Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation

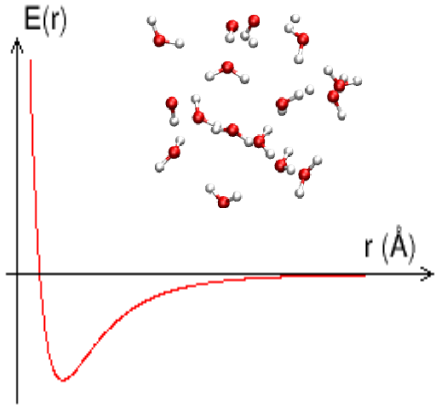
Fabien Cailliez, Pascal Pernot
Laboratoire de Chimie Physique, CNRS UMR8000

Molecular simulations and forcefields

- Molecular simulations:
  - Sampling of representative configurations of the system (MD or MC)
  - Computation of macroscopic properties (laws of statistical physics)

- Forcefield: mathematical expression of the interatomic potential as a function of the nuclei positions

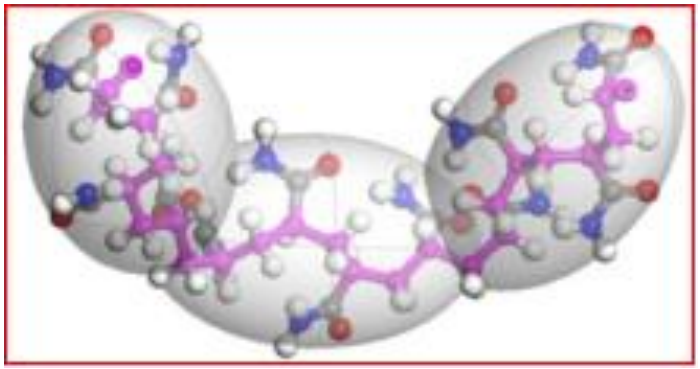
$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r} + \cdots$$

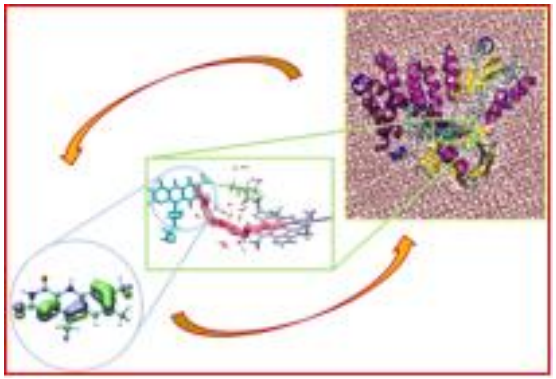


- Forcefield parameters:
  - Physical interpretation
  - Calibrated to reproduce some reference (theoretical or experimental) data

Uncertainties in molecular simulations

- Importance of monitoring uncertainties in molecular simulations:
  - Molecular simulation used as a decision tool (industrial applications)
    → Confidence interval for the prediction needed
  - Multi-scale simulations
    → Transfer of uncertainties along the different scales





Uncertainties and errors in molecular simulations

- Numerical uncertainties:
  - Limited sampling of the configuration space
  - Use of efficient but “approximate” algorithms
  - Can be monitored and reduced « easily »

- Parametric uncertainties:
  - Forcefield parameters calibrated over uncertain experimental data
  - How to evaluate them? → statistical calibration
  - What is their impact on the computed properties? → uncertainty propagation

- Modeling errors:
  - Forcefields may be unable to reproduce some reference data
  - Can we correct them?
  - Transferability issue for the correction of model inadequacy?

Brief survey of the litterature

- 2008: Cooke & Schmidler, Biophysical Journal, 95: 4497–4511
  Cailbration of dielectric constant in peptides to better reproduce helical folding of peptides
- 2011: Cailliez & Pernot,, J. Chem. Phys. 134, 054124
  Bayesian calibration of a LJ forcefield for Argon and UP in MD simaultions
- 2012: Angelikopoulos *et al.*, J. Chem. Phys. 137, 144103 (2012)
  Bayesian calibration of a LJ forcefield for Argon using GP surrogate models
- 2012: Rizzi *et al.*, Multiscale model. Simul. 10(4):1428–1492
  Estimation of forcefield parameters of a water model an UP in MD simulations using Polynomial Chaos
- 2013: Rizzi *et al.*, J. Chem. Phys. 138, 194105
  Statistical calibration of LJ parameters of monoatomic ions
- 2014: Cailliez *et al.*, J. Comp. Chem., 35, 130–149
  Estimation of forcefield parameters of a water model and UP in MD simulations using GP surrogate models
- 2016: Wu *et al.*, Phil. Trans. R. Soc. A 374: 20150032
  Hierarchical modeling to calibrate LJ parameters among heterogeneous data
- 2017 : Pernot & Cailliez, AIChE J:63(10): 4642-4665
  Review of statistical calibration/prediction models handling data inconsistency and model inadequacy

Outline of the presentation

- Bayesian calibration framework
- Statistical calibration of a forcefield for Argon:
  - Comparing numerical and parametric uncertainties
  - Transferability issues
- Calibration of a water forcefield:
  - How to decrease the computational burden?
  - Use of surrogate models and Efficient Global Optimization strategies
- How to deal with model inadequacy?
- Conclusions

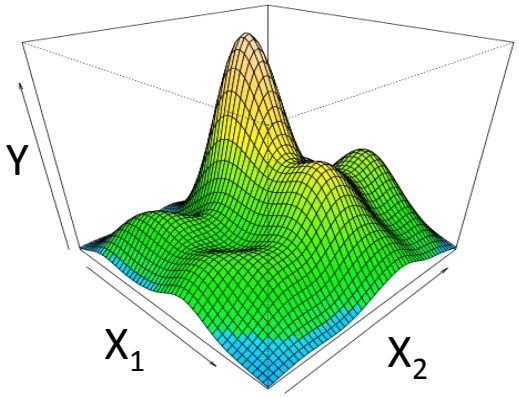
Outline of the presentation

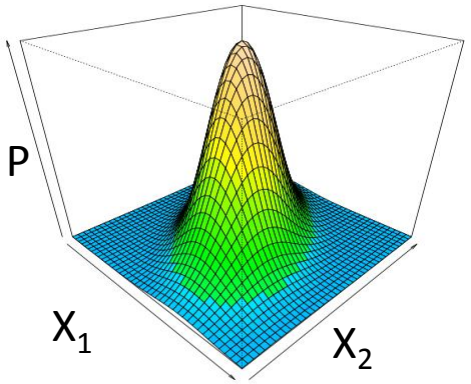
- Bayesian calibration framework
- Statistical calibration of a forcefield for Argon:
  - Comparing numerical and parametric uncertainties
  - Transferability issues
- Calibration of a water forcefield:
  - How to decrease the computational burden?
  - Use of surrogate models and Efficient Global Optimization strategies
- How to deal with model inadequacy?
- Conclusions

Statistical calibration and uncertainty propagation

Calibration data:
$Y = Y(X_1, X_2)$

PDF:
$P(\Omega|\{Y_{i,iexp}\})$





$\theta = \{X_1, X_2\}$

Bayesian calibration

$$P(\theta|\{Y_{i,exp}\}) \propto \underbrace{P(\{Y_{i,exp}\}|\theta)}_{\text{Likelihood}} \times \underbrace{P(\theta)}_{\text{Prior}}$$

Independant measurements and Gaussian hypothesis for residuals:

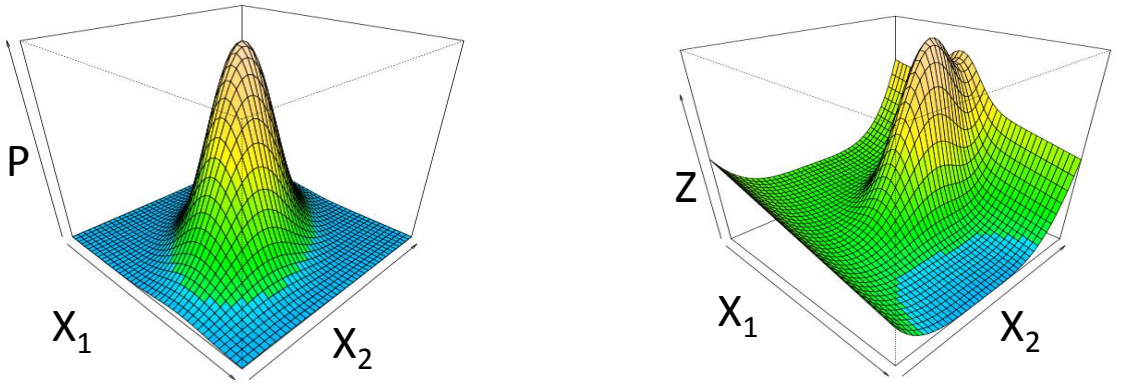
$$P(\{Y_{i,exp}\}|\theta) = \prod_i \left(\frac{1}{u_i \sqrt{2\pi}} \exp\left(-\frac{(Y_i - Y_{i,exp})^2}{2u_i^2} \right) \right)$$

$$u_i^2 = u_{i,exp}^2 + u_{i,mod}^2$$

Statistical calibration and uncertainty propagation

PDF:
$P(\theta|\{Y_{i,iexp}\})$

Results of a simulation:
$Z = Z(X_1, X_2)$



Uncertainty propagation

Monte Carlo sampling of the PDF and estimation of the model:

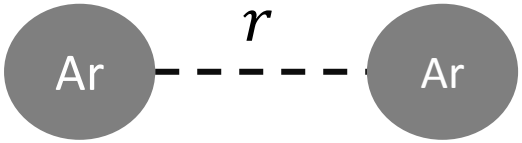
$$\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z(\theta_i) \qquad u_Z^2 = var(Z_i)$$

Outline of the presentation

- Bayesian calibration framework
- Statistical calibration of a forcefield for Argon:
  - Comparing numerical and parametric uncertainties
  - Transferability issues
- Calibration of a water forcefield:
  - How to decrease the computational burden?
  - Use of surrogate models and Efficient Global Optimization strategies
- How to deal with model inadequacy?
- Conclusions

A simple test-case: forcefield for Argon

- Two-parameters Lennard-Jones forcefield for Argon:



$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

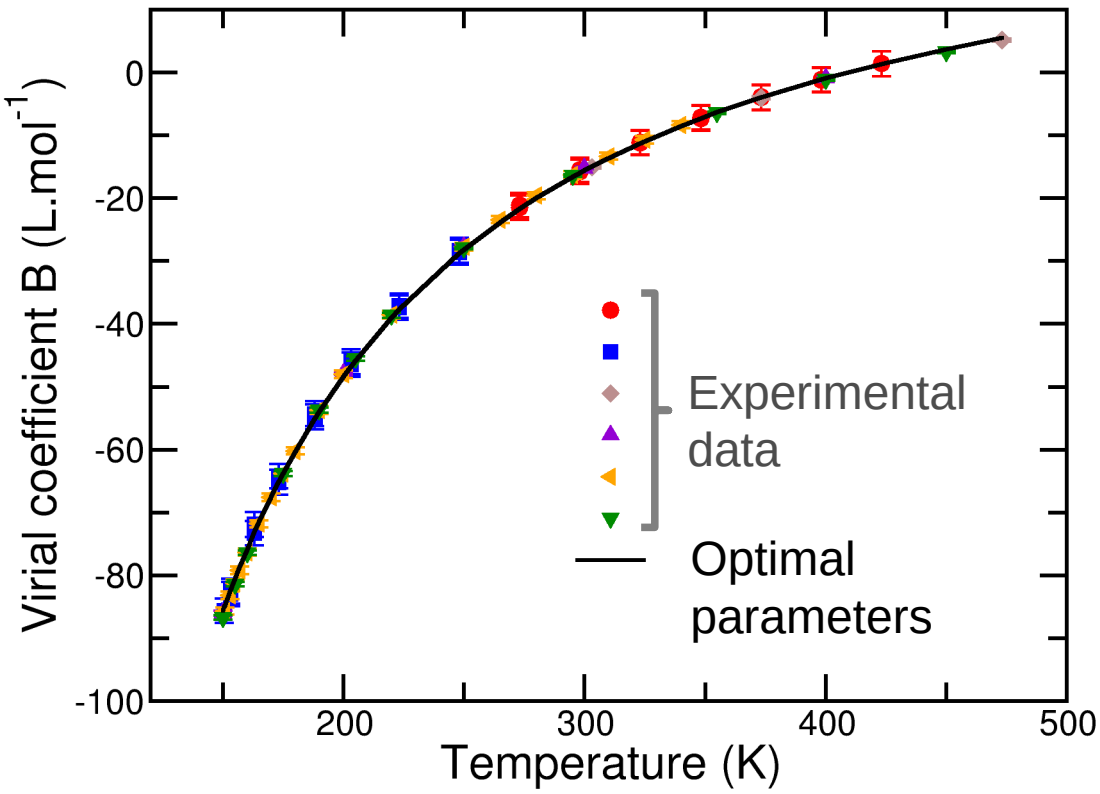
- Statistical calibration:
  - *Uniform prior*
  - Experimental data for calibration:

    2nd virial coefficient B from 150K to 450K

    (lit.: $\sigma = 3.405\text{Å}$; $\varepsilon = 119\text{K}$)

- Specificities of this calibration:
  - Only two parameters
  - Analytical expression linking B to the parameters σ and ε
    → Analytical PDFs and $u_{i,mod} = 0$

Calibration on 2nd virial coefficient

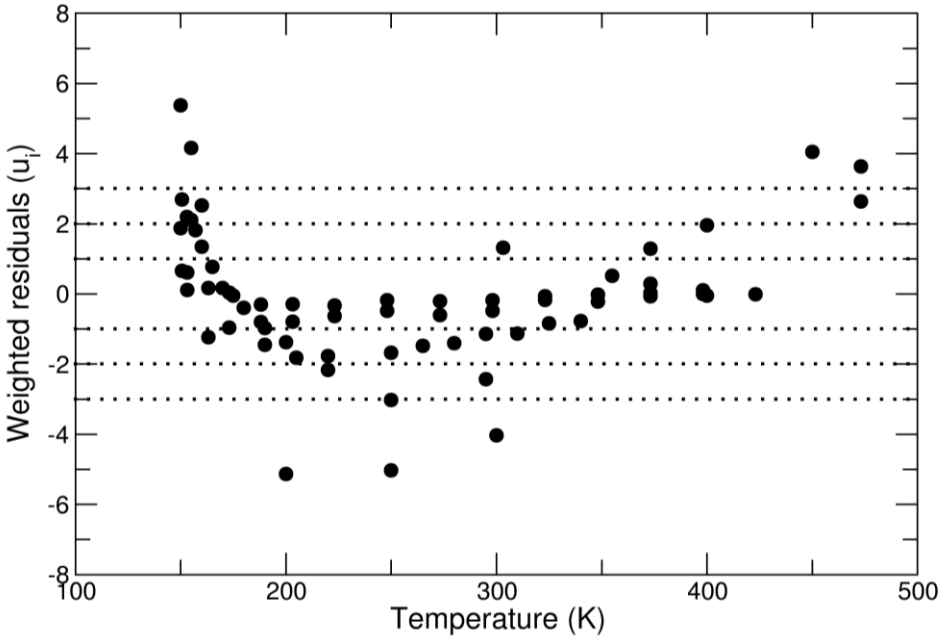


$$\sigma = 3.422 \pm 0.002 \text{ Å}$$
$$\varepsilon = 119.61 \pm 0.09 \text{ K}$$

(lit.: $\sigma = 3.405\text{Å}$; $\varepsilon = 119\text{K}$)

- Various sources of measurements
- At first sight: successfull calibration and small uncertainties

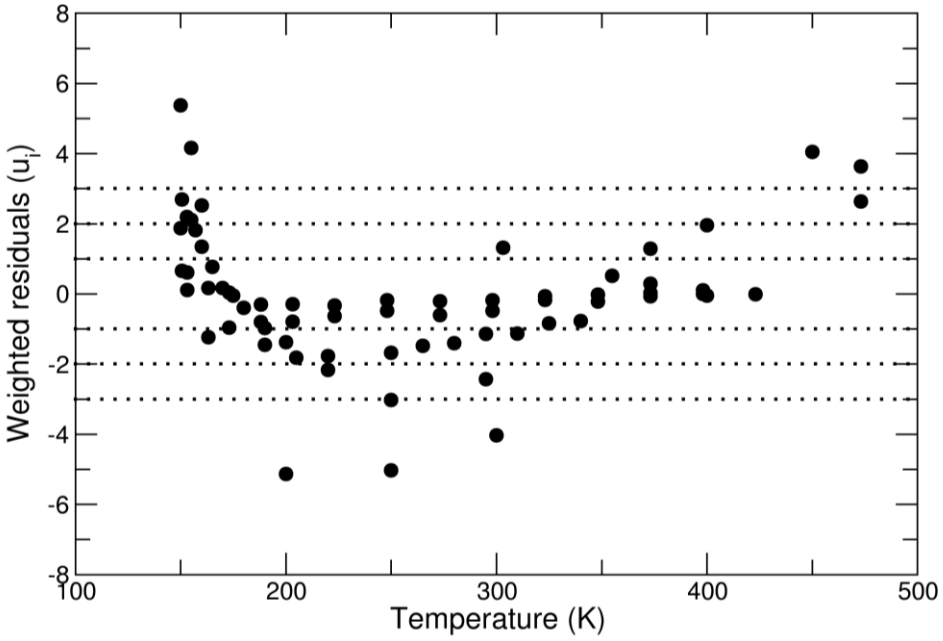
Calibration on 2nd virial coefficient



$$w_i = \frac{B(\bar{\sigma}, \bar{\varepsilon}, T_i) - B_{i,exp}}{u_i}$$

- Gaussian hypothesis for the residues violated
- Possible origins of the problem:
  - Inadequacy of the model
  - Inconsistency between some experimental data
  - Underestimated experimental uncertainties

Calibration on 2nd virial coefficient



$$w_i = \frac{B(\bar{\sigma}, \bar{\varepsilon}, T_i) - B_{i,exp}}{u_i}$$

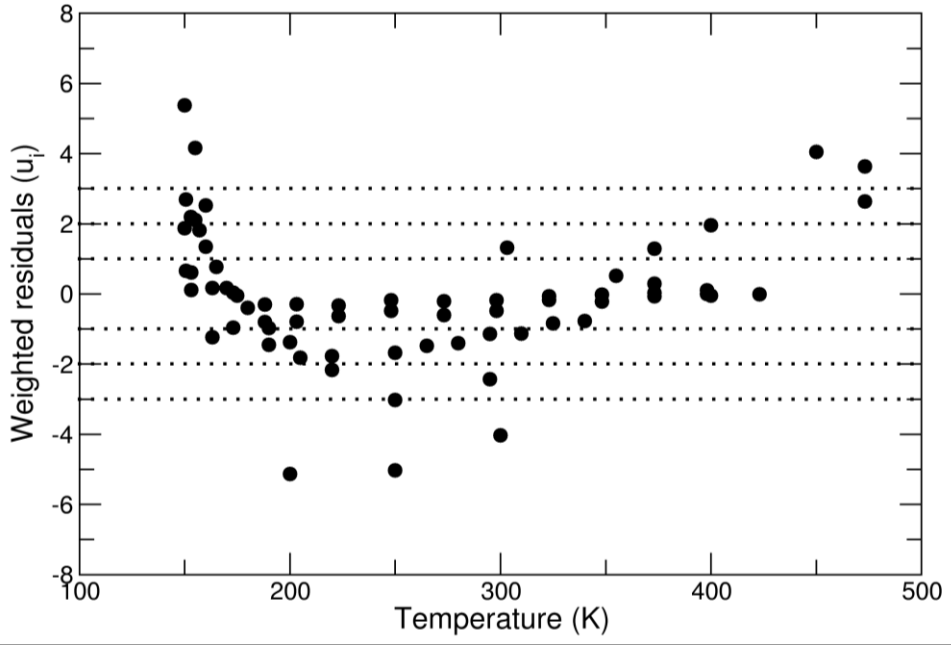
- Gaussian hypothesis for the residues violated
- Possible origins of the problem:
  - Inadequacy of the model $u_i'^2 = u_{i,exp}^2 + u_{i,m}^2$
  - Inconsistency between some experimental data
  - Underestimated experimental uncertainties

Additive uncertainties on model predictions

$u_m = \mathcal{N}(0, s^2)$

⟶ not transferable for prediction to other quantities

Calibration on 2nd virial coefficient



$$w_i = \frac{B(\bar{\sigma}, \bar{\varepsilon}, T_i) - B_{i,exp}}{u_i}$$

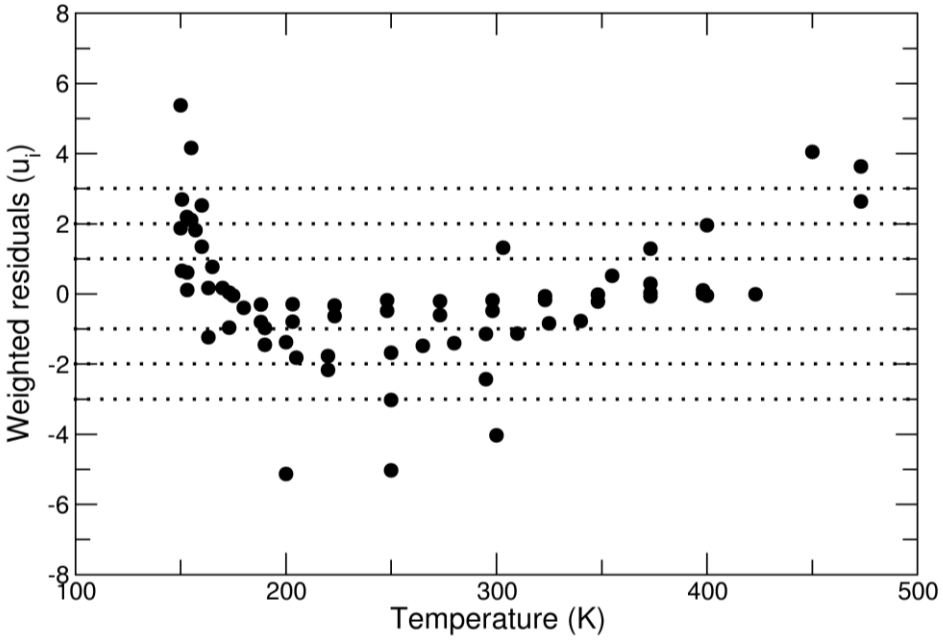
Fixed Laboratory Effects Model (FLEM):

- Each experimental set bears an unknown but constant bias.
- Modification of the data by adding a term (to be calibrated) that depends on the experimental data set.

$$B'_{i,exp}(\sigma, \varepsilon) = B_{i,exp}(\sigma, \varepsilon) + \lambda_S(i)$$

- Gaussian hypothesis for the residues violated
- Possible origins of the problem:
  - Inadequacy of the model
  - Inconsistency between some experimental data
  - Underestimated experimental uncertainties

Calibration on 2nd virial coefficient



$$w_i = \frac{B(\bar{\sigma}, \bar{\varepsilon}, T_i) - B_{i,exp}}{u_i}$$

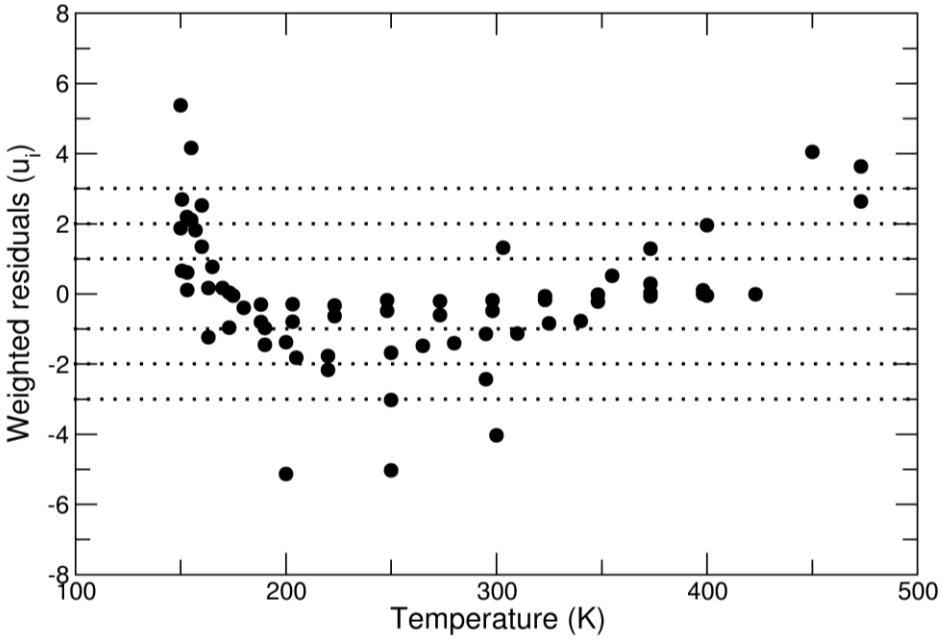
Random Laboratory Effects Model (RLEM):
Each experimental data bears a supplementary unknown bias

$$u_{\text{add}} = \mathcal{N}(0, s^2)$$

- Gaussian hypothesis for the residues violated
- Possible origins of the problem:
  - Inadequacy of the model
  - Inconsistency between some experimental data
  - Underestimated experimental uncertainties

$$u_i'^2 = u_{i,exp}^2 + u_{i,\text{add}}^2$$

Calibration on 2nd virial coefficient

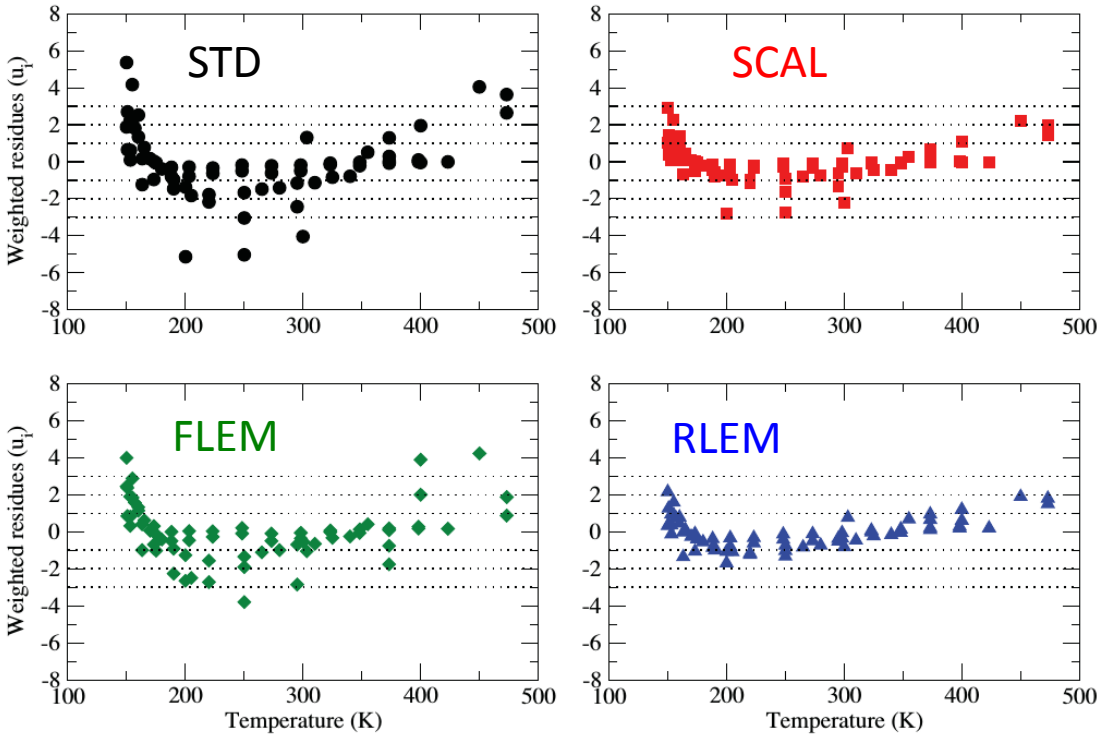


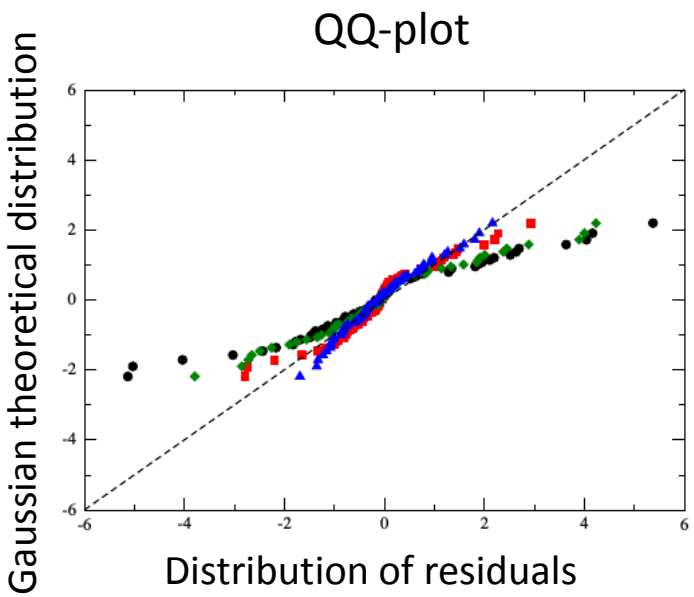
$$w_i = \frac{B(\bar{\sigma}, \bar{\varepsilon}, T_i) - B_{i,exp}}{u_i}$$

- Gaussian hypothesis for the residues violated
- Possible origins of the problem:
  - Inadequacy of the model
  - Inconsistency between some experimental data
  - Underestimated experimental uncertainties: $u'_i = s \times u_{i,exp}$

SCAL:
Scaling of experimental uncertainties by an a priori unknown factor

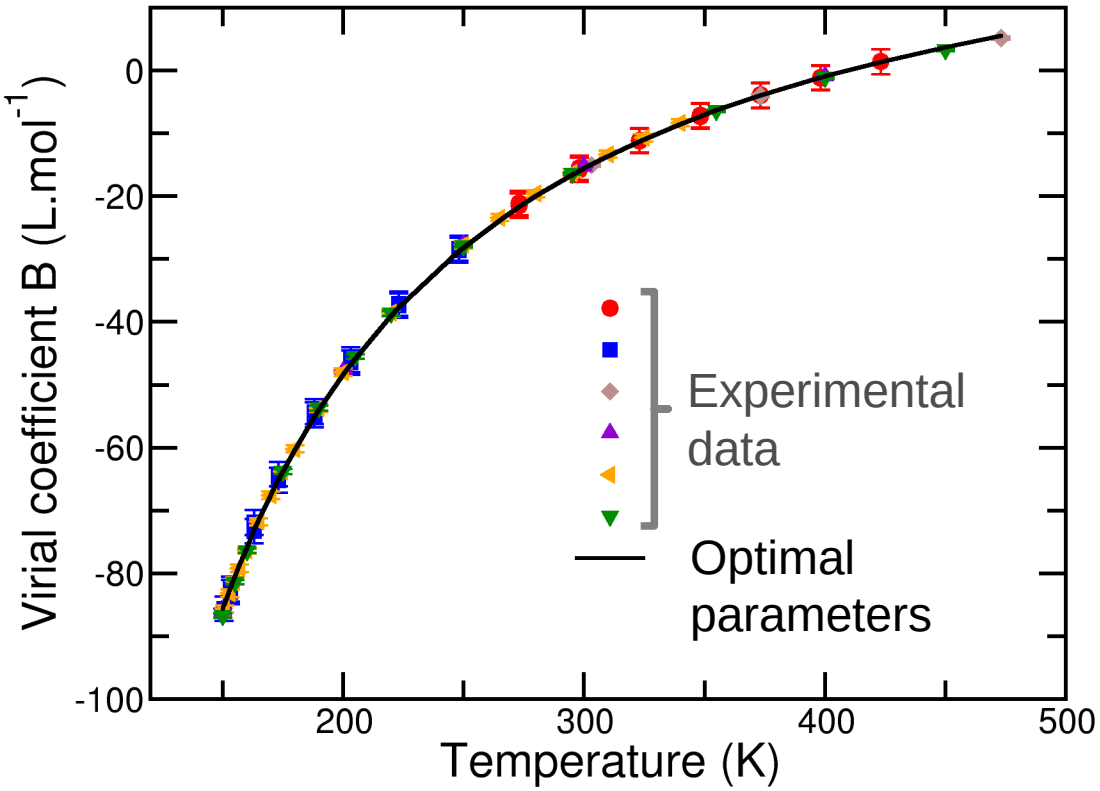
Various calibration schemes





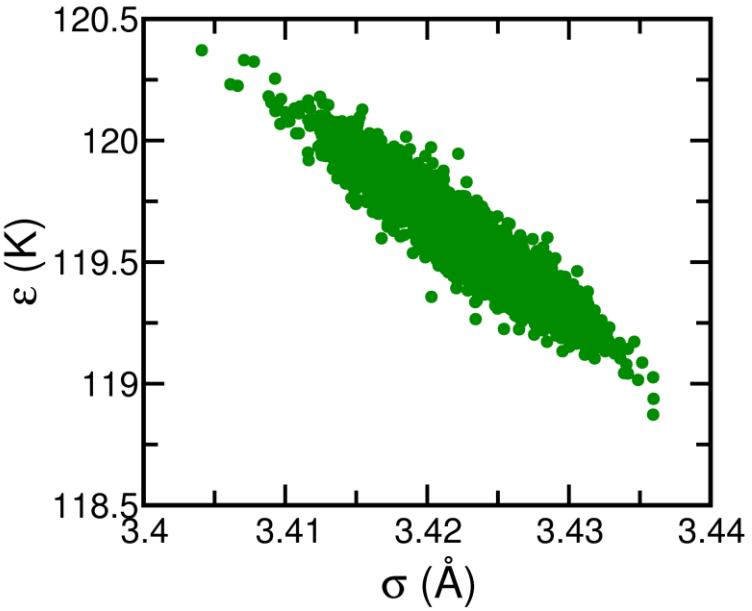
- SCAL and RLEM models achieve statistical consistency in the residuals
- Long-range correlation remains at different temperatures (limitation of the LJ interaction model)

Results of the SCAL calibration



- Small scaling factor for uncertainties (<2)
- Small uncertainties on the parameters
- Strong correlation between σ and ε

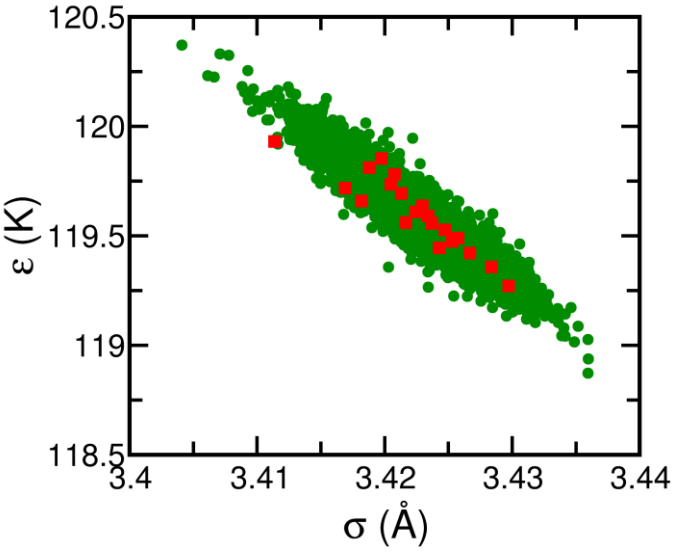
Markov chain over the PDF

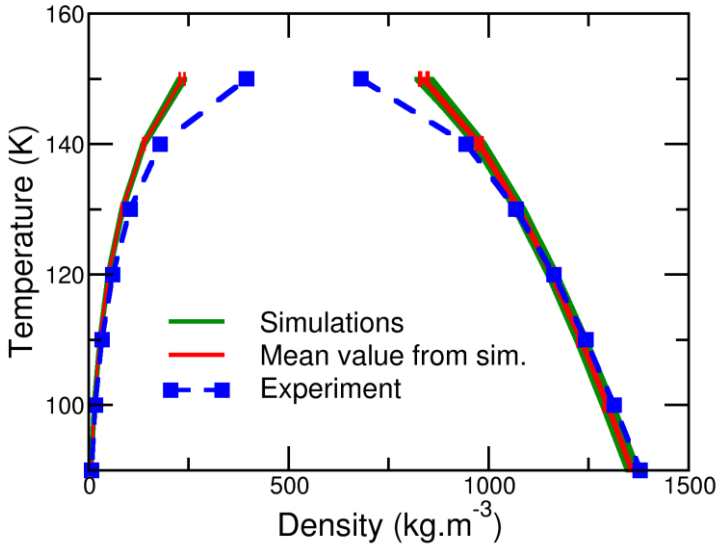


σ = 3.422 ± 0.004 Å
ε = 119.61 ± 0.18 K

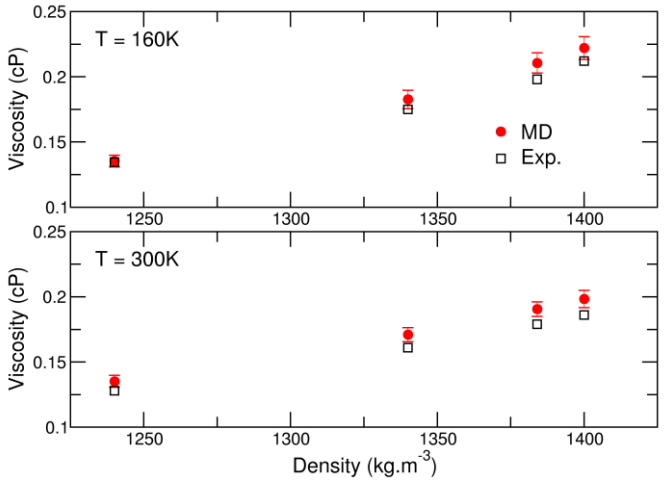
Values from literature:
σ = 3.405 Å, ε = 119.8 K

Uncertainty propagation in molecular simulation

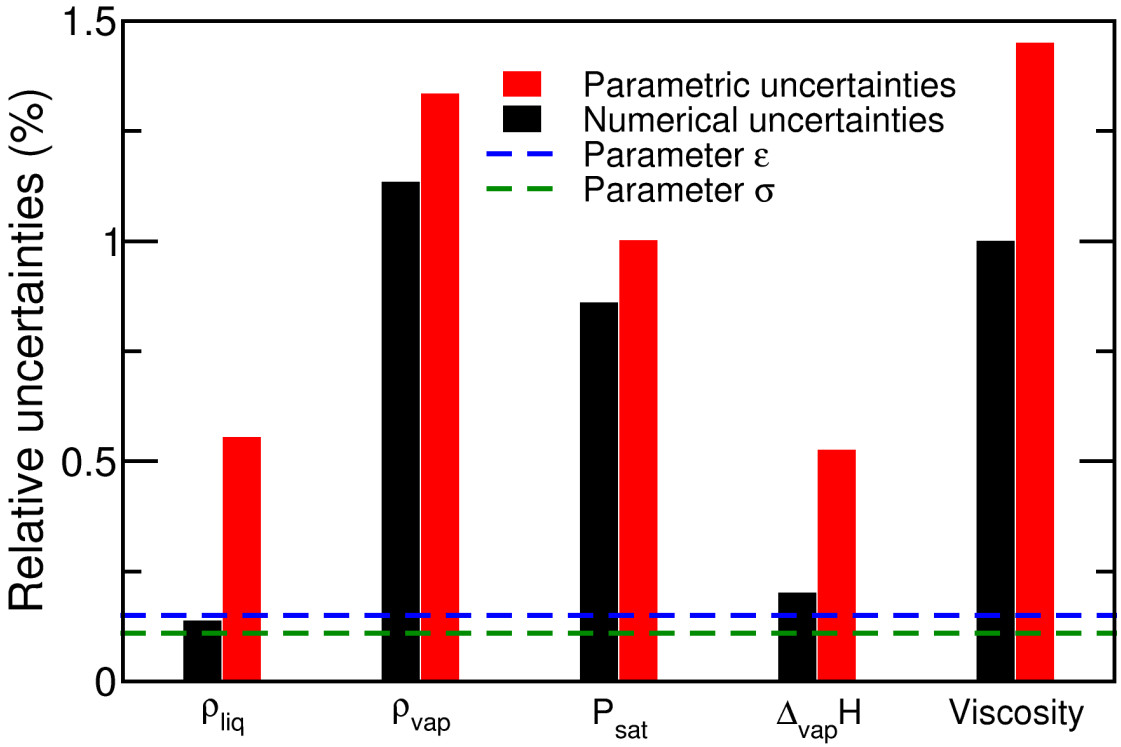




- LHS sample of the PDF of (σ, ε)
- Computation of L/V phase diagrams and liquid viscosities
- Parametric uncertainties remain small and do not allow to reconcile computed and experimental values



Numerical/parametric uncertainties



- Parameters uncertainties amplified by molecular simulation
- Parametric uncertainties bigger than numerical uncertainties

A first conclusion

- What we have learnt from this LJ system:
  - An operative methodology for statistical calibration and uncertainty propagation
  - Parametric uncertainties small but greater than numerical uncertainties
  - Taking into account parametric uncertainty is not sufficient to have quantitative transferability to other properties

Cailliez and Pernot, *J. Chem. Phys.* **134**, 054124 (2011)

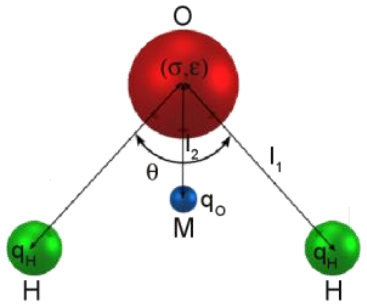
- What to do next:
  - Increase the complexity of the forcefield?
  - Is the method tractable when calibration data requires molecular simulation to be evaluated?
  - How to deal with model inadequacy?

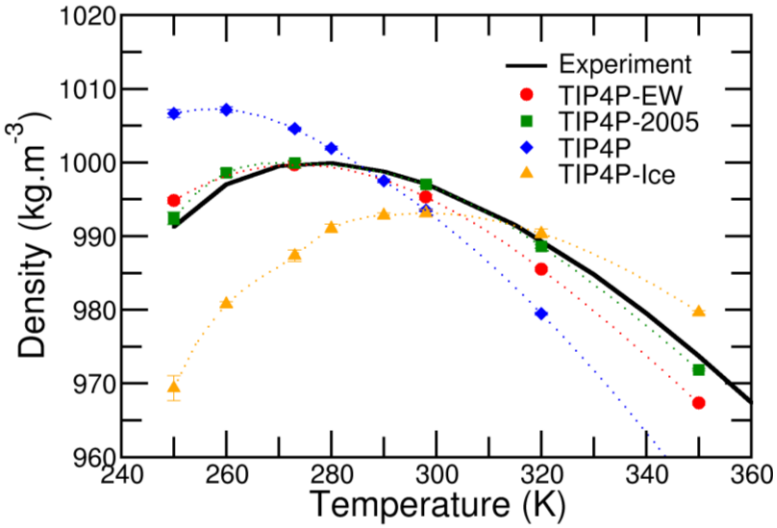
Outline of the presentation

- Bayesian calibration framework
- Statistical calibration of a forcefield for Argon:
  - Comparing numerical and parametric uncertainties
  - Transferability issues
- Calibration of a water forcefield:
  - How to decrease the computational burden?
  - Use of surrogate models and Efficient Global Optimization strategies
- How to deal with model inadequacy?
- Conclusions

TIP4P water forcefield: strategy of calibration

- TIP4P water forcefield:
  - Various parameter sets available in the literature
  - 4 parameters: $\sigma, \varepsilon, q_H, l_2$
  - Calibration data: liquid water density at 5 temperatures from 253K to 350K
  - Molecular simulations needed to compute the calibration data





- Strategy of calibration:
  - Use of surrogate models (GP) to avoid the use of expensive molecular simulations

Which surrogate?

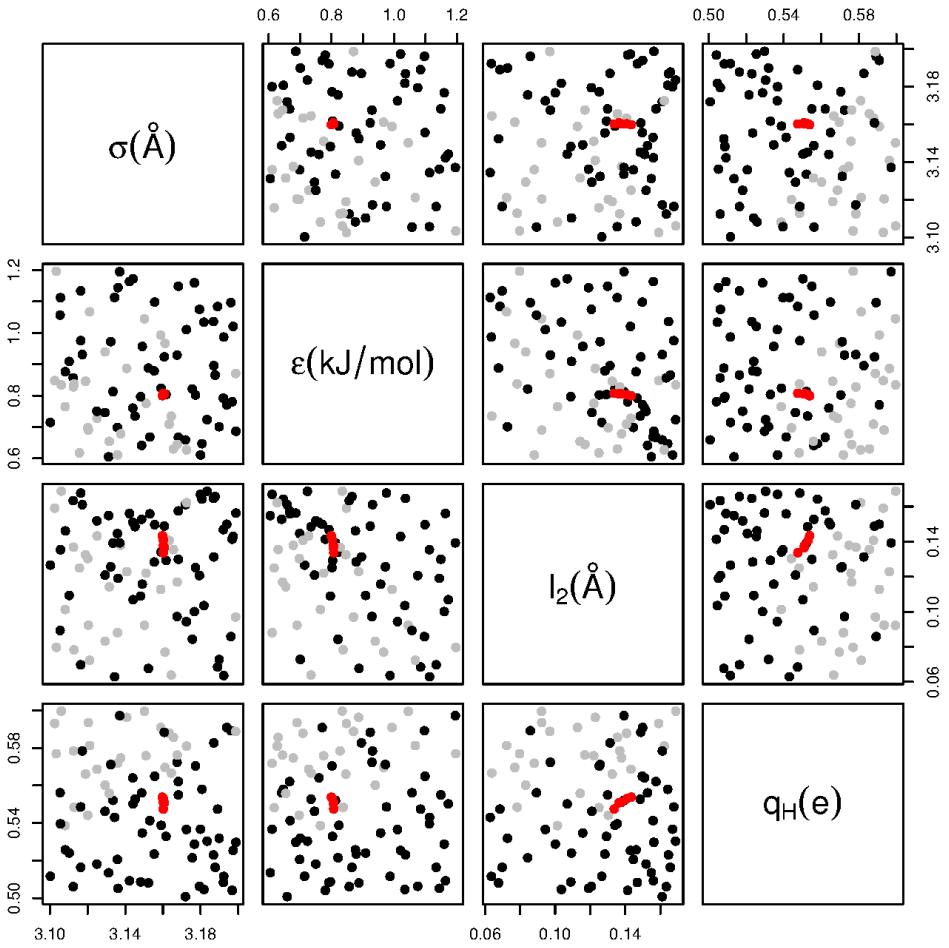
- Function to be optimized in the calibration process:

$$F(\theta) = \sum_{i=1}^{5} \frac{\left(Y_i(\theta) - Y_{i,exp}\right)^2}{u^2}$$

Representative simulation uncertainty

- Option 1: Use a GP as a « direct » surrogate model for F
  - ++ : efficient optimization strategies using GP surrogate models (EGO)
  - - - : F is a quickly-varying function of θ with large amplitude
- Option 2: Use GP models for each observable Y_i and then a composite surrogate model for F
  - ++ : Y_i are smooth and rather monotonous functions of θ
  - - - : The surrogate model for F is not a GP

Building of the surrogate models



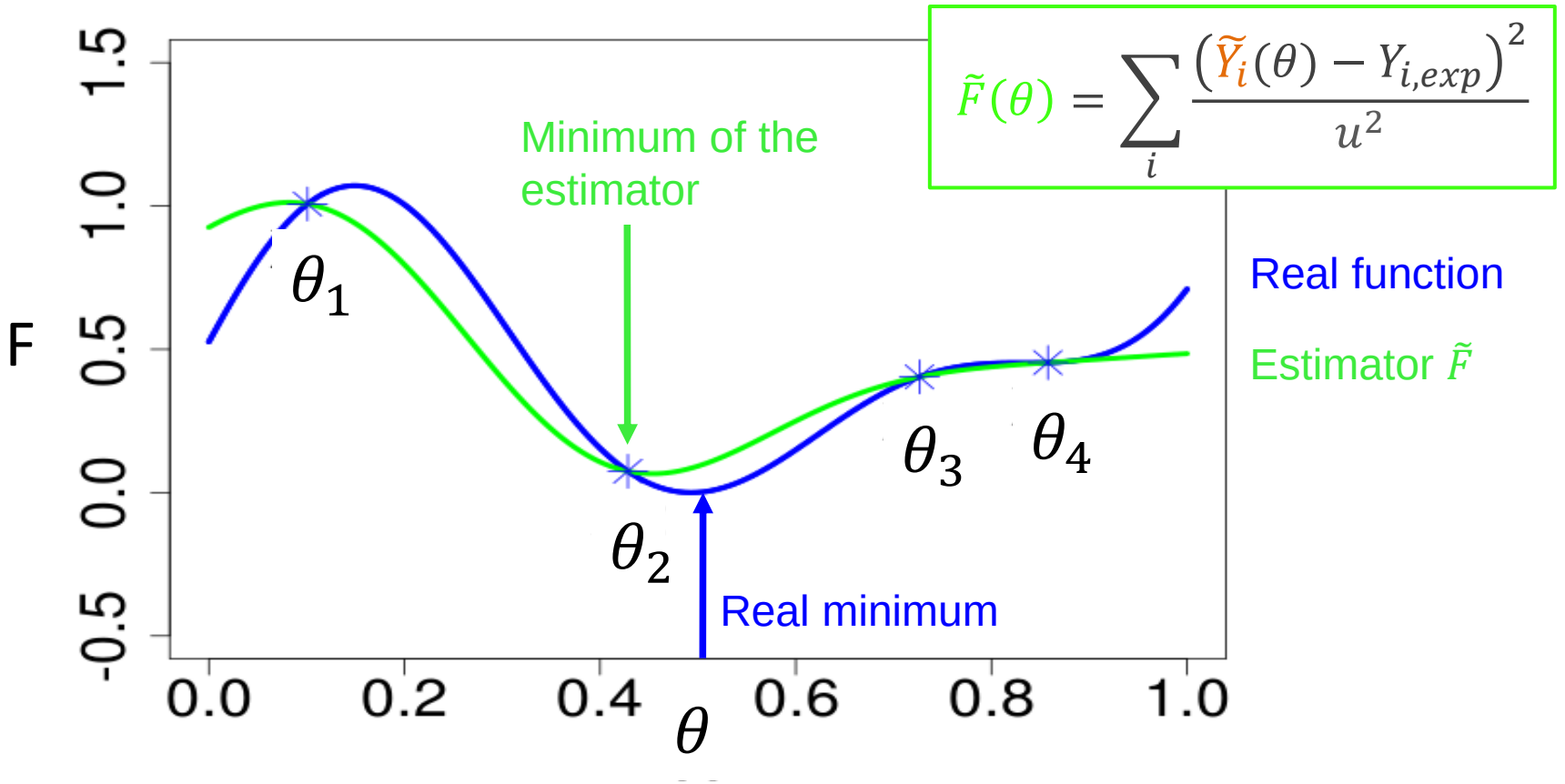
D1: black & grey dots D2: black dots

- Initial sampling of the parameter space:
  - Maximin LHS
  - 84 parameter sets (D1)
- Surrogate models:
  - One for each property
  - Leave-one-out predictivity coefficients:

$$Q_2 \geq 92\%$$

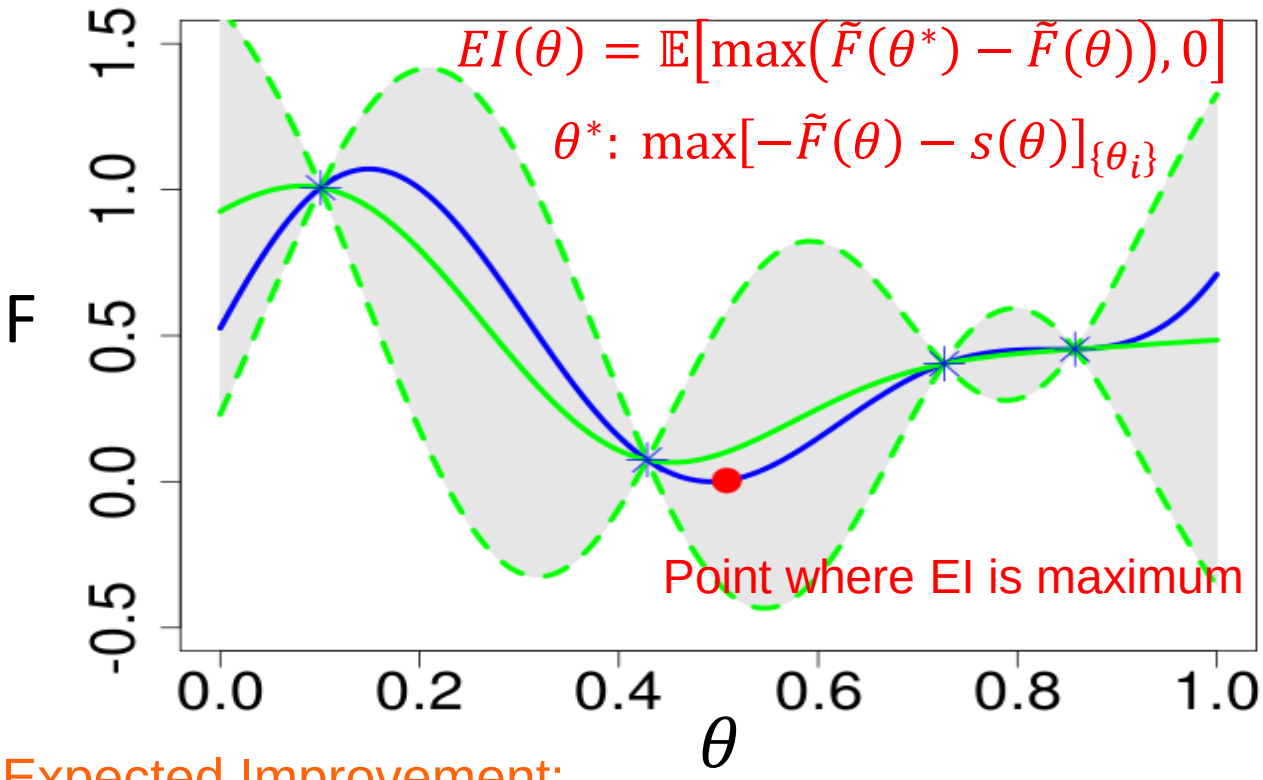
- Some parameter sets lead to badly converged simulations:
  - "Glassy-like" water
  - Subsample of 57 parameter sets (D2)

Calibration with GP surrogate models



- The minimum region of the estimator does not necessarily reproduce accurately the real minimum
- Iterative improvement of the estimator of the PDF:
  Use of “Efficient Global Optimization” (EGO) algorithms

Efficient Global Optimization



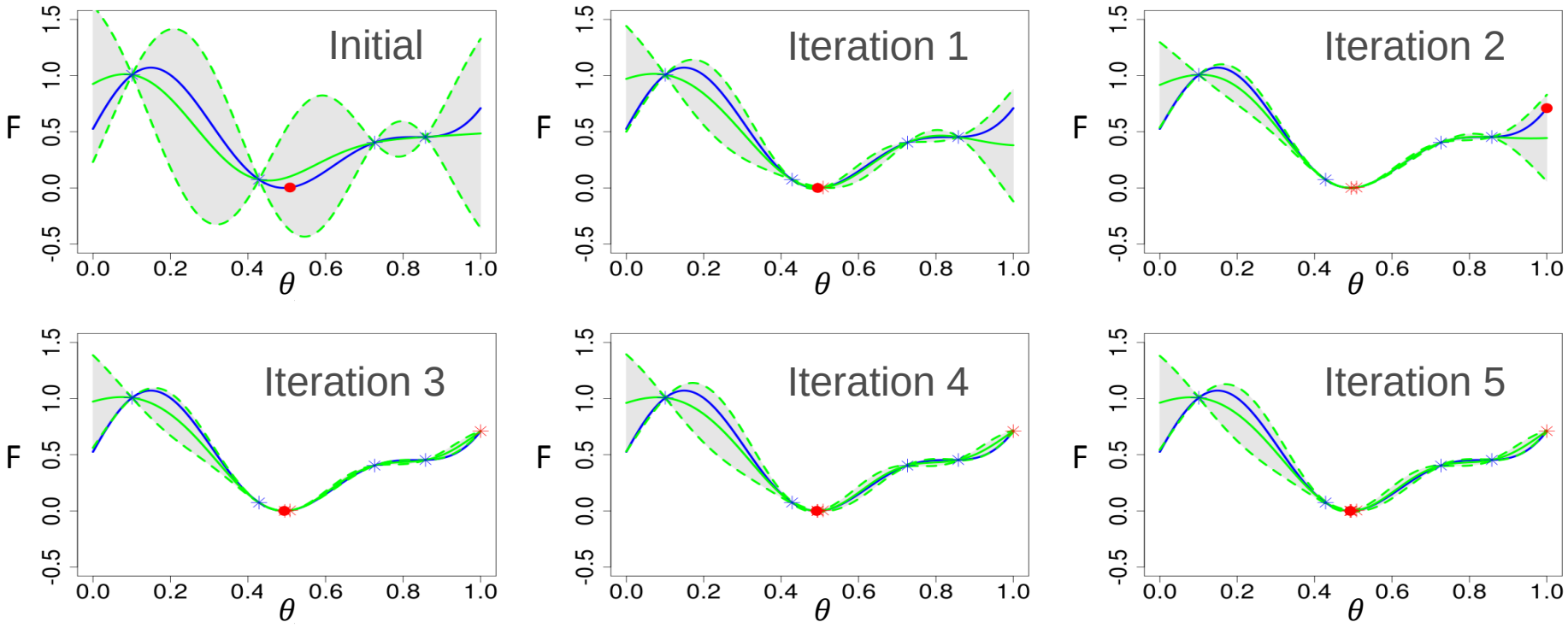
Huang *et al.*, 2006, *J. Glob. Opt.*, **34**: 441

Real function

Estimator

- Expected Improvement:
  use of the uncertainty prediction $s(\theta)$ of the surrogate model $\tilde{F}$
- Two technical difficulties:
  - GP optimised on stochastic data
  - $\tilde{F}$ not a GP: EI computed numerically

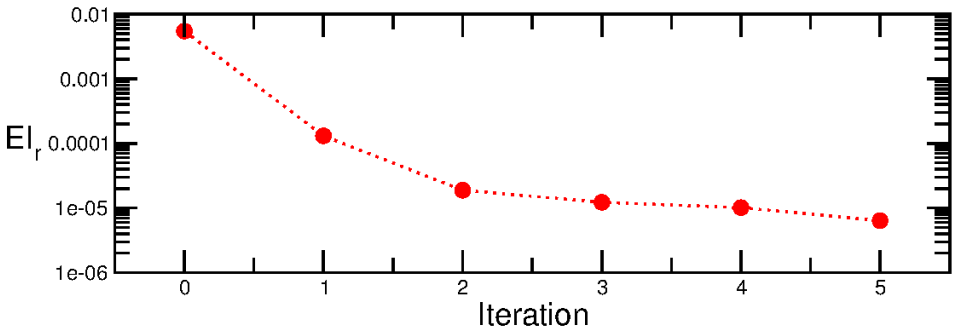
Efficient Global Optimization

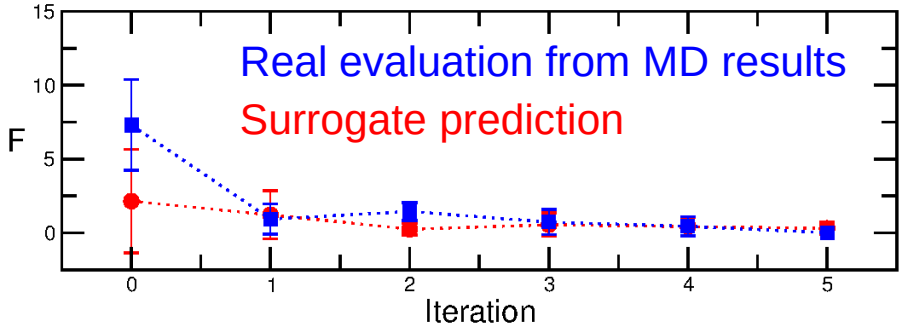


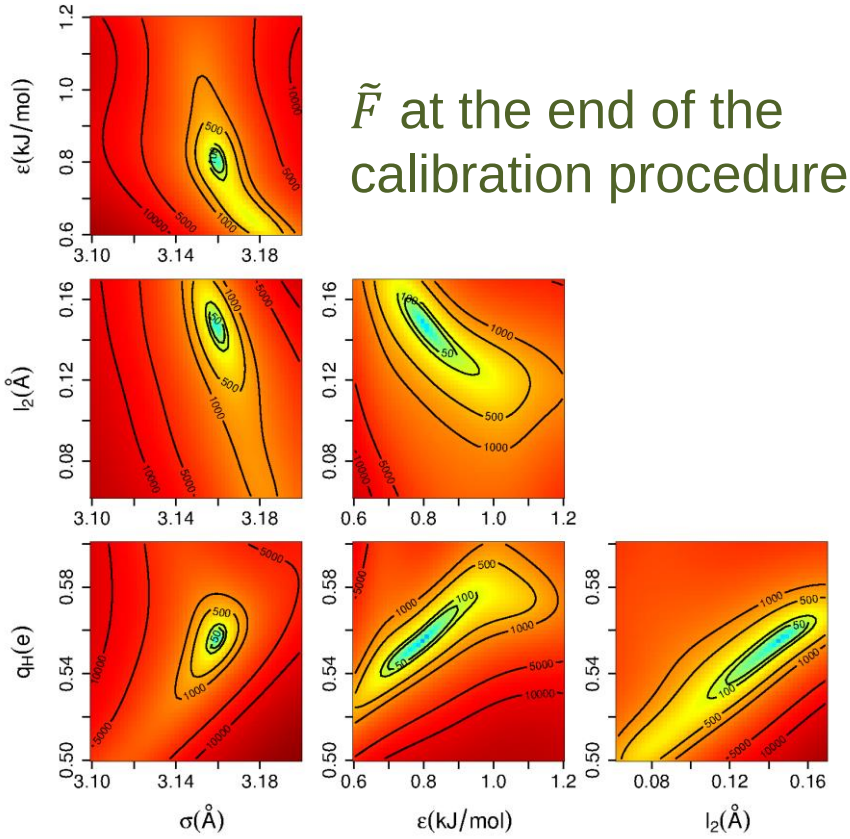
- Iterative procedure
- End of the procedure:

$$EI_r = \frac{EI}{\max(\tilde{F}) - \min(\tilde{F})} < 10^{-5}$$

EGO convergence



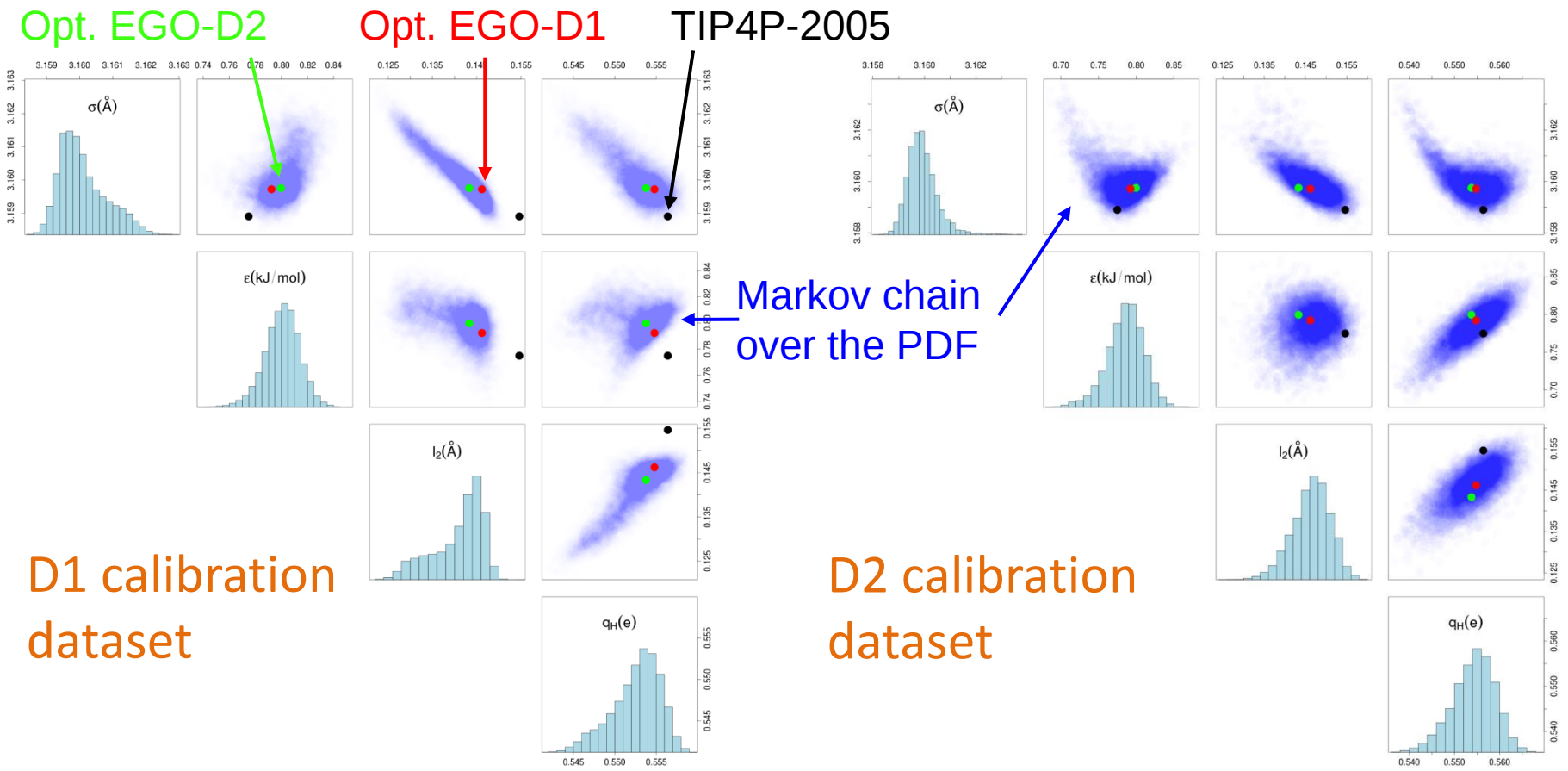




$\tilde{F}$ at the end of the calibration procedure

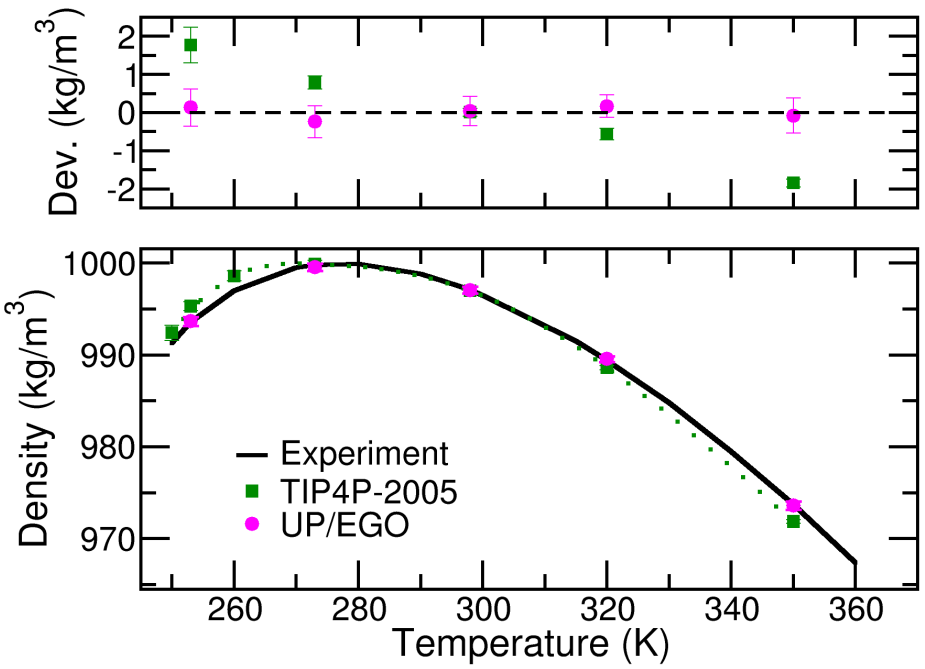
- Rapid convergence of the EGO
- Rapid improvement of the prediction around the optimum
- Single well defined optimum of the score function

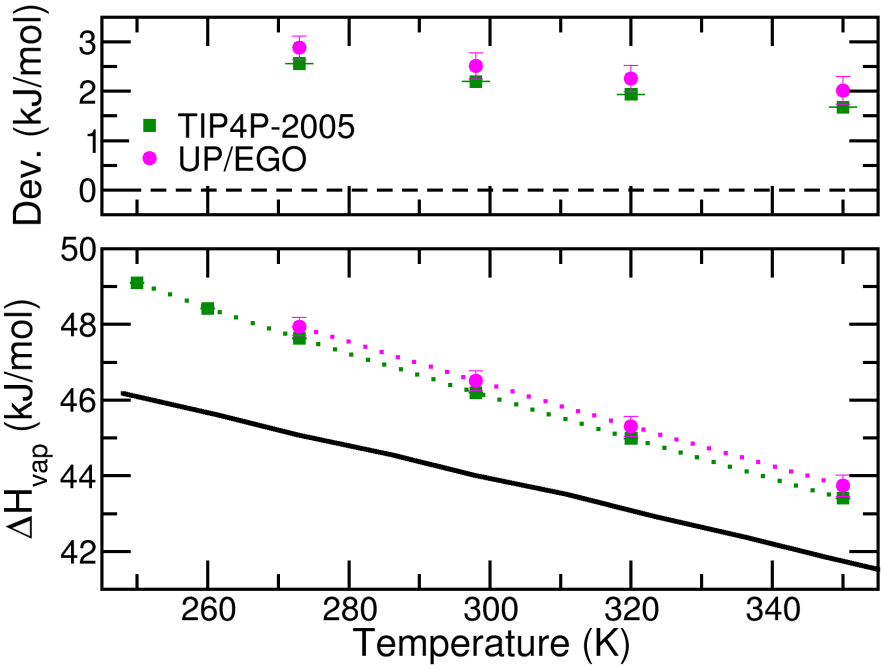
Results of the calibration



- D1 calibration dataset: ≈90 parameter sets used
- D1 and D2 calibration: similar results – low sensitivity to badly converged simulations

Parametric uncertainties – TIP4P forcefield





- Uncertainty propagation using kriging surrogate models for density and vaporization enthalpy
- Parametric uncertainties bigger than numerical uncertainties

A second conclusion

- Similar conclusions as for the argon case regarding parametric uncertainties:
  - At least as big as numerical uncertainties
  - Taking into account parametric uncertainty is not sufficient to have quantitative transferability to other properties
- Use of surrogate models:
  - Extensive exploration of parameter space at lower cost
  - Global sensitivity analysis (reduce parameter space dimension)
  - Global optimisation of the parameters possible

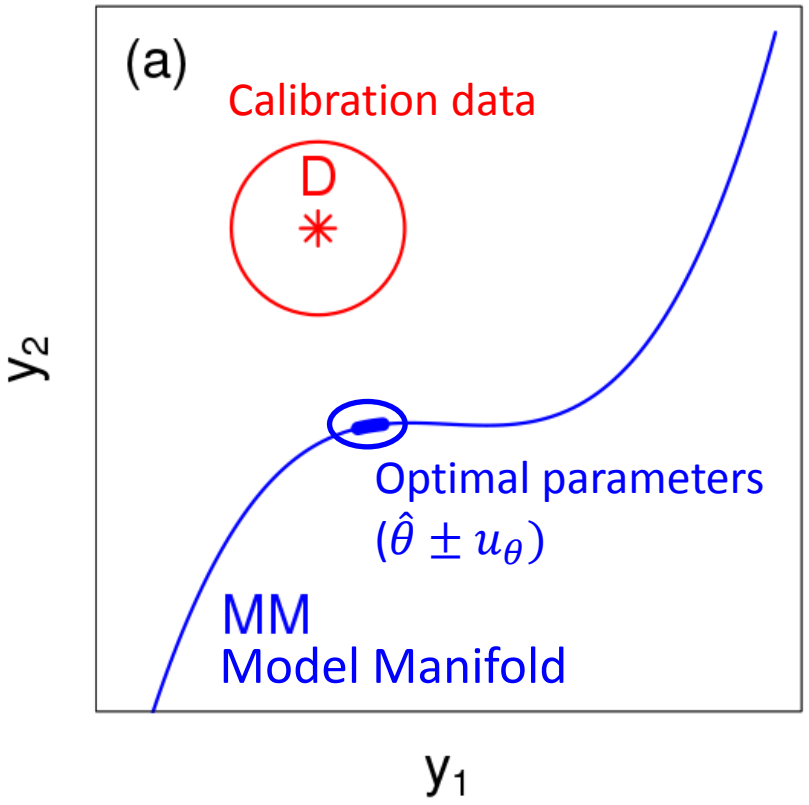
Cailliez, Bourasseau, and Pernot, *J. Comp. Chem.* ***35****: 130-149 (2014)*

- Limitations and unresolved issues:
  - Reducing the cost of the optimization procedure
  - How to deal with model inadequacy?

Outline of the presentation

- Bayesian calibration framework
- Statistical calibration of a forcefield for Argon:
  - Comparing numerical and parametric uncertainties
  - Transferability issues
- Calibration of a water forcefield:
  - How to decrease the computational burden?
  - Use of surrogate models and Efficient Global Optimization strategies
- How to deal with model inadequacy?
- Conclusions

The issue of model inadequacy



(Deterministic) model

$$y_i = \mathcal{M}(x_i, \theta) + R_i + e_i$$

Calibration data

Residual

Experimental uncertainty

- Inadequacy remains at the calibration stage
- Prediction inefficient even taking into account parameters uncertainties

Solving model inadequacy on synthetic data

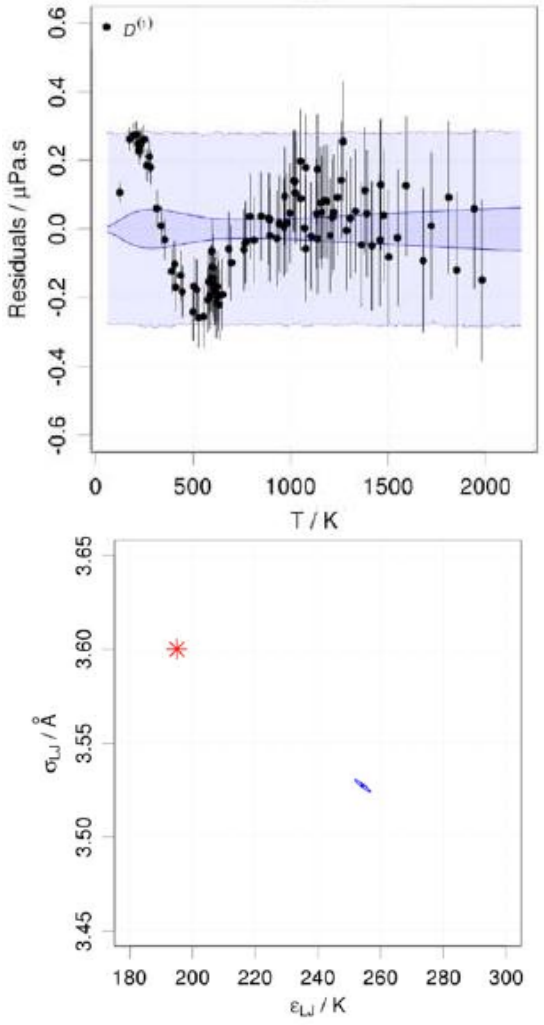
- Krypton described by a Lennard-Jones potential: $\theta = \{\sigma = 3.6\text{Å}; \varepsilon = 195\text{K}\}$
- Gas-phase viscosity: Chapman-Enskog model:

$$\eta = \mathcal{M}(T, \sigma, \varepsilon) = 2.6693 \frac{\sqrt{mT}}{\sigma^2 \Omega}$$

$$\Omega = \frac{A}{(T^*)^B} + \frac{C}{\exp(DT^*)} + \frac{E}{\exp(FT^*)} \qquad T^* = T/\varepsilon$$

- Synthetic data:
  - 100 data points for various T, generated with a modified value of C in CE formula
  - Generation of synthetic « experimental » uncertainties

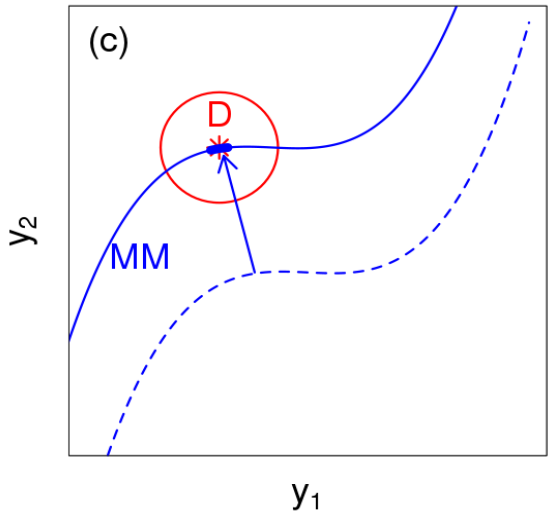
Standard calibration

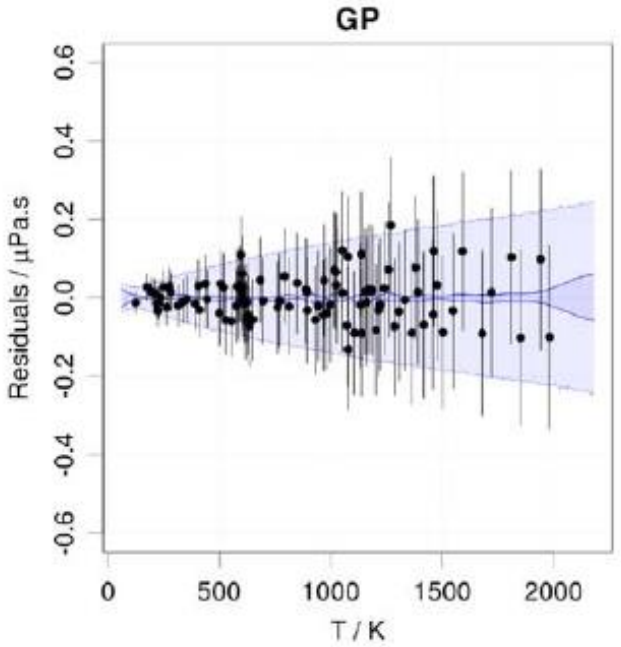


Correcting the model (GP)

- Add a discrepancy term to correct model errors:

$$y_i = \mathcal{M}(x_i, \theta) + GP(x_i, \theta_K) + e_i$$





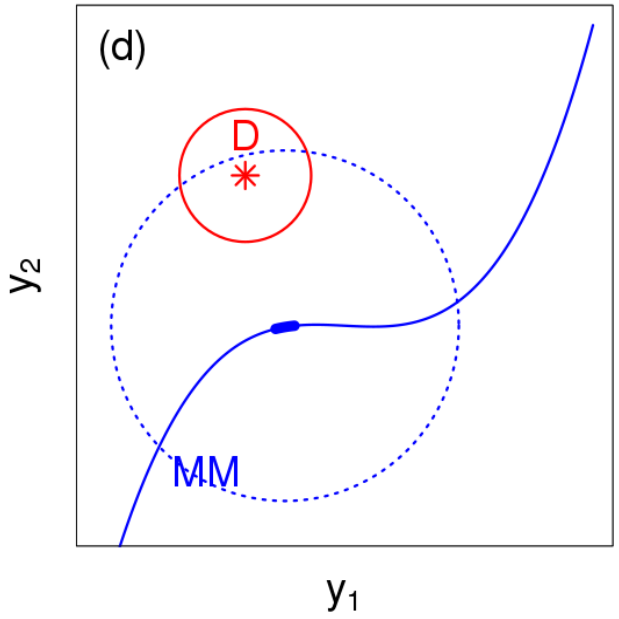
- Inadequacy issue is resolved but…
- Non-transferability of the correction to the prediction of another type of data

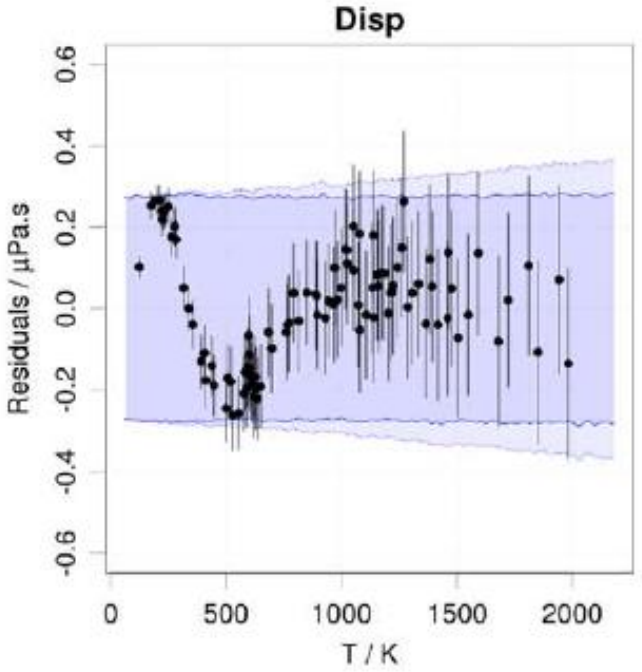
Kennedy & O'Hagan (2001), J. Roy. Stat. Soc. B, 63: 425-464

Correction at the prediction level (Disp)

- Adding a stochastic term to the model to increase the uncertainty of the prediction:

$$y_i = \mathcal{M}(x_i, \theta) + e_D + e_i \qquad e_D = \mathcal{N}(0, s^2)$$

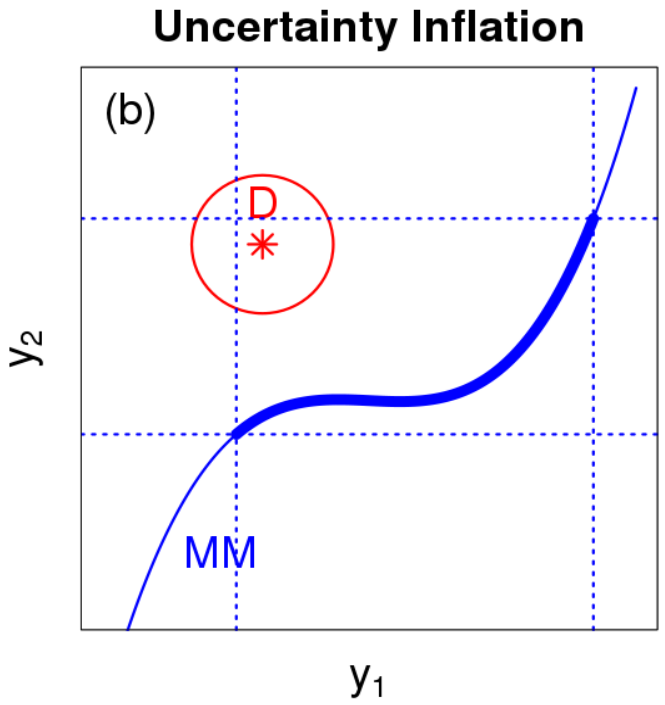




- Justified if no trend in the residuals
- Non-transferable to the prediction of another property

Increase parameter uncertainties

- Optimizing the covariance matrix Σ_θ of the parameters
- Variance inflation (VarInf):
  - scaling the covariance matrix obtained from standard calibration: $\Sigma'_\theta = s \times \Sigma_\theta$
- Hierarchical Bayesian framework* (Hier):
  - Divide the dataset D in series D_i
  - Calibrate parameters θ_i for each D_i
  - Find hyperparameters to reproduce the distribution of θ_i : $\theta_i \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$
- « Direct » stochastic modeling** (ABC):
  - $\mathcal{M}(\theta) \longrightarrow \mathcal{M}(\theta, \Sigma_\theta)$
  - Optimize $p(\theta, \Sigma_\theta | D)$

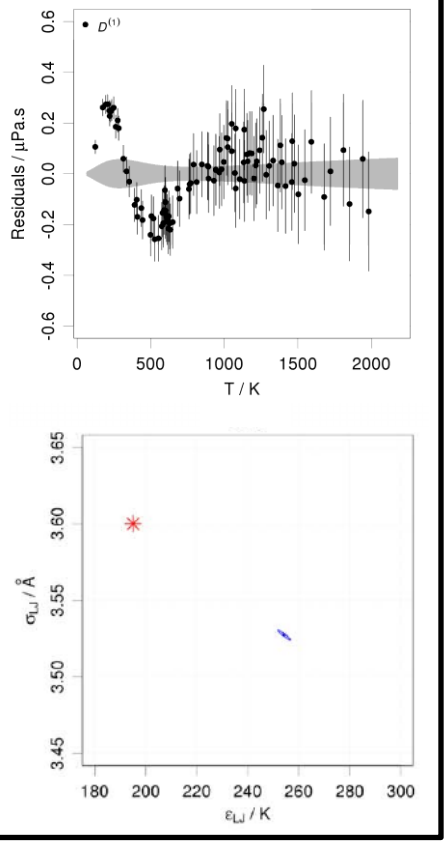


* Wu et al. (2015), Phil. Trans. R. Soc. A 374: 20150032

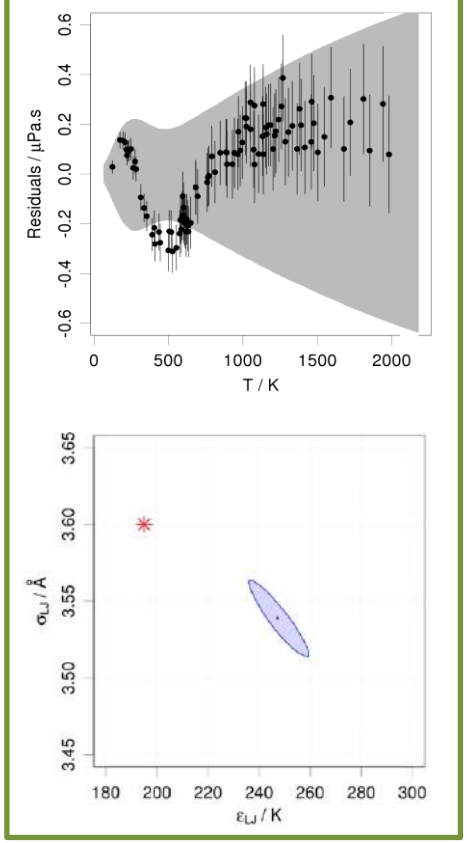
** Csilléry et al., *Trends Ecol Evol.* 2010;25:410–418.

Increase parameter uncertainties

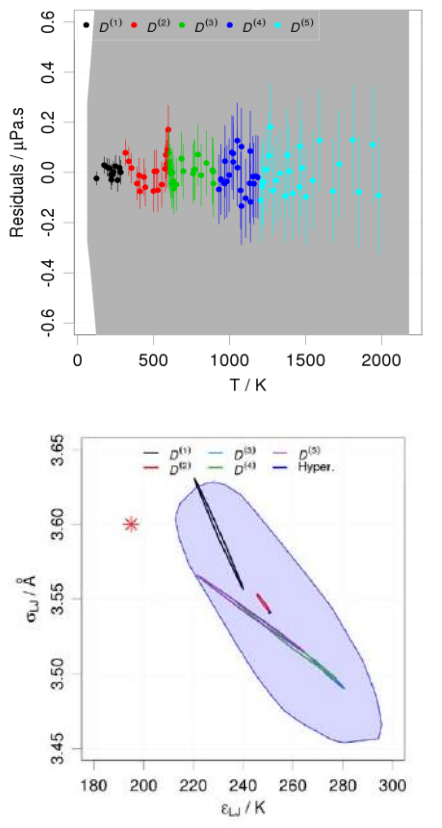
Standard calibration



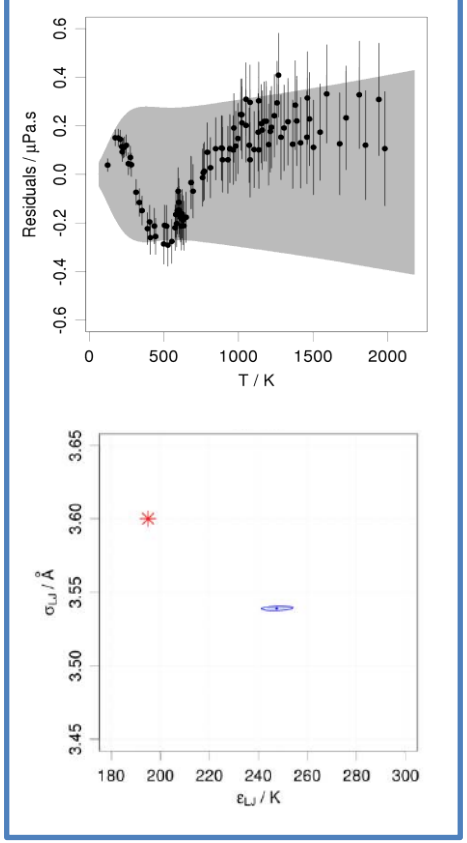
VarInf calibration



Hier calibration



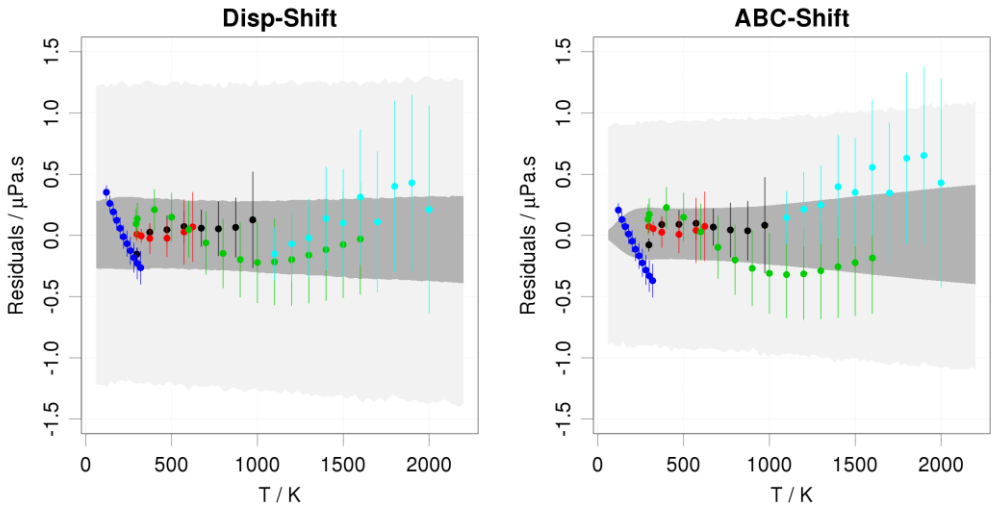
ABC calibration

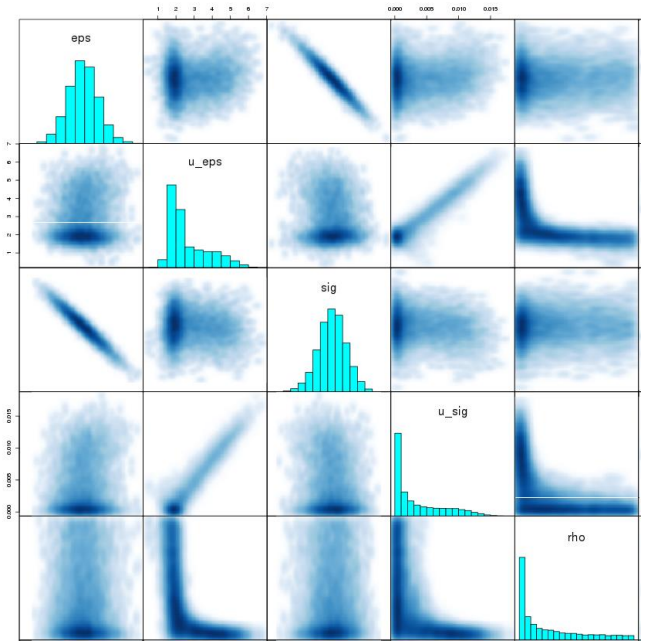


- VarInf and Hier calibrations: overestimated prediction bands
- ABC calibration: most reasonable option

A real test-case on experimental data

- Experimental data for Krypton viscosity





- When transferability to other properties is not an issue, Disp correction is OK
- ABC calibration:
  - reasonable but problems of multimodality of the solutions
  - Might be improved…

Pernot & Cailliez (2017), AIChE J:63(10): 4642-4665

Concluding remarks

- Study of parameters uncertainties in molecular simulations is still in its infancy
- Bayesian calibration is an adequate framework to determine forcefield parameters and their uncertainties
- Surrogate models and Efficient Global Optimization strategies can be used to alleviate the computational burden of the calibration
- Parametric uncertainties may be the main source of uncertainties in the calculation of fluid properties
- Model inadequacy and transferability to various properties are major issues for quantitative reliable predictions

Acknowledgements

- Laboratoire de Chimie Physique :
  - Pascal Pernot
  - Arnaud Bourasseau
  - Manuel Lopez-Ortiz
  - Jean-Marie Teuler
  - Bernard Rousseau

- Computational facilities :