Supplementary Information for "Application of machine-learning algorithms to predict the transport properties of Mie fluids"

Justinas Šlepavičius,[1] Alessandro Patti,[1,2] James L. McDonagh,[3] and Carlos Avendaño[1, a)]

[1)]*Department of Chemical Engineering, School of Engineering, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*

[2)]*Department of Applied Physics, University of Granada, Fuente Nueva s/n, 18071 Granada, Spain*

[3)]*IBM Research Europe, The Hartree Centre STFC Laboratory Sci-Tech Daresbury Warrington* [b)]

(Dated: 9 June 2023)

[a)]Electronic mail: carlos.avendano@manchester.ac.uk

[b)]Current address: Ladder Therapeutics doing business as Serna Bio, Lab F37, Stevenage Bioscience Catalyst, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2FX, UK

I. UNCERTAINTY

To measure the uncertainty we use 2 approaches. The first is to measure the standard deviation (STD) between the self-diffusion coefficients of 5 independent runs, while the second approach is through the calculation of the uncertainty in the slope of the MSD, which also reflects the quality of the MSD. We calculated the uncertainty in the slope using the equation

$$\sigma_{\text{slope}} = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-R^2}{N-2}} \tag{1}$$

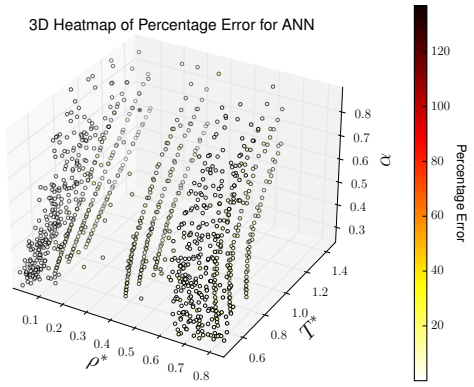
where x, y are the dependent and independent variables, respectively, and N is the number of data points in the MSD[1]. For the first method, it has been observed that the errors, given by one STD, are $\approx 0.05\%$, while the uncertainty obtained from the slope was in the range 0.5%-15%.

II. HEAT MAPS

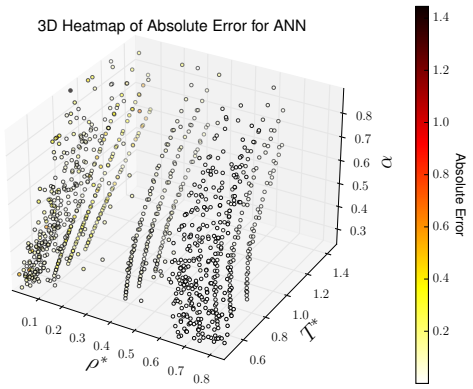
The range of applicability of the ML models is tested *via* the use of heat maps in the phase space. In this section we present the heat maps of the ML algorithms tested on the full data set. To check that the bias on the simulations due to the selection of densities in the supercritical (SC) region, we have run 50 additional simulations in the SC region in random locations in our range of densities and temperatures. We looked at both the percentage and absolute errors, as the high variability in the magnitude of D^* makes different methods more effective. The heat maps are presented through Figure S1. For both black-box methods the percentage error is low everywhere, except for a couple of points from the additional SC data. Absolute error is the highest in the vapour phase for all 3 methods. FOR SR in the liquid phase the error is quite high, again presenting the need for the cohesion parameter

III. PREDICTION OF LJ SELF-DIFFUSION COEFFICIENTS FROM MEIER *ET AL.*

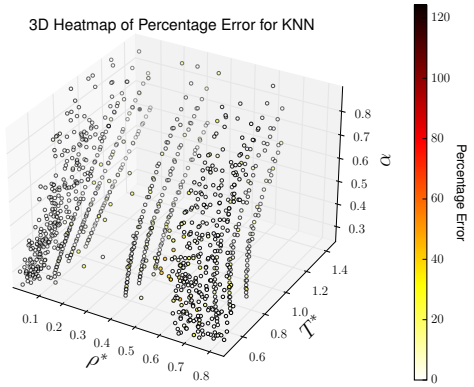
Most research conducted on LJ particles suffers from two main limitations when compared to the methodology employed in this work. Firstly, previous studies on LJ fluids tend to focus on a cutoff distance of $r_{\text{cut}} = 2.5\sigma$, which is considerably shorter than the cutoff distance utilized in our research ($r_{\text{cut}} = 6\sigma$). Secondly, investigations into pure LJ fluids typically employ relatively small particle numbers, with $N = 1372$ being the largest system size commonly used. These limitations



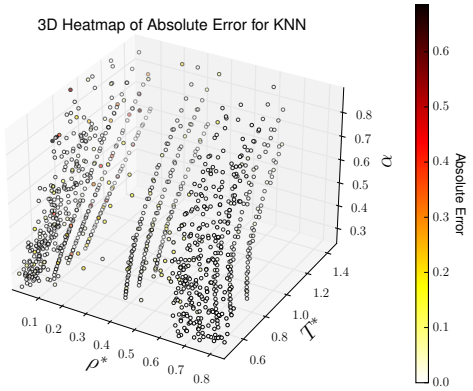
(a) Heat Map for Percentage Error for ANN



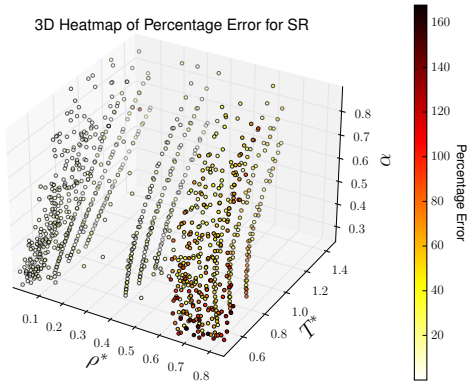
(b) Heat Map for Absolute Error for ANN



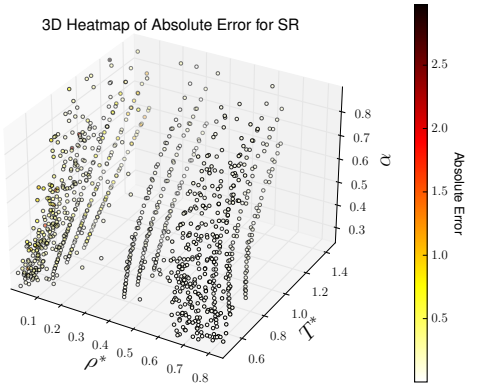
(c) Heat Map for Percentage Error for KNN



(d) Heat Map for Absolute Error for KNN

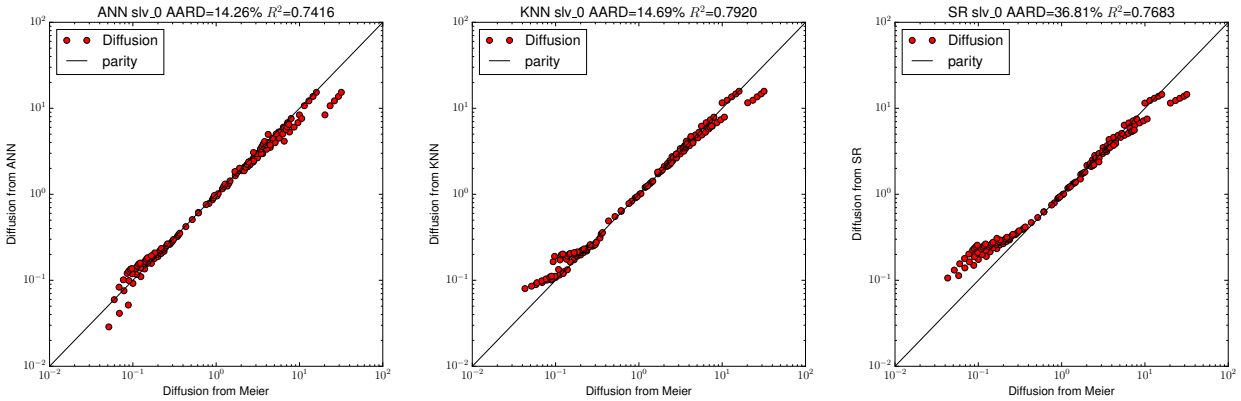


(e) Heat Map for Percentage Error for SR



(f) Heat Map for Absolute Error for SR

FIG. S1: Heat maps of the accuracy of the ML to predict the self-diffusion coefficient in the phase space. The analysis of the accuracy is represented by both the absolute average relative (AARD) and mean-absolute error (MAE). The results correspond to (a-b) ANN, (c-d) KNN, and (e-f) SR.



(a) Parity plot of ANN for Meier *et al.*

(b) Parity plot of KNN for Meier *et al.*

(c) Parity plot of SR for Meier *et al.*

FIG. S2: Prediction of self-diffusion coefficient of Lennard-Jones particles using the ML algorithms reported in this work. The results correspond to (a) ANN, (b) KNN, and (c) SR. The reference data is taken from the work by Meier *et al.*[2].

fail to capture subtle effects associated with dense liquid systems.

The work by Meier *et al.*[2] reports data on self-diffusion coefficients using a larger cutoff radius. In their study, the authors utilized $N = 1372$ particles and a cutoff distance of $r_{\text{cut}} = 5.5\sigma$. We have created parity plots, displayed in Figure S2, which demonstrate the accuracy of our predictions when compared to the data reported in their work.

Our ML models effectively capture the data presented by Meier *et al.* The most significant deviations from parity are observed in the low and high D^* regions, and the reasons for these discrepancies are similar. In regions of high D^* (corresponding to high T^* and low ρ^*), the particles are sparsely distributed to the extent that the 0.5σ difference in cutoff distance significantly impacts the average number of interactions per particle. This, in turn, directly affects the mean square displacement (MSD) of the particles, resulting in lower values of D^*. Conversely, the opposite effect is observed in the low D^* region, where the particles are densely packed, and the reduced cutoff distance (r_{cut}) leads to a decrease in the number of particles within the interaction shell. This reduction affects the particle displacement, resulting in overestimated predictions of D^*. However, the ML models perform well in predicting D^* values between these extremes.

IV. PARITY PLOTS OF THE PREDICTION OF *D* FOR REAL FLUIDS FOR ALL ML MODELS

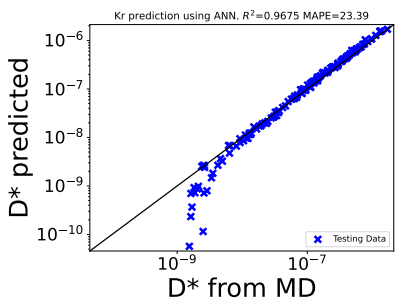
The prediction of the self-diffusion coefficients for real fluids using the ML algorithms developed in this work, namely ANN, KNN, and SR, is presented in this section. The systems correspond to Kr, CH_4, and CO_2. The results are presented in Figure S3.

REFERENCES

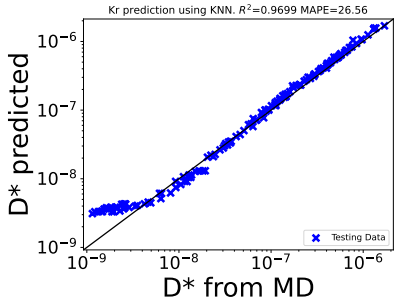
[1] A. F. Siegel and M. R. Wagner, in *Practical Business Statistics* (Elsevier, 2022) pp. 313–370.

[2] K. Meier, A. Laesecke, and S. Kabelac, J. Chem. Phys. **121**, 9526 (2004).

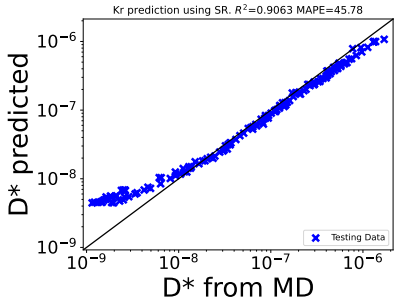
[3] J. P. Allers, F. H. Garzon, and T. M. Alam, Phys. Chem. Chem. Phys. **23**, 4615 (2021).



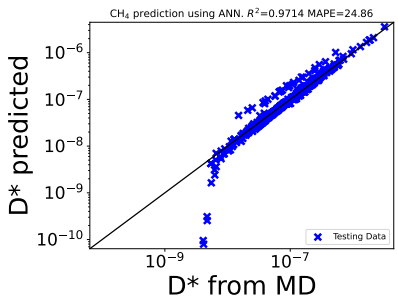
(a) Parity plot of ANN for Kr



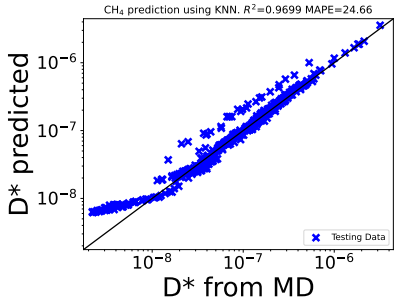
(b) Parity plot of KNN for Kr



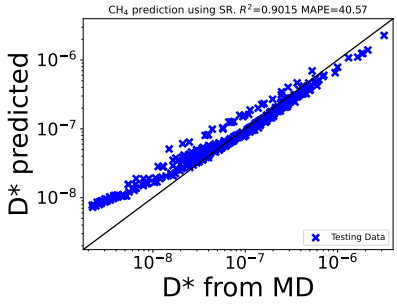
(c) Parity plot of SR for Kr



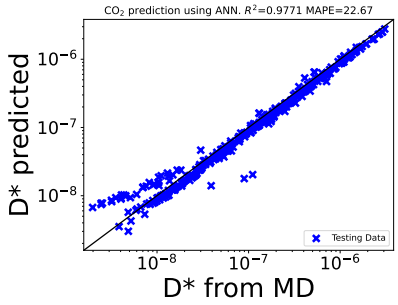
(d) Parity plot of ANN for CH_4



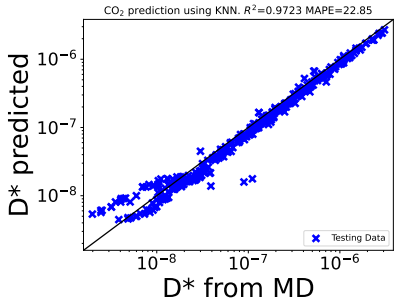
(e) Parity plot of KNN for CH_4



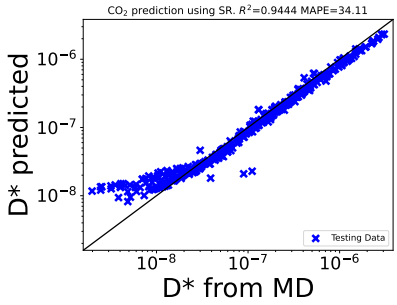
(f) Parity plot of SR for CH_4



(g) Parity plot of ANN for CO_2



(h) Parity plot of KNN for CO_2



(i) Parity plot of SR for CO_2

FIG. S3: Results for self-diffusion coefficients of (a-c) krypton Kr, (d-f) methane CH4, and (g-i) carbon dioxide CO_2 predicted by ANN, KNN and SR. The results are compared to the experimental results collected by Allers et al.[3].