Multicriteria optimization of molecular force field models

M. T. Horsch,¹ K. Stöbener,¹,² S. Werth,¹ J. Vrabec,³ and H. Hasse¹

¹Laboratory of Engineering Thermodynamics, University of Kaiserslautern, Germany
²Fraunhofer Institute for Industrial Mathematics, Kaiserslautern, Germany
³Thermodynamics and Energy Technology, University of Paderborn, Germany

Ulam Computer Simulation Workshop
Lviv, June 23, 2017
Reliable molecular force field models

- Physically realistic modeling of intermolecular interactions
- Separate contributions due to repulsive and dispersive as well as electrostatic interactions

Physics (qualitative accuracy)

- No blind fitting, but parameters of effective pair potentials are adjusted to experimental data
- Physical realism facilitates reliable interpolation and extrapolation

Engineering (quantitative reliability)
Literature models adjusted to bulk VLE data

2CLJQ models:
- 2 LJ centers
- 1 quadrupole

Fit of parameters $\sigma$, $\varepsilon$, $L$, $Q$ to VLE data of 29 fluids by Stoll et al.

Deviation:
- $\delta\rho' \approx 1\%$
- $\delta\rho^s \approx 5\%$

No interfacial properties were considered for the parameterization.
Surface tension: Long-range correction

Long range correction from the density profile, following Janeček.¹⁻³

Angle averaging expression for multi-site models, following Cook and Rowlinson⁴,⁵ as well as Lustig.³,⁶

Surface tension: Long-range correction

Long range correction from the density profile, following Janeček.

Angle averaging expression for multi-site models, following Cook and Rowlinson as well as Lustig.

Two-center LJ fluid (2CLJ)

$T = 0.979 \varepsilon$

Janeček-Lustig term

no angle averaging

no correction at all

large systems “1”: molecular dynamics

http://www.ls1-mardyn.de/
Validation of molecular force fields

2CLJQ: Two LJ centers + quadrupole

Fit to bulk properties
About 20 % overestimation of the surface tension

Validation of molecular force fields

2CLJQ: Two LJ centers + quadrupole

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

2CLJD: Two LJ + dipole

$\Delta \gamma / \%$

Number of fluids

$< 5$ 5-10 10-15 15-20 $> 20$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$

$\gamma / \text{mN m}^{-1}$

$T / \text{K}$


# Validation of molecular force fields

## Non-polar: 1CLJ
- Neon (Ne)
- Argon (Ar)
- Krypton (Kr)
- Xenon (Xe)
- Methane (CH₄)

## Quadrupolar: 2CLJQ
- Fluorine (F₂)
- Chlorine (Cl₂)
- Bromine (Br₂)
- Iodine (I₂)
- Nitrogen (N₂)
- Oxygen (O₂)
- Carbon dioxide (CO₂)
- Carbon sulfide (CS₂)
- Ethane (C₂H₆)
- Ethylene (C₂H₄)
- Acetylene (C₂H₂)
- Propene (CH₃CH=CH₂)
- Propylene (CH₂=CH=CH₂)

## Dipolar: 2CLJD
- Carbon monoxide (CO)
- R11 (CFCl₂)
- R12 (CF₂Cl₂)
- R13 (CF₃Cl)
- R13B1 (CFBrF)
- R22 (CHFCl)
- R23 (CHF₂)
- R41 (CHF)
- R123 (CHCl₂CF₂)
- R124 (CHFClCF₂)
- R125 (CH₂CF₂)
- R134a (CH₂F₂)
- R141b (CH₃CF₂)
- R142b (CH₃CF₂Cl)
- R143a (CH₃CF₃)
- R152a (CH₃CHF₂)
- R40 (CH₂Cl)
- R40B1 (CH₂Br)
- CH₃
- R30B1 (CH₂BrCl)
- R20 (CHCl₂)
- R20B3 (CHBr₃)
- R21 (CHFCl₂)


## Multicentric United Atom Models
- Isobutane (C₄H₁₀)
- Cyclohexane (C₆H₁₂)
- Methanol (CH₃OH)
- Ethanol (C₂H₅OH)
- Formaldehyde (CH₂O)
- Dimethyl ether (CH₃OCH₃)
- Acetone (C₃H₆O)
- Ammonia (NH₃)
- Methylamine (NH₂CH₃)
- Dimethylamine (CH₃NHCH₃)
- R227ea (CF₃CHFCCF₃)
- Sulfur dioxide (SO₂)
- Ethylene oxide (C₂H₄O)
- Dimethyl sulfide (CH₃-S-CH₃)
- Hydrogen cyanide (HCN)
- Acetonitrile (CH₃CN)
- Thiophene (SC₄H₄)
- Nitromethane (CH₃NO₂)
- Phosgene (COCl₂)
- Benzene (C₆H₆)
- Toluene (C₆H₅CH₃)
- Chlorobenzene (C₆H₅Cl)
- Dichlorobenzene (C₆H₄Cl₂)
- Cyclohexanol (C₆H₁₀OH)
- Cyclohexanone (C₆H₁₀O)
- Cyanogen (C₃N₄)
- Cyanogen chloride (CCIN)
- Formic acid (CH₂O₂)
- Ethylene glycol (C₃H₆O₂)
- TIP4P/2012 water (H₂O)
- Hydrazine (N₂H₄)
- Monomethylhydrazine (CH₃N₂H)
- Dimethylhydrazine (C₂H₅N₂)
- Perfluorobutane (C₆F₁₄)
- Ethyl acetate (C₃H₆O₂)
- HMDSO (C₆H₁₂OSi₂)
- D₄ (C₄H₈O₂Si₂)

### Additional Data
- Carbon dioxide (CO₂)
- Oxygen (O₂)
- Nitrogen (N₂)
- Ethane (C₂H₆)
- Ethylene (C₂H₄)
- Propene (CH₃CH=CH₂)
- Propylene (CH₂=CH=CH₂)
- Methane (CH₄)
- Ethylene oxide (C₂H₄O)

### Additional Chemicals
- Acetonitrile (NC₃)
- Cyanogen (C₃N₄)
- Cyanogen chloride (CCIN)
- Thiophene (SC₄H₄)
- Nitromethane (CH₃NO₂)
- Phosgene (COCl₂)
- Benzene (C₆H₆)
- Toluene (C₆H₅CH₃)
- Chlorobenzene (C₆H₅Cl)
- Dichlorobenzene (C₆H₄Cl₂)
- Cyclohexanol (C₆H₁₀OH)
- Cyclohexanone (C₆H₁₀O)
Multicriteria optimization requires characterizing the whole model class.
Surface tension of 2CLJQ and 2CLJD fluids

- Systematic exploration of the four-dimensional model parameter space
- Correlation of $\gamma$ by critical scaling expressions (2CLJQ, 2CLJD, Mie-6)
Computation of the Pareto set\textsuperscript{1, 2}

**Multicriteria optimization problem**

Simultaneously minimized objective functions $f_\xi$ with $\xi \in \{\rho', \rho^s, \gamma\}$ given by

\[
f_\xi = \left\langle \delta \xi^2 \right\rangle_{0.55T_c^{\exp} < T < 0.95T_c^{\exp}} = \lim_{N \to \infty} \frac{1}{N+1} \sum_{i=0}^{N} \left(1 - \frac{\xi^{\text{sim}}(T)}{\xi^{\text{exp}}(T)}\right)^2 T/T_c = 0.55 + 0.4i/N \]

(here: $N = 9$).

**Sandwiching**

Alternating construction of inner (reachable) and outer (unreachable) approximations, in regions where the Pareto set is locally convex.

**Hyperboxing**

In non-convex regions (hyperboxes), Pascoletti-Serafini scalarization is used to formulate an appropriately constrained single-criterion problem.

Computation of the Pareto set

- **Model parameters**
  - LJ size parameter $\sigma$
  - LJ energy parameter $\varepsilon$
  - Model elongation $L$
  - Quadrupole moment $Q$

- **Optimization criteria**
  - Saturated liquid density $\rho'$
  - Saturated vapor pressure $p^s$
  - Vapor-liquid surface tension $\gamma$

Dimension of Pareto set $d \leq a$.  

Dimension of Pareto set $d \leq b - 1$.

$$d = \min(a, b - 1).$$  

(here, $a = 4$)  

(here, $b = 3$)  

(here, $d = 2$)

---

Multicriteria molecular model optimization\textsuperscript{1, 2}

Representation of objective and parameter spaces by patch plots:

Pareto-optimal 2CLJQ models of molecular oxygen

\textsuperscript{1}\textit{Stöbener et al.}, \textit{Fluid Phase Equilib.} 373, 100, \textbf{2014}; \textsuperscript{2}\textit{Stöbener et al.}, \textit{Fluid Phase Equilib.} 408, 141, \textbf{2016}. 
Multicriteria molecular model optimization\textsuperscript{1, 2}

Requirements for the criteria follow the priorities of the target application:

Restrictions imposed on 2CLJQ models of molecular oxygen

\textsuperscript{1}Stöbener et al., \textit{Fluid Phase Equilib.} 373, 100, \textbf{2014}; \textsuperscript{2}Stöbener et al., \textit{Fluid Phase Equilib.} 408, 141, \textbf{2016}.
Multicriteria molecular model optimization$^{1, 2}$

Requirements for the criteria follow the priorities of the target application:

Model accuracy for ten quadrupolar fluids\textsuperscript{1, 2}

![Graph showing model accuracy for ten quadrupolar fluids](image)

- Surface tension: avg. 18 \%
- Vapor pressure: avg. 2.2 \%
- Sat. liq. density: avg. 0.3 \%

Model accuracy for ten quadrupolar fluids\(^1\)–\(^3\)

- surface tension: avg. 18 % → 12 %
- vapor pressure: avg. 2.2 % → 3.5 %
- sat. liq. density: avg. 0.3 % → 1.4 %

Comparison between model classes\textsuperscript{1}

Carbon dioxide: Mie–$n,6$ potential $u(r) = \frac{n}{n-6} \left( \frac{n}{6} \right)^{6} \epsilon \left[ \left( \frac{\sigma}{r} \right)^n - \left( \frac{\sigma}{r} \right)^6 \right]$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Parameter and objective space comparisons for carbon dioxide model.}
\end{figure}

Comparison between model classes

Carbon dioxide: Mie–$n,6$ potential ./ other model classes

---

deviation of saturated liquid density (%)  

---

deviation of vapor pressure (%)  

---

Paradigm shift in molecular modeling

The art of molecular modeling

An expert modeling artist designs and publishes
- a single optimized model for a particular fluid,
- according to his choice of criteria (often unknown to the public),
- users are passive, they have to live with the artists' decision.

Molecular modeling as a technology

For well-characterized model classes and multiple optimization criteria,
- the dependence of thermodynamic properties on the model parameters is determined and correlated,
- the deviation between model properties and real fluid behavior is characterized, and the Pareto set is published,
- users can design their own tailored model with minimal effort.