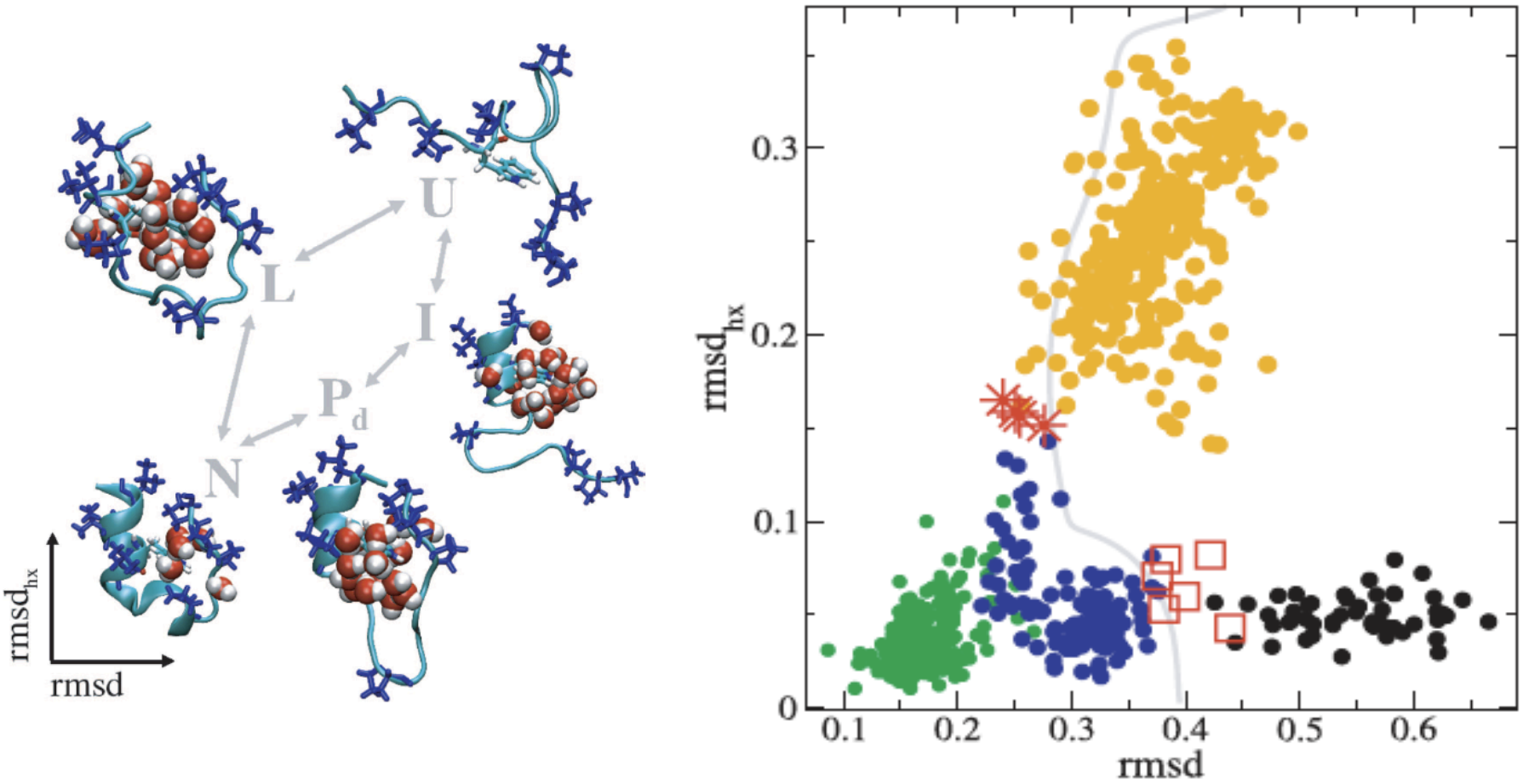
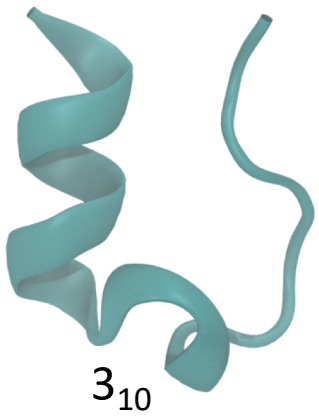
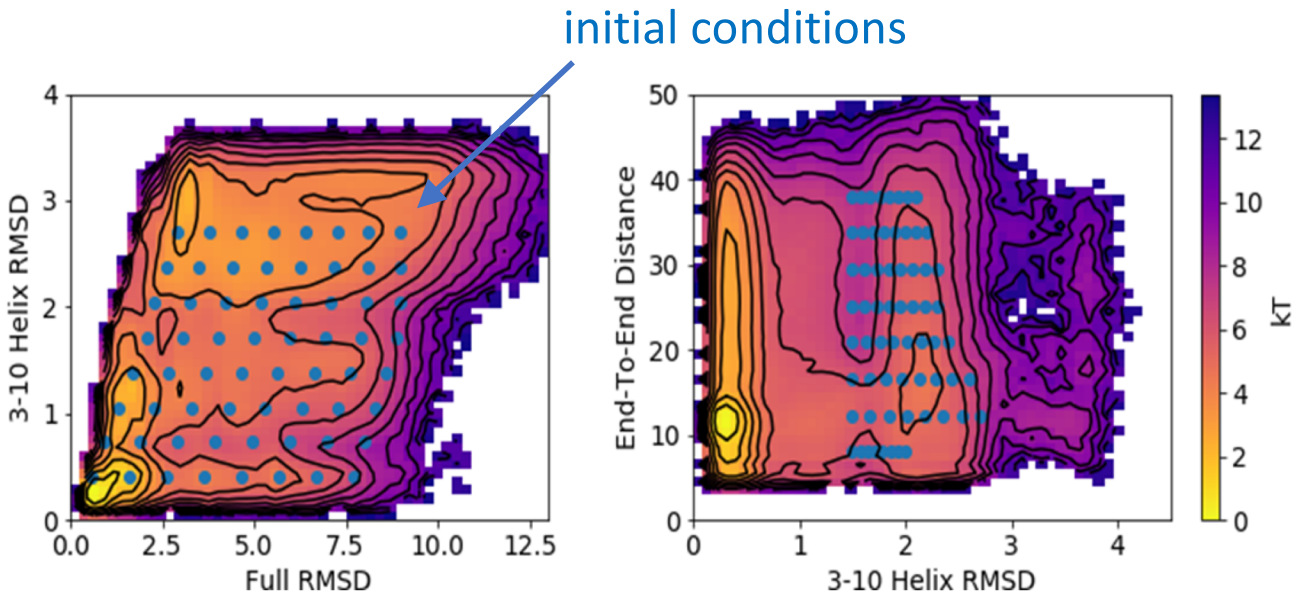


Figure from Juraszek and Bolhuis, PNAS 103, 15859-15864 (2006).

Generating the data set





- Represent system with CHARMM36m force field;
- Distribute initial points in RMSD and 3-10 helix RMSD space;
- Launch 14 independent 30 ns trajectories from each of the points.
- Based on comparison with umbrella sampling, add 2 more 30 ns trajectories from points in α helix RMSD and end-to-end distance space;

- 1024 trajectories for a total of 30 μs

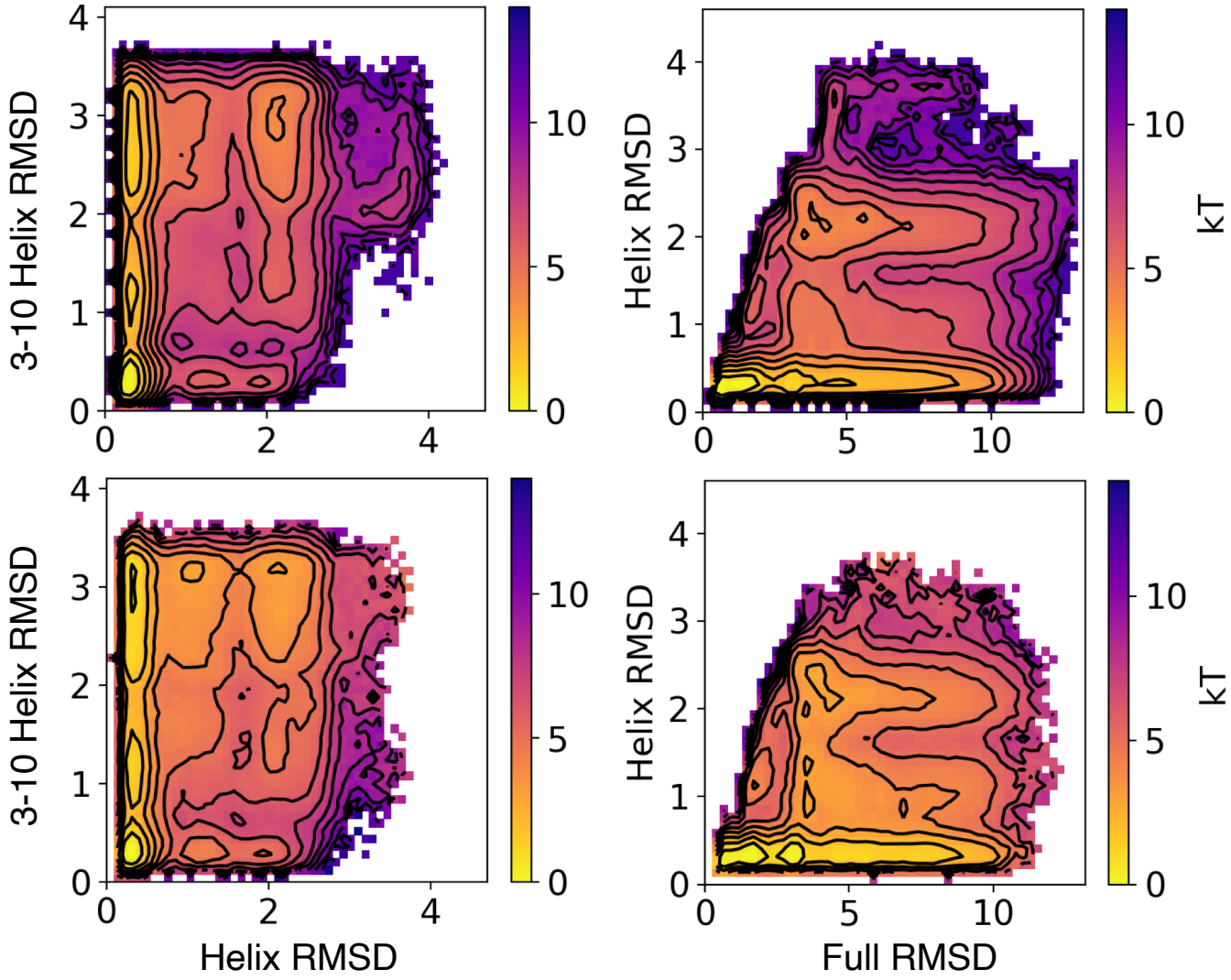
Strahan, Antoszewski, Lorpaiboon, Vani, Weare, Dinner (2021) JCTC 17, 2948-2963

Our Galerkin method gives good agreement with independent simulations.

Dynamical Galerkin Approximation (DGA)

$(T^\tau)^\dagger w = w$

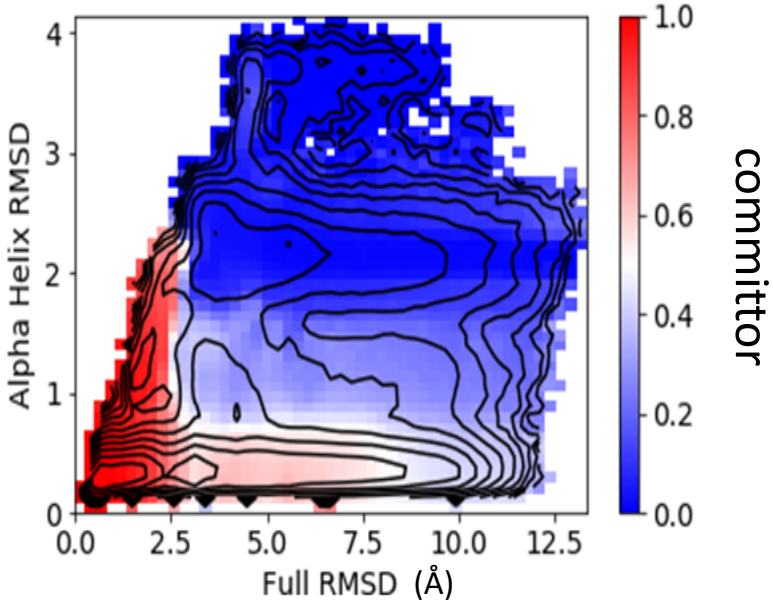
Replica Exchange Umbrella Sampling

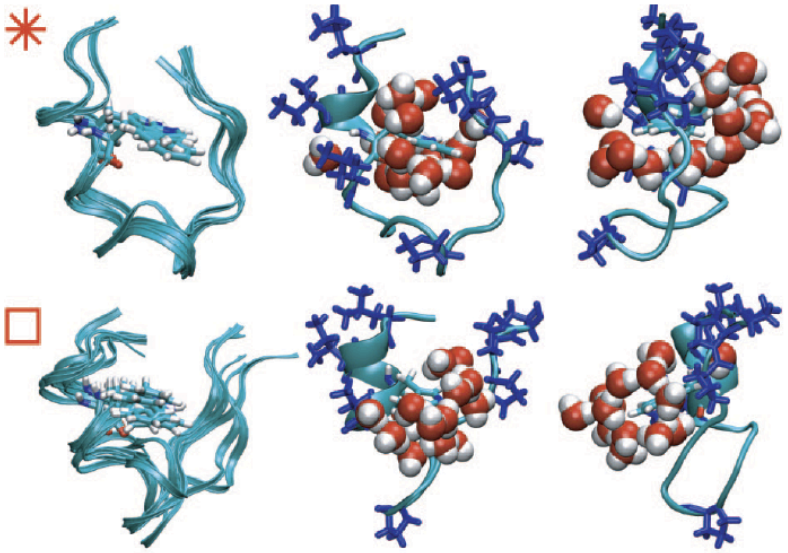


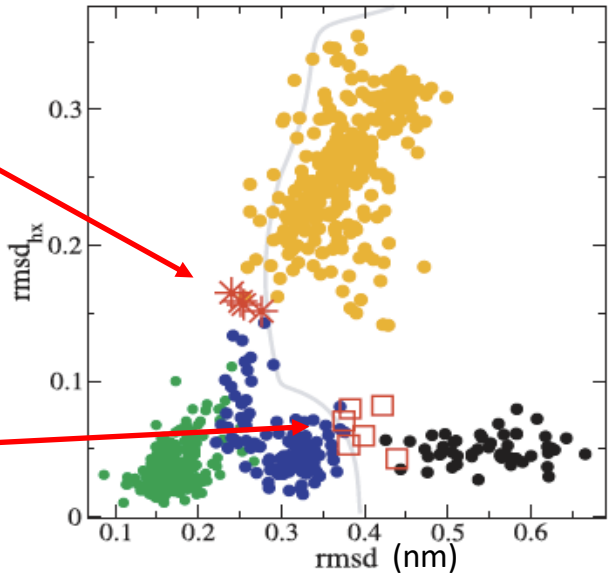
Comparison of committors to transition states identified by Juraszek and Bolhuis, PNAS 103, 15859-15864 (2006)

We compute committors
for all structures (>6M) in 153D
and then for visualization project to 2D.

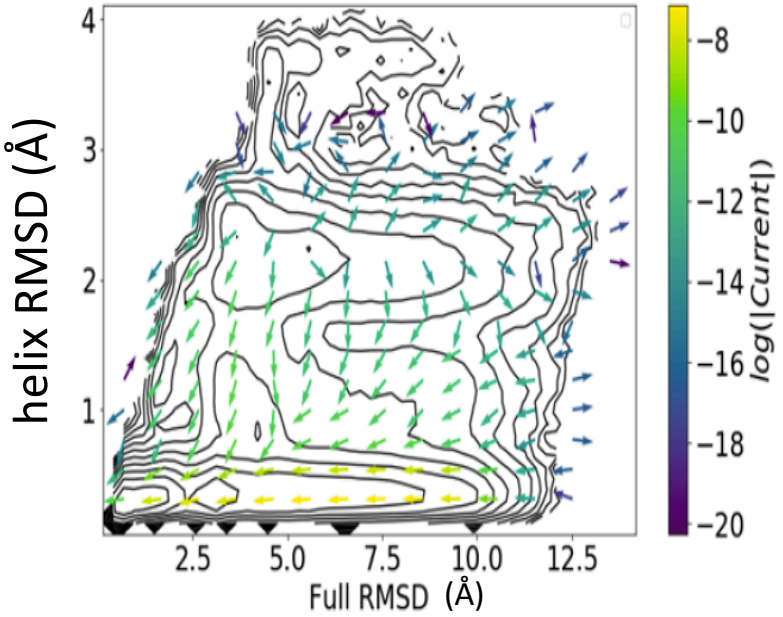
Contours show PMF with 1 k_BT spacing.







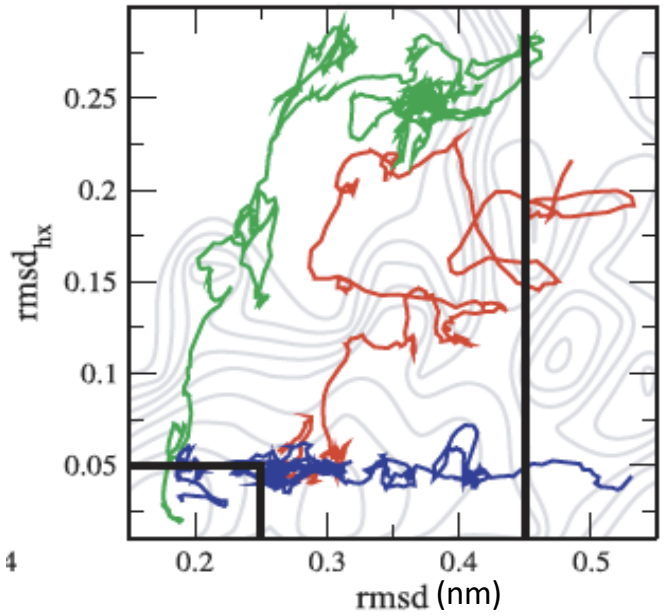
Reactive currents for trp-cage folding

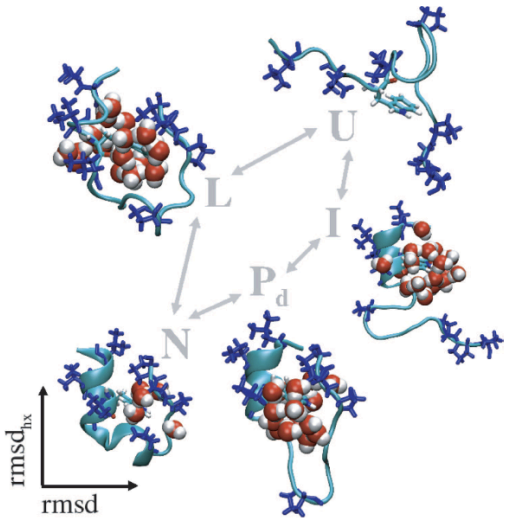


(color scale shows magnitude of the current, contours show PMF with 1 k_BT spacing)

$$R_{AB} = \{\mathbf{1}_C(x)\mathcal{L}[\mathbf{1}_{C^c}q_+](x) - \mathbf{1}_{C^c}(x)\mathcal{L}[\mathbf{1}_C q_+](x)\}\, q_-(x)\pi(x)dx$$
$$= \int_{\partial C} J_{AB}(x)\cdot n_C(x)d\sigma_C = \int_{\partial C^\theta} J^\theta_{AB}(s)\cdot n_{C^\theta}d\sigma_{C^\theta}$$

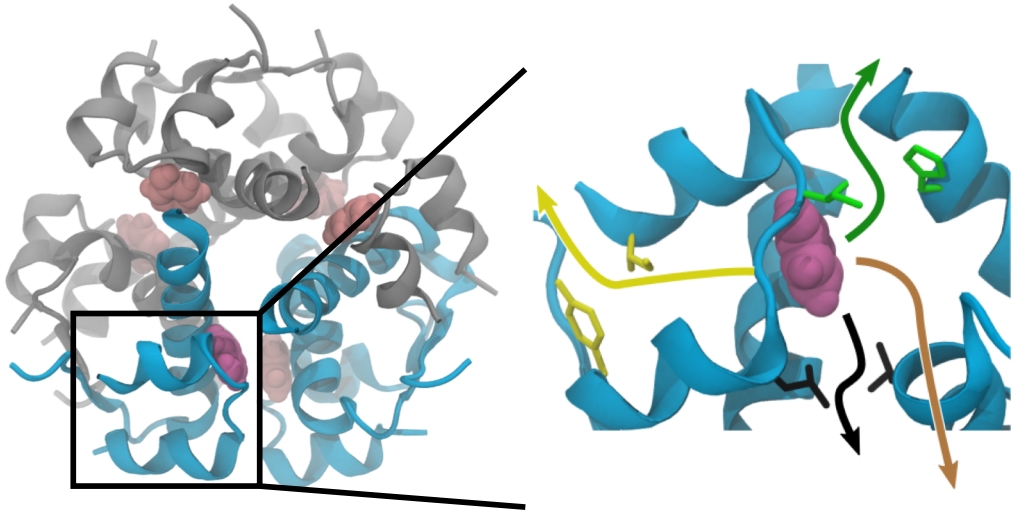
The first calculation of this key TPT quantity for a molecular simulation clearly shows the two pathways and allows quantification of their contributions.



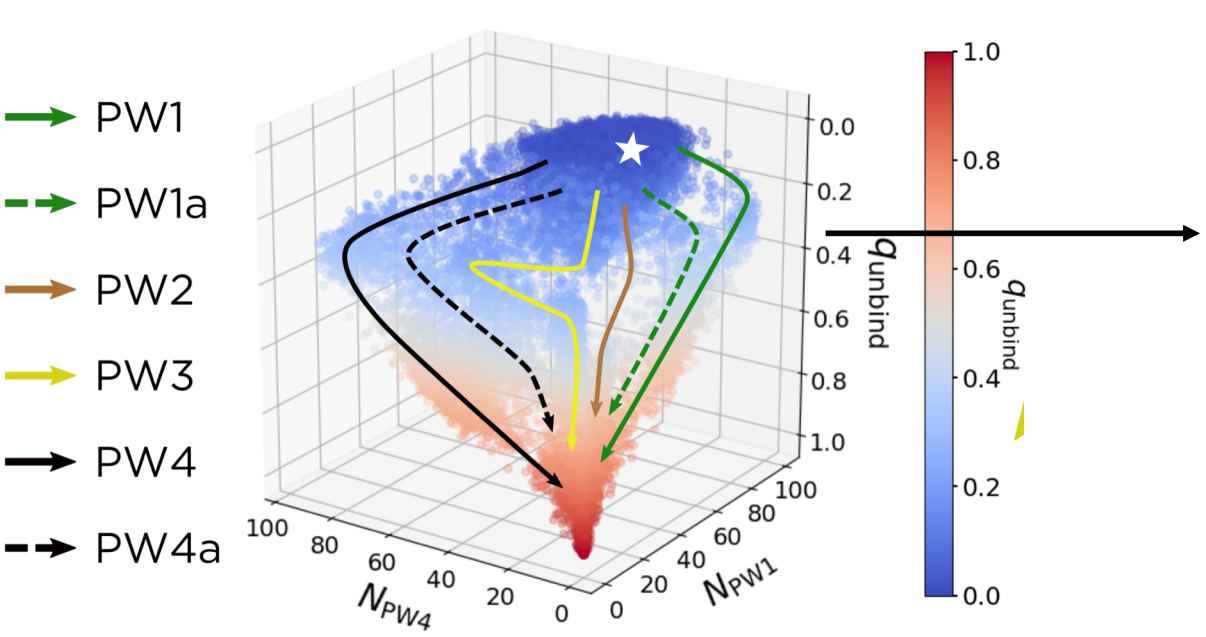


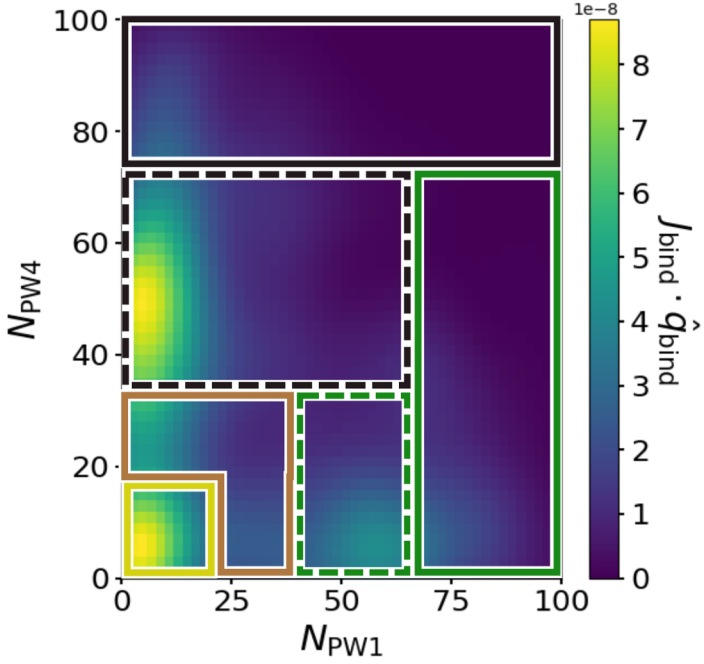
Juraszek and Bolhuis,
PNAS 103, 15859-15864
(2006)

DGA enables us to compute the relative fluxes through six pathways of phenol release from the insulin hexamer.



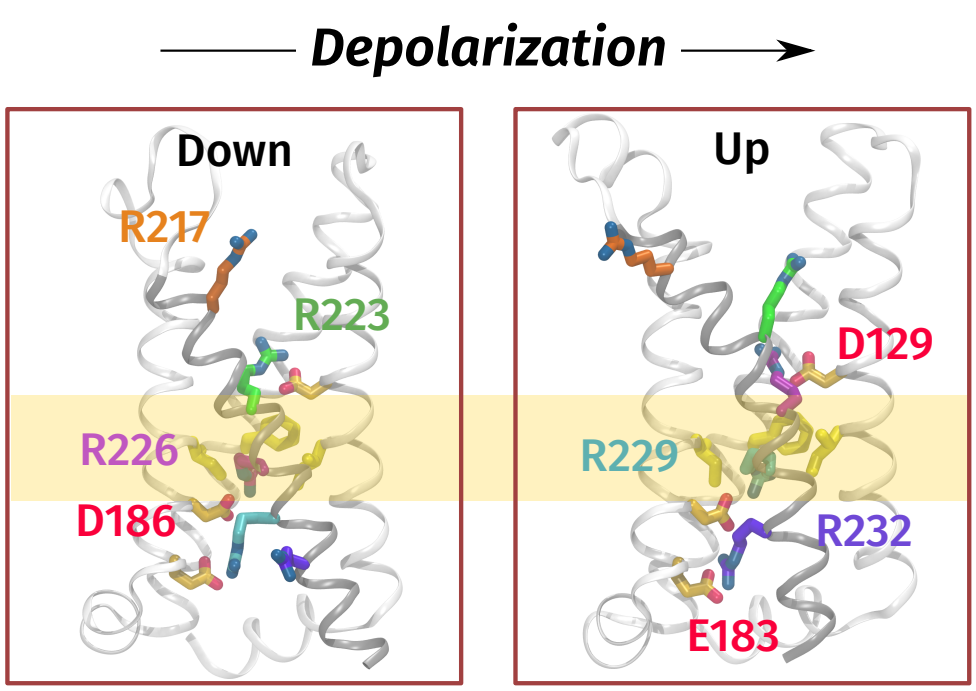
PW1: 11.2% PW1a: 11.2%
PW2: 16.3% PW3: 12.6%
PW4: 13.7% PW4a: 35.0%

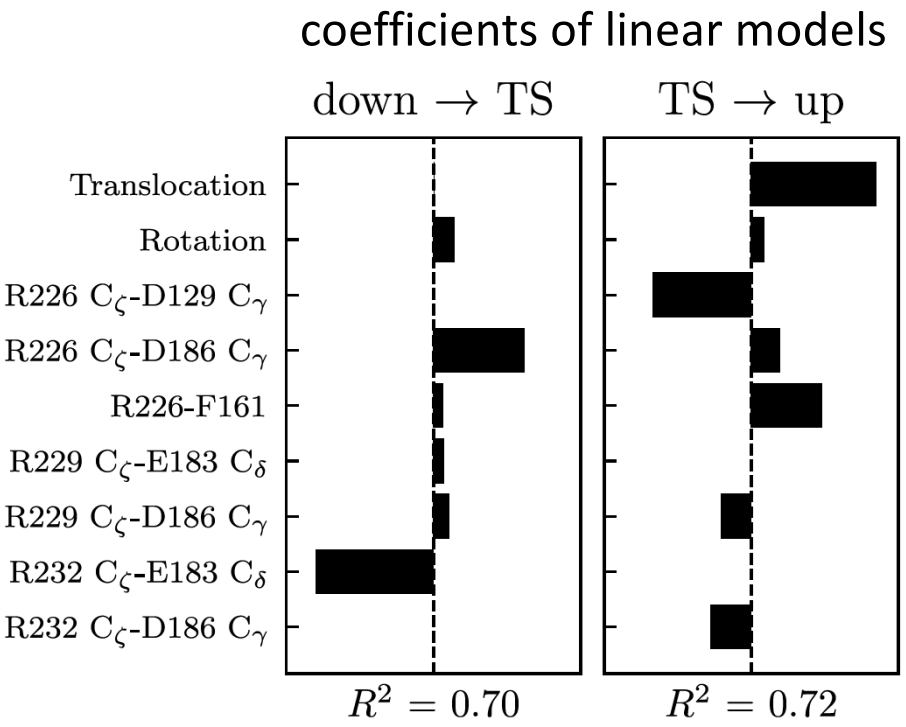


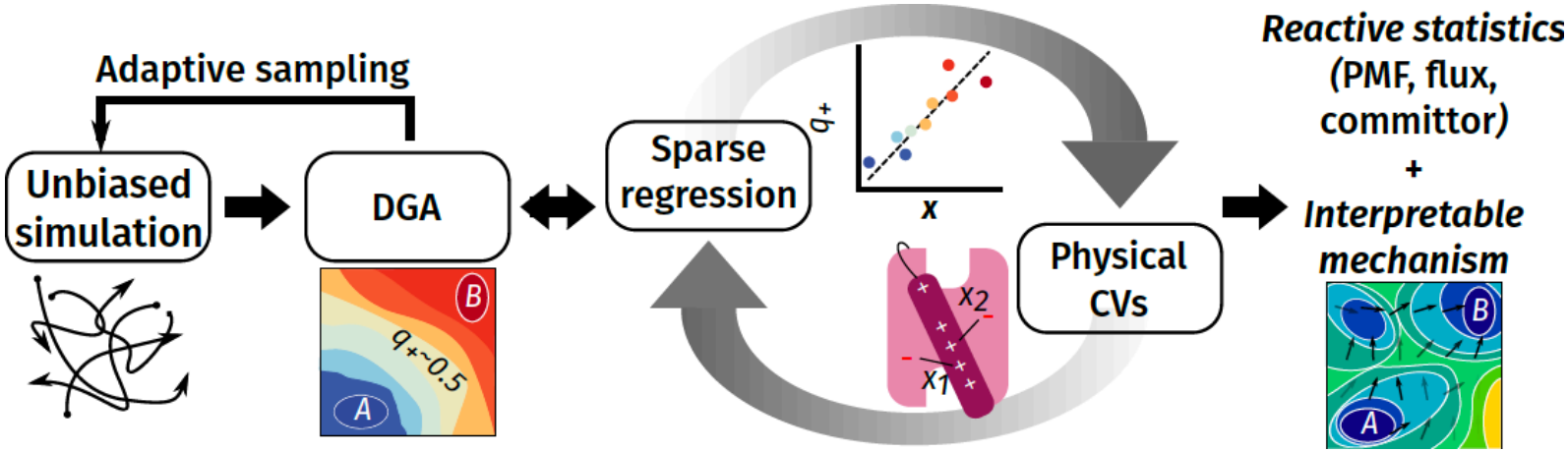


Antoszewski, Lorpaiboon, Strahan, Dinner, J Phys Chem B 125, 11637−11649 (2021)
See also Lorpaiboon, Weare, Dinner, J Chem Phys 157, 094115 (2022) for an augmented TPT.

DGA reveals a stepwise mechanism for voltage sensing.







Guo, Shen, Roux, Dinner, biorxiv.org (2022)

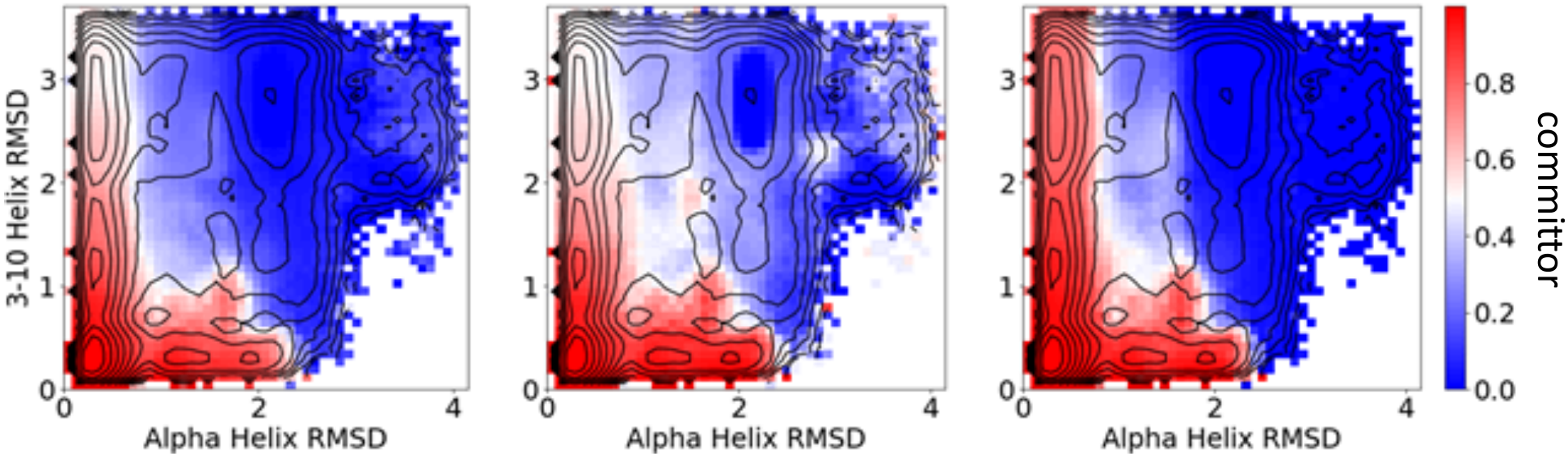
The choice of basis matters.

multiply 153 pairwise distances between selected Cα atoms by $d_A d_B/(d_A+d_B)^2$ and orthogonalize

standard for Markov State Models (MSMs)

indicator functions on pairwise distances

indicator functions on TICA coordinates



trp-cage committors computed for all structures (>6M) and then projected to 2D
(color scale shows committors, contours show PMF with 1 $k_B T$ spacing)

Outline

- A Galerkin approach
  - Applications to protein dynamics
- A neural-network approach
  - Numerical experiments illustrating key features
  - Application to a model of sudden stratospheric warming
- Addressing the two-trajectory requirement

Motivated by the equations we want to solve

$$(\mathcal{S}^\tau - \mathcal{I})u(x) = -\mathbb{E}_x\left[\int_0^{\tau\wedge T} h(X_s)ds\right] \text{ for } x \in D = (A \cup B)^c$$

$$u(x) = g(x) \text{ for } x \notin D$$

we consider the loss functions

$$C_{\text{FKE}} = \left\| \left((\mathcal{S}^\tau - \mathcal{I})u_\theta + \mathbb{E}_x\left[\int_0^{\tau\wedge T} h(X_s)ds\right] \right) \mathbb{1}_D \right\|_\mu^2$$

$$C_{\text{BC}} = \|(u_\theta - g)\mathbb{1}_{D^c}\|_\mu^2 .$$

and seek

$$\theta^* = \arg\min_\theta [C_{\text{FKE}} + \lambda C_{\text{BC}}].$$

Because there are terms that are quadratic in $\mathcal{S}^\tau u_\theta$, we need two trajectories from each initial point for unbiased estimates:

$$\bar{C}_{\mathrm{FKE}} = \frac{1}{n}\sum_{i=1}^{n}\left(u_\theta(X^{i,1}_{k^{i,1}\Delta}) - u_\theta(X^i_0) - \Delta\sum_{s=0}^{k^{i,1}} h(X^{i,1}_{s\Delta})\right) \times \left(u_\theta(X^{i,2}_{k^{i,2}\Delta}) - u_\theta(X^i_0) - \Delta\sum_{s=0}^{k^{i,2}} h(X^{i,2}_{s\Delta})\right)\mathbb{1}_D(X^i_0)$$

$$\bar{C}_{\mathrm{BC}} = \frac{1}{n}\sum_{i=1}^{n}\left(u_\theta(X^i_0) - g(X^i_0)\right)^2 \mathbb{1}_{D^c}(X^i_0).$$

We represent u_θ by a neural network and minimize

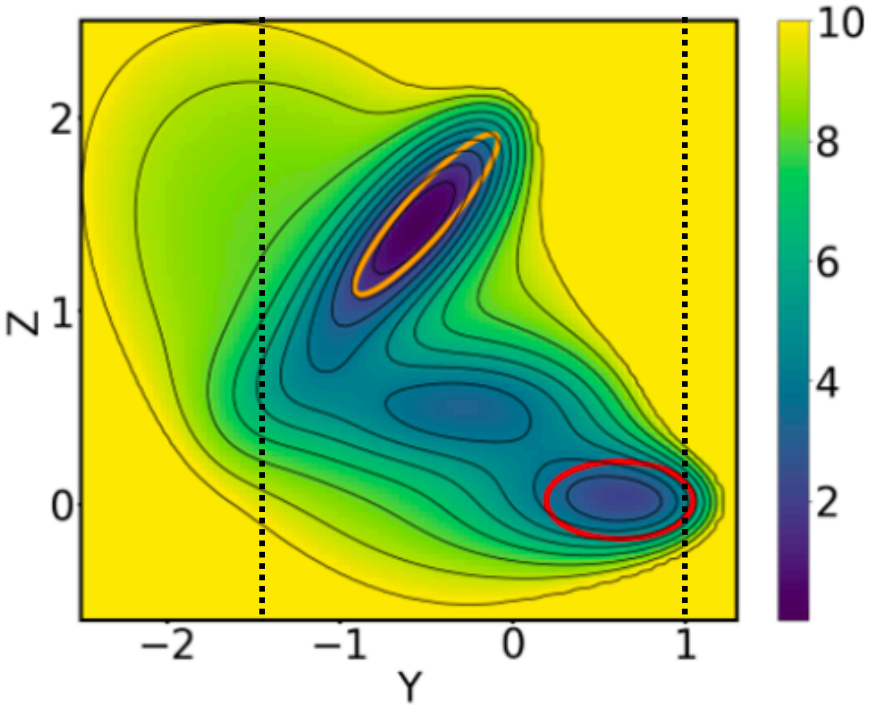
$$\bar{C} = \bar{C}_{\mathrm{FKE}} + \lambda\bar{C}_{\mathrm{BC}}$$

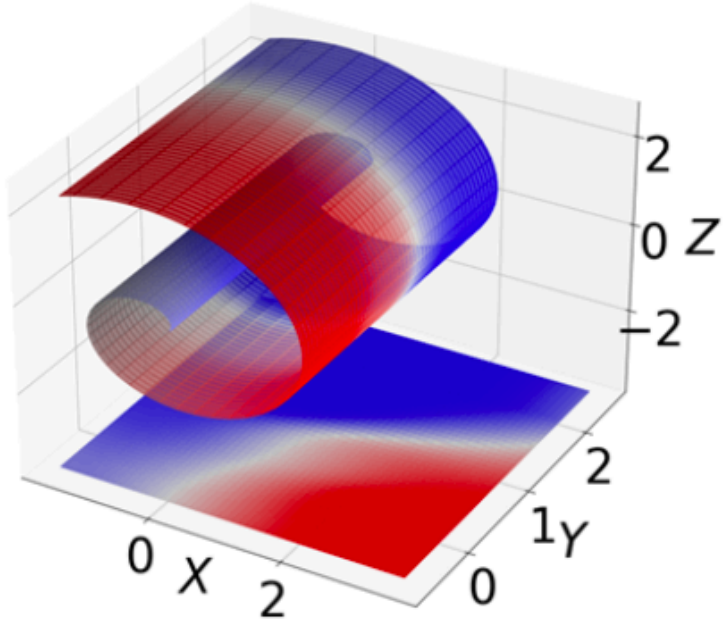
Our loss has several advantages over previous ones:

- it allows computation of any statistic that can be cast in Feynman-Kac form;
- it does not require explicit knowledge of the model;
- it does not require microscopic reversibility;
- it allows for use of arbitrary lag times;
- it allows use of an arbitrary sampling distribution.

Strahan, Finkel, Dinner, Weare, arxiv:2208.01717 (2022)

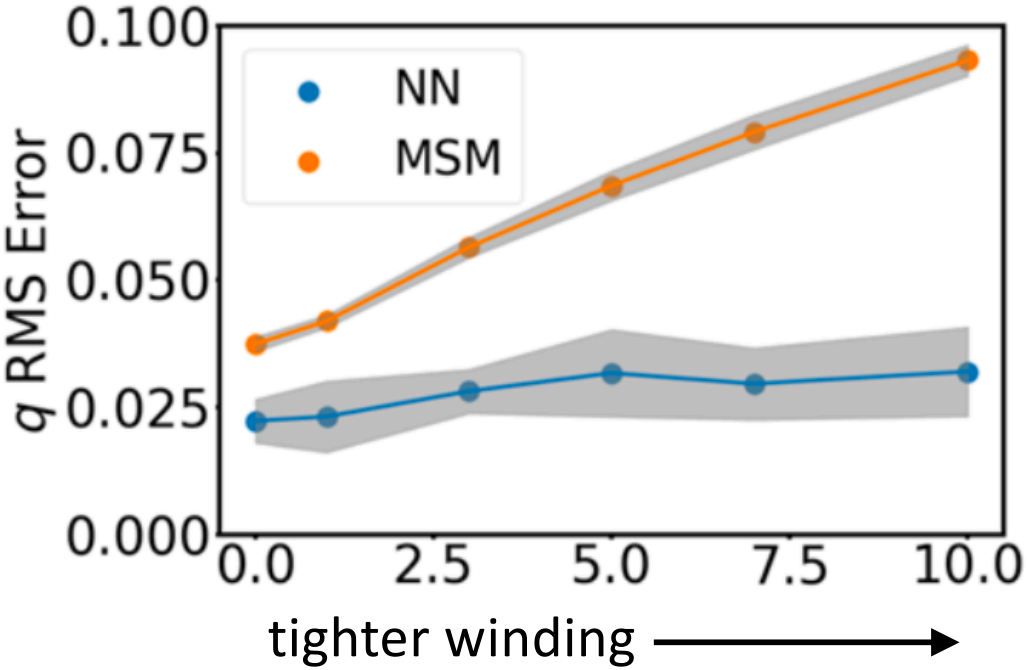
To illustrate these advantages, we consider dynamics on the Müller-Brown potential mapped to the Swiss roll.

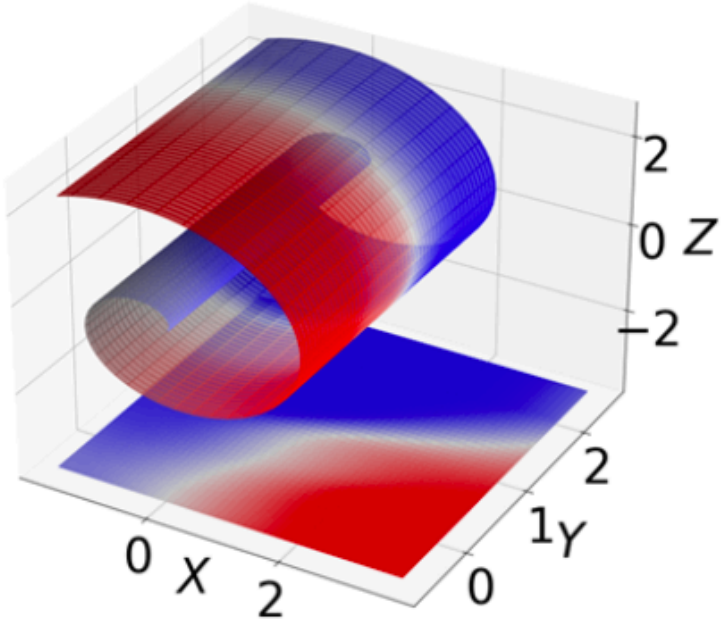




- Overdamped Langevin dynamics initiated from 30,000 points drawn uniformly from between the dashed lines.
- Map the trajectories onto the Swiss Roll so the algorithm must learn the manifold of the dynamics.
- Compare a 3-input, 3 30-sigmoid hidden layers, 1-sigmoid output neural network with a 300 state MSM and a grid-based reference.

To illustrate these advantages, we consider dynamics on the Müller-Brown potential mapped to the Swiss roll.

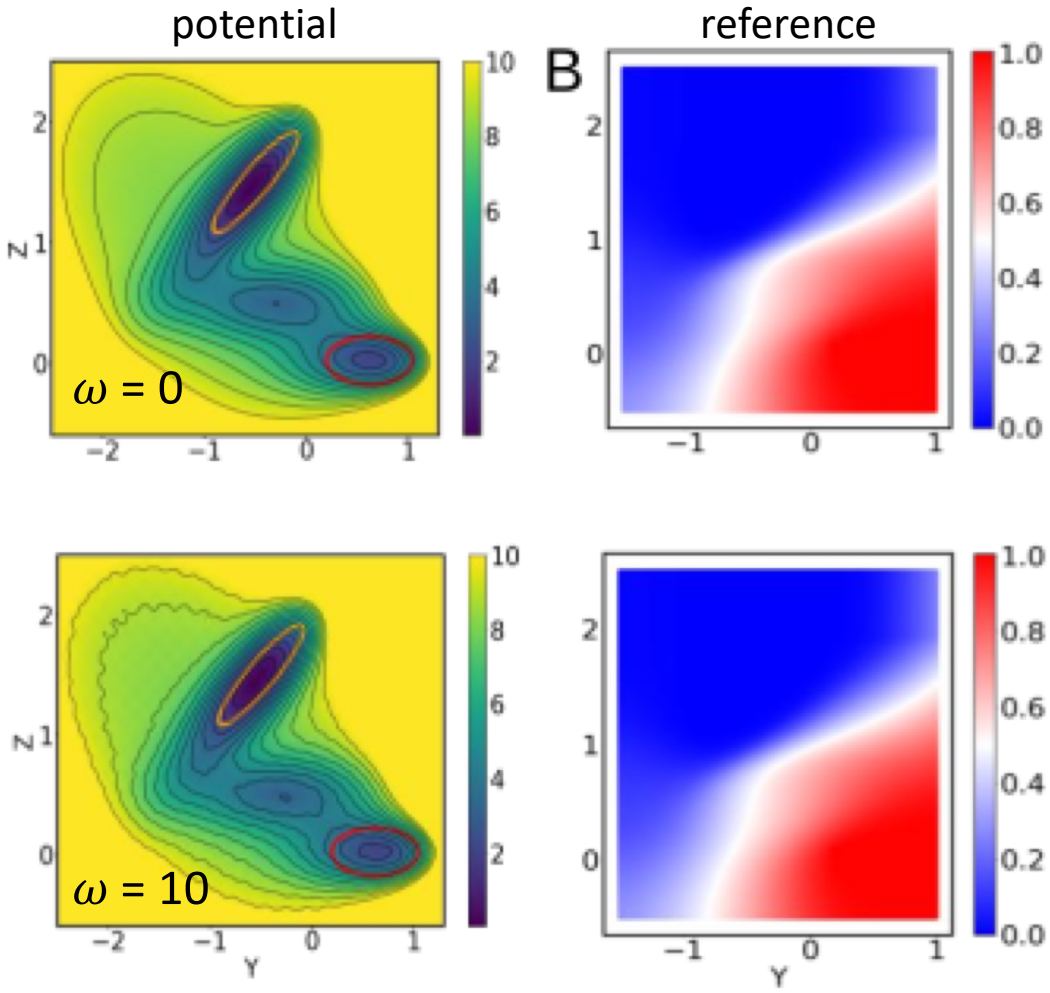


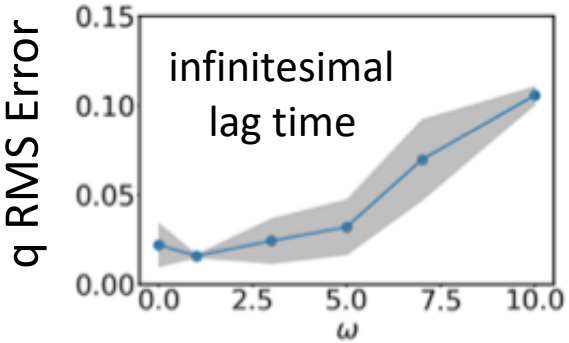


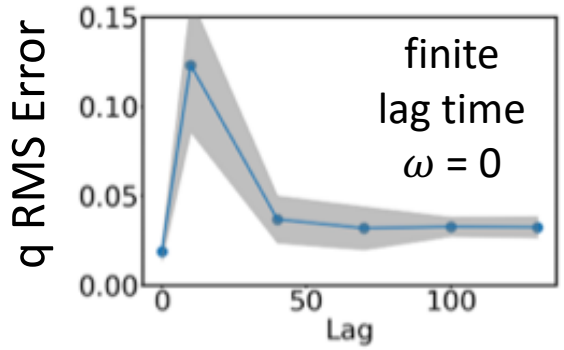
- Overdamped Langevin dynamics initiated from 30,000 points drawn uniformly from between the dashed lines.
- Map the trajectories onto the Swiss Roll so the algorithm must learn the manifold of the dynamics.
- Compare a 3-input, 3 30-sigmoid hidden layers, 1-sigmoid output neural network with a 300 state MSM and a grid-based reference.

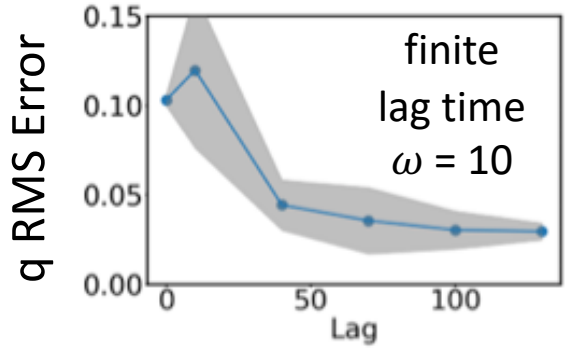
Infinitesimal (or short) vs finite (or long) lag time

Compare with $\|\mathcal{L}q\|_\mu^2$, which gives rise to
$\|\nabla q\|_\pi^2$ and $\|q(\tau) - q(0)\|_\pi^2$ for overdamped reversible dynamics



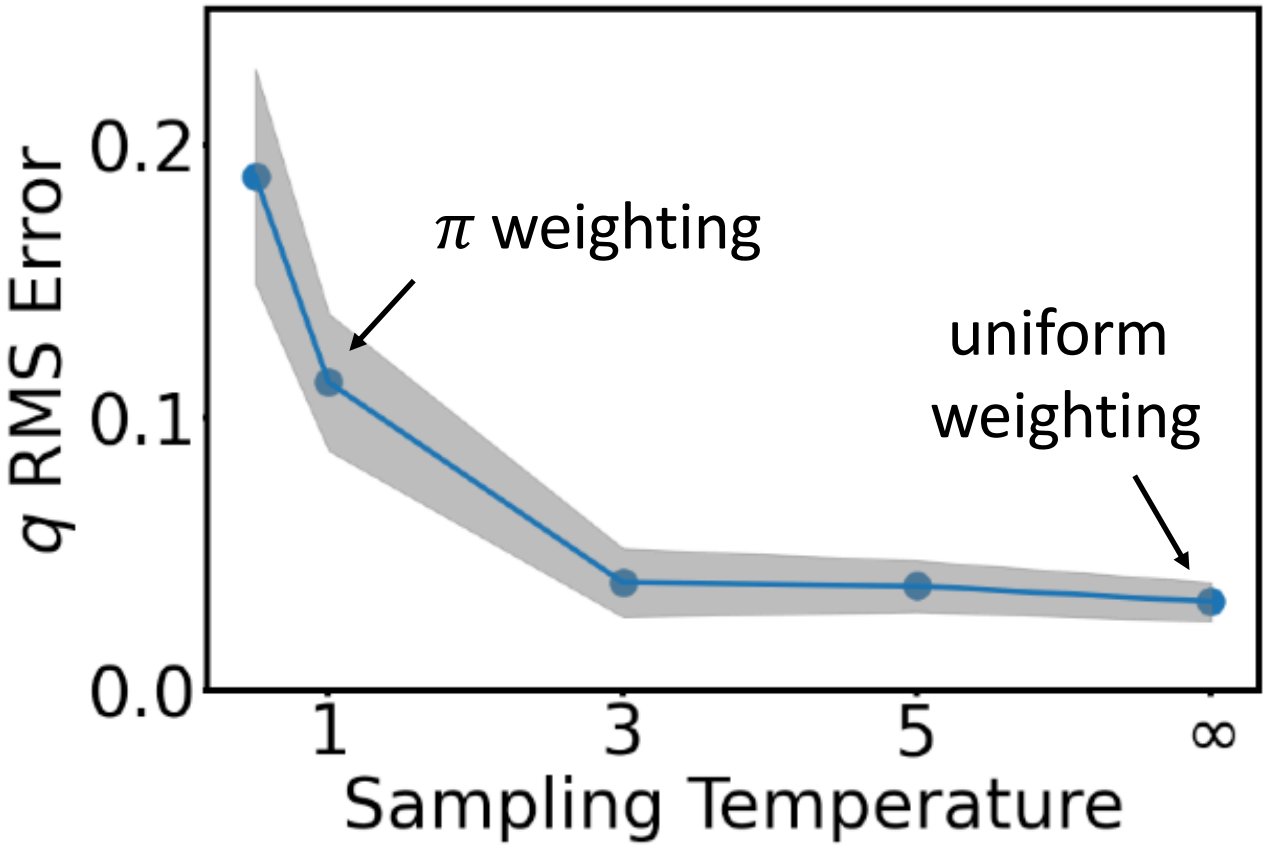






Choice of sampling distribution (arbitrary μ vs. π)

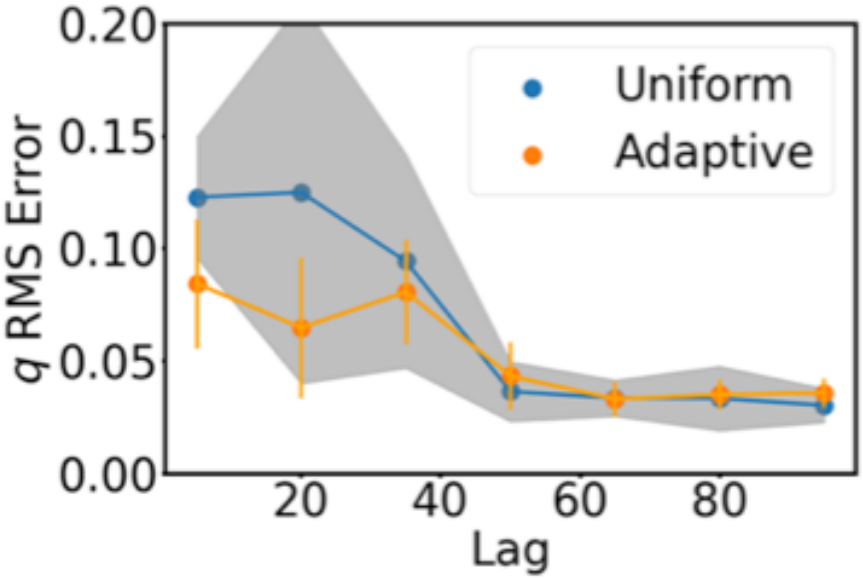
A uniform distribution of initial points is better than the stationary distribution.

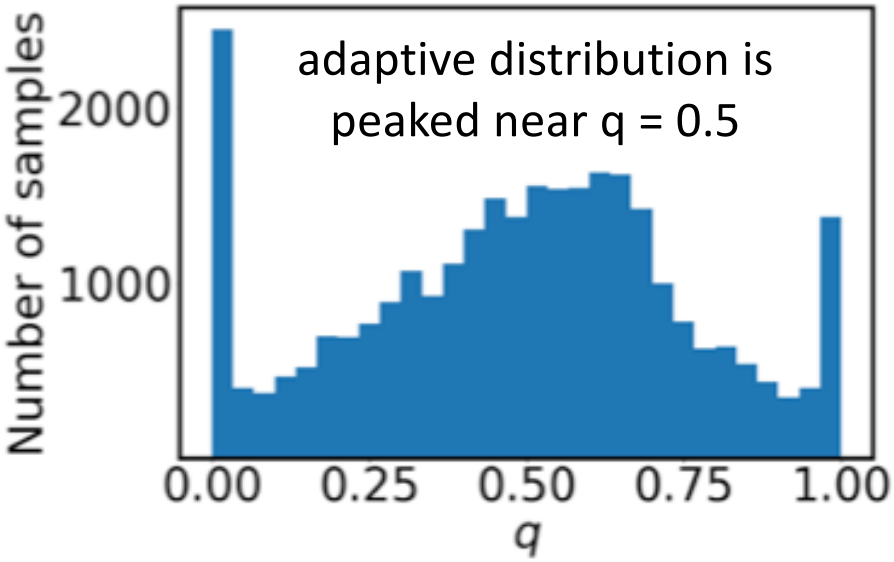


Choice of sampling distribution (arbitrary μ vs. π)

Adaptively sample with probability

$$P_\ell = \frac{\int \left((\mathcal{T}_{D^c}^{\tau} - \mathcal{I}) u_\theta(x) - \mathbb{E}_x \left[\int_0^{\tau \wedge T} h(X_s) ds \right] \right)^2 \mathbb{1}_{S_\ell}(u_\theta(x)) \mu(x) dx}{\int \mathbb{1}_{S_\ell}(u_\theta(x)) \mu(x) dx}$$

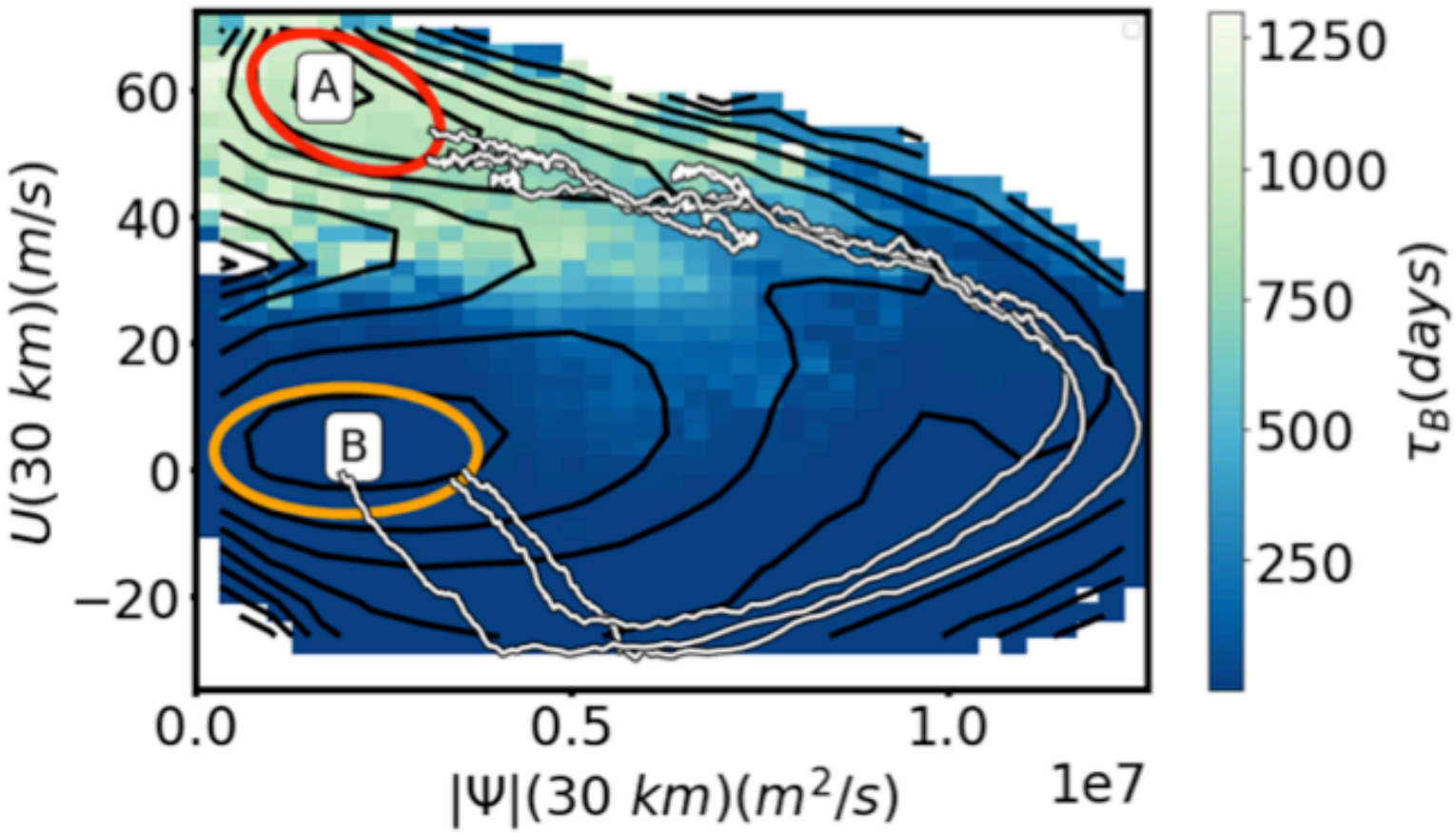




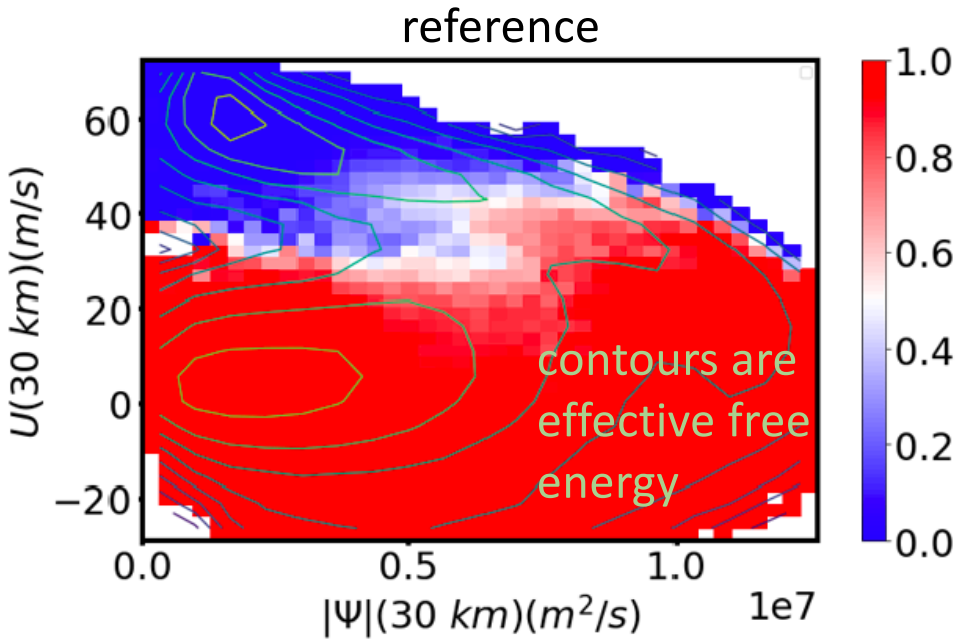
The Holton-Mass model describes stratospheric flow in terms of the interaction of two fields:

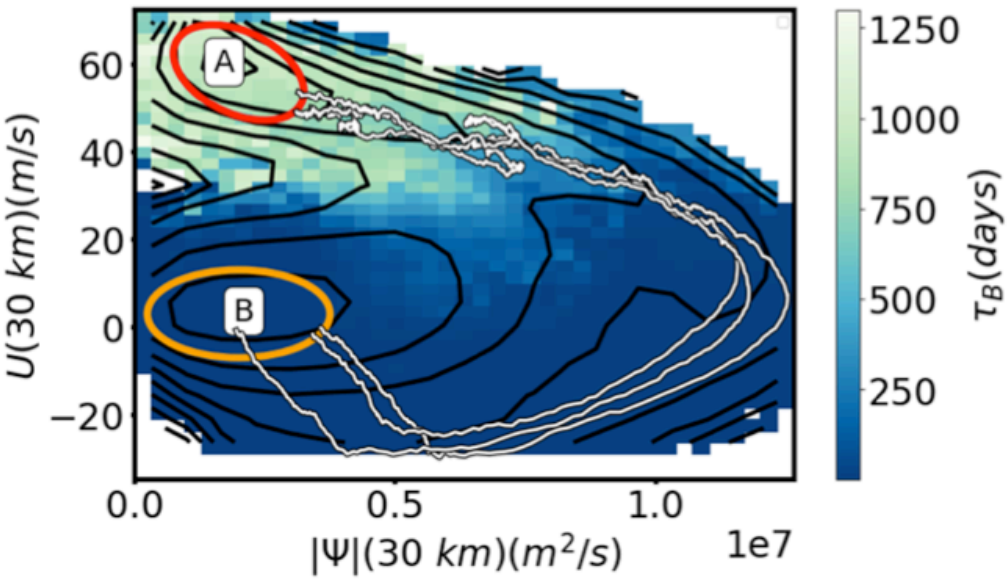
- the mean flow $\bar{u}$;
- the perturbation streamfunction ψ'.

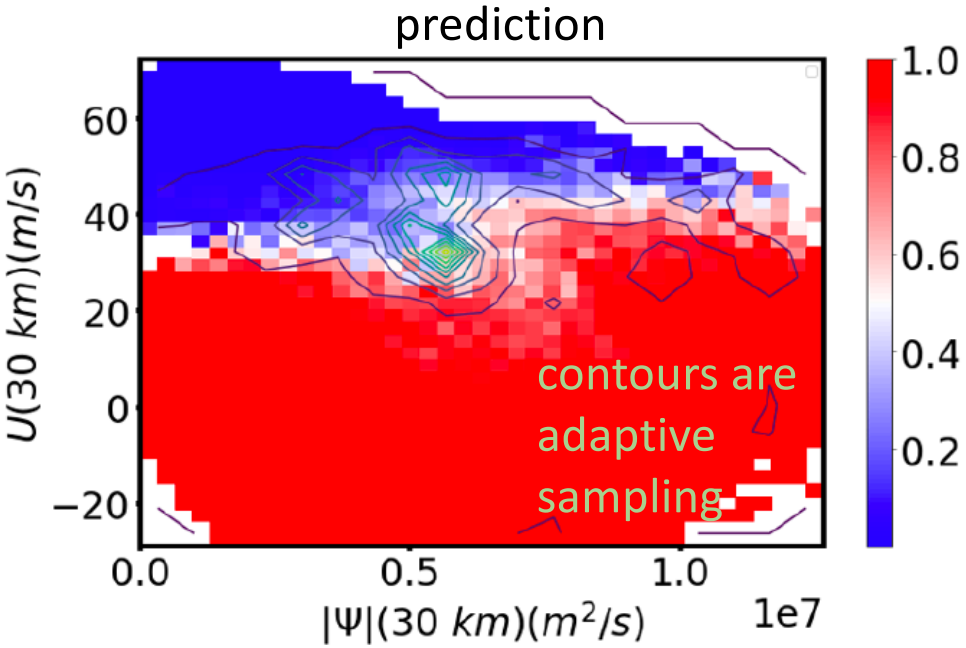
These fields are spatially discretized, resulting in a 75-dimensional state space.

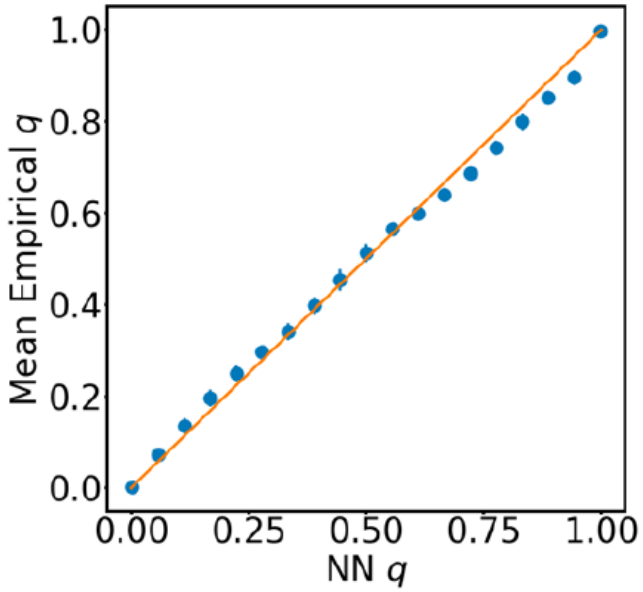


Holton-Mass model committor





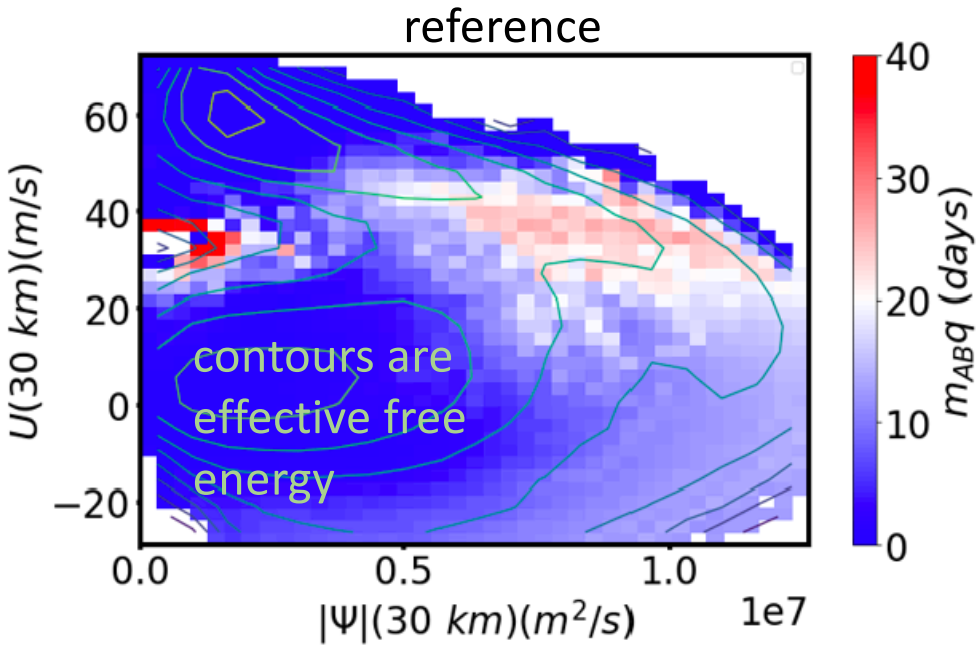


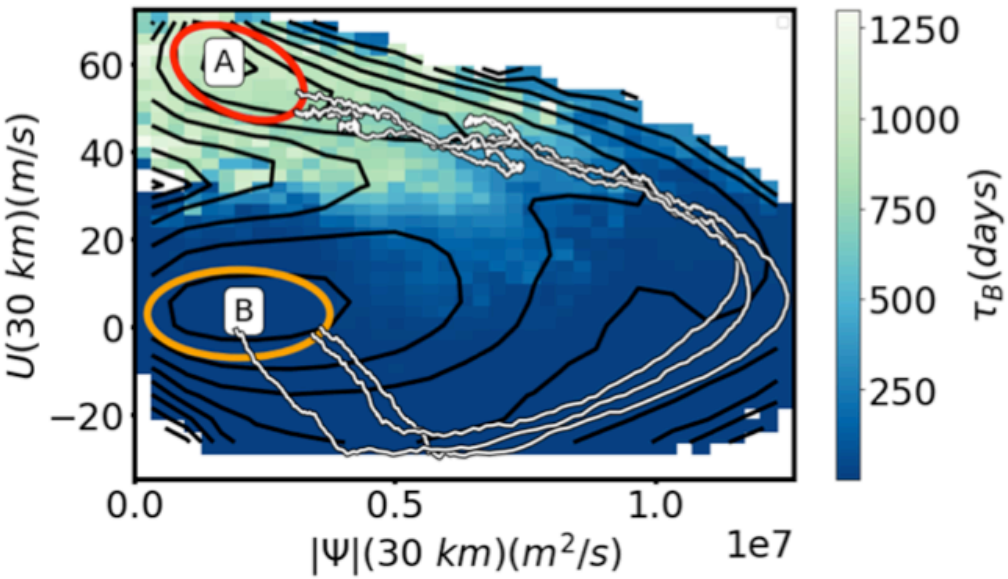


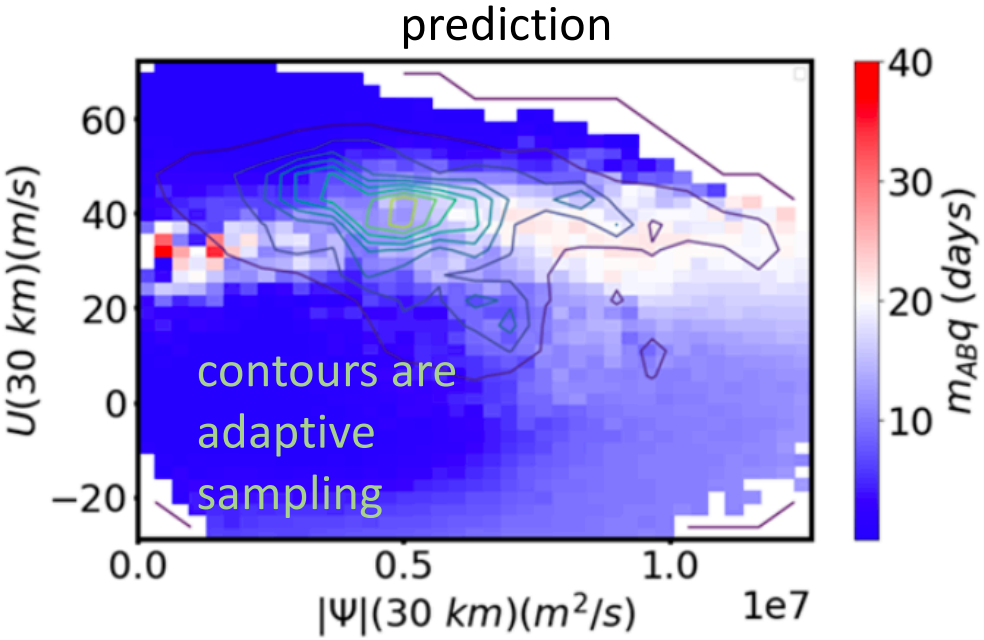
τ = 10 days

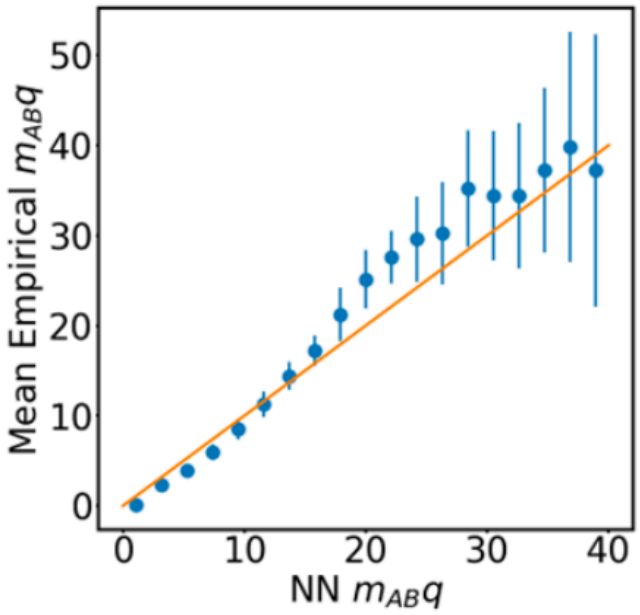
75 inputs, 5 50-ReLU hidden layers, 1-sigmoid output

Holton-Mass model lead time









τ = 10 days

75 inputs, 5 50-ReLU hidden layers, 1-sigmoid output

Our loss has several advantages over previous ones:

- it allows computation of any statistic that can be cast in Feynman-Kac form;
- it does not require explicit knowledge of the model;
- it does not require microscopic reversibility;
- it allows for use of arbitrary lag times;
- it allows use of an arbitrary sampling distribution.

But it requires two trajectories from each initial condition.

Outline

- A Galerkin approach
  - Applications to protein dynamics
- A neural-network approach
  - Numerical experiments illustrating key features
  - Application to a model of sudden stratospheric warming
- Addressing the two-trajectory requirement

The gradient of our norm is

$$\langle (\mathcal{I} - \mathcal{S}^\tau)\, u_\theta - r, \nabla_\theta u_\theta \rangle_\mu - \langle (\mathcal{I} - \mathcal{S}^\tau)\, u_\theta - r, \mathcal{S}^\tau \nabla_\theta u_\theta \rangle_\mu \,,$$

where $r(x) = -\mathbb{E}_x \left[\int_0^{\tau \wedge T} h(X^s)\, ds \right]$.

In TD methods for reinforcement learning, a "semigradient" descent is often used (i.e., the second term is dropped).

Notice that the first term is the exact gradient of the loss

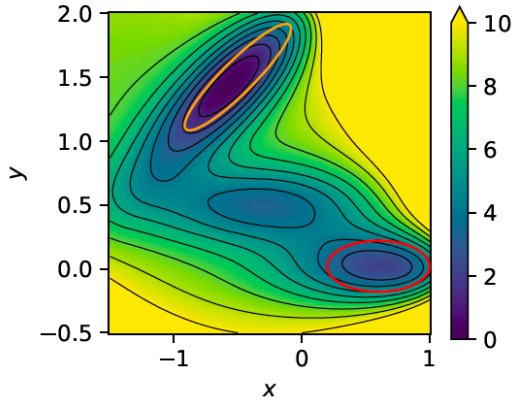
$$\frac{1}{2} \| u_{\theta'} - \mathcal{S}^\tau u_\theta - r \|_\mu^2,$$

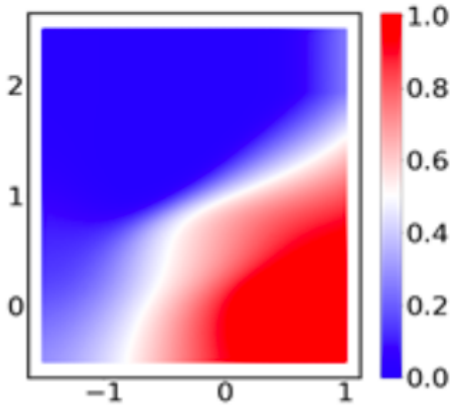
evaluated at $\theta' = \theta$. This suggests the (Richardson) iteration

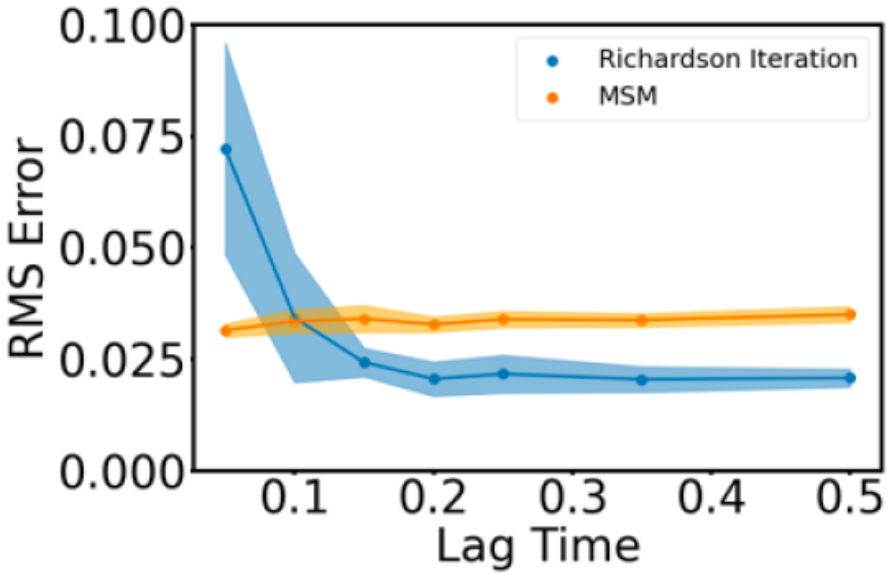
$$u_{\theta^{s+1}} \approx \mathcal{S}^\tau u_{\theta^s} + r.$$

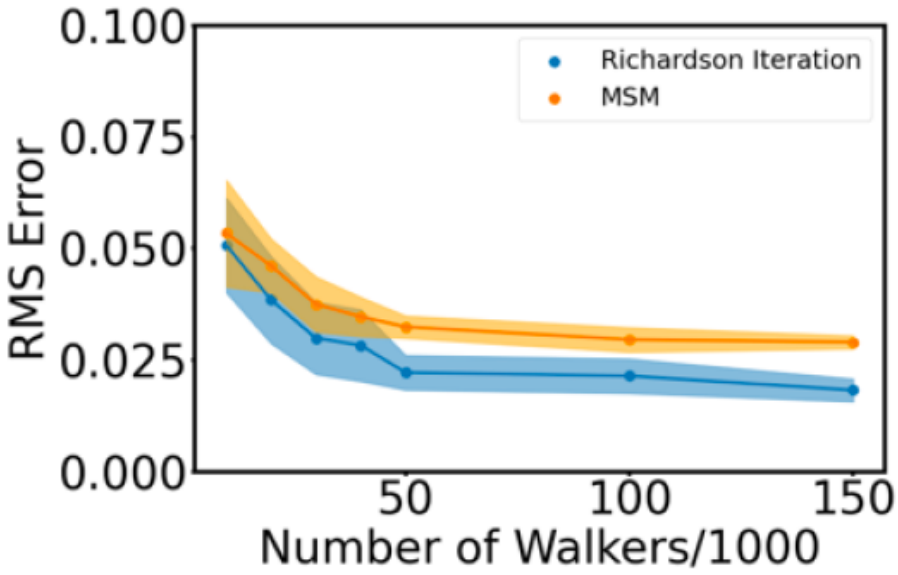
Müller-Brown committor (2D)

Comparison of Richardson iteration with a 400-state MSM

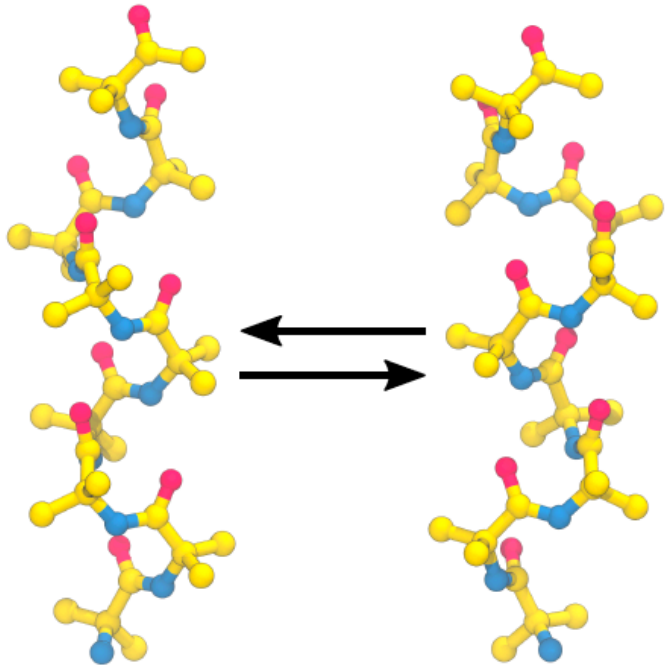


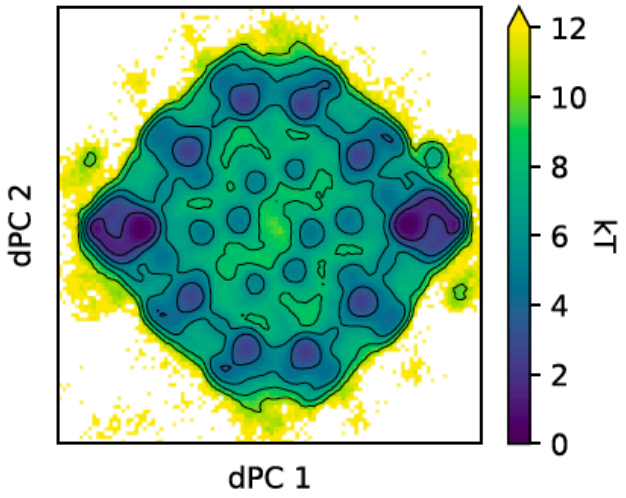






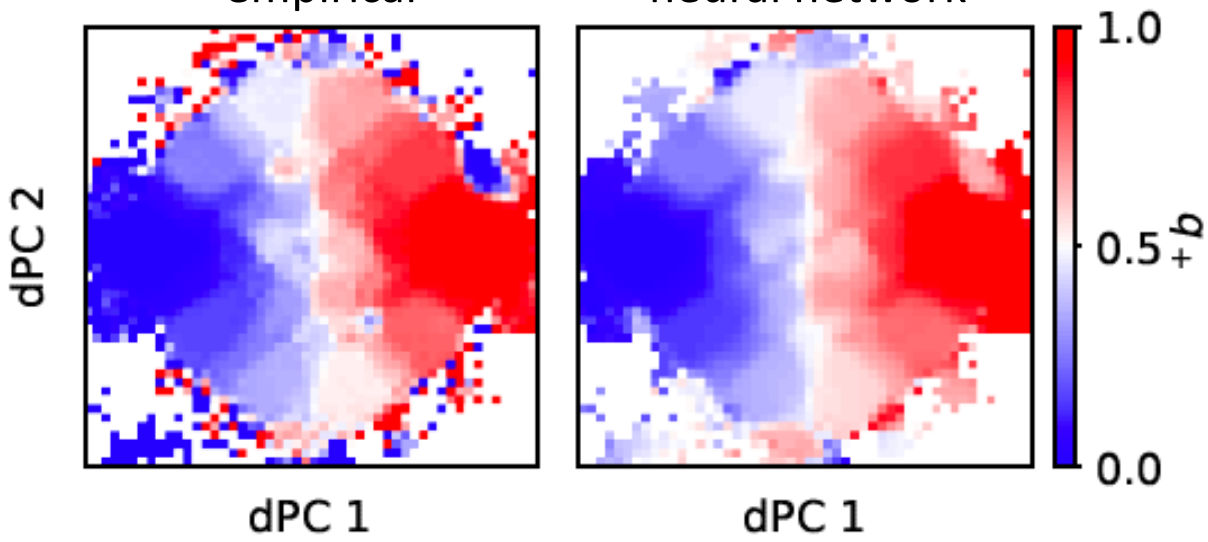
Committor for a peptide of 9 α-aminoisobutyric acids

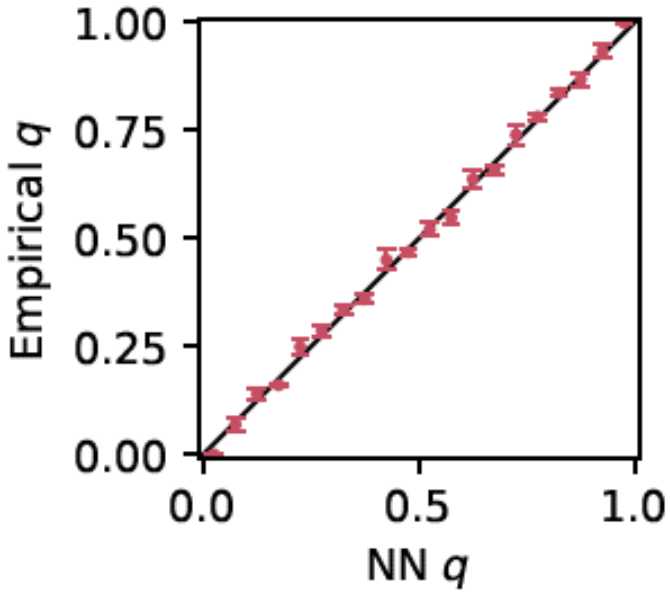




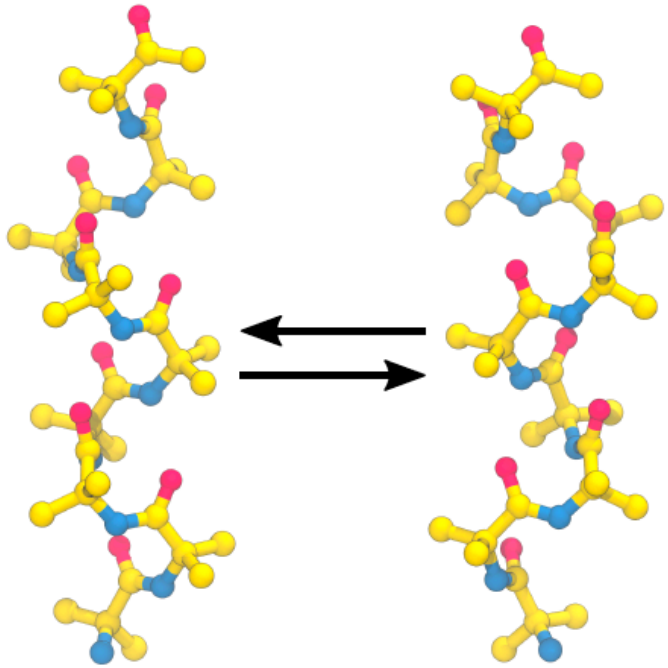
empirical

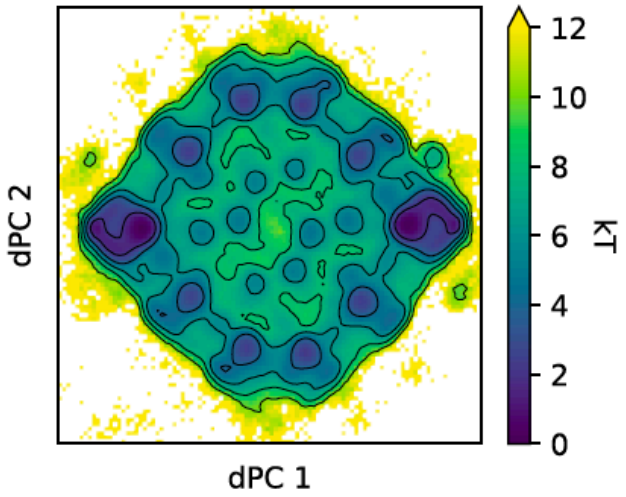
neural network

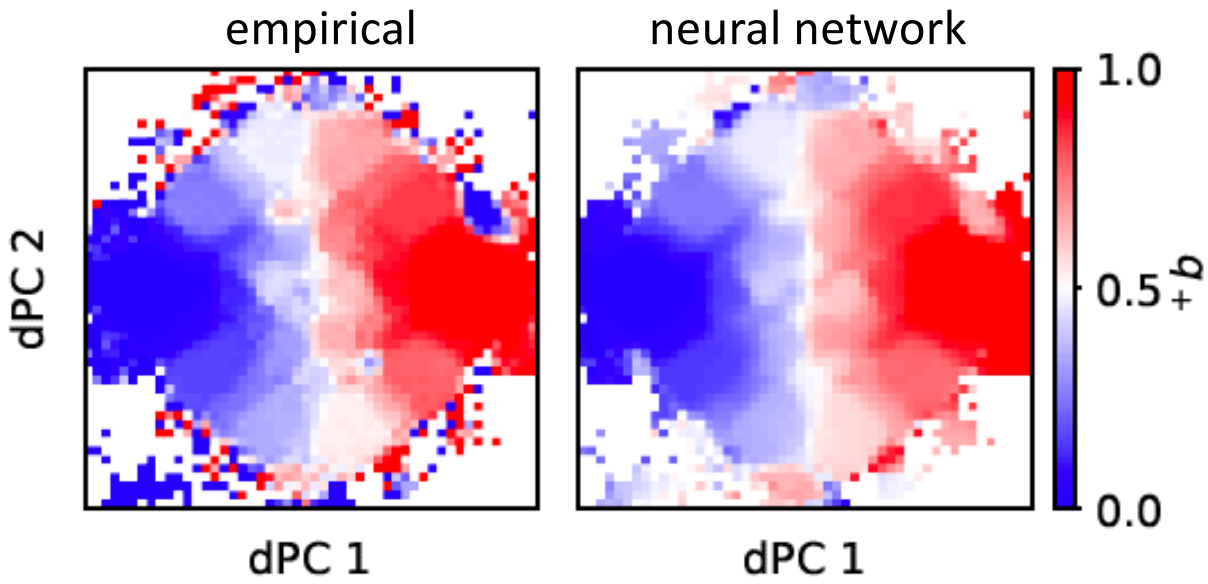


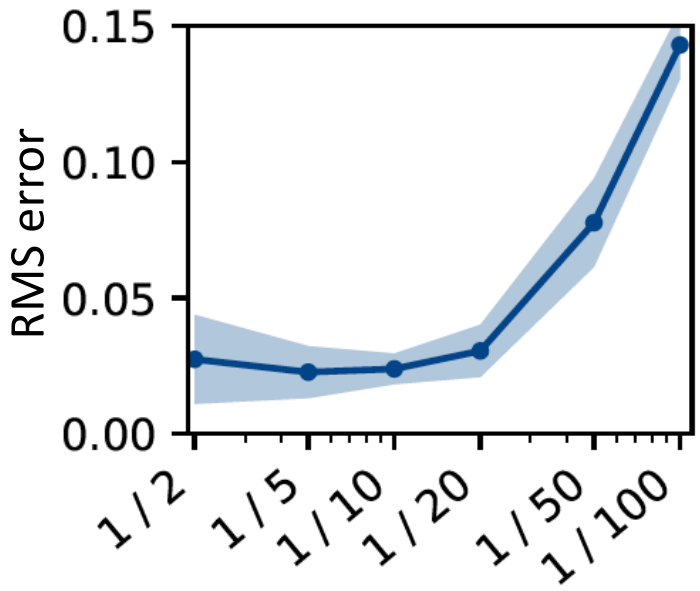


Committor for a peptide of 9 α-aminoisobutyric acids









Summary

We can obtain rare-event statistics by solving equations of the transition operator.

We have used both Galerkin approximation (basis expansion) and variational approaches based on neural networks to solve these equations.

Our methods can be applied to trajectories sampled at finite lag times, without knowledge of the underlying model or its dynamics (which can be irreversible).

Dynamical Galerkin Approximation

- Thiede, Giannakis, Dinner, Weare, JCP 150, 244111 (2019)
- Strahan, Antoszewski, Lorpaiboon, Vani, Weare, Dinner, JCTC 17, 2948 (2021)

Applications

- Antoszewski, Lorpaiboon, Strahan, Dinner JPC B 125, 11637-11649 (2021)
- Guo, Shen, Roux, Dinner, biorxiv.org (2022)

Neural networks

- Strahan, Finkel, Dinner, Weare, arxiv:2208.01717 (2022)
- Strahan, Guo, Dinner, Weare, in preparation.