Multicriteria optimization of molecular force field models

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

[1]*Laboratory of Engineering Thermodynamics, University of Kaiserslautern, Germany*

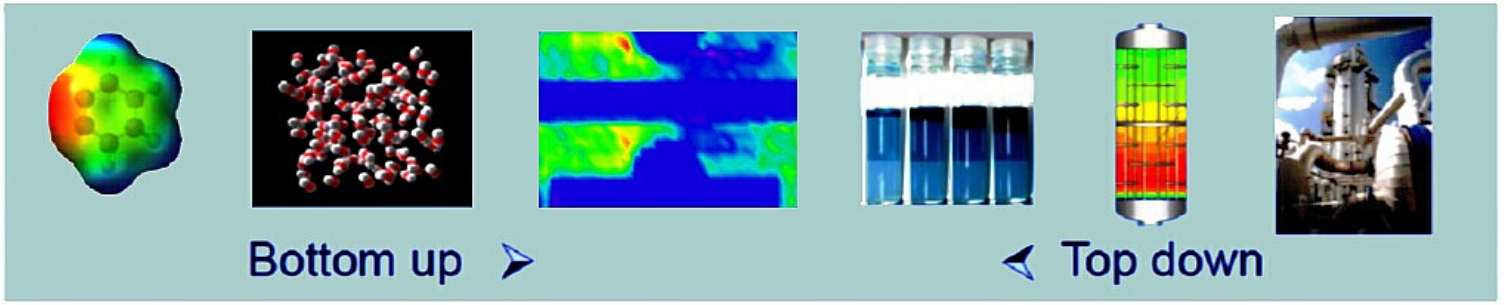
[2]*Fraunhofer Institute for Industrial Mathematics, Kaiserslautern, Germany*

[3]*Thermodynamics and Energy Technology, University of Paderborn, Germany*

Computational Molecular Engineering

Ulam Computer Simulation Workshop
Lviv, June 23, 2017

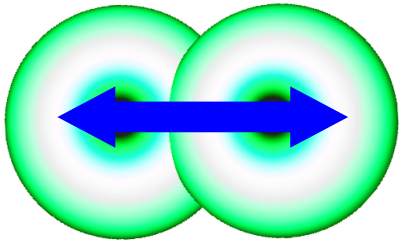
Reliable molecular force field models



Physics (qualitative accuracy) ⟷ **Engineering (quantitative reliability)**

- Physically realistic modeling of intermolecular interactions
- Separate contributions due to repulsive and dispersive as well as electrostatic interactions
- No blind fitting, but parameters of *effective pair potentials* are adjusted to experimental data
- Physical realism facilitates reliable interpolation and extrapolation

Literature models adjusted to bulk VLE data



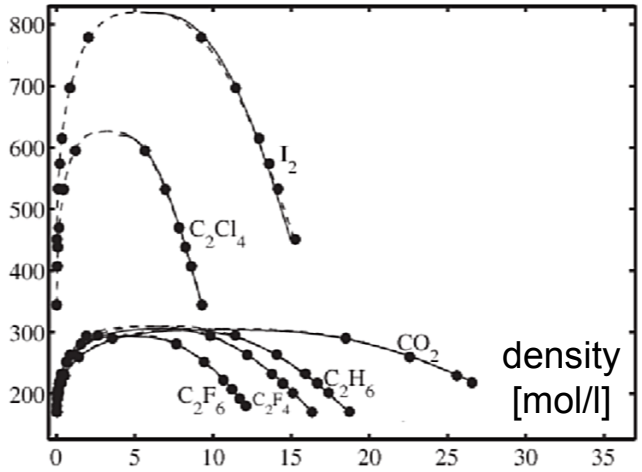
2CLJQ models:

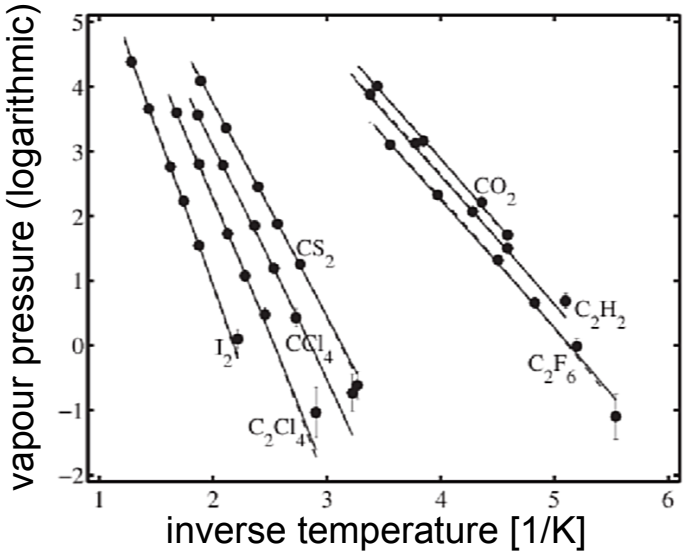
- 2 LJ centers
- 1 quadrupole

Fit of parameters σ, ε, L, Q to VLE data of 29 fluids by Stoll *et al.*

Deviation:

- $\delta\rho'$ ≈ 1 %
- δp^s ≈ 5 %





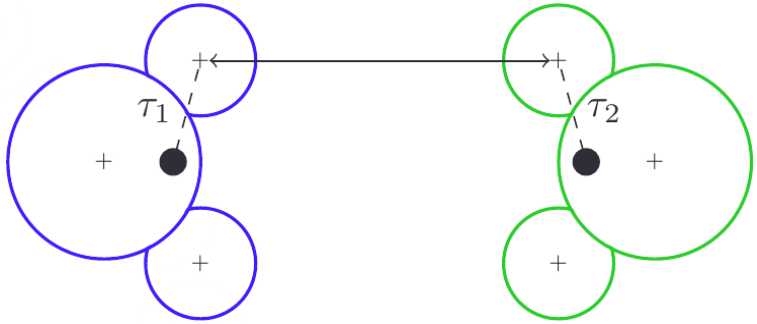
Literature models by J. Stoll, H. Hasse, J. Vrabec *et al.*, **2001** onwards.

- simulation
— DIPPR correlation

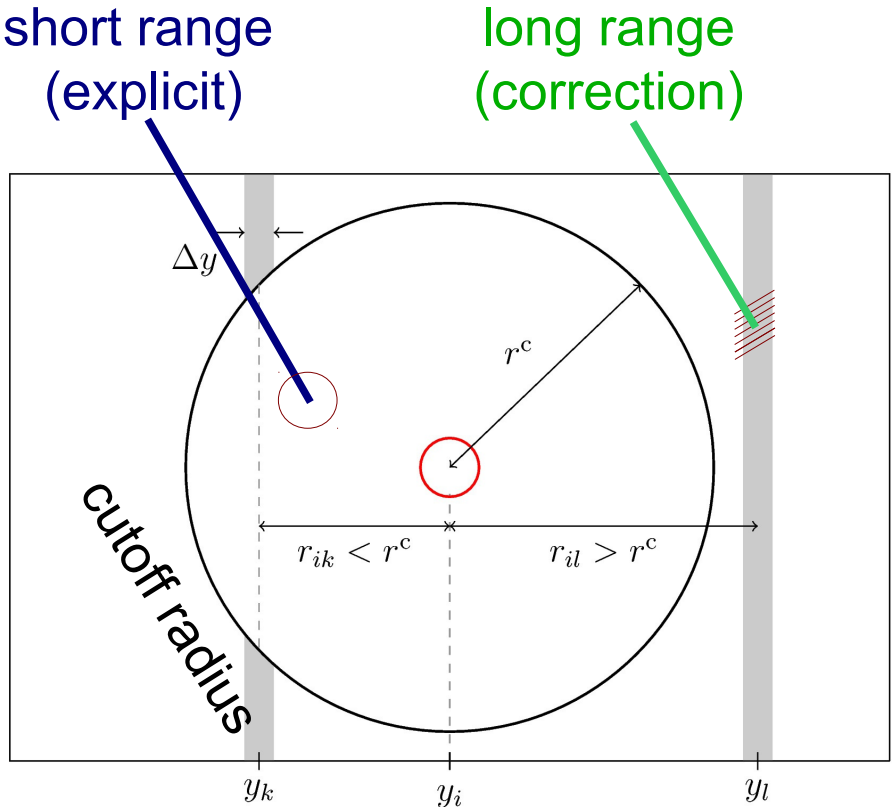
No interfacial properties were considered for the parameterization.

Surface tension: Long-range correction

Long range correction from the density profile, following Janeček.[1–3]



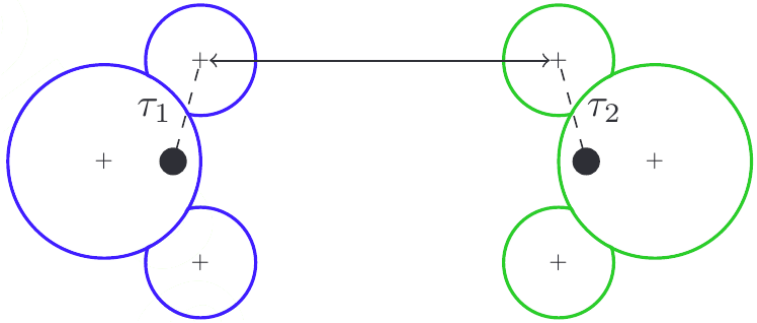
Angle averaging expression for multi-site models, following Cook and Rowlinson[4, 5] as well as Lustig.[3, 6]



[1]Janeček, *J. Phys. Chem. B,* 110, 6264, **2006**; [2]Goujon *et al.*, *J. Chem. Theory Comput.* 11, 4573, **2015**; [3]Werth *et al.*, *Mol. Phys.* 112, 2227, **2014**; [4]Cook and Rowlinson, *Proc. Roy. Soc. A* 219, 405, **1953**; [5]Werth *et al.*, Mol. Phys. 113, 3750, 2015; [6]Lustig, *Mol. Phys.* 65, 175, **1988**.

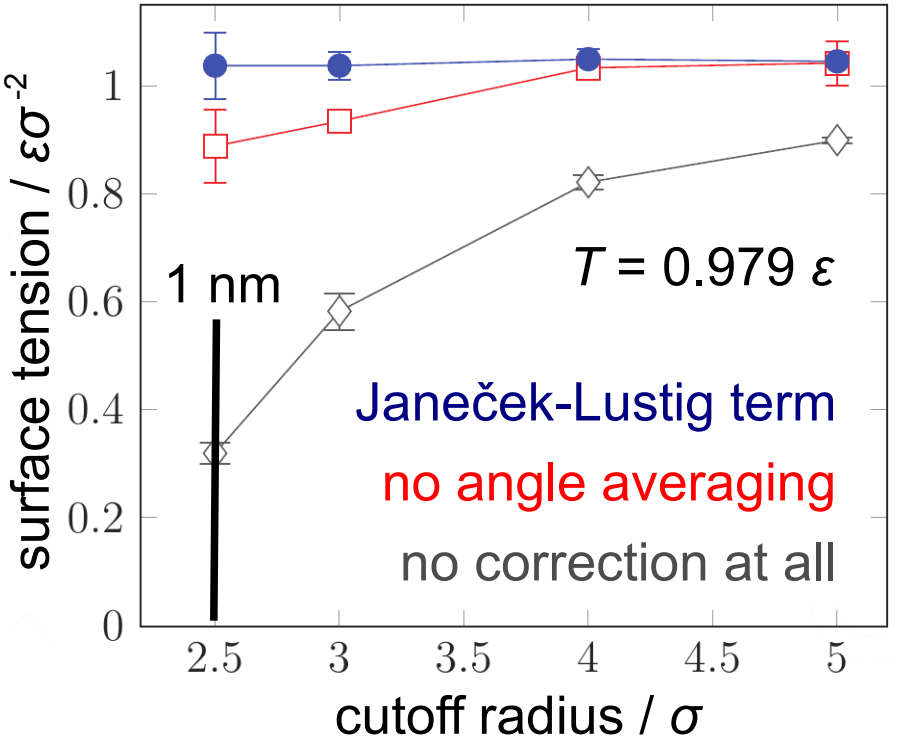
Surface tension: Long-range correction

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)

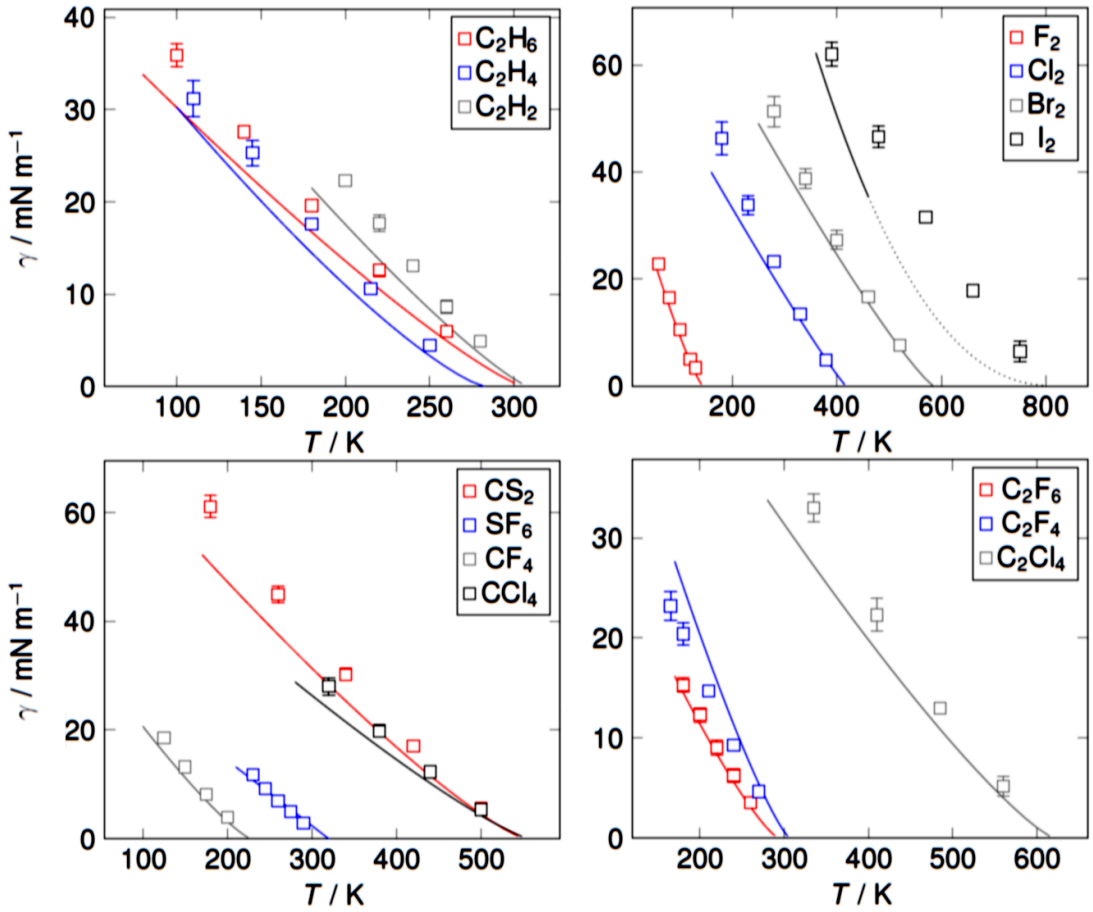


large systems “1”: molecular dynamics

http://www.ls1-mardyn.de/

Validation of molecular force fields

2CLJQ: Two LJ centers + quadrupole[1]



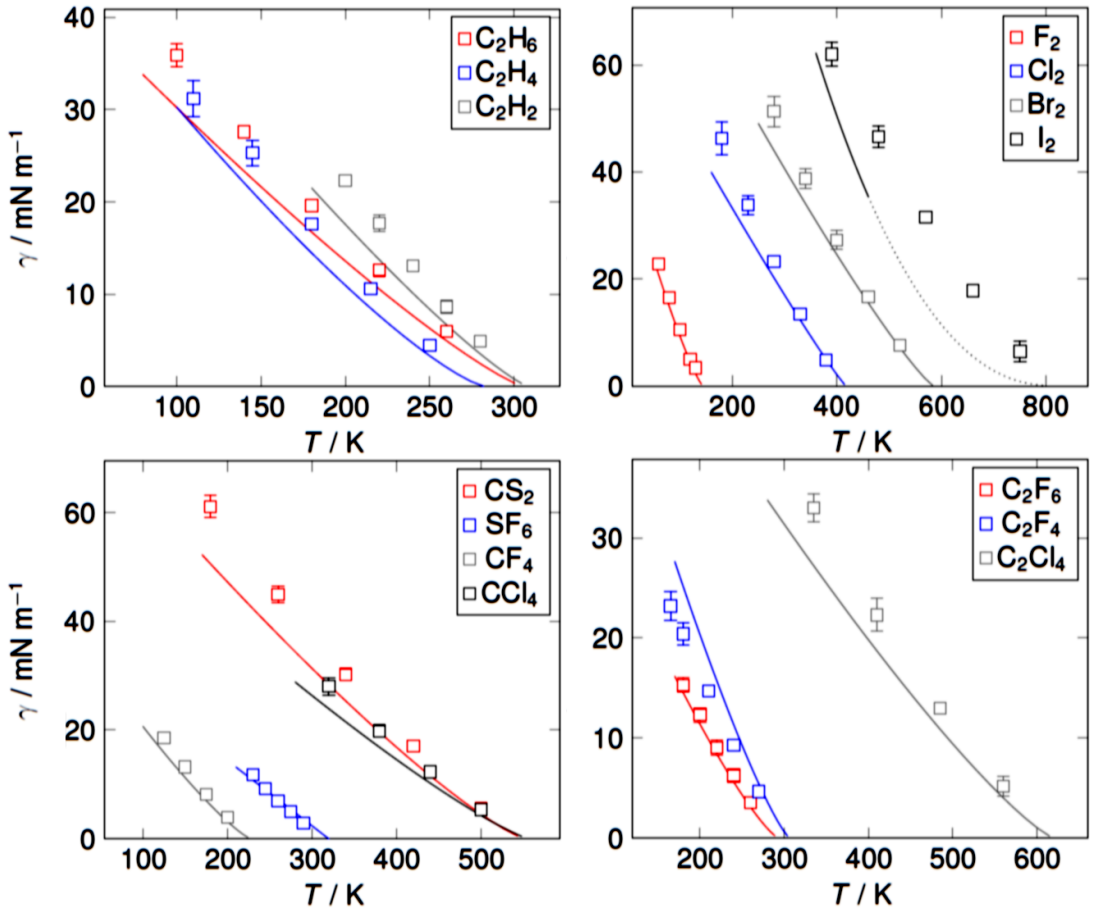
Fit to bulk properties

About 20 % overestimation of the surface tension

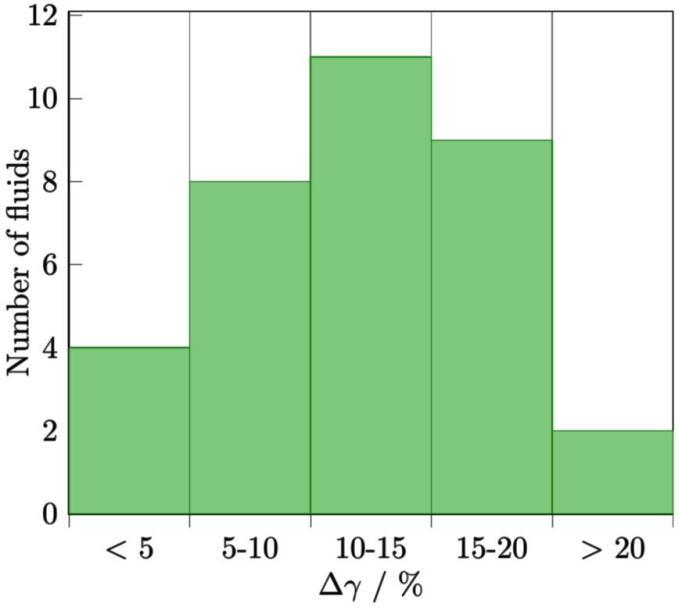
[1]S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse, *Chem. Eng. Sci.* 121, 110–117, **2015**.

Validation of molecular force fields

2CLJQ: Two LJ centers + quadrupole[1]



2CLJD: Two LJ + dipole[2]



[1] S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse, *Chem. Eng. Sci.* 121, 110–117, **2015**;

[2] S. Werth, M. Horsch, H. Hasse, *J. Chem. Phys.* 144, 054702, **2016**.

Validation of molecular force fields

Non-polar: 1CLJ

Neon (Ne)
Argon (Ar)
Krypton (Kr)
Xenon (Xe)
Methane (CH_4)

Quadrupolar: 2CLJQ

Fluorine (F_2)
Chlorine (Cl_2)
Bromine (Br_2)
Iodine (I_2)
Nitrogen (N_2)

+ 20 %

Oxygen (O_2)
Carbon dioxide (CO_2)
Carbon sulfide (CS_2)
Ethane (C_2H_6)
Ethylene (C_2H_4)
Acetylene (C_2H_2)
R116 (C_2F_6)
R1114 (C_2F_4)
R1110 (C_2Cl_4)
Propadiene (CH_2=C=CH_2)
Propyne (CH_3-C≡CH)

Propylene (CH_3-CH=CH_2)
R846 (SF_6)
R14 (CF_4)
R10 (CCl_4)
R113 ($CFCl_2$-CF_2Cl)
R114 (CF_2Cl-CF_2Cl)
R115 (CF_3-CF_2Cl)
R134 (CHF_2-CHF_2)
R150B2 (CH_2Br-CH_2Br)
R114B2 ($CBrF_2$-$CBrF_2$)
R1120 (CHCl=CCl_2)

Literature models by J. Stoll, H. Hasse, J. Vrabec *et al.*, **2001 – 2016**

Dipolar: 2CLJD

Carbon monoxide (CO)
R11 ($CFCl_3$)
R12 (CF_2Cl_2)
R13 (CF_3Cl)
R13B1 ($CBrF_3$)
R22 (CHF_2Cl)
R23 (CHF_3)
R41 (CH_3F)
R123 ($CHCl_2$-CF_3)
R124 (CHFCl-CF_3)
R125 (CHF_2-CF_3)
R134a (CH_2F-CF_3)
R141b (CH_3-$CFCl_2$)
R142b (CH_3-CF_2Cl)
R143a (CH_3-CF_3)
R152a (CH_3-CHF_2)
R40 (CH_3Cl)
R40B1 (CH_3Br)
CH_3I
R30B1 (CH_2BrCl)
R20 ($CHCl_3$)
R20B3 ($CHBr_3$)
R21 ($CHFCl_2$)

+ 12 %

R32 (CH_2F_2)
R30 (CH_2Cl_2)
R30B2 (CH_2Br_2)
CH_2I_2
R12B2 (CBr_2F_2)
R12B1 ($CBrClF_2$)
R10B1 ($CBrCl_3$)
R161 (CH_2F-CH_3)
R150a ($CHCl_2$-CH_3)
R140 ($CHCl_2$-CH_2Cl)
R140a (CCl_3-CH_3)
R130a (CH_2Cl-CCl_3)
R160B1 (CH_2Br-CH_3)
R150B2 ($CHBr_2$-CH_3)
R131b (CH_2F-CCl_3)
R123B1 ($CHClBr$-CF_3)
R112a (CCl_3-CF_2Cl)
R1141 (CHF=CH_2)
R1132a (CF_2=CH_2)
R1140 (CHCl=CH_2)
R1122 (CHCl=CF_2)
R1113 (CFCl=CF_2)
R1113B1 (CFBr=CF_2)

Multicentric United Atom Models

Isobutane (C_4H_{10})
Cyclohexane (C_6H_{12})
Methanol (CH_3OH)
Ethanol (C_2H_5OH)
Formaldehyde (CH_2=O)
Dimethyl ether (CH_3-O-CH_3)
Acetone (C_3H_6O)
Ammonia (NH_3)
Methylamine (NH_2-CH_3)
Dimethylamine (CH_3-NH-CH_3)
R227ea (CF_3-CHF-CF_3)
Sulfur dioxide (SO_2)
Ethylene oxide (C_2H_4O)

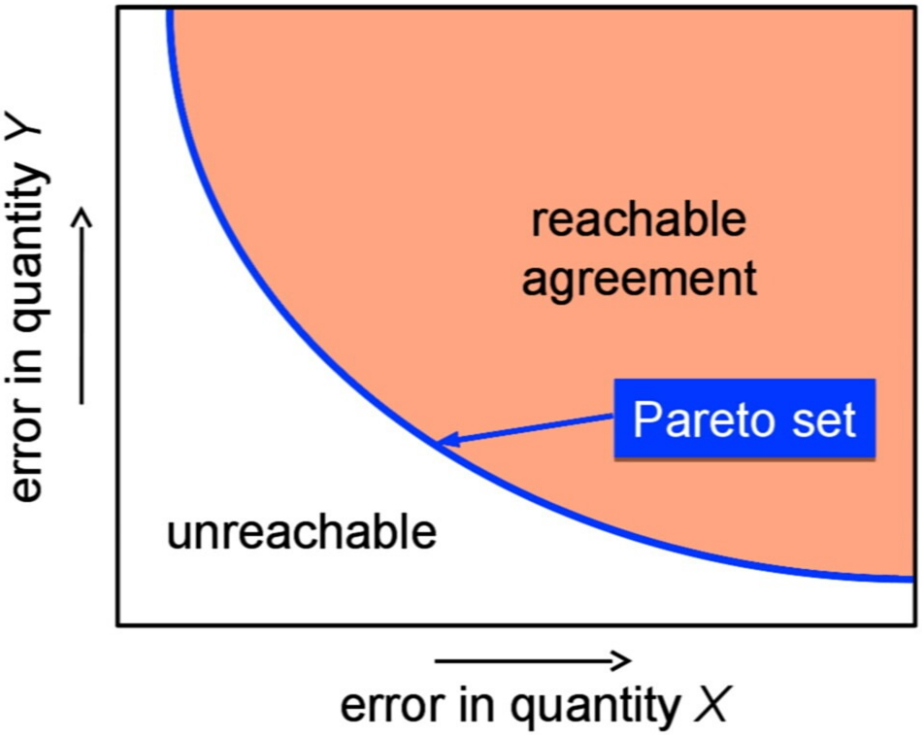
Dimethyl sulfide (CH_3-S-CH_3)
Hydrogen cyanide (HCN)
Acetonitrile (NC_2H_3)
Thiophene (SC_4H_4)
Nitromethane (CH_3NO_2)
Phosgene ($COCl_2$)
Benzene (C_6H_6)
Toluene (C_7H_8)
Chlorobenzene (C_6H_5Cl)
Dichlorobenzene ($C_6H_4Cl_2$)
Cyclohexanol ($C_6H_{11}OH$)
Cyclohexanone ($C_6H_{10}O$)

Cyanogen (C_2N_2)
Cyanogen chloride (CClN)
Formic acid (CH_2O_2)
Ethylene glycol ($C_2H_6O_2$)
TIP4P/2012 water (H_2O)
Hydrazine (N_2H_4)
Monomethylhydrazine (CH_6N_2)
Dimethylhydrazine ($C_2H_8N_2$)
Perfluorobutane (C_4F_{10})
Ethyl acetate ($C_4H_8O_2$)
HMDSO ($C_6H_{12}OSi_2$)
D4 ($C_8H_{24}O_4Si_4$)

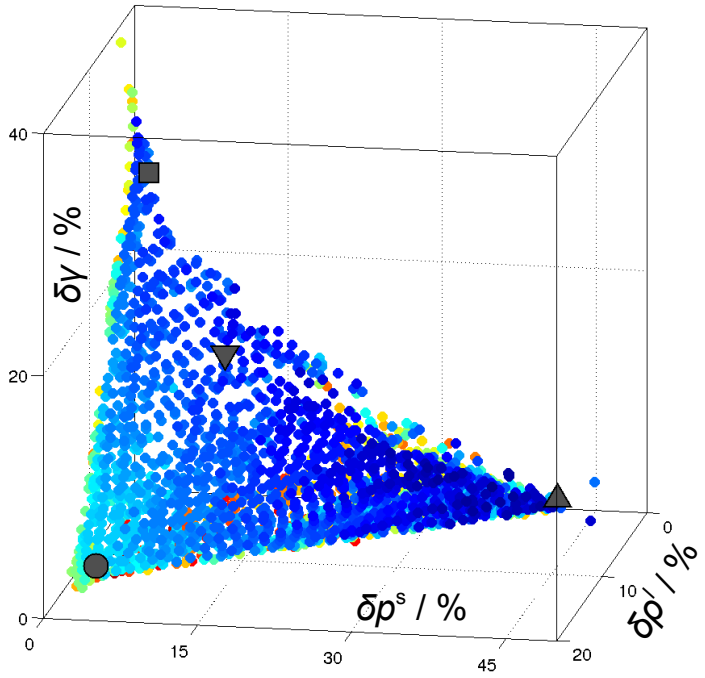
+ 22 %

Optimization with multiple objectives

Pareto optimality criterion



Multiple objectives

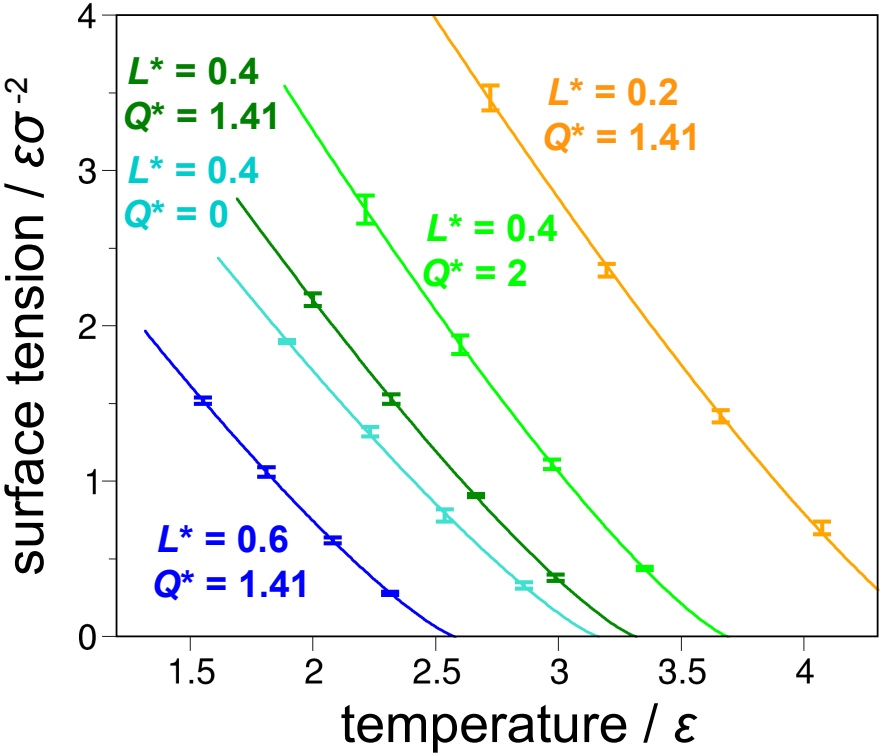


(2CLJQ for carbon dioxide)

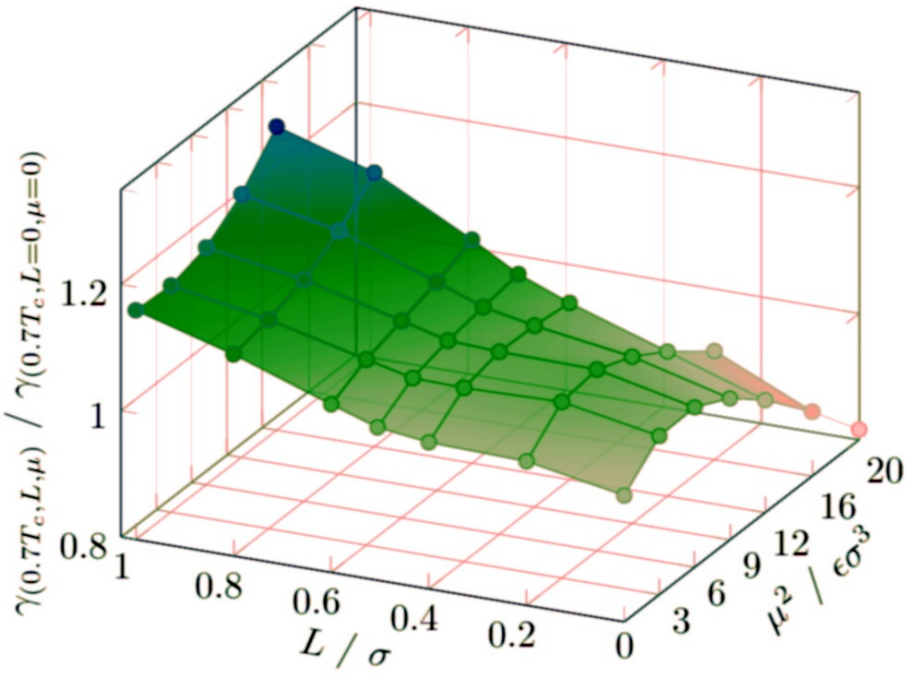
Multicriteria optimization requires characterizing the whole model class.

Surface tension of 2CLJQ and 2CLJD fluids

Two LJ + quadrupole (2CLJQ)



Two LJ + dipole (2CLJD)



- Systematic exploration of the four-dimensional model parameter space
- Correlation of γ by critical scaling expressions (2CLJQ, 2CLJD, Mie-6)

Computation of the Pareto set[1, 2]

Multicriteria optimization problem

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

$$f_\xi = \left\langle \delta \xi^2 \right\rangle_{0.55 T_c^{exp} < T < 0.95 T_c^{exp}} = \lim_{N \to \infty} \frac{1}{N+1} \sum_{i=0}^{N} \left(1 - \frac{\xi^{sim}(T)}{\xi^{exp}(T)} \right)^2_{T/T_c = 0.55 + 0.4 i/N} \quad \text{(here: } N = 9\text{)}.$$

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.

[1]M. Bortz *et al.*, *Comput. Chem. Eng.* 60, 354, **2014**; [2]Stöbener *et al.*, *Fluid Phase Equilib.* 411, 33, **2016**.

Computation of the Pareto set[1, 2]

a model parameters

(here, $a = 4$)

- LJ size parameter σ
- LJ energy parameter ε
- Model elongation L
- Quadrupole moment Q

Dimension of Pareto set $d \leq a$.

b optimization criteria

(here, $b = 3$)

- Saturated liquid density ρ'
- Saturated vapor pressure p^s
- Vapor-liquid surface tension γ

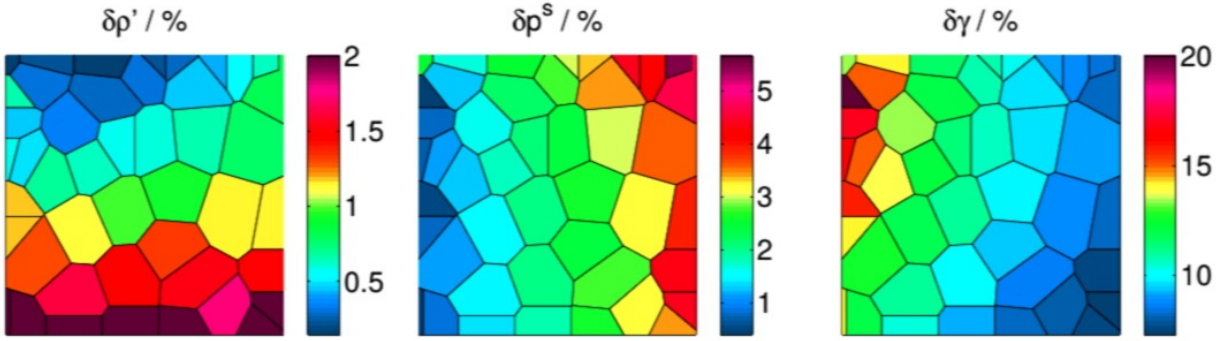
Dimension of Pareto set $d \leq b - 1$.

$$\boldsymbol{d = \min(a, b - 1).}$$ (here, $d = 2$)

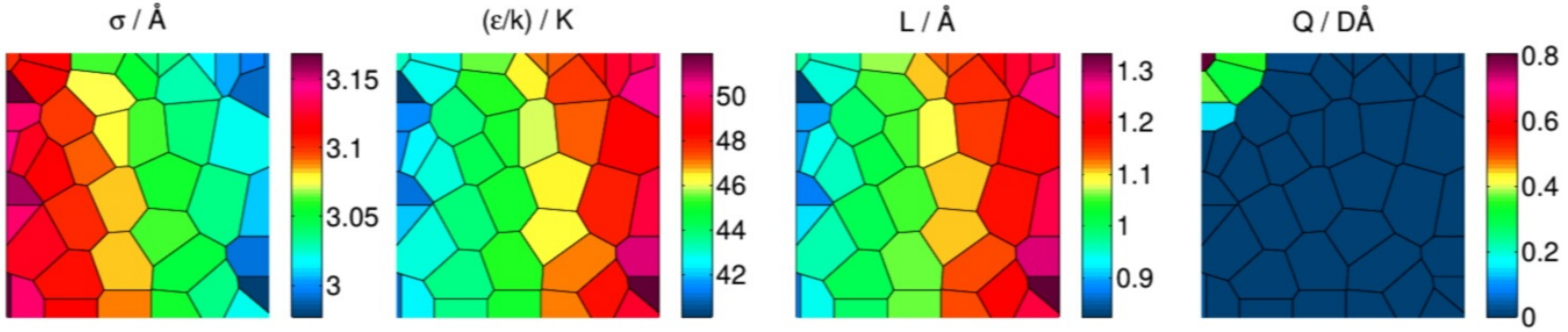
[1]M. Bortz *et al.*, *Comput. Chem. Eng.* 60, 354, **2014**; [2]Stöbener *et al.*, *Fluid Phase Equilib.* 411, 33, **2016**.

Multicriteria molecular model optimization[1, 2]

Representation of objective and parameter spaces by patch plots:



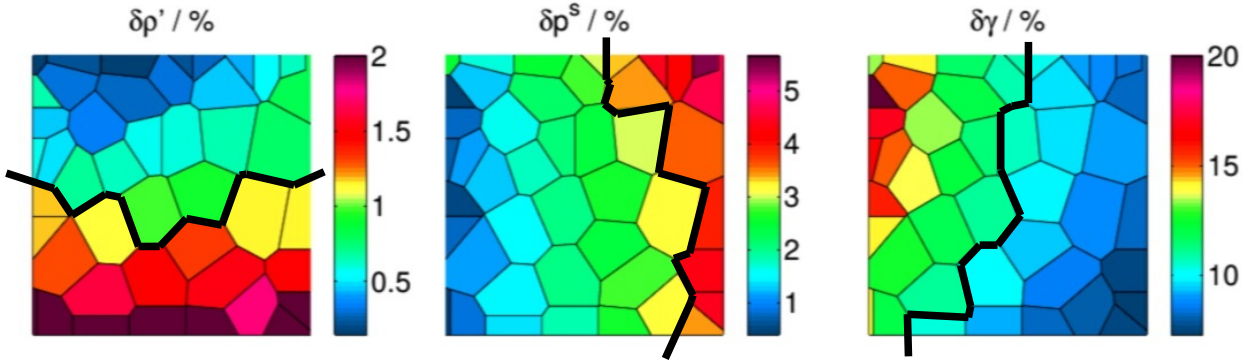
Pareto-optimal 2CLJQ models of molecular oxygen



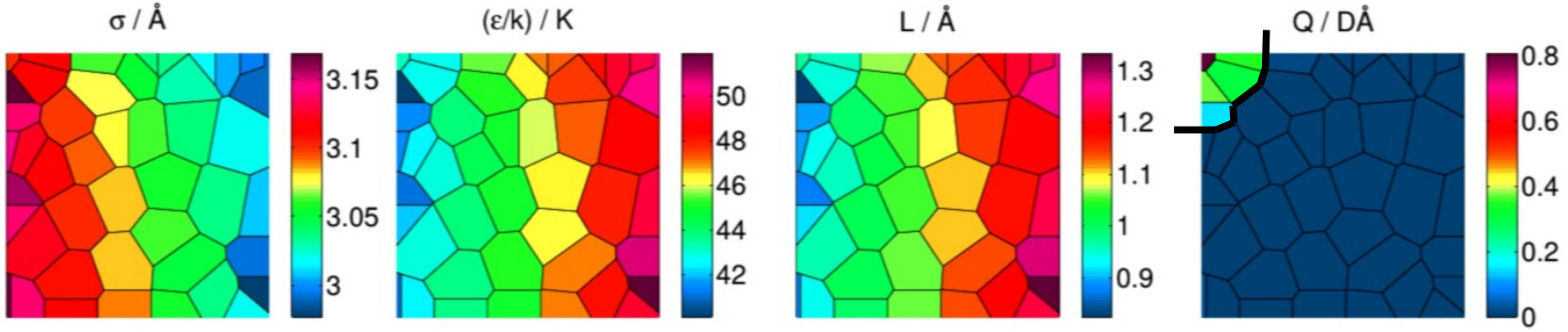
[1]Stöbener *et al.*, *Fluid Phase Equilib.* 373, 100, **2014**; [2]Stöbener *et al.*, *Fluid Phase Equilib.* 408, 141, **2016**.

Multicriteria molecular model optimization[1, 2]

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



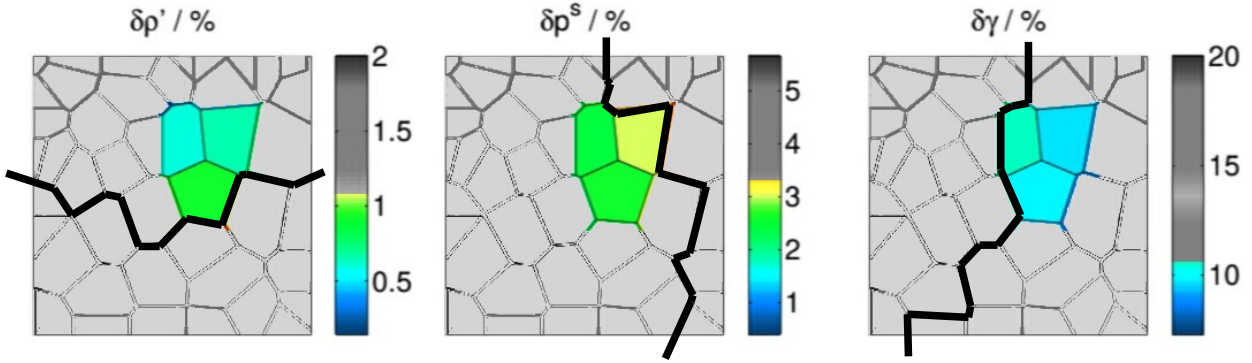
Restrictions imposed on 2CLJQ models of molecular oxygen



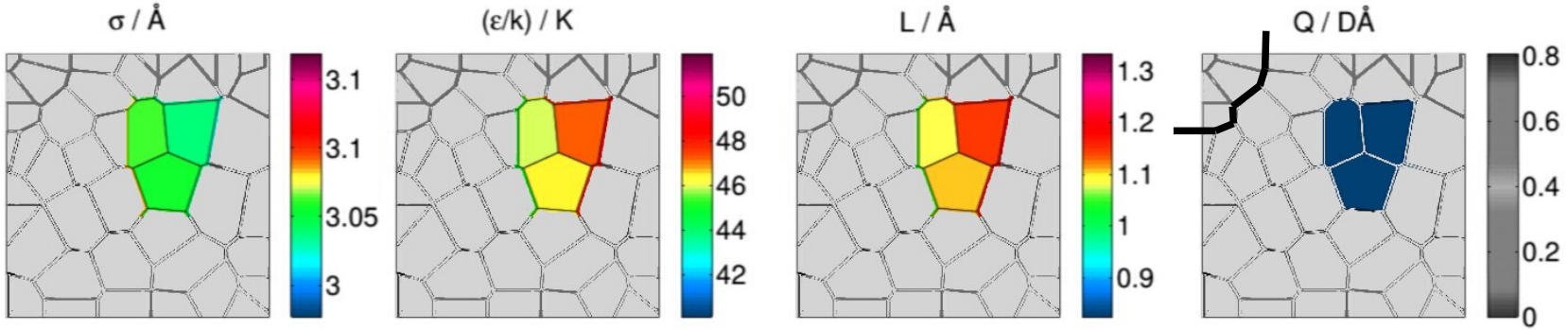
[1]Stöbener *et al.*, *Fluid Phase Equilib.* 373, 100, **2014**; [2]Stöbener *et al.*, *Fluid Phase Equilib.* 408, 141, **2016**.

Multicriteria molecular model optimization[1, 2]

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

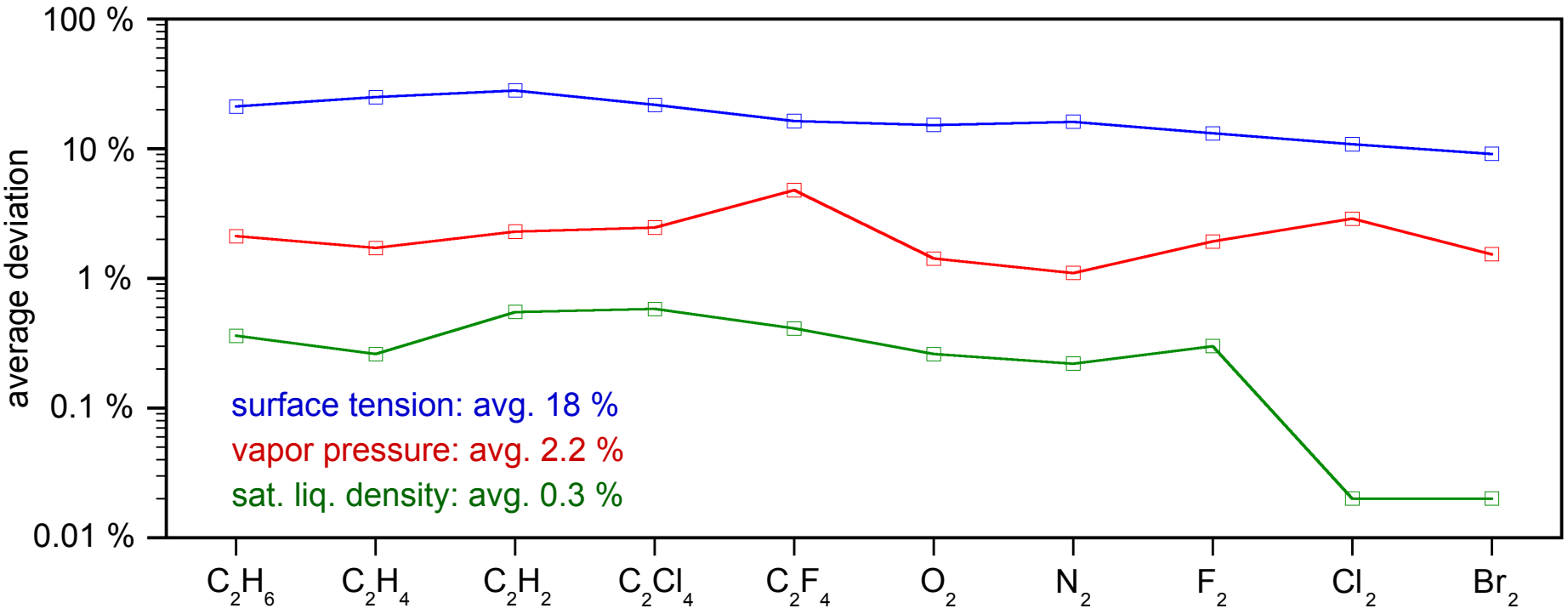


2CLJ models of molecular oxygen fulfilling all requirements



[1]Stöbener *et al.*, *Fluid Phase Equilib.* 373, 100, **2014**; [2]Stöbener *et al.*, *Fluid Phase Equilib.* 408, 141, **2016**.

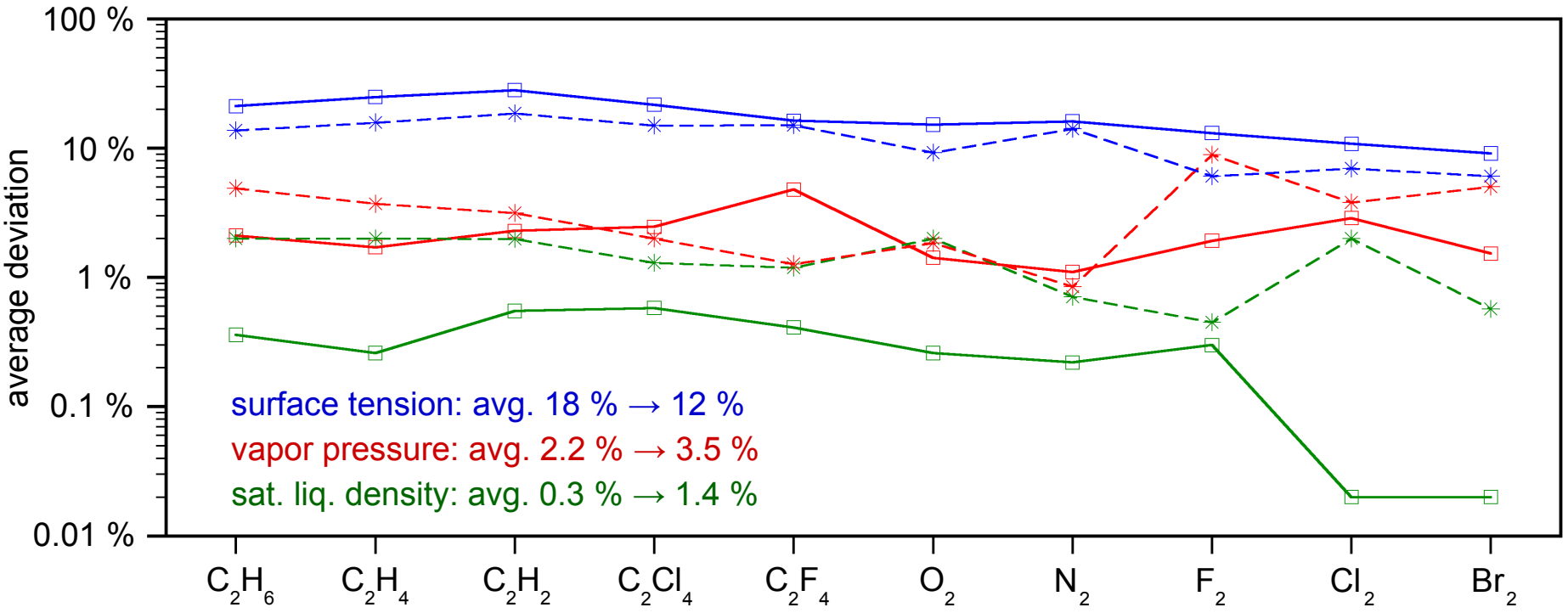
Model accuracy for ten quadrupolar fluids[1, 2]



[1]J. Vrabec, J. Stoll, H. Hasse, *J. Phys. Chem. B* 105(48), 12126–12133, **2001**;

[2]S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse, *Chem. Eng. Sci.* 121, 110–117, **2015**.

Model accuracy for ten quadrupolar fluids[1–3]



[1]J. Vrabec, J. Stoll, H. Hasse, *J. Phys. Chem. B* 105(48), 12126–12133, **2001**;

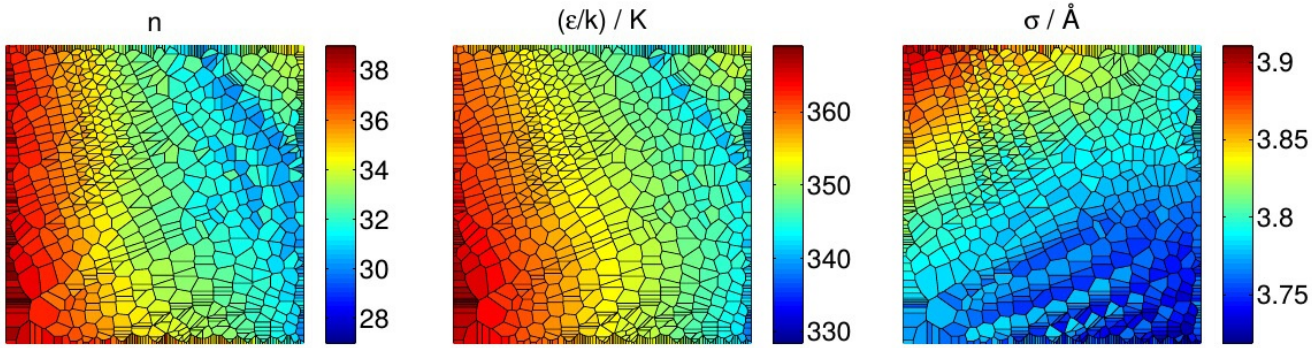
[2]S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse, *Chem. Eng. Sci.* 121, 110–117, **2015**;

[3]K. Stöbener, P. Klein, M. Horsch, K.-H. Küfer, H. Hasse, *Fluid Phase Equilib.* 411, 33–42, **2016**.

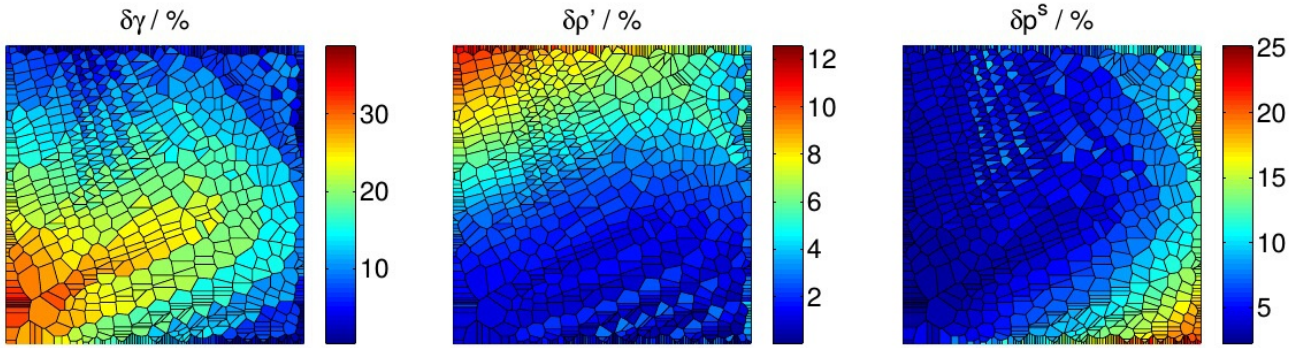
Comparison between model classes[1]

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

parameter space



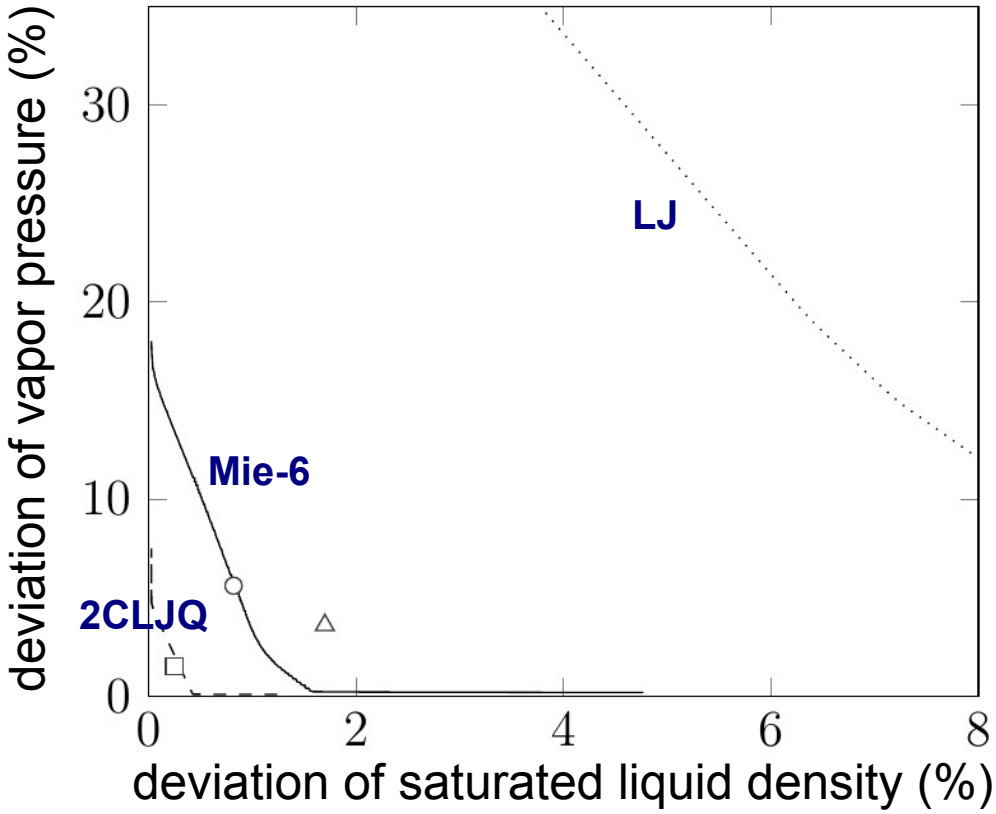
objective space



[1]S. Werth, K. Stöbener, M. Horsch, H. Hasse, *Mol. Phys.* 115(9–12), 1017–1030, **2017**.

Comparison between model classes[1]

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



LJ potential (2 parameters)

Mie–*n*,6 potential (3 parameters)

2CLJQ potential (4 parameters)

□ Vrabec *et al.* (2001) 2CLJQ model

△ Avendaño *et al.* (2011) Mie–*n*,*m* model

[1]S. Werth, K. Stöbener, M. Horsch, H. Hasse, *Mol. Phys.* 115(9–12), 1017–1030, **2017**.

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**.