# Obtaining coarse-grained models from multiscale data

Sebastian Krumscheid

CSQI – MATHICSE, École Polytechnique Fédérale de Lausanne

April 17, 2015

EPSRC
Engineering and Physical Sciences Research Council

MARIE CURIE ACTIONS

multiflow
FP7-PEOPLE-2007-1-1-ITN

erc

joint work with G. A. Pavliotis (IC, Mathematics) and S. Kalliadasis (IC, Chemical Engineering)

# Data-driven coarse-graining

## What are we interested in?

- Many Dynamical systems in the natural sciences are characterised by the presence of processes that occur across several length and time scales, e.g. atmosphere-ocean system, biological systems, materials and molecular dynamics, etc.
- Full multiscale system is cumbersome to analyse: high-dimensional, nonlinear coupling, small scale vs. large scale effects, etc. Sometimes, it is not even fully known.
- Commonly, only the evolution of a few selected degrees of freedom is of main interest, which are often observable.

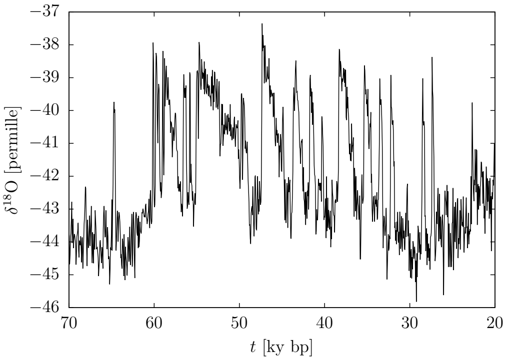
### Idea: Data-Driven Coarse-Graining

Use data (observations) of full system to find an adequate simplified **low-dimensional** coarse-grained model that retains the essential dynamic characteristics of the degrees of freedom of interest.

# Data-driven coarse-graining: A motivating example

## The paleoclimatic record

- Celebrated (partial) record of $\delta^{18}O$ ($\approx$ proxy for temperature) from the NGRIP ice core during last glacial period [Anderson et al. Nature, 2004]




- Temperature is a single degree of freedom of an arguably high-dimensional climate model.

# Inference for coarse-grained dynamical systems

## Abstract framework

- Consider a dynamical system $Z^\varepsilon$ evolving according to

$$\frac{dZ^\varepsilon}{dt} = F(Z^\varepsilon)\,, \quad Z^\varepsilon(t) \in \mathcal{Z}$$

- Decompose the state space into subspaces $\mathcal{X}$ and $\mathcal{Y}$:

$$\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}\,, \quad \dim(\mathcal{X}) \ll \dim(\mathcal{Y})$$

$\mathcal{X}$: state space of degrees of freedom of interest

### Data-Driven Coarse-Graining

Use data $X^\varepsilon = P_{\mathcal{X}} Z^\varepsilon$ to find a stochastic coarse-grained system

$$dX = f(X;\theta)\,dt + g(X;\theta)\,dW_t\,, \quad X(t) \in \mathcal{X}$$

such that $X \approx X^\varepsilon$ (in an appropriate sense).

### But ...

Inverse problem for $\theta$ based on $X^\varepsilon = P_{\mathcal{X}} Z^\varepsilon$ not straightforward!

# Failure of classic approaches: Multiscale diffusions

A toy example: homogenization [Pavliotis, Stuart. Springer. 2008]

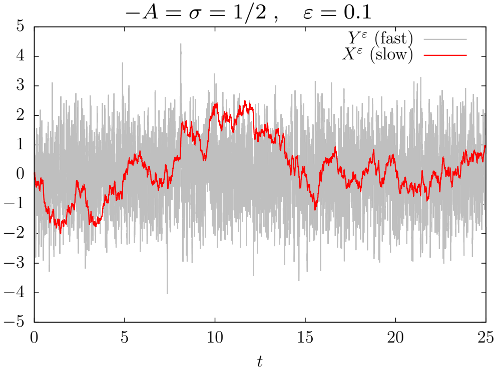
## Dynamical System

$$dX^\varepsilon = \left( AX^\varepsilon + \frac{\sqrt{\sigma}}{\varepsilon} Y^\varepsilon \right) dt \,,$$

$$dY^\varepsilon = -\frac{1}{\varepsilon^2} Y^\varepsilon \, dt + \frac{\sqrt{2}}{\varepsilon} \, dV_t$$

## Coarse-Grained System

$$dX = AX \, dt + \sqrt{2\sigma} \, dW_t$$




- Coarse-grained system rigorously obtained via Homogenization theory
- Commonly used parametric estimators for SDEs are MLE and QVP:

$$\hat{A}_{\mathrm{MLE}} = -0.026 \not\approx -0.5 = A \,, \quad \hat{\sigma}_{\mathrm{QVP}} = 0.026 \not\approx 0.5 = \sigma \,.$$

## Failure for multiscale systems on the diffusive time scale

[Pavliotis, Stuart. 2007], [Papavasiliou, Pavliotis, Stuart. 2009], [Azencott, Beri, Timofeyev. 2010]

Standard estimators for **coarse-grained** system based on observations from full **multiscale** system are (asymptotically) biased

- For the toy example: $\lim_{\varepsilon \to 0} \lim_{T \to \infty} \hat{A}_{\mathrm{MLE}}(\varepsilon, T) = A + \sigma$

## General abstract nonsense or practically relevant?

- A practitioner believes the "true" coarse-grained model is

$$dX = f(X)\, dt + g(X)\, dW_t$$

- An estimator is derived from this model: $\mathcal{E}(X)$
- One does not observe $X$, but a perturbed version $X^\varepsilon$ instead.
- Is the estimator robust w.r.t. the perturbation? Does it hold that

$$\mathcal{E}(X^\varepsilon) \to \mathcal{E}(X)\,, \quad \text{if} \quad X^\varepsilon \to X \quad \text{as } \varepsilon \to 0\,?$$

# Derivation of general purpose estimator

- Let $X_\xi$ denote the solution to coarse-grained Itô SDE

$$dX = f(X)\,dt + g(X)\,dW_t\ , \quad X(0) = \xi \in \mathbb{R}^d\ ,$$

  with $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $g\colon \mathbb{R}^d \to \mathbb{R}^{d\times r}$
- Both $f$ and $G := gg^T \in \mathbb{R}^{d\times d}$ depend on unknown parameters $\theta \equiv (\theta_1, \ldots, \theta_n)^T \in \Theta \subset \mathbb{R}^n$:

$$f(x) \equiv f(x;\theta) := \sum_{j=1}^{n} \theta_j f_j(x) \quad \text{and} \quad G(x) \equiv G(x;\theta) := \sum_{j=1}^{n} \theta_j G_j(x)$$

- For any function $\phi \in C_b^2(\mathbb{R}^d)$ and any $t > 0$, Itô's formula implies

$$\mathbb{E}\Big(\phi\big(X_\xi(t)\big)\Big) - \phi(\xi) = \sum_{j=1}^{n} \theta_j \int_0^t \mathbb{E}\Big((\mathcal{L}_j\phi)\big(X_\xi(s)\big)\Big)\,ds$$

  with generators $\mathcal{L}_j\phi = f_j \cdot \nabla\phi + \frac{1}{2}G_j : \nabla\nabla\phi$

- This can be written as

$$a(\xi)^T\theta = b_c(\xi) \ ,$$

$b_c(\xi) := \mathbb{E}\big(\phi\big(X_\xi(t)\big)\big) - \phi(\xi)$ and $a(\xi) := \Big(\int_0^t \mathbb{E}\Big((\mathcal{L}_j\phi)\big(X_\xi(s)\big)\Big)\, ds\Big)_{1\leq j\leq n} \in \mathbb{R}^n$

- Equation $a(\xi)^T\theta = b_c(\xi)$ is ill-posed
- Since the equation is valid for any trial point $\xi$, we can overcome this shortcoming by considering multiple trial points $(\xi_i)_{1\leq i\leq m}$, thus

$$A\theta = b \ ,$$

with $A := \big(a(\xi_i)^T\big)_{1\leq i\leq m} \in \mathbb{R}^{m\times n}$ and $b := \big(b_c(\xi_i)\big)_{1\leq i\leq m} \in \mathbb{R}^m$

- Define estimator as least-squares solutions

$$\hat{\theta} := \arg\min_{x\in\mathcal{S}} \|x\|_2^2 \ , \quad \mathcal{S} := \big\{x \in \mathbb{R}^n\colon\ \|Ax - b\|_2^2 = \min\big\}$$

# The toy example revisited: Does it work at all?

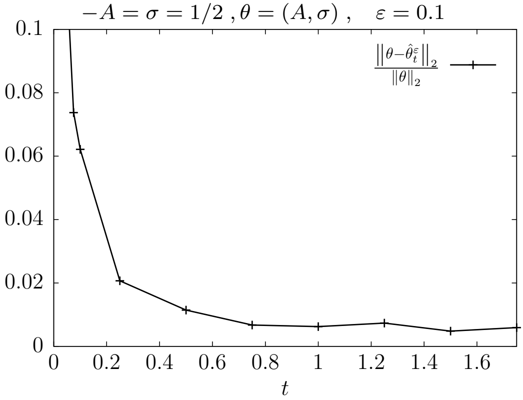
## Dynamical System

$$dX^\varepsilon = \left( AX^\varepsilon + \frac{\sqrt{\sigma}}{\varepsilon} Y^\varepsilon \right) dt \,,$$

$$dY^\varepsilon = -\frac{1}{\varepsilon^2} Y^\varepsilon \, dt + \frac{\sqrt{2}}{\varepsilon} \, dV_t$$

## Coarse-Grained System

$$dX = AX \, dt + \sqrt{2\sigma} \, dW_t$$




## Observations

- consistent parameter estimation seems possible
- sufficiently large $t$ removes multiscale bias:

$$\text{multiscale bias} \approx \sigma\big(|A| + t^{-1}\big)\varepsilon^2 + \mathcal{O}(\varepsilon^4)$$

# Just lucky? A more complex system.

## Dynamical System

$$dX^\varepsilon = \left(\frac{Y^\varepsilon}{\varepsilon}\sqrt{\sigma_a + \sigma_b (X^\varepsilon)^2} + (A-\sigma_b)X^\varepsilon - B(X^\varepsilon)^3\right) dt\ , \quad Y^\varepsilon \text{ as before}$$

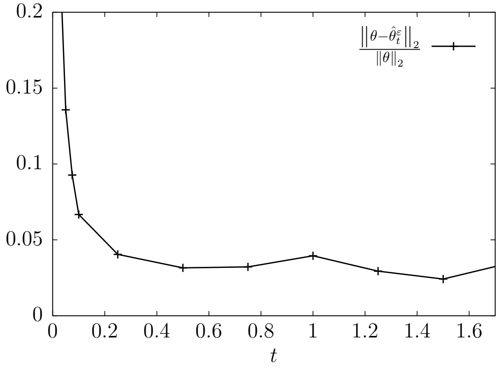
## Coarse-Grained System

$$dX = (AX - BX^3)\, dt + \sqrt{2(\sigma_a + \sigma_b X^2)}\, dW_t$$

- True values:
  $\theta = (A, B, \sigma_a, \sigma_b)$ with

  $$A = 1\ , \quad \sigma_a = 0.49$$
  $$B = 2\ , \quad \sigma_b = 0.81$$

- $\varepsilon = 0.1$



# What about rigorous results? Not just lucky! [K. arXiv, 2014]

## Robustness

### Assumptions

A1 $X^\varepsilon \Rightarrow X$ as $\varepsilon \to 0$ in $C([0,T],\mathbb{R}^d)$

A2 Sampling errors in discretely sampled observations vanish as $h \to 0$

A3 Error of approximating time integrals by numerical quadrature vanishes as $\delta \to 0$

A4 Error of approximating expectations by finite averages vanishes as $N \to \infty$

### Proposition (Robustness)

*Under assumptions A1–A4, the estimator is robust with respect to multiscale perturbations, in the sense that*

$$\lim_{\varepsilon \to 0} \hat{\theta}(X^\varepsilon) = \theta \ , \quad \textit{a.s.}$$

*for any* $t > 0$ *and* $\phi \in C_b^2(\mathbb{R}^d)$.

# Many more Examples [K., Pavliotis, Kalliadasis. MMS, 2013], [Kalliadasis, K., Pavliotis. arXiv , 2014]

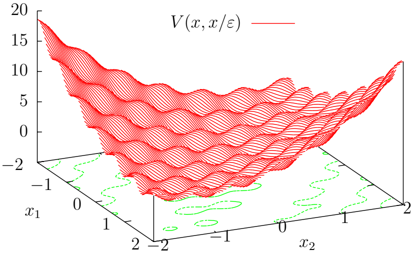
Theory covers problem of obtaining coarse-grained models for:

# Many more Examples [K., Pavliotis, Kalliadasis. MMS, 2013], [Kalliadasis, K., Pavliotis. arXiv , 2014]

Theory covers problem of obtaining coarse-grained models for:

- multiscale problems with multidimensional coarse-grained models

Brownian Motion in two-scale Potential $x \mapsto V(x, x/\varepsilon)$



# Many more Examples [K., Pavliotis, Kalliadasis. MMS, 2013], [Kalliadasis, K., Pavliotis. arXiv , 2014]

Theory covers problem of obtaining coarse-grained models for:

- multiscale problems with multidimensional coarse-grained models
- stochastic PDEs

Burgers equation in a small noise regime:

$$du_\varepsilon = \Big( (\partial_x^2 + 1) u_\varepsilon + \tfrac{1}{2} \partial_x {u_\varepsilon}^2 + \varepsilon^2 \nu u_\varepsilon \Big)\, dt + \varepsilon \mathcal{Q}\, d\mathcal{W}_t$$

Study solutions of $\mathcal{O}(\varepsilon)$ on times scales $\mathcal{O}(1/\varepsilon^2)$: diffusive rescaling $v_\varepsilon$ s.t. $\varepsilon v_\varepsilon(\varepsilon^2 t) = u_\varepsilon(t)$

# Many more Examples [K., Pavliotis, Kalliadasis. MMS, 2013], [Kalliadasis, K., Pavliotis. arXiv , 2014]

Theory covers problem of obtaining coarse-grained models for:

- multiscale problems with multidimensional coarse-grained models
- stochastic PDEs
- deterministic systems with stochastic limit:
  - Kac–Zwanzig models: "particle in a heat bath"
  - deterministic model of Brownian motion
  - fast deterministic chaos
  - ...
- etc.

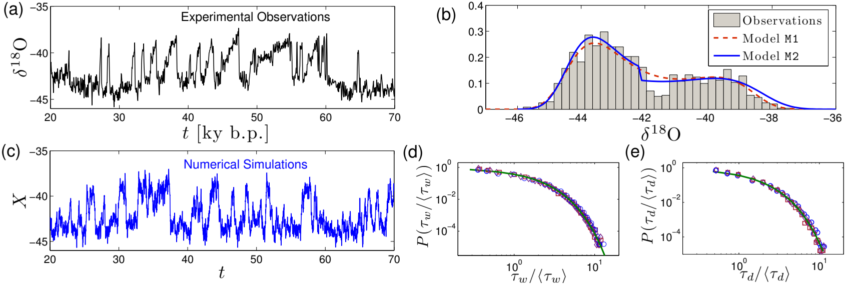
Kac–Zwanzig model:
A distinguished particle moves in a potential $V$ and interacts with $M$ heat bath particles:

$$H = \frac{1}{2}P^2 + V(Q) + \frac{1}{2}\sum_{j=1}^{M}\frac{{p_j}^2}{m_j} + \frac{1}{2}\sum_{j=1}^{M} k_j(q_j - Q)^2$$

# Data-driven coarse-graining: A real-world application

## The paleoclimatic record revisited [K., Pradas, Kalliadasis, Pavliotis. 2015]

- A robust estimation procedure provides more confidence when studying real-world phenomena based on data that may be prone to effects from multiple scales.




- Example of model-based analysis: average time between Dansgaard–Oeschger events:
  - $\tau_{\text{DO}}$ = average time to exit from warm state + average time to exit from cold state : model `M1` $\tau_{\text{DO}} \approx 1.60$ ky and model `M2` $\tau_{\text{DO}} \approx 1.51$ ky
  - Previously reported value in the literature (various physical arguments and/or complex models): $1.5$ ky.

# A general purpose procedure for data-driven coarse-graining

## Take Home Message

- Multiscale effects in data can result in inconsistent (i.e. false) parametric estimators for coarse-grained models,
- Using a robust scheme however, it is possible to obtain simplified low-dimensional models from available data.
- Question: Can you rule out the presence of multiscale effects in your data? If not, then use classic parametric estimators carefully.

## Generalisations and Extensions

- Additional data contamination by noise, e.g. via filtering techniques
- Passage to fully nonparametric setting ($\checkmark$)
- Applications in (computational) molecular dynamics
- Is a Bayesian approach helpful?