

# *Dimension Reduction for Systems with Slow Relaxation*

**Shankar C. Venkataramani, Raman C. Venkataramani & Juan M. Restrepo**

**Journal of Statistical Physics**

1

ISSN 0022-4715

Volume 167

Combined 3-4

J Stat Phys (2017) 167:892-933

DOI 10.1007/s10955-017-1761-7

**Volume 167 • Numbers 3-4 • May 2017**

## Journal of Statistical Physics

**Special Issue: Dedicated to the Memory  
of Leo Kadanoff**

**Guest Editors: Susan N. Coppersmith ·  
Sidney R. Nagel**

10955 • ISSN 0022-4715  
167(3-4) 417-1080 (2017)

 Springer

 Springer

**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Dimension Reduction for Systems with Slow Relaxation In Memory of Leo P. Kadanoff

Shankar C. Venkataramani<sup>1</sup>  ·  
Raman C. Venkataramani<sup>2</sup> · Juan M. Restrepo<sup>3</sup>

Received: 29 September 2016 / Accepted: 1 March 2017 / Published online: 18 March 2017  
© Springer Science+Business Media New York 2017

**Abstract** We develop reduced, stochastic models for high dimensional, dissipative dynamical systems that relax very slowly to equilibrium and can encode long term memory. We present a variety of empirical and first principles approaches for model reduction, and build a mathematical framework for analyzing the reduced models. We introduce the notions of universal and asymptotic filters to characterize ‘optimal’ model reductions for sloppy linear models. We illustrate our methods by applying them to the practically important problem of modeling evaporation in oil spills.

It is with immense gratitude that we dedicate this article to Leo Kadanoff. Two of the authors (SV and JR) first met Leo as postdocs. Our lives would have been very different if not for the outsize role that Leo played in our professional development and also in our personal growth. His door, and his mind, were always open. He reminded us to ask questions, about how humility gave the courage to know what we knew and we did not. He taught us to be fearless about pursuing a wide range of interests. The fun he had with science was infectious, and the skill he had to ask the right questions is impossible to match. Our experiences were by no means unique. There are hundreds of people whose lives Leo touched in the same way. So many of his informal seminar or lunch questions turned into full research enterprises. It is no exaggeration to say that at some point people needed only to know that Leo had been the one to ask the question in order to assure themselves that their scientific investigations were worthwhile. The last time that one of us saw Leo was in May of 2015. Coincidentally, it was at a talk on the subject of oil spill modeling, and included some of the rudimentary ideas that grew into this paper. Leo came down to the (new) James Franck Institute for the talk. He was just as sharp as ever, and he made sure that the graduate students in the audience got all the physical intuition that the speaker elided over, by interjecting appropriately. It was classic Leo. How we miss him!

✉ Shankar C. Venkataramani  
shankar@math.arizona.edu

Raman C. Venkataramani  
ramanv@ieee.org

Juan M. Restrepo  
restrepo@math.oregonstate.edu

<sup>1</sup> Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA

<sup>2</sup> Seagate Technology, 389 Disc Dr., Longmont, CO 80503, USA

<sup>3</sup> Department of Mathematics, Oregon State University, Corvallis, OR 97330, USA

**Keywords** Dimension reduction · Slow relaxation · Sloppy models · Mori–Zwanzig projection · Multi-scale · Aging · Weathering · Glassy systems · Oil spills

## 1 Introduction

The methods of equilibrium statistical physics are remarkably successful in characterizing the thermodynamic limit of Hamiltonian systems with many interacting degrees of freedom. Considering adiabatic perturbations of the equilibrium state, one can also compute the transport coefficients that characterize the linear response to external forcing in non-equilibrium states.

For many out-of-equilibrium systems, the relaxation to equilibrium is exponential and is governed by linear response theory. There are interesting examples that do not fit this paradigm—*Glassy systems* relax very slowly (typically logarithmically) [7]. Such systems include spin glasses [6], forced crumpling [42] and the stick-slip phenomenon and evolution of frictional strength [45]. In these systems, the relaxation to a *putative equilibrium* state is extremely slow and it is unclear if an equilibrium can *ever be reached*.

A striking example in this direction is the phenomenon of *aging* [1,7]. A glassy system is allowed to relax over a long time scale and is then driven by an external perturbation for a length of time  $t_w$ , the waiting time. The perturbation is then removed, and the system is allowed to relax once again. In this situation, the subsequent relaxation dynamics depends not only on the (initially perturbed) macroscopic state of the system, but also on the waiting time  $t_w$ , showing that the relaxation of the system depends on details of its micro-state, and that the micro-state has *memory*, i.e. it can ‘record’ aspects of the history of the system. In particular the dynamics can distinguish identical macro-states that result from distinct preparations (e.g. different waiting times) of an initially “relaxed” or “equilibrium” sample.

There are other consequences of slow relaxation to equilibrium. In amorphous materials below the glass transition temperature  $T_g$ , the slow relaxation manifests itself as a slow change in the effective properties of the material, a process called *physical aging*. As we discuss below, a similar phenomenon is very important in the modeling of oil spills, although the mechanisms are different and crude oil is not a “glassy system” in the usual understanding of this term.

The evolution of oil in the environment is also called *Weathering*. Weathering is informally defined as the mechanical or chemical changes that occur in compounds, such as rock, due to exposure. Its distinguishing feature is that it occurs over very long times. Weathering is a process that is distinct from glassy dynamics, nevertheless, it shares the feature of slow relaxation with the latter.

Our interest in capturing such weathering dynamics arose from work developing a model for oil spill transport in the ocean [43,49,50]. Crude oil is made up of hundreds, even thousands of different chemical compounds. Oil will react chemically due to its complex chemistry, its their exposure to the elements, and due to biological action [30,51]. These changes occur at many time scales (from hours to decades) without any clear scale separation [22].

Oil is both a major economic commodity and a particularly dangerous pollutant so there is great urgency in addressing a spill as soon as it happens. A transport model for an oil spill is thus not very useful if it can only model the asymptotic state. This is because the buoyancy, surface tension, and viscosity of oil droplets depend on their chemical composition and are changing, due to exogenous and endogenous reactions, at different rates at different times.

These changes on the micro-scale have dramatic consequences on the large scale dynamics of expansive oil spills, and on how these oil spills are going to be contained. Petroleum and chemical engineers are thus keenly interested in understanding and effectively modeling these changes, and one of the goals in this work is to develop methods to construct reduced dimensional models to this end (see [22,51] for a review of existing modeling approaches). We seek to develop low dimensional, stochastic models for multi-scale, dissipative dynamics, that can be applied to the practically important problems.

The other goal of this work is just as important—we want to develop intuition and new approaches for problems in statistical physics including (i) coarse graining for systems with slow relaxation/long memory and (ii) robust predictions for *sloppy models* [9,62].

Recently, several techniques have been proposed to address the issue of dimension reduction for complex systems using tools from non-equilibrium statistical mechanics [13,20,27,52,59]. These techniques are based on earlier work by Mori [44], Zwanzig [64,65] and Kawasaki [33]. The Mori–Zwanzig projection operator formalism [32,66] decomposes the high-dimensional phase space of a system into resolved (or observed) variables and unresolved/unobserved degrees of freedom. The key idea is to project the full dynamics on the phase space (more properly, to project the Liouville equation for the evolution of probability measures on phase space) on to dynamics on the resolved degrees of freedom using statistical information to take an expectation over the unresolved degrees of freedom [66]. This procedure results in a *generalized Langevin equation* (GLE)

$$\dot{X} = f(X) + \int_0^t K(X(t-s), s)ds + \eta(X, t),$$

where  $X \in \mathbb{R}^n$  is the set of  $n$  resolved degrees of freedom,  $K(X(t-s), s)$  is the *memory kernel*, that quantifies the information in the history of the resolved variables  $X$  on its subsequent evolution, and  $\eta \in \mathbb{R}^n$  the “noise” is governed by the *orthogonal dynamics* [20,28].

In situations with an invariant measure on phase space, and one uses a linear Mori projection [44], the resulting GLE is linear, and the memory kernel and the noise covariance are related by the fluctuation–dissipation theorem [36]. If the memory kernel  $H$  decays exponentially one can truncate the memory integral. Approximating the memory kernel  $H(s)$  in terms of particular families of functions [32,35,60,66] gives rise to autonomous, stochastic differential equations that approximate the evolution of the resolved degrees of freedom  $z(t)$ , and for a particular “nonlinear” Zwanzig projection which gives the evolution of the conditional expectation of the current state, i.e an optimal predictor for the resolved variables [13,14]. In this work, we will explore to what extent these methods carry over to situations without a non-trivial invariant measure on phase space.

This paper is organized as follows. In Sect. 2 we review some of the literature on modeling the evaporation of crude oil, and present a simplified model for this process. In Sect. 3 we review the Mori–Zwanzig projection operator formalism focusing on the discrete time setting as in Darve et al. [20]. Section 4 reviews some basic ideas in signal processing and then discusses various approaches to dimension reduction/stochastic modeling using autonomous/shift-invariant filters. In Sect. 5 we develop an analytic framework that allows us to design non-autonomous/time-varying reduced models for systems with slow relaxation. In Sect. 6 we apply our methods to weathering in oil spills and present a concluding discussion in Sect. 7.

## 2 A Dynamical System with Slow Relaxation

There is considerable interest in the evaporation process of crude oil, since this is an important process in the initial stages of an oil spill. Typical crude oil spills in the ocean can lose up to 40% or more of their oil to evaporation in the first few days before other important processes, e.g. the emulsification of oil in water, have a significant effect [24].

Theoretical approaches to modeling oil evaporation [23, 40, 53] are based on modifying equations for the (much better understood process of) evaporation of water [54]. Thermodynamics governs the process by which water molecules from the liquid enter the vapor phase at an air–water interface. However, this is not the limiting factor which determines the rate of evaporation. Rather, the rate of evaporation is regulated by the saturation of the air boundary-layer near the water surface. Indeed, dry air can hold up to a certain (temperature dependent) maximum amount of water vapor (the relative humidity cannot be more than 100%), and once the boundary layer reaches this level of saturation, the rate of evaporation is essentially governed by how quickly the water vapor can be transported (turbulent diffusion, laminar flow, etc) away from the water surface. Using similar ideas, the evaporation of each compound in oil can be modeled as

$$E/C \approx KTS$$

where  $E$  is the evaporation rate,  $C$  the concentration of the compound,  $K$  is a mass transfer rate,  $T$  is a coefficient that characterizes the turbulent/laminar transport of the vapor away from the interface and  $S$  is a factor that depends on the saturation of the boundary layer by the evaporating fluid (cf. Eq. (1) in [24]). The quantity  $\alpha = KTS$  is the evaporation rate constant, and various theoretical/empirical approaches to for obtaining expressions for  $T$  and  $S$  are discussed in detail by Fingas [24].

Based on these considerations, we investigate a simple, generic dynamical system that exhibits slow relaxation as a model for the evaporation of ‘oil’, thought of as a composite with many individual species. This system is solvable and thus offers a benchmark for measuring the performance of reduced dimensional models.

We will assume that ‘oil’ consists of  $I$  distinct species with concentrations  $c_i(t)$ ,  $i = 1, 2, \dots, I$  each decaying at a constant rate  $\alpha_i = K_i T S_i$ , to obtain

$$\partial_t c_i(t) = -\alpha_i c_i(t), \quad \alpha_i > 0, 1 \leq i \leq I. \quad (1)$$

We can also think of this system as describing non-interacting eigenmodes in a dissipative system that is relaxing to its equilibrium.  $I \gg 1$  will be assumed very large. It is in this situation that a low-dimensional approximation of the system is particularly useful. We can chose the indexing so that  $\alpha_{i+1} > \alpha_i$  (identifying “chemically distinct” species with the same evaporation rate as a single virtual species). The species with  $\alpha$  close to zero decay very slowly, so this model is similar in spirit to the approach of Amir et al. [1] for studying aging in glasses.

Note that, every concentration  $c_i$  is decaying in time, so the eventual state is one with no oil—the invariant measure is a singular measure corresponding to a point mass at 0. This system is dissipative, so the methods of equilibrium statistical mechanics of Hamiltonian systems do not directly apply. The projection operator formalism, however, *does not rely* on the underlying dynamics on phase space being Hamiltonian, and the Mori–Zwanzig approach is thus applicable [12].

We consider the case of a single source of measurements (i.e. a scalar function  $M(t)$ ). We assume that the observation  $M(t)$  is a weighted average of the concentrations  $c_i$ :

$$M(t) = \sum_i \beta_i c_i(t) = \sum_i \beta_i c_i(0) e^{-\alpha_i t}.$$

It is impractical/impossible to separately measure the concentrations/amounts  $c_i$  of *all* the individual species. One approach that naturally suggests itself is to use the measured quantity  $M(t)$  to extract the various decay rates  $\alpha_i$  using nonlinear fitting. *This approach will not work* [62]. Indeed, the discrete oil evaporation model (1) is a prototypical example of a sloppy model [62]. It is identical to the model for the mixture of radioactive nuclides considered in Waterfall et al. [62], and as they show, one cannot hope to extract the decay rates  $\alpha_i$ ,  $i = 1, 2, \dots, I$  from the measured function  $M(t)$  [56, 62].

We will therefore consider the complementary limit, where the number of distinct species distinct species  $I \gg 1$ . We will not attempt to identify the individual species in the mixture; rather we will use ideas from signal processing and statistical mechanics to make predictions for quantities that are robust [9] and insensitive to the precise composition of the oil. To this end, we will assume that  $\beta_i$  and  $\alpha_i$  depend “smoothly” on  $i$ , i.e. we can interpolate the values of  $\alpha_i$  and  $\beta_i$  to get functions that only vary on scales  $\Delta i \gg 1$ . Thus, we can replace the discrete index  $i$  by a continuous variable  $w$

$$\alpha_i = \alpha_{\min}(1 - w) + \alpha_{\max}w, \quad (2)$$

where  $\alpha_{\min} = \alpha_1$  is the evaporation rate of the least volatile species and  $\alpha_{\max} = \alpha_I$  is the evaporation rate of the most volatile species. A natural time scale for the system is

$$t_0 = \frac{1}{\alpha_{\max} - \alpha_{\min}}, \quad (3)$$

corresponding to the time by which the relative concentration

$$\frac{c_I(t_0)}{c_1(t_0)} = \frac{1}{e} \frac{c_I(0)}{c_1(0)}.$$

We will now obtain the equations for the continuum limit  $I \rightarrow \infty$  for time scales  $t \gtrsim t_0$  i.e. on time scales on which the relative concentrations of the various species vary significantly. Let  $i(w)$  connote a smooth, monotonic interpolation of the inverse of the function

$$w(i) = \frac{\alpha_i - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}},$$

and  $\rho(w, t)$  the smooth interpolation of the function

$$\rho(w(i), t) = e^{\alpha_{\min} t} \beta_i c_i(t) \cdot \frac{\alpha_{\max} - \alpha_{\min}}{\alpha_{i+1} - \alpha_i}.$$

A direct calculation now yields

$$M(t) = \sum_i \beta_i c_i(t) = e^{-\alpha_{\min} t} \sum_i \rho(w(i), t) [w(i+1) - w(i)].$$

Using the differential equation for  $c_i$ , we get

$$\partial_t \rho(w(i), t) = -(\alpha_{\min} + (\alpha_{\max} - \alpha_{\min})w(i)) \rho(w(i), t).$$

We rescale to a non-dimensional time  $\tilde{t} = \frac{t}{t_0} = (\alpha_{\max} - \alpha_{\min})t$  and consider the case  $\alpha_{\min} = 0$  to describe slow relaxation. Indeed, for crude oil in the environment,  $\alpha_{\min} \approx 0$  and it is in the scale of 1/decade.



Taking the (naive) continuum limit in the two previous equations, and dropping the tildes on  $\tilde{t}$  with the understanding that henceforth  $t$  is dimensionless, we obtain

$$\partial_t \rho(w, t) = -w\rho(w, t), \quad M(t) = \int_0^1 \rho(w, t) dw. \quad (4)$$

We will refer to this equation, along with an appropriate *random* initial condition  $\rho(w, 0)$  as the *linear evaporation process*. The only assumptions that went into the derivation of the continuum limit are, (1)  $I \gg 1$ ; (2)  $\alpha_i$  and  $\beta_i$  vary on scales  $\Delta i \gg 1$ . The continuum limit equations are *independent* of the precise details of  $\alpha_i$  and  $\beta_i$ , which are absorbed into the change of variable from  $c_i(t)$  to  $\rho(w, t)$ .

For a complete specification of the problem, we need to characterize the statistics of the initial measure  $\rho(w, 0)$ , which is random, reflecting the uncertainties in the initial composition of the oil. In particular, the statistics of  $\rho(w, 0)$  should be inferred from the statistics of the concentrations  $c_i(0)$  (of the discrete species), but one needs some attention to manner in which we take the limit  $I \rightarrow \infty$  so that we do not get a deterministic (instead of random) limit for  $\rho(w, 0)$ . This would indeed be the case, from the central limit theorem, if the various  $c_i(0)$  were i.i.d random variables with finite variance and distributions independent of  $I$ . This issue is somewhat subtle, and we address it in Sect. 5.

We can explicitly solve (4) to get  $\rho(w, t) = \rho(w, 0)e^{-wt}$ . This is the Schrödinger picture of the evolution of the system, in which the measure associated with the state of the oil changes with time.

The Heisenberg picture arises from considering *linear observables* given by pairing the measure  $d\mu^t \equiv \rho(w, t)dw$  with (an appropriate subset of) continuous functions on  $[0, 1]$ . Every continuous function  $g$  is associated with a linear observable  $G$  given by

$$G(t) = \int g(w)\rho(w, t)dw.$$

The total mass  $M(t)$  is thus the observable associated with the constant function  $f(x) = 1$ . In the Heisenberg picture we have

$$G(t) = \int g(w)\rho(w, t)dw = \int g(w)e^{-wt}\rho(w, 0)dw = \int g^t(w)d\mu^0(w). \quad (5)$$

The evolution of the continuous function associated with an observable  $G$  is given by  $g^t(w) = g(w)e^{-wt}$  so that  $\partial_t g^t(w) = -wg^t(w)$ .

We note the contrast between our system (4), and the typical situation of a dynamical system  $\dot{x} = f(x)$  on a high-dimensional phase space  $\Sigma$ . For the dynamical system, the Heisenberg picture is given by evolving continuous functions on  $\Sigma$  through  $g^t(x) = g(\varphi(x, t))$  where  $\varphi$  is the solution map, i.e  $\varphi(x, 0) = x$  and  $\partial_t \varphi(x, t) = f(\varphi(x, t))$ . In particular, the constant function 1 is an invariant under this evolution.  $g(x) = 1$  for all  $x$  in  $\Sigma$  implies that  $g^t(x) = 1$  for all  $x$  and  $t$ . In contrast, the evolution in (5) which gives  $g^t(w) = e^{-wt}$ . Consequently, the continuum limit (4) is *not the Liouville equation* for a dynamical system.

We will henceforth work in a discrete time setting, that can be viewed as a Takens delay-coordinate embedding [55] of the continuous time system. There are many reasons to do this, including the difficulty in parametrizing continuous time stochastic processes [15, 39] and the fact that, for our application to oil spills, the sensor data is only obtained at discrete time intervals. We can recast (4) and (5) as maps in discrete time by defining  $t = n\tau$ ,  $\rho_n(w) \equiv$



$\rho(w, n\tau)$ . With these substitutions,

$$\begin{aligned}\rho_{n+1}(w) &= \Lambda^T \rho_n(w), \\ g^{(n+1)\tau}(w) &= \Lambda g^{n\tau}(w),\end{aligned}\tag{6}$$

where  $\Lambda$  is a bounded operator on  $C([0, 1])$  that takes a continuous function  $g(w)$  to  $\Lambda g(w) \equiv e^{-w\tau} g(w)$ , and  $\Lambda^T$  is the adjoint on the dual space of measures on  $[0, 1]$ . Note that  $\Lambda$  extends naturally to a self-adjoint operator on  $L^2([0, 1])$  also defined by  $\Lambda h(w) = e^{-w\tau} h(w)$  for all  $h \in L^2$ , so we can also consider the (larger) set of observables given by  $L^2$  functions on  $[0, 1]$ . In this case the density  $\rho$  is also in  $L^2$  and the evolution is given by a self-adjoint operator  $\Lambda^T = \Lambda$  on  $L^2([0, 1])$ . Although the maps in (6) are not the transfer operator [3] (respectively the Koopman operator [10]) corresponding to a dynamical system, they have the same formal structure so we will attempt to use discrete time projection operator techniques for model reduction [15, 38, 39].

### 3 The Mori–Zwanzig Projection Formalism

We first present a short review of the discrete-time Mori–Zwanzig projection formalism following the presentation in Darve et al. [20]. The setup is as follows:  $\mathcal{H}$  is a Hilbert space and  $\Lambda : \mathcal{H} \rightarrow \mathcal{H}$  is a linear operator on this space. We can think of  $\Lambda$  as  $e^{\tau\mathcal{L}}$  where  $\tau$  is a ‘time-step’ and  $\mathcal{L}$  is the Liouville operator (the generator) evolving measures on phase space  $\Sigma$  and  $\mathcal{H} \subseteq \mathcal{M}(\Sigma)$  so every element of  $\mathcal{H}$  can be interpreted as a (signed) measure on  $\Sigma$ . Linear observables are given by linear operators  $g : \mathcal{H} \rightarrow \mathbb{R}$ , so the set of linear observables is the dual  $\mathcal{H}^* = \mathcal{H}$ . A (general, nonlinear) observable is any (measurable) function of a finite collection of linear observables, so the observables form an algebra of mappings  $\mathcal{H} \rightarrow \mathbb{R}$ . Finally,  $P : \mathcal{H} \rightarrow \mathcal{H}$  is an orthogonal projection and  $Q = I - P$  is the complementary projection. We will use the bra-ket notation and represent states (elements of  $\mathcal{H}$ ) by ket-vectors and linear observables by bra-vectors.

We consider the discrete time dynamical system  $|\rho_{n+1}\rangle = \Lambda|\rho_n\rangle$ . We decompose  $|\rho_n\rangle = |\xi_n\rangle + |\eta_n\rangle$  where  $|\xi_n\rangle = P|\rho_n\rangle$  (the observations) and  $|\eta_n\rangle = Q|\rho_n\rangle$ . An elementary argument by induction shows that

$$\begin{aligned}|\rho_n\rangle &= |\xi_n\rangle + |\eta_n\rangle = |\xi_n\rangle + Q\Lambda(|\xi_{n-1}\rangle + |\eta_{n-1}\rangle) \\ &= |\xi_n\rangle + Q\Lambda|\xi_{n-1}\rangle + (Q\Lambda)^2(|\xi_{n-2}\rangle + |\eta_{n-2}\rangle) \\ &= |\xi_n\rangle + Q\Lambda|\xi_{n-1}\rangle + (Q\Lambda)^2|\xi_{n-2}\rangle + \cdots + (Q\Lambda)^n|\xi_0\rangle + (Q\Lambda)^n Q|\rho_0\rangle.\end{aligned}\tag{7}$$

For an observable  $G_n = \langle g|\rho_n\rangle$ , we therefore obtain

$$G_n = \sum_{k=0}^n \langle g|(Q\Lambda)^k|\xi_{n-k}\rangle + \langle g|(Q\Lambda)^n Q|\rho_0\rangle.\tag{8}$$

In the Heisenberg picture,  $G_n = \langle g_n | \rho_0 \rangle$ , and from (8) it follows that

$$\begin{aligned} \langle g_n | &= \sum_{k=0}^n \langle g | (Q\Lambda)^k P \Lambda^{n-k} + \langle g | Q(\Lambda Q)^n \\ &= \langle g | P \Lambda^n + \sum_{k=1}^n \langle g | Q(\Lambda Q)^{k-1} \Lambda P \Lambda^{n-k} + \langle g | Q(\Lambda Q)^n \\ &= \langle g | P \Lambda^n + \sum_{k=1}^n \langle F_{k-1} | \Lambda P \Lambda^{n-k} + \langle F_n |, \end{aligned} \quad (9)$$

where we have defined  $\langle F_k | = \langle g | Q(\Lambda Q)^k$ . Equation (9) is identical to Eq. (6) in Ref. [20]. It follows that  $\langle F_k | \xi_j \rangle = 0$  for all  $j$  and  $k$  since  $QP = 0$ . For this reason,  $\langle F_k |$  is usually treated as ‘noise’, although, in principle, one can characterize  $\langle F_k |$  through solutions of the the *orthogonal dynamics* [20, 28] (see also Appendix 2)

$$\langle F_{n+1} | = \langle F_n | \Lambda Q, \quad \langle F_0 | = \langle g | Q. \quad (10)$$

Equation (9) is the discrete time Mori–Zwanzig decomposition and (11) below is the adjoint, which evolves the states instead of the observables. These equations are identities and are often taken as starting points for building methods to estimate quantities that are not directly observed, in terms of quantities  $\xi_n, \xi_{n-1}, \dots, \xi_0$  that have been observed by time  $n$ .

A problem of significant interest is *prediction*, i.e. estimating  $\xi_n$  using the information available at time  $n-1$ , which are the quantities  $\xi_{n-1}, \xi_{n-2}, \dots, \xi_0$ . Using (7) with  $n \rightarrow n-1$  and  $|\xi_n\rangle = P\Lambda|\rho_{n-1}\rangle$ , we obtain

$$|\xi_n\rangle = \underbrace{P\Lambda|\xi_{n-1}\rangle}_{\text{Markovian}} + \underbrace{\sum_{k=2}^n P\Lambda(Q\Lambda)^{k-1}|\xi_{n-k}\rangle}_{\text{memory}} + \underbrace{P\Lambda(Q\Lambda)^{n-1}Q|\rho_0\rangle}_{\text{noise}}, \quad (11)$$

where the right hand side is decomposed into the Markovian term, the “optimal” estimate of  $|\xi_n\rangle$  given the current state  $|\xi_{n-1}\rangle$ , the memory term that encodes the dependence on the past observations  $|\xi_{n-2}\rangle, |\xi_{n-3}\rangle, \dots, |\xi_0\rangle$ , and the noise, which is orthogonal to  $|\xi_j\rangle$  and depends on the microscopic details of the initial condition, i.e it depends on  $|\rho_0\rangle$  and not just on  $|\xi_0\rangle$ . Of course, an important caveat here is that the interpretation of the decomposition as Markovian, memory and noise terms relies on the origins of this procedure in the near-equilibrium statistical mechanics context, and *it is by no means clear that this interpretation is valid for the evaporation process* (4).

Nonetheless, the equation is a formally exact decomposition of a PDE with stochastic initial conditions into a part that only depends on a subset of the degrees of freedom, *the resolved variables*, along with an exact expression for the remainder. In what follows, we will lump the Markovian and the memory terms into a single quantity, so the distinction between the resulting two terms is whether or not they only depend on the observed quantities  $|\xi_j\rangle$ , or the entire (microscopic) initial condition  $|\rho_0\rangle$ .

### 3.1 The Memory Kernel for the Weathering of Oil

We now return to the evaporation model. We assume there is a single observed quantity,  $M_n = \int \rho_n dw$ . We take  $\mathcal{H} = L^2([0, 1])$ , the space of square integrable functions on  $[0, 1]$ . Defining  $|1\rangle$  to denote the constant function  $g(x) = 1$ , we have

$$M_n = \langle 1 | \rho_n \rangle, \quad \langle 1 | 1 \rangle = 1.$$

The orthogonal projection on to the one-dimensional space spanned by the constant functions is given by  $P = |1\rangle\langle 1|$ . Consequently,  $|\xi_n\rangle = P|\rho_n\rangle = |1\rangle\langle 1|\rho_n\rangle = M_n|1\rangle$  and  $M_n = \langle 1|\xi_n\rangle$ . Using this in (11) and projecting on to constants gives

$$M_n = \sum_{k=1}^n h_k M_{n-k} + \beta_n, \quad (12)$$

where  $h_k = \langle 1 | (\Lambda Q)^{k-1} \Lambda | 1 \rangle$  and  $\beta_n = \langle 1 | (\Lambda Q)^n | \rho_0 \rangle$  is the ‘noise’ that depends explicitly on the microscopic initial condition  $|\rho_0\rangle$ . We can also obtain the same equation from the (usual) Mori–Zwanzig decomposition in (9) by taking  $\langle g | = \langle 1 | \Lambda$ .

Equation (12) is exact. In particular, it holds for  $|\rho_0\rangle = |1\rangle$ , in which case the noise vanishes,  $\beta_n \equiv 0$  for all  $n$ . If  $|\rho_0\rangle = c|1\rangle$ , for some constant  $c$ ,  $M_0 = \langle 1 | c | 1 \rangle = c$  and we can explicitly solve (4) to obtain

$$M_n = M_0 \int_0^1 e^{-wn\tau} dw = \begin{cases} M_0 \frac{1-e^{-n\tau}}{n\tau} & n \geq 1, \\ M_0 & n = 0. \end{cases} \quad (13)$$

Consequently, the memory kernel  $h_k$  is determined by

$$\frac{1 - e^{-n\tau}}{n\tau} = \sum_{k=1}^n h_k \frac{1 - e^{-(n-k)\tau}}{(n-k)\tau} \quad \text{for all } n \geq 1. \quad (14)$$

We can solve for  $h_k$  using the  $\mathcal{Z}$ -transform (equivalently the generating function). Let  $\hat{M}(z) = \sum_{n=0}^{\infty} M_n z^{-n}$  and  $\hat{H}(z) = \sum_{n=0}^{\infty} h_n z^{-n}$ . The sum defining  $\hat{M}(z)$  converges for  $z$  outside the unit disk since  $M_n$  is clearly a decreasing sequence. We can compute the sum explicitly to obtain

$$\hat{M}(z) = M_0 \sum_{n=0}^{\infty} \int_0^1 e^{-wn\tau} z^{-n} dw = M_0 \int_0^1 \frac{ze^{w\tau}}{ze^{w\tau} - 1} dw = \frac{M_0}{\tau} \log \left[ \frac{ze^{\tau} - 1}{z - 1} \right].$$

Multiplying (14) by  $z^{-n}$  and summing on  $n \geq 1$  gives

$$\hat{M}(z) - M_0 = \hat{M}(z) \hat{H}(z), \quad (15)$$

and rearranging yields

$$H(z) = \left[ 1 - \frac{M_0}{\hat{M}(z)} \right] = \left[ 1 - \frac{\tau}{\log(e^{\tau} z - 1) - \log(z - 1)} \right]. \quad (16)$$

$H$  is analytic outside the unit circle and has a branch point singularity at  $z = 1$ . Expanding about  $z = \infty$  gives

$$H(z) = z^{-1} \frac{1 - e^{-\tau}}{\tau} + z^{-2} \frac{(1 - e^{-\tau})((\tau - 2) + (\tau + 2)e^{-\tau})}{2\tau^2} + \dots,$$

so that the coefficients  $h_k$  can be explicitly computed. Our interest is in the long time behavior of  $h_k$ , which can be deduced from the  $z \rightarrow 1$  behavior of  $\hat{H}(z)$ .  $\hat{H}(z)$  has a logarithmic branch point at  $z = 1$ . In particular, this implies that the series for  $\hat{H}(z)$  does not converge for any  $z$  with  $|z| < 1$ , so that the sequence  $h_k$  decays slower than the exponential  $e^{-\epsilon k}$  for any  $\epsilon > 0$ .

The transfer operator methods in Flajolet and Odlyzko [25] (Theorem 3A and comments on pp. 231–232) imply, in fact, that

$$h_k \sim \frac{1}{k \log^2(k)} \quad \text{as } k \rightarrow \infty, \quad (17)$$

so that  $h_k$  decays algebraically. Although  $\sum h_k$  converges to  $\hat{H}(1) = 1$  (by (16)), the partial sums go to 1 extremely slowly,  $\left|1 - \sum_{k=1}^N h(k)\right| \sim \log(N)^{-1}$ .

The memory kernel  $h_k$  thus has a fat tail. The algebraic decay of  $h_k$  is a reflection of the extremely slow relaxation in  $\rho(w, t) = \rho_0(w)e^{-wt}$  for species with  $w$  close to zero. In general, the initial condition  $\rho_0(w)$  has an effect for times of order  $1/w$ , so the initial condition is not “forgotten” for long times, leading to the fat tails and slow decay of correlations.

## 4 Dimension Reduction, Stochastic Modeling and Filtering

We are interested in *model reduction*, i.e. in developing low dimensional (approximate) models for predicting the behavior of high dimensional complex systems, e.g. the linear evaporation process (4). Before we describe our work on this problem, we first review some basic terminology from signal processing, and then present a roadmap to guide the reader through our various approaches to the problem of prediction/model reduction for (4).

### 4.1 Filtering, Estimation and Prediction

In our context, the general prediction/estimation problem along with data assimilation is the following: The sequence  $\{\rho_k\}$  describes the “state” of the system (4) sampled at discrete times  $t = k\tau$ . We are given a sequence of noisy measurements  $\tilde{M}_k = \langle 1 | \rho_k \rangle + \sigma \gamma_k$  where the  $\gamma_k$  are uncorrelated normal variates. What is the “best” prediction for  $M_n = \langle 1 | \rho_n \rangle$  in terms of the measurements  $\tilde{M}_k$  for  $k < n$ ? Abstractly, the optimal estimate is given by a conditional expectation

$$\tilde{M}_n = \mathbb{E}[M_n | \tilde{M}_{n-1}, \tilde{M}_{n-2}, \dots, \tilde{M}_1, \tilde{M}_0].$$

We seek a concrete representation for the optimal estimator, i.e. a (sequence of) explicit functions  $F_n$  such that

$$\mathbb{E}[M_n | \tilde{M}_{n-1}, \tilde{M}_{n-2}, \dots, \tilde{M}_1, \tilde{M}_0] \approx F_n(\tilde{M}_{n-1}, \tilde{M}_{n-2}, \dots, \tilde{M}_j, \dots).$$

with  $\tilde{M}_j = 0$  for  $j \leq 0$ . We will call such functions  $F_n$  *filters* or *predictors*. We can classify filters by the following properties:

1. The filter is *autonomous* or *shift-invariant* if  $F_n \equiv F$  independent of  $n$ .
2. If  $F_n$  only depends on  $\tilde{M}_{n-1}, \tilde{M}_{n-2}, \dots, \tilde{M}_{n-L}$  for some finite  $L$ , then it is a *finite impulse response (FIR)* filter with  $L$  *taps*. Otherwise, the filter is an *finite impulse response (IIR)* filter that uses information from the entire time history of the time series  $\tilde{M}_k$ .
3. A filter  $F_n$  is *linear* if it is given by a linear function of its arguments.
4. A filter  $F_n$  is *genie-aided* if it has access to more information than is available in  $\tilde{M}_{n-1}, \tilde{M}_{n-2}, \dots$ . Such filters cannot be built in practice. Nonetheless, as with the Maxwell demon, this fictional construct is useful because it allows us to bound the best-case behavior of constructible filters.
5. A filter  $F_n$  is *empirical* or *data-driven* if it is obtained through regression on many realizations of the underlying random process  $\tilde{M}_k$ .

**Table 1** A summary of the various filters we will consider in this work, along with a description of their features

Filter	Definition	Linear	Data driven	Other features
Memory-kernel/transfer function based methods				
MZ filter	Eq. (18)	Yes	No	Needs all history
Truncated MZ (FIR)	Eq. (19)	Yes	No	
MZ-Pade	Eq. (20)	Yes	No	
Harmonic filter	Eq. (23)	No	No	
Statistical regression based methods				
Linear Oracle	Sect. 4.4	Yes	Yes	Genie-aided
Empirical linear	Eq. (24)	Yes	Yes	Averaged over runs
Empirical harmonic	Eq. (25)	No	Yes	Averaged over runs
Methods that exploit slow decay of correlations				
Asymptotic filter	Eq. (33)	Yes	No	Non-autonomous
Universal filter	Eq. (35)	Yes	No	Unstable
Extended asymptotic	Eq. (37)	No	No	Hidden variables

All but the Asymptotic filter are shift-invariant, and all but the MZ filter have finitely many taps

Table 1 describes our various approaches to building filters for the linear evaporation process (4). At the gross level, there are three distinct approaches. The first approach is based on the memory kernel (16), or equivalently, the single realization corresponding to  $|\rho_0\rangle = |1\rangle$  given by (13). These filters are described in Sects. 4.2 and 4.3. The second approach, discussed in Sect. 4.4, is empirical and relies of estimating coefficients in filter functions using statistical regression on independent realizations of the random process (4). The final approach is non-empirical, and exploits the slow relaxation inherent in the process  $M_k$ . In this case, the slow decay of the memory kernel (17) is beneficial, rather than detrimental, contrary to intuition. This approach is discussed in Sect. 5.

We also note that solving the filtering/prediction problem is very closely related to obtaining reduced models for the high-dimensional system (4). If  $F_n$  is a (close to) optimal filter, then the process

$$\widehat{M}_n = F_n(\widehat{M}_{n-1}, \widehat{M}_{n-2}, \dots, \widehat{M}_j, \dots) + \theta_n,$$

where the quantities  $\theta_n$  are stochastic with the appropriate statistics, is a good surrogate for the high dimensional process that generates  $M_n$ . This reduction is particularly efficient if the filter  $F_n$  is shift-invariant and has finitely many taps. Indeed, this is the framework in which the Mori–Zwanzig projection operator formalism is used to build reduced models for various high-dimensional systems [15, 29, 39].

A natural question is: “*Why consider multiple approaches?*” We do this because we have good analytical understanding of the ‘high-dimensional’ dynamics of (4), so we have a good theoretical basis for assessing the performance of many of the popular approaches to stochastic modeling/dimension reduction. We are able to evaluate the relative merits of the various assumptions/approximations that are inherent in the different approaches. Finally, we are able to develop an analytic framework that gives new approaches to model reduction for high dimensional systems with long term memory, and one that is applicable to practical problems.

In order to assess the performance of our various filters, as well as to generate the empirical filters by regression, we need realizations of the evaporation process (4) with random initial data  $|\rho_0\rangle$  drawn from an appropriate distribution. We numerically generate such realizations as follows:

**Algorithm I** Generating synthetic data

1. We discretize the interval  $[0, 1]$  into  $I$  equal intervals of size  $\Delta = 1/I$ . For our simulations we take  $I = 1000$ .
2. We assign the initial mass distribution by picking  $I$  independent random variables  $u_i$ ,  $i = 1, 2, \dots, I$  uniformly distributed random variables on  $[0, 1]$  and then normalize to set

$$\rho_0(i) = \frac{u_i}{\Delta \sum_{i=1}^I u_i}.$$

By symmetry, the marginal distributions of the quantities  $\rho_0(i)$  are identical, but they do depend on  $I$ , the “total number of species”. They are however not independent. By construction  $\sum_i \rho_0(i) = I$ . To the interval  $[(i-1)\Delta, i\Delta]$ , indexed by  $i$ , we associate the decay rate

$$w(i) = \left(i - \frac{1}{2}\right) \Delta, \quad i = 1, 2, \dots, I,$$

corresponding to the middle of the interval.

3. We pick  $\tau = \log(3/2)$  so that  $\rho_n(i) = \rho_0(i) \left(\frac{2}{3}\right)^{nw(i)}$ .
4. We compute  $M_n = \sum \rho_n(i) \Delta$  for  $1 \leq n \leq N$ , where we choose  $N$  such that the assumption that we are discretizing a continuum density using  $I$  intervals is still valid. This requires that there are at least  $\sim 10$  intervals for which the density  $\rho_n(i)$  has not decayed down to zero. This gives the rule of thumb  $\log(3/2)w(10)N \sim 1$  so that  $N \sim \frac{I}{10 \log(3/2)} \sim \frac{I}{4}$ . We can thus safely take  $N = 200$ .  $\square$

The numerical procedure is very close in spirit to the original discrete model (1) with  $I$  distinct species. The one difference is that, since we are discretizing a continuum limit, we can pick the decay rates  $w(i)$  on the basis of our discretization, and not through any relation with the “true” decay rates of the components of oil. This also ties in with the idea that the individual decay rates in the mixture cannot be identified, and our methods have to be robust to possible changes in the underlying “bare” decay rates.

Through this procedure we obtain many random realizations of (a discretization of) the system in (4) with  $M_0 = 1$ . We use the computed values of  $M_k$  as the “measurements”  $\bar{M}_k$  in estimating  $M_n$  from the measurements for  $k < n$ . In particular, we will assume there is no measurement noise.

## 4.2 The Mori Projection and Linear Autonomous Estimators

Equation (12) is *exact* (see also (41) in Appendix 1) and gives a stochastic reduced model of the system (4) on replacing the quantities  $\beta_n$  (determined by the microscopic initial condition  $|\rho_0\rangle$ ) with a stochastic process  $\theta_n$ , typically a Gaussian process, that has the same “statistics”, i.e. we match the means and the covariances

$$\mathbb{E}[\beta_n] = \mathbb{E}[\theta_n] = 0, \quad \mathbb{E}[\beta_n^T \beta_m] = \mathbb{E}[\theta_n^T \theta_m], \quad \text{for all } m, n \geq 0,$$

where the expectations  $\mathbb{E}$  for  $\beta$  are over a natural measure for the initial conditions and the expectations for  $\theta$  are over the measure underlying the stochastic process  $\theta_n$ . Equation (12) thus gives the stochastic model

$$M_n = \sum_{k=1}^n h_k M_{n-k} + \theta_n.$$

Consequently, we also have the associated prediction/filtering algorithm

$$\tilde{M}_n = \sum_{k=1}^n h_k \tilde{M}_{n-k}, \quad (18)$$

which we will call the MZ filter. The MZ filter is shift-invariant (autonomous), but nonetheless evaluating the sum in (18) requires us to keep track of the entire history of  $\tilde{M}_n$ . As we argued above, since  $h_k$  has a fat tail, one cannot simply truncate the sum at a fixed  $L$  and expect to get good results.

Figure 1 compares the performance of three potential estimators. The first estimator does not use any data assimilation, so the predicted sequence  $\tilde{M}_n$  is given by (13). The second estimator truncates the sum in the Mori–Zwanzig decomposition at  $L = 6$  where the value 6 has no particular significance and is chosen purely for the purposes of illustrating the effects of truncating the sum. Naive truncation gives the estimator  $\tilde{M}_n \approx \sum_{k=1}^L h_k M_{n-k}$ , which is biased at  $O(n^{-1})$ , because the quantities  $h_k M_{n-k}$  have positive means and their sum over  $n - L \leq k < n$  is  $O(n^{-1})$ . We can attempt to eliminate this bias by an *ad hoc* “renormalization” of the weights

$$h'_k = \frac{h_k}{\sum_{k=1}^L h_k}, \quad \tilde{M}_n = \sum_{k=1}^L h'_k \tilde{M}_{n-k} \quad (19)$$

so that  $\sum_{k=1}^L h'_k = \sum_{k=1}^{\infty} h(k) = 1$ , and the estimator has a bias  $O(n^{-2})$  and is thus ‘better’ in the limit  $n \rightarrow \infty$ . We will call this the *truncated FIR* (Finite impulse response) filter in contrast to the third estimator, (18) which is an *IIR* (Infinite impulse response) filter that is obtained from the Mori–Zwanzig decomposition and incorporates the entire history of  $M_n$ .

For each estimator, we define the “inferred noise” or the one-step prediction error  $\varepsilon_n$  as the difference  $|\tilde{M}_n - M_n|$  between the estimate  $\tilde{M}_n$  using information available at time  $n - 1$  and the (random) value  $M_n$  (“the truth”) for a realization. We display these differences for a single ‘typical’ realization in Fig. 1. Figure 2 show the averaged error over many realizations.

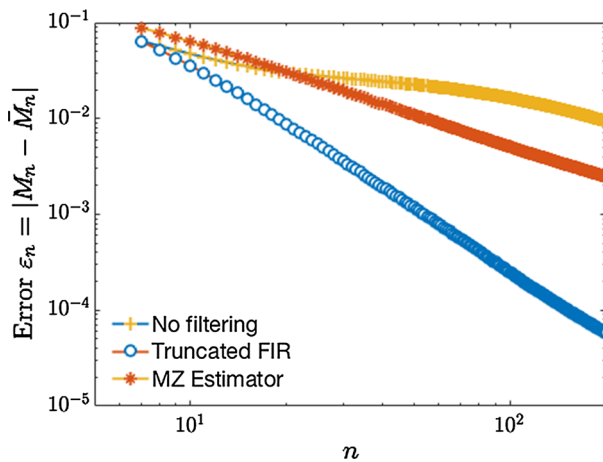
Very surprisingly, the MZ estimator using the entire history does worse than the truncated MZ estimator with renormalization and, on average, *also worse than the estimator without any data assimilation*.

An alternative approach to truncating the sum in the Mori–Zwanzig estimator is to work in the  $\mathcal{Z}$ -transform domain and approximate  $\hat{H}(z)$  in (15) by a rational function in  $z^{-1}$  [46], i.e.  $\hat{H}(z) \approx p(z^{-1})/q(z^{-1})$  where  $p$  and  $q$  are polynomials of degrees less than or equal to  $L$ , and we normalize by requiring that  $q(0) = 1$ . Since  $h_0 = 0$ , it follows that  $p(0) = 0$ . The  $\mathcal{Z}$ -transform of the sequence  $\tilde{M}_n$  of estimates given by

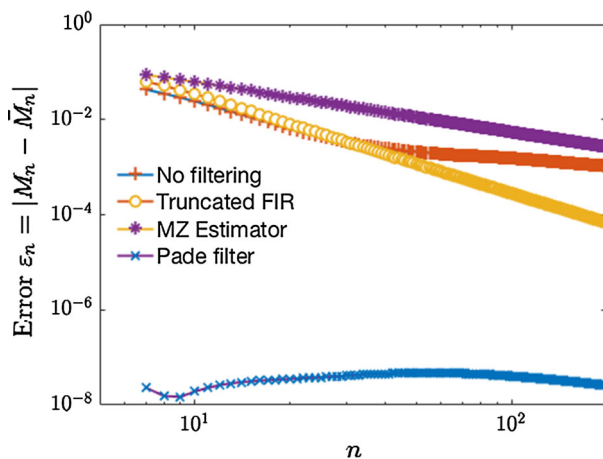
$$\sum_{n=0}^{\infty} \tilde{M}_n z^{-n} \equiv \hat{\tilde{M}}(z) \approx \frac{M_0}{1 - \frac{p(z^{-1})}{q(z^{-1})}} = M_0 \frac{q(z^{-1})}{q(z^{-1}) - p(z^{-1})}.$$

If we define  $b(z^{-1}) = q(z^{-1}) - p(z^{-1})$ ,  $b$  is also a polynomial with the normalization  $b(0) = 1$ . Thus, we get an auto-regressive AR( $L$ ) model [8, 37]  $b(z^{-1}) \hat{\tilde{M}}(z) = M_0 q(z^{-1})$ . Writing  $b(z^{-1}) = 1 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_j z^{-j}$  and  $q(z^{-1}) = 1 + q_1 z^{-1} + q_2 z^{-2} + \dots + q_j z^{-j}$ , we have the estimator





**Fig. 1** The inferred error for the three estimators used on one random realization. The estimator with no data assimilation (yellow) is given by (13), the renormalized FIR filter (blue) is given by (19) and the Mori–Zwanzig estimator (red) that uses the entire history is given by (18) (Color figure online)



**Fig. 2** The inferred error averaged over 100 realizations. These curves are very stable and do not vary discernibly between sets of 100 independent realizations. Three of the estimators are the same as in Fig. 1. We also compare the FIR filter generated by the [6,6] Padé approximant of the Mori–Zwanzig transfer function  $H(z)$ . The one-step prediction error for the Padé filter is substantially smaller than the errors in the other estimators

$$\bar{M}_n = M_0 q_n - \sum_{k=1}^n b_k \bar{M}_{n-k},$$

where we have used the convention  $b_k = 0$  (resp.  $q_k = 0$ ) for indices  $k$  greater than the degrees of the respective polynomials. The sum on the right hand side therefore has no more than  $L$  non-zero terms. Of course, if the estimator is ‘good’, then  $\bar{M}_n$  the estimate for  $M_n$  using information available prior to time  $n$  is close to the true value  $M_n$ . In fact, one might argue that the Mori–Zwanzig decomposition (12) is exact for the “true” sequence  $M_n$  and thus one would do better (or certainly not much worse) by replacing the estimates  $\bar{M}_{n-k}$  by

their measured values  $M_{n-k}$  since we are assuming there is no measurement error. This gives the estimator

$$\bar{M}_n = M_0 q_n - \sum_{k=1}^n b_k M_{n-k}. \quad (20)$$

One can view this as an alternate renormalization of the weights  $h_k$  in a truncated MZ estimator, one that is perhaps better justified and less ad hoc than the choice  $h'_k = h_k (\sum_{k=0}^L h_k)^{-1}$  from above. In Fig 2, we present the results using the [6,6] Padé approximant of  $H(z)$  about  $z = \infty$  to obtain the 6th order polynomials  $p$  and  $q$ , which then give a 6 tap filter (20) for predicting  $M_n$  from  $M_{n-1}, M_{n-2}, \dots, M_{n-6}$ . Very surprisingly, the Padé filter, which is ostensibly designed to approximate the MZ estimator through a filter with finitely many delays, *performs significantly better* than the MZ estimator (18).

### 4.3 Nonlinear Filters

Since the evaporation process (4) is linear, the equations do not have a natural scale for the mass  $M$ . In particular, the total amounts of oil is an extensive quantity whose (scale-free) logarithmic derivative  $\partial_t M(t)/M(t)$  should only depend on the relative fractions of the different species and not their total amounts. It follows that the dynamics, and consequently also the estimators, should be homogeneous of degree 1, i.e if the estimator of  $M_n$  is given by  $\bar{M}_n = F_n(M_{n-1}, M_{n-2}, \dots, M_1, M_0)$ , then  $F_n$  must satisfy  $F_n(\lambda M_{n-1}, \dots, \lambda M_1, \lambda M_0) = \lambda F_n(M_{n-1}, \dots, M_1, M_0)$  for all  $\lambda > 0$ . It is easy to verify that all the linear estimators in Sect. 4.2 have this property.

Note also that the Mori–Zwanzig decomposition (12) is exact, and further, the noise in the equation is exactly zero for choices of the initial condition  $|\rho_0\rangle$  in the range of the projection  $P = |1\rangle\langle 1|$ . Consequently, the estimators in the previous section were designed to (exactly or approximately) recover the sequence  $M_n$  for initial conditions in the range of  $P$ . We can also seek nonlinear estimators with the same property.

The motivation for considering nonlinear estimators that the process  $M_n$  behaves (roughly) like  $1/n$ , so no finite-lag ARMA filter can generate such a process. Indeed, the impulse response of such a filter is determined by its poles [46] and consists of sums of sequences of the form  $n^l z_j^{-n}$  where  $l < \text{order of the pole } z_j$  of the filter transfer function. On the other hand, for the sequence in (13),  $M_n^{-1} \sim n$  with exponentially small corrections. The sequence  $M_n^{-1} = n$  can indeed be generated by an ARMA filter [8], in particular, by a 2nd order pole at  $z_j = 1$ .

This naturally leads us to filter  $M_n^{-1}$  by seeking weights  $v_k$  such that

$$\frac{1}{M_n} = \sum_{k=1}^n \frac{v_k}{M_{n-k}}$$

exactly for the sequence (13), and then use this as a starting point for truncation as in the previous Sect. 4.2. We have, for  $|\rho_0\rangle = |1\rangle$ ,

$$\frac{1}{M_n} = \begin{cases} \frac{n\tau}{M_0(1-e^{-n\tau})} = \frac{n\tau}{M_0} + \frac{n\tau e^{-n\tau}}{M_0} + \frac{n\tau e^{-2n\tau}}{M_0} + \dots & n \geq 1, \\ \frac{1}{M_0} & n = 0. \end{cases} \quad (21)$$

The  $\mathcal{Z}$ -transform of the sequence  $M_n^{-1}$  is given by

$$T(z) = \sum_{n=0}^{\infty} z^{-n} \frac{1}{M_n} = \frac{1}{M_0} \left[ 1 + \frac{z^{-1}\tau}{(1-z^{-1})^2} + \sum_{j=1}^{\infty} \frac{z^{-1}\tau e^{-j\tau}}{(1-e^{-j\tau}z^{-1})^2} \right].$$

Note that, in contrast to the  $\mathcal{Z}$ -transform for  $M_n$ , which has logarithmic branch points at  $z = 1$  and  $z = e^{-\tau}$ , the  $\mathcal{Z}$ -transform for  $M_n^{-1}$  has poles of order 2 at  $z = e^{-j\tau}$ ,  $j = 0, 1, 2, \dots$  and an essential singularity at  $z = 0$ .

As before, we can compute the sequence  $v_k$  through its  $\mathcal{Z}$ -transform by

$$S(z) = \sum_{k=0}^{\infty} v_k z^{-k} = z \left[ 1 - M_0^{-1} T(z)^{-1} \right].$$

This form is not directly useful, since we do not know the zeros of  $T(z)$ , which correspond to the poles of  $S(z)$ , which in turn determine the asymptotic behavior of  $v_n$ .

We will instead use an alternative approach that directly determines rational approximations to  $T(z)$  which can then be used to generate finite lag estimators for  $M_n^{-1}$ . Since the series expansion for  $\frac{1}{M_n}$  in (21) converges exponentially, we can truncate the sum at order  $e^{-mn\tau}$  to obtain

$$M_0 T(z) \approx 1 + \frac{z^{-1}\tau}{(1-z^{-1})^2} + \frac{z^{-1}\tau e^{-\tau}}{(1-e^{-\tau}z^{-1})^2} + \dots + \frac{z^{-1}\tau e^{-m\tau}}{(1-e^{-m\tau}z^{-1})^2},$$

a rational function approximation, where the error in this approximation is uniformly bounded on the unit circle. Given a rational approximation to  $T(z) = M_0^{-1} q(z^{-1})/b(z^{-1})$ , we can design a linear predictor for  $M_n^{-1}$  as above (cf. (20)). We illustrate this method with explicit calculations for  $m = 1$ :

$$\begin{aligned} T(z) &\approx \frac{1}{M_0} \left[ 1 + \frac{z^{-1}\tau}{(1-z^{-1})^2} + \frac{z^{-1}\tau e^{-\tau}}{(1-e^{-\tau}z^{-1})^2} \right] \\ &= \frac{1}{M_0} \frac{((1-z^{-1})^2(1-e^{-\tau}z^{-1})^2 + z^{-1}\tau[(1-z^{-1})^2 + (1-e^{-\tau}z^{-1})^2])}{(1-z^{-1})^2(1-e^{-\tau}z^{-1})^2} \\ &\equiv \frac{1}{M_0} \frac{q(z^{-1})}{b(z^{-1})}, \end{aligned} \quad (22)$$

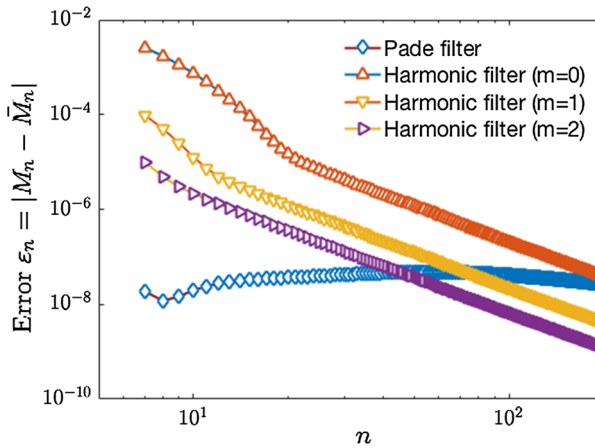
where the last line defines the polynomials  $b$  and  $q$  through the expressions on the middle line. The polynomials  $b$  and  $q$  are normalized,  $b(0) = q(0) = 1$ , and have degree  $2m + 2$  (in general) corresponding to the  $m + 1$  quadratic factors from the poles of order 2 in the rational approximation of  $T(z)$ . By the same arguments as in Sect. 4.2, we get the following estimator for  $M_n^{-1}$ :

$$\bar{M}_n^{-1} = M_0^{-1} q_n - \sum_{k=1}^n b_k M_{n-k}^{-1}, \quad (23)$$

with the convention that  $b_j = q_j = 0$  for  $j > 2m + 2$  so that the sum on the right hand side has at most  $2m + 2$  non-zero terms, i.e. 4 terms in the case in (22) with  $m = 1$ . We will refer to the estimator in (23) as a *harmonic filter*, because for the case  $m = 0$ , the filter reduces to

$$\bar{M}_n^{-1} = 2M_{n-1}^{-1} - M_{n-2}^{-1} \quad \text{for } n \geq 3,$$

i.e. the middle value  $M_{n-1}$  is the harmonic mean of the extreme values  $M_{n-2}$  and  $M_n$ . Figure 3 shows the comparison between the performance of the pade-truncated MZ estimator



**Fig. 3** The one step error for the harmonic filters with  $m = 0, 1$  and  $2$  corresponding to predictions using  $L = 2, 4$  and  $6$  respectively. We also show the inferred error of the Padé filter which was the ‘best’ among the various linear filters that we considered in 4.2

(20) with 6 taps, and the harmonic estimator (23) with  $m = 0, 1$  and  $2$ . The Padé truncated estimator performs better for small  $n$ , while the Harmonic predictors have comparable or superior performance for  $n \gtrsim 30$ .

#### 4.4 Empirical Filters: Data-Driven Stochastic Parameterization

A powerful approach to stochastic parameterization is through data-driven model reduction [11, 16, 29, 39]. In this approach, the coefficients in a parametric or semi-parametric model are determined by comparison with data. For this procedure, it is crucial that one begins with an appropriate form for the reduced model. Indeed a model with too many parameters can overfit the training data, i.e. the coefficients can become sensitive to the random noise in the data, which will then lead to poor predictions from the reduced model. We can view the finite lag linear and harmonic estimators in the previous two sections (see Eqs. (20) and (23)) as useful *ansatzes* for building data driven ‘empirical’ estimators. In this approach, the coefficients of the polynomials  $q$  and  $b$  are not inferred from the particular sequence corresponding to the initial condition  $|\rho_0\rangle = M_0|1\rangle$ , but rather, are estimated from many (random) realizations of the time series for  $M_n$  with the initial condition  $|\rho_0\rangle$  sampled from an appropriate distribution.

The appropriate state space model is  $M_n = \sum_{k=1}^n h_k M_{n-k} + \beta_n$  where  $\beta_n = \langle 1 | (\Lambda Q)^n | \rho_0 \rangle$  is now a *non-stationary* random process (see Sect. 5). Nonetheless, as a crude approximation, we can assume the AR( $L$ ) model

$$\bar{M}_n = M_0 q_n + \sum_{k=1}^{\min(n, L)} h'_k M_{n-k} + \sigma(\{M_{n-1}, M_{n-2}, \dots, M_0\}) \theta_n,$$

where  $q_k, h'_k$  are fixed (in  $n$ ) renormalized weights that are zero for  $n > L$ , and  $\theta_n$  are i.i.d. normal variates, giving a stochastic parameterization [2] of  $M_n$ . The non-stationarity of the noise process is modeled through the variance parameter  $\sigma$  that could, in principle, depend on the state as encoded by the entire history  $\{M_n, M_{n-1}, \dots, M_0\}$ . This is a common idea in regression analysis, called *variance inflation* [37]. It is an uncontrolled approximation because we are insisting that the covariance matrix for the fluctuations be diagonal.

We estimate  $\sigma$  by (i) using the homogeneity of the underlying process (4), and (ii) assuming that, for  $n > L$ ,  $\sigma_n$  only depends on the ‘recent’ past of  $M_n$  so it is only a function of the quantities  $M_n, M_{n-1}, \dots, M_{n-L}$  that appear in the sum  $\sum_{k=1}^L h'_k M_{n-k}$ . This still leaves open a range of possibilities, and to the extent any of these approximations are valid, the results should not depend on precisely how we choose to parameterize the variance (we give a *post facto* justification for the insensitivity to the particular approximations through the analysis in Sect. 5 below). We will thus make the “simple” choice

$$\sigma(\{M_{n-1}, \dots, M_0\}) \propto M_{n-1} + M_{n-2} + \dots + M_{n-L}.$$

for parameterizing the variance in terms of the state. With this choice, we can estimate the weights  $h'_k$  by the following Monte-Carlo procedure:

**Algorithm II: Filters determined by statistical regression**

1. Pick initial conditions  $|\rho_0^{(j)}\rangle$  for  $j = 1, 2, \dots, J$  by sampling from an appropriate distribution.
2. For each initial condition  $|\rho_0^{(j)}\rangle$ , generate the sequence  $M_k^{(j)}$  for  $k = 1, 2, \dots, N$ . This is the “training data set”.
3. Find the weights  $h'_k$  by minimizing the sum of the normalized squared residuals  $\sum_{j=1}^J \sum_{n=L+1}^N \left[ \frac{M_n^{(j)} - \sum_{k=1}^L h'_k M_{n-k}^{(j)}}{\sum_{k=1}^L M_{n-k}^{(j)}} \right]^2$ , where the outer sum is over different realizations, and the inner sum is over all subsequences of  $L$  consecutive values of  $M_k^{(j)}$ . The resulting equations are of course the analogs of the *Yule–Walker* equations [8, 61, 63] for the design of AR filters for stationary random processes. In our case, the covariances  $\mathbb{E}[M_n M_k]$  are estimated by averaging over time in each realization and also averaging over an ensemble of realizations.
4. Once we determine the parameters  $h'_k$ , we can then determine the parameters  $q_n$  by minimizing  $\sum_{j=1}^J \left[ M_n^{(j)} - \sum_{k=1}^n h'_k M_{n-k}^{(j)} - M_0 q_n \right]^2$ , where we assume that the solutions have been scaled such that  $M_0$  is the same for the  $L$  independent realizations. Note that we are only averaging over different realizations, and not over time, so we do not have the issue of estimating the variance of a non-stationary noise process. Minimizing over the choice of  $q_j$  yields

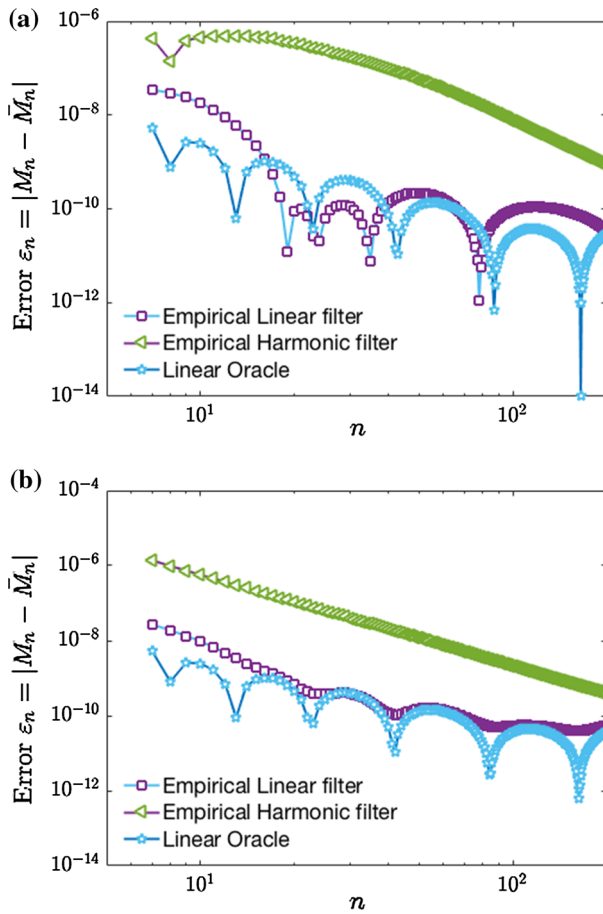
$$q_n = \frac{1}{J M_0} \sum_{j=1}^J \left[ M_n^{(j)} - \sum_{k=1}^n h'_k M_{n-k}^{(j)} \right]$$

5. An obvious modification of this method also applies to determine the regression coefficients for estimating  $M_n^{-1}$  from its history. Since  $(M + \Delta)^{-1} \approx M^{-1} - \Delta/M^2$  if  $\Delta/M \ll 1$ , we can postulate the variance parameterization for the filter

$$\bar{M}_n^{-1} = q'_n + \sum_{k=1}^{\min(n, L)} v'_k M_{n-k}^{-1} + \sigma'(\{M_{n-1}, M_{n-2}, \dots, M_0\}) \theta'_n$$

$$\text{as } \sigma' \propto (M_{n-1} + M_{n-2} + \dots + M_{n-L})/M_n^2 \sim M_{n-1}^{-1} + M_{n-2}^{-1} + \dots + M_{n-L}^{-1}. \quad \square$$

We display the results of this procedure in Fig. 4. We obtain the empirical AR( $L$ ) estimators  $(q_k, h'_k)$  and  $(q'_k, v'_k)$  with  $L = 6$  by averaging the residuals over 100 independent realizations, each starting with  $M_0 = 1$  and run for 200 steps. The resulting filters are given by

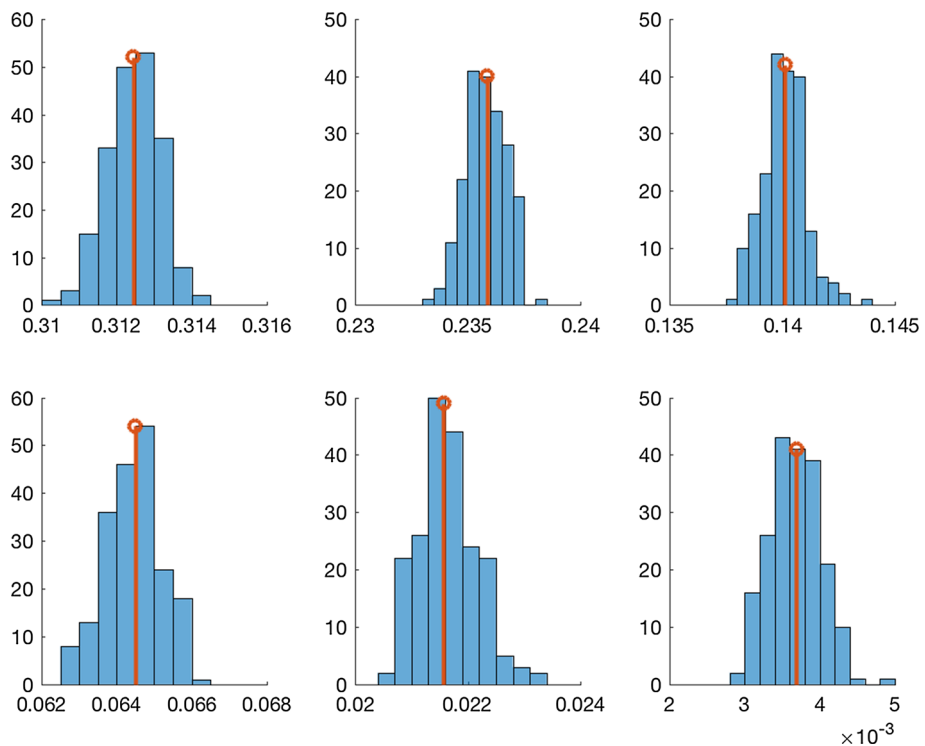


**Fig. 4** The performance of the empirical filters determined by regression using many model runs. **a** One-step prediction error for a single realization. The dips in the error correspond to times where the error changes sign, and the number of such changes is naturally related to the number of taps in the filter. **b** The average performance of the empirical filters over 100 random realizations. The empirical linear filter performs better than the empirical harmonic filter for all  $n$ . It also tracks the average performance of (the realization dependent) linear oracle, except at the dips, which persist even upon averaging

$$\begin{aligned} \bar{M}_n = & 5.2218M_{n-1} - 11.3232M_{n-2} + 13.0495M_{n-3} - 8.4286M_{n-4} \\ & + 2.8926M_{n-5} - 0.4120M_{n-6} \end{aligned} \quad (24)$$

$$\begin{aligned} \bar{M}_n^{-1} = & 4.9161M_{n-1}^{-1} - 10.1034M_{n-2}^{-1} + 11.1366M_{n-3}^{-1} - 6.9607M_{n-4}^{-1} \\ & + 2.3446M_{n-5}^{-1} - 0.3332M_{n-6}^{-1} \end{aligned} \quad (25)$$

We test the performance of these filters on 100 new realizations, that were not part of the training set, and also construct, for each realization, a *linear oracle*, i.e. a filter that has knowledge of the future, by minimizing the sum of the squared residuals over all subsequences of consecutive values of  $M_n$  for this realization. By construction, the linear oracle has the smallest possible error among all linear filters for the given realization, and is thus a good benchmark for measuring the performance of any given linear filter.



**Fig. 5** Distribution of the poles of the linear oracles for 200 independent realizations. The abscissa in each plot is  $1 - z_j$ ,  $j = 1, 2, \dots, 6$  where  $z_j$  is the  $j$ th smallest pole. The empirical filter is obtained by averaging the residuals over time and also the first 100 realizations. It has poles at 0.6876, 0.7641, 0.8599, 0.9355, 0.9784 and 0.9963 respectively and these locations are marked on the corresponding histograms. Note that the relative variation in the pole location, among the various realizations, gets larger, exponentially in  $j$ , showing that the pole locations are sloppy parameters

The process  $M_n$  is not stationary, and certainly not ergodic, so there is no reason to expect that we can replace an ensemble average over different realizations by a time average. Also, as discussed above, the procedure for computing the linear oracle cannot be done ‘online’ (i.e. as the data  $M_n$  is being generated) because we need the entire history of the sequence  $M_n$  to compute it. Nonetheless, as we illustrate in Fig. 5, the variations between the linear oracles for different realizations are small, and they all agree with the filter generated by averaging the residuals over time and realizations. This is also reflected in the fact that the averaged filter performs nearly as well as the linear oracle, for data *that were not part of the training set*.

## 5 Asymptotic Filters

In the previous section we considered various approaches to building a stochastic parametrization of the process  $M_n$ . The numerics revealed some counter-intuitive results. The numerics also demonstrate, unequivocally, that the ‘best’ stochastic parameterization came from the data-driven empirical linear filter. In this section, we develop a framework for the analyz-



ing the continuum model (4) in a probabilistic setting with random initial conditions drawn from a distribution. We also characterize the noise process  $\beta_n$  in the Mori–Zwanzig projection (12), and this allows us to rigorously analyze the stochastic parametrization/filtering schemes from Sect. 4. Through this analysis we provide an explanation for the observations from numerical simulations, and also present some new theoretical insights into stochastic parametrization/prediction for systems with slow relaxation/algebraic decay of correlations.

In our numerical discretization of (4), we take

$$\rho_0^I(w) = \sum_{i=1}^I \Delta \gamma_i \Delta \left( w - \left( \frac{i}{I} - \frac{1}{2I} \right) \right), \quad (26)$$

where  $\Delta = 1/I$  and the quantities  $\gamma_i$  are i.i.d non-negative random variables with mean  $\mu_\gamma = \mathbb{E}[\gamma]$  and variance  $\sigma_\gamma^2 = \mathbb{E}[(\gamma - \mathbb{E}[\gamma])^2]$ . If  $\phi, \psi$  are continuous functions on  $[0, 1]$ , direct calculations show that

$$\mathbb{E} \left[ \int_0^1 \rho_0(w) \phi(w) dw \right] = \mathbb{E}[\gamma] \sum_{i=1}^I \phi \left( \frac{i}{I} - \frac{1}{2I} \right) \Delta \approx \mu_\gamma \int_0^1 \phi(x) dx,$$

and

$$\begin{aligned} \mathbb{E} \left[ \iint \phi(w) \psi(w') \rho_0(w) \rho_0(w') dw dw' \right] &= \left[ \mathbb{E}[\gamma] \sum_{i=1}^I \phi \left( \frac{i}{I} - \frac{1}{2I} \right) \Delta \right]^2 \\ &\quad + \mathbb{E}[(\gamma - \mathbb{E}[\gamma])^2] \sum_{i=1}^I \phi \left( \frac{i}{I} - \frac{1}{2I} \right) \psi \left( \frac{i}{I} - \frac{1}{2I} \right) \Delta^2 \\ &\approx \mu_\gamma^2 \iint \phi(w) \psi(w') dw dw' + \frac{\sigma_\gamma^2}{I} \int \phi(w) \psi(w). \end{aligned}$$

The distribution for  $\gamma$  can potentially depend on  $I$ . We will assume that  $\mu_\gamma \rightarrow 1$  and  $\frac{\sigma_\gamma^2}{I} \rightarrow \bar{\sigma}^2$  as  $I \rightarrow \infty$ . If the distribution of the weights  $\gamma_I$  is independent of  $I$ , then either  $\bar{\sigma}^2 = 0$ , the *central limit theorem* scaling, or  $\bar{\sigma}^2 = \infty$ , and neither situation is appropriate for modeling natural oil where we expect that there is some finite variance associated with the uncertainty in the composition of oil. Thus, we need to sample from initial conditions with  $0 < \bar{\sigma}^2 < \infty$ .

For any prescribed value  $0 < \bar{\sigma}^2 < \infty$ , we can indeed find a family of  $I$ -dependent distributions that satisfy these conditions, by appropriately truncating and rescaling a distribution that has finite mean but infinite variance. The details are presented in Appendix 3.

Under the conditions  $\mu_\gamma \rightarrow 1$  and  $\frac{\sigma_\gamma^2}{I} \rightarrow \bar{\sigma}^2$  as  $I \rightarrow \infty$ , we have a distribution of initial conditions  $|\rho_0\rangle$  defined by the weak limits of sequences of the form (26). For any pair of observables  $\phi$  and  $\psi$ , we have

$$\begin{aligned} \mathbb{E}[\langle \phi | \rho_0 \rangle] &= \langle \phi | 1 \rangle \equiv \langle \phi | \mathbb{E}[|\rho_0\rangle] \\ \mathbb{E}[\langle \phi | \rho_0 \rangle \langle \psi | \rho_0 \rangle] &= \langle \phi | 1 \rangle \langle 1 | \psi \rangle + \sigma^2 \langle \phi | \psi \rangle \equiv \langle \phi | \mathbb{E}[|\rho_0\rangle] \langle \rho_0 | \rangle | \psi \rangle. \end{aligned} \quad (27)$$

These equations characterize the first two moments of the distribution on the initial conditions for the continuum limit (4). Alternatively, we could have postulated that the distribution of the initial conditions had these expressions for the first and second moments by requiring that the distribution be invariant under all the (mathematically simplifying but not physically well motivated) exchanges  $w \rightarrow w'$  that interchanges the amounts of two distinct species in

the oil. This naturally leads to  $\mathbb{E}|\rho_0\rangle \propto |1\rangle$  and  $\mathbb{E}|\rho_0\rangle\langle\rho_0|$  is a linear combination of  $|1\rangle\langle 1|$  and the identity operator on  $\mathcal{H}$  as these are the only invariant operators under the permutation  $w \rightarrow w'$ . The homogeneity of the model (4) allows us to eliminate one parameter from the distribution of initial conditions by rescaling the initial mass to  $M_0 = 1$ . The symmetry argument therefore allows for a two parameter family of distribution of initial conditions  $|\rho_0\rangle$  characterized by

$$\mathbb{E}[\langle\phi|\rho_0\rangle] = \langle\phi|1\rangle, \quad \mathbb{E}[\langle\phi|\rho_0\rangle\langle\psi|\rho_0\rangle] = \zeta^2\langle\phi|1\rangle\langle 1|\psi\rangle + \bar{\sigma}^2\langle\phi|\psi\rangle, \quad \zeta^2 \geq 1.$$

In this view, we have given an explicit construction for how to sample initial conditions from a one-parameter subfamily (corresponding to  $\zeta^2 = 1$ ) of such measures, as (subsequential) weak limits of measures consisting of finitely many point masses.

We now characterize the noise process  $\beta_n$  in (12). In the Schrödinger picture the mass at time  $n$  is given by the observable  $\langle g| = \langle 1|\Lambda$  acting on the state  $|\rho_{n-1}\rangle$  at time  $n - 1$ . Using this with (11), we get

$$h_k = \langle 1|\Lambda(Q\Lambda)^k|1\rangle, \quad \beta_n = \langle 1|(\Lambda Q)^n|\rho_0\rangle.$$

Consequently,

$$\mathbb{E}[\beta_n] = \mathbb{E}[\langle 1|(\Lambda Q)^n|\rho_0\rangle] = \langle 1|(\Lambda Q)^n|1\rangle = 0,$$

and

$$\begin{aligned} \mathbb{E}[\beta_n\beta_m] &= \langle 1|(\Lambda Q)^n\mathbb{E}[|\rho_o\rangle\langle\rho_o|](Q\Lambda)^m|1\rangle \\ &= \langle 1|(\Lambda Q)^n(P + \bar{\sigma}^2)(Q\Lambda)^m|1\rangle \\ &= \bar{\sigma}^2\langle 1|(\Lambda Q)^{n+m-1}\Lambda|1\rangle \\ &= \bar{\sigma}^2h_{n+m-1}. \end{aligned}$$

We have used  $PQ = 0$ ,  $Q^2 = Q$ . This is a fluctuation–dissipation relation for the system in (4), and not unexpectedly, it *does not have the form of the usual fluctuation–dissipation theorem* for a Hamiltonian system with short memory [19]. Note also, that

$$\begin{aligned} \mathbb{E}[\beta_nM_j] &= \langle 1|(\Lambda Q)^n\mathbb{E}[|\rho_o\rangle\langle\rho_o|]\Lambda^j|1\rangle \\ &= \langle 1|(\Lambda Q)^n(P + \bar{\sigma}^2)\Lambda^j|1\rangle \\ &= \bar{\sigma}^2\langle 1|(\Lambda Q)^n\Lambda^j|1\rangle. \end{aligned}$$

These expectations are non-zero in general; we compute them explicitly in Appendix 2.

This result is somewhat unexpected. It is certainly true that  $\langle F_n|\xi_n\rangle = \langle F_n|P\Lambda^n|\rho_0\rangle = 0$  (see discussion before (10)) for any initial measure  $\rho_0$ , but this does not imply that  $\mathbb{E}[\beta_nM_j] = \mathbb{E}[\langle F_n|\rho_0\rangle\langle\rho_0|\xi_n\rangle] = 0$ . The issue is that the observable  $g^t$  given by the constant function  $g^0(x) = 1$  has non-trivial evolution (see discussion after (5)) in contrast to the situation for dynamical systems, so the usual intuition does not apply. The “noise” is no longer uncorrelated with the observations, and this explains why the projection formalism *does not give an optimal prediction/stochastic parametrization procedure* for (4). The empirical filters, albeit still linear, and with far fewer taps, do perform better than the MZ estimator, because, by construction, the ensemble average of the product of the noise and past observations is zero.

## 5.1 Optimal Time-Varying Filters

We now address the question of why the empirical linear filter performed almost as well as the linear oracles. For the process  $M_n = \langle 1 | \Lambda^n | \rho_0 \rangle$ , we have

$$\mathbb{E}[M_n] = \mathbb{E}[\langle 1 | \Lambda^n | \rho_0 \rangle] = \langle 1 | \Lambda^n | 1 \rangle = \frac{1 - e^{-n\tau}}{n\tau},$$

and

$$\mathbb{E}[M_n M_j] = \langle 1 | \Lambda^n | \mathbb{E}[\langle \rho_0 | \langle \rho_0 | \Lambda^j | 1 \rangle] = \mathbb{E}[M_n] \mathbb{E}[M_j] + \bar{\sigma}^2 \langle 1 | \Lambda^{n+j} | 1 \rangle.$$

By regression, there is indeed an optimal  $\text{AR}(L)$  filter of the form

$$M_n = q_n M_0 + h_1^{(n)} M_{n-1} + h_2^{(n)} M_{n-2} + \cdots + h_L^{(n)} M_{n-L} + \theta_n,$$

where the innovation  $\theta_n$  is orthogonal to  $M_{n-1}, M_{n-2}, \dots, M_{n-L}$  and  $M_0 = 1$ . Note that the optimal filter is allowed to (and as we see below, in general does) depend on  $n$ , so it is not autonomous. The orthogonality condition gives the Yule–Walker equations

$$\mathbb{E}[M_n M_{n-k}] = q_n \mathbb{E}[M_{n-k}] + \sum_{j=1}^L h_j^{(n)} \mathbb{E}[M_{n-j} M_{n-k}], \quad (28)$$

$$\mathbb{E}[M_n] = q_n + \sum_{j=1}^L h_j^{(n)} \mathbb{E}[M_{n-j}]. \quad (29)$$

Multiplying the second row by  $\mathbb{E}[M_{n-k}]$  and subtracting from the first row gives

$$\frac{1 - e^{(2n-k)\tau}}{(2n-k)\tau} = \sum_{j=1}^L h_j^{(n)} \frac{1 - e^{(2n-k-j)\tau}}{(2n-k-j)\tau}, \quad k = 1, 2, \dots, L.$$

Note that these equations are independent of  $\bar{\sigma}^2$ , so they do not require small noise. For  $n - L \gg \tau^{-1} \sim O(1)$ , we can ignore the exponentially small quantities and get the matrix system  $v = Ah$  where the vector  $v$  and matrix  $A$  are as defined below,

$$\begin{pmatrix} \frac{1}{2n-1} \\ \frac{1}{2n-2} \\ \vdots \\ \frac{1}{2n-L} \end{pmatrix} = \begin{pmatrix} \frac{1}{2n-2} & \frac{1}{2n-3} & \cdots & \frac{1}{2n-L-1} \\ \frac{1}{2n-3} & \frac{1}{2n-4} & \cdots & \frac{1}{2n-L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2n-L-1} & \frac{1}{2n-L-2} & \cdots & \frac{1}{2n-2L} \end{pmatrix} \begin{pmatrix} h_1^{(n)} \\ h_1^{(n)} \\ \vdots \\ h_L^{(n)} \end{pmatrix}. \quad (30)$$

The coefficient matrix  $A$  for this system is a variant of the classical *Hilbert matrix*, a well known example of an ill-conditioned matrix. The condition number of this matrix is  $\sim (2n)^L / L!$  which can be enormous, and we cannot solve the system in a numerically stable manner, although, of course, a unique solution does exist. We compute solutions to this system in Appendix 4 to obtain

$$\begin{aligned} h_j^{(n)} &= \prod_{i \neq j}^L \frac{i}{i-j} \prod_{i=1}^L \frac{2n-i-j}{2n-i} \\ &= (-1)^{j-1} \binom{L}{j} + (-1)^j \frac{L^2}{2n} \binom{L-1}{j-1} + O(n^{-2}). \end{aligned} \quad (31)$$

We can fix  $L$ , the order of the filter, and look at the behavior of the filter coefficients as expansions in  $n$ . Picking  $L = 6$  (corresponding to predicting  $M_n$  using  $M_{n-1}, M_{n-2}, \dots, M_{n-6}$ ) for illustration, we get

$$\begin{aligned} h_1^{(n)} &= 6 - \frac{36}{2n-1}, \\ h_2^{(n)} &= -15 + \frac{630}{2n-1} - \frac{225}{n-1}, \\ h_3^{(n)} &= 20 - \frac{3360}{2n-1} + \frac{2100}{n-1} - \frac{1200}{2n-3}, \\ h_4^{(n)} &= -15 + \frac{7560}{2n-1} - \frac{6300}{n-1} + \frac{6300}{2n-3} - \frac{450}{n-2}, \\ h_5^{(n)} &= 6 - \frac{7560}{2n-1} + \frac{7560}{n-1} - \frac{10080}{2n-3} + \frac{1260}{n-2} - \frac{180}{2n-5}, \\ h_6^{(n)} &= -1 + \frac{2772}{2n-1} - \frac{3150}{n-1} + \frac{5040}{2n-3} - \frac{840}{n-2} + \frac{210}{2n-5} - \frac{3}{n-3}. \end{aligned} \quad (32)$$

The (non-autonomous) filter

$$\bar{M}_n = \sum_{j=1}^6 h_j^{(n)} M_{n-j} \quad (33)$$

with the coefficients given by (32) is the *Asymptotic filter* with 6 taps. Clearly, the filter coefficients converge as  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} h_j^{(n)} = (-1)^{j-1} \binom{L}{j}$$

This is a *post facto* justification for why we could average over  $n$ , in addition to averaging over independent ensembles, in determining the coefficients of the empirical filter (Sect. 4.4).

## 5.2 Universal Filters for Slowly Decaying Correlations

There is a satisfying intuitive explanation for the form of the asymptotic limit filter. Consider the problem of finding coefficients  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_L$  such that the asymptotic growth for the linear combination  $\sum_{i=0}^L \alpha_i (n-i)^{-1}$  is as small as possible, where we normalize the coefficients by requiring that  $\alpha_0 = 1$ . It is clear that for generic choices of  $\alpha_i$ , the decay rate of the combination is  $O(n^{-1})$ , but we can do better by judicious choices of  $\alpha$ . For example,  $\alpha_0 = 1, \alpha_1 = -1$  and the rest of the  $\alpha_i = 0$  gives a decay rate  $O(n^{-2})$ . The smallest possible asymptotic behavior comes from the coefficients  $\alpha_i$  set equal to a row of Pascal's triangle with alternating signs, as we can see from an inductive argument. For this choice of  $\alpha_i$ , we have

$$\sum_{i=0}^L \binom{L}{i} \frac{(-1)^i}{n-i} = \frac{L!}{n(n-1)(n-2) \cdots (n-L)} \sim \frac{L!}{n^{L+1}},$$

and for any other choice of the coefficients, the decay of the linear combination is slower. So it is indeed to be expected that if the optimal filter coefficients converge  $\lim_{n \rightarrow \infty} h_j^{(n)} = \alpha_j$ , then  $\alpha_j = (-1)^j \binom{L}{j}$ . The coefficient  $q_n$  is determined by (28) as

$$q_n = \mathbb{E}[M_n] - \sum_{j=1}^L h_j^{(n)} \mathbb{E}[M_{n-j}] \sim \frac{L!}{n^{L+1}} + o(n^{-L-1}).$$

Thus  $q_n$  decays very rapidly so that it can be set to zero for  $n > L \sim O(1)$ .

These filter coefficients are ‘universal’ for all processes with slowly decaying correlations. Indeed, for a slowly decaying function  $f(x)$  (say one consisting of nonpositive powers of  $\log(x)$  and of  $x$ ), we have

$$\sum_{i=0}^L \binom{L}{i} (-1)^i f(n-i) \sim \frac{d^L}{dx^L} f\left(n - \frac{L}{2}\right),$$

and one cannot get better asymptotic decay with constant coefficient linear combinations of  $L+1$  consecutive terms. It is thus tempting to suggest that all processes with slowly decaying correlations can be stochastically parameterized by

$$[(1-R)^L f]_n = \sum_{i=0}^L \binom{L}{i} (-1)^i f_{n-i} = \sigma_n \theta_n, \quad (34)$$

where  $R$  is the right shift operator on sequences,  $[Rf]_n = f_{n-1}$ ,  $\theta_n$  are independent normal variates and the variance parameter  $\sigma_n$  has statistics that can be estimated from data. Eq. (34) is the model reduction that is associated with the filter

$$\tilde{M}_n = \sum_{j=1}^L (-1)^{j-1} \binom{L}{j} M_{n-j}, \quad (35)$$

which is an *universal* filter for processes with slowly decaying correlations, since it is expected to work just as well for any such process.

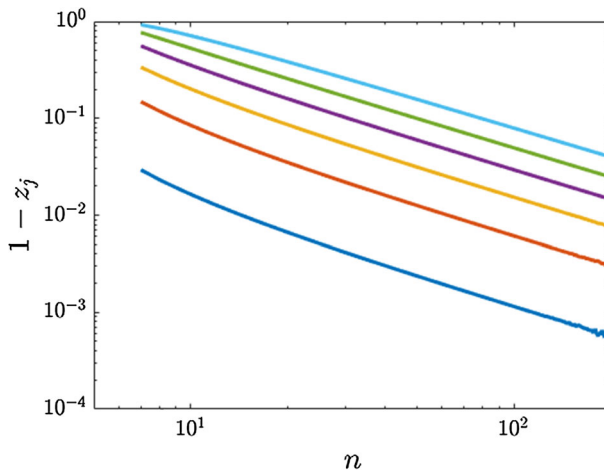
The universal stochastic parametrization (34) does not depend on the correlation structure of the process that is being modeled, besides requiring that it decay algebraically. Thus one does not expect this model to track a realization of the underlying process without additional data assimilation. Nor does one expect that an ensemble of solutions of (34) with appropriate statistics for  $\theta_n$  necessarily reproduce the statistics of an ensemble of realizations of the underlying process. Indeed, the transfer function of the universal filter is  $(1-z^{-1})^{-L}$  and has a pole of order  $L$  at  $z=1$ . The filter is thus unstable, and has homogenous solutions  $f_n = n^j$  for  $j=0, 1, 2, \dots, L-1$  which do not decay to 0.

We can attempt to remedy these shortcomings by going to higher order in the solutions of the Yule–Walker equations (33). The  $n$ -dependent corrections to the limiting filter coefficients do reflect the particular correlation function  $\mathbb{E}[M_n M_j] \sim 1/(m+j)$  for the evaporation process and are thus *not universal*. The correction depends explicitly on  $n$  so including these corrections will make the filter non-autonomous. However, these corrections make the filter stable. Figure 6 shows the poles of the filter transfer function

$$H^{(n)}(z) = \frac{1}{1 - \sum_{j=1}^L h_j^{(n)} z^{-j-1}},$$

corresponding to a shift invariant filter obtained by “freezing” the time index  $n$ . Note that all the poles are real, less than 1, and approach 1 as  $n \rightarrow \infty$  from inside the unit circle.

If we order the poles by  $z_j \leq z_{j+1}$ , we have  $z_j \approx 1 - c_j n^{-1}$ , and the constants  $c_j$  are (roughly) geometrically distributed, i.e  $\log(c_{j-1}/c_j) \sim O(1)$  (strong level repulsion) for



**Fig. 6** The poles  $z_j$ ,  $j = 1, 2, \dots, L$  of the asymptotic filter (33) as a function of  $n$  for  $L = 6$ . To show the convergence to 1 we plot  $1 - z_j \approx -\log(z_j)$ , and these quantities are all positive indicating that the filters are stable. Also, for each of the poles, we see that  $z_j \approx 1 - c_j n^{-1}$  from the slope of the corresponding graph. These curves also demonstrate that our model is sloppy, as evident from the level repulsion between the poles  $z_j$  at every fixed value of  $n$

all  $j$ , indicating that the poles are sloppy parameters [62] for the description of the linear evaporation process (4). The filter coefficients  $h_k^{(n)}$  are determined by

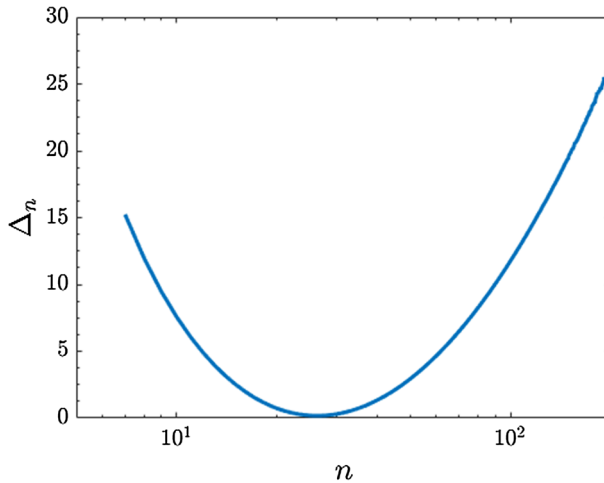
$$1 - \sum_{k=1}^L h_k^{(n)} z^{-j-1} = \prod_{j=1}^L \left(1 - \frac{z_j}{z}\right) \approx \prod_{j=1}^L \left(1 - \frac{1 - c_j n^{-1}}{z}\right)$$

so that the filter coefficients are  $n$  dependent, *symmetric* functions of the quantities  $c_j$ . Although the poles are sloppy, the filter coefficients themselves are robust [62]. We first learned this principle, *viz.* symmetric functions of random quantities are computable in terms of the low order moments of their distribution, and are hence robust, in work with Leo Kadanoff on the extremal distribution of points for the Thomson problem in 2D domains [4].

We can now give an analytical explanation for the reason that the empirical linear filter performed almost as well as the genie-aided linear oracle, and thus is demonstrably a near-optimal linear filter, among all shift-invariant linear filters with  $L$  (a given number of) taps. For an interval of time  $1 \leq k \leq N$ , we can pick an autonomous filter that is (approximately) optimal for the entire range by using the corrections in (33) with  $n = \bar{k}$  being an appropriate “averaged” time index over the interval of interest. This will give a filter with fixed coefficients and  $L$  taps, that is guaranteed to be stable. Since the corrections in (33) are  $O(n^{-1})$  it is not unreasonable to expect that

$$\frac{1}{\bar{k}} \sim \frac{1}{N} \sum_{j=L}^N \frac{1}{j} \sim \frac{\log(N)}{N},$$

so that  $\bar{k}$  is the harmonic mean of the time interval. For  $N = 200$ , we would estimate  $\bar{k} \approx 37.75$ . Since the quantity of interest is the deviation of the  $j$ th pole from 1. We define the discrepancy



**Fig. 7** The discrepancy  $\Delta_n$  between the poles of the empirical filter and the poles of the asymptotic filter (33). The empirical filter was constructed as described in Sect. 4.4 by running the process for  $N = 200$  time steps and averaging the residuals over 100 realizations. The poles of the empirical filter are at 0.9963, 0.9784, 0.9355, 0.8599, 0.7641 and 0.6876 respectively

$$\Delta_n = \sum_{j=1}^L \left| \log \left( \frac{1 - z_j^{(n)}}{1 - \tilde{z}_j} \right) \right|^2,$$

where  $z_j^{(n)}$  is the  $j$ th pole of the asymptotic filter (33) (shown in Fig. 6) and  $\tilde{z}_j$  is the  $j$ th pole of the empirical filter (see Fig. 5). Figure 7 shows the discrepancy  $\Delta_n$  as a function of  $n$ . Minimizing the discrepancy, we would infer that  $\bar{\kappa} \approx 26.43$  which is on the same scale, although a little smaller than our estimate of 37.75.

## 6 Multilayer Stochastic Models for the Evaporation of Oil

The asymptotic filter (33) is not autonomous, and building a reduced model using this filter will similarly give a non-autonomous stochastic parametrization. We can use a standard ‘trick’ to recast time-dependent systems as autonomous systems on a larger phase space [47]. In particular, by enlarging our “phase-space” to include an additional dynamical variable  $\kappa$  that tracks  $n$ , we have the autonomous stochastic parametrization

$$\begin{aligned} M_n - \sum_{j=1}^L h_j^{(\kappa_n)} M_{n-j} &= \sigma_n \theta_n, \\ \kappa_n &= \kappa_{n-1} + 1 + s_n \theta'_n, \\ \sigma_n^2 &\simeq \bar{\sigma}^2 \frac{L!}{\kappa_n^{L+1}}, \end{aligned} \quad (36)$$

where  $\theta_n, \theta'_n$  are i.i.d process of normal variates, the dynamical variable  $\kappa$  tracks the “microstructure” of the oil composition in terms of its “age,” and the filter coefficients  $h_j^{(\kappa)}$  are given by (31) with  $n = \kappa$ .



The reduced model in (36) has the form on a multilayer stochastic model (MSM) [11, 35, 41], where the quantity  $M_n$  is directly observable and the quantity  $\kappa_n$  is hidden. To use this model for stochastic parametrization, we can specify  $\bar{\sigma}$  and set  $s_n = 0$  so that  $\kappa_n = n$ . Alternatively, given noisy measurements  $\tilde{M}_n$ , we can estimate the state  $(M_n, \kappa_n)$  and a parameter  $\bar{\sigma}^2$  jointly by a nonlinear filtering algorithm, for example by the extended Kalman filter [31]. We will present these results in a later publication.

A crude version of this nonlinear filtering approach is through approximating  $\kappa_n$  in terms of  $M_{n-1}, M_{n-2}, \dots$ . In conjunction with (36) and (31), this will give a stochastic parametrization for  $M_n$  without additional ‘hidden’ variables. Since  $\mathbb{E}[M_k] \approx \frac{M_0}{k\tau}$ , we can define an “instantaneous age”  $\mu_n$  by

$$\frac{1}{\mu_n - 1} \equiv 1 - \frac{M_{n-1}}{M_{n-2}} \approx 1 - \frac{\mathbb{E}[M_{n-1}]}{\mathbb{E}[M_{n-2}]} = \frac{1}{n-1},$$

so that  $\mu_n$  gives an estimate of  $n$  based on the “recent past”  $M_{n-1}$  and  $M_{n-2}$ . The instantaneous age is a fluctuating quantity, and we can estimate the age  $\kappa_n$  by smoothing  $\mu_n$  through  $\kappa_n = (1 - \delta)(\kappa_{n-1} + 1) + \delta\mu_n$ , which is the appropriate filtering strategy for the quantity  $\kappa_n$  evolving as in (36) (the ‘model’), with fluctuating estimates given by  $\mu_n$  (the ‘measurements’).

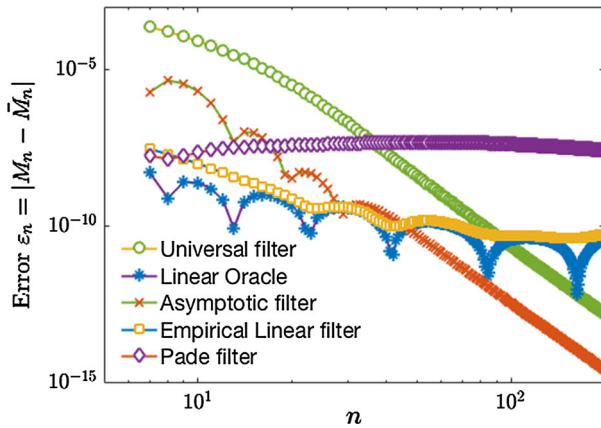
Using  $\kappa_n$  in place of  $n$  in (33) gives a nonlinear filtering algorithm for  $M_n$ , that we will call an *extended asymptotic filter*. It is a nonlinear modification of (33) that makes it autonomous. Explicitly, the extended asymptotic filter with  $L$ -taps is given by  $\kappa_0 = 0$  and

$$\begin{aligned} \mu_n &= \frac{M_{n-1}}{M_{n-2} - M_{n-1}} + 1, \\ \kappa_n &= \kappa_{n-1} + 1 + \delta(\mu_n - \kappa_{n-1}) \\ h_j^{(\kappa_n)} &= \prod_{i \neq j}^L \frac{i}{i-j} \prod_{i=1}^L \frac{2\kappa_n - i - j}{2\kappa_n - i}, \quad j = 1, 2, 3, \dots, L \\ \tilde{M}_n &= \sum_{j=1}^L h_j^{(\kappa_n)} M_{n-j} \end{aligned} \quad (37)$$

This filter is non-empirical, nonlinear, autonomous, independent of realization and is not genie-aided (does not require knowledge of the future). It contains a parameter  $\delta$  that we can set, with  $\delta = 0$  corresponding to a very stable, but non-responsive filter, while  $\delta = 1$  is a very responsive, but potentially numerically unstable filter. We compare the performance of the extended asymptotic filter (37) with  $\delta = 1$ , the universal filter (35) with the empirical filters from 4.4 and the Padé filter from 4.2, when applied to synthetic data. The results are shown in Fig. 8.

Outside of an initial transient, the extended asymptotic filter (37) is clearly better than the competing methods. The one-step prediction error has the optimal scaling  $\varepsilon_n \sim n^{-L-1}$  and by the arguments from before, we would expect this to be the smallest possible scale for the error from a filter with  $L$  taps.

To illustrate the practical application of the extended asymptotic filter, we apply it to empirical evaporation curves for various types of crude oils. Fingas [23] has measured the evaporation curves for about 200 different oils (crudes, fuel oils, diesels, etc.) under a variety of conditions and found that the important parameters are the time of evaporation and the ambient temperature. The time and temperature dependence of the evaporation curves are best fit by one of the following two equations (Eqs. (10) and (11) in [23]):



**Fig. 8** Comparison between the one-step prediction errors for the reduced models given by the empirical and the asymptotic filters. The empirical filter is generated by averaging the squared residuals over 100 realizations, and the performance of each filter is assessed by averaging errors over another set of 100 realizations. For large  $n$ , the one-step prediction error of the universal filter (34) is small; nonetheless it is unstable, and cannot be used without incorporating additional information from measurements. The extended asymptotic filter is run with  $\delta = 1$  and the corresponding curve is indicated as ‘Asymptotic filter’ in the legend

$$\begin{aligned} \%E &= (0.165(\%D) + 0.045(T - 15)) \log(t) \quad \text{and} \\ \%E &= (0.0254(\%D) + 0.01(T - 15))\sqrt{t} \end{aligned}$$

for oils that follow a “logarithmic” (respectively “square-root”) equation where  $\%E$  is the percentage of oil evaporated at time  $t$  in minutes,  $\%D$  is the percentage (by weight) of the crude oil that is distilled at  $180^\circ\text{C}$  and  $T$  is the ambient temperature in degrees Celsius. We can convert the empirical evaporation curve into the total remaining mass  $M(t)$  by

$$M(t) = 1 - \frac{\%E}{100}.$$

The fitting functions (the logarithmic and square-root equations) are clearly not valid if  $t$  is too small, since  $\%E$  and/or its time derivative blows up as  $t \rightarrow 0$ . They are also not valid for very large  $t$  as  $\%E$  cannot be more than a 100%. Finally, they are in arbitrary “empirical” units. We will non-dimensionalize, as in Sect. 2 by the (unknown!) evaporation rate of the most volatile component in the oil, and also modify (regularize) the small time behavior of the functions to ensure that  $M(0) = 1$ ,  $\frac{d}{dt}M(0) < \infty$ . The regularized, nondimensional functions  $M(t)$  are thus in one of two forms:

$$M(t) = 1 - a \log(1 + t/t_0) \quad \text{and} \quad (38)$$

$$M(t) = 1 - a(\sqrt{1 + t/t_0} - 1) \quad (39)$$

where  $t$ ,  $t_0$  (small scale cutoff) and  $a \ll 1$  are all dimensionless. These equations necessarily have a limited range of validity since we need  $M(t) \geq 0$  for all  $t$ . Solving Eqs. (38) and (39) for  $M(T_{\max}) = 0$ , we estimate the ranges of validity by  $T_{\max} \sim t_0 e^{1/a}$  for the logarithmic equation and  $T_{\max} \sim t_0/a^2$  for the square-root equation.

The parameters  $a$  and  $t_0$  are related by the following argument. Since we nondimensionalize time by the evaporation rate  $\alpha_{\max}$  of the most volatile species, the evaporation rate should satisfy

$$-\dot{M}(t) \lesssim \alpha_{\max} M(t) = M(t).$$

Using this inequality for (38) and (39) at  $t = 0$  gives  $a/t_0 \lesssim 1$ . We can also compute  $\frac{d^2}{dt^2} \log(M(t))$  for the linear evaporation process (4) to obtain

$$\frac{d^2}{dt^2} \log(M(t)) = \frac{\int w^2 \rho(w, t) dw \int \rho(w, t) dw - \left( \int w \rho(w, t) dw \right)^2}{\left( \int \rho(w, t) dw \right)^2} \geq 0$$

by the Cauchy–Schwarz inequality. Note that, this relation has to hold for *every realization*, and not just in an averaged sense. This relation *does not hold* for the logarithmic equation (38) for  $t \geq T_{\text{crit}} \approx \frac{1}{e} T_{\text{max}}$ , although by the time it breaks down, we have  $M(T_{\text{crit}}) \approx a \ll 1$ .

For the square-root equation (39), the relation fails to hold for  $t \geq T_{\text{crit}} \approx \frac{1}{4} T_{\text{max}}$ . The empirical square-root fit should therefore break down *well before* the total mass hits zero. Indeed, this should already occur by the time  $\%E \approx 50\%$ . This is an experimentally testable prediction, and checking this will help assess the validity of modeling assumptions that lead to (4). On the other hand, assuming (4) is a good microscopic model, the empirical fits in (38) and (39) can not good models for the “truth” unless  $t \lesssim T_{\text{crit}}$ . Consequently, the ability of a filter to track/predict these functions accurately is not necessarily a positive feature. Rather, we would hope that the filters “discover” that beyond a certain point, the assumed “truth” actually is not.

A final point relates to the role of sampling. In our analysis,  $M_n$  is given by  $M_n = M(n\tau)$  corresponding to a sampling interval  $T_{\text{sampling}} = \tau/\alpha_{\max}$ . In practice, the sampling time is determined by experimental/technical considerations and cannot be freely specified. While  $\tau \sim O(1)$  is the ideal situation, so that the time-series data resolves the dynamics on the fastest scales in the problem, the practically achievable value of  $\tau$  can be “large” and will therefore introduce an additional nondimensional parameter in the discrete time problem. The extended asymptotic filter (37) and the universal filter (35), however, are independent of  $\tau$ . This is a very desirable feature since it allows us to use the same filter independent of the sampling interval.

To assess the performance of the filters on “real” data, we use the following numerical procedure:

### Algorithm III: Data assimilation and filtering/prediction

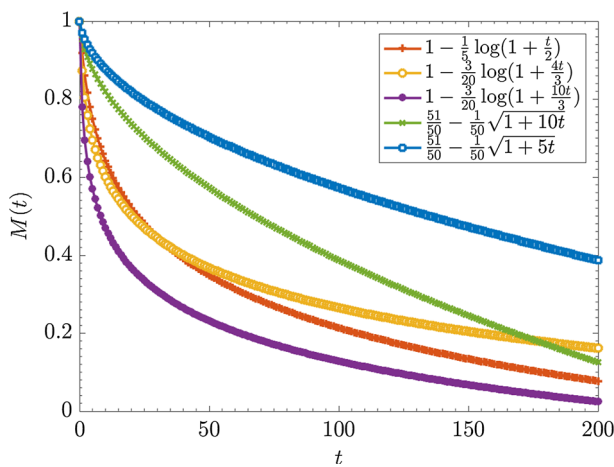
1. Pick parameters  $a$  and  $t_0$  such that  $a/t_0 \lesssim 1$  and  $T_{\text{max}} \gtrsim 200$ . The parameter values we use are listed in Table 2.
2. Generate time series  $M_n$  using the various parameter values for  $a$  and  $t_0$  in (38) and (39) with  $0 \leq t = n\tau \leq 200$ , where the sampling interval is  $\tau = 1, 4$  or  $10$ . This procedure corresponds to sampling the curves in Fig. 9 at equally spaced intervals. Depending on the sampling rate we get between 20 and 200 samples for each curve.
3. For various time series (different functions, parameters and sampling rates), we compute the one step prediction error for the empirical linear filter (24), the universal filter (35) and the extended asymptotic filter (37) with  $\delta = 1$ . These results are shown in Fig. 10.

□

Comparing the vertical scales in Fig. 10, we see that, for sufficiently large number of samples, the universal and extended asymptotic filters have a smaller one step prediction error than the empirical filter, but for  $\lesssim 30$  samples, the empirical filter has a smaller prediction error. This reflects an initial transient, and is similar to the behavior in Fig. 8 where we compared the performance of the filters on synthetic data generated by numerically simulating the process in (4).

**Table 2** Parameter values, model equations and limits on the range of validity for the numerically generated evaporation curves

$T_{max}$	$a/t_0$	$a$	$t_0$	$T_{crit}$	Model equation
Logarithmic evaporation curve					
250	0.1	0.2	2	$\sim 107$	$M(t) = 1 - \frac{1}{5} \log(1 + \frac{t}{2})$
500	0.2	0.15	0.75	$\sim 216$	$M(t) = 1 - \frac{3}{20} \log(1 + \frac{4t}{3})$
200	0.5	0.15	0.3	$\sim 87$	$M(t) = 1 - \frac{3}{20} \log(1 + \frac{10t}{3})$
Square-root evaporation curve					
250	0.2	0.02	0.1	$\sim 65$	$M(t) = \frac{51}{50} - \frac{1}{50} \sqrt{1 + 10t}$
500	0.1	0.02	0.2	$\sim 130$	$M(t) = \frac{51}{50} - \frac{1}{50} \sqrt{1 + 5t}$

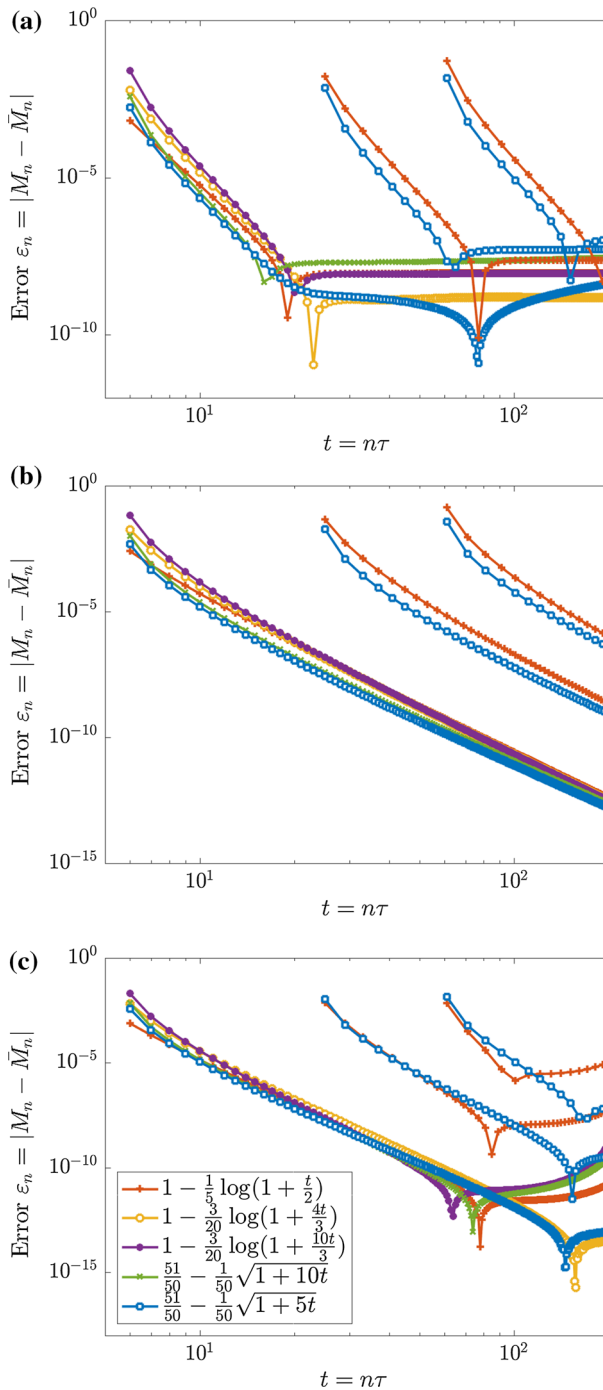


**Fig. 9** The evaporation curves generated by the logarithmic and square-root equations (38) and (39). The parameters/equations for the curves are given in Table 2

The error for the universal filter (Fig. 10b) is monotonically decreasing in  $n$ , the number of samples, roughly as a power law. In contrast, the error has interesting temporal structure for the empirical filter (Fig. 10a) and the extended asymptotic filter (Fig. 10c). Since we are plotting the absolute value of the error on a logarithmic scale, the downward spikes in the error are signatures of the times where the error changes sign.

For the empirical filter, the error decreases until it hits a “floor”, roughly between  $10^{-9}$  and  $10^{-8}$ . This floor is *independent of the time series and the sampling rate*, so it is inherent to the filter. In contrast, the error for the extended asymptotic filter has spikes at fixed times  $t = n\tau$ , independent of the sampling interval  $\tau$ , so these spikes reflect features in the signal, and not the structure of the filter. Indeed, these spikes *correlate well* with the critical times  $T_{crit}$  beyond which the given time series *cannot be realized in any solution* of (4). Beyond the spikes, the one-step prediction error for the extended asymptotic filter increases, in contrast to the other two filters. These observations support the following conclusions:

1. The universal filter has very small error as  $n \rightarrow \infty$ , but is not very discriminating. It tracks all functions with slowly decaying correlations, whether or not they come from solutions of (4).



**Fig. 10** One step prediction errors. The filters are applied to time series obtained from the functions in Table 2, sampled at intervals  $\tau$  of 1, 4 and 10 time units. **a** Empirical linear filter. **b** Universal filter. **c** Extended asymptotic filter with  $\delta = 1$ . Note the difference in the vertical scales

2. In tracking solutions of (4), the empirical linear filter is very discriminating/nearly optimal among all linear filters with fixed coefficients and  $L$  (a given number of) taps. However, it has a floor for its error reflecting the fact that we have ignored the  $L + 1$ th eigenvalue of the underlying sloppy model in the construction of this filter. Because of the level repulsion between the eigenvalues of matrices from sloppy models [62] the resulting error floor is pretty small when  $L$  is moderate.
3. The extended asymptotic filter is (essentially) time varying so it has additional freedom which can be exploited to make filters that are both discriminating, and circumvent the above argument for the error floor.

The numerical evidence for this picture is strong motivation to try and formalize this intuition into rigorous mathematical statement and proofs.

## 7 Discussion

In this work, we have developed nonlinear, stochastic, reduced models for the evaporation process (4), both empirically, i.e. in a data driven manner, and reductively from first principles, using the Mori–Zwanzig projection operator formalism, and by solving the appropriate Yule–Walker equations. The underlying system (4) has slow relaxation, and long memory, so it is of interest to see what intuition can be gleaned from our results, that might be generally applicable to other non-equilibrium systems with slow relaxation.

Hamiltonian dynamics naturally supports an invariant measure on phase space, and the corresponding Liouvillian is *skew-symmetric* on the  $L^2$  space for this invariant measure. In contrast, our system (4) has no physically relevant invariant measure, is not the Liouvillian for any dynamical system, and is *symmetric* rather than skew-symmetric on an appropriate  $L^2$  space. However, the evaporation process (4) is representative of systems with long memory and slow decay of correlations. So, we believe that our results do give some intuition for this class of non-equilibrium systems.

It is very surprising that a direct application of the Mori–Zwanzig projection operator formalism yields poor results for this system, considering the fact that the projection operator formalism is *exact* and is thus a natural starting point to make approximations. On the other hand, the MZ projection operator formalism, because it is exact, is constrained in ways that an empirical approximation is not. In this sense, a MZ decomposition, as in (12), has lots of parameters (the memory kernel  $h_k$ ) and thus has the potential to overfit the training data, i.e. the sequence  $\mathbb{E}[M_n]$  given by (13). As a consequence of this overfitting, the resulting model has poor predictive power. This argument suggests that the projection operator formalism is perhaps best suited for systems with an exponential decay of correlations, but perhaps not so well suited for systems with long memory [34].

Another lesson, that is reinforced by our results for the model process (4), is the importance of picking the right ansatz and the right parameters for data-driven model reduction. In particular, both the MSM approach [11, 35] and the NARMAX approach [15, 39] to model identification use power series, and they assume that the higher order terms are smaller than the lower order ones. In our example, the model is homogeneous, so these approaches will suggest that  $M_n$  should be approximated as a linear combination of  $M_{n-1}, \dots, M_{n-L}$ . However, our results in Sects. 4.2 and 4.3 indicate that, for situations with slow relaxation, one could do better by taking more general nonlinear homogeneous combinations, e.g. harmonic averages.

A key contribution of this work is development of the notions of *universal* and *asymptotic* filters. Indeed, the existence of these filters corroborates with theoretical ideas on *universality* for systems with slow decay of correlations [7] and experimental observation of universality in the glass transition [21]. These filters are also applicable in practice. Absent any other information, the universal filter (34) is the optimal first order predictor [14] for systems with slow relaxation. However, this filter is unstable, so for particular problems, one has to go beyond this universal filter and develop asymptotic filters (e.g. (33) and (37)), i.e. stable, necessarily non-autonomous filters given by solutions of the appropriate Yule–Waler equations. These filters are asymptotic in that they converge to the universal filter as  $n \rightarrow \infty$ . We can view the asymptotic filters as analogs of the  $t$ -damping equation in [14], albeit for our case where the memory is not short.

The extended asymptotic filter given by (37) is a practical computational tool. It allows us to identify a single quantity  $\kappa_n$  which accounts for the dependence of the evolution of macroscopic quantity  $M_n$  on the microscopic state  $|\rho_n\rangle$ . For this reason, we will interpret  $\kappa_n$  as the *microscopic age* of an oil distribution  $\rho_n(w)$ . This interpretation of  $\kappa$  naturally follows from

$$M_n = \int e^{-nw\tau} \rho_0(w) dw,$$

$$\frac{M_{n-1}}{\kappa_{n-1}} \approx M_{n-1} - M_n = \int (1 - e^{-w\tau}) e^{-(n-1)w\tau} \rho_0(w) dw.$$

Thus  $\kappa_n$  is determined by the microscopic density distribution  $\rho_n$  through

$$\kappa_n \approx \frac{\int \rho_n(w) dw}{\int (1 - e^{-w\tau}) \rho_n(w) dx} \quad (40)$$

This relation can now be used to develop computationally efficient reduced models for the weathering of oil. For example, (4) models a discrete release where oil is initially released at time  $t = 0$ , and no further oil is added to the spill. In this situation, we can estimate  $\kappa_n$  to obtain  $\kappa_n \approx n$  which is indeed our motivation for defining the dynamical variable  $\kappa_n$  in (36). We can also consider the case of a continuous spill, in which case we have the microscopic model (cf. (6))

$$\rho_n(w) = \Lambda \rho_n(w) + u_n(w)$$

where  $u_n$  represents the (deterministic or random) distribution of the oil added between time  $n - 1$  and  $n$ . Using the relation (40), we can thus obtain effective equations for  $\kappa_n$  to describe a continuous spill. This equation will replace the middle equation in (36). We can now use the asymptotic filter coefficients in (33) along with the model definition (36) to obtain reduced stochastic models for a variety of oil spill scenarios. We will discuss these methods elsewhere.

As we have discussed previously, we can interpret our methods as a computational framework to study the evolution of an autonomous, linear, high-dimensional system (Eq. (1)) in the sloppy-model universality class [62]. There has been considerable work on understanding the geometry of the parameter landscape in sloppy models, and in identifying robust combinations of parameters that can be extracted from data [56]. In contrast to earlier work, which is in a *static framework* (for an autonomous sloppy model, one attempts to model the system with another autonomous model, albeit one with robust parameters), our work suggests that the robust parameters (in our case the filter coefficients  $h_k^{(n)}$ ) might themselves evolve, even when the underlying model is autonomous. We hope to explore the consequences of this idea for other complex/nonlinear sloppy models.



Another direction we intend to pursue is to develop these ideas in a *non-parametric* setting [5], as applied to spatially extended systems with slow relaxation. In particular, we want to combine our techniques for slow relaxation with the Nonlinear Laplacian spectral analysis (NLSA) method [18, 26] which combines ideas from singular spectral analysis (see [58], [57]) with the ideas of Coifman and Lafon [17] who introduced Diffusion maps (Laplacian eigenmaps). Diffusion maps can be thought of as a powerful generalization of the Takens delay-coordinate embedding [55] to a form that is applicable to extended and high dimensional systems. The NLSA methods apply to large scale problems, but require short memory. Conversely, our methods in this work generates low dimensional reduced models from spatially homogeneous systems, but can handle long memory. A deeper comparison between these methods might reveal ways in which both of these techniques could be improved and made practical on large scale problems, such as those that arise from processes that are modeled by evolution equations with spatial dependence.

**Acknowledgements** S.V. would like to acknowledge the many, very illuminating discussions with Kevin Lin who was very generous with his time and his ideas. We are grateful to an anonymous referee for pointing out the potential connections between our work and the sloppy models universality class. This viewpoint turns out to be particularly fruitful. This work was funded in part by a Grant from GoMRI. We also received support from NSF-DMS-1109856 and NSF-OCE-1434198.

## Appendix 1: The Memory Kernel for Multiple Observables

One other comment is that we can indeed compute the memory kernel explicitly for the evaporation process (4), not just for the case with one observable, the mass  $M_n$ , but also more generally if we have a vector-valued linear observable  $\Phi$ , i.e  $l$  scalar-valued observables  $\Phi = \{\phi^1, \phi^2, \dots, \phi^l\}^T$ . Each scalar linear observable  $\phi^i$  is given by an element of  $\mathcal{H}^*$ , and we will denote the corresponding bra-vector by  $\langle \phi^i |$ . Using the Gram-Schmidt procedure if necessary, we can assume that the vectors  $\langle \phi^i |$  given an orthonormal basis for their span, a  $l$ -dimensional subspace of  $\mathcal{H}^*$ . The orthogonal projection  $P^* : \mathcal{H}^* \rightarrow \mathcal{H}^*$  onto this subspace is given by

$$P^* = |\phi^1\rangle\langle\phi^1| + |\phi^2\rangle\langle\phi^2| + \dots + |\phi^l\rangle\langle\phi^l|.$$

It follows that  $\langle \phi^i | P^* = \langle \phi^i | P = \langle \phi^i |$  and (11) gives

$$\langle \phi^i | \xi_{n+1} \rangle = \sum_{k=0}^n \sum_{j=1}^l \langle \phi^i | \Lambda(Q\Lambda)^k | \phi^j \rangle \langle \phi^j | \xi_n \rangle + \langle \phi^i | (\Lambda Q)^{n+1} | \rho_0 \rangle.$$

The quantities  $\langle \phi^i | \xi_{n+1} \rangle$  are the entries of the “vector” observable  $\Phi_{n+1}$ . Defining the matrices  $H_k$  by  $(H_k)_{ij} = \langle \phi^i | \Lambda(Q\Lambda)^k | \phi^j \rangle$  for  $k = 0, 1, 2, \dots$  and the (column) vectors  $\beta_n$  by the entries  $\beta_n^i = \langle \phi^i | (\Lambda Q)^{n+1} | \rho_0 \rangle$ , we have the Mori-Zwanzig decomposition

$$\Phi_{n+1} = \sum_{k=0}^n H_k \Phi_{n-k} + \beta_n. \quad (41)$$

If  $|\rho_0\rangle$  is in the span of  $|\phi^i\rangle$ , then  $Q|\rho_0\rangle = 0$  so that the noise  $\beta_n$  is identically zero. Taking  $|\rho_0\rangle = |\phi^1\rangle, |\phi^2\rangle, \dots, |\phi^l\rangle$  in turn, and collecting the corresponding column vectors  $\Phi_n$  into a  $l \times l$  matrix  $\Xi_n$ , we have

$$(\mathcal{E}_n)_{ij} = \langle \phi^i | \Lambda^n | \phi^j \rangle = \int_0^1 \phi^i(w) e^{-nw\tau} \phi^j(w) dw$$

is a symmetric matrix for each  $n$ , and

$$\mathcal{E}_{n+1} = \sum_{k=0}^n H_k \mathcal{E}_{n-k}, \quad \text{for } n = 0, 1, 2, \dots$$

As before, we can determine the memory kernel  $H_k$  using the  $\mathcal{Z}$ -transform. Defining the matrices

$$\hat{\mathcal{E}}(z) = \sum_{n=0}^{\infty} z^{-n} \mathcal{E}_n, \quad \hat{H}(z) = \sum_{n=0}^{\infty} z^{-n} H_n$$

we get

$$\hat{H}(z) = z(I - \hat{\mathcal{E}}(z)^{-1}).$$

The matrix  $\mathcal{E}_n$  is symmetric for all  $n$ , so that  $H_n$  is also symmetric for all  $n$ . We expect that the norm  $\|\mathcal{E}_n\|$  typically decays no faster than  $1/n$ . This is true for instance if the constant functions are in the range of  $P$ , or more generally if there are continuous functions  $\psi$  with  $\psi(0) > 0$  in the range of  $P$ . In this case, we expect that the norm of  $H_n$  decays no faster than  $1/(n \log^2(n))$  indicating again that, generically, one expects fat tails in the memory kernel for the system (4) if we use the Mori–Zwanzig decomposition based on any finite set of linear observables.

## Appendix 2: Orthogonal Dynamics

We will now compute the statistics of the noise process  $\beta_n$  in the Mori–Zwanzig decomposition (12) with the usual approach through the study of the projection equation (9) and the orthogonal dynamics (10). Since the orthogonal dynamics are linear, it suffices to solve the system

$$\langle F_0 | = \langle \delta(w-x) | Q, \quad \langle F_{n+1} | = \langle F_n | \Lambda Q, \quad n = 0, 1, 2, \dots$$

where  $x \in [0, 1]$  is fixed. A calculation reveals that, for any continuous function  $\phi$ ,

$$\langle F_0 | \phi \rangle = \langle \delta(w-x) | \phi \rangle - \langle \delta(w-x) | P | \phi \rangle = \phi(x) - \int_0^1 \phi(w) dw.$$

We will thus associate  $\langle F_0 |$  with the “function”  $F_0(w) = \delta(w-x) - 1$ . We can follow this computation to solve the orthogonal dynamics equations recursively. For example,

$$\begin{aligned} \langle F_1 | \phi \rangle &= \langle \delta(w-x) - 1 | \Lambda | \phi \rangle - \langle \delta(w-x) - 1 | \Lambda P | \phi \rangle \\ &= \int_0^1 (\delta(w-x) - 1) e^{-w\tau} \phi(w) dw - \int_0^1 (\delta(w-x) - 1) e^{-w\tau} dw \int_0^1 \phi(w) dw \\ &= \int_0^1 \left[ e^{-x\tau} \delta(w-x) - e^{-w\tau} - e^{-x\tau} + \frac{1-e^{-\tau}}{\tau} \right] \phi(w) dw, \end{aligned}$$

so that  $\langle F_1 |$  corresponds to the function  $F_1(w) = e^{-x\tau} \delta(w-x) - e^{-w\tau} - e^{-x\tau} + \frac{1-e^{-\tau}}{\tau}$ . Using the fact that  $Q$  and  $\Lambda$  are self-adjoint operators on  $\mathcal{H}$ , and further  $\langle \psi | \Lambda | \phi \rangle =$

$\int \psi(w)e^{-w\tau}\phi(w)dw$  so that  $\Lambda$  is diagonal on the “basis”  $\{\delta(w-x)\}_{\{x \in [0,1]\}}$ , an inductive argument shows that  $F_n(w) = e^{-nx\tau}\delta(w-x) + \Psi_n(w;x)$  where  $\Psi_n$  is a smooth, symmetric function  $\Psi_n(w;x) = \Psi_n(x;w)$ . We will use these conclusions to verify the full solution for  $\langle F_n|$  that we obtain below by independent means.

Consider the  $\mathcal{Z}$ -transform  $\langle \hat{F}| = \sum z^{-n}\langle F_n|$ . The orthogonal dynamics imply

$$\langle \hat{F}|(1 - z^{-1}\Lambda Q) = \langle \hat{F}| - z^{-1}\langle \hat{F}|\Lambda + z^{-1}\langle \hat{F}|\Lambda|1\rangle|1\rangle = \langle F_0|.$$

Using the ansatz  $\hat{F}(z, x, w) = \hat{A}(z, x)\delta(w-x) + \hat{\Psi}(z, x, w)$  corresponding to a decomposition of  $F_n$  into its singular and regular parts, we get the pair of equations

$$(1 - z^{-1}e^{-x\tau})\hat{A} = 1, \\ \hat{\Psi} - z^{-1}e^{-x\tau}\hat{\Psi} + z^{-1}e^{-w\tau}\hat{A} + z^{-1}\int e^{-x\tau}\hat{\Psi}dx = -1,$$

where we have suppressed the arguments  $(z, x, w)$  for  $\hat{A}$  and  $\hat{\Psi}$  for clarity. We can solve the first equation to obtain

$$\hat{A} = \frac{1}{1 - z^{-1}e^{-x\tau}}.$$

Using this in the second equation, we obtain

$$\hat{\Psi} = -\frac{1}{(1 - z^{-1}e^{-x\tau})(1 - z^{-1}e^{-w\tau})} - \frac{z^{-1}C(z, w)}{1 - z^{-1}e^{-x\tau}},$$

where  $C(z, w) = \int e^{-x\tau}\hat{\Psi}dx$  is determined in terms of the required solution  $\hat{\Psi}$  self-consistently. Multiplying by  $e^{-x\tau}$  and integrating in  $x$ , and solving the resulting equation for  $C(z, w)$ , we obtain

$$C(z, w) = -\frac{z(\tau - \log(1 - e^\tau z) + \log(1 - z))}{(1 - z^{-1}e^{-w\tau})(\log(1 - z) - \log(1 - e^\tau z))}.$$

Using this result in the computation for  $\hat{\Psi}$  gives

$$\hat{\Psi}(z, x, w) = \frac{\tau}{(1 - z^{-1}e^{-w\tau})(1 - z^{-1}e^{-x\tau})(\log(1 - z) - \log(1 - e^\tau z))}.$$

This gives a complete solution of orthogonal dynamics equation by

$$\hat{F}(z) = \frac{\delta(w-x)}{1 - z^{-1}e^{-x\tau}} + \frac{\tau}{(1 - z^{-1}e^{-w\tau})(1 - z^{-1}e^{-x\tau})(\log(1 - z) - \log(1 - e^\tau z))}.$$

The singular part of  $F_n$  is therefore  $e^{-n\tau x}\delta(w-x)$  as we noted above. Further, the regular part  $\hat{\Psi}$  is symmetric in  $w$  and  $x$ , implying this property for each of the functions  $\Psi_n$ . Finally, for an observable given by a continuous function  $g$ , the solution to the orthogonal dynamics is given by

$$\sum_{n=0}^{\infty} z^{-n}\langle g|Q(\Lambda Q)^n|\phi\rangle = \int_0^1 \int_0^1 g(x)\hat{F}(z, w, x)\phi(w)dw dx.$$

For the observable  $M_n$ , the prediction for the total mass at the next time step, we have  $\langle g \rangle = \langle 1 | \Lambda$ . The  $\mathcal{Z}$ -transform of the memory kernel is given by

$$\begin{aligned} H(z) &= \sum_{k \geq 1} z^{-k} h_k = z^{-2} \sum_{k \geq 1} z^{-k+2} \langle 1 | (\Lambda Q)^{k-1} \Lambda | 1 \rangle \\ &= z^{-1} \langle 1 | \Lambda | 1 \rangle + z^{-2} \sum_{n \geq 0} z^{-n} \langle 1 | \Lambda Q (\Lambda Q)^n \Lambda | 1 \rangle \\ &= z^{-1} \frac{1 - e^{-\tau}}{\tau} + z^{-2} \int_0^1 \int_0^1 e^{-x\tau} \hat{F}(z, w, x) e^{-w\tau} dw dx \\ &= \left[ 1 - \frac{\tau}{\log(e^\tau z - 1) - \log(z - 1)} \right]. \end{aligned}$$

The  $\mathcal{Z}$ -transform of the expected values of the noise sequence  $\beta_n$  is given by

$$\sum z^{-n} \mathbb{E}[\beta_n] = \sum z^{-n} \mathbb{E}[\langle F_n | \rho_0 \rangle] = \int_0^1 \int_0^1 e^{-x\tau} \hat{F}(z, w, x) dw dx = 0$$

and the correlations between the noise  $\beta_n$  and the mass  $M_j$  are given by  $\mathbb{E}[\beta_n M_j] = \bar{\sigma}^2 \langle 1 | (\Lambda Q)^n \Lambda^j | 1 \rangle$  (see Sect. 5). Taking the (two index)  $\mathcal{Z}$ -transform, noting that  $\beta_0 = 0$ , we have

$$\begin{aligned} \sum_{n \geq 1} \sum_{j \geq 0} z^{-n} \zeta^{-j} \mathbb{E}[\beta_n M_j] &= \bar{\sigma}^2 \sum_{n \geq 1} \sum_{j \geq 0} z^{-n} \zeta^{-j} \langle 1 | (\Lambda Q)^n \Lambda^j | 1 \rangle \\ &= \bar{\sigma}^2 z^{-1} \sum_{n \geq 0} \sum_{j \geq 0} z^{-n} \zeta^{-j} \langle 1 | \Lambda Q (\Lambda Q)^n \Lambda^j | 1 \rangle \\ &= \bar{\sigma}^2 z^{-1} \int_0^1 \int_0^1 \frac{e^{-x\tau} \hat{F}(z, w, x)}{1 - \zeta^{-1} e^{-w\tau}} dw dx \\ &= \frac{\bar{\sigma}^2}{\tau} z^{-1} \left[ \frac{1 - z^{-1} e^{-\tau}}{(1 - z^{-1})(1 - z^{-1} e^{-\tau})} \right. \\ &\quad \left. + \frac{\log\left(\frac{1 - z^{-1}}{1 - z^{-1} e^{-\tau}}\right) \left( \log\left(z \log\left(\frac{\zeta - 1}{\zeta e^\tau - 1}\right) - \zeta \log\left(\frac{z - 1}{z e^\tau - 1}\right) \right) \right)}{(z - \zeta) \log\left(\frac{z - 1}{e^\tau z - 1}\right)} \right]. \end{aligned}$$

It is not true that  $\mathbb{E}[\beta_n M_j] = 0$  if  $n > j$ , as one would expect in the Mori–Zwanzig decomposition for a system with an invariant measure. In particular,

$$\mathbb{E}[\beta_2 M_1] = \frac{\bar{\sigma}^2 (1 - e^{-\tau}) ((\tau - 2) + (\tau + 2) e^{-\tau})}{2\tau^2} \neq 0$$

### Appendix 3: Sampling Initial Conditions

For any prescribed value  $0 < \bar{\sigma}^2 < \infty$ , we can indeed find a family of  $I$ -dependent distributions such that

$$\mu_\gamma \rightarrow 1, \frac{\sigma_\gamma^2}{I} \rightarrow \bar{\sigma}^2 \text{ as } I \rightarrow \infty$$

by appropriately truncating and rescaling a distribution that has finite mean but infinite variance. For example, the function

$$f(x) = \begin{cases} \frac{9}{10} & 0 \leq x \leq \frac{2}{3}, \\ \frac{9}{10} \left(\frac{3x}{2}\right)^{5/2} & x > \frac{2}{3} \end{cases}$$

satisfies  $f \geq 0$  on  $(0, \infty)$  and  $\int_0^\infty f(x)dx = 1$ , so  $f$  is indeed a nonnormalized density on  $(0, \infty)$ . Further  $\int_0^\infty xf(x)dx = 1$  and  $\int_0^L x^2 f(x)dx \sim \sqrt{\frac{32}{75}L}$  for  $L \gg 1$ . We can therefore define a sequence of  $I$  dependent distributions by truncating the support of  $f$  and renormalizing to have unit mass, i.e.

$$f_I(x) = \begin{cases} c_I f(x) & 0 \leq x \leq L_I \\ 0 & x > L_I, \end{cases}$$

where  $L_I$  is any sequence satisfying  $L_I \geq 2/3$  for all  $I$ ,  $L_I \nearrow \infty$  and  $\sqrt{\frac{32}{75}L_I} \rightarrow \bar{\sigma}^2$  as  $I \rightarrow \infty$ . Given such a sequence  $L_I$ , the normalization  $c_I$  is determined by  $\int_0^{L_I} f_I(x)dx = 1$  so that  $c_I \rightarrow 1$ .

## Appendix 4: Asymptotic Solutions of the Yule–Walker Equations

We seek a solution to (30) as an asymptotic series in  $n$ , i.e. solutions of the form

$$h_j^{(n)} = a_j^0 + \frac{1}{n}a_j^1 + \frac{1}{n^2}a_j^2 + \cdots. \quad (42)$$

The difficulty in solving this system is evident if we expand the coefficient matrix  $A$  as a power series in  $n$ :

$$A = \frac{1}{2n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} + \frac{1}{4n^2} \begin{pmatrix} 2 & 3 & \cdots & L+1 \\ 3 & 4 & \cdots & L+2 \\ \vdots & \vdots & \ddots & \vdots \\ L+1 & L+2 & \cdots & 2L \end{pmatrix} + \cdots.$$

Assuming  $L \geq 3$ , the two matrices displayed in the expansion of  $A$  are singular. The first matrix has rank 1, the second has rank 2. Indeed the first  $L-1$  matrices in the expansion of  $A$  are all singular and their (row) nullspaces are nested

$$v^T \begin{pmatrix} 2 & 3 & \cdots & L+1 \\ 3 & 4 & \cdots & L+2 \\ \vdots & \vdots & \ddots & \vdots \\ L+1 & L+2 & \cdots & 2L \end{pmatrix} = 0 \implies v^T \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = 0,$$

and so on. The determinant of  $A$  is thus very close to zero ( $\det\{A\} \sim O(n^{-L^2})$  as we see below) so it is not clear that we have solutions for  $h^{(n)}$  where the leading order behavior stays  $O(1)$  instead of diverging with  $n$ . Proving the boundedness of  $h^{(n)}$  and determining the  $O(1)$  solution thus requires consideration of  $L$  solvability conditions given by the vectors that span the common (row)-nullspaces of the initial  $j$  terms in the expansion of  $A$  for  $j = 1, 2, \dots, L-1$ . Higher order terms will require even longer expansion of the matrices and more solvability conditions.

In the general case of a process with slowly decaying correlations, it is still true that the matrix of coefficients in the Yule–Walker equation is nearly singular, and one does have to go through the process described above to find optimal, reduced dimensional, models for such systems. For the evaporation process (4) however, the coefficient matrix has a special structure, that we exploit to find the solutions for the optimal filter  $h^{(n)}$ . The matrix  $A$  is a *Cauchy matrix* [48] i.e its entries are of the form  $A_{ij} = 1/(x_i - y_j)$ . In particular, we can choose  $x_i = 2n - i$  and  $y_j = j$ . The determinant of a Cauchy matrix  $A_{ij} = 1/(x_i - y_j)$  is given by [48]

$$\det \{A\} = \frac{\prod_{i>j} (x_i - x_j)(y_j - y_i)}{\prod_i \prod_j (x_i - y_j)}.$$

For the particular matrix  $A$  from above, the terms in the numerator are all bounded by  $L$  and the terms in the denominator are all  $\approx 2n$  if  $n \gg L$ . Consequently,  $\det A \sim O(n^{-L^2})$ . The matrix  $\hat{A}_m$  obtained by replacing the  $m$ th column of  $A$  by the vector  $v_i = \frac{1}{2n-i}$  is also a Cauchy matrix  $\hat{A}_{ij} = 1/(x_i - \hat{y}_j)$ , with the same choice  $x_i = 2n - i$  and

$$\hat{y}_j = \begin{cases} y_j & j \neq m, \\ 0 & j = m. \end{cases}$$

Cramer's rule now yields,

$$h_m^{(n)} = \frac{\det \{\hat{A}\}}{\det \{A\}} = \prod_{i \neq m} \frac{i}{i - m} \prod_i \frac{2n - i - m}{2n - i}. \quad (43)$$

## References

1. Amir, A., Oreg, Y., Imry, Y.: On relaxations and aging of various glasses. *Proc. Natl Acad. Sci. U.S.A.* **109**, 1850–1855 (2012)
2. Arnold, H.M., Moroz, I.M., Palmer, T.N.: Stochastic parametrizations and model uncertainty in the Lorenz '96 system. *Philos. Trans. R. Soc. Lond. A* **371**, 20120510 (2013)
3. Baladi, V.: Positive Transfer Operators and Decay of Correlations, vol. 16. *Advanced Series in Nonlinear Dynamics*. World Scientific, Singapore (2000)
4. Berkenbusch, M.K., Claus, I., Dunn, C., Kadanoff, L.P., Nicewicz, M., Venkataramani, S.C.: Discrete charges on a two dimensional conductor. *J. Stat. Phys.* **116**, 1301–1358 (2004)
5. Berry, T., Harlim, J.: Forecasting turbulent modes with nonparametric diffusion models: learning from Noisy data. *Physica D* **320**, 57–76 (2016)
6. Bouchaud, J.-P., Cugliandolo, L.F., Kurchan, J., Mezard, M.: Out of equilibrium dynamics in spin-glasses and other glassy systems. In: *Spin Glasses and Random Fields*, pp. 161–223. World Scientific, Singapore (1998)
7. Bouchaud, J.-P.: Aging in glassy systems: new experiments, simple models, and open questions. In: Cates, M.E., Evans, M. (eds.) *Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability and Flow*, pp. 285–304. Institute of Physics, Bristol (2000)
8. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, Hoboken (2015)
9. Brown, K.S., Sethna, J.P.: Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* **68**, 021904 (2003)
10. Budisic, M., Mohr, R., Mezic, I.: Applied Koopmanism. *Chaos* **22**(4), 047510 (2012)
11. Chekroun, M.D., Kondrashov, D., Ghil, M.: Predicting stochastic systems by noise sampling, and application to the El Niño-southern oscillation. *Proc. Natl Acad. Sci. U.S.A.* **108**, 11766–11771 (2011)
12. Chorin, A.J., Hald, O.H.: *Stochastic Tools in Mathematics and Science*, vol. 58. *Texts in Applied Mathematics*. Springer, New York (2014)
13. Chorin, A.J., Hald, O.H., Kupferman, R.: Optimal prediction and the Mori–Zwanzig representation of irreversible processes. *Proc. Natl Acad. Sci. U.S.A.* **97**, 2968–2973 (2000)

14. Chorin, A., Hald, O., Kupferman, R.: Optimal prediction with memory. *Physica D* **166**, 239–257 (2002)
15. Chorin, A.J., Lu, F.: Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. *Proc. Natl Acad. Sci. U.S.A.* **112**, 9804–9809 (2015)
16. Chorin, A., Stinis, P.: Problem reduction, renormalization, and memory. *Commun. Appl. Math. Comput. Sci.* **1**, 1–27 (2007)
17. Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmonic Anal.* **21**, 5–30 (2006)
18. Comeau, D., Zhao, Z., Giannakis, D., Majda, A.J.: Data-driven prediction strategies for low-frequency patterns of north pacific climate variability. *Clim. Dyn.* **48**(5), 1855–1872 (2015)
19. Crisanti, A., Ritort, F.: Violation of the fluctuation–dissipation theorem in glassy systems: basic notions and the numerical evidence. *J. Phys. A* **36**, R181 (2003)
20. Darve, E., Solomon, J., Kia, A.: Computing generalized langevin equations and generalized Fokker–Planck equations. *Proc. Natl Acad. Sci. U.S.A.* **106**, 10884–10889 (2009)
21. Dixon, P.K., Wu, L., Nagel, S.R., Williams, B.D., Carini, J.P.: Scaling in the relaxation of supercooled liquids. *Phys. Rev. Lett.* **65**, 1108–1111 (1990)
22. Fingas, M.F.: A literature review of the physics and predictive modelling of oil spill evaporation. *J. Hazard. Mater.* **42**, 157–175 (1995)
23. Fingas, M.: Modeling evaporation using models that are not boundary-layer regulated. *J. Hazard. Mater.* **107**, 27–36 (2004)
24. Fingas, M.: Modeling oil and petroleum evaporation. *J. Pet. Sci. Res.* **2**(3), 104–115 (2013)
25. Flajolet, P., Odlyzko, A.: Singularity analysis of generating functions. *SIAM J. Discret. Math.* **3**, 216–240 (1990)
26. Giannakis, D., Majda, A.J.: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl Acad. Sci. U.S.A.* **109**, 2222–2227 (2012)
27. Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17**, R55 (2004)
28. Givon, D., Kupferman, R., Hald, O.H.: Existence proof for orthogonal dynamics and the Mori–Zwanzig formalism. *Isr. J. Math.* **145**, 221–241 (2005)
29. Harlim, J., Kang, E.L., Majda, A.J.: Regression models with memory for the linear response of turbulent dynamical systems. *Commun. Math. Sci.* **11**(2), 481–498 (2013)
30. Hoult, D.P. (ed.): *Oil on the Sea: Proceedings of a Symposium on the Scientific and Engineering Aspects of Oil Pollution of the Sea*. Springer, New York (1969)
31. Jazwinski, A.H.: *Stochastic Processes and Filter Theory*. Academic Press, New York (1970)
32. Kampen, N.V.: *Stochastic Processes in Physics and Chemistry*, 3rd edn. North-Holland Personal Library. North Holland, Amsterdam (2007)
33. Kawasaki, K.: Simple derivations of generalized linear and nonlinear Langevin equations. *J. Phys. A* **6**, 1289 (1973)
34. Kawasaki, K.: Theoretical methods dealing with slow dynamics. *J. Phys.* **12**, 6343 (2000)
35. Kondrashov, D., Chekroun, M., Ghil, M.: Data-driven non-Markovian closure models. *Physica D* **297**, 33–55 (2015)
36. Kubo, R.: The fluctuation–dissipation theorem. *Rep. Prog. Phys.* **29**, 255 (1966)
37. Kutner, M., Nachtsheim, C., Neter, J., Li, W.: *Applied Linear Statistical Models*. McGraw-Hill/Irwin, Chicago (2004)
38. Lin, K., Lu, F.: Stochastic parametrization, filtering, and the Mori–Zwanzig formalism. Preprint (2017)
39. Lu, F., Lin, K.K., Chorin, A.J.: Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation. *Physica D* **340**, 46–57 (2017)
40. Mackay, D., Matsugu, R.S.: Evaporation rates of liquid hydrocarbon spills on land and water. *Can. J. Chem. Eng.* **51**, 434–439 (1973)
41. Majda, A.J., Harlim, J.: Physics constrained nonlinear regression models for time series. *Nonlinearity* **26**, 201 (2013)
42. Matan, K., Williams, R.B., Witten, T.A., Nagel, S.R.: Crumpling a thin sheet. *Phys. Rev. Lett.* **88**, 076101 (2002)
43. Moghimi, S., Ramírez, J.M., Restrepo, J.M., Venkataramani, S.C.: Mass exchange dynamics of surface and subsurface oil in shallow-water transport. Preprint (2017)
44. Mori, H.: Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **33**, 423–455 (1965)
45. Oded, B., Rubinstein, S.M., Fineberg, J.: Slip-stick and the evolution of frictional strength. *Nature* **463**, 76–9 (2010)
46. Oppenheim, A.V., Schaffer, R.W., Buck, J.R.: *Discrete-Time Signal Processing*, 2nd edn. Prentice-Hall Signal Processing Series. Prentice Hall, Englewood Cliffs (1999)
47. Ott, E.: *Chaos in Dynamical Systems*. Cambridge University Press, New York (1993)

48. Polya, G., Szegő, G.: Problems and Theorems in Analysis II: Theory of Functions, Zeros, Polynomials, Determinants, Number Theory, Geometry, Classics in Mathematics. Springer, New York (1998)
49. Restrepo, J.M., Venkataramani, S.C., Dawson, C.: Nearshore sticky waters. *Ocean Model.* **80**, 49–58 (2014)
50. Restrepo, J.M., Ramírez, J.M., Venkataramani, S.C.: An oil fate model for shallow waters. *J. Marine Sci. Eng.* **3**, 1504–1543 (2015)
51. Spaulding, M.L.: A state-of-the-art review of oil spill trajectory and fate modeling. *Oil Chem. Pollut.* **4**, 39–55 (1988)
52. Stinis, P.: Renormalized Mori–Zwanzig-reduced models for systems without scale separation. *Proc. R. Soc. Lond. Ser. A* **471**, 20140446 (2015)
53. Stiver, W., Mackay, D.: Evaporation rate of spills of hydrocarbons and petroleum mixtures. *Environ. Sci. Technol.* **18**(11), 834–840 (1984)
54. Sutton, O.G.: Wind structure and evaporation in a turbulent atmosphere. *Proc. R. Soc. Lond. Ser. A* **146**, 701–722 (1934)
55. Takens, E.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.-S. (eds.) *Dynamical Systems and Turbulence*, pp. 366–381. Springer, Berlin (1981)
56. Transtrum, M.K., Machta, B.B., Sethna, J.P.: Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E* **83**, 036701 (2011)
57. Vautard, R., Yiou, P., Ghil, M.: Singular-spectrum analysis: a toolkit for short, Noisy chaotic signals. *Physica D* **58**, 95–126 (1992)
58. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* **35**, 395–424 (1989)
59. Venturi, D., Cho, H., Karniadakis, G.: Mori–Zwanzig approach to uncertainty quantification. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) *Handbook of Uncertainty Quantification*. Springer, Heidelberg (2016)
60. Venturi, D., Karniadakis, G.: Convolutionless Nakajima–Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proc. R. Soc. Math. Phys. Eng. Sci.* **470**, 20130754–20130754 (2014)
61. Walker, G.: On periodicity in series of related terms. *Proc. R. Soc. Lond. Ser. A* **131**, 518–532 (1931)
62. Waterfall, J.J., Casey, F.P., Gutenkunst, R.N., Brown, K.S., Myers, C.R., Brouwer, P.W., Elser, V., Sethna, J.P.: Sloppy-model universality class and the Vandermonde matrix. *Phys. Rev. Lett.* **97**, 150601 (2006)
63. Yule, G.U.: On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philos. Trans. R. Soc. Lond. Ser. A* **226**, 267–298 (1927)
64. Zwanzig, R.: Problems in nonlinear transport theory. In: Garrido, L. (ed.) *Systems Far from Equilibrium*, vol. 132. *Lecture Notes in Physics*. Springer, Berlin (1980)
65. Zwanzig, R.: Nonlinear generalized Langevin equations. *J. Stat. Phys.* **9**, 215–220 (1973)
66. Zwanzig, R.: *Nonequilibrium Statistical Mechanics*. Oxford University Press, New York (2001)