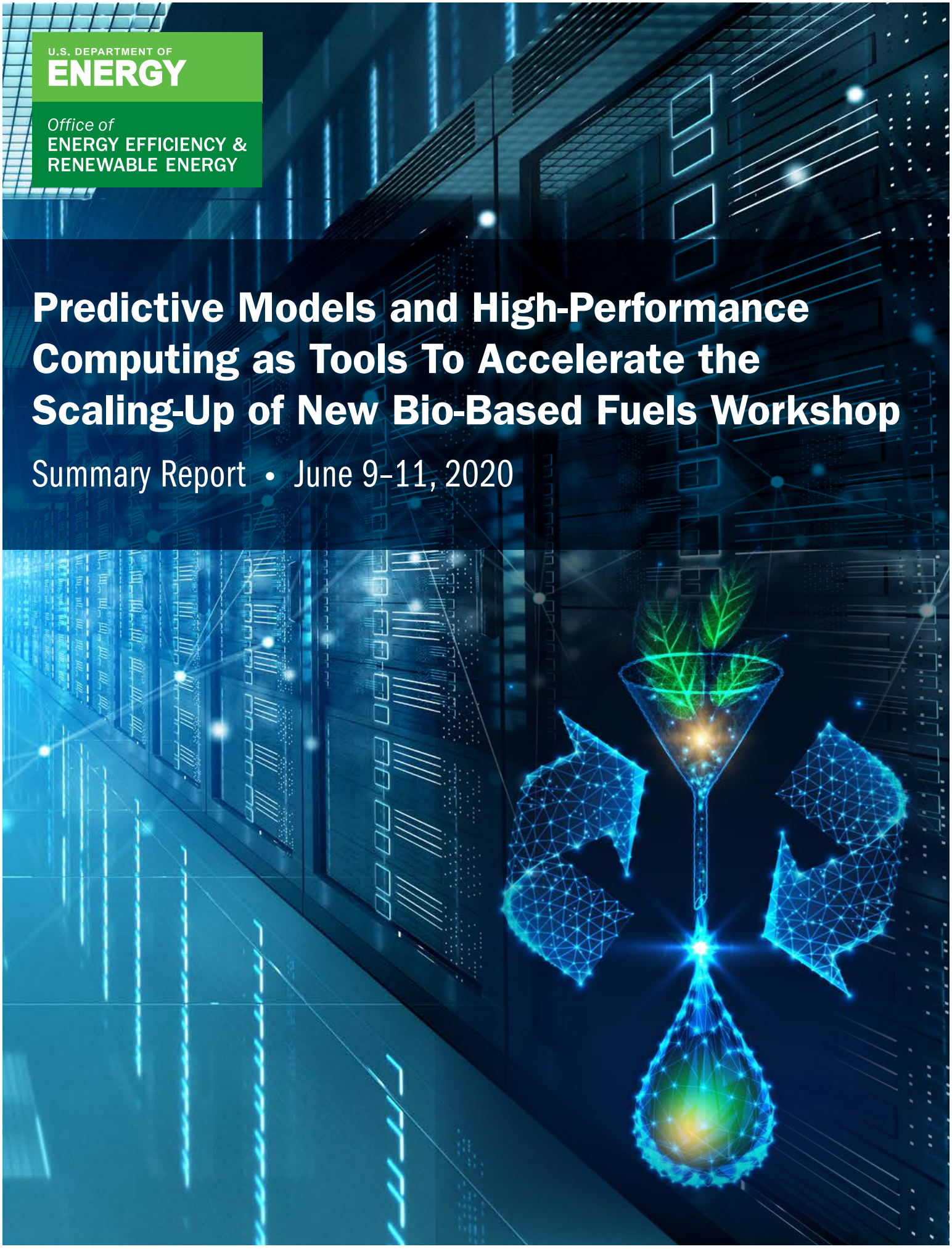


U.S. DEPARTMENT OF
ENERGY

Office of
ENERGY EFFICIENCY &
RENEWABLE ENERGY

Predictive Models and High-Performance Computing as Tools To Accelerate the Scaling-Up of New Bio-Based Fuels Workshop

Summary Report • June 9–11, 2020

(This page intentionally left blank)

Summary report from the virtual "Predictive Models and High-Performance Computing as Tools to Accelerate the Scaling-Up of New Bio-Based Fuels" workshop held on June 9–11, 2020.

Workshop and summary report sponsored by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy's Bioenergy Technologies Office.

Summary report prepared by Lauren Illing, Remy Biron, and Joshua Messner.

Foreword

The U.S. Department of Energy's (DOE's) Office of Energy Efficiency and Renewable Energy (EERE) invests in a diverse portfolio of technologies to ensure domestic energy security, continued economic competitiveness, environmental sustainability, and the availability of cleaner fuels and power.

This report summarizes the input received from attendees of a public workshop sponsored by DOE-EERE on June 9–11, 2020. The views and opinions of the workshop attendees, as summarized in this document, do not necessarily reflect those of the U.S. government or any agency thereof, nor do their employees make any warranty, expressed or implied, or assume any liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe upon privately owned rights.

The authors would like to thank the workshop participants for their contributions during the workshop and to also acknowledge the team of moderators, facilitators, scribes, and logistics support staff who assisted with planning and administering the event.

List of Acronyms

AI	artificial intelligence
BETO	Bioenergy Technologies Office
CCPC	Consortium for Computational Physics and Chemistry
CFD	computational fluid dynamics
ChemCatBio	Chemical Catalysis for Bioenergy Consortium
CPFD	computational particle fluid dynamics
DCS	distributed control system
DOE	U.S. Department of Energy
EERE	Office of Energy Efficiency and Renewable Energy
FCIC	Feedstock-Conversion Interface Consortium
HPC	high-performance computing
IBR	integrated biorefinery
ML	machine learning
NREL	National Renewable Energy Laboratory
PDU	process development unit
PRD	priority research direction
R&D	research and development
TEA	techno-economic analysis

Executive Summary

Investments in integrated biorefinery scale-up have exceeded $2 billion in the United States by private industry and the U.S. Department of Energy (DOE), and yet these facilities are still struggling to have reliable and continuous operations at commercial production levels. The "Predictive Models and High-Performance Computing as Tools to Accelerate the Scaling-Up of New Bio-Based Fuels" workshop (hereafter "the workshop") gathered bioenergy stakeholders across diverse sectors and technology areas, including representatives from industry, academia, and government. One hundred seventy-five participants provided their input to DOE's Bioenergy Technologies Office (BETO) regarding best practices for utilizing mathematical modeling tools across multiple scales to reduce technology uncertainty and accelerate scaling-up of biorefinery/chemical production equipment and optimize operations.

This workshop summary report summarizes stakeholder input related to discussion objectives including:

- Understanding how modeling tools can be effectively utilized in conjunction with operational data from bench, pilot, and demonstration facilities to augment and accelerate scale-up and integration efforts
- Identifying synergies, challenges, and gaps in current modeling efforts and data collection methods across technologies
- Utilizing artificial intelligence and machine-learning computational technologies together with well-instrumented systems to assess and validate models and develop transfer functions to represent scale-up accurately
- Developing best practices and publicly available multiscale modeling tools to reliably represent integrated equipment and operations involved in biorefineries.

BETO anticipates that the participant discussions, in addition to this workshop summary report, will have a role in the acceleration and commercialization of integrated manufacturing processes, with the following potential impacts:

- Provide validated computational tools, helping to achieve performance goals
- Improve understanding of state-of-the-art modeling tools at multiple scales for bioenergy and manufacturing technologies currently in progress
- Enhance scientific understanding to make mathematical models predictive and to gain insights into potential operational issues
- Extend mathematical models into biorefinery/manufacturing operations to guide control schemes and optimization strategies.

Table of Contents

List of Figures

List of Tables

Introduction

Advanced biofuels have the potential to enable the United States to use our enormous, renewable resources to avoid approximately 10% of U.S. carbon emissions and displace up to 27% of U.S. petroleum use, while also creating domestic jobs across many sectors of our economy.

Two U.S. Department of Energy (DOE) Bioenergy Technologies Office (BETO)-funded reports, "Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply" (the "Billion-Ton Study") (2005) and the "U.S. Billion-Ton Update" (2011), project that the United States could sustainably produce more than 1 billion tons of biomass—without competing with the food, feed, and fiber supply. This projection does not include the potential for algal biofuels, which could be significant.

DOE-BETO Background

BETO's vision is "a thriving and sustainable bioeconomy fueled by innovative technologies" and its mission is to develop "transformative and revolutionary bioenergy technologies for a sustainable nation." To accomplish these ends, BETO focuses on forming public–private partnerships with key stakeholders to produce advanced bioenergy and bioproducts from biomass. BETO focuses on reducing risks by developing technologies that enable industry investment and deployment at scale. Similarly, BETO's strategic goal is to enable the use of America's abundant biomass and waste resources for advanced biofuels, bioproducts, and biopower by:

- Identifying and developing biofuel pathways and end uses
- Lowering the cost of production through increased efficiency, productivity, and yields
- Completing applied research and development (R&D) on complex, real-world systems and integrating engineering processes for promising new advanced bioenergy technologies
- While maintaining or enhancing economic, environmental, and social sustainability.

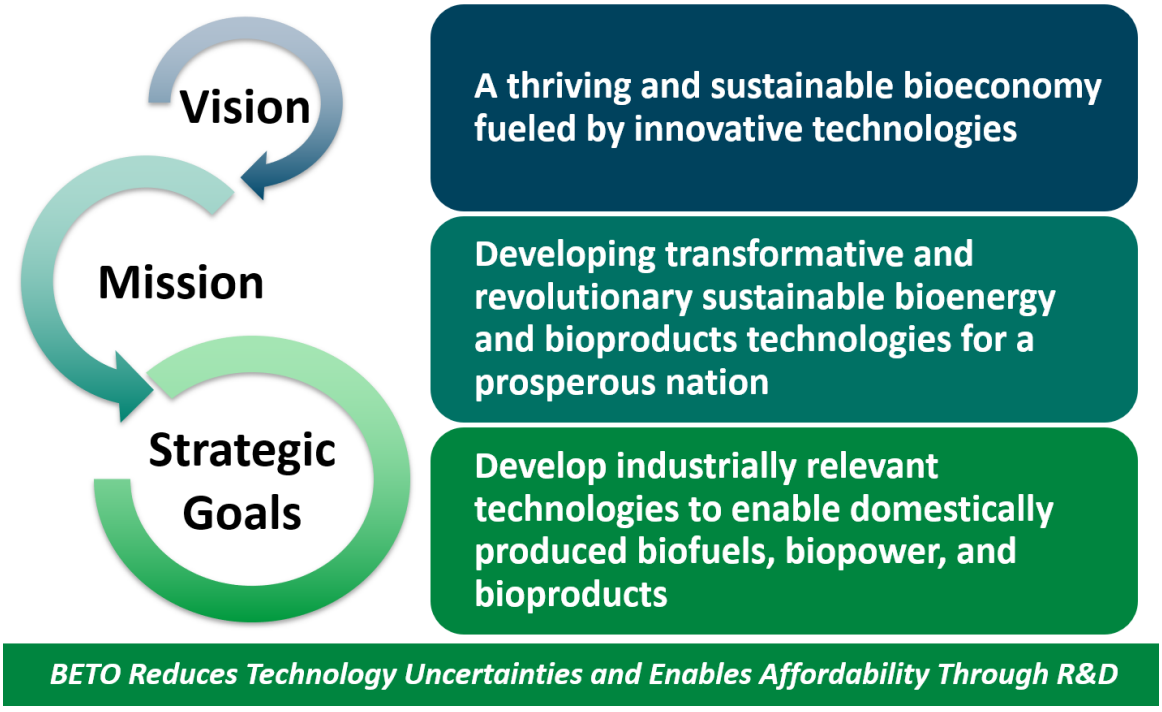


Figure 1. Bioenergy Technologies Office's mission and vision

Existing Computational Modeling Efforts

DOE-BETO is extensively involved with bioenergy-related computational modeling efforts. Currently, these efforts are primarily part of the BETO consortia, process development units, and high-performance computing. At the beginning of the "Predictive Models and High-Performance Computing as Tools to Accelerate the Scaling-Up of New Bio-Based Fuels" workshop (hereafter "the workshop"), BETO provided attendees with an overview of these efforts to establish a baseline for discussions.

Consortia

DOE does have several consortia currently working on various bioenergy-related models. Although all of the consortia have their hands in computational modeling, there are three that focus on this work: the Feedstock-Conversion Interface Consortium (FCIC), the Chemical Catalysis for Bioenergy Consortium (ChemCatBio), and the Consortium for Computational Physics and Chemistry (CCPC).

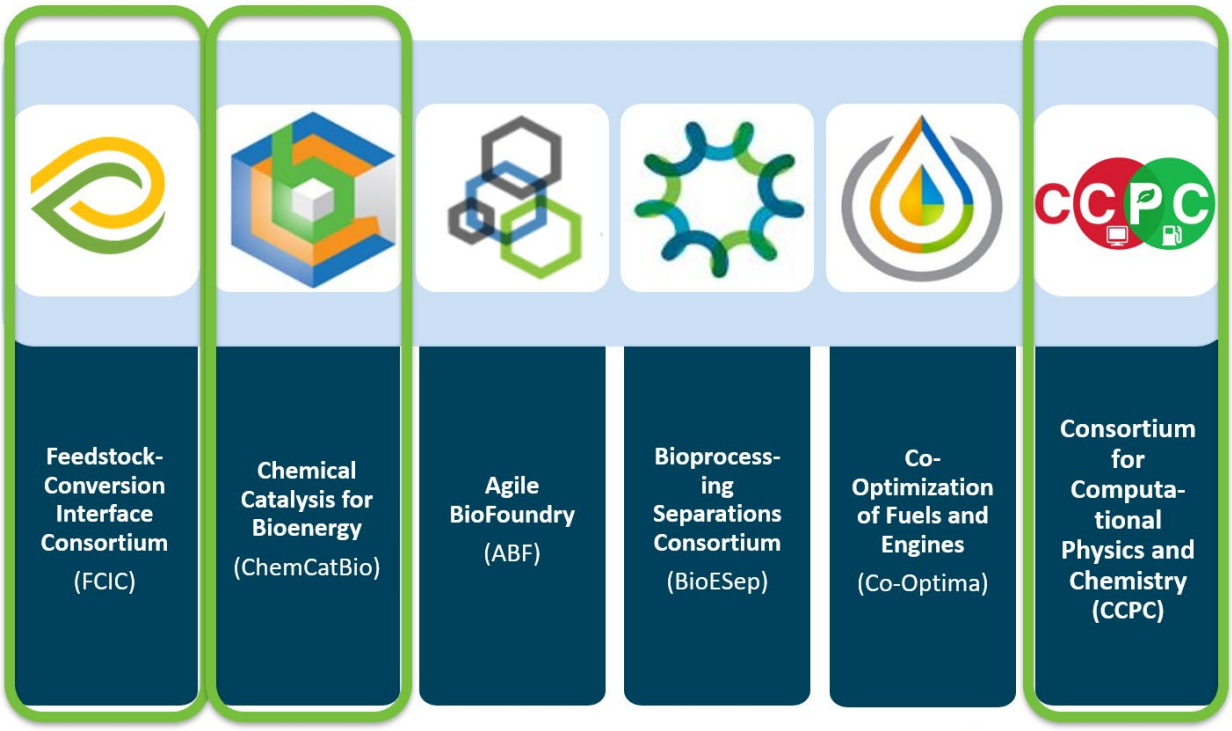


Figure 2. Bioenergy Technologies Office's consortia (computational modeling-focused consortia marked with green)

FCIC

The Feedstock-Conversion Interface Consortium (FCIC) objectives are to quantify, understand, and manage variability in biomass from field through downstream conversion and to understand how biomass composition, structure, and behavior impacts system performance. Some of the goals include providing first principles-based data related to unit operations and providing transfer functions to bridge scales from bench all the way through a pioneer biorefinery.

ChemCatBio

The Chemical Catalysis for Bioenergy Consortium (ChemCatBio) aims to accelerate the development of catalysts and related technologies for the production of fuels and chemicals derived from biomass and waste

feedstocks. The consortium maintains the ChemCatBio Data Hub, where account holders can submit their own data to share with others and/or search a repository of data, including interactive tools, databases, and models.

CCPC

Consortium for Computational Physics and Chemistry (CCPC) brings together unique computational modeling facilities and experience to address conversion challenges. From molecular-scale chemistry to process-scale operations, the CCPC integrates experimental and computational capabilities to address real-world complexities and enable the advancement of renewable energy technologies. Bringing together theory, sophisticated computational methods, and high-performance computing, CCPC provides an understanding of material properties and functionality. Additionally, the consortium assists collaborators with predicting performance based on fundamental principles and optimizing the design of new materials and unit operations.

Process Development Units

In addition to consortia, BETO has a number of process development units (PDUs) used to aid in scaling-up various bioenergy technologies and processes and, in some cases, work alongside BETO's consortia.

The Integrated Biorefinery Research Facility at the National Renewable Energy Laboratory (NREL) is a biochemical conversion pilot plant that accommodates bench-to-pilot-scale processes for converting biomass into a variety of fuels and chemicals at process throughputs of up to one ton of dry biomass per day.

The Thermal and Catalytic Process Development Unit at NREL has state-of-the-art equipment for thermochemical process development and testing, ranging from catalyst and feedstock characterizations to bench-scale reactors, all the way up to pilot plants.

The Biomass Feedstock National User Facility at Idaho National Laboratory offers technology to overcome biomass challenges during scale-up and integration of biomass feedstock preprocessing facilities.

The Advanced Biofuels and Bioproducts Process Development Unit at Lawrence Berkley National Laboratory enables researchers to evaluate, adapt, develop, demonstrate, and transfer commercially viable biorefining technologies for biomass deconstruction and advanced biofuels, proteins/enzymes, and bio-based chemicals production.

Lastly, Pacific Northwest National Laboratory's Modular Hydrothermal Liquefaction System includes three operations sections (or skids) to enable larger-scale hydrothermal processing methods to convert wet-waste feedstocks into biofuels. The hydrothermal liquefaction skid uses high temperature and pressure to break down sludge into simpler compounds, turning them into bio-crude and bio-oil, which can then be further processed into biofuels and products.

High-Performance Computing

In addition to the PDUs and the consortia, BETO has access to the supercomputing at several national labs, including stacks at both NREL (Eagle Computing System) and Oak Ridge National Laboratory (Summit). Eagle is a DOE Office of Energy Efficiency and Renewable Energy (EERE)-specific high-performance computer with a peak performance of 8 petaFLOPS. Summit is the smartest and most powerful scientific supercomputer in the world, with a 200-petaFLOP performance capacity.

Figure 3. Oak Ridge National Laboratory's high-performance computer

Workshop Purpose

The workshop's purpose was to understand how modeling tools can be effectively utilized in conjunction with operational data to augment and accelerate scale-up and integration efforts and reduce scaling risk. Through the use of highly instrumented bench-, pilot-, and demonstration-scale facilities along with high-performance computing (HPC), artificial intelligence (AI), and machine learning (ML), it may be possible to reduce technology uncertainty and accelerate scale-up.

Over the past decade, federal and private investments in the scale-up of cellulosic integrated biorefineries have exceeded $2 billion. A majority of these facilities have either been shut down or are struggling to have reliable and continuous operations. Furthermore, mathematical models representing the total end-to-end process flow of an integrated biorefinery do not currently exist.

To date, computational modeling efforts have successfully incorporated particle-scale models of biomass feedstocks into reactor models, but typically this is done using a simplified approach with a submodel of reduced order (or lower dimension). There could potentially be benefits to including the full particle model, but that is computationally intensive. There has been a lot of work at universities trying to link atomic-scale models to reactor-/process-scale models.

BETO began the workshop describing four observations made over the past several years: (1) feedstocks are not well understood, (2) solids handling is not well understood, (3) understanding equipment is essential, and (4) understanding integration of equipment is essential. These observations were the foundation of all modeling discussions conducted with workshop participants during breakout sessions. The general descriptions follow:

(1) Feedstocks are not well understood.

BETO has observed that feedstock variation should not be underestimated, and associated quality and system specifications should all be understood prior to design.

Verenium Biofuels' final technical report emphasizes this fact. This was one of BETO's large-scale demonstration integrated biorefinery projects designed to utilize bagasse to make cellulosic ethanol. The report states:

> *"One of the single most important lessons learned from the project was the data gathered from compositional [analysis] of the feedstocks. The key message regarding feedstock composition is that it will vary. It will vary over the course of the growing season; it will vary due to harvest method; it will vary due to on-site staging/storage methods; and finally, it will change due to handling methods immediately prior to the conversion process."*

(2) Solids handling is not well understood.

Solids handling should not be underestimated and should be fully understood prior to design. This includes abrasion, bridging, channeling, particle size distribution, effects on rate of reaction, ash content, neutralization effects, ability to flow, and the angle of repose, to name a few.

The BETO FCIC images (Figures 4 and 5) illustrate that when improper construction materials are used in feedstock handling, like the 316 stainless steel used on the auger in Figure 5, significant wear can occur. This wear can occur in pipes, reactors, pumps, conveyors, and really anything that comes in contact with solids—particularly biomass solids—during a scaled-up process. Additionally, you can see how the milled corn stover in Figure 4 can bind and lock up a piece of equipment, bringing a facility to a standstill when the solids handling system is not properly designed.

Figure 4. Channeling of corn stover
Source: Jenike & Johansen

Figure 5. Wear on stainless steel 316 auger from corn stover (left: light wear; right: significant wear)
Source: FCIC

(3) Understanding equipment is essential.

When utilizing equipment designed for a different purpose, it is critical to understand the similarities and—more importantly—the differences, and how to account for these differences. "Commercially available" equipment needs to be treated as a new technology if it has not been demonstrated in a new function or at a new scale. Greater emphasis needs to be placed on the risks associated with scaling-up critical unit operations from proof-of-concept to lab-scale to "next"-scale designs; additional data validation, modeling, and piloting efforts should be investigated prior to design of an integrated facility.

(4) Understanding integration of equipment is equally essential.

Combining two well-known unit operations in a new way can increase complexity and lead to operational issues. Also, engineers need to understand the systems readiness level, which quantifies the maturity of technology integration in a system and overall system readiness—basically how an unproven unit operation will affect the overall process, which is only as strong as its weakest link.

Effective utilization of mathematical models and artificial intelligence with machine learning to validate and improve system and subsystems models using highly instrumented bench-, pilot-, and demonstration-scale facilities may reduce technology uncertainty and accelerate scale-up of biorefinery manufacturing facilities.

Plenary Presentation Summaries

In order to set a basis of understanding for the workshop participants and to highlight the current work being performed in this space, two subject-matter experts presented. The first presentation, "Modeling and Simulation for Bioenergy Applications: Challenges, Successes, and Lessons Learned," was given by Dr. Peter Ciesielski, a scientist at NREL. The second plenary presentation, titled "Scientific Artificial Intelligence and Machine Learning," was given by Dr. Steven Lee, Program Manager for Applied Mathematics and Artificial Intelligence at the Office of Science's Advanced Scientific Computing Research.

Modeling and Simulation for Bioenergy Applications: Challenges, Successes, and Lessons Learned

Dr. Peter Ciesielski began with an outline of just how difficult predictive modeling and scale-up of bioenergy applications can be due to incorrect assumptions and attempted shortcuts when working with biomass and bio-based intermediates. The assumption that handling and comminution methods designed for other solid materials will also work for biomass is not correct and can lead to issues in attempts to model biomass feedstocks. Biomass feedstock particle sizes are inconsistent and composition can vary even within a single bale, so trying to employ the same handling and comminution methods as might work for a feedstock like coal will not work. To the same effect, attempts to utilize upgrading and conversion technologies originally developed for petroleum feedstocks will not have the same results when applied to biomass-derived feedstocks such as pyrolysis oil. Dr. Ciesielski also discussed the difficulties encountered in modeling work when scientists assume that modeling and simulation frameworks and engineering heuristics developed for other feedstocks will translate directly to biomass; these assumptions lead to inaccurate results. He went on to explain that the hierarchical structure and variability of biomass feedstocks differentiate it from other commonly used feedstocks.

The hierarchical structure of biomass gives rise to emergent properties that dictate the behavior of biomass feedstocks in handling, preprocessing, and conversion operations. He explained that emergent properties are difficult or impossible to characterize by experiment or simulation performed at individual length/time scales, which ultimately makes it difficult to scale up bioenergy applications. An example of an emergent property is the nanomechanics of cellulose. He discussed how kink defects formed by mechanical stress in cellulose fibrils can be linked to the breakdown of bonds within these fibrils. They used the CHARMM carbohydrate forcefield model to simulate the kink defects in their work. They then showed that following partial enzymatic digestion, the sites of these kink defects were targeted for hydrolysis initiation by the enzymes more readily than other sites, which leads to greater conversion of cellulose. These findings provide new fundamental insights to evaluate the impacts of preprocessing and pretreatments.

Dr. Ciesielski provided an additional example focused on thermochemical conversion in catalytic fast pyrolysis. He again discussed the highly variable nature of biomass feedstocks and how the different microstructures of various species of biomass dictate intraparticle heat and mass transfer. In addition to the microstructure within the particles, even varied particle shape and size can affect conversion times, fluidization behavior, and flowability, which makes modeling these systems particularly difficult. He discussed a critical question in developing biomass models: “If the feedstock attributes change, how does the model change?” They have modeled aspen and pine particles to simulate heat-transfer scenarios as well as the reaction/diffusion process for lignin extraction. The work is further described in a paper by Thornburg et al., “Mesoscale Reaction-Diffusion Phenomena Governing Lignin-First Biomass Fractionation.” He stated that by developing detailed particle models, they enabled the prediction of yield, product distribution, and required conversion times as a function of feedstock attributes, including particle size, shape, microstructure, and composition.

Dr. Ciesielski then discussed how coupling to reactor simulations and ensemble calculations enables highly accurate predictive reactor models. He looked at NREL’s fluidized bed reactor and discussed how reactor-scale simulations and particle-scale simulations can be used together to achieve an accurate product yield prediction within 1%–2%. The differences between the two simulation scales gives insight into different aspects of the conversion process. Whereas the reactor-scale simulations account for hydrodynamics and are used to estimate residence-time distributions, the particle-scale simulations are used to account for feedstock-specific effects and calculate conversion.

One of the next steps in bioenergy process simulation could come from the ability to directly import complex 3D geometries for finite element analysis. Dr. Ciesielski gave the example of modeling a catalyst particle’s microstructure. By taking a transmission electron micrograph of the catalyst particle microstructure, they can make a meshed geometry that can then be used directly in finite element analysis or simulations that provide researchers results that more closely relate to real-world conditions. Something else that can help researchers

gain insight into what happens in real-world scenarios is the development of reduced-order models. These models can help researchers extract intrinsic kinetics from experimental data, which then enables process-scale models that can investigate reactor designs and operating conditions.

He closed by discussing conclusions and lessons learned through his work. He suggested to start by modeling the feedstock itself, not the process. This helps avoid embedded assumptions (e.g., black box with arrows), and models are generally applicable to many processes. He also suggested starting from scratch and using measurements and data to parameterize the model. Models are rooted in reality rather than assumptions or convenient math. Additionally, he suggested appreciating the complexity of organic systems because there are not "plug-and-play" tools that will make things easier. Finally, he proposed modeling variability distributions rather than averages, and noted that ensemble calculations are critical because one particle is not representative of the entire feedstock.

Scientific Artificial Intelligence and Machine Learning

Dr. Steven Lee discussed the next steps for advancing AI and ML, as well as how these technologies can be harnessed to advance science and energy research. He provided an overview explaining that AI and ML for science broadly refers to the development and use of AI-enabled facilities, infrastructure, and technologies in a manner that accelerates scientific discovery, increases scientific competitiveness, and creates innovative scientific and operational capabilities. He broke down the basics of AI, ML, and deep learning and described the types of tasks that can be enabled through these technologies. He posed that AI could be developed and used to assist, augment, or automate human skills to complete tasks associated with vision and perception, natural language processing, search and planning, problem solving, and knowledge reasoning. Additionally, ML is the foundational basis for AI and these algorithms make predictions, decisions, and estimates from building a mathematical model or "learning" based on probabilities, samples, or training data. The types of tasks that ML is best utilized with are classification, clustering, regression, simplified or surrogate models, feature extraction, and pattern recognition. He also described deep learning, which includes neural network-trained approaches for tasks such as spam filtering, fraud/anomaly detection, and image analysis. All three of these technologies can be harnessed to advance scientific research but are still in relatively early stages of development. In order to further these technologies and enable the next generation of machine-learning methods and AI capabilities Dr. Lee described six priority research directions (PRDs) that need to be focused on. These PRDs were classified as "foundations" and "capabilities." The three "foundation" PRDs focus on the idea that ML should be domain-aware, interpretable, and robust. This means that it should leverage and respect scientific domain knowledge, provide explainable and understandable results, and produce stable, well-posed, and reliable formulations. The "capabilities" PRDs focus on data-intensive scientific machine learning, ML-enhanced simulations, and intelligent automation and decision support.

To demonstrate the benefits of employing AI in science, Dr. Lee described an experiment performed by researchers at Brookhaven National Lab and Lawrence Berkeley National Lab. In this experiment, an AI algorithm was integrated into an X-ray-scattering instrument and used to study nanomaterials. The AI was able to autonomously reconstruct a nanomaterial's shape faster and more accurately than when performed manually. This achievement is just one example of how an AI ecosystem can be harnessed to perform experimental tasks more effectively to enable new science.

One of the PRDs for the capabilities of future ML algorithms is to enable handling data-intensive tasks. Due to the data-intensive nature of many scientific applications, scientific ML methods require development of improved methods for statistical learning in these high-dimensional systems involving noisy and complex data sets. The development of these next-generation algorithms will need to incorporate approaches to find structure hidden within complex high-dimensional data and be able to sample efficiently within high-dimensional parametric model spaces.

Machine learning will also give an advantage in computational model and algorithm development, as a combination of scientific computing with learned adaptivity will help create more-efficient simulations. It can

also be used for *in situ* parameter tuning, which would help further improve scientific models and simulations. Progress in this space will require the development of new methods to quantify trade-offs and optimally manage the interplay between traditional and machine-learning models and implementations. Regarding intelligent automation and decision support, there are challenges in managing the interplay between automation and human decision-making that machine learning is well suited to handle. Machine learning can help complex systems such as power grids or cybersecurity. As these types of systems continue to grow more complex, AI and machine learning can help guide data acquisition and ensure data quality and adequacy. Scientific machine-learning methods can be also be utilized to improve system resilience or responsiveness.

AI and machine learning are rapidly evolving technologies, and there is tremendous national and international interest in AI and machine-learning development and use for science. DOE has hosted "AI for Science" town halls, and in Europe, the Alan Turing Institute has launched a £38.8-million research program focused on AI for science and government. The fiscal year 2020 Office of Science Funding Opportunity Announcements and Lab Calls are generating numerous new ideas in this space, and the Office of Science's planning and AI-ecosystems approach focuses on DOE's strengths and assets, including modern facilities, national laboratories, and high-performance computing capabilities.

Stakeholder Input

The purpose of the "Predictive Models and High-Performance Computing as Tools to Accelerate the Scaling-Up of New Bio-Based Fuels" workshop was to understand how modeling tools can be effectively utilized in conjunction with operational data to augment and accelerate scale-up and integration efforts.

The workshop was designed to improve BETO's understanding of process modeling activities at multiple scales for bioenergy technologies currently underway, identify existing gaps and development areas for future modeling efforts to accelerate reliable scaling-up and integration, and promote information exchange and share lessons learned. It also sought to identify gaps in experimental data to enhance modeling and opportunities for enhanced scientific understanding to make models of different scales more reliable. Further, the workshop was meant to determine if effective utilization of mathematical models and artificial intelligence with machine learning in conjunction with highly instrumented bench-, pilot-, and demonstration-scale facilities can be used to validate and improve system and subsystem models.

This virtual workshop used a web conferencing and brainstorming platform called MeetingSphere to gather input from participants. Participants were divided into four breakout groups, each covering the same set of topics and questions. The participant input was exported from MeetingSphere at the end of the workshop and compiled into this workshop report.

Most brainstorming discussions included a prioritization activity. The prioritization results described in table format were created using the following methodology:

- Averaged sum of impact and likelihood scores from all four breakout groups
- Identified top-ranked responses
- Adjusted for duplication
- Edited for clarity.

In some cases, the top-ranked responses did not include the top response from each of the four groups. To ensure that each group's top priority was represented in this summary, we list the top-ranked response for each group in addition to the table of aggregated priorities across groups.

Understanding the Current State of Technology

The first breakout session topic was identified with the recognition that there are not enough data or enough data in the same format to feed a system of models. Additionally, the bioenergy industry faces modeling challenges related to dealing with lack of physical properties data, new compounds (including biological ones), and variation in data.

Participants identified a few themes from their discussions on this topic. For example, there are some valuable new standardized sensor technologies and analytic techniques available for the industry to use. Another fairly recent development is the application of machine learning at various scales, from atomic to commercial facilities. Additionally, the best type of model to use will depend upon the user's ultimate purpose. There are varying perspectives regarding whether or not to begin with a simple model and then build its complexity. A commonly expressed idea was that bioenergy researchers need standardized data and models because lab-scale and plant-scale systems can—and do—differ.

Current Process Modeling Efforts

Participants provided their input to develop an inventory of current process modeling tools. Workshop organizers asked for participants to consider both simple and rigorous tools. Simple, or traditional, modeling tools are defined as models that explain a single aspect in a fundamental way or those that may not require advanced computing resources. Rigorous modeling tools are defined as models that include the integration of multiple models and/or those that include significant complexity, thus requiring more computing resources.[1]

Workshop participants mentioned many of the tools BETO currently uses. For example, full-process models (Aspen) are being used for techno-economic analysis efforts and to identify critical research needs. CCPC is using complex multiphase flow with interphase exchanges (MFiX) using computational fluid dynamics (CFD) models. Computational particle fluid dynamics (CPFD) models by Barracuda, coupled with Aspen, have been used for scale-up and commercial-scale prediction; both Aspen and CPFD models have been validated to pilot and commercial data. There have been machine-learning techniques applied to catalyst modeling with density functional theory for catalyst surface chemistry. This has been done in CCPC and ChemCatBio to help guide experimentalists on materials to study for catalyst innovations in bioenergy.

General Approach

Stakeholders provided input regarding the general approach to support process modeling. For example, they emphasized the need for developing fundamentals first before diving into rigorous models. Additionally, they suggested that scientific ML should try to capture data from multiple sources, including HPC models to describe a process (this is referred to as "multifidelity analysis" in applied math). Also, ML-guided interpretation of experimental characterization data (spectroscopy and microscopy) may be based on simulated models.

There are a wide variety of machine-learning tools being applied to (for example) microbial strain engineering to accelerate improvements in bioproduction titers, rates, and yields. Ideas have been proposed to combine genome-scale metabolic models, computational fluid dynamics, and data from arrays of bioreactor sensors to better predict microbial bioreactor scale-up from the submilliliter to thousands-of-liters scale. Machine learning could be layered and integrated into these approaches. There are many dimensions that could be explored, such as bioreactor geometries, microbial genotypes (and communities thereof), feedstock, and controlling strategies (e.g., feed rates/locations, mixing, aeration, where to place sensors, how to extract products). Computational fluid dynamics models could use concepts like voxels, each containing an independent (parallel evolution) but potentially interacting microbial genotypes/metabolic states that flow through the reactor.

[1] Definitions developed in reference to *Rigorous Simulation: Its Theory and Applications,* Adam Duracz (https://old.liu.se/elliit/artiklar-3/1.741918/AdamDuracz.pdf)

Classification on different scales may be needed—molecular-scale, mesoscale, and reactor-scale models. Challenges are different at each scale, and scale integration itself is a unique problem. Connecting information between scales with ML-based reduced-order models could be useful. There have been great developments in models at the different scales; however, the integration between scales requires a lot of work. Molecular refers to anything from the sub-nanoscale to nanoscale; mesoscale is anything from the nanoscale up to microscale. The biggest flexibility lies in "mesoscale." Mesoscale is gray territory and also important for connections. Atomistic to mesoscale is also very difficult (e.g., for simulating catalytic conversion reactions) and ML can play an important role in this effort. There is a lot of work done at NREL for reactor-scale CFD modeling of low-temperature conversion as well (e.g., pretreatment, enzymatic hydrolysis, and aerobic bioreactors). Reduced-order assumptions are used bridge the connection between scales. NREL are Idaho National Laboratory are collaborating on scale-bridging between mesoscale and macroscale in particle mechanics for FY21.

Process Modeling Tools Currently in Use or in Development

This section lists data collection tools as well as simple and rigorous modeling tools. Participant input related to general approaches and bioenergy applications is also included. All tools mentioned in this report are merely an inventory of software and models familiar to the bioenergy stakeholders who participated in this breakout session discussion. This is not intended to be a comprehensive list, nor is it an endorsement of any tool listed.

Generally, participant identified five categories of tools:

1. Computational fluid dynamics codes (e.g., Fluent, OpenFOAM)
2. Fundamental kinetic and thermodynamic models
3. Machine-learning tools (e.g., for catalyst development and thermodynamics)
4. Process models (AspenPlus and SuperPro)
5. Cloud-based computing platforms (e.g., Amazon Web Services, Azure, Google Cloud).

Workshop participants mentioned several mathematical and statistical data modeling methods, including recurrent neural network (for predicting process deviations), discrete element method (for high-performance particle flow), artificial neural network, support vector machines, k-nearest neighbor, decision tree, and random forest. The tools that participants use or are familiar with in the bioenergy industry are listed here:

Data Collection Tools

- JMP Software from SAS for statistical discovery
- LabKey for research data management
- iFIX from GE Digital for sensor data acquisition
- PI System from OSIsoft for real-time operational data collection and storage
- Internet of Things (IoT) sensors for data collection
- Blockchain for production attributes.

Modeling Tools

- Adaptive mesh refinement
- AspenTech is the main process modeling tool in use, with embedded kinetic models.
    - Includes Aspen Plus (steady-state) and Aspen Custom Modeler (steady-state and dynamic)

- Anura3D software for modeling large deformation and soil–water–structure interaction using the material point method
- Athena Visual Studio for process modeling
    - This is the primary tool used at Dow Chemical for reaction kinetics and models used for process scale-up
- Combination of techno-economic analysis (TEA) tools like SuperPro or AspenTech with Python-based ML/AI libraries
    - For example, researchers have developed AI/ML models based on CPFD and Aspen-based simulations and data validation
- ChemCAD for chemical process simulation
- Chemkin-Pro for gas phase oxidation/secondary reactions
- COMSOL Multiphysics for modeling and simulating
- Custom tools can be developed using standard, open-ended software and coding languages including MATLAB, Python, R, and Excel
    - Excel supports linear regressions
    - There are statistics and machine-learning toolboxes available, as well as is a deep-learning toolbox, for MATLAB
    - Python-based machine-learning tools from the SciPy library (e.g., scikit-learn, TPOT)
    - R-based random forest regressions and artificial neural networks supported by open-source platforms such as TensorFlow, Keras, and Theano
    - All of these work with DOE supercomputers and are compatible with Jupyter notebook
- Fluent fluid simulation software from Ansys
    - There are many CFD codes such as Fluent that are routinely used by industry for scale-up of reacting turbulent fluid flows with heat and mass transfer
- High-performance computing software from Hewlett Packard Enterprise Development
- LIGGGHTS discrete element method particle simulation code (open-source)
- Molecular dynamics tools (various available)
- OpenFOAM CFD software (open-source)
    - CFD utilized for simulations of unit operations (e.g., bubble column)
- ROCKY advanced particle simulation software using discrete element method. TensorFlow machine-learning platform (open-source)
- Stress analysis and dynamic analysis codes such as ANSYS and NASTRAN

  - These codes are used to model structure behavior and fluid structure interaction in industrial equipment

- Two software development efforts in the DOE system have upgrades toward exascale and high-performance computing platforms
  - MFiX by the National Energy Technology Laboratory: CFD multiphase flow
  - MOOSE by Idaho National Laboratory: finite element model code.

Bioenergy Applications

Workshop attendees were also asked to identify where modeling could be utilized in bioenergy applications. The following summarizes their inputs:

- The data required for developing fundamental models can be obtained by collaborating with equipment manufacturers, where they may be willing working with national labs to generate the required data
- Operational data of pilot plants or commercial industries (AI could be very useful to adjust the parameters of simple models—often heat- and mass-transport coefficients in interphases)
- Thermophysical property data for bio-derived oxygenated molecules to support process simulation calculations (i.e., energy balance and phase equilibrium)
- Machine-learning models are being developed for BETO to predict performance of catalytic fast pyrolysis based upon mass spectral measurements of vapor-phase products
- Reduced-order models are being used to perform sensitivity analysis of the complex multivariate nature of bioenergy processes
- Feedstock supply models can feed conversion process models and/or TEA able to provide geospatial info on specific feedstock types, location, and lowest cost delivered to reactor throat
- Solubility data and activity coefficient models for bio-derived oxygenated molecules for product separations.

One approach that attendees suggested for using simple models is developing custom, “home-brewed” tools that can be used to quantify reactor data quality by leveraging existing literature criteria. Poor-quality data with issues like non-isothermal conditions result in poor models and surprises in scale-up. Correlations in the literature can be used to quantify these effects and ensure that the data are able to be leveraged for kinetic models. Important criteria to quantify include isothermal temperature, flow characteristics such as wall channeling during plug flow, uniform residence time distribution, and catalyst wetting. An open-source tool could be useful, as these correlations are hard to employ and result in inconsistent usage without a consistent tool or framework for researchers to use.

Participants explained that distributed control system (DCS) data can be used in software provided by Pavilion, Emerson Neural Networks, and Allen Bradley to generate embedded machine learning. Operators have installed and serviced DCS for ethanol plants. This has been successful in controlling variables that have been illusive like moisture control in dried distillers’ grains on drum dryers, feed rate into distillation, and reflux. Participants also mentioned that they have been able to combine lab data with DCS data in real time to push alerts from automatically calculated data points that alert operators what to do in real time. Also, there are ways to make lab data push into DCS systems and control valves automatically, although workshop participants have only done it in the pilot phase. In particular, an advantage of this technology is the great amount of data generated by industry facilities that can feed these machine-learning techniques. For example,

data can be pulled every second from 3,000 tags. The data can be exported and pushed into other software in real time.

Breakout session participants noted that implementing DCS-embedded machine learning can be expensive, especially for an entire plant. One stakeholder estimated that the cost ranged from $500,000 to $3 million to implement plantwide. For specific parts of the plant that have a particular business case (i.e., dryers), the cost could be as low as $50,000. Simply compiling and organizing the data could cost between $20,000 to $60,000. Participants noted that these types of projects can face significant resistance among facility decision makers, especially if the leadership does not have prior experience with these efforts. Generally, the industry is more concerned with basic automation rather than "nearly AI"—projects like sequencing, alarm management, and interlocks. Another participant commented that an added benefit of this type of project is that the data can be saved in the cloud and used for industrywide research.

Best Practices for Developing or Updating Biotechnology Data Sets

Response categories include developing data collection methods, data standardization, and data repositories; leveraging published data; and tool development.

The top-ranked ideas from the aggregated input are listed in Table 1. The top-ranked ideas from each group include:

- Design a comprehensive protocol and standard for data collection, documentation, and submission from different sources
- Have an accessible repository of all available data
- Establish an open/public database on unit operations data, especially data such as capital costs and operating costs. This could be a baseline for all the TEAs.
- Have a consistent framework for reporting data to allow combining of data sets from various sources/experimental sets.

Table 1. Best Practices for Developing or Updating Biotechnology Data Sets

1	Address limited existing data by developing an accessible repository of all available data
2	Develop a consistent framework for reporting data to allow combining of data sets from various sources/experimental sets (expand upon previous FCIC efforts).
3	Identify and agree on key physical properties required for the scale-up
4	Need better characterization of biomass with thermodynamically consistent components
5	Support for high-throughput activities with standardized data-collection protocols
6	Leverage the large amount of data from scientific publications
7	Support data measurement independent of specific process applications
8	Work directly with experimentalists to design experiments and automated data acquisition systems to produce data sets that are readily applicable to ML methods
9	Develop tools to rapidly obtain compositional data for biocrudes or biointermediate streams.

Trade-Offs in Tool Selection

Breakout session participants discussed the pros and cons of simple and rigorous models. The input provided is listed in the following sections:

Pros of Simple Models

- Can be validated against experimental data
- Accessible to broader range of users
- Can be easier to understand by peers
- Inexpensive
- Relationships between model variables are clearly exposed
- Can be used as building blocks for rigorous models
- Provide insight for non-experts
- Straightforward to develop and verify with experimental data
- Easier to interpret
- Quicker, easier to create and understand
- Less computationally expensive
- Takes less time and resources to develop (compared to rigorous models)
- Less expensive to maintain and continually improve
- Allow for screening of alternatives without excessive use of time and other resources
- Can be applied to many applications (i.e., different industries)
- Less-complex data may lead to the ability to use for multiple case applications
- Less likely to overfit
- Potential for high-throughput analysis/results
- Great for exploration
- Provides scientifically rigorous results
- Collaboration: easy to share with others
- Use across disciplines
- Oftentimes the tools have wider access amongst collaborators
- More likely to be used consistently across multiple entities

- Good when data or process uncertainty is such that use of a highly detailed model would give a false sense of the level of detail that is understood
- Can provide better understanding of physics and chemistry
- Can provide energy and mass balance
- Easier to verify correctness
- More robust/consistent
- Can be more robust if one does not have all the information
- Better scalability
- Easy to use, fast
- Interpretability
- Easy to integrate as a submodel in a rigorous model
- Repeatability
- Can be used to model entire production process
- Easy to identify limitations
- Can be used capture key process fundamentals.

Cons of Simple Models

- May overlook important variables, parameters
- Simplified model may miss crucial second- or third-order interactions
- Gloss over complexity, inaccurate simplifications
- Cannot recognize complex patterns
- Difficult to remove dependencies related to complexity of application (not always useful)
- If the data input and complexity is not sufficient, the result is useless
- Accuracy may be limited
- Can be dangerous because they are easy to misuse
- May be too simple to capture the complexity of the system, leading to incorrect conclusions and actual implementations
- Does not work well in integrated process
- May not capture full physics/chemistry/biology
- They oversimplify the chemistry or biology and are valid over limited range

- Sometimes neglect key physical phenomena
- Typically, simple models do not account for critical material attributes like particle size, porosity, density, and their impacts on reactor size requirements
- Often valid only for a small range of operating conditions
- Limited ability to direct decision-making across large scales
- Less detail means less-robust techno-economic analysis
- Often not realistic to real systems
- Can be very risky when extrapolating results to new systems or conditions
- Potentially making the wrong simplifications
- Limited applicability
- May not scale
- Big approximations, conclusions made without understanding the approximations and assumptions, can be very misleading, inflexible.

Pros of Rigorous Models

- Should be closest approximation of reality
- May represent chemistry/biology more accurately
- More likely to capture important phenomena
- Could be applicable across a broader range of input parameters
- More likely to be predictive rather than extrapolative
- Could be used for unsteady processes
- Can be incorporated in other rigorous models
- Can provide “correct” results for use in process design and plant design
- Can provide new insights into novel phenomena
- Can be more readily validated against experiment because of capturing all of the physical phenomena
- Suitable for integrated process
- Can be used to develop low-order simple models
- Account for critical material attributes and critical process parameters more accurately
- Valid across a variety of operating conditions because they use physics-based predictive models
- Higher accuracy and confidence with results

- Assuming your assumptions are correct, and robustness is built in, can be the most accurate at predicting
- Can provide the understanding of mechanisms
- Can be used to interrogate specific aspects of complex systems
- Rigorous models tend to avoid more assumptions (than simple models) and are amenable to sensitivity analysis, which can enable the development of simple models
- Can provide important information about connections between different systems
- Can analyze systems across large scales (e.g., regions, supply chains)
- Can capture full detail for complex processes (even the complexity of biomass variability)
- Can provide information that is more representative of real processes
- Probably more inclusive and produce better results if done well
- Enables specific design guidance for reactor design and process control
- Provide more opportunity for transformative discovery, as opposed to incremental improvements
- Address complexity when needed and valuable
- The "computational cost issue" can be elegantly solved by multiscale models (macroscale simplifications supported by microscale behavior)
- Can handle complex mathematical relationships
- Can understand the phenomenon with details that cannot be observed or characterized
- More accurate and reliable, based on fewer assumptions and approximations, assumptions and approximations are known, can be scaled reliably and understandably
- Very helpful in scale-up
- Can capture not-so-obvious patterns
- Can be embedded into simpler frameworks for others to use
- Once validated, rigorous models can be extrapolated to wider range
- Easier to "lock" the tool because it takes expertise to use.

Cons of Rigorous Models

- Computationally expensive; cost increases with every additional equation and relationship
- Expensive (time/cost) to develop, maintain, and improve
- Limited to use by specialists
- Requires expertise and deeper scientific knowledge to use correctly

- Too complex to understand by peers
- Difficult to debug
- Will require expensive high-performance computing
- Can provide a false sense of security to users
- May be more difficult to validate/verify correctness
- Can require large data sets, which may not be available
- Tend to be process-specific and not widely applicable
- Tend to be more difficult and expensive to verify with experimental data
- Can be difficult to integrate into TEA/life cycle analysis models
- May require more data that are not available
- Results are not always easy to interpret
- May be challenging to replicate/reproduce
- Can be difficult to set up; upfront analysis may be more labor-intensive than actual implementation
- Confuse complexity for detail and accuracy
- Longer run time required
- Difficult to collect parameters for model
- Can be more sensitive to boundary conditions
- Can lose physical understanding
- Resulting model may be limited to a single application
- Potential waste of time and resources creating a rigorous model for a potentially useless situation
- Too many parameters involved and are more interdependent to each other. A small variance could lead a big error in results.
- Difficult to scale to process or system level
- Can be challenging to debug, verify
- Can be difficult to leverage model and share for collaboration
- In the bioenergy space, the rigorous models are only as good as data sets used to validate the model
- The prediction heavily relies on the constitutive laws
- Some models rely on emerging technologies
- Interpretability can be a challenge.

Best Practices for Developing Rigorous Models

Breakout session participants discussed methods for developing rigorous models, as well as how to determine whether models should be developed utilizing industry learnings, past issues, and first principles. Response categories included assumptions, data needs, direct model integration, identifying failures, measurement methods, and validating models with data.

Participants described their suggested best practices as the following:

- *Base models on data:* Models are used successfully in control system algorithms, but these are for processes that are very well understood with a lot of prior data.
- *Improve solids models:* Solids tracking in process models tend to be not as robust as those for liquids and gases.
- *Understand nonlinearity:* Models have failed when phenomena are not well understood and when there is nonlinearity involved.
- *Understand outliers:* Develop component-based characterization of biomass composition with broad applicability for process stimulation. It is the outliers that cause the problems.
- *Rely on first principles:* Rigorous thermodynamics and kinetics properties can be developed from first-principles quantum chemical calculations and molecular dynamics simulations using carefully benchmarked force fields. These properties can then be used in larger, more rigorous models for scale-up.

The top-ranked ideas among all breakout groups are listed in Table 2. The top-priority ideas identified by each group were:

- Models are used successfully in control system algorithms, but these are for processes that are very well understood with a lot of prior data. These models can also fail in unsteady operations.
- Bridge first-principles chemical kinetics to mesoscale processes like mass and heat transport.
- Verify the model predictions with actual experimental data.
- Identify the critical failure/reasons for biofuel industrial.

Table 2. Best Practices for Developing Rigorous Models

1	Identify the critical failure/reasons for biofuel industrial
2	Validate first-principle methods with good industrial data
3	Hybrid first-principles ML modeling instead of pure/blind ML modeling
4	Ensure that the data quality to build the model is good by developing tools to quantify reactor data quality, leveraging existing literature criteria
5	Learn from past scale-up failures
6	Solids tracking in some process models tend to be weak

7	Do not assume lab results are necessarily representative of scaled-up systems
8	Better measurement methods and protocols for sampling materials at different stages in processing
9	Bridge first-principles chemical kinetics to mesoscale processes like mass and heat transport
10	Perform sensitivity analyses and/or uncertainty analyses as models are developed

Tool Development and Paths Forward

The second breakout session of the workshop began with an independent exercise for participants to compile examples of current models that have been developed utilizing (or in conjunction with) operational data from bench, pilot, and demonstration facilities to augment and accelerate scale-up and integration efforts (including related to biotechnology or outside of the field). The group then reviewed the list together and discussed the weaknesses and constraints of the various process-modeling tools. After prioritizing constraints based on impact and likelihood, each group discussed and prioritized strengths and benefits of current modeling tools.

Current Process Modeling Tool Constraints

Response categories included "upgrading" (advancing or improving) models, accurate model development, data accuracy, standardization, validation, data availability, measuring key parameters, and model development. The challenges were summarized in the session recap as follows:

- *Lack of data:* ML and AI require a lot of consistent training data sets
- *Not based on first principles*: Not reliable outside of the narrow range of input parameters on which models are trained
- *Difficult to use*: ML and AI systems require expertise beyond that of typical process engineers
- *Complex kinetics*: Bio-deconstruction kinetics may be difficult to represent in any mode other than batch
- *Uncertainty*: Challenge of incorporating uncertainty into ML systems.

The top-ranked ideas among all breakout groups are listed in Table 3. The top-priority ideas identified by each group were:

- Integrated biorefinery (IBR) inputs (feedstocks) and outputs (products) are different depending on the process and models do not always capture these complexities
- In general, the industry lacks commercial validation data
- Data quality impacts the model usefulness
- Developing suitable kinetic models for process modeling is very challenging and requires a large, coordinated experimental effort. This can be especially burdensome to experimentalists, whose charter is often to achieve certain productivity metrics rather than produce a data set amenable to kinetic parameter extraction.

Table 3. Current Process Modeling Tool Constraints

1	Development of suitable kinetic models for process modeling requires a large, coordinated experimental effort
2	Lack of validation data
3	Data quality impacts the model usefulness
4	Interpretability
5	Some key parameters that are critical to models can be difficult to measure accurately; having to make assumptions or rely on secondary measurements (e.g., pressure drop) to validate the model is challenging.

Current Process Modeling Tool Strengths and Benefits

Regarding the strengths and benefits of current process modeling tools, workshop breakout session participants mentioned ideas related to challenge identification, process improvement, data quality, model/data utilization, process improvements, and subject-matter expert engagement. The top-ranked ideas among all breakout groups are listed in Table 4. The top-priority ideas identified by each group were:

- Identifying bottlenecks to direct research
- At commercial scale, investigating and de-risking new processes, or even existing processes, with new feedstocks or operating conditions is much faster, cheaper, and less risky *in silico*
    - However, this requires a very robust and well-validated modeling framework to be useful
- Using cross-scale validated process models as design tools
- Improving process control
    - AI teaches the control system in a plant how to improve process by taking real-time data and combining them with legacy data to improve process control.

Table 4. Current Process Modeling Tool Strengths and Benefits

1	Cross-scale validated process models can be used as design tools
2	Improved process control
3	Good quality process models could lead to more credible TEAs
4	Engage the subject-matter experts who developed the model(s)
5	Integration of traditional models into a single workflow exposes and improves relationships between linked processes
6	Improved input data for process models

7	Identifying bottlenecks to direct research
8	Incorporation of domain physics to improve accuracy
9	Eliminating unfeasible designs such as reactors with impossible mass transfer at scale
10	Identification of most-impactful physical factors for process intensification.

Challenges Related to Developing New Models

Breakout session groups pivoted from prioritizing the constraints and strengths of current modeling tools towards challenges related to developing new models. During this discussion, workshop participants provided input related to bridging rigorous and simple models, collaboration, data availability, data quality, funding, and model validation.

One participant described the primary barrier as quantity and quality of data, which requires measurements of input parameters for a process and the output. Machine learning connects the two and builds a model to predict output, given input values. The input in a reactor could be such variables as flow rates, temperatures, and composition. Output could include such variables as yield of biofuels or selectivity of products.

The top-ranked ideas among all breakout groups are listed in Table 5. The top-priority ideas identified by each group highlighted the need for interdisciplinary research teams and physical property input data. The four groups voted the following challenges as the highest combined likelihood of overcoming and impact to advancing the industry:

- Greater collaboration between rigorous modelers, process modelers, and experimentalists
- Bringing together teams of researchers with different skill and knowledge sets
- In bioenergy, with every new feedstock or new catalyst, there is typically a lack of the kinetic rate information or other critical data on which to build the model. The complexity is a challenge.
- Lack of physical property data.

Table 5. Challenges Related to Developing New Models

1	Collaboration between teams of researchers with different skill and knowledge sets
2	Funding for model development
3	Poor quality or incomplete bench-scale experimental data
4	Lack of physical property data
5	Lack of the kinetic rate information or other critical data for new feedstocks and catalysts on which to build the model
6	Lack of physical understanding of pathways and processes
7	Develop translational methodology to capture rigorous high-performance computing models in simpler models accessible to practitioners

8	Funding to do experiments to generate needed data
9	Model calibration and validation
10	Integration of new tools and approaches into commercial software.

Using Models and Operations Data to Predict Scale-Up Requirements

Workshop participants provided input to assist BETO with understanding how models in concert with operations data can be utilized to predict scale-up requirements to move to the next scale. Breakout session groups discussed various approaches, including full-spectrum improvements, identifying focus areas, model development, model improvements, model validation, and process improvements.

One participant explained that it is critical for the industry to have appropriate physical models for transport of material, wear and tear on equipment, and more robust and better measurements for control. These models could serve to address "the most important failure in biomass conversion processes when scaling-up, [which] is material handling, transport, and particle size reduction." In this stakeholder's experience, the uniformization, pretreatment, handling of the material, and its variability at the beginning of the process does not scale with wear and tear on equipment, and these impacts are not yet accurately captured by models.

Similarly, another participant suggested that many biomass processing failures were using material handling and process technologies that worked for other materials. Many times, this may have been done as a cost-saving approach. Potentially, revisiting the physics and science behind the primary designs and inputting better-modeled parameters of the new biomass materials could provide guidance about what interventions are necessary for robust control.

The top-ranked ideas among all breakout groups are listed in Table 6. The top-priority ideas identified by each group were:

- Incorporate "lessons learned" into models to account for differences between bench-scale data and observations at large scale. Support with first-principle science.
- Use data from bench scale to develop models for pilot and demonstration facilities.
- The models can be used to increase confidence for commercial decisions. Extending from experimental data, a model that shows the fundamentals behind the experiment results can give confidence to moving forward with a technology to the next step.
- Modeling suggests what data collection is important, operations feeds back data that validates and improves models, then better models provide better designs.

Table 6. Using Models and Operations Data to Predict Scale-Up Requirements

1	Modeling suggests what data collection is important, operations feeds back data that validates and improves models, then better models provide better designs.
2	Use data from bench scale to develop models for pilot and demonstration facilities.
3	Capturing key process parameters/variables from operational data to facilitate predicting scale-up parameters via a model.

4	The models can be used to increase confidence for commercial decisions. Extending from experimental data, a model that shows the fundamentals behind the experiment results can give confidence to moving forward with a technology to the next step.
5	Build leverageable toolbox of model frameworks for standard kinetic model development to accelerate consistent kinetic-model development (e.g., fixed bed, pore diffusion framework).
6	Understanding mechanics of various reactor designs, models can be used to predict reactor performance to identify correct reactor type at scale-up.
7	Sensitivity analysis can help identify where experimentation can be the most impactful.
8	All operating conditions and feedstock characterizations need to be preserved to be used as boundary conditions and validation for models.
9	Modeling can be used to help experimentalists determine operating conditions and reduce experimental time.
10	Models are better if based in first principles and physically meaningful parameters are fit by appropriate data from operations.

Best Practices for Combining Multiple-Scale Models

As a final topic for the second breakout session, workshop participants discussed their experience with creating a fully modeled biotechnology process combining electro-, nano-, meso-, and CFD-scale models. While providing input regarding the level of combination that modelers can reasonably expect, participants discussed the importance of collaboration, coupling smaller-scale data to larger-scale properties, identifying challenges, integrating models to fill gaps and weaknesses, integrating multiple models, linking experiments and models, linking models based on common properties and interactions, model improvements, training data, and using experimental data to link models.

One participant remarked that connecting software models is not the biggest challenge, but rather the industry should be dedicating resources toward further developing the theory of upscaling. While discussing the critical consistency between scales, another attendee mentioned that "integrated designs require consistency between the parts, but they also tend to require higher funding levels. A balanced approach between smaller, medium, and larger funding opportunity options is critical."

The top-ranked ideas among all breakout groups are listed in Table 7. The top-priority ideas identified by each group were:

- At each scale, find critical parameters and properties, then apply those lessons in the next lower- or higher-scale system. Validation is necessary at each scale.
- Use physics-based insight to couple the relevant physics between different models.
- Handle feedstock complexity at the particle/mesoscale and couple that to bulk hydrodynamics at the reactor scale.
- Incorporate or translate multiple-scale models as plug-and-play submodels inside commercial process simulators.

Table 7. Best Practices for Combining Multiple-Scale Models

1	Incorporate useful training data
2	Involve experts from different disciplines to connect models at different scales
3	Incorporate or translate multiple-scale models as plug-and-play submodels inside commercial process simulators
4	Handle feedstock complexity at the particle/mesoscale and couple that to bulk hydrodynamics at the reactor scale
5	Use physics-based insight to couple the relevant physics between different models
6	Associate well-defined experimental conditions metadata with actual data
7	Use links between single-scale models and other scale models to address missing pieces and weaknesses
8	Use direct simulations of experiments to improve experiment–model connections
9	Develop subroutines within a multiscale framework that evaluate models at specific scales when the global model encounters an uncertainty.
10	Use machine-learning/AI approaches to decide what areas can benefit from extra detail.

Collaboration Considerations

During the third breakout session, workshop participants provided their input to identify which bioenergy technologies have well-developed process models, determine if a library of models/"modeling toolbox" would be useful and achievable, and make suggestions regarding how a potential toolbox could be created.

Throughout the workshop, breakout groups commented about the importance of involving interdisciplinary experts for modeling efforts, including experimental researchers, first-principle scientists, and modelers. The collaboration discussed in the third breakout session explored approaches for encouraging this type of team building and covered collaboration between researchers and commercial enterprises as well as the government and industry in general.

To begin this breakout session, participants were asked to vote on whether a public biotechnology process modeling toolbox should be created to allow for speedier model development. The unanimous response was "yes, it should be created."

Technologies and Unit Operations with Well-Developed Process Models

During an independent activity, workshop participants provided their input about what well-developed process models exist in the bioenergy industry.

One participant noted that most major chemical-producing chemical companies have proprietary well-developed process models that they have used to design and operate their plants. Examples of these models can be found in Hewlett Packard's Cray computer literature from the 1980s and 1990s, when the first HPC applications for process modeling were done.

Participants also suggested that almost all biochemical and thermochemical process models are still "primitive." This may be due to using models from other fields that are not applicable and usually just plain

wrong. BETO consortia (CCPC and FCIC) are funding model development, and as these efforts make progress, it becomes even more clear how much work is still needed and how inapplicable other models were. This participant went on to explain that the existing funding "has gone a long way to take us from the models that are in the literature and used by other industries to models that are more applicable and more predictive for biomass-related processes."

Further, participants noted that the DOE Advanced Manufacturing Office's RAPID Institute, which is focused on modular chemical process intensification, has a project (Project 2.7) to develop an Aspen+ process model for an autothermal biomass fast pyrolysis process. The work is in progress for the fluidized bed reactor.

The participants consider the following technologies and unit operations to have well-developed process models:

- Gas-phase catalytic upgrading
- Gas to ethanol (process commercialized by LanzaTech)
- Dry and wet mill starch-based ethanol processes
- Metabolic modeling (process commercialized by Zymergen)
- Cellulosic biomass conversion: ethanol, drop-in fuels
- Biomass pyrolysis/bio-oil upgrading (however, there are still important knowledge gaps)
- Biopower energy plants (electrical production)
- Fast pyrolysis for some range of variability of biomass feedstocks (particularly at the research level and with respect to the size and shape of pine)
- Fixed-bed and riser-type catalytic upgrading unit ops for thermochemical process (at the research level and limited to catalysts for which there are kinetics available)
- Syngas to products (in general, there are good models for technologies that are similar to more conventional processes; they are often intermediates to products)
- Enzymatic hydrolysis models (reasonably predictive, but largely based on correlations rather than first principles)
- Pretreatment models for existing methods are reasonable
- Thermochemical conversion models are reasonable (in some cases).

Technologies and Unit Operations That Need Further Process Model Development

After developing an inventory of technologies and unit operations with well-developed process models, workshop participants provided their input on where further efforts should be focused. This list identifies potential areas for improvement in process modeling.

One participant remarked that all of the existing models should be developed further, explaining that "none of the models in the biomass area are good enough because of our lack of understanding of biomass chemistry." Another attendee commented that in general, the bioenergy industry needs to dedicate more effort into improved validation and experimental theory/computation coupling. Yet another stakeholder suggested that the industry needs to define how ML/AI tools can be applied across modeling spaces, further explaining that successes for ML/AI have been positive but generally more so within one unit operation or one technical step, while it is still unclear how to effectively implement across operations. Lastly, participants generally indicated

the importance of developing models that may be leveraged by others so that the research community can use the models for their specific process (this could potentially be achieved with model "hooks").

- Continuous operation of nonlinear reaction systems such as biomass deconstruction
- In-line measurements informing PDU models in the national labs (potentially with guidance from industry)
- Enzymatic or organism-based processing (rather than just new chemical approaches to conversion of biomass)
- Waste plastic digestion models
- Capture impacts of non-homogeneity in large-scale fermentation systems
- Incorporate genomics and metabolomics modeling into bioprocessing, especially given the large amounts of data being generated for these areas
- Separations: solid–liquid, liquid–liquid, multiple phases (e.g., solid, liquid, vapor, and gas)
- Catalyst performance, including heat and mass transfer and momentum transport impacts on conversion
- Multiple-phase transport
- High-solid mixing, heat, and mass transfer
- Slurry-phase and bubble-phase reactor modeling
- Combining dynamic physical and mechanical properties and chemical reactions in biomass size reduction
- Biomass particle attrition at various process conditions (e.g., initial particle size, moisture, anatomical components, chemical composition)
- Linking processing modeling tools for operations within other models needed to determine viability, such as life cycle assessment and techno-economic assessments
- Linking available tools to help give a complete view of a process (e.g., combining tools for identifying feedstocks, scale-up, performance, cost, and environmental impact)
- Downstream operations from conversion
- Feedstock (biomass) characterization
- Co-feeding biomass and biocrude into existing refinery infrastructure
- Multiscale models that increase accuracy, especially as processes change and can predict outside known parameter space
- ML/AI-assisted design of experiments, as well as connection between scales
- Biomass to intermediates via most thermochemical or biochemical routes
- Separations process modeling
- Feedstock feeding and handling models are needed

- Reactive comminution pretreatments for biochemical conversion (e.g., deacetylation with mechanical refining)
- Scaling-up and scaling-down bioreactors and biorefineries (including collection and transport of biomass as well as economics of scaling of biorefineries)
- Biological transformations in pilot/demo/commercial bioreactors
- Higher-fidelity multiphysics models for all operations using HPC to accommodate increased computational cost
- ML methods for underlying physics so expensive calculations are not repeated
- Model integration, model calibration/validation, system-level multiphysics models connecting process and natural resources to address environmental impact
- Thermophysical properties for bio-derived molecules
- Phase equilibrium modeling for systems involving oxygenated molecules (abundant in biotechnology processes and relatively few in chemical processes)
- Intensified separation models for adsorption equilibria (gas or liquid) and membrane-based separation
- Translational technology to reduce rigorous reaction kinetics models to simple reaction kinetics models
- Time-domain-specific models (e.g., aging, degradation, time-dependent properties, degree of reuse for recycled stocks).

Creating a Library of Validated Mathematical Models and Software Tools

This discussion focused on the potential role for BETO in supporting the creation of a library of validated mathematical models and software tools relevant to biorefinery and manufacturing operations. Participants provided input related to consortia, data availability, development of interchangeable unit operations framework, encouraging collaboration, funding, and making information available publicly.

One participant commented that "BETO should require project performers to explicitly discuss their plans for validating their individual models, and the integration thereof (hierarchically) across scales. This should also include basic operational tactics like version controlling their model components." Another attendee noted that it would be beneficial to engage with commercial developers such as AspenTech, Siemens, AVEVA, and Chemstations. They explained that this approach would have the potential of being relatively accessible to the research community, compared to high-fidelity models that require supercomputers.

The top-ranked ideas among all breakout groups are listed in Table 8. The top-priority ideas identified by each group were:

- Gather and release current and future biotech models built within the national labs for public access—much like NREL's TEA models
- Host a website listing available tools/models and links/contact info for accessing them
- Establish consortia for continued development, update, maintenance, support, training, etc. of any tools. It is critical that these efforts are long-term and not just one-time sharing.
- Develop and publish a conceptual framework for process modeling and scale-up for biotechnology processes such that modelers can connect their models into that framework.

Table 8. Creating a Library of Validated Mathematical Models and Software Tools: Potential Role for BETO

1	Gather and release current and future biotech models built within the national labs for public access
2	Establish long-term consortia for continued development, update, maintenance, support, training, etc. of any tools.
3	Pilot efforts to demonstrate a framework that can organize and easily swap in/out unit operations
4	Host a website listing available tools/models and links/contact info for accessing them
5	Develop and publish a conceptual framework for process modeling and scale-up for biotechnology processes such that modelers can connect their models into that framework
6	Interface with corporate experts (i.e., bring companies like Aspen to the table)
7	Facilitate the implementation of web application programming interfaces and web apps
8	Providing a funding opportunity dedicated to this goal
9	Use DOE genome and other centers to provide data for development of bio-based models
10	Incentivize data-sharing from experimentalists/pilot plants, modeler-led.

Existing Modeling Collaborations

This discussion focused on establishing a baseline collaboration among model developers, process flow sheeting vendors, and other stakeholders.

- AspenTech works very closely with industry and used to work very closely with HPC vendors.
- DOE consortia, user facilities, manufacturing institutes, and other initiatives (e.g., CCPC, Co-Optimization of Engines and Fuels, Advanced Biofuels and Bioproducts Process Developing Unit, RAPID) encourage collaboration between industry and national laboratories.
    - CCPC is a successful collaboration that is multi-lab and multiscale (atomic-, meso-, and reactor-scale); they are trying to improve industry engagement. There is a lot of collaboration among modelers in bioenergy in the CCPC and this includes collaboration with other BETO experimentalists. Some collaboration has occurred with industry, and industry advisory panels have been used to guide the direction.
    - The RAPID Institute has sponsored multiple process intensification projects in which model developers, software vendors, and experimentalists are working together to develop, evaluate, and raise technology readiness levels for various intensified processes including multiple biotechnology processes.

- Stakeholders share data with model developers so that models can be constructed based on actual conditions. For example, a machine-learning laboratory research project has collaboration between the modeler and the subject-matter expert.
- Researchers often take process flow modeling software from the petroleum industry (e.g., Aspen) and build their own in-house biotechnology functionality into them.
- Process flow sheeting vendors are generally only involved when operations are scaling-up from bench to pilot.
- R&D projects done by universities and national laboratories (including those funded by BETO) often receive in-kind licenses from vendors, and industry often participates in an advisory capacity.

Identifying Opportunities for Collaboration

Breakout session participants were asked to provide input related to potential opportunities to improve and enable relationships between modeling efforts, model developers, scientists, engineers, and other bioenergy stakeholders. Discussions covered topics related to interagency collaboration, Funding Opportunity Announcement (FOA) requirements, user facilities, working groups, workshops, and data collection.

The top-ranked ideas among all breakout groups are listed in Table 9. The top-priority ideas identified by each group were:

- Create a user facility focused on model development and published code
- Funding opportunities that include labs, developers, and commercial modelers (i.e., Aspen) that emphasize collaboration and minimize matching funds.
- Funding with conditions: require partnership including targeted groups (researchers, industry, software, and modelers) in order to consider proposals
- Working groups (specific topics as well as integration efforts).

Table 9. Identifying Opportunities for Collaboration

1	Funding with conditions: require partnership including targeted groups (researchers, industry, software, and modelers) in order to consider proposals
2	Further fund successful inter-lab consortia like CCPC and allow for more academia involvement through Funding Opportunity Announcements
3	Sponsor workshops and symposia
4	Create a user facility focused on model development and published code
5	Facilitate linkages between different DOE programs to garner expertise from various domains
6	Look at published code in a similar way as we look at a published manuscript (as a deliverable for project funding)
7	Provide a list of existing models and capabilities and highlight needs for future model development
8	Engage with software vendors beyond the common request for cost-share contributions of in-kind licenses.

International Lessons Learned

The closing workshop discussion focused on what the United States could potentially learn from international efforts on model development.

Some participants noted that other countries have placed a greater emphasis on workforce development, especially training scientists to develop software for leading-edge computing (both HPC and web). One attendee suggested that 20+ years ago, the United States was the unquestioned leader in model development, but funding for model development was reduced and universities deemphasized programming for getting a science degree. At the same time, scientists in the United States put an overreliance on commercial software rather than custom-built models.

Attendees also noted that the European Union was proactive in policy guidance related to AI. Also, in biotechnology, the European Union has several multi-institutional collaborative efforts that involve both national labs, academia, and industry that span for years.

Appendix A: Agenda

Agenda: Workshop on Predictive Models and High Performance Computing as Tools to Accelerate the Scaling-Up of New Bio-Based Fuels U.S. Department of Energy (DOE), Bioenergy Technologies Office (BETO) June 9–11, 2020 *(Agenda subject to change)*		
Tuesday, June 9, 2020		
Time (EDT)	**Agenda Item**	**Speaker**
10:30 a.m. – 11:00 a.m.	Pre-Discussion Questions, Coffee, and Networking Activities	
11:00 a.m. – 11:30 a.m.	Welcome and Introduction	Josh Messner, DOE BETO
11:30 a.m. – 11:45 a.m.	Introduction to Workshop Software	Lauren Illing, BCS LLC
11:45 a.m. – 12:15 p.m.	Break	
12:15 p.m. – 1:15 p.m.	Modeling and Simulation for Bioenergy Applications: Challenges, Successes, and Lessons Learned	Dr. Peter Ciesielski, National Renewable Energy Laboratory
	Scientific Artificial Intelligence/Machine Learning	Dr. Steven Lee, Advanced Scientific Computing Research, DOE
1:15 p.m. – 1:30 p.m.	Breakout Session Introduction	BETO
1:30 p.m. – 2:30 p.m.	Break	
2:30 p.m. – 4:30 p.m.	Breakout Session 1	
Wednesday, June 10, 2020		
Time (EDT)	**Agenda Item**	**Speaker**
10:30 a.m. – 11:00 a.m.	Pre-Discussion Questions, Coffee, and Networking Activities	
11:00 a.m. – 11:30 a.m.	Welcome and Day 1 Recap	BETO / Rapporteurs – one volunteer from each breakout
11:30 a.m. – 1:30 p.m.	Breakout Session 2 – Part 1	
1:30 p.m. – 2:30 p.m.	Break	
2:30 p.m. – 4:30 p.m.	Breakout Session 2 – Part 2	
Thursday, June 11, 2020		
Time (EDT)	**Agenda Item**	**Speaker**
10:30 a.m. – 11:00 a.m.	Pre-Discussion Questions, Coffee, and Networking Activities	
11:00 a.m. – 11:30 a.m.	Welcome and Day 2 Recap	BETO / Rapporteurs – one volunteer from each breakout
11:30 a.m. – 1:30 p.m.	Breakout Session 3	
1:30 p.m. – 2:30 p.m.	Break	
2:30 p.m. – 3:30 p.m.	Breakout Session Summaries and Conclusions	Rapporteurs – one volunteer from each breakout
3:30 p.m. – 3:45 p.m.	Thank you	BETO

Appendix B: Participant List

First Name	Last Name	Title	Company/ Organization	City	State	E-mail
Zia	Abdullah	Biomass Program Manager	NREL	Golden	CO	zia.abdullah@nrel.gov
Ezinne	Achinivu	Ms.	Sandia National Lab	Berkeley	CA	eachinivu@lbl.gov
Bruce	Adkins	Senior Research Scientist	Oak Ridge National Laboratory	Oak Ridge	TN	adkinsbd@ornl.gov
Syed	Ahsan	Postdoctoral Researcher	National Renewable Energy Laboratory	Lakewood	CO	syed.ahsan@nrel.gov
Olumoye	Ajao	Biorefinery Specialist	Natural Resources Canada - CanmetENERGY	Varennes	QC	olumoye.ajao@canada.ca
Ehsan	Akbari Fakhrabadi	Mr.	The University of Toledo	Toledo	OH	ehsan.akbarifakhrabadi@utoledo.edu
Sneha	Akhade	Dr.	LLNL	Livermore	CA	akhade1@llnl.gov
Srinivasulu	Ale	Associate Professor (Geospatial Hydrology)	Texas A&M AgriLife Research	Vernon	TX	sriniale@ag.tamu.edu
Mouloud	Amazouz	Project Manager	NRCan	Varennes	QC	mouloud.amazouz@canada.ca
Carol	Babb	Vice President	ICF	Golden	CO	carol.babb@icf.com
Raymond	Berrios	Overnight Assistant Manger	28	Lido Beach	NY	Ray.berrios2014@gmail.com
Vivek	Bharadwaj	Scientist	NREL	Golden	CO	vivek.bharadwaj@nrel.gov
Hans	Blaschek	Professor Emeritus	University of Illinois	Champaign	IL	Blaschek@illinois.edu

Jason	Bromley	Director, Process Integration	LanzaTech	Skokie	IL	jason.bromley@lanzatech.com
Paul	Bryan	Dr.	Independent Consultant (representing ESN)	Pinole	CA	separ8r@gmail.com
Trey	Buechler	Process Automation Engineer II	Poet Design & Construction	Sioux Falls	SD	trey.buechler@poet.com
Daniel	Burciaga	President & CEO	ThermoChem Recovery Int'l, Inc. (TRI)	Baltimore	MD	dburciaga@tri-inc.net
Pralhad	Burli	Economist	Idaho National Laboratory	Idaho Falls	ID	pralhad.burli@inl.gov
Eric	Bylaska	Dr.	PNNL/Chemical Physics Theory Team	Richland	WA	eric.bylaska@pnnl.gov
Luis	Carias	Program Data Analyst	DOE	Washington	DC	Luis.Carias@ee.doe.gov
Benjamin	Caudle	Graduate Research Assistant	Texas Tech University	Lubbock	TX	ben.caudle@ttu.edu
John	Caupert	Executive Director	National Corn to Ethanol Research Center	Edwardsville	IL	jcaupert@ethanolresearch.com
Ravi	Chandran	CTO	ThermoChem Recovery International	Baltimore	MD	rchandran@tri-inc.net
Chau-Chyun	Chen	Professor	Texas Tech University	Lubbock	TX	chauchyun.chen@ttu.edu
Rao	Chitikela	Consultant — Water, Energy, & EHSS	RC-WEE Solutions	Dublin	OH	srao.chitikela@gmail.com
Leland	Cogliani	Mr.	Lewis Burke Associates	Washington	DC	leland@lewis-burke.com
Jason	Crater	Director of Fermentation	Genomatica, Inc.	San Diego	CA	jcrater@genomatica.com

Joe	Cresko	Chief Engineer	DOE, AMO	Washington	DC	joe.cresko@ee.doe.gov
Michael	Crowley	Dr.	NREL	Denver	CO	michael.crowley@nrel.gov
Ali	Demirci	Professor	Pennsylvania State University	University Park	PA	demirci@psu.edu
David	Dixon	Robert Ramsay Chair	The University of Alabama	Tuscaloosa	AL	dadixon@ua.edu
Jie	Dong	Assistant Professor	Southern Illinois University Edwardsville	Edwardsville	IL	jdong@ethanolresearch.com
Jim	Dooley	Chief Technology Officer	Forest Concepts, LLC	Auburn	WA	jdooley@forestconcepts.com
Herb	Doty	Technical Specialist	General Motors	Fenton	MI	herb.doty@gm.com
Mahmood	Ebadian	Research Associate	University of British Columbia	Vancouver	BC	mahmood.ebadian@ubc.ca
Mohamed	El Koujok	Dr.	CanmetENERGY-Natural Resources Canada (NRCan)	Varennes	QC	MOHAMED.ELKOUJOK@CANADA.CA
Michael	Fatigati	Mr.	Vantaggio Renewable Energy	Dayton	OH	mfatigati@pacbell.net
Maoqi	Feng	CEO	Polykala Technologies LLC	San Antonio	TX	contact@polykalatech.com
Leonard	Fitts	Burlington City School District	Interim School Superintendent	Moorestown	NJ	drleonardm@aol.com
Roger	Ford	President	Eureka Energy Corporation	Pikeville	KY	roger@eureka-energy.com
Marie-Odile	Fortier	Assistant Professor	University of California, Merced	Fresno	CA	mfortier2@ucmerced.edu
Kimberly	Frazier	President	Qmberli Analytics	Washington	DC	qdp@qmberli.com

Manuel	Garcia-Perez	Professor	Washington State University	Richland	WA	mgarcia-perez@wsu.edu
Greg	Gonzales	Combat Veteran	Walla Walla Community College Student WWCC Equality Club	Walla Walla	WA	usmcgonz3521@gmail.com
Maximilian	Gorensek	Senior Fellow	Savannah River National Laboratory	Aiken	SC	maximilian.gorensek@srnl.doe.gov
Prasad	Gupte	Technology Manager	DOE	Washington	DC	prasad.gupte@ee.doe.gov
Yong	Han	Group Leader	Lawrence Livermore National Laboratory	Livermore	CA	han5@llnl.gov
Peter	He	Associate Professor	Auburn University	Auburn	AL	qhe@auburn.edu
Rodolfo	Hernández-Chaverri	Sr.	State University a Distance	Alajuela	FL	rohernandez@uned.ac.cr
Nathan	Hillson	Staff Scientist	Berkeley Lab	Berkeley	CA	njhillson@lbl.gov
Jamie	Holladay	Manager	Pacific Northwest National Laboratory	Richland	WA	jamie.holladay@pnnl.gov
Hongqiang	Hu	Chemical Engineer	Idaho National Laboratory	Idaho Falls	ID	hongqiang.hu@inl.gov
Marianthi	Ierapetritou	Bob and Jane Gore Centennial Chair	University of Delaware	Newark	DE	mgi@udel.edu
Elchin	Jafarov	Scientist	Los Alamos National Laboratory	Los Alamos	NM	elchin@lanl.gov
Kshitij C	Jha	Founder	Biena Tech	Akron	OH	kcjha@bienatech.com
Wencheng	Jin	Postdoc	Idaho National Laboratory	Idaho Falls	ID	wencheng.jin@inl.gov
Vijaya Gopal	Kakani	Professor	Oklahoma State University	Stillwater	OK	v.g.kakani@okstate.edu

Rahul Reddy	Kancharla	Graduate Research Fellow	Idaho National Laboratory	Idaho Falls	ID	rahulreddy.kancharla@inl.gov
Nazih	Kassem	Student	Cornell University	Ithaca	NY	nk644@cornell.edu
Youngwoon	Kim	Dr.	University of California, Merced	Merced	CA	ykim257@ucmerced.edu
James	Kubicki	Professor	UTEP	El Paso	TX	jdkubicki@utep.edu
Mourad	Lakbiri	Sr.	MASEN	Marrakech	TAB	mourad.lakbiri@outlook.com
Steven	Lee	Program Manager for Applied Mathematics	DOE Advanced Scientific Computing Research	Germantown	MD	Steven.Lee@science.doe.gov
Jeremy	Leong	Technology Manager	U.S. Department of Energy, Advanced Manufacturing Office	Washington	DC	jeremy.leong@ee.doe.gov
Jiaoyan	Li	Postdoc Researcher	Idaho National Lab	Ammon	ID	jiaoyan.li@inl.gov
Matthew	Liberatore	Professor	University of Toledo	Toledo	OH	matthew.liberatore@utoledo.edu
Yingqian	Lin	Postdoctoral Researcher	Idaho National Lab	Idaho Falls	ID	yingqian.lin@inl.gov
James	Lischeske	Research Engineer	National Renewable Energy Laboratory	Arvada	CO	james.lischeske@nrel.gov
Izet	Loloci	Business Development Manager	Blessed Investment	Tirana	AL	iloloci@blessedinvestment.com
Jorge	Lopez Ordovas	Mr.	EBRI, Aston University	Birmingham	AL	lopezorj@aston.ac.uk
Yimin	Lu	Mr.	Idaho National Laboratory	Atlanta	GA	Yimin.Lu@inl.gov
John	Lukas	Senior Consultant	ICF	Woodinville	WA	john.lukas@icf.com

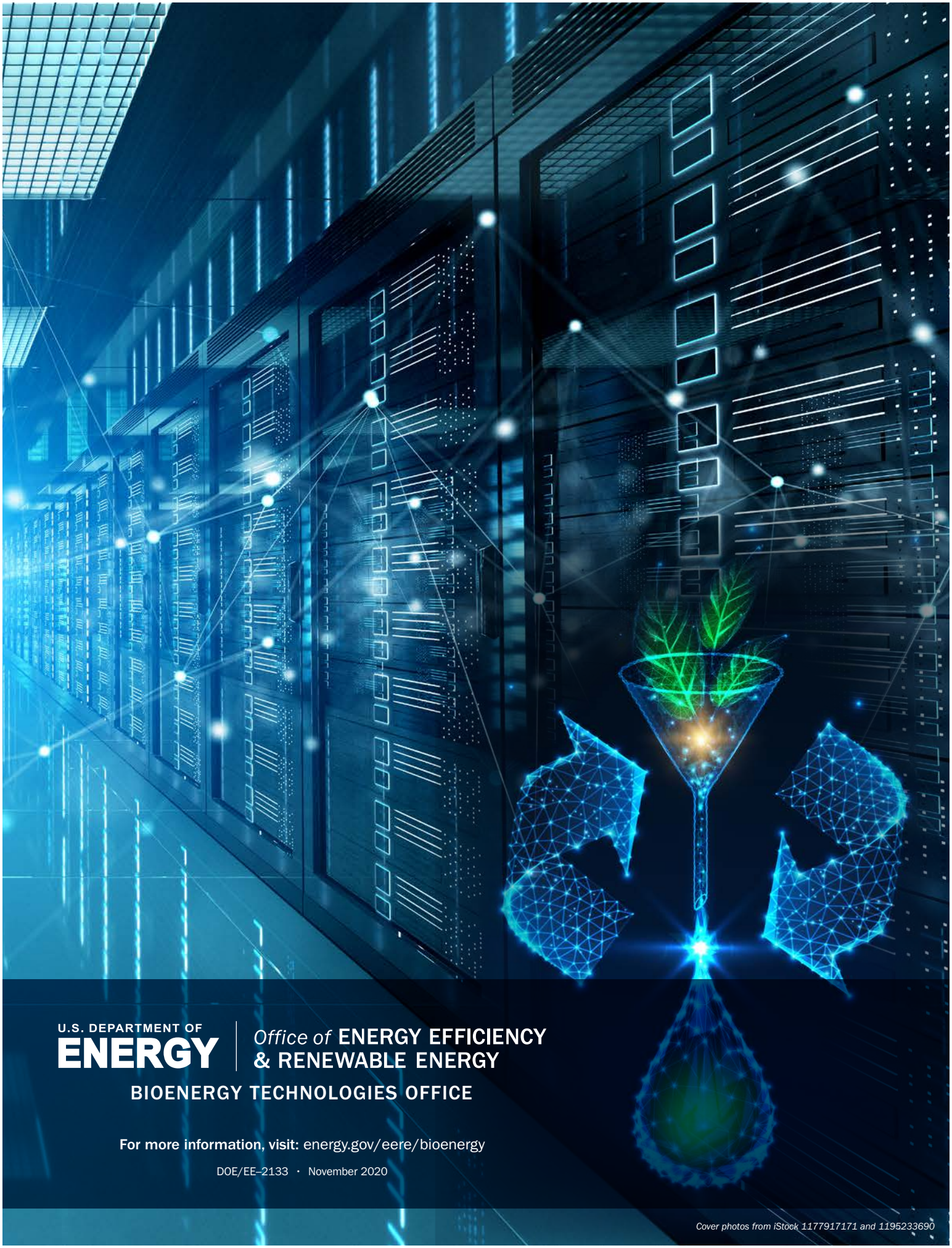
Ehsan	Mahdinia	Dr.	Penn State	State College	PA	mahdinia@psu.edu
Aldo	Martínez Pérez	Chemical Engineer	ITA	Aguascalientes	AGS	aldo_2094@hotmail.com
John	McGowen	Director of Operations	Arizona State University Arizona Center for Algae Technology and Innovation	Mess	AZ	John.mcgowen@asu.edu
Carson	Merkwan	Manager	Direct Automation	Sioux Falls	SD	Carsonm@direct-automation.com
Peter	Miao	Global Workflow Consultant - Flowsheeting Tools	Shell Global Solutions	Houston	TX	peter.miao@shell.com
Maria	Mikitka	Budget Analyst	Department of Energy	Washington	DC	maria.mikitka@hq.doe.gov
Jeffrey	Miller	Project Management	Mobility Management & Marketing	Olympia	WA	MiChargeCommunity@gmail.com
Raul	Miranda	Team Lead	U.S. Department of Energy/Office of Science	Germantown	MD	raul.miranda@science.doe.gov
Dave	Newport	VP, Manager of Projects	ThermoChem Recovery International	Baltimore	MD	dnewport@tri-inc.net
Quang	Nguyen	Mr.	Idaho National Laboratory	Idaho Falls	ID	quang.nguyen@inl.gov
Mark	Nimlos	Separations and Polymer Science Group Leader	National Renewable Energy Laboratory	Golden	CO	mark.nimlos@nrel.gov
Fredrick	Nundu	Chief Executive Officer	Tender Mercies Foundation	South Acworth	NH	nunduf@yahoo.com
Mariefel	Olarte	Research Chemical Engineer	Pacific Northwest National Laboratory	Richland	WA	mariefel.olarte@pnnl.gov

Oluwafemi	Oyedeji	Dr.	Oak Ridge National Laboratory	Oak Ridge	TN	oyedejia@ornl.gov
Asanga	Padmaperuma	Lab Relationship Manager	PNNL	Richland	WA	asanga.padmaperuma@pnnl.gov
Ignasi	Palou-Rivera	Technology Platform Director	RAPID/AIChE	New York	NY	ignap@aiche.org
Sunkyu	Park	Associate Professor	North Carolina State University	Raleigh	NC	sunkyu_park@ncsu.edu
Jim	Parks	PI of the CCPC	Oak Ridge National Laboratory	Knoxville	TN	parksjeii@ornl.gov
Vikram	Pattarkine	CEO	PEACE USA	Mechanicsburg	PA	vikram@peaceusa.net
Brennan	Pecha	Dr.	National Renewable Energy Lab	Golden	CO	brennan.pecha@nrel.gov
David	Punchard	Founding partner	Avespa	Jupiter	FL	dpunchard@avespa.com
Prem Raj	Pushpakaran	Professor	NITC	Calicut	MA	drpremrajp@nitc.ac.in
Rakesh	Radhakrishnan	Tech to Market Advisor	ARPA-E	Washington	DC	rakesh.radhakrishnan@hq.doe.gov
Ahmed	Ragab	Research Scientist	Natural Resources Canada	Varennes	QC	ahmed.ragab@canada.ca
Robert	Rallo	Division Director	Pacific Northwest National Laboratory	Richland	WA	robert.rallo@pnnl.gov
Kalai	Ramea	Research Scientist	Palo Alto Research Center	Palo Alto	CA	kramea@parc.com
Magdalena	Ramirez	Distinguished Research Scientist	Idaho National Laboratory	Idaho Falls	ID	magdalena.ramirezcorredores@inl.gov
Mark	Randall	Pres/CEO	T2 Energy	Encinitas	CA	mrandall@t2energy.com

Vi	Rapp	Research Scientist	Berkeley Lab	Berkeley	CA	vhrapp@lbl.gov
Linda	Rawson	President	DynaGrace Enterprises, Inc.	Morgan	UT	linda@dynagrace.com
Allison	Robinson	Dr.	TDA Research, Inc.	Golden	CO	arobinson@tda.com
Manuel	Rodriguez-Valdez	ENERGY PROJECT MANAGER	GRANJAS CARROLL DE MEXICO	Perote	VER	mrodriguezv@altosano.com
William	Rogers	Senior Research Engineer	National Energy Technology Laboratory	Morgantown	WV	william.rogers@netl.doe.gov
Abhishek	Roy	Senior Data Scientist	Cargill	Wayzata	MN	abhishek_roy@cargill.com
Basudeb	Saha	President	RiKarbon Inc	Newark	DE	bsaha@rikarbon.com
Greg	Shipley	CEO	BioEnergy Design	Paso Robles	CA	Greg@bioenergydesign.com
David	Sievers	Senior Engineer	National Renewable Energy Laboratory	Golden	CO	david.sievers@nrel.gov
Deborah	Sills	Assistant Professor	Bucknell University	Lewisburg	PA	deborah.sills@bucknell.edu
Hari	Sitaraman	Researcher: Mechanical Engineering	NREL	Golden	CO	hariswaran.sitaraman@nrel.gov
Siva	Sivasubramanian	Consultant	Independent Consultant	Wayland	MA	sivasubramanian_m@yahoo.com
Sarah	Smith	Sr. Scientific Engineering Associate	Lawrence Berkeley Lab	Berkeley	CA	sjsmith@lbl.gov
Trevor	Smith	Senior Project Engineer	Boston Government Services, LLC	Golden	CO	trevor.smith@ee.doe.gov
Jim	Spaeth	Bioenergy Program Manager	U.S. Department of Energy	Golden	CO	jim.spaeth@ee.doe.gov

Ranjan	Srivastava	Professor & Dept. Head	University of Connecticut	Storrs	CT	rs@uconn.edu
Peter	St John	Staff Scientist	NREL	Golden	CO	peter.stjohn@nrel.gov
Anne	Starace	Dr.	NREL	Golden	CO	anne.starace@nrel.gov
Vitalii	Starchenko	Dr.	Oak Ridge National Laboratory	Oak Ridge	TN	starchenkov@ornl.gov
Hanifrahmawan	Sudibyo	Mr.	Cornell University	Ithaca	NY	hs987@cornell.edu
Deepti	Tanjore	Director, Advanced Biofuels and Bioproducts Process Development Unit	Lawrence Berkeley National Laboratory	Emeryville	CA	dtanjore@lbl.gov
David	Tew	Program Director	ARPA-E, US Department of Energy	Washington	DC	david.tew@hq.doe.gov
Timothy	Theiss	Group Leader, Renewable Energy Systems	Oak Ridge National Laboratory	Oak Ridge	TN	theisstj@ornl.gov
Nicholas	Thornburg	Postdoctoral Research Engineer	National Renewable Energy Laboratory	Golden	CO	nicholas.thornburg@nrel.gov
Mike	Thorson	Senior Chemical Engineer	PNNL	Richland	WA	michael.thorson@pnnl.gov
Cliff	Tsay	Principal consultant	Quick's Net Consulting	San Jose	CA	cliff.tsay@quixnet.biz
Sheila	Van Cuyk	Program Manager	Los Alamos National Laboratory	Los Alamos	NM	ancuyk@lanl.gov
Arun	Veeramany	Scientist	Pacific Northwest National Laboratory	Richland	WA	arun.veeramany@pnnl.gov

Dion	Vlachos	Professor	Univ. of Delaware	Newark	DE	vlachos@udel.edu
Jin	Wang	Professor	Auburn University	Auburn	AL	wang@auburn.edu
Lloyd	Wilson	Professor and Center Director	Texas A&M University	Beaumont	TX	lt-wilson@aesrg.tamu.edu
James	Winkler	Manager, Metagenomics	LanzaTech, Inc	Skokie	IL	james.winkler@lanzatech.com
Arthur	Wiselogel	Dr.	BGS	Golden	CO	art.wiselogel@ee.doe.gov
Michael	Wolcott	Regents Professor	Washington State University	Pullman	WA	wolcott@wsu.edu
Brandon	Wood	Staff Scientist	LLNL	Livermore	CA	brandonwood@llnl.gov
Yidong	Xia	Computational R&D Scientist	Idaho National Laboratory	Idaho Falls	ID	yidong.xia@inl.gov
Feng	Yan	Professor	University of Alabama	Tuscaloosa	AL	fyan4@ua.edu
Neal	Yancey	Mr.	Idaho National Laboratory	Idaho Falls	ID	neal.yancey@inl.gov
Bin	Yang	Associate Professor	Washington State University	Richland	WA	bin.yang@wsu.edu
Ethan	Young	Researcher	National Renewable Energy Laboratory	Golden	CO	ethan.young@nrel.gov
Bill	Yu	Professor	Case Western Reserve University	Cleveland	OH	xxy21@case.edu
Yan	Zhang	Research Director	National Corn to Ethanol Research Center	Edwardsville	IL	yzhang@ethanolresearch.com
Mingxia	Zhou	Postdoc	Argonne National Laboratory	Lemont	IL	mingxia@anl.gov



U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

BIOENERGY TECHNOLOGIES OFFICE

For more information, visit: energy.gov/eere/bioenergy

DOE/EE–2133 · November 2020

Cover photos from iStock 1177917171 and 1195233690