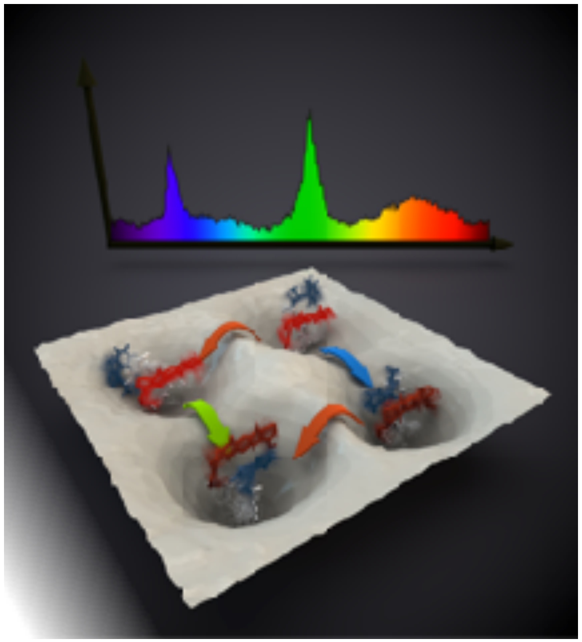
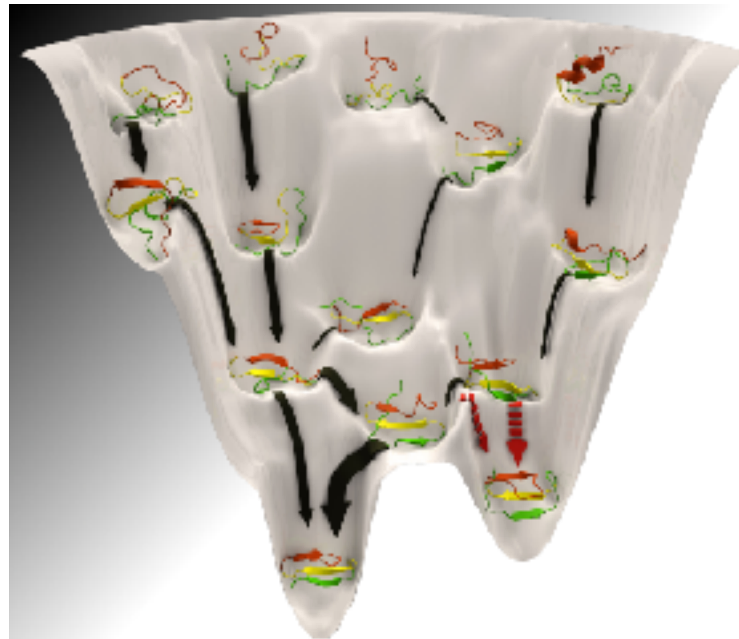
Model optimization and selection:
Variational Approach for Markov Processes (VAMP)





Frank Noé (FU Berlin)
frank.noe@fu-berlin.de

Computational Molecular Biology

Freie Universität Berlin

Motivation

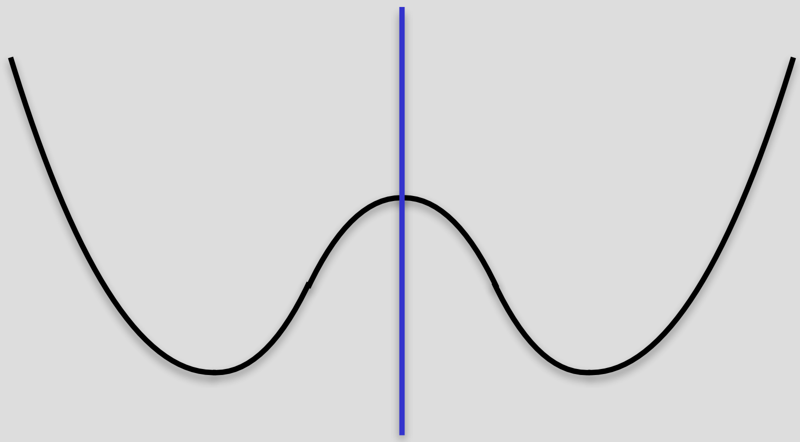
Which parameters?

$(x_1, x_2, \ldots, x_T)$

$(s_1, s_2, \ldots, s_T)$

transition matrix?

How many states?



2 ?
10 ?
1000 ?

Which features?

Ca-coordinates ?
distances ?
contacts ?

Parameter optimization problem

Hyperparameter optimization / model selection problem

Solving model selection problem requires two ingredients:

1) A **score** to rank models (MSMs, TICA, etc) by goodness

    ==> Variational principle

2) A **statistical validation method** to avoid overfitting

    ==> Cross-validation https://en.wikipedia.org/wiki/Cross-validation_(statistics)

Slow processes

Backward propagator

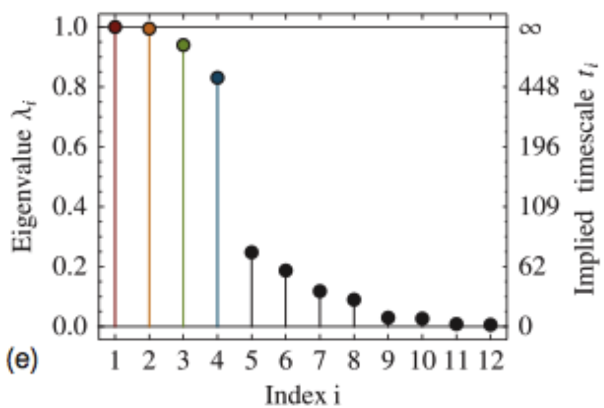
$$\rho_\tau = \mathcal{T}(\tau)\rho_0$$

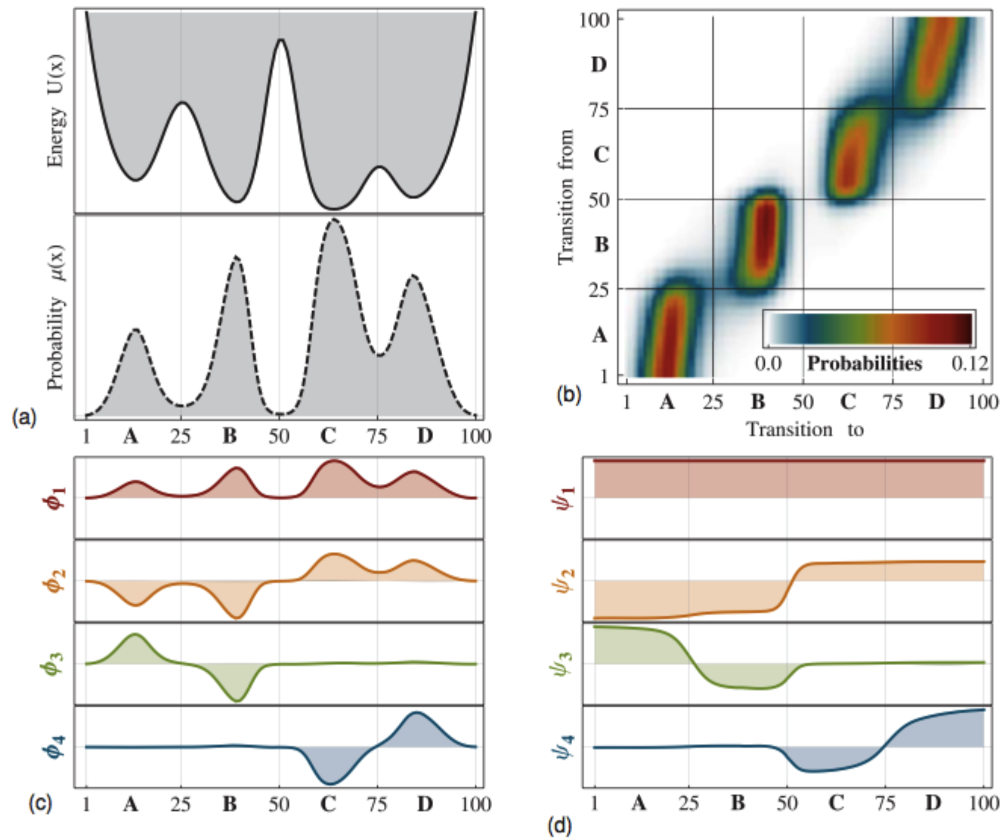
Spectral decomposition

$$\rho_\tau = \sum_{i=1}^{\infty} \mathrm{e}^{-\tau\kappa_i} \langle \psi_i \mid \rho_0 \rangle \psi_i$$

Processes:

Eigenvalues / timescales κ_i^{-1}





Schütte et al: **J. Comput. Phys.** (1999), Prinz et al.: **J. Chem. Phys.** 134, p174105 (2011)

Part I : Variational score

***Variational theorem**: For any $m \geq 1$, the first m eigenfunctions $\psi_1, \ldots, \psi_m$ are the solution of the following optimization problem*

$$R_m = \max_{f_1,\ldots,f_m} \sum_{i=1}^{m} \mathbb{E}_\mu \left[f_i\left(\mathbf{x}_t\right) f_i\left(\mathbf{x}_{t+\tau}\right)\right],$$

$$\text{s.t.} \quad \mathbb{E}_\mu \left[f_i\left(\mathbf{x}_t\right)^2 \right] = 1,$$

$$\mathbb{E}_\mu \left[f_i\left(\mathbf{x}_t\right) f_j\left(\mathbf{x}_t\right)\right] = 0, \ \textit{for } i \neq j,$$

where $\mathbf{E}_\mu[\cdot]$ denotes the expected value with $\mathbf{x}_t$ sampled from the stationary density μ and the maximum value is the generalized Rayleigh quotient, or Rayleigh trace $R_m = \sum_{i=1}^{m} \lambda_i$.

Noé and Nüske, **MMS** 11, 635-655 (2013)
Nüske et al, **JCTC** 10, 1739-1752 (2014)

Part I : Variational score

Variational theorem: *For any $m \geq 1$, take a set of functions $f_1, ..., f_m$ and the covariance matrices $\mathbf{C}(0)$ and $\mathbf{C}(\tau)$ with elements:*

$$c_{ij}(0) = \mathbb{E}_t\left[f_i(\mathbf{x}_t)\, f_j(\mathbf{x}_t)\right]$$
$$c_{ij}(\tau) = \mathbb{E}_t\left[f_i(\mathbf{x}_t)\, f_j(\mathbf{x}_{t+\tau})\right]$$

where $\mathbb{E}_t[\cdot]$ denotes the ergodic time average. Now perform an eigenvalue decomposition

$$\mathbf{C}(\tau)\mathbf{b}_k = \mathbf{C}(0)\mathbf{b}_k\hat{\lambda}_k$$
$$\mathbf{K}\mathbf{b}_k = \mathbf{b}_k\hat{\lambda}_k,$$

where we have used the abbreviation $\mathbf{K} = \mathbf{C}(0)^{-1}\mathbf{C}(\tau)$.

Then, the maximization of the Rayleigh trace:

$$R_m = \max_{f_1,\ldots,f_m} \sum_{k=1}^{m} \hat{\lambda}_k = \sum_{i=1}^{m} \lambda_i.$$

solves the first m eigenfunctions $\psi_1, \ldots, \psi_m$ of the transfer operator by

$$\psi_k(\mathbf{x}) = \sum_i b_{ki} f_i(\mathbf{x}).$$

Noé and Nüske, **MMS** 11, 635-655 (2013)
Nüske et al, **JCTC** 10, 1739-1752 (2014)

Part I : Variational score

***Variational theorem**: For any $m \geq 1$, take a set of functions $f_1, ..., f_m$ and the covariance matrices $\mathbf{C}(0)$ and $\mathbf{C}(\tau)$ with elements:*

$$c_{ij}(0) = \mathbb{E}_t \left[f_i(\mathbf{x}_t) \, f_j(\mathbf{x}_t) \right]$$
$$c_{ij}(\tau) = \mathbb{E}_t \left[f_i(\mathbf{x}_t) \, f_j(\mathbf{x}_{t+\tau}) \right]$$

where $\mathbb{E}_t [\cdot]$ denotes the ergodic time average. Now perform an eigenvalue decomposition

$$\mathbf{C}(\tau)\mathbf{b}_k = \mathbf{C}(0)\mathbf{b}_k \hat{\lambda}_k$$
$$\mathbf{K}\mathbf{b}_k = \mathbf{b}_k \hat{\lambda}_k,$$

where we have used the abbreviation $\mathbf{K} = \mathbf{C}(0)^{-1}\mathbf{C}(\tau)$.

Then, the maximization of the kinetic variance:

$$K_m = \max_{f_1,...,f_m} \sum_{k=1}^{m} \hat{\lambda}_k^2 = \sum_{i=1}^{m} \lambda_i^2.$$

solves the first m eigenfunctions $\psi_1, \ldots, \psi_m$ of the transfer operator by

$$\psi_k(\mathbf{x}) = \sum_i b_{ki} f_i(\mathbf{x}).$$

Noé and Nüske, **MMS** 11, 635-655 (2013)
Nüske et al, **JCTC** 10, 1739-1752 (2014)
*Noé and Clementi, **JCTC** 11, 5002—5011 (2015)

Variational Approach for Markov processes (VAMP)

We have the trajectory

$$\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T)$$

We define features $f_1, ..., f_n$ that are candidates for the eigenfunctions or singular functions. For each configuration we thus get a n-dimensional feature vector:

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x}))$$

We estimate covariance matrices, for example using the direct estimator:

$$\mathbf{C}_{\mathbf{f}}^{00}(\mathbf{X}) = \frac{1}{T-\tau}\sum_t^{T-\tau} \mathbf{f}(\mathbf{x}(t))\, \mathbf{f}^\top(\mathbf{x}(t))$$

$$\mathbf{C}_{\mathbf{f}}^{0\tau}(\mathbf{X}) = \frac{1}{T-\tau}\sum_t^{T-\tau} \mathbf{f}(\mathbf{x}(t))\, \mathbf{f}^\top(\mathbf{x}(t+\tau))$$

$$\mathbf{C}_{\mathbf{f}}^{\tau\tau}(\mathbf{X}) = \frac{1}{T-\tau}\sum_t^{T-\tau} \mathbf{f}(\mathbf{x}(t+\tau))\, \mathbf{f}^\top(\mathbf{x}(t+\tau)).$$

and $\mathbf{C}_\tau$ (either TICA or MSMs).

The estimates of $\mathbf{C}^{00}$ and $\mathbf{C}^{\tau\tau}$ are sometimes regularized by adding $\lambda\mathbf{I}$, an identity matrix scaled by a small parameter λ (also called Shrinkage or Ridge estimator). This is to ensure that $\mathbf{C}^{00}$ and $\mathbf{C}^{\tau\tau}$ are invertible.

Variational Approach for Markov processes (VAMP)

VAMP-2 Score

$$R_2(\mathbf{f}, \mathbf{X}, k) = \left\| \mathbf{C}_{\mathbf{f}}^{00}(\mathbf{X})^{-\frac{1}{2}} \; \mathbf{C}_{\mathbf{f}}^{0\tau}(\mathbf{X}) \; \mathbf{C}_{\mathbf{f}}^{\tau\tau}(\mathbf{X})^{-\frac{1}{2}} \right\|_{V(k)}$$

Where

$$\|\mathbf{A}\|_{V(k)}^2 = \sum_{i=1}^{k} \sigma_i^2$$

Wu and Noé, **arXiv:**1707.04659 (2017)

where $\mathbf{\Sigma} = (\sigma_1, ..., \sigma_n)$ are the singular values (in descending order) of the matrix decomposition

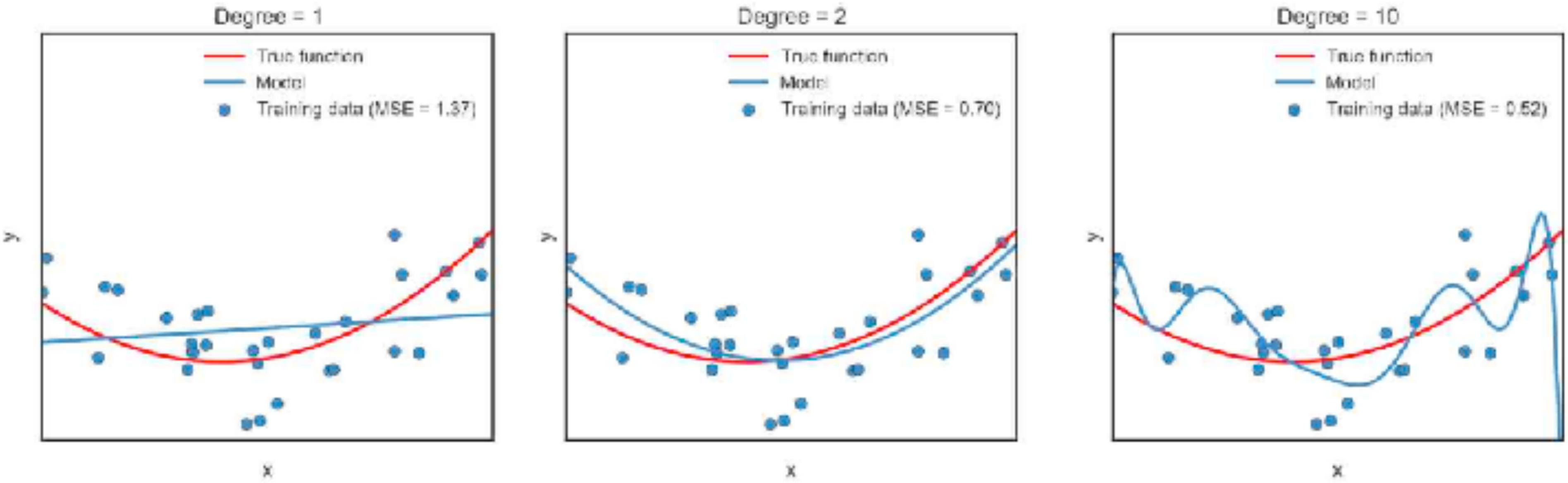
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

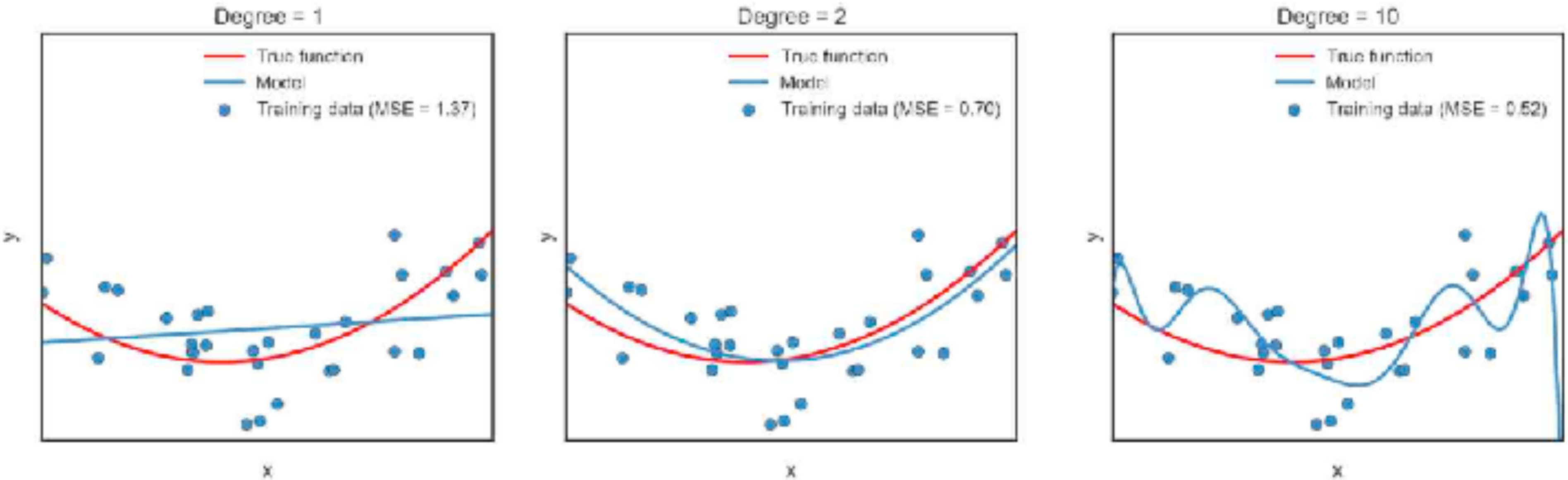
If dynamics are **reversible** (detailed balance), $\mathbf{C}_{\mathbf{f}}^{0\tau}$ is symmetric, $\mathbf{U} = \mathbf{V}$, the singular values are equal to the eigenvalues $\sigma_i = \lambda_i$, and the VAMP-2 norm is the kinetic distance:

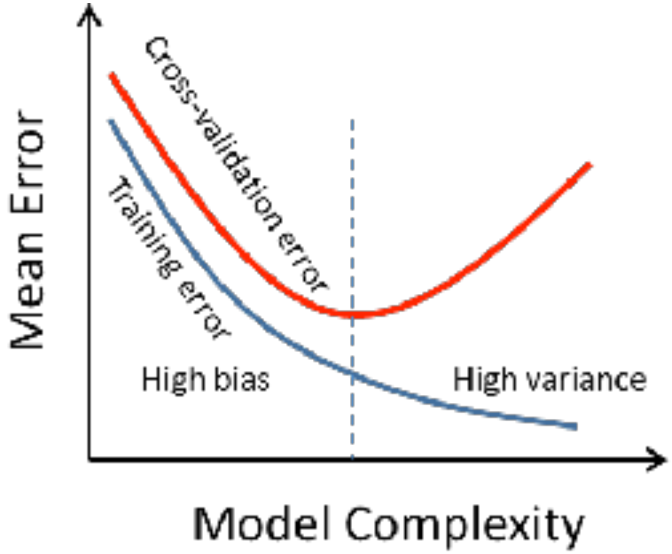
$$\|\mathbf{A}\|_{V(k)}^2 = \sum_{i=1}^{k} \lambda_i^2$$

Noé and Clementi, **JCTC** 11, 5002-5011 (2015)

Part II : Statistical validation




Degree = 1
True function
Model
Training data (MSE = 1.37)
y
x
Degree = 2
True function
Model
Training data (MSE = 0.70)
y
x
Degree = 10
True function
Model
Training data (MSE = 0.52)
y
x


Mean Error
Cross-validation error
Training error
High bias
High variance
Model Complexity

Cross-validation

Training:

Use data $\mathbf{X}_{\text{train}}$ and optimize $\mathbf{f}$:

$$\mathbf{f} = \arg\max_{\boldsymbol{f}} R_2(\boldsymbol{f}, \mathbf{X}_{\text{train}}, k)$$

The training score is then given by

$$R_2^{\text{train}}(\mathbf{f}, k) = R_2(\mathbf{f}, \mathbf{X}_{\text{train}}, k)$$

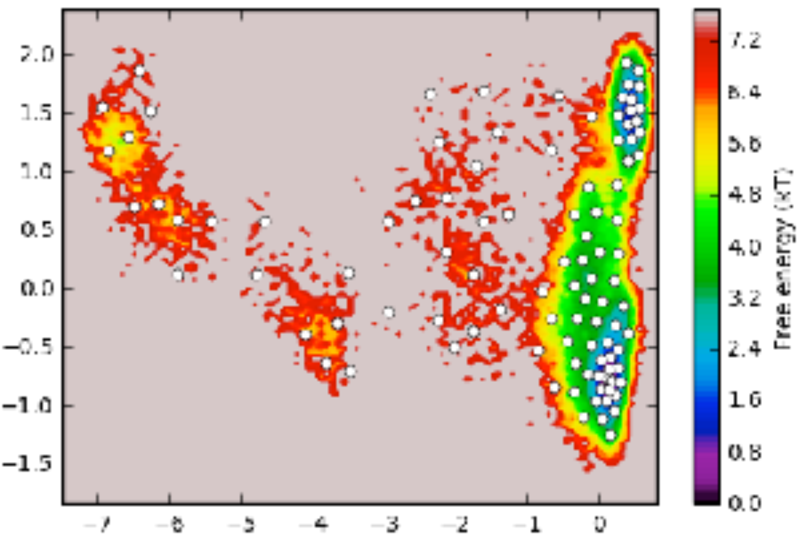
Validation:

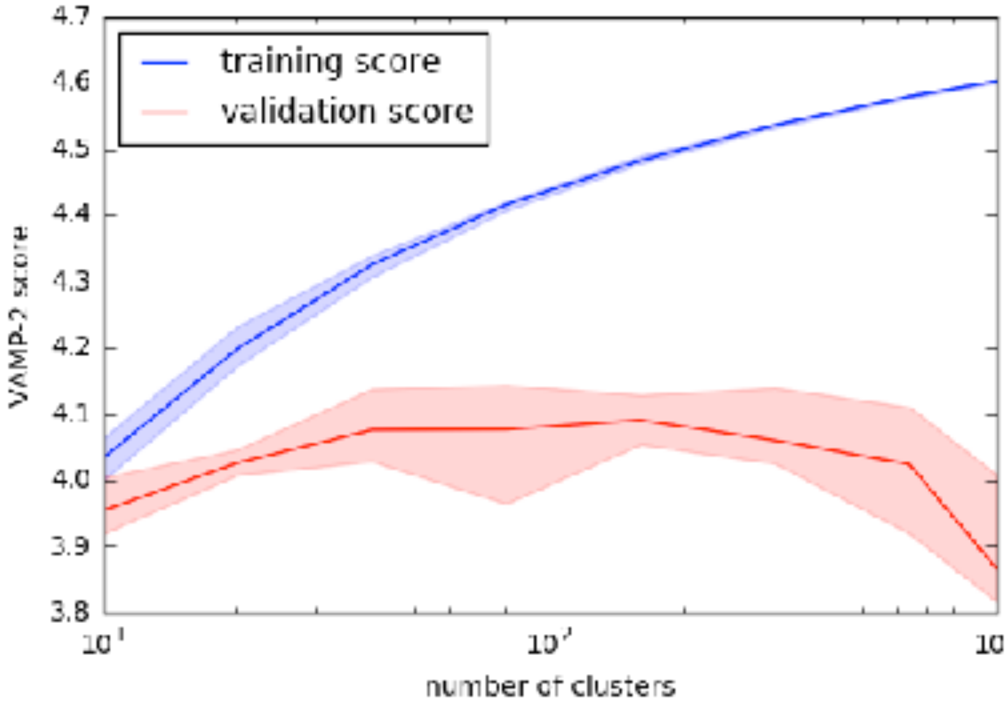
Use independent data $\mathbf{X}_{\text{val}}$ and compute validation score:

$$R_2^{\text{val}}(\mathbf{f}, k) = R_2(\mathbf{f}, \mathbf{X}_{\text{val}}, k)$$

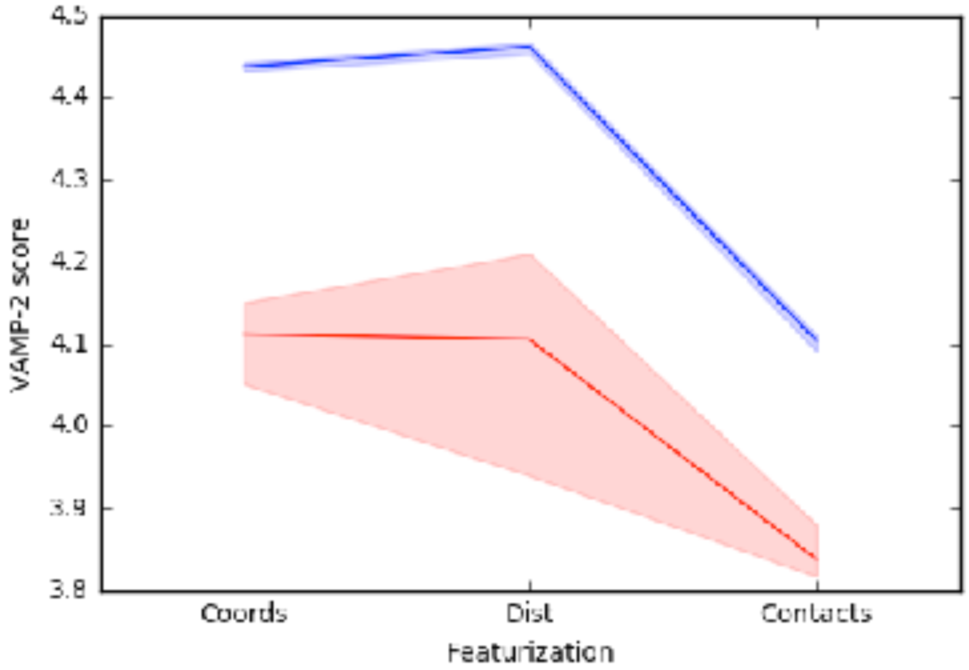
Note that we keep the transformation $\mathbf{f}$ learned in training, but we apply it on new data $\mathbf{X}_{\text{val}}$. $R_2^{\text{val}}(\mathbf{f}, k)$ can be used to score different models, e.g. compared different sets of functions.

How many states in BPTI?

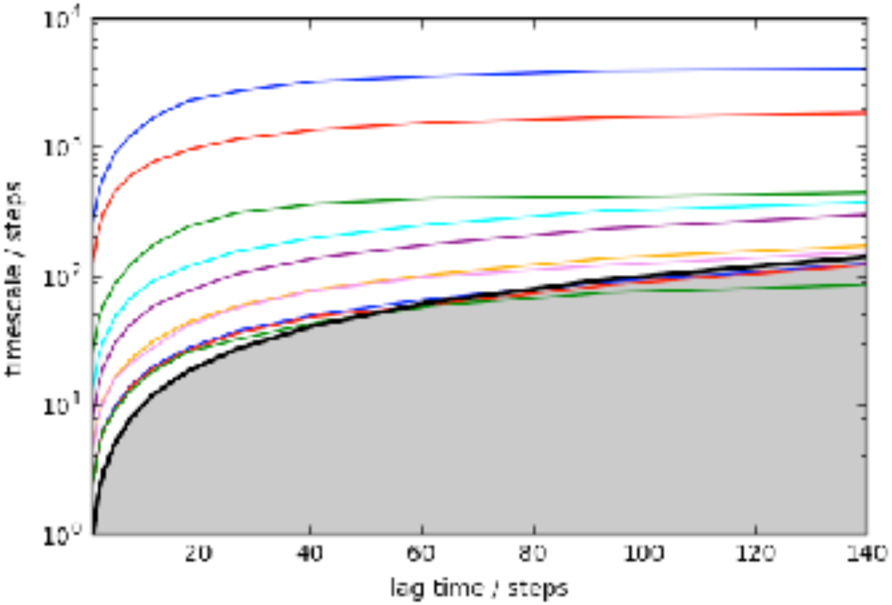




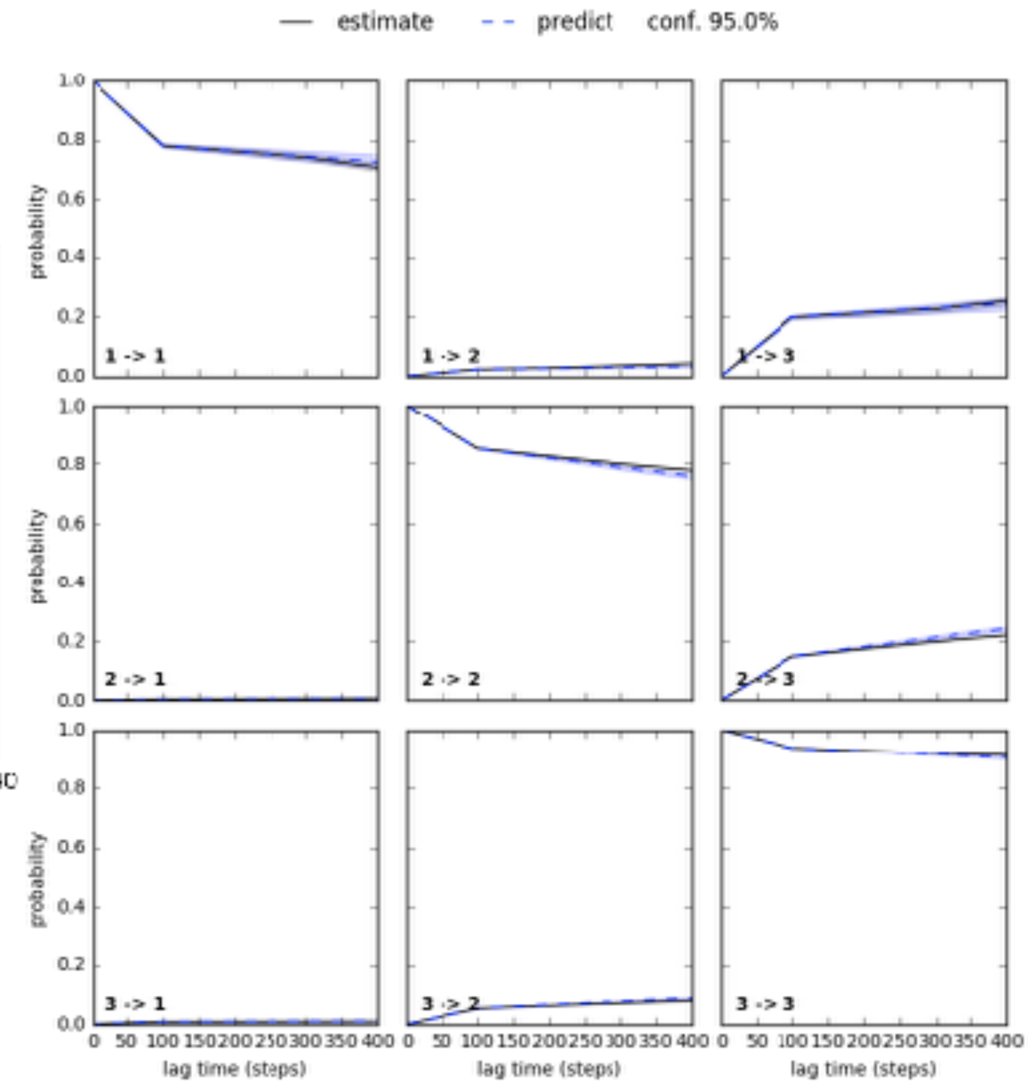
Which features for BPTI?



Validation

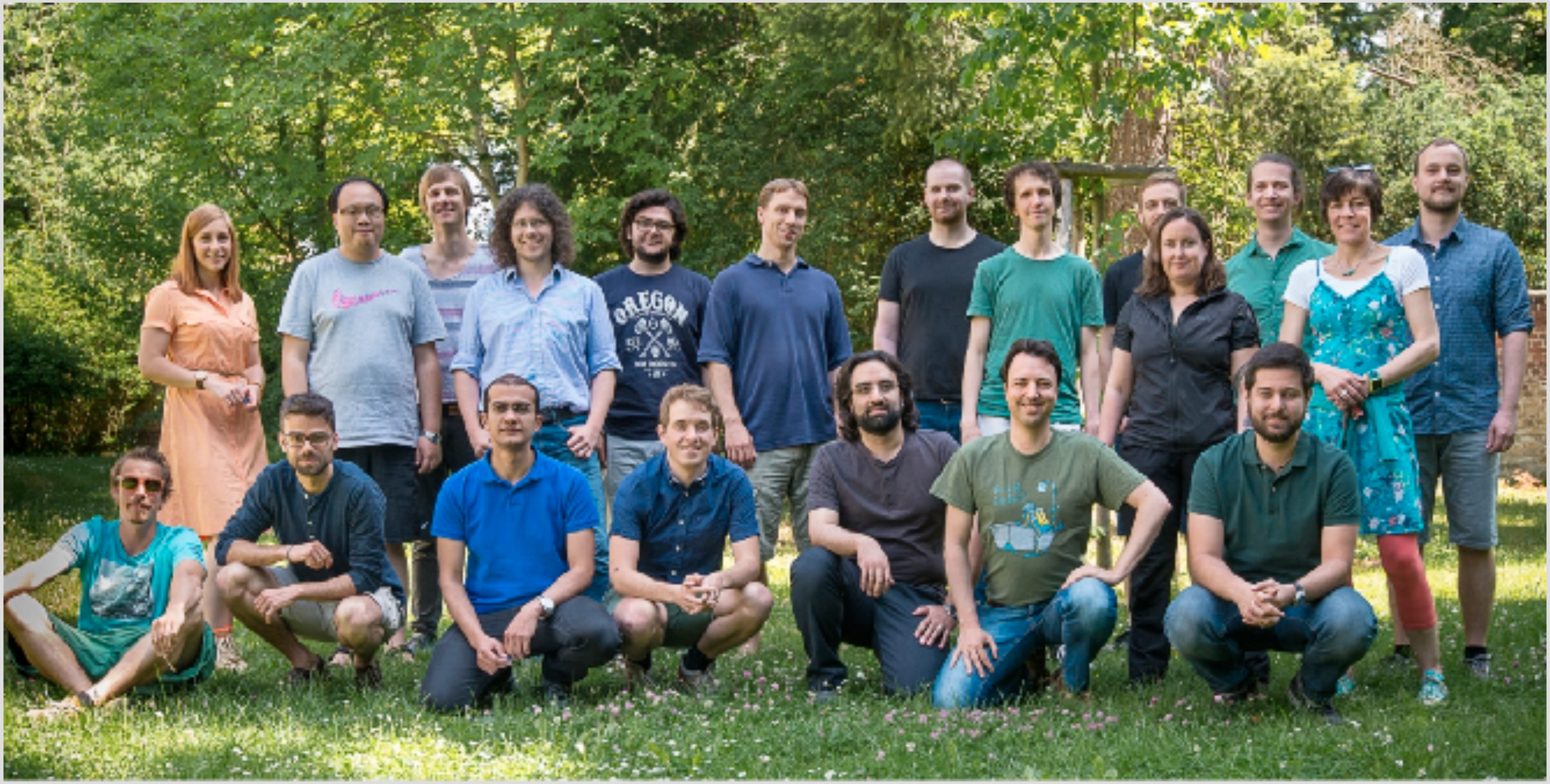


timescale / steps
20
40
60
80
100
120
140
lag time / steps



estimate
predict
conf. 95.0%
probability
1 -> 1
1 -> 2
1 -> 3
2 -> 1
2 -> 2
2 -> 3
3 -> 1
3 -> 2
3 -> 3
lag time (steps)

Acknowledgements



Collaborations

Cecilia Clementi (Rice University)
Christof Schütte (FU Berlin)
Eric Vanden-Eijnden (Courant NY)
Thomas Weikl (MPI Potsdam)
Edina Rosta (King's College London)
Bettina Keller (FU Berlin)
Vijay Pande (Stanford)
Volker Haucke (FMP Berlin)
Stephan Sigrist (FU Berlin)
Oliver Daumke (MDC)
John Chodera (MSKCC NY)
Gianni de Fabritiis (Barcelona)

Funding

European Research Council
Established by the European Commission
DFG
SFB 958
SFB 740
EINSTEIN Foundation.de