Engineering 7 (2021) 1231–1238

Contents lists available at ScienceDirect

Engineering

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

Research
AI Energizes Process Manufacturing—Perspective

Hybrid Data-Driven and Mechanistic Modeling Approaches for Multiscale Material and Process Design

Teng Zhou [a,b,*], Rafiqul Gani [c,d], Kai Sundmacher [a,b]

[a] Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg D-39106, Germany
[b] Process Systems Engineering, Otto-von-Guericke University Magdeburg, Magdeburg D-39106, Germany
[c] PSE for SPEED Co., Ltd., Allerod DK 3450, Denmark
[d] Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 9 September 2020
Revised 22 October 2020
Accepted 15 December 2020
Available online 1 April 2021

Keywords:
Data-driven
Surrogate model
Machine learning
Hybrid modeling
Material design
Process optimization

ABSTRACT

The world's increasing population requires the process industry to produce food, fuels, chemicals, and consumer products in a more efficient and sustainable way. Functional process materials lie at the heart of this challenge. Traditionally, new advanced materials are found empirically or through trial-and-error approaches. As theoretical methods and associated tools are being continuously improved and computer power has reached a high level, it is now efficient and popular to use computational methods to guide material selection and design. Due to the strong interaction between material selection and the operation of the process in which the material is used, it is essential to perform material and process design simultaneously. Despite this significant connection, the solution of the integrated material and process design problem is not easy because multiple models at different scales are usually required. Hybrid modeling provides a promising option to tackle such complex design problems. In hybrid modeling, the material properties, which are computationally expensive to obtain, are described by data-driven models, while the well-known process-related principles are represented by mechanistic models. This article highlights the significance of hybrid modeling in multiscale material and process design. The generic design methodology is first introduced. Six important application areas are then selected: four from the chemical engineering field and two from the energy systems engineering domain. For each selected area, state-of-the-art work using hybrid modeling for multiscale material and process design is discussed. Concluding remarks are provided at the end, and current limitations and future opportunities are pointed out.

1. Introduction

Materials can be broadly classified into two categories: functional process materials and end-user materials or products. The former are used during the manufacturing process (e.g., solvents, catalysts, adsorbents), while the latter are typically end-user chemicals; that is, their functions go beyond the factory gates. Functional process materials with tailored properties are often at the heart of advances in the process industry because the choice of these materials affects not only the economic cost, but also the environmental, health, and safety performance of the processes. As described by Grossmann and Westerberg [1], a modern process system can be decomposed into multiple scales at which different physical and/or chemical phenomena take place. The lowest scale involves all decisions that are linked to the structures of molecules or materials used in the process—for example, liquid solvents and solid adsorbents for chemical separations, heterogeneous catalysts for reactions, or refrigerants and phase change materials (PCMs) for energy transfer and conversion. Historically, materials have been discovered by means of the experimental trial-and-error method. This method is slow and inefficient, given the large size of the material design space. Due to the recent development of theoretical model-based methods, it is now popular and efficient to employ computer-aided approaches to guide material selection and design. On the other hand, it should be noted that there are always strong interactions

* Corresponding author.
E-mail address: zhout@mpi-magdeburg.mpg.de (T. Zhou).

https://doi.org/10.1016/j.eng.2020.12.022

between material selection and the operation of the processes where the material is used. For this reason, all the scales involved in a process system should be considered simultaneously, making integrated material and process design essential [2].

The first step of multiscale modeling is to connect the molecular scale to the phase scale, where the main task is to model and predict the macroscopic properties (e.g., diffusion coefficient, thermal conductivity, enthalpy, and Gibbs free energy) of the fluid mixtures based on the atomic- or molecular-level information. In principle, quantum chemical computation, molecular simulation, and equations of state (EoS) can provide these predictions. However, these calculations are computationally expensive and system dependent. Fortunately, due to the ever-increasing availability of experimental and theoretical data, it is now popular and efficient to model the properties of molecules and materials from their structures via descriptor-based empirical models [3]. Since these models describe the system property or behavior purely based on data correlations, they are known as data-driven models. Mathematical representations such as linear, polynomial, artificial neural network (ANN), Gaussian process, and kriging are widely used in data-driven property modeling [4]. A recent review on the methods and applications of data-driven approaches for the discovery and optimal design of various types of materials can be found in Ref. [3]. Knowing the macroscopic properties of the system, it is then possible to derive the constitutive relations (e.g., kinetics and phase equilibria) and implement them into the mass, energy, and momentum conservation laws of each process unit. Taking into account the connectivity between different units, one can finally scale the system upward from the phase level to the process level. Since the constitutive relations and conservation laws are derived from physical knowledge, they are known as first-principle or mechanistic models. The combination of data-driven and mechanistic models makes the solving of multiscale material and process design problems efficient and much faster. This model-combination strategy is known as hybrid modeling, and the resulting entire model is called a hybrid model.

In fact, there are three different types of hybrid model structures. As summarized in Ref. [5], the first parallel structure (type I) is actually a summation of two parts: a mechanistic term and a data-driven term. Within the model, well-known knowledge is represented with the mechanistic term, while the data-driven part describes the features that are unknown or expensive to understand. This type of hybrid model can lead to improved prediction accuracy [6,7]. Besides the parallel structure, there are two other consecutive hybrid structures where the data-driven model is arranged either before (type II) or after (type III) the mechanistic model. Compared with the structure of type III, the hybrid structure of type II is much more popular in process engineering and has already been used in many applications in multiscale material and process design [8–10]. In this hybrid structure, empirical data-driven models are first used to predict the properties of materials, which are then substituted into the mechanistic process models. By doing so, it is possible to successfully bridge the gap between material and process scales and efficiently perform integrated material and process design.

Due to the popularity and increasing number of applications, researchers have contributed several reviews on hybrid modeling. For example, Zendehboudi et al. [11] provide a review on hybrid modeling for process optimization, control, and monitoring. McBride et al. [12] emphasize the importance of hybrid modeling for separation process design. Yang et al. [13] highlight the significance of hybrid modeling in smart manufacturing. However, to the best of our knowledge, there is a lack of work summarizing the applications of hybrid modeling in multiscale material and process design. In this article, we first describe the general principles and design methodology of hybrid modeling for integrated material and process design. Two solution strategies are then introduced to solve hybrid-model-based material and process design problems. Later, we select six representative areas where hybrid modeling either has already been successfully applied or can potentially be used to simultaneously design the material and the process. For each area, we briefly review the state-of-the-art work and point out the current limitations and possible opportunities. Concluding remarks are provided at the end of this article.

2. Methodology

As mentioned in the introduction, there are two different types of materials: functional process materials and end-user materials or products. Fig. 1 illustrates the scheme of hybrid modeling for material and product design. Since functional process materials are typically used in the process industry, the design objective is to find energy-efficient and environmentally friendly processes. In comparison, the task of end-user material design is to promote a better life for end-users and a more sustainable society. As depicted in Fig. 1, the application domain of process systems engineering (PSE) has been broadened from the traditional analysis, simulation, and optimization of chemical processes to include the optimal design of molecules and materials [14]. Under this trend, the computer-aided material (or product) and process design method [15,16] has played an important role in the optimal design of various functional materials and chemical products.

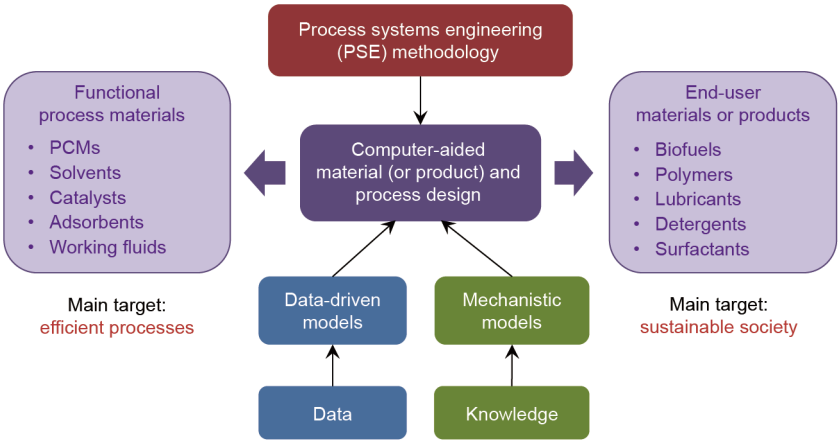

Fig. 1. Schematic diagram of hybrid modeling for computer-aided material (or product) and process design.

The increasing amount of available data makes data-driven models very important tools to predict material or product properties based on their structural and composition information [3]. On the other hand, mechanistic models are usually employed to describe the phenomena or principles of the process in which the material is used, due to the well-known underlying physics. By combining these two types of models, the material/product and the process system can be optimally designed by the formulation and solution of a mathematical optimization problem. Since the optimal design of end-user chemical products has been well reviewed by Uhlemann et al. [17] and Fung et al. [18], this article focuses on process materials design.

For better illustration, we herein classify process functional materials further into two categories: molecular materials that are typically composed of single or multiple molecules (e.g., solvents and working fluids (WFs)) and solid materials such as heterogeneous catalysts and adsorbents. Fig. 2 shows the scheme of hybrid modeling for integrated functional material and process design. The problem can be described as follows: Given a batch or continuous process using a functional material (e.g., solvent and adsorbent), find the optimal material structure and process operating conditions that will lead to the best process performance. This is a typical optimization problem, and the design variables include material selections and process conditions. The objective function is normally defined as a process performance index, such as the total annual cost, total energy consumption, or environmental effects of the process. As illustrated in Fig. 2, in order to calculate the objective function from the given design variables, we need ① property models that relate material structures to material properties and ② process models that relate the properties and process operating conditions to the overall process performance. As described in the introduction, the property models are usually empirical or data-driven, while the process models are most likely knowledge or mechanism-based. Zhou et al. [3] summarize a large number of publicly accessible structure and property databases for molecules and various solid materials. These data are very useful for building data-driven property models. With the data-driven property models and the mechanistic process models, one can successfully predict the performance of the process where the material is used. Once this forward problem (property prediction and process simulation) is successfully completed, the best material structure and process conditions can then be identified by solving a reverse material and process design problem.

Two solution strategies, decomposed and integrated design methods, can be employed to solve the reverse design problem. The decomposed design method solves material and process design problems in a sequential way. That is, desirable material properties are first defined by analyzing the process characteristics, and the optimal material is then identified to match these properties by solving a computer-aided material design problem. If the material is simply a molecule (e.g., a solvent or WF), empirical or semi-empirical data-driven models (e.g., the very popular group contribution (GC)-based models) are usually available for the prediction of various properties of molecules. In this case, the material design problem can be easily handled with the well-known computer-aided molecular design (CAMD) methodology [19]. Papadopoulos et al. [15] and Austin et al. [20] provide comprehensive overviews on the CAMD method, software/tools, and solution techniques. On the other hand, if the material is solid and has complex structures (e.g., adsorbents and catalysts), complex relationships usually exist between the material structure and its properties, which are difficult to model using traditional correlation methods. Thanks to the development of machine learning (ML) and deep learning methods [19,21], these complex relationships can now be efficiently modeled. For the optimal design of solid functional materials using ML models, interested readers can refer to the work of Zhou et al. [3]. After promising materials are identified, process design and optimization can then be performed for each material to find the best-matching process conditions.

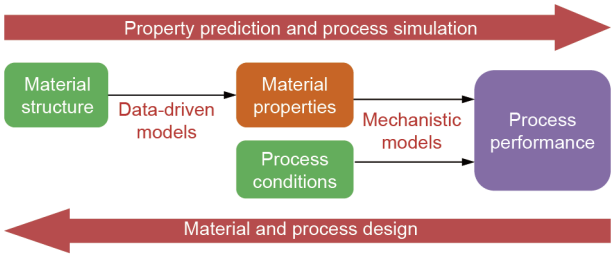

Fig. 2. Schematic diagram of hybrid modeling for integrated functional material and process design.

Decomposed design methods solve molecular/material design and process design problems sequentially [22,23]. Despite the high efficiency of these methods, this approach can lead to suboptimal solutions due to the following two reasons [24]: ① It is often difficult to know in advance which material property dominates the process performance; and ② the specification of process conditions strongly influences the selection of the material, while the selected material, on the other hand, reversely determines the optimal operating conditions of the process. Decomposed material and process design cannot reasonably capture this interdependent relationship.

In contrast to the decomposed design method, the integrated design method attempts to simultaneously identify the best material and process conditions [8,25]. This is usually performed by formulating and solving a mixed-integer nonlinear programming (MINLP) optimization problem [9], because both discrete variables representing material structures and continuous variables (i.e., process operating conditions) are involved, and most of the property and process models are inherently nonlinear. It is worth noting that when the mixed design space is large and the property and process models are very complicated, a successful solution of the MINLP problem significantly relies on good initial estimates. In addition, in most cases, process functional materials are pure substances. When mixtures of substances (e.g., solvent blends) are being designed, additional constraints related to the mixture composition and properties must be considered, which makes the integrated design problem much more challenging [26,27].

3. Applications

In the following sections, six representative areas are selected in which hybrid modeling either has already been successfully applied or can potentially be used to simultaneously design the material and process. For each area, the modeling strategy, solution method, main results, current limitations, and future opportunities are highlighted. The selected areas include solvent-based reaction and separation processes, adsorption separation processes, membrane separation processes, heterogeneous catalytic processes, organic Rankine cycle (ORC) processes, and thermal energy storage (TES) processes.

3.1. Solvent-based reaction and separation processes

Solvents are important functional materials in the chemical industry for promoting reactions and separations. The effect of a solvent on a reaction or separation process depends fully on the

solvent properties [28], such as the solvation ability and enthalpy of vaporization. First-principle methods (e.g., density functional theory (DFT) computation) can predict various solvent properties [29]. However, such methods are computationally expensive to use in multiscale solvent and process design. In order to tackle these complex design problems, data-driven models are typically used to predict solvent properties [30–33]. Combining them with suitable process models, it is possible to perform integrated solvent and process design. For example, Zhou et al. [9] propose an empirical data-driven model using conductor-like screening model for real solvents (COSMO-RS)-based solvent descriptors to correlate the solvent kinetic effects on a Diels–Alder reaction. By combining this data-driven kinetic model with mechanistic process models representing the reactor, distillation column, and heat exchanger, an MINLP-based integrated solvent and process design problem was formulated and solved to maximize the economic profit of the reaction process. The result showed that the optimal process included a less efficient (in terms of reaction efficiency) solvent that benefited the solvent–product separation more.

Accurate phase equilibrium prediction is very important for solvent selection in separation processes. Traditional predictive thermodynamic models, such as predictive Soave–Redlich–Kwong (PSRK) [34], are mathematically complicated because they are highly nonlinear and sometimes contain implicit equations. Replacing them with much simpler and explicit data-driven models can be very beneficial in order to reduce the computational demand in the integrated design of the solvent and the process system. Valencia-Marquez et al. [35] use the hybrid modeling method to perform an integrated ionic liquid (IL) and process design study for post-combustion CO_2 capture. A total of 394 experimental CO_2 solubilities in different ILs under various conditions were collected from the literature. An empirical correlation model using temperature, pressure, and the IL molecular weight as inputs was regressed to predict the CO_2 solubility in ILs. By implementing this data-driven model into the mechanistic process models, the solvent and absorption process were simultaneously optimized. Despite much simplification, the data-driven model did not take into account the structural effect of ILs on CO_2 solubility. Recently, Song et al. [30] establish a GC-based ANN model to accurately predict the CO_2 solubility in various ILs at different temperatures and pressures based on 10 116 experimental CO_2 solubility data. This model can well capture the complex relationship between IL structure and CO_2 solubility. We are now implementing this ANN model into a rigorous rate-based absorption model to perform integrated IL and process design for CO_2 capture, and it is expected that the first-hand results will be disclosed soon.

Hybrid modeling is used not only for pure solvent design, but also for mixture solvent design. McBride and Sundmacher [26] use a thermomorphic solvent (TMS) mixture consisting of dimethylformamide and decane to separate the homogeneous catalyst from the product after a reaction through temperature-controlled phase splitting in a decanter. The reaction being investigated was the hydroformylation of long-chain alkenes. Reactor and process design for this reaction is typically performed for a specific composition of the solvent mixture, without considering the economic impact of catalyst leaching. In order to reduce the complexity of process design, linear models were regressed to describe the partitions of multicomponents between two liquid phases in the decanter based on data obtained from rigorous liquid–liquid equilibrium (LLE) calculations. In addition, using the limited available experimental data, another quadratic correlation was fitted for estimating catalyst loss based on the composition of the TMS mixture. By combining the linear LLE expression and the quadratic catalyst leaching correlation with the mechanistic process models, the total cost of the hydroformylation process was optimized to obtain the optimal solvent composition as well as the best process operations. The optimization results showed that catalyst loss had a significant effect on the process cost, and that the frequently investigated solvent composition should be altered to increase catalyst retention. This important observation would not have been obtained if rigorous mechanistic models (instead of hybrid models) had been employed in the optimization.

3.2. Adsorption separation processes

Chemical separation can be performed with a liquid solvent. On the other hand, the separation can also be accomplished with a solid material using adsorption or membrane separation technology. Adsorption separation is usually implemented by pressure-swing adsorption (PSA) or temperature-swing adsorption processes. These processes are composed of two or more beds that interact with each other in a cyclic manner following a sequence of steps. Due to their dynamic and spatial distribution characteristics, adsorption processes are normally governed by a set of time- and space-dependent partial differential equations. To the best of our knowledge, there is no systematic work on integrated adsorbent and adsorption process design as yet. Most of the existing works have screened promising adsorbents by means of high-throughput molecular simulations and performed process optimization for each of the top materials. For example, Hasan et al. [36] pre-screen a list of promising zeolites from a large zeolite database based on their CO_2/N_2 adsorption Henry selectivity (the ratio of Henry constants of gases in the adsorbent) obtained from grand canonical Monte Carlo simulations. For each of the top zeolites, CO_2 and N_2 adsorption isotherms were generated, based on which the PSA process for post-combustion carbon capture was optimized. First et al. [37] and Liu et al. [38] extend this material screening and process optimization method to natural gas purification and H_2S separation, respectively. Despite the efficiency of the method, it should be noted that a single property criterion (e.g., the adsorption selectivity) is insufficient to reflect the complex effects of an adsorbent on the process performance. For this reason, many researchers have attempted to propose more comprehensive and reliable evaluation metrics [39,40]. However, Khurana and Farooq [41] found that all of the molecular simulation-based indicators were not well correlated to the real optimized process performance of the adsorbent. Obviously, there is a gap between the proposed indices and the predictions of how the adsorbents will perform under fully optimized process conditions. This gap can be only addressed by integrated material and process design.

In order to simultaneously design the material and the process in an integrated manner, it is first necessary to relate the adsorbent structure to the adsorption isotherm, because this relation is required for process simulation and optimization. The easiest way is to build data-driven correlations to predict the isotherm model parameters from the adsorbent structures. By combining this data-driven model with adsorption process models, it is then possible to perform integrated adsorbent and process design. In the past decade, metal–organic frameworks (MOFs), as an important type of porous material, have been shown to possess great potential for many applications, and particularly for gas separation. Many researchers have built data-driven models to predict the separation performance, as represented by various performance metrics, of MOFs based on high-throughput molecular simulation data [42–46]. Despite this progress, there is still a lack of a quantitative relationship model connecting the MOF structure with the adsorption isotherms. Two main reasons can be suggested. First, the building blocks for constructing MOFs are not easy to define and, importantly, it is difficult to systematically determine which combinations of building blocks can result in stable MOF structures. Second, different adsorbate–adsorbent pairs may exhibit

different types of isotherms, making them difficult to describe with a generic mathematical form.

3.3. Membrane separation processes

In addition to adsorption, membrane separation is widely used for separation tasks. In the conventional development of membrane plants, the selection of membrane materials and the design of membrane-based separation processes are decomposed [47]. However, recent developments in membrane science have shown that a good membrane material needs to balance the tradeoff between permeability and selectivity [48,49]. This tradeoff is only accessible through the optimal design and evaluation of the membrane separation processes.

The major difficulty in simultaneous membrane material and process design has been the lack of a mechanistic model describing the influence of the material synthesis protocol on the membrane properties. Rall et al. [8] propose a hybrid-model-based methodology to simultaneously perform membrane synthesis and process design. A layer-by-layer (LbL) nanofiltration membrane was selected to perform water desalination. For the LbL nanofiltration membrane, the authors built an ANN model to predict the pore radius and layer charge based on the fabrication parameters of the membrane, including the number of polyelectrolyte bi-layers applied (N_{layer}) and the NaCl concentration in the polyelectrolyte coating solution (c_{NaCl}). Two other ANNs were developed for the prediction of salt retention and permeability using the pore radius and layer charge as the inputs. By combining the three ANN models, the authors successfully linked the membrane structural parameters (N_{layer} and c_{NaCl}) with the membrane separation properties indicated by the salt retention and permeability. With these two properties, mechanistic models considering mass balances, constitutive equations, a pump model, and cost models were developed for the membrane process design. By integrating the ANN models with the mechanistic models, N_{layer} and c_{NaCl} as the material structural variables and the feed flow rate as the process variable were simultaneously optimized to minimize the total process cost. The resulting problem was solved by a deterministic global optimization algorithm. It was found that a better separation performance at a lower cost could be achieved through the integrated design, compared with that obtained from a decomposed material and process design. This hybrid modeling method is generalizable. However, it should be noted that the data-driven models were developed based on experimental data for the specific LbL nanofiltration membrane. New data will be required when new types of membranes are to be designed. Moreover, membranes can indeed achieve high-purity separations at a relatively low energy cost. However, they are not suitable for large-scale separations with high feed fluxes. Tula et al. [50] suggest a hybrid distillation–membrane separation scheme for efficient chemical separations. Due to the efficiency in combining different types of models, the hybrid modeling strategy can also play an important role in designing such hybrid separation processes.

3.4. Heterogeneous catalytic processes

Catalysts are widely used in the chemical industry. Most catalysts used in large-scale industrial processes are heterogeneous, with the gas or liquid reaction taking place on the surface of the catalyst [51]. First-principle catalyst design investigates the reaction mechanism, quantifies the rates of elementary steps, and finally establishes a micro-kinetic model [52–54]. The unknown activation energies of the elementary reactions can be linearly correlated with the reaction enthalpy changes through the so-called Polanyi scaling relationship [55]. The reaction enthalpies are further determined by the adsorption or binding energies of key reaction species/intermediates on the catalyst surface. By doing so, the reactivity can finally be related to a few binding energies. The plot of the catalyst activity versus these binding energies is known as the volcano curve [52,55,56]. Given a specific catalyst, these energies, as catalyst descriptors, can be directly obtained through DFT computations [57]. For catalyst design, it is possible to first optimize the catalyst descriptors and then synthesize catalysts to match the optimal descriptor values.

The importance of performing integrated catalyst and reactor design was demonstrated very early by Jacobsen et al. [58]. Using DFT calculations, the researchers proved that, for ammonia synthesis, the maximal reactivity versus the binding energy of nitrogen (i.e., the peak of the volcano) is sensitive to the reactor temperature. Thybaut et al. [59] build a micro-kinetic model and implemented it into a simplified one-dimensional reactor model for the oxidative coupling of methane reaction. The catalyst descriptors and reactor operating conditions were simultaneously optimized to maximize the product yield. Through the optimization, the researchers successfully identified the desired characteristics of the catalyst and the reactor conditions matching the optimal virtual catalyst. Despite this significant progress, the researchers did not really design and synthesize a catalyst, due to the lack of a relationship model linking the catalyst composition and structure to the catalyst descriptors. Fortunately, this kind of relation can now be appropriately described by data-driven models based on data obtained from DFT computations for an optimally designed set of catalyst samples. In fact, data-driven or ML methods have already been used for heterogeneous catalyst design [60]. Ref. [2] provides an overview on the recent developments in data-driven catalyst design. Given this advance, it is expected that hybrid modeling approaches will play an important role in multiscale catalyst and reactor design.

3.5. Organic Rankine cycle processes

Material and process design appears not only in chemical processes, but also in energy processes. The ORC uses low-temperature heat to generate electricity [61]. To efficiently make use of diverse low-temperature heat sources, the ORC process must be tailored to the specific application by optimally designing both the WF and the process operations [62]. Traditionally, WF selection and ORC process optimization are performed sequentially [63]. Such an approach is efficient, but can lead to suboptimal solutions. Schilling et al. [64] present a method for the integrated design of the ORC process and the WF. The perturbed-chain statistical associating fluid theory (PC-SAFT) EoS model [65] is employed to describe the thermodynamic behavior of the WF. A data-driven GC model is used to estimate EoS parameters of the WF from its molecular structure, resulting in the so-called GC-PC-SAFT model [66]. This thermodynamic model is then combined with a mechanistic process model to formulate an MINLP problem. By solving the optimization problem, the optimal WF and ORC process conditions can be identified simultaneously.

Even though the integrated WF and ORC process design problem has been successfully solved, modeling the complex thermodynamic behavior often gives rise to high nonlinearities and even implicit functions that can deteriorate the optimization performance. Considering the long time required to evaluate thermodynamic properties and their derivatives using rigorous models, the idea of local thermodynamic models has been developed and used since the early 1980s [67]. These simplified approximation models were regressed from the data generated from rigorous thermodynamic relationships. Schweidtmann et al. [68] establish a simple data-driven model that learns the WF thermodynamic properties from data generated by a rigorous EoS. By implementing this data-driven property model into the mechanistic process model,

the ORC process was optimized. It was found that, for a pre-defined WF, the central processing unit (CPU) time for the global optimization of the process was reduced to 2.9 h using hybrid models. In contrast, the initial formulation based on the original EoS could not be solved to global optimality within 12 h. Huster et al. [10] further improve the approach by considering WF selection in addition to the process optimization. An ORC process for the waste heat recovery of a diesel truck was considered. A total of 37 suitable WFs were preselected from 122 fluids available in the thermodynamic library CoolProp [69], based on the heat sink temperature. ANN models for predicting the thermodynamic and transport properties of WFs were trained for each of the 37 WF candidates based on data retrieved from CoolProp. By combining these ANN models with mechanistic process models, deterministic global optimization of the ORC process was finally performed for every WF to find the best-matching pressure levels and WF flow rate. It was found that monoaromatic hydrocarbons are a very promising group of WFs for the given application.

3.6. Thermal energy storage processes

Thermal energy (e.g., solar radiation) is widely available and easy to access, which can be stored in the form of latent heat, sensible heat, or both. In comparison with sensible heat storage, latent heat storage (LHS) is more attractive due to the much higher energy density [70]. LHS can be accomplished through solid–liquid and liquid–gas phase transformations. The solid–liquid transition has been proven to be more attractive for use in large-scale TES due to its small volume changes. In such a system, a PCM absorbs heat and melts, and the PCM releases heat when it solidifies. The selection of PCM plays an important role in the development of high-efficiency TES systems. A suitable melting point (T_m) that matches the specific application is the prerequisite for selecting PCMs. For example, materials with a T_m between 0 to 5 °C can be used for food preservation, while a T_m over 60 °C is suitable for solar hot-water generation and industrial waste heat recovery. In addition to a suitable melting point, PCMs must exhibit other desirable properties, such as a high thermal conductivity and heat of fusion as well as a low viscosity and corrosivity [71]. Organic PCMs feature moderate melting points. However, their thermal conductivity is quite low; besides, they are usually volatile and flammable [72]. Inorganic salts have extremely high melting points and suffer from corrosion and supercooling [73]. These drawbacks of the conventional PCMs promote the development of new high-performance TES materials.

Organic salts, also known as ILs, are composed of organic cations and organic or inorganic anions. These compounds are chemically and thermally stable, nonvolatile, and nonflammable. Importantly, their properties can be well tuned by changing the cations, anions, and/or substituents on the ions, which makes ILs designable materials. It has been demonstrated that a well-designed IL can possess a higher heat of fusion than the commercialized PCMs [74,75]. Notably, even though ILs are commonly known to be liquids at or near room temperature, the official definition of ILs uses the boiling point of water as the reference; that is, ILs are organic salts that are liquid below 100 °C [76]. Their wide range of melting temperatures, ranging from negative to 100 °C, and the ability to tailor their properties make ILs very promising PCMs. Despite that, except for a few scattered works that have experimentally investigated the possibility of using ILs as potential PCMs [74,75,77], there is a lack of model-based systematic selection or structural design of ILs for TES applications.

Besides a well-selected PCM, increasing the heat transfer area between the PCM and the heat transfer fluid (HTF) is another critical issue. Usually, this can be achieved by making small PCM spheres encapsulated by, for example, stainless steel and fixing these PCM capsules uniformly into a heat storage tank where the HTF can flow in and out. Such a TES system is governed by a set of heat balances in the form of partial differential equations. Recently, our group has started to investigate integrated IL and TES process design using the hybrid modeling method. Data-driven models are first used to predict various IL properties (heat of fusion, thermal conductivity, melting point, etc.). By combining these data-driven property models with the mechanistic heat balance equations, an optimization problem is formulated and solved to identify the best IL structure and system operating conditions. The result proves that the optimally designed IL shows a higher TES performance than the traditional PCM, paraffin wax, in terms of the average TES power of the system.

4. Concluding remarks

To meet the requirements for more efficient and flexible modern processes, it is necessary to consider multiple design issues at different scales; thus, such processes suffer from multiscale complexity. Due to the strong interaction between the microscopic material scale and the macroscopic process scale, the optimal design of materials and that of the process system must be considered simultaneously. However, solving such a multiscale design problem is very challenging because of the need to integrate multiple models at different scales. Data-driven models have started to play a significant role in material science due to their ability to learn data and behaviors without knowing the underlying physical mechanisms. However, to solely use data-driven models is usually not practical because including the process decision variables as well will result in a large-dimensional design problem, which requires an excessive amount of data for model regression. Hybrid modeling combines the advantages of data-driven and mechanistic models. It has paved a way to solve challenging multiscale design problems. In hybrid modeling, the process-related principles are represented with mechanistic models, while the material properties that are expensive to determine can be described by data-driven models. This article highlighted the significance of hybrid modeling in multiscale material and process design by first introducing the generic design methodology and then discussing the state-of-the-art work in six selected application areas. For those areas where hybrid modeling has not yet been successfully applied, we pointed out opportunities and directions for applying this method. It is our hope that the provided insights and discussions will inspire the further development of this method and stimulate its application in more areas.

Despite the significant progress that has been made, care must be taken when performing hybrid modeling. Some limitations and opportunities need to be pointed out. First, for molecular materials, there are already very sound empirical models to predict their properties, such as the very popular GC models. However, for complex solid materials, such as heterogeneous catalysts and adsorbents, there is a lack of structure-based property prediction models, which certainly deserve more attention and efforts. Second, due to the low extrapolation ability of data-driven models, the optimal design of experiments must be carefully performed in order to increase the generalization ability of these models. Furthermore, in order to reduce the problem dimension (i.e., the amount of training data), data-driven models should only represent a phenomenon with an unclear physical mechanism or property that is expensive to obtain. Third, most property and process models are highly nonlinear, making the resulting MINLP problems difficult to solve to high-quality solutions using standard solvers. Even though a few deterministic global optimization algorithms have been developed to solve problems involving data-driven models [78,79], they are generally limited to problems with

relatively small sizes. For large-scale material and process design problems using hybrid models, more robust global optimization methods with acceptable computational cost still need to be developed.

Acknowledgement

The authors acknowledge the financial support from the Max Planck Society, Germany, for the Computer-Aided Material and Process Design (CAMPD) project.

Compliance with ethics guidelines

Teng Zhou, Rafiqul Gani, and Kai Sundmacher declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Grossmann IE, Westerberg AW. Research challenges in process systems engineering. AIChE J 2000;46(9):1700–3.
[2] Shi H, Zhou T. Computational design of heterogeneous catalysts and gas separation materials for advanced chemical processing. Front Chem Sci Eng 2021;15(1):49–59.
[3] Zhou T, Song Z, Sundmacher K. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. Engineering 2019;5(6):1017–26.
[4] McBride K, Sundmacher K. Overview of surrogate modeling in chemical process engineering. Chem Ing Tech 2019;91(3):228–39.
[5] Von Stosch M, Oliveira R, Peres J, Feyo de Azevedo S. Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput Chem Eng 2014;60:86–101.
[6] Zhang S, Wang F, He D, Jia R. Batch-to-batch control of particle size distribution in cobalt oxalate synthesis process based on hybrid model. Powder Technol 2012;224:253–9.
[7] Zahedi G, Lohi A, Mahdi KA. Hybrid modeling of ethylene to ethylene oxide heterogeneous reactor. Fuel Process Technol 2011;92(9):1725–32.
[8] Rall D, Schweidtmann AM, Aumeier BM, Kamp J, Karwe J, Ostendorf K, et al. Simultaneous rational design of ion separation membranes and processes. J Membr Sci 2020;600:117860.
[9] Zhou T, McBride K, Zhang X, Qi Z, Sundmacher K. Integrated solvent and process design exemplified for a Diels–Alder reaction. AIChE J 2015;61(1):147–58.
[10] Huster WR, Schweidtmann AM, Mitsos A. Working fluid selection for organic Rankine cycles via deterministic global optimization of design and operation. Optim Eng 2020;21(2):517–36.
[11] Zendehboudi S, Rezaei N, Lohi A. Applications of hybrid models in chemical, petroleum, and energy systems: a systematic review. Appl Energy 2018;228:2539–66.
[12] McBride K, Sanchez Medina EI, Sundmacher K. Hybrid semi-parametric modeling in separation processes: a review. Chem Ing Tech 2020;92(7):842–55.
[13] Yang S, Navarathna P, Ghosh S, Bequette BW. Hybrid modeling in the era of smart manufacturing. Comput Chem Eng 2020;140:106874.
[14] Stephanopoulos G, Reklaitis GV. Process systems engineering: from Solvay to modern bio- and nanotechnology: a history of development, successes and prospects for the future. Chem Eng Sci 2011;66(19):4272–306.
[15] Papadopoulos AI, Tsivintzelis I, Linke P, Seferlis P. Computer-aided molecular design: fundamentals, methods, and applications. Ref Module Chem Mol Sci Chem Eng 2018.
[16] Zhang L, Babi DK, Gani R. New vistas in chemical product and process design. Annu Rev Chem Biomol Eng 2016;7(1):557–82.
[17] Uhlemann J, Costa R, Charpentier JC. Product design and engineering—past, present, future trends in teaching, research and practices: academic and industry points of view. Curr Opin Chem Eng 2020;27:10–21.
[18] Fung KY, Ng KM, Zhang L, Gani R. A grand model for chemical product design. Comput Chem Eng 2016;91:15–27.
[19] Alshehri AS, Gani R, You F. Deep learning and knowledge-based methods for computer-aided molecular design—toward a unified approach: state-of-the-art and future directions. Comput Chem Eng 2020;141:107005.
[20] Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: an introduction and review of tools, applications, and solution techniques. Chem Eng Res Des 2016;116:2–26.
[21] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: recent applications and prospects. NPJ Comput Mater 2017;3:54.
[22] Zhou T, Song Z, Zhang X, Gani R, Sundmacher K. Optimal solvent design for extractive distillation processes: a multiobjective optimization-based hierarchical framework. Ind Eng Chem Res 2019;58(15):5777–86.
[23] Lek-utaiwan P, Suphanit B, Douglas PL, Mongkolsiri N. Design of extractive distillation for the separation of close-boiling mixtures: solvent selection and column optimization. Comput Chem Eng 2011;35(6):1088–100.
[24] Kossack S, Kraemer K, Gani R, Marquardt W. A systematic synthesis framework for extractive distillation processes. Chem Eng Res Des 2008;86(7):781–92.
[25] Burger J, Papaioannou V, Gopinath S, Jackson G, Galindo A, Adjiman CS. A hierarchical method to integrated solvent and process design of physical CO_2 absorption using the SAFT-γ Mie approach. AIChE J 2015;61(10):3249–69.
[26] McBride K, Sundmacher K. Data driven conceptual process design for the hydroformylation of 1-dodecene in a thermomorphic solvent system. Ind Eng Chem Res 2015;54(26):6761–71.
[27] Austin ND, Samudra AP, Sahinidis NV, Trahan DW. Mixture design using derivative-free optimization in the space of individual component properties. AIChE J 2016;62(5):1514–30.
[28] Reichardt C. Solvents and solvent effects: an introduction. Org Process Res Dev 2007;11(1):105–13.
[29] Reichardt C, Welton T. Solvents and solvent effects in organic chemistry. 4th ed. Weinheim: Wiley-VCH; 2011.
[30] Song Z, Shi H, Zhang X, Zhou T. Prediction of CO_2 solubility in ionic liquids using machine learning methods. Chem Eng Sci 2020;223:115752.
[31] Zhao Y, Gao J, Huang Y, Afzal RM, Zhang X, Zhang S. Predicting H_2S solubility in ionic liquids by the quantitative structure–property relationship method using $S_{\sigma\text{-profile}}$ molecular descriptors. RSC Adv 2016;6(74):70405–13.
[32] Zhao Y, Huang Y, Zhang X, Zhang S. A quantitative prediction of the viscosity of ionic liquids using $S_{\sigma\text{-profile}}$ molecular descriptors. Phys Chem Chem Phys 2015;17(5):3761–7.
[33] Zhou T, Jhamb S, Liang X, Sundmacher K, Gani R. Prediction of acid dissociation constants of organic compounds using group contribution methods. Chem Eng Sci 2018;183:95–105.
[34] Holderbaum T, Gmehling J. PSRK: a group contribution equation of state based on UNIFAC. Fluid Phase Equilib 1991;70(2–3):251–65.
[35] Valencia-Marquez D, Flores-Tlacuahuac A, Vasquez-Medrano R. An optimization approach for CO_2 capture using ionic liquids. J Clean Prod 2017;168:1652–67.
[36] Hasan MMF, First EL, Floudas CA. Cost-effective CO_2 capture based on *in silico* screening of zeolites and process optimization. Phys Chem Chem Phys 2013;15(40):17601–18.
[37] First EL, Hasan MMF, Floudas CA. Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. AIChE J 2014;60(5):1767–85.
[38] Liu T, First EL, Faruque Hasan MM, Floudas CA. A multi-scale approach for the discovery of zeolites for hydrogen sulfide removal. Comput Chem Eng 2016;91:206–18.
[39] Wiersum AD, Chang JS, Serre C, Llewellyn PL. An adsorbent performance indicator as a first step evaluation of novel sorbents for gas separations: application to metal–organic frameworks. Langmuir 2013;29(10):3301–9.
[40] Leperi KT, Chung YG, You F, Snurr RQ. Development of a general evaluation metric for rapid screening of adsorbent materials for postcombustion CO_2 capture. ACS Sustainable Chem Eng 2019;7(13):11529–39.
[41] Khurana M, Farooq S. Adsorbent screening for postcombustion CO_2 capture: a method relating equilibrium isotherm characteristics to an optimum vacuum swing adsorption process performance. Ind Eng Chem Res 2016;55(8):2447–60.
[42] Qiao Z, Peng C, Zhou J, Jiang J. High-throughput computational screening of 137953 metal–organic frameworks for membrane separation of a $CO_2/N_2/CH_4$ mixture. J Mater Chem A 2016;4(41):15904–12.
[43] Altintas C, Erucar I, Keskin S. High-throughput computational screening of the metal organic framework database for CH_4/H_2 separations. ACS Appl Mater Interfaces 2018;10(4):3668–79.
[44] Aghaji MZ, Fernandez M, Boyd PG, Daff TD, Woo TK. Quantitative structure-property relationship models for recognizing metal organic frameworks (MOFs) with high CO_2 working capacity and CO_2/CH_4 selectivity for methane purification. Eur J Inorg Chem 2016;2016(27):4505–11.
[45] Gharagheizi F, Tang D, Sholl DS. Selecting adsorbents to separate diverse near-azeotropic chemicals. J Phys Chem C 2020;124(6):3664–70.
[46] Shi Z, Yang W, Deng X, Cai C, Yan Y, Liang H, et al. Machine-learning-assisted high-throughput computational screening of high performance metal–organic frameworks. Mol Syst Des Eng 2020;5(4):725–42.
[47] Abetz V, Brinkmann T, Dijkstra M, Ebert K, Fritsch D, Ohlrogge K, et al. Developments in membrane research: from material via process design to industrial application. Adv Eng Mater 2006;8(5):328–58.
[48] Park HB, Kamcev J, Robeson LM, Elimelech M, Freeman BD. Maximizing the right stuff: the trade-off between membrane permeability and selectivity. Science 2017;356(6343):eaab0530.
[49] Labban O, Liu C, Chong TH, Lienhard JH. Relating transport modeling to nanofiltration membrane fabrication: navigating the permeability–selectivity trade-off in desalination pretreatment. J Membr Sci 2018;554:26–38.
[50] Tula AK, Befort B, Garg N, Camarda KV, Gani R. Sustainable process design & analysis of hybrid separations. Comput Chem Eng 2017;105:96–104.
[51] Srivastava RD. Heterogeneous catalytic science. Boca Raton: CRC Press; 1988.
[52] Cheng J, Hu P. Utilization of the three-dimensional volcano surface to understand the chemistry of multiphase systems in heterogeneous catalysis. J Am Chem Soc 2008;130(33):10868–9.
[53] Lausche AC, Hummelshøj JS, Abild-Pedersen F, Studt F, Nørskov JK. Application of a new informatics tool in heterogeneous catalysis: analysis of methanol dehydrogenation on transition metal catalysts for the production of anhydrous formaldehyde. J Catal 2012;291:133–7.

[54] Wang Z, Cao XM, Zhu J, Hu P. Activity and coke formation of nickel and nickel carbide in dry reforming: a deactivation scheme from density functional theory. J Catal 2014;311:469–80.
[55] Bligaard T, Nørskov JK, Dahl S, Matthiesen J, Christensen CH, Sehested J. The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis. J Catal 2004;224(1):206–17.
[56] Wang Z, Wang HF, Hu P. Possibility of designing catalysts beyond the traditional volcano curve: a theoretical framework for multi-phase surfaces. Chem Sci 2015;6(10):5703–11.
[57] Schumann J, Medford AJ, Yoo JS, Zhao ZJ, Bothra P, Cao A, et al. Selectivity of synthesis gas conversion to C_{2+} oxygenates on fcc(111) transition-metal surfaces. ACS Catal 2018;8(4):3447–53.
[58] Jacobsen CJH, Dahl S, Boisen A, Clausen BS, Topsøe H, Logadottir A, et al. Optimal catalyst curves: connecting density functional theory calculations with industrial reactor design and catalyst selection. J Catal 2002;205(2):382–7.
[59] Thybaut JW, Sun J, Olivier L, Van Veen AC, Mirodatos C, Marin GB. Catalyst design based on microkinetic models: oxidative coupling of methane. Catal Today 2011;159(1):29–36.
[60] Goldsmith BR, Esterhuizen J, Liu JX, Bartel CJ, Sutton C. Machine learning for heterogeneous catalyst design and discovery. AIChE J 2018;64(7):2311–23.
[61] Quoilin S, Broek MVD, Declaye S, Dewallef P, Lemort V. Techno-economic survey of organic Rankine cycle (ORC) systems. Renew Sustain Energy Rev 2013;22:168–86.
[62] Lampe M, Stavrou M, Bücker HM, Gross J, Bardow A. Simultaneous optimization of working fluid and process for organic Rankine cycles using PC-SAFT. Ind Eng Chem Res 2014;53(21):8821–30.
[63] Wang ZQ, Zhou NJ, Guo J, Wang XY. Fluid selection and parametric optimization of organic Rankine cycle using low temperature waste heat. Energy 2012;40(1):107–15.
[64] Schilling J, Lampe M, Gross J, Bardow A. 1-Stage CoMT-CAMD: an approach for integrated design of ORC process and working fluid using PC-SAFT. Chem Eng Sci 2017;159:217–30.
[65] Gross J, Sadowski G. Perturbed-chain SAFT: an equation of state based on a perturbation theory for chain molecules. Ind Eng Chem Res 2001;40(4):1244–60.
[66] Mourah M, NguyenHuynh D, Passarello JP, de Hemptinne JC, Tobaly P. Modelling LLE and VLE of methanol + *n*-alkane series using GC-PC-SAFT with a group contribution k_{ij}. Fluid Phase Equilib 2010;298(1):154–68.
[67] Chimowitz EH, Anderson TF, Macchietto S, Stutzman LF. Local models for representing phase-equilibria in multicomponent, nonideal vapor–liquid and liquid–liquid systems. 1. Thermodynamic approximation functions. Ind Eng Chem Process Des Dev 1983;22(2):217–25.
[68] Schweidtmann AM, Huster WR, Lüthje JT, Mitsos A. Deterministic global process optimization: accurate (single-species) properties via artificial neural networks. Comput Chem Eng 2019;121:67–74.
[69] Bell IH, Wronski J, Quoilin S, Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. Ind Eng Chem Res 2014;53(6):2498–508.
[70] Abhat A. Low temperature latent heat thermal energy storage: heat storage materials. Sol Energy 1983;30(4):313–32.
[71] Sharma A, Tyagi VV, Chen CR, Buddhi D. Review on thermal energy storage with phase change materials and applications. Renew Sustain Energy Rev 2009;13(2):318–45.
[72] Sharma RK, Ganesan P, Tyagi VV, Metselaar HSC, Sandaran SC. Developments in organic solid–liquid phase change materials and their applications in thermal energy storage. Energy Convers Manage 2015;95:193–228.
[73] Kenisarin M, Mahkamov K. Solar energy storage using phase change materials. Renew Sustain Energy Rev 2007;11(9):1913–65.
[74] Terasawa N, Tsuzuki S, Umecky T, Saito Y, Matsumoto H. Alkoxy chains in ionic liquid anions; effect of introducing ether oxygen into perfluoroalkylborate on physical and thermal properties. Chem Commun 2010;46(10):1730–2.
[75] Zhu J, Bai L, Chen B, Fei W. Thermodynamical properties of phase change materials based on ionic liquids. Chem Eng J 2009;147(1):58–62.
[76] Plechkova NV, Seddon KR. Applications of ionic liquids in the chemical industry. Chem Soc Rev 2008;37(1):123–50.
[77] Vijayraghavan R, Rana UA, Elliott GD, MacFarlane DR. Protic ionic solids and liquids based on the guanidinium cation as phase-change energy-storage materials. Energy Technol 2013;1(10):609–12.
[78] Boukouvala F, Floudas CA. ARGONAUT: algorithms for global optimization of constrained grey-box computational problems. Optim Lett 2017;11(5):895–913.
[79] Schweidtmann AM, Mitsos A. Deterministic global optimization with artificial neural networks embedded. J Optim Theory Appl 2019;180(3):925–48.