

Article

Not peer-reviewed version

Modeling of Consolidated Bioprocessing for Bioproduction—A Comprehensive Review

Mark Korang Yeboah* and [Dirk Söffker](#)

Posted Date: 14 November 2025

doi: 10.20944/preprints202511.0920.v1

Keywords: consolidated bioprocessing; polynomial models; response surface methodology; regression models; neural network



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Modeling of Consolidated Bioprocessing for Bioproduction—A Comprehensive Review

Mark Korang Yeboah * and Dirk Söffker

Chair of Dynamics and Control, University of Duisburg–Essen, Lotharstraße, 47057, Duisburg, Germany

* Correspondence: mark.korangyeboah@uni.due.de (M.K.Y)

Abstract

With the rising global population and increasing energy demands, sustainable bioproducts such as bioethanol offer essential alternatives to fossil fuels. Unlike first-generation bioethanol derived from food crops like corn, second-generation bioethanol is produced from lignocellulosic (LC) biomass, a non-food resource that addresses sustainability concerns. Consolidated Bioprocessing (CBP) integrates enzyme production, hydrolysis, and fermentation in a single step, using either microbial consortia or engineered microorganisms, reducing costs and simplifying the process compared to separate hydrolysis and fermentation (SHF) and simultaneous saccharification and fermentation (SSF). However, CBP systems are complex due to the dynamic interactions between microbial consortia, metabolic pathways, and process conditions. Addressing these complexities requires advanced modeling techniques that effectively capture non-linear relationships and optimize process parameters. Machine learning-based models have the potential to advance the field of CBP by enabling data-driven approaches to capture complex bioprocess dynamics, improve prediction accuracy, and optimize bioproduct production in CBP systems, thus paving the way toward commercial viability. This review gives an actual overview of relevant key processes CBP, the current state of modeling CBP, its limitations, and the emerging role of machine learning (ML) as a solution to CBP's modeling challenges. It details recent modeling techniques for CBP, including polynomial models, response surface methodologies, with detailed discussions on regression models and neural network models. In this paper, a summarized review of first-order principle-based modeling approaches as well as data-driven modeling approaches is included, emphasizing advancements that contribute to the scalability and efficiency of CBP for bioproduct production. This review provides new perspectives and insights on the modeling of consolidated bioprocessing for utilizing low-cost lignocellulosic biomass in bioproduction.

Keywords: consolidated bioprocessing; polynomial models; response surface methodology; regression models; neural network

1. Introduction

The growing global demand for energy and the increasing concerns about climate change and fossil fuel depletion have accelerated the search for renewable and sustainable alternatives [1,2]. Among various options, biofuels, particularly bioethanol, have gained significant attention due to their potential to reduce greenhouse gas emissions and reliance on petroleum-based fuels [3]. However, using first-generation bioethanol, derived from food crops such as corn and sugarcane, raises ethical and environmental concerns related to food security and land use [4,5]. This has driven a shift toward second-generation bioethanol, produced from lignocellulosic biomass (LCB), an abundantly available non-food resource from agricultural residues, forest waste, and municipal solid waste [6,7].

Despite its promise, lignocellulosic biomass presents substantial processing challenges because of its complex and recalcitrant structure. The tight cellulose, hemicellulose, and lignin matrix resists enzymatic breakdown, necessitating pretreatment to unlock fermentable sugars [8–10]. Conventional

bioconversion involves multiple steps, including pretreatment, enzymatic hydrolysis, and fermentation, each of which requires specific conditions and adds to the cost and inefficiency of the overall process [7,11]. Consolidated bioprocessing (CBP) addresses these issues by integrating enzyme production, hydrolysis, and fermentation into a single step, reducing process time, cost, and complexity. It can utilize engineered microbes or microbial consortia capable of performing all required functions, offering a promising pathway for cost-effective lignocellulosic bioconversion [3,10,12].

Despite the advantages of CBP, its industrial adoption remains limited due to several technical challenges. These include low product yields, limited strain robustness, inconsistent process performance at scale, and difficulty optimizing the complex interdependent stages of the CBP [13,14]. Addressing these limitations requires precise control and deep understanding of the bioprocess. Traditional empirical modeling techniques such as the response surface methodology (RSM) have been used for process optimization, but they often fail to capture the non-linear and dynamic nature of CBP systems [12,15]. This is where advanced computational approaches, particularly machine learning (ML) and artificial intelligence (AI), offer powerful alternatives. These tools can model complex relationships among variables, optimize conditions, and predict system behavior, ultimately enabling more innovative and reliable bioprocess development [14,15].

This review presents a comprehensive overview of CBP, focusing on the key processes involved, the current state of its modeling, and the limitations that hinder its broader application. It explores traditional and modern modeling strategies used in CBP, including polynomial and response surface methodologies, and advanced approaches such as regression analysis and neural networks. The integration of first principle-based models and data-driven models is discussed, showing how recent advancements, especially those involving machine learning, help to address challenges in prediction, optimization, and scalability. By synthesizing these developments, this review offers new and recent perspectives on the role of modeling in advancing CBP for the efficient and sustainable conversion of low-cost lignocellulosic biomass into valuable bioproducts.

2. Research Advances in CBP for Bioproduction

Consolidated Bioprocessing presents a fully integrated bioprocess in which the generation of enzymes, enzymatic hydrolysis of the released cellulose, and fermentation of the glucose to ethanol all take place in a single process step (Figure 1) [6,13,14]. This process can be accomplished in two different ways: either by cultivating together a community of microbes in efficient reactor systems that have niches for each community member and divide the labor of hydrolysis and fermentation among different species or by using a genetically modified organism (GMO) in conventional bioreactors that can both hydrolyze cellulose and ferment sugars to ethanol [9,13]. This is a way to improve the economic viability of second-generation bioproducts because it shortens the fermentation period and lowers capital and operating expenses. It presents the highest level of process integration for bioprocessing [6,9]. Due to ongoing fermentation, this mechanism also prevents the accumulation of inhibitory quantities of reducing sugars (glucose), which improves saccharification and fermentation to produce bioproducts [3,12,15].

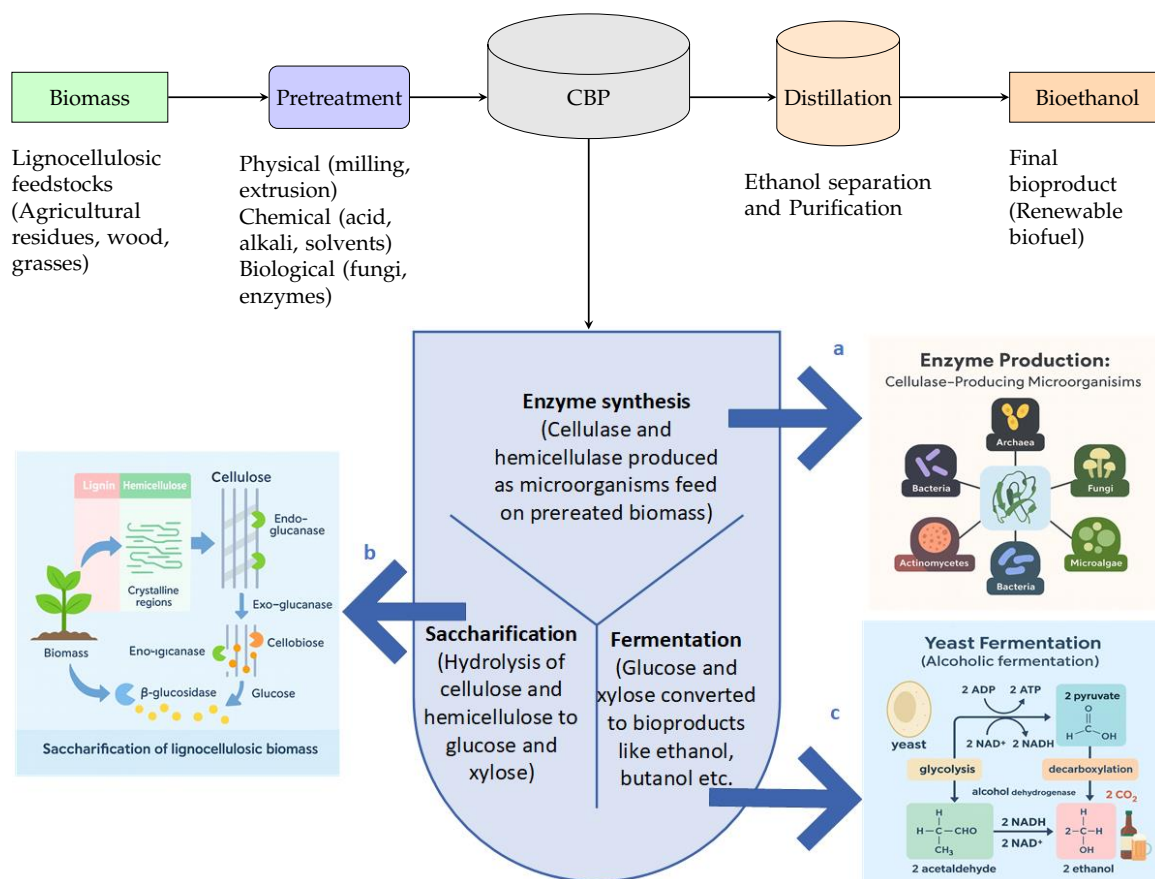


Figure 1. A schematic diagram of consolidated bioprocessing for the production of bioethanol. Redrawn from: a [16], b [17], c [18].

2.1. Enzyme Synthesis

Microbes, such as bacteria, yeast, and algae, can convert organic materials into various types of biofuels, including biodiesel, bioethanol, biogas, and biohydrogen. Microorganisms grow on a substrate of moist solid particles in a process known as solid-state fermentation [19–21]. Water droplets, films, and gas phases in the crevices between the substrate particles provide the microbes with an air supply. Food is obtained from the particle substrate as well as from nutrients dissolved in water droplets or films [3,19]. Given the use of simple cultivation equipment, lower capital costs, greater yields per reactor volume, decreased contamination by bacteria, low effluent generation, and minimal requirements for aeration and agitation, solid-state fermentation has replaced submerged fermentation in the manufacture of some chemicals and enzymes from lignocellulosic (LC) substrate [12,19]. In Table 1, the variety of microbes that can produce certain enzymes for producing biofuels is shown.

Cellulose can be broken down by many bacteria, actinomycetes, and filamentous fungi. Filamentous fungi are widely used in commercial enzyme production because the number of enzymes generated is typically greater than that produced by yeast or bacteria [9,21,22]. In this regard, solid-state fermentation, which mimics the natural environment of these microorganisms, has shown to be an appropriate method for producing these enzymes [23]. The vast majority of cellulases and hemicellulases utilized in industrial applications are generated by filamentous fungi, which are able to secrete enzyme complexes containing significant concentrations of these enzymes [13,20,24,25].

The natural habitats of recognized cellulolytic bacteria, such as *Ruminococcus* and *Clostridium*, are ruminants and insects, respectively [22,26]. These bacteria have shown good and developed fermentation in ruminants, where raw biomass is fed directly and digested in different stomachs [27,28]. As a result, the development made possible by these ruminants and other anaerobic systems, also known as natural CBP, allows for impending advancements in biomass conversion in CBP [24,28,29].

The ability of the host organism to regulate the expression and secretion of these enzymes in response to biomass substrates is crucial for effective CBP performance. Currently, the most researched prospective CBP-enabling bacteria is the obligate anaerobe *Clostridium thermocellum*, which also has the fastest growth rates with crystalline cellulose [19,22,30]. The use of a synthetic co-cultivation system with characterized microbes with specialized roles for the synergistic conversion of biomass/complex substrates, as well as bacterial consortia or microbial consortia that include yeast and fungi, appear as an ideal CBP agent [24,31–33].

One of the key aspects of CBP is the in situ synthesis of cellulolytic and hemicellulolytic enzymes by the microbial host. Unlike traditional bioprocessing methods that require externally produced enzymes, CBP relies on engineered or naturally occurring microorganisms to produce these enzymes during fermentation, thereby reducing costs and simplifying the production workflow. In CBP, efficient enzyme synthesis is critical for the degradation of lignocellulosic biomass into fermentable sugars [34]. Microorganisms such as *Clostridium thermocellum* [35], *Trichoderma reesei* [36], and genetically engineered strains of *Saccharomyces cerevisiae* [24] or *Escherichia coli* [26] have been employed to produce key enzymes like endo- β -glucanase, β -glucosidase, and exo- β -glucanase [20,27].

These enzymes work synergistically to break down cellulose into glucose, which can then be fermented into biofuels or biochemicals. Synergistic activity is necessary for effective saccharification because no single enzyme can break down the cellulose chain into a monomeric unit [33]. Exo and endo- β -glucanases act from both the reducing and the non-reducing ends and exhibit significant synergism [20,37]. By preventing the buildup of cellobiose, β -glucosidase converts cellobiose into glucose, overcoming catabolic repression. The first three phases of cellulase-mediated hydrolysis are

1. Cellulase enzymes attach to the cellulose's surface and adhere to it,
2. Biotransformation of cellulose to fermentable sugars, and
3. Desorption of cellulase.

These processes are primarily controlled by substrate concentration, enzyme dosage, and reaction conditions [38]. For instance, *Trichoderma* produces a higher amount of endo- and exo- β -glucanases and a lower amount of β -glucosidase [38,39]. When *Aspergillus* is present, it generates more endo- β -glucanase, β -glucosidase, and less exo- β -glucanase [17,21,33]. The reducing sugar yield and reaction rates are raised at low substrate concentrations, but they are decreased at high substrate concentrations. The end-product inhibition of the cellulase enzyme causes a reduction in sugar yield and reaction rates at high substrate concentrations [39,40]. While increasing the output of reducing sugar, a high enzyme dose also sharply raises the cost of processing. To solve the problems, one method is to choose the best parameters, such as temperature, pH, and incubation time at low enzyme doses. Additionally, cellulases are negatively impacted by lignin [23,41,42]. Enzyme non-productive adsorption and irreversible binding have an impact on the entire process because they reduce the accessibility of cellulose to cellulase [29,34,37].

Table 1. Functions of lignocellulose-degrading enzymes and their associated microorganisms

Enzymes	Specific types	Function	Microorganisms (Bacterial and Fungal Species)	References
Cellulases	Endoglucanase (EG)	Breaks internal bonds in the cellulose chain, creating new chain ends.	<i>Clostridium sp.</i> , <i>T. reesei</i> , <i>Cellulomonas sp.</i> , <i>T. viride</i> , <i>Thermomonospora sp.</i> , <i>A. niger</i> , <i>Bacillus sp.</i> , <i>P. helicum</i> , <i>Streptomyces sp.</i> , <i>P. betulinii</i> , <i>R. flavefaciens</i> , <i>A. nidulans</i> , <i>Pedobacter sp.</i> , <i>A. fumigatus</i> , <i>F. succinogenes</i> , <i>A. oryzae</i> , <i>R. albus</i> , <i>M. grisea</i> , <i>Mucilaginibacter sp.</i> , <i>N. crassa</i> , <i>F. gramineum</i>	[17,23,25,39,43]
	β -Glucosidase (BG)	Hydrolyzes cellobiose into glucose molecules. Works in synergy with cellulases and hemicellulases to ensure complete sugar release.		
	Exoglucanase (CBH)	Cleaves cellulose from the ends of the chains, releasing cellobiose.		
Hemicellulases	Xylanases	Breaks down xylan (a major component of hemicellulose) by hydrolyzing β -1,4-xylosidic bonds. Converts xylan into shorter oligosaccharides and xylooligosaccharides.	<i>Bacillus sp.</i> , <i>A. niger</i> , <i>P. bryantii</i> , <i>P. betulinii</i> , <i>R. flavefaciens</i> , <i>B. cinerea</i> , <i>P. xylanivorans</i> , <i>A. nidulans</i> , <i>F. succinogenes</i> , <i>A. fumigatus</i> , <i>R. albus</i> , <i>A. oryzae</i> , <i>B. succinogenes</i> , <i>M. grisea</i> , <i>Pedobacter sp.</i> , <i>F. gramineum</i> , <i>Mucilaginibacter sp.</i>	[9,20,33,37]
	Endo- β -1,4-glucanase	Hydrolyzes random internal β -1,4-glycosidic bonds in glucans, including cellulose and hemicellulose. Produces smaller oligosaccharides and enhances accessibility for other enzymes.		
	β -Xylosidase	Cleaves β -1,4-linked xylooligosaccharides into individual xylose units. Complements xylanase by breaking down shorter xylo-oligomers into simple sugars.		
	α -Galactosidase	Hydrolyzes α -1,6-linked galactose residues from galactomannans and other hemicelluloses. Removes side chains from mannans and arabinogalactans, making them easier to degrade.		
	Acetyl esterase	Removes acetyl groups from xylan and other hemicelluloses. Makes xylan more accessible to xylanases by breaking down ester linkages.		
Mannanase	Breaks down mannans (a type of hemicellulose) by hydrolyzing β -1,4-mannosidic bonds. Converts mannans into mannose and oligosaccharides.			
Lignases	Laccase (LaC)	Oxidizes lignin using oxygen, creating radicals for degradation.	<i>A. lipoferum</i> , <i>D. squalens</i> , <i>B. subtilis</i> , <i>G. applanatum</i> , <i>C. basilensis</i> , <i>T. reesei</i> , <i>R. ornithinolytica</i> , <i>T. longibrachiatum</i> , <i>Prevotella sp.</i> , <i>M. tremellosus</i> , <i>Pseudomonas sp.</i> , <i>P. chrysosporium</i> , <i>Pseudobutyrvibrio sp.</i> , <i>C. subvermispora</i> , <i>P. cinnabarinus</i> , <i>Pleurotus sp.</i> , <i>P. rivulosus</i> , <i>Pseudobutyrvibrio sp.</i>	[30,31,41,42]
	Lignin peroxidase (LiP)	Breaks down non-phenolic lignin structures using H ₂ O ₂ .		
	Manganese peroxidase (MnP)	Uses Mn ³⁺ to degrade lignin and open aromatic rings.		
	Versatile peroxidase (VP)	Combines the functions of LiP and MnP to degrade lignin.		

Genetic engineering plays a vital role in enhancing enzyme synthesis in CBP systems [44]. Through synthetic biology and metabolic engineering, researchers have introduced heterologous enzyme genes into industrial strains to broaden their catalytic capabilities. Promoters, signal peptides, and secretion pathways are optimized to increase enzyme yield and activity under fermentation conditions [32,45]. Moreover, efforts are ongoing to develop microbial consortia or co-culture systems, where different organisms contribute complementary enzymatic activities, further boosting the overall efficiency of biomass conversion [31,32]. However, challenges remain in balancing enzyme production with microbial growth and product synthesis. Overproduction of enzymes can impose metabolic burdens on the host, affecting its viability and productivity [46]. Additionally, the complex regulation of enzyme gene expression in response to diverse and often recalcitrant biomass feedstocks demands finely tuned control systems [40]. Despite these challenges, advances in systems biology, omics technologies, and bioprocess optimization are steadily improving the feasibility and economic viability of CBP, improving the sustainable large-scale biofuel production [45].

2.2. Glucose production (hydrolysis)

The majority of cellulose fibers are made of parallel 1,4-linked glucose polymers that are linear and rigid due to significant hydrogen bonding, i.e., the glucose polymers in cellulose are unbranched chains, and the chains are stiff and resist bending due to the extensive hydrogen bonding between the glucose unit [47,48]. These polymers are organized in a crystalline array that is water-resistant, and because cellulose is resistant to hydrolysis by weak acids or alkalis, it can be transformed into fermentable glucose by strong acids and/or high temperatures [34,47]. Such methods have not been widely used because strong acids have to be retrieved, and due to the fact that diluted acids at high temperatures can cause "browning" reactions and a discharge of inhibitors, which diminish glucose yields and can obstruct future ethanol fermentations [48]. As a result, enzymatic hydrolysis employing a mixture of cellulases and cellobiases has been the preferred approach [22,30,40,49].

Cellulases, however, can only digest loose ends of amorphous cellulose at the surface of fibers (exo- β -1,4 glucanases) or sporadic flexible sections of this material (endo- β -1,4 glucanases). Cellobiases are required to convert this disaccharide into glucose because both kinds of enzymes create cellobiose. Owing to the style of the previous delignification operation, which involves severe treatments that increase the number of amorphous sites or loose ends available for cellulase attack, the total hydrolysis rate is slow [24,50,51]. Adsorption of the enzymes to the finite number of target sites is the "rate-limiting" step; hence, increasing the enzyme concentration as in typical enzyme reactions will not enhance the rate of hydrolysis. As a result, cellulose hydrolysis is still sluggish, necessitating the use of numerous reactors. However, in CBP, fermentation, enzymatic hydrolysis, and enzyme synthesis can all be done simultaneously in the fermentation broth. Consolidated bioprocessing is the preferable method of hydrolysis because it eliminates the product inhibition of cellulases by glucose [34,49].

During CBP, microorganisms generate enzymes to break down biomass and create monomeric sugars for their existence. In a natural setting, microorganisms utilize biomass-degrading enzymes in two different ways: the cellulosomal and the free enzyme systems [50,52]. The free enzyme system uses secreted enzymes in their separate forms to react with biomass substrates. The free enzyme system is employed in biorefineries for biomass conversion because it is simple to replicate. The intricate structure of cellulose and hemicellulose necessitates the use of a variety of enzymes to break down different sorts of bonds during the hydrolysis of LC biomass. These molecular scissors enzymes extract particular monomeric sugars from complicated polysaccharides [29,53,54].

2.3. Microbial fermentation

Using fungi, yeast, or bacteria, microbial fermentations transform sugars generated from LC biomass into biofuels or biochemicals [50,55,56]. This procedure can be carried out in conjunction with enzymatic hydrolysis (simultaneous saccharification and fermentation, SSF), independently from enzymatic hydrolysis (separate hydrolysis and fermentation, SHF), or by combining both processes using microorganisms (consolidated bioprocessing, CBP) [51–53]. It is possible to ferment glucose and xylose individually or together in a process called co-fermentation [50,54]. In terms of cost reductions, consolidated bioprocessing is thought to be superior to other fermentation processes. To enable the fermentation of xylose, numerous difficult attempts have been made to genetically change *Saccharomyces cerevisiae* and *Zymomonas mobilis* [29,43,55].

Microorganisms like *Clostridium thermocellum* and *Clostridium phytofermentans* have the ability to directly ferment LC material into biofuels. An anaerobic thermophilic bacterium called *C. thermocellum* ferments at a temperature of about 60 °C. It generates a cellulosome, an enzyme complex that is more effective at degrading cellulose than free enzymes, as demonstrated using pure cellulose samples [49,53,55]. Ethanol and acetic acid make up the majority of *C. thermocellum*'s fermentation products. Despite being an intriguing CBP microbe, *C. thermocellum* is constrained by its poor tolerance for fermentation products [6,39,52]. It can consume nearly all of the LC biomass's sugars and produces free enzymes that can degrade cellulose and hemicellulose. Ethanol is the predominant byproduct of the fermentation of sugar, whereas acetic acid is the minor byproduct. Additionally, *C. phytofermentans*

has a low tolerance for byproducts of fermentation. Another area of interest for researchers is CBP yeast. Here, the goal is to introduce innovative cellulase and hemicellulase genes into a yeast strain that produces biofuel [25,38,54,56].

2.4. Challenges in sugar utilization and bio-product formation

One of the major challenges in consolidated bioprocessing (CBP) is the efficient and balanced utilization of mixed sugars derived from lignocellulosic biomass [57]. Lignocellulose contains both hexose (e.g., glucose) and pentose sugars (e.g., xylose, arabinose), but many native CBP microorganisms lack the ability to metabolize pentoses effectively. Even in engineered strains, glucose repression, a phenomenon in which the presence of glucose inhibits the uptake and metabolism of other sugars, can lead to inefficient sugar utilization [35,44]. This sequential consumption of sugars reduces overall fermentation efficiency and prolongs processing time, negatively impacting productivity and yield [58,59].

Nowadays, the main process configuration for producing LC biofuels is SHF. It is also an innovative practice to use cofermentation microorganisms to ferment enzymatic hydrolysate that comprises both glucose and xylose in simultaneous saccharification and cofermentation (SSCF) [51–53]. The fermentation of xylose occurs significantly more slowly than that of glucose. Xylose transportation depends on glucose transporters because *S. cerevisiae* lacks specialist xylose transporters, which is why slow xylose fermentation has been quantitatively examined by [40,55]. However, glucose transporters prefer glucose over xylose because of their high affinity for glucose. Therefore, xylose fermentation always happens before glucose fermentation by *S. cerevisiae* (and the majority of other microbes) [38]. After the consumption of glucose, there is a sizable quantity of biofuel (such as ethanol) present in the fermentation broth, which has a strong potential to stifle the metabolism of the xylose fermentation [49,56]. Other fermentation metabolites produced during the fermentation of glucose, in addition to ethanol, can be extremely important in preventing the fermentation of xylose. During fermentation, some microbial strains may occasionally take both glucose and xylose simultaneously (without glucose catabolite suppression) (e.g. *Thamnidium elegans* under aerobic conditions) [43,54,57].

The toxicity of sugar degradation intermediates and fermentation byproducts is also a major issue. Compounds such as furfural, hydroxymethylfurfural (HMF), and acetic acid are generated during biomass pretreatment and can inhibit microbial metabolism [8,60,61]. These inhibitors interfere with sugar transport, enzyme activity, and cellular respiration, thereby impairing both sugar utilization and bio-product formation [61,62]. Overcoming this challenge requires developing robust CBP strains with high tolerance to inhibitors or engineering detoxification pathways that can neutralize or bypass the effects of these compounds [63]. The metabolic burden associated with simultaneous enzyme production, sugar catabolism, and bio-product synthesis also poses a significant bottleneck. In CBP, the microbial host must perform multiple tasks within a single system, often leading to resource competition between growth, enzyme expression, and product formation pathways [60,63]. This metabolic load can decrease the efficiency of one or more processes, resulting in suboptimal yields [51]. Engineering strategies such as dynamic pathway regulation, compartmentalization of metabolic functions, or the use of co-culture systems are being explored to alleviate this burden and improve overall process balance [59,64].

Achieving high titers, rates, and yields of target bio-products remains a major obstacle in CBP development. Many promising microorganisms exhibit low productivity under industrial conditions due to factors such as suboptimal enzyme secretion, limited precursor availability, or feedback inhibition in product pathways. Moreover, the complexity of lignocellulosic substrates can introduce variability in sugar release and fermentation performance. Addressing these challenges requires an integrated approach that combines metabolic engineering, process optimization, and systems biology to fine-tune cellular pathways and enhance the robustness and efficiency of CBP systems for commercial-scale bio-production.

2.5. Experimental approaches for optimizing CBP systems

Optimizing CBP systems requires a strategic combination of experimental and statistical approaches to enhance enzyme production, substrate utilization, and overall process efficiency. Various optimization techniques, including classical one-factor-at-a-time (OFAT) methods and advanced statistical models such as response surface methodology (RSM), are employed to identify key process variables and their interactions [65]. Factorial designs, such as the Plackett–Burman design (PBD) for screening significant parameters and the Central Composite Design (CCD) for response optimization, allow for precise control over fermentation conditions, leading to improved yields and cost efficiency [66]. Additionally, bioreactor scaling studies under controlled and uncontrolled conditions provide insights into the kinetic behavior of microbial strains and the feasibility of large-scale production [67]. By integrating these experimental approaches, CBP systems can be fine-tuned to maximize microbial efficiency and enzyme productivity, paving the way for industrial applications in biofuel and bioproduct manufacturing [58,68].

A co-immobilized cultivation system incorporating *Trichoderma reesei*, *Aspergillus niger*, and *Zymomonas mobilis* demonstrated improved enzymatic hydrolysis and fermentation efficiency for bioethanol production [69]. The synergistic action of *T. reesei* and *A. niger* increased saccharification enzyme activity, leading to a cellulose conversion rate of 46.27% and a reducing sugar yield of 2.57 g/L. The immobilization of *Z. mobilis* in alginate beads further enhanced fermentation, achieving a bioethanol yield of 0.56 g/L under optimal conditions [69]. The integration of saccharification and fermentation within a single bioreactor simplified the process, improving bioethanol conversion efficiency. Experimental comparisons reveal that co-immobilized cultivation outperforms suspension cultivation in saccharification enzyme activity, confirming its potential for bioethanol production [68]. The modified bioreactor design successfully reduced operational complexity and costs, although challenges such as fluctuating medium volumes remain [69,70]. Future studies should explore agricultural waste as a feedstock and optimize cultivation conditions to enhance ethanol yields. While the study relied on empirical optimization, integrating computational modeling could improve process predictability, scalability, and overall efficiency.

Metabolic engineering also plays a crucial role in improving CBP-based bioethanol production. A co-culture of *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* achieved an ethanol yield of 38 g/L from 92 g/L of avicel, eliminating organic acid byproducts that typically inhibit fermentation [36]. The study demonstrated that metabolic constraints, rather than cellulolytic efficiency, were the primary limiting factors in ethanol production. With approximately 90% avicel hydrolysis and ethanol titers reaching 80% of theoretical yields, these findings support the potential of genetic engineering in optimizing CBP systems [35]. Further improvements in CBP systems have been observed using membrane-based bioreactors (MBMs) [36]. The combination of *Trichoderma reesei*, *Saccharomyces cerevisiae*, and *Scheffersomyces stipitis* enabled high ethanol yields from undetoxified dilute-acid pretreated wheat straw. Ethanol concentrations reached 7.2 g/L, with 79% of the ethanol diffusing into the gas phase for in situ removal, minimizing toxicity and improving downstream processing efficiency. This approach highlights the advantages of co-culturing cellulolytic and fermentative microorganisms for industrial bioethanol production. Further optimization of strain selection and media composition could enhance the applicability of MBM technology in large-scale operations [36,68].

A stable artificial symbiotic system was developed for ABE (acetone-butanol-ethanol) fermentation through CBP using *Clostridium cellulovorans* and *Clostridium beijerinckii* [71]. This mixed-culture approach facilitated simultaneous saccharification and fermentation of alkali-extracted deshelled corn cobs (AECC) without the need for external enzyme addition. The system efficiently degraded 68.6 g/L of AECC and produced 11.8 g/L of solvents (2.64 g/L acetone, 8.30 g/L butanol, and 0.87 g/L ethanol) within 80 hours, demonstrating a significant improvement over initial conditions. Real-time polymerase chain reaction (PCR) analysis of 16S ribosomal ribonucleic acid (rRNA) gene sequences provided insights into microbial interactions, revealing cooperative and competitive dynamics between *C. cellulovorans* and *C. beijerinckii*, which contributed to improved biobutanol production [71]. The study

also highlights the potential of artificial microbial symbiosis in enhancing biofuel production from lignocellulosic biomass via CBP. The co-culturing approach leveraged the metabolic complementarities of *C. cellulovorans* for cellulose degradation and *C. beijerinckii* for solventogenesis, eliminating the need for externally added butyrate. However, despite its success, the underlying molecular mechanisms that govern interspecies interactions remain unclear, necessitating further studies to fully elucidate the metabolic pathways involved [72]. A summary of different experimental approaches for CBP optimization is shown in Table 2. The empirical optimization of CBP yields good results, but integrating advanced modeling techniques such as RSM, ANN and hybrid modeling can enhance process predictability and scalability. These findings underscore the importance of microbial synergy in biofuel research and provide a foundation for optimizing CBP strategies for cost-effective and sustainable bioethanol and biobutanol production [70,71].

Table 2. Summary of some experimental (Trial-and-error/OFAT) approaches for CBP optimization

Microbial consortia	Substrate	Bioproduct	Yield/Productivity	Reference
Co-culture of <i>Clostridium beijerinckii</i> and <i>Clostridium cellulovorans</i>	Alkali-extracted deshelled corn cobs	Acetone, butanol, ethanol (ABE)	2.64 g/L acetone, 8.30 g/L butanol, 0.87 g/L ethanol; Productivity = 11.8 g/L of ABE solvents	[71]
<i>T. reesei</i> BCRC 31863, <i>A. niger</i> BCRC 3113, <i>Z. mobilis</i> BCRC 10809	Carboxymethyl-cellulose	Bioethanol	Productivity = 0.56 g/L; Reducing sugar conversion = 11.2%	[69]
<i>Clostridium thermocellum</i> and <i>Thermoanaerobacterium saccharolyticum</i>	Avicel	Bioethanol, acetate, lactate	Productivity = 38 g/L of bioethanol	[35]
<i>Trichoderma reesei</i> , <i>Saccharomyces cerevisiae</i> , and <i>Scheffersomyces stipitis</i>	Wheat straw	Bioethanol	Yield = 67%	[36]
<i>Saccharomyces cerevisiae</i> and <i>C. phytofermentans</i>	α -cellulose	Bioethanol	Productivity = 22 g/L bioethanol	[73]
<i>Trichoderma reesei</i> and <i>Candida molischiana</i>	α -cellulose	Bioethanol	Yield = 15%	[74]
<i>Clostridium thermocellum</i> and <i>Clostridium thermolacticum</i>	Micro-crystallized cellulose (MCC)	Bioethanol	Yield = 75%	[75]
<i>Phlebia radiata</i> and <i>Saccharomyces cerevisiae</i>	Waste lignocellulose material	Bioethanol	Productivity = 32.4 g/L	[56]
<i>Acremonium cellulolyticum</i> and <i>Saccharomyces cerevisiae</i>	Solka-Floc (SF)	Bioethanol	Concentration = 8.7– 46.3 g/L	[76]
<i>Acetivibrio thermocellus</i> and <i>Thermoclostridium stercorarium</i>	Mixture of cellulose and Xylan	Bioethanol	Concentration = 40.4 mM	[77]

3. Review of recent modeling approaches for CBP

Modeling in the context of CBP refers to the formulation of computational or mathematical representations that capture biological, chemical, and physical dynamics within a CBP system. These models serve for various purposes, such as understanding system behavior, optimizing process performance, and predicting outcomes under different operating conditions. Broadly, CBP modeling approaches can be categorized into two main types: mechanistic (or first principle-based) models [78] and data-driven models [79]. Mechanistic models rely on established biological, chemical, or physical laws to represent microbial metabolism, enzyme kinetics, and reactor dynamics. These include deterministic models such as Monod kinetics, structured models, and computational fluid dynamics (CFD), as well as stochastic models for simulating system variability [78,80]. In contrast, data-driven models, namely machine learning (ML) and statistical methods like response surface methodology (RSM), do not require detailed prior knowledge of system mechanisms; instead, they infer relationships directly from experimental data [81]. Response surface methodology, though sometimes referred to as a modeling tool, is more accurately a statistical technique for process optimization that generates empirical models, often polynomial equations, to approximate system responses [82]. A third emerging category involves hybrid models, which integrate first principle-based and data-driven elements to leverage the strengths of both approaches [83]. This structured framework will guide the review of recent advancements in CBP modeling, where each modeling category is discussed in detail, highlighting its applications, strengths, and limitations in biofuel process design and optimization.

Modeling of CBP is currently focused on enhancing system efficiency or yield [72,74], and predicting performance [84] under varying process conditions. Researchers have employed deterministic and stochastic models to simulate microbial growth, enzyme production dynamics, substrate degradation, and product formation. Kinetic models, such as Monod-type and structured models, have been utilized to describe the metabolic interactions and resource allocation of CBP-relevant microorganisms like *Clostridium thermocellum* and engineered yeast strains. Additionally, computational fluid dynamics (CFD) and optimization techniques have been applied to predict mixing efficiency, reactor performance, and process bottlenecks [70,80]. To improve the predictive capabilities of CBP models, researchers are integrating multi-scale and hybrid modeling approaches. These models combine genome-scale metabolic network reconstructions with process-level simulations to understand CBP systems' metabolic fluxes and energy trade-offs [1,2,72]. Machine learning algorithms are also being explored to analyze experimental datasets and optimize fermentation conditions in silico. Such approaches allow identifying critical parameters affecting yield and productivity, facilitating targeted genetic and process engineering strategies. Overall, recent modeling efforts are advancing the ability to design and optimize CBP systems for industrial-scale biofuel production, ultimately supporting the transition to more sustainable and economically viable bioprocesses [85,86].

Mathematical modeling has emerged as a valuable tool for optimizing microbial consortia in CBP. The co-culture of *C. phytofermentans*, *Saccharomyces cerevisiae* cdt-1, and *Candida molischiiana* demonstrated long-term stability under controlled oxygen diffusion, producing 22 g/L ethanol from 100 g/L α -cellulose [73]. The addition of exogenous cellulases significantly enhanced ethanol output, emphasizing the importance of enzyme activity regulation. Computational models incorporating ecological theory and metabolic flux analysis could provide insights into microbial interactions, helping to design stable and efficient CBP consortia [73,87]. The potential of modeling extends to optimizing environmental conditions for CBP. In a study by [74] using *Trichoderma reesei* and *Candida molischiiana*, temperature shifts from 30°C to 50°C after 36 hours of cultivation increased reducing sugar and glucose production by 95% and 70%, respectively. This optimization significantly improves ethanol yields, demonstrating the impact of dynamic parameter adjustments. Computational simulations could further refine these conditions, identifying the optimal temperature and pH profiles to maximize the conversion of ethanol [74,87].

The integration of thermophilic and cellulolytic bacteria, such as *Acetivibrio thermocellus* and *Thermoclostridium stercorarium*, has been shown to enhance cellulose degradation and ethanol fermentation [77]. The combined cellulase systems result in an ethanol yield nearly double that of single-strain fermentations. Modeling approaches that incorporate metabolic pathways and enzyme expression dynamics can help optimize these microbial interactions, improving CBP process efficiency and scalability [77]. Empirical optimization methods such as one-factor-at-a-time (OFAT) approaches and trial-and-error experimentation are still widely used in CBP research. However, these methods often fail to account for interactions between multiple process variables, leading to suboptimal results [87]. Computational tools like RSM and ANN provide more efficient alternatives by identifying hidden patterns and optimizing process parameters with minimal experimental effort. Implementing these techniques could significantly improve CBP scalability and cost-effectiveness, accelerating the transition from laboratory research to industrial bioethanol production [70,77,86].

Development in modeling approaches for CBP broadens in scope and sophistication, moving beyond traditional kinetic and stoichiometric approaches to embrace more integrated, multi-scale frameworks. Although foundational models, such as Monod-based kinetics and structured growth simulations, remain vital for understanding microbial metabolism and enzyme dynamics, the field has shifted toward hybrid modeling that couples genome-scale metabolic reconstructions with process-level simulations [88,89]. These approaches aim to capture the interplay between microbial physiology and environmental variables, particularly for key organisms like *Clostridium thermocellum* and engineered yeasts [35]. Additionally, the modeling of microbial consortia has gained attention, with ecological theory and metabolic flux analysis being applied to co-cultures for improved stability

and productivity [90]. Advanced computational tools, including machine learning, response surface methodology (RSM), and artificial neural networks (ANN), are being leveraged to optimize fermentation parameters and predict outcomes under complex, dynamic conditions [87]. Despite these advances, challenges remain in accurately simulating lignocellulose degradation and multi-organism interactions at scale. Nonetheless, the integration of modeling with experimental design continues to drive progress toward scalable, efficient CBP systems tailored for industrial biofuel production [85].

3.1. Polynomial Models

Polynomial models are among the simplest mathematical approaches used to describe bioprocess systems, including CBP. These models utilize polynomial equations to approximate relationships between variables such as substrate concentration, enzyme activity, microbial growth, and product formation. By fitting experimental data to polynomial functions, researchers can derive empirical models that capture general trends in the system behavior [91,92]. Polynomial models are particularly useful for their ease of implementation, low computational demands, and ability to represent nonlinear relationships over limited ranges of process parameters. They are often employed in the initial stages of process modeling or optimization, where a quick and approximate understanding of system behavior is sufficient [93].

Despite their simplicity, polynomial models have significant limitations when applied to complex bioprocesses like CBP. These models are inherently empirical and lack the mechanistic basis to account for the intricate biological interactions occurring within the system. As a result, they are often inadequate for accurately predicting process performance outside the range of experimental data used for model calibration [91,92]. For instance, polynomial models may fail to capture enzyme-substrate dynamics, metabolic shifts, or inhibition effects that are critical for CBP processes. Moreover, as the degree of the polynomial increases to improve accuracy, the models become susceptible to overfitting, reducing their reliability and robustness for practical applications [86,94].

Polynomial models were evaluated to determine their suitability for modeling CBP for bioethanol production using encoded microbial consortia (EMC) datasets [95]. The models included linear, quadratic, cubic, and quartic polynomials, tested on experimental secondary datasets with 96 unrefined and 73 refined data points. The performance of these models was suboptimal, as reflected in performance metrics such as coefficient of determination (R^2) and root mean square error (RMSE). For the 96-point dataset, the R^2 value was 0.552 for the linear model but dropped to -2.78×10^{13} , -4.97×10^{27} , and -4.33×10^{29} for quadratic, cubic, and quartic models, respectively. Similar trends were observed with the 73-point refined dataset, where the linear model achieved an R^2 of 0.814, better than 96 unrefined datasets, and a similar drop in subsequent polynomial models. The RMSE values further supported this, with the linear models having moderate RMSE values (12.6 and 7.56 for the 96 and 73-point datasets, respectively) compared to high RMSE values for higher-order polynomial models, such as 1.16×10^{16} for the 93 dataset and 1.06×10^{16} for the 73 dataset for quartic models [95].

The study by [94] provides a clear illustration of the limitations of polynomial models when applied to CBP systems. The analysis reveals that standard polynomial equations, particularly higher-order models, failed to capture the complex, non-linear interactions inherent in CBP processes. This was evidenced by negative R^2 values and F-statistics of zero, with corresponding p-values of 1, statistical indicators of overfitting and lack of model significance. These results suggest that, in this specific case, polynomial models offered limited predictive power and provided little insight into variable relationships [94]. These findings underscore the limitations of polynomial regression in modeling CBP with EMC for bioethanol production. While linear models display some explanatory power, the overall inadequacy of all polynomial models suggests the need for alternative approaches. Techniques such as machine learning, which can accommodate the intricate, non-linear relationships inherent in CBP systems, or first principle-based modeling frameworks, may offer superior performance. Future studies should focus on these advanced methodologies to enhance model accuracy and utility for optimizing CBP processes.

Consequently, polynomial models are not sufficient for providing reliable predictions in CBP systems, particularly under dynamic or large-scale conditions [94,95]. The inability to incorporate first principle-based insights into biological processes limits their applicability for scenario analysis, optimization, and scale-up studies. Researchers have recognized the need for more advanced modeling approaches, such as kinetic [67,96], hybrid [89,97], or multi-scale models [68,83], which can better capture the complexity of CBP systems. According to [93,94], while polynomial models may serve as a preliminary tool for exploring data trends, they must be complemented or replaced by first principle-based frameworks to ensure accurate predictions and guide process improvements effectively.

Current use of polynomial models in CBP research reflects their role as simple, low-cost tools for preliminary data analysis, but their limitations with respect to overfitting, poor extrapolation, and inability to capture complex nonlinear relationships have become increasingly evident. While linear models can approximate basic trends, higher-order polynomials often perform poorly due to overfitting and their limited physical interpretability. As such, these models are not suitable for capturing the complex, nonlinear, and dynamic nature of CBP systems, particularly under industrially relevant conditions [78]. The focus in recent methodological developments has shifted toward more sophisticated modeling approaches, such as kinetic, hybrid, and multi-scale frameworks, that better represent biological processes and interactions [98]. There is also growing interest in machine learning techniques, which offer improved predictive power and the ability to uncover hidden patterns in experimental data. Overall, while polynomial models still serve a limited exploratory role, they are being rapidly outpaced by mechanistically informed and data-driven alternatives that are more capable of supporting CBP optimization and scale-up [87].

3.2. Response Surface Methodology

Response surface methodology (RSM) is a powerful statistical and mathematical technique used for process optimization and therefore modeling of bioprocess systems, including CBP. Response surface methodology is designed to explore the relationships between multiple independent variables and one or more response variables, providing a structured approach to optimize process conditions [65,99]. By fitting experimental data to second-order polynomial models, RSM generates a response surface that predicts how changes in process parameters, such as temperature, pH, substrate concentration, and inoculum size, influence outcomes like enzyme activity, microbial growth, or product yield. This approach is particularly useful for identifying optimal operating conditions while minimizing experimental effort through a systematic design of experiments (DoE), such as Box-Behnken or Central Composite Designs [82,100]. One of the key strengths of RSM is its ability to handle multiple interacting variables simultaneously, providing insights into their combined effects on bioprocess performance. In CBP systems, where enzyme synthesis, saccharification, and microbial fermentation occur in a single step, RSM can help identify process bottlenecks and interactions that impact overall efficiency [65,66]. For instance, RSM can reveal optimal substrate loading and fermentation conditions that maximize biofuel production while minimizing inhibitory effects [98]. The graphical representation of response surfaces and contour plots enables researchers to visualize the relationships between variables and responses, facilitating informed decision-making and targeted process adjustments [93,100].

A response surface methodology (RSM)-based optimization of saccharification variables for [TEA][H₂SO₄]-pretreated *Saccharum spontaneum* biomass under one-pot consolidated bioprocessing (OPCB) demonstrated substantial improvements in sugar yield and cellulose conversion [101]. The key parameters (enzyme load, temperature, time, and pH) were systematically optimized, with temperature identified as the most influential factor. The study achieved a 2.7-fold increase in sugar yield (up to 531.0 mg/g biomass) and 93.7% cellulose conversion efficiency. The model shows high predictive accuracy, with close agreement between experimental and predicted values [66,101]. Response surface methodology was also employed in [99] to model and optimize CBP for bacterial alkaline phosphatase (ALP) production. A rotatable central composite design (RCCD) was specifically used to evaluate the effects of key components of the medium, molasses, ammonium nitrate (NH₄)₂NO₃, and potassium chloride (KCl), on ALP productivity. Following initial screening via the Plackett–Burman design (PBD),

RCCD enabled a more precise determination of the optimal concentrations of these significant factors by fitting a second-order polynomial regression model [99]. The experimental design consisted of eighteen trials, including factorial, axial, and center points, to assess interactions among the selected variables systematically. This statistical approach results in a substantial increase in ALP production, achieving up to a 94-fold improvement over the basal medium. The findings highlight the effectiveness of RSM in optimizing fermentation parameters for enhanced bioprocess efficiency, demonstrating its potential for industrial-scale ALP production [99].

Using RSM, Selvakumar et al [102] optimized ethanol production from *Manihot esculenta* Crantz YTP1 stems via consolidated bioprocessing (CBP) involving *Cellulomonas fimi* and *Zymomonas mobilis*. Acid pretreatment using acetic, nitric, and mixed acids was optimized for delignification, achieving 85% lignin removal and 87.45% cellulose release. Response surface methodology (RSM) and artificial neural networks (ANN) were employed to refine process parameters, pH, temperature, agitation, and time, leading to maximum cellulase activity (11.63 IU/mL) and ethanol production (9.39 g/L). To evaluate the statistical reliability of the model, an analysis of variance (ANOVA) was conducted. The results indicate a statistically significant quadratic model, with high R^2 values (0.9679 for cellulase and 0.9691 for ethanol), confirming the model's robustness and predictive validity [102]. The study identified optimal conditions (pH 5, 31 °C, 150 rpm, 24 h) with a desirability score of 1.00, which was experimentally validated, affirming RSM's effectiveness in predicting outcomes and supporting the scalability of CBP for ethanol production [102]. However, while RSM offers significant advantages in process optimization, it is not without limitations. It relies on polynomial equations, which limits its predictive power when modeling complex systems. In highly dynamic and nonlinear CBP processes, RSM may fail to capture intricate biological and metabolic interactions beyond the range of experimental data. Additionally, its reliance on empirical fitting shows the lack of mechanistic insights into underlying processes, limiting its utility for extrapolation or scenario analysis [66,100]. Despite these limitations, RSM remains a valuable tool for preliminary process optimization and identifying critical process parameters, often serving as a complement to more advanced kinetic and first principle-based modeling approaches in CBP research [82,96]

Response surface methodology is currently a well-established approach for optimizing CBP-related processes, particularly effective in identifying key parameter interactions and determining local optima for variables such as temperature, pH, and substrate concentration. It enables efficient experimental design and reduces the number of trials needed to achieve meaningful process improvements, making it valuable in early-phase development and pilot-scale studies [100]. However, due to its foundation in empirical second-order polynomial equations, it lacks the capacity to capture the complex biological mechanisms, nonlinear feedback, and dynamic behavior inherent in consolidated bioprocessing systems [65]. This limits its predictive power, especially outside the experimental design space or in rapidly changing process conditions. The methodological focus has therefore shifted toward integrating response surface methodology with more advanced tools, such as machine learning algorithms and first principle-based or kinetic models, to enhance robustness and predictive accuracy [98]. These hybrid approaches represent a growing trend in CBP modeling, where RSM serves as a preliminary or complementary technique rather than a standalone solution.

3.3. Machine Learning-Based Modeling of CBP

Machine learning (ML)-based modeling has emerged as a powerful approach to address CBP dynamics, providing a data-driven alternative to traditional modeling methods. Machine learning techniques, such as artificial neural networks (ANNs), decision trees, and support vector machines (SVM), can analyze large experimental datasets to identify hidden patterns and relationships among process variables [79,103,104]. Unlike polynomial or kinetic models, ML methods do not require explicit assumptions about system behavior, making them highly adaptable to nonlinear and dynamic processes typical of CBP [96,105]. For instance, ML models can predict biofuel yields, enzyme production rates, and substrate consumption based on temperature, pH, substrate concentration, and microbial growth conditions, offering improved accuracy over empirical methods [106,107].

One of the key advantages of ML-based models is their ability to handle high-dimensional, multivariate datasets and account for complex interactions between process variables [103,107]. In CBP systems, where multiple biological processes such as enzyme production, saccharification, and fermentation occur simultaneously, ML algorithms can efficiently integrate data from various sources, including experimental results, omics data, and process monitoring outputs [81,108]. Moreover, ML techniques such as deep learning and ensemble methods enable continuous learning and adaptation, enhancing their predictive capabilities as more data becomes available [105,107,109]. However, despite their potential, ML models require large and high-quality datasets for training and validation, and their black-box nature often limits interpretability. To address these challenges, researchers are integrating ML with first principle-based models to improve both prediction accuracy and process understanding, enabling more reliable optimization and control of CBP systems [103,106].

In the context of CBP machine learning-based modeling, the choice of learning paradigm directly shapes both the predictive capacity of the model and the degree of reliance on domain expertise. Supervised learning approaches, such as support vector machines or neural networks, have been widely applied to predict key process outcomes (e.g., ethanol yield, enzyme activity), but their accuracy depends strongly on the availability and quality of annotated experimental datasets, which are often scarce in CBP research. Unsupervised methods, including clustering and dimensionality reduction techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), are instead valuable for exploring high-dimensional omics or process datasets, where prior labeling is not feasible, enabling the discovery of latent structures guiding hypothesis generation. Semi-supervised strategies provide a compromise when annotated data are limited, improving robustness by leveraging larger pools of unlabeled data. Thus, beyond technical distinctions, the selection of supervised, unsupervised, or semi-supervised learning reflects a trade-off between data availability and the extent to which expert knowledge is required to develop, interpret, and validate models tailored for CBP applications [108,110,111].

3.3.1. Regression Models

Regression models are a fundamental subset of machine learning techniques applied to model and predict outcomes in CBP. These models aim to establish mathematical relationships between input variables (e.g., substrate concentration, enzyme activity, pH, and temperature) and output responses (e.g., biofuel yield, microbial growth, or product formation). Linear regression is often the simplest approach, where the response variable is expressed as a weighted sum of predictor variables [112]. Although linear regression can be useful for initial analyses, CBP processes are inherently nonlinear, requiring the use of more sophisticated regression models such as polynomial regression, ridge regression, and lasso regression to improve accuracy [67]. These techniques allow researchers to model nonlinear trends and penalize overfitting, leading to more robust predictions [103]. Advanced regression methods, such as support vector regression (SVR), decision tree regression, and random forest regression, have proven particularly effective in handling the nonlinear and complex nature of CBP systems [97]. For instance, SVR maps input data into higher-dimensional spaces using kernel functions, enabling the capture of intricate relationships between variables [112]. Similarly, decision tree-based regression models, such as random forests and gradient-boosted trees, partition data into subsets based on variable importance and predict outcomes with high accuracy [104]. These methods are advantageous for CBP modeling as they can manage multicollinearity, variable interactions, and missing data while maintaining interpretability. Such regression models can predict CBP process performance under various conditions, offering insights into optimal parameters for enhancing biofuel yields and process efficiency [81].

In the modeling of consolidated bioprocessing using regression models by [95], data preprocessing plays a crucial role in improving model accuracy by minimizing noise caused by inconsistent readings, measurement errors, and outliers. A residual plot analysis guided the removal of 23 out of 96 secondary data points, effectively reducing the residual interval from (-60, 60) to (-3, 3), thereby enhancing prediction reliability. The refined dataset of 73 data points demonstrates significantly improved model

performance, as indicated by reduced RMSE and MSE values, along with an increased R^2 value of 1.000, signifying near-perfect model fit. Among the 24 models assessed, Gaussian Process Regression (GPR) models consistently ranked among the top six across all evaluation metrics, highlighting their ability to capture the underlying patterns in bioethanol production data. The comparison of predicted versus actual responses further confirms that the refined dataset yielded predictions closer to the ideal prediction line, affirming the effectiveness of data preprocessing in regression modeling for consolidated bioprocessing [84,95]. Further evaluation of Gaussian process regression (GPR) models by [95] shows that matern 5/2 GPR performs best with 53 original training data points, while exponential GPR performs well with 2559 seeded and refined synthetic data points. Despite the superior training performance of the 53 data points model, evidenced by lower RMSE, MSE, and MAE values, the test results for both models were comparably close. However, the larger dataset (2559 seeded and refined synthetic data points) demonstrates higher prediction speed, indicating a trade-off between accuracy and computational efficiency [105]. The predicted versus true response plots and residual plots further reinforce these findings, as the 53/10 data points (10 original data points for testing) model exhibits residuals closer to the zero line, with an interval of (-6,6) for training and (-2,2) for testing, whereas the 2559/10 data points model shows wider residual intervals of (-60,60) and (-15,15), respectively. A narrow residual interval suggests that the model produces more precise and unbiased predictions [113]. These findings highlight the importance of selecting appropriate dataset sizes and kernel functions in GPR modeling to balance predictive accuracy and generalization capability. Additionally, separating training, validation, and test datasets ensure unbiased model evaluation, preventing overfitting and enhancing the reliability of predictions on unseen data [84,95].

Despite their advantages, regression models also face limitations when applied to CBP systems. While they provide accurate predictions within the range of available data, their performance can decline when extrapolating to untested conditions. Furthermore, regression models often require large datasets to achieve reliable results, as small or noisy datasets can lead to underfitting or overfitting. To address these challenges, researchers are integrating regression models with other machine learning techniques, such as artificial neural networks or hybrid models, to leverage the strengths of multiple approaches [88,97]. By combining regression analysis with experimental and first principle-based knowledge, these models can provide a more comprehensive understanding of CBP systems, improving process optimization and scalability for industrial applications [106,112].

Regression-based modeling of CBP has reached a stage where accurate predictions are feasible within the bounds of available experimental data, particularly through advanced methods such as support vector regression, decision tree models, and Gaussian process regression [67]. These models have shown strong performance in identifying key process parameters and forecasting outcomes such as bioethanol yield, especially when paired with careful data preprocessing and feature selection. However, generalization to untested conditions remains a challenge, as purely data-driven models often lack mechanistic grounding and can struggle with extrapolation. Current methodological developments focus on integrating regression models with other machine learning approaches, such as neural networks, or embedding them within hybrid frameworks that combine empirical data with mechanistic insights [114]. The emphasis of future research should be on improving the robustness, scalability, and interpretability of the model to support practical CBP optimization and decision making in broader operational scenarios [114,115].

3.3.2. Neural Network Models

Neural network models have gained significant attention in recent years for their ability to model complex and nonlinear processes in CBP. These models, inspired by the structure and function of biological neural networks, consist of interconnected layers of neurons that process input data and generate predictions [105,116]. Artificial neural networks (ANNs) are particularly suited for CBP due to their ability to learn intricate patterns and relationships between input variables, such as enzyme activity, substrate concentration, microbial growth, and fermentation conditions, and output responses like product yield and process efficiency [93]. Through training on experimental

datasets, ANNs can accurately model highly nonlinear behaviors that are difficult to capture using traditional regression or polynomial models [88,93]. A key strength of neural network models lies in their flexibility and scalability. Advanced architectures, such as deep neural networks (DNNs) with multiple hidden layers, enable the modeling of complex bioprocesses with numerous interacting variables [116]. For CBP, deep learning approaches can integrate large-scale datasets, including process monitoring data, omics data (e.g., genomics and metabolomics), and environmental conditions, to improve prediction accuracy [109,117]. Techniques such as backpropagation and optimization algorithms adjust the weights and biases of the neural network during training, allowing the model to minimize prediction errors and improve performance [88]. Neural networks can also be adapted to new data over time, making them well-suited for dynamic and evolving CBP processes where variability in biological systems is common [116]. Alternatively, transfer learning is a common approach for this kind of task.

Artificial neural network modeling is applied to optimize cellulase activity and ethanol yield, focusing on selecting an optimal neural network architecture [102]. Various topologies were tested, and the most effective model was a 4-8-1 topology with a hyperbolic tangent (Tanh) function in the hidden layer and a linear function in the output layer. The training process minimized the root mean square error (RMSE) and maximized the coefficient of determination (R^2), demonstrating high predictive accuracy. The ANN model achieved R^2 values of 0.9950, 0.9960, and 0.9851 for training, validation, and test datasets in cellulase activity prediction, while ethanol yield predictions yielded R^2 values of 0.9797, 0.9907, and 0.9804. The final ANN model shows overall R^2 values of 0.9908 for cellulase and 0.9794 for ethanol, confirming its strong ability to predict process outcomes based on input parameters [102]. A comparison between RSM and ANN modeling demonstrates that while both approaches effectively optimize process parameters, ANN provides more accurate predictions with lower RMSE and absolute average deviation (AAD%) values [102]. The ANN model exhibited better data fitting, reducing prediction error compared to RSM. The final optimal conditions result in cellulase activity of 11.63 ± 1.23 IU/mL and ethanol yield of 9.39 ± 0.33 g/L with fewer experimental runs. Although RSM remains a widely used optimization tool, ANN proves to be a superior alternative for process modeling due to its higher precision and ability to capture complex, non-linear relationships in cellulase and ethanol production through CBP [102].

The performance evaluation of ANN models trained on 53 and 2559 data points reveals that models trained with the smaller and experimental secondary dataset achieve better accuracy compared to the refined and seeded synthetic dataset [95]. Using a consistent learning rate of 0.01 and a single hidden layer, the model with 20 neurons demonstrated the lowest MSE for training (0.0180 for 53 data points) and validation (0.3110 for 53 data points). However, the model with 10 neurons produced the lowest MSE for testing (3.2670 for 53 data points), followed by the model with 20 neurons (4.0710 for 53 data points), indicating its robustness in predicting unseen data. In contrast, ANN models trained on 2559 data points exhibited significantly higher MSE values, suggesting the presence of outliers in the refined and seeded synthetic dataset [95]. The correlation coefficient (R) further highlights the difference in model performance, with values exceeding 0.9 for ANN models trained on 53 data points, indicating a strong positive linear correlation, whereas models trained on 2559 data points show a moderate correlation. Overall, the ANN models trained on the smaller experimental dataset achieve superior predictive accuracy and reliability, emphasizing the importance of careful dataset selection and preprocessing in neural network modeling for consolidated bioprocessing [84,95].

Despite their predictive power, neural network models come with certain challenges. They often require large, high-quality datasets for effective training, as insufficient or noisy data can lead to overfitting or poor generalization [108]. Additionally, neural networks are often criticized for their "black-box" nature, as they do not provide direct insights into the underlying biological mechanisms of CBP processes [88]. This lack of interpretability can be a limitation when trying to understand system dynamics or troubleshoot process inefficiencies. To address this, researchers are combining neural networks with first principle-based models and techniques such as explainable AI (XAI) to improve

interpretability while retaining predictive accuracy [118,119]. Overall, neural network models hold immense potential for enhancing CBP optimization, enabling more efficient and scalable biofuel and biochemical production.

Neural network models are currently at the forefront of CBP modeling due to their capacity to capture complex, nonlinear relationships among numerous process variables. Recent studies show that well-structured artificial neural networks (ANNs), even with limited but high-quality datasets, can achieve high predictive accuracy for outcomes like cellulase activity and ethanol yield [117,120]. Methodological progress has focused on optimizing neural network architectures, training parameters, and data preprocessing techniques to enhance robustness and reliability. However, their performance is highly sensitive to data quality and quantity. The “black-box” nature of neural networks limits mechanistic interpretability [117,120]. The current focus is shifting toward hybrid approaches, combining ANNs with first principle-based models or explainable AI, to improve transparency and applicability in real-world bioprocesses. These developments suggest that while neural networks are highly capable for CBP optimization, their integration with interpretable frameworks is key to broader adoption and scalability [116,117]. A summary of recent modeling approaches used in CBP optimization are shown in Table 3.

Table 3. Summary of recent modeling approaches for CBP

Modeling Approach	Microorganisms	Substrate	Bioproduct	Performance Metrics	Reference
RSM	<i>Hangateiclostridium thermocellum</i> KSMK1203 and consortium of <i>Cellulomonas fimi</i> MTCC 24 and <i>Zymomonas mobilis</i> MTCC 92	Pre-treated <i>Allium ascalonicum</i> leaves	Bioethanol	$R^2 = 0.9397$	[37]
	<i>Cellulomonas fimi</i> MTCC 24 and <i>Zymomonas mobilis</i> MTCC 92	Thermo-chemo pretreated <i>Manihot esculenta</i> Crantz YTP1 stem	Cellulase	$R^2 = 0.9679$, RMSE = 0.7943	[102]
Bioethanol			$R^2 = 0.9691$, RMSE = 1.0526		
ANN	<i>Cellulomonas fimi</i> MTCC 24 and <i>Zymomonas mobilis</i> MTCC 92	Thermo-chemo pretreated <i>Manihot esculenta</i> Crantz YTP1 stem	Cellulase	$R^2 = 0.990$, RMSE = 0.5151	[102]
			Bioethanol	$R^2 = 0.979$, RMSE = 0.6575	
	18 different microorganisms	Secondary dataset with different cellulosic substrates	Bioethanol	$R^2 = 0.995$, MSE = 2.529	[95]
		Seeded synthetic dataset with different cellulosic substrates	Bioethanol	$R^2 = 0.636$, MSE = 114.713	
GPR	18 different microorganisms	Secondary dataset with different cellulosic substrates	Bioethanol	$R^2 = 0.9700$, RMSE = 0.2445	[84]
		Seeded synthetic dataset with different cellulosic substrates	Bioethanol	$R^2 = 0.980$, RMSE = 1.826	[95]

3.4. Summary of the State of the Art in First-Order Principles and Data-Driven Modeling of CBP

The current state of the art in modeling CBP can be broadly categorized into two main approaches: first-order principle models and data-driven models. First-order principle models, grounded in

mechanistic and deterministic frameworks, are based on the understanding of fundamental biological, chemical, and physical processes that govern CBP systems. These models incorporate detailed kinetics, such as enzyme-substrate interactions, microbial growth rates, and metabolic fluxes, to predict system behavior [67,89,97]. Commonly used frameworks include Monod kinetics [78], Michaelis-Menten enzyme models [121], and genome-scale metabolic network reconstructions [122]. By leveraging these principles, first-order models provide mechanistic insights into enzyme activity, substrate hydrolysis, and product formation, enabling researchers to optimize CBP parameters based on a clear understanding of the system's dynamics [85,123]. While first-order principle models offer interpretability and biological relevance, they are often limited by their complexity and the need for extensive parameterization. The accuracy of these models relies heavily on precise knowledge of kinetic constants, reaction pathways, and system constraints, which are often difficult to obtain for CBP processes involving multiple microbial species and environmental factors [123]. Additionally, these models struggle to capture the stochastic and dynamic nature of biological systems, particularly at industrial scales where variability in feedstock composition and environmental conditions is significant. This limitation has led to the growing adoption of data-driven models [89,97].

Data-driven models, such as regression methods, neural networks, support vector machines, and ensemble learning techniques, have demonstrated significant potential for modeling and optimizing CBP systems. These approaches perform well at handling high-dimensional, nonlinear, and multivariate datasets, which are often encountered in CBP research [84]. Machine learning models can predict biofuel yields, enzyme production rates, and substrate utilization based on historical data, making them particularly useful for process optimization and scenario analysis [79,81]. Moreover, advancements in deep learning and hybrid modeling frameworks allow researchers to integrate omics data, process monitoring information, and experimental results to improve predictive accuracy. Data-driven approaches complement first-order principle models, providing a robust solution for modeling CBP processes under varying conditions [89,109].

One of the major challenges in bioprocess modeling and optimization is the efficient collection and standardization of available data, as highlighted by [79]. The growing number of studies on consolidated bioprocessing often involve highly variable experimental conditions, making it difficult to integrate data for comprehensive analysis. A more detailed and standardized approach to reporting experimental data is necessary to facilitate data management and improve predictive modeling. In the study contacted by [95], while the microbial consortium was encoded for machine learning, key factors such as the specific strain type (wild or engineered) and the ratio of microorganisms in the consortium were not included in the dataset. These omissions likely contributed to prediction errors, as different microbial strains exhibit varying efficiencies in bioethanol production based on substrate composition. Additionally, the pre-treatment methods applied to lignocellulosic biomass, which significantly impact lignin removal and cellulose accessibility, were not encoded in the training data. The exclusion of pre-treatment efficiency, which directly influences bioethanol yield, may have introduced further inaccuracies in model predictions. Addressing these limitations by incorporating detailed microbial and pre-treatment parameters into future datasets will enhance the accuracy and reliability of bioprocess modeling in consolidated bioprocessing research [95].

Despite their advantages, data-driven models face challenges such as the need for large, high-quality datasets for training and the "black-box" nature of certain algorithms, which can limit interpretability. However, recent innovations in explainable AI and hybrid modeling approaches are bridging the gap between first principle-based and machine learning models [89,124]. By combining the predictive power of machine learning with the interpretability of first-order principles, hybrid models are emerging as a promising solution for CBP modeling. These frameworks leverage the strengths of both approaches, enabling accurate predictions while maintaining the ability to extract biological insights, troubleshoot inefficiencies, and optimize process conditions effectively [81,123].

A comparative table summarizing the advantages and disadvantages of first-order principle and data-driven modeling approaches in CBP, using the qualitative scoring system (++ very strong, + strong, 0 neutral, – weak, -- very weak) is shown in Table 4.

Table 4. Comparison of first-principle vs. data-driven modeling approaches in CBP

Criteria	First principle-based models	Data-driven models
Interpretability and mechanistic insight	++	–
Amount of data required	+	--
Predictive accuracy under known conditions	++	+
Ability to update with new experimental results	–	++
Computational complexity	–	0
Handling multivariate interactions	0	++
Suitability for early-stage research	++	0
Need for system understanding	++	–
Ease of implementation	–	+

Interpretability remains a fundamental distinction between first principle-based and data-driven modeling approaches in CBP. First-order models are inherently transparent, as they are based on established biochemical and physical laws, offering direct insights into how variables like enzyme kinetics or substrate concentrations influence system outcomes [125,126]. This makes them particularly valuable for diagnosing performance issues or guiding rational process optimization [127]. In contrast, data-driven models, especially deep learning frameworks, are often regarded as "black boxes" due to their limited ability to explain internal decision-making processes. While they may perform well in predictive accuracy, their lack of interpretability can hinder the extraction of biological meaning and make it difficult to uncover the root causes of observed phenomena [113,127]. Another key difference lies in their dependency on prior system knowledge and their adaptability. First principle-based models require detailed mechanistic data, such as kinetic rates, metabolic pathways, and system stoichiometry, which may be challenging to obtain for complex or poorly understood CBP systems. Data-driven models bypass this need by learning directly from empirical data, making them better suited for exploratory analyses or cases with limited mechanistic insight [125,126,128]. Furthermore, data-driven models offer superior flexibility; they can be quickly updated or retrained with new datasets, improving over time as more information becomes available. In contrast, first principle-based models often need substantial reformulation or reparameterization to remain valid under changing conditions [129]. This adaptability, along with their ability to model nonlinear, high-dimensional interactions, gives data-driven approaches a clear edge in dynamic and heterogeneous CBP environments [130].

First-order and data-driven models differ significantly in their data requirements and ability to handle varying data quality. First-order models often generate reliable predictions from relatively small datasets, provided the necessary kinetic and mechanistic parameters are well-defined [125]. Their reliance on established biological principles allows them to maintain predictive power even in data-sparse environments [113,128]. In contrast, data-driven models typically require large, high-quality datasets for high performance. When data is limited, imbalanced, or noisy, these models are more susceptible to underfitting, overfitting, and reduced generalization, unless robust preprocessing, feature selection, and regularization techniques are applied [131]. When dealing with complex, multivariate datasets, such as those involving omics data or dynamic process variables, data-driven models hold a clear advantage. Their architectures are specifically designed to manage high-dimensional input spaces and uncover hidden patterns within large-scale data. This makes them ideal for modern

CBP applications integrating genomics, proteomics, and environmental monitoring [113,130]. While first-order models can, in theory, incorporate such data, doing so often requires extensive customization and simplifying assumptions that may reduce model fidelity. Additionally, parameter tuning is generally more labor-intensive in first principle-based models, as it requires manual estimation of numerous interdependent variables [129,132]. In contrast, data-driven models benefit from automated optimization routines, which can streamline the calibration process, albeit with significant computational demands. Both approaches, however, struggle with missing or incomplete data, with data-driven models requiring careful imputation strategies and first-order models being limited by the completeness of mechanistic understanding [132].

First-order models also perform well in providing insights into biological mechanisms, offering a clear understanding of cause-and-effect relationships within a CBP system, and supporting hypothesis-driven research [133]. However, they may struggle with scalability to industrial conditions due to their reliance on assumptions that might not hold in large-scale, variable environments [127,131]. On the other hand, data-driven models are better suited for industrial scalability, as they can adapt to real-time operational data and learn from it. Still, they often come with higher computational costs, especially when employing advanced techniques like deep learning. While first-order models are less resource-intensive, particularly when using simplified kinetics or steady-state assumptions, data-driven models require significant computational infrastructure for training and validation [128,133]. A summary of different modeling approaches that can be applied in CBP research, highlighting their potential applications is shown in Table 5.

Table 5. Summary of modeling approaches relevant to CBP research with potential applications

Model Type	Description and Potential Application in CBP
Deterministic models	Using ordinary differential equations (ODEs) to simulate microbial growth, enzyme production, substrate degradation, and product formation. Suitable for controlled systems and can help design predictive bioprocess control strategies.
Stochastic models	Incorporating random variables to account for biological noise and fluctuations in microbial behavior. Useful for microbial consortia, variability in feedstock composition, and uncertain process conditions.
Kinetic models (Monod, structured models)	Description of enzyme kinetics, microbial metabolism, and growth dynamics. Can be extended to include co-culture dynamics and substrate competition in CBP systems.
Computational Fluid Dynamics (CFD)	Simulation of reactor hydrodynamics, mixing patterns, mass transfer, and heat exchange. Can be used to optimize large-scale CBP bioreactors and reduce process bottlenecks.
Multi-scale modeling	Integration of genome-scale metabolic models with process-level dynamics to understand intracellular fluxes and system behavior at different scales. Potentially useful to link metabolic engineering with reactor performance in CBP.
Hybrid models	Combination of mechanistic (first-principle) models with data-driven approaches like support vector machines or random forests to improve prediction accuracy and interpretability. Hybrid models are useful to predict CBP outcomes under novel feedstocks.
Reinforcement learning models	Utilizing reward-based algorithms to optimize process parameters dynamically. Can be applied to adaptive control of CBP processes, e.g., feeding strategies or environmental adjustments.
Evolutionary algorithms	Optimization techniques inspired by natural selection. Can be used to optimize multi-objective CBP process parameters, microbial community composition, or pathway design.

4. Summary and Conclusions

Consolidated bioprocessing (CBP) offers a cost-effective, sustainable route to valorize lignocellulosic biomass by collapsing enzyme production, hydrolysis, and fermentation into a single operation, reducing unit operations relative to SHF/SSF. The literature synthesized here focuses on lignocellulosic CBP for bioethanol and related organic products and surveys modeling approaches spanning first-principles (kinetic/stoichiometric balances, polynomial/response-surface surrogates) to data-driven methods (regularized regression and neural networks). Across enzyme synthesis, hydrolysis, and fermentation, persistent bottlenecks remain in CBP, such as incomplete pentose co-utilization, product/solids inhibition at high loadings, scale-dependent mass-transfer limits, and limited strain robustness under industrial stressors. Purely mechanistic models provide interpretability and scale awareness but underfit strongly non-linear, time-varying couplings among hydrolysis, transport, and metabolism. Purely data-driven models capture complex mappings and enable soft sensing yet lack physical guarantees and extrapolability across feedstocks and operating windows. This complementarity explains the growing relevance of hybrid modeling, where mechanistic structure enforces stoichiometry, thermodynamic feasibility, and transport constraints, while learned components absorb residual kinetics, inhibition cross-terms, and context-specific regulation with uncertainty quantification suitable for control and scale-up. Representative case studies report improved predictive accuracy, tighter control, and faster design cycles relative to single-paradigm baselines.

In recent years, kinetic/stoichiometric reconstructions augmented with machine-learning surrogates, transfer learning, and active-learning design have been developed to accelerate parameter estimation, soft sensing, and control in bioprocessing. Despite these advances, key gaps persist. Most studies remain laboratory-bound, multi-scale data (molecular→bioreactor) are rarely integrated coherently, and multi-omics (genomics, transcriptomics, proteomics, metabolomics) is underexploited for parameterization and validation. From recent developments it follows that incorporating real-time spectroscopy and soft sensors (for sugars, inhibitors, and redox/metabolic state) with adaptive and robust model-predictive control under industrially realistic conditions is essential to mitigate catabolite repression, product/solids inhibition, and consortium drift in CBP. Priority directions include physics-constrained hybrid frameworks that learn residual rates from curated multi-omics and time-series data (standardized protocols and open, well-annotated datasets spanning lab-to-pilot scales for reproducible benchmarking) and integration of digital twins with techno-economic and life-cycle assessments to target cost and risk-dominant levers (solids handling, uptime, in-situ product removal). Ultimately, coupling mechanistic interpretability with data-driven predictive accuracy (and closing the loop with real-time sensing and adaptive control under industrially realistic conditions) will be pivotal to advance CBP from proof-of-concept to reliable, large-scale production of bioethanol and higher-value bioproducts.

Author Contributions: Conceptualization, M.K.Y. and D.S.; formal analysis, M.K.Y. and D.S.; investigation, M.K.Y.; supervision, D.S.; original draft writing, M.K.Y.; review—writing and revision, M.K.Y. and D.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Acknowledgments: This work is partly supported by DAAD, Germany.

Conflicts of 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.

References

1. Sebestyén, V. Renewable and Sustainable Energy Reviews: Environmental impact networks of renewable energy power plants. *Renewable and Sustainable Energy Reviews* **2021**, *151*, 111626.
2. Nastasi, B.; Markovska, N.; Puksec, T.; Duić, N.; Foley, A. Renewable and sustainable energy challenges to face for the achievement of Sustainable Development Goals. *Renewable and Sustainable Energy Reviews* **2022**, *157*, 112071.

3. Li, Z.; Waghmare, P.R.; Dijkhuizen, L.; Meng, X.; Liu, W. Research advances on the consolidated bioprocessing of lignocellulosic biomass. *Engineering Microbiology* **2024**, *4*, 100139. <https://doi.org/https://doi.org/10.1016/j.engmic.2024.100139>.
4. Bertrand, E.; Dussap, C.G. First Generation Bioethanol: Fundamentals—Definition, History, Global Production, Evolution. In *Liquid Biofuels: Bioethanol*; Springer, **2022**; pp. 1–12.
5. De Almeida, M.A.; Colombo, R. Production chain of first-generation sugarcane bioethanol: characterization and value-added application of wastes. *BioEnergy Research* **2023**, *16*, 924–939.
6. Althuri, A.; Gujjala, L.K.S.; Banerjee, R. Partially consolidated bioprocessing of mixed lignocellulosic feedstocks for ethanol production. *Bioresource Technology* **2017**, *245*, 530–539. <https://doi.org/https://doi.org/10.1016/j.biortech.2017.08.140>.
7. Roukas, T.; Kotzekidou, P. From food industry wastes to second generation bioethanol: a review. *Reviews in Environmental Science and Bio/Technology* **2022**, *21*, 299–329.
8. Wei Kit Chin, D.; Lim, S.; Pang, Y.L.; Lam, M.K. Fundamental review of organosolv pretreatment and its challenges in emerging consolidated bioprocessing. *Biofuels, bioproducts and biorefining* **2020**, *14*, 808–829.
9. Olguin-Maciél, E.; Singh, A.; Chable-Villacis, R.; Tapia-Tussell, R.; Ruiz, H.A. Consolidated Bioprocessing, an Innovative Strategy towards Sustainability for Biofuels Production from Crop Residues: An Overview. *Agronomy* **2020**, *10*. <https://doi.org/10.3390/agronomy10111834>.
10. Guo, J.; Liu, D.; Xu, Y. Perspectives and advances in consolidated bioprocessing strategies for lignin valorization. *Sustainable Energy & Fuels* **2024**, *8*, 1153–1184.
11. Moonsamy, T.A.; Mandegari, M.; Farzad, S.; Görgens, J.F. A new insight into integrated first and second-generation bioethanol production from sugarcane. *Industrial Crops and Products* **2022**, *188*, 115675.
12. Singhania, R.R.; Patel, A.K.; Singh, A.; Haldar, D.; Soam, S.; Chen, C.W.; Tsai, M.L.; Dong, C.D. Consolidated bioprocessing of lignocellulosic biomass: Technological advances and challenges. *Bioresource Technology* **2022**, *354*, 127153. <https://doi.org/https://doi.org/10.1016/j.biortech.2022.127153>.
13. Periyasamy, S.; Beula Isabel, J.; Kavitha, S.; Karthik, V.; Mohamed, B.A.; Gizaw, D.G.; Sivashanmugam, P.; Aminabhavi, T.M. Recent advances in consolidated bioprocessing for conversion of lignocellulosic biomass into bioethanol – A review. *Chemical Engineering Journal* **2023**, *453*, 139783. <https://doi.org/https://doi.org/10.1016/j.cej.2022.139783>.
14. Re, A.; Mazzoli, R. Current progress on engineering microbial strains and consortia for production of cellulosic butanol through consolidated bioprocessing. *Microbial Biotechnology* **2023**, *16*, 238–261.
15. Gupte, A.P.; Di Vita, N.; Myburgh, M.W.; Cripwell, R.A.; Basaglia, M.; van Zyl, W.H.; Viljoen-Bloom, M.; Casella, S.; Favaro, L. Consolidated bioprocessing of the organic fraction of municipal solid waste into bioethanol. *Energy Conversion and Management* **2024**, *302*, 118105.
16. Korsá, G.; Konwarh, R.; Masi, C.; Ayele, A.; Haile, S. Microbial cellulase production and its potential application for textile industries. *Annals of Microbiology* **2023**, *73*, 13.
17. Ma, X.; Li, S.; Tong, X.; Liu, K. An overview on the current status and future prospects in *Aspergillus* cellulase production. *Environmental Research* **2024**, *244*, 117866. <https://doi.org/https://doi.org/10.1016/j.envres.2023.117866>.
18. CareerPower. Anaerobic respiration: definition, equation and examples. <https://www.careerpower.in/school/biology/anaerobic-respiration>, 2024. Accessed: 26 September 2025.
19. Kumar, V.; Ahluwalia, V.; Saran, S.; Kumar, J.; Patel, A.K.; Singhania, R.R. Recent developments on solid-state fermentation for production of microbial secondary metabolites: Challenges and solutions. *Bioresource Technology* **2021**, *323*, 124566.
20. Althuri, A.; Mohan, S.V. Sequential and consolidated bioprocessing of biogenic municipal solid waste: a strategic pairing of thermophilic anaerobe and mesophilic microaerobe for ethanol production. *Bioresource technology* **2020**, *308*, 123260.
21. Ramos, M.D.; Sandri, J.P.; Claes, A.; Carvalho, B.T.; Thevelein, J.M.; Zangirolami, T.C.; Milessi, T.S. Effective application of immobilized second generation industrial *Saccharomyces cerevisiae* strain on consolidated bioprocessing. *New biotechnology* **2023**, *78*, 153–161.
22. Wen, Z.; Ledesma-Amaro, R.; Lu, M.; Jin, M.; Yang, S. Metabolic engineering of *Clostridium cellulovorans* to improve butanol production by consolidated bioprocessing. *ACS synthetic biology* **2020**, *9*, 304–315. <https://doi.org/10.1021/acssynbio.9b00331>.
23. Bhardwaj, N.; Kumar, B.; Agrawal, K.; Verma, P. Current perspective on production and applications of microbial cellulases: a review. *Bioresources and Bioprocessing* **2021**, *8*, 95.

24. Minnaar, L.; den Haan, R. Engineering natural isolates of *Saccharomyces cerevisiae* for consolidated bioprocessing of cellulosic feedstocks. *Applied Microbiology and Biotechnology* **2023**, *107*, 7013–7028.
25. Bhati, N.; Shreya.; Sharma, A.K. Cost-effective cellulase production, improvement strategies, and future challenges. *Journal of Food Process Engineering* **2021**, *44*, e13623.
26. Weimer, P.J. Degradation of cellulose and hemicellulose by ruminal microorganisms. *Microorganisms* **2022**, *10*, 2345.
27. Balla, A.; Silini, A.; Cherif-Silini, H.; Bouket, A.C.; Boudechicha, A.; Luptakova, L.; Alenezi, F.N.; Belbahri, L. Screening of cellulolytic bacteria from various ecosystems and their cellulases production under multi-stress conditions. *Catalysts* **2022**, *12*, 769.
28. Fu, R.; Han, L.; Li, Q.; Li, Z.; Dai, Y.; Leng, J. Studies on the concerted interaction of microbes in the gastrointestinal tract of ruminants on lignocellulose and its degradation mechanism. *Frontiers in Microbiology* **2025**, *16*, 1554271.
29. Wen, Z.; Li, Q.; Liu, J.; Jin, M.; Yang, S. Consolidated bioprocessing for butanol production of cellulolytic Clostridia: development and optimization. *Microbial biotechnology* **2020**, *13*, 410–422.
30. Singh, L.K.; Chaudhary, G. *Advances in Biofeedstocks and Biofuels, Biofeedstocks and Their Processing*; John Wiley & Sons, **2016**.
31. Jiang, Y.; Lv, Y.; Wu, R.; Lu, J.; Dong, W.; Zhou, J.; Zhang, W.; Xin, F.; Jiang, M. Consolidated bioprocessing performance of a two-species microbial consortium for butanol production from lignocellulosic biomass. *Biotechnology and Bioengineering* **2020**, *117*, 2985–2995.
32. Gao, H.; Li, M.; Yang, L.; Jiang, Y.; Jiang, W.; Yu, Z.; Zhang, W.; Xin, F.; Jiang, M. Spatial niche construction of a consortium-based consolidated bioprocessing system. *Green Chemistry* **2022**, *24*, 7941–7950.
33. Srivastava, M.; Srivastava, N.; Ramteke, P.W.; Mishra, P.K. *Approaches to enhance industrial production of fungal cellulases*; Springer, **2019**; p. 200.
34. Schlembach, I.; Hosseinpour Tehrani, H.; Blank, L.M.; Büchs, J.; Wierckx, N.; Regestein, L.; Rosenbaum, M.A. Consolidated bioprocessing of cellulose to itaconic acid by a co-culture of *Trichoderma reesei* and *Ustilago maydis*. *Biotechnology for biofuels* **2020**, *13*, 1–18.
35. Argyros, D.A.; Tripathi, S.A.; Barrett, T.F.; Rogers, S.R.; Feinberg, L.F.; Olson, D.G.; Foden, J.M.; Miller, B.B.; Lynd, L.R.; Hogsett, D.A.; et al. High Ethanol Titrers from Cellulose by Using Metabolically Engineered Thermophilic, Anaerobic Microbes. *Applied and Environmental Microbiology* **2011**, *77*, 8288–8294, [<https://journals.asm.org/doi/pdf/10.1128/aem.00646-11>]. <https://doi.org/10.1128/AEM.00646-11>.
36. Brethauer, S.; Studer, M.H. Consolidated bioprocessing of lignocellulose by a microbial consortium. *Energy Environ. Sci.* **2014**, *7*, 1446–1453. <https://doi.org/10.1039/C3EE41753K>.
37. Kavitha, S.; Gajendran, T.; Saranya, K.; Selvakumar, P.; Manivasagan, V.; Jeevitha, S. An insight - A statistical investigation of consolidated bioprocessing of *Allium ascalonicum* leaves to ethanol using *Hangateiclostridium thermocellum* KSMK1203 and synthetic consortium. *Renewable Energy* **2022**, *187*, 403–416. <https://doi.org/https://doi.org/10.1016/j.renene.2022.01.047>.
38. Fang, H.; Deng, Y.; Pan, Y.; Li, C.; Yu, L. Distributive and collaborative push-and-pull in an artificial microbial consortium for improved consolidated bioprocessing. *AIChE Journal* **2022**, *68*, e17844, [<https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.17844>]. <https://doi.org/https://doi.org/10.1002/aic.17844>.
39. Shahab, R.L.; Luterbacher, J.S.; Brethauer, S.; Studer, M.H. Consolidated bioprocessing of lignocellulosic biomass to lactic acid by a synthetic fungal-bacterial consortium. *Biotechnology and Bioengineering* **2018**, *115*, 1207–1215. <https://doi.org/https://doi.org/10.1002/bit.26541>.
40. Sharma, J.; Kumar, V.; Prasad, R.; Gaur, N.A. Engineering of *Saccharomyces cerevisiae* as a consolidated bioprocessing host to produce cellulosic ethanol: Recent advancements and current challenges. *Biotechnology Advances* **2022**, *56*, 107925. <https://doi.org/https://doi.org/10.1016/j.biotechadv.2022.107925>.
41. Cagide, C.; Castro-Sowinski, S. Technological and biochemical features of lignin-degrading enzymes: a brief review. *Environmental Sustainability* **2020**, *3*, 371–389.
42. Atiwesh, G.; Parrish, C.C.; Banoub, J.; Le, T.A.T. Lignin degradation by microorganisms: A review. *Biotechnology Progress* **2022**, *38*, e3226.
43. Amoah, J.; Ishizue, N.; Ishizaki, M.; Yasuda, M.; Takahashi, K.; Ninomiya, K.; Yamada, R.; Kondo, A.; Ogino, C. Development and evaluation of consolidated bioprocessing yeast for ethanol production from ionic liquid-pretreated bagasse. *Bioresource Technology* **2017**, *245*, 1413–1420. SI on Advances in Industrial Bioprocesses and Products- Genetic and Metabolic Engineering Interventions, <https://doi.org/https://doi.org/10.1016/j.biortech.2017.05.171>.

44. Cunha, J.T.; Romani, A.; Inokuma, K.; Johansson, B.; Hasunuma, T.; Kondo, A.; Domingues, L. Consolidated bioprocessing of corn cob-derived hemicellulose: engineered industrial *Saccharomyces cerevisiae* as efficient whole cell biocatalysts. *Biotechnology for biofuels* **2020**, *13*, 138.
45. Yan, Q.; Fong, S.S. Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. *Frontiers in microbiology* **2017**, *8*, 2060.
46. Banner, A.; Toogood, H.S.; Scrutton, N.S. Consolidated bioprocessing: synthetic biology routes to fuels and fine chemicals. *Microorganisms* **2021**, *9*, 1079.
47. Zeng, M.; Pan, X. Insights into solid acid catalysts for efficient cellulose hydrolysis to glucose: progress, challenges, and future opportunities. *Catalysis Reviews* **2022**, *64*, 445–490.
48. Sari, N.K.; Ernawati, D.; Sari, K.N. Optimization of Glucose from *Saccaromyces cereviciae* Liquid Waste Using the Acid Hydrolysis Process. *Nusantara Science and Technology Proceedings* **2024**, pp. 13–17.
49. Lu, J.; Lv, Y.; Jiang, Y.; Wu, M.; Xu, B.; Zhang, W.; Zhou, J.; Dong, W.; Xin, F.; Jiang, M. Consolidated Bioprocessing of Hemicellulose-Enriched Lignocellulose to Succinic Acid through a Microbial Cocultivation System. *ACS Sustainable Chemistry & Engineering* **2019**.
50. Liu, Y.J.; Li, B.; Feng, Y.; Cui, Q. Consolidated bio-saccharification: Leading lignocellulose bioconversion into the real world. *Biotechnology Advances* **2020**, *40*, 107535. <https://doi.org/https://doi.org/10.1016/j.biotechadv.2020.107535>.
51. Li, J.; Chen, B.; Gu, S.; Zhao, Z.; Liu, Q.; Sun, T.; Zhang, Y.; Wu, T.; Liu, D.; Sun, W.; et al. Coordination of consolidated bioprocessing technology and carbon dioxide fixation to produce malic acid directly from plant biomass in *Myceliophthora thermophila*. *Biotechnology for Biofuels* **2021**, *14*, 1–13.
52. Dempfle, D.; Kröcher, O.; Studer, M.H.P. Techno-economic assessment of bioethanol production from lignocellulose by consortium-based consolidated bioprocessing at industrial scale. *New biotechnology* **2021**, *65*, 53–60.
53. He, F.; Chen, J.; Gong, Z.; Xu, Q.; Yue, W.; Xie, H. Dissolution pretreatment of cellulose by using levulinic acid-based protic ionic liquids towards enhanced enzymatic hydrolysis. *Carbohydrate Polymers* **2021**, *269*, 118271. <https://doi.org/10.1016/j.carbpol.2021.118271>.
54. Kumar, V.; Fox, B.G.; Takasuka, T.E. Consolidated bioprocessing of plant biomass to polyhydroxyalkanoate by co-culture of *Streptomyces* sp. SirexAA-E and *Priestia megaterium*. *Bioresource Technology* **2023**, *376*, 128934. <https://doi.org/https://doi.org/10.1016/j.biortech.2023.128934>.
55. Tsai, S.L.; Sun, Q.; Chen, W. Advances in consolidated bioprocessing using synthetic cellulosomes. *Current Opinion in Biotechnology* **2022**, *78*, 102840. <https://doi.org/https://doi.org/10.1016/j.copbio.2022.102840>.
56. Mattila, H.; Kačar, D.; Mali, T.; Lundell, T. Lignocellulose bioconversion to ethanol by a fungal single-step consolidated method tested with waste substrates and co-culture experiments. *AIMS Energy* **2018**, *6*, 866–879. <https://doi.org/10.3934/energy.2018.5.866>.
57. Mazzoli, R. Current progress in production of building-block organic acids by consolidated bioprocessing of lignocellulose. *Fermentation* **2021**, *7*, 248.
58. Sharma, A.; Saini, H.; Thakur, B.; Soni, R.; Soni, S.K. Consolidated bioprocessing of biodegradable municipal solid waste for transformation into biofertilizer formulations. *Biomass Conversion and Biorefinery* **2024**, *14*, 20923–20937.
59. Qiu, Y.; Wu, M.; Bao, H.; Liu, W.; Shen, Y. Engineering of *Saccharomyces cerevisiae* for co-fermentation of glucose and xylose: current state and perspectives. *Engineering Microbiology* **2023**, *3*, 100084.
60. Panagopoulos, V.; Boura, K.; Dima, A.; Karabagias, I.K.; Bosnea, L.; Nigam, P.S.; Kanellaki, M.; Koutinas, A.A. Consolidated bioprocessing of lactose into lactic acid and ethanol using non-engineered cell factories. *Bioresource Technology* **2022**, *345*, 126464.
61. Li, B.; Wang, L.; Wu, Y.J.; Xia, Z.Y.; Yang, B.X.; Tang, Y.Q. Improving acetic acid and furfural resistance of xylose-fermenting *Saccharomyces cerevisiae* strains by regulating novel transcription factors revealed via comparative transcriptomic analysis. *Applied and Environmental Microbiology* **2021**, *87*, e00158–21.
62. Bhavana, B.K.; Mudliar, S.N.; Bokade, V.; Debnath, S. Effect of furfural, acetic acid and 5-hydroxymethylfurfural on yeast growth and xylitol fermentation using *Pichia stipitis* NCIM 3497. *Biomass Conversion and Biorefinery* **2024**, *14*, 4909–4923.
63. Singh, N.; Gupta, A.; Mathur, A.S.; Barrow, C.; Puri, M. Integrated consolidated bioprocessing for simultaneous production of Omega-3 fatty acids and bioethanol. *Biomass and bioenergy* **2020**, *137*, 105555.
64. Bar-Peled, L.; Kory, N. Principles and functions of metabolic compartmentalization. *Nature metabolism* **2022**, *4*, 1232–1244.

65. Upadhyay, A.; Kovalev, A.A.; Zhuravleva, E.A.; Pareek, N.; Vivekanand, V. Enhanced production of acetic acid through bioprocess optimization employing response surface methodology and artificial neural network. *Bioresource Technology* **2023**, *376*, 128930.
66. Breig, S.J.M.; Luti, K.J.K. Response surface methodology: A review on its applications and challenges in microbial cultures. *Materials Today: Proceedings* **2021**, *42*, 2277–2284.
67. Forster, T.; Vázquez, D.; Müller, C.; Guillén-Gosálbez, G. Machine learning uncovers analytical kinetic models of bioprocesses. *Chemical Engineering Science* **2024**, *300*, 120606. <https://doi.org/https://doi.org/10.1016/j.ces.2024.120606>.
68. Maitra, S.; Singh, V. A consolidated bioprocess design to produce multiple high-value platform chemicals from lignocellulosic biomass and its techno-economic feasibility. *Journal of Cleaner Production* **2022**, *377*, 134383.
69. Liu, Y.K.; Yang, C.A.; Chen, W.C.; Wei, Y.H. Producing bioethanol from cellulosic hydrolyzate via co-immobilized cultivation strategy. *Journal of bioscience and bioengineering* **2012**, *114*, 198–203.
70. Olson, D.G.; McBride, J.E.; Shaw, A.J.; Lynd, L.R. Recent progress in consolidated bioprocessing. *Current opinion in biotechnology* **2012**, *23*, 396–405.
71. Wen, Z.; Wu, M.; Lin, Y.; Yang, L.; Lin, J.; Cen, P. Artificial symbiosis for acetone-butanol-ethanol (ABE) fermentation from alkali extracted deshelled corn cobs by co-culture of *Clostridium beijerinckii* and *Clostridium cellulovorans*. *Microbial cell factories* **2014**, *13*, 1–11.
72. Mbaneme-Smith, V.; Chinn, M.S. Consolidated bioprocessing for biofuel production: recent advances. *Energy and Emission Control Technologies* **2015**, *3*, 23–44. <https://doi.org/10.2147/EECT.S63000>.
73. Zuroff, T.R.; Xiques, S.B.; Curtis, W.R. Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic *Clostridium phytofermentans*/yeast co-culture. *Biotechnology for biofuels* **2013**, *6*, 1–12.
74. Bu, Y.; Alkotaini, B.; Salunke, B.K.; Deshmukh, A.R.; Saha, P.; Kim, B.S. Direct ethanol production from cellulose by consortium of *Trichoderma reesei* and *Candida molischiana*. *Green Processing and Synthesis* **2019**, *8*, 416–420. <https://doi.org/doi:10.1515/gps-2019-0009>.
75. Xu, L.; Tschirner, U. Improved ethanol production from various carbohydrates through anaerobic thermophilic co-culture. *Bioresource technology* **2011**, *102*, 10065–10071.
76. Park, E.Y.; Naruse, K.; Kato, T. One-pot bioethanol production from cellulose by co-culture of *Acremonium cellulolyticus* and *Saccharomyces cerevisiae*. *Biotechnology for Biofuels* **2012**, *5*, 1–11.
77. Wang, N.; Yan, Z.; Liu, N.; Zhang, X.; Xu, C. Synergy of Cellulase Systems between *Acetivibrio thermocellus* and *Thermoclostridium stercorarium* in Consolidated-Bioprocessing for Cellulosic Ethanol. *Microorganisms* **2022**, *10*. <https://doi.org/10.3390/microorganisms10030502>.
78. Jin, Q.; Wu, Q.; Shapiro, B.M.; McKernan, S.E. Limited mechanistic link between the Monod equation and methanogen growth: a perspective from metabolic modeling. *Microbiology spectrum* **2022**, *10*, e02259–21.
79. Cheng, Y.; Bi, X.; Xu, Y.; Liu, Y.; Li, J.; Du, G.; Lv, X.; Liu, L. Artificial intelligence technologies in bioprocess: Opportunities and challenges. *Bioresource Technology* **2023**, *369*, 128451. <https://doi.org/https://doi.org/10.1016/j.biortech.2022.128451>.
80. Sadino-Riquelme, M.; Donoso-Bravo, A.; Zorrilla, F.; Valdebenito-Rolack, E.; Gómez, D.; Hansen, F. Computational fluid dynamics (CFD) modeling applied to biological wastewater treatment systems: An overview of strategies for the kinetics integration. *Chemical Engineering Journal* **2023**, *466*, 143180.
81. Khanal, S.K.; Tarafdar, A.; You, S. Artificial intelligence and machine learning for smart bioprocesses. *Bioresource Technology* **2023**, *375*, 128826. <https://doi.org/https://doi.org/10.1016/j.biortech.2023.128826>.
82. Madhuvanthi, S.; Jayanthi, S.; Suresh, S.; Pugazhendhi, A. Optimization of consolidated bioprocessing by response surface methodology in the conversion of corn stover to bioethanol by thermophilic *Geobacillus thermoglucosidasius*. *Chemosphere* **2022**, *304*, 135242.
83. Gopalakrishnan, S.; Johnson, W.; Valderrama-Gomez, M.A.; Icten, E.; Tat, J.; Ingram, M.; Shek, C.F.; Chan, P.K.; Schlegel, F.; Rolandi, P.; et al. COSMIC-dFBA: A novel multi-scale hybrid framework for bioprocess modeling. *Metabolic Engineering* **2024**, *82*, 183–192.
84. Yeboah, M.K.; Asiedu, N.Y.; Dogbe, S.; Addo, A. Performance of Machine Learning Based-Modelling Approach in Consolidated Bioprocessing with Microbial Consortium for Bioethanol Production. *Industrial Biotechnology* **2024**, *20*, 77–97, [<https://doi.org/10.1089/ind.2024.0003>]. <https://doi.org/10.1089/ind.2024.0003>.
85. Dempfle, D.B. Model-based scale-up of a continuously operated consolidated bioprocess based on a microbial consortium for the production of ethanol. PhD thesis, École Polytechnique Fédérale de Lausanne (EPFL), **2022**. <https://doi.org/https://infoscience.epfl.ch/bitstreams/89988351-e1ca-4306-994f-ab1b209478b0/download>.

86. Rendón-Castrillón, L.; Ramírez-Carmona, M.; Ocampo-López, C.; Gómez-Arroyave, L. Mathematical model for scaling up bioprocesses using experiment design combined with Buckingham Pi theorem. *Applied Sciences* **2021**, *11*, 11338.
87. Mangipudi, S.; Reddy, D.G.V.; Ranganathan, P. Computational tools in bioprocessing. In *Current Developments in Biotechnology and Bioengineering*; Elsevier, **2022**; pp. 211–231.
88. Agharafeie, R.; Ramos, J.R.C.; Mendes, J.M.; Oliveira, R. From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. *Fermentation* **2023**, *9*. <https://doi.org/10.3390/fermentation9100922>.
89. Mahanty, B. Hybrid modeling in bioprocess dynamics: Structural variabilities, implementation strategies, and practical challenges. *Biotechnology and Bioengineering* **2023**, *120*, 2072–2091.
90. Liang, Y.; Ma, A.; Zhuang, G. Construction of environmental synthetic microbial consortia: based on engineering and ecological principles. *Frontiers in Microbiology* **2022**, *13*, 829717.
91. Abdelgalil, S.; Soliman, N.; Abo-Zaid, G.; Abdel-Fattah, Y. Bioprocessing strategies for cost-effective large-scale production of bacterial laccase from *Lysinibacillus macroides* LSO using bio-waste. *International Journal of Environmental Science and Technology* **2021**, pp. 1–20.
92. Sriraman, V.; Johnrajan, J.; Yazhini, K.; Rathinasabapathi, P. Bioprocess engineering essentials: cultivation strategies and mathematical modeling techniques. In *Industrial microbiology and biotechnology: a new horizon of the microbial world*; Springer, **2024**; pp. 247–276.
93. de Andrade Bustamante, R.; de Oliveira, J.S.; Dos Santos, B.F. Modeling biosurfactant production from agroindustrial residues by neural networks and polynomial models adjusted by particle swarm optimization. *Environmental Science and Pollution Research* **2023**, *30*, 6466–6491.
94. do Nascimento, J.F.C.; dos Reis, B.D.; de Baptista Neto, Á.; Lerin, L.A.; de Oliveira, J.V.; de Paula, A.V.; Remonato, D. Comparing a polynomial DOE model and an ANN model for enhanced geranyl cinnamate biosynthesis with Novozym® 435 lipase. *Biocatalysis and Agricultural Biotechnology* **2024**, *58*, 103240.
95. Yeboah, M.K. Modelling of Consolidated Bioprocessing for Bioethanol Production: Encoded Microbial Consortium for Machine Learning. Master thesis, Kwame Nkrumah University of Science and Technology, Kumasi, Department of Agricultural and Biosystems Engineering, **2024**.
96. Monteiro, M.; Fadda, S.; Kontoravdi, C. Towards advanced bioprocess optimization: A multiscale modelling approach. *Computational and Structural Biotechnology Journal* **2023**, *21*, 3639–3655.
97. O'Brien, C.M.; Zhang, Q.; Daoutidis, P.; Hu, W.S. A hybrid mechanistic-empirical model for in silico mammalian cell bioprocess simulation. *Metabolic Engineering* **2021**, *66*, 31–40.
98. Tsafarakidou, P.; Manthos, G.; Zagklis, D.; Mema, J.; Kornaros, M. Assessment of substrate load and process pH for bioethanol production—Development of a kinetic model. *Fuel* **2022**, *313*, 123007.
99. Abdelgalil, S.A.; Soliman, N.A.; Abo-Zaid, G.A.; Abdel-Fattah, Y.R. Dynamic consolidated bioprocessing for innovative lab-scale production of bacterial alkaline phosphatase from *Bacillus paralicheniformis* strain APSO. *Scientific reports* **2021**, *11*, 6071.
100. Djimtoingar, S.S.; Derkyi, N.S.A.; Kuranchie, F.A.; Yankyera, J.K. A review of response surface methodology for biogas process optimization. *Cogent Engineering* **2022**, *9*, 2115283.
101. Vaid, S.; Sharma, S.; Dutt, H.C.; Mahajan, R.; Bajaj, B.K. One pot consolidated bioprocess for conversion of *Saccharum spontaneum* biomass to ethanol-biofuel. *Energy Conversion and Management* **2021**, *250*, 114880.
102. Selvakumar, P.; Kavitha, S.; Sivashanmugam, P. Optimization of process parameters for efficient bioconversion of thermo-chemo pretreated *Manihot esculenta* Crantz YTP1 stem to ethanol. *Waste and Biomass Valorization* **2019**, *10*, 2177–2191.
103. Mondal, P.P.; Galodha, A.; Verma, V.K.; Singh, V.; Show, P.L.; Awasthi, M.K.; Lall, B.; Anees, S.; Pollmann, K.; Jain, R. Review on machine learning-based bioprocess optimization, monitoring, and control systems. *Bioresource technology* **2023**, *370*, 128523.
104. Razzak, S.A.; Alam, M.S.; Hossain, S.Z.; Rahman, S.M. Tree-based machine learning for predicting *Neochloris oleoabundans* biomass growth and biological nutrient removal from tertiary municipal wastewater. *Chemical Engineering Research and Design* **2024**, *210*, 614–624.
105. Serrano-Bermúdez, L.M. Bioprocesses in the Era of Artificial Intelligence. *Revista Colombiana de Biotecnología* **2025**, *27*, 1–4.
106. Duong-Trung, N.; Born, S.; Kim, J.W.; Schermeyer, M.T.; Paulick, K.; Borisyak, M.; Cruz-Bournazou, M.N.; Werner, T.; Scholz, R.; Schmidt-Thieme, L.; et al. When bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. *Biochemical Engineering Journal* **2023**, *190*, 108764. <https://doi.org/https://doi.org/10.1016/j.bej.2022.108764>.

107. Helleckes, L.M.; Hemmerich, J.; Wiechert, W.; von Lieres, E.; Grünberger, A. Machine learning in bioprocess development: from promise to practice. *Trends in biotechnology* **2023**, *41*, 817–835.
108. Baako, T.M.D.; Kulkarni, S.K.; McClendon, J.L.; Harcum, S.W.; Gilmore, J. Machine learning and deep learning strategies for Chinese hamster ovary cell bioprocess optimization. *Fermentation* **2024**, *10*, 234.
109. Krishna, V.V.; Pappa, N.; Rani, S.J.V. Deep learning based soft sensor for bioprocess application. In Proceedings of the 2021 IEEE second international conference on control, measurement and instrumentation (CMI). IEEE, **2021**, pp. 155–159.
110. Imamoglu, E. Artificial Intelligence and/or Machine Learning Algorithms in Microalgae Bioprocesses. *Bioengineering* **2024**, *11*, 1143.
111. Singh, A.; Singhal, B. Role of machine learning in bioprocess engineering: current perspectives and future directions. *Design and Applications of Nature Inspired Optimization: Contribution of Women Leaders in the Field* **2023**, pp. 39–54.
112. Luo, Y.; Kurian, V.; Ogunnaike, B.A. Bioprocess systems analysis, modeling, estimation, and control. *Current Opinion in Chemical Engineering* **2021**, *33*, 100705.
113. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18.
114. Rogers, A.W.; Song, Z.; Ramon, F.V.; Jing, K.; Zhang, D. Investigating ‘greyness’ of hybrid model for bioprocess predictive modelling. *Biochemical Engineering Journal* **2023**, *190*, 108761.
115. Agharafeie, R.; Oliveira, R.; Ramos, J.R.C.; Mendes, J.M. Application of hybrid neural models to bioprocesses: A systematic literature review. *Authorea Preprints* **2023**.
116. Agharafeie, R.; Ramos, J.R.C.; Mendes, J.M.; Oliveira, R. From shallow to deep bioprocess hybrid modeling: Advances and future perspectives. *Fermentation* **2023**, *9*, 922.
117. Yang, C.T.; Kristiani, E.; Leong, Y.K.; Chang, J.S. Big data and machine learning driven bioprocessing—recent trends and critical analysis. *Bioresource technology* **2023**, *372*, 128625.
118. Angelov, P.P.; Soares, E.A.; Jiang, R.; Arnold, N.I.; Atkinson, P.M. Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2021**, *11*, e1424.
119. Das, A.; Rad, P. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* **2020**.
120. Rodriguez-Granose, D.; Jones, A.; Loftus, H.; Tandeski, T.; Heaton, W.; Foley, K.T.; Silverman, L. Design of experiment (DOE) applied to artificial neural network architecture enables rapid bioprocess improvement. *Bioprocess and biosystems engineering* **2021**, *44*, 1301–1308.
121. Srinivasan, B. A guide to the Michaelis–Menten equation: steady state and beyond. *The FEBS journal* **2022**, *289*, 6086–6098.
122. Clark, T.J.; Guo, L.; Morgan, J.; Schwender, J. Modeling plant metabolism: from network reconstruction to mechanistic models. *Annual Review of Plant Biology* **2020**, *71*, 303–326.
123. Narayanan, H.; Luna, M.F.; von Stosch, M.; Cruz Bournazou, M.N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M. Bioprocessing in the digital age: the role of process models. *Biotechnology journal* **2020**, *15*, 1900172.
124. Rozov, S. Machine Learning and Deep Learning methods for predictive modelling from Raman spectra in bioprocessing. *arXiv preprint arXiv:2005.02935* **2020**.
125. Nazemzadeh, N.; Malanca, A.A.; Nielsen, R.F.; Gernaey, K.V.; Andersson, M.P.; Mansouri, S.S. Integration of first-principle models and machine learning in a modeling framework: An application to flocculation. *Chemical Engineering Science* **2021**, *245*, 116864.
126. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2020**, *10*, e1379.
127. Krishnan, M. Against interpretability: a critical examination of the interpretability problem in machine learning. *Philosophy & Technology* **2020**, *33*, 487–502.
128. Zhang, Z.; Zhang, H. Application of Big Data Analysis in Intelligent Industrial Design Using Scalable Computational Model. *Scalable Computing: Practice and Experience* **2025**, *26*, 1180–1195.
129. Isoko, K.; Cordiner, J.L.; Kis, Z.; Moghadam, P.Z. Bioprocessing 4.0: a pragmatic review and future perspectives. *Digital Discovery* **2024**, *3*, 1662–1681.
130. Yang, S.; Wan, M.P.; Chen, W.; Ng, B.F.; Dubey, S. Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. *Applied Energy* **2020**, *271*, 115147.

131. Habib, M.K.; Ayankoso, S.A.; Nagata, F. Data-driven modeling: concept, techniques, challenges and a case study. In Proceedings of the 2021 IEEE international conference on mechatronics and automation (ICMA). IEEE, 2021, pp. 1000–1007.
132. Han, Z.; Gao, C.; Liu, J.; Zhang, J.; Zhang, S.Q. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608* 2024.
133. Penloglou, G.; Kiparissides, A. Advanced Modeling of Biomanufacturing Processes. *Processes* 2024, 12, 387.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.