



OPEN

Data-driven time-dependent state estimation for interfacial fluid mechanics in evaporating droplets

Sahar Andalib[✉], Kunihiro Taira & H. Pirouz Kavehpour

Droplet evaporation plays crucial roles in biodiagnostics, microfabrication, and inkjet printing. Experimentally studying the evolution of a sessile droplet consisting of two or more components needs sophisticated equipment to control the vast parameter space affecting the physical process. On the other hand, the non-axisymmetric nature of the problem, attributed to compositional perturbations, introduces challenges to numerical methods. In this work, droplet evaporation problem is studied from a new perspective. We analyze a sessile methanol droplet evolution through data-driven classification and regression techniques. The models are trained using experimental data of methanol droplet evolution under various environmental humidity levels and substrate temperatures. At higher humidity levels, the interfacial tension and subsequently contact angle increase due to higher water uptake into droplet. Therefore, different regimes of evolution are observed due to adsorption–absorption and possible condensation of water which turns the droplet from a single component into a binary system. In this work, machine learning and data-driven techniques are utilized to estimate the regime of droplet evaporation, the time evolution of droplet base diameter and contact angle, and level of surrounding humidity. Droplet regime is estimated by classification algorithms through point-by-point analysis of droplet profile. Decision tree demonstrates a better performance compared to Naïve Bayes (NB) classifier. Additionally, the level of surrounding humidity, as well as the time evolution of droplet base diameter and contact angle, are estimated by regression algorithms. The estimation results show promising performance for four cases of methanol droplet evolution under conditions unseen by the model, demonstrating the model's capability to capture the complex physics underlying binary droplet evolution.

Wetting and spreading of liquid on a solid surface is an omnipresent phenomenon in nature and engineering technologies such as biodiagnostics, inkjet printing, microfabrication, spray cooling, and agricultural irrigation^{1–13}. Droplet evaporation has gained increasing attention over the past 25 years^{14–19}. The original work²⁰ on the coffee-ring effect was followed by the comprehensive study²¹ to theoretically, experimentally, and numerically calculate the evaporation rate of sessile water droplets. Evaporation of a single component sessile droplet is influenced by various factors such as substrate temperature²², environment pressure²³, surfactant concentration²⁴, substrate thermal conductivity^{25–27}, and surrounding gas²⁸. When an organic fluid droplet evaporates on a solid surface, thermocapillary instabilities known as hydrothermal waves (HTWs) are created due to surface tension gradient along the interface²⁹.

Although the evolution of a single component droplet is mainly understood, the physics becomes complex when there is more than one component in the droplet. Three stages were reported for evaporation of water–ethanol mixture droplets where the first stage corresponded to evaporation of a more volatile component while the last stage was responsible for evaporation of a less volatile component³⁰. The humidity of the surrounding plays a crucial role in evaporation of a binary sessile droplet. Two separate studies^{31,32} observed a rise in contact angle of binary mixtures which suggested possible condensation of water on droplet. Adsorption of water in ethanol and ethanol/water mixtures is also reported by time-resolved infrared spectroscopy³³. Three different techniques of optical visualization, infrared thermography, and acoustic high-frequency echography were employed in a comprehensive study³⁴ to examine the evaporation of butanol, ethanol, water/butanol, and water/ethanol droplets. Their results showed that due to the high hygroscopic power of ethanol, the humidity of the environment had a noticeable effect on the evolution of pure ethanol droplets. Heterogeneous thermal patterns alongside the evolution of acoustic reflection coefficient proved that ethanol droplet undergoes continuous

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA.
✉email: sandalib@ucla.edu

water loading. The combined influence of ambient temperature and relative humidity on early stages (i.e. pinned contact angle) of ethanol droplet was also examined³⁵. In another study³⁶ the water loaded onto ethanol droplet was quantified by gas injection chromatography (GIC) under controlled ambient temperature and relative humidity. The observed reduction in ethanol concentration was attributed to both ethanol evaporation as well as water intake on the drop. They concluded that at low relative humidity, the main mechanism for water intake was that of adsorption–absorption, though at high relative humidity water condensation plays a more dominant role. While the changes in relative humidity are commonly imposed environmental conditions, controlling ambient temperature to tune the effect of humidity is rather an expensive method for practical applications. The effects of relative humidity on methanol droplet evaporation was regulated by adjusting the temperature of the substrate^{37,38} which is less expensive compared to controlling ambient temperature. It was concluded that increasing substrate temperature maintains the liquid–gas interface temperature above the dew point which in turn limits the water condensation on the drop. A regime map was also proposed based on droplet evolution under various environmental conditions.

As the number of components in the droplet increases, the underlying physics becomes more complex. Recently, multi-component droplet evaporation revealed new phenomena such as spontaneous nucleation of oil microdroplets, phase transition, and multi-component diffusion^{39–42}. Such intricate physics with numerous parameters in play makes experimental studies sophisticated and time-consuming while requiring advanced equipment to finely control the environmental condition. On the other hand, the highly non-axisymmetric nature of the problem due to compositional inhomogeneities brings up significant challenges to numerical models.

Machine learning methods have emerged as powerful tools for analyzing a wide range of fluid mechanics problems^{43,44} such as turbulence^{45–49}, phase transition⁵⁰, ignition⁵¹, vortex vibrations^{52,53}, and aerodynamics disturbances⁵⁴. Image processing and pattern recognition techniques have been employed to analyze the remaining stains after evaporation of sessile droplets. Analysis of patterns in dried drops of biological fluids revealed a lot of information for medical diagnostics^{55–59}. Two different studies^{60,61} showed that the chemistry of fluid and substrate can be identified by recognition of patterns in the stains. The aforementioned studies utilized machine learning techniques to indirectly measure and detect the mechanisms inside droplets from the footprint after evaporation. Here, we introduce a data-driven approach to directly analyze the dynamics of binary droplet evaporation (induced by transfer of a second component present in the atmosphere).

In the present study, we use data-driven classification and regression algorithms to analyze the time-dependent behavior of a methanol droplet at different levels of environmental humidity and temperature of the substrate. Water uptake into droplet through adsorption–absorption and possibly condensation turns methanol droplet into a binary system. Based on the environmental condition, a droplet evolves in different regimes: evaporation-dominated, transition, or condensation-dominated. The capability of the proposed model is evaluated by estimating four different parameters, namely: regime of droplet evaporation (through classification algorithms), level of surrounding humidity (through regression algorithms), time evolution of droplet base diameter (through regression algorithms), and time evolution of droplet contact angle (through regression algorithms). First, a classification algorithm is trained to estimate the regime of droplet evaporation through analysis of diameter and contact angle evolution over time. The objective of the model is to detect the regime of droplet evolution with even a single data point at a specific time. Second, a regression algorithm is utilized to detect the humidity of the surrounding by analyzing droplet evolution. The high hygroscopic nature of methanol allows greater amount of water uptake into the droplet in humid environments. The higher content of water in droplet increases the contact angle and alters the rate of change of volume. The regression model analyzes these changes and reversely estimates the humidity. Last, given the condition of the surrounding, the continuous evolution of macroscopic parameters of droplet, i.e., diameter and contact angle, is estimated. Our method shows great potential in opening up new paths to analyze more complicated multi-component droplet evolution and interfacial fluid mechanics in general. Estimating the evolution of different parameters of droplet is a crucial step in designing high quality and high-resolution finish products in droplet-based biodiagnostics, inkjet printing, and microfabrication technologies. Our proposed method provides necessary information on evaporation of organic liquid droplets under various environmental conditions with simple and easy-to-use algorithms without the need to perform complicated simulations. We show that the regime of droplet, relative humidity of surrounding, and time evolution of diameter and contact angle can be estimated under various conditions. In real-world applications anticipating the regime of droplet is of great importance as in many instances, the occurrence of one regime or the other should be avoided. For example, a droplet sitting on a surface forever is not ideal for high resolution of printing or biosensing.

Results

Physics of droplet evaporation. Droplet evaporation is influenced by numerous factors including liquid/substrate properties as well as environmental conditions^{25–28,35,36,62}. We analyze the evolution of a sessile methanol droplet through macroscopic parameters: volume, V , diameter, D , contact angle, θ , and time, t , under controlled relative humidity of surrounding, RH , and substrate temperature, T (see Fig. 1a, inset). The variables are nondimensionalized as: $t^* = t/t_f$, $V^* = V/V_0$, $D^* = D/D_0$, $\theta^* = \theta/\theta_0$, $T^* = T/T_0$, $RH^* = RH/RH_0$, where $\theta_0 = 90^\circ$, $T_0 = 35^\circ\text{C}$, $RH_0 = 100\%$, and t_f, V_0, D_0 , that are experimentally measured, stand for total evaporation time, initial volume, and initial diameter, respectively. The experiments are conducted in a chamber with controlled humidity and on a substrate with controlled temperature (Fig. 1a). Details of experimental procedure are given in “Methods” section. Three regimes of droplet evolution are observed under various relative humidity of the surrounding and substrate temperature (Fig. 1b, top left) namely: evaporation-dominated, transition, and condensation-dominated. Three sub figures represent the evolutions of θ^* , V^* , and D^* over t^* . Nondimensional plots are reported to better visualize different evolution patterns.

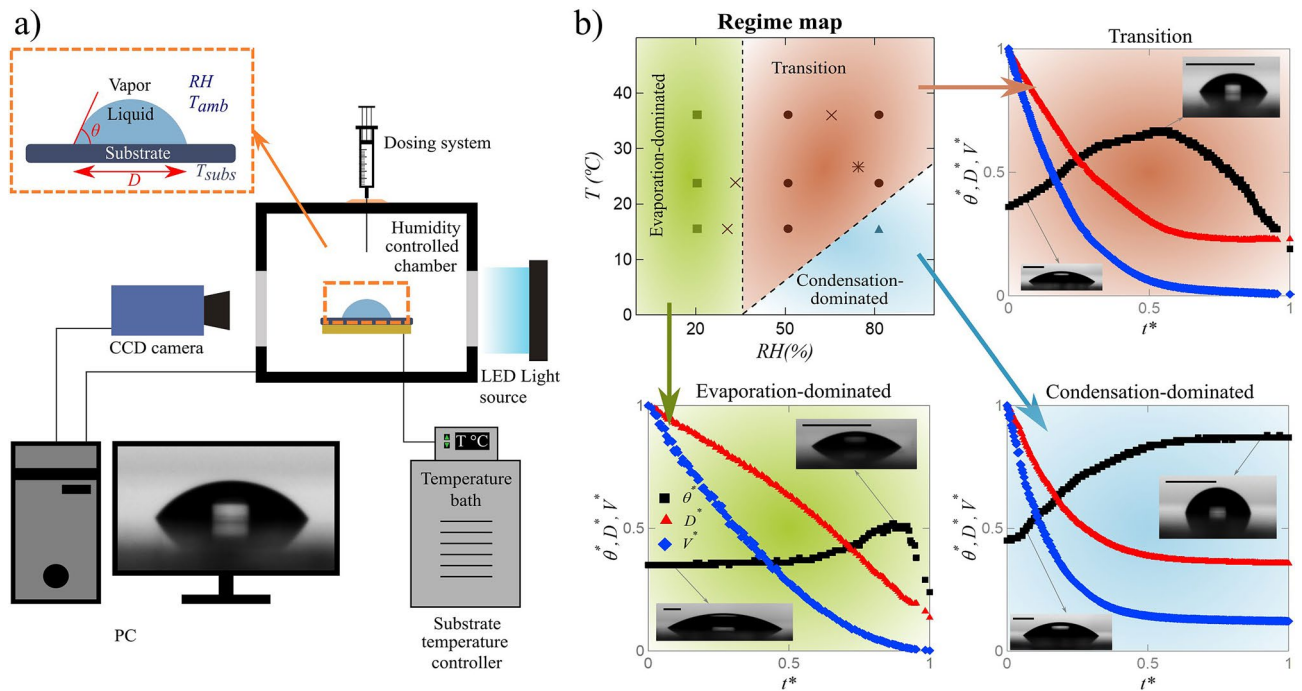


Figure 1. (a) Schematic of the experimental setup with macroscopic parameters of droplet shown in the inset; (b) regime map of droplet evaporation (top-left) under various relative humidity (RH) of surrounding and substrate temperature (T), evolutions of nondimensional contact angle, volume, and diameter versus time for evaporation-dominated (bottom-left), transition (top-right), and condensation-dominated (bottom-right) regimes. Each scale bar in droplet images represents a length of 1 mm. The schematic is made using free and open-source software Inkscape (Harrington⁶³).

At low relative humidity (the green-shaded region in Fig. 1b), change in substrate temperature does not alter the qualitative evolution of droplet. In this regime, the contact angle stays constant for most of droplet lifetime followed by a slight increase and a sharp decrease towards the end (Fig. 1b, bottom left). The modest rise in contact angle is attributed to the interplay of high evaporation rate of methanol and receding speed at the triple line³¹. Diameter and volume monotonically decrease during droplet lifespan.

Due to the high hygroscopic nature of methanol, at higher relative humidity, water vapor transfers into the droplet at the liquid–gas interface. Water adsorbing–absorbing and possibly condensing on the interface is reported in previous studies^{30,33–37}. The growth in the concentration of water content changes the interfacial tensions and results in higher contact angle^{34,36,37}. Unlike low relative humidity, substrate temperature plays a determining role in the regime of droplet evolution at high relative humidity of the surrounding. In the transition regime (red-shaded region), contact angle rises to a maximum value before gradually decreasing towards the end of droplet lifetime. Increasing contact angle demonstrates water uptake into droplet while methanol is evaporating. At the point of maximum contact angle, most of methanol has already evaporated and droplet consists mainly of water. However, some studies revealed that a small amount of residual methanol remains until the end of droplet lifetime^{31,32}. Even though diameter and volume decrease monotonically, two obvious slopes are observed in their evolutions (Fig. 1b, top right). The two slopes correspond to two stages: the initial stage when merely methanol evaporates and the second stage when water mainly evaporates at a slower rate.

When the humidity of the environment is high and the substrate temperature is sufficiently low, another regime is observed. In condensation-dominated (blue-shaded) regime, contact angle monotonically increases until it reaches a plateau. Both diameter and volume converge to a non-zero value. Lower substrate temperature enhances water uptake through condensation by dropping the liquid–gas interface temperature below that of dew point³⁷. In this regime, droplet comes to a quasi-steady state with a remaining droplet consisting mainly of water^{31,32,34,36}.

Regime classification. We have used a classification algorithm to detect the regime of droplet evaporation as sketched in Fig. 1b. The classifier is trained with data on contact and diameter at each specific point in time and then classifies the regime of droplet evaporation. Dependence of variables is shown by the correlation matrix in Fig. 2a where RG stands for the regime of droplet evaporation. Diameter and volume are coupled for a spherical cap sessile droplet through the relation $V = (\pi/3)(D/2)^3(2 + \cos\theta)(1 - \cos\theta)^2$ which assumes slow quasi–static evaporation. t^* , D^* , and θ^* are used as input variables and RG is the target variable. It is observed that the contact angle is highly proportional to humidity because the higher the humidity, the higher the amount of water uptake into drop. Higher water content increases the interfacial tension at the triple line which results in higher contact angles.

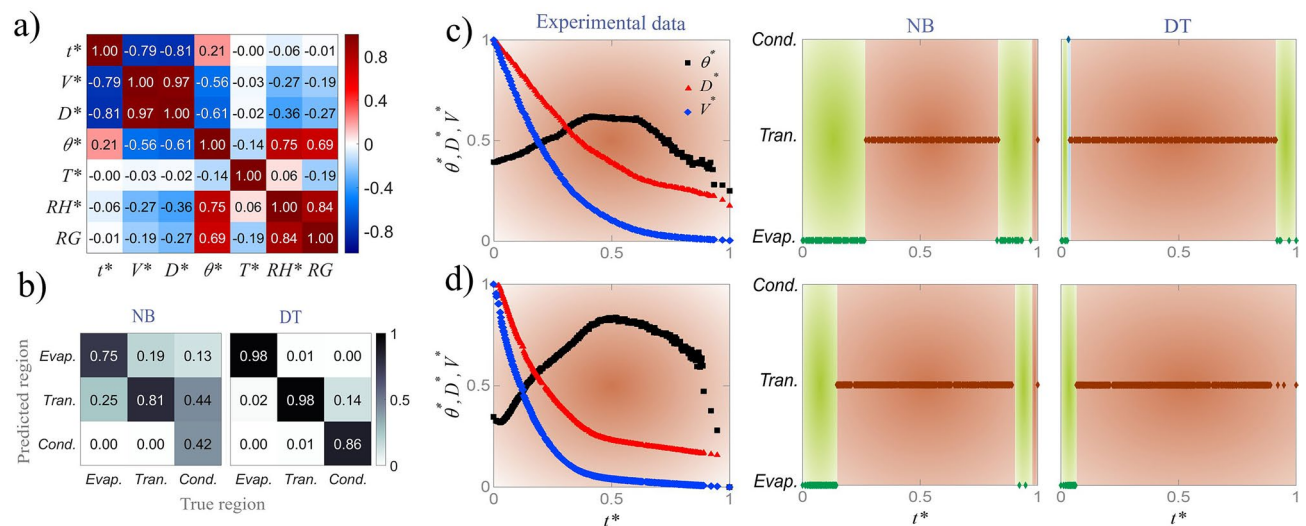


Figure 2. Results of regime classification: **(a)** correlation matrix for parameters in droplet evaporation; **(b)** results of test set for regime detection illustrated with confusion matrices for Naïve Bayes (NB) and decision tree (DT) algorithms; **(c)** point-by-point validation results with NB and DT classifiers for regime detection with experimental data of droplet evaporation for $RH = 80\%$ and $T = 35^\circ\text{C}$; **(d)** point-by-point results of estimation set with NB and DT classifiers for regime detection with experimental data of droplet evaporation for experiment #4 ($RH = 75\%$ and $T = 25^\circ\text{C}$) in estimation set. The colors in **(c,d)** correspond to the regime colors used in Fig. 1b.

Classifier	Accuracy	RG	Precision	Recall	F-score
NB	0.75	Evap.	0.74	0.65	0.6
		Trans.	0.81	0.78	0.79
		Cond.	0.46	0.93	0.61
DT	0.96	Evap.	0.99	0.97	0.97
		Trans.	0.97	0.96	0.96
		Cond.	0.83	0.89	0.85

Table 1. Regime classification results on test set: standard metrics for two different classifiers.

The framework for detection of the regime of droplet evaporation characterizes the behavioral pattern of droplet by learning the values of contact angle and diameter at each specific point in time and classifying them to each regime. The model then labels the test and validation sets based on similar evolution observed previously during training. The ratio of training to test set is 80–20%. Two classifiers of Naïve Bayes (NB) and decision tree (DT) are trained and confusion matrices are used to compare the performance of classifiers on the test set (Fig. 2b). Precision, recall, F-score, and overall accuracy values (shown in Table 1) provide a comprehensive evaluation of the performance of each classifier on the test set. Based on the results shown in the Fig. 2b and Table 1, DT outperforms NB for all regimes of the test set. It is also observed that the detection of the condensation-dominated regime is challenging for both classifiers. Detection of evaporation-dominated regime reaches an F score value of 0.97 with decision tree classifier. For NB, around half (43%) of the points in the condensation-dominated regime are classified as transition regime (see Fig. 2b). Replacing diameter with volume slightly improves the results for both classifiers (6% on average) for this regime. This is due to a more discernible evolution of V^* compared to D^* towards the end of droplet lifetime for this regime. However, since measuring diameter is a more direct approach and also more convenient for the user, the model is trained with diameter.

The capability of the model to detect the regime of droplet evaporation under each specific condition (RH and T) is evaluated through a validation set. The validation step is performed on a single experiment at each time that is held out during training/testing. Validation results, averaged for each condition, are presented in Table 2. The accuracy and recall are the same for validation because at each condition there is only one true regime. Figure 2c illustrates a sample of validation for $RH = 80\%$ and $T = 35^\circ\text{C}$. The classifier assigns a region for each point in time based on the value of contact angle and diameter. The true regime for this condition is the transition regime (red). The red regions on the two plots on the right side of Fig. 2c represent the regions that are classified correctly as transition regime and the green regions show the regions that are incorrectly classified as the evaporation-dominated regime. It is observed that both classifiers correctly detect the region of the majority of the data points although DT demonstrates better performance. NB struggles at the beginning and end of droplet lifetime. This issue is less pronounced for DT.

Classifier	RH (%)	15 °C	23 °C	35 °C
NB	20	0.72 (± 0.04)	0.73 (± 0.08)	0.78 (± 0.06)
	50	0.80 (± 0.06)	0.62 (± 0.11)	0.63 (± 0.02)
	80	0.41 (± 0.10)	0.88 (± 0.06)	0.83 (± 0.08)
DT	20	0.98 (± 0.03)	0.96 (± 0.08)	0.96 (± 0.08)
	50	0.94 (± 0.10)	0.94 (± 0.03)	0.98 (± 0.03)
	80	0.74 (± 0.23)	0.97 (± 0.05)	0.94 (± 0.07)

Table 2. Regime classification results on validation set under each specific *RH* and *T* condition with two different classifiers.

Experiment	<i>RH</i> (%)	<i>T</i> (°C)	NB		DT	
			Evap.	Trans.	Evap.	Trans.
1 (X)	30	15	0.54	0.46	0.67	0.33
2 (X)	33	23	0.61	0.39	0.79	0.21
3 (X)	65	35	0.32	0.68	0.16	0.84
4 (*)	75	25	0.18	0.82	0.07	0.93

Table 3. Results of regime estimation for four experiments under new conditions unseen by the model. The true regime of each experiment is shown on the regime map in Fig. 1b.

The capability of the model to estimate the regime of droplet is evaluated on an estimation data set from conditions that are unseen by the model and do not contribute to the model training, testing, and validation. The values of *RH* and *T* for these conditions are randomly selected in the range of $20\% < RH < 80\%$ and $15\text{ °C} < T < 35\text{ °C}$ (shown with cross and star marks on the regime map of Fig. 1b). The model classifies each experiment under each regime with different ratios as shown in Table 3. For example, Experiment 1 with *RH* of 30% and *T* of 15 °C is close to the boundary of evaporation-dominated and transition regimes. With NB, 54% of the data points in this experiment are classified as evaporation-dominated regime and 46% as transition regime. This is expected due to the location of Experiment 1 on the regime map. It should be noted that the lines on the regime map are approximate boundaries. Figure 2d demonstrates regime estimation of all data points of experiment 4 with both classifiers. The shown evolution of contact angle is not similar to any of the evolutions shown in Fig. 1b subfigures. In fact, contact angle decreases at the beginning and then starts rising. Experiment #4 falls in transition regime. That is the reason that the profile for experimental data in Fig. 2d is colored in red for the true regime of experiment #4. As it is seen, NB and DT correctly classify 84 and 93% of the data points in experiment 4 to transition regime. The green regions represent the data points that are incorrectly classified as evaporation-dominated regime.

Relative humidity estimation. In this section, we show the ability of the model to detect environmental humidity by analyzing the evolution of contact angle and diameter through regression algorithms. Polynomial regression with four different orders (linear, quadratic, third-order, and fourth-order) and regression tree are used for training. The coefficient of determination (R^2) increases from 0.66 for linear up to 0.93 for fourth-order polynomial regression. The test results for all five regression methods are shown in Fig. 3a. The horizontal axis shows the true value of RH^* and the vertical axis shows the estimated values averaged over all the points for each RH^* . As it can be seen, the higher the order of polynomial regression, the closer the average estimation to the ground-truth value and the smaller the error bar. Furthermore, regression tree performs more accurately compared to all polynomial regression methods. Model performance under each specific condition through validation set is shown in Fig. 2b. The consistent colors throughout Fig. 3a–c represent different regression methods and different markers are used for each substrate temperature in Fig. 3b. Based on the results shown in Fig. 3b, all methods except linear regression produce reasonably accurate results. The validation results get closer to actual values as the order of the polynomial regression increases. Nonetheless, it must be noted that higher order polynomial regression increases the computational cost as well as the chance of over-fitting. Performance of regression tree is comparable to third-order and fourth-order polynomial. The capability of different regression methods to estimate new humidity values by analyzing the time evolution of contact angle and diameter is presented in Fig. 3c. The new relative humidity values (30, 33, 65, and 75%) are randomly selected in the range of 20–80%. It is noteworthy that the model has not seen any data of droplet evolution under these *RH* values during training, testing, or validation. It is seen that, unlike testing and validation where increasing the order of polynomial or complexity of the model (i.e., regression tree) produces more accurate results, higher order polynomials do not result in better estimation of unseen conditions. As a matter of fact, linear, quadratic, and third-order polynomials estimate more accurately. This is a common issue when the model fits the training data very well and it negatively affects the model performance on the new data set. Figure 3c clearly indicates over-fitting with fourth-order polynomial.

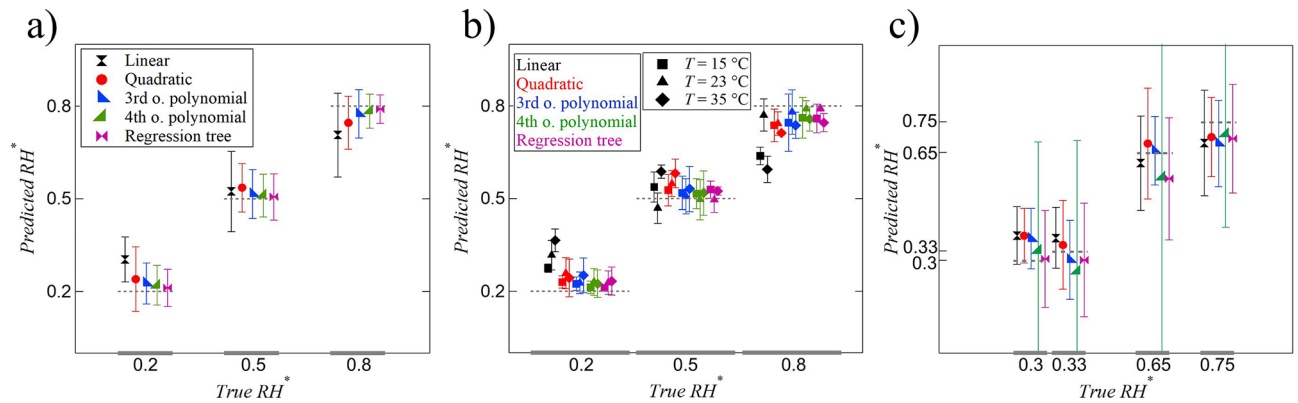


Figure 3. Relative humidity regression results: (a) test set; (b) validation; (c) estimation. Markers for all colors in (b) represent different temperatures as shown in legend. Markers and colors in (c) are the same as in (a).

Diameter and contact angle estimation. In this section, the capability of the model to estimate the continuous evolution of contact angle and diameter over time is evaluated through regression algorithms. This means that the model estimates the evolution of diameter and contact angle at each time increment (approximately one second apart). The input variables include T^* , RH^* , t^* , and θ^* (or D^*) and the target variable is D^* (or θ^*). By increasing the order of the polynomial, the coefficient of determination, R^2 , for training improves from 0.87 to 0.99, and from 0.78 to 0.96 for diameter and contact angle estimation, respectively. The performance of five different regression methods on the test set is presented in Fig. 4a. The first row represents the results when D^* is the target variable and the second row illustrates the results for θ^* as the target variable. The closer the distribution of data to the diagonal line in these plots, the better the performance of the model on the test set. Based on the results shown in Fig. 4a, the diameter test results saturate after third order polynomial while for contact angle the performance keeps improving when increasing the degree of polynomial from third to fourth.

The validation results are summarized in Fig. 4b for D^* (top) and θ^* (bottom) in terms of R^2 and root mean square error (*rmse*). With D^* being the target variable, an average R^2 of 0.8 or higher and average *rmse* less than 0.1 are achieved for all nine conditions. Going from linear to quadratic to third order polynomial increases and decreases the value of R^2 and *rmse*, respectively. The profiles of R^2 and *rmse* exhibit saturation, and further increase in the order of the polynomial does not improve model performance on validation data. This is consistent with the test results where model performance saturates at third order. Furthermore, regression tree demonstrates accuracy comparable to third-order and fourth-order polynomials. By comparing the range of axes in Fig. 4b-top with bottom, it is obvious that R^2 values are generally lower (hardly reaching 0.7) and error is higher when θ^* is the target variable. In fact, there are a few instances where the average R^2 turns negative, suggesting that the overall estimation of the model is worse than an estimation with a constant average value.

The performance of the model on estimating the evolution of θ^* and D^* versus time under four new conditions that did not contribute to model training, testing, or validation is shown in Fig. 4c. One value of R^2 and *rmse* are reported for each condition (or experiment) which shows the overall quality of the fit. Higher coefficients of determination and lower *rmse* values demonstrate the better performance of the model in estimating D^* than θ^* evolution. Based on the results shown in Fig. 4c, third order polynomial regression has the best performance in estimating diameter. The results become less accurate with fourth-order polynomial which suggests over-fitting. It is interesting to note that regression tree, which had higher accuracy during testing and validation, is outperformed even by linear regression during estimation. The evolution of diameter versus time estimated by quadratic regression for Experiment 3 is depicted in Fig. 4d. As it can be seen, even with a quadratic regression, the model estimates the evolution of D^* quite accurately for an unseen condition. Considering the range of values on axes of Fig. 4c-top and bottom, estimating the evolution of contact angle is more challenging for the model. Unlike estimating diameter, increasing the order of polynomials has a negligible effect. The accuracy of the model stays almost constant for linear, quadratic, and third order. However, it worsens drastically for fourth-order polynomial due to over-fitting the data. The R^2 and *rmse* values for contact angle estimation with fourth-order polynomial fall outside the range shown in the plot. Since the estimation of θ^* is generally more challenging for the model, the effect of over-fitting is more noticeable compared to D^* estimation. The overall better performance of the model for diameter estimation compared to contact angle estimation is due to the fact that diameter evolution is relatively smooth and therefore easier to estimate whereas contact angle evolution changes substantially under different conditions. Figure 4d-bottom illustrates contact angle evolution over time estimated with third-order polynomial for Experiment 1 which is in the evaporation-dominated regime. Although the maximum value of contact angle is underestimated with 3rd order polynomial, the time corresponding to the maximum contact angle is estimated accurately. Also, there is a good agreement between the estimated and actual contact angle evolution during most of droplet lifetime with R^2 value of 0.74 and *rmse* of 0.05.

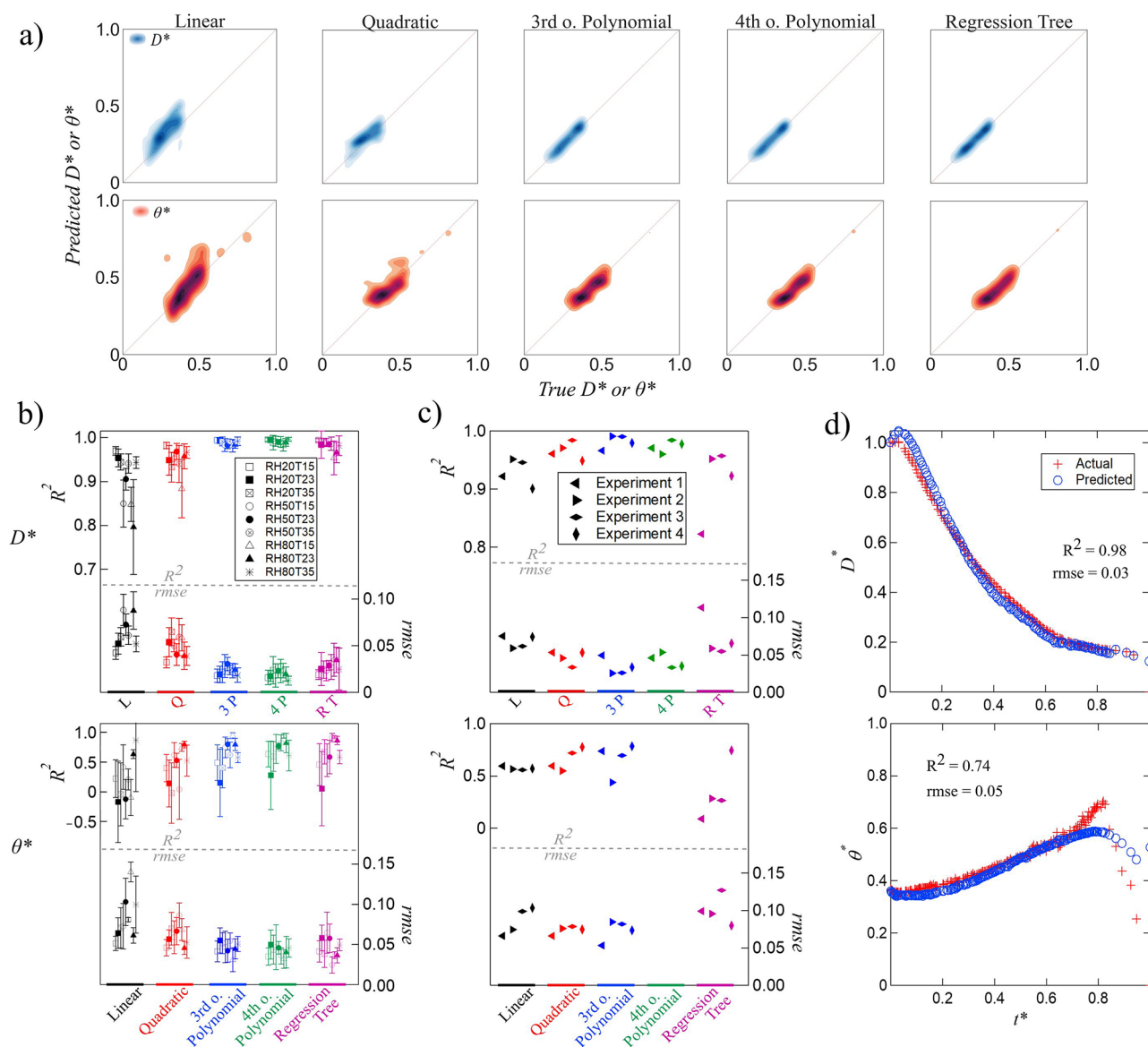


Figure 4. Diameter (D^*) and contact angle (θ^*) regression results: **(a)** test set; **(b)** validation set; **(c)** estimation set; **(d)** diameter estimation with quadratic regression for E3 (top) and contact angle estimation with third-order polynomial regression for E1 (bottom).

Discussion

In this study, we have analyzed the complex physics of sessile droplet evaporation (which starts with a single component and turns into a binary system due to the transfer of a second component i.e., water) through machine learning, classification and regression algorithms. Four different parameters pertaining to droplet evaporation, namely: regime of droplet evaporation, level of surrounding humidity, time evolution of droplet base diameter, and time evolution of droplet contact angle, are estimated. Point-by-point analysis of droplet profile enables time-dependent estimations given a limited number of data points. This means that the model does not need the entire evolution profile of a droplet to make an estimation. Instead, only a few (or even a single) data points are (is) sufficient for estimation, although more data points result in more accurate estimation. The model estimation capability is then assessed on the data that do not contribute to training, testing, or validation.

Two different classifiers are utilized to estimate the regime of droplet evaporation. NB is chosen as a simple easy-to-interpret algorithm, while DT serves as a more powerful algorithm. As expected, DT outperforms NB due to a more robust internal structure at the expense of computational cost and transparency. Both classifiers showed impressive performance (minimum 75% accuracy) on estimating the regime of droplet. Knowledge on droplet evaporation regime is necessary for compatible designs in numerous industries such as droplet-based biosensors or ink-jet printing.

Additionally, level of surrounding humidity, time evolution of droplet base diameter, and time evolution of droplet contact angle are estimated through regression techniques. Polynomial regressors, as well as regression

tree, are trained through point-by-point analysis of droplet evolution. The model performance improved by increasing the order of the polynomial and using regression tree for training, test, and validation sets. However, when estimating the new conditions unseen by the model, fourth-order polynomial and regression tree suffered from data over-fitting. The best performance of the model is achieved by third-order polynomial. In general, the model estimation results are more accurate when estimating diameter evolution compared to contact angle estimation. This is due to smoother, hence easier to estimate, evolution of diameter with time. The sharp changes in θ^* under different conditions make the estimation of its evolution challenging for the model. Information of this type is of great importance for technologies such as ink-jet printing, or droplet-based biodiagnostics where the estimations can provide critical information on the base diameter or contact angle of the droplet at a specific time.

In the current work, behavior of a sessile droplet transitioning from a single fluid into a binary mixture following transfer of a second component from the atmosphere is studied through data-driven techniques. The model demonstrated promising performance detecting the regime of droplet evolution, the humidity level surrounding droplet, and time evolution of diameter and contact angle. The current case study demonstrates the capability of the proposed model to analyze complex interfacial fluid mechanics problems through machine learning algorithms. Although we have estimated four parameters of droplet evaporation by this model, these kinds of techniques can be expanded to perform a wide range of estimations. Our preliminary study opens up new ways to study binary or multi-component droplet evolution which might lead to better analyzing the complex physics of the problem.

Methods

Experimental setup and procedures. The experiments are carried out in a chamber (with dimensions $127 \times 127 \times 76 \text{ mm}^3$) of drop shape analyzer (DSA 100) from KRÜSS. The relative humidity inside the chamber is controlled and the temperature is kept at room temperature (i.e., $23 \text{ }^\circ\text{C}$). The bottom side of the chamber is equipped with a Peltier plate (electrical system and a temperature bath) to control the substrate temperature. The top of the chamber has a small hole for passing the syringe. Both sides of the chamber have transparent windows for visualization purposes. An LED light is used for illumination and a CCD camera is utilized to capture the time evolution of droplet profile (Fig. 1a). The whole setup is mounted on an anti-vibration table to eliminate environmental disturbances.

Methanol is purchased from Fisher Scientific with a purity of 99.8%. The glass substrates are coated with a very thin PDMS (polydimethylsiloxane) layer to achieve spherical and reproducible droplets with measurable contact angles. In order to ensure that methanol does not interact with the PDMS coating, multiple methanol droplets are successively deposited at the same location on the substrate and let evaporate. No change is observed in the initial contact angle of droplets or in their evolution during evaporation. The relative humidity inside the chamber and the temperature of the substrate are set to desired values and enough time is passed to make sure quasi-steady state is achieved. Three values of relative humidity: 20%, 50%, and 80% alongside three values of substrate temperature: $15 \text{ }^\circ\text{C}$, $23 \text{ }^\circ\text{C}$ (room temperature), and $35 \text{ }^\circ\text{C}$ are tested. A drop of methanol is gently deposited on the glass substrate. Droplet volume is under $5 \mu\text{l}$ in order to keep droplet size below capillary length. The corresponding volume to the capillary length for methanol droplet is around $45 \mu\text{l}$. The evaporation process of methanol droplet is recorded by the CCD camera at 50 frames per second. All experiments under each relative humidity and substrate temperature condition are repeated five to ten times to ensure the reproducibility of the data. A KRÜSS Drop Shape Analyzer software is utilized to measure the time evolution of contact angle (θ), base diameter (D), and volume (V). Due to the observed differences between the right and left contact angle values, elliptical fit is used for contact angle and volume measurements. For better analysis of droplet behavior, all dimensional parameters are nondimensionalized as followed: $D^* = D/D_0$, $V^* = V/V_0$, $t^* = t/t_f$; where D_0 , V_0 , and t_f are initial diameter, initial volume, and total evaporation time, respectively.

Data acquisition. The data for the model is generated by experiments of methanol droplet evaporation under various environmental conditions. Nine different conditions are created by a combination of three levels of surrounding relative humidity: 20%, 50%, and 80% with three substrate temperatures: $15 \text{ }^\circ\text{C}$, $23 \text{ }^\circ\text{C}$, and $35 \text{ }^\circ\text{C}$. For each of these nine conditions, the relative humidity in the chamber and substrate temperature is set to the desired values and enough time is passed to ensure quasi-steady state. Then a droplet of methanol is gently deposited on the substrate and the evolution of droplet base diameter, contact angle, and volume is recorded over time with a CCD camera of Drop Shape Analyzer (see Fig. 1a inset). Each point in time with corresponding contact angle, diameter, and volume is considered a data point for the model.

Data partitioning and processing. The data generated under nine different conditions mentioned in the previous section are used to train, test, and validate the model. The droplet evaporation experiments for each of these nine conditions are repeated five to ten times to make sure that the results are consistent and reproducible. We carried out 60 droplet evaporation experiments under these nine conditions which created a total number of 10,850 data points that are used for training, testing and validating the model. There are additional 761 data points that are used to assess the performance of the model at the final step. The environmental conditions under which these data points are generated are completely different than the other 10,850 data points used for training, testing, and validating the model.

Training and testing. The data for training, testing, and validation are generated by methanol droplet evaporation under nine different conditions which are the combinations of three levels of surrounding humidity: 20%, 50%, and 80% with three substrate temperatures: $15 \text{ }^\circ\text{C}$, $23 \text{ }^\circ\text{C}$, and $35 \text{ }^\circ\text{C}$. The behavior of droplet under the same condition is similar between experiments however it is not exactly the same due to microscopic defects

of the surface, initial conditions, etc. The data from a single experiment is held out at the beginning of training as a validation set. The remaining data from 59 experiments is partitioned into a training set and a test set with 80–20% ratio. It should be noted that multiple training to test set ratios are examined to ensure the convergence of the model which is discussed in the “Performance criteria” section.

Cross validation. A 10-fold cross-validation is carried out. The results for the test set show one of 10 scenarios from the test set. Cross-validation is performed to ensure all the results for training and test set are similar to each other and there is no anomaly in the data.

Validation. Once the model is trained and tested, it is then examined on the validation set that was held out at the beginning of training. The validation set that is kept out, is the data pertaining to one whole experiment out of 60 experiments. This means that the model has not seen the data for the specific experiment in the validation set. Validation is performed to specifically evaluate the performance of the model under each specific condition of surrounding humidity and substrate temperature. This procedure is repeated 60 times until each experiment is held out once and validated. Validation results are averaged over all experiments under each specific RH and T condition.

State estimation. There are additional 761 data points, called estimation set, that are generated by four experiments of methanol droplet evaporation. It must be noted that the machine has not seen any data of droplet evaporation under these new conditions. The data in the estimation set does not contribute to the framework during training, testing, or validation. The surrounding humidity and substrate temperature for these new conditions are randomly selected to be in the range within which the model is trained (i.e., $20\% < RH < 80\%$ and $15\text{ }^\circ\text{C} < T < 35\text{ }^\circ\text{C}$).

Performance criteria. The performance criteria of the model are reported by standard metrics. For classification, the confusion matrix of the model is used as well as the precision, recall, F score, and overall accuracy values for each regime. For validation, the accuracy values are the same as recall values for each combination of RH and T because there is only one true regime for each validation set. For regression methods, coefficient of determination (R^2) is reported to show how well the model fits the data in the training set. For RH testing, validation, and estimation, the actual RH of the environment is compared against the estimated value of RH . When D^* (or θ^*) is the target variable, the test results are demonstrated as estimated values versus actual values. The validation and estimation results are reported by R^2 that shows the quality of the fit; the proportion of variance in the target variable which is predictable from the input variables; and root mean squared error ($rmse$) which represents how much the estimation is off on average when estimating the average target variable. We have also tested the model convergence for both classification and regression algorithms. We trained the model by considering only fractions of the available data i.e., 80%, 70%, 60%, and 50%, and observed that the results for the test set remain unchanged. This proves the fact that the number of data points used for machine learning of this specific problem is converged.

Classifiers. We have used Naïve Bayes Classifier^{64–67} as a simple and easy-to-interpret algorithm. Since the algorithm is simple, there is less chance for over-fitting the data, it is faster and needs a smaller memory footprint. However, the restrictive underlying assumptions compromise its accuracy for real case scenarios when the variables are not fully independent of each other.

Bagged Decision Tree^{68–70} with 250 trees is also used. It is a powerful classifier with built-in support for cross-validation and a specialized function to measure feature importance. However, it results in complex models that are not very transparent. It is often hard to understand how it makes estimations.

Received: 20 February 2021; Accepted: 28 May 2021

Published online: 30 June 2021

References

- Bonn, D., Eggers, J., Indekeu, J., Meunier, J. & Rolley, E. Wetting and spreading. *Rev. Mod. Phys.* **81**, 739 (2009).
- Plawsky, J. *et al.* Nano- and microstructures for thin-film evaporation—a review. *Nanos. Microsc. Thermophys. Eng.* **18**, 251–269 (2014).
- Andalib, S. & Kavehpour, P. Fluorescence microscopy of precursor films in evaporating droplets. *APS Div. Fluid Dyn. Meet. Abstr.* **20**, Q11-003 (2017).
- De Gans, B.-J., Duineveld, P. C. & Schubert, U. S. Inkjet printing of polymers: State of the art and future developments. *Adv. Mater.* **16**, 203–213 (2004).
- Kawase, T., Siringhaus, H., Friend, R. H. & Shimoda, T. Inkjet printed via-hole interconnections and resistors for all-polymer transistor circuits. *Adv. Mater.* **13**, 1601–1605 (2001).
- Park, J. & Moon, J. Control of colloidal particle deposit patterns within picoliter droplets ejected by ink-jet printing. *Langmuir* **22**, 3506–3513 (2006).
- Smith, P. J. & Shin, D. H. *Inkjet-Based Micromanufacturing* (Wiley, 2012).
- Shedd, T. & Pautsch, A. Spray impingement cooling with single- and multiple-nozzle arrays. Part II. Visualization and empirical models. *Int. J. Heat Mass Transf.* **48**, 3176–3184 (2005).
- Sodtke, C. & Stephan, P. Spray cooling on micro structured surfaces. *Int. J. Heat Mass Transf.* **50**, 4089–4097 (2007).
- Molle, B., Tomas, S., Hendawi, M. & Granier, J. Evaporation and wind drift losses during sprinkler irrigation influenced by droplet size distribution. *Irrig. Drain.* **61**, 240–250 (2012).
- Hurth, C. *et al.* Biomolecular interactions control the shape of stains from drying droplets of complex fluids. *Chem. Eng. Sci.* **137**, 398–403 (2015).

12. Trantum, J. R., Wright, D. W. & Haselton, F. R. Biomarker-mediated disruption of coffee-ring formation as a low resource diagnostic indicator. *Langmuir* **28**, 2187–2193 (2012).
13. Gulka, C. P. *et al.* Coffee rings as low-resource diagnostics: Detection of the malaria biomarker plasmodium falciparum histidine-rich protein-ii using a surface-coupled ring of ni (ii) nta gold-plated polystyrene particles. *ACS Appl. Mater. Interfaces* **6**, 6257–6263 (2014).
14. Cazabat, A.-M. & Guena, G. Evaporation of macroscopic sessile droplets. *Soft Matter* **6**, 2591–2612 (2010).
15. Shahidzadeh-Bonn, N., Rafai, S., Azouni, A. & Bonn, D. Evaporating droplets. *J. Fluid Mech.* **549**, 307 (2006).
16. Brutin, D. & Starov, V. Recent advances in droplet wetting and evaporation. *Chem. Soc. Rev.* **47**, 558–585 (2018).
17. Andalib, S. & Kavehpour, P. Effect of evaporation rate on stick-slip behavior of the contact line. *Bull. Am. Phys. Soc.* **63**, 20 (2018).
18. Andalib, S. & Kavehpour, P. Experimental investigation of evaporation rate of a droplet on behavior of the contact line. In *19th International Coating Science and Technology Symposium* (2018).
19. Zang, D., Tarafdar, S., Tarasevich, Y. Y., Choudhury, M. D. & Dutta, T. Evaporation of a droplet: From physics to applications. *Phys. Rep.* **804**, 1–56 (2019).
20. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
21. Hu, H. & Larson, R. G. Evaporation of a sessile droplet on a substrate. *J. Phys. Chem. B* **106**, 1334–1344 (2002).
22. Mollaret, R., Sefiane, K., Christy, J. & Veyret, D. Experimental and numerical investigation of the evaporation into air of a drop on a heated surface. *Chem. Eng. Res. Des.* **82**, 471–480 (2004).
23. Sefiane, K. & Tadrif, L. Experimental investigation of the de-pinning phenomenon on rough surfaces of volatile drops. *Int. Commun. Heat Mass Transfer* **33**, 482–490 (2006).
24. Sefiane, K. The coupling between evaporation and adsorbed surfactant accumulation and its effect on the wetting and spreading behaviour of volatile drops on a hot surface. *J. Petrol. Sci. Eng.* **51**, 238–252 (2006).
25. Ristenpart, W., Kim, P., Domingues, C., Wan, J. & Stone, H. A. Influence of substrate conductivity on circulation reversal in evaporating drops. *Phys. Rev. Lett.* **99**, 234502 (2007).
26. David, S., Sefiane, K. & Tadrif, L. Experimental investigation of the effect of thermal properties of the substrate in the wetting and evaporation of sessile drops. *Colloids Surf. A* **298**, 108–114 (2007).
27. Sobac, B. & Brutin, D. Thermal effects of the substrate on water droplet evaporation. *Phys. Rev. E* **86**, 021602 (2012).
28. Sefiane, K., Wilson, S. K., David, S., Dunn, G. J. & Duffy, B. R. On the effect of the atmosphere on the evaporation of sessile droplets of water. *Phys. Fluids* **21**, 062101 (2009).
29. Sobac, B. & Brutin, D. Thermocapillary instabilities in an evaporating drop deposited onto a heated substrate. *Phys. Fluids* **24**, 032103 (2012).
30. Sefiane, K., Tadrif, L. & Douglas, M. Experimental study of evaporating water–ethanol mixture sessile drop: Influence of concentration. *Int. J. Heat Mass Transf.* **46**, 4527–4534 (2003).
31. Sefiane, K., David, S. & Shanahan, M. E. Wetting and evaporation of binary mixture drops. *J. Phys. Chem. B* **112**, 11317–11323 (2008).
32. Liu, C., Bonaccorso, E. & Butt, H.-J. Evaporation of sessile water/ethanol drops in a controlled environment. *Phys. Chem. Chem. Phys.* **10**, 7150–7157 (2008).
33. Innocenzi, P. *et al.* Evaporation of ethanol and ethanol–water mixtures studied by time-resolved infrared spectroscopy. *J. Phys. Chem. A* **112**, 6512–6516 (2008).
34. Chen, P. *et al.* Evaporation of binary sessile drops: Infrared and acoustic methods to track alcohol concentration at the interface and on the surface. *Langmuir* **32**, 9836–9845 (2016).
35. Fukatani, Y. *et al.* Effect of ambient temperature and relative humidity on interfacial temperature during early stages of drop evaporation. *Phys. Rev. E* **93**, 043103 (2016).
36. Kita, Y. *et al.* Quantifying vapor transfer into evaporating ethanol drops in a humid atmosphere. *Phys. Chem. Chem. Phys.* **20**, 19430–19440 (2018).
37. Andalib, S., Alshehri, A. & Kavehpour, P. Combined effect of relative humidity and substrate temperature on evaporation of methanol droplet. *J. Coat. Technol. Res.* **16**, 1691–1698 (2019).
38. Andalib, S., Alshehri, A. & Kavehpour, P. Effect of a second component in organic droplet evaporation: Initially present versus absorbed during the process. *APS Di. Fluid Dyn. Meet. Abstr.* **M04–031**, 20 (2019).
39. Tan, H. *et al.* Evaporation-triggered microdroplet nucleation and the four life phases of an evaporating ouzo drop. *Proc. Natl. Acad. Sci.* **113**, 8642–8647 (2016).
40. Diddens, C. *et al.* Evaporating pure, binary and ternary droplets: Thermal effects and axial symmetry breaking. *J. Fluid Mech.* **823**, 470–497 (2017).
41. Li, Y. *et al.* Evaporation-triggered segregation of sessile binary droplets. *Phys. Rev. Lett.* **120**, 224501 (2018).
42. Lohse, D. & Zhang, X. Physicochemical hydrodynamics of droplets out of equilibrium. *Nat. Rev. Phys.* **2**, 426–443 (2020).
43. Brenner, M., Eldredge, J. & Freund, J. Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids* **4**, 100501 (2019).
44. Brunton, S. L., Noack, B. R. & Koumoutsakos, P. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508 (2020).
45. Duriez, T., Brunton, S. L. & Noack, B. R. *Machine Learning Control-Taming Nonlinear Dynamics and Turbulence* Vol. 116 (Springer, 2017).
46. Singh, A. P., Medida, S. & Duraisamy, K. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.* **55**, 2215–2227 (2017).
47. Duraisamy, K., Iaccarino, G. & Xiao, H. Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377 (2019).
48. Fukami, K., Fukagata, K. & Taira, K. Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120 (2019).
49. Fukami, K., Fukagata, K. & Taira, K. Assessment of supervised machine learning methods for fluid flows. *Theoret. Comput. Fluid Dyn.* **34**, 497–519 (2020).
50. Jadrlich, R., Lindquist, B. & Truskett, T. Unsupervised machine learning for detection of phase transitions in off-lattice systems i foundations. *J. Chem. Phys.* **149**, 194109 (2018).
51. Popov, P. P. *et al.* Machine learning-assisted early ignition prediction in a complex flow. *Combust. Flame* **206**, 451–466 (2019).
52. Raissi, M., Wang, Z., Triantafyllou, M. S. & Karniadakis, G. E. Deep learning of vortex-induced vibrations. *J. Fluid Mech.* **861**, 119–137 (2019).
53. Verma, S., Novati, G. & Koumoutsakos, P. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl. Acad. Sci.* **115**, 5849–5854 (2018).
54. Hou, W., Darakananda, D. & Eldredge, J. D. Machine-learning-based detection of aerodynamic disturbances using surface pressure measurements. *AIAA J.* **57**, 5079–5093 (2019).
55. Yakhno, T. A. *et al.* The informative-capacity phenomenon of drying drops. *IEEE Eng. Med. Biol. Mag.* **24**, 96–104 (2005).
56. Killeen, A. A. *et al.* Protein self-organization patterns in dried serum reveal changes in b-cell disorders. *Mol. Diagn. Ther.* **10**, 371–380 (2006).
57. Shabalin, V. & Shatokhina, S. Diagnostic markers in the structures of human biological liquids. *Singapore Med. J.* **48**, 440 (2007).
58. Brutin, D., Sobac, B., Loquet, B. & Sampol, J. Pattern formation in drying drops of blood. *J. Fluid Mech.* **667**, 85 (2011).
59. Hamadeh, L. *et al.* Machine learning analysis for quantitative discrimination of dried blood droplets. *Sci. Rep.* **10**, 1–13 (2020).

60. Kim, N. *et al.* Identification of fluid and substrate chemistry based on automatic pattern recognition of stains. *Anal. Methods* **4**, 50–57 (2012).
61. Gorr, H. M., Xiong, Z. & Barnard, J. A. Pattern recognition for identification of lysozyme droplet solution chemistry. *Colloids Surf. B* **115**, 170–175 (2014).
62. Bhardwaj, R., Longtin, J. P. & Attinger, D. A numerical investigation on the influence of liquid properties and interfacial heat transfer during microdroplet deposition onto a glass substrate. *Int. J. Heat Mass Transf.* **50**, 2912–2923 (2007).
63. Harrington, B. Inkscape. <http://www.inkscape.org> (2004–2005).
64. Buntine, W. L. *A theory of learning classification rules*. Ph.D. thesis, university of technology, sydney (1990).
65. Kononenko, I. Comparison of inductive and Naive Bayesian learning approaches to automatic knowledge acquisition. In *Current Trends in Knowledge Acquisition* (B. Wielinga, ed) (1990).
66. Langley, P. In *European Conference on Machine Learning* 153–164 (Springer, 1993).
67. Langley, P. & Sage, S. Induction of selective Bayesian classifiers. In *Uncertainty Proceedings 1994* 399–406 (Elsevier, 1994).
68. Swain, P. H. & Hauska, H. The decision tree classifier: Design and potential. *IEEE Trans. Geosci. Electron.* **15**, 142–147 (1977).
69. Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **21**, 660–674 (1991).
70. Rao, J. S. & Potts, W. J. Visualizing bagged decision trees. *AAAI Knowl. Discov. Databases Proc.* **20**, 243–246 (1997).

Acknowledgements

We thank Dr. Nima Jabbari for helpful discussions and preparation of Fig. 4a. We also express our gratitude to Prof. Bud Homby for his invaluable comments on an earlier version of the manuscript. This work was supported by a fellowship from the Mechanical and Aerospace Engineering Department at the University of California Los Angeles.

Author contributions

S.A., K.T., and H.P.K. designed research, K.T. and H.P.K. supervised research, S.A. performed research, analyzed data, and wrote the paper. All authors reviewed the manuscript.

Funding

This research was supported by the National Science Foundation under Grant No. CBET-2032533.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021