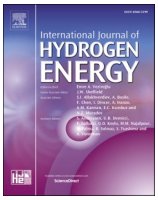
International Journal of Hydrogen Energy 255 (2026) 156385



Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

journal homepage: www.elsevier.com/locate/he





A similarity-based CFD–machine-learning framework for the design and scale-up of fluidized-bed reactors: application on biomass fast pyrolysis

M. Baldelli *, L. Bartolucci, S. Cordiner, P. Mele, V. Mulone

Department of Industrial Engineering, Tor Vergata University of Rome, Via del Politecnico 1, Rome, 00133, Italy

ARTICLE INFO

Keywords:
Machine learning
CFD
Fluidized-bed reactor
Fast pyrolysis
Scale-up
Reduced order model

ABSTRACT

The design and scale-up of fluidized-bed thermochemical reactors are challenging due to the strong coupling between hydrodynamics, heat transfer, and reaction kinetics, as well as the limited transferability of models trained at a single scale. This work proposes a physics-informed CFD–machine learning framework for reactor design and similarity-based scale-up using reduced-order surrogate models. A structured CFD campaign generates a consistent dataset across reactor sizes and operating conditions. Starting from the governing equations, a dimensionless feature space is derived through similarity analysis and used to train Gaussian Process models that predict thermo-fluid reactor descriptors and lumped product yields. Feature relevance is then exploited to support constrained design and to relax full similarity requirements in scale-up applications. Demonstrated on cellulose fast pyrolysis in a fluidized-bed reactor, the methodology improves the interpretability and consistency of surrogate predictions across multiple scales and is formulated to be extendable to other gas-solid thermochemical system.

1. Introduction

Thermochemical conversion is increasingly recognized as a viable route for valorizing biogenic residues and waste-derived streams into energy carriers and fuels, contributing to decarbonization and improved resource efficiency [1,2]. Among the available pathways, pyrolysis remains one of the most established processes for converting solid feedstocks into condensable liquids, non-condensable gases, and char [3,4].

In addition to the production of liquid fuels and chemical intermediates, thermochemical conversion of biomass is increasingly investigated as a route for sustainable hydrogen production [5,6]. Biomass pyrolysis can serve as an intermediate step in hydrogen production pathways, where the pyrolysis vapors or bio-oil are subsequently upgraded through catalytic steam reforming or gasification to generate hydrogen-rich syngas [7,8]. Despite sustained progress at laboratory and pilot scale, the translation of pyrolysis technologies to industrially relevant throughputs remains challenging [9]. In particular, scale-up must preserve the coupled hydrodynamic, thermal, and reactive environments that govern devolatilization and secondary reactions, while accommodating constraints in heat supply, residence-time distributions, and solids handling that intensify with reactor size [10,11].

Fluidized-bed reactors (FBRs) are widely adopted for pyrolysis due to their strong heat and mass transfer, vigorous mixing, and near-isothermal behavior under appropriate operating conditions [12,13]. These features make FBRs attractive for scale-up but also render their behavior highly sensitive to changes in geometry and operating conditions [14]. Variations in reactor diameter, distributor design, and superficial gas velocity can significantly alter bubbling regimes, solids circulation, and gas–solid contacting, with direct consequences for heating rates, vapor residence times, and product selectivity [2]. As a result, geometric scale-up alone rarely ensures similarity of the governing physics, and design rules derived at one scale may fail when extrapolated to another [15,16].

From a design perspective, reliable scale-up requires quantitative understanding of how hydrodynamics, heat transfer, and reaction kinetics interact in both the dense bed and the freeboard. Key observables include temperature uniformity, pressure-drop dynamics, residence times, and indicators of unstable fluidization [11,16,17]. While experimental campaigns remain essential, they are costly to extend across the multidimensional design space defined by reactor size, bed inventory, gas velocity, and thermal boundary conditions. Consequently, modeling approaches are increasingly used to complement experiments and

This article is part of a special issue entitled: EFCHC25 conference (Iulianelli) published in International Journal of Hydrogen Energy.

* Corresponding author.
E-mail address: matteo.baldelli@uniroma2.it (M. Baldelli).

https://doi.org/10.1016/j.ijhydene.2026.156385
Received 23 March 2026; Received in revised form 8 June 2026; Accepted 1 July 2026
Available online 3 July 2026

0360-3199/© 2026 The Authors. Published by Elsevier Ltd on behalf of Hydrogen Energy Publications LLC. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

reduce the number of full-scale trials.

Computational Fluid Dynamics (CFD) has become a cornerstone for analyzing gas–solid hydrodynamics and heat transfer in fluidized beds, providing spatially resolved insight into phase distribution, bubbling behavior, and thermal fields [18,19]. However, the high computational cost of CFD limits its direct application to broad parametric studies and systematic scale-up analyses [20,21]. In parallel, machine-learning (ML) techniques have been widely explored to construct fast surrogate models for predicting reactor performance once trained on experimental or numerical data [22–24]. Recent studies have demonstrated the potential of ML for a broader range of applications in reacting multiphase systems, including the development of data-driven closure models, prediction of transport phenomena, reaction kinetics modeling, product yield estimation, and optimization of reactor operating conditions [25–27]. By leveraging large experimental and CFD-generated datasets, ML methods provide a complementary framework for extracting hidden relationships between hydrodynamic, thermal, and reactive variables, thereby accelerating the analysis and design of complex thermochemical conversion processes [28]. While such models can accurately capture nonlinear relationships within the sampled domain, their predictive capability often deteriorates when applied across reactor geometries or scales not represented in the training data [29].

This limitation motivates integrated CFD–ML strategies, in which CFD is used to generate consistent datasets across a wide design space and ML is employed to construct reduced-order models suitable for rapid design exploration [30–32]. For fluidized-bed pyrolysis, however, there remains a need for surrogate frameworks explicitly formulated around hydrodynamic and reactive similarity, so that predictions retain physical meaning when extrapolating across scales rather than merely interpolating within a single geometry [33,34].

The work presented here addresses whether a CFD-generated, physics-informed surrogate framework can be formulated to support reactor design and similarity-based scale-up of fluidized-bed thermochemical systems, while preserving the dominant hydrodynamic, thermal, and reactive regimes across changes in size and operating conditions. Specifically, it investigates whether a reduced-order representation based on dimensionless similarity parameters can improve the interpretability and transferability of machine-learning surrogates beyond the scales directly sampled in the training data.

The main contribution of this work is methodological. A structured CFD–machine-learning framework is proposed in which (i) governing equations and boundary conditions are systematically transformed into a dimensionless feature space; (ii) Gaussian-Process (GP) surrogate models are trained on these similarity parameters to predict both thermo-fluid reactor descriptors and lumped product yields; and (iii) feature relevance is explicitly exploited to relax full similarity constraints and enable similarity-aware design and scale-up. Biomass fast pyrolysis in a fluidized-bed reactor is used as a representative case study to demonstrate the approach. While the application focuses on fast pyrolysis, the methodology is formulated independently of the specific reaction network and is intended as a general CFD-informed surrogate modeling strategy for gas–solid thermochemical reactors.

2. Materials and methods

In this section, the proposed data-driven methodology is outlined, along with the associated experimental and numerical setups. First, the experimental procedure is detailed, and then the CFD simulation campaign adopted to generate the numerical dataset is introduced. Finally, the data-driven framework is presented, including the definition of the input features and the mathematical basis of the optimization algorithms employed for reactor design and scale-up.

2.1. Experimental setup

Experiments were conducted in a laboratory-scale bubbling fluidized-bed reactor operated under inert conditions and validated at three pyrolysis temperatures (400–500 °C). The setup is used to validate the CFD framework and is not the focus of the study. For the sake of completeness, a brief description of the experimental device and test procedure is provided.

2.1.1. Materials

Microcrystalline cellulose spherical pellets (CELLETS® 700–1000 μm), supplied by Pharmatrans Sanaq AG (CH), were used as feedstock for the pyrolysis process. The bed material consisted of spherical quartz sand (177–250 μm; bulk density 1450 kg/m^3) supplied by Valli Granulati S.r.l. Acetone and hexane (HPLC grade, 99%) used for liquid handling and extraction were purchased from Merck (Darmstadt, Germany).

2.1.2. Reactor setup

Experiments were conducted in a laboratory-scale bubbling fluidized-bed reactor developed at the SCERG laboratories of the Tor Vergata University of Rome (Fig. 1). The reactor operates in semi-continuous mode (up to 500 g h^{-1}) and consists of a vertical AISI 304 stainless-steel tube (42 mm internal diameter, 300 mm height), equipped with a porous distributor plate. Temperature was monitored along the reactor axis using K-type thermocouples, and nitrogen was used as fluidizing gas, with flow rate controlled via a mass flow controller. Biomass feeding was ensured by a screw feeder, while heating was provided by a multi-zone electrical furnace. The pressure drop across the bed was continuously monitored. Downstream of the reactor, a cyclone separated char, followed by a condensation system for liquid products and a series of cold traps for further vapor recovery. The remaining non-condensable gases were finally analyzed using an online micro-GC system. For further details, the interested reader is referred to Supplementary Note S1 and to Ref. [35].

2.1.3. Experimental procedure

Before each run, the reactor is heated to the target operating temperature and nitrogen is flowed to maintain an inert atmosphere. The fluidizing gas flow rate is set to ensure three times the minimum fluidization velocity, 3 u_{mf}, where u_{mf} is estimated at reaction temperature using the semi-empirical Ergun correlation [36]. The experimental campaign and the corresponding validation of the CFD framework were conducted at three set-point temperatures (400, 450, and 500 °C). For each operating condition, product yields were determined after processing 500 g of biomass. Bio-oil and biochar yields were measured gravimetrically, while the non-condensable gas yield was quantified through online gas analysis. To maintain consistency with the numerical results, the yields were normalized to unity by redistributing the total measured weight loss proportionally between biochar and bio-oil based on their respective yields.

2.1.4. Analytical methods

The semi-quantitative chemical composition of cellulose pyrolysis oil was analyzed using a Shimadzu GC/MS system (QP2010SE) equipped with an Equity® 5 capillary column (30 m × 0.25 mm × 0.25 μm). Additional details about the analytical protocol can be found here [4,37]. Before analysis, the single-phase bio-oil obtained from the 500 °C runs was extracted using a protocol based on EPA Method 3546. 2.5 g of bio-oil were dissolved in 12.5 mL of acetone and 12.5 mL of hexane, mixed thoroughly, and left in a separating funnel overnight. Although acetone and hexane are fully miscible, the addition of pyrolysis oil caused phase separation into two distinct layers. After liquid–liquid extraction, the phases were filtered through a 0.45 μm PTFE syringe filter, and 0.5 mL of each extract was diluted with 4.5 mL of acetone. In this study, only the acetone-extracted phase was considered for comparison with the numerical results, and only the species included in the numerical model were evaluated. The full list of identified compounds is provided in Table 1. However, for methodology validation, only a subset

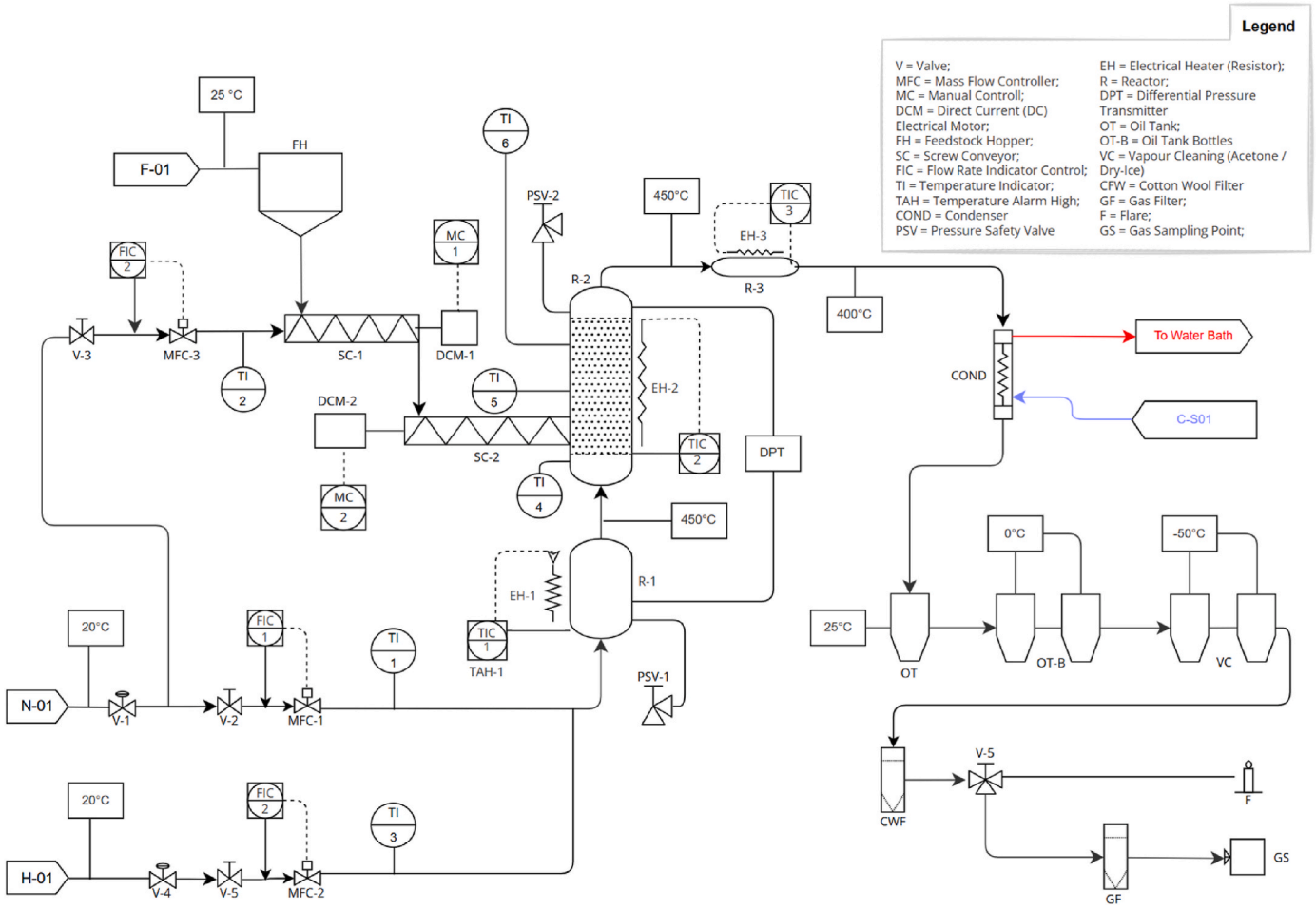


Fig. 1. P&ID of the FBR experimental set-up.

Table 1
Bio-oil composition.

Compound name	Relative abundance (Peak Area %)
Levoglucosan (LVG)	59.7
1,4:3,6-Dianhydro-.alpha.-D-glucopyranose	8.8
1,6-Anhydro-.beta.-D-glucofuranose	4.8
Furfural (FF)	3.8
5-Hydrxoymethylfurfural (5-HMF)	2.6
Furfural, 5-methyl-	1.2
H-Pyran-4-one, 3,5-dihydroxy-2-methyl-	1.1
Levoglucosenone	1.1
2(3H)-Furanone	1.1
Others	15.7

of the identified compounds was analyzed.

2.2. CFD simulation setup

This section describes the CFD simulation setup and the design of the numerical campaign. The numerical framework is first introduced, while the full governing equations and constitutive closures are included in the Supplementary Information for brevity. The computational reactor domain and boundary conditions are then described, followed by an overview of the entire numerical campaign.

2.2.1. Numerical framework

The CFD simulations were conducted using the Euler–Euler reacting multiphase framework available in OpenFOAM v2506 [38], adopted here to describe the coupled gas–solid hydrodynamics, heat transfer, and chemical conversion in the fluidized-bed reactor. In this approach, gas and solid phases are modeled as interpenetrating continua, while biomass conversion is represented through the detailed pyrolysis mechanism proposed by Ranzi et al. [39,40].To keep the main manuscript focused on the proposed methodology, the full governing equations, closure relations, and numerical details are provided in the Supplementary Note S2.

2.2.2. Domain and boundary condition definition

In the proposed numerical setup, the FBR is simulated in a 2D framework using a rectangular computational domain with length H and width D, representing the reactor's height and diameter, respectively, see Fig. 2(a). At the start of each simulation, the bed of sand is initialized at the bottom of the domain, occupying a height of H_{bed}. Fluidization is achieved by injecting nitrogen from the lower boundary at a superficial velocity u_0. The reactor is fed with a continuous cellulose mass flow rate, $\dot{m}_b$, through a lateral inlet located above the initial sand bed surface, while the outlet boundary at a constant pressure p_0 is placed below the top of the reactor. Reactor walls are maintained at a uniform temperature T_{pyro}, equal to the selected pyrolysis set point, and no-slip conditions are imposed along the reactor walls. The grid resolution is determined through a mesh-independence study balancing accuracy and computational cost. A set of uniform meshes of a domain with H equal to 0.7 m and D equal to 0.08 m, with different cell sizes, was evaluated. The variation of selected quantities, such as height-averaged gas and solid volume fraction, was found to be negligible for cells with side lengths smaller than 1.5×10^{-3} m, as shown in Fig. 2(b).

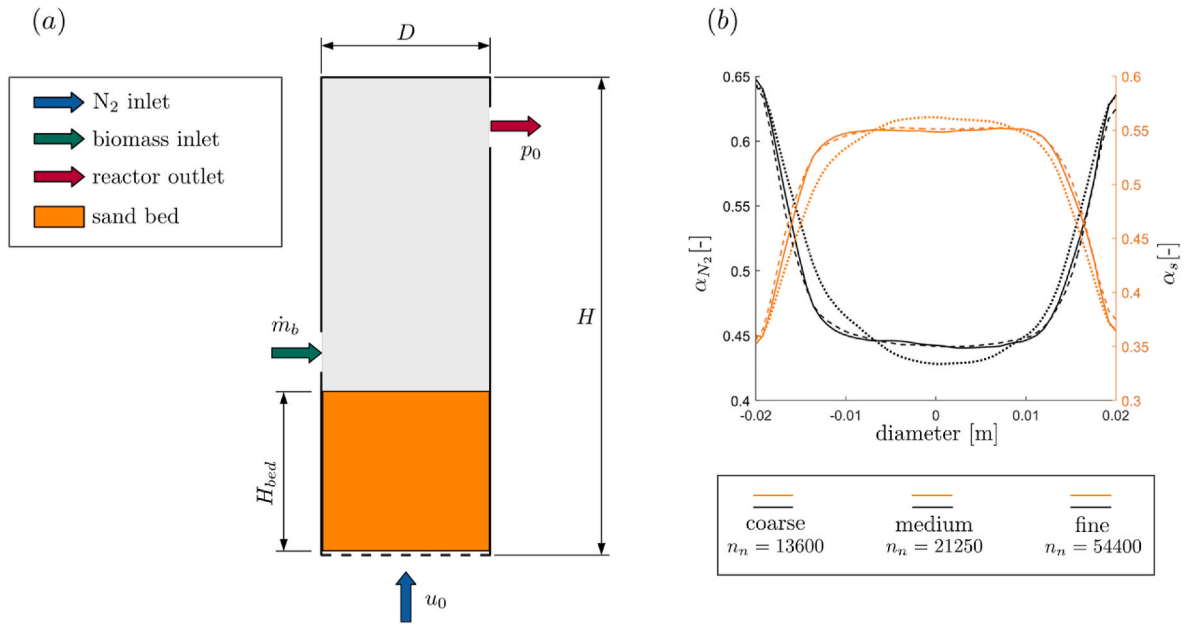


Fig. 2. CFD setup and grid sensitivity. (a) Sketch of the 2D rectangular domain of a FBR with height H and diameter D. The carrier gas (N_2) enters from the bottom of the reactor at velocity u_0 (blue arrow), while biomass is fed from a side with a mass flow rate $\dot{m}_b$ (green arrow). The reactor outlet is located at the top right and is set at pressure p_0 (red arrow). The walls are maintained at constant temperature T_{pyro}. The sand bed is initialized to fill an equivalent height of H_{bed}. **(b)** Comparison of the height-averaged volume fraction of sand, α_s, and nitrogen, α_{N_2}, for different mesh grids and for simulations without reactions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.2.3. Numerical campaign

Machine learning techniques require large training and test datasets. To meet this requirement, an extensive CFD simulation campaign is carried out to systematically explore the relevant design and operating space, generating a consistent dataset of input parameters and corresponding reactor responses. To ensure that the predictions of the proposed ML operate largely within the interpolation regime, the CFD dataset is generated by sampling reactor sizes to cover a wide geometric range, from lab-scale configurations up to industrially relevant scales. In addition, the operating condition parameters (*e.g.* T_{pyro}, u_0, biomass particle diameter d_b) are selected within physically meaningful bounds representative of fast pyrolysis of residual biomass. All the varied parameters and the corresponding ranges are defined in Table 2.

The parameter space was explored using a Latin Hypercube Sampling (LHS) method to define each simulated case and ensure well-distributed coverage of the multidimensional domain, within a maximum budget of 120 CFD simulations. Moreover, simulations are performed for 40 s of physical time to capture the transient evolution and achieve a statistically representative condition; see Supplementary Note S3 for the statistical analysis.

Consistent with the objectives of this work, the CFD database includes both reactive and non-reactive simulations. Specifically, 40 simulations accounted for the pyrolysis reaction scheme, while the remaining 80 cases are under non-reactive conditions. This choice served two purposes: first, to allow the surrogate ML models to learn hydrodynamic similarity and the impact of fluidization independently of chemistry; second, to limit the computational cost, particularly for the largest reactor sizes, where reactive simulations become expensive. In fact, even for larger reactors, purely thermo-fluid dynamics simulations require only a few hours of wall-clock time, whereas enabling reactive biomass chemistry can increase the computational cost to the order of weeks. All simulations were performed on a workstation using 16 CPU cores (AMD Ryzen Threadripper 7980X) in parallel with 250 GB of RAM.

Table 2
CFD parameter ranges.

Parameter	Range	Unit
Reactor diameter D	0.04 – 0.8	m
Reactor height H	0.34 – 3	m
Biomass flow rate $\dot{m}_{bio}$	0.3 – 100	kg/s
Bed height H_{bed}	0.1 – 1.7	m
Pyrolysis temperature T_{pyro}	350 – 650	°C
N_2 inlet velocity u_0	0.03 – 0.4	m/s
Biomass particle diameter d_b	0.2 – 0.8	mm
Bed particle density ρ_s	2700	kg/m^3
bed particle diameter. d_s	0.2 – 0.3	mm

2.3. Data-driven approach

This section outlines the ML methodology adopted in this work, together with the associated optimization workflow and feature-engineering strategy used to parametrize the problem. After defining the feature in a compact, physically meaningful form, the ML models are introduced, and finally, the optimization procedure is described.

2.3.1. Dimensional analysis and feature definition

Similarity theory was used to derive a compact, physics-based feature space by nondimensionalizing the governing equations and boundary conditions. The resulting dimensionless parameter set captures the coupled influence of hydrodynamics, heat transfer, and reaction kinetics, providing a scale-independent representation of reactor behavior. The full derivation is reported in the Supplementary Note S4.

The resulting set $\boldsymbol{\pi}$ of the proposed dimensionless features for the simulation setup used in this work, is summarized below

$$\boldsymbol{\pi} = \Big[\mathrm{Re_D}, \mathrm{Fr}, \mathrm{Eu}, \mathrm{Ste}, \mathrm{Pr}, \mathrm{Da}_{bc}, \mathrm{Da}_{bg}, \mathrm{Da}_{bt}, \mathrm{Da}_{tg}, \gamma_{bc}, \gamma_{bg}, \gamma_{bt}, \gamma_{tg}, \rho_b^*, \rho_s^*, d_b^*, d_s^*, \dot{m}_b^*, u_0^*, H^*, H_{bed}^* \Big]. \quad (1)$$

The nondimensionalization is performed using the reactor diameter D as the reference length scale, the minimum fluidization velocity u_{mf} as the reference velocity scale, the nitrogen density $\rho_{g,0} = \rho_g(T_{pyro})$ as the reference density, the gas dynamic viscosity $\mu_{g,0} = \mu_g(T_{pyro})$ as the reference viscosity, and T_{pyro} and p_0 as the reference temperature and

pressure, respectively. The superscript * indicates a dimensionless parameter and the definition of the dimensionless number is provided in Table 3. In particular, $Da_{bc}, Da_{bg}, Da_{bt}$, and Da_{tg} denote the Damköhler numbers for biomass → char, biomass → gas, biomass → tar, and tar → gas reactions (see Supplementary Note S5), respectively.

The nondimensionalization procedure adopted in this work is not unique: the choice of reference scales is inherently problem-dependent and determines which dimensionless groups carry the most direct physical meaning. Here, the characteristic spatial scale is taken as the reactor diameter, consistent with the primary objectives of reactor design and scale-up. Alternative scalings are equally admissible. If the reactor height H is instead chosen as the reference length, the natural time scale becomes H/u, which is an approximation of the mean gas residence time; under this convention, the Damköhler number acquires an immediate interpretation as the ratio of the characteristic reaction time to the residence time. A particle-based scaling, employing the particle diameter d_s as the reference length, shifts the emphasis to particle-scale transport and renders the Reynolds number the most informative parameter for characterizing interphase transfer.

These frameworks are not independent: any scaling can be recovered from another through a straightforward algebraic transformation. Replacing one reference length with another simply multiplies the affected dimensionless groups by the corresponding length ratio; for instance, substituting D with H rescales the relevant groups by the factor D/H. Consequently, results derived under one convention can be unambiguously recast in any other without loss of generality.

As a final remark on the parameter set $\boldsymbol{\pi}$, its dimensionality could be further reduced by introducing dimensionless groups such as the Archimedes or Stokes numbers, which incorporate some of the parameters already defined. However, in the present work these parameters are intentionally kept separate to preserve the explicit representation of the underlying physical effects.

2.3.2. Data-driven techniques and workflow

A two-stage GP surrogate model is developed to replace CFD simulations. The surrogate model maps the $\boldsymbol{\pi}$ set to the biomass pyrolysis product yields $\left(y_c, y_t, y_g\right)$ through an intermediate layer of 6 thermo-fluid parameters $\boldsymbol{\theta} \in \mathbb{R}^6$ defined in Table 4. Therefore, the complete data-driven pipeline is

$$\boldsymbol{\pi} \xrightarrow{\text{GP-stage1}} \boldsymbol{\theta} \xrightarrow{\text{GP-stage2}} (\zeta_c, \zeta_t) \xrightarrow{\text{ALR}^{-1}} \left(y_c, y_t, y_g\right),$$

where $\zeta_c = \log\left(y_c / y_g\right)$ and $\zeta_t = \log\left(y_t / y_g\right)$ is the additive log-ratio (ALR) transform of the product yields taking into account that $y_c + y_t + y_g = 1$. Under this transformation, the targets (ζ_c, ζ_t) are unconstrained, and the corresponding yields are recovered via the inverse ALR mapping, which enforces the mass conservation by construction. The automatic relevance determination (ARD) algorithm is applied to both GP-stages to rank feature importance, enabling dimensionality reduction and improving GP performance [41,42]. Moreover, a feature preprocessing step is performed before the model's training. In particular, features with dynamic ranges exceeding two orders of magnitude are log-transformed to reduce scale differences. In our setup only the Damköhler numbers satisfy this criterion. All features are then standardized to zero mean and unit variance. In all the GP models the Matérn 5/2 kernel function is employed where the hyperparameters are estimated by maximizing the log marginal likelihood.

Table 3
Dimensionless number definition.

Symbols	Equation	Definition
Re_D	$\frac{\rho_{g,0} u_0 D}{\mu_{g,0}}$	Reynolds number
Fr	$\frac{gD}{u_{mf}^2}$	Froude number
Eu	$\frac{p_0}{\rho_{g,0} u_{mf}^2}$	Euler number
Ste	$\frac{c_{p,g} T_0}{\Delta H_r}$	Stefan-like number
Pr	$\frac{\mu_{g,0} c_{p,g}}{\lambda_g}$	Prandtl number
Da	$\frac{\nu D}{u_{mf}} \dot{r}$	Damköhler number
γ	$\frac{E_a}{R T_{\text{pyro}}}$	Thermal severity

The training procedure consists of two nested levels: an external level and an internal level. At the external level, each GP stage is validated independently using k-fold cross-validation (*i.e.* 5-fold). Within each fold, an internal bagging procedure is adopted to reduce prediction variance. Specifically, for each GP stage, an ensemble of 15 GP models is trained, with each model fitted on a different bootstrap-resampled version of the training partition of that fold. The final prediction and the associated uncertainty for each fold are then obtained by aggregating the ensemble outputs. It is important to note that GP-stage1 and GP-stage2 are defined on different datasets: the former uses the full set of 120 CFD simulations, whereas the latter uses only the subset of 40 simulations including pyrolysis reactions. In both cases, during k-fold cross-validation, training is performed only on the training partition associated with the current fold. The first 5 s of simulation are excluded from the averaging procedure in order to remove the initial transient and ensure statistically representative results.

A final remark on the use of 2D simulations instead of 3D CFD simulations is needed. Despite the process's intrinsic 3D characteristics, this is not expected to substantially affect the proposed framework. The GP-stage1 features are not directly linked to the detailed local dynamics, whereas the GP-stage2 parameters, although dynamics-dependent, are defined as space-time-averaged quantities. As a result, the transition to 3D simulations would mainly induce systematic shifts in these averaged values, rather than alter the underlying surrogate relationships. See Supplementary Note S6 for further details.

2.3.3. Reactor design: constrained multi-objective Pareto optimization

The product distribution $\left(y_c, y_t, y_g\right)$ can define a multi-objective optimization problem. For instance, in fast-pyrolysis applications, y_t should be maximized, while the minimization of y_c and y_g is required. However, not every point in the design space corresponds to a physically viable reactor: geometrical and thermo-fluid conditions must also satisfy engineering constraints. The two GP surrogates are therefore coupled to identify the constrained Pareto front, *i.e.* the set of designs that are simultaneously Pareto-optimal in yield space and feasible with respect to thermo-fluid and geometrical bounds.

The optimization domain is defined as the hyper-rectangular bounding box of the training data

$$\Psi = \prod_j^{\dim(\boldsymbol{\pi})} \left[\pi_{1,j}, \pi_{2,j}\right],$$

where $\pi_{1,j}$ and $\pi_{2,j}$ are the desired limits for the jth feature. A set of N candidate designs is extracted from Ψ by LHS. For every candidate $\boldsymbol{\psi} \in \Psi$, both the yield prediction $\left(\hat{y}_c, \hat{y}_t, \hat{y}_g\right) = ALR^{-1}[\mathrm{GP}_2(\mathrm{GP}_1(\boldsymbol{\psi}))]$ and the thermo-fluid response $\hat{\boldsymbol{\theta}} = \mathrm{GP}_1(\boldsymbol{\psi})$ are evaluated, where GP_1 and GP_2 denote the GP-stage1 and GP-stage2 surrogates, respectively, and the hat symbol indicates a predicted quantity. A candidate is labeled *feasible* if all the thermo-fluid predictions lie within the prescribed bound:

$$\mathscr{F} = \{\, \boldsymbol{\psi} \in \Psi \mid \hat{\theta}_k(\boldsymbol{\psi}) \in \left[\theta_{1,k}, \theta_{2,k}\right] \quad \forall\, k \in \left[1, \dim(\boldsymbol{\theta})\right] \}.$$

Because the GP prediction degrade outside the training data, an extrapolation penalty is imposed based on the Euclidean normalized distance of each candidate from its nearest training point. If the distance is larger than a prescribed maximum value, the candidate is discarded.

Table 4
Thermo-fluid parameters. The table summarizes the definition of the dimensionless parameter set $\boldsymbol{\theta}$. Parameters are computed from CFD solution fields.

Symbol	Equation	Definition
m_{idx} mix index	$\frac{1}{u_{mf}}\left(\frac{\sum_{\text{bed}}\alpha_s\left(u_{s,y}-\langle u_{s,y}\rangle\right)^2}{\sum_{\text{bed}}\alpha_s}\right)^{1/2}$	weighted standard deviation of the axial sand velocity
$\sigma_T T$ uniformity	$\frac{1}{\langle T_{\text{bed}}\rangle}\left(\frac{\sum_i^n\left(T_{\text{bed},i}-\langle T_{\text{bed}}\rangle\right)^2}{n}\right)^{1/2}$	standard deviation of the temperature in the bed
δ_p pressure drop	$\frac{\langle \Delta p_{\text{bed}}\rangle}{\rho_{g,0}u_{mf}^2}$	pressure drop across the bed
τ_g residence time	$\frac{H}{u_{mf}/\alpha_g}\frac{1}{D/u_{mf}}$	scaled gas residence time
f_b bubble frequency	$\frac{N_{trans}}{n_x\Delta t}\frac{D}{u_{mf}}$	N_{trans} number of emulsion-to-void transition (threshold $\alpha_{emulsion}=0.15$) summed over n_x (number of cells along the diameter) in the Δt time window
D_b bubble diameter	$D_b=\frac{\sum_j D_{eq,j}A_j}{\sum_j A_j}$	$D_{eq}=\sqrt{(4A/\pi)}$ and $A=\alpha\lvert C\rvert\Delta x^2$ where C is the number of connected cells such that $\alpha_{emulsion}<0.15$

The feasible, extrapolation-safe candidates are subjected to non-dominated sorting with respect to the three yield objectives (maximize $\hat{y}_t$ and minimize $\hat{y}_c$ and $\hat{y}_g$), providing the constrained Pareto front. Candidates in the Pareto front are those for which no yield can be improved without either worsening another yield or violating a constraint.

2.3.4. Similarity-based reactor scale-up

Complete dimensional similarity, *i.e.* matching all the $\boldsymbol{\pi}$ groups, is not possible when the reactor size changes, because the dimensionless constraints over-determine the independent physical design variables $\boldsymbol{\varphi}=\left[T_{\text{pyro}},H,H_{\text{bed}},D,u_0,d_s,d_b,\dot{m}_b,\rho_s\right]$. The proposed scale-up procedure relaxes this constraint by exploiting the ARD relevance analysis from the GP surrogates. In particular, for a given base-case reactor, the scaled possible designs are obtained by minimizing the deviation of the K most important features (together with the product yield distribution) with respect to the base case, while allowing the remaining $\dim(\boldsymbol{\pi})-K$ features to vary freely.

Given a base-case reactor characterized by the dimensionless group $\boldsymbol{\pi}_0$ with target yields $\boldsymbol{y}_*=\left(y_{c^*},y_{t^*},y_{g^*}\right)$ and a scale factor s on a size metric $m(\boldsymbol{\pi})$ (*e.g.* bed volume, biomass flow rate), the scale-up is formulated as a constrained minimization problem over the design variables:

$$\boldsymbol{\pi}^*=\operatorname{argmin} S(\boldsymbol{\pi}) \quad \text{subject to} \quad m(\boldsymbol{\pi})=s\,m(\boldsymbol{\pi}_0)$$

where $\boldsymbol{\pi}^*$ represents the scaled reactor, and the objective function $S(\boldsymbol{\pi})$ reads

$$S(\boldsymbol{\pi})=w_y\underbrace{\sqrt{\frac{1}{3}\sum_k\left(\hat{y}_k-y_{k^*}\right)^2}}_{\text{yield mismatch}}+w_\pi\underbrace{\sqrt{\sum_j w_j\delta_j^2}}_{\text{weighted }\pi\text{-deviation}}+\lambda_p\underbrace{\sum_{j\in\mathcal{J}}\max\left(\lvert\delta_j\rvert-\tau_\pi,0\right)^2}_{\text{penalty on top }K\text{ features}}. \tag{2}$$

Here, $\delta_j=\left(\pi_j(\boldsymbol{\pi})-\pi_j(\boldsymbol{\pi}_0)\right)/\max\left(\lvert\pi_j(\boldsymbol{\pi}_0)\rvert\right)$ is the relative deviation of each feature from the reference, w_y and w_π are weighting coefficient, λ_p is the penalty coefficient, and τ_π is the maximum tolerated relative deviation for the important $\boldsymbol{\pi}$ features, and w_j denote the ARD importance of the jth feature. The first term drives the surrogate-predicted yields of the scaled-up reactor towards the reference values. The second term concentrates the dimensionless similarity on the features most relevant for the yield prediction, as identified by the ARD. The third term penalizes violations of the hard tolerance on the top K groups, while the remaining groups are constrained only softly through the importance-weighted term, providing the geometric degrees of freedom needed for scale-up.

The objective $S(\boldsymbol{\pi})$ is non-smooth and involves discrete variables, which precludes direct gradient-based optimizations. The problem is therefore solved in two steps. First, a global exploration was performed by evaluating $S(\boldsymbol{\pi})$ on a certain number of candidates generated by LHS, with sampling in the $\dim(\boldsymbol{\pi})-$ dimensional space, with bound extended by an arbitrary factor of $\sqrt{s}$ on geometric and flow variables to accommodate larger scale reactor. Candidates not satisfying the size constraint $\lvert m(\boldsymbol{\pi})-s\,m(\boldsymbol{\pi})\rvert/m(\boldsymbol{\pi}_0)\le 0.05$ or the physical constraint $H>H_{\text{bed}}$ and $H>D$ are discarded. Second, the best candidates from the global search are refined locally by minimizing $S(\boldsymbol{\pi})$ using Powell's conjugate-direction method [43]. The size constraint was enforced during refinement via a quadratic penalty.

Fig. 3 provides a synoptic overview of the complete methodology, integrating data preparation, surrogate modeling, constrained multi-objective optimization, and similarity-based scale-up into a unified workflow.

3. Results and discussion

This section first presents the validation of the CFD setup against experimental data, along with the CFD results. Next, it shows the performance of the proposed ML models and demonstrates them numerically on biomass fast pyrolysis examples, including their use in creating design maps for FBRs and their potential for scale-up applications.

3.1. Experimental validation and CFD results

Although high-level modeling of pyrolysis is relatively straightforward, accurately describing the coupled physical and chemical phenomena governing fluid–particle and particle–particle interactions in fluidized-bed reactors remains challenging. To support the methodological approach proposed in this work for fast pyrolysis, and to ensure the reliability of the presented results, an experimental validation of the chemical reaction scheme used in the numerical simulations is initially reported, while an experimental validation of the hydrodynamics is reported in Supplementary Note S9.

The FBR used for validation is the experimental system described in Section 2.1. The reactor is initially filled with a sand bed, with a granular diameter of $d_s=2.2\times10^{-4}$ m, occupying approximately one-third of the total reactor height, and it is fluidized with nitrogen while being heated to the selected pyrolysis temperature. Once thermal equilibrium is reached, the reactor is fed with a mass flow of 500 g/h of pure cellulose with a granular diameter of $d_b=8.5\times10^{-4}$ m. The residence time of the produced gases and the fluidization regime are controlled by adjusting the velocity of the fluidizing gas, which is set to $3u_{mf}$ at the chosen pyrolysis temperature. The test is conducted for 1 h, thus processing a total of 500 g of cellulose. Afterwards, the products are analyzed to determine their mass fractions and compositions. In total, three tests were carried out at T_{pyro} of 400, 450, and 500 °C, respectively. The same conditions are then considered from a numerical point

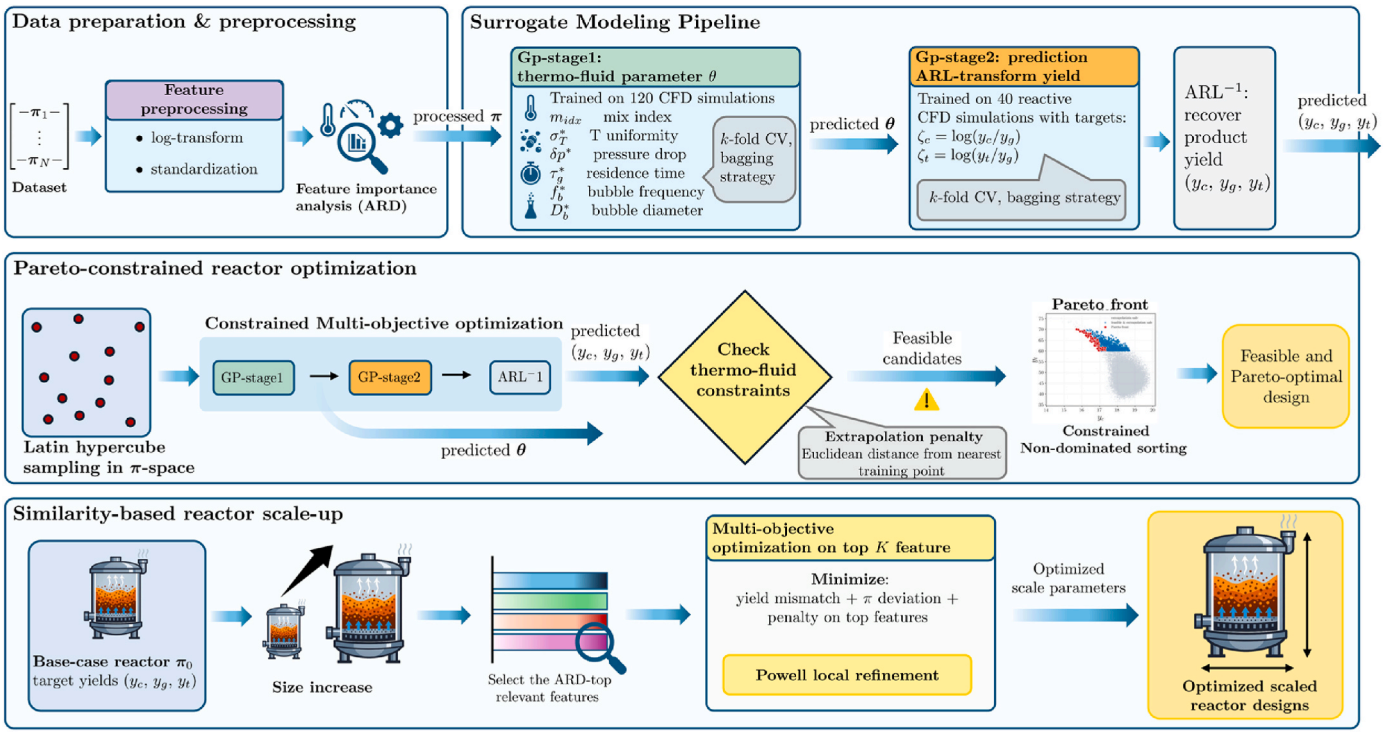


Fig. 3. Methodology. Synoptic overview of the complete methodology, integrating data preparation, surrogate modeling, constrained muclttttt optimization, and similarity-based scale-up into a unified workflow.

of view, and the results, in terms of product distribution and composition, are compared with the experiments, see Fig. 4.

In Fig. 4(a), the product distribution (biochar, bio-oil, and non-condensable gas yields) is shown for comparison. The gas yield has a relative error of 25% between CFD and experiments; this reflects the very small absolute contribution of the gas fraction, which is always the least abundant product. As a result, the absolute deviation remains limited and within measurement uncertainty, making it acceptable. The predicted bio-oil and biochar yields have maximum relative deviations of about 6%. These results suggest that the CFD simulation and the chemical mechanism used effectively reproduce the main trends and the temperature dependence of char, tar, and gas yields under the tested conditions with good accuracy. Regarding the oil composition, the species included in the chemical kinetics model are lumped, which does not allow for a direct comparison. Nevertheless, two compounds are described by the numerical and experimental approaches, levoglucosan (LVG) and hydroxymethylfurfural (HMF). In Fig. 4(b), the calculated and experimental data for LVG are compared for $T_{\text{pyro}} = 500\,^{\circ}C$, and an excellent agreement can be observed with a maximum deviation of $\simeq$ 13 % for the HMF, which accounts for less than 3% of the total oil composition. All the species not explicitly described by the two methods were grouped under the category '*others*'. For the gas composition, a strong consistency was observed across all the main species considered (*i.e.* CO and CO_2) and for all the temperatures analyzed, with a relative error between 3% and 4%, see Fig. 4(c). The influence of temperature on the production trends is generally well captured, confirming the reliability of the implemented numerical model (both from a physics and chemical point of view). A larger deviation is observed in the prediction of hydrogen and methane. Nevertheless, over the investigated temperature range, these species represent, on average, less than 10% of the total gas produced, and the corresponding differences remain within the experimental measurement uncertainty.

The validation results indicate that the implemented physico-chemical CFD model reproduces the experimental data with satisfactory accuracy. Good quantitative agreement for the dominant product fractions is obtained, and the model consistently captures the observed temperature-dependent trends. While deviations remain for some components, the predicted yield distributions follow the experimental evolution closely across the investigated conditions. Moreover, the proposed workflow is readily generalizable to individual chemical species once sufficiently detailed compositional information is available for model calibration and comparison.

3.2. Data-driven model: fast pyrolysis applications

3.2.1. Feature analysis

Although the feature importance analysis relies on trained models, it is presented here before discussing the GPs' training and test results, as it provides useful insight for interpreting the model behavior discussed in the following sections. To further validate the proposed feature analysis, the ARD approach was compared with the permutation importance method, thereby assessing its effectiveness and reliability for the considered models.

In Fig. 5(a), the GP-stage1 feature importance is reported. The features are ranked from the top to the bottom according to their impact on the GP model. Since the GP-stage1 maps the dimensionless feature vector $\boldsymbol{\pi}$ into the thermo-fluid parameters $\boldsymbol{\theta}$, it effectively links the reactor characteristics to the thermal and fluidization regimes. The three most influential features, *i.e.* d_s^*, H_{bed}^*, and u_0^*, are key parameters in FBRs since they govern the onset of fluidization, the bed hydrodynamics, the intensity of gas–solid mixing and heat transfer. The sand particle diameter, together with its density ρ_s^*, determines the minimum fluidization velocity and influences particle mixing and bubble formation. The bed height regulates the solid inventory and the gas–solid contact time, whereas the superficial gas velocity controls the fluidization regime and the intensity of particle agitation. Together, these parameters determine the effectiveness of gas–solid mixing and heat transfer, which are critical for reactor performance. The Froude number, Fr, characterizes the relative importance of inertial and gravitational forces and therefore affects particle suspension, bubble rise dynamics, and the overall hydrodynamics of the fluidized bed. Finally, the Euler number

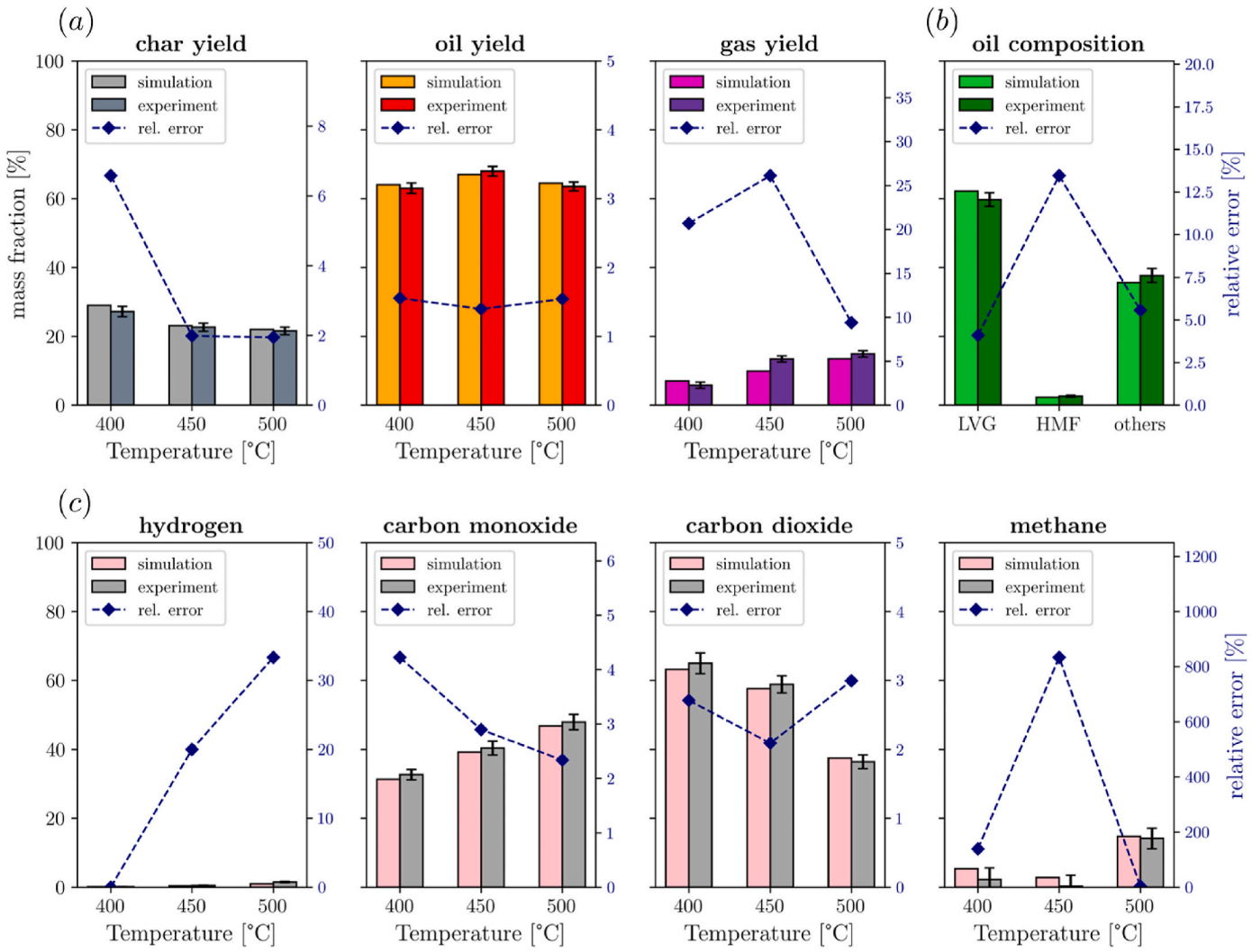


Fig. 4. Experimental validation. (a) Comparison between the product distribution obtained from the numerical simulation (light colors) and experimental campaign (dark colors) for three different temperatures $T_{\text{pyro}} = 400, 450, 500°\text{C}$. (b) Comparison between the oil composition for $T_{\text{pyro}} = 500\ °C$. (c) Gas composition comparison between numerical (pink bars) and experimental results (gray bars). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

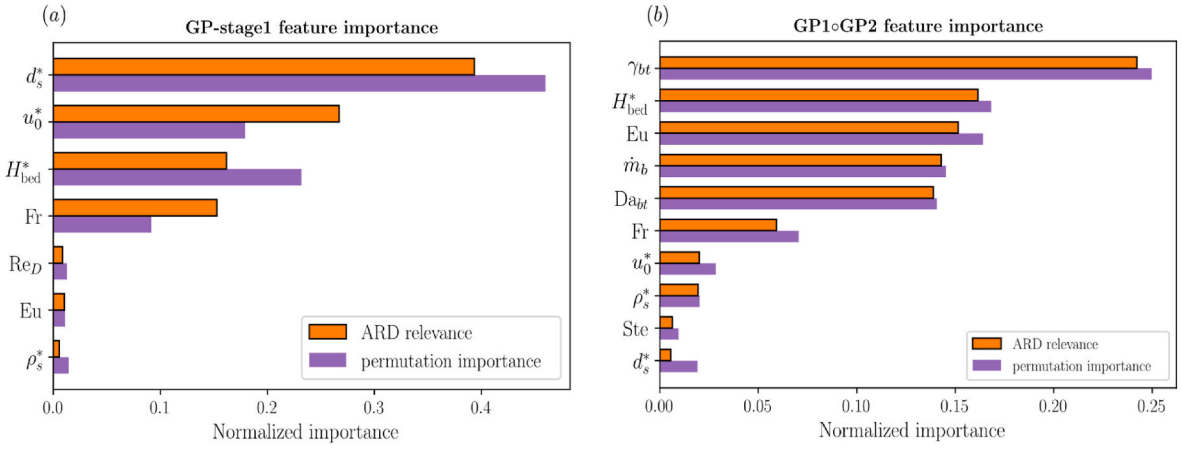


Fig. 5. Feature analysis. The ARD and permutation importance methods are compared in the analysis of feature impact on the proposed GP models. **(a)** GP-stage1 feature importance analysis. **(b)** Importance analysis on the GP models compositions.

Eu characterizes the pressure drop across the bed. The only thermo-fluid parameter that is not explicitly captured by the dominant features is σ_t^*, which represents the thermal uniformity of the bed. This is mainly due to the fact that, in the simulations presented, the sand bed temperature

remains nearly uniform on average, so that σ_t^* exhibits only limited variability ($\leq 5\%$) across the dataset. Nevertheless, this quantity can still be considered partially represented by Re_D. In addition to its hydrodynamic role, Re_D contains the gas density, which depends on T_{pyro}, and therefore is also able to indirectly account for a thermal effect, partially compensating for the limited ARD weight of σ_T^*.

The ARD analysis for the composite GP models is reported in Fig. 5 (b). Here, the reactions are also considered, therefore, temperature plays a fundamental role. In fact, two of the most important features are γ_{bt} and Da_{bt} that are directly correlated to the pyrolysis temperature and reaction rate. It is important to note that the emergence of only one representative γ term and one representative Damköhler number is not the result of an a priori selection, but rather a direct outcome of the ARD procedure. This reflects the fact that the γ terms essentially act as descriptors of the thermal level of the process and therefore share the same temperature dependence, differing mainly by constant scaling factors. Similarly, the Damköhler numbers, although associated with different reaction pathways, exhibit the same functional dependence on temperature and operating conditions, since they all describe the ratio between characteristic reaction and transport timescales. As a result, the

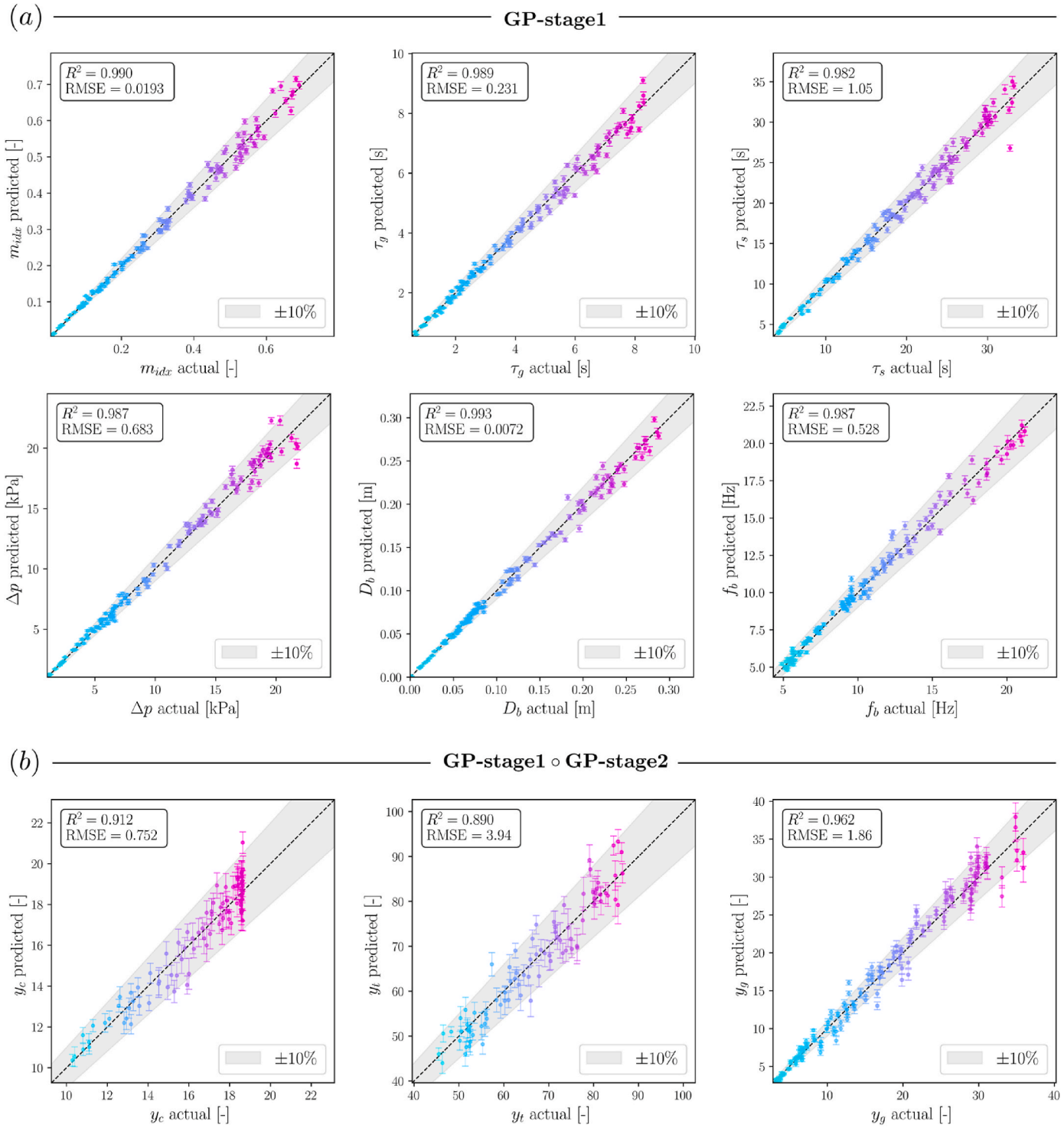


Fig. 6. GP model performance. (a) Predictive performance of the GP-stage1 model. **(b)** Predictive performance of the composite GP model.

ARD procedure naturally identifies γ_{bt} and Da_{bt} as representative descriptors of thermal severity and reaction–transport coupling, respectively. Accordingly, the biomass flow rate $\dot{m}_b$ also emerges as an influential feature. By changing the feed rate of the amount of cold biomass introduced into the bed, local temperature gradients are modified, which in turn affects the particle heating rate. Finally, the influence of Ste and ρ_b^* reflects the amount of thermal energy required by the biomass to sustain chemical conversion, which impacts the extent of the devolatilization and the resulting product distribution.

3.2.2. Prediction performance of GP models

Fig. 6 assesses the predictive capability of the proposed GP surrogate model. The results for GP-stage1, shown in Fig. 6(a), highlight that the predictions closely follow the diagonal line across all the thermo-fluid quantities, indicating that the surrogate is able to accurately capture the relationship between the dimensionless reactor parameters and the corresponding thermo-fluid operating conditions. In particular, most of the predictions fall within ±10 % deviation band, suggesting that the GP model provides a reliable approximation of the CFD simulations across the explored parameter space. The associated uncertainty estimates are generally small and consistent with the observed prediction errors, demonstrating the effectiveness of the calibration of the GP model using the nested cross validation and bagging procedures.

Fig. 6(b) reports the predictive performance of the composite surrogate, obtained by sequentially coupling GP-stage1 with GP-stage2 to predict the pyrolysis product distribution. As expected, the predictive accuracy slightly deteriorates compared to the single-stage model due to uncertainty propagation between the two surrogate levels. Nevertheless, the predictions remain well aligned with the reference values, and most of the points still lie within the ±10% error bounds. This result indicates that the proposed hierarchical surrogate structure is able to preserve good predictive performance even when the intermediate thermo-fluid quantities are not known exactly but are instead estimated by the first GP layer.

3.2.3. Fluidized-bed reactor design

To illustrate the potential of the proposed surrogate framework as a reactor design tool, the trained composite GP model was used to

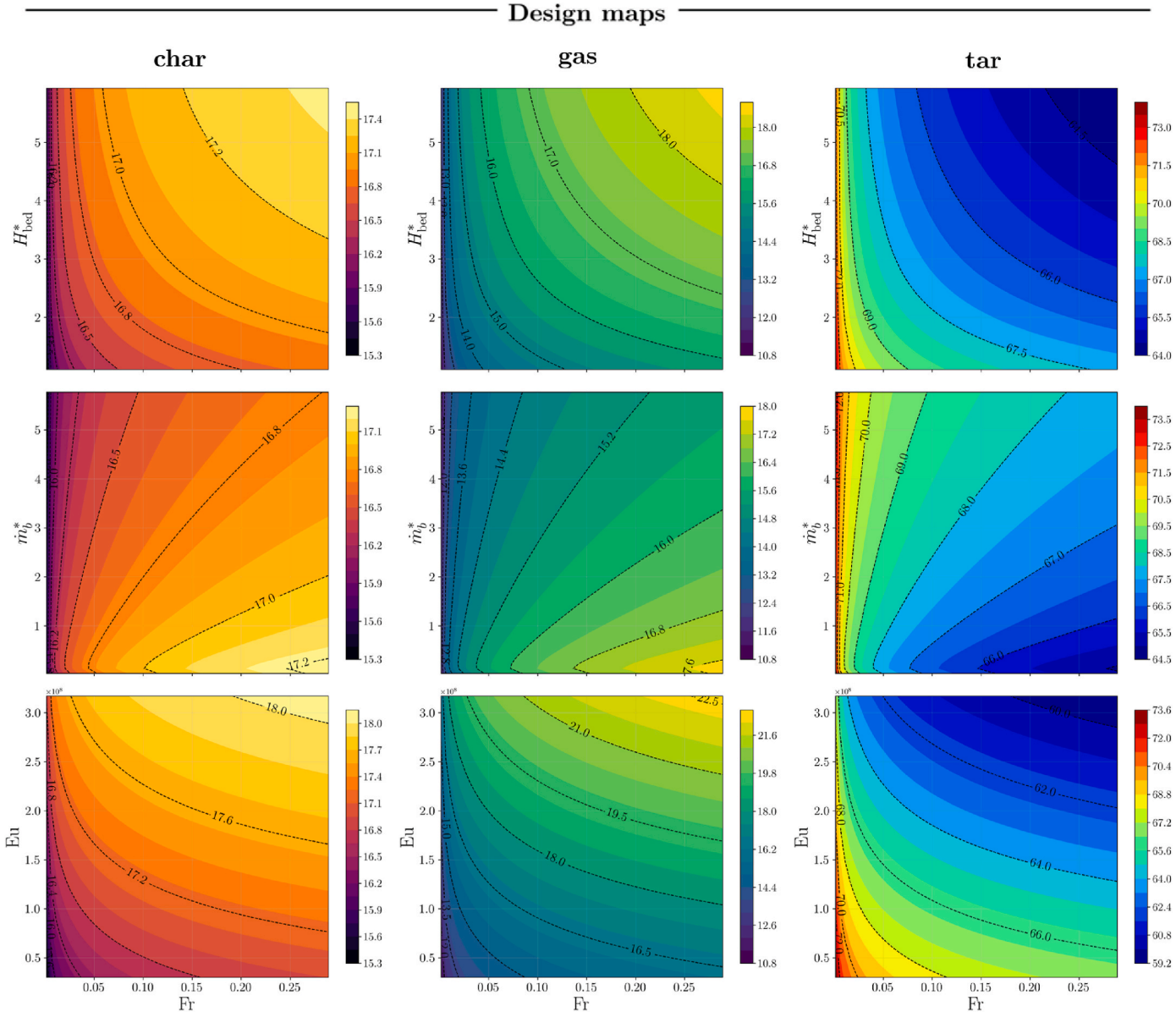


Fig. 7. Design maps: hydrodynamics. Product distribution maps as a function of $\mathrm{Fr}, \mathrm{Eu}, H^*_{\mathrm{bed}}$, and $\dot{m}_b$. The remaining dimensionless parameters are fixed at the median values of the dataset.

generate parametric design maps over selected subsets of the dimensionless feature space. Unlike CFD simulations, which are computationally expensive and therefore impractical for systematic exploration of the design domain, the surrogate model allows rapid evaluation of thousands of operating conditions. This capability enables the identification of relationships between reactor parameters and product distribution, as well as the visualization of design trade-offs across multiple physical regimes.

In the following analysis, the GP surrogate was evaluated over structured grids in some of the feature space while keeping the remaining variables fixed at representative reference values derived from the training dataset. The resulting response surfaces describe the predicted yields of the lumped products (char, tar, and gas) as functions of the selected parameters.

For example, Fig. 7 reports the design maps obtained from the composite GP surrogate by varying selected hydrodynamic similarity parameters. In particular, the predicted product yields are shown as functions of the Froude number, Fr, combined with the dimensionless bed height H^*_{bed}, the dimensionless biomass feed rate $\dot{m}^*_b$, and the Euler number Eu. Since these quantities, not only are some of the most ARD-top features, but also mainly control the fluidization regime, pressure losses, solids inventory, and gas-solid interaction, the maps provide a compact representation of the hydrodynamic contribution to reactor performance.

A first general trend emerging from Fig. 7 is that increasing Fr shifts the product distribution away from tar and toward char and gas. This behavior is consistently observed in all three rows of the figure: as Fr increases, the predicted char and gas yields increase, whereas the tar fraction decreases. From a physical standpoint, higher Fr corresponds to stronger inertial effects relative to gravity, which are associated with more intense fluidization, enhanced gas-solid contacting, and a larger tendency toward vapor conversion and secondary degradation of condensable species. The effect of H^*_{bed} is comparatively weaker but still systematic. At fixed Fr, increasing the dimensionless bed height leads to a moderate increase in char and gas yields and to a corresponding decrease in tar. This suggests that larger bed inventories promote longer

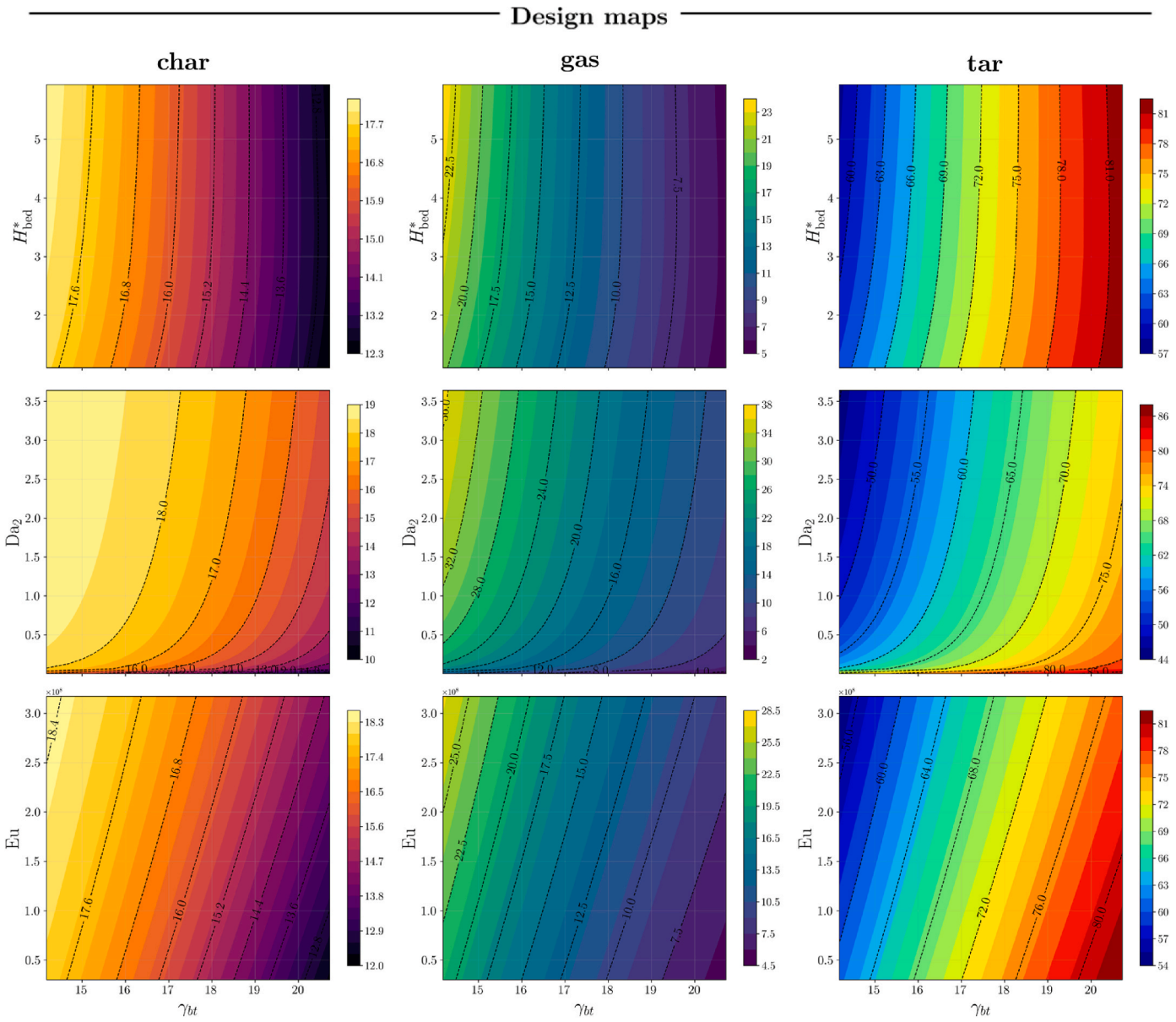


Fig. 8. Design maps: thermo-fluid parameters. Product distribution maps as a function of $\mathrm{Fr}, \mathrm{Da}, \gamma_1, H^*_{\mathrm{bed}}$. The remaining dimensionless parameters are fixed at the median values of the dataset.

interaction between solids and vapors within the dense region, thereby enhancing conversion of condensable intermediates. A different trend is observed for the dimensionless biomass feed rate $\dot{m}_b^*$. While the influence of Fr remains dominant, increasing $\dot{m}_b^*$ tends to slightly reduce char and gas production and to preserve a larger tar fraction. This can be interpreted as a thermal dilution effect: a higher feed rate introduces more cold biomass relative to the available bed heat capacity, locally reducing the temperature and limiting the extent of secondary vapor conversion while preserving a high heating rate.

Finally, the maps involving Eu show that the pressure-drop-related hydrodynamic regime also affects the predicted selectivity. Higher Eu, especially when combined with larger Fr, is associated with increased char and gas yields and reduced tar production. Overall, Fig. 7 confirms that the surrogate model captures a coherent hydrodynamic trend: regimes characterized by more intense fluidization and stronger gas-solid interaction tend to reduce condensable yields and favor the formation of non-condensable products and solid residue.

Fig. 8 extends the analysis by introducing thermal and chemical parameters alongside the hydrodynamic ones. In particular, the first two rows show the predicted product yields as functions of the dimensionless bed height H_{bed}^* and the Euler number Eu, respectively, both combined with the thermal parameter γ_{bt}. The third row instead illustrates the interaction between the Froude number and the Damköhler number associated with the chemical conversion process. Compared with Fig. 6, these maps provide a broader representation of the coupled thermo-fluid-dynamic and kinetic effects governing reactor performance.

A first important observation is the strong sensitivity of the product distribution to γ_{bt}, confirming the ARD analysis results. Since γ_{bt} is inversely proportional to the pyrolysis temperature, increasing γ_{bt} corresponds to decreasing thermal severity. The maps therefore indicate that lower temperatures favor char formation, while higher temperatures promote devolatilization and enhance the production of condensable vapors and permanent gases. This behavior is consistent with the expected response of biomass fast pyrolysis, in which increasing temperature accelerates conversion and shifts the product distribution away from the solid residue.

When γ_{bt} is combined with H_{bed}^*, the thermal effect remains dominant, while the bed height introduces a secondary modulation of the product yields. At fixed temperature, increasing H_{bed}^* tends to slightly increase char and gas fractions and reduce tar yield, suggesting that a larger bed inventory promotes a longer interaction between vapors and the dense bed region, thereby enhancing secondary conversion effects. However, compared with the variation induced by γ_{bt}, the influence of bed height remains more limited.

A similar trend is observed in the maps combining Eu and γ_{bt}. Also in this case, the temperature-related parameter controls the main variation in the response surfaces, while the Euler number introduces a more moderate correction associated with the hydrodynamic regime and the pressure-drop characteristics of the bed. In particular, higher Eu values tend to reduce the tar fraction and slightly increase char and gas production, especially under conditions corresponding to higher pyrolysis temperatures. This suggests that pressure-drop-related hydrodynamic effects may intensify gas-solid interaction and promote the conversion of condensable intermediates.

The third row highlights the combined effect of hydrodynamic intensity and reaction timescale through the joint variation of Fr and the Damköhler number. In this case, both parameters exert a clear influence on the predicted yields. The maps show that the reactor response is governed by the balance between transport-related effects, represented by Fr, and chemical conversion rates, represented by the Da. Their combined variation identifies different operating regimes, ranging from conditions favoring the preservation of condensable products to regimes in which more extensive secondary conversion leads to higher gas and char formation.

Overall, Fig. 8 further confirms that the proposed surrogate framework is not limited to purely hydrodynamic interpretations but can also be used to explore the coupled influence of fluidization, thermal severity, and chemical kinetics on reactor behavior. This highlights the usefulness of the methodology as a rapid design-support tool for identifying meaningful trends within a complex multidimensional parameter space.

It is important to note that, in practical reactor design, the objective is rarely to find an optimal configuration in an absolute sense. Rather, the design task typically consists in identifying solutions that satisfy a set of application-specific requirements simultaneously, such as geometric limitations, target throughput, admissible operating conditions, and minimum product-performance thresholds among others. Fig. 9 demonstrates how the proposed constrained-design methodology addresses this class of problems, operating as a flexible decision-support tool that allows different design criteria to be imposed depending on the specific industrial objective.

For the case reported in Fig. 9, the constrained-design procedure described in Section 2.3 was applied by prescribing a set of representative requirements on both reactor geometry and process performance. In particular, constraints were imposed on the reactor dimensions (H^*), on the minimum desired oil yield, on the superficial gas velocity (u_0^*), and on the biomass flow rate to be processed ($\dot{m}_b$). The list of the imposed constraints can be found in Table 5. Starting from the admissible design domain, a large number of candidate reactors were generated and evaluated through the surrogate models. These candidates, represented by the gray points in Fig. 9, define the broader set of possible reactor configurations explored by the method.

A first selection step was then performed by retaining only those configurations that satisfied the imposed product-distribution requirement, namely the minimum desired oil yield. These reactors are identified by the blue points and represent the subset of designs that are potentially suitable for the target application from a process-performance perspective. Subsequently, a second filtering and optimization stage was applied by accounting for the additional physical and operational constraints embedded in the constrained-design procedure. The resulting red points therefore identify the subset of feasible and Pareto-optimal reactor configurations, *i.e.* those that simultaneously satisfy the prescribed design requirements, remain consistent with the thermo-fluid and operational limitations of the system, and cannot be further improved with respect to the selected objectives without worsening at least one competing criterion. From an engineering standpoint, each of the reactors represented by the red points can be interpreted as a valid Pareto-optimal solution for the selected application scenario. In other words, the proposed optimization method does not return a single universally optimal reactor, but rather a set of candidate optimal designs compatible with the imposed objectives and constraints. This is an important feature of the approach, since in real design practice the final selection is often performed by the engineer on the basis of additional considerations not explicitly included in the optimization, such as construction complexity, economic cost, plant integration, or operational flexibility. In Fig. 9(b), the design parameters and the product yields for five Pareto-optimal reactors, are reported. The figure highlights how very different reactors offer the same performance in term of the desired yields. In fact, despite noticeable differences in reactor sizes and operating conditions, the predicted yields remain nearly unchanged across the selected Pareto-optimal solutions. Further details on the reactor designs obtained can be found in Supplementary Note S7, while Supplementary Note S8 discusses the versatility of the proposed methodology when combined with other ML techniques.

Overall, Fig. 9 is intended as an illustrative example of how the proposed methodology can be used in a design-oriented context. Different constraints, objectives, or selection criteria could be introduced depending on the intended application, meaning that the same framework and optimization method can be adapted to different reactor-design problems rather than being limited to the specific case presented here.

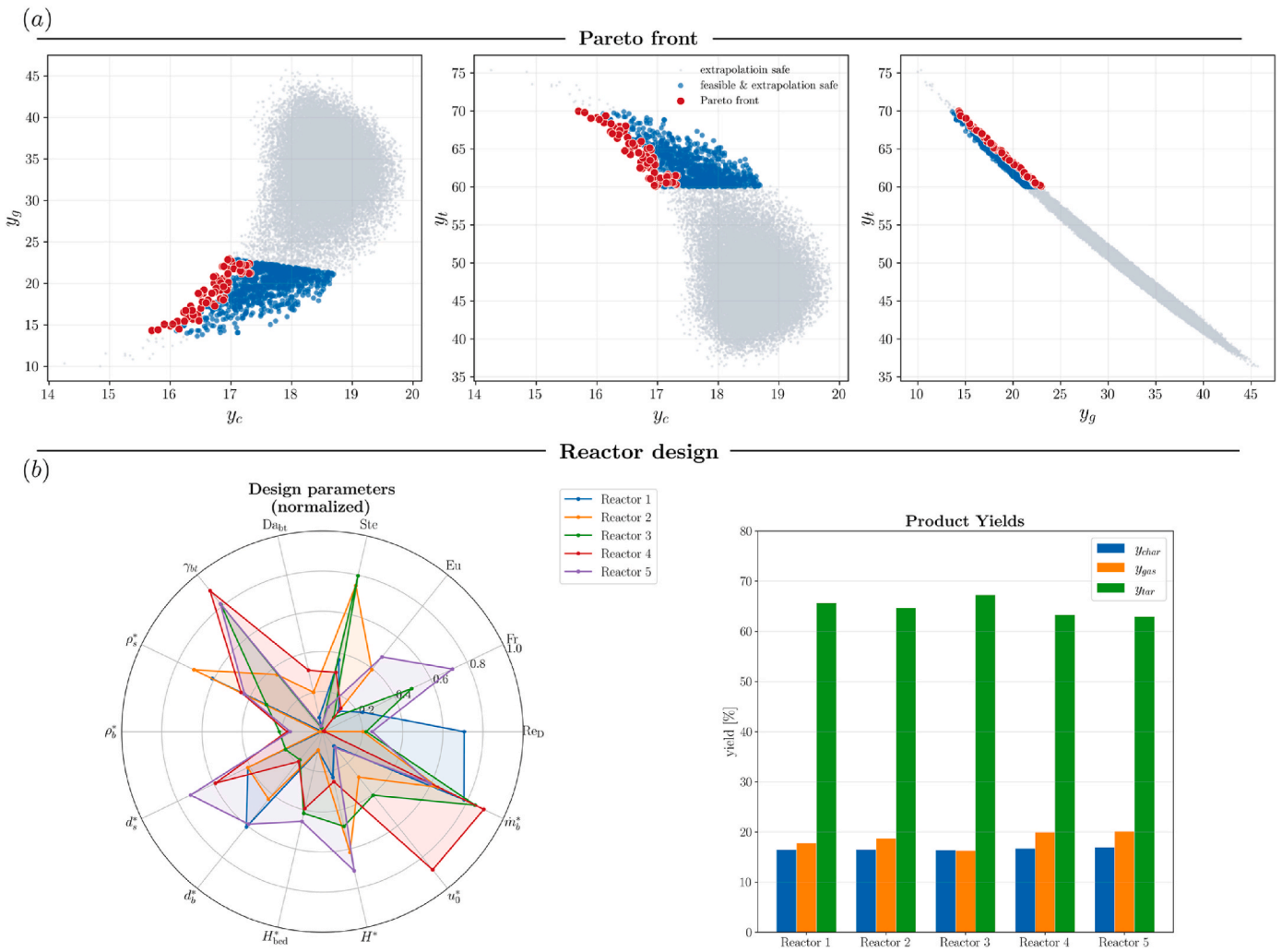


Fig. 9. Constrained design. (a) Pareto-optimal reactor resulting from the proposed constrained design approach. For a given set of constraint: gray dots represent all the reactor configuration that satisfy the physical and geometrical constraints; blue dots represent the reactor that satisfy also the product distribution constraint; red dots represent the Pareto-optimal reactors. **(b)** Design parameters for 5 Pareto-optimal reactors (left) and the corresponding product yields (right). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 5
Example of constrained parameters.

Parameter	Range
y_t	0.6 – 0.7
H^*	4.0 – 6.0
u_0^*	2.0 – 3.0
m_b^*	4.0 – 5.0

3.2.4. *Fluidized-bed reactor scale-up*

To demonstrate the applicability of the similarity-based scale-up methodology described in Section 2.3.4, the procedure was applied to a representative reactor configuration selected from the CFD dataset and used as a base case. This example is presented as an illustrative demonstration of the proposed similarity-based procedure, rather than as a general validation of predictive scale-up accuracy.

The objective of the scale-up exercise was to identify reactor configurations capable of processing a larger biomass throughput while preserving the key dimensionless characteristics governing the reactor behavior. In particular, two scale-up scenarios were considered, corresponding to doubling and quadrupling the dimensionless biomass flow rate. The optimization procedure was therefore applied by imposing the scaled biomass throughput while constraining a subset of dimensionless parameters to remain unchanged with respect to the base configuration. Specifically, the parameters γ_{bt}, ρ_s^*, ρ_b^*, H^*, and H_{bed}^* were kept constant, ensuring that the scaled reactors preserved the same thermal severity, particle properties, and relative geometric proportions of the base case.

The selection of these constraints is motivated by physical considerations related to the dominant mechanisms governing fluidized-bed fast pyrolysis and to the results obtained from the feature importance analysis. In particular, maintaining the same value of γ_{bt} ensures that the thermal severity of the process remains unchanged, thereby preserving the temperature-dependent reaction kinetics. Similarly, fixing the solid densities ρ_s^* and ρ_b^* guarantees that the particle properties remain identical to those of the base case. Finally, keeping the dimensionless geometric parameters H^* and H_{bed}^* constant preserves the relative proportions between reactor height, bed inventory, and reactor diameter, thereby maintaining similar gas–solid contact patterns and hydrodynamic structure within the fluidized bed. Under these constraints, the scale-up algorithm searched the design space for reactor configurations minimizing the deviation from the reference feature set while satisfying the imposed throughput requirement.

The results of the scale-up procedure are summarized in Fig. 10, which reports the candidate reactor configurations ranked according to the objective function defined in Section 2.3.4. Each solution corresponds to a reactor design that satisfies the prescribed constraints while

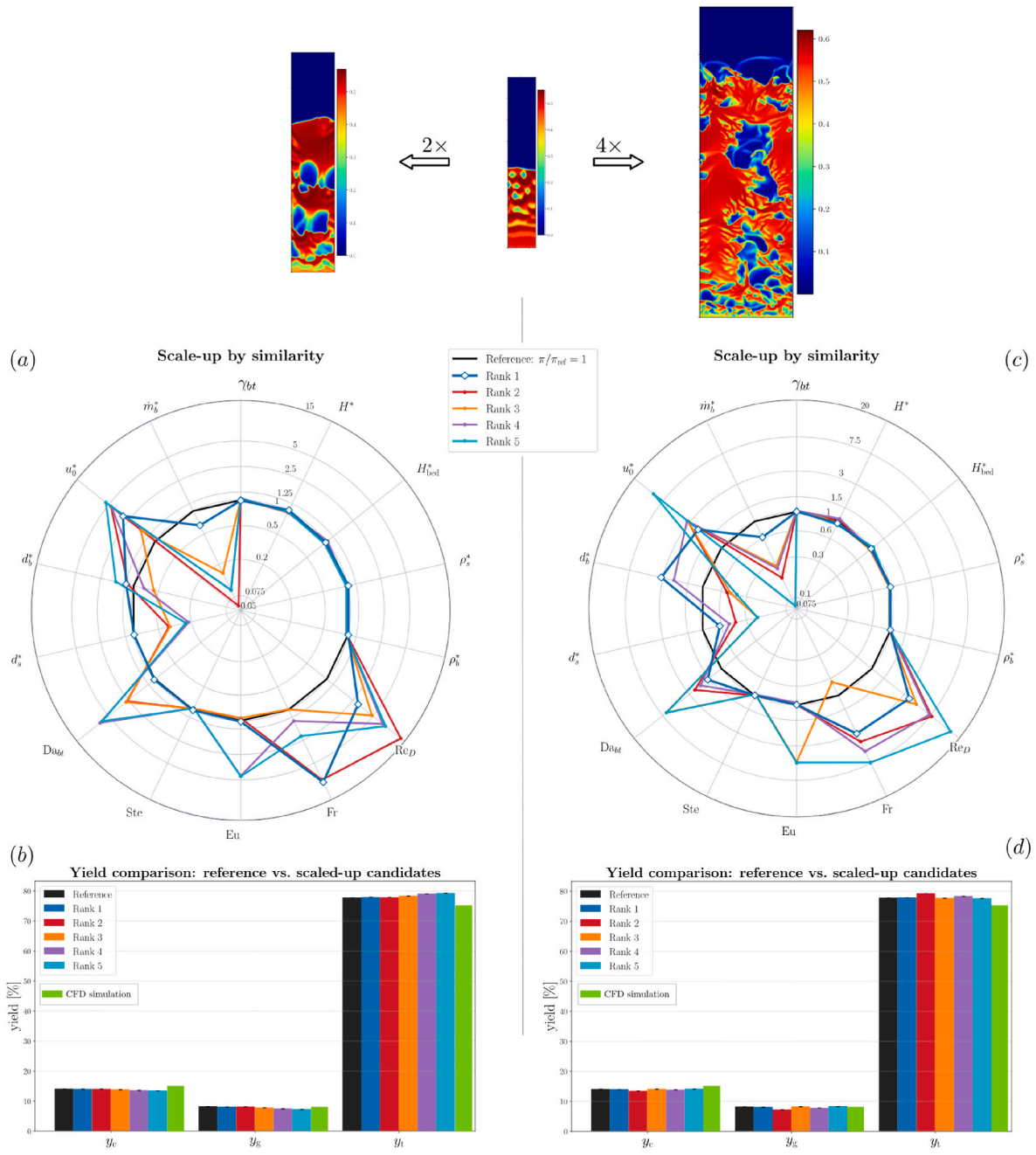


Fig. 10. Similarity scale-up. (a,c) Radar charts of the dimensionless reactor parameters for the configurations identified by the scale-up procedure. All values are reported relative to the base-case scenario, represented by the black curve. **(b,d)** Product-yield distributions for the five reactor configurations satisfying the scale-up criterion. The light-green bars denote the product distribution for the CFD-simulated Rank 1 solution. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

processing the increased biomass flow rate. The ranking reflects the degree to which the resulting configuration preserves the similarity conditions with respect to the base reactor.

For the best-ranked configuration (Rank 1) in Fig. 10, it can be observed that the similarity constraints are satisfied with minimal deviation from the reference state. In particular, the Eu remains essentially unchanged with respect to the base configuration, even though it was not explicitly imposed as a constraint in the optimization procedure. This behavior indicates that the optimization naturally converged toward a hydrodynamic regime characterized by a similar pressure-drop distribution and fluidization behavior. Furthermore, although the Stefan number was not explicitly imposed as a constraint in the scale-up procedure, the GP framework naturally identified scaled reactor configurations for which this parameter remained nearly constant. This behavior suggests that, by preserving the physical properties of the biomass, the model implicitly captured the associated thermochemical similarity between the base case and the scaled reactors. It is also worth noting that the Damköhler numbers differ from that of the base reactor. This variation should not be interpreted as a change in the intrinsic chemical kinetics of the system. Rather, it arises from the modification of the reactor geometry and operating conditions required to achieve the desired scale-up. In particular, the change in reactor diameter D and dimensionless superficial velocity u_0^* affects the characteristic transport timescale appearing in the Damköhler number definition, leading to a different value of this parameter while the underlying kinetic model remains unchanged.

To further assess the reliability of the proposed procedure, the Rank 1 reactors identified for the scale-up scenarios corresponding to biomass-throughput increases by factors of 2 and 4 are simulated by means of CFD. The corresponding product distributions are reported in Fig. 10(b), where the scaled reactors are directly compared, in term of the yield, with the base case. The results show that the scaled reactors obtained through the proposed methodology are able to reproduce a product distribution close to that of the reference reactor. In particular, although the char fraction is generally overestimated and the tar fraction correspondingly underestimated, the overall distribution remains similar to the base-case behavior for both the scaling factors. This indicates that the proposed scale-up procedure is capable of preserving the main conversion trends and therefore provides a meaningful route to identify scaled reactor designs with comparable process performance.

Overall, the results demonstrate that the proposed scale-up methodology is capable of identifying reactor configurations that preserve the dominant similarity conditions of the base system while accommodating significant variations in throughput. The example reported in Fig. 10 illustrates how the method can be used to systematically explore feasible scale-up solutions and to quantify the degree of similarity between scaled reactors and their reference configuration.

4. Conclusions

The design and scale-up of fluidized-bed thermochemical reactors remain inherently complex, as reactor performance emerges from the strong coupling between fluidization hydrodynamics, heat transfer, and reaction kinetics. While high-fidelity CFD provides detailed insight into these phenomena, its direct application to broad parametric exploration and systematic scale-up studies is constrained by computational cost. In this context, this work proposes a structured CFD–machine-learning framework to support reactor design and similarity-based scale-up through reduced-order, physics-informed surrogate models.

The primary contribution of the study is methodological. A CFD-informed surrogate framework has been developed in which governing equations and boundary conditions are systematically transformed into a dimensionless feature space derived from similarity analysis. Gaussian-Process surrogate models trained on these features reproduce the dominant thermo-fluid and reactive trends observed in the CFD simulations within the explored design space, while offering improved interpretability and transferability compared with purely data-driven approaches. The use of feature-relevance analysis further enabled a controlled relaxation of full similarity constraints, providing a practical route to similarity-aware design and scale-up.

The framework was applied to biomass fast pyrolysis in a fluidized-bed reactor as a representative case study. Design maps generated from the surrogate models illustrated how hydrodynamic, thermal, and kinetic parameters jointly influence product distribution and demonstrated the approach's potential for rapid exploration of multidimensional design spaces. The constrained-design example showed how application-specific requirements—such as geometric limits, throughput, and target yields—can be systematically incorporated into the decision process.

The methodology was also demonstrated on a similarity-based scale-up problem, in which reactor configurations corresponding to increased biomass throughput were identified while preserving selected key dimensionless parameters. Additional CFD simulations of the best-ranked scaled configurations confirmed that the approach can identify designs that maintain product distributions close to that of the reference reactor, although some systematic deviations, such as a tendency to overpredict char and underpredict tar, were observed. These results highlight both the potential and the current limitations of the surrogate-based scale-up procedure and indicate directions for further refinement.

Overall, the results show that the proposed CFD–ML framework provides a physically grounded and flexible methodology for supporting reactor design and scale-up while significantly reducing the need for exhaustive CFD campaigns. Although the approach was demonstrated for biomass fast pyrolysis in a fluidized-bed reactor, its formulation is independent of the specific reaction network and reactor application. As such, the framework is intended as a general strategy for CFD-informed surrogate modeling of gas–solid thermochemical systems, to be further extended and validated as additional data and applications become available.

Declaration of generative AI technologies in the manuscript preparation process

During the preparation of this work the authors used Claude.ai in order to proof reading and editing the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

M. Baldelli: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **L. Bartolucci:** Supervision, Writing – review & editing. **S. Cordiner:** Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing. **P. Mele:** Data curation, Writing – original draft, Writing – review & editing. **V. Mulone:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

M.B. and S.C. acknowledge the financial support from Project ECS 0000024 Rome Technopole, –CUP E83C22003240001, NRP Mission 4 Component 2 Investment 1.5, Funded by the European Union-NextGenerationEU.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijhydene.2026.156385.

References

[1] International Energy Agency. Renewables 2024, Analysis and Forecast to 2030. 2024.
[2] Sorino D, et al. Pyrolysis of mixed contaminated plastic wastes: assessing the influence of polymers composition, temperature and residence time. Waste Manag 2025;201:114793.
[3] Li J, et al. Recent advances in plastic waste pyrolysis for liquid fuel production: critical factors and machine learning applications. Appl Energy 2023;346:121350.
[4] Bartolucci L, Cordiner S, Mele P, Mulone V. Defatted spent coffee grounds fast pyrolysis polygeneration system: lipid extraction effect on energy yield and products characteristics. Biomass Bioenergy 2023;179:106974.
[5] Huang B-S, Chen H-Y, Chuang K-H, Yang R-X, Wey M-Y. Hydrogen production by biomass gasification in a fluidized-bed reactor promoted by an Fe/CaO catalyst. Int J Hydrogen Energy 2012;37:6511–8.
[6] Kuo J-H, Lin C-L, Ho C-Y. Effect of fluidization/gasification parameters on hydrogen generation in syngas during fluidized-bed gasification process. Int J Hydrogen Energy 2022;47:40656–63.
[7] Lopez G, et al. Hydrogen generation from biomass by pyrolysis. Nat Rev Methods Primers 2022;2:20.
[8] Go KS, Son SR, Kim SD, Kang KS, Park CS. Hydrogen production from two-step steam methane reforming in a fluidized bed reactor. Int J Hydrogen Energy 2009;34:1301–9.
[9] Matsen JM. Scale-up of fluidized bed processes: principle and practice. Powder Technol 1996;88:237–44.

[10] Couto N, Silva VB, Bispo C, Rouboa A. From laboratorial to pilot fluidized bed reactors: analysis of the scale-up phenomenon. Energy Convers Manag 2016;119:177–86.
[11] Kalo L, Pant HJ, Upadhyay RK. Validation of the glicksman scaling law for gas–solid conical fluidized beds using the radioactive particle tracking technique. Ind Eng Chem Res 2020;59:20943–52.
[12] Dang Q, Hu W, Rover M, Brown RC, Wright MM. Economics of biofuels and bioproducts from an integrated pyrolysis biorefinery. Biofuel Bioprod Biorefining 2016;10:790–803.
[13] Gebreslassie BH, Slivinsky M, Wang B, You F. Life cycle optimization for sustainable design and operations of hydrocarbon biorefinery via fast pyrolysis, hydrotreating and hydrocracking. Comput Chem Eng 2013;50:71–91.
[14] Weih N, et al. Operational experience with miscanthus feedstock at the bioliq® fast pyrolysis plant. J Anal Appl Pyrolysis 2024;177:106338.
[15] Knowlton TM, Karri SBR, Issangya A. Scale-up of fluidized-bed hydrodynamics. Powder Technol 2005;150:72–7.
[16] Glicksman, L. R. Scaling Relationships for Fluidized Beds.
[17] Glicksman LR, Hyre M, Woloshun K. Simplified scaling relationships for fluidized beds. Powder Technol 1993;77.
[18] Mellin P, Zhang Q, Kantarelis E, Yang W. An Euler-Euler approach to modeling biomass fast pyrolysis in fluidized-bed reactors - focusing on the gas phase. Appl Therm Eng 2013;58:344–53.
[19] Ranganathan P, Gu S. Computational fluid dynamics modelling of biomass fast pyrolysis in fluidised bed reactors, focusing different kinetic schemes. Bioresour Technol 2016;213:333–41.
[20] Venier CM, et al. Computational modeling of biomass fast pyrolysis in fluidized beds with eulerian multifluid approach. Fluids 2024;9.
[21] Wen K, Guo L, Xia Z, Cheng S, Chen J. A hybrid simulation method integrating CFD and deep learning for gas–liquid bubbly flow. Chem Eng J 2024;495:153515.
[22] Li J, Suvarna M, Pan L, Zhao Y, Wang X. A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification. Appl Energy 2021;304:117674.
[23] Park J-G, Jun H-B, Heo T-Y. Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models. Appl Energy 2021;298:117250.
[24] Molinaro R, Singh J-S, Catsoulis S, Narayanan C, Lakehal D. Embedding data analytics and CFD into the digital twin concept. Comput Fluids 2021;214:104759.
[25] Swaminathan N, Parente A. Machine Learning and its Application to Reacting Flows: ML and Combustion. Springer; 2023.
[26] Zou X, et al. Generative artificial intelligence and hybrid models to accelerate LES in reactive flows: application to hydrogen/methane combustion. arXiv preprint arXiv:2507.08426 2025.
[27] Zhu L-T, et al. Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors. Ind Eng Chem Res 2022;61:9901–49.
[28] Qin L, Zhou L, Li Z, Lei H, Zhu L-T. CFD analysis of operating parameters and machine learning-aided optimization of reactor performance in a polyethylene pyrolysis fluidized bed. Fuel 2026;421:138977.
[29] Baldelli M, Bartolucci L, Cordiner S, Mulone V. Pyrolysis of biomass in fluidized-bed reactors: towards a data-driven approach for reactor design and control. In: Journal of Physics: Conference Series, 3143. IOP Publishing; 2025. 012010.
[30] van Ommen JR, Teuling M, Nijenhuis J, van Wachem BGM. Computational validation of the scaling rules for fluidized beds. Powder Technol 2006;163:32–40.
[31] Baldelli M, Bartolucci L, Cordiner S, Mulone V. Fluidized bed reactors for energy recovery from biomass and wastes: a data-driven approach towards the development of digital twins for real time control and monitoring. In: Journal of Physics: Conference Series, vol. 2893. IOP Publishing; 2024. 012014.
[32] Jalalifar S, et al. A hybrid SVR-PSO model to predict a CFD-based optimised bubbling fluidised bed pyrolysis reactor. Energy 2020;191:116414.
[33] Hu C, et al. A deep-learning model for predicting spatiotemporal evolution in reactive fluidized bed reactor. Renew Energy 2024;225:120245.
[34] Ebrahimi M, Siegmann E, Prieling D, Glasser BJ, Khinast JG. An investigation of the hydrodynamic similarity of single-spout fluidized beds using CFD-DEM simulations. Adv Powder Technol 2017;28:2465–81.
[35] Bartolucci L, Cordiner S, Mele P, Mulone V. Energy efficiency and hydrogen demand of atmospheric fast catalytic hydropyrolysis of residual lignocellulosic biomass. In: Journal of physics: conference series, 3143. IOP Publishing; 2025. 012005.
[36] Kunii D, Levenspiel O. Fluidization Engineering. Elsevier; 2013.
[37] Weller HG, Tabor G, Jasak H, Fureby C. A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput Phys 1998;12:620–31.
[38] Ranzi E, Debiagi PEA, Frassoldati A. Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note I: kinetic mechanism of biomass pyrolysis. ACS Sustainable Chem Eng 2017;5:2867–81.
[39] Ranzi E, et al. Chemical kinetics of biomass pyrolysis. Energy Fuel 2008;22:4292–300.
[40] Mørup M, Hansen LK. Automatic relevance determination for multi-way models. J Chemometr: J Chemometr Soc 2009;23:352–63.
[41] Liu K, Li Y, Hu X, Lucu M, Widanage WD. Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries. IEEE Trans Ind Inf 2019;16:3767–77.
[42] Hestenes MR. Conjugate direction methods. In: Conjugate Direction Methods in Optimization. Springer; 1980. p. 81–149.
[43] Buckley AG. Constrained minimization using powell's conjugacy approach. SIAM J Numer Anal 1976;13:520–35.