Fitting timeseries by continuous-time Markov chains: A quadratic programming approach

D.T. Crommelin, E. Vanden-Eijnden

Courant Institute of Mathematical Sciences, New York University
New York, USA

Abstract

Construction of stochastic models that describe the effective dynamics of observables of interest is an useful instrument in various fields of application, such as physics, climate science, and finance. We present a new technique for the construction of such models. From the timeseries of an observable, we construct a discrete-in-time Markov chain and calculate the eigenspectrum of its transition probability (or stochastic) matrix. As a next step we aim to find the generator of a continuous-time Markov chain whose eigenspectrum resembles the observed eigenspectrum as closely as possible, using an appropriate norm. The generator is found by solving a minimization problem: the norm is chosen such that the object function is quadratic and convex, so that the minimization problem can be solved using quadratic programming techniques. The technique is illustrated on various toy problems as well as on datasets stemming from simulations of molecular dynamics and of atmospheric flows.

Key words: Markov chains, embedding problem, inverse problems, timeseries analysis
PACS: 05.10.Gg, 02.50.Ga, 05.45.Tp, 36.20.Ey, 92.60.Bh, 02.30.Zz
1991 MSC: 60J27, 60J22, 62M10, 62M05, 49N45

1 Introduction

Inverse modeling is a powerful tool for the investigation of complex systems in various fields of application. If models derived from first principles are either very complicated or just non-existing, inverse modeling - constructing models

Email addresses: crommelin@cims.nyu.edu (D.T. Crommelin), eve2@cims.nyu.edu (E. Vanden-Eijnden).

from available data - can be an interesting alternative. A typical situation where inverse modeling can be very useful is the case when a complex system yields relatively simple macroscopic behavior that may be captured with low-order models, but where the derivation of such low-order models from first principles is very difficult or impossible. Such situations arise for example in molecular dynamics, econometrics, or climate science. In this study we focus on the construction of low-order stochastic models in the form of continuous-time Markov chains from timeseries.

To make things more precise and clarify the issues, let us start with a simpler problem. Consider a continuous-time Markov chain on a finite state-space $\mathcal{S}$ (for a general introduction to the theory of Markov chains, see [1] and [2]; continuous-time Markov chains are treated in detail in [3]). This chain is completely determined by its generator L, that is, the matrix with nonnegative off-diagonal elements, nonpositive diagonal elements and zero row sums ($\sum_y L(x,y) = 0 \ \forall x \in \mathcal{S}$) such that

$$\lim_{t \to 0+} \frac{\mathbb{E}_x f(X_t) - f(x)}{t} = \sum_{y \in \mathcal{S}} L(x,y)\, f(y), \tag{1}$$

for all suitable test functions $f : \mathcal{S} \to \mathbb{R}$; here $\mathbb{E}_x$ denotes the expectation conditional on $X_{t=0} = x$ and X_t denotes a sample path of the continuous-time Markov chain. Assume that the chain is ergodic and stationary, and denote by $\{X_t\}_{t \in \mathbb{R}}$ the equilibrium trajectory of the chain in a given realization. Given a sampling of X_t at discrete time-intervals, $\{X_{t_j}\}_{j \in \mathbb{Z}}$, with $t_j = jh$, $h > 0$, how can we reconstruct the generator L of the chain? Since h is fixed and may be rather large, (1) cannot be used. Yet, one can define the transition probability matrix $P^{(h)}$ whose elements $P^{(h)}(x,y)$ give the probability to have gone from state x to state y after a time-interval h:

$$P^{(h)}(x,y) = \lim_{N \to \infty} \frac{\sum_{j=-N}^{N} \mathbf{1}(X_{jh} = x) \mathbf{1}(X_{(j+1)h} = y)}{\sum_{j=-N}^{N} \mathbf{1}(X_{jh} = x)}, \tag{2}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, $\mathbf{1}(X_{jh} = x) = 1$ if $X_{jh} = x$ and $\mathbf{1}(X_{jh} = x) = 0$ otherwise. By ergodicity, this limit exists and is unique, and $P^{(h)}$ and L are related as

$$P^{(h)} = \exp(hL), \qquad L = h^{-1} \log P^{(h)}. \tag{3}$$

This relation offers a way to reconstruct L from $P^{(h)}$; $\log P^{(h)}$ can be computed by using e.g. the spectral decomposition of $P^{(h)}$ (see (7) below).

Unfortunately, the above procedure is not practical for the actual problems that we want to address in this paper in which:

(1) the discrete sampling is finite, i.e. we are only given $\{X_{t_j}\}_{j=0,\dots N}$ corresponding to $t_j \in [0, Nh]$ for some $N < \infty$;

(2) the underlying process X_t may not be Markov since, in a typical application, X_t is the timeseries of an observable which may display memory effects.

These features lead to the following practical difficulty. The matrix $\tilde{P}^{(h)}$ computed via

$$\tilde{P}^{(h)}(x,y) = \frac{\sum_{j=0}^{N-1} \mathbf{1}(X_{jh} = x)\mathbf{1}(X_{(j+1)h} = y)}{\sum_{j=0}^{N-1} \mathbf{1}(X_{jh} = x)}, \tag{4}$$

is, by construction, a stochastic matrix satisfying $\sum_y \tilde{P}^{(h)}(x,y) = 1 \ \forall x$ and $\tilde{P}^{(h)}(x,y) \geq 0 \ \forall x, y$. However, $L = h^{-1} \log \tilde{P}^{(h)}$ will, in general, have some negative or even complex off-diagonal elements and therefore will not be acceptable since it is not a generator of a continuous-time Markov chain. This issue is in fact related to the following famous (and open) *embedding problem* for Markov chains: not every discrete-in-time Markov chain (such as the one associated with $\tilde{P}^{(h)}$) has an underlying continuous-time chain, and the necessary and sufficient conditions for this to be the case are unknown. It is known what conditions a matrix must satisfy in order to be a generator (real, nonnegative off-diagonal elements; zero row sums); also, if L is a generator then all matrices $P^{(h)} = \exp(hL)$ with $h \geq 0$ are stochastic matrices. However, it is *not* known what exact conditions a stochastic matrix P must satisfy so that it can be written as $P = \exp(hL)$ with L a generator and $h \geq 0$ (i.e., the conditions for P to be embeddable). The subset of $n \times n$ embeddable matrices within the set of all $n \times n$ stochastic matrices has a very complicated geometrical structure (except if $n = 2$); in particular, it is non-convex. For more details on the embedding problem we refer to [4,5,6,7].

The main result of this paper is an efficient algorithm which gets around the difficulty posed by the embedding problem and permits to re-construct a true generator L from the observed $\tilde{P}^{(h)}$. This is done via the solution of a variational problem: we find the generator L such that $\exp(hL)$ is the closest to the measured stochastic matrix $\tilde{P}^{(h)}$ in the sense that their spectrum are the closest. If $h^{-1}\log(\tilde{P}^{(h)})$ is a generator (i.e. if $\tilde{P}^{(h)}$ is embeddable), then the procedure gives $L = h^{-1}\log(\tilde{P}^{(h)})$. If $h^{-1}\log(\tilde{P}^{(h)})$ is not a generator, then the procedure gives the true generator L which is the closest to $h^{-1}\log(\tilde{P}^{(h)})$ in the sense above. The resulting models give a much more faithful representation of the statistics and dynamics of the given timeseries than models constructed by only aiming to reproduce the equilibrium distribution and, possibly, a decorrelation time. Notice also that this construction is different from other procedures that have been proposed in the literature to go around the embedding problem. For instance, when a given $P^{(h)}$ does not have an exact underlying generator because $\log(P^{(h)})$ has negative off-diagonal elements, Israel *et al.* [6] propose to set the negative off-diagonal elements to zero and change the diagonal elements such that the condition of zero row sums is satisfied. The approximation method we describe in this paper is less ad-hoc

and, we believe, more attractive as it directly aims at reproducing key characteristics of the Markov chain like the leading modes of the eigenspectrum of the measured $\tilde{P}^{(h)}$. Bladt and Sørensen [7] use maximum likelihood estimation to find a generator given a timeseries. Their numerical procedure to find the maximum likelihood estimator seems efficient but rather costly; moreover, the estimator may not exist in various cases. The latter problem becomes particularly urgent if the sampling frequency is low (i.e., if h is large), or if the iterates of the estimator approach the boundary of the set of generators (which is likely to happen if the given data has no exact underlying generator). These are circumstances that typically show up in applications; the approach we present here is capable of dealing with them.

The remainder of this paper is organised as follows: in section 2 we recall the spectral decomposition of a generator L (section 2.1) and we give the variational problem used to determine the Markov chain generator with an eigenspectrum that matches the observed spectrum as closely as possible (section 2.2). This variational problem involves the minimization of a quadratic object function subject to linear constraints, i.e. it leads to a quadratic programming problem, for which there are well-established solution methods. The overall numerical procedure and its cost are presented in section 2.3, and some error estimates are given in section 2.4. In section 3, the algorithm is illustrated by using it to obtain approximate generators for toy problems with and without exact underlying generators. In sections 4 and 5 we apply the method to data from applications: one timeseries generated by a model simulating the dynamics of an alanine dipeptide molecule in vacuum (section 4), and another one obtained from a model that describes large-scale atmospheric flow (section 5). In both case, the data is non-Markovian at short time-intervals; we show how to obtain good results by using information from longer time intevals. Concluding remarks and possible generalizations are given in section 6.

2 A quadratic programming approach

2.1 Preliminaries: Spectral representation of the generator

Assume that the number of states in $\mathcal{S}$ is n, and let $\{\psi_k, \phi_k, \lambda_k\}_{k=1,\ldots,n}$ with $\psi_k = (\psi_k(1), \ldots, \psi_k(n))$ and $\phi_k = (\phi_k(1), \ldots, \phi_k(n))^T$, be the complete biorthogonal set of eigenmodes and eigenvalues of a generator L,

$$L\phi_k = \lambda_k \phi_k, \qquad \psi_k L = \lambda_k \psi_k, \qquad \psi_k \phi_l = \delta_{kl}, \tag{5}$$

ordered such that $\operatorname{Re}\lambda_k \geq \operatorname{Re}\lambda_{k+1}$ for all k. The first eigenmode ($k = 1$) has special properties, since it contains the invariant distribution $\mu(x)$ (assumed

to be unique): $\psi_1 = \mu$, $\phi_1 = (1, \ldots, 1)^T$ and $\lambda_1 = 0$. Assuming that all eigenvalues have multiplicity one, L can be represented as

$$L(x, y) = \sum_{k=1}^{n} \lambda_k \, \phi_k(x) \, \psi_k(y) \tag{6}$$

The spectral representation of L allows one to easily compute the exponential of L. Indeed, if $P^{(h)} = \exp(hL)$, then $P^{(h)}$ can be represented as

$$P^{(h)}(x, y) = \sum_{k=1}^{n} \Lambda_k \, \phi_k(x) \, \psi_k(y) \tag{7}$$

where $\Lambda_k = \exp(\lambda_k h)$. Conversely, if $\{\psi_k, \phi_k, \Lambda_k\}_{k=1,\ldots,n}$ is the complete bi-orthogonal set of eigenmodes and eigenvalues of $P^{(h)}$ so that (7) holds, then $L = h^{-1} \log P^{(h)}$ can be represented as in (6) with $\lambda_k = h^{-1} \log \Lambda_k$.

2.2 A quadratic variational problem

Suppose that we have constructed from (4) the stochastic matrix $\tilde{P}^{(h)}$ associated with the finite sampling $\{X_{t_j}\}_{j=0,\ldots,N}$, $t_j = jh$, $h > 0$, $N < \infty$. As explained in the introduction, the stochastic matrix $\tilde{P}^{(h)}$ may have no true generator associated with it. This means that if $\{\tilde{\psi}_k, \tilde{\phi}_k, \tilde{\Lambda}_k\}$ is the complete bi-orthogonal set of eigenmodes and eigenvalues of $\tilde{P}^{(h)}$ so that $\tilde{P}^{(h)}$ can be represented as

$$\tilde{P}^{(h)}(x, y) = \sum_{k=1}^{n} \tilde{\Lambda}_k \, \tilde{\phi}_k(x) \, \tilde{\psi}_k(y), \tag{8}$$

the matrix

$$\tilde{L}(x, y) = \sum_{k=1}^{n} \tilde{\lambda}_k \, \tilde{\phi}_k(x) \, \tilde{\psi}_k(y) \qquad \text{where } \tilde{\lambda}_k = h^{-1} \log \tilde{\Lambda}_k \tag{9}$$

will in general have negative or even complex off-diagonal element and will therefore not qualify as a generator of a continuous-time Markov chain.

To get around this difficulty, we propose to find a true generator $L(x, y)$ optimal with respect to the data at hand by minimizing the following object function under variation of L:

$$\sum_{k=1}^{n} \left(\alpha_k |\tilde{\psi}_k L - \tilde{\lambda}_k \tilde{\psi}_k|^2 + \beta_k |L\tilde{\phi}_k - \tilde{\lambda}_k \tilde{\phi}_k|^2 + \gamma_k |\tilde{\psi}_k L \tilde{\phi}_k - \tilde{\lambda}_k|^2 \right) \tag{10}$$

subject to the constraints

$$L(x, x) = - \sum_{\substack{y \in \mathcal{S} \\ y \neq x}} L(x, y) \qquad \forall \; x \in \mathcal{S}. \tag{11}$$

and

$$L(x,y) \geq 0 \qquad \forall\, x,y \in \mathcal{S} \;\text{ with }\; x \neq y. \tag{12}$$

In (10) α_k, β_k, γ_k are weights, typically of the form $\alpha_k = \tilde{\alpha}_k\,|\tilde{\lambda}_k\tilde{\psi}_k|^{-2}$, $\beta_k = \tilde{\beta}_k\,|\tilde{\lambda}_k\tilde{\phi}_k|^{-2}$, $\gamma_k = \tilde{\gamma}_k\,|\tilde{\lambda}_k|^{-2}$. If we then pick $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1\ \forall\, k$, relative errors of the same order in the eigenvectors and eigenvalues of L will give contributions of similar magnitude in E.

The constraint (11) can be straightforwardly accounted for explicitly in (10), and one is then left with a quadratic functional which can be compactly written as:

$$E(L) = \tfrac{1}{2}\langle L, HL\rangle + \langle F, L\rangle + E_0 \tag{13}$$

Here

$$\begin{aligned}\langle L, HL\rangle &= \sum_{\substack{x,y,x',y'\in\mathcal{S}\\ x\neq y, x'\neq y'}} L(x,y)H(x,y,x',y')L(x',y')\\ \langle F, L\rangle &= \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} F(x,y)L(x,y)\end{aligned} \tag{14}$$

and we defined

$$\begin{aligned}&H(x,y,x',y')\\ &\quad= 2\sum_{k=1}^{n}\Big(\alpha_k\left(\delta(x,x')+\delta(y,y')-\delta(x',y)-\delta(x,y')\right)\tilde{\psi}_k(x)\bar{\tilde{\psi}}_k(x')\\ &\qquad+\beta_k\,\delta(x,x')\big(\tilde{\phi}_k(x)-\tilde{\phi}_k(y)\big)\big(\bar{\tilde{\phi}}_k(x')-\bar{\tilde{\phi}}_k(y')\big)\\ &\qquad+\gamma_k\,\tilde{\psi}_k(x)\bar{\tilde{\psi}}_k(x')\left(\bar{\tilde{\phi}}_k(x')-\bar{\tilde{\phi}}_k(y')\right)\Big)\end{aligned} \tag{15}$$

$$\begin{aligned}F(x,y) = \sum_{k=1}^{n}\Big(&\alpha_k\,\tilde{\lambda}_k\left(\tilde{\psi}_k(x)\bar{\tilde{\psi}}_k(x)-\tilde{\psi}_k(y)\bar{\tilde{\psi}}_k(x)\right)\\ &+\beta_k\,\tilde{\lambda}_k\,\phi_k(x)\left(\bar{\phi}_k(x)-\bar{\phi}_k(y)\right)\\ &+\gamma_k\,\tilde{\lambda}_k\,\bar{\tilde{\psi}}_k(x)\big(\bar{\tilde{\phi}}_k(x)-\bar{\tilde{\phi}}_k(y)\big)\\ &+\ \text{complex conjugate}\ \Big)\end{aligned} \tag{16}$$

$$E_0 = \sum_{k=1}^{n}\tilde{\lambda}_k\bar{\tilde{\lambda}}_k\Big(\sum_x\left(\alpha_k\,\tilde{\psi}_k(x)\bar{\tilde{\psi}}_k(x)+\beta_k\,\tilde{\phi}_k(x)\bar{\tilde{\phi}}_k(x)+\gamma_k\right)\Big), \tag{17}$$

where the bar denotes complex conjugate. (13) is a quadratic object function to be minimized over all $L(x,y)$ with $x,y\in\mathcal{S}$ and $x\neq y$ subject to (12). Since there are n^2-n off-diagonal elements in L, E should be thought of as a function on $\mathbb{R}^{n^2-n}$. E is also convex since it is straightforward to check that the level sets of E are ellipsoids in $\mathbb{R}^{n^2-n}$ centered around the point in $\mathbb{R}^{n^2-n}$ associated with the off-diagonal elements of $\tilde{L}$ given in (9). Thus $\tilde{L}$ is the absolute minimizer of (13) with $E(\tilde{L})=0$. $\tilde{L}$ is also the minimizer of (13) subject to (12) if $\tilde{P}^{(h)}$ is embeddable, since in this case $\tilde{L}$ satisfies (12). If $\tilde{P}^{(h)}$ is not embeddable, $\tilde{L}$ is not the minimum of E subject to (12), but the minimization problem still

has a unique solution because because the domain for L defined by (12) is a convex domain in $\mathbb{R}^{n^2-n}$. The corresponding minimizer $L_{\min}$ is the unique true generator "closest" to $\tilde{L}$.

One can use other object functions than (10) to formulate the search for an optimal generator as a minimization problem. In particular, it is easy to show that the minimizer of

$$E' = \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} |L(x,y) - \tilde{L}(x,y)|^2, \tag{18}$$

over all matrices L such that $L(x,y) \geq 0$ if $x \neq y$ must be

$$L(x,y) = \max(\text{Re } \tilde{L}(x,y), 0) \qquad \text{if } x \neq y, \tag{19}$$

which can be supplemented by $L(x,x) = -\sum_{y\neq x} L(x,y)$ to obtain a generator. The disadvantage of using (18), however, is that the information on the eigenvectors and eigenvalues of $\tilde{L}$ enters only very indirectly (18), which means that the spectrum of (19) may be rather different from those of $\tilde{L}$. In contrast, (10) precisely aim at reproducing the spectrum of $\tilde{L}$ as closely as possible, which is important since this spectrum embeds the most important features of the dynamics of X_t.

2.3 Numerical procedure and computational cost

The minimization of (13) subject to (12) defines a standard quadratic problem which can be solved via well-established numerical methods (see e.g. [8], [9]). For this study we use the internal quadratic programming algorithm of Matlab; various other software packages are also available to solve this type of problems. As input for the Matlab algorithm the matrix H and the vectors F are needed, as well as an initial guess for L.

As mentioned earlier, for a Markov chain with n states, the minimization problem is of dimension $n^2 - n$ and has $n^2 - n$ inequality constraints, see (12). The computational cost of a quadratic programming algorithm may become prohibitive when the number of states in the chain is large. One particularly interesting way to reduce the size of the minimization problem, and thereby the computational cost, is to restrict the class of allowed Markov chain generators. For example, when considering a certain system there may be physical grounds for allowing only non-zero transition rates from one state to a selected few other states. This will be the case e.g. if the state-space $\mathcal{S}$ inherits some topology from the physical space where the actual dynamics takes place (as will be the case in the examples treated in sections 4 and 5). If for instance, each state is connected to $m < n$ other states only, yielding a mn-dimensional

minimization problem. We will discuss this possibility in more detail in section 6.

2.4 Error estimates

The computational strategy that we propose has two sources of errors as far as the fitting of the observed timeseries by the continuous-time Markov chain is concerned. The first stems from the fact that the timeseries may not be Markov. In this case, the Markov assumption itself is a source of error. This error may dominate in many applications, but it is hard to quantify sytematically. Therefore we shall not dwell on this issue in this section and postpone its discussion till sections 4 and 5 where we investigate it via numerical experiments and show how our computational strategy allows to overcome errors due to non-Markovianity of the timeseries.

The second source of error is that, even if the observed timeseries is Markov and the exact matrix $P^{(h)}$ given by (2) is embeddable, in general $\tilde{P}^{(h)} \neq P^{(h)}$ due to finite sampling. As a result the observed spectrum will be different from the actual spectrum of the chain. This is the error that we quantify in this section.

2.4.1 Central limit theorem and error estimate on $\tilde{L} = h^{-1} \log \tilde{P}^{(h)}$

Let $P^{(h)} = \exp(hL)$ be the true transition probability matrix underlying the timeseries from which we have constructed $\tilde{P}^{(h)}$ according to (4). Assume that the chain has been sampled at $N = \lfloor T/h \rfloor$ successive points with uniform time interval h, consistent with the process being observed on a fixed window of time $T > 0$ independent of the lag h. The error on $\tilde{P}^{(h)}$ obeys a version of the central limit theorem, see [10]: as $T \to \infty$,

$$\sqrt{T}\Big(\tilde{P}^{(h)} - P^{(h)}\Big) \to \sqrt{h}\, Q^{(h)}, \tag{20}$$

in probability, where $Q^{(h)}$ is a Gaussian matrix with mean zero and covariance

$$\mathbb{E}\, Q^{(h)}(x,y) Q^{(h)}(x',y') = \frac{P^{(h)}(x,y)}{\mu(x)} \Big(\delta(y,y') - P^{(h)}(x,y')\Big)\delta(x,x'). \tag{21}$$

For $\tilde{L} = h^{-1} \log \tilde{P}^{(h)}$ we have

$$\begin{aligned} \tilde{L} - L &= h^{-1} \log\Big[1 + \Big(P^{(h)}\Big)^{-1}\Big(\tilde{P}^{(h)} - P^{(h)}\Big)\Big] \\ &\approx h^{-1} \Big(P^{(h)}\Big)^{-1}\Big(\tilde{P}^{(h)} - P^{(h)}\Big) \end{aligned} \tag{22}$$

since $\tilde{P}^{(h)} \approx P^{(h)}$ as T becomes large. This implies that as $T \to \infty$,

$$\sqrt{T}\big(\tilde{L} - L\big) \to \frac{1}{\sqrt{h}} \left(P^{(h)}\right)^{-1} Q^{(h)}, \tag{23}$$

where we have used (20), and $Q^{(h)}$ is the same matrix as above.

2.4.2 Error estimates on the eigenspectrum

Using standard matrix perturbation theory one can use (20) to derive the following convergence estimates on the eigenspectrum $\{\tilde{\psi}_k, \tilde{\phi}_k, \tilde{\lambda}_k\}$ of $\tilde{L} = h^{-1} \log \tilde{P}^{(h)}$: as $T \to \infty$,

$$\begin{cases} \sqrt{T}\big(\tilde{\lambda}_k - \lambda_k\big) \to \dfrac{1}{\sqrt{h}}\, e^{-\lambda_k h} \psi_k Q^{(h)} \phi_k, \\ \sqrt{T}\big(\tilde{\psi}_k - \psi_k\big) \to \sqrt{h}\, \psi_k R_k^{(h)} Q^{(h)} \\ \sqrt{T}\big(\tilde{\phi}_k - \phi_k\big) \to \sqrt{h}\, R_k^{(h)} Q^{(h)} \phi_k \end{cases} \tag{24}$$

in probability. Here $\{\psi_k, \phi_k, \lambda_k\}$ is the set of eigenvectors and eigenvalues of L, $Q^{(h)}$ is the Gaussian matrix defined before, and

$$R_k^{(h)}(x,y) = \sum_{q \neq k} \left(e^{\lambda_k h} - e^{\lambda_q h}\right)^{-1} \phi_q(x) \psi_q(y).$$

For $k = 1$ a stronger estimate can be derived, since by construction $\tilde{\lambda}_1 = 0$ and $\tilde{\phi}_1 = (1, \ldots, 1)^T$ (i.e., the errors on $\tilde{\lambda}_1$ and $\tilde{\phi}_1$ are always zero). For $\tilde{\psi}_1$ we have, as $T \to \infty$,

$$\sqrt{T}\left(\tilde{\psi}_1 - \mu\right) \to \sqrt{h}\, b^{(h)} \tag{25}$$

in probability, where $b^{(h)}$ is a Gaussian vector with mean zero and covariance

$$\begin{aligned} \mathbb{E}\, b^{(h)}(x) b^{(h)}(y) &= \mu(x)\big(\delta(x,y) - \mu(y)\big) \\ &\quad + \sum_{q \neq 1} \frac{e^{\lambda_q h}}{1 - e^{\lambda_q h}} \big(\psi_q(x)\phi_q(y)\mu(y) + \psi_q(y)\phi_q(x)\mu(x)\big) \end{aligned} \tag{26}$$

3 Toy examples

3.1 A toy example with exact generator

As a first test, we generate a timeseries for a simple 4-state Markov chain that has an exact underlying generator, and use the timeseries to reconstruct the

generator. This allows us to assess the convergence of the algorithm with the length of the timeseries. We choose the true generator to be

$$L_{\text{exact}} = \begin{pmatrix} -0.6 & 0.4 & 0.2 & 0 \\ 0.5 & -1.2 & 0.4 & 0.3 \\ 0.3 & 0 & -0.6 & 0.3 \\ 0 & 0.1 & 0.2 & -0.3 \end{pmatrix} \tag{27}$$

A sample path is generated from this generator and sampled at timelag $h = 1$. The eigenvalues of L_{exact} are 0, -0.40, -0.89, -1.42, so $h = 1$ is of the order of the characteristic timescale of the system. We use timeseries with 10^n data points, with $n = 3, \ldots, 7$. Thus, the stochastic matrix $\tilde{P}$ is calculated from 10^n data points, and the eigenspectrum of $\tilde{P}$ is used in the object function (10). For all calculations, the object function coefficients are set to $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all $k = 1, 2, 3, 4$.

The generators obtained by minimizing (10) subject to (12) are denoted by $L_{\text{min}}^{10^n}$; the generators constructed from the same 10^n data points according to (9) are denoted by $\tilde{L}^{10^n}$. From the $\tilde{L}^{10^n}$ we also construct $\tilde{L}_g^{10^n}$ using (19). Thus the $L_{\text{min}}^{10^n}$ and $\tilde{L}_g^{10^n}$ always are true generator, whereas the $\tilde{L}^{10^n}$ may not (they usually have one or more negative off-diagonal elements).

In figures 1 and 2 we show two different error measures for the various matrices. Figure 1 is a graph of the value of E, the proposed object function, using the eigenspectrum of L_{exact} as reference spectrum. This error measure indicates how well the eigenspectrum of the exact underlying generator (L_{exact}) is recovered by the various approximations of the generator. In figure 2 a different error measure was used: it shows the distance E' from L_{exact} using the norm (18). Both figures show the convergence to L_{exact} of the approximations with increasing length of timeseries. Increasing the length by a factor 10 reduces both error measures with about a factor 10.

As an example, for $n = 4$ the actual matrices are as follows:

$$L_{\text{min}}^{10^4} = \begin{pmatrix} -0.612 & 0.416 & 0.196 & 0.000 \\ 0.533 & -1.278 & 0.441 & 0.304 \\ 0.309 & 0.002 & -0.623 & 0.312 \\ 0.009 & 0.079 & 0.195 & -0.283 \end{pmatrix} \tag{28a}$$

$$\tilde{L}^{10^4} = \begin{pmatrix} -0.612 & 0.417 & 0.198 & -0.002 \\ 0.534 & -1.279 & 0.441 & 0.304 \\ 0.310 & 0.001 & -0.624 & 0.313 \\ 0.009 & 0.079 & 0.195 & -0.284 \end{pmatrix} \tag{28b}$$

$$\tilde{L}_g^{10^4} = \begin{pmatrix} -0.615 & 0.417 & 0.198 & 0.000 \\ 0.534 & -1.279 & 0.441 & 0.304 \\ 0.310 & 0.001 & -0.624 & 0.313 \\ 0.009 & 0.079 & 0.195 & -0.284 \end{pmatrix} \tag{28c}$$

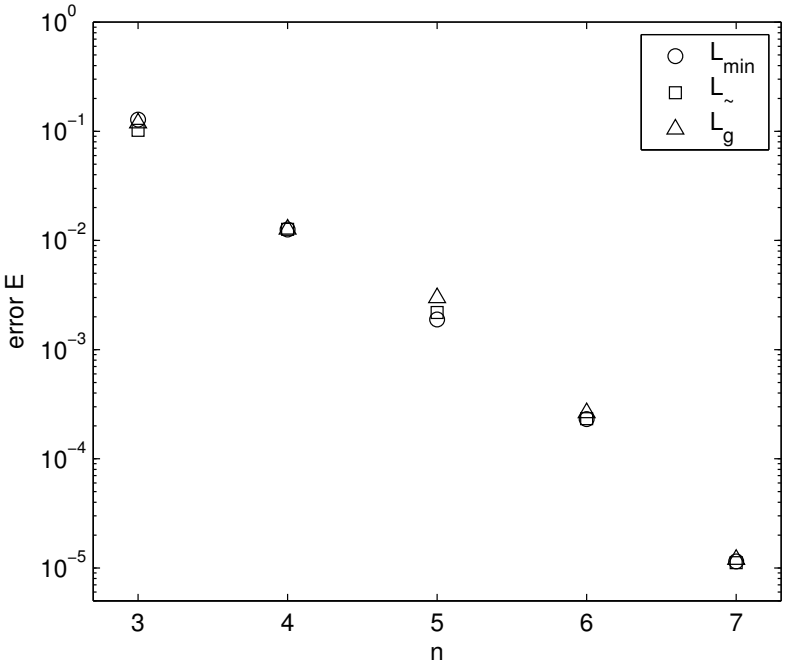

Figure 1. Errors of the approximations of L_{exact}. Shown are the values of E, the proposed object function, using the eigenspectrum of L_{exact} as reference spectrum. This indicates how well the eigenspectrum of L_{exact} is recovered by the approximations. The length of the timeseries used to find the approximations is 10^n points, with n on the horizontal axis.

Notice that in this first toy example, the procedure that we propose to construct $L_{\min}$ by constrained minimization of (13) does not produce results that are significantly better than using the generator in (19). This, however, is mainly due to the fact that the timeseries is Markov and has an underlying generator. The full power of our procedure will become more apparent in situations where the timeseries has no underlying generator, such as the ones we consider in the next section as well as in sections 4 and 5.

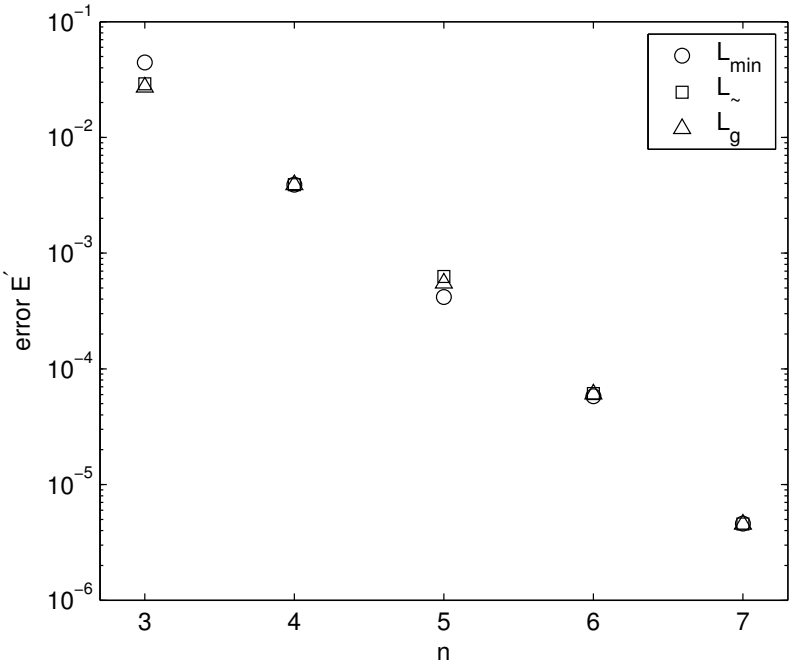

Figure 2. Errors of the approximations of L_{exact}. Shown are the values of E', the distance to L_{exact} using the norm (18). The length of the timeseries used to find the approximations is 10^n points, with n on the horizontal axis.

In this section we illustrate the algorithm by using it on another simple example, but this time without underlying generator (i.e. for a $P^{(h)}$ that is not embeddable). Specifically, we take

$$P^{(h)}(x,y) = \begin{cases} \dfrac{c_x}{|x-y|} & \text{if } \; x \neq y \\ c_x & \text{otherwise} \end{cases} \tag{29}$$

where the c_x are normalisation constants such that $\sum_y P^{(h)}(x,y) = 1 \;\forall\, x$ (thereby ensuring that $P^{(h)}$ is a stochastic matrix). We choose the state-space to have 10 states: $x, y \in \{1, 2, \ldots, 10\}$. The value of the lag h is irrelevant here; we set it to 1 (taking another value would only correspond to a time rescaling of the Markov chain). The eigenvalues $\tilde{\Lambda}_k$ of $P^{(h)}$ are all real; four are on the negative real axis. Therefore $P^{(h)}$ has no exact underlying generator. For the reference eigenvalues $\tilde{\lambda}_k$ we take $\log|\tilde{\Lambda}_k|$ if $\tilde{\Lambda}_k$ is negative and real; $\log\tilde{\Lambda}_k$ otherwise.

Two different sets of weights were used for this calculation. As the first set we take $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all k. Thus, there is no extra emphasis on the invariant distribution ($k = 1$) or the next leading eigenmodes. The results of the minimization are presented in the form of comparisons between the eigenmodes $\{\psi_k, \phi_k, \lambda_k\}$ of the generator $L_{\min}$ that came out of the minimization, and the reference set $\{\tilde{\psi}_k, \tilde{\phi}_k, \tilde{\lambda}_k\}$. The leading four ψ_k and $\tilde{\psi}_k$, as well as the eigenvalues λ_k and $\tilde{\lambda}_k$ are shown in figures 3 and 4. To quantify the difference between the two sets eigenmodes, the relative errors $|\psi_k - \tilde{\psi}_k|/|\tilde{\psi}_k|$, $|\phi_k - \tilde{\phi}_k|/|\tilde{\phi}_k|$ and $|\lambda_k - \tilde{\lambda}_k|/|\tilde{\lambda}_k|$ are shown in figure 5 (the errors of ϕ_1 and λ_1 are zero by construction and are therefore not shown).

The second set of weights is equal to the previous set, except that $\alpha_1, \alpha_2, \alpha_3$, β_2, β_3, γ_2 and γ_3 all have been multiplied by 100. Thus, errors in the leading three eigenmodes (including the invariant distribution) are penalized more heavily than previously. As a result, the relative errors for those modes (both eigenvectors and eigenvalues) are much smaller than in the previous calculation, see figure 6. In a figure, they are indistinguishable by eye from the leading three reference eigenmodes (figure not shown).

The results show that the algorithm works very well for finding a generator matrix whose eigenmodes resemble the reference eigenmodes closely, even though the stochastic matrix (29) has no exact underlying generator. By increasing the weight on the leading eigenmodes the match between the reproduced and reference leading eigenmodes becomes very good.

Constructing a generator using (19) gives bad results in this case. The invariant

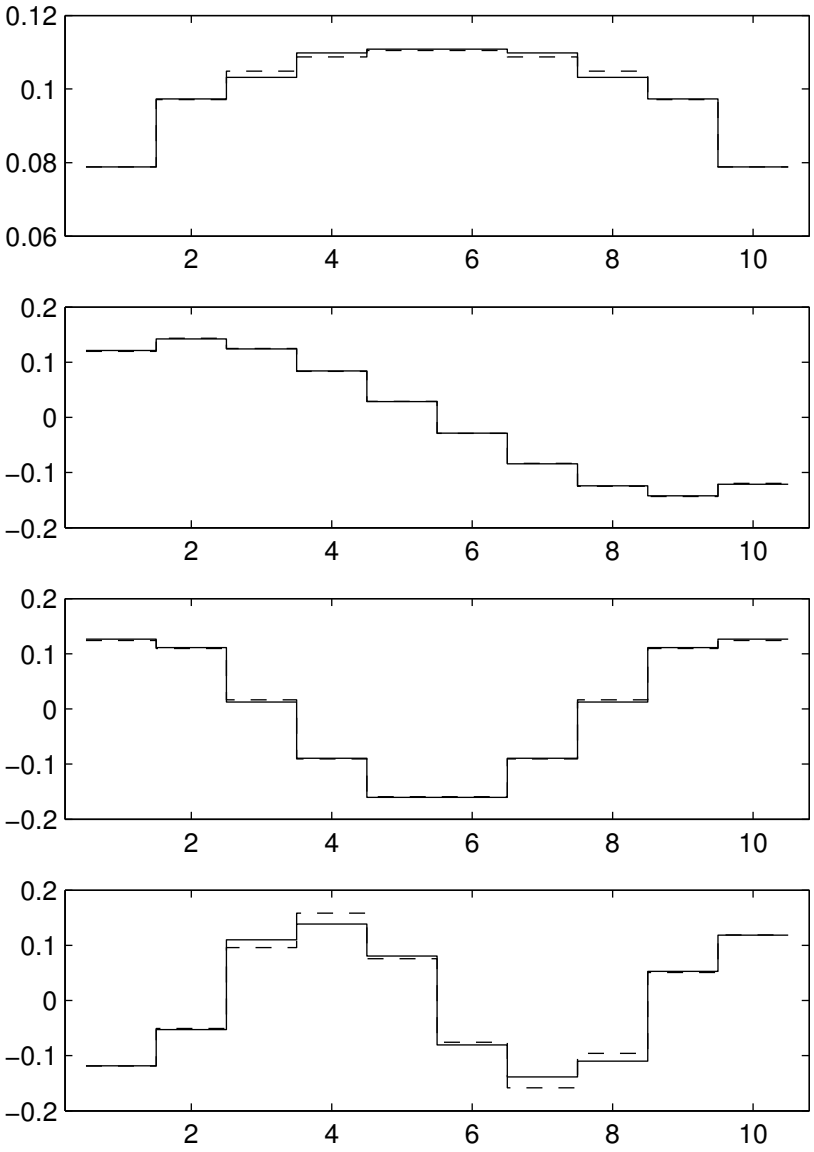

Figure 3. Results of the Markov chain toy example without exact underlying generator, using weights $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all k. Shown are the leading four eigenvectors ψ_k of optimal generator $L_{\min}$ (solid) and reference vectors $\tilde{\psi}_k$ (dashed). Top to bottom: $k = 1, 2, 3, 4$.

distribution, ψ_1, is well captured but the eigenvectors and eigenvalues with $k > 1$ have big errors. For example, the leading reference eigenvalues are $\tilde{\lambda}_2 = -0.60, \tilde{\lambda}_3 = -1.08$; the generator obtained from (19) gives $\lambda_2 = -1.53, \lambda_3 = -2.37$. By contrast, the generator obtained with our proposed object function has relative errors for λ_2 and λ_3 of about 10^{-2} using the first set of weights (figure 5) and about 10^{-4} using the second set of weights (figure 6).

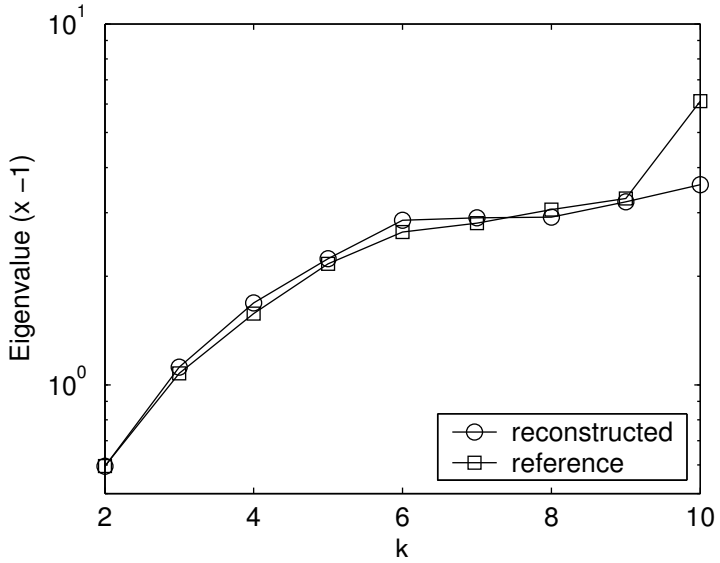

Figure 4. Results of the Markov chain toy example without exact underlying generator, using weights $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all k. Shown are the eigenvalues (real part, multiplied by –1) of $L_{\min}$ and reference eigenvalues.

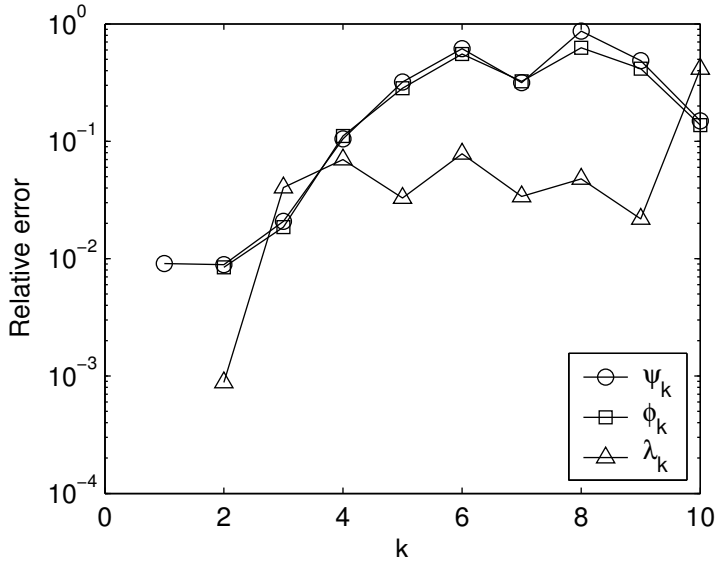

Figure 5. Relative errors $|\psi_k - \tilde{\psi}_k|/|\tilde{\psi}_k|$, $|\phi_k - \tilde{\phi}_k|/|\tilde{\phi}_k|$ and $|\lambda_k - \tilde{\lambda}_k|/|\tilde{\lambda}_k$ of the Markov chain toy example (without exact underlying generator) with weights $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all k. By construction, the errors on $\tilde{\lambda}_1$ and $\tilde{\phi}_1$ are zero, and are therefore not shown.

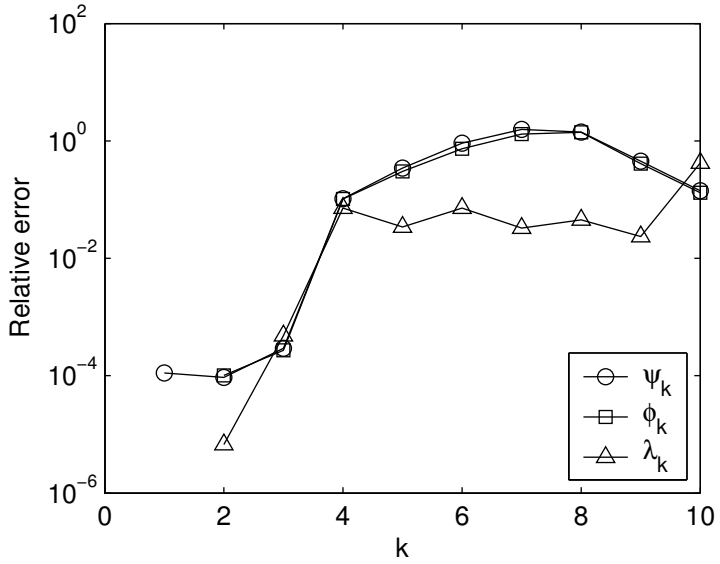

Figure 6. Relative errors $|\psi_k - \tilde{\psi}_k|/|\tilde{\psi}_k|$, $|\phi_k - \tilde{\phi}_k|/|\tilde{\phi}_k|$ and $|\lambda_k - \tilde{\lambda}_k|/|\tilde{\lambda}_k$ of the Markov chain toy example (without exact underlying generator) with extra weight on the leading eigenmodes (see text). By construction, the errors on $\tilde{\lambda}_1$ and $\tilde{\phi}_1$ are zero, and are therefore not shown.

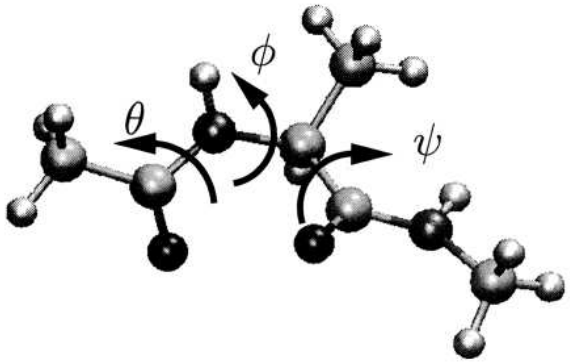

Figure 7. Schematic representation of the alanine dipeptide (CH_3-CONH-$CHCH_3$-CONH-CH_3). The backbone dihedral angles are labeled by ϕ (Φ in text): C-N-C-C and ψ (Φ in text): N-C-C-N. The picture is taken from [14].

4 Application to a timeseries from molecular dynamics

In this section, we test our numerical procedure on data from a numerical simulation of the alanine dipeptide molecule. The ball and stick model of alanine dipeptide is shown in figure 4: the molecule has backbone degree of freedoms (dihedral angles Φ and Ψ), three methyl groups (CH_3), as well as polar groups (N-H, C=O). The data we used was generated by a molecular dynamic simulation of alanine dipeptide in vacuum using of the full atomic representation of the molecule with the CHARMM29 force field [11]. Out of this data, we extract the timeseries of one backbone dihedral (or torsion) angle, traditionally denoted as Φ (see figure 4), which is 5×10^6 points long (time interval between consecutive points equals 0.1 picosecond). Other strategies to arrive at reduced descriptions of molecular dynamics can be found for example in [12,13].

A histogram of the distribution of the angle Φ, figure 8, shows three peaks (two well-prononced ones around $\Phi = -150°$ and $\Phi = -90°$, and one much more shallow around $\Phi = 75°$) corresponding to three long-lived conformation states that characterize alanine dipeptide (see e.g. [15,14,16]). Since the three sets can be distinguished in the histogram of Φ, we want to find a generator that correctly describes the statistics and dynamics of Φ alone. This is a particularly difficult test, since typically both torsion angles Φ and Ψ are used in attempts to find a reduced description of the macrostate dynamics of the molecule (as, for instance, in [13]). We bin the data (i.e. Φ) into 10 bins of 36° each, thereby obtaining a state space $\mathcal{S}$ with 10 states. The timeseries is binned accordingly.

The eigenvalue spectrum $\tilde{\lambda}_k$ is calculated for various lags h, by constructing $\tilde{P}^{(h)}$ from (8), calculating its spectrum $\tilde{\Lambda}_k$ and using (9). In figure 9 we show

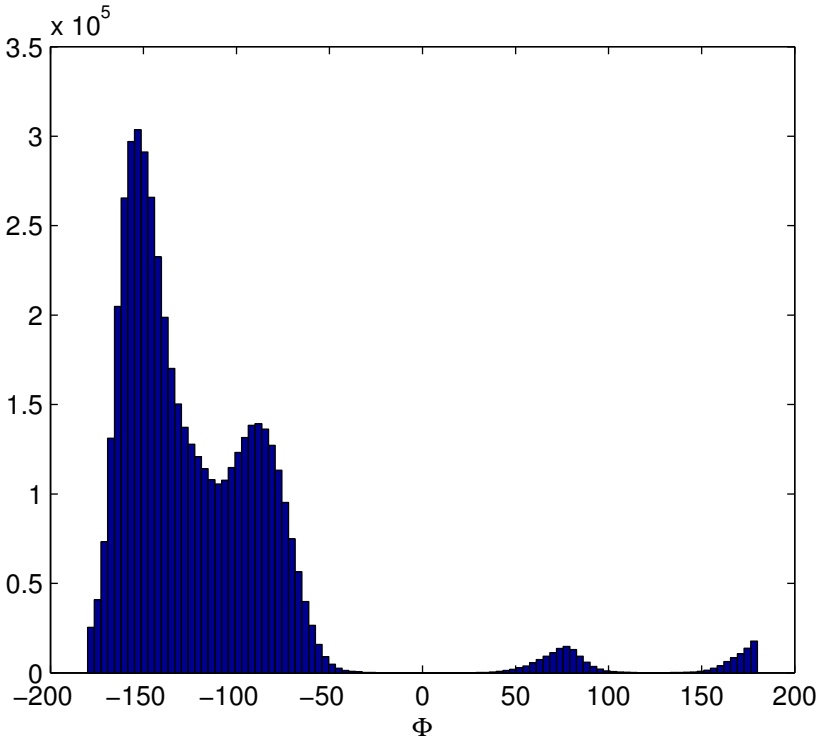

Figure 8. Histogram for the torsion angle Φ of the simulated alanine dipeptide molecule. The domain is periodic. Three metastable sets are visible, one with approximate range $\Phi \in (50°, 100°)$, another one with $\Phi \in (-180°, -110°) \cup (150°, 180°)$ and a third one with $\Phi \in (-110°, -50°)$.

$\tilde{\lambda}_k$ for $h = 0.1\, h_n$ picoseconds, $1 \leq h_n \leq 100$. The value of $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$ can be seen to vary with h due to the non-Markov nature of the data. From figure 9 as well as from the shape of the autocorrelation function (shown in figure 14) we infer that in order to get the long timescale behavior right, one should set $\tilde{\lambda}_2 = -0.007$ and $\tilde{\lambda}_3 = -0.4$. The other eigenvalues, as well as all eigenvectors, are taken from $P^{(h)}$ at $h = 0.1$ ps.

To find the optimal generator, the weights of the object function are set to $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ for all k. The eigenvalues λ_k and (leading) eigenvectors ψ_k and ϕ_k of the resulting generator are shown in figures 10, 11 and 12. The leading eigenmodes are well reproduced by the generator, including the invariant distribution (ψ_1). From the structure of the second and third eigenvectors it is visible that the two slow timescales of the system are related to transitions between the three metastable sets. A graphical representation of the generator itself is given by showing $\log[\mu(x) L_{\min}(x, y)]$ (no summation over x; $x \neq y$) in figure 13. Showing $L_{\min}$ itself would not be informative as the figure would be dominated by jump rates out of states with very low probability. From figure 13 one can see that a) transitions are mainly local (the highest transition rates are for jumps from x to $x \pm 1$) and b) the Markov chain is nearly time-reversible (it is close to detailed balance, as can be seen from the

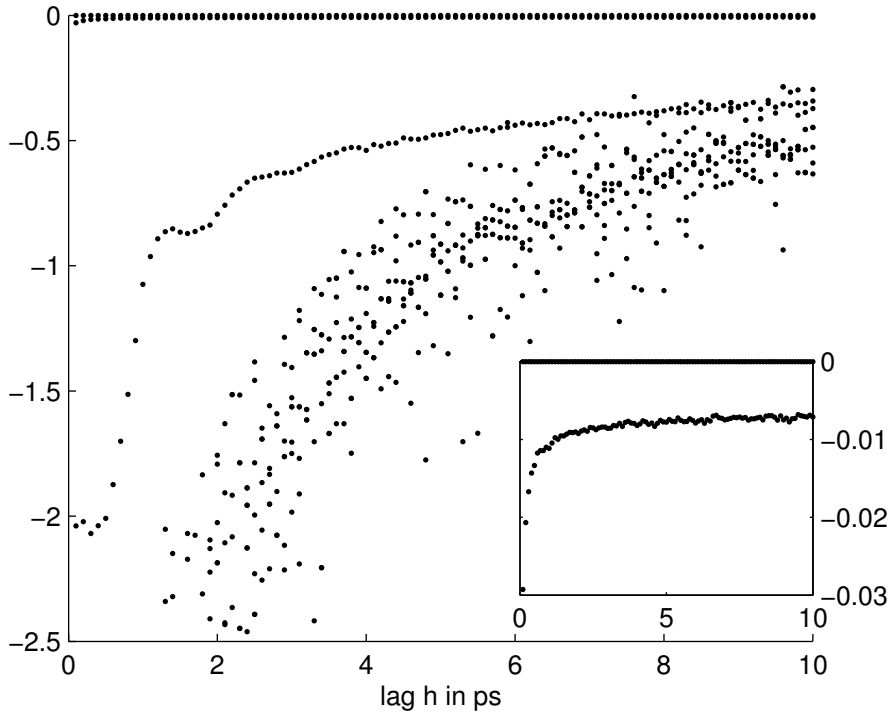

Figure 9. Real parts of the eigenvalues $\tilde{\lambda}_k$ calculated from the molecular simulation data using different lags. In the inset we have zoomed in on the values of $\tilde{\lambda}_2$ (by construction, $\tilde{\lambda}_1 = 0$ for all h).

near-symmetry of $\mu(x)L_{\min}(x,y)$).

The autocorrelation function (ACF) of the data and of the Markov chain generator are shown in figure 14. The rate of decorrelation at longer timescale ($>$ 20 ps) of the Markov chain is a bit higher than that of the data. This is related to a minor error in the reconstructed second eigenvalue, $\lambda_2 < \tilde{\lambda}_2$. Notwithstanding, the overall shape of the ACF is well reproduced by the Markov chain, which is quite remarkable since the ACF involves two very different timescales – very rapid decay at the beginning, much slower afterwards. Overall, the reconstructed Markov chain is capable of correctly describe the dynamics of the torsion angle Φ in alanine dipeptide despite the severe coarsening that a representation of the molecule by this angle alone represents.

Note that, in order to reproduce auto- and cross-correlations functions of a given timeseries correctly with a reconstructed optimal generator, it is not enough to reproduce the eigenvalues correctly – the eigenvectors are important as well. It is easy to show that the ACF, for example, can be written as

$$\begin{aligned} \mathbb{E}\,(X_{t+\tau}\,X_t) &= \sum_{k=1}^{n} e^{t\,\lambda_k} \sum_{x\in\mathcal{S}} \mu(x)\,x\,\phi_k(x) \sum_{y\in\mathcal{S}} y\,\psi_k(y) \\ &= \sum_{k=1}^{n} e^{t\,\lambda_k}\,c_k \end{aligned} \tag{30}$$

(assuming here, for simplicity, zero mean and unit variance for X_t). The ACF is thus determined by both the eigenvalues λ_k and the constants c_k, which are in turn determined by the eigenvectors.

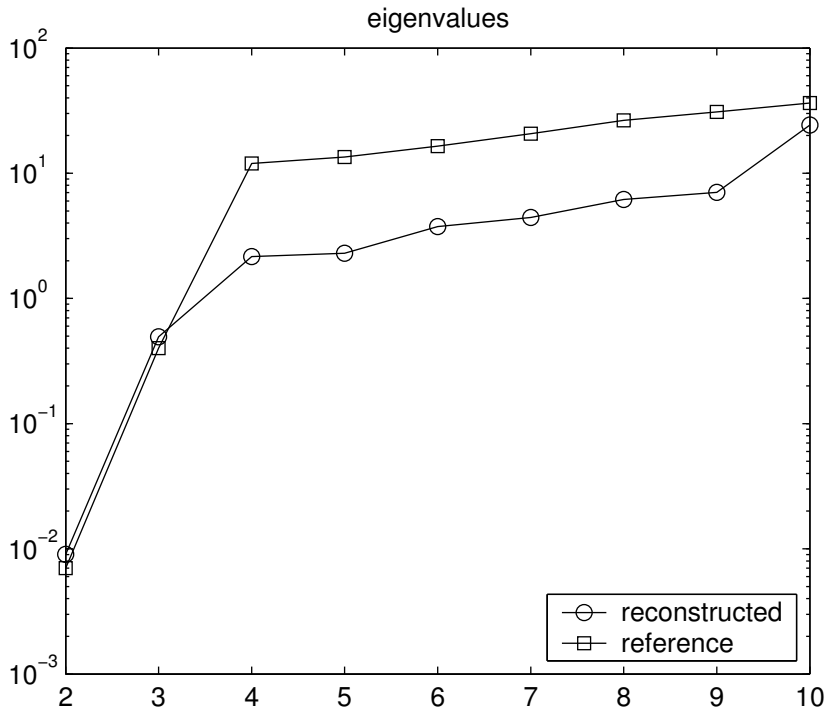

Figure 10. Real parts ($\times -1$) of the eigenvalues λ_k and $\tilde{\lambda}_k$ (reconstructed and reference) of the molecular data example.

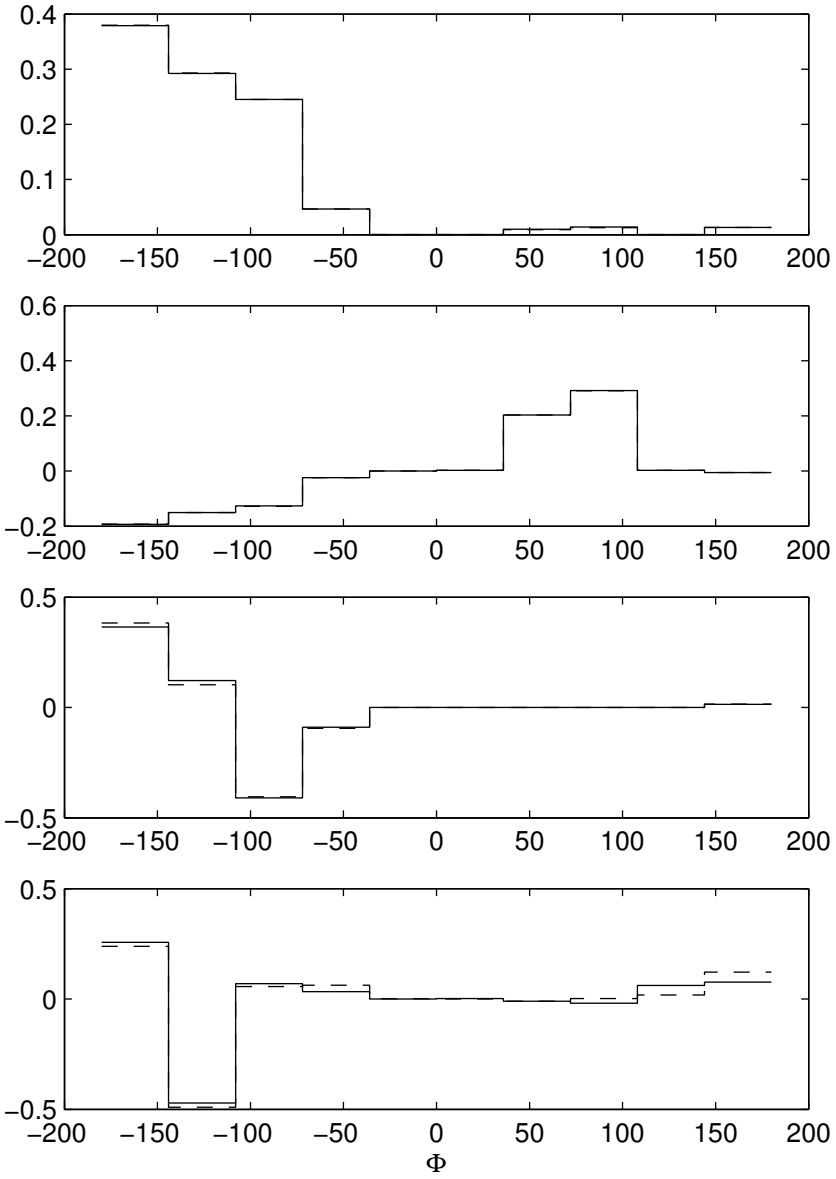

Figure 11. Leading eigenvectors ψ_k and $\tilde{\psi}_k$ (reconstructed, solid and reference, dashed) of the molecular data example. From top to bottom: k=1,2,3,4.

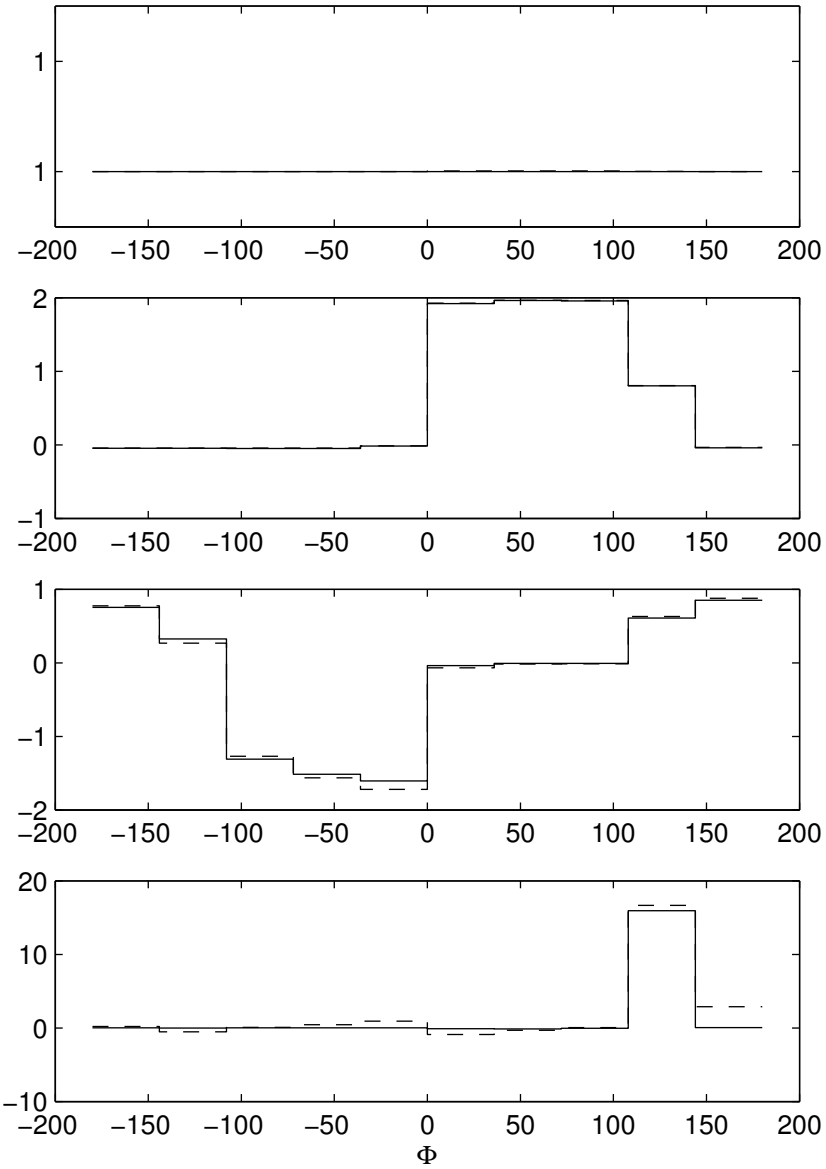

Figure 12. Leading eigenvectors ϕ_k and $\tilde{\phi}_k$ (reconstructed, solid and reference, dashed) of the molecular data example. From top to bottom: k=1,2,3,4.

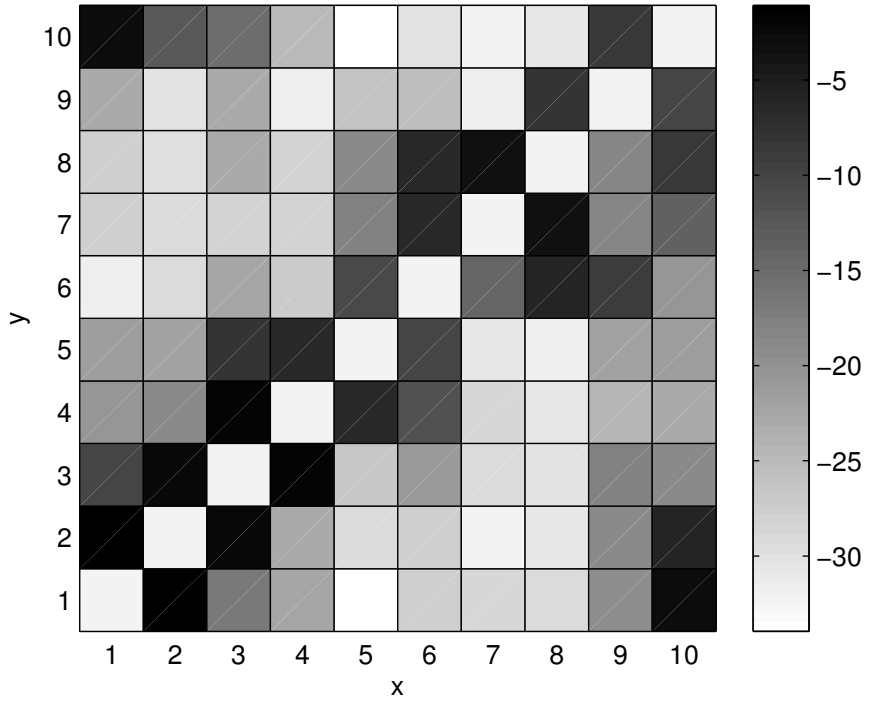

Figure 13. Graphical representation of the optimal generator $L_{\min}$ for the molecular data. Shown is $\log[\mu(x)L_{\min}(x,y)]$ (no summation over x). Showing $L_{\min}$ itself would not be informative as it would be dominated by jump rates out of states with very low probability. The state $x = 1$ corresponds to $\Phi \in (-180°, -144°)$, $x = 2$ to $\Phi \in (-144°, -108°)$, etcetera. Only the off-diagonal elements are shown; the diagonal elements are fully determined by the off-diagonal ones. The Markov chain is nearly time-reversible (near-symmetry of $\mu(x)L_{\min}(x,y)$) and shows locality of transitions (dominant transition rates from x to $x \pm 1$).

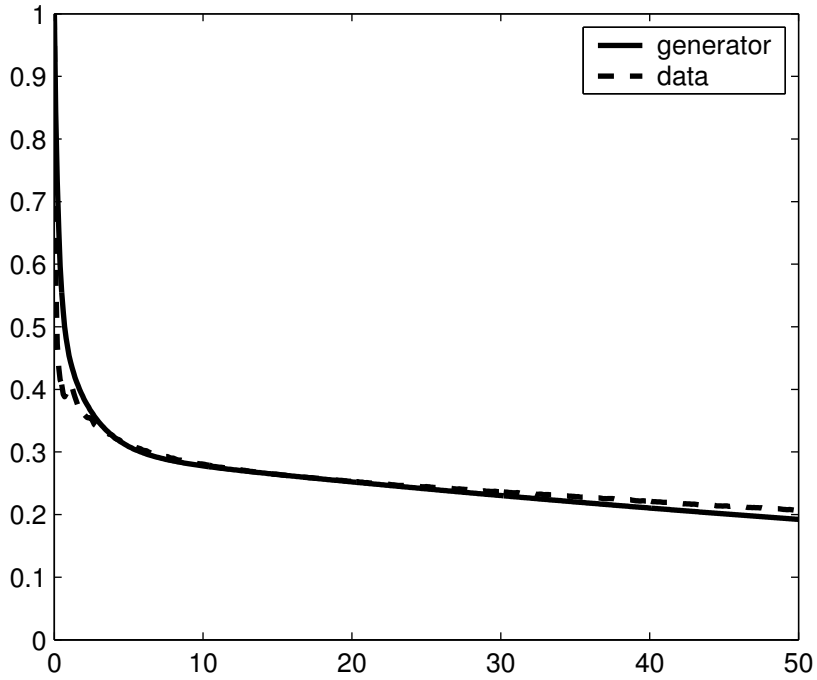

Figure 14. Autocorrelation function for Φ, from optimal Markov chain generator and directly from data. The unit of time is picoseconds. The overall shape of the ACF is well reproduced by the Markov chain, which is quite remarkable since the ACF involves two very different timescales – very rapid decay at the beginning, much slower afterwards.

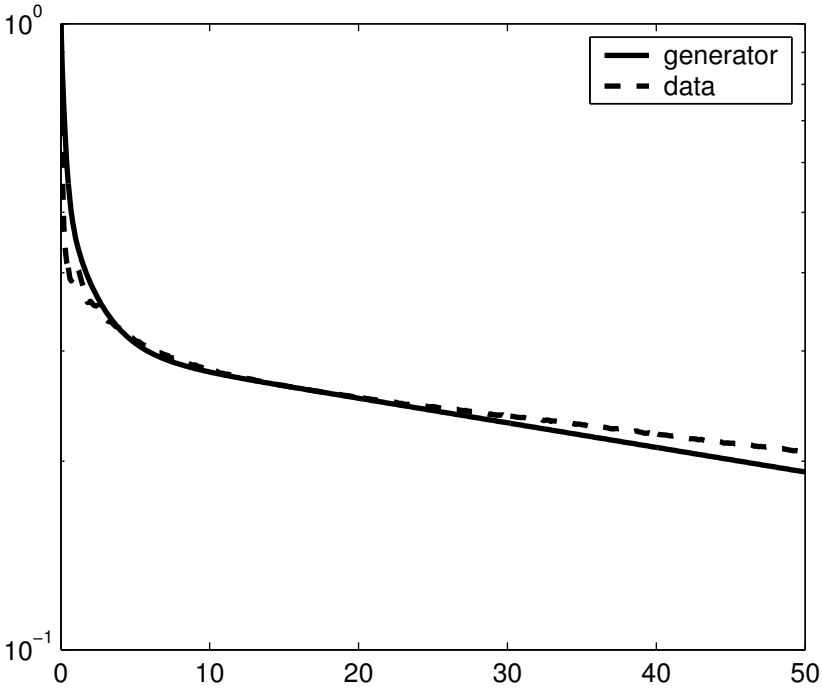

Figure 15. Same as in figure 14 in linear-log scale.

5 Application to a timeseries from an atmospheric model

In this section we use a timeseries of a model for large-scale atmospheric flow in the Northern Hemisphere. The model describes barotropic flow over realistic topography, and generates a realistic climate using 231 variables (wavenumber truncation T21). A more detailed description of the model, its physical interpretation and its dynamics is given in [17]. As is usually the case in models for large-scale flow, much of the interesting dynamics is captured by a fairly low number of leading modes of variability (Principal Components, or PCs). Describing the dynamics of the leading PCs without explicitly invoking the other, trailing PCs is a challenging and well-known problem; one that we aim to tackle here using a continuous-time Markov chain description (see also [18], [19] and [20] for different approaches to arrive at a reduced description of the dynamics of the leading modes of variability of the same atmospheric model). Since we resolve only one or two variables out of a total of 231, without a clear timescale separation between resolved and unresolved variables, non-Markov effects are important in this situation. It is far from obvious that a Markov chain can be successful at all under these circumstances.

Although the use of continuous-time Markov chains is rare in atmosphere-ocean science, the use of discrete-time Markov chains is not. Examples can be found in [21], [22], [23], [24], [25] and many more studies.

5.1 1-dimensional situation

For the 1-dimensional case the leading Principal Component (PC1) is used; in section 5.2 we consider the 2-dimensional case, using PC1 and PC3. A total of 10^7 datapoints is available, with a timestep $h = 1$ which is interpreted as 1 day.

The state space for PC1 is discretized into 10 bins, which we interpret as the 10 states of the state-space $\mathcal{S}$. The timeseries is binned accordingly, and we calculate $\tilde{P}^{(h)}$ from (8) and its eigenspectrum for all lags $1 \leq h \leq 200$. The real parts of the (leading) eigenvalues $\tilde{\lambda}_k$ are shown in figure 16. Eigenvalues with $\mathrm{Re}\tilde{\lambda}_k < -0.1$ are not shown, since we are primarily interested in the leading eigenvalues. As can be seen, $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$ (both are real) drop in value over the range $1 \leq h \leq 10$, then go up again. Only for long lags ($h \approx 200$ for $\tilde{\lambda}_2$, $h \approx 100$ for $\tilde{\lambda}_3$) do they reach values that are consistent with the ACF for PC1 (shown in figure 20): $\tilde{\lambda}_2 = -0.007$, $\tilde{\lambda}_3 = -0.03$. For even longer lags, the estimates for $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$ eventually become overwhelmed by sampling error. For $\tilde{\lambda}_3$ this can be seen to happen for $h > 100$ (figure 16), for $\tilde{\lambda}_2$ it lies beyond the limits of the figure.

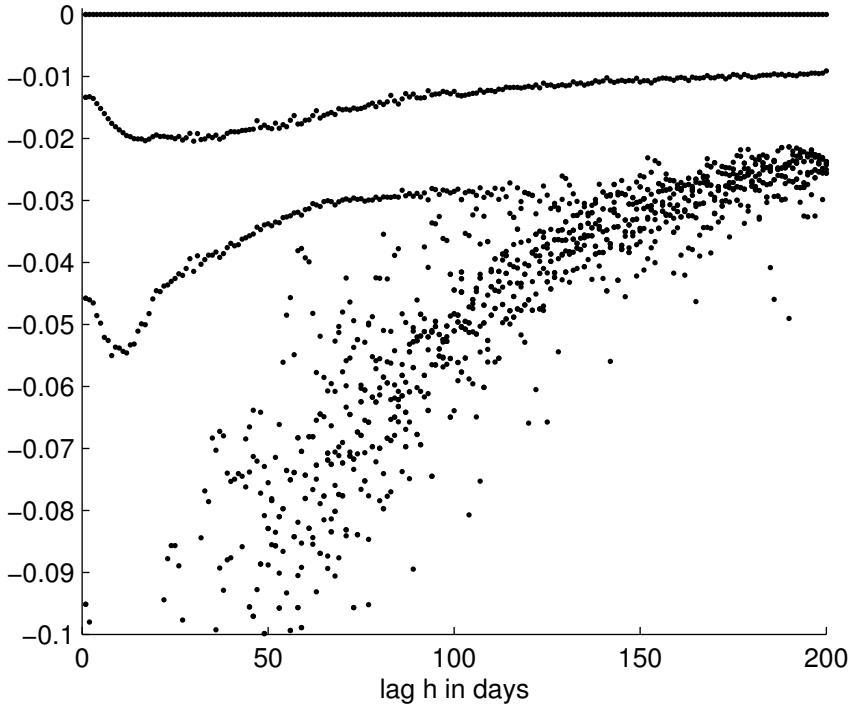

Figure 16. Real parts of the eigenvalues $\tilde{\lambda}_k$ calculated from the atmospheric dataset (PC1).

With figure 16 in mind, we set $\tilde{\lambda}_2 = -0.007$, $\tilde{\lambda}_3 = -0.03$ by hand, and use the spectrum of $\tilde{P}^{(h)}$ at $h = 100$ for the other eigenvalues and for all eigenvectors. Other than for the molecular data in the previous section, the lag $h = 1$ is too low in this case, because $h = 1$ is below the slowest of the fast timescales of the system. If data from $h = 1$ is used, these fast timescales negatively affect the effective description of the slow dynamics. At $h = 100$ this is no longer a problem.

The object function weights are set to $\tilde{\alpha}_1 = \tilde{\gamma}_2 = \tilde{\gamma}_3 = 100$ and $\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1$ otherwise. The generator obtained from the minimization has leading eigenmodes that match the reference spectrum quite well, see figure 17 for the eigenvalues and figures 18 and 19 for the leading four ψ_k and ϕ_k. The ACF is also well reproduced, see figures 20 and 21. In figure 22, the generator is represented by showing $\log[\mu(x)L_{\min}(x, y)]$ (no summation over x; $x \neq y$). This generator, unlike the one obtained for the molecular data, is far from time-reversibility ($\mu(x)L_{\min}(x, y)$ is not nearly symmetric), and nonlocal transitions (from x to $y > x + 1$ or $y < x - 1$) are important.

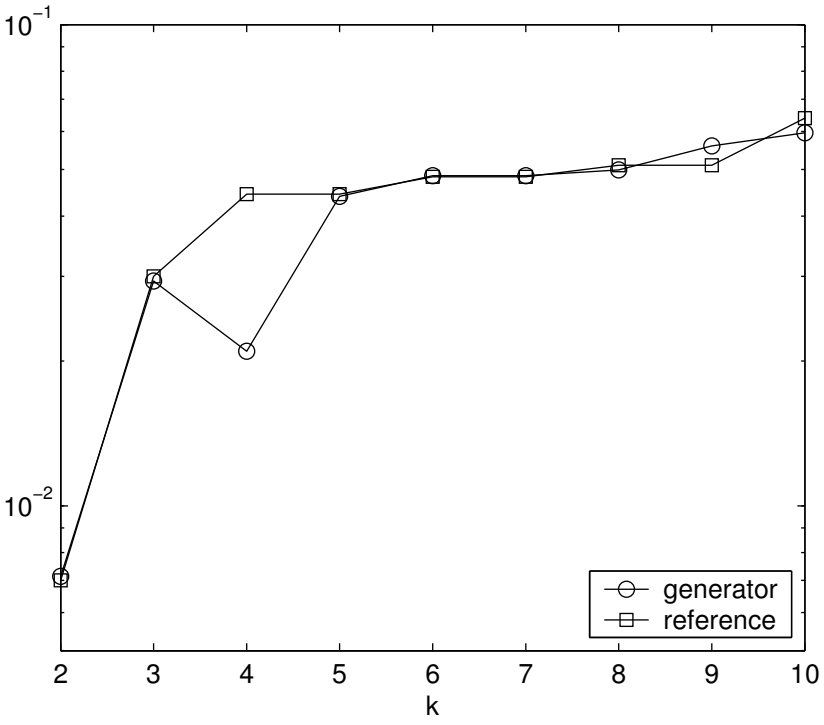

Figure 17. Real parts ($\times -1$) of the eigenvalues λ_k and $\tilde{\lambda}_k$ (reconstructed and reference) of the atmospheric data. All $\tilde{\lambda}_k$ were taken from $P^{(h)}$ at $h = 100$, except for $k = 2, 3$ where we set $\tilde{\lambda}_2 = -0.007$, $\tilde{\lambda}_3 = -0.03$ (see text).

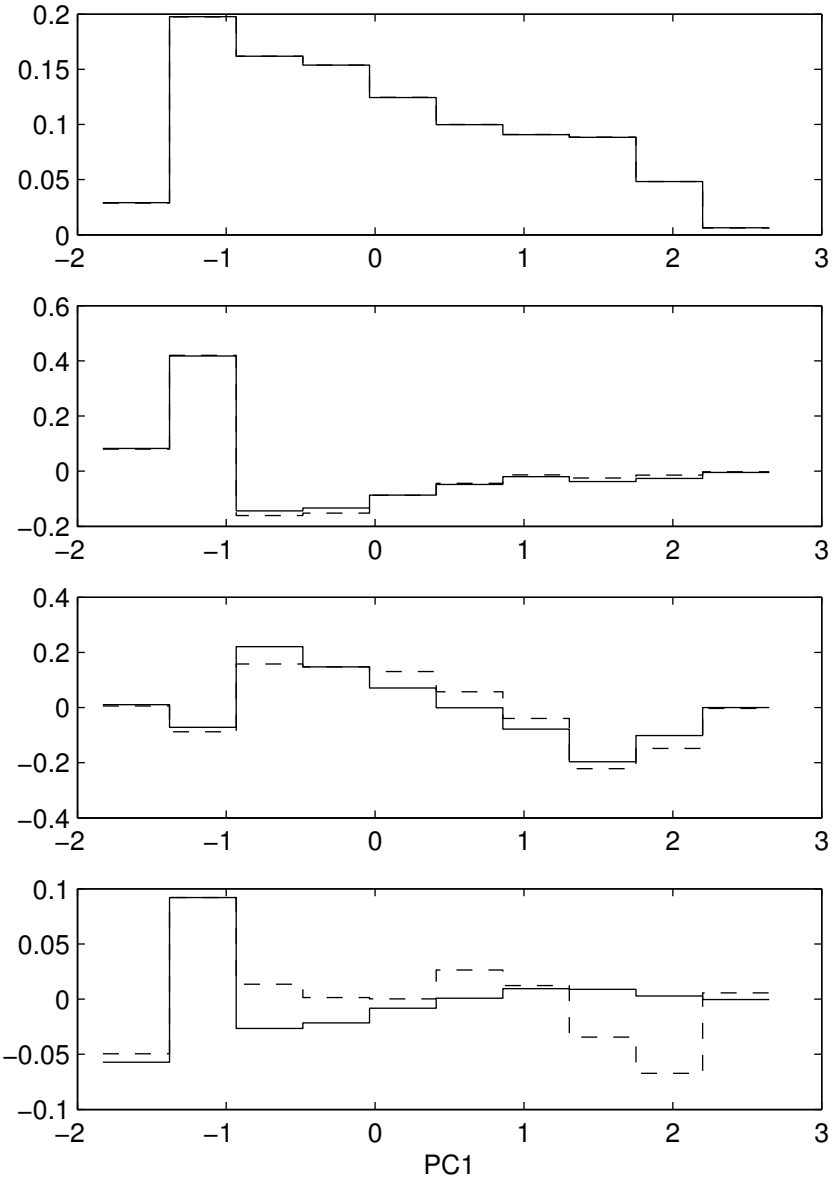

Figure 18. Leading eigenvectors ψ_k and $\tilde{\psi}_k$ (reconstructed, solid and reference, dashed) of the 1-dimensional atmospheric data example. From top to bottom: k=1,2,3,4. All $\tilde{\psi}_k$ were taken from $P^{(h)}$ at $h = 100$.

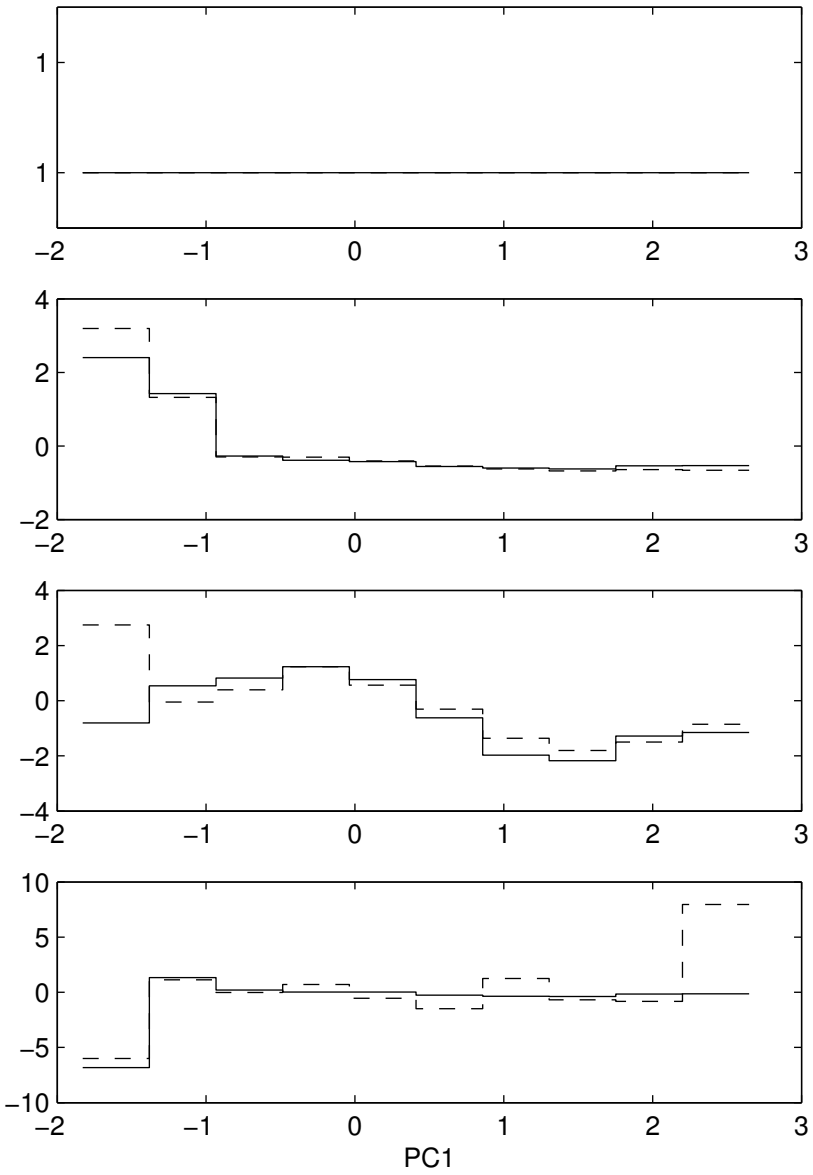

Figure 19. Leading eigenvectors ϕ_k and $\tilde{\phi}_k$ (reconstructed, solid and reference, dashed) of the 1-dimensional atmospheric data example. From top to bottom: k=1,2,3,4. All $\tilde{\phi}_k$ were taken from $P^{(h)}$ at $h = 100$.

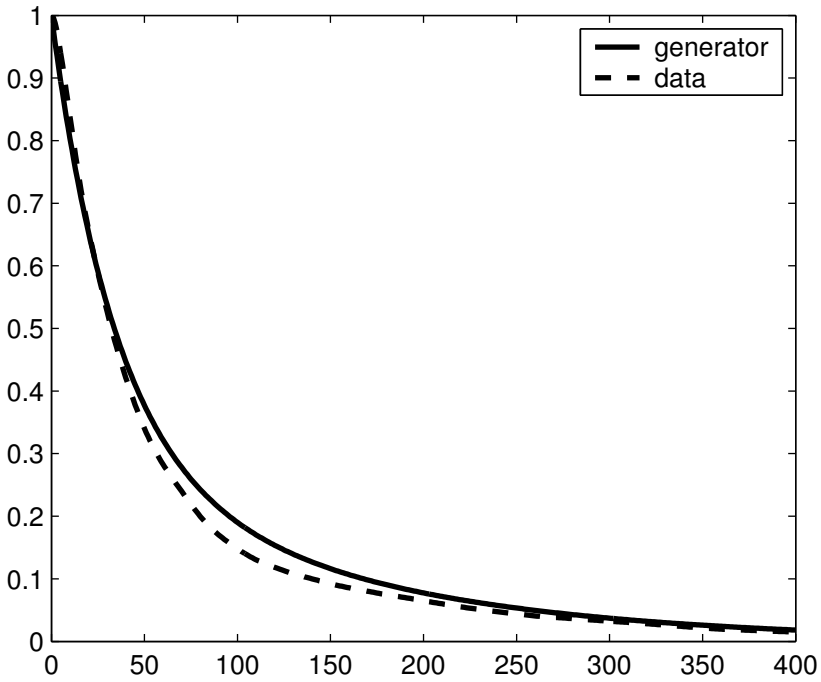

Figure 20. Autocorrelation function for PC1, from optimal Markov chain generator and directly from data. Timelags are in days. The generator was obtained with most eigenmodes estimated at $h = 100$, see text.

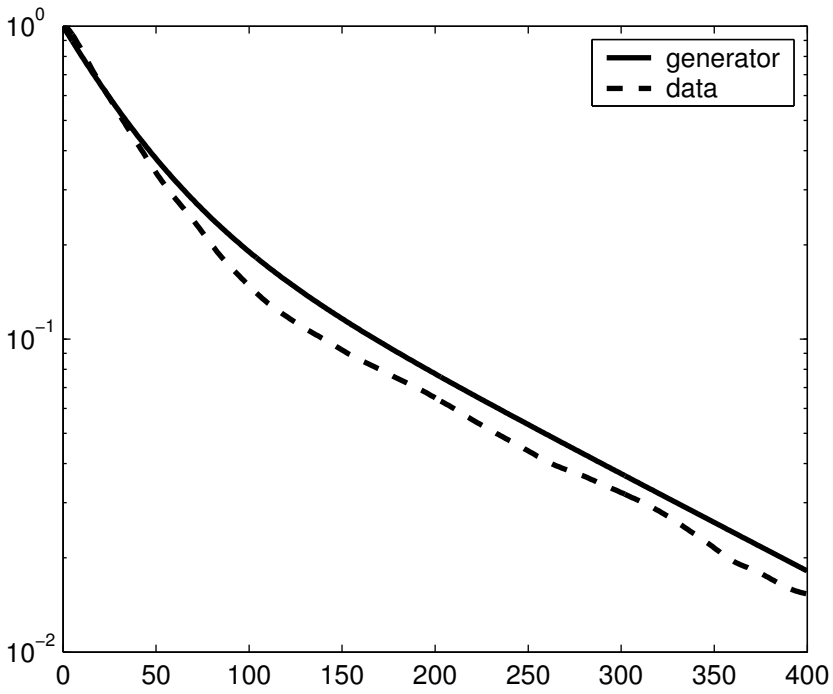

Figure 21. Same as figure 20 in linear-log scale

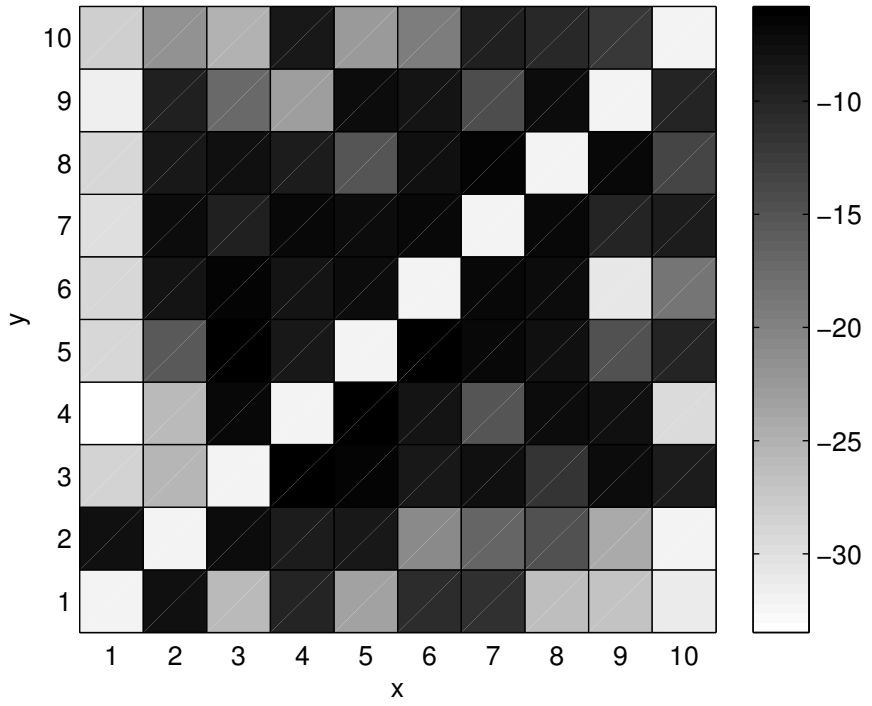

Figure 22. Graphical representation of the optimal generator $L_{\min}$ for the atmospheric data (1-d). Shown is $\log[\mu(x)L_{\min}(x,y)]$ (no summation over x). Showing $L_{\min}$ itself would not be informative as it would be dominated by jump rates out of states with very low probability. The state $x = 1$ corresponds to PC1 < -1.38, $x = 2$ to $-1.38 <$ PC1 < -0.93, etcetera. Only the off-diagonal elements are shown; the diagonal elements are fully determined by the off-diagonal ones.

For the 2-dimensional case (timeseries for PCs 1 and 3) we calculate $\tilde{P}^{(h)}$ from (4) with $h = 50$, using 5×5 bins. As in the 1-dimensional case, too small values of h yield leading eigenmodes of $\tilde{P}^{(h)}$ with which the ACFs cannot be reproduced correctly, even if the generator eigenspectrum matches the reference spectrum perfectly. Because of the higher number of bins than used in the 1-dimensional case, sampling errors for $\tilde{P}^{(h)}$ are too large at $h = 100$. Therefore we use a smaller lag, $h = 50$. The leading eigenvalue $\tilde{\lambda}_2$ is again adjusted to -0.007 (just as in the 1-dimensional case); $\tilde{\lambda}_3 = -0.033$ at $h = 50$ and is not further adjusted. Two of the bins remain empty, so effectively the state space is discretized into 23 bins. The object function weights are the same as in the 1-dimensional case ($\tilde{\alpha}_k = \tilde{\beta}_k = \tilde{\gamma}_k = 1 \ \forall\, k$, except $\tilde{\alpha}_1 = \tilde{\gamma}_2 = \tilde{\gamma}_3 = 100$).

Figure 23 shows the eigenvalues λ_k of the resulting generator $L_{\min}$, as well as the reference values $\tilde{\lambda}_k$. Figures 24 and 25 show the leading eigenvectors. The leading eigenvectors and eigenvalues are well reproduced, in particular the invariant distribution ψ_1. The ACFs are shown in figure 26; the generator captures the decay of both ACFs rather well.

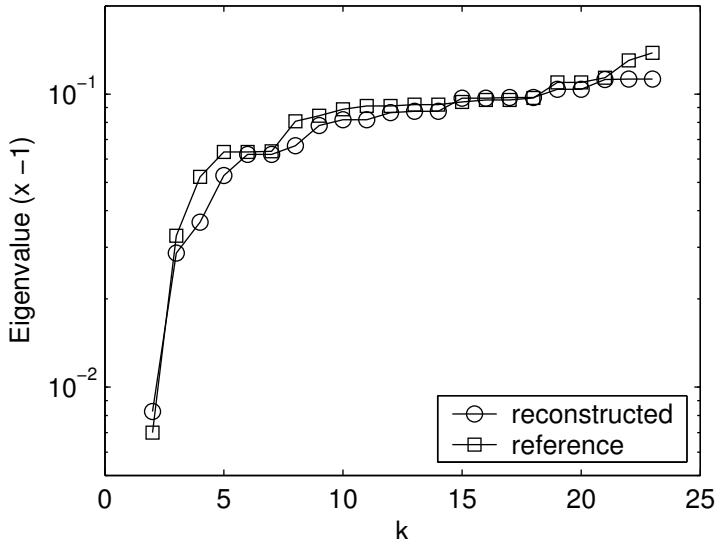

Figure 23. Eigenvalues λ_k and $\tilde{\lambda}_k$ (reconstructed and reference) of 2-dimensional atmospheric data example.

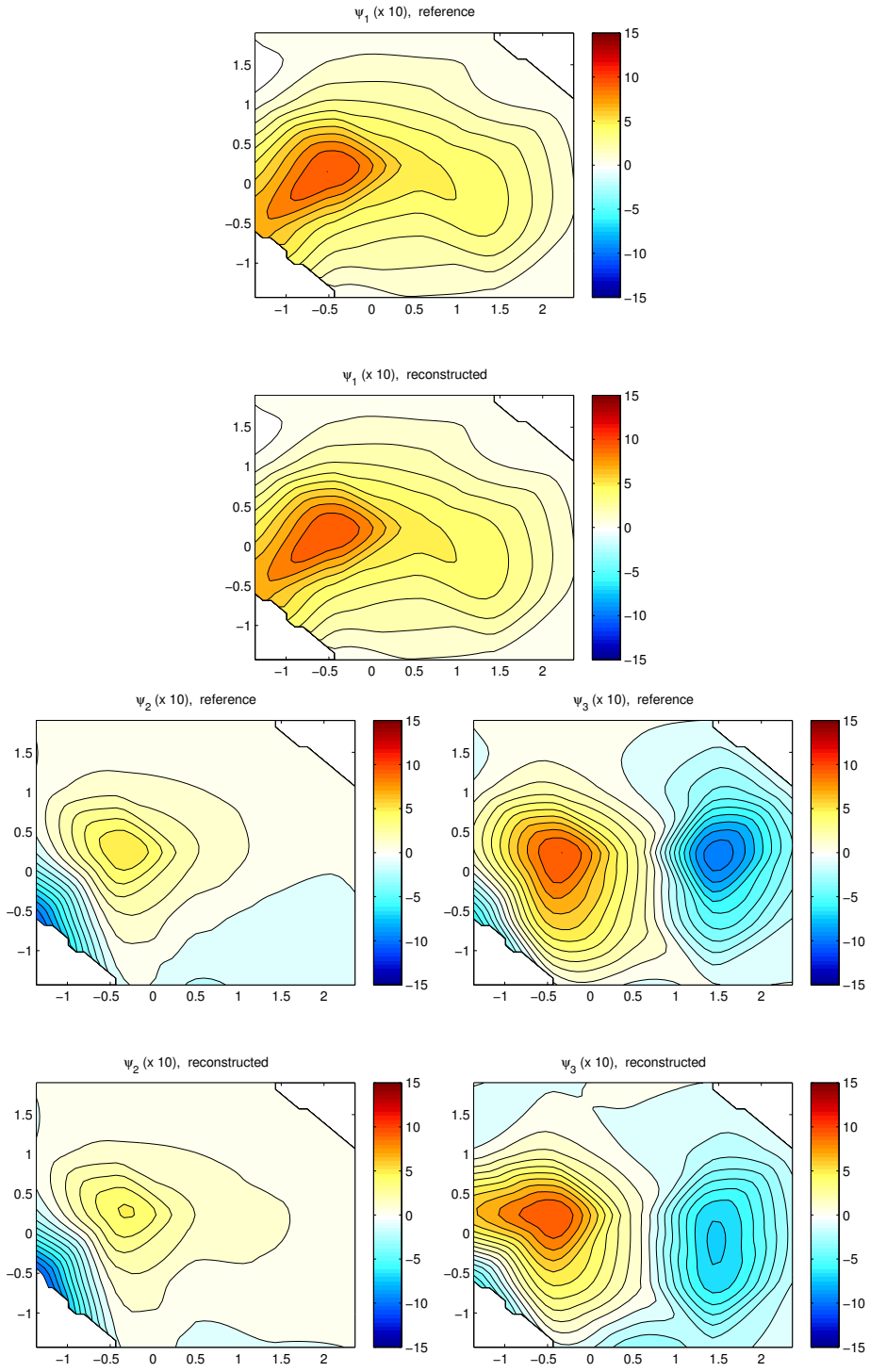

Figure 24. Leading eigenvectors ψ_k and $\tilde{\psi}_k$ (reconstructed and reference) of 2-dimensional atmospheric data example.

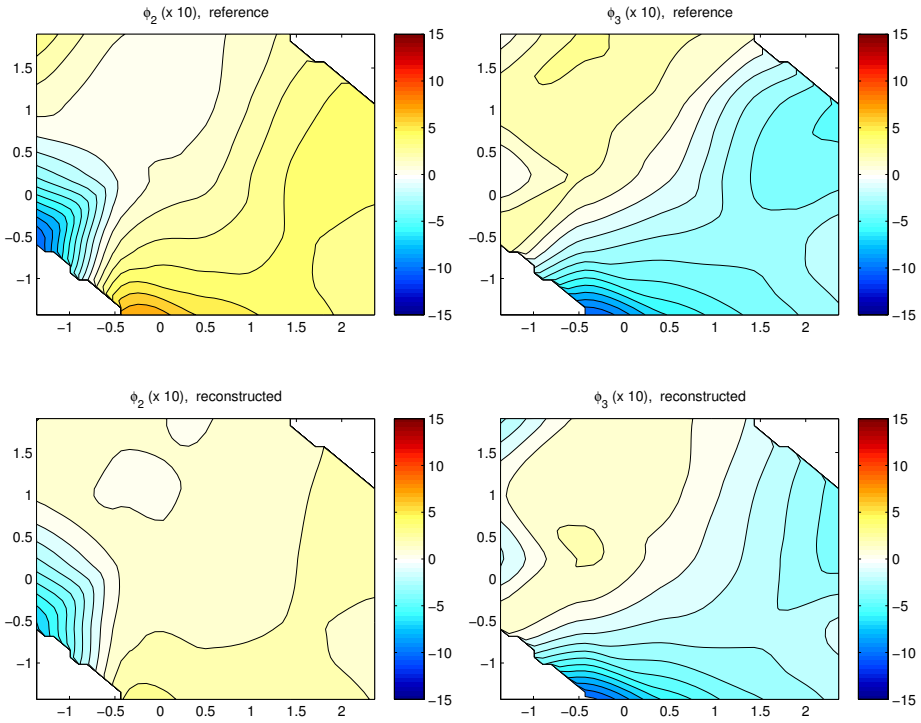

Figure 25. Leading eigenvectors ϕ_k and $\tilde{\phi}_k$ (reconstructed and reference) of 2-dimensional atmospheric data example.

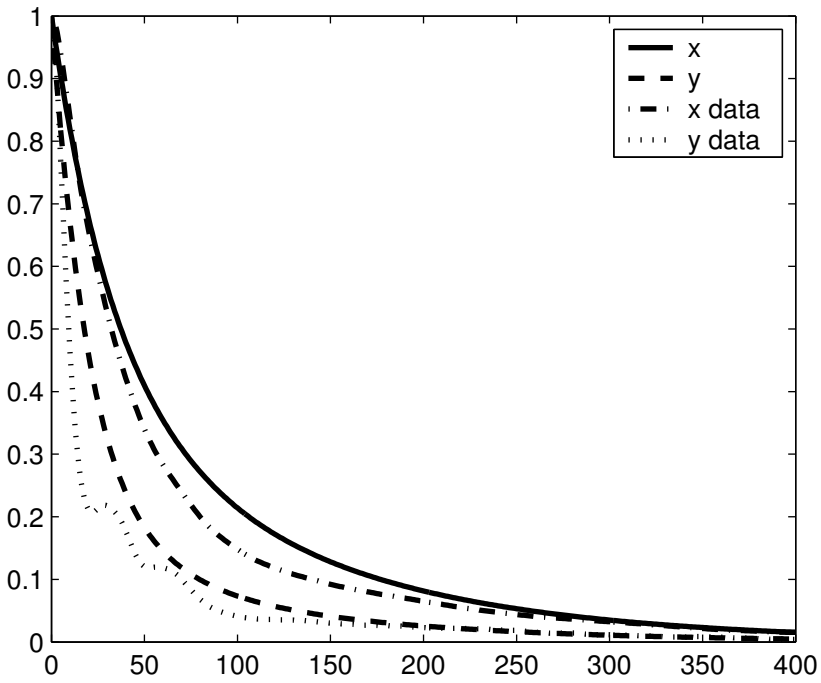

Figure 26. Autocorrelation functions for PCs 1 (x) and 3 (y), from optimal Markov chain generator and directly from data.

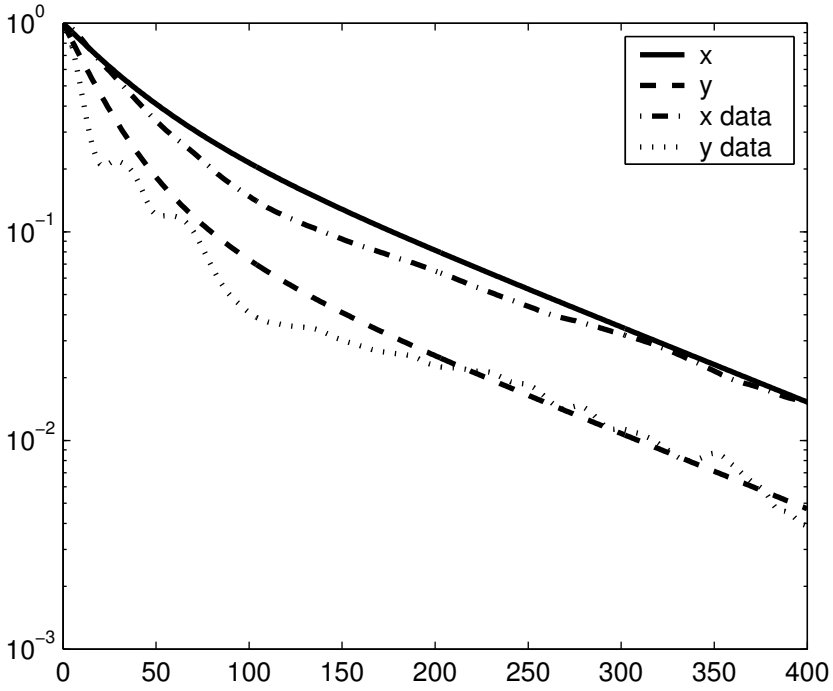

Figure 27. Same as in figure 26 in linear-log scale. Notice that the procedure allows to reproduce the ACF rather well, including the fact that the decay occur on two different time-scales: fast at the beginning, then more slowly afterwards.

6 Conclusion and generalization

We have presented a new method to fit data from timeseries by a continuous-time Markov chain via the solution of a quadratic programming problem. The key element in the approach is the matching of the eigenspectrum of the generator with the eigenspectrum observed from the data. Since the eigenspectrum completely determines the generator and its associated Markov process, matching it with the observed spectrum, if succesful, amounts to the most complete reconstruction of the continuous-time Markov chain from the data. Matching of the observed invariant distribution is part of this reconstruction.

The Markov chain embedding problem implies that one cannot expect given timeseries to have always an exact underlying generator (in practice, the stochastic matrices calculated from data hardly ever have an exact underlying generator). Therefore, optimal generators have to be found by minimizing an object function that measures the difference between the desired and the actual spectral properties of the generator. We have proposed an object function that is quadratic and convex, and does not require calculation of the eigenspectrum of the generator. Minimization is therefore computationally cheap and easy (quadratic programming). The reconstruction method was illustrated with several examples. We used timeseries generated by Markov chains with and without underlying generator, as well as timeseries from two applications (molecular dynamics and atmospheric flow) with problems of non-Markov effects at short timelags. The algorithm gave good results; in all examples the leading eigenvectors and eigenvalues of the reconstructed generator resembled the observed eigenmodes (very) closely.

In its most general setting, imposing no additional conditions on L other than that it be a true generator, the minimization problem is of dimension $n^2 - n$ if the state-space of the Markov chain consists of n states. However, by restricting the class of Markov chain generators, the dimensionality of the problem can be reduced without reducing the size of the state-space. For example, the assumption of detailed balance, $\mu(x)L(x,y) = \mu(y)L(y,x)$, eliminates half of the variables from the minimization problem: only the matrix elements $L(x,y)$ with $x > y$ need to be determined. Detailed balance is a non-trivial assumption; for instance, the optimal generator found for the molecular data is close to detailed balance, whereas the generators for the atmospheric data are not. Another example of restricting the class of generators would be to impose the structure of a birth-death process. Such a process is characterised by a generator of the type:

$$\sum_{y \in \mathcal{S}} L(x,y) f(y) = \sum_{j=1}^{m} \nu_j(x)(f(x+e_j) - f(x)) \tag{31}$$

where $\nu_j(x) \geq 0$ are constants, and e_j are such that $x + e_j \in \mathcal{S}$ if $\nu_j(x) \neq 0$

and, typically, m is (much) smaller than n. One would then fix the e_j and minimize the object function (13) over all possible $\nu_j(x)$ subject to $\nu_j(x) \geq 0$. The dimensionality of this minimization problem is mn, which is substantially smaller than $n^2 - n$ if $m \ll n$. In terms of implementation, it is a completely straightforward generalization of what was done in this paper.

Notice in particular that a structure like (31) for the generator is quite natural for systems in which jumps can only occur between states x and y which correspond to neighboring bins in physical space. These considerations lead us to the possibility of generalizing the reconstruction procedure, outlined in this paper for finite state Markov chains, to diffusion processes. An appropriate discretization of the Fokker-Planck operator using e.g. finite-differences or finite-elements will convert the problem of reconstructing the drift and diffusion coefficients into the problem of reconstructing a generator with a structure similar to the one in (31). The procedure proposed here can be used to tackle the latter problem, and thereby reconstruct the drift and diffusion in spatially discretized form. We intend to explore this approach to the reconstruction of diffusion processes in a future study.

Acknowledgments

We thank Christian Franzke for providing us with the timeseries of the atmospheric model, and Paul Maragakis for making his molecular simulation data available. Helpful comments and suggestions from Andy Majda are gratefully acknowledged. We also thank Weinan E, Paul Maragakis, Weiqing Ren, and Richard Tsai for helpful discussions when this project was at an early stage. This work was sponsored in part by NSF through Grants DMS01-01439, DMS02-09959, DMS02-39625 and DMS-0222133, and by ONR through Grants N-00014-04-1-0565 and N-00014-96-1-0043.

References

[1] J. Norris, Markov chains, Cambridge University Press, 1997.

[2] O. Häggström, Finite Markov chains and Algorithmic Applications, Cambridge University Press, 2002.

[3] W. Anderson, Continuous-Time Markov chains, Springer-Verlag, 1991.

[4] J. Kingman, The imbedding problem for finite Markov chains, Z. Wahrsch. 1 (1962) 14–24.

[5] B. Singer, S. Spilerman, The representation of social processes by Markov models, Am. J. Sociol. 82 (1976) 1–54.

[6] R. Israel, J. Rosenthal, J. Wei, Finding generators for Markov chains via empirical transtition matrices, with applications to credit ratings, Math. Finance 11 (2001) 245–265.

[7] M. Bladt, M. Sørensen, Statistical inference for discretely observed Markov jump processes, J. R. Statist. Soc. B 67 (2005) 395–410.

[8] P. Gill, W. Murray, M. Wright, Practical Optimization, Academic Press, 1981.

[9] J. Nocedal, S. Wright, Numerical Optimization, Springer, 1999.

[10] T. Anderson, L. Goodman, Statistical inference about Markov chains, Ann. Math. Statist. 28 (1957) 89–110.

[11] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comp. Chem. 4 (1983) 187–217.

[12] W. Huisinga, C. Schütte, A. Stuart, Extracting macroscopic stochastic dynamics: model problems, Comm. Pure Appl. Math. 56 (2003) 234–269.

[13] G. Hummer, I. G. Kevrekidis, Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations, J. Chem. Phys. 118 (2003) 10762.

[14] P. G. Bolhuis, C. Dellago, D. Chandler, Reaction coordinates of biomolecular isomerization, Proc. Natl. Acad. Sci 97 (2000) 5877–5882.

[15] J. Apostolakis, P. Ferrara, A. Caflisch, Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water, J. Chem. Phys. 10 (1999) 2099–2108.

[16] W. Ren, E. Vanden-Eijnden, P. Maragakis, W. E, Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide, J. Chem. Phys., to appear.

[17] D. Crommelin, Regime transitions and heteroclinic connections in a barotropic atmosphere, J. Atmos. Sci. 60 (2003) 229–246.

[18] F. Selten, An efficient description of the dynamics of barotropic flow, J. Atmos. Sci. 52 (1995) 915–936.

[19] F. Kwasniok, The reduction of complex dynamical systems using principal interaction patterns, Physica D 92 (1996) 28–60.

[20] C. Franzke, A. Majda, E. Vanden-Eijnden, Low-order stochastic mode reduction for a realistic barotropic model climate, J. Atmos. Sci. 62 (2005) 1722–1745.

[21] K. C. Mo, M. Ghil, Cluster analysis of multiple planetary flow regimes, J. Geophys. Res. 93 (1988) 10927–10952.

[22] R. Vautard, K. C. Mo, M. Ghil, Statistical significance test for transition matrices of atmospheric Markov chains, J. Atmos. Sci. 47 (1990) 1926–1931.

[23] R. Pasmanter, A. Timmermann, Cyclic Markov chains with an application to an intermediate ENSO model, Nonlin. Proc. Geophys. 10 (2003) 197–210.

[24] G. Lacorata, R. Pasmanter, A. Vulpiani, Markov chain approach to a process with long-time memory, J. Phys. Oceanogr. 33 (2003) 293–298.

[25] D. Crommelin, Observed nondiffusive dynamics in large-scale atmospheric flow, J. Atmos. Sci. 61 (2004) 2384–2396.