

RESEARCH ARTICLE

Robust interpolation for dispersed gas-droplet flows using statistical learning with the fully Lagrangian approach

C. P. Stafford  | O. Rybdylova

Advanced Engineering Centre, School of Architecture, Technology and Engineering, University of Brighton, Brighton, UK

Correspondence

C. P. Stafford, Advanced Engineering Centre, School of Architecture, Technology and Engineering, University of Brighton, Brighton BN2 4GJ, UK.
Email: c.stafford@brighton.ac.uk

Funding information

UK Research and Innovation,
Grant/Award Number: MR/T043326/1

Summary

A novel methodology is presented for reconstructing the Eulerian number density field of dispersed gas-droplet flows modelled using the fully Lagrangian approach (FLA). In this work, the nonparametric framework of kernel regression is used to accumulate the FLA number density contributions of individual droplets in accordance with the spatial structure of the dispersed phase. The high variation which is observed in the droplet number density field for unsteady flows is accounted for by using the Eulerian-Lagrangian transformation tensor, which is central to the FLA, to specify the size and shape of the kernel associated with each droplet. This procedure enables a high level of structural detail to be retained, and it is demonstrated that far fewer droplets have to be tracked in order to reconstruct a faithful Eulerian representation of the dispersed phase. Furthermore, the kernel regression procedure is easily extended to higher dimensions, and inclusion of the droplet radius within the phase space description using the generalised fully Lagrangian approach (gFLA) additionally enables statistics of the droplet size distribution to be determined for polydisperse flows. The developed methodology is applied to a range of one-dimensional and two-dimensional steady-state and transient flows, for both monodisperse and polydisperse droplets, and it is shown that kernel regression performs well across this variety of cases. A comparison is made against conventional direct trajectory methods to determine the saving in computational expense which can be gained, and it is found that 10^3 times fewer droplet realisations are needed to reconstruct a qualitatively similar representation of the number density field.

KEYWORDS

fully Lagrangian approach, kernel regression, polydisperse droplets

1 | INTRODUCTION

Gas-droplets flows are ubiquitous within science and engineering, and are central to many industrial applications and environmental phenomena, including facilitating the decarbonisation of propulsion systems for transportation,¹ determining the spread of airborne droplets containing a pathogen,² and modelling droplet condensation within

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *International Journal for Numerical Methods in Fluids* published by John Wiley & Sons Ltd.

atmospheric clouds.³ Of primary interest in the characterisation and description of such systems is the variation of droplet number density throughout the flows, with this being a key determinant in physical factors such as reaction rates and probability of infection. The mechanisms that are responsible for the variation in droplet number density have received much attention, with studies considering both the influence of the carrier flow structures and droplet physics upon the clustering and dispersion of droplets.

Whilst experimental studies continue to be an important informant of droplet behaviour through the provision of quantities such as the distribution of droplet velocities and sizes, it is useful to be able to study specific flow configurations without having to use an experimental setup. Mathematical and numerical modelling is able to fulfil this role with the aid of computational simulations, and provides a wealth of detailed information that can be difficult to obtain experimentally, such as knowledge of the instantaneous droplet number density field. Such modelling is able to provide detailed insight into the physical phenomena which occur in droplet-laden flows, and constitutes a topic of growing research interest.

Within the realm of modelling gas-droplet flows two distinct approaches exist for handling the droplet phase. The first treats the droplets as a continuum in a similar manner to the gas carrier flow, and solves a set of coupled partial differential equations which govern the evolution of the droplet phase field variables, specifically the number density, mean velocity, and kinetic stresses. By its nature, the resolution of this approach is limited by the fidelity of the computational grid upon which the equations are solved, and consequently the resultant Eulerian description of the droplet phase is restricted in the level of detail that can be represented.⁴ In contrast, droplets can instead be treated individually, with the equation of motion for each droplet being solved along their separate Lagrangian trajectories. This method is able to capture a considerable level of detail, with the behaviour of droplets in different regions of the flow being accurately accounted for through their individual treatment, however the drawback to the Lagrangian approach lies in construction of the Eulerian droplet number density field from the trajectories. In order to obtain statistical convergence of the resulting Eulerian description, the number of droplets required is often prohibitive, being of the order of $\mathcal{O}(10^3)$ in every grid cell for a well resolved flow.⁵

A point of contention between the Eulerian and Lagrangian descriptions of the droplet phase is in the capturing of important physical phenomena such as the crossing trajectories effect,⁶ in which the non-negligible inertia of droplets can cause their trajectories to intersect in physical space. This has the consequence of the droplet phase field variables becoming multi-valued; a consideration that is important to account for in the modelling approach. Such instances of multi-valuedness can be observed in the structure of the dispersed phase continuum where they manifest as the envelope of intersecting trajectories, which is referred to as a *fold*. Eulerian descriptions of the droplet field are by construction single-valued at a given point in space, and therefore unable to include the full detail of the crossing trajectories effect in their standard form. To address this, studies have sought to account for this behaviour directly within the Eulerian representation of the droplet phase field variables by means of developments such as a family of moment methods⁴ and multi-fluid models.⁷ In contrast, the Lagrangian description of droplets along their trajectories is naturally able to handle the intersection of trajectories, and therefore also the ensuing multi-valued nature of the droplet field, making it an ideal candidate for the development of further models.

To utilise the detail intrinsic to the Lagrangian description and address the issue of needing to simulate prohibitively high numbers of droplets, a body of work has been developed upon the assumption that the droplet phase can also be treated as a continuum with no self-stresses.⁸ This has been pioneered by Osipov,⁹ and involves tracking the number density along individual trajectories by utilising the conservation of mass in Lagrangian form, a procedure known as the fully Lagrangian approach (FLA; some works refer to this as the fully Lagrangian method, or Osipov's method). A number of subsequent studies have built upon this foundation, including its application to a gas flow through a turbine,⁵ examination of the moments of droplet number density,¹⁰ comparison to experimental spray data,¹¹ applicability to turbulent flows,¹² inclusion of momentum coupling effects,¹³ and most recently, extension to the consideration of polydisperse flows.¹⁴ Despite these advances, the FLA remains a developing concept which presents a variety of challenges within its use, chief of which is the accurate reconstruction of the droplet number density field from trajectory data.

In this paper, the use of a statistical learning approach is proposed to accumulate the contributions from individual droplets, and reproduce the Eulerian number density field in a robust manner which is applicable across a range of different flows. To accomplish this, kernel regression is used, with the averaged nature of this procedure enabling it to account for the multi-valued nature of the contributions along folds, and thereby construct a single-valued representation of the number density field. This extends the linear interpolation approach used in a previous study,¹⁴ and makes accurate representation of the dispersed phase-field possible for a wider class of flows. The developed methodology is applied to a range of one-dimensional and two-dimensional steady-state and transient flows, for both monodisperse and polydisperse droplets, and it is shown that kernel regression performs well across this variety of cases.

The remainder of the paper is structured as follows. Section 2 outlines the formulation of the fully Lagrangian approach, and details the assumptions involved. Section 3 introduces the concept of kernel regression, how it can be applied to the calculation of number density, and how the information provided by the fully Lagrangian approach can be used to ensure contributions from trajectories are accumulated in a physically consistent manner. It is further demonstrated how the procedure can be generalised into a higher-order phase space that includes droplet radius to enable modelling of polydisperse flows. Additionally, aspects of the computational implementation are detailed. Section 4 describes application of the kernel regression approach to a wide class of flow configurations and presents the numerical results. Section 5 appraises the observed simulation results, and discusses the potential of the procedure for application to more complex flows in terms of computational efficiency.

2 | THE FULLY LAGRANGIAN APPROACH

2.1 | General formulation for polydisperse droplets

Representation of the dispersed phase in a Lagrangian sense is considered through the use of an appropriate governing equation for individual droplets or particles. Henceforth, in this paper droplets are taken to be liquid and can be subject to evaporation, whilst particles are assumed to be solid and non-evaporating; the more general case of droplets is usually focused upon. In this work, droplets are considered to be spherical, within a dilute suspension, and are also assumed to act as point masses. Under these conditions, droplets with position $\mathbf{x}_d(t)$, velocity $\mathbf{v}_d(t)$ and radius $r_d(t)$ are modelled individually along trajectories according to the arbitrary governing laws and initial conditions at time t_0

$$\dot{\mathbf{x}}_d(t) = \mathbf{f}(\mathbf{x}_d(t), \mathbf{v}_d(t), r_d(t), t), \quad \mathbf{x}_d(t_0) = \mathbf{x}_0, \dot{\mathbf{x}}_d(t_0) = \mathbf{v}_0, \quad (1a)$$

$$\dot{r}_d(t) = \varphi(\mathbf{x}_d(t), \mathbf{v}_d(t), r_d(t), t), \quad r_d(t_0) = r_0, \quad (1b)$$

where the subscript d denotes variables that are sampled along the Lagrangian droplet trajectories, $\mathbf{f}(\mathbf{x}, \mathbf{v}, r, t)$ is the force per unit mass acting upon a droplet, and $\varphi(\mathbf{x}, \mathbf{v}, r, t)$ is the rate of radial change for a droplet with Eulerian position \mathbf{x} , velocity \mathbf{v} , and radius r at time t , and droplets have initial position \mathbf{x}_0 , velocity \mathbf{v}_0 , and radius r_0 at time t_0 . Without loss of generality, both \mathbf{f} and φ are assumed to be dependent upon all of \mathbf{x} , \mathbf{v} , r , and t , although φ may also be dependent upon other thermodynamic parameters which are related to the droplet evaporation.

The FLA is based on the assumption that the dispersed phase can be represented as a continuum,⁹ with trajectories characterised by their initial position \mathbf{x}_0 and time t . Recent work has generalised this concept by extending the phase space to include the dependence on initial droplet radius r_0 , referred to as the generalised fully Lagrangian approach (gFLA).¹⁴ Then the dispersed phase probability density $p(\mathbf{x}_d, r_d, t)$ can be interpreted as the number density in this wider phase space, and is calculated directly along the trajectories of individual droplets by considering the Lagrangian form of conservation of mass⁹

$$p(\mathbf{x}_d, r_d, t) = \frac{p(\mathbf{x}_0, r_0, t_0)}{|\det(\mathbf{J}(\mathbf{x}_0, r_0, t))|}, \quad (2)$$

where the Jacobian tensor $\mathbf{J}(\mathbf{x}_0, r_0, t)$ is specified in block matrix form as¹⁴

$$\mathbf{J}(\mathbf{x}_0, r_0, t) = \begin{bmatrix} \mathbf{J}^{xx} & \mathbf{J}^{xr} \\ \mathbf{J}^{rx} & J^{rr} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{x}_d}{\partial \mathbf{x}_0} & \frac{\partial \mathbf{x}_d}{\partial r_0} \\ \frac{\partial r_d}{\partial \mathbf{x}_0} & \frac{\partial r_d}{\partial r_0} \end{bmatrix}. \quad (3)$$

The procedure for deriving Equation (2) directly from the behaviour of individual droplets described by Equation (1) is outlined in Appendix A, along with the implications of the physical assumptions involved. The physical interpretation of \mathbf{J} is the local elemental deformation along a trajectory with respect to its initial state at (\mathbf{x}_0, r_0, t) . This is equivalent to the Eulerian-Lagrangian transformation that describes the influence of variables transported by droplets on the local Eulerian field, meaning that \mathbf{J} is able to provide information about the phase space deformation of the droplet field. The blocks of the Jacobian given by Equation (3) represent this deformation in the associated part of phase space; \mathbf{J}^{xx} is the

spatial deformation given by the standard FLA, J^{rr} is the deformation in radial space, and the remaining blocks provide the corresponding cross components. The evolution equations for each block can be obtained by taking partial derivatives of the governing Equation (1) with respect to (\mathbf{x}_0, r_0) , yielding the system

$$\dot{\mathbf{J}}^{xx} = \frac{\partial \mathbf{f}_d}{\partial \mathbf{x}} \cdot \mathbf{J}^{xx} + \frac{\partial \mathbf{f}_d}{\partial \mathbf{v}} \cdot \dot{\mathbf{J}}^{xx} + \frac{\partial \mathbf{f}_d}{\partial r} J^{rx}, \quad (4a)$$

$$\dot{\mathbf{J}}^{xr} = \frac{\partial \mathbf{f}_d}{\partial \mathbf{x}} \cdot \mathbf{J}^{xr} + \frac{\partial \mathbf{f}_d}{\partial \mathbf{v}} \cdot \dot{\mathbf{J}}^{xr} + \frac{\partial \mathbf{f}_d}{\partial r} J^{rr}, \quad (4b)$$

$$\dot{J}^{rx} = \frac{\partial \varphi_d}{\partial \mathbf{x}} \cdot \mathbf{J}^{xx} + \frac{\partial \varphi_d}{\partial \mathbf{v}} \cdot \dot{\mathbf{J}}^{xx} + \frac{\partial \varphi_d}{\partial r} J^{rx}, \quad (4c)$$

$$\dot{J}^{rr} = \frac{\partial \varphi_d}{\partial \mathbf{x}} \cdot \mathbf{J}^{xr} + \frac{\partial \varphi_d}{\partial \mathbf{v}} \cdot \dot{\mathbf{J}}^{xr} + \frac{\partial \varphi_d}{\partial r} J^{rr}. \quad (4d)$$

The general form of the initial conditions at time t_0 pertaining to Equations (4) are

$$\mathbf{J}^{xx}(\mathbf{x}_0, r_0, t_0) = \mathbf{I}, \quad \dot{\mathbf{J}}^{xx}(\mathbf{x}_0, r_0, t_0) = \partial \mathbf{v}_0 / \partial \mathbf{x}_0, \quad (5a)$$

$$\mathbf{J}^{xr}(\mathbf{x}_0, r_0, t_0) = \mathbf{0}, \quad \dot{\mathbf{J}}^{xr}(\mathbf{x}_0, r_0, t_0) = \mathbf{0}, \quad (5b)$$

$$J^{rx}(\mathbf{x}_0, r_0, t_0) = \mathbf{0}, \quad (5c)$$

$$J^{rr}(\mathbf{x}_0, r_0, t_0) = 1, \quad (5d)$$

where \mathbf{I} is the identity matrix. The system (4) then describes the evolution of the blocks of \mathbf{J} whilst accounting for the coupling that exists between the different blocks. Note that this coupling only occurs between blocks in the same column of the Jacobian in Equation (3), with the different columns being associated with independent systems of equations. This formulation can also be extended to provide a more complete description of the droplet thermophysical behaviour by including temperature and mass as phase space variables instead of only the droplet radius. The initial conditions (5) are found by applying the Jacobian as defined in Equation (3) to the initial conditions associated with the governing Equations (1). It is assumed here that the initial droplet position \mathbf{x}_0 and velocity \mathbf{v}_0 are independent of radius r_0 , and further that the initial droplet radius r_0 is also independent of position \mathbf{x}_0 . Following,⁵ however, it is noted that the initial droplet velocity \mathbf{v}_0 varies with position \mathbf{x}_0 depending upon both the nature of the carrier flow and how droplets are introduced into the flow, and identification of the correct initial condition for $\partial \mathbf{v}_0 / \partial \mathbf{x}_0$ is crucial to obtaining the correct behaviour for the evolution of \mathbf{J} .

For the standard FLA, in which the droplet radius r_d is not considered, the probability density $p(\mathbf{x}, r, t)$ reduces to the number density $n(\mathbf{x}, t)$, which is a function of only physical space \mathbf{x} . For the monodisperse cases in Section 4.1 the analysis is accordingly focused upon $n(\mathbf{x}, t)$, whilst the polydisperse flows in Section 4.2 consider the more general case of $p(\mathbf{x}, r, t)$.

2.2 | Formulation for simplified physical models

Whilst the gFLA formulation in Section 2.1 is applicable to general equations of motion and evaporation for spherical droplets modelled as point masses in a dilute suspension, by further assuming that the carrier flow density is much less than that of the dispersed phase, the applicability is then restricted to gas-droplet flows. Making the additional assumption of a low droplet Reynolds number enables the droplet momentum to be modelled using a linear drag law. For a simplified physical model of evaporation, all heat at the droplet surface is taken to be spent on evaporation, then the droplet evaporation rate $\varphi = \varphi(r, t)$ is only dependent on the radius r , and independent of position \mathbf{x} and velocity \mathbf{v} . Under these conditions, the associated nondimensional forms of \mathbf{f} and φ from Equations (1) are given by

$$\mathbf{f}(\mathbf{x}, \mathbf{v}, r, t) = \frac{1}{St_0^* r^2} (\mathbf{u}(\mathbf{x}, t) - \mathbf{v}), \quad (6a)$$

$$\varphi(r, t) = -\frac{\delta}{2r}, \quad (6b)$$

where S_0^{t*} is a reference Stokes number corresponding to droplets of characteristic radius r_{d0}^* , $\mathbf{u}(\mathbf{x}, t)$ is the carrier flow velocity and δ is the rate of change of the droplet surface area. In this case the equations of evolution for system (4) simplify to

$$\dot{\mathbf{j}}^{xx} = \frac{1}{S_0^{t*} r_d^2} \frac{\partial \mathbf{u}_d}{\partial \mathbf{x}} \cdot \mathbf{j}^{xx} - \frac{1}{S_0^{t*} r_d^2} \mathbf{j}^{xx} - \frac{2}{S_0^{t*} r_d^3} (\mathbf{u}_d - \dot{\mathbf{x}}_d) \cdot \mathbf{j}^{xx}, \quad (7a)$$

$$\dot{\mathbf{j}}^{xr} = \frac{1}{S_0^{t*} r_d^2} \frac{\partial \mathbf{u}_d}{\partial \mathbf{x}} \cdot \mathbf{j}^{xr} - \frac{1}{S_0^{t*} r_d^2} \mathbf{j}^{xr} - \frac{2}{S_0^{t*} r_d^3} (\mathbf{u}_d - \dot{\mathbf{x}}_d) \cdot \mathbf{j}^{xr}, \quad (7b)$$

$$\dot{j}^{xx} = \frac{\delta}{2r_d^2} j^{xx}, \quad (7c)$$

$$\dot{j}^{rr} = \frac{\delta}{2r_d^2} j^{rr}, \quad (7d)$$

where $\mathbf{u}_d(t) = \mathbf{u}(\mathbf{x}_d(t), t)$ and $\partial \mathbf{u}_d(t)/\partial \mathbf{x} = \partial \mathbf{u}(\mathbf{x}_d(t), t)/\partial \mathbf{x}$ respectively denote that the fluid velocity and fluid velocity gradient are evaluated along the droplet trajectories. The initial conditions for the system (7) remain the same as those given by Equations (5).

2.3 | Computational advantage

The most salient advantage of using the FLA framework to compute the number density field is the increase in computational efficiency over conventional box-counting methods which can be realised, principally due to effective use of the information provided by the Eulerian-Lagrangian transformation. This aspect of the method has been focused upon in several works, and stems from the observation that an accuracy of $<0.1\%$ in calculation of the Eulerian number density field is only achieved using box-counting methods when $\mathcal{O}(10^4)$ droplets are present in each Eulerian cell within a simulation.^{5,12} In contrast, due to the ability of the FLA to provide the number density along trajectories, and the facet of the Eulerian-Lagrangian transformation that trajectory data can be meaningfully extrapolated onto the local spatial region, the FLA only requires one trajectory per Eulerian cell to achieve the same level of accuracy.¹² This advantage means that despite the additional computation of the Jacobian as it evolves along each trajectory according to Equations (4), the FLA is still approximately 20 times more efficient than box-counting methods for two-dimensional simulations.^{5,11} Another analysis found that a box-counting approach for number density requires $\sim 10^6$ trajectories to be computed, whereas the FLA only needs $\sim 10^3$ trajectories to obtain the same level of detail, however this is offset by the additional expense of calculating the Jacobian along trajectories which entails a fortyfold increase in computational cost for three-dimensional simulations, and also by linear interpolation of the trajectory number density data onto the Eulerian grid which takes almost as long as the numerical integration along trajectories.¹⁵ Therefore, despite providing a 10^3 saving on the number of trajectories computed when compared against the box-counting approach, the overall efficiency improvement of the FLA for calculation of the Eulerian number density field is reduced to around tenfold. It is also noted that in implementations of the FLA, neglecting droplet collisions in the dilute phase limit results in the independence of trajectories, which further enables effective parallelisation and computational efficiency.¹⁶

2.4 | Numerical considerations

The gFLA formulation in Section 2.1 is valid for the regime of dilute droplet flows, and makes the important assumption that the droplet field is continuous between trajectories and exhibits smooth spatial variation. Such an assumption is, however, at odds with the multi-valued nature that characterises the trajectory crossing experienced by inertial droplets. The strength of the standard FLA formulation is that it is able to retain this detail that is inherent to individual trajectories within its description of the spatial droplet field, however this is manifested in the elemental volume of the Eulerian-Lagrangian transformation, given by $\det(\mathbf{J}^{xx}(\mathbf{x}_0, t))$, becoming zero at the point of trajectory crossings.⁵ By virtue of Equation (2), the droplet number density is then singular at these points. This phenomenon is in fact necessary to maintain consistency with the assumption of continuity for the dispersed phase, and is simply how the FLA framework incorporates the information of crossing trajectories into the concept of a smoothly defined field. It has further been

demonstrated in previous work that such singularities in number density are integrable,¹⁷ meaning that the local Eulerian number density field within a defined region remains finite.¹²

Notwithstanding the fact that the FLA presents a theoretically sound approach for number density calculation, in practice the issue lies in how to accumulate the multi-valued number density contributions along trajectories into an Eulerian field representation in a physically correct manner. An interpolation procedure is only valid within a region of the droplet field which is single-valued, meaning that careful treatment of the multi-valued regions is paramount to ensuring that the calculation of an Eulerian number density field is meaningful. Previous works have addressed this by using the fact that the sign of the Jacobian determinant J changes every time a droplet crosses the trajectory of another droplet, effectively providing a means of keeping track of which layer of the droplet field a given trajectory is in. Since each layer of the droplet field is single-valued with non-intersecting trajectories, interpolation can be used to calculate the Eulerian number density within each individual layer, and then due to the number density field being additive, the contributions from each layer at a given point are summed together to obtain the total number density.^{12,15,18} This requires that droplets are indexed in such a way that a distinction is made every time a fold is crossed, with the most straightforward means of doing this being to keep count of the number of times this occurs for each droplet, then all droplets with the same count index form one layer of the droplet field.¹⁸ Such a procedure enables the reproduction of a number density field which although single-valued, is able to contain the detail of different layers within the droplet field; a feature which is difficult to replicate in Eulerian-based simulations.⁴ The drawback of indexing droplets in this manner is that sufficiently many need to be present within each layer of the droplet field for the interpolation methods to be well defined and able to work correctly. In practice this is only an issue when trajectories initially cross and there are few droplets within the newly created layer of the droplet field, however this still presents a situation which requires appropriate treatment to ensure that the contribution of these droplets is accounted for in the number density calculation. Notwithstanding this, the physical accuracy which can be achieved through the use of a Lagrangian approach highlights the potential ability of the FLA to accurately simulate industrially relevant systems.

3 | INTERPOLATION OF THE NUMBER DENSITY FIELD

The primary point of contention in developing useful computational tools which make use of the FLA is that the number density provided by the method is along trajectories, whereas in many numerical methods and applied contexts it is most often the Eulerian field information which is required. In this sense, the scattered number density data which the FLA produces requires the use of an appropriate interpolation procedure for accumulation onto an Eulerian grid. It has been noted that simple linear interpolation is sufficient for a steady-state flow configuration, however for transient flows, that contain vortices, a more comprehensive accumulation algorithm is required.¹⁶ This is a subject which has received surprisingly little attention to date, yet has the potential to open up the FLA methodology to a wider and more general class of flows.

3.1 | Existing procedures for number density calculation

The classical box-counting approach to calculating the number density field in dispersed multiphase flows simply involves the accumulation of mass contributions from all the droplets within an Eulerian cell, then weighting them by the cell volume. This is also referred to as the nearest neighbour method,¹⁹ and can be interpreted as reassigning the contribution from a droplet to act at the centre of the cell in which it is located. Such an approach is associated with a discontinuous droplet weighting function, since it takes no account of how far a droplet is located from the cell centre, meaning that this information is lost in the accumulation process, and resulting in the generation of artificial noise. It is this noise which can cause a high variation in the computed number density field, and necessitates the requirement for such a high number of droplets using the box-counting approach in order to produce a stable value for the number density.

An improvement to the box-counting approach is achieved by considering a droplet weighting function with C^0 continuity which is able to account for the location of a droplet relative to the Eulerian grid on which the accumulation of contributions is made. One procedure that utilises such a methodology is the cloud-in-cell (CIC) approach,²⁰ in which the weighting function takes the form of a triangular-shaped kernel. This is equivalent to assigning contributions from a droplet to the corner nodes of the Eulerian cell in which it resides by using linear interpolation, which importantly ensures that the constraints of continuity and momentum conservation are respected by the accumulation procedure,

whilst also maintaining numerical stability. The CIC approach is a widely adopted method of accumulating contributions in dispersed multiphase flow simulations due to its simplicity of implementation, and, owing to the fact that the weighting function retains C^0 continuity of the droplet field, requires fewer droplets than the nearest neighbour method to achieve a stable result for the number density field on an identical Eulerian grid. The domain of influence for the droplet weighting function for a Cloud-In-Cell approach is, however, naturally limited to the Eulerian cell in which the droplet is located. This means that many droplets per cell are still required to achieve statistical convergence of the number density field, and the computational expense associated with this can be prohibitive within simulations of industrial scale processes.

Within the broad class of meshfree methods which exist for interpolation of scattered data, kernel-based methods have seen a variety of applications, in particular for fluid dynamics in the form of smoothed particle hydrodynamics (SPH).²¹ Whilst the rationale for SPH is well established, the normalisation condition required of the weighting kernel is such that a sufficient number of particles must be inside the kernel for the accumulation procedure to remain valid, which has implications for the simulation of compressible flows in which the spatial distribution of particles is non-uniform. Although the SPH framework has been extended to handle weakly-compressible multiphase flows through the use of adaptive spatial resolution, this has only been shown to be accurate in the case where the density ratio between the two phases is less than 10,²² whilst typical gas-droplet flows have a much higher density ratio. Consequently the ability to effectively treat the clusters and voids that are characteristic of dispersed multiphase flow using the SPH framework remains limited.

Previous work has highlighted the ability of Voronoï tessellations to calculate the droplet number density field from droplet location data, by recognising that the area of the Voronoï cells is the inverse of the local droplet number density.²³ Notably, it was demonstrated that it is possible to identify regions of clusters and voids within the droplet field through categorisation of the area of Voronoï cells. This methodology enables information about the Lagrangian dynamics of individual droplets to be included within the representation of the Eulerian number density field, meaning that a certain level of detail about the structures in the droplet field can be retained. However, since the influence of a given droplet is taken to be uniform over its associated Voronoï cell, the resulting Eulerian number density field is discontinuous along the Voronoï cell boundaries. In this sense, the construction of the number density field using Voronoï tessellations requires a sufficiently large number of droplets to obtain a smooth result, an aspect shared with the CIC approach. Despite this, Voronoï tessellations have been incorporated into a solver as a full hydrodynamics scheme,²⁴ and subsequently developed into Voronoï particle hydrodynamics (VPH).²⁵ The advantages offered by such a solver are however offset by the additional complexity of the formulation, and need to generate the tessellation for the entire droplet field at every point in time.

An option which mirrors the thinking behind the use of Voronoï tessellations is that of the Delaunay triangulation, with these two procedures being closely related as the dual graphs of each other. This is motivated by the use of Delaunay triangulations in scattered data interpolation, with the triangulation effectively providing a mesh over which commonly employed interpolation procedures can be used between the known values at data points. The desirable property unique to the Delaunay triangulation is that it is constructed to satisfy the empty circumcircle criterion, which ensures maximization of the internal angles of triangles, with the reasoning that interpolation routines will exhibit greater accuracy and stability over more uniformly shaped triangles. These aspects have seen the Delaunay triangulation developed as a means of constructing density fields from scattered data points,²⁶ and extended to a hydrodynamical method with the Delaunay tessellation field estimator (DTFE) technique.²⁷ However, the shortcoming of such an approach is that the Delaunay triangulation necessarily changes discontinuously at certain instants due to the empty circumcircle criterion,²⁵ meaning that the DTFE is not well suited as the basis of numerical solver.

3.2 | Requirements of a suitable interpolation procedure

The immediate advantage offered by the FLA methodology is that the exact value of the droplet number density field is already known along trajectories before the use of any interpolation procedure, thereby eliminating the noise that is inherent in procedures such as nearest neighbour or CIC interpolation. The challenge then remains to find an appropriate means of interpolating these values that will retain sufficient accuracy in describing the structures of the droplet field. Existing work has demonstrated that linear interpolation between neighbouring trajectories is able to capture the smooth, and often monotonic, variation of the number density field in steady-state flows well, however the considerably more complex structure of the droplet field in transient flows, for which folds occur and multiple layers are inherent, means that such an approach is not applicable.¹⁴ To effectively extend the concept of linear interpolation to transient flows requires the establishment of a more general framework.

Whilst the ability to extend the Delaunay triangulation to points in n -dimensional space does makes it a possible method for interpolating number densities obtained using the gFLA onto a mesh that spans both physical and radial space, using a triangulation-based approach to interpolate FLA data has a number of distinct disadvantages. Principally, the computational expense of having to construct a triangulation over all droplets within each layer of the droplet field, when coupled with the subsequent interpolation procedure of the number density data, would likely negate most of the saving offered by computing the number density along trajectories for the fewer droplets required using the FLA. To compound this, triangulation of droplets would not be able to distinguish and properly resolve voids in the droplet field when using linear interpolation within triangles, which would necessitate the use of a higher order interpolation method that is able to capture the spatial gradient of the number density field, and thereby further increase the computational expense. Finally, whilst knowledge of the Jacobian tensor provides information about the Eulerian-Lagrangian transformation of trajectory variables onto the local Eulerian field for each droplet, and thereby detail of the structures within the droplet field, a triangulation-based procedure is unable to utilise this information in its reconstruction of the number density field.

The potential drawbacks of triangulation-based methods for accumulating the FLA number density data help to elucidate the desirable properties that a suitable interpolation routine should possess. To maintain consistency with the Lagrangian nature of the FLA methodology, a meshfree method which accumulates contributions without prior knowledge of the spatial relationship between droplets would be most appropriate in terms of numerical efficiency, and realising the optimal computational saving which can be achieved using the FLA. In this sense, the contribution of a given droplet should not be dependent on that of other droplets, and the procedure for the accumulation of droplet contributions should also be valid regardless of the number of droplets that contribute to the Eulerian field at a given point. Such a requirement is further justified in the context of the FLA by considering that since droplet contributions are first accumulated within layers of the droplet field, then interpolating between droplets becomes an issue when a layer contains sufficiently low numbers of droplets, and is not possible in the case when the number of droplets is less than or equal to the dimensionality of the simulation. In particular, it is possible that a certain layer may contain only a single droplet, however it is still desirable that the contribution from this droplet is included in the reconstruction of the Eulerian number density field.

Additionally, an interpolation scheme should be able to account for the fact that the number density field for inertial droplets within transient flows has defined clusters and voids, and in particular that droplets in voids can be sparse. It is therefore desirable that the influence of those droplets that are present in voids is able to extend over the surrounding domain to a greater degree, in order to reflect that there are fewer other droplets in the immediate vicinity from which contributions to the Eulerian droplet field can be accumulated. The nature of this demand is most appropriately met through the use of a meshfree method in which a degree of control over the domain of influence that individual droplet contributions have can be exercised.

3.3 | Kernel regression

Within the context of accumulating the number density data obtained from the FLA, the problem of capturing clusters and voids in the droplet field at a competitive computational cost is addressed here through the means of kernel regression. The motivation behind this choice is that, as a form of statistical learning, the averaged nature of such an approach enables it to effectively handle spatial locations at which only a single data point makes a contribution. This is seen by considering the form of the Nadaraya-Watson estimator¹⁹

$$n(\mathbf{x}, t) = \frac{\sum_{i=1}^N K_H(\mathbf{x}, \mathbf{x}_d^i) n(\mathbf{x}_d^i, t)}{\sum_{j=1}^N K_H(\mathbf{x}, \mathbf{x}_d^j)}, \quad (8)$$

in which $n(\mathbf{x}_d^i, t)$ is the instantaneous number density along the trajectory \mathbf{x}_d associated with droplet i at time t as obtained from the FLA, $n(\mathbf{x}, t)$ is the Eulerian number density field, $K_H(\mathbf{x}, \mathbf{x}_d^i)$ is a kernel which provides a weighting for the contribution of the droplet \mathbf{x}_d^i to the accumulation point \mathbf{x} , and N is the number of droplets for which the range of support of the associated kernel includes the point \mathbf{x} . In Equation (8), the numerator provides the weighted sum of the Lagrangian number density contributions $n(\mathbf{x}_d^i, t)$ at \mathbf{x} , whilst the denominator is equal to the sum of the weighting factors given by the kernels that contribute at \mathbf{x} . Consequently, in the accumulation procedure for $n(\mathbf{x}, t)$ it is seen that the contributions from $n(\mathbf{x}_d^i, t)$ are effectively weighted by weights w^i defined by

$$w^i = \frac{K_H(\mathbf{x}, \mathbf{x}_d^i)}{\sum_{j=1}^N K_H(\mathbf{x}, \mathbf{x}_d^j)}. \quad (9)$$

In this sense, the Nadaraya-Watson estimator (8) can be considered as an expansion in renormalised radial basis functions w^i ,¹⁹ and it is the appearance of the denominator in Equation (9) which enables a representative value for $n(\mathbf{x}, t)$ to be obtained even in the case when $N = 1$. This circumvents the requirement for a specific number of droplets to be within the kernel in order to satisfy its normalisation condition, albeit at the introduction of a statistical error arising due to the low sample size. Nonetheless, for areas of extreme sparsity in inertial droplet fields, this provides an appropriate means by which to infer the behaviour of $n(\mathbf{x}, t)$ without having to resort to using an exhaustive number of droplets in simulations.

For specification of the kernel, a range of functions including Gaussian, Epanechnikov, and tricube kernels are commonly used. In the present work, a Gaussian form is chosen due to its possession of C^∞ continuity and convenient geometrical interpretation. In the one-dimensional case, this is specified by

$$K_h(\mathbf{x}, \mathbf{x}_d^i) = \exp \left[-\frac{1}{2} \frac{\|\mathbf{x} - \mathbf{x}_d^i\|^2}{h^2} \right], \quad (10)$$

where h is the smoothing length of the kernel. In this form only the Euclidean distance between the droplet position \mathbf{x}_d^i and accumulation point \mathbf{x} is used to determine the relative contribution of droplets, resulting in a spherical kernel. A more generalised approach can be taken by using the concept of structured kernels, in which distinct smoothing lengths along arbitrary axes of alignment can be defined.¹⁹ For the specific case of a multivariate Gaussian form, such a kernel is given by

$$K_{\mathbf{H}}(\mathbf{x}, \mathbf{x}_d^i) = \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}_d^i)^\top \cdot \mathbf{H}^{-1} \cdot (\mathbf{x} - \mathbf{x}_d^i) \right], \quad (11)$$

where \mathbf{H} is the bandwidth matrix that contains information about the smoothing lengths in different directions. The particular choice of $\mathbf{H} = h^2 \mathbf{I}$ corresponds to a spherical kernel, in which case Equation (11) accordingly reduces to Equation (10).

Control over the kernel in simulations is specified solely by the choice of the smoothing length for spherical kernels, or the bandwidth matrix for structured kernels, and this is dependent on the specific context in which the method is being applied. In general, however, it is important that the smoothing length(s) are large enough to create a field representation which varies smoothly in space, yet not so large that excessive smoothing is applied, in which case detail from the original data will be lost. Therefore, a crucial aspect in the application of kernel regression is to ensure that care is taken in the selection of an appropriate smoothing length for a given situation.

3.4 | Utilisation of the Eulerian-Lagrangian transformation tensor

The main strength of using kernel regression for accumulation of the FLA number density data lies within the ability to use information from the FLA to specify the smoothing length h or bandwidth matrix \mathbf{H} of the kernels, in order to gain a finer degree of control over the reconstruction of the number density field than is offered by a constant-sized kernel. The rationale for this procedure is in alignment with the more general framework of metric learning methods for kernel regression, in which the kernel is adaptively defined according to the local behaviour of the interpolation data.^{28,29} In the case of the standard FLA formulation this can be accomplished by using the Eulerian-Lagrangian transformation tensor $\mathbf{J}^{\mathbf{x}}$ to specify the kernel size and shape associated with each droplet, which enables the highly varying sparsity that is observed in the droplet number density field of unsteady flows to be accounted for. Such a procedure is consistent with the meshfree nature of kernel methods, since it respects the individuality of Lagrangian trajectories by having a distinct bandwidth matrix for each droplet \mathbf{x}_d^i rather than each Eulerian position \mathbf{x} . Furthermore, using the Eulerian-Lagrangian transformation to determine the bandwidth matrix is also theoretically sound, as it makes use of the physical interpretation of $\mathbf{J}^{\mathbf{x}}$ in describing how the influence of a given droplet is exerted locally upon the surrounding domain, and accordingly specifies the range and shape of support for the kernel that is used to accumulate the number density contribution from that droplet to the associated Eulerian field representation. The physical correctness of this can be inferred directly from the FLA using Equation (2), by considering that if the number density associated with an individual droplet is high, the domain over which it acts must be reduced to conserve the mass contribution of that droplet along its trajectory. The necessary information about the domain size is seen to be provided by the Jacobian determinant $|\det(\mathbf{J}^{\mathbf{x}})|$, and directly associating this with the domain of support of a kernel-based method then naturally incorporates this physical

information within the numerical accumulation procedure for constructing the Eulerian number density field. Additionally, specifying the kernel using $\mathbf{J}^{\mathbf{x}}$ is equivalent to interpreting the smoothing length(s) as a representative lengthscale(s) for a given kernel. This is consistent with the introduction of a lengthscale within the definition of number density, a concept which has been emphasised in previous work as being necessary to provide a meaningful representation of the number density field.^{30,31}

Using the Eulerian-Lagrangian transformation tensor $\mathbf{J}^{\mathbf{x}}$ to determine the range and shape of support of the kernel for a droplet also provides a way of mitigating against the occurrence of singularities in the Lagrangian number density data that the FLA provides. This is because whenever a singularity in number density occurs, the corresponding value of $\det(\mathbf{J}^{\mathbf{x}})$ passes smoothly through zero, and this is associated with the domain over which the local Eulerian-Lagrangian transformation applies also reducing to zero. Therefore using $\mathbf{J}^{\mathbf{x}}$ to specify the kernel at this point will result in a kernel with a zero range of support, that is, mathematically equivalent to the Dirac delta function. Since the Eulerian number density field is constructed over a discrete grid with finite grid spacing, once the range of support of a kernel becomes small enough in the case when $|\det(\mathbf{J}^{\mathbf{x}})|$ is sufficiently small, the contribution from the droplet at that point will fall below the size of an Eulerian cell and no longer extend to cover any grid points, and will therefore not be included in the Eulerian field representation. Thus specification of the kernel in accordance with $\mathbf{J}^{\mathbf{x}}$ offers a means of effectively filtering out or limiting the propagation of the non-physical contributions associated with the high number density values that are inherent to the FLA, and thereby maintaining a more realistic description of the number density field in regions of the droplet field that contain crossing trajectories.

In practice, such a procedure can be applied in the cases of both spherical and structured kernels. For a spherical kernel, the smoothing length h is determined by setting the size of the spherical domain of the kernel to be equal to the elemental volume deformation $|\det(\mathbf{J}^{\mathbf{x}})|$ associated with a droplet along its trajectory. For a flow configuration with a spatial domain of \mathbb{R}^d , this determines the smoothing length for the kernel associated with a droplet at a given time t as

$$h(\mathbf{x}_0, t) = h_0 |\det(\mathbf{J}^{\mathbf{x}}(\mathbf{x}_0, t))|^{1/d}, \quad (12)$$

where h_0 is the initial smoothing length associated with the droplet at time t_0 and position \mathbf{x}_0 , and d is the number of spatial dimensions in the system under consideration. Using Equation (12), the smoothing length $h(\mathbf{x}_0, t)$ is automatically updated at each timestep, thereby adjusting the size of the kernel directly in accordance with the value of $|\det(\mathbf{J}^{\mathbf{x}})|$ at that time. This methodology is consistent with the concept of a variable smoothing length in SPH, where the simplest approach to updating h utilises the number density $n(\mathbf{x}, t)$ such that³²

$$h(\mathbf{x}, t) = h_0 \left(\frac{n(\mathbf{x}_0, t_0)}{n(\mathbf{x}, t)} \right)^{1/d}. \quad (13)$$

For monodisperse droplets, it is seen that substitution of the Lagrangian continuity Equations (2) into (13) demonstrates the equivalence between Equations (12) and (13) as procedures for adaptively scaling $h(\mathbf{x}_0, t)$, which highlights the rational physical basis for Equation (12).

In the case of a structured kernel, a positive semi-definite bandwidth matrix \mathbf{H} is constructed by taking

$$\mathbf{H}(\mathbf{x}_0, t) = h_0^2 \mathbf{J}^{\mathbf{x}}(\mathbf{x}_0, t) \cdot \mathbf{J}^{\mathbf{x}\top}(\mathbf{x}_0, t). \quad (14)$$

It is appropriate to use the initial one-dimensional smoothing length h_0 within the definition of this structured kernel due to the initial condition on the Eulerian-Lagrangian transformation tensor of $\mathbf{J}^{\mathbf{x}}(\mathbf{x}_0, t_0) = \mathbf{I}$ in Equations (5), from which it follows by Equation (14) that the initial bandwidth matrix $\mathbf{H}_0 = \mathbf{H}(\mathbf{x}_0, t_0) = h_0^2 \mathbf{I}$, and is therefore a spherical kernel. The bandwidth matrix $\mathbf{H}(\mathbf{x}_0, t)$ can then subsequently be updated at each timestep using Equation (14), adjusting the size, shape, and orientation of the kernel directly in accordance with the Eulerian-Lagrangian transformation tensor $\mathbf{J}^{\mathbf{x}}$ along a trajectory at that time. Specification by means of Equation (14) results in an ellipsoidal kernel which represents the deformation of the initial spherical kernel at that point along a trajectory, with the lengths of the semi-axes and angles of rotation from the Cartesian axes determined by the components of $\mathbf{J}^{\mathbf{x}}$ at that point. Importantly, this procedure can be physically interpreted as specification of the shape and size of the range of support for the kernel associated with an individual droplet in accordance with the spatial structures of the Eulerian number density field. For instance, in regions of high droplet number density where droplets cluster in elongated structures, the ellipsoids associated with the kernels for these droplets will be aligned along these structures, and this ensures that the number density values along

trajectories are projected only onto the areas of the Eulerian number density field which have similar physical behaviour in terms of droplet clustering. Therefore using the Eulerian-Lagrangian transformation tensor to specify the kernel provides a data-driven approach to the accumulation of droplet contributions in a manner which is physically consistent with the droplet number density field, and requires no further assumptions to be made beyond specification of the initial smoothing length h_0 . A demonstration that specification of \mathbf{H} by Equation (14) defines a kernel for which the domain size varies proportionately to $|\det(\mathbf{J}^{\mathbf{x}})|$ is provided in Appendix B.

3.5 | Extension of the phase space to include droplet radius

The kernel regression procedure can straightforwardly be extended to include the droplet radius r_d within the phase space coordinate vector, enabling the droplet size distribution to be determined for polydisperse flows using the gFLA by means of Equation (8), where now instead of $n(\mathbf{x}, t)$ it is $p(\mathbf{x}, r, t)$ that is being reconstructed. In this case, it is inappropriate to use a spherical kernel, since the scales of the domains for which physical space \mathbf{x} and radial space r are relevant will be different. For the same reason, and also due to the higher dimensionality meaning that the same number of droplets will be more sparsely distributed in (\mathbf{x}, r) space, consideration of the structures in the droplet probability density field $p(\mathbf{x}, r, t)$ is no longer as important, and it is sufficient to vary the kernel size based on just the magnitude of the phase space Jacobian determinant $\det(\mathbf{J}(\mathbf{x}_0, r_0, t))$. In this case a structured kernel of the same form as in Equation (11) is used, however it is now written in terms of the phase space coordinate $\xi = (\mathbf{x}, r)$ and phase space trajectories $\mathbf{z}_d = (\mathbf{x}_d, r_d)$ as

$$K_H(\xi, \mathbf{z}_d^i) = \frac{1}{\sqrt{\det(\mathbf{H})}} \exp \left[-\frac{1}{2} (\xi - \mathbf{z}_d^i)^\top \cdot \mathbf{H}^{-1} \cdot (\xi - \mathbf{z}_d^i) \right], \quad (15)$$

with the bandwidth matrix \mathbf{H} now being defined by

$$\mathbf{H}(\mathbf{x}_0, r_0, t) = \begin{bmatrix} h_{x0}^2 |\det(\mathbf{J}^{\mathbf{x}\mathbf{x}}(\mathbf{x}_0, t))|^{2/d} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & h_{r0}^2 |J^r(r_0, t)|^2 \end{bmatrix}, \quad (16)$$

where h_{x0} and h_{r0} are the initial smoothing lengths in physical and radial space respectively. Whilst the resulting kernel is still ellipsoidal, it remains aligned with the Cartesian axes rather than rotating to align with any structures in the droplet number density field. Furthermore, the kernel remains spherical across all spatial dimensions \mathbf{x} , with separate control of h_{x0} and h_{r0} then allowing for the range of support for the kernel in physical and radial space to be determined independently. In this manner the contributions from droplets can be appropriately accumulated to produce the Eulerian probability density field in (\mathbf{x}, r) space, and thereby a description of the droplet size distribution at each point \mathbf{x} in space.

3.6 | Numerical implementation

In practice, a Gaussian kernel has an infinite range of support with non-zero contributions at every Eulerian grid point in the domain, although these will quickly become negligible at a certain distance away from a droplet. Since kernel regression normalises the resultant Eulerian number density field using only those droplets which contribute at a given point, it is computationally favourable to limit the extent of the range of support for droplets by imposing an artificial range of compact support. This can be realised in the general case of the structured multivariate Gaussian kernel (15) by using the Mahalanobis distance associated with a given kernel to determine which Eulerian grid points will receive non-negligible contributions from that kernel. Specifically, employing the often used 3-sigma rule of the Gaussian distribution, which states that 99.8% of the contributions from a given droplet lie within three smoothing lengths of its position, the condition for Eulerian grid points to receive a contribution from that droplet is

$$\sqrt{(\mathbf{x} - \mathbf{x}_d^i)^\top \cdot \mathbf{H}^{-1} \cdot (\mathbf{x} - \mathbf{x}_d^i)} \leq 3. \quad (17)$$

This procedure then automatically accounts for the shape and orientation of the kernel, and cuts off contributions beyond the isocontour of the kernel at a distance of three smoothing lengths from the droplet position. The number of

gridpoints that a given kernel contributes to is then drastically reduced from the full domain, significantly enhancing the computational economy of the kernel regression process. The same principle can also be applied for the gFLA by replacing \mathbf{x} and \mathbf{x}_d in Equation (17) with ξ and \mathbf{z}_d respectively.

Having a compact support for a kernel is advantageous when it comes to accumulating contributions from different droplets onto the Eulerian field. Since kernel regression is a meshfree method, and the kernel is specified uniquely for each individual droplet, it is convenient to consider the extent of all the contributions made by a single droplet onto the Eulerian gridpoints. As the kernel possesses a compact support given by Equation (17), the subset of gridpoints which receive a non-zero contribution from a given droplet can then be determined. In the case of a uniform Cartesian grid, it is judicious to use the minimum bounding box which encloses the region defined by the Mahalanobis distance using Equation (17). This can be realised for the multivariate Gaussian kernel (15), since the maximum extent of the range of compact support in the k^{th} coordinate direction is given by

$$s_k^{\text{max}} = 3\sqrt{H_{kk}}, \quad (18)$$

in which summation over k is not implied, and the coefficient of three arises due to invoking the three-sigma rule in defining the compact support. In this manner, the subset of the domain to which a single droplet makes contributions at a given point in time can be easily determined from the uniquely defined kernel associated with that particular droplet, enabling efficient accumulation of these values onto the Eulerian field. This is particularly simple as the use of a droplet search algorithm is not required, and takes advantage of the Lagrangian nature of kernel regression along with the fact that the Eulerian gridpoints are uniformly spaced. For more general or dynamically generated grids a further procedure is needed to determine the gridpoints which lie within the Mahalanobis distance of a droplet, adding to the computational expense of the interpolation procedure.

Kernel regression is able to construct a representative Eulerian number density field that respects conservation of droplet phase properties due to the manner in which the accumulation from individual droplets is performed. Specifically, the weights w^j for the Nadaraya-Watson estimator defined by Equation (9) normalise the droplet contributions at a given gridpoint regardless of the number of droplets which contain that gridpoint within their compact support, meaning that the influence of a given droplet on the accumulated number density field varies. This is exemplified for droplets with a sufficiently high number density, in which case the domain of the associated kernel will become correspondingly small and not extend to any of the neighbouring gridpoints, thus effectively filtering the contribution of that droplet from the Eulerian field at that instant in time. Despite this, kernel regression is still able to produce a physically consistent Eulerian field, and in particular is well suited to pairing with FLA number density data as it provides an estimator for the Eulerian number density field that is independent of the number of sampled trajectories. This is in keeping with the ethos of the FLA, in that by representing the dispersed phase as a continuum, only a subset of the trajectories required by a conventional accumulation procedure are needed to build an accurate description of the number density field.

For the case when the full shape and size of the kernel are specified using $\mathcal{J}^{\mathbf{x}\mathbf{x}}$ by means of Equation (14), an additional consideration must be made when droplets are in the vicinity of the envelope of folds in the droplet phase and the number density is almost singular. Due to $|\det(\mathcal{J}^{\mathbf{x}\mathbf{x}})|$ becoming zero when a trajectory crosses the envelope of folds, the shape of the local deformation associated with $\mathcal{J}^{\mathbf{x}\mathbf{x}}$ becomes elongated as the droplet approaches the envelope, and eventually collapses in the dimension perpendicular to the envelope so that it is fully aligned at the point where the droplet crosses the envelope. The consequence of defining the kernel using Equation (14) is then that as the droplet approaches and moves away from the fold envelope, the kernel becomes excessively elongated far beyond the extent of local structures in the droplet phase, and makes non-physical contributions to the regression procedure at distant gridpoints. It is therefore necessary to impose a restriction on the maximum extent to which a kernel can elongate in order to ensure that the accumulation procedure maintains stability and produces a representative reconstruction of the number density field. This is performed in the present work by simply restricting the kernel shape such that each of the semi-axes of the ellipsoidal domain upon which the kernel is defined cannot be greater than three times the value of the smoothing length h associated with the equivalent spherical kernel at that point, as determined by Equation (12). In practice, this limit on the elongation factor has been found to represent an acceptable trade off between capturing the anisotropic structures that are present in the droplet number density field, whilst preserving an accurate physical description that does not contain spurious numerical effects. In the event that it is necessary to limit the length of one of these semi-axes, the others are accordingly rescaled so that the value of $|\det(\mathcal{J}^{\mathbf{x}\mathbf{x}})|$ at that point on the given trajectory remains unchanged, thereby enabling droplets in the vicinity of the fold envelope to contribute to the kernel regression procedure in a manner which remains representative of the local structures in the droplet phase.

The important issue of selecting a suitable value for h_0 must be made at the beginning of a simulation, and in general is best specified depending upon the initial spatial distribution of droplets. For a seeding of droplets across an interval over which the Eulerian number density field can be considered constant, it is convenient to take h_0 as the same for all droplets, and the simplest way of setting this value is then to define it as the average inter-droplet spacing across the interval, denoted Δx_{d0} . Since h_0 is equivalent to the standard deviation in the one-dimensional Gaussian kernel (10), this then ensures that the compact support defined using the three-sigma rule provides sufficient coverage of the computational domain for the given droplet seeding. It should be noted that the specification of $h_0 = \Delta x_{d0}$ is only a guideline however, and whilst this is used across most of the cases in Section 4, some variation may be required in order to obtain either higher fidelity or smoothness of the Eulerian number density field as desired. In practice the range of permissible values for the initial smoothing length of a Gaussian kernel with respect to the average inter-droplet spacing is $\Delta x_{d0}/3 \ll h_0 \ll 2\Delta x_{d0}$, where the lower bound is the minimum size at which kernels with compact support are able to effectively provide coverage of the computational domain, and the upper bound represents the highest degree of smoothing which might reasonably be needed. Furthermore, in the case of an expanding spray in which droplets are initially injected through a narrow inlet but then spread to span a far wider region, the initial spatial distribution of droplets may no longer be an adequate metric for basing h_0 upon, and a larger value will be necessary in order to achieve sufficient coverage of the computational domain with the given droplet seeding.

In terms of the gFLA, h_{x0} is defined in the same way as h_0 , whilst the radial space smoothing length h_{r0} is specified in a similar manner in terms of the initial inter-droplet size spacing, denoted Δr_{d0} . The range of validity for the radial space smoothing length differs compared to h_{x0} in practice due to the less pronounced occurrence of coherent structures in the Eulerian droplet probability density field p , and an appropriate set of values is given by $0.5 \Delta r_{d0} \ll h_{r0} \ll \Delta r_{d0}$ for a Gaussian kernel with compact support. Analogously to h_{x0} , the lower bound provides greater resolution of the droplet size distribution in the Eulerian probability density field, whilst the upper bound offers a higher degree of smoothing across the droplet sizes.

4 | NUMERICAL SIMULATIONS

4.1 | Monodisperse droplets

It is instructive to first focus upon application of the kernel regression procedure outlined in Section 3 to monodisperse droplets for which evaporation effects are not considered, and the extension of the droplet distribution into radial space to account for the droplet size therefore does not have to be included. In the following, a range of flow configurations of varying dimensionality and complexity are used to illustrate the performance of the methodology.

4.1.1 | One-dimensional unsteady droplet motion

To begin with, a one-dimensional case of an unsteady droplet velocity field is considered,¹⁷ which operates under the assumption that droplets are characterised by a large inertia ($St \gg 1$). Droplets are initially released with a non-uniform velocity distribution, and their motion and Jacobian evolve according to the equations

$$\ddot{x}_d = 0, \quad x_d(t_0) \in [0, 1], \dot{x}_d(t_0) = 1 - x_d^2(t_0), \quad (19a)$$

$$\ddot{j}^{xx} = 0, \quad J^{xx}(t_0) = 1, \dot{j}^{xx}(t_0) = -2x_d(t_0). \quad (19b)$$

Equations (19) admit the analytical solution

$$x_d(t) = x_0 + (1 - x_0^2)t, \quad (20a)$$

$$J^{xx}(t) = 1 - 2x_0t. \quad (20b)$$

The associated number density $n(x_d, t)$ evolves according to Equation (2) with initial condition $n(x_0, t) = n_0$, and the corresponding solution in this case is given by

$$n(x_d, t) = \frac{n_0}{|1 - 2x_0t|}. \quad (21)$$

From Equation (20a) it follows that a fold is formed at $t = 0.5$, with the envelope of droplet trajectories along this being

$$x = t + \frac{1}{4t}, \quad t \geq 0.5. \quad (22)$$

In this case the Eulerian number density field can also be obtained as

$$n(x, t) = \begin{cases} n_0 / \sqrt{1 - 4tx + 4t^2} & t < x < 1 \text{ or } 1 < x < t, \\ 2n_0 / \sqrt{1 - 4tx + 4t^2} & 0.5 < t < x \text{ and } x > 1, \end{cases} \quad (23)$$

where $n_0 = n(x_0, t_0)$, and in which it is assumed that droplets moving towards the envelope of the trajectories form a separate layer of the fold to those droplets moving away from the envelope, and that these layers belong to different non-interacting continua. The total number density in the region is simply double that in the remainder of the x - t domain which is occupied by droplets, resulting in the piecewise nature of Equation (23). Along the fold envelope (22) the droplet number density becomes infinitely high, however this singularity remains integrable.¹⁷ The analytical solution (23) is graphically depicted in x - t space in Figure 1A, and features a discontinuity along the internal boundary of the multi-valued region given by the second sub-function of Equation (23). The region of the domain occupied by the droplet phase is reconstructed on a grid at time intervals of $\Delta t = 0.01$ with 100 uniformly spaced points of separation Δx at each time, and uses an initial seeding of 101 droplets positioned uniformly on the interval $x_0 \in [0, 1]$, corresponding to an initial average inter-droplet spacing of $\Delta x_{d0} = 0.01$.

To assess the efficacy of the kernel regression procedure in Section 3, the Eulerian number density field is reconstructed from the Lagrangian number density (21) using the Nadaraya-Watson estimator in Equation (8). Since this case is one-dimensional, the kernel given in Equation (10) is utilised, with the smoothing length h specified by Equation (12). This produces the result shown in Figure 1B. For comparison purposes, it is appropriate to consider the relative error

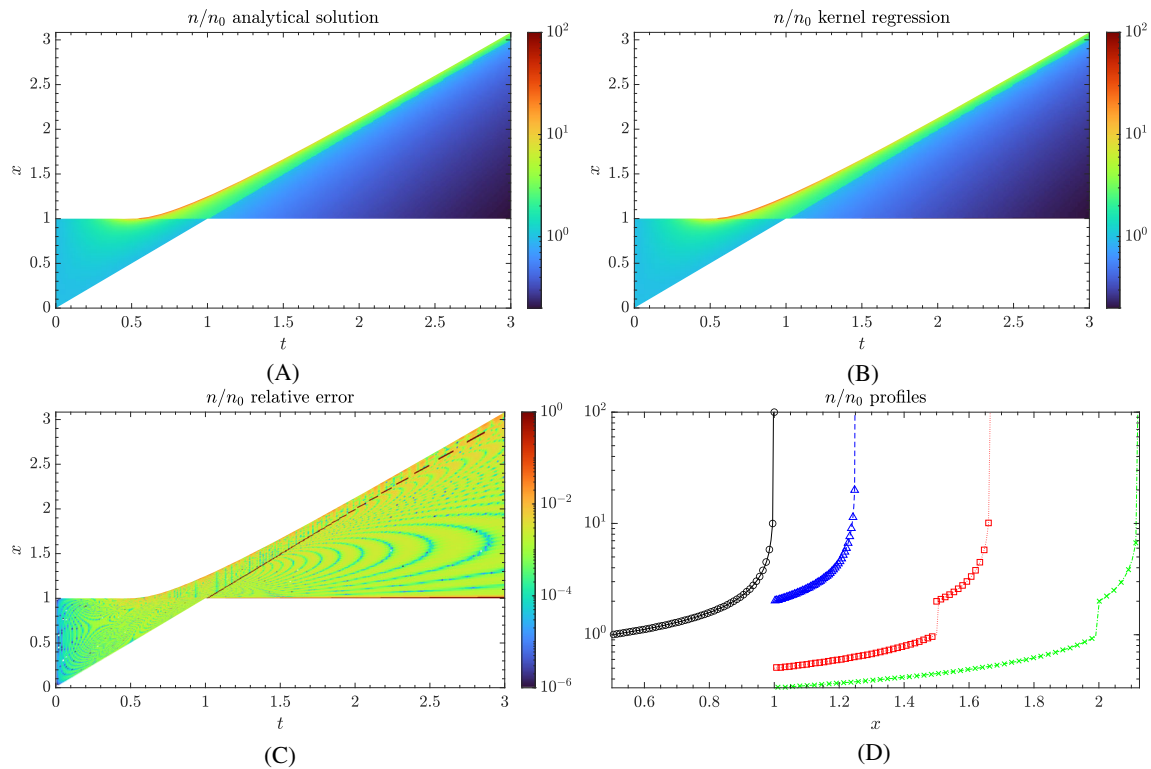


FIGURE 1 (A) Analytical solution for n/n_0 as given in Equation (23); (B) n/n_0 obtained from kernel regression; (C) Relative error in n/n_0 between figures (A) and (B); (D) Comparison of kernel regression (symbols) with analytical solution (lines) for profiles of n/n_0 at selected instants in time: $[-\circ-]$ $t = 0.5$, $[-\triangle-]$ $t = 1.0$, $[-\square-]$ $t = 1.5$, $[-\times-]$ $t = 2.0$. [Colour figure can be viewed at wileyonlinelibrary.com]

between kernel regression and the analytical solution, and this is presented in Figure 1C. It is seen that the relative error in the kernel regression is generally below 10^{-2} throughout the x - t domain, except at the fold envelope (22) and along the boundary of the multi-valued region. This more marked error arises due to the smoothing property of kernels, in that the influence of a given droplet is distributed over a small region of the domain, with the consequence that representation of steep gradients and discontinuities in the number density field cannot be exactly captured using this procedure. Therefore it is expected that some additional error will result from the use of kernel regression in reconstructing the number density field, however an appropriate choice of initial smoothing length h_0 can mitigate against this,²¹ with a smaller kernel generally being preferable subject to n remaining smooth. In this case, the smoothing length is set proportional to the variable grid spacing at each point in time, with the initial value of $h_0 = \Delta x_{d0}/3$ being chosen in order to achieve a high degree of fidelity in the reconstructed Eulerian number density field. The overall accuracy of the kernel regression is seen to be good, as can be inferred from examination of the spatial profiles of n at selected instances in time, depicted in Figure 1D. The only discernible loss of accuracy is in the vicinity of the discontinuity between the single-valued and multi-valued regions of Equation (23) for $t \geq 1$, as can be seen in the jump in the profiles for $y = 1.5$ and $y = 2$. Moreover, it is seen that whilst n remains finite in the vicinity of the fold envelope (22), kernel regression is able to capture the increase in number density well, with the error only becoming greater than 10^{-2} along the fold itself.

4.1.2 | Two-dimensional fan spray injection in cross-flow

Consider a vertical injection of droplets with velocity magnitude v^* into a horizontal cross flow with constant velocity $\mathbf{u} = (1, 0)$. Taking the case of $St = 1$, the droplet and Jacobian equations of evolution, given by Equations (1a) and (4a) respectively, are

$$\dot{\mathbf{x}}_d = \mathbf{u} - \dot{\mathbf{x}}_d, \quad (24a)$$

$$\dot{\mathbf{J}}^{\mathbf{xx}} = -\beta \mathbf{J}^{\mathbf{xx}}, \quad (24b)$$

where the initial conditions for \mathbf{x}_d in Equation (24a) are given by

$$\begin{aligned} x_{d,1}(t_0) &\in [-\epsilon, \epsilon], & \dot{x}_{d,1}(t_0) &= v^* \sin\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right), \\ x_{d,2}(t_0) &= 0, & \dot{x}_{d,2}(t_0) &= v^* \cos\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right), \end{aligned} \quad (25)$$

and the initial conditions for $\mathbf{J}^{\mathbf{xx}}$ in Equation (24b) are given by

$$\begin{aligned} \mathbf{J}^{\mathbf{xx}}(t_0) &= \mathbf{I}, \\ j_{11}^{\mathbf{xx}}(t_0) &= \frac{1}{\epsilon} \frac{\pi}{4} v^* \cos\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right), \\ j_{12}^{\mathbf{xx}}(t_0) &= -\frac{1}{\epsilon} \frac{\pi}{4} v^* \sin\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right) + \frac{1 - \dot{x}_{d,1}(t_0)}{\dot{x}_{d,2}(t_0)}, \\ j_{21}^{\mathbf{xx}}(t_0) &= -\frac{1}{\epsilon} \frac{\pi}{4} v^* \sin\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right), \\ j_{22}^{\mathbf{xx}}(t_0) &= \frac{1}{\epsilon} \frac{\pi}{4} v^* \sin^2\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right) / \cos\left(\frac{\pi x_{d,1}(t_0)}{4\epsilon}\right) - 1, \end{aligned} \quad (26)$$

in which the expressions for $j_{12}^{\mathbf{xx}}(t_0)$ and $j_{22}^{\mathbf{xx}}(t_0)$ are derived from Reference 5

$$\frac{\partial}{\partial x_{0,2}} = -\frac{\dot{x}_{d,1}(t_0)}{\dot{x}_{d,2}(t_0)} \frac{\partial}{\partial x_{0,1}} + \frac{1}{\dot{x}_{d,2}(t_0)} \frac{\partial}{\partial t}. \quad (27)$$

The system (24) with initial conditions (25) and (26) admits the analytical solution

$$\begin{aligned}
 x_{d,1}(t) &= x_{d,1}(t_0) + (1 - \dot{x}_{d,1}(t_0)) \exp(-t), \\
 x_{d,2}(t) &= x_{d,2}(t_0)(1 - \exp(-t)), \\
 \dot{x}_{d,1}(t) &= 1 - (1 - \dot{x}_{d,1}(t_0)) \exp(-t), \\
 \dot{x}_{d,2}(t) &= \dot{x}_{d,2}(t_0) \exp(-t), \\
 \mathbf{J}^{\mathbf{x}\mathbf{x}}(t) &= \mathbf{I} + \mathbf{J}^{\mathbf{x}\mathbf{x}}(t_0)(1 - \exp(-t)), \\
 \mathbf{j}^{\mathbf{x}\mathbf{x}}(t) &= \mathbf{j}^{\mathbf{x}\mathbf{x}}(t_0) \exp(-t).
 \end{aligned} \tag{28}$$

From Equation (28), it follows that the associated number density $n(\mathbf{x}_d, t)$ with initial condition $n(\mathbf{x}_0, t) = n_0$ evolving according to Equation (2) is given by

$$n(\mathbf{x}_d, t) = \frac{n_0}{|1 + (1 - \exp(-t))(j_{11}^{\mathbf{x}\mathbf{x}}(t_0) + j_{22}^{\mathbf{x}\mathbf{x}}(t_0)) + (1 - \exp(-t))^2(j_{11}^{\mathbf{x}\mathbf{x}}(t_0)j_{22}^{\mathbf{x}\mathbf{x}}(t_0) - j_{12}^{\mathbf{x}\mathbf{x}}(t_0)j_{21}^{\mathbf{x}\mathbf{x}}(t_0))|}. \tag{29}$$

It is then possible to numerically evaluate the Eulerian number density field in this case using Equation (29) and the parametric solutions given in Equation (28). The droplet field includes a multi-valued region containing separate layers of a fold, with the envelope of this fold occurring along the top edge of the spray, and the other boundary of the multi-valued region presenting an internal discontinuity within the spray. The analytical solution from Equations (28) to (29) is graphically depicted as a steady-state distribution in x - y space in Figure 2A, where the droplet number density field is reconstructed on a uniform Cartesian grid with spacing of $\Delta x = \Delta y = 0.0025$. In this example the value of the parameter ϵ used to define the interval in which droplets are injected in Equation (25) is taken to be $\epsilon = 0.05$, with a total of 101 droplets being injected uniformly over $\mathbf{x} \in [-\epsilon, \epsilon], 0$ at the start of the simulation, giving a value of $\Delta x_{d0} = 0.001$.

Numerical reconstruction of the steady-state droplet field from the Lagrangian number density (29) using the Nadaraya-Watson estimator (8) is shown in Figure 2B. The structured kernel in Equation (11) is used for this

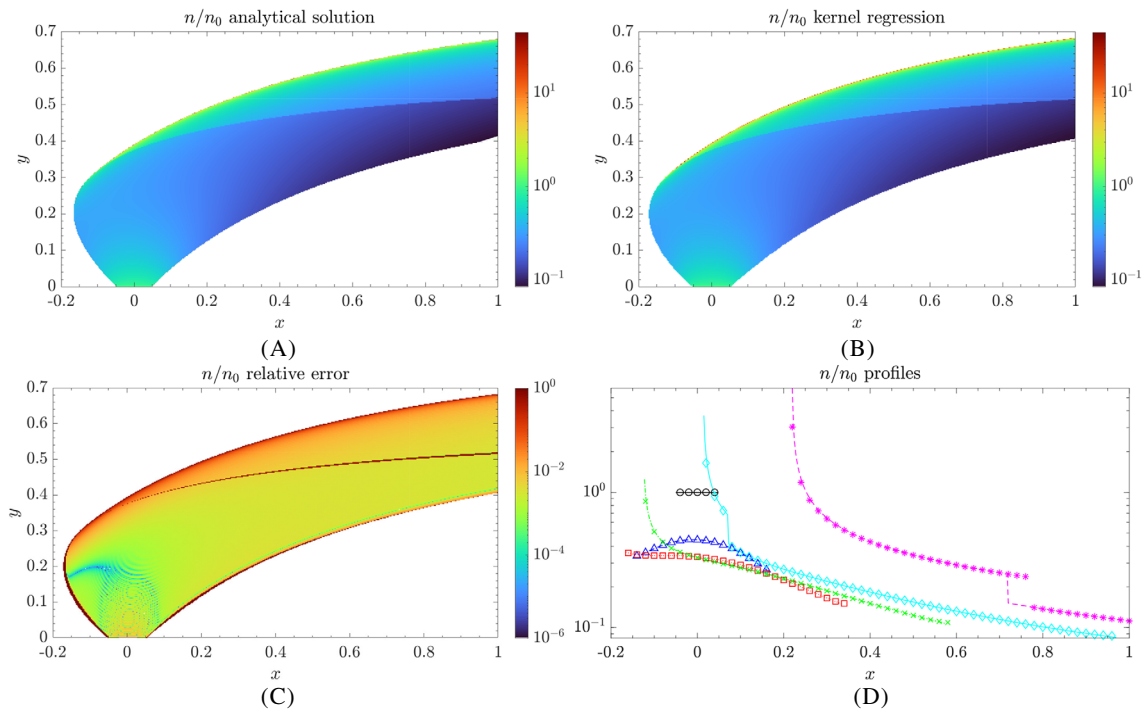


FIGURE 2 (A) Analytical solution for n/n_0 obtained from Equations (28) and (29); (B) n/n_0 obtained from kernel regression; (C) Relative error in n/n_0 between figures (A) and (B); (D) Comparison of kernel regression (symbols) with analytical solution (lines) for profiles of n/n_0 at selected y locations: $[-\circ-]$ $y = 0$, $[-\triangle-]$ $y = 0.1$, $[-\square-]$ $y = 0.2$, $[-\times-]$ $y = 0.3$, $[-\diamond-]$ $y = 0.4$, $[-*-]$ $y = 0.5$. [Colour figure can be viewed at wileyonlinelibrary.com]

multidimensional case, with the bandwidth matrix \mathbf{H} being specified by Equation (14). The relative error of kernel regression is shown in Figure 2C, and as with the one-dimensional case in Section 4.1.1 generally remains below 10^{-2} except in the vicinity of the fold envelope, along the discontinuity between the multi-valued and single-valued regions of the spray, and along the edges of the spray. Again, it is the perceptible size of the kernel which causes these more marked errors to arise, and this is especially apparent along the edges of the spray where the kernel extends beyond the region occupied by droplets into the surrounding area. This effect is balanced against the need to keep the kernel large enough to produce a smooth number density field from the sample of trajectories used in the simulation, once again requiring careful choice of the initial smoothing length h_0 . In this instance, the default value of $h_0 = \Delta x_{d0}$ is used to achieve this. Notwithstanding this, the overall accuracy of kernel regression can be observed from the profiles of n at selected y values in Figure 2D, and is seen to generally be good with the exception of the discontinuity along the edge of the multi-valued region in the spray. This is observed in the profile for $y = 0.5$, where kernel regression does not capture the exact location of the discontinuity as a result. Aside from this, the physical behaviour and increases in number density near the fold envelope are well captured.

4.1.3 | Flow around a cylinder: Steady-state case ($Re = 20$)

To determine the capabilities of the kernel regression procedure in a context which is more representative of a general engineering flow, the case of a two-dimensional gas-droplet flow around a cylinder is considered. This is a prototypical problem, and exhibits distinct flow regimes depending upon the flow Reynolds number Re ; specifically a symmetric steady-state flow at low Re , and a periodic von Kármán vortex street beyond the critical value of Re at which the transition to unsteadiness occurs.

The evolution of the carrier flow is described using the Navier-Stokes equations for an incompressible fluid

$$\nabla \cdot \mathbf{u} = 0, \quad (30a)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{Re} \nabla^2 \mathbf{u}. \quad (30b)$$

The cylinder is taken to have radius R which is used as the representative lengthscale, and the uniform free-stream velocity at which fluid enters the inlet of the domain is chosen as the associated velocity scale U . The Reynolds number is then defined in this work as $Re = UR/\nu$, in contrast to much of the existing literature in which the cylinder diameter is chosen as the characteristic lengthscale. In this context, the critical Reynolds number at which the transition to unsteadiness occurs is $Re_c = 23.5$.

To investigate droplet behaviour in the steady-state regime, the case of $Re = 20$ is considered. Since the underlying carrier flow is symmetrical about the cylinder in the direction normal to the flow, the droplet distribution also exhibits this symmetry. Droplets are injected at $x/R = -5$ over the interval $y/R \in [-4, 4]$ with the free-stream carrier flow velocity $\mathbf{u}/U = (1, 0)$. In the steady-state cases, a total of 101 droplets are injected over this interval in a square-law profile at the start of the simulation to give an average inter-droplet spacing of $\Delta x_{d0} = 0.08$, with the droplet seeding increasing towards the centreline $y/R = 0$. Interaction of the droplets with the cylinder is not included within the simulation, and any droplets which do come into contact with the cylinder are subsequently omitted from the calculation of the number density field at later times.

The droplet trajectories evolve according to a linear drag law, which for the case of monodisperse non-evaporating droplets determines the equations for motion and Jacobian evolution from Equations (1a) to (4a) as

$$\dot{\mathbf{x}}_d = \frac{1}{St} (\mathbf{u}(\mathbf{x}_d(t), t) - \dot{\mathbf{x}}_d), \quad (31a)$$

$$\dot{\mathbf{J}}^{\mathbf{xx}} = \frac{1}{St} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}(\mathbf{x}_d(t), t) \cdot \mathbf{J}^{\mathbf{xx}} - \frac{1}{St} \mathbf{J}^{\mathbf{xx}}. \quad (31b)$$

The associated initial conditions are given by

$$\mathbf{x}_d(t_0) = (-5, [-4, 4]), \quad \dot{\mathbf{x}}_d(t_0) = (1, 0), \quad (32a)$$

$$\mathbf{J}^{\mathbf{xx}}(\mathbf{x}_0, t_0) = \mathbf{I}, \quad \dot{\mathbf{J}}^{\mathbf{xx}}(\mathbf{x}_0, t_0) = \mathbf{0}, \quad (32b)$$

where the initial condition on $\mathbf{J}^{\mathbf{x}\mathbf{x}}$ is due to the uniformity of the droplet phase at the point of injection. The Lagrangian number density $n(x_d, t)$ associated with a droplet is then calculated using Equation (2) without the appearance of the droplet radius r_d as a parameter.

Equations (31) are solved using a customisation of the Lagrangian particle tracking library in OpenFOAM, and can be coupled to the required solver for the carrier flow; in this case pimpleFoam has been used to solve Equations (30). The domain size for the simulations is $-20 \leq x/R \leq 30$ and $-20 \leq y/R \leq 20$ in a Cartesian coordinate system centred on the cylinder. Further details of the numerical setup, mesh independence, and validation can be found in Reference 14.

In order to reconstruct the Eulerian droplet number density field using the Nadaraya-Watson estimator (8) in the general multidimensional case, the structured kernel in Equation (11) is used with the bandwidth matrix \mathbf{H} specified by Equation (14). The reconstruction is done on a uniform Cartesian grid with a spacing of $\Delta x/R = \Delta y/R = 0.04$. This is considered for three different droplet sizes which correspond to $St = 0.1, 1, \text{ and } 10$, and the number density field n reconstructed using kernel regression is displayed in Figure 3A,C,E respectively for each of these cases. The profiles of the number density field at selected values of x are also displayed in Figure 3B,D,F for these respective values of St .

Both the extent and number density distribution of the droplet field is seen to vary markedly with St , with the common feature to all cases being the wake behind the cylinder which is devoid of droplets. For $St = 0.1$ the droplets follow the flow relatively closely, and the two regions of the droplet field formed by droplets travelling above and below the cylinder meet downstream of the cylinder after a distance of $\sim 8R$. In contrast, the droplet field at the higher values of St remains separated into two distinct regions beyond $x = 25R$. The other key difference between the cases lies within the rate of variation in number density of the droplet field. For $St = 0.1$ the droplet field is largely uniform away from the cylinder, and only experiences rapid variation along the edges of the cylinder wake, whilst $St = 1.0$ and $St = 10$ display a larger region of variation in the number density field, but with the change being progressively more gradual.

The profiles in Figure 3B,D,F provide a means of assessing the efficacy of kernel regression against the exact values of number density obtained from the FLA. This is possible since in the case of a steady-state flow it is permissible to interpolate the values of number density along trajectories between time points,¹⁴ and therefore obtain an accurate

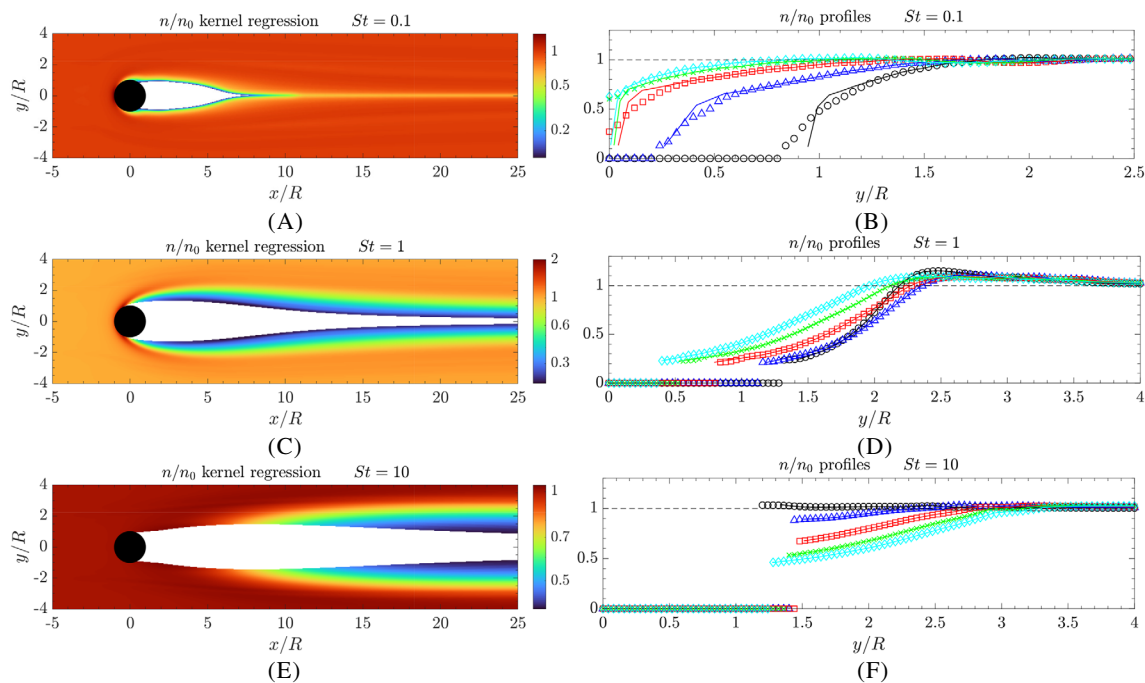


FIGURE 3 Reconstruction of n/n_0 using kernel regression in monodisperse steady-state flow around a cylinder at $Re = 20$ for: (A) $St = 0.1$; (C) $St = 1$; (E) $St = 10$; Profiles of n/n_0 obtained using kernel regression (symbols) at selected locations [\circ] $x/R = 3$, [\triangle] $x/R = 6$, [\square] $x/R = 9$, [\times] $x/R = 12$, [\diamond] $x/R = 15$ compared against linear interpolation (lines) of FLA number density data for: (B) $St = 0.1$; (D) $St = 1$; (F) $St = 10$. [Colour figure can be viewed at wileyonlinelibrary.com]

descriptor of the number density field along given cross-section profiles independently of kernel regression. Owing to the steady-state behaviour of the flow, the number density varies smoothly without the presence of folds and also largely monotonically, and these characteristics essentially mean that linear interpolation of the number density values along trajectories onto a chosen profile provides an exact solution against which kernel regression can be compared. Of particular note is the case for $St = 0.1$, in which the kernel regression performs well across a large part of the flow, but loses accuracy in the region of rapid variation along the edge of the cylinder wake, as depicted in Figure 3B. This is due to the perceptible size of the kernel causing the regression procedure to introduce a certain level of smoothing into the result, and therefore limiting the level of variation in number density which can be accurately captured. However, this is balanced against the kernel needing to be large enough to provide a smooth representation of the number density field for a given initial droplet seeding within a simulation, and as with the previous examples emphasises the importance of judicious selection of the initial smoothing length h_0 . In these cases the default value of $h_0 = \Delta x_{d0}$ is selected to provide a reasonable coverage of the computational domain for the chosen initial seeding of 101 droplets which is used. Increasing the initial droplet seeding would enable a smaller value of h_0 to smoothly reconstruct the number density field, which would in turn provide greater accuracy in the regions of rapid variation at lower St . Nonetheless, at higher St the more gradual variation in number density for the different x profiles is successfully accounted for by kernel regression across the entire droplet phase, and the procedure provides a high level of accuracy as demonstrated in Figure 3D,F.

An illustration of the computational efficiency afforded by kernel regression can be made for this steady state approach by comparing to the number density field obtained using direct trajectory methods. In this instance, the Cloud-In-Cell approach that was outlined in Section 3.1 is used, with the accumulation made upon the same Eulerian grid as used for kernel regression. Since the number density is calculated directly in the CIC approach, a sufficient number of droplets are required in each grid cell to produce a stable result, and this is achieved by a higher initial seeding of droplets across the injection interval $y/R \in [-4, 4]$. In contrast to the FLA which uses 101 droplets within this range, the CIC approach requires 10,001 uniformly spaced droplets across the interval for the resultant Eulerian number density field to sufficiently converge, and furthermore the injection rate of the droplets also has to be increased by a factor of 10. Thus overall the number of droplet realisations required by the CIC approach is found to be 10^3 times more than that by the FLA, consistent with previous works. The number density fields produced by this procedure are displayed in Figure 4A,C,E for each of $St = 0.1, 1, \text{ and } 10$ respectively, and it can be observed that the number density distribution is largely similar to that obtained from the FLA in the corresponding cases shown in Figure 3A,C,E. The much higher seeding of trajectories needed for the CIC approach does, however, result in some differences to the profile of the wake behind the cylinder, and this is best seen through a direct comparison of the two procedures as illustrated in Figure 4B,D,F. This depicts the relative error between the number density fields produced by the kernel regression and CIC approaches, and it is seen that this is largest along the edge of the wake, where the error can exceed 10^{-1} . Away from the wake however, the error generally varies between 10^{-2} and 10^{-1} , which is indicative of the level of accuracy that kernel regression of FLA data is able to achieve using 10^3 times fewer droplet realisations, and the associated computational speedup that comes with this.

4.1.4 | Flow around a cylinder: Transient case ($Re = 100$)

For consideration of droplet behaviour in the transient regime, the case of $Re = 100$ is used. The underlying carrier flow becomes periodic at this level of unsteadiness, and forms the well-known phenomenon of a von Kármán vortex street. The configuration is identical to that in Section 4.1.3 except that the value of Re is higher and the initial droplet seeding consists of 81 droplets that are injected over the interval $y/R \in [-4, 4]$ with uniform spacing at each timestep. This corresponds to an initial average inter-droplet spacing of $\Delta x_{d0} = 0.1$, however the initial smoothing length is kept as $h_0 = 0.08$ as for the steady-state case, in order to achieve better resolution of the transient structures in the droplet number density field. In practice, this value of $h_0 = 0.8\Delta x_{d0}$ also represents the lower size limit of initial smoothing length that will result in a smooth number density field for transient configurations, and as such has been found to be the optimum value for achieving this trade-off in these cases. As before, the carrier flow evolves according to Equations (30) and the droplets and Jacobian are governed by Equations (31). The three distinct droplet sizes corresponding to $St = 0.1, 1, \text{ and } 10$ are again considered, with the number density field n reconstructed using kernel regression at $t = 40$ displayed in Figure 5A,C,E respectively for each case, and the associated profiles of the number density field at selected values of x given in Figure 5B,D,F.

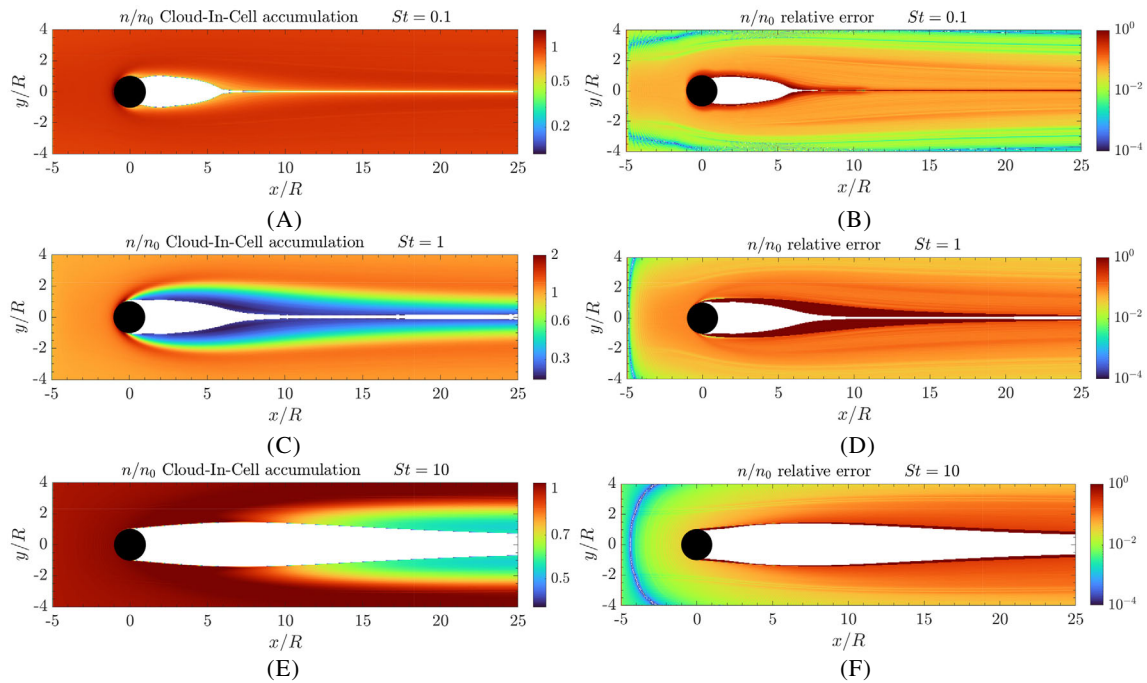


FIGURE 4 Reconstruction of n/n_0 using cloud-in-cell accumulation in monodisperse steady-state flow around a cylinder at $Re = 20$ for: (A) $St = 0.1$; (C) $St = 1$; (E) $St = 10$; Relative error in n/n_0 between kernel regression and cloud-in-cell accumulation for: (B) $St = 0.1$ (from Figures 3A and 4A); (D) $St = 1$ (from Figures 3C and 4C); (F) $St = 10$ (from Figures 3E and 4E). [Colour figure can be viewed at wileyonlinelibrary.com]

The droplet behaviour is seen to reflect that of the carrier flow, with clear build-ups and voids in the number density field that are directly influenced by the structure of the vortices that form in the wake of the cylinder. The crucial feature of the dispersed phase behaviour is the effect of different levels of droplet inertia on the number density field at the different values of St . For $St = 0.1$ the voids in the droplet number density field are closely aligned with the location of vortices in the carrier flow due to the relatively low droplet inertia, as observed in Figure 5A. This is consistent with Maxey's centrifuging mechanism,³³ which argues that low inertia droplets are ejected from areas of high vorticity. Of greater interest are the areas of higher droplet number density, which form along distinct curves between the vortices. The representation of this in Figure 5A is a demonstration that kernel regression is able to capture this level of detail accurately in the reconstructed number density field, and successfully account for the complex behaviour of the droplet field in transient flows. At higher St , the voids of low droplet number density that occur around vortices become much larger, in addition to the appearance of a wake downstream of the cylinder in which no droplets are present, as seen in Figure 5C,E. For $St = 1$ in (5c), the droplet field evolves as alternating regions in which droplets occur followed by voids that contain no droplets, with the region in which droplets are present characterised by a fold line of high number density along the leading edge and an area of lower number density along the trailing edge. For $St = 10$ this behaviour is also observed, with the fold lines exhibiting a higher number density than for $St = 1$, and the trailing edge of the droplet field containing both regions of lower number density and distinct layers of folds as observed from the discontinuities in the droplet field in Figure 5E. The successful capturing of these fold layers shows that kernel regression is able to retain this level of detail from the FLA number density data, and account for this phenomenon at a far lower computational cost than direct trajectory methods.¹⁵

The x profiles of the number density field in Figure 5B,D,F exhibit the different levels of variation in number density that the kernel regression procedure is able to reproduce. For intervals where there is no data for a given profile, the number density is zero as there are no droplets present at that point. It can be seen that as St increases, the maximum number density within the flow becomes higher, and the maximum spatial gradient of the number density also increases. In particular, for $St = 1$ and $St = 10$ the profiles in the vicinity of fold lines display rapid variation in number density, however kernel regression is capable of providing a smooth representation of this behaviour, which serves to illustrate the flexibility of this procedure for use in complex flow configurations.

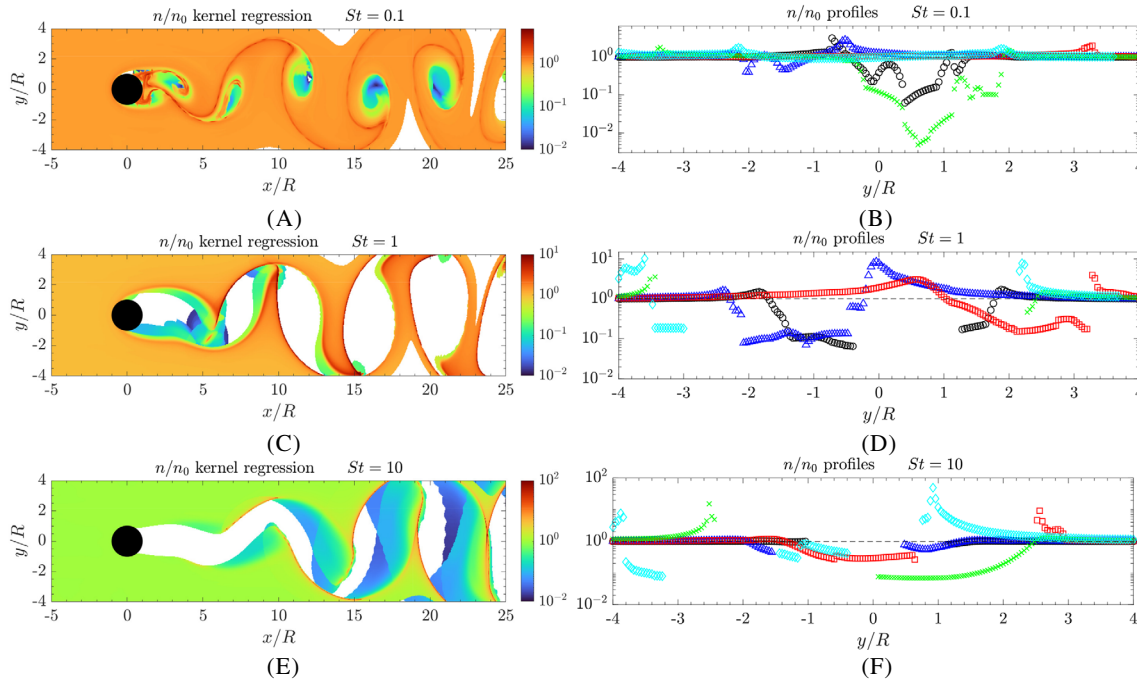


FIGURE 5 Reconstruction of n/n_0 using kernel regression in monodisperse transient flow around a cylinder at $Re = 100$ and time $t = 40$ for: (A) $St = 0.1$; (C) $St = 1$; (E) $St = 10$; Profiles of n/n_0 obtained using kernel regression at selected locations [\circ] $x/R = 3$, [\triangle] $x/R = 6$, [\square] $x/R = 9$, [\times] $x/R = 12$, [\diamond] $x/R = 15$ for: (B) $St = 0.1$; (D) $St = 1$; (F) $St = 10$. [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Polydisperse droplets

Generalisation to the case of polydisperse droplets enables investigation of the droplet size distribution within a flow. Extension of the kernel regression framework into radial space is outlined in Section 3.5, and in the following some of the flow configurations from the monodisperse cases in Section 4.1 are used to illustrate the efficacy of this procedure.

To specify the different droplet sizes within simulations, an appropriate probability distribution is defined at the outset as a function of the initial droplet radius $r_0 = r_d(t_0)$. Following previous work,¹⁴ a log-normal distribution that is the same at all initial locations $\mathbf{x}_0 = \mathbf{x}_d(t_0)$ is assumed,

$$p(\mathbf{x}_0, r_0, t_0) = \frac{1}{r_0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\ln(r_0) - \mu)^2}{2\sigma^2}\right], \quad (33)$$

in which the mean and standard deviation parameters are chosen to be $\mu = 0.16$ and $\sigma = 0.4$ respectively. The droplet size r_0 in Equation (33) is nondimensionalised by r_{d0}^* , which is a reference droplet radius corresponding to the peak of the distribution. Whilst a log-normal distribution is representative of the spread of droplet sizes in some applications,³⁴ an alternative distribution, for instance Rosin-Rammler, can easily be specified if it is more physically appropriate.

Of interest in polydisperse droplet flows is not only the distribution of droplet sizes, but also the spatial distribution and average size statistics across all sizes of droplet. These can be determined by considering the moments of p , which utilises its interpretation as a probability density field. Specifically, the number density n , average radius \bar{r} , and radius variance $\overline{r'r'}$ can be obtained using the definitions

$$n(\mathbf{x}, t) = \int_r p(\mathbf{x}, r, t) dr, \quad (34a)$$

$$\bar{r}(\mathbf{x}, t) = \frac{1}{n(\mathbf{x}, t)} \int_r r p(\mathbf{x}, r, t) dr, \quad (34b)$$

$$\overline{r'r'}(\mathbf{x}, t) = \frac{1}{n(\mathbf{x}, t)} \int_r (r - \bar{r})(r - \bar{r}) p(\mathbf{x}, r, t) dr. \quad (34c)$$

The averaged field variables in Equations (34) are evaluated within simulations by numerically integrating the probability density p obtained from the kernel regression procedure across all droplet sizes in accordance with the various radius weightings used. Here attention is restricted to just the moments of p given in Equations (34), however it is possible to define further higher-order statistics of the droplet size distribution as required by the case under consideration, for instance the skewness and kurtosis may be of interest in transient flows. It is also possible to obtain relevant statistics for industrial spray systems, for instance the Sauter mean diameter, from p using a similar procedure.

4.2.1 | One-dimensional quiescent flow with evaporating droplets

To illustrate the behaviour of evaporating droplets in a simple case, a one-dimensional flow of droplets in quiescent air with uniform temperature is considered, as previously used in Reference 14. Droplets are initially located at $x_0 = 0$, with velocity $v_0 = 1$, radius $r_0 \in [0, 4]$ and the probability density $p(\mathbf{x}_d(t_0), r_d(t_0), t_0)$ as specified by Equation (33). For the case of $St_0^* = 1$, the droplet motion and evaporation governed by Equations (1) become

$$\ddot{x}_d = -\frac{1}{r_d^2}\dot{x}_d, \quad x_d(t_0) = 0, \dot{x}_d(t_0) = 1, \quad (35a)$$

$$\dot{r}_d = -\frac{\delta}{2r_d}, \quad r_d(t_0) \in [0, 4], \quad (35b)$$

and the corresponding Jacobian evolution given in Equations (4) becomes

$$\dot{j}^{xx} = -\frac{1}{r_d^2}j^{xx} + \frac{2}{r_d^3}\dot{x}_d j^{rx}, \quad (36a)$$

$$\dot{j}^{xr} = -\frac{1}{r_d^2}j^{xr} + \frac{2}{r_d^3}\dot{x}_d j^{rr}, \quad (36b)$$

$$\dot{j}^{rx} = \frac{\delta}{2r_d^2}j^{rx}, \quad (36c)$$

$$\dot{j}^{rr} = \frac{\delta}{2r_d^2}j^{rr}, \quad (36d)$$

with the initial conditions as given in Equations (5) along with $j^{xx}(x_0, r_0, t_0) = 0$. The systems (35) and (36) admit analytical solutions along trajectories as detailed in Reference 14. The probability density along trajectories is calculated using Equation (2), however to reconstruct the probability density field it is necessary to use multidimensional kernel regression as outlined in Section 3.5, with the corresponding phase space for this case being $\xi = (x, r)$. For this case, the droplet evaporation rate is specified by $\delta = 1$, and a total of 100 droplet sizes defined uniformly over the range $r_0 \in [0, 4]$ are injected at the start of the simulation, giving an initial inter-droplet size spacing of $\Delta r_{d0}/r_{d0}^* = 0.04$. The initial radial space smoothing length is set as $h_{r0} = 0.5\Delta r_{d0}$ to achieve a good resolution over the droplet size distribution, whilst as all droplets are started from the same location at $x = 0$ and Δx_{d0} is consequently not defined in this case, the initial physical space smoothing length is also taken to be $h_{x0} = h_{r0}/r_{d0}^* = 0.5\Delta r_{d0}/r_{d0}^*$ for consistency. The instantaneous phase space probability density field p is reconstructed on a uniform Cartesian grid with spacing $\Delta x = 0.04$ and $\Delta r/r_{d0}^* = 0.04$

The complete Eulerian probability density field in (x, r) space is depicted in Figure 6A, and it is seen that kernel regression is able to reproduce a smoothly varying distribution across both physical space and radial space, enabling the full behaviour of the evaporation process to be examined. In Figure 6B, the droplet size distribution profiles at selected x locations are shown. The profile at $x = 0$ is the initial probability density distribution specified in Equation (33), with the profiles at subsequent locations reflecting the decrease in probability density as droplets evaporate whilst they travel through the domain. Figure 6C depicts the droplet spatial distribution profiles at selected droplet sizes, with the peak probability occurring for $r = 1$ as also determined by Equation (33). The averaged field variables obtained from the moments of p using Equations (34) are given in Figure 6D. The number density n initially shows a minor increase at small x , before droplets begin to fully evaporate and n subsequently decreases exponentially. In contrast, the average radius \bar{r} initially decreases at small x , before gradually growing and reaching its peak value at $x \approx 5.5$. This is indicative of the small droplets which form the peak of the distribution (33) evaporating more quickly than larger droplets, causing the

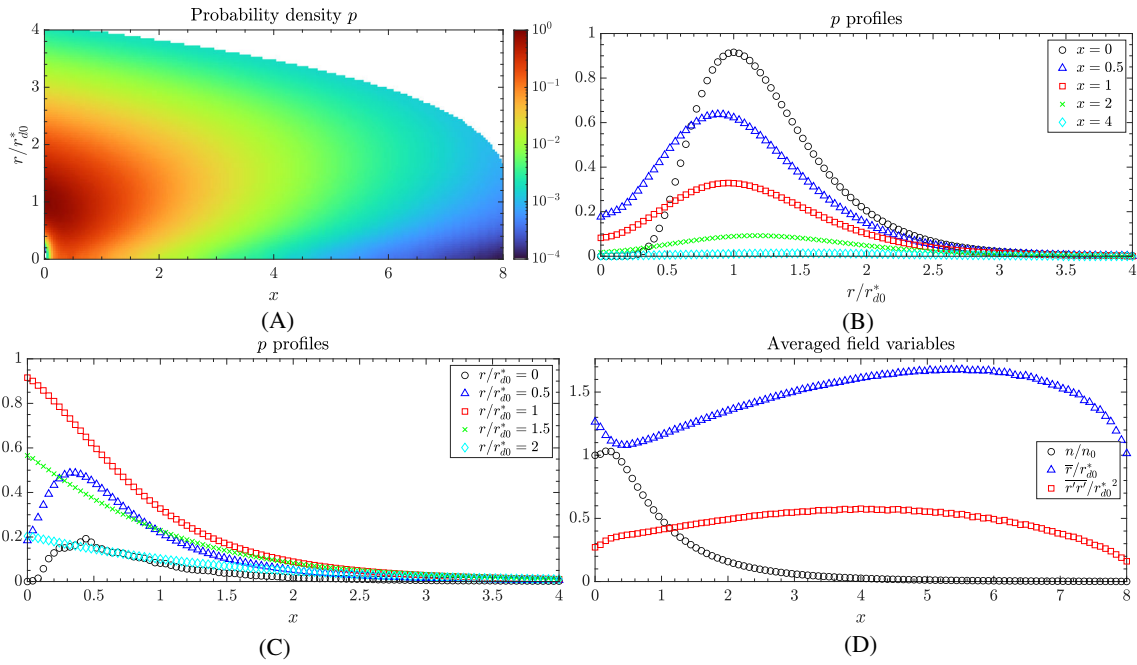


FIGURE 6 (A) Reconstruction of the probability density field p using kernel regression for a one-dimensional quiescent flow with evaporating droplets; (B) Profiles of p at selected x locations; (C) Profiles of p at selected values of r/r_{d0}^* ; (D) Averaged field variables: $[\circ]$ n/n_0 , $[\triangle]$ \bar{r}/r_{d0}^* , $[\square]$ $\overline{r'r'}/r_{d0}^{*2}$. [Colour figure can be viewed at wileyonlinelibrary.com]

average size to increase during this period. Following this, \bar{r} decreases as the larger droplets evaporate. Finally, the variance in droplet radius $\overline{r'r'}$ slowly increases as droplets travel across the domain, reaching a peak value at $x \approx 4$. This reflects the increased spread of the profiles displayed in Figure 6B at successive locations, before $\overline{r'r'}$ slowly decreases across the remainder of the domain. Although a simple example, this case serves to demonstrate the wealth of information that kernel regression is able to efficiently extract from the trajectory data in polydisperse flows.

4.2.2 | Two-dimensional fan spray injection in cross-flow

This case is an extension of the configuration that was considered in Section 4.1.2 to polydisperse evaporating droplets, and has also been previously studied in Reference 14. For the case of $St_0^* = 1$, droplet motion and evaporation are governed by

$$\ddot{\mathbf{x}}_d = \frac{1}{r_d^2} (\mathbf{u} - \dot{\mathbf{x}}_d), \quad (37a)$$

$$\dot{r}_d = -\frac{\delta}{2r_d}, \quad r_d(t_0) \in [0, 4], \quad (37b)$$

where $\mathbf{u} = (1, 0)$ as before, and the initial conditions for \mathbf{x}_d and $\dot{\mathbf{x}}_d$ are identical to those in Equation (25). The corresponding Jacobian components evolve according to

$$\dot{\mathbf{j}}^{xx} = -\frac{1}{r_d^2} \mathbf{j}^{xx} - \frac{2}{r_d^3} (\mathbf{u} - \dot{\mathbf{x}}_d) \cdot \mathbf{J}^x, \quad (38a)$$

$$\dot{\mathbf{j}}^{xr} = -\frac{1}{r_d^2} \mathbf{j}^{xr} - \frac{2}{r_d^3} (\mathbf{u} - \dot{\mathbf{x}}_d) J^{rr}, \quad (38b)$$

$$\dot{\mathbf{j}}^x = \frac{\delta}{2r_d^2} \mathbf{j}^{rx}, \quad (38c)$$

$$\dot{\mathbf{j}}^{rr} = \frac{\delta}{2r_d^2} J^{rr}, \quad (38d)$$

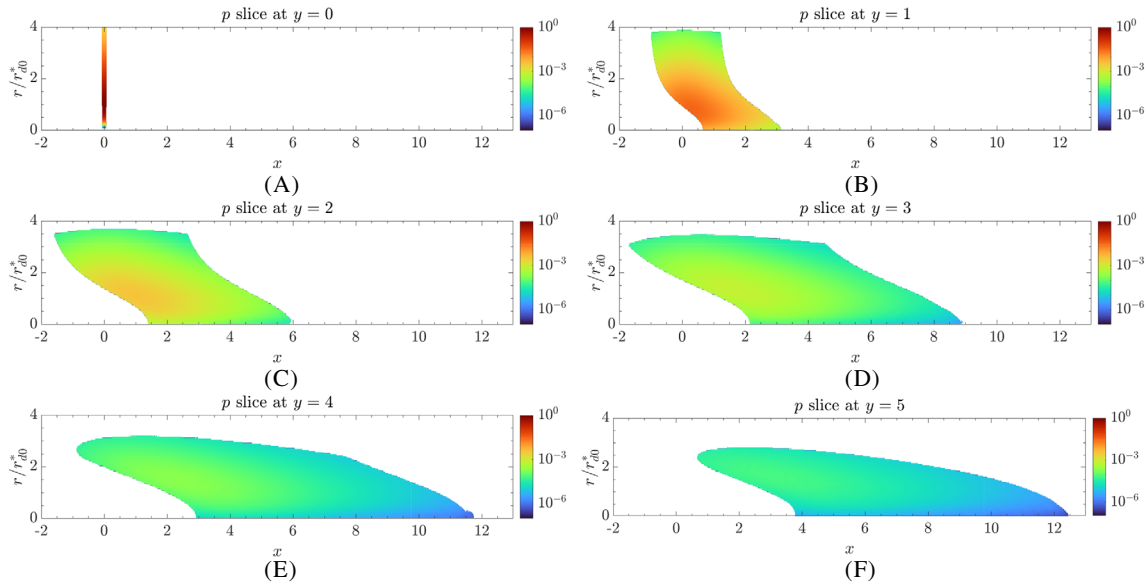


FIGURE 7 Slices of the probability density field p obtained using kernel regression for a two-dimensional fan spray injection in cross-flow at selected y locations: (A) $y = 0$; (B) $y = 1$; (C) $y = 2$; (D) $y = 3$; (E) $y = 4$; (F) $y = 5$. [Colour figure can be viewed at wileyonlinelibrary.com]

where the initial conditions are as in Equations (5), and those for $\mathbf{J}^{\mathbf{x}\mathbf{x}}$ being the same as in Equation (26). As in the monodisperse case, the systems (37) and (38) admit analytical solutions along trajectories, and these are given in Reference 14. Equation (2) is used to calculate the probability density along trajectories, with multidimensional kernel regression subsequently used to reconstruct the probability density field over the phase space $\xi = (x, y, r)$ through the use of Equations (15) and (16). In this configuration, the phase space probability density field p is reconstructed on a uniform Cartesian grid with spacing $\Delta x = \Delta y = 0.01$ and $\Delta r/r_{d0}^* = 0.04$, with the droplet evaporation rate being determined by $\delta = 1$. As in the monodisperse case, $\epsilon = 0.05$ is used to define the interval for droplet injection, and in this case at the start of the simulation 100 droplet sizes defined uniformly over the range $r_0 \in [0, 4]$ are injected at each of 101 locations uniformly spread over the interval $\mathbf{x}_0 \in [-\epsilon, \epsilon], 0$. This yields initial inter-droplet spacings of $\Delta x_{d0} = 0.001$ and $\Delta r_{d0}/r_{d0}^* = 0.04$ in position and size respectively, however due to the expanding nature of the spray in the spatial domain a larger initial physical space smoothing length of $h_{x0} = 5\Delta x_{d0}$ is required to achieve adequate coverage of the droplet field, whilst the default initial radial space smoothing length of $h_{r0} = \Delta r_{d0}$ is used.

For the case of a multidimensional polydisperse droplet flow this enables the collection of a wealth of information about the steady-state droplet distribution across the spatial dimensions and range of droplet sizes, and Figures 7 and 8 are used to illustrate this, with the variation in droplet size distribution in one spatial direction (x and y) depicted as slices in the other spatial direction (y and x respectively). This shows both how the spatial distribution of droplets varies as they travel away from the injection interval, and the range of droplet sizes which are present at a given location, but also importantly the distribution of droplet sizes within that range. It is seen that as droplets travel away from the injection interval in both the x and y directions, the range of droplet sizes decreases as the larger droplets evaporate, and the droplets become more dispersed in an asymmetrical profile which is reflective of the cross-flow configuration for the case. Of note is Figure 8A for the profile at $x = -1$, which illustrates the droplets that initially travel against the flow in the x direction from the injection interval. It is observed that only sufficiently large droplets appear at this location, which is a consequence of them having enough inertia to maintain their initial momentum, in contrast to smaller droplets which are more responsive to the flow and do not reach $x = -1$. This highlights the ability of the kernel regression procedure to provide detailed insight into the evolution of polydisperse droplet flows.

Of particular interest within polydisperse droplet flows is the number density n obtained using Equation (34a). This is depicted for this case in Figure 9A, and shows the full extent of the droplet spatial distribution across all droplet sizes, with the number density decreasing as droplets move away from the injection interval and evaporate. Profiles of n at selected y locations are shown in Figure 9B, and illustrate the rapid decrease in number density close to the injection interval. The distribution for the average droplet radius \bar{r} calculated using Equation (34b) is shown in Figure 9C, and it is observed

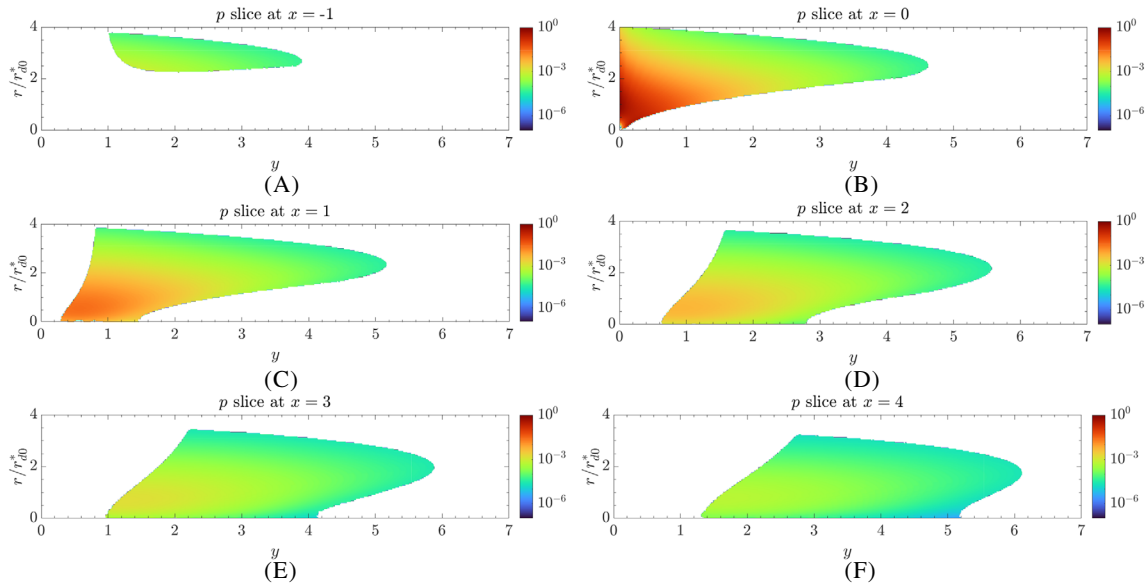


FIGURE 8 Slices of the probability density field p obtained using kernel regression for a two-dimensional fan spray injection in cross-flow at selected x locations: (A) $x = -1$; (B) $x = 0$; (C) $x = 1$; (D) $x = 2$; (E) $x = 3$; (F) $x = 4$. [Colour figure can be viewed at wileyonlinelibrary.com]

that this decreases across the flow with increasing x position. This is consistent with the observation from Figure 8A, in which only the largest droplets are able to travel against the carrier flow from the injection interval to reach the profile at $x = -1$. The profiles of \bar{r} in Figure 9D reflect the fact that at increasing y locations, the droplet field fills a wider part of the domain in the x direction, whilst both the peak value and rate of decrease in \bar{r} are reduced. The droplet radius variance distribution $\overline{r'r'}$ in Figure 9E is obtained from Equation (34c) and illustrates that the highest spread in droplet size is found in the centre region of the droplet field, with less variation found along the lower and upper edges of the spray, where in accordance with Figure 9C smaller and larger droplets are respectively more dominant. The profiles of $\overline{r'r'}$ in Figure 9F show that the highest droplet variance occurs at $y \approx 3$, whilst the spread of variance in the x direction increases with greater values of y position. Together this collection of information exemplifies how kernel regression is able to produce stable distributions for the averaged field variables from trajectory data provided by the gFLA, and highlights its suitability for application to more general gas-droplet flows.

4.2.3 | Flow around a cylinder: Steady-state case ($Re = 20$)

For investigation into the behaviour of polydisperse droplets in a more general flow, the example of two-dimensional gas-droplet flow around a cylinder that was described in Sections 4.1.3 and 4.1.4 is returned to. The carrier flow is governed by Equations (30) as previously, and now 41 different droplet sizes within the range $St \in [0.1, 10]$ are introduced at each of 101 injection points across the interval $y/R \in [-4, 4]$. The probability density is initialised in accordance with the distribution (33). In the following the focus is on the ability of kernel regression to reconstruct the droplet size distribution within the flow, and therefore the use of non-evaporating droplets is maintained. To achieve this, the droplet motion and Jacobian evolution are described as in the monodisperse case by Equations (31), and supplemented with the droplet evaporation rate $\varphi = 0$ and equations and associated initial conditions for the other blocks in Equations (4) and (5) respectively to yield $\mathbf{J}^{xr} = \mathbf{0}$, $\mathbf{J}^x = \mathbf{0}$ and $J^{rr} = 1$.

As with the previous polydisperse droplet examples, reconstruction of the droplet probability density field p follows the procedure outlined in Section 3.5, with the structured kernel defined as in Equation (15) and the bandwidth matrix \mathbf{H} given by Equation (16). The probability density field is accumulated onto a uniform Cartesian grid with a spatial spacing of $\Delta x/R = \Delta y/R = 0.04$ as in Section 4.1.3, and radial spacing $\Delta r/r_{d0}^* = 0.133$. The initial inter-droplet spacings in location and size are given by $\Delta x_{d0}/R = 0.08$ and $\Delta r_{d0}/r_{d0}^* = 0.1$ respectively, and the initial smoothing lengths set as $h_{x0} = \Delta x_{d0}$ and $h_{r0} = 0.667\Delta r_{d0}$ accordingly.

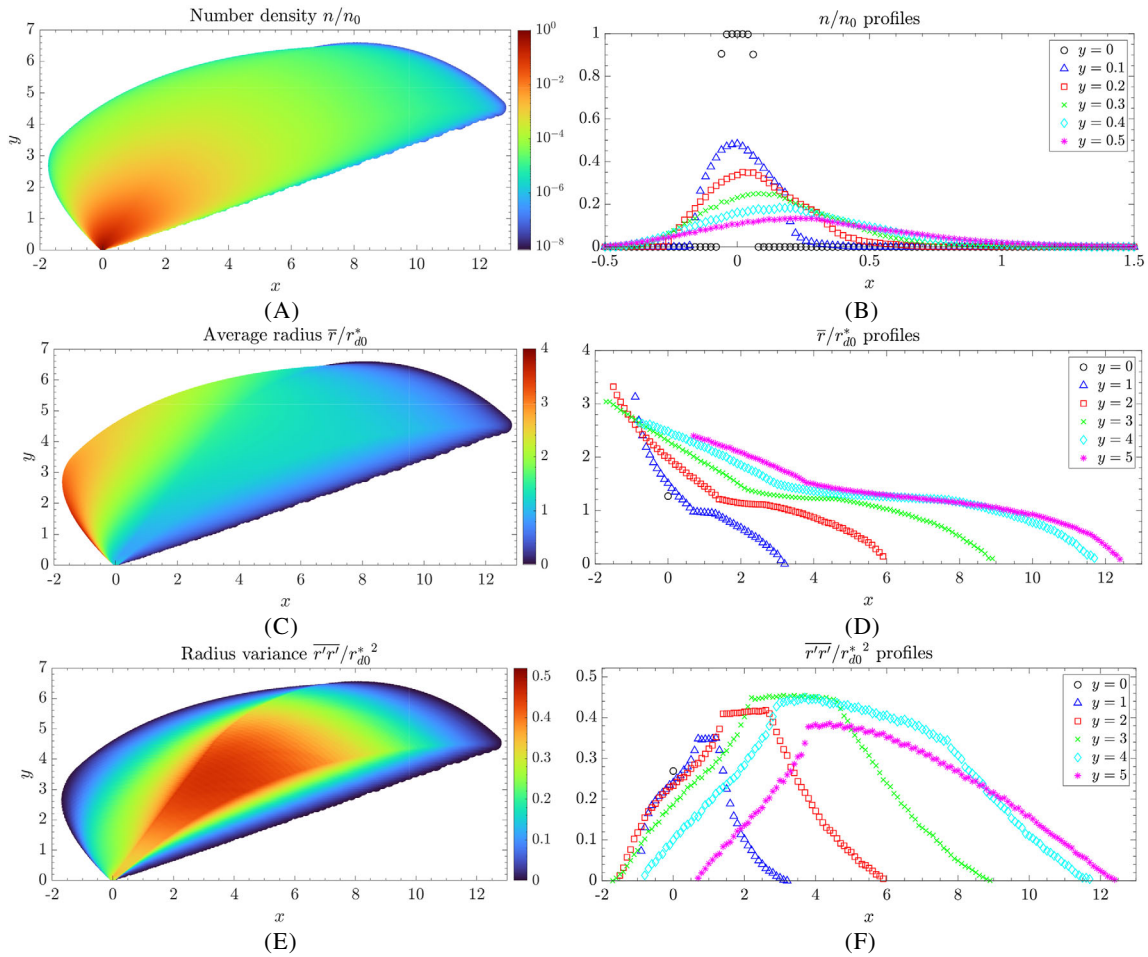


FIGURE 9 Reconstruction of the averaged field variables using kernel regression for a two-dimensional fan spray injection in cross-flow: (A) Number density n/n_0 ; (C) Average radius \bar{r}/r_{d0}^* ; (E) Radius variance $\overline{r^2}/r_{d0}^{*2}$; Profiles of the averaged field variables at selected y locations: (B) n/n_0 ; (D) \bar{r}/r_{d0}^* ; (F) $\overline{r^2}/r_{d0}^{*2}$. [Colour figure can be viewed at wileyonlinelibrary.com]

The kernel regression procedure again enables the detail of the droplet distribution across the spatial dimensions and range of droplet sizes to be obtained, and in this case the variation in droplet size distribution in the y direction is shown in Figure 10 as slices in the y direction to demonstrate this. Since the flow is steady-state, the distribution behaviour that is observed across the slices varies in a straightforward manner, with the width of the wake behind the cylinder decreasing as the distance from the cylinder at $x/R = 0$ is increased. In terms of the droplet size distribution, it can clearly be seen that at a given distance from the cylinder, the width of the wake is less for smaller droplets, agreeing with the behaviour previously observed in Figure 7 for monodisperse droplets at selected values of St . Indeed, it is possible to extract the spatial number density field at a given droplet radius from the reconstructed field p , and choosing the values of r corresponding to $St = 0.1, 1$, and 10 reproduces the distributions displayed in Figure 3A,C,E respectively. Furthermore, at any point in the domain, it can be seen in Figure 10 that there is a clear distribution across the range of droplet sizes, which is reflective of the initial size distribution specified by Equation (33).

The number density n is displayed in Figure 11A, and shows the behaviour of the droplet spatial distribution across all droplet sizes. It is seen that the wake region behind the cylinder with no droplets present corresponds to that obtained in Figure 3A for the smallest droplet size of $St = 0.1$. Additionally, the number density varies smoothly and in general gradually when moving from the centreline $y/R = 0$ to the edges of the domain. This can be seen in the profiles of n at selected y locations displayed in Figure 11B, and shows the variation in number density at different distances behind the cylinder. The average radius distribution \bar{r} shown in Figure 11C illustrates that larger droplets accumulate as the droplet field splits into two distinct regions when moving past the cylinder, whilst the region of the droplet field that is downstream of the cylinder wake only contains the smallest droplets. The profiles of \bar{r} in Figure 11D show that at these selected x

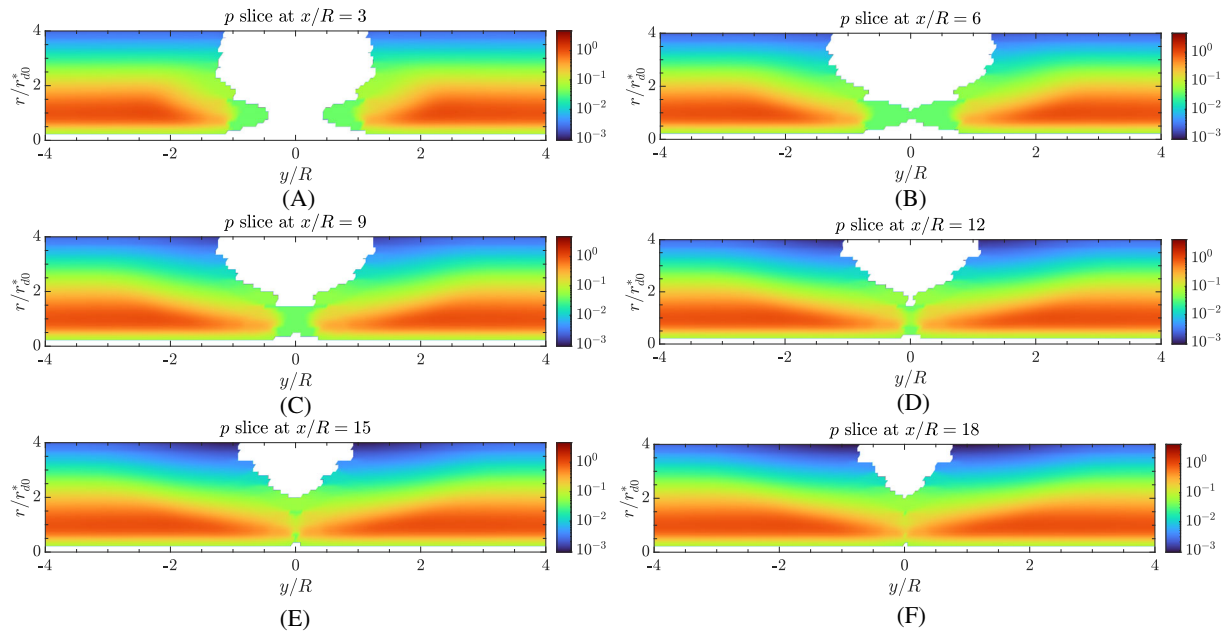


FIGURE 10 Slices of the probability density field p obtained using kernel regression for polydisperse steady-state flow around a cylinder at $Re = 20$ for selected x locations: (A) $x/R = 3$; (B) $x/R = 6$; (C) $x/R = 9$; (D) $x/R = 12$; (E) $x/R = 15$; (F) $x/R = 18$. [Colour figure can be viewed at wileyonlinelibrary.com]

locations the average droplet size does not vary much within the range of droplet sizes, and is uniformly lower towards the centreline $y/R = 0$. Both the distribution and profiles of variance in droplet radius $\overline{r'r'}$, as depicted in Figure 11E,F respectively, are qualitatively similar to the behaviour of \overline{r} , with the highest variance observed at the point where the droplet field parts to pass around the cylinder.

4.2.4 | Flow around a cylinder: Transient case ($Re = 100$)

For the transient regime of polydisperse droplet flow around a cylinder, $Re = 100$ is used as before in Section 4.1.4, and otherwise the configuration is identical to that in Section 4.2.3. In terms of the kernel regression procedure, the probability density field is accumulated onto a uniform Cartesian grid with a spatial spacing of $\Delta x/R = \Delta y/R = 0.04$ and radial spacing $\Delta r/r_{d0}^* = 0.133$ as before, with the initial smoothing lengths again specified as $h_{x0} = \Delta x_{d0}$ and $h_{r0} = 0.667\Delta r_{d0}$.

Again it is primarily the droplet spatial distribution which is of interest, and following the previous section the variation in droplet size distribution in the y direction is shown in Figure 12 as slices in the y direction at $t = 30$. Whilst the probability density distribution in Figure 12A displays only minor deviations from being symmetrical at a distance of $x/R = 3$ behind the cylinder, at increasing distances this symmetry breaks down as the vortex street becomes established. At a sufficient distance, the probability density distribution at a given slice then exhibits regions with either no droplets or a high probability density depending where in the periodic cycle of the flow the slice is located, for example at $x/R = 15$ as displayed in Figure 12E. Due to the transient nature of the flow, the effect of specifying the initial probability density using Equation (33) is less apparent than in the previous steady-state example, however the highest probability density values are still found among the smaller droplet sizes, with fewer larger droplets occurring. A wake is seen to persist for some distance behind the cylinder for only the largest droplets, but eventually ceases to exist once the periodic flow has become fully established. As with the steady-state case, the spatial number density field at a given droplet radius can be extracted from the reconstructed field p , whereby selecting the values of r corresponding to $St = 0.1, 1, \text{ and } 10$ retrieves the distributions in Figure 8A,C,E respectively.

Figure 13A displays the number density n at $t = 30$, exhibiting the cumulative effect of all droplet sizes on the droplet spatial distribution. It is observed that all droplet sizes are ejected from the vortices by virtue of their inertia, with clustering occurring along curves which are distinct for different droplet sizes. This is clearly an accumulation of the behaviour previously displayed in Figure 8A,C,E for different droplet sizes in the monodisperse case, and exemplifies the ability of

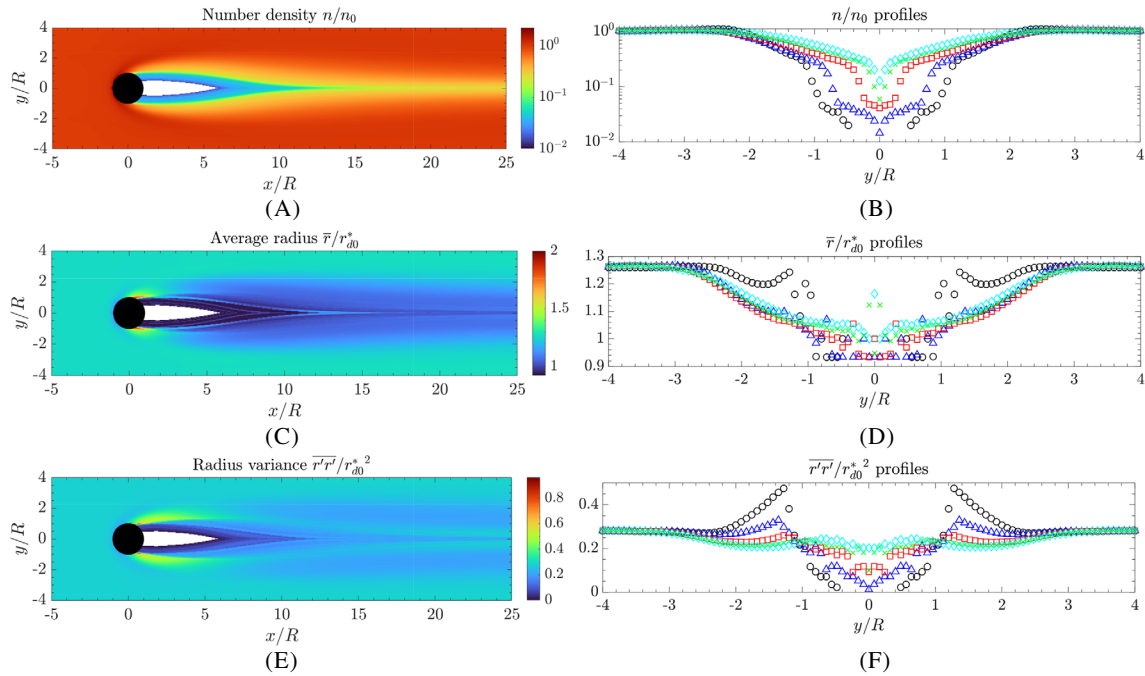


FIGURE 11 Reconstruction of the averaged field variables using kernel regression for steady-state polydisperse droplet flow around a cylinder at $Re = 20$: (A) Number density n/n_0 ; (C) Average radius \bar{r}/r_{d0}^* ; (E) Radius variance $\overline{r'r'}/r_{d0}^{*2}$; Profiles of the averaged field variables at selected x locations: [O] $x/R = 3$, [Δ] $x/R = 6$, [\square] $x/R = 9$, [\times] $x/R = 12$, [\diamond] $x/R = 15$; (B) n/n_0 ; (D) \bar{r}/r_{d0}^* ; (F) $\overline{r'r'}/r_{d0}^{*2}$. [Colour figure can be viewed at wileyonlinelibrary.com]

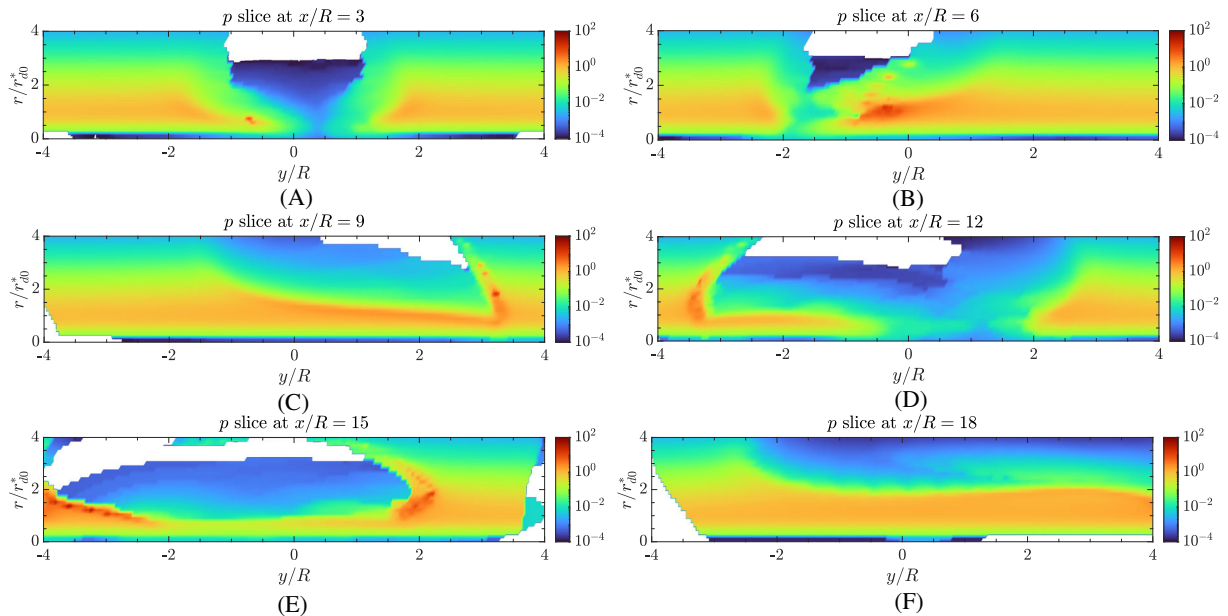


FIGURE 12 Slices of the probability density field p obtained using kernel regression for polydisperse transient flow around a cylinder at $Re = 100$ and time $t = 30$ for selected x locations: (A) $x/R = 3$; (B) $x/R = 6$; (C) $x/R = 9$; (D) $x/R = 12$; (E) $x/R = 15$; (F) $x/R = 18$. [Colour figure can be viewed at wileyonlinelibrary.com]

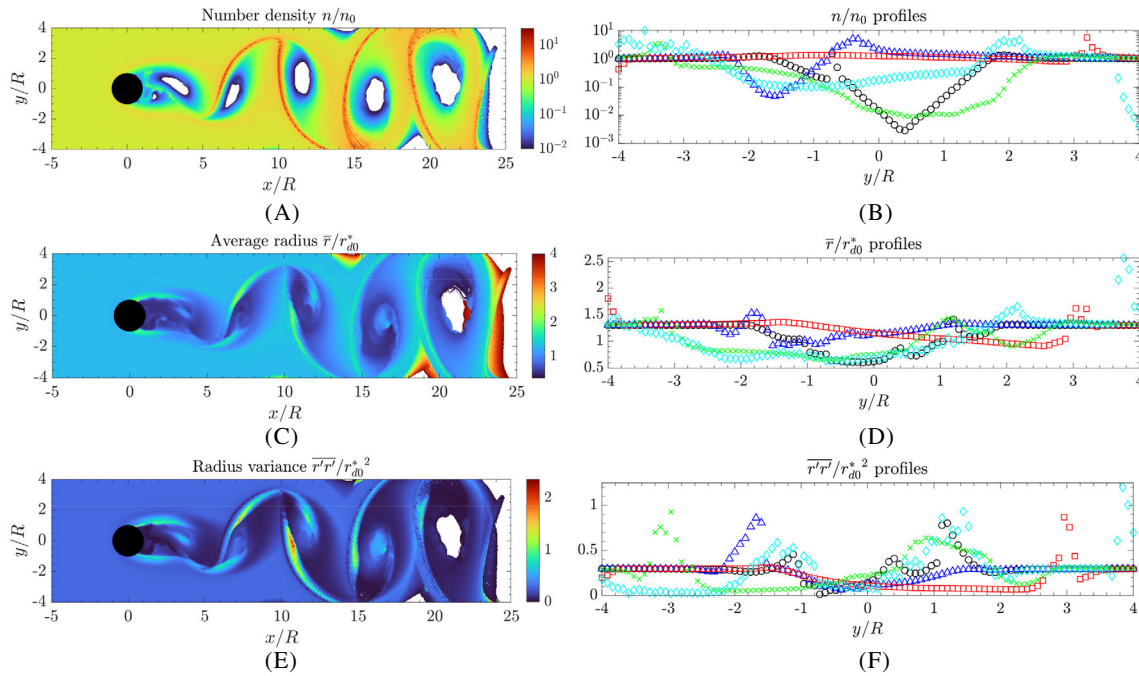


FIGURE 13 Reconstruction of the averaged field variables using kernel regression for transient polydisperse droplet flow around a cylinder at $Re = 100$ and time $t = 30$: (A) Number density n/n_0 ; (C) Average radius \bar{r}/r_{d0}^* ; (E) Radius variance $\overline{r'r'}/r_{d0}^{*2}$; Profiles of the averaged field variables at selected x locations: $[\circ]$ $x/R = 3$, $[\triangle]$ $x/R = 6$, $[\square]$ $x/R = 9$, $[\times]$ $x/R = 12$, $[\diamond]$ $x/R = 15$; (B) n/n_0 ; (D) \bar{r}/r_{d0}^* ; (F) $\overline{r'r'}/r_{d0}^{*2}$. [Colour figure can be viewed at wileyonlinelibrary.com]

kernel regression to account for this range of behaviour in the process of reconstructing the number density field. The variation in number density at different x profiles is displayed in Figure 13B, and it is seen that the behaviour of a given profile depends largely on its location with respect to the periodic cycle of the flow. For example, the number density at $x/R = 9$ is largely uniform for this cross-section of the droplet field, however the profile closer to the cylinder at $x/R = 3$ shows much higher variation as it passes through one of the vortices at this particular snapshot in time. The distribution and profiles of the average droplet radius \bar{r} at $t = 30$ are shown in Figure 13C,D respectively, and evidence that larger droplets accumulate along the distinct curves between the vortices, with only smaller droplets existing in the vicinity of the vortices. As with the steady-state case in Figure 11, the distribution and profiles of variance in droplet radius $r'r'$ at $t = 30$ and illustrated in Figure 13E,F respectively are qualitatively similar to the behaviour of \bar{r} in this case.

5 | CONCLUSIONS

This paper has presented a novel methodology for reconstructing the droplet number density field using the probability density along trajectories calculated by the generalised fully Lagrangian approach (gFLA; the fully Lagrangian approach taking into account the effect of droplet evaporation), and the technique of kernel regression to accumulate the contributions from individual droplets onto an Eulerian grid. This has been applied to a range of steady-state and transient flow configurations for both monodisperse and polydisperse droplets, and it has been demonstrated that the kernel regression procedure can reliably produce a smooth representation of the droplet probability density field and average field variables both in space and across the range of droplet sizes.

This approach has several advantages. The strongest benefit of kernel regression is making use of the meshfree nature of the method to define the kernel using the Jacobian tensor from the gFLA, so that individual droplet contributions are extrapolated to the surrounding Eulerian region in accordance with the local structure of the droplet number density field. This results in a consistent procedure that enables the retention of a high level of detail in the number density field from a relatively small sample of droplets, in contrast to a conventional cloud-in-cell approach which employs elementary averaging and therefore requires both more trajectories and a more frequent injection rate to reliably reproduce

the number density field. It is demonstrated that kernel regression yields a stable result with 10^3 times fewer droplet realisations than the CIC approach, and is therefore able to achieve the potential reduction in the computational cost required to obtain a smooth yet qualitatively accurate continuum representation for inertial droplet field variables. The informed use of data in such a way is becoming increasingly important, as the storage of datasets becomes impractical for highly resolved simulations due to the amount of information involved. Another consequence of being a meshfree method is that discontinuities in the droplet field are automatically dealt with by kernel regression, such as at the edge of the droplet region, and in this manner arbitrarily complex flows can be treated without requiring prior knowledge of the extent of the domain which is occupied by droplets.

Additionally, varying the size of the kernel in this way means that areas of sparse droplet number density can be effectively dealt with. Particle-based methods such as smoothed particle hydrodynamics require a minimum number of particles to satisfy the normalisation condition of the kernel, limiting their use to weakly compressible flows in which the number density of particles is roughly constant. In contrast, kernel regression satisfies the normalisation condition of the kernel regardless of how many particles are within the compact support, making the approach applicable to voids in which there are relatively few particles. Furthermore, specification of the kernel using the Jacobian provides a means of obtaining a smooth representation for regions containing few droplets, as the low number density associated with droplets in these regions results in the compact support of the kernel being proportionately large, and extending to cover the areas between sparsely distributed droplets.

Computationally, the procedure can be implemented by considering the contributions a given droplet makes to grid-points within its compact support, rather than having to identify all the droplets which contribute at a given gridpoint. When the reconstruction is performed upon a Cartesian grid this avoids the need for a droplet search algorithm, and is therefore relatively efficient. It is also conceptually trivial to extend the kernel regression procedure to account for the droplet size distribution by extending the dimensionality of the kernel, and reconstructing the probability density field on a higher-dimensional grid.

The representation of the number density field produced by kernel regression is, however, a locally constant estimator, and is unable to predictively extrapolate beyond the range of number densities which are sampled along trajectories. This is a shortcoming that is also inherent in linear interpolation, and means that in areas of rapid variation in number density, kernel regression will not return the exact value of the maximum or minimum local number density unless there is a droplet at that point. Furthermore, by its nature kernel regression is a statistical procedure, and is able to achieve a smooth representation of the droplet field only at the expense of some loss of absolute accuracy. This generally also occurs in regions of rapid variation in number density, with the smoothing characteristic of kernel regression meaning that for a given smoothing length and number of droplets only a certain level of variation can be captured. In practice this is most restrictive at a low droplet inertia, as has been observed in this work for low St , whilst the accuracy remains good for higher values of St . In general, kernel regression requires far fewer droplets than conventional direct trajectory methods to achieve a smooth Eulerian representation of the droplet phase at a reasonable level of accuracy, however increasing the number of droplet trajectories that are sampled will improve the accuracy if needed.

The stability of kernel regression and its versatility across a variety of flow configurations clearly demonstrate the efficacy of this procedure, and its suitability for upscaling and application to more general spray systems. In particular, the ability to accurately determine the complete droplet size distribution and its associated statistics at any spatial location within the droplet field make this methodology relevant to the simulation of many industrial and environmental systems, in which detailed knowledge of the droplet behaviour is paramount to the control and optimisation of such processes.

ACKNOWLEDGMENTS

The authors are grateful to the UKRI Future Leaders Fellowship (Grant MR/T043326/1) for their financial support, and would also like to acknowledge Prof A. N. Osipov for his helpful suggestions.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

C. P. Stafford  <https://orcid.org/0000-0002-5312-7422>

REFERENCES

1. Costa M, Piazzullo D. Biofuel powering of internal combustion engines: production routes, effect on performance and CFD modeling of combustion. *Front Mech Eng.* 2018;4:9. doi:10.3389/fmech.2018.00009
2. Bourouiba L, Dehandschoewercker E, Bush JWM. Violent expiratory events: on coughing and sneezing. *J Fluid Mech.* 2014;745:537-563. doi:10.1017/jfm.2014.88
3. Sidin RSR, Ijzermans RHA, Reeks MW. A Lagrangian approach to droplet condensation in atmospheric clouds. *Phys Fluids.* 2009;21(10):106603. doi:10.1063/1.3244646
4. Laurent F, Vié A, Chalons C, Fox RO, Massot DM. A hierarchy of Eulerian models for trajectory crossing in particle-laden turbulent flows over a wide range of Stokes numbers. Tech. rep. 2012.
5. Healy D, Young J. Full Lagrangian methods for calculating particle concentration fields in dilute gas-particle flows. *Proc R Soc A: Math Phys Eng Sci.* 2005;461(2059):2197-2225. doi:10.1098/rspa.2004.1413
6. Yudine M. Physical considerations on heavy-particle diffusion. *Adv Geophys.* 1959;6:185-191.
7. Chaisemartin DS, Fréret L, Kah D, et al. Eulerian models for turbulent spray combustion with polydispersity and droplet crossing. *Comptes Rendus Mécanique.* 2009;337(6-7):438-448. doi:10.1016/j.crme.2009.06.016
8. Marble FE. Dynamics of dusty gases. *Annu Rev Fluid Mech.* 1970;2(1):397-446. doi:10.1146/annurev.fl.02.010170.002145
9. Osipov A. Lagrangian modelling of dust admixture in gas flows. *Astrophys Space Sci.* 2000;274(1/2):377-386. doi:10.1023/A:1026557603451
10. Ijzermans RHA, Meneguz E, Reeks MW. Segregation of particles in incompressible random flows: singularities, intermittency and random uncorrelated motion. *J Fluid Mech.* 2010;653:99-136. doi:10.1017/S0022112010000170
11. Zaripov TS, Gilfanov AK, Begg SM, Rybdylova O, Sazhin SS, Heikal MR. The fully Lagrangian approach to the analysis of particle/droplet dynamics: implementation into ANSYS fluent and application to gasoline sprays. *Atom Sprays.* 2017;27(6):493-510. doi:10.1615/AtomizSpr.2017019354
12. Papoutsakis A, Rybdylova OD, Zaripov TS, Danaïla L, Osipov AN, Sazhin SS. Modelling of the evolution of a droplet cloud in a turbulent flow. *Int J Multiphase Flow.* 2018;104:233-257. doi:10.1016/j.ijmultiphaseflow.2018.02.014
13. Rybdylova O, Zaripov T, Li Y. Two-way coupling of the fully Lagrangian approach with OpenFOAM for spray modelling. In: Zaleski S, ed. *Proc ILASS 2019.* ILASS (Institute for Liquid Atomization and Spray Systems); 2019:244656.
14. Li Y, Rybdylova O. Application of the generalised fully Lagrangian approach to simulating polydisperse gas-droplet flows. *Int J Multiphase Flow.* 2021;142:103716. doi:10.1016/j.ijmultiphaseflow.2021.103716
15. Mishchenko AV, Godenko EA, Izmodenov VV. Lagrangian fluid approach for the modeling of peculiarities of the interstellar dust distribution in the astrospheres/heliosphere. *Month Notice R Astronom Soc.* 2019;491(2):2808-2821. doi:10.1093/mnras/stz3193
16. Govindarajan B, Leishman JG, Gumerov NA. Evaluation of particle clustering algorithms in the prediction of brownout dust clouds. *Ann Forum Proc AHS Int.* 2011;1:325-347.
17. Osipov AN. Investigation of regions of unbounded growth of the particle concentration in disperse flows. *Fluid Dyn.* 1984;19(3):378-385. doi:10.1007/BF01093900
18. Lebedeva NA, Osipov AN, Sazhin SS. A combined fully Lagrangian approach to mesh-free modelling of transient two-phase flows. *Atom Sprays.* 2013;23(1):47-69. doi:10.1615/AtomizSpr.2013006269
19. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer; 2009.
20. Laux S. On particle-mesh coupling in Monte Carlo semiconductor device simulation. *IEEE Trans Comput-Aided Des Integrat Circuits Syst.* 1996;15(10):1266-1277. doi:10.1109/43.541446
21. Price DJ. Smoothed particle hydrodynamics and magnetohydrodynamics. *J Comput Phys.* 2012;231(3):759-794. doi:10.1016/j.jcp.2010.12.011
22. Wang ZB, Chen R, Wang H, Liao Q, Zhu X, Li SZ. An overview of smoothed particle hydrodynamics for simulating multiphase flow. *App Math Model.* 2016;40(23-24):9625-9655. doi:10.1016/j.apm.2016.06.030
23. Monchaux R, Bourgoïn M, Cartellier A. Preferential concentration of heavy particles: a Voronoi analysis. *Phys Fluids.* 2010;22(10):103304. doi:10.1063/1.3489987
24. Serrano M, Español P, Zúñiga I. Voronoi fluid particle model for Euler equations. *J Stat Phys.* 2005;121(1-2):133-147. doi:10.1007/s10955-005-8414-y
25. Heß S, Springel V. Particle hydrodynamics with tessellation techniques. *Month Notice R Astronom Soc.* 2010;406(4):2289-2311. doi:10.1111/j.1365-2966.2010.16892.x
26. Schaap WE, van der Weygaert R. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. *Astron Astrophys.* 2000;363(3):29-32.
27. Pelupessy FI, Schaap WE, van de Weygaert R. Density estimators in particle hydrodynamics. *Astron Astrophys.* 2003;403(2):389-398. doi:10.1051/0004-6361:20030314
28. Huang R, Sun S. Kernel regression with sparse metric learning. *J Intell Fuzzy Syst.* 2013;24(4):775-787. doi:10.3233/IFS-2012-0597
29. Weinberger KQ, Tesauro G. Metric learning for kernel regression. In: Meila M, Shen X, eds. *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics. 2 of Proceedings of Machine Learning Research.* PMLR; 2007:612-619.
30. Monchaux R, Bourgoïn M, Cartellier A. Analyzing preferential concentration and clustering of inertial particles in turbulence. *Int J Multiphase Flow.* 2012;40:1-18. doi:10.1016/j.ijmultiphaseflow.2011.12.001
31. Papoutsakis A, Gavaises M. A model for the investigation of the second-order structure of caustic formations in dispersed flows. *J Fluid Mech.* 2020;892:A4. doi:10.1017/jfm.2020.176

32. Fraga Filho CAD. *Smoothed Particle Hydrodynamics*. Springer; 2019.
33. Maxey MR. The gravitational settling of aerosol particles in homogeneous turbulence and random flow fields. *J Fluid Mech.* 1987;174:441-465. doi:10.1017/S0022112087000193
34. Kooij S, Sijs R, Denn MM, Villermaux E, Bonn D. What determines the drop size in sprays? *Phys Rev X.* 2018;8(3):031019. doi:10.1103/PhysRevX.8.031019

How to cite this article: Stafford CP, Rybdylova O. Robust interpolation for dispersed gas-droplet flows using statistical learning with the fully Lagrangian approach. *Int J Numer Meth Fluids.* 2023;1-35. doi: 10.1002/fld.5225

APPENDIX A. VERIFICATION OF THE LAGRANGIAN DROPLET PHASE CONTINUITY EQUATION FOR EVAPORATING DROPLETS

Consider droplets with position $\mathbf{x}_d(t)$, velocity $\mathbf{v}_d(t)$ and radius $r_d(t)$ for which the evolution of trajectories and evaporation are governed by Equations (1). The general continuum representation of the droplet phase can then be described in terms of the one-particle distribution function $w(\mathbf{x}, \mathbf{v}, r, t)$ by the collisionless kinetic equation⁹

$$\frac{\partial w}{\partial t} = -\frac{\partial}{\partial \mathbf{x}} \cdot [\mathbf{v}w] - \frac{\partial}{\partial \mathbf{v}} \cdot [\mathbf{f}w] - \frac{\partial}{\partial r} [\varphi w]. \quad (\text{A1})$$

To obtain the droplet phase continuity equation in the subset of phase space given by $\xi = (\mathbf{x}, r)$, the droplet mean field variables defined as the velocity-averaged moments of $w(\mathbf{x}, \mathbf{v}, r, t)$ are considered. Specifically, the probability density $p(\mathbf{x}, r, t)$ and average velocity $\bar{\mathbf{v}}(\mathbf{x}, r, t)$ are defined as

$$p(\mathbf{x}, r, t) = \int_{\mathbf{v}} w(\mathbf{x}, \mathbf{v}, r, t) d\mathbf{v}, \quad (\text{A2a})$$

$$\bar{\mathbf{v}}(\mathbf{x}, r, t) = \frac{1}{p(\mathbf{x}, r, t)} \int_{\mathbf{v}} \mathbf{v}w(\mathbf{x}, \mathbf{v}, r, t) d\mathbf{v}. \quad (\text{A2b})$$

Then integrating the kinetic Equation (A1) over \mathbf{v} yields

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial \mathbf{x}} \cdot [\bar{\mathbf{v}}p] - \frac{\partial}{\partial r} [\bar{\varphi}p], \quad (\text{A3})$$

where the velocity derivative term in Equation (A1) vanishes as it is assumed that $w \rightarrow 0$ as $\mathbf{v} \rightarrow \pm\infty$, and the velocity-averaged evaporation rate $\bar{\varphi}(\mathbf{x}, r, t)$ is defined as

$$\bar{\varphi}(\mathbf{x}, r, t) = \frac{1}{p(\mathbf{x}, r, t)} \int_{\mathbf{v}} \varphi(\mathbf{x}, \mathbf{v}, r, t) w(\mathbf{x}, \mathbf{v}, r, t) d\mathbf{v}. \quad (\text{A4})$$

Equation (A3) is the Eulerian form of the droplet phase continuity equation that governs the transport of $p(\mathbf{x}, r, t)$. To proceed to the Lagrangian form of Equation (A3), it is appropriate to consider the phase space trajectories $\mathbf{z}_d = (\mathbf{x}_d, r_d)$ in ξ that evolve according to the average velocity $\bar{\mathbf{v}}$ and velocity-averaged evaporation rate $\bar{\varphi}$

$$\dot{\mathbf{x}}_d = \bar{\mathbf{v}}(\mathbf{x}_d, r_d, t), \quad \mathbf{x}_d(t_0) = \mathbf{x}_0, \quad (\text{A5a})$$

$$\dot{r}_d = \bar{\varphi}(\mathbf{x}_d, r_d, t), \quad r_d(t_0) = r_0. \quad (\text{A5b})$$

The corresponding system form of Equations (A5) is then given by

$$\dot{\mathbf{z}}_d = \bar{\mathbf{S}}(\mathbf{z}_d, t), \quad \mathbf{z}_d(t_0) = \xi_0, \quad (\text{A6})$$

where $\bar{\mathbf{S}}(\xi, t) = (\bar{\mathbf{v}}(\mathbf{x}, r, t), \bar{\varphi}(\mathbf{x}, r, t))$. Thus Equation (A3) can be written in the phase space ξ as

$$\frac{\partial}{\partial t} p(\xi, t) = -\frac{\partial}{\partial \xi} \cdot \left[\bar{\mathbf{S}}(\xi, t) p(\xi, t) \right]. \quad (\text{A7})$$

The appropriate form of the Lagrangian derivative corresponding to motion along the phase space trajectories \mathbf{z}_d is given by

$$\frac{D}{Dt} := \frac{\partial}{\partial t} + \bar{\mathbf{S}}(\xi, t) \cdot \frac{\partial}{\partial \xi}, \quad (\text{A8})$$

whereupon Equation (A7) can now be interpreted along the phase space trajectory \mathbf{z}_d and written as

$$\frac{D}{Dt} p(\mathbf{z}_d, t) = -p(\mathbf{z}_d, t) \frac{\partial}{\partial \xi} \cdot \bar{\mathbf{S}}(\mathbf{z}_d, t). \quad (\text{A9})$$

Equation (A9) can be formally solved along the phase space trajectory \mathbf{z}_d , and along with the initial condition $p(\mathbf{z}_d(t_0), t_0) = p(\xi_0, t_0)$ yields the solution

$$p(\mathbf{z}_d, t) = p(\xi_0, t_0) \exp \left[-\int_{t_0}^t \frac{\partial}{\partial \xi} \cdot \bar{\mathbf{S}}(\mathbf{z}'_d, t') dt' \right], \quad (\text{A10})$$

where $\mathbf{z}'_d = \mathbf{z}_d(t')$, and it is assumed that $p(\xi_0, t_0) > 0$ in accordance with the probability density $p(\xi, t)$ being interpreted as a strictly positive quantity. Re-introducing the interpretations $\xi = (\mathbf{x}, r)$, $\bar{\mathbf{S}} = (\bar{\mathbf{v}}, \bar{\varphi})$ leads to the solution in the familiar physical variables

$$p(\mathbf{x}_d, r_d, t) = p(\mathbf{x}_0, r_0, t_0) \exp \left[-\int_{t_0}^t \frac{\partial}{\partial \mathbf{x}} \cdot \bar{\mathbf{v}}(\mathbf{x}'_d, r'_d, t') + \frac{\partial}{\partial r} \cdot \bar{\varphi}(\mathbf{x}'_d, r'_d, t') dt' \right], \quad (\text{A11})$$

where $\mathbf{x}'_d = \mathbf{x}_d(t')$ and $r'_d = r_d(t')$. Equation (A11) is the solution for the probability density $p(\mathbf{x}_d, r_d, t)$ as it evolves along trajectories. However, this assumes that droplet trajectories are governed by the velocity averaged Equations (A5) through the velocity field $\bar{\mathbf{v}}(\mathbf{x}, r, t)$ and evaporation rate $\bar{\varphi}(\mathbf{x}, r, t)$. Now consider the Jacobian tensor defined by

$$\mathbf{J}(\xi_0, t) = \frac{\partial \mathbf{z}_d(t)}{\partial \xi_0}. \quad (\text{A12})$$

Then taking the partial derivative $\partial/\partial \xi_0$ of the phase space trajectory \mathbf{z}_d described by Equation (A6), the evolution of $\mathbf{J}(\xi_0, t)$ is determined by

$$\dot{\mathbf{J}}(\xi_0, t) = \frac{\partial \bar{\mathbf{S}}}{\partial \xi}(\mathbf{z}_d, t) \cdot \mathbf{J}(\xi_0, t). \quad (\text{A13})$$

Consider also the Jacobian determinant $\det(\mathbf{J}(\xi_0, t))$. Then Jacobi's formula from linear algebra states that

$$\frac{d}{dt} \det(\mathbf{J}(\xi_0, t)) = \det(\mathbf{J}(\xi_0, t)) \operatorname{tr} \left(\mathbf{J}^{-1}(\xi_0, t) \cdot \dot{\mathbf{J}}(\xi_0, t) \right). \quad (\text{A14})$$

Utilising the evolution Equation (A13) for $\mathbf{J}(\xi_0, t)$ therefore leads to

$$\begin{aligned} & \frac{d}{dt} \det(\mathbf{J}(\xi_0, t)) \\ &= \operatorname{tr} \left(\mathbf{J}^{-1}(\xi_0, t) \cdot \frac{\partial \bar{\mathbf{S}}}{\partial \xi}(\mathbf{z}_d, t) \cdot \mathbf{J}(\xi_0, t) \right) \det(\mathbf{J}(\xi_0, t)), \\ &= \operatorname{tr} \left(\mathbf{J}(\xi_0, t) \cdot \mathbf{J}^{-1}(\xi_0, t) \cdot \frac{\partial \bar{\mathbf{S}}}{\partial \xi}(\mathbf{z}_d, t) \right) \det(\mathbf{J}(\xi_0, t)), \\ &= \operatorname{tr} \left(\frac{\partial \bar{\mathbf{S}}}{\partial \xi}(\mathbf{z}_d, t) \right) \det(\mathbf{J}(\xi_0, t)), \\ &= \frac{\partial}{\partial \xi} \cdot \bar{\mathbf{S}}(\mathbf{z}_d, t) \det(\mathbf{J}(\xi_0, t)), \end{aligned} \quad (\text{A15})$$

where the third line follows since the trace operator is invariant under cyclic permutations of products of arguments, and the final line is due to the trace of the gradient being equivalent to the divergence. Equation (A15) can be formally solved along the phase space trajectory \mathbf{z}_d , which presents the general solution

$$|\det(\mathbf{J}(\xi_0, t))| = |C| \exp \left[\int_{t_0}^t \frac{\partial}{\partial \xi} \cdot \bar{\mathbf{S}}(\mathbf{z}'_d, t') dt' \right], \quad (\text{A16})$$

in which no restriction is made upon the sign of the arbitrary constant C . Since C is determined from the initial condition on $\det(\mathbf{J}(\xi_0, t))$, this allows for the case where $\det(\mathbf{J}(\xi_0, t))$ is negative to be accounted for in the description provided by Equation (A16). Applying the general initial condition of $\det(\mathbf{J}(\xi, t_0)) = 1$ (which follows from the definition of \mathbf{J} in Equation (A12) and the initial condition upon \mathbf{z}_d in Equation (A6) at time t_0) then yields

$$|\det(\mathbf{J}(\xi_0, t))| = \exp \left[\int_{t_0}^t \frac{\partial}{\partial \xi} \cdot \bar{\mathbf{S}}(\mathbf{z}'_d, t') dt' \right]. \quad (\text{A17})$$

Re-introducing the interpretations $\xi = (\mathbf{x}, r)$, $\bar{\mathbf{S}} = (\bar{\mathbf{v}}, \bar{\varphi})$ leads to the equivalent form in physical variables

$$|\det(\mathbf{J}(\mathbf{x}_0, r_0, t))| = \exp \left[\int_{t_0}^t \frac{\partial}{\partial \mathbf{x}} \cdot \bar{\mathbf{v}}(\mathbf{x}'_d, r'_d, t') + \frac{\partial}{\partial r} \cdot \bar{\varphi}(\mathbf{x}'_d, r'_d, t') dt' \right]. \quad (\text{A18})$$

Comparing the solutions for $p(\mathbf{x}_d, r_d, t)$ and $\det(\mathbf{J}(\mathbf{x}_0, r_0, t))$ in Equations (A11) and (A18) respectively, it is seen that conservation of mass along trajectories in (\mathbf{x}, r) space can be expressed as

$$\frac{p(\mathbf{x}_d, r_d, t)}{p(\mathbf{x}_0, r_0, t_0)} = \frac{1}{|\det(\mathbf{J}(\mathbf{x}_0, r_0, t))|}, \quad (\text{A19})$$

from which the Lagrangian form of the continuity equation expressed in Equation (2) immediately follows.

Note that since this formulation hinges upon use of droplet trajectories described by the velocity field $\bar{\mathbf{v}}(\mathbf{x}, r, t)$, the resultant continuum description inherently assumes the invertibility of $\mathbf{J}(\mathbf{x}_0, r_0, t)$ at a given point in time, that is $\det(\mathbf{J}(\mathbf{x}_0, r_0, t)) \neq 0$. Physically this corresponds to the single-valuedness of droplet velocities at a given point \mathbf{x} , which in reality is only representative of fluid tracer droplets in the limit $St \rightarrow 0$. Consequently, this description is unable to qualitatively account for the crossing trajectories effect of sufficiently inertial droplets at finite St , for which occurrences of $\det(\mathbf{J}(\mathbf{x}_0, r_0, t)) = 0$ are observed along the envelope of folds in the droplet velocity field, and the probability density $p(\mathbf{x}_d, r_d, t)$ becomes singular. Similarly, $\det(\mathbf{J}(\mathbf{x}_0, r_0, t))$ becomes negative once the associated trajectory crosses the envelope of a fold. However, these phenomena do not violate the physical correctness of Equation (2) as a conservation law, and indeed do not preclude using values of $\det(\mathbf{J}(\mathbf{x}_0, r_0, t))$ calculated from simulation data of inertial droplet trajectories. This can be understood by considering the geometric meaning of $\det(\mathbf{J}(\mathbf{x}_0, r_0, t))$ as the volume scaling factor of the linear transformation described by the Jacobian $\mathbf{J}(\mathbf{x}_0, r_0, t)$, with the associated sign simply showing whether the transformation preserves or reverses orientation. In the context of calculating $p(\mathbf{x}_d, r_d, t)$ the orientation of the elemental volume is not relevant, and therefore using $|\det(\mathbf{J}(\mathbf{x}_0, r_0, t))|$ to evaluate the probability density by means of Equation (2) remains physically consistent in the case that $\det(\mathbf{J}(\mathbf{x}_0, r_0, t)) < 0$.

APPENDIX B. DEMONSTRATION OF THE KERNEL DOMAIN SIZE VARYING PROPORTIONATELY TO THE JACOBIAN DETERMINANT

In order to account for the high degree of variation in the number density field for inertial droplets, it is desired that reconstruction using a kernel-based method is able to vary the size of the kernel domain of support according to the local droplet number density. Within the FLA methodology, the physical interpretation of the Jacobian $\mathbf{J}^{\mathbf{x}\mathbf{x}}$ as the Eulerian-Lagrangian transformation provides the necessary information regarding the extent of the local Eulerian droplet field over which the influence of an individual droplet acts. It is therefore necessary to specify the kernel such that its domain of support is proportionate to $|\det(\mathbf{J}^{\mathbf{x}\mathbf{x}})|$.

For the specific choice of the multivariate Gaussian kernel in Equation (11), it can be shown that the domain size for a compact support is proportional to $\sqrt{|\det(\mathbf{H})|}$. Then for the positive semi-definite specification of $\mathbf{H} = h_0^2 \mathbf{J}^{\mathbf{x}\mathbf{x}} \cdot \mathbf{J}^{\mathbf{x}\mathbf{x}\top}$ in

Equation (14), we have that

$$\sqrt{\det(\mathbf{H})} = \sqrt{\det(h_0^2 \mathbf{J}^{\mathbf{x}\mathbf{x}} \cdot \mathbf{J}^{\mathbf{x}\mathbf{x}\top})} = \sqrt{h_0^{2d} \det(\mathbf{J}^{\mathbf{x}\mathbf{x}}) \det(\mathbf{J}^{\mathbf{x}\mathbf{x}\top})} = h_0^d \sqrt{\det(\mathbf{J}^{\mathbf{x}\mathbf{x}}) \det(\mathbf{J}^{\mathbf{x}\mathbf{x}})} = h_0^d |\det(\mathbf{J}^{\mathbf{x}\mathbf{x}})|. \quad (\text{B1})$$

Thus the domain of support for the kernel is proportionate to $|\det(\mathbf{J}^{\mathbf{x}\mathbf{x}})|$ as required, meaning that the kernel is able to adaptively scale in accordance with the local Eulerian-Lagrangian transformation so that the spatial structures of the droplet field, which accompany the variation in number density, are captured.