Computer Physics Communications 185 (2014) 3228–3239

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

# Automated parameterization of intermolecular pair potentials using global optimization techniques

Andreas Krämer [a,b,*], Marco Hülsmann [b], Thorsten Köddermann [b], Dirk Reith [a,b]

[a] Department of Electrical Engineering, Mechanical Engineering and Technical Journalism, Bonn-Rhein-Sieg University of Applied Sciences, Grantham-Allee 20, 53757 Sankt Augustin, Germany

[b] Department of Simulation Engineering, Fraunhofer Institute for Algorithms and Scientific Computing, Schloß Birlinghoven, 53757 Sankt Augustin, Germany

ARTICLE INFO

*Article history:*
Received 24 March 2014
Received in revised form
26 August 2014
Accepted 29 August 2014
Available online 16 September 2014

*Keywords:*
Optimization
Molecular simulation
Molecular modeling
Automation
Force field parameters

ABSTRACT

In this work, different global optimization techniques are assessed for the automated development of molecular force fields, as used in molecular dynamics and Monte Carlo simulations. The quest of finding suitable force field parameters is treated as a mathematical minimization problem. Intricate problem characteristics such as extremely costly and even abortive simulations, noisy simulation results, and especially multiple local minima naturally lead to the use of sophisticated global optimization algorithms. Five diverse algorithms (pure random search, recursive random search, CMA-ES, differential evolution, and taboo search) are compared to our own tailor-made solution named CoSMoS. CoSMoS is an automated workflow. It models the parameters' influence on the simulation observables to detect a globally optimal set of parameters. It is shown how and why this approach is superior to other algorithms. Applied to suitable test functions and simulations for phosgene, CoSMoS effectively reduces the number of required simulations and real time for the optimization task.



## 1. Introduction

Nowadays, molecular dynamics and Monte-Carlo simulations are indispensable in various areas, including thermodynamic properties of fluids [1], transport processes in liquids [2], protein folding [3], polymer properties [4] or pharmaceutics [5]. They are likely to become even more important due to the rapid development and affordability of powerful computers.

As possibilities open up, there is a growing need for accurate molecular models that are tailored for specific applications with quantitatively matching capabilities. The parameterization of the models, which are given as force field equations, is the most critical part of the modeling process. While *intramolecular* parameters can be derived from quantum mechanics, *intermolecular* parameters, e.g. the Lennard-Jones parameters, have to be adjusted empirically. In practice, this means that the latter have to be tuned in order to reproduce macroscopic physical observables, such as density, diffusion coefficients, viscosity, vapor pressure and heat of vaporization. This so-called calibration is the principal bottleneck in the modeling process.

The first reason for the high cost of the calibration is the lengthy computation time required for a single simulation. Simulations have to be iteratively repeated with changing parameter settings in order to minimize the loss between simulated and experimental observables. Second, the simulation observables are calculated as statistical averages and hence they are noisy. Third, simulations can terminate without any useful result, for instance when the simulation system becomes unstable for a certain parameter combination. Fourth, the objective loss functional may have plenty of local optima which is unfavorable for the discovery of a global optimum as well.

Consequently, the calibration requires sophisticated optimization algorithms that are capable of detecting a globally optimal set of intermolecular force field parameters automatically and within an acceptable amount of time. Hence, they have to scale well on multi-core computers, be robust with respect to noise as well as abortive function evaluations and prevent preliminary convergence to local optima. Each of the above criteria is indispensable for a generic force field optimization workflow.

* Corresponding author at: Department of Electrical Engineering, Mechanical Engineering and Technical Journalism, Bonn-Rhein-Sieg University of Applied Sciences, Grantham-Allee 20, 53757 Sankt Augustin, Germany. Tel.: +49 22418659644.
*E-mail addresses:* andreas.kraemer@h-brs.de (A. Krämer), marco.huelsmann@scai.fraunhofer.de (M. Hülsmann), thorsten.koeddermann@scai.fraunhofer.de (T. Köddermann), dirk.reith@h-brs.de (D. Reith).

http://dx.doi.org/10.1016/j.cpc.2014.08.022
0010-4655/

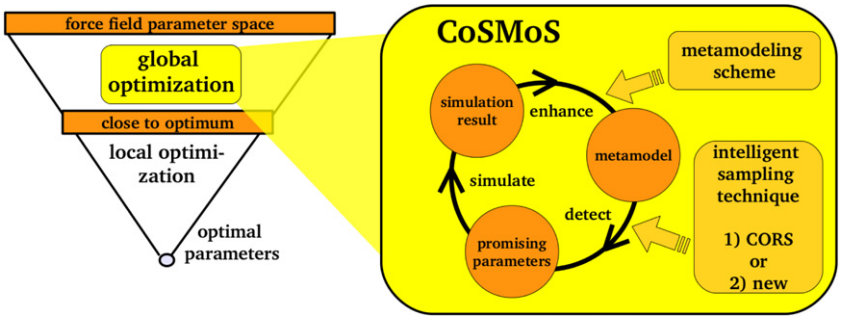


**Fig. 1.** Desired optimization workflow to detect optimal force field parameters. The global CoSMoS optimization is based on metamodeling. The metamodel is utilized to detect a promising parameter vector. The simulation observables are calculated and integrated into the metamodel in order to enhance its accuracy.

Although several groups have worked on developing accurate force fields, the authors do not know any approach that meets all of the criteria. The most significant contributions are briefly highlighted in the next section.

### 1.1. Related work

The Nelder–Mead simplex method [6] was one of the first optimization algorithms used for automated force field design [7,8]. It is robust with respect to noise and derivative-free but does not guarantee convergence, not even to a local optimum. Up to a hundred sequential function evaluations were needed to solve calibration problems for hydrocarbons at a single temperature. The obtained parameters failed to produce similarly good results for other temperatures [9].

Later, Ungerer et al. [10,11] used the gradient-based Gauß–Newton method, obtaining accurate force fields for small molecules. In the meantime, a similar method by Stoll [12] was successfully applied to other molecules such as cyclohexanol [13] and acetonitrile [14]. These methods are efficient in the immediate neighborhood of an optimum but they require a suitable initial guess for the objective parameters. Otherwise, they converge to a non-optimal solution.

The recently proposed gradient-based optimization workflow (GROW) by Hülsmann et al. [15] has extended this approach by various descent methods like *steepest descent, conjugate gradients,* and *trust region*, allowing the initial parameters to be situated farther away from the optimum. GROW has proven to be a useful and reliable tool in diverse applications: test functions [16], small molecules [17], and ionic liquids [18]. Nevertheless, suitable starting parameters are necessary for convergence. Parallel iterations are not possible for these gradient-based methods since iterations depend on each other.

Alternatively, metamodels were used for the optimization task. Metamodels are also called *response surface models* in the literature [19]. The terms are used interchangeably throughout the present paper, denoting multimodal interpolations or regressions with cheap evaluations. Maaß et al. [20] have studied the global influence of force field parameters for ethylene oxide. Metamodels, based on 80 random parameter sets, were created and studied with the interactive tool DesParO [21]. A set of parameters was selected manually. Most recently, Hülsmann et al. [22,23] have shown that global optimization in connection with GROW is likely to be a generalizable workflow for optimizing intermolecular parameters entirely automatically from scratch (cf. Fig. 1). The parameters provided by DesParO were not optimal, but the subsequent GROW optimization turned them into an excellent force field within 14 steepest descent iterations. However, the interactivity of the global metamodeling as well as the great number of overall function evaluations are unfavorable for a generic calibration tool, as the overall optimization process took weeks [22]. At that point, it became clear that the global part of the optimization process could be enhanced in terms of efficiency and automation. Consequently, in this work, research was focused on appropriate global optimization techniques.

### 1.2. Developing an appropriate global optimization strategy

Global optimization is concerned with two goals: *exploration* of the search space and *exploitation* of previous function evaluations. Various techniques have been developed, for instance simulated annealing [24,25], evolutionary methods (including swarm optimization and genetic algorithms) [26–29], taboo search [30,31], multistart [32], direct search [33,34], and finally metamodel-based optimization [19,35–37]. The last are particularly suited for functions with computationally expensive evaluations as they involve the results of all previous evaluations [35].

Different types of global optimizers were tested by the authors with respect to the given calibration problem. In particular, six different algorithms were selected for comparison: pure random search (PRS), recursive random search (RRS) [38], the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [39,40], differential evolution (DE) [27], a special taboo search (TS) algorithm [30], and Constrained Optimization using Response Surfaces (CORS) [19,37]. Each of them represents a particular class of global optimizers.

In contrast to the random sampling favored by DesParO, modern metamodel-based optimizers, like CORS, rely on intelligent sampling techniques. In fact, the metamodel is exploited to concentrate sampling onto more interesting domains of the search space. Hence, the accuracy of the model grows especially in these domains and, furthermore, the sampling points are likely to approach an optimal set of parameters. Taken together, metamodeling and optimization complement and stimulate each other.

CoSMoS follows this approach as a global optimization tool for the **C**alibration **o**f molecular force fields by **S**imultaneous **M**odeling **o**f **S**imulated data (cf. Fig. 1). Currently, three metamodeling schemes and two different intelligent sampling techniques are implemented in CoSMoS: The CORS sampling technique and a new method developed by the authors. Additional components of CoSMoS are a suitable normalization of the metamodels, different loss functionals for the optimization, a parallelization framework and a way to make use of abortive simulations. New features like other sampling techniques or metamodeling schemes can easily be integrated through the modular program structure. CoSMoS uses the GROW interface to molecular simulations. Taken together, it is tailored to meet all of the above-mentioned criteria and satisfy the need for an appropriate globally convergent force field calibration tool. It can be used as a pre-optimizer for gradient based optimization (cf. Fig. 1) or as a stand-alone solution.

The numerical results on test functions and a case study on phosgene illustrate that metamodel-driven optimization is a viable