Optimization and Engineering
https://doi.org/10.1007/s11081-024-09907-2

RESEARCH ARTICLE

Multi-objective Bayesian shape optimization of an industrial hydrodynamic separator using unsteady Eulerian-Lagrangian simulations

A. P. Roberts[1] · **A. A. M. Rahat**[2] · **D. S. Jarman**[3] · **J. E. Fieldsend**[1] · **G. R. Tabor**[1]

Received: 8 March 2024 / Revised: 26 May 2024 / Accepted: 10 July 2024

Abstract

The shape of a hydrodynamic particle separator has been optimized using a parallelized and robust formulation of Bayesian optimization, with data from an unsteady Eulerian flow field coupled with Lagrangian particle tracking. The uncertainty due to the mesh, initial conditions, and stochastic dispersion in the Eulerian-Lagrangian simulations was minimized and quantified. This was then translated across to the error term in the Gaussian process model and the minimum probability of improvement infill criterion. An existing parallelization strategy was modified for the infill criterion and customized to prefer exploitation in the decision space. In addition, a new strategy was developed for hidden constraints using Voronoi penalization. In the approximate Pareto Front, an absolute improvement over the base design of 14% in the underflow collection efficiency and 10% in the total collection efficiency was achieved, which resulted in the filing of a patent.* The corresponding designs were attributed to the effective distribution of residence time between the trays via the removal of a vertical plume. The plume also reduced both efficiencies by creating a flow path in a direction that acted against effective settling. The concave down and offset tray shapes demonstrated the value of Bayesian optimization in producing useful and non-intuitive designs.

Keywords Industrial hydrodynamic separator · Multi-objective Bayesian shape optimization · Multi-surrogate parallelization · Voronoi failure penalization · Eulerian-Lagrangian one-way coupling · uRANS · k-ω SST

*Roberts A.P. et al. (2024). *A separator for separating solids from a solid–liquid mixture*, UK patent application number 2400481.4, European patent application number 24151725.9.

Extended author information available on the last page of the article

Published online: 17 August 2024

List of symbols

α, β, ϵ	Constants in monotonic beta cumulative distribution function
α_p	Volume fraction of particles
α_{mpoi}	Minimum probability of improvement scalarization
$\alpha_\omega, \beta_\omega, \sigma_\omega, \sigma_k$	Blended coefficients in k-ω SST model
β^*, a_1, c_1	Fixed coefficients in k-ω SST model
$\mathbf{x}, \mathbf{y}$	Decision vector, objective vector
$\mathbf{x}^*$	Next best decision
X_f, X_s	Set of failed decisions, set of successful decisions
$\mathbf{\Omega}$	Fluid rotation
δ, κ, q	Terms in parallelization strategy
$\eta_{capture}$	Capture collection efficiency after 10 residence times
η_{under}	Underflow collection efficiency after 10 residence times
η_{total}	Total collection efficiency after 10 residence times
ϕ_{pi}, Φ_p	Penalization for objective i, penalization with minimum radius
Ψ	Set of decisions to penalize
ξ, ψ	Gaussian random draw, uniform random draw
σ_{under}	Standard deviation in underflow collection efficiency
σ_{total}	Standard deviation in total collection efficiency
A, a	Tray major radius at top, at bottom (or function)
B, b	Tray minor radius at top, at bottom (or function)
c	Center location function
D	Tray body diameter
G	Rosin-Rammler cumulative distribution function
g	Monotonic beta cumulative distribution function
H	Component of direction vector
i, j	Objective number, decision number
N	Number of objectives
m_{in}	Mass introduced at 0 residence times
m_{over}	Mass collected at overflow after 10 residence times
m_{under}	Mass collected at underflow after 10 residence times
$m_{surfaces}$	Mass collected on surfaces after 10 residence times
m_{vessel}	Mass collected in vessel after 10 residence times
p	Kinematic pressure
s	Tray number
Y_c, y_c	Tray center offset at top, tray center offset at bottom
z_{max}, z_{min}	Location of top of tray, location of bottom of tray
$\mathbf{x}_{pareto}$	Pareto set decision
$\mathbf{x}_{penalize}$	Penalized decision
$\delta_{benching}$	Benching gap
$\delta_{inter-tray}$	Inter-tray gap
Δt	Timestep
$\overline{\mathbf{u}}$	Steady component of fluid velocity
$\mathbf{F}_D, \mathbf{F}_L, \mathbf{F}_G$	Drag force, lift force, gravity force
$\mathbf{g}$	Acceleration due to gravity

$\mathbf{u}_{rel}, \mathbf{u}_{fluct}$	Relative velocity, fluctuating velocity
$\mathbf{u}, \mathbf{u}_p$	Fluid velocity, particle velocity
$\mathcal{P}^*, \mathcal{F}^*$	Approximate Pareto set, approximate Pareto front
χ	Input decision space
$X_{evaluated}$	Evaluated decisions
μ_i	Mean of objective i
ν, ν_t, ν_{eff}	Kinematic viscosity, turbulent viscosity, effective viscosity
ω_{log}	Specific turbulence dissipation rate in log-layer
$\overline{d}_p, n$	Coefficients in Rosin-Rammler CDF
ρ, ρ_p	Fluid density, particle density
σ_i^2	Variance in objective i
θ	Rotation angle
ε_i	Homoscedastic error variance for objective i
$\widetilde{\mathcal{L}}$	Approximation to Lipschitz constant
$\widetilde{\mathcal{M}}$	Approximation to best value seen so far
C_D, C_L	Particle drag coefficient, particle lift coefficient
d_p	Particle diameter
F_1	Blending function in k-ω SST model
k	Turbulent kinetic energy
m_i	Replacement term in the MPoI for objective i
P	Probability
P_k	Turbulent kinetic energy production
Re_D	HeadCell Reynolds number
T	Residence time
t	Time
x, y, z	Cartesian coordinate
U_{nozzle}	Nozzle velocity
m_p	Particle mass
$\mathcal{G}, \mathcal{U}$	Gaussian distribution, uniform distribution
y_i	Objective number i
$\rho_{spearman}$	Spearman coefficient
$\tau_{kendall}$	Kendall coefficient
Φ_f	Voronoi penalization
y^+	Dimensionless wall distance

1 Introduction

In wastewater treatment plants, solid particles are an unavoidable component of urban and municipal wastewater. The removal of these solids is necessary to reduce clogging, draining, and cleaning in tanks and pipes, whilst also protecting mechanical equipment from abrasion and wear. The headworks of a treatment plant therefore employs a grit-removal system.

Grit-removal systems are usually designed on the premise of removing clean inorganic particles. Removing particles finer than 75 μm in these systems is normally omitted in their design as they pose a lower risk to downstream equipment. However, suspended solids can lead to the development of sludge deposits and anaerobic conditions if left untreated, which can become a maintenance issue.

During average flow conditions, most grit passes through the grit removal system at a relatively low flowrate. However, during periods of high flowrate, such as during a storm, the quantity of grit captured reduces substantially. It is therefore important that the grit removal system not only operates efficiently during low flow conditions but also under sustained high flows.

The advantage of using Computational Fluid Dynamics (CFD) combined with Bayesian Optimization (BO) to design grit removal systems is it enables realistic simulations of the physical process of hydrodynamic separation, whilst ensuring manufacturing or geometric constraints are considered alongside multiple competing objectives.

1.1 HeadCell system

The HeadCell® (Hydro International Ltd) is a vortex separator used to remove solid particles through a multi-tray hydraulically-driven vortex system. An appealing aspect of its design is the small carbon footprint, having no electrical requirements and no moving parts.

The HeadCell system is outlined in Fig. 1. The design consists of a stack of hydraulically independent polymer trays, which are submerged in a concrete chamber. Screened sewage enters the influent duct and passes into the grit chamber. The duct directs the flow into the distribution header to evenly distribute the influent tangentially into the modular multiple-tray system. The tangential inflow for each tray establishes a vortex flow pattern causing solids to fall into a boundary layer on each tray. Grit settles out by gravity along the sloped surface of each tray and then solids are swept to the center opening which allows them to fall to an underflow collection sump. Degritted effluent flows out of the trays, over a weir, and into an overflow.

The settled solids from the HeadCell are continuously pumped from the grit sump to an open vortex grit washing system e.g. the SlurryCup®. The SlurryCup is used for the removal of organic material from inorganics. The inorganic material is then passed on for further processing, e.g. to the SpiralSnail® and this contains a helical screw used for grit dewatering.

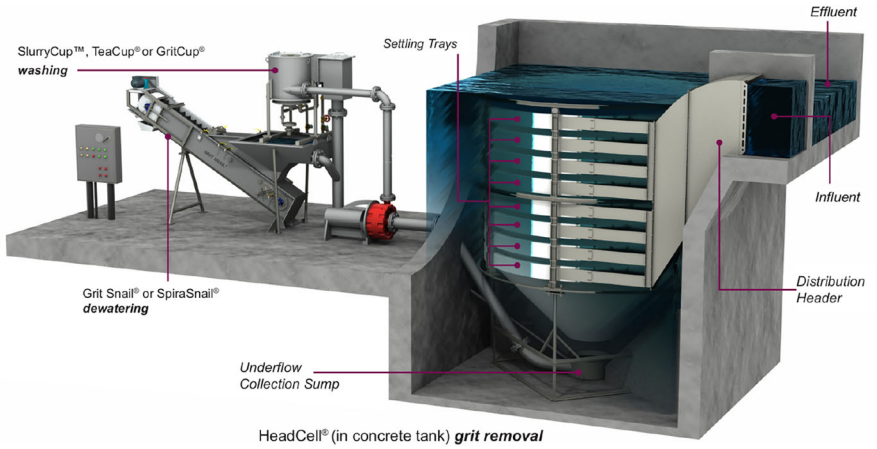

Fig. 1 Schematic of the grit-removal system using the HeadCell apparatus

It is thought that the shape of the HeadCell trays could be optimized to maximize grit collection away from the overflow and maximize grit concentration towards the underflow.

1.2 Previous related studies

Vortex separators are used in several industrial applications, and so have received significant research interest, particularly regarding combining CFD with optimization. Various approaches to the optimization have been investigated including a response surface methodology (Sun et al. 2017), artificial neural networks (Elsayed and Lacor 2011) and genetic algorithms (Sun and Yoon 2018). CFD simulations can be initialized using a steady state result followed by a transient simulation (Sun et al. 2017). One-way coupling of the fluid field to the particles is utilized on the assumption that the particles do not affect the flow field due to the low particle loading (Sun et al. 2017; Elsayed and Lacor 2011). The need for transient simulations is due to the inherent instability of the cyclone flow field, such that a time-resolved Eulerian-Lagrangian approach is often used to include the effect of vortex precession (Sun and Yoon 2018).

The choice of turbulence model can have a substantial impact on the flow field within the simulation of the vortex separator. The literature largely consists of examples of cyclone separation for strongly swirling flows such as those found in the Stairmand cyclone. For these cases, usually either the Reynolds stress model (Launder et al. 1975) or the Spalart-Shur rotation correction (Shur et al. 2000) is applied. This is because the swirl number created by the vortex is around 2.945 based on the geometry (Hartley 1994; Montavon et al. 2000).

For optimization, the work presented here uses the code developed by the current authors in previous studies (Rahat et al. 2017). The type of problem being solved is known as a black-box optimization, specifically derivative-free optimization, in the form of Gaussian process regression (Audet and Hare 2017). It follows the approach of

combining high-fidelity CFD simulations that provide the objectives, with surrogate models that describe the objectives within a decision space (Pretsch et al. 2023). We have previously used homoscedastic Gaussian Process (GP) models without a significant error term to generate several different infill criteria. This is suitable for cases where low uncertainty is present in the CFD simulations, such as steady single-phase flow in a duct (Daniels et al. 2022). Choosing the appropriate infill criterion is essential to the success of the optimization and the best strategy depends on the problem (Rahat et al. 2017). However, it was not clear how uncertainty in the CFD simulations should be treated in the optimization process.

Another issue that was raised by the current authors was the problem of parallelization (Daniels et al. 2022). The developers of previous work have looked at synchronous (González et al. 2016) and asynchronous (Alvi et al. 2019) forms of this approach. Some use local penalization while others do not. Lipschitz-based penalization can be applied to GP models with an error term, however the Lipschitz constant may become underestimated (González et al. 2016). Where there is uncertainty in the objectives, the rival to the Lipschitz-based penalization approach is an acquisition function based on expected hypervolume improvement, but for noisy observations (Daulton et al. 2021).

In addition to the issue of parallelization, there is the problem of how to deal with any mesh or convergence failures. This is usually done by aborting (stopping the optimization), restarting (retrying a specified number of times), continuing (by stepping toward the failed target), or recovering (giving a large objective function value) (Griffin et al. 2006). This problem of crashing has the following characteristics: non-quantifiable (i.e. binary), unrelaxable (i.e. it must be satisfied), simulation-based (i.e. it cannot be known apriori without running a simulation), and also hidden (i.e. it does not appear in the problem formulation). As only violations can be detected by raising an error flag or exception it is called a hidden constraint (Le Digabel and Wild 2023). Others reject failed points from the training set, replace failed points based on non-failed points, or predict the failed region (Bussemaker et al. 2024; Tran et al. 2021). However, it is not clear if any of these approaches are appropriate for BO or under which circumstances each is applicable.

1.3 Paper overview

This study attempts to apply insights gained in prior research in optimization to an industrial problem. This requires considering performance increases for collection efficiency, as well as manufacturing or geometric constraints that would raise costs if not respected (Guo et al. 2024). Performance improvement for the present work is focused on the collection of fine particles at high flow rates. The HeadCell is efficient at removing relatively large particles, however separating smaller particles is more challenging, as identified in prior work (Ji et al. 2023). In addition, the literature review in Sect. 1.2 highlighted several difficulties in applying optimization to uncertain (multi-phase and unsteady) and expensive CFD problems:

- How to minimize the uncertainty in the CFD problem itself
- How to define the objectives and parameterization in the CFD model

– Which infill criterion should be used given the uncertainty in the CFD
– How to include uncertainty from the CFD as the error term in the GP model
– How to parallelize the optimization algorithm
– How to handle hidden constraints and under which circumstances this applies

In order to tackle these problems, Sect. 2 of this paper covers the CFD methodology, followed by the parameterization and objectives in Sects. 3.1 and 3.2. Section 3.3 covers sampling the decision space, Sect. 3.4 defines the infill criterion, and then the implications for the GP model are outlined in Sect. 3.5. Section 3.6 addresses the issue of parallelization, while Sect. 3.7 addresses hidden constraints. The problem formulation is summarized in Sect. 3.8 alongside searching the acquisition function. Section 4 presents the results and discussion, which is followed by conclusions and future work in Sect. 5.

2 CFD methodology

2.1 Standard *k-ω* SST formulation

The fluid flow was described using the unsteady Reynolds-Averaged Navier-Stokes (uRANS) equations. The particular form of the equations used was that given in the open-source C++ toolbox OpenFOAM (Weller et al. 1998). The continuity equation for incompressible flow is given by Eq. 1 and the unsteady momentum equation is Eq. 2.

$$\nabla \cdot \mathbf{u} = 0 \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{u} \otimes \mathbf{u}) = -\nabla p + \nabla \cdot \left\{ (\nu + \nu_t) \left[\nabla \mathbf{u} + (\nabla \mathbf{u})^\top \right] \right\} \tag{2}$$

In the present case the magnitude of the mean swirl number created by the vortex is relatively low, around 0.183, and is comparable with the swirl number that can be simulated with the standard k-ω Shear Stress Transport (SST) model (Engdar and Klingmann 2002). Furthermore, the application of the Spalart-Shur rotation correction made a difference of less than 2% to the objectives in this study. Hence, the k and ω transport equations are Eqs. (3) and (4) (Menter et al. 2003).

$$\frac{\partial k}{\partial t} + \nabla \cdot (\mathbf{u} k) = \nabla \cdot [(\nu + \sigma_k \nu_t) \nabla k] - P_k - \beta^* k \omega \tag{3}$$

$$\frac{\partial \omega}{\partial t} + \nabla \cdot (\mathbf{u} \omega) = \nabla \cdot [(\nu + \sigma_\omega \nu_t) \nabla \omega] - \frac{\alpha_\omega}{\nu_t} P_k - \beta_\omega \omega^2 + 2(1 - F_1) \frac{\sigma_{\omega 2}}{\omega} \nabla k \cdot \nabla \omega \tag{4}$$

The Finite Volume Method was used to integrate Eqs. (2), (3) and (4). Equation (1) is handled iteratively via decoupled fields for pressure and velocity (Weller et al. 1998). A second-order linear upwind difference scheme was adopted for the convection term in Eq. (2). The convection terms in Eqs. (3) and (4) are from first-order upwind differencing. All gradient and diffusion terms in Eqs. (2), (3), and (4) were discretized using linear interpolation (central differencing). The first-order Eulerian temporal scheme

was then used to realize the transient terms in Eqs. (2), (3), and (4). Greenshields and Weller (2022) provide further information on the numerical implementation used in OpenFOAM.

2.2 Lagrangian particle tracking

When the particle density is greater than the fluid density, drag and gravity are the largest forces, whilst shear-induced lift is included due to the large number of walls in the separator (Shi et al. 2018). The pressure gradient force, Basset history force and virtual mass force will be small, as the oscillation frequency of the fluid is low (Sommerfeld et al. 2008).

$$m_p \frac{\mathrm{d}\mathbf{u}_p}{\mathrm{d}t} = \mathbf{F}_D + \mathbf{F}_G + \mathbf{F}_L \tag{5}$$

$\mathbf{F}_D$, $\mathbf{F}_G$, and $\mathbf{F}_L$, are the drag, gravitation/buoyancy, and lift forces exerted on the particles. Once these forces are calculated, the trajectories of the particles are obtained using integration of the particle velocity $\mathbf{u}_p$ via Eq. (6).

$$m_p \frac{\mathrm{d}\mathbf{u_p}}{\mathrm{d}t} = \frac{3}{4}\frac{\rho}{\rho_p}\frac{m_p}{d_p} C_D \mathbf{u}_{rel}|\mathbf{u}_{rel}| + m_p \mathbf{g}\left(1 - \frac{\rho}{\rho_p}\right) + \frac{\rho_f}{2}\frac{\pi}{4} d_p^3 C_L \left(\mathbf{u}_{rel} \times \boldsymbol{\Omega}\right) \tag{6}$$

$\mathbf{u}_{rel}$ is the relative velocity between the phases and is given by Eq. (7).

$$\mathbf{u}_{rel} = \mathbf{u} - \mathbf{u}_p = (\overline{\mathbf{u}} + \mathbf{u}_{fluct}) - \mathbf{u}_p \tag{7}$$

The way in which the fluctuating component of the fluid velocity, $\mathbf{u}_{fluct}$, in Eq. (7) affects the particle tracking is via a stochastic dispersion model (Gosman & Ioannides 1983). This gives a random magnitude and random spherical direction if Eqs. (8), (9), and (10) are considered together.

$$\mathbf{u}_{fluct} = \sqrt{\frac{2k}{3}}\xi\left[H\cos(\theta),\, H\sin(\theta),\, \psi\right] \tag{8}$$

$$\xi = \mathcal{G}(0,1) \quad \text{and} \quad \psi = \mathcal{U}(0,1) \tag{9}$$

$$H = \sqrt{(1-(2\psi-1)^2)} \quad \text{also} \quad \theta = 2\pi\psi \tag{10}$$

Sommerfeld et al. (2008) describe the forces used in the Lagrangian formulation should further details be required.

2.3 Two-phase system

The working fluid is water at 20°C. The particles are sand with a density of 2650 kg/m^3, which matches that used in the experiment. Sand is preferred as the primary function of the HeadCell is the removal of solid inorganics. The Particle Size Distribution (PSD) is governed by the diameters between 75 and 212 μm, with a median diameter (D50)

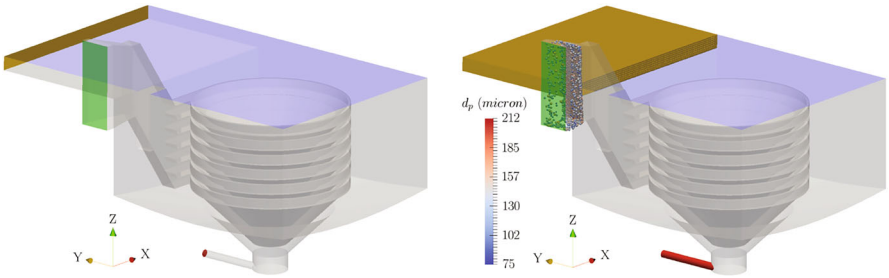

Fig. 2 Validation case—Eulerian boundary conditions (left): inlet (green), underflow (red), overflow (orange), free surface (blue), and walls (grey); Lagrangian boundary conditions (right): particles introduced at the inlet (green), particles captured at underflow (red), and overflow (orange), particles rebound from the free surface (blue) and walls (grey)

of 125 μm. This distribution was chosen as it is a mid-range sand grade that still makes the HeadCell susceptible to the loss of finer particles. The Rosin-Rammler Cumulative Distribution Function (CDF) in Eq. (11) was fitted through the data provided by the laboratory giving $\overline{d_p}$ = 140 mm and n = 9.

$$G = 1 - \exp\left(-\frac{d_p}{\overline{d_p}}\right)^n \tag{11}$$

2.4 Case formulation

Prior work in optimization has generally analyzed parts of systems as opposed to complete systems, with a suitable choice of boundary conditions to model the parts not considered (Daniels et al. 2020). Since the duct is a separate optimization problem, it was removed for BO, and replaced with a nozzle velocity, which along with the tray diameter, sets the Reynolds number in Eq. (12).

$$Re_D = \frac{DU_{nozzle}}{\nu} \tag{12}$$

However, since validation data was only available at the system level, the duct was included to check that the physics of the simulation was correct. This required the use of two separate CFD cases: one for validation as shown in Fig. 2 and one for optimization as shown in Fig. 3.

The left portion of Figs. 2 and 3 shows the boundary conditions used in the scale model of the HeadCell for the Eulerian computations for the validation and BO cases respectively. The boundary conditions included an inlet velocity, overflow pressure, free surface slip, underflow zero gradient, and no-slip walls. Four flowrates were used for validation to match the experiment. A Reynolds number of 1.36×10^6 was used for BO. This was chosen as it is the design condition that all experiments must achieve and is a high flow rate case.

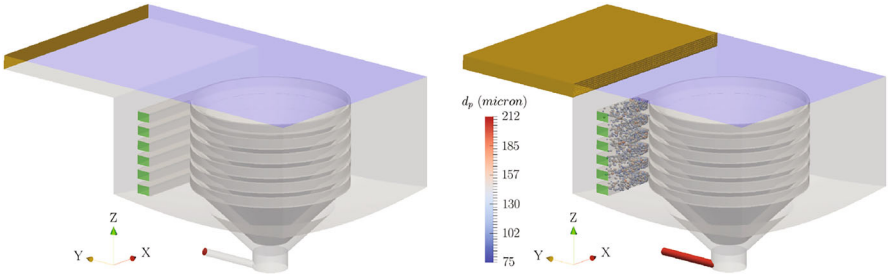

Fig. 3 BO case—Simplified equivalent of Fig. 2 for optimization

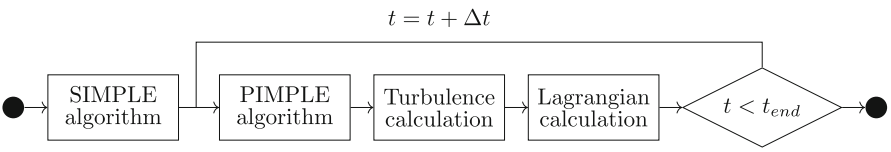

Fig. 4 Eulerian-Lagrangian one-way coupling

The right portions of Figs. 2 and 3 show the boundary conditions for the Lagrangian computations. The inlet introduces the particles, whilst particles that leave the domain in the direction of the overflow or the underflow escape in the volume occupied by a cuboid and a cut cylinder respectively. The PSD is applied spatially uniformly across the inlet in a 60×175 grid, injecting 10,500 particles with random diameters. The inlet, walls, and free surface were set to rebound with a coefficient of restitution of 0.95 and a friction coefficient is 0.2. Details on the implementation of the boundary conditions in OpenFOAM can be found in Greenshields and Weller (2022).

The Eulerian timestep was chosen based on resolving one vortex rotation with around 7 timesteps (Sun et al. 2017). The period of the vortex rotation was computed from the nozzle velocity and the tray body diameter ($D = 908$ mm), which resulted in an Eulerian timestep of 0.28 s at the design flow rate. The unsteady Eulerian-Lagrangian coupling was required to resolve the vortex core precession.

The timestep is applied to the Eulerian-Lagrangian one-way coupled simulation as shown in Fig. 4. A solver was written for this based on the work of a previous author (Kasper 2017). The one-way coupling means the influence of the particles on the fluid flow is neglected due to the low volume fraction of particles, as $\alpha_p < 10^{-6}$ (Elghobashi 1994).

Computations were initialized using the SIMPLE algorithm (Semi-Implicit Method for Pressure Linked Equations). The Eulerian phase takes advantage of the PIMPLE algorithm which is a blend of SIMPLE and PISO (Pressure-Implicit with Splitting of Operators). The PIMPLE algorithm is stable for Courant numbers greater than 1. The integration of the Lagrangian equations takes place at a much finer temporal resolution than the Eulerian equations, as the cell-based Courant number cannot exceed 1. A further limit was set based on the particle response time (Sommerfeld et al. 2008), and the smaller of the numerical and physical timesteps was chosen.

2.5 Collection efficiencies

The water volume of a scale model of the HeadCell with 6 trays is 877 L. One residence time is found by the division of the water volume by the flowrate. At the design flowrate 10 residence times is therefore 210.3 s, which is the physical time used in the simulations.

The underflow efficiency is given by Eq. (13) and the total efficiency is given by Eq. (14). m_{in} is the total mass of particles introduced at the inlet. m_{under}, m_{over}, and m_{vessel} are the masses of particles collected in the underflow volume, overflow volume, and in the vessel after 10 residence times respectively.

$$\eta_{under} = \frac{m_{under}}{m_{in}} \tag{13}$$

$$\eta_{total} = \frac{m_{in} - m_{over}}{m_{in}} = \frac{m_{vessel} + m_{under}}{m_{in}} \tag{14}$$

2.6 Mesh generation

The mesh was generated using the Cartesian mesh generation software cfMesh (Juretic 2020). This was chosen as it allowed the handling of non-conformal geometries. The non-conformity is caused by the Boolean union of the tray geometry (created in Python with numpy-stl) and extruded by their normal vectors in Salome (de France 2019) with primitive CAD in Salome.

In order to generate parametric meshes, the geometry was described using Standard Triangle Language (STL) surfaces. The parts of the geometry that were parameterized included the 6 trays (all identical), the 6 lips (all identical, a copy of the upper section of the tray), and the benching (which is a copy of the lower section of the tray). Due to the parametrization of the benching, it was also necessary to update the grit pot height, vessel, and basin, although these are not directly part of the optimization process. A slice through the mesh with a zoomed-in view is shown in Fig. 5.

The target y^+ for all the walls was between 30 and 100, with an average of around 30 for the parameterized surfaces, hence requiring the use of wall functions. When the first grid point is within the logarithmic layer, the k-ω SST model can still produce valid results, within the limitations of logarithmic wall functions. In addition, placing the first point in the viscous sub-layer created meshes that led to excessive calculation times and convergence problems. Due to these limitations, wall functions were preferred as shown in other optimization studies (Sun and Yoon 2018).

A grid independence check was performed, varying the maximum cell size by $\pm 10\%$. This resulted in meshes of 9.08 million, 10.86 million, and 13.24 million cells. The results indicated that the middle mesh showed an absolute difference of 1.41% compared to the fine mesh for total efficiency and a 0.13% absolute difference compared to the fine mesh for underflow efficiency. To avoid the generation of invalid meshes, maximum cell sizes in increments of 0.01mm were used (between 13.52mm and 13.56mm). A further routine was written based on minimizing the maximum non-orthogonality and skewness using multiple calls to an existing mesh improvement algorithm in cfMesh (Juretic 2020) with progressively relaxed surface iterations.

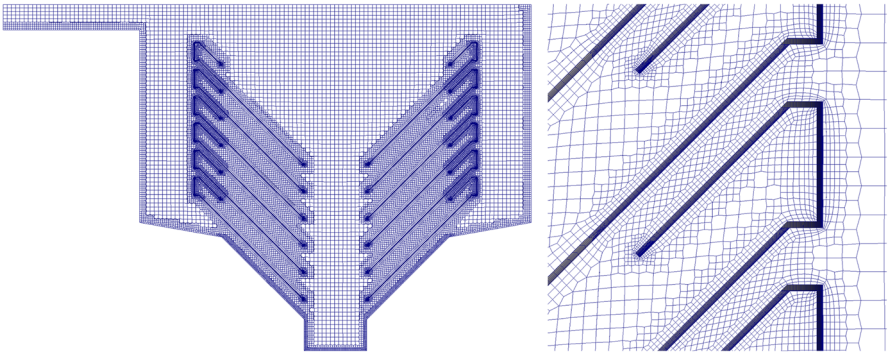

Fig. 5 Slice through the mesh (left), a zoomed view of the inter-tray gap (right)

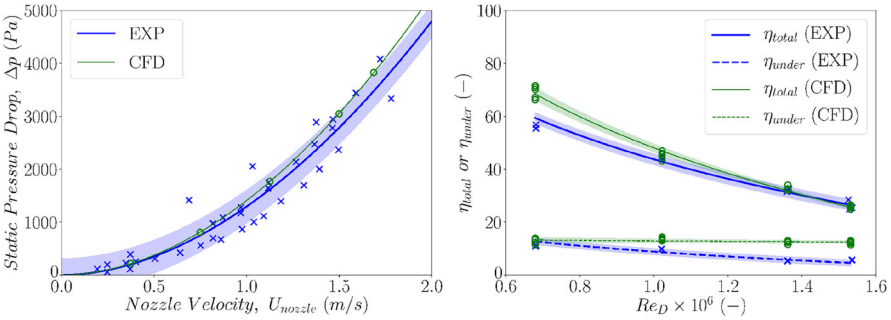

Fig. 6 Static pressure drop is time-averaged over 1 residence time using the PIMPLE algorithm (left); collection efficiencies are obtained from a time-resolved flow field solving the particle tracking simultaneously every fluid timestep for 10 residence times (right)

With a mesh size of 10.86 million cells, using 128 cores per simulation for 4,000 SIMPLE iterations and around 748 PIMPLE-Lagrangian timesteps at 45 iterations per timestep, one run took around 20 h of wall clock time.

2.7 Case validation

It is necessary to validate the CFD model using the base case, which has a linear tray shape, against known values from the laboratory experiments. The validation was done using a time-resolved flow field and included the duct in the system. In Fig. 6 the range and number of points were influenced by the fact that less data was available for the trend in the collection efficiencies due to the length of the experimental procedure. However, the pressure drop is a faster measurement to make, so more data was available for the fit. Figure 6 shows that the pressure drop has a relative difference of less than 10%, whilst the total and underflow efficiencies are within 10% absolute difference compared with the experimental data.

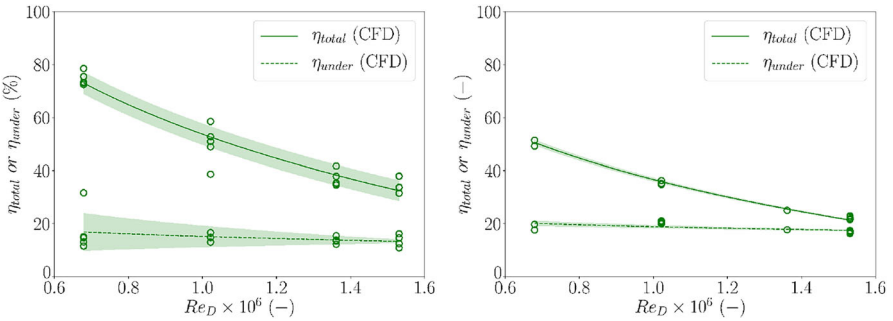

Fig. 7 SIMPLE (stopped at 3000, 3500, 4000, 4500, and 5000 iterations) then the Lagrangian particle tracking for 10 residence times (left); PIMPLE with Lagrangian particle tracking for 10 residence times (right)—5 averaging windows are shown for the initial conditions (2020, 2010, 2000, 1990, and 1980)

Table 1 Standard deviation in objectives—maximum cell sizes were 13.56, 13.55, 13.54, 13.53, and 13.52 mm; averaging windows for initial conditions were 2020, 2010, 2000, 1990, and 1980 on identical meshes; stochastic dispersion averaging was done on 5 samples with identical meshes and initial conditions

Uncertainty cause	σ_{under} (-)	σ_{total} (-)
Mesh Size + Initial Conditions + Stochastic Dispersion	1.41	1.60
Initial Conditions + Stochastic Dispersion	0.60	0.82
Stochastic Dispersion	0.20	0.35

2.8 Case simplification for tray optimization

Despite the requirement to remove the duct for BO, time-resolved flow fields were preferred to steady-state solutions. The reason for this is shown in Fig. 7, where the particle tracking results are very sensitive to the number of iterations and this is because no single steady-state solution exists with the SIMPLE algorithm, however running the PIMPLE algorithm alongside the Lagrangian particle tracking every timestep reduces the uncertainty via improved convergence.

2.9 Quantification of uncertainty

Table 1 shows that the stochastic dispersion is the smallest uncertainty when the initial conditions and the mesh are identical. The next smallest influence is the initial conditions as it is uncertain which iteration the SIMPLE algorithm will stop at. The largest cause of uncertainty is the mesh size (in this case the maximum cell size automatically chosen to avoid an invalid mesh). Based on Table 1, the standard deviation in the objectives can be expected to be of the order of 2%, which is primarily due to uncertainty regarding the mesh having an effect on the Eulerian and Lagrangian fields. To minimize the effect of the initial conditions, the SIMPLE algorithm was stopped when the spatial arithmetic average velocity was judged to be closest to the iterative average.

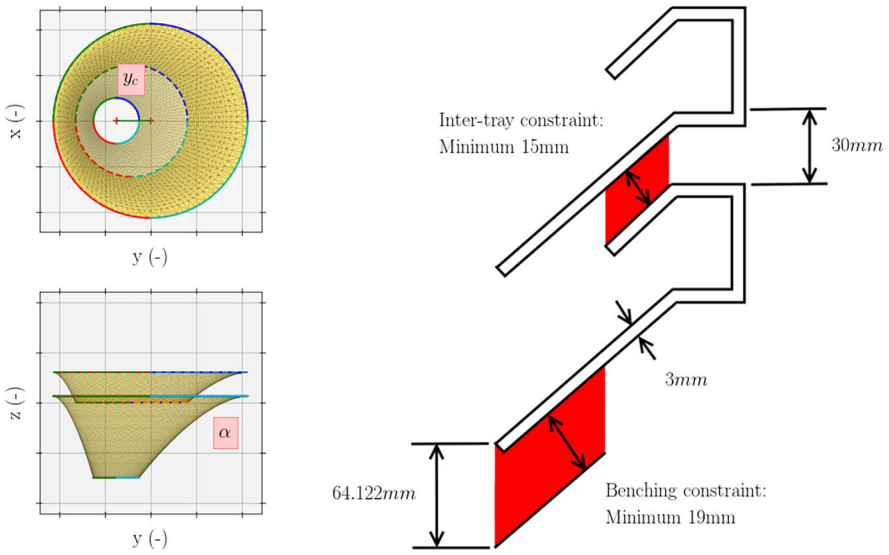

Fig. 8 Surface parameterization (left), zoomed view of geometric constraints (right)

3 Optimization methodology

3.1 Parameterization

Figure 8 shows the definitions of the two parameters used to parameterize the trays y_c and α, giving a decision vector in Eq. (15). α is a parameter used in a monotonic beta CDF. This monotonicity ensures that the tray surface has no cavities or abrupt changes in shape, which could not be manufactured using the rotational molding process. y_c is defined at the base of the tray and is the offset of the center of the tray from the geometric center defined by the top of the tray.

$$\mathbf{x} = [y_c, \alpha]^\top \quad \text{with} \quad \mathbf{x} \in \chi \tag{15}$$

Equation (16) shows the monotonic beta CDF. This sets the variation of the radii with vertical distance (Eqs. 17 and 18) and the center location with vertical distance (Eq. 19).

$$g(z', \alpha, \beta) = \frac{\int_0^{z'} \epsilon^{\alpha-1}(1-\epsilon)^{\beta-1} d\epsilon}{\int_0^1 \epsilon^{\alpha-1}(1-\epsilon)^{\beta-1} d\epsilon} \quad \text{where} \quad \alpha > 0, \quad \beta = 1 \quad \text{and} \quad 0 \le z' \le 1 \tag{16}$$

$$a(z', \alpha, \beta) = (a - A)g(z', \alpha, \beta) + A \quad \text{where} \quad A = 440\,\text{mm} \quad \text{and} \quad a = 50\,\text{mm} \tag{17}$$

$$b(z', \alpha, \beta) = (b - B)g(z', \alpha, \beta) + B \quad \text{where} \quad B = 440\,\text{mm} \quad \text{and} \quad b = 50\,\text{mm} \tag{18}$$

$$c(z', \alpha, \beta) = (y_c - Y_c)g(z', \alpha, \beta) + Y_c \quad \text{where} \quad Y_c = 0\,\text{mm} \quad \text{and} \quad y_c > 0 \tag{19}$$

Equations (17) and (18) show that the radius at the base of the tray was set to 50 mm. A radius of 50 mm was needed to explore a larger decision space, whilst not being so small that it created a blockage risk, as shown in Fig. 9. Equations (17)–(19) are subsequently used to define the shape of the tray given in Cartesian coordinates (Eqs. 20, 21 and 22).

$$x(z', \alpha, \beta, \theta) = a(z', \alpha, \beta)\cos(\theta) \quad \text{where} \quad 0 \leq \theta \leq 2\pi \tag{20}$$

$$y(z', \alpha, \beta, \theta) = b(z', \alpha, \beta)\sin(\theta) + c(z', \alpha, \beta) \quad \text{where} \quad 0 \leq \theta \leq 2\pi \tag{21}$$

$$z(z') = (z_{max} - z_{min})z' + z_{min} \tag{22}$$

It was also necessary to set geometric constraints which take the form of an inter-tray normal distance and a benching-tray normal distance as shown in Fig. 8. The inter-tray gap of 15 mm was determined by trial and error and was set based on the minimum value that gave a successful mesh generation. The reason for this is that if the surfaces are too close, then the mesh degenerates due to lack of resolution. The benching gap was applied based on the prior Hydro recommended minimum, thus limiting the blockage due to particle build-up in this area.

3.2 Multi-objective optimization

The goals of the present work are to simultaneously maximize the underflow collection efficiency (Eq. 13) and the total collection efficiency (Eq. 14). However, since the libraries used to search the decision space operate only on minimization, the search must be converted from maxima into a search for minima (by multiplying the objectives by -1) as shown by Eqs. (23) and (24).

$$\min_{\mathbf{x}\in\chi} y_1(\mathbf{x}) = \min_{\mathbf{x}\in\chi} -\eta_{under}(\mathbf{x}) \tag{23}$$

$$\min_{\mathbf{x}\in\chi} y_2(\mathbf{x}) = \min_{\mathbf{x}\in\chi} -\eta_{total}(\mathbf{x}) \tag{24}$$

3.3 Latin hypercube sampling

To minimize the uncertainty in the subsequent GP model, it was decided that the minimum Euclidean distance between the sampled points should be maximized in order to provide the greatest coverage of the available decision space (known as maximin in the literature). For small-size problems, (such as the current case) a relatively large number of exchanges are affordable, such that maximin variants can be used. However, for large-size problems, it may converge very slowly and require a tremendous number of exchanges (Jin et al. 2005). The decision space is first defined by the geometric constraints illustrated in Fig. 8 and then by the bounds. The bounds were chosen as $y_c \in [0, 170\text{mm}]$ and $\alpha \in [0.29, 1.15]$. This space was used to obtain 21 Latin Hypercube Samples (LHS) (Jones et al. 1998), which required on average 30 samples to achieve (as the geometric constraints would eliminate around 9 samples).

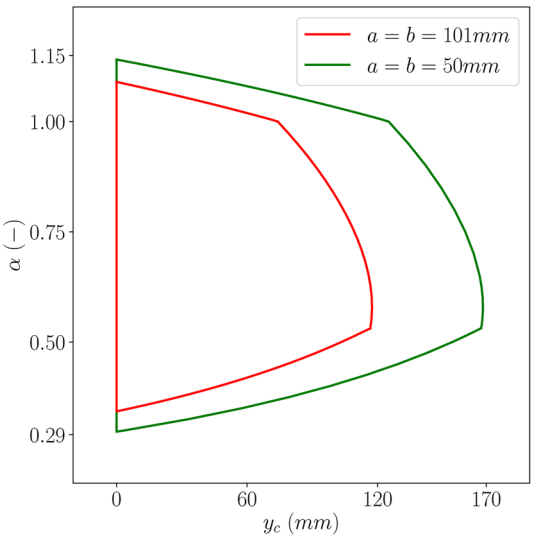

Fig. 9 Extent of constraint space using a=b=101 mm and a=b=50 mm

The comparison of maximin versus a random seed is shown in Fig. 10, illustrating the improved coverage.

Figure 11 shows the comparison between the LHS (shown as crosses) and the base case (shown as a dotted red line). The objectives were obtained using the CFD methodology in Sect. 2 and the parameterization in Sect. 3.1. The blue crosses indicate samples that are either dominated by or mutually non-dominant with the approximate Pareto front. The red cross indicates the approximate Pareto front of samples which are not dominated by any other sample observed so far. The grey dotted line in Fig. 11 shows that the two objectives are largely proportional to each other. This is because the total collection efficiency is a function of both vessel and underflow collection efficiencies as shown in Eq. (14).

3.4 Infill criterion

In a multi-surrogate approach, independent GP models are assumed for each objective i, where N is the number of objectives. The infill criterion used was the Minimum Probability of Improvement (MPoI) (Rahat et al. 2017). The primary reason for using MPoI is it can deal with uncertainty in measurements analytically. Comparisons with other infill criteria are possible, however, this was not done in the present work due to the expensive nature of the CFD problem. Nevertheless, for constrained problems, MPoI has been shown to have comparable convergence with respect to the Probability of Improvement (PI) and Expected Hypervolume Improvement (EHVI) (Grapin et al. 2022).

The multi-objective infill criterion is defined based on the improvement upon any solution as given by Eqs. (25)–(27). A dominance comparison is used to arrive at

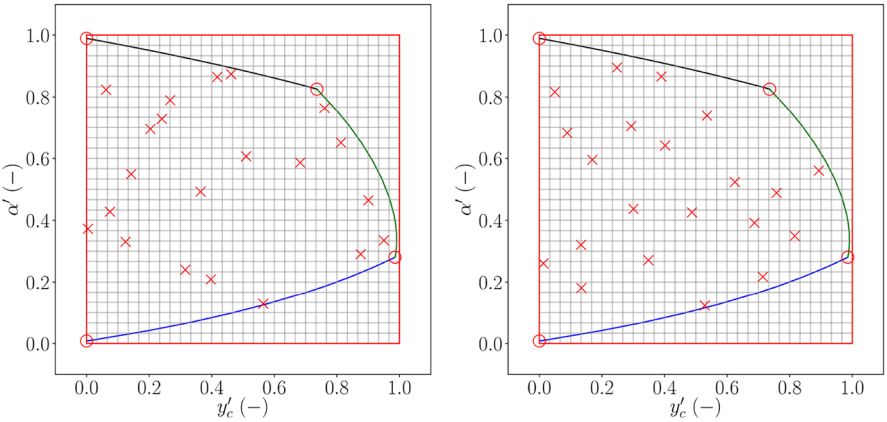

Fig. 10 Arbitrary random seed (left), a random seed that maximizes the minimum Euclidean distance (right)—red lines show the bounds, the red crosses are the LHS data, and the red circles are the 4 corner cases used; the black line is the inter-tray constraint for convex-down trays; the green line is the inter-tray constraint for concave-down trays and the blue line is the benching gap constraint for concave-down trays; the grid is 30 × 30 square

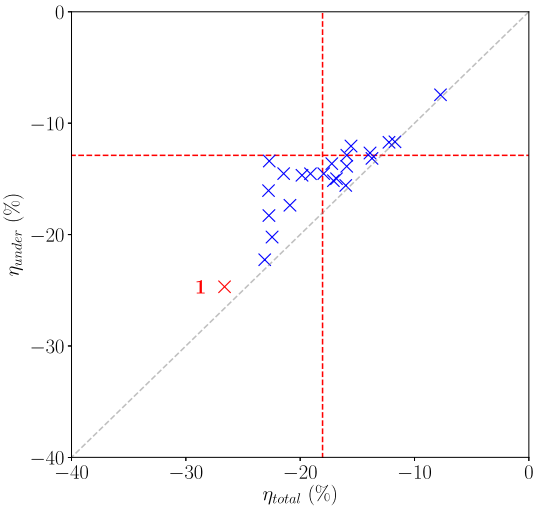

Fig. 11 21 LHS in 2 objectives—approximate Pareto front identified by numbered crosses, the base design is located at the dotted red line (with base radius = 50 mm)

the approximate Pareto set (Rahat et al. 2017), where $\mathbf{x}_{pareto}$ is a decision in the approximate Pareto set $\mathcal{P}^*$, which results in the approximate Pareto front $\mathcal{F}^*$.

$$P(\mathbf{x} \prec \mathbf{x}_{pareto} \text{ or } \mathbf{x}_{pareto} \,||\, \mathbf{x}) = 1 - P(\mathbf{x}_{pareto} \prec \mathbf{x}) \tag{25}$$

$$1 - P(\mathbf{x}_{pareto} \prec \mathbf{x}) = 1 - \prod_{i=1}^{N} \frac{1}{2}\left[1 + \text{erf}\left(\frac{m_i(\mathbf{x}_{pareto}, \mathbf{x})}{\sqrt{2}}\right)\right] \tag{26}$$

$$m_i(\mathbf{x}_{pareto}, \mathbf{x}) = \frac{\mu_i(\mathbf{x}) - \mu_i(\mathbf{x}_{pareto})}{\sqrt{\varepsilon_i^2 + \sigma_i^2(\mathbf{x})}} \tag{27}$$

The MPoI scalarization is Eq. (28), which looks for decisions that have a high probability of dominating the approximate Pareto set, by measuring the minimum probability between the decision and the approximate Pareto set.

$$\alpha_{mpoi}(\mathbf{x}, \mathcal{P}^*) = \min_{\mathbf{x}_{pareto} \in \mathcal{P}^*} (1 - P(\mathbf{x}_{pareto} \prec \mathbf{x})) \tag{28}$$

Thus, in a multi-objective optimization, the next solution to evaluate is the solution that gives the largest MPoI across the approximate Pareto set $\mathcal{P}^*$, i.e. the decision expected to make the largest improvement in the MPoI.

$$\mathbf{x}^* = \arg\max_{\mathbf{x} \in \chi} \left(\alpha_{mpoi}(\mathbf{x}, \mathcal{P}^*)\right) \tag{29}$$

3.5 Gaussian process model

In terms of the GP model, a Gaussian likelihood is preferred because:

- Any attempt to quantify the distribution will be undersampled so the distribution with the largest entropy must be chosen due to a lack of knowledge.
- Gaussian distributions allow evaluation of the acquisition function in closed form.

In this instance, the function used in the process of finding Maximum Likelihood Estimates (MLE) (Frazier 2018) was minimizing the Negative Log Marginal Likelihood (NLML). The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was used with 10 restarts to optimize the kernel hyperparameters (Rahat et al. 2017).

Epistemic uncertainty was encoded via a non-isotropic variance hyperparameter in the GP model. The Matern 5/2 covariance kernel was used as it is suitable for simulating realistic functions (Snoek et al. 2012). No attempt was made to restrict the variance in the GP model, as no information was available that could have been used to restrict it. Hence it had a range between 1×10^{-4} and 1×10^{8}.

Aleatoric uncertainty in the objectives of the CFD model required an error term in the GP model (Jones et al. 1998). This took the form of a fixed hyperparameter. Alternative names for the error term are nugget or jitter and it is assumed to be directly proportional to the variance of the uncertainty quantified in Sect. 2.9 (Bostanabad et al. 2018). It was found that a homoscedastic standardized error variance of 0.1 was sufficient and any further reductions in the error term gave uncorrelated GP models.

Upper and lower bounds normalization produced good fits to the CFD data when the lengthscale hyperparameter was bounded by the minimum and maximum Euclidean distance in the data. In addition, to prevent over-sampling, a robust approach was also taken to prevent the lengthscale dropping below 1×10^{-4}.

In order to measure the ability of the GP model to predict values that have not been part of the training data, the leave-one-out cross-validation was conducted using two measurements of success. Figure 12 shows the leave-one-out cross-validation where the model is as trained on $n - 1$ data points leaving out one test point. At the test point, the objective is predicted and compared with the known objective. The success of the fit is measured by Spearman's rank correlation coefficient, $\rho_{spearman}$, and Kendall's rank correlation coefficient, $\tau_{kendall}$ (Zwillinger and Kokoska 2000).

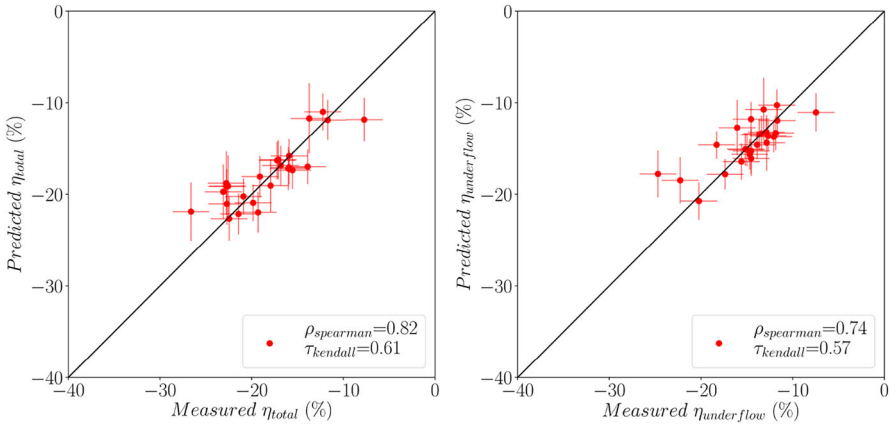

Fig. 12 Leave-one-out cross-validation with standardized error variance $\varepsilon_{i\in N} = 0.1$—this results in an epistemic plus aleatoric predicted standard deviation that is similar to the maximum aleatoric standard deviation in the measured data (around 2%)

Figure 12 shows that $\rho_{spearman}$ for the measured and predicted values is a strong positive correlation ($\rho_{spearman} > 0.6$). $\tau_{kendall}$ also shows a strong positive correlation between the measured and predicted data ($\tau_{kendall} > 0.4$) for both objectives. The standard deviation present in the data (i.e. 2%) is comparable with the standard deviation from the model, and the results are broadly in agreement. The predicted-measured charts do not fall perfectly on a straight line perhaps because of the uncertainties such as the mesh, initial conditions, and stochastic dispersion as quantified in Table 1. Phenomena such as vortex core precession, plume flow, and resuspension could also have an influence on both objectives. These transient effects may mean the objectives are not simply governed by the tray geometry alone but also by the frequency of the velocity fluctuations. However, the sensitivity to this effect is beyond the scope of this work, as the geometry is being optimized, not the transient conditions.

3.6 Parallelization

The parallelization penalization is done using the hard local penalizer, with $q = -5$, $\kappa = 1$ and $\delta = 1 \times 10^{-10}$ for each objective, i, with $\mathbf{x}_{penalize}$ as a penalized location as shown in Eq. (30) (Alvi et al. 2019).

$$\phi_{pi}\left(\mathbf{x}, \mathbf{x}_{penalize}\right) = \left[\left(\frac{||\mathbf{x} - \mathbf{x}_{penalize}|| \; + \; \delta}{\frac{|\mu_i(\mathbf{x}_{penalize}) - \widetilde{\mathcal{M}}_i|}{\widetilde{\mathcal{L}}_i} + \kappa \frac{\sigma_i(\mathbf{x}_{penalize})}{\widetilde{\mathcal{L}}_i}}\right)^q + 1^q\right]^{1/q} \tag{30}$$

Double the normal Lipschitz constant was used due to the likely underestimation of the Lipschitz constant due to the standard deviation in the objectives (González et al. 2016). The estimate of the best value, $\widetilde{\mathcal{M}}$, can be under-estimated if data points are used. This would lead to a large radius as $\mu - \widetilde{\mathcal{M}}$ would be large, which is dangerous

as it may exclude optimal points. Therefore $\widetilde{\mathcal{M}}$ is given a value from the model using the points already evaluated as shown in Eq. (31) (González et al. 2016). This allows the radius to shrink around the optimal point (Alvi et al. 2019). However, the standard deviation in the GP model means it cannot reach zero.

$$\widetilde{\mathcal{M}_i} = \max \mu_i(\mathbf{x}) \quad \text{with} \quad \mathbf{x} \in X_{evaluated} \tag{31}$$

A problem arises in applying Eq. (30) to the multi-surrogate MPoI case, in that multiple radii of exclusion are possible as there is more than one Lipschitz constant, more than one best value, and more than one standard deviation at the penalization point. Continuing the theme of maximizing exploitation, in order to choose the penalization value that is closest to the serial case (and is therefore minimally penalizing the landscape), the penalization with the minimum radius of exclusion is selected, i.e. the one giving the largest value for the penalization function as shown in Eq. (32). A local Lipschitz constant is used in order to make the penalization sensitive to the gradient at the particular location (Alvi et al. 2019).

$$\begin{aligned}\Phi_p(\mathbf{x}, \Psi) &= \prod_{\mathbf{x}_{penalize} \in \Psi} \left[\max_{i \in N} \left\{ \phi_{pi}(\mathbf{x}, \mathbf{x}_{penalize}) \right\} \right] \\ &= \prod_{\mathbf{x}_{penalize} \in \Psi} \left[\max_{i \in N} \left\{ \left[\left(\frac{||\mathbf{x} - \mathbf{x}_{penalize}|| \; + \; \delta}{\frac{|\mu_i(\mathbf{x}_{penalize}) - \widetilde{\mathcal{M}_i}|}{\widetilde{\mathcal{L}_i}} + \kappa \frac{\sigma_i(\mathbf{x}_{penalize})}{\widetilde{\mathcal{L}_i}}} \right)^q + 1^q \right]^{1/q} \right\} \right]\end{aligned} \tag{32}$$

This therefore modifies the acquisition function to Eq. (33).

$$\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{X}} \left(\Phi_p(\mathbf{x}, \Psi) \; \alpha_{mpoi}(\mathbf{x}, \mathcal{P}^*) \right) \tag{33}$$

The serial approach will produce better optimization results than the parallel approach, since at any point during the optimization the GP model uses all the information strictly available at the time, whilst the parallel approach chooses decisions based on both complete and incomplete CFD runs. However, if the function evaluations take a long time (such as 20 h in this case) and end-to-end optimization time is important, then the parallel approach becomes attractive. By default, it is recommended to keep the ratio of allowed parallelism to total trials relatively small ($< 10\%$) in order to not hurt optimization performance too much (Daulton et al. 2020), but the reasonable ratio can differ depending on the specific setup. In the present case, batches of 5 samples are used with 100 simulations which is a 5% ratio. Using this parallelization approach and with around 20 h per CFD simulation, the wall clock time can be reduced from around 3.5 months to around 18 days.

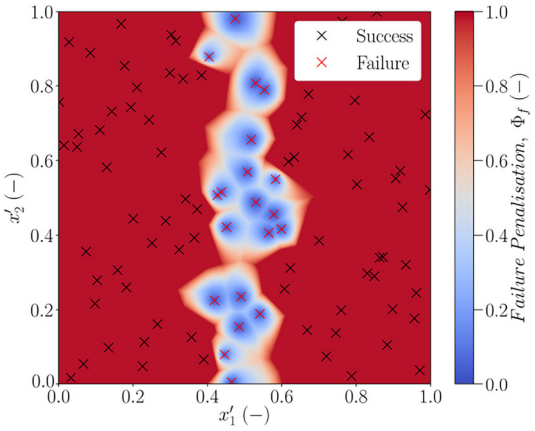

Fig. 13 Voronoi penalization for a fictitious problem—failures have been artificially triggered to occur between decisions $x_1' = 0.4$ and $x_1' = 0.6$

3.7 Hidden constraints

The parallelized optimization approach is evaluated with a CFD model, which is an expensive 3D Eulerian-Lagrangian simulation. The first problem with this is that even if a mesh is successfully generated, it may not be of sufficient quality to allow the solution to converge. In the current work, it is assumed that a mesh will be successful (since the mesh optimization process is robust and the geometric constraints on the mesh are well-known). However, if the mesh or solution results in non-convergence, then the decision must be penalized. A second problem is that hardware failures can occur, such that the CFD fails to return objectives, or the BO is stopped completely. Both of these are known as hidden constraints in the optimization literature (Le Digabel and Wild 2023).

If a solution fails to converge during the BO, an additional sample will be added per failure and the decision that caused the failure would be penalized. However, failure to converge is serious and would require restructuring the CFD model or the mesh. Hardware failures would result in the CFD model being paused, where it would wait for a manual restart. If the BO is killed, it can be restarted with the saved results file posing as the initial samples in the next run.

For the convergence failures, the acquisition function can be penalized using parameter-free Voronoi tessellation. The penalization function is defined by Eq. (34), where X_s are previously successful decisions and X_f are previously failed decisions. An illustration of the shape of this function in 2D is shown in Fig. 13 for an artificial problem.

$$\Phi_f\left(\mathbf{x}, X_s, X_f\right) = \min_{\mathbf{x}\in\mathbb{X}}\left(1, \frac{\min(||\mathbf{x} - X_f||)}{\min(||\mathbf{x} - X_s||)}\right) \tag{34}$$

This modifies the acquisition function in Eq. (33) to a different form, as given by Eq. (35). The Bayesian framework shown in Rahat et al. (2017) was altered to suit the approach in the current work.

$$\mathbf{x}^* = \underset{\mathbf{x}\in\chi}{\arg\max} \left[\Phi_f \left(\mathbf{x}, X_s, X_f \right) \Phi_p(\mathbf{x}, \Psi)\, \alpha_{mpoi}(\mathbf{x}, \mathcal{P}^*) \right] \tag{35}$$

3.8 Optimization problem formulation

To summarize the problem formulation, the objective vector to be minimized is as shown in Sect. 3.2 and restated as Eq. (36).

$$\mathbf{y}(\mathbf{x}) = [-\eta_{under}(\mathbf{x}), -\eta_{total}(\mathbf{x})] \quad \text{with} \quad \mathbf{x} \in \chi \tag{36}$$

The decision vector is shown in Sect. 3.1, restated as Eq. (37).

$$\mathbf{x} = [y_c, \alpha]^\top \quad \text{with} \quad \mathbf{x} \in \chi \tag{37}$$

The bounds of the decision vector are in Sect. 3.3, given as the constraint in Eq. (38).

$$(y_c \in [0, 170\text{mm}]) \wedge (\alpha \in [0.29, 1.15]) \tag{38}$$

Further geometric constraints on the decision space are that the benching gap and inter-tray gaps cannot be less than 19 mm and 15 mm respectively which are stated in Sect. 3.1 and also as the constraint in Eq. (39).

$$(\delta_{benching} \geq 19\text{mm}) \wedge (\delta_{inter-tray} \geq 15\text{mm}) \tag{39}$$

The next solution to evaluate is based on the MPoI scalarization α_{mpoi} (see Sect. 3.4) using the parallel penalization Φ_p (see Sect. 3.6) and the failure penalization Φ_f (see Sect. 3.7) as given by Eq. (35).

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is used to search the landscape of the acquisition function by starting in random locations within the bounds and using the geometric and hidden constraints. When a limit of $50,000 \times n_{dim}$ samples is used the time taken for this search is around 40 min, and increasing the number of samples did not improve the MPoI values, since a plateau was reached. Due to the error term in the GP model, decisions may be selected more than once, which is allowed for in the data handling.

4 Results and discussion

Figure 14 shows the result of the BO process. The approximate Pareto set hypervolume was defined as the volume of objective space which is dominated by the approximate Pareto set and dominates a reference vector. The reference vector was chosen as the objective vector from the base design. The approximate Pareto set size was defined as the number of decisions in the approximate Pareto set. Should a decision be removed from the approximate Pareto set, it was recorded as an ejection. Finally, the approximate Pareto set indices illustrate the lifespan of a particular decision in the approximate Pareto set.

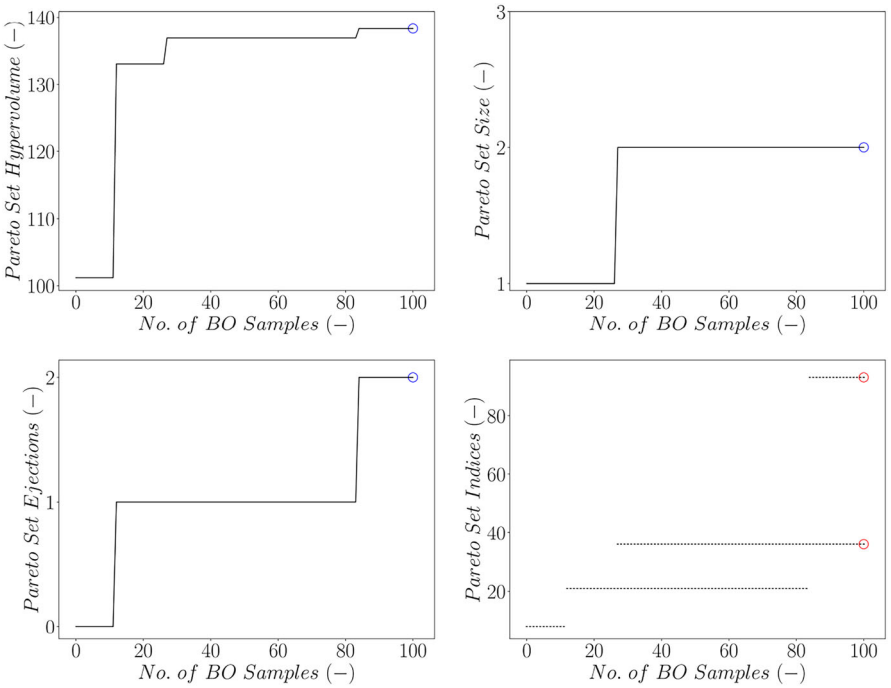

Fig. 14 Convergence of the approximate Pareto set vs number of BO samples

It is clear that convergence of the approximate Pareto set hypervolume has been achieved. The number of designs in the approximate Pareto set was found to be 2, whilst 2 ejections had taken place. The approximate Pareto set is also fairly stable, having had the same set of 2 decisions for around 20 Bayesian samples.

Ideally, multiple runs of the Bayesian optimization would be performed and a summary attainment surface would be created. However, considering that one Bayesian run can take around 18 days even with the parallelization, the time for 3–5 such runs would be the order of months to complete. It was also considered that the standard deviation in the objectives of the CFD model was the primary source of uncertainty, not the location of the LHS, the fit of the GP model, or the uncertainty related to the CMA-ES search pattern. Running multiple BO runs is therefore unlikely to obtain any further information.

With regards to the GP model of the objectives, Fig. 15 shows there is strong agreement between the objectives regarding the likely location of the approximate Pareto front. One drawback is that the shape of the GP models in Fig. 15 produced a cliff-edge landscape, which was perhaps over-exploited. It may be possible to modify the acquisition function to achieve a greater balance between exploitation and exploration. However, the standard deviation in both models is fairly low, being lowest around the LHS and the Bayesian samples. The solid red and dotted red lines denote the inter-tray and benching geometric constraints respectively. Beyond this boundary no solutions are possible, therefore the uncertainty is raised where there can never be any data.

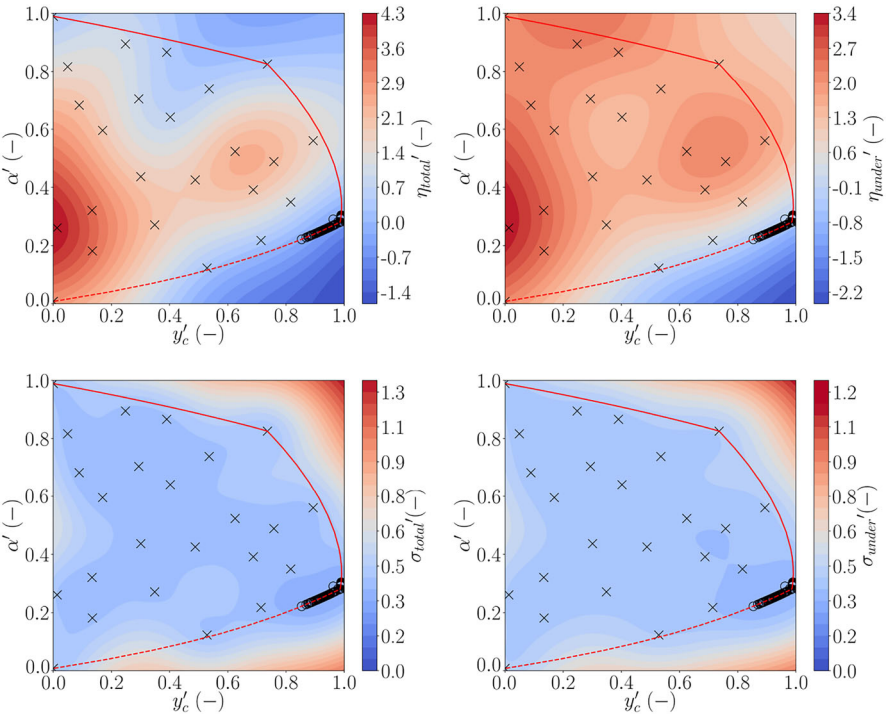

Fig. 15 GP model—total efficiency (top left), underflow efficiency (top right), standard deviation in total efficiency (bottom left) and standard deviation in underflow efficiency (bottom right); the crosses are LHS and the circles are Bayesian samples

Figures 16 and 17 show the approximate Pareto front and Pareto set that was achieved. The approximate Pareto front demonstrates that the improvement over the base design was 14% for the underflow collection efficiency and 10% for the total collection efficiency. In addition, the approximate Pareto set is located on the benching constraint limit. This means that the benching gap is more vital than the inter-tray gap. This is not as intuitive as it first appears, as the geometric constraints could have suggested that it is better to minimize both the inter-tray gap and the benching gap.

The approximate Pareto set does not indicate the most robust solution, which would require running the designs in multiple installations and conditions. A barrier to robustness is that the uRANS Eulerian-Lagrangian approach is expensive and still contains a degree of uncertainty. Some authors claim a speed-up with a quasi-Eulerian-Eulerian approach, which also eliminates stochastic effects (Li et al. 2020), however, these methods rely on calibration for accuracy. Despite this, Fig. 18 shows that the decisions in the approximate Pareto set are within around 18 mm for y_c and 0.05 for α, and the shapes appear similar. The objectives are also within 1%, which is easily within the uncertainty present in the CFD.

Both concavity and convexity have been identified as playing an essential role in improving the performance of hydrocyclones (Ji et al. 2023). Concavity in particular (or what could be termed concave down in the case of Fig. 18) is associated with better

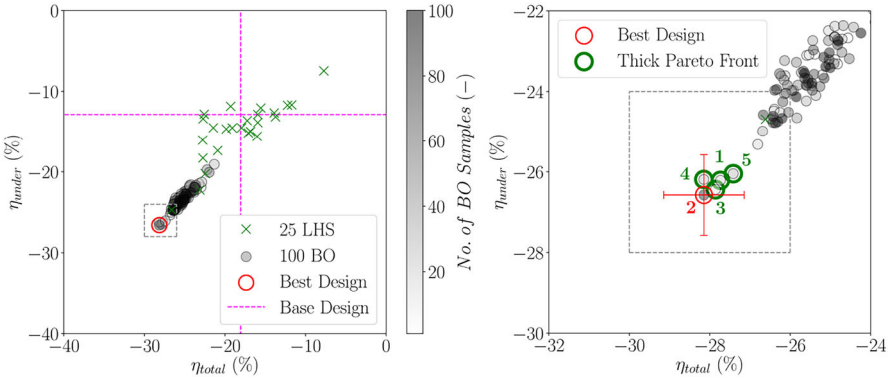

Fig. 16 The best design compared with the base design (left) and the thick approximate Pareto front (right)

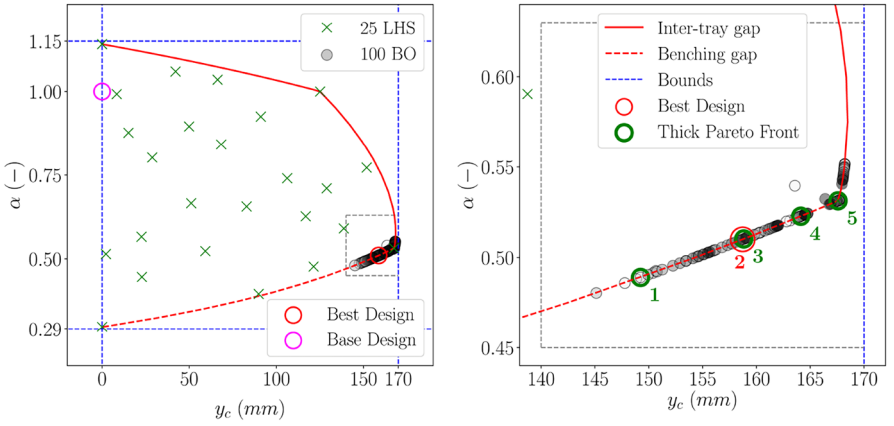

Fig. 17 The best design compared with the base design (left) and the approximate Pareto set from the thick approximate Pareto front (right)

continuity at the bottom of the device and a more stable flow field (Li et al. 2022). The improvement in the total collection efficiency of hydrocyclones due to such shapes has also been reported in previous studies (Pandey et al. 2022; Li et al. 2022).

From Fig. 16, it is important to understand the composition of the total collection efficiency in the base case. This is largely made up of two contributions, the particles that are in suspension and the particles that are captured by the underflow and the surfaces. Equation (40) defines a capture efficiency, which is plotted in Fig. 19 after 10 residence times plus one timestep.

$$\eta_{capture} = \frac{m_{surfaces} + m_{under}}{m_{in}} \tag{40}$$

It is clear from this that the total collection efficiency in the base case is composed of a substantial percentage of particles in suspension, up to 14% at the lowest Reynolds number. However, for the optimum case, the amount of mass in suspension is no

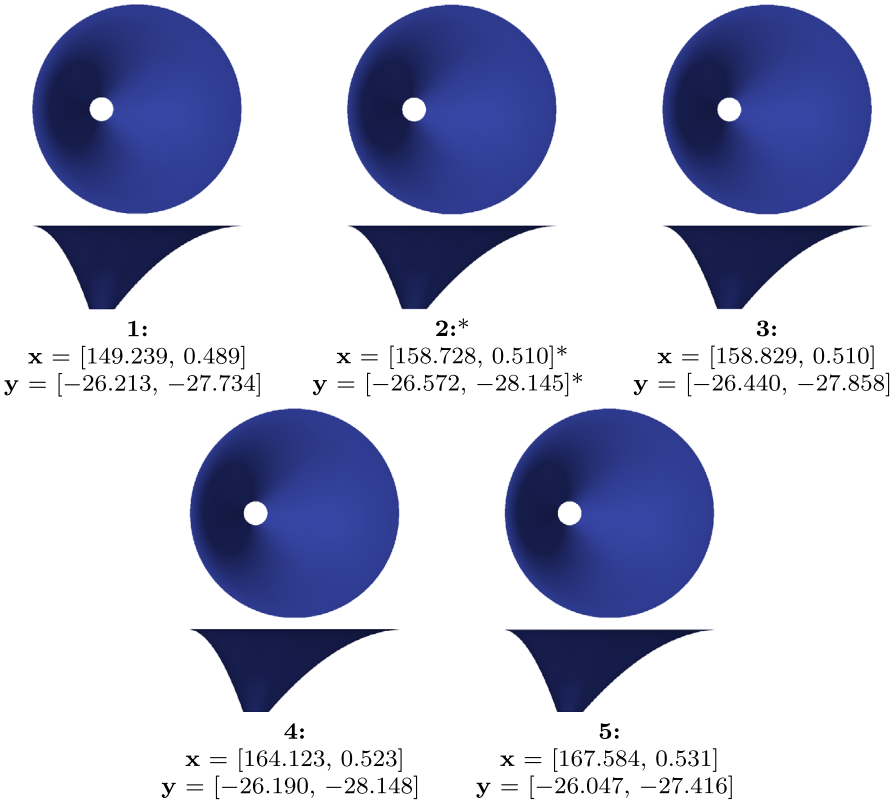

Fig. 18 Approximate Pareto set trays—for decisions **x** and objectives **y**; each of the objectives are within 1% of the best design (* = best design)

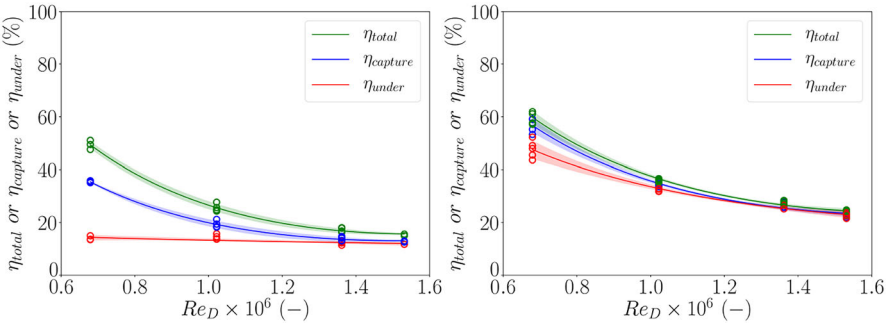

Fig. 19 Collection efficiencies vs Reynolds number – base (left), optimum (right)

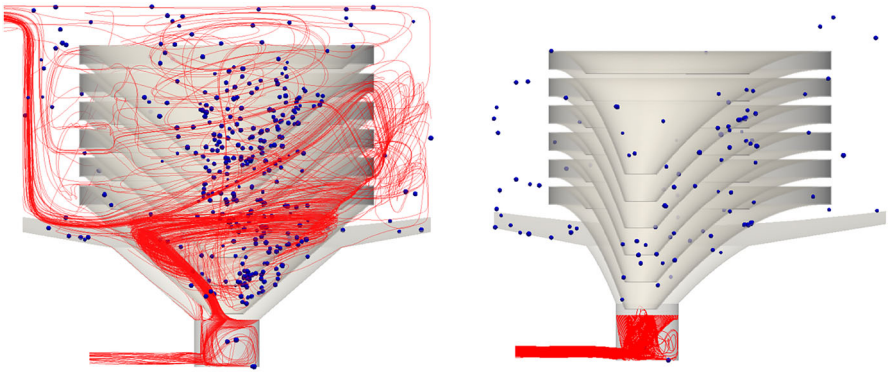

Fig. 20 Instantaneous discrete streamlines and particles at 10 residence times—base (left), optimum (right) for $Re_D = 1.36 \times 10^6$

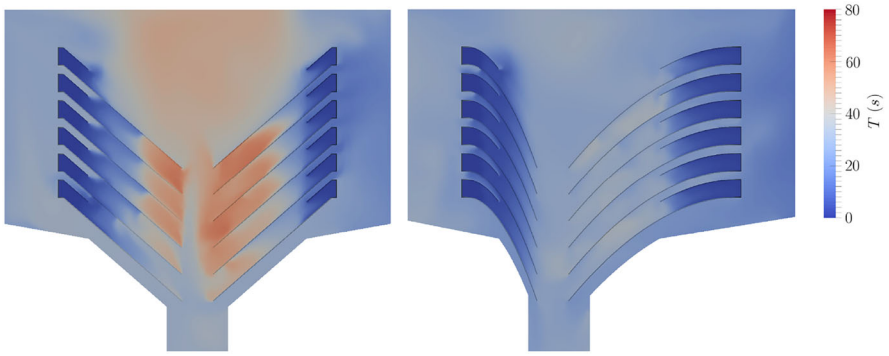

Fig. 21 Instantaneous LMA of fluid at 10 residence times—base (left), optimum (right) for $Re_D = 1.36 \times 10^6$

more than 3% at the lowest Reynolds number. This is primarily due to the substantial increase in the underflow collection efficiency, as the particles are sent directly to the underflow and do not remain in suspension. The optimization can achieve such improvements partly because of a well-defined decision space (see Sect. 3.1). If over-constrained it cannot reach effective locations. Conversely, if the decision space is too large, it requires too many samples to find optimal results. The effectiveness of the optimization is also a function of minimizing the uncertainty in the CFD data by using the PIMPLE algorithm (see Sect. 2.8) and also by the data having a good fit to the chosen GP model (see Sect. 3.5).

Figure 20 gives the physical basis for why the capture efficiency is so substantially improved. Shown here are particle locations after 10 residence times and streamlines emanating from a cross located at the top of the grit pot (as it is positioned in the base case). It is clear from the streamlines in the base case that a plume is responsible for suspending the particles in solution and sending them out of the device rather than capturing them in the underflow and on the surfaces.

In addition to this, it is possible to provide the Local Mean Age (LMA) of the fluid, which is analogous to residence time. The equation for the LMA of the fluid is given by the steady-state scalar T in Eq. (41) (Bartak et al. 2001).

$$\nabla \cdot (\mathbf{u}T) - \nabla^2 \nu_{eff} T = 1 \tag{41}$$

The result of the LMA of the fluid (or T in Eq. 41) for the design condition is shown in Fig. 21. This adds further evidence to the fact that the base design has a very much raised residence time for the fluid. In addition, the optimum design not only lowers this but also provides a much more uniform distribution of residence time from tray to tray. The conclusion that improved performance of hydrodynamic separators can be achieved by reducing the residence time is in alignment with previous work (Ji et al. 2023). In addition, the finding of improved separation efficiency by offsetting the vortex using an eccentric geometry is also in agreement with studies in the past (Yao et al. 2022).

5 Conclusions and future work

5.1 Conclusions

A hydrodynamic separator has been optimized using unsteady CFD coupled with Lagrangian particle tracking combined with a parallelized and robust formulation of BO. The major challenge was how to handle the uncertainty in the data which was due to stochastic dispersion, initial conditions, and mesh quality in the CFD. This resulted in the use of the MPoI infill criterion and also a homoscedastic error term in the GP models. A very exploitative parallelization strategy was devised to cope with the cliff-edge landscape in the objective function, alongside handling hidden constraints due to software or hardware failures.

The approximate Pareto front showed consistent designs that were all concave down and offset towards the inlet side of the trays. The increase in performance achieved was an absolute improvement over the base design of 14% in the underflow collection efficiency and 10% in the total collection efficiency. The physical explanation of the increase in performance of the new designs was attributed to having a design that uniformly distributes the residence time across the trays, whilst also substantially reducing its magnitude. This meant that most of the total collection efficiency was composed of particles sent directly to the underflow. However, in the base design, the presence of a plume meant a substantial percentage of the particles were being held in suspension and did not reach the underflow in the simulated time period. This result is considered novel, useful, and non-intuitive such that it has been taken forward for a patent application (Roberts et al. 2024).

5.2 Future work

Future work should be focused on further minimizing the uncertainty in the CFD simulations, perhaps using alternative approaches. These uncertainties include the transient nature of the flow, the mesh generation, initial conditions, and stochastic

dispersion. Furthermore, although the geometry has been optimized for a particular installation scenario, it is not yet clear if differing situations would show similar performance improvements. However, this would require multiple evaluations for each perceived variable, which would substantially raise the expense. A case for using lower-fidelity CFD could be made on this basis.

Acknowledgements The authors would like to thank Steven Daniels for laying the foundations for the CFD-BO integration in the prior EPSRC project and for the discussions during the current project. In addition, thanks are due to George de Ath for the discussions on Bayesian optimization. Finally, the discussions, experimental assistance, and design expertise of Paul LeCornu were also invaluable to the project. The CFD-BO simulations were performed on the GW4 Isambard UK National Tier 2 HPC facility at the Met Office, Exeter, UK (http://gw4.ac.uk/isambard/). This system is operated by GW4 and the UK Met Office and funded by the EPSRC (EP/T022078/1).

Author contributions Conceptualization: all; Methodology: all; Formal analysis and investigation: A.P.R.; Writing—original draft preparation: A.P.R.; Writing—review and editing: all; Funding acquisition: D.S.J., J.E.F., G.R.T.; Resources: A.A.M.R., D.S.J., J.E.F., G.R.T.; Supervision: A.A.M.R., D.S.J., J.E.F., G.R.T.

Funding This work was supported by Innovate UK and the EPSRC (partnership number 11477).

Data availibility No data for public archival are reported in this study. This study does not report any data of this kind. For the purpose of open access, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising.

Declarations

Ethical approval and consent to participate This work does not report on or involve any animals, humans, human data, human tissue, or plants, such that ethics approval or consent to participate is not applicable.

Consent for publication The information contained in this study does not involve any individual person's data in any form, such that consent for publication is not applicable.

References

Alvi AS, Ru B, Callies J, Roberts SJ, Osborne MA (2019) Asynchronous batch Bayesian optimisation with improved local penalisation. In: 36th international conference on machine learning, ICML 2019, pp 373–387

Audet C, Hare W (2017) Derivative-free and blackbox optimization. Springer Series in Operations Research and Financial Engineering. https://doi.org/10.1007/978-3-319-68913-5. Accessed 14 May 2024

Bartak M, Cermak M, Clarke JA, Denev J, Drkal F, Lain M, Macdonald IA, Majer M, Stankov P (2001) Experimental and numerical study of local mean age of air. In: 7th international IBPSA conference, Rio de Janeiro, Brazil, pp. 773–780

Bostanabad R, Kearney T, Tao S, Apley DW, Chen W (2018) Leveraging the nugget parameter for efficient gaussian process modeling. Int J Numer Methods Eng 114(5):501–516. https://doi.org/10.1002/nme.5751

Bussemaker JH, Saves P, Bartoli N, Lefebvre T, Nagel B (2024) Surrogate-based optimization of system architectures subject to hidden constraints. https://hal.science/hal-04462408. Accessed 14 May 2024

Daniels SJ, Rahat AAM, Tabor GR, Fieldsend JE, Everson RM (2020) Shape optimisation of the sharp-heeled Kaplan draft tube: performance evaluation using computational fluid dynamics. Renew Energy 160:112–126. https://doi.org/10.1016/j.renene.2020.05.164

Daniels SJ, Rahat AAM, Tabor GR, Fieldsend JE, Everson RM (2022) Application of multi-objective Bayesian shape optimisation to a sharp-heeled Kaplan draft tube. Optim Eng 23(2):689–716. https://doi.org/10.1007/s11081-021-09602-6

Daulton S, Balandat M, Bakshy E (2020) Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. In: 34th conference on neural information processing systems, NeurIIPS 2020. https://doi.org/10.48550/arXiv.2006.05078

Daulton S, Balandat M, Bakshy E (2021) Parallel Bayesian optimization of multiple noisy objectives with expected hypervolume improvement. In: 35th conference on neural information processing systems, NeurIIPS 2021. https://doi.org/10.48550/arXiv.2105.08195

de France E (2019) Salome version 9.3.0. https://www.salome-platform.org. Accessed 14 May 2024

Elghobashi S (1994) On predicting particle-laden turbulent flows. Appl Sci Res 52:309–329. https://doi.org/10.1007/BF00936835

Elsayed K, Lacor C (2011) Modeling, analysis and optimization of aircyclones using artificial neural network, response surface methodology and cfd simulation approaches. Powder Technol 212(1):115–133. https://doi.org/10.1016/j.powtec.2011.05.002

Engdar U, Klingmann J (2002) Investigation of two-equation turbulence models applied to a confined axis-symmetric swirling flow. In: Proceedings of the ASME 2002 pressure vessels and piping conference. Computational technologies for fluid/thermal/structural/chemical systems with industrial applications, vol 2, pp 199–206. https://doi.org/10.1115/PVP2002-1590

Frazier PI (2018) A tutorial on Bayesian optimization. https://doi.org/10.48550/arXiv.1807.02811. Accessed 14 May 2024

González J, Dai Z, Hennig P, Lawrence N (2016) Batch Bayesian optimization via local penalization. In: Proceedings of the 19th international conference on artificial intelligence and statistics, AISTATS 2016, pp 648–657. https://doi.org/10.48550/arXiv.1505.08052

Gosman AD, Ioannides E (1983) Aspects of computer simulation of liquid-fueled combustors. J Energy 7(6):482–490. https://doi.org/10.2514/3.62687

Grapin R, Diouane Y, Morlier J, Bartoli N, Lefebvre T, Saves P, Bussemaker J (2022) Regularized infill criteria for multi-objective Bayesian optimization with application to aircraft design. In: AIAA aviation 2022 forum. https://doi.org/10.2514/6.2022-4053

Greenshields C, Weller H (2022) Notes on computational fluid dynamics: general principles. CFD Direct Ltd, Reading, UK. https://doc.cfd.direct/notes/cfd-general-principles/. Accessed 14 May 2024

Griffin JD, Eldred MS, Martinez-Canales ML, Watson J-P, Kolda TG, Giunta AA (2006) Dakota developers manual. https://digital.library.unt.edu/ark:/67531/metadc881219/. Accessed 14 May 2024

Guo M, Yang L, Son H, Le DK, Manickam S, Sun X, Yoon J-Y (2024) An overview of novel geometrical modifications and optimizations of gas-particle cyclone separators. Sep Purif Technol 329:125136. https://doi.org/10.1016/j.seppur.2023.125136

Hartley C (1994) Measurement of flow velocities within a hydrocyclone using laser doppler anemometry. Technical report FTN/X/82, AEA, Power Fluidics, BNFL

Ji L, Paul P, Shanbhag B, Dixon I, Kuang S, He L (2023) Emerging application of hydrocyclone in biotechnology and food processing. Sep Purif Technol 309:122992. https://doi.org/10.1016/j.seppur.2022.122992

Jin R, Chen W, Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. J Stat Plan Inference 134:268–287. https://doi.org/10.1016/j.jspi.2004.02.014

Jones D, Schonlau M, Welch W (1998) Efficient global optimization of expensive black-box functions. J Glob Optim 13:455–492. https://doi.org/10.1023/A:1008306431147

Juretic F (2020) cfMesh, advanced meshing tool (version from OpenFOAM 2012). https://cfMesh.com. Accessed 14 May 2024

Kasper R (2017) Particle simulation with OpenFOAM: introduction, fundamentals and applications, German OpenFOAM User Meeting 2017 (GOFUN2017). https://www.foamacademy.com/wp-content/uploads/2016/11/GOFUN2017_ParticleSimulations_slides.pdf. Accessed 14 May 2024

Launder BE, Reece GJ, Rodi W (1975) Progress in the development of a Reynolds-stress turbulence closure. J Fluid Mech 68(3):537–566. https://doi.org/10.1017/S0022112075001814

Le Digabel S, Wild S (2023) A taxonomy of constraints in black-box simulation-based optimization. Optim Eng. https://doi.org/10.1007/s11081-023-09839-3

Li D, Wei Y, Marchisio D (2020) QEEFoam: a Quasi-Eulerian-Eulerian model for polydisperse turbulent gas-liquid flows. Implementation in OpenFOAM, verification and validation. Int J Multiph Flow 136:103544

Li W, Huang Z, Li G, Ye C (2022) Effects of different cylinder roof structures on the vortex of cyclone separators. Sep Purif Technol 296:121370. https://doi.org/10.1016/j.seppur.2022.121370

Menter F, Kuntz M, Langtry RB (2003) Ten years of industrial experience with the SST turbulence model. Turbul Heat Mass Transf 4:625–632

Montavon C, Grotjans H, Hamill I, Phillips H, Jones I (2000) Mathematical modelling and experimental validation of flow in a cyclone. In: BHR conference on cyclone technologies, Warwick

Pandey S, Saha I, Prakash O, Mukherjee T, Iqbal J, Roy A, Wasilewski M, Brar L (2022) CFD investigations of cyclone separators with different cone heights and shapes. Appl Sci 12:4904. https://doi.org/10.3390/app12104904

Pretsch L, Arsenyev I, Czech C, Duddeck F (2023) Interdisciplinary design optimization of compressor blades combining low- and high-fidelity models. Struct Multidiscip Optim. https://doi.org/10.1007/s00158-023-03516-w

Rahat AAM, Everson RM, Fieldsend JE (2017) Alternative infill strategies for expensive multi-objective optimisation. In: GECCO 2017—proceedings of the 2017 genetic and evolutionary computation conference, pp 873–880. https://doi.org/10.1145/3071178.3071276

Roberts AP, Jarman DS, Fieldsend JE, Tabor GR (2024) A separator for separating solids from a solid-liquid mixture. UK patent application number 2400481.4, European patent application number 24151725.9

Shi B, Ding L, Liu Y, Yang J, Song S, Wu H, Wang W, Gong J (2018) Hydrate slurry flow property in w/o emulsion systems. R Soc Chem Adv 8:11436–11445. https://doi.org/10.1039/C7RA13495A

Shur ML, Strelets MK, Travin AK, Spalart PR (2000) Turbulence modeling in rotating and curved channels: assessing the Spalart-Shur correction. AIAA J 38(5):784–792. https://doi.org/10.2514/2.1058

Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst 4:2951–2959

Sommerfeld M, van Wachem B, Oliemans R (2008) Best practice guidelines for computational fluid dynamics of dispersed multiphase flows. In: European Research Community on Flow, Turbulence and Combustion (ERCOFTAC 2008). Version 1

Sun X, Kim S, Yang SD, Kim HS, Yoon JY (2017) Multi-objective optimization of a Stairmand cyclone separator using response surface methodology and computational fluid dynamics. Powder Technol 320:51–65. https://doi.org/10.1016/j.powtec.2017.06.065

Sun X, Yoon JY (2018) Multi-objective optimization of a gas cyclone separator using genetic algorithm and computational fluid dynamics. Powder Technol 325:347–360. https://doi.org/10.1016/j.powtec.2017.11.012

Tran A, Eldred M, McCann S, Wang Y (2021) srMO-BO-3GP: a sequential regularized multi-objective constrained Bayesian optimization for design applications. https://doi.org/10.48550/arXiv.2007.03502. Accessed 14 May 2024

Weller HG, Tabor G, Jasak H, Fureby C (1998) A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput Phys 12(6):620–631. https://doi.org/10.1063/1.168744

Yao Y, Huang Z, Zhou T, Li J, Cheng L, Zhang M, Yang H, Lyu J (2022) Double-eccentric design for the vortex finder of a cyclone separator. Ind Eng Chem Res. https://doi.org/10.1021/acs.iecr.2c02054

Zwillinger D, Kokoska S (2000) CRC standard probability and statistics tables and formulae. Chapman and Hall, New York

Authors and Affiliations

A. P. Roberts[1] · A. A. M. Rahat[2] · D. S. Jarman[3] · J. E. Fieldsend[1] · G. R. Tabor[1]

✉ A. P. Roberts
a.p.roberts@exeter.ac.uk

A. A. M. Rahat
a.a.m.rahat@swansea.ac.uk

D. S. Jarman
djarman@hydro-int.com

J. E. Fieldsend
j.e.fieldsend@exeter.ac.uk

G. R. Tabor
g.r.tabor@exeter.ac.uk

1 Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK

2 Department of Computer Science, Swansea University, Swansea, UK

3 Hydro International UK Ltd, Clevedon, UK