Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Hybrid Modelling of Segmented Flow Extraction Process for Digital Twin Development in Critical Metals Recovery

Arun Pankajakshan, Konstantinos Katsoulas, Malik Olasinde, Cong Chao, Eric S. Fraga, Panagiota Angeli and Federico Galvanin*

Department of Chemical Engineering, University College London, Torrington Place, WC1E 7JE, London, United Kingdom
* Corresponding Author: f.galvanin@ucl.ac.uk.

ABSTRACT

Critical metals are indispensable in renewable, low-carbon, and hydrogen technologies due to their unique catalytic and electrochemical properties. They are primarily sourced through mining, which is associated with significant environmental impacts and geopolitical risks due to the uneven global distribution of ore deposits. As a result, efficient recovery of these metals from secondary sources such as electronic waste has become increasingly important. In this context, liquid-liquid extraction (LLE) has emerged as a promising separation technique due to its high selectivity and scalability. The development of intensified, continuous-flow LLE in small channels offers further advantages in terms of mass transfer efficiency, solvent utilization, and process sustainability, making it an attractive approach for the recovery of critical metals. A flow pattern known as segmented flow further enhances mass transfer in LLE in small channels. This work presents a hybrid modelling approach for developing a predictive model of a segmented flow LLE process, intended for digital twin implementation in critical metals recovery. Within the hybrid modelling framework, mass transfer is modelled using a lumped approach, which allows to treat mass transfer independently from flow hydrodynamics. Further, hydrodynamic and mass transfer models are developed in parallel using Gaussian process (GP)-based active learning (AL) and model-based design of experiments (MBDoE), respectively. Prior knowledge of flow regimes is used in developing the hydrodynamic model. The method was tested in an *in silico* case study and shown to efficiently develop reliable models for segmented flow extraction in small channels.

Keywords: hybrid modelling, active learning, model-based design of experiments, critical metals, extraction, segmented flow

INTRODUCTION

Critical metals are essential to the functioning of modern technological systems, with applications ranging from lithium-ion batteries in smartphones and laptops to rare earth elements (REEs) used in permanent magnets for wind turbines and electric vehicles. These vital metals are used in many sectors including communications, renewable energy, transport and life sciences (Department for Business & Trade, 2026). The primary sources of critical metals are economically viable ore deposits, whose distribution across the globe is greatly uneven. In addition, the process of mining to obtain critical metals from their ore deposits is a highly resource-intensive process and can cause enormous damage to the environment. As a result, the supply of critical metals is increasingly vulnerable due to geopolitical, environmental and economic constraints. This risk in supply chain has pushed the recovery of critical metals from secondary sources such as electronic waste (e-waste) (Sun et al., 2017).

The recovery of critical metals from e-waste usually proceeds through pyrometallurgical or hydrometallurgical routes. Hydrometallurgical routes are preferred due to their flexibility, high recovery rate and sustainable nature. Hydrometallurgical processes essentially involve dissolving e-waste in acid solutions followed by selective separation of critical metals via liquid-liquid extraction (LLE). LLE involves contacting two immiscible liquids to selectively transfer solutes between them. Intensified, continuous flow LLE in small channels offers further

advantages such as enhanced mass transfer, less material consumption, precise control of operating conditions, real-time monitoring and ease of automation and digitization. Based on the way in which the immiscible liquids are contacted, in small channels, LLE can proceed through different flow patterns. One of the possible flow patterns is segmented flow which is a two-phase flow pattern in which one immiscible liquid forms discrete segments within a continuous phase of the other liquid (Garstecki et al., 2006; Mac Giolla Eain et al., 2013). Segmented flow is particularly attractive due its regular hydrodynamics and enhanced mass transfer, supporting the development of predictive models and advanced process control. Consequently, implementing digital twins of segmented flow LLE in small channels can accelerate the development of critical metals recovery from e-waste. A digital twin is a software model (Ljung, 1999) of a physical system which interacts with the physical system continuously and acts as a reliable virtual entity of the real world described by the physical system. Compared to the traditional simulation models, digital twins are adaptive and responsive real time simulations enabled via fast computation, advanced sensing and automation. Digital twins are becoming increasingly important to automate decision making related to productivity, cost, energy consumption, and sustainability for complicated systems. The core of such digital twins are mathematical models of physical systems, which are instrumental for simulation and prediction.

The aim of this study is in developing a digital twin model of a segmented flow LLE system for recovering critical metals from e-waste. The geometry of segmented flow pattern and the resulting hydrodynamic characteristics strongly influence the mass transfer between the two liquid phases. Consequently, modelling segmented flow requires to develop both hydrodynamic and mass transfer models.

Hydrodynamic modelling of segmented flow in small channels has traditionally relied on a combination of mechanistic and semi-empirical approaches. Mechanistic models, based on computational fluid dynamics (CFD) simulations, offer detailed insight but are computationally expensive. Semi-empirical models, on the other hand, use fitted correlations for flow parameters and are typically calibrated for specific fluid systems and operating conditions. They often fail to reliably predict system behaviour under new operating conditions which limit their application as digital twins. In this work we combine two powerful modelling techniques – probabilistic modelling via Gaussian processes (GPs) and model-based design of experiments (MBDoE) (Geremia et al., 2026) methods to develop a hybrid modelling framework for the segmented flow. In the proposed framework, hydrodynamic models are developed using GP and active learning (AL) (Lewis & Catlett, 1994), while MBDoE methods are employed to precisely estimate the parameters of the mass transfer model. The framework is tested on a simulated case study and is further shown to be effective in developing hybrid models of segmented flow LLE processes.

2. BACKGROUND AND METHODOLOGY

As shown in Fig. 1, segmented flow in small channels is characterised by repeated volumes of one phase (dispersed phase) in a continuous volume formed by the other phase (continuous phase). The dispersed phase is called plug and the continuous phase volume between two plugs is called slug. A single plug and slug constitute a unit cell. The size of plug, slug and unit cell are characterised by their lengths L_p, L_s and L_u, respectively. Moreover, plugs are assumed to be cylindrical in shape with two hemispherical caps attached at both ends.

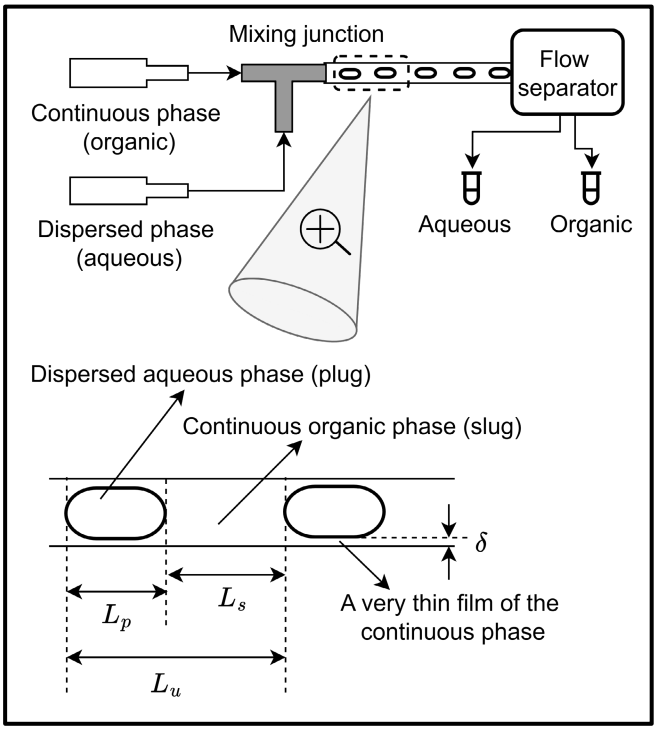

Figure 1. Schematic showing a segmented flow contactor set up, the segmented flow pattern and hydrodynamic characteristics of the segmented flow pattern.

In this work we consider a segmented flow LLE process in which the mass transfer by extraction happens from the dispersed aqueous phase to the continuous organic phase. Typically, the rate of mass transfer is affected by three main factors – the velocity and mixing patterns within each of the phases, the concentration driving forces and the interfacial area available for mass transfer. These three factors are influenced by several variables. Some of them for example, the design of the

mixing junctions (T-shaped junction or Y shaped junction, rectangular or circular cross section), the channel diameter and length and the properties of the contacted liquids are fixed for a specific system. While some variables such as the mixture velocity, the flow rate fraction of any of the two phases or the initial concentration of metals in the aqueous phase can be varied to study their effects on the plug-slug geometries as well as on the rate of mass transfer.

2.1 Fundamentals of modelling segmented flow extraction in small channels

Mathematical models of segmented flow LLE in small channels consist of a hydrodynamic model, a model for the pressure drop and a mass transfer model. Important hydrodynamic parameters related to segmented flow LLE are: a) primary flow geometries such as L_p, L_s and L_u, b) film thickness (δ), c) the dispersed phase holdup (ε_d) and d) the specific interfacial area (a). As shown in Fig. 1, a thin film of the continuous phase separates the slugs from the channel wall. The dispersed phase holdup is the volume fraction of the dispersed phase within one unit cell of the segmented flow:

$$\varepsilon_{\mathrm{d}} = \frac{\text{volume of dispersed phase in one unit cell}}{\text{total volume of the unit cell}} \tag{1}$$

The specific interfacial area is the area available for mass transfer per volume of the unit cell, and is defined as:

$$a = \frac{\text{total surface area of the dispersed phase}}{\text{total volume of the unit cell}} \tag{2}$$

The hydrodynamic modelling involves developing mathematical relationships between the inputs or the operating conditions to the primary flow geometries as well as to the film thickness. Typically, the relationship between inputs and film thickness is modelled using physics-based correlations such as Bretherton-type scaling laws, often extended to account for fluid properties and operating conditions (Mac Giolla Eain et al., 2013). Such correlations are often functions of dimensionless numbers such as the capillary number and the Weber number, depending on whether the flow is in the visco-capillary regime or in the visco-inertial regime (Mac Giolla Eain et al., 2013). Given L_p, L_s, L_u and δ, and assuming a plug geometry consisting of a cylindrical core with two hemispherical caps, expressions for ε_d and a can be derived using simple geometric relations. The rate of mass transfer between the dispersed phase and the continuous phase depends on the driving force, i.e., the metal concentration difference between the two phases, and the interfacial area available for mass transfer. Assuming an ideal plug flow behaviour (producing plugs of same dimension with no axial mixing) and negligible concentration gradient along the radial direction, the mass transfer along the channel length can be described by:

$$u_{\mathrm{mix}} \frac{\mathrm{d}C_{\mathrm{aq}}}{\mathrm{d}z} = -\frac{K_{\mathrm{aq}} a}{\varepsilon_{\mathrm{d}}} \left(C_{\mathrm{aq}}(z) - C_{\mathrm{aq}}^{\mathrm{eq}}\right) \tag{3}$$

In Eq. (3), $u_{mix}, K_{aq}, C_{aq}(z)$ and C_{aq}^{eq} are the mixture velocity, aqueous phase mass transfer coefficient, concentration of metal in the aqueous phase at a distance z from the channel inlet and the equilibrium concentration of metal, respectively. The mixture velocity is defined as the ratio of the total volumetric flow rate (Q_T) and the cross-sectional area of the channel with internal diameter d_{int}. The total volumetric flow rate is defined as the sum of volumetric flow rates of the continuous phase (Q_c) and the dispersed phase (Q_d). The equilibrium concentration is defined as:

$$C_{\mathrm{aq}}^{\mathrm{eq}} = \frac{C_{\mathrm{aq}}^{\mathrm{in}}}{D(Q_{\mathrm{d}}/Q_{\mathrm{c}})+1} \tag{4}$$

In Eq. (4), D, C_{aq}^{in}, Q_d and Q_c are the distribution coefficient, initial concentration of metal in the aqueous phase, volumetric flow rate of the dispersed phase and the volumetric flow rate of the continuous phase, respectively. The distribution coefficient is usually evaluated from batch equilibrium experiments. Pressure drop is a key parameter in designing segmented flow contactors as it provides crucial information regarding energy consumption. However, it becomes more significant in designing scale-out segmented flow systems (Garciadiego Ortega et al., 2018) where multiple channels are involved. For this reason, in this study, a pressure drop model is not considered.

2.2 A hybrid modelling framework for segmented flow extraction in small channels

The mass transfer model in the form provided in Eq. (3) imposes two challenges. The first is its direct dependency on hydrodynamic terms such as the specific interfacial area and the dispersed phase holdup. This prevents the independent development (estimation and validation) of the mass transfer model as the predictive quality of the mass transfer model strongly depends on the accurate evaluation of ε_{d} and a using the hydrodynamic model. Secondly, the aqueous phase mass transfer coefficient K_{aq} in Eq. (3) relates the mass flux through the interface to the concentration driving force. Consequently, when the interfacial area changes (i.e., the plug shape changes), for example due to variations in the flow rate, the value of mass transfer coefficient changes. For this reason, it is a local property that depends on the specific hydrodynamic conditions and cannot be uniquely estimated as a single model parameter over a range of operating conditions. As a solution to both the challenges, we lump the mass transfer coefficient, the specific interfacial area and the dispersed phase holdup into a single parameter k' given by:

$$k' = K_{\mathrm{aq}} a / \varepsilon_{\mathrm{d}} \tag{5}$$

The lumped parameter k' physically signifies mass transfer per unit interfacial area per unit volume of the dispersed phase. This parameter can be treated as constant (Garciadiego-Ortega et al., 2020) across a wide range of operating conditions that affect the primary plug-slug geometries, and subsequently the interfacial area and holdup. However, the parameter k' can be affected by the mixture velocity, which promotes internal circulation within the plugs and slugs and enhances mass transfer even without changing the plug geometry. Therefore, analogously to the Arrhenius equation describing the temperature dependency of reaction rate constants, we employ a simple scaling rule to account for velocity dependency of k', given by:

$$k' = k'_0\left(\frac{u_{\text{mix}}}{u_{\text{mix}}^{\text{ref}}}\right)^p \tag{6}$$

In Eq. (6), k'_0 is the lumped mass transfer coefficient corresponding to a reference mixture velocity u_{mix}^{ref}. The exponent p in Eq. (6) represents the combined sensitivity of the effective mass transfer rate to mixture velocity, accounting for velocity-induced changes in interfacial area, phase holdup, and local mass transfer. By substituting Eq. (5) and (6) in Eq. (3), the mass transfer model can be modified as:

$$u_{\text{mix}}\frac{\mathrm{d}C_{\text{aq}}}{\mathrm{d}z} = -k'_0\left(\frac{u_{\text{mix}}}{u_{\text{mix}}^{\text{ref}}}\right)^p\left(C_{\text{aq}}(z) - C_{\text{aq}}^{\text{eq}}\right) \tag{7}$$

Now, Eq. (7) contains two parameters k'_0 and p that can be uniquely estimated so that Eq. (7) holds true for a wide range of operating conditions. This approach allows the mass transfer model to be developed independently of the hydrodynamic model. We suggest the application of closed loop experimentation via optimal feedback loop (with minimum data points) to develop both hydrodynamic and mass transfer model. To this end, we employ GP modelling via AL for the development of hydrodynamic model and MBDoE for the mass transfer model development. Details of these methods are given below.

2.2.1 Gaussian processes and active learning for hydrodynamic model development

Consider a small channel segmented flow LLE system in which the aqueous phase forms the plug and the organic phase forms the slug. Let the system is affected by two inputs – mixture velocity and the continuous phase flow rate fraction (ϕ_c), which is defined as the ratio of continuous phase volumetric flow rate over the total volumetric flow rate. Now, an experiment i is identified by the input vector $\mathbf{u}_i = [u_{i1}, u_{i2}]$ and the outputs - $\mathbf{y}_{h_i} = \left[y_{h_{i1}}, y_{h_{i2}}\right]$ and y_{m_i}. Here, u_{i1} and u_{i2} respectively denote the values of mixture velocity and continuous phase flow rate fraction that define experiment i. Similarly, $y_{h_{i1}}, y_{h_{i2}}$ and y_{m_i} respectively denote the observed values of plug length [m], unit cell length [m] and concentration of metal [mol/L] in the dispersed aqueous phase at the outlet of the channel of length L_z. It is assumed that the observations $y_{h_{i1}}, y_{h_{i2}}$ and y_{m_i} are corrupted by independent and identically distributed Gaussian noise with zero mean and standard deviations of σ_1, σ_2 and σ_c, respectively.

Consider a dataset from n experiments consisting of inputs $\mathbf{u}_1, \dots, \mathbf{u}_{\text{n}}$ and outputs $\mathbf{y}_{h_1}, \dots, \mathbf{y}_{h_n}$ and $y_{m_1}, \dots, y_{m_n}$. The relationship between $\mathbf{u}$ and $\mathbf{y}_h$, which is the relationship between the inputs and the primary plug-slug geometries is usually modelled as empirical correlations validated using polynomial regression. Disadvantages of this approach are the lack of proper quantification of the uncertainty in model predictions and the lack of accurate predictions under new operating conditions, which were not used in calibration. As a better approach that guarantees accurate estimation of prediction uncertainties, and that captures non-linear system behaviour, we model the relationship between $\mathbf{u}$ and $\mathbf{y}_h$ using a multiple-output GP model (Bonilla et al., 2007).

In a multiple-output GP modelling framework, several related output variables are modelled together by assuming each output having an underlying hidden function called the latent function. Mathematically, the observed value of output l at input $\mathbf{u}_i$ can be represented as:

$$y_{h_{il}} = f_l(\mathbf{u}_i) + \epsilon_l, \epsilon_l \sim \mathcal{N}\left(0, \sigma_l^2\right) \tag{8}$$

In Eq. (8), f_l is the latent function for output l and ϵ_l is the error term. In the GP framework, the latent functions are treated as random functions drawn from a GP, meaning that for any set of inputs $\mathbf{u}_1, \dots, \mathbf{u}_n$, the latent function values follow a multivariate Gaussian distribution.

$$[f_{11}, \dots, f_{\text{n}1}, f_{12}, \dots, f_{\text{n}2}; \mathbf{u}_1, \dots, \mathbf{u}_{\text{n}}] \sim \mathcal{N}(\mathbf{m}, \mathbf{K}) \tag{9}$$

In Eq. (9), $\mathbf{m}$ and $\mathbf{K}$ represent the mean vector and the covariance matrix, respectively. The covariance between any two latent function values is given by:

$$Cov\left(f_l(\mathbf{u}), f_j(\mathbf{u}')\right) = \kappa_{lj}^f \kappa_u(\mathbf{u}, \mathbf{u}') \tag{10}$$

The RHS of Eq. (10) is called a kernel function. Essentially, it assigns covariance between the latent function values based on the covariance or similarities observed among inputs and outputs. According to Eq. (10), similarity between inputs is captured by $\kappa_u(\mathbf{u}, \mathbf{u}')$ whereas similarity between outputs is captured by κ^f. We employed the intrinsic coregionalization model (ICM) kernel proposed by Bonilla et al., (2007) for the product $\kappa_{lj}^f \kappa_u(\mathbf{u}, \mathbf{u}')$, with the radial basis function (RBF) kernel for $\kappa_u(\mathbf{u}, \mathbf{u}')$.

The GP model training involves estimating the optimal values of hyperparameters of the kernel function that minimises a loss function. We used the negative marginal log-likelihood as the loss function. Once trained with few points, the GP model can be used to predict the plug and unit length for any unseen input data point. Along with these predictions, the model also provides an estimate of

uncertainty associated with the predictions. This uncertainty will be largest in regions less explored, where the model is least confident. Therefore, selecting new experiments in the largest uncertainty region and retraining the model with those new data points helps to reduce the model prediction uncertainty over the experimental iterations. This iterative approach of training the model with informative data points is known as active learning (AL) (Lewis & Catlett, 1994). Mathematically, AL is an optimisation problem of the form:

$$\max_{\mathbf{u}_{\mathrm{lb}} \le \mathbf{u} \le \mathbf{u}_{\mathrm{ub}}} \psi\{\mathrm{Var}[f(\mathbf{u})]\} \tag{11}$$

where, $Var[f(\mathbf{u})]$ is the posterior predictive variance of the GP model at conditions defined by the vector $\mathbf{u}$ and ψ is a scalar function of the variance vector. We have chosen ψ as the ℓ^2-norm of the variance vector. The AL procedure was applied iteratively to develop a reliable GP model that maps the mixture velocity and continuous phase flow rate fraction to the plug and unit lengths.

2.2.2 Model-based design of experiments for mass transfer model development

The development of mass transfer model presented in Eq. (7) essentially requires precise estimation of the two model parameters lumped in the N_θ (here, 2) dimensional vector $\theta = [k_0', p]$. Let us revisit the dataset formed from n experiments. The mass transfer data consists of the set of inputs $\mathbf{u}_1, \ldots, \mathbf{u}_{\mathrm{n}}$ and outputs $y_{m_1}, \ldots, y_{m_n}$. Each input vector corresponds to an experimental condition specified by the values of mixture velocity and continuous phase flow rate fraction, while the output variables denote the metal concentration in the dispersed phase at the channel outlet. With the availability of such a dataset, a parameter estimation can be performed to obtain the best values of the model parameters that minimises the discrepancy between predicted and observed concentration values. We employed maximum likelihood method of parameter estimation to obtain the optimal parameter estimates $\hat{\theta}$.

The uncertainty in estimation of $\hat{\theta}$ can be quantified as the parameter covariance matrix $\mathbf{V}_\theta$, which in turn can be approximated as the inverse of observed Fisher information matrix $\mathbf{H}_\theta$ (Geremia et al., 2026).

$$\mathbf{H}_\theta = \left[\mathbf{V}_\theta^0\right]^{-1} + \sum_{\mathrm{i}=1}^{n} \mathbf{S}_{\mathrm{i}}^{\mathrm{T}} (\sigma_{\mathrm{c}}^2)^{-1} \mathbf{S}_{\mathrm{i}} \tag{12}$$

In Eq. (12), $\mathbf{S}_{\mathrm{i}}$ represents the parameter sensitivity matrix with dimensions equal to number of output variables times the number of model parameters. Since, we have the case with only one output variable, $\mathbf{S}_{\mathrm{i}}$ is a row vector whose elements represent the first derivatives of the output variable with respect to the model parameters, evaluated at the estimated parameter values and at the condition investigated in experiment i. The parameter covariance matrix $\mathbf{V}_\theta$ is evaluated as the inverse of $\mathbf{H}_\theta$. The statistical precision in which the parameters are estimated can be evaluated using a *t*-test. From the parameter covariance matrix, the test statistic *t*-value for the *t*-test can be computed for each parameter estimate as:

$$t_{\mathrm{i}} = \frac{\hat{\theta}_{\mathrm{i}}}{t_{N-N_\theta}(1-\alpha/2)\sqrt{\mathbf{V}_{\theta_{\mathrm{ii}}}}} \tag{13}$$

When the computed *t*-value for an estimate is greater than a reference *t*-value, $t_{ref} = t_{N-N_\theta}(1-\alpha/2)$, that parameter is identified as precisely estimated, otherwise the test fails and new experiments need to be designed to improve the statistical precision. This can be achieved by solving the MBDoE optimisation problem, which suggests new condition that minimises a scalar function ξ of the expected parameter covariance matrix $\hat{\mathbf{V}}_\theta$.

$$\min_{\mathbf{u}_{\mathrm{lb}} \le \mathbf{u} \le \mathbf{u}_{\mathrm{ub}}} \xi\left[\hat{\mathbf{V}}_\theta(\mathbf{u})\right] \tag{14}$$

Popular choices of ξ are the alphabetic design criteria with A-, D- and E-designs, respectively minimising the trace, determinant and maximum eigenvalue of the expected covariance matrix. Like AL, MBDoE can be also performed iteratively in closed loop until all the model parameters are precisely estimated.

2.3 Case study

A simulation study was designed to test the proposed hybrid modelling approach. For the simulation study, the extraction of uranium U(VI) by tributyl phosphate (TBP) (100 %) (Garciadiego-Ortega et al., 2020) in a channel of internal diameter 2 mm was selected as the case study. In their study, Garciadiego-Ortega et al. (2020) have developed the models provided in Eq. (15) and (16) to predict the dimensionless plug and slug lengths. They also validated these models over the range of operating conditions provided in Table 1.

Table 1: Experimental design space in terms of variable bounds.

Input	Bounds
Mixture velocity (cm/s)	[1.062, 4.24]
Continuous phase flow rate fraction (–)	[0.2, 0.5]

In this study, the models with the coefficients in Eq. 15 and 16 are assumed as the "true models" representing the hydrodynamic component of the segmented flow LLE system. The design space provided in Table 1 was chosen to be the design space in this simulation study.

$$\ln\left(\frac{L_{\mathrm{p}}}{d}\right) = -2.56 - 0.872\ln\left(\frac{Q_{\mathrm{c}}}{Q_{\mathrm{T}}}\right) - 0.36\ln\left(\frac{\mu_{\mathrm{d}}}{\mu_{\mathrm{c}}}\right) - 0.280\ln(\mathrm{Ca_c}) + 0.099\ln\left(\frac{\mathrm{Re_c}}{\mathrm{Ca_c}}\right) \tag{15}$$

$$\ln\left(1 - \frac{L_{\mathrm{p}}}{L_{\mathrm{u}}}\right) = -0.82 + 1.181\ln\left(\frac{Q_{\mathrm{c}}}{Q_{\mathrm{T}}}\right) - 0.099\ln\left(\frac{\mu_{\mathrm{d}}}{\mu_{\mathrm{c}}}\right) -$$

$$0.130\ln(\mathrm{Ca}_c) \quad (16)$$

In their study, Garciadiego-Ortega et al. (2020) have used the following correlation proposed by Mac Giolla Eain et al., (2013) to calculate the film thickness:

$$\frac{2\delta}{d_{\mathrm{int}}} = 0.35(\mathrm{Ca}_c)^{0.354}(\mathrm{We}_c)^{0.097} \quad (17)$$

According to Mac Giolla Eain et al., (2013), Eq. (17) is only valid within a specific range of operating conditions defined by the dimensionless numbers listed in Table 2. However, a subregion of the chosen design space in Table 1 resulted in values of the dimensionless numbers outside the feasible bounds provided in Table 2.

Table 2: Dimensionless numbers and their ranges over which the film thickness correlation in Eq. (17) is valid.

Dimensionless number	Description	Desired range
Capillary (Ca)	$\mathrm{Ca}_c = \mu_c u_{mix}/\gamma$	0.002 – 0.119
Weber (We)	$\mathrm{We}_c = \rho_c u_{mix}^2 d/\gamma$	0.047 – 0.697
Reynolds (Re)	$\mathrm{Re}_c = \rho_c u_{mix} d/\mu_c$	14.46 – 100.96

Therefore, in the simulation study, the AL was solved as a constrained optimization problem with the constraint $0.047 \leq \mathrm{We} \leq 0.697$. Based on the work by Garciadiego-Ortega et al., (2020), the true values of the parameters of the lumped mass transfer model - k_0' and p were chosen as 0.055 and 0.8, respectively with $u_{mix}^{ref} = 1.062$ cm/s. For *in silico* data generation, it is assumed that the plug length, unit length and concentration data are corrupted by independent and identically distributed Gaussian noise with zero mean and standard deviations of 0.0004 m, 0.0004 m and 0.001 mol/L, respectively. The physical properties of the fluids such as viscosity (μ), density (ρ) and interfacial tension (γ) required to calculate Eq. (15-17) and the dimensionless numbers in Table 2 as well as the value of the distribution coefficient D were obtained from Garciadiego-Ortega et al., (2020). The channel length L_z was assumed as 0.5 m.

RESULTS AND DISCUSSION

Initial experiments to calibrate both the hydrodynamic and mass transfer models were generated using Latin hypercube sampling (LHS) applied in the design space provided in Table 1. Four initial experiments were designed using the LHS method. The outputs (plug length, unit length and metal concentration in the dispersed phase at the channel outlet) corresponding to these experiments were obtained by simulating the true models and adding noise to them. The initial hydrodynamic dataset was used to train the GP model, and the initial mass transfer dataset was used to estimate parameters of the lumped mass transfer model. The results of parameter estimation of the mass transfer model using the initial dataset (i.e., at the end of 4 DoE experiments) are shown in Table 3.

Table 3: Parameter estimation results after DoE and AL. The reference *t*-values after DoE and AL are 2.91 and 1.71 respectively.

Parameter	Estimate	*t*-value	Evaluation
k_0'	0.060 ±0.03	**1.98**	End of DoE
p	0.704 ±0.49	**1.43**	
k_0'	0.058 ±0.005	10.39	End of AL
p	0.734 ±0.08	8.35	

As shown in Table 3, the parameter estimates obtained from the initial DoE exhibited relatively large uncertainty. The *t*-values of the parameter estimates were smaller than the reference *t*-value (2.91). At this stage, one can perform either AL, MBDoE or a multi-objective optimization using AL and MBDoE to design next set of experiments for developing the segmented flow model. In this work, we focussed on the development of hydrodynamic model. Thus, after the DoE, a campaign of AL experiments, consisting of 20 new experiments were designed iteratively to reduce the prediction uncertainty of the GP model. In this simulated experimental campaign, AL was solved as a constrained optimization problem with the constraint equation $0.047 \leq \mathrm{We} \leq 0.697$.

The experiments designed in the AL campaign along with the initial LHS experiments are shown in Fig. 2. As shown in the Fig. 2, in the constrained AL problem, a low velocity region appeared to be infeasible as it resulted in Re and We below their validated range. A physical significance of this observation is that the correlation for film thickness provided in Eq. (17) applies to the visco-inertial regions, where the velocity or inertia force start to have an influence on the film thickness. However, at low velocity, this influence is negligible with capillary and viscous forces dominating (Mac Giolla Eain et al., 2013). Hence, even the small weighting of the Weber number is not appropriate in calculating film thickness at low velocities. As a result, this low velocity region proved to be the infeasible region in the constrained AL problem. It can be also inferred from Fig. 2 that the constrained AL problem always resulted in feasible solutions, however, very close to the feasible boundary. This is because AL picks new experiments in the less explored regions.

The experiment numbers showing the exact iteration of experiments are provided as labels in Fig. 2. It is also understood from Fig. 2 that the AL campaign was successful in exploring the design space. Further evidence of the good performance of the AL campaign in reducing the prediction uncertainty of the GP model is

provided in Fig. 3.

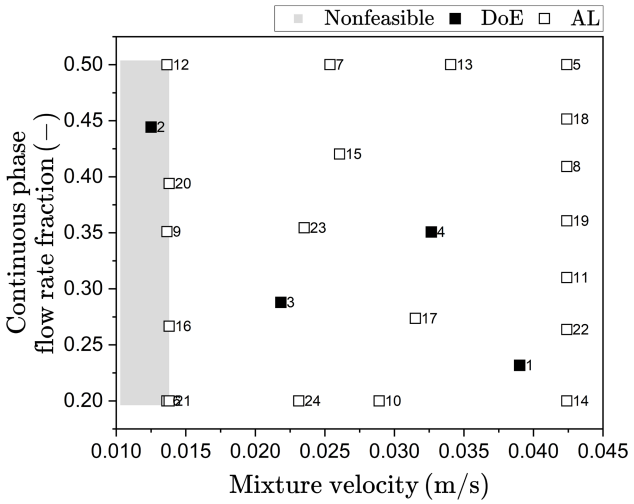

Figure 2. The simulated experimental campaign showing the DoE (filled squares) and AL (open squares) experiments. Labels adjacent to the scatter points denote the experiment number. The nonfeasible design space is shown using grey scatter points.

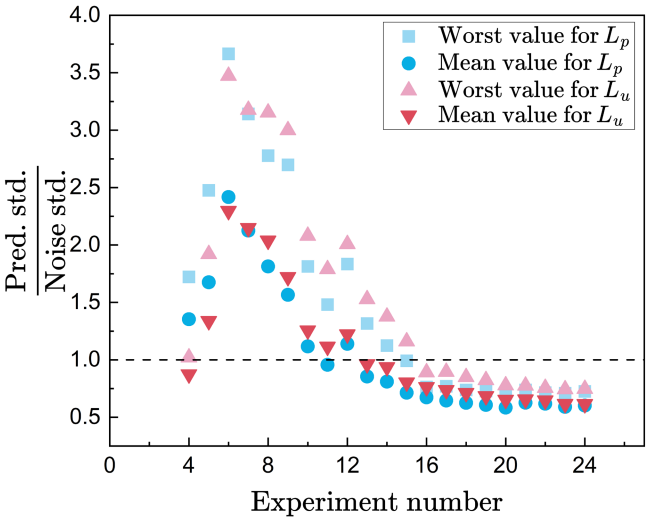

Figure 3. Convergence of the AL method in terms of the ratio of standard deviation of prediction to that of noise.

As shown in Fig. 3, for each output (L_p and L_u), the ratio of GP predictive standard deviation to the corresponding observation noise standard deviation was used as a performance indicator of the AL method. Specifically, the worst and mean values of this ratio evaluated

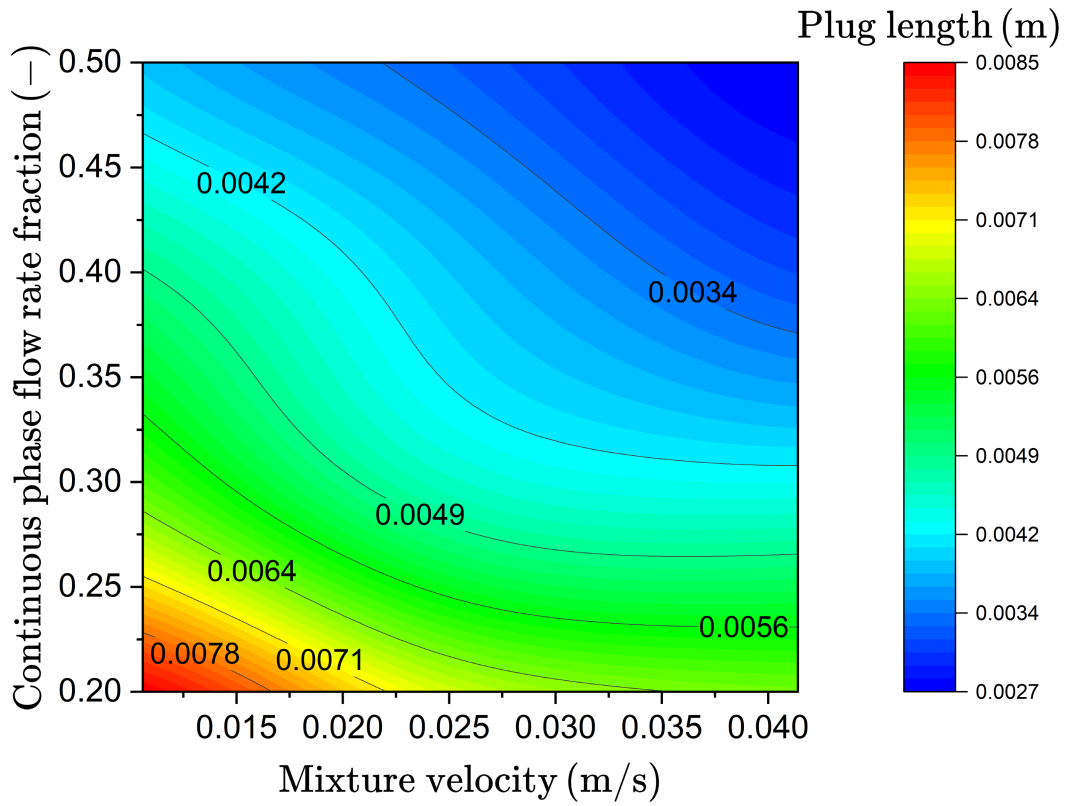

Figure 4. Prediction of plug length by the GP model developed using constrained AL method. The numbers on the contour lines represent the predicted values of plug length.

posteriori (after each design and execution) across the design space were used as the performance indicators. As expected, the ratio decreased gradually as new informative training point was picked in each iteration of the AL method. Ideally, one can stop performing the AL experimental campaign when the ratio of the standard deviations drops below 1 (marked by the dashed line in Fig. 3), indicating that the uncertainty in predicting L_p and L_u by the GP model can be fully explained by the uncertainty in the random nature of measurements.

Besides the prediction uncertainty, the prediction behaviour of the final GP model (after 20 AL experiments) developed through the constrained AL method is illustrated in Fig. 4.

As shown in Fig. 4, the plug length increases with decreasing mixture velocity and continuous phase flow rate fraction. A decrease in the continuous phase flow fraction increases the proportion of the dispersed phase in the channel, resulting in larger plugs. Also, as the mixture velocity decreases, the drag force which helps detach the plugs at the inlet decreases, again producing longer plugs (Garstecki et al., 2006). The behaviour of the GP model is quite appealing as its predictions of the plug length align well with known physical phenomena affecting plug formation and its size.

CONCLUSION

A hybrid modelling approach has been proposed to develop a digital twin model of a segmented flow LLE process to recover critical metals. In the proposed method, a lumped mass transfer model was developed to decouple the mass transfer and hydrodynamic models. Furthermore, the proposed approach enables the closed-loop development of both hydrodynamic and mass transfer models via GP-based AL and MBDoE, respectively. When prior constraints regarding flow regimes or hydrodynamic characteristics are available, this knowledge can be incorporated to guide the development of the GP model describing hydrodynamic features such as plug and slug lengths. Although the approach decouples the hydrodynamic and mass transfer components of the segmented flow model, it allows both models to be developed simultaneously through multi-objective optimization of AL and MBDoE. The approach was tested on an *in silico* case study and proved to be very efficient in developing a combined data-driven and knowledge-based model of segmented flow LLE in small channels.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding received from the UK Engineering and Physical Sciences Research Council (EPSRC) through the grant EP/Z533798/1, which supported this work.

AUTHOR IDENTIFIERS

Author ORCIDs:
Pankajakshan A: 0000-0002-3365-2668
Katsoulas K: 0009-0004-9833-0127
Olasinde M: 0000-0001-6001-1404
Fraga ES: 0000-0002-5732-6082
Angeli P: 0000-0001-9575-4259
Galvanin F: 0000-0002-3296-1581

NOMENCLATURE

a	Specific interfacial area
C_{aq}	Concentration of metal in the aq. phase
D	Distribution coefficient
d_{int}	Channel internal diameter
f	Latent function
K_{aq}	Aqueous phase mass transfer coefficient
k'	Lumped mass transfer coefficient
L_p, L_s, L_u	Plug, slug and unit cell length
L_z	Channel length
Q	Volumetric flow rate
u	Input variable
u_{mix}	Mixture velocity
y	Output or response variable
z	Axial coordinate
δ	Film thickness
ε_d	Dispersed phase holdup
κ^f	Kernel function for cross-covariance
κ_u	Kernel function for input covariance
σ	Standard deviation
$\mathbf{H}_\theta$	Observed Fisher information matrix
$\mathbf{K}$	Covariance matrix of the GP kernel
$\mathbf{m}$	Mean vector of the GP kernel
$\mathbf{S}$	Parameter sensitivity matrix
$\mathbf{V}_\theta$	Parameter covariance matrix
θ	Model parameter vector

REFERENCES

1. Williams O. A switched gaussian process for estimating disparity and segmentation in binocular stereo. Advances in Neural Information Processing Systems 19 :1497-1504 (2007). https://doi.org/10.7551/mitpress/7503.003.0192
2. Department for Business & Trade. (2026, January 23). *Vision 2035: Critical Minerals Strategy*. https://www.gov.uk/government/publications/uk-critical-minerals-strategy/vision-2035-critical-minerals-strategy
3. Garciadiego Ortega E, Tsaoulidis D, Angeli P. Predictive model for the scale-out of small channel

two-phase flow contactors. Chemical Engineering Journal 351:589-602 (2018). https://doi.org/10.1016/j.cej.2018.06.020
4. Garciadiego-Ortega E, Tsaoulidis D, Pineda M, Fraga ES, Angeli P. Hydrodynamics and mass transfer in segmented flow small channel contactors for uranium extraction. Chemical Engineering and Processing - Process Intensification 153:107921 (2020). https://doi.org/10.1016/j.cep.2020.107921
5. Garstecki P, Fuerstman MJ, Stone HA, Whitesides GM. Formation of droplets and bubbles in a microfluidic t-junction—scaling and mechanism of break-up. Lab Chip 6:437 (2006). https://doi.org/10.1039/b510841a
6. Geremia M, Macchietto S, Bezzo F. A review on model-based design of experiments for parameter precision – open challenges, trends and future perspectives. Chemical Engineering Science 319:122347 (2026). https://doi.org/10.1016/j.ces.2025.122347
7. Lewis DD, Catlett J. Heterogeneous uncertainty sampling for supervised learning. Machine Learning Proceedings 1994 :148-156 (1994). https://doi.org/10.1016/b978-1-55860-335-6.50026-x
8. Ljung, L. (1999). *System Identification: Theory for the User* (T. Kailath, Ed.; Second Edition). Prentice Hall PTR.
9. Mac Giolla Eain M, Egan V, Punch J. Film thickness measurements in liquid–liquid slug flow regimes. International Journal of Heat and Fluid Flow 44:515-523 (2013). https://doi.org/10.1016/j.ijheatfluidflow.2013.08.009
10. Sun Z, Cao H, Xiao Y, Sietsma J, Jin W, Agterhuis H, Yang Y. Toward sustainability for recovery of critical metals from electronic waste: the hydrochemistry processes. ACS Sustainable Chem. Eng. 5:21-40 (2016). https://doi.org/10.1021/acssuschemeng.6b00841