

RESEARCH ARTICLE

Open Access

# Estimation of error in observables of coarse-grained models of atomic systems

John Tinsley Oden<sup>\*</sup>, Kathryn Farrell and Danial Faghihi

<sup>\*</sup>Correspondence:  
oden@ices.utexas.edu  
Institute for Computational  
Engineering and Sciences The  
University of Texas at Austin, 201  
East 24th St, Stop C0200 POB 4.102,  
78712 Austin, TX, USA

## Abstract

**Background:** The use of coarse-grained approximations of atomic systems is the most common methods of constructing reduced-order models in computational science. However, the issue of central importance in developing these models is the accuracy with which they approximate key features of the atomistic system. Many methods have been proposed to calibrate coarse-grained models so that they qualitatively mimic the atomic systems, but these are often based on heuristic arguments.

**Methods:** A general framework for deriving a *posteriori* estimates of modeling error in coarse-grained models of key observables in atomistic systems is presented. Such estimates provide a new tool for model validation analysis. The connection of error estimates with relative information entropy of observables and model predictions is explained for so-called misspecified models. The relationship between model plausibilities and Kullback-Leibler divergence between the true parameters and model predictions is summed up in several theorems.

**Results:** Numerical examples are presented in this paper involving a family of coarse-grained models of a polyethylene chain of united atom monomers. Numerical results suggest that the proposed methods of error estimation can be very good indications of the error inherent in coarse-grained models of observables in the atomistic systems. Also, new theorems relating the Kullback-Leibler divergence between model predictions and observations to measures of model plausibility are presented.

**Conclusions:** A formal structure for estimating errors produced by coarse-graining atomistic models is presented. Numerical examples confirm that the estimates are in agreement with exact errors for a simple class of materials. Errors measured in the  $D_{KL}$ -divergence can be related to computable model plausibilities. The results should provide a powerful framework for assessing the validity and accuracy of coarse-grained models.

**Keywords:** Molecular dynamics, Coarse-grained models, Adjoint systems, Information entropy

## Background

### Coarse-grained-reduced order models

The most common method of constructing reduced-order models in all of computational science involves the use of coarse-grained models of atomic systems, whereby systems of atoms are aggregated into “beads”, or “super atoms”, or molecules to reduce the number of degrees of freedom and to lengthen the time scales in which the evolution of events are simulated.

The use of coarse-grained (CG) approximations has been prevalent in molecular dynamics (MD) simulations for many decades. Comprehensive reviews of a large segment of the literature on CG models was recently given by Noid [1] and Li *et al.* [2], and an application to semiconductor nano-manufacturing is discussed in Farrell *et al.* [3]. The issue of central importance in developing CG models is the accuracy with which they approximate key features of the atomistic system. Many methods have been proposed to calibrate CG models so that they qualitatively mimic the all-atom (AA) systems, but these are often based on heuristic arguments.

In this paper, we develop *a posteriori* estimates of error in CG approximations of observables in the AA system. We focus on standard molecular dynamics models of micro-canonical ensemble (NVE) thermodynamics, and we call upon the theory of model adaptivity and error estimation laid down in [4] and [5]. In this particular setting, new estimates are also obtained when the information entropy of Shannon [6] is used as a quantity of interest. This leads to methods for estimating CG-model parameters that involve the Kullback-Leibler divergence between probability densities of observables in the AA and CG systems.

In the final *Results and discussion* section of this presentation, we review several statistical properties of parametric models, including asymptotic properties of misspecified models and generalizations of the Bernstein-von Mises theorem advanced by Kleijn and van der Vaart [7]. There, the fundamental role of the Kullback-Leibler distance (the  $D_{KL}$ ) between the true probability distribution and the observations accessible by the model is reviewed. We present results in the form of theorems that relate the  $D_{KL}$  to measures of model plausibility that arise from Bayesian approaches to model selection. The relationships of the *a posteriori* estimates to the statistical interpretations are summarized in concluding remarks.

### Preliminaries, conventions and notations

We generally approach the problem of developing computer models of large atomic systems through the use of any of several hardened molecular dynamics (MD) codes or through equivalent Monte Carlo approximations invoking the ergodic hypothesis. For a system of  $n$  atoms, the Hamiltonian is

$$H(\mathbf{r}^n, \mathbf{p}^n) = \sum_{\alpha=1}^n \frac{m_{\alpha}}{2} \mathbf{p}_{\alpha} \cdot \mathbf{p}_{\alpha} + u(\mathbf{r}^n), \quad (1)$$

where  $\mathbf{r}^n = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\}$  is the set of atomic coordinate vectors,  $\mathbf{p}^n = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$  is the set of particle momentum vectors,  $m_{\alpha}$  the atomic mass of the  $\alpha$ -th atom and  $u(\mathbf{r}^n)$  is the potential energy or interaction potential. Then  $(\mathbf{r}^n, \mathbf{p}^n)$  defines a point or microstate in the phase space  $\Gamma_{AA}$  of the all atom (AA) model. In typical MD simulations, the potential is, for example, of the form

$$u(\mathbf{r}^n) = V_{\text{bond}}(\mathbf{r}^n) + V_{\text{angle}}(\mathbf{r}^n) + V_{\text{dihedral}}(\mathbf{r}^n) + V_{\text{non-bonded}}(\mathbf{r}^n) + V_{\text{coulomb}}(\mathbf{r}^n), \quad (2)$$

where

$$V_{\text{bond}}(\mathbf{r}^n) = \sum_{i=1}^{N_b} \frac{1}{2} k_{ri} |\mathbf{r}_i - \mathbf{r}_{0i}|^2, \quad (3a)$$

$$V_{\text{angle}}(\mathbf{r}^n) = \sum_{i=1}^{N_a} \frac{1}{2} k_{\theta_i} (\theta_i - \theta_{0i})^2, \quad (3b)$$

$$V_{\text{dihedral}}(\mathbf{r}^n) = \sum_{i=1}^{N_{nb}-1} \sum_{j=1}^{N_{nb}} \frac{V_{ji}}{2} [1 + (-1)^{j-1} \cos(j\phi_i)], \quad (3c)$$

$$V_{\text{non-bonded}}(\mathbf{r}^n) = \sum_{i=1}^{N_{nb}} \sum_{j>i} 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^\alpha - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^\beta \right), \quad r_{ij} \leq r_2, \quad (3d)$$

$$V_{\text{columb}}(\mathbf{r}^n) = \sum_{i=1}^{N_q-1} \sum_{j>i}^{N_q} 4\epsilon_0 \frac{q_i q_j}{r_{ij}}. \quad (3e)$$

Here covalent bonds are represented by the harmonic potential (1a), changes in bond angles by (1b), torsional potentials by changes in dihedral angles (1c), Lennard-Jones non-bonded potentials by (1d), with  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  and  $r_c$  the cut-off radius,  $(\alpha, \beta)$  typically = (12, 6), and Coulomb potentials between charges  $q_i$  at  $\mathbf{r}_i$  and  $q_j$  at  $\mathbf{r}_j$  (1e). These forms are typical of those implemented in popular MD codes, although several other common potentials could be added. The parameters of the potential model are given by the vector of physical coefficients:  $\{k_i, k_{\theta_i}, V_{ji}, \phi_i, \epsilon_{ij}, \sigma_{ij}, r_c, \dots\}$ . In general, atomic properties and values of parameters for the full all-atom system are supplied by systems calibrated using experimental data or quantum mechanics predictions (see, e.g. the OPLS data in [8,9]).

Given the Hamiltonian (1), Hamilton's equations of motion are:

$$\frac{\partial H}{\partial \mathbf{p}_\alpha} = \dot{\mathbf{r}}_\alpha, \quad \frac{\partial H}{\partial \mathbf{r}_\alpha} = -\dot{\mathbf{p}}_\alpha, \quad 1 \leq \alpha \leq n. \quad (4)$$

In MD, it is assumed that the atomic system evolves according to the laws of Newtonian mechanics, so we set  $\mathbf{p}_\alpha = m_{(\alpha)} \dot{\mathbf{r}}_\alpha$ , and the second Hamiltonian equation in (4) reduces to the system of equations

$$m_{\alpha\beta} \ddot{r}_{\beta i}(t) + \partial_{\alpha i} u(\mathbf{r}^n(t)) - f_{\alpha i}(t) = 0, \quad 1 \leq \alpha, \beta \leq n, \quad 1 \leq i \leq 3, \quad (5)$$

where repeated indices are summed throughout their range,  $m_{\alpha\beta} = m_{(\alpha)} \delta_{\alpha\beta}$  is the mass of atom  $\alpha$ , superimposed dots indicate time derivations,  $r_{\beta i}$  is the component of  $\mathbf{r}_\beta$  in direction  $i$ ,  $\partial_{\alpha i} = \partial / \partial r_{\alpha i} u(\mathbf{r}^n(t))$  is the total interatomic potential of the system given, e.g., by (2), and  $f_{\alpha i}(t)$  is the  $i$ th component of applied force on atom  $\alpha$  at time  $t$ . We will add initial conditions,  $\dot{r}_{\beta i}(0) = v_{\beta i}$ , and  $r_{\beta i}(0) = r_{\beta i}^0$ , where  $v_{\beta i}$  and  $r_{\beta i}^0$ , for now, are assumed to be given.

Molecular dynamical equations of the form (5) are typical of those in standard MD codes that are numerically integrated with randomly-sampled initial conditions over time intervals to approximate systems with constant energy and fixed volume and fixed number of particles corresponding to so-called micro-canonical ensembles. Without loss in generality, we confine this development to such thermodynamic scenarios noting that straightforward extensions to, say, constant temperature settings, are covered by replacing (5) with appropriate "thermostat" models, such as the Langevin or Nose-Hoover formulations (see e.g. [10]). The general approach is then applicable to canonical ensembles and more general statistical thermodynamics settings.

A fundamental concept in the molecular theory of matter is that macroscopic properties actually observed in experiments, the so-called *observables*, must be the result of

averages over a time interval  $\tau$  of some phase function  $q(\mathbf{r}^n, \mathbf{p}^n)$  that depends on the phase-point positions  $(\mathbf{r}^n, \mathbf{p}^n)$  in phase space  $\Gamma_{AA}$ , as all measurements require a finite duration. Moreover, for thermodynamic systems in equilibrium, this average, denoted  $\langle q \rangle$ , must be independent of the starting time  $t_0$  and it must attain a value from an essentially infinite time duration. Thus, our goal in constructing the all-atom model is to compute observables of the form (cf [10,11]),

$$\langle q \rangle = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_{t_0}^{t_0+\tau} q(\mathbf{r}^n(t), \mathbf{p}^n(t)) dt. \tag{6}$$

Here we shall confine our attention to phase functions that depend only on the configurations of systems in thermodynamic equilibrium. Our *quantities of interest* are then written,

$$Q_r = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_{t_0}^{t_0+\tau} q(\mathbf{r}^n(t)) dt. \tag{7}$$

In all but the simplest applications, it is impossible to solve the dynamic system (5) owing to its enormous size. Therefore, reduced-order models must be developed. The process involves aggregating groups of atoms into beads or molecules or “super atoms” so as to create a coarse-grained (CG) molecular model. The CG model has  $N$  coordinate vectors  $\mathbf{R}^N(t) = \{\mathbf{R}_1(t), \mathbf{R}_2(t), \dots, \mathbf{R}_N(t)\}$ ,  $N < n$ ; and the corresponding equations of motion are

$$M_{AB} \ddot{R}_{Bi}(t) + \partial_{Ai} U(\mathbf{R}^N(t), \boldsymbol{\theta}) - F_{Ai}(t) = 0, \quad 1 \leq A, B \leq N, \quad 1 \leq i \leq 3. \tag{8}$$

$M_{AB}$  defining the CG mass matrix,  $\partial_{Ai} = \partial/\partial R_{Ai}$ ,  $U(\cdot, \cdot)$  the interaction potential energy of the CG system,  $\boldsymbol{\theta}$  a vector of parameters defining the CG model, and  $F_{Ai}(t)$  the  $i$ th component of applied force at bead  $A$  at time  $t$ . Initial conditions are  $\dot{R}_{Ai}(0) = V_{Ai}$ , and  $R_{Ai}(0) = R_{Ai}^0$ . The unknown parameters with a potential of the form (2) are denoted, for example,

$$\boldsymbol{\theta} = (K_{Ri}, R_{0i}, K_{\theta i}, \theta_{0i}, V_{0i}, \varepsilon_{ii}, \sigma_{ii}, \dots), \tag{9}$$

the notation, in analogy with (1), being chosen to indicate parameters of the CG model.

It is important to establish a kinematic and algebraic relation between coordinates of particles in the AA system and those in the CG system. A very large literature exists on various coarse-graining mapping schemes, and choices of the appropriate map from the AA to the CG system or vice versa are often based on heuristic methods (see e.g. [2]). Our general approach can be adapted to any such well-defined AA-to-CG or CG-to-AA map, but for definiteness, we describe one such family of mappings.

Let  $\mathcal{J}_A$  be the index set of AA-coordinate labels of atoms aggregated into a single bead  $A$  with CG-coordinate vector  $\mathbf{R}_A$  emanating from the origin to a reference point labeled  $A$  within the bead (e.g.  $\mathbf{R}_A$  could be chosen to define the center of mass,  $\mathbf{R}_A = \sum_{\alpha \in \mathcal{J}_A} m_\alpha \mathbf{r}_\alpha / M_A$ ,  $M_A = \sum_{\alpha \in \mathcal{J}_A} m_\alpha$ ). Let  $\mathbf{a}_{A\alpha}(t)$  be a vector from the reference center of bead  $A$  to the end point of AA-coordinate vector  $\mathbf{r}_\alpha(t)$ ,  $\alpha \in \mathcal{J}_A$ . Let  $G_\alpha^A$  be component of the  $n \times N$  array,

$$G_\alpha^A = \begin{cases} 1 & \text{if } \alpha \in \mathcal{J}_A \\ 0 & \text{if otherwise.} \end{cases} \tag{10}$$

Then we have,

$$\mathbf{r}_\alpha = \sum_{A=1}^N G_\alpha^A (\mathbf{R}_A + \mathbf{a}_{A\alpha}), \quad 1 \leq \alpha \leq n. \tag{11}$$

Here we assume that each AA coordinate vector  $\mathbf{r}_\alpha$  belongs to only one bead identified with CG vector  $\mathbf{R}_A$ , but this is not a necessary assumption. In the case of bonded systems in which  $\mathbf{r}_\alpha$  is associated with, say, two index sets  $\mathcal{J}_A$  and  $\mathcal{J}_B$ , we simply choose either  $\mathcal{J}_A$  or  $\mathcal{J}_B$  as the representative of  $\mathbf{r}_\alpha$  and associate  $\mathbf{r}_\alpha$  with only one bead to avoid double counting.

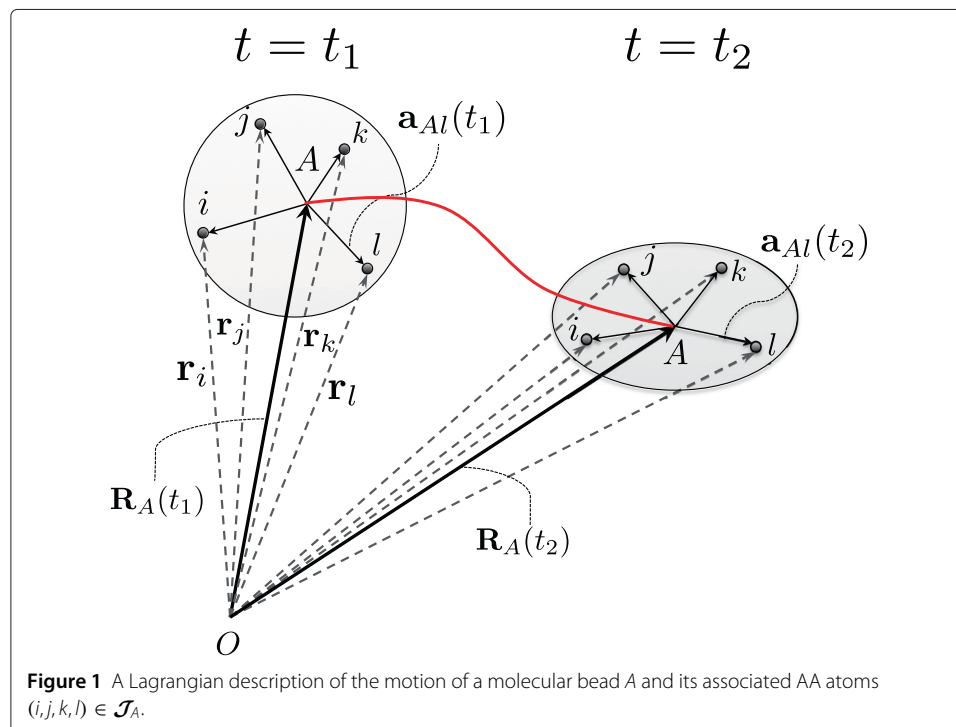
These kinematical conventions are illustrated in Figure 1. The AA coordinates  $\mathbf{r}_\alpha$  of atoms assigned to molecular bead  $A$  remain with that bead throughout all possible motions of the CG system. The Boolean array  $G_\alpha^A$  merely adjusts labels of the AA system to agree with labels assigned beads in the CG system.

Returning to (7), it is clear that the CG approximation of the quantities of interest (QoI) are of the form,

$$Q_{\mathbf{R}}(\boldsymbol{\theta}) = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_{t_0}^{t_0+\tau} q(G(\mathbf{R}^N(t), \boldsymbol{\theta})) dt \tag{12}$$

where  $G$  is the AA-to-CG map defined in (11), where we denote  $\mathbf{r}^n = G(\mathbf{R}^N(t), \boldsymbol{\theta})$ , and where we specifically present the dependence of the QoI on the CG potential parameters  $\boldsymbol{\theta}$ .

Now it is obvious that the evolution of the CG system defined by the coordinate vectors  $\mathbf{R}_A(t)$  satisfying (8) do not satisfy the “true” equations of motion (5). Indeed, if  $\mathbf{R}^N(t) = \{\mathbf{R}_1(t), \mathbf{R}_2(t), \dots, \mathbf{R}_N(t)\}$  is the set of  $N$ -vectors satisfying (8), they constrain the motion



of the AA system via (11), so upon introducing the constrained motions into (5) we obtain the residual,

$$\begin{aligned} \rho_{\alpha i}(\boldsymbol{\theta}, t) &= m_{\alpha\beta} \sum_{A=1}^N G_{\beta}^A (\ddot{\mathbf{R}}_{Ai} + \ddot{a}_{A\beta i})(t) \\ &\quad + \partial_{\alpha i} u(G(\mathbf{R}^N(t), \mathbf{a}^N(t))) - \sum_{A=1}^N G_{\beta}^A \omega_{\alpha}^{\beta} F_{Ai}(t), \\ &1 \leq \alpha, \beta \leq n, \quad 1 \leq i \leq 3, \end{aligned} \tag{13}$$

where  $\sum_{\beta} \omega_{\alpha}^{\beta} = 1$ ,  $\sum_{\alpha \in \mathcal{J}_A} \mathbf{f}_{\alpha} = \mathbf{F}_A$ , and

$$\begin{aligned} G(\mathbf{R}^N(t), \mathbf{a}^N(t)) &= \left\{ \sum_A G_1^A(\mathbf{R}_A + \mathbf{a}_{A1}), G_2^A(\mathbf{R}_A + \mathbf{a}_{A2}) \right. \\ &\quad \left. \dots, G_n^A(\mathbf{R}_A + \mathbf{a}_{An}) \right\}, \\ &= \mathbf{r}^n(t). \end{aligned} \tag{14}$$

On the left-hand side of (13), we have taken note that the residual depends on the CG-model parameters  $\boldsymbol{\theta}$ , not explicitly presented on the right for simplicity. The AA force  $\mathbf{f}_{\alpha}$  at coordinate  $\alpha$  is a fraction  $\omega_{\alpha}^{\beta}$  of the CG force  $F_{Ai}$ . In general, as a first-order approximation, one can take the vectors  $\mathbf{a}_{A\alpha}$  as time-independent constant vectors equal to their value in a given reference configuration ( $\mathbf{a}_{A\alpha}(t) = \mathbf{a}_{A\alpha}(t_0)$ ;  $\dot{\mathbf{a}}_{A\alpha} = \ddot{\mathbf{a}}_{A\alpha} = \mathbf{0}$ ). Then the approximate residual results of the form

$$\hat{\rho}_{\alpha i}(\boldsymbol{\theta}, t) = G_{\alpha}^A m_{\alpha\beta} \ddot{\mathbf{R}}_{Ai} + \partial_{\alpha i} u(G(\mathbf{R}^N(t), \mathbf{a}^N(t_0))) - f_{\alpha i}, \tag{15}$$

with  $f_{\alpha i} = \omega_{\alpha}^{\beta} G_{\beta}^A F_{Ai}$ , and with repeated indices summed,  $1 \leq \alpha, \beta \leq n, 1 \leq A \leq N$ .

It is possible to define a reverse or “push back” relationship that assigns to every AA coordinate vector  $\mathbf{r}_{\alpha}$  the CG coordinate vector  $\mathbf{R}_A$  of the bead to which  $\mathbf{r}_{\alpha}$  belongs. Given  $\mathbf{r}_{\alpha}$ , set

$$\mathbf{R}_A = G_A^{\alpha}(\mathbf{r}_{\alpha} - \mathbf{a}_{A\alpha}), \quad 1 \leq A \leq N, \tag{16}$$

where  $G_A^{\alpha}$  is the transpose of  $G_{\alpha}^A$  (no sum on  $\alpha$ ).

Thus, for each time  $t \in [t_0, \tau]$ , one can select a sample  $\omega(t)$  of AA coordinates  $\{\mathbf{r}_1^*(t), \mathbf{r}_2^*(t), \dots, \mathbf{r}_n^*(t)\}$  and employ (16) to generate the corresponding CG coordinates  $\{\mathbf{R}_1^*(t), \mathbf{R}_2^*(t), \dots, \mathbf{R}_N^*(t)\}$ . We use the simplified notation,

$$\mathbf{R}^N(t) = G(\omega(t)) \tag{17}$$

to define the image of this sample in the CG system.

## Methods

### Weak forms of the dynamical problem

It is convenient to recast the molecular dynamics problem into a weak or variational form. Toward this end, we introduce the Banach spaces

$$\begin{aligned} \mathcal{V} &= \left\{ \mathbf{r}^n(t); \mathbf{r}_{\alpha} \in \Omega \subset \mathbb{R}^3 \times H^2(0, \tau); 1 \leq \alpha \leq n; \right. \\ \left. \|\mathbf{r}^n(\cdot)\|^2 &= \int_0^{\tau} [\ddot{\mathbf{r}}_{\alpha} \cdot \ddot{\mathbf{r}}_{\alpha} + \dot{\mathbf{r}}_{\alpha} \cdot \dot{\mathbf{r}}_{\alpha} + \mathbf{r}_{\alpha} \cdot \mathbf{r}_{\alpha}] d\tau < \infty \right\}, \end{aligned} \tag{18}$$

and the semilinear and linear forms  $\mathcal{B} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}, \mathcal{F} : \mathcal{V} \rightarrow \mathbb{R}$ , given by

$$\mathcal{B}(\mathbf{r}^n; \mathbf{v}^n) = \int_0^\tau [m_{\alpha\beta} \ddot{\mathbf{r}}_\alpha(t) \cdot \mathbf{v}_\beta(t) + \partial_{\alpha i} u(\mathbf{r}^n(t)) v_{\alpha i}(t)] dt + I_0(\mathbf{v}^n), \tag{19}$$

$$\mathcal{F}(\mathbf{v}^n) = \int_0^\tau f_{\alpha i}(t) v_{\alpha i}(t) dt + v_{\alpha i}(0) m_{\alpha\beta} V_\beta - \dot{v}_{\alpha i}(0) m_{\alpha\beta} r_{\beta}^0, \tag{20}$$

where

$$I_0(\mathbf{v}^n) = v_{\alpha i}(0) m_{\alpha\beta} \dot{r}_\beta(0) - \dot{v}_{\alpha i}(0) m_{\alpha\beta} r_\beta(0). \tag{21}$$

The notation  $\mathcal{B}(\cdot; \cdot)$  is intended to mean that  $\mathcal{B}(\cdot; \cdot)$  is possibly nonlinear in entries to the left of the semi-colon and linear in the entries to the right of it.

The problem of finding  $\mathbf{r}^n \in \mathcal{V}$  such that

$$\mathcal{B}(\mathbf{r}^n; \mathbf{v}^n) = \mathcal{F}(\mathbf{v}^n) \quad \forall \mathbf{v}^n \in \mathcal{V}, \tag{22}$$

is equivalent to (8) in the sense that every solution of (8) with appropriate initial conditions, satisfies (22), and any sufficiently smooth solution of (22) satisfies (8).

**The adjoint problem**

Let

$$\mathcal{B}'(\mathbf{r}^n; \mathbf{z}^n, \mathbf{v}^n) = \lim_{\theta \rightarrow 0} \theta^{-1} [\mathcal{B}(\mathbf{r}^n + \theta \mathbf{z}^n; \mathbf{v}^n) - \mathcal{B}(\mathbf{r}^n; \mathbf{v}^n)] \tag{23}$$

and

$$Q'(\mathbf{r}^n; \mathbf{v}^n) = \lim_{\theta \rightarrow 0} \theta^{-1} [Q(\mathbf{r}^n + \theta \mathbf{v}^n) - Q(\mathbf{r}^n)], \tag{24}$$

where  $Q$  is a functional on  $\mathcal{V}$ , and both  $\mathcal{B}'(\cdot; \cdot)$  and  $Q'(\cdot; \cdot)$  are assumed to exist and be finite (i.e.  $\mathcal{B}(\cdot; \cdot)$  and  $Q(\cdot)$  are Gateaux differentiable). Then the adjoint or dual problem associated with (22) is

Find  $\mathbf{z}^n = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\} \in \mathcal{V}$  such that

$$\mathcal{B}'(\mathbf{r}^n; \mathbf{z}^n, \mathbf{v}^n) = Q'(\mathbf{r}^n; \mathbf{v}^n) \quad \forall \mathbf{v}^n \in \mathcal{V}. \tag{25}$$

Introducing (19) into (23) gives, after some manipulations,

$$\begin{aligned} \mathcal{B}'(\mathbf{r}^n; \mathbf{z}^n, \mathbf{v}^n) &= \int_0^\tau (m_{\alpha\beta} \ddot{z}_{\beta i} - H_{\alpha i \beta j}(\mathbf{r}^n(t)) z_{\beta j}) v_{\alpha i} dt \\ &\quad + \dot{m}_{\alpha\beta} z_{\beta i}(\tau) v_{\alpha i}(\tau) \\ &\quad - m_{\alpha\beta} \dot{z}_{\beta i}(\tau) v_{\alpha i}(\tau), \end{aligned} \tag{26}$$

where  $H_{\alpha i \beta j}$  is the Hessian,

$$H_{\alpha i \beta j}(\mathbf{r}^n(t)) = \frac{\partial^2 u(\mathbf{r}^n(t))}{\partial r_{\alpha i} \partial r_{\beta j}}. \quad 1 \leq \alpha, \beta \leq n, 1 \leq i, j \leq 3. \tag{27}$$

Likewise, if  $Q(\mathbf{r}^n) = \int_0^\tau q(\mathbf{r}^n(t)) dt$ , then

$$Q'(\mathbf{r}^n(t); \mathbf{v}^n) = \int_0^\tau \partial_{\alpha i} q(\mathbf{r}^n(t)) v_{\alpha i} dt. \tag{28}$$

Note that (25) is solved “backward in time;” the forward problem (22) is solved for  $\mathbf{r}^n(t)$ , which determines the coefficients in  $\mathcal{B}'(\mathbf{r}^n; \mathbf{z}^n, \mathbf{v}^n)$  which marches the adjoint solution from  $t = \tau$  to  $t = 0$ . The dynamical system corresponding to (25) is:

$$m_{\alpha\beta} \ddot{z}_{\beta i} + H_{\alpha i \beta j}(\mathbf{r}^n(t)) z_{\beta j}(t) = \partial_{\alpha i} q(\mathbf{r}^n(t)). \tag{29}$$

**Theory of a posteriori estimation of modeling error**

Let us now review the theory of *a posteriori* estimation of modeling error advanced in [4] and expanded in [5]. We consider an abstract variational problem of finding an element  $u$  in a topological vector space  $\mathcal{V}$  such that,

$$\mathcal{B}(u; v) = \mathcal{F}(v), \quad \forall v \in \mathcal{V}, \tag{30}$$

where  $\mathcal{B}(\cdot; \cdot)$  is a semilinear form from  $\mathcal{V} \times \mathcal{V}$  into  $\mathbb{R}$  and  $\mathcal{F}$  is a linear functional on  $\mathcal{V}$ . Problem (30) is equivalent to the problem of finding a solution  $u$  of the problem  $A(u) = F$  in the dual space  $\mathcal{V}'$ , where  $A$  is the map induced by  $\mathcal{B}(\cdot; \cdot) : \langle A(u), v \rangle = \mathcal{B}(u; v) = \mathcal{F}(v) = \langle \mathcal{F}, v \rangle$ ,  $\langle \cdot; \cdot \rangle$  denoting duality pairing in  $\mathcal{V}' \times \mathcal{V}$ . Assuming (30) is solvable for  $u$ , we wish to compute the value  $Q(u)$  of a functional  $Q : \mathcal{V} \rightarrow \mathbb{R}$  representing a quantity of interest, or an observable of interest.

We assume that the semilinear form  $\mathcal{B}(\cdot; \cdot)$  and the functional  $Q(\cdot; \cdot)$  are three times Gateaux differentiable on  $\mathcal{V}$  with respect to  $u$ . In particular, the following limits exist (recall (23) and (24)),

$$\left. \begin{aligned} \mathcal{B}'(u; w, v) &= \lim_{\theta \rightarrow 0} \theta^{-1} [\mathcal{B}(u + \theta w, v) - \mathcal{B}(u, v)] \\ Q'(u; v) &= \lim_{\theta \rightarrow 0} \theta^{-1} [Q(u + \theta v) - Q(u)] \end{aligned} \right\} \tag{31}$$

with similar definitions of higher-order derivatives, e.g.  $\mathcal{B}''(u; w_1, w_2, v)$ ,  $\mathcal{B}'''(u; w_1, w_2, w_3, v)$ ,  $Q''(u; w, v)$ ,  $Q''(u; v_1, v_2, v_3)$ , etc. See [5] for details.

The adjoint problem associated with (30) and the quantity of interest  $Q$  consists of finding  $z \in \mathcal{V}$  such that

$$\mathcal{B}'(u; z, v) = Q'(u; v), \quad \forall v \in \mathcal{V}. \tag{32}$$

Now let  $u_0$  be an arbitrary element selected in  $\mathcal{V}$ . The residual functional (or “residuum”) associated with  $u_0$  is defined as the semilinear functional  $\mathcal{R} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ ,

$$\mathcal{R}(u_0; v) = \mathcal{F}(v) - \mathcal{B}(u_0; v), \tag{33}$$

which, for each  $u_0 \in \mathcal{V}$ , is a linear functional on  $\mathcal{V}$ .

Obviously, if  $u_0 = u$ , the solution of (30),  $\mathcal{R}(u; v) = 0 \quad \forall v \in \mathcal{V}$ . Thus,  $\mathcal{R}(u_0; v)$  describes the degree to which the vector  $u_0$  fails to satisfy the central problem (30).

We now recall the basic theorem in [5]:

**Theorem 1.** *Let the semilinear form  $\mathcal{B}(\cdot; \cdot)$  in (30) and the quantity of interest  $Q$  be three-times continuously Gateaux differentiable on  $\mathcal{V}$ . Let  $u_0$  be an arbitrary element of  $\mathcal{V}$ . Then the error in  $Q(u)$  produced by replacing  $u$  by  $u_0$  is given by:*

$$Q(u) - Q(u_0) = \mathcal{R}(u_0; z) + \Delta \tag{34}$$

where  $\Delta$  is a remainder involving higher-order terms in  $e_0 = u - u_0$  and  $\varepsilon_0 = z - z_0$ ,  $z_0$  being an approximation of  $z$ .

An explicit form of  $\Delta$  is given in the appendix.

If  $u_0$  is not an arbitrary vector taken from  $\mathcal{V}$  but is a solution of a surrogate problem approximating (30) (such as a coarse-grained model approximating an AA model), then it often happens that  $\Delta$  is negligible compared to the residual. Then (34) reduces to the approximation,

$$Q(u) - Q(u_0) \approx \mathcal{R}(u_0; z). \tag{35}$$



This relation is the basis for many successful methods of *a posteriori* error estimation of both modeling error and numerical error. Whenever  $\mathcal{B}(\cdot; \cdot)$  is a bilinear form and  $Q(\cdot)$  is linear,  $\Delta \equiv 0$ .

### A–posteriori estimation of error in CG approximations

The CG approximations of the “ground truth” AA system are characterized by a parametric class  $\mathcal{P}(\theta)$  of molecular dynamics models, one model corresponding to each choice of the vector  $\theta$  in a space  $\Theta$  of parameters defining the CG intermolecular potential  $U(\mathbf{R}^N(t), \theta)$ . For a given value of  $\theta$ , observables of interest in states of thermodynamic equilibrium of the CG system are typically generated as averages of samples of the observables taken over subintervals  $[t_k, t_{k+1}] \subset [0, \tau]$ , for a distribution of initial conditions (see, e.g. [10]).

If we employ the approximation (35) to the AA and CG models, then an estimate of the error in CG approximations of the observable is immediate. Let  $q(\mathbf{r}^n)$  be a phase function whose ensemble average  $\langle q \rangle$  is an observable of interest, denoted  $Q_{\mathbf{r}}$  as in (7). The CG approximation is  $Q_{\mathbf{R}}(\theta)$  and the error, given by (35), is

$$\varepsilon(\theta) = Q_{\mathbf{r}} - Q_{\mathbf{R}}(\theta) \approx \mathcal{R}(\mathbf{R}^N(\theta); \mathbf{z}^n), \tag{36}$$

where  $\mathcal{R}(\mathbf{R}^N(\theta); \mathbf{z}^n) = \int_0^\tau \rho_{\alpha i}(\theta, t) z_{\alpha i}(t) dt$ ,  $\rho_{\alpha i}(\theta, t)$  is the residual in (13) (or (15)), and  $\mathbf{z}^n$  is the solution to the corresponding adjoint problem, and is generally unknown. If  $\mathbf{Z}^n$  is an approximation of  $\mathbf{z}^n$ , then

$$|\varepsilon(\theta)| \leq |\mathcal{R}(\mathbf{R}^N(\theta); \mathbf{Z}^n)| + C(\theta) \|\mathbf{z}^n - \mathbf{Z}^n\| \tag{37}$$

with

$$C(\theta) = \sup_{\mathbf{v}^n \in \mathcal{V}} \frac{\|\mathcal{R}(\mathbf{R}^N(\theta, t))\|_{\mathcal{V}'}}{\|\mathbf{v}^n\|_{\mathcal{V}}} \tag{38}$$

$\|\cdot\|_{\mathcal{V}'}$  being the norm on the dual space  $\mathcal{V}'$ . The problem of error estimation thus reduces to one of developing efficient procedures to compute the residual ( $\rho$ ) and to compute reasonable approximations of  $\mathbf{z}^n$ .

It is clear that a quantitative estimate such as (36) (or an approximation with  $\mathbf{z}^n$  replaced by  $\mathbf{Z}^n$ ) could be a powerful tool for determining validity of the CG model or in designing validation experiments for CG models. In theory, it also provides a basis for selecting optimal parameters for a given model so as to manage  $\varepsilon(\theta)$ . We elaborate on this notion in the final part of the *Results and discussion* section.

## Results and discussion

### Numerical example: estimation of error in CG-approximation of a polyethylene chain

We describe in this section an application of a posterior error estimation described in the previous section, involving CG approximations of a well-known model of polyethylene. For the base “AA” model, we consider a united atom model of a polyethylene chain containing 200 CH<sub>2</sub> (methyl) monomers, meaning we have aggregated hydrogen and carbon atoms into an “AA” bead for simplicity. The united-atom coordinates  $r_i$  define the locations of each particle on an  $r$ -axis, and the displacement is denoted  $u_i(t)$ . As an additional simplification, we assume that the interatomic potential is characterized by harmonic bonds of the form (1a), with parameter  $k_l = k = 350 \text{ kCal/mol}$ , bond length of  $l = 1.5 \text{ \AA}$ , and atomic mass  $m = 14.026 \text{ gr/mol}$ . Initially, each united atom is separated by bond

length  $l$ , the initial velocities are zero, and the system is assigned an initial displacement field  $u_i(0) = f(r_i)$ , where  $f(r_i) = 1.2e^{-0.1r_i(0)}$ . Under these conditions, the AA system (5) reduces to

$$m\ddot{\mathbf{u}} - \mathbf{k}\mathbf{u} = \mathbf{0}, \quad \mathbf{u}(0) = \mathbf{u}_0, \quad \dot{\mathbf{u}}(0) = \mathbf{v}_0, \tag{39}$$

where  $\mathbf{m}$  and  $\mathbf{k}$  are the mass matrix and the stiffness matrix of the united atom system and are of the form

$$\mathbf{m} = m \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \quad \mathbf{k} = k \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & \ddots & & \\ & \ddots & \ddots & \ddots & -1 \\ & & \ddots & -1 & 2 \end{bmatrix}. \tag{40}$$

A family  $\mathcal{M}$  of CG approximations of this model is obtained by aggregating the atoms into beads, with CG models in  $\mathcal{M}$  distinguished by the number  $P$  of atoms per CG bead. The resulting CG system (8) is of the form

$$M\ddot{\mathbf{U}} - \mathbf{K}\mathbf{U} = \mathbf{0}, \quad \mathbf{U}(0) = \mathbf{U}_0, \quad \dot{\mathbf{U}}(0) = \mathbf{V}_0, \tag{41}$$

with the mass of each bead set to  $M = Pm$  and the bond stiffness  $K = \alpha k/P; \alpha \in \mathbb{R}^+$ .

Upon solving (41) for the CG displacement trajectory  $\mathbf{U}(t)$ , we compute the residual trajectory

$$\boldsymbol{\rho} = m\ddot{\mathbf{U}}_{\text{CG}} - \mathbf{k}\mathbf{U}_{\text{CG}}, \tag{42}$$

where  $\mathbf{U}_{\text{CG}}(t)$  is the projection  $\Pi\mathbf{U}(t)$  onto AA atom locations.

The bilinear and linear forms described in (19)-(21) reduce, in this case, to

$$\mathcal{B}(\mathbf{u}; \mathbf{v}) = \int_0^\tau \mathbf{v}^T (\mathbf{m}\ddot{\mathbf{u}} - \mathbf{k}\mathbf{u}) dt - \mathbf{v}^T(0)\mathbf{m}\dot{\mathbf{u}}(0) - \dot{\mathbf{v}}^T(0)\mathbf{m}\mathbf{u}(0), \tag{43}$$

$$\mathcal{F}(\mathbf{v}) = -\mathbf{v}^T(0)\mathbf{m}\mathbf{v}_0 - \dot{\mathbf{v}}^T(0)\mathbf{m}\mathbf{u}_0. \tag{44}$$

and (25) yielding

$$\begin{aligned} \mathcal{B}'(\mathbf{u}; \mathbf{v}, \mathbf{z}) &= \int_0^\tau (\mathbf{m}\dot{\mathbf{z}} - \mathbf{k}\mathbf{z})^T \mathbf{v} dt \\ &\quad + (\mathbf{m}\mathbf{z}(\tau))^T \dot{\mathbf{v}}(\tau) - (\mathbf{m}\dot{\mathbf{z}}(\tau))^T \mathbf{v}(\tau) \\ &= \mathcal{Q}'(\mathbf{v}). \end{aligned} \tag{45}$$

As an example of a QoI, we take  $Q_r$  to be the locally-averaged displacement,

$$Q_r = \int_0^\tau \zeta(t) dt; \quad \zeta(t) = \frac{1}{N_0} \sum_{i \in \mathcal{N}} u_i(t); \quad \mathcal{N} = \{i : x_i \leq \beta l; \beta \in \mathbb{R}^+\}, \quad N_0 = \text{card } \mathcal{N}, \tag{46}$$

for which the strong form of the dual problem is

$$m\ddot{\mathbf{z}} - \mathbf{k}\mathbf{z} = \mathbf{q}, \quad \mathbf{m}\dot{\mathbf{z}}(\tau) = \mathbf{0}, \quad \mathbf{z}(\tau) = \mathbf{0}, \tag{47}$$

where  $N_0$  is the number of united atoms considered in set  $\mathcal{N}$  and  $\mathbf{q}$  is the vector defined such that  $Q(\mathbf{u}) = \mathbf{q}^T \mathbf{u}$ . Given the QoI (46),  $\mathbf{q}$  will be as a  $n \times 1$  vector,

$$q_i = \begin{cases} 1 & \text{if } x_i \leq \beta l \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, n. \tag{48}$$

The residual function  $\mathcal{R}(\cdot, \cdot)$  of (36) in this example is of the form

$$\mathcal{R}(\mathbf{U}_{CG}, \mathbf{z}^n) = \int_0^\tau \eta_t(t) dt; \quad \eta_t(t) = \sum_{i=1}^n \mathbf{z}_i(t) \cdot \rho_i(t) dt. \tag{49}$$

The estimated error in the QoI is then

$$\mathcal{E}_{\text{est.}} = \mathcal{R}(\mathbf{U}_{CG}, \mathbf{z}^n) \approx Q_{\mathbf{r}} - Q_{\mathbf{R}}, \tag{50}$$

where  $Q_{\mathbf{R}} = \int_0^\tau \eta_t(t) dt$  and then the exact error is

$$\mathcal{E}_{\text{exact}} = \mathcal{E}_{\text{est.}} + \Delta, \tag{51}$$

$\Delta$  being the remainder in (34). Since the forms in (43)-(46) are linear in their respective arguments, the exact remainder  $\Delta$  should be zero, but the error introduced by the numerical integration schemes employed generally leads to an additional numerical error  $\Delta_{\Delta t} \neq 0$ . We employ a converted Runge-Kutta algorithm here to integrate (39), (41), and (47).

The results of the coarse-grained model for the case of  $P = 4$  are presented in Figure 2. Figure 2a shows the coarse-scale displacement  $\mathbf{U} = \mathbf{U}(t)$  at different times, obtained from solution of (41) over the time domain  $t \in [0, \tau]$ . The local residual of Figure 2b is then computed from (42). Figure 2c shows the solution of  $\mathbf{z}(t)$  at different times. It is observed that the adjoint solution propagates in time in the opposite direction to the primal solution, (47) being integrated backward in time.

It is known that in general, the solution of the base model is not available. However, in order to show the effectiveness of the method presented here, the equations of motion for the united atom system is also solved in this example. Having the solution of the united atom model,  $\mathbf{u}(t)$ , the evolution of the exact  $\zeta$  and estimated  $\eta_t$  over time is shown in Figure 2d.

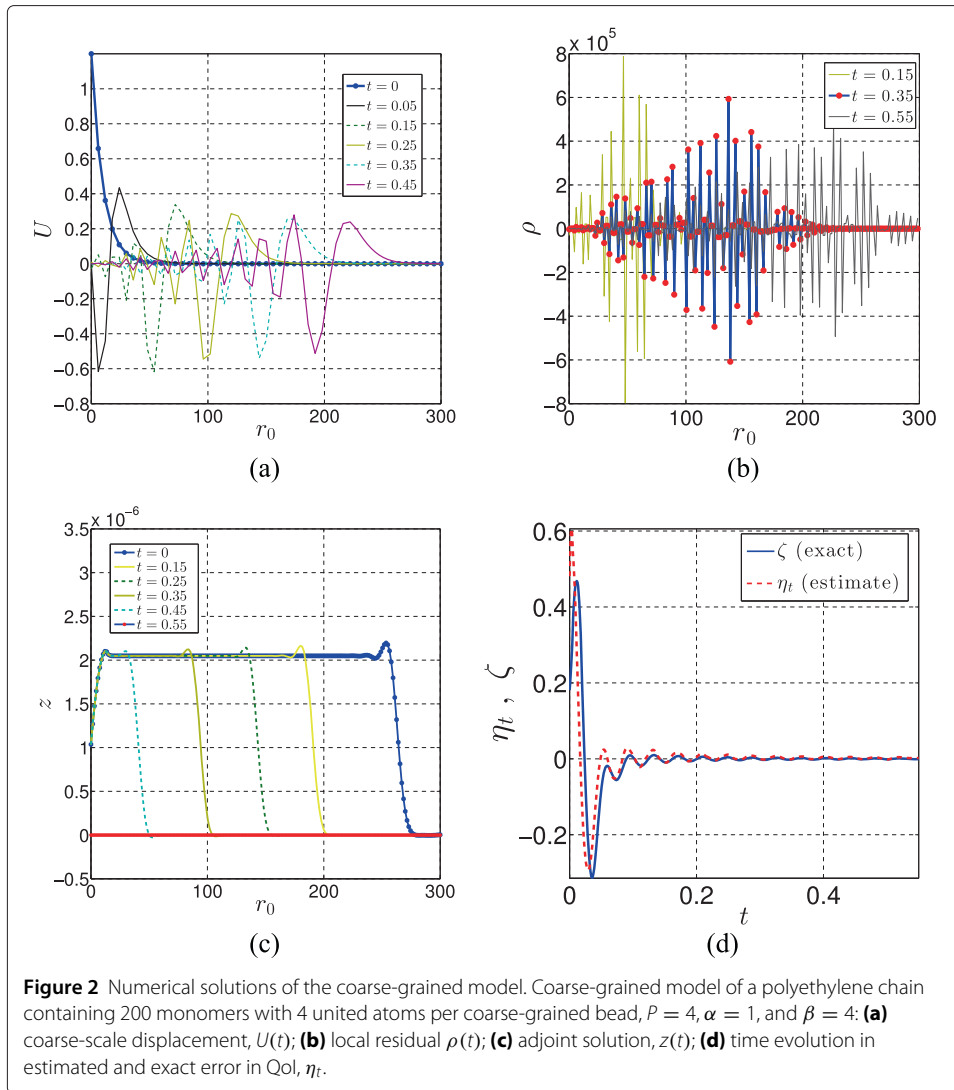
Numerical approximations to the exact error are compared with the estimated error for various CG approximation of the united atom model in Figure 3a. The estimated error  $\mathcal{R} = \mathcal{R}(\mathbf{U}_{CG}(\boldsymbol{\theta}); \mathbf{z}^n)$  for various values of  $P$ , with  $\theta = \alpha k/P$ , are indicated in Figure 3b for  $\alpha = 1$ . The computed estimated error ( $\mathcal{E}_{\text{est}} = \mathcal{R}$ ) versus the parameter  $\alpha$  are indicated in Figure 3c.

In general, the solution of the base model is not available, but the effectiveness of the method presented here is determined by comparing the CG solutions with the exact united atom model. The exact  $\zeta$  and estimated  $\eta_t$  over time are shown in Figure (2d).

**Maximum entropy principle for atomic systems**

Among features of the AA system that could qualify as quantities of interest, we consider a special measure of uncertainty content embodied in the so-called information entropy. In 1948, Shannon [6] introduced the concept of information entropy as a real-valued function  $H(p)$  of probability distributions (densities)  $p$  as a logical measure of uncertainty content in  $p$  that satisfied four rather straight forward “common-sense” desiderata (see also [12] for full details). For a discrete pdf  $p = \{p_1, p_2, \dots, p_n\}$ , the entropy is defined by

$$H(p) = - \sum_{i=1}^n p_i \log p_i, \tag{52}$$



and for a continuous density,  $p \in L^2(\mathbb{R})$ , we write

$$H(p) = - \int_{\mathbb{R}} p(y) \log p(y) dy. \tag{53}$$

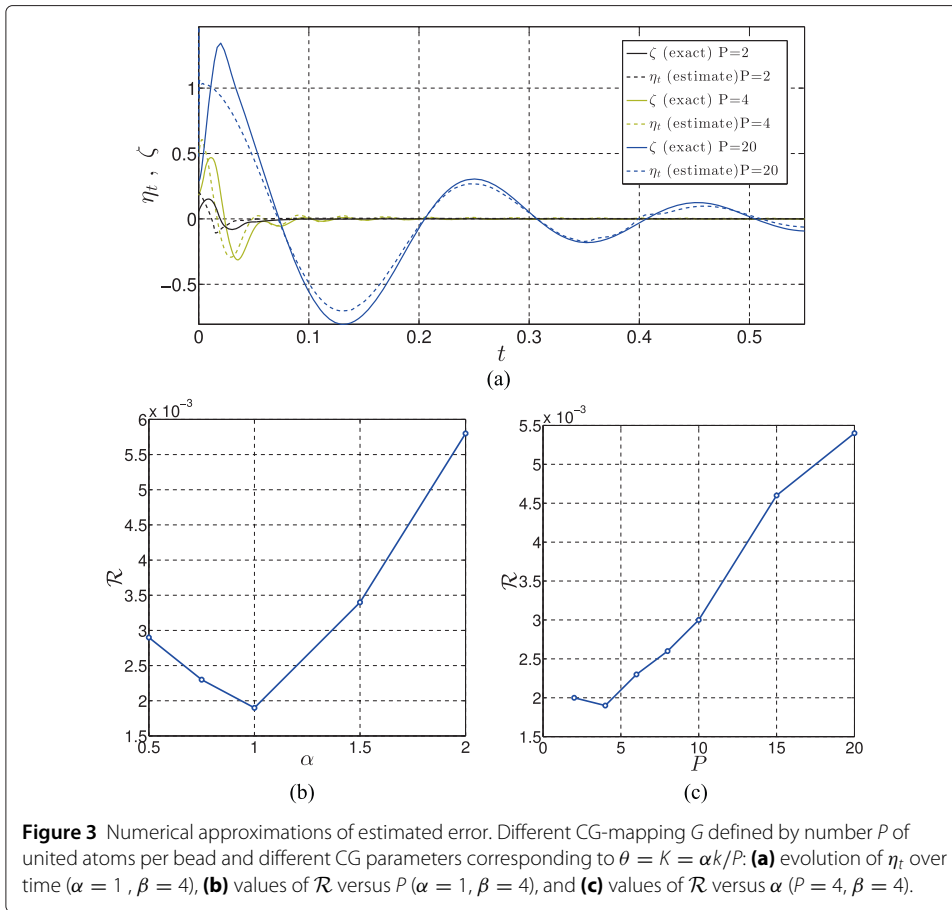
Given two probability densities  $p$  and  $q$ , with non-empty support of domains, the relative entropy between  $p$  and  $q$  is given by the Kullback-Leibler divergence,

$$\begin{aligned} D_{KL}(p||q) &= \int_{\mathbb{R}} p(y) \log \frac{p(y)}{q(y)} dy \\ &= H(p, q) - H(p), \end{aligned} \tag{54}$$

where  $H(p, q) (= - \int_{\mathbb{R}} p \log q dy)$  is the cross entropy and it is understood that  $0 \log \frac{0}{0} = 0$  and  $0 \log \frac{0}{q} = 0$ .

Shannon's principle of maximum entropy asserts that in the set  $\mathcal{P}$  of all possible probability distributions relevant to a random field, the correct probability  $p$  corresponds to the maximum entropy:

$$H(p) = \max_{q \in \mathcal{P}} H(q). \tag{55}$$



### Errors in information entropy

The connection with the statistical mechanics characterization of the AA and CG models can be established by choosing as a quantity of interest the infinite-time average of the phase function  $q(\mathbf{r}^n(t))$  over  $[0, \infty]$ . For this we invoke the ergodic hypothesis,

$$Q_{\mathbf{r}} = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_{t_0}^{t_0+\tau} q(\mathbf{r}(t)) dt \tag{56}$$

$$= \int_{\Gamma} \rho(\mathbf{r}^n) q(\mathbf{r}^n) d\mathbf{r}^n \tag{57}$$

$$= \langle q \rangle \tag{58}$$

$\rho(\mathbf{r}^n)$  being the distribution function for the ensemble under study and  $\Gamma$  the corresponding phase space subdomain. The corresponding CG approximation is

$$Q_{\mathbf{R}}(\theta) = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_{t_0}^{t_0+\tau} q(G(\mathbf{R}^N(t); \theta)) dt \tag{59}$$

$$= \int_{\Gamma} \rho(\mathbf{r}^n) q(G(\mathbf{R}^N; \theta)) d\mathbf{r}^n \tag{60}$$

where the notation  $G(\mathbf{R}^N; \theta)$  represents the relation (11). Setting

$$q(\mathbf{r}^n) = \log \rho(\mathbf{r}^n), \tag{61}$$

gives immediately

$$Q_{\mathbf{r}} - Q_{\mathbf{R}}(\boldsymbol{\theta}) = D_{KL}(\rho(\mathbf{r}^n) \parallel \rho(G(\mathbf{R}^N; \boldsymbol{\theta}))), \tag{62}$$

where  $D_{KL}(\cdot \parallel \cdot)$  is the Kullback-Leibler divergence defined in (54). Thus, if  $\mathbf{z}^n$  is the equilibrium solution of (25) with  $Q'(\mathbf{r}^n; \mathbf{v}^n) = \int_0^\tau \partial_{\alpha_i} q(\mathbf{r}^n(t)) \nu_{\alpha_i} dt$ , then

$$D_{KL}(\rho(\mathbf{r}^n) \parallel \rho(G(\mathbf{R}^N; \boldsymbol{\theta}))) \cong \mathcal{R}(\mathbf{R}^N(\boldsymbol{\theta}); \mathbf{z}^n). \tag{63}$$

The specification (60) of the CG approximation (with  $\rho(\mathbf{r}^n)$  as opposed to  $\rho(\mathbf{R}^N(\boldsymbol{\theta}))$ ) requires some explanation. In interpreting (60), one assumes the role of an observer who resides in the AA system and, instead of the true phase function  $q(\mathbf{r}^n)$ , observes a corrupted version for each choice of  $\boldsymbol{\theta}$  constrained to reside only in microstates accessible by the CG-model. This is also the interpretation of the residual described in (13) and (15). It is also noted that the estimate (63) is reminiscent of the minimum relative entropy method suggested by Shell [13].

A fundamental question arises at this point: given estimates (36) or (63), is it possible to find a special parameter vector  $\boldsymbol{\theta}^*$  that makes the error  $\epsilon(\boldsymbol{\theta}^*) = 0$ ? This question is related to the so-called well-specification or misspecification of the CG model. We believe the answer to this question is generally “no.”

**Model misspecification and statistical analysis**

A fundamental concept in the mathematical statistics literature on parametric models is the notion of a well-specified model, one that has the property that a special parameter vector  $\boldsymbol{\theta}^*$  exists that the model  $\mathcal{P}$  maps into the *truth*; i.e. the true observational data. If no such parameter exists, the model is said to be *misspecified*.

More generally, we consider a space  $\mathcal{Y}$  of physical observables (in our case, the values of appropriate observables sampled from the AA model) and a set  $\mathbb{M}(\mathcal{Y})$  of probability measures  $\mu$  on  $\mathcal{Y}$ . As always, a target quantity of interest  $Q : \mathbb{M} \rightarrow \mathbb{R}$  is identified (e.g.  $Q(\mu) = \mu[X \geq a]$ ,  $X$  being a random variable and  $a$  a threshold value). We seek a particular measure  $\mu^*$  which yields the “true” value of the quantity of interest  $Q(\mu^*)$ . We wish to predict  $Q(\mu^*)$  using a parametric model  $\mathcal{P} : \Theta \rightarrow \mathbb{M}(\mathcal{Y})$ ,  $\Theta$  being the space of parameters. Again, if a  $\boldsymbol{\theta}^* \in \Theta$  exists such that  $\mathcal{P}(\boldsymbol{\theta}^*) = \mu^*$ , the model is said to be well-specified; otherwise, if  $\mu^* \notin \mathcal{P}(\Theta)$ , the model is misspecified. See, e.g., Geyer [14], Kleijn and van der Vaart [7], Freedman [15], Nickl [16]. In our model discussed in Section ‘Preliminaries, conventions and notations’, we seek a parameter  $\boldsymbol{\theta}^*$  of the CG model such that  $\epsilon(\boldsymbol{\theta}^*)$  of (36) is zero, an unlikely possibility for most choices of  $Q$ .

To recast the issue of error estimation into a statistical setting, we presume that our goal is to determine (predict) a probability distribution of a random variable, an observable  $q$  in the AA system, using a CG model  $\mathcal{P}$ , given a set  $y_1, y_2, \dots, y_n$  of *iid* (independent, identically-distributed) random variables representing samples  $y_i = q(\omega_i)$  ( $\omega_i = \mathbf{r}_i^n$  is meant to denote a particular point in phase space). We denote by  $\pi(y_i|\boldsymbol{\theta})$  the conditional probability density  $p$  of the distance between the random data  $y_i$  and the parameter-to-observation map  $d_i(\boldsymbol{\theta})$ ,

$$p(y_i - d_i(\boldsymbol{\theta})) = \pi(y_i|\boldsymbol{\theta}); \quad i = 1, 2, \dots, n, \tag{64}$$

where  $\pi(y_i|\boldsymbol{\theta})$  is the  $i$ th component of the likelihood function. The joint density of the data vector  $\mathbf{y}^n = y_1, y_2, \dots, y_n$  is then,

$$\pi_n(y_1, y_2, \dots, y_n|\boldsymbol{\theta}) = \pi(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n \pi(y_i, \boldsymbol{\theta}). \tag{65}$$

The log-likelihood function is

$$L_n(\boldsymbol{\theta}) = \log \pi(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^n \log \pi(y_i|\boldsymbol{\theta}). \tag{66}$$

Let  $\pi(\boldsymbol{\theta})$  be any prior probability density on the parameters  $\boldsymbol{\theta}$  (computed, for instance, using the maximum entropy method of Jaynes [12], as described for CG models in [3]); then the posterior density satisfies,

$$\pi_n(\boldsymbol{\theta}|\mathbf{y}) = \pi(y_1, y_2, \dots, y_n|\boldsymbol{\theta})\pi(\boldsymbol{\theta})/Z(\boldsymbol{\theta}), \tag{67}$$

where  $Z(\boldsymbol{\theta}) = \int_{\Theta} \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$  is the model evidence.

The following definitions and theorems follow from these relations:

- The Maximum Likelihood Estimate (MLE) is the parameter  $\hat{\boldsymbol{\theta}}^n$  that maximizes  $L_n(\boldsymbol{\theta})$ :

$$\hat{\boldsymbol{\theta}}^n = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}). \tag{68}$$

- The Maximum A Posterior Estimate (MAP) is the parameter  $\tilde{\boldsymbol{\theta}}^n$  that maximizes the posterior pdf:

$$\tilde{\boldsymbol{\theta}}^n = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \pi_n(\boldsymbol{\theta}|\mathbf{y}). \tag{69}$$

- The Bayesian Central Limit Theorem for well-specified models under commonly satisfied smoothness assumptions (also called the Bernstein-von Mises Theorem [7,16,17]) asserts that

$$\pi_n(\boldsymbol{\theta}|\mathbf{y}) \xrightarrow{\mathcal{P}} \mathcal{N}(\boldsymbol{\theta}^*; \mathbf{I}^{-1}(\boldsymbol{\theta}^*)), \tag{70}$$

where convergence is convergence in probability,  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes a normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ ,  $\hat{\boldsymbol{\theta}}$  is the generalized MLE, and  $\mathbf{I}(\boldsymbol{\theta})$  is the Fisher information matrix,

$$I_{ij}(\boldsymbol{\theta}) = - \sum_{k=1}^n \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \pi(y_k|\boldsymbol{\theta}) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \tag{71}$$

- Given a set of parametric models,  $\mathcal{M} = \{\mathcal{P}_1(\boldsymbol{\theta}_1), \mathcal{P}_2(\boldsymbol{\theta}_2), \dots, \mathcal{P}_m(\boldsymbol{\theta}_m)\}$ , the posterior plausibility of model  $j$  is defined through the applications of Bayesian arguments by (see [3,18])

$$\rho_j = \pi(\mathcal{P}_j|\mathbf{y}, \mathcal{M}) = \frac{\int_{\Theta_j} \pi(\mathbf{y}|\boldsymbol{\theta}_j, \mathcal{P}_j, \mathcal{M})\pi(\boldsymbol{\theta}_j|\mathcal{P}_j, \mathcal{M})d\boldsymbol{\theta}_j\pi(\mathcal{P}_j|\mathcal{M})}{\pi(\mathbf{y}|\mathcal{M})} \tag{72}$$

with  $\sum_{j=1}^m \rho_j = 1$ , and the largest  $\rho_j \in [0, 1]$  corresponds to the most plausible model for data  $\mathbf{y} \in \mathcal{Y}$ .

Finally, we come to the case of misspecified parametric models in which  $\mu^* \notin \mathcal{P}(\Theta)$ ; i.e. no parameter  $\boldsymbol{\theta}^*$  exists such that the truth  $\mu^* = \mathcal{P}(\boldsymbol{\theta}^*)$ . This situation, we believe, is by far the most common encountered in the use of CG models.

We remark that in the (rare?) case of a well-specified CG model, for any continuous functional  $Q : \Theta \rightarrow \mathbb{R}$  and if  $\Theta$  is compact, if  $\theta^*$  is the unique minimizer of  $Q$  and if

$$\sup_{\theta \in \Theta} |Q(\theta; y_1, y_2, \dots, y_n) - Q(\theta)| \xrightarrow{\mathcal{P}} 0, \tag{73}$$

as  $n \rightarrow \infty$ , then the sequence

$$\hat{\theta}^n = \operatorname{argmin}_{\theta \in \Theta} Q_n(\theta; y_1, y_2, \dots, y_n) \tag{74}$$

converges to  $\theta^*$  in probability as  $n \rightarrow \infty$ . This is proved in Nickl [16]. In particular, under mild assumptions on the smoothness of the log-likelihood  $L_n(\theta)$ ,

$$Q(\theta^*) - Q(\theta) = -D_{KL}(\pi(\cdot|\theta) \parallel \pi(\cdot|\theta^*)) \tag{75}$$

$D_{KL}(\cdot|\cdot)$  being the Kullback-Leibler distance defined in (54). By Jensen’s inequality (see, e.g. [16]),  $Q(\theta^*) \leq Q(\theta) \forall \theta \in \Theta$ ; i.e.  $\theta^*$  is the minimizer of  $Q$ .

The asymptotic results for the finite misspecified case is summed up in the powerful result of Kleijn and van der Vaart [7,19]: let  $g(\mathbf{y})$  denote the probability density associated with the true distribution  $\mu^*$ . Then the posterior density  $\pi_n(\theta|\mathbf{y})$  converges in probability to the normal distribution,

$$\pi_n(\theta|\mathbf{y}) \xrightarrow{\mathcal{P}} \mathcal{N}(\theta^\dagger, \mathbf{V}^{-1}(\theta^\dagger)), \tag{76}$$

where

$$\mathbf{V}_{ij}(\theta) = -\mathbb{E}_g \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} D_{KL}(\cdot | \pi(\cdot|\theta)) \right]_{\theta=\theta^\dagger}. \tag{77}$$

Thus, the best approximation to  $g$  in  $\mathcal{P}(\Theta)$  is the model with the parameter

$$\theta^\dagger = \operatorname{argmin}_{\theta \in \Theta} D_{KL}(g \parallel \pi(\cdot|\theta, \mathcal{P}, \mathcal{M})) \tag{78}$$

$\mathcal{M}$  being a class of parametric models to which  $\mathcal{P}$  belongs.

It is easily shown that  $\theta^\dagger$  is a maximum likelihood estimate, i.e. it maximizes the expected value of the log-likelihood relative to the true density  $g$ :

$$\begin{aligned} \theta^\dagger &= \operatorname{argmin}_{\Theta} \left[ \int_{\mathcal{Y}^n} g(\mathbf{y}) \log g(\mathbf{y}) \, d\mathbf{y} - \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta) \, d\mathbf{y} \right] \\ &= \operatorname{argmin}_{\Theta} \left[ - \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta) \, d\mathbf{y} \right] \\ &= \operatorname{argmax}_{\Theta} \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta) \, d\mathbf{y} \\ &= \operatorname{argmax}_{\Theta} \mathbb{E}_g [\log \pi(\mathbf{y}|\theta)], \end{aligned} \tag{79}$$

where the negative self-entropy  $\int g \log g \, d\mathbf{y}$  was eliminated since it does not depend on  $\theta$  and therefore does not affect the optimization.

**Plausibility- $D_{KL}$  theory**

Let us now suppose that we have two misspecified models,  $\mathcal{P}_1$  and  $\mathcal{P}_2$ . We may compare these models in the Bayesian setting through the concept of model plausibility: if  $\mathcal{P}_1$  is more plausible than  $\mathcal{P}_2$ ,  $\rho_1 > \rho_2$ . In the maximum likelihood setting, the model that yields a probability measure closer to  $\mu^*$  is considered the “better” model. That is, if

$$D_{KL}(g \parallel \pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})) < D_{KL}(g \parallel \pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})), \tag{80}$$



it can be said that model  $\mathcal{P}_1$  is better than model  $\mathcal{P}_2$ . The theorems presented here define the relationship between these two notions of model comparison.

However, Bayesian and frequentist methods fundamentally differ in the way they view the model parameters. Bayesian methods consider parameters to be stochastic, characterized by probability density functions, while frequentist approaches seek a single, deterministic parameter value. To bridge this gap in methodology, we note that considering parameters as deterministic vectors, for example  $\theta_0$ , is akin to assigning them delta functions as their posterior probability distributions, which result from delta prior distributions. In this case, the model evidence is given by

$$\pi(\mathbf{y}|\mathcal{P}_i, \mathcal{M}) = \int_{\Theta} \pi(\mathbf{y}|\theta, \mathcal{P}_i, \mathcal{M})\delta(\theta - \theta_0) d\theta = \pi(\mathbf{y}|\theta_0, \mathcal{P}_i, \mathcal{M}). \tag{81}$$

In particular, if we consider the optimal parameter  $\theta_i^\dagger$  for model  $\mathcal{P}_i$ ,  $\pi(\mathbf{y}|\mathcal{P}_i, \mathcal{M}) = \pi(\mathbf{y}|\theta_i^\dagger, \mathcal{P}_i, \mathcal{M})$ . We can take the ratio of posterior model plausibilities,

$$\frac{\rho_1}{\rho_2} = \frac{\pi(\mathbf{y}|\mathcal{P}_1, \mathcal{M})\pi(\mathcal{P}_1|\mathcal{M})}{\pi(\mathbf{y}|\mathcal{P}_2, \mathcal{M})\pi(\mathcal{P}_2|\mathcal{M})} = \frac{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})\pi(\mathcal{P}_1|\mathcal{M})}{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})\pi(\mathcal{P}_2|\mathcal{M})} = \frac{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})}{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})} O_{12}, \tag{82}$$

where  $O_{12} = \pi(\mathcal{P}_1|\mathcal{M})/\pi(\mathcal{P}_2|\mathcal{M})$  is the prior odds and is often assumed to be one. With these assumptions in force, we present the following theorems.

**Theorem 2.** *Let (82) hold. If  $\mathcal{P}_1$  is more plausible than  $\mathcal{P}_2$  and  $O_{12} \leq 1$ , then (80) holds.*

*Proof.* If  $\mathcal{P}_1$  is more plausible than  $\mathcal{P}_2$ ,

$$1 < \frac{\rho_1}{\rho_2} = \frac{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})}{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})} O_{12} \leq \frac{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})}{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})} \tag{83}$$

Equivalently, the reciprocal of the far right-hand side is less than one, so

$$\log \frac{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 0. \tag{84}$$

Since  $g(\mathbf{y})$  is a probability measure, it is always non-negative. Thus

$$g(\mathbf{y}) \log \frac{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 0 \Rightarrow \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \frac{\pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M})} d\mathbf{y} < 0. \tag{85}$$

This can be expanded into

$$\int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M}) d\mathbf{y} - \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M}) d\mathbf{y} < 0, \tag{86}$$

which means

$$- \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta_1^\dagger, \mathcal{P}_1, \mathcal{M}) d\mathbf{y} < - \int_{\mathcal{Y}^n} g(\mathbf{y}) \log \pi(\mathbf{y}|\theta_2^\dagger, \mathcal{P}_2, \mathcal{M}) d\mathbf{y}. \tag{87}$$

By adding the quantity  $\int_{\mathcal{Y}^n} g \log g d\mathbf{y}$  to both sides, the desired result (80) immediately follows.  $\square$

This theorem demonstrates that if model  $\mathcal{P}_1$  is “better” than model  $\mathcal{P}_2$  in the Bayesian sense, it is also a “better” deterministic model in the sense of (80). However, the reverse

implication requires much stronger conditions. The assertion (80) can be equivalently written as

$$\int_{\mathcal{Y}^n} g(\mathbf{y}) \log \frac{\pi(\mathbf{y}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} d\mathbf{y} < 0. \tag{88}$$

For this inequality to hold, the relationship

$$\frac{\pi(\mathbf{y}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 1 \tag{89}$$

does not necessarily need to be true for every point  $\mathbf{y} \in \mathcal{Y}^n$ .

One perhaps naive way to proceed is to invoke the Mean Value Theorem: if  $|\mathcal{Y}^n| < \infty$  and under suitable smoothness conditions, there exists some  $\bar{\mathbf{y}} \in \mathcal{Y}^n$  such that

$$\int_{\mathcal{Y}^n} g(\mathbf{y}) \log \frac{\pi(\mathbf{y}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\mathbf{y}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} d\mathbf{y} = |\mathcal{Y}^n| g(\bar{\mathbf{y}}) \log \frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})}. \tag{90}$$

Then, combining (88) and (90) yields,

$$|\mathcal{Y}^n| g(\bar{\mathbf{y}}) \log \frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 0. \tag{91}$$

Since  $|\mathcal{Y}^n| > 0$  and  $g(\mathbf{y}) > 0$ ,

$$\log \frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 0 \Rightarrow \frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})} < 1 \Rightarrow \frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})} > 1. \tag{92}$$

If  $O_{12} \geq 1$ ,

$$\frac{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})}{\pi(\bar{\mathbf{y}}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M})} O_{12} > 1 \Rightarrow \frac{\rho_1}{\rho_2} > 1. \tag{93}$$

Thus  $\mathcal{P}_1$  is more plausible than  $\mathcal{P}_2$  for given data  $\bar{\mathbf{y}}$ .

In summary, we have:

**Theorem 3.** *If  $D_{KL}(g||\pi(\mathbf{y}|\boldsymbol{\theta}_1^\dagger, \mathcal{P}_1, \mathcal{M})) < D_{KL}(g||\pi(\mathbf{y}|\boldsymbol{\theta}_2^\dagger, \mathcal{P}_2, \mathcal{M}))$  and if  $|\mathcal{Y}^n| < \infty$  and if (90) holds, then there exists a  $\bar{\mathbf{y}} \in \mathcal{Y}^n$  such that  $\mathcal{P}_1$  is more plausible than  $\mathcal{P}_2$ , given that  $O_{12} \geq 1$ .*

### Conclusions

The formal structure of a posterior estimates for errors in quantities of interest in CG approximations of atomistic systems is given by (36) if the CG model is sufficiently close to the AA model in some sense, and this error depends upon the CG model parameter  $\boldsymbol{\theta}$ . Numerical experiments presented in Section ‘Numerical example: estimation of error in CG-approximation of a polyethelyne chain’ involving a family of CG models of a polyethylene chain of united atom monomers suggest that these estimates can be very good indications of the error inherent in CG models of observables in the AA system.

In section ‘Errors in information entropy’ an example of special interest arises in the comparison of the information entropy of AA and CG models. This leads to estimates (62) and (63) involving the Kullback-Leibler divergence,  $D_{KL}$ .

When the CG model is misspecified in a statistical sense, which is generally the case, the “ $D_{KL}$ -distance” between the AA truth and the best possible approximation of any CG model is defined by choosing  $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$ , the minimizer of the  $D_{KL}$  as indicated in (78). Under special assumptions, one can relate the  $D_{KL}$  distance to Bayesian posterior model

plausibility, as stated in our Theorem 2, which provides sufficient conditions for the most plausible model among a class of models to be in fact closest to the AA model in the  $D_{KL}$  sense. The possible role of estimates such as (36), (62), and (63) in model validation should be noted.

For each map  $G : AA \rightarrow CG$  of the type defined by (10), a set  $\mathcal{M}$  of parametric model classes  $\{\mathcal{P}_1(\theta_1), \mathcal{P}_2(\theta_2), \dots, \mathcal{P}_m(\theta_m)\}$  is defined, each with undetermined and possibly random parameter vectors  $\theta_i$ . For a calibration scenario, AA calibration data  $y_c = \{y_1, y_2, \dots, y_n\}$  are sampled, and a series of Bayesian updates is performed using an expanded form of Bayes’s rule that recognizes prior choices of the set  $\mathcal{M}$  and the class  $\mathcal{P}_j$  within  $\mathcal{M}$ :

$$\pi(\theta_j | y_c, \mathcal{P}_j, \mathcal{M}) \propto \pi(y_c | \theta_j, \mathcal{P}_j, \mathcal{M}) \pi(\theta_j | \mathcal{P}_j, \mathcal{M}), \quad 1 \leq j \leq m \tag{94}$$

The marginalization of the right-hand side of this relation is the model evidence, which serves as a likelihood function for a higher level of Bayes’s rule. The corresponding posteriors are the model plausibilities of (72). We remark that the notion of model plausibilities is an extension of the idea of Bayes factors prevalent in statistic literature (see e.g. [12] for discussion of the ideas) and was introduced to the best of our knowledge in [18]. The development of algorithms involving Bayesian plausibilities to study model selection in CG models of complex atomic system is discussed in [3,20].

It has been demonstrated, the most plausible model in a set will, under stated assumptions, involve parameters that minimize the  $D_{KL}$ –distance between the model and the so-called truth parameters. Whether that “best” model is valid for the intended purpose depends on tolerances set of error in key observables, the QoIs of the validation scenario (see [3]).

**Appendix**

A surrogate pair of equations approximating (30) and (32) may be embodied in the problem of finding the pair  $(u_0, z_0) \in \mathcal{V} \times \mathcal{V}$  such that

$$\left. \begin{aligned} \mathcal{B}_0(u_0; v) &= \mathcal{F}_0(v) \quad \forall v \in \mathcal{V} \\ \mathcal{B}'_0(u_0; z_0, v) &= Q'_0(u_0; v) \quad \forall v \in \mathcal{V} \end{aligned} \right\}. \tag{A-1}$$

The remainder  $\Delta$  in (34) can, in this case, be shown (see [4]) to be given by:

$$\begin{aligned} \Delta &= \frac{1}{2} \int_0^1 \{ \mathcal{B}''(u_0 + \theta e_0; e_0, z_0 + \theta \varepsilon_0) \\ &\quad - Q''(u_0 + \theta e_0; e_0, e_0, \varepsilon_0) \} d\theta \\ &\quad + \frac{1}{2} \int_0^1 \{ Q'''(u_0 + \theta e_0; e_0, e_0, \varepsilon_0) - 3\mathcal{B}''(u_0 + \theta e_0; e_0, \varepsilon_0) \\ &\quad - \mathcal{B}'''(u_0 + \theta e_0; e_0, e_0, e_0, z_0 + \theta) \\ &\quad - \mathcal{B}'''(u_0 + \theta e_0; e_0, e_0, e_0, z_0 + \theta \varepsilon_0) \theta (1 - \theta) \} d\theta, \end{aligned} \tag{A-2}$$

where

$$e_0 = u - u_0 \quad \text{and} \quad \varepsilon_0 = z - z_0. \tag{A-3}$$

The theory and estimates reduce to finite element *a posteriori* error estimates in the special case in which  $u_0 = u_h$  and  $z_0 = z_h$  are finite element approximation of solutions  $(u, z)$  to partial differential equations (see e.g. [21]).

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

The theory and numerical simulation represent joint work by all authors. All authors read and approved the final manuscript.

**Acknowledgements**

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-5C0009286. The authors benefited from suggestions of Eric Wright, who read an early draft of this work.

Received: 24 November 2014 Accepted: 12 February 2015

Published online: 03 May 2015

**References**

1. Noid WG (2013) Perspective: coarse-grained models for biomolecular systems. *J Chem Phys* 139:090901
2. Li Y, Abberton BC, Kroger M, Liu WK (2013) Challenges in multiscale modeling of polymer dynamics. *Polymers* 5(2):751–832. doi:10.3390/polym5020751
3. Farrell K, Oden JT (2014) Calibration and validation of coarse-grained models of atomic systems: Application to semiconductor manufacturing. *Comput Mech* 54(1):3–19. doi:10.1007/s00466-014-1028-y
4. Oden JT, Prudhomme S (2002) Estimation of modeling error in computational mechanics. *J Comput Phys* 182(2):496–515
5. Oden JT, Prudhomme S, Romkes A, Bauman PT (2006) Multiscale modeling of physical phenomena: adaptive control of models. *SIAM J Sci Comput* 28(6):2359–2389
6. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423623656
7. Kleijn BJK, van der Vaart A (2002) The asymptotics of misspecified bayesian statistics. In: Mikosch T, Janzura M (eds). *Proceedings of the 24th European Meeting of Statisticians*. Prague, Czech Republic
8. Jorgensen WL, Tirado-Rives J (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110(6):1657–1666
9. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118(45):11225–11236
10. Frenkel D, Smit B (2001) *Understanding molecular simulation: from Algorithms to applications*, Computational science. 2nd edn, Vol. 1. Academic Press, San Diego
11. Haile JM (1997) *Molecular dynamics simulation*. John Wiley and Sons, NY
12. Jaynes ET (2003) *Probability theory: the logic of science*. Cambridge University Press, Cambridge
13. Shell MS (2008) The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J Chem Phys* 129(14):144108
14. Geyer CJ (2003). 5601 Notes: the sandwich estimator. School of Statistics, University of Minnesota
15. Freedman DA (2006) On the so-called “Huber sandwich estimator” and “robust standard errors”. *Am Stat* 34:299–302
16. Nickl R (2012). *sSTATISTICAL THEORY*. Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge
17. Kleijn BJK (2004). *Bayesian asymptotics under misspecification*. PhD thesis, Free University Amsterdam
18. Beck JL, Yuan K-V (2004) Model selection using response measurements: Bayesian probabilistic approach. *J Eng Mech* 130(2):192–203
19. Kleijn BJK (2012) van der Vaart AW (2012) The Bernstein-von-Mises theorem under misspecification. *Electronic J Stat* 6:354–381. doi:10.1214/12-EJS675
20. Farrell K, Oden JT, Faghihi D (2015) A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. *J Comput Phys* 295:189–208. ISSN 0021-9991
21. Becker R, Rannacher R (2001) An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica* 10:1–102. doi:10.1017/S0962492901000010

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)