BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

DOE Bioenergy Technologies Office (BETO) 2021 Project Peer Review

Feedstock to Function Tool: Improving biobased product and fuel development through adaptive technoeconomic and performance modeling

FEEDSTOCK TO FUNCTION

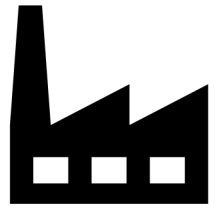
March 23, 2021

Systems Development and Integration (SDI)

Vi Rapp, Corinne Scown, and Kyle Niemeyer

Lawrence Berkeley National Laboratory

Project Overview

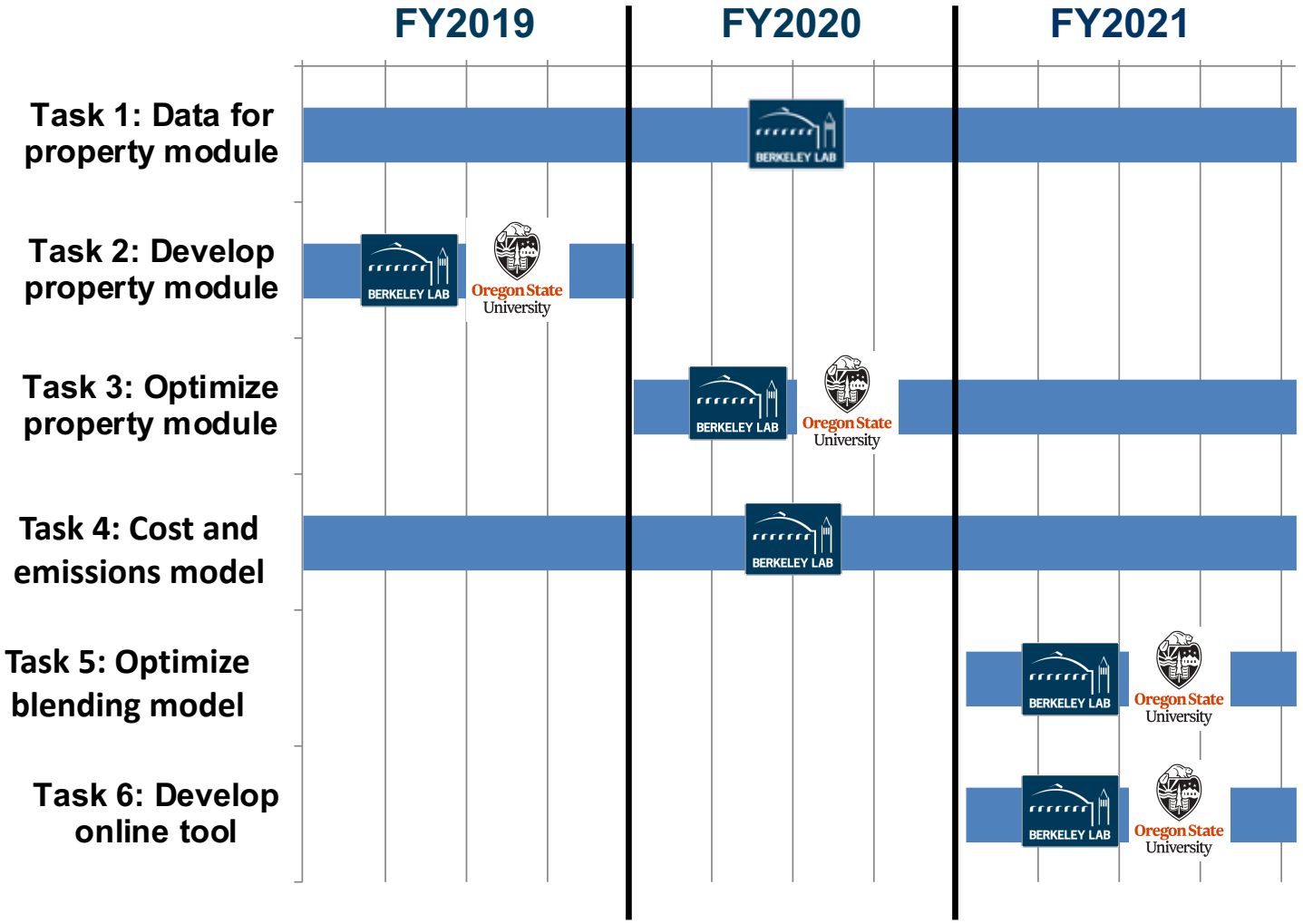
Opportunity: Reduce cost and risk by rapidly screening bio-derived molecules to replace or substitute petrochemical intermediates, fuels, and chemicals

Objective: Develop a 'Feedstock to Function' tool that predicts biomass derived molecule properties and evaluates the cost, benefits, and risk of promising molecules early in the R&D cycle to enable faster, less expensive bioprocess optimization and scale-up

Goals

- Demonstrate webtool concept with molecules to enhance or replace aviation fuels
- Leverage published datasets and machine learning to predict:
  - fundamental aviation fuel properties
  - cost and emissions of promising biobased molecules

1 – Management: Schedule and Risks



Risk	Mitigation Strategy
Data inaccuracy and inconsistency	Identify multiple datasets containing overlapping values; use median or average values for small discrepancies; manually evaluate data with large discrepancies
Machine learning algorithm unable to predict properties to within 15% of experimental values	Investigate feature selection methods, and consider a wide breadth of algorithms and through optimization to achieve at least 15% accuracy
Machine learning methods unable to predict mass/energy balances and costs	Use commercial software to provide reliable cost estimates and explore alternate methods to accommodate non-linear effects

1 – Management: Project Team and Communication

Berkeley Lab

PI
Vi Rapp

Co-PI
Corinne Scown

Applied Math
Ana Comesana

Biological Engineer
Nawa Baral

Software Developer
Tyler Huntington

Oregon State University

OSU PI
Kyle Niemeyer

Chemical Engineer
Morgan Mayer

Website Developer
Lauren Shareshian

Communication & Collaboration

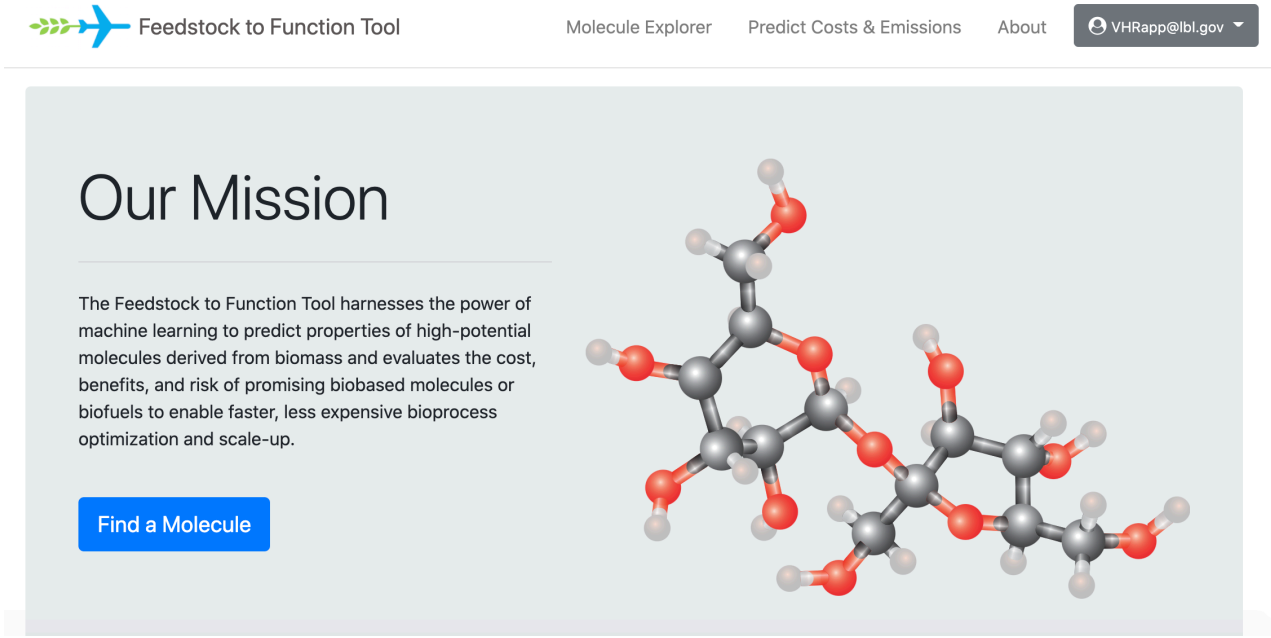
- The Center of Excellence for Alternative Jet Fuels and Environment (ASCENT)
- Co-Optima Program
- Existing Biojet efforts: Sandia, NREL, PNNL, Dayton University, Georgia Tech
- Industry (e.g. chemical)
- Naval Air Warfare Center Weapons Division

This project uses high performance computing resources located at Oak Ridge National Laboratory and provided by the Bioenergy Technologies Office.

2 – Approach

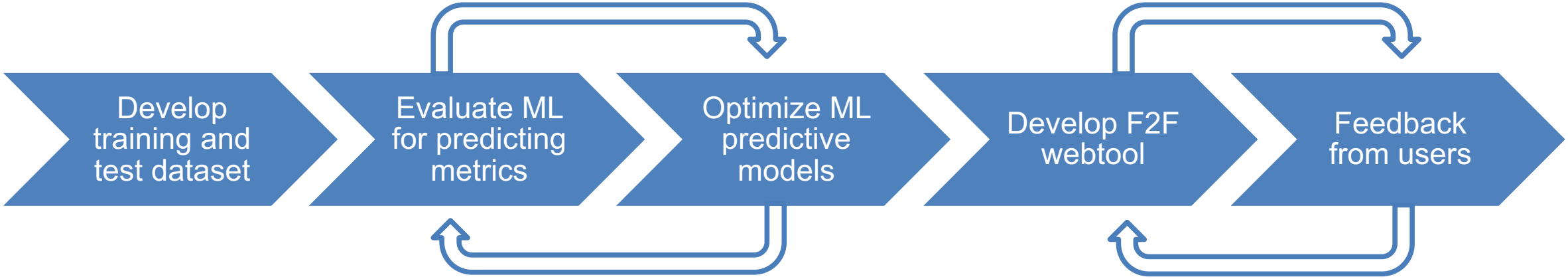
Develop an opensource tool that rapidly screens bio-derived molecules to replace or substitute petrochemical intermediates, fuels, and chemicals

- Accelerate innovation and development of commercially viable, high-performance biofuels, bioproducts
  - Support BETO with growing America's energy future
  - Addresses BETO MYPP barriers (ADO-C, Ct-J, Ct-N)



- Potentially reduce molecule exploration from months or years to days or weeks
- Rapidly screen for viable molecules to accelerate R&D efforts, reduce expenditure, and enable more productive and successful experimentation

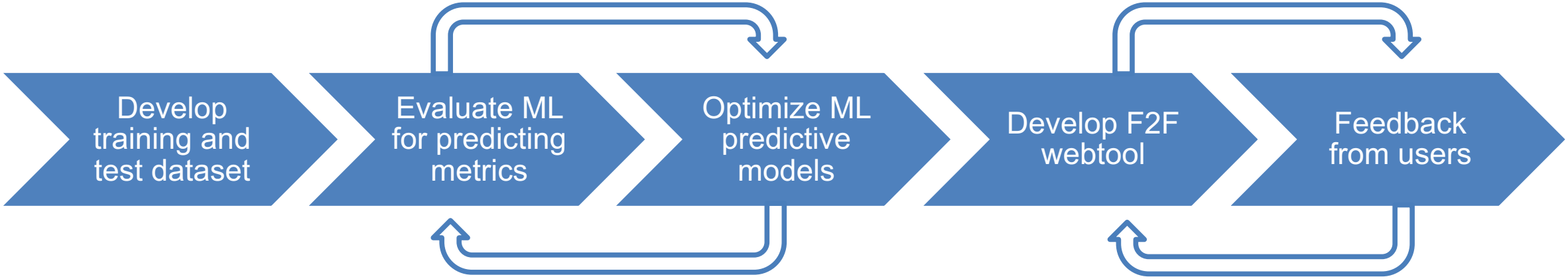
2 – Approach



Technical Approach

- Collect and clean training data set (PubChem, Co-Optima database, NIST, trials from process simulation outputs using commercial software, and other published data)
- Develop machine learning (ML) model for predicting:
  - Fundamental properties of pure (neat) molecules and blends for aviation fuels
  - Minimum selling price (MSP) and greenhouse gas (GHG) emissions of select aviation fuels
- Optimize and validate accuracy of ML models by comparing to published data
- Publish free webtool that allows users to explore viable molecules and associated costs and emissions with production and scale-up

2 – Approach



Metrics of Success

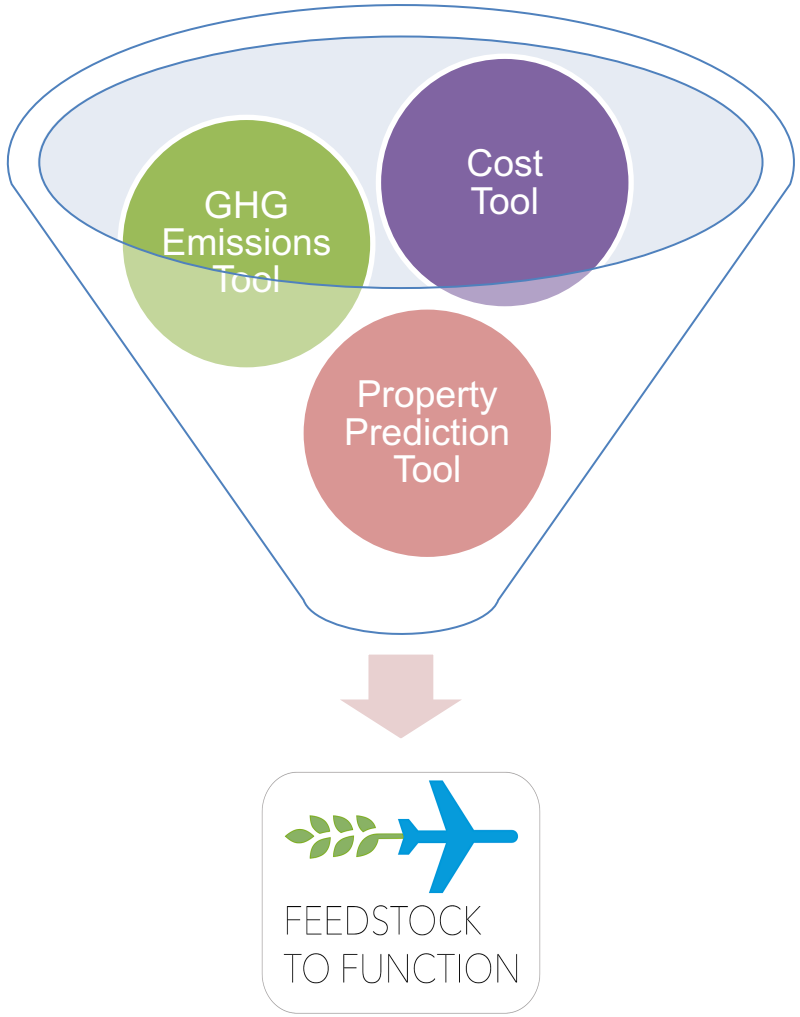
- Predict properties of select jet fuels within 15% of published values (Go/No-Go)
- Predict costs within $1.00/gal and emissions within 10% of commercial software
- Publish online F2F tool that predicts properties for potential alternative jet fuel molecules

Challenges

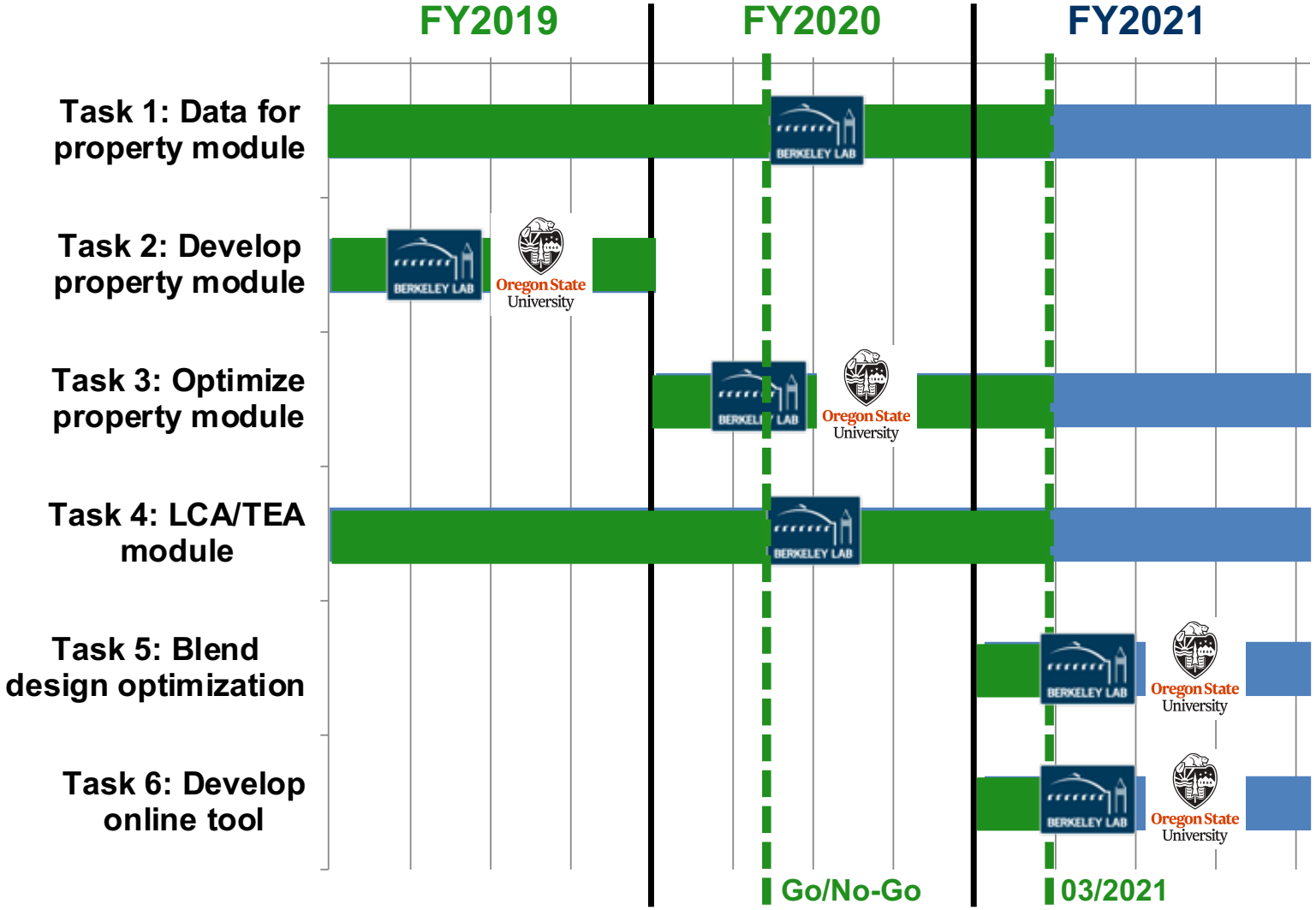
- Finding enough data to develop accurate machine learning (ML) models
- Validating accuracy and reliability of compiled data
- Validating ML methods can accurately predict mass/energy balances and costs

3 – Impact

- Current approach hinders innovation as it requires multiple tools, domain expertise for each tool, and potentially months to years of development for property validation
- The ‘Feedstock to Function’ tool will encourage innovation and accelerate early R&D by:
  - enabling users to rapidly and seamlessly screen molecule properties, cost, and emissions of viable molecules
  - identifying new molecules and opportunities previously unexplored
  - establishing a framework for a designing new classes of molecules amenable to novel production pathways
- Early and ongoing feedback from industry informs tool improvements to maximize impact and use



Progress and Outcomes

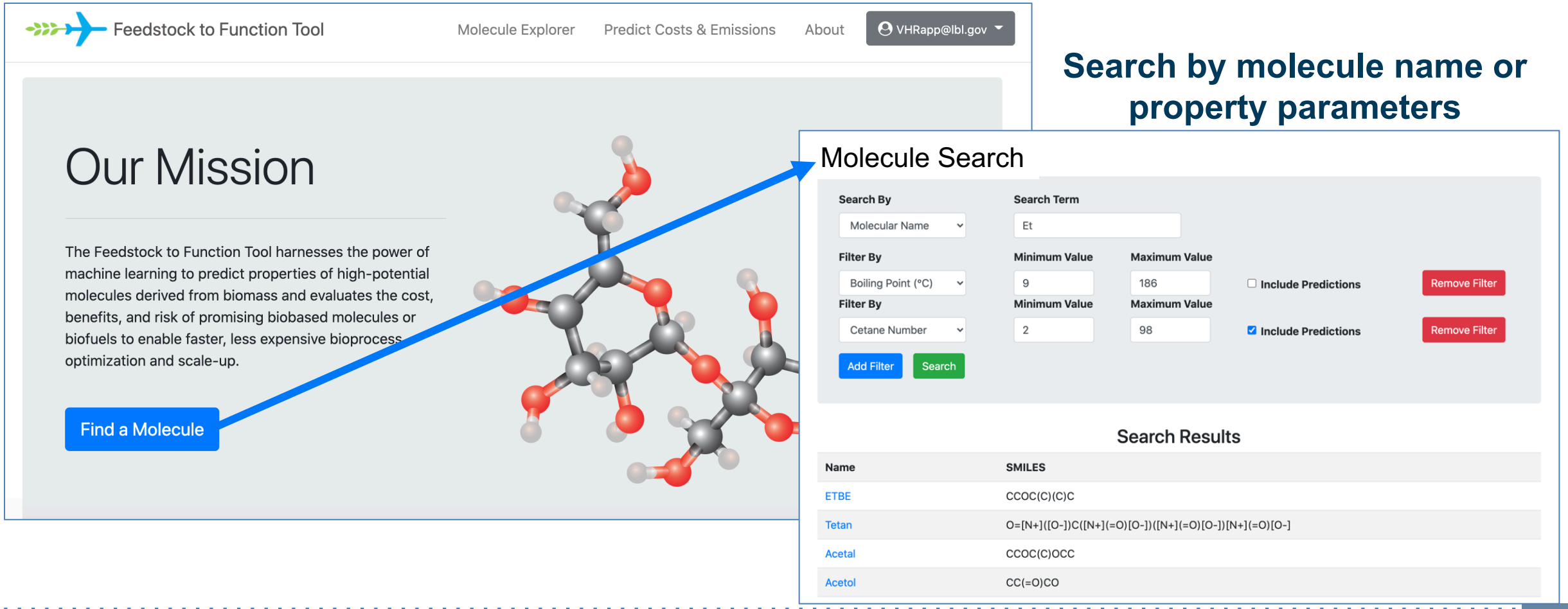


- ✓ FY2019: Demonstrate accuracy of pure (neat) fuel property predictions, minimum selling price, and emissions.
- ✓ FY2020: Demonstrate accuracy of fuel blend property predictions, minimum selling price, and emissions.
- ✓ Go/No-Go: Property prediction of F-24 Blend within 15% for flash point and cetane number (Go/No-Go Milestone)
- FY2021: Release online F2F tool that predicts properties, costs, and emissions for potential alternative jet fuel molecules.

4 – Progress and Outcomes

Published preliminary webtool for evaluation and feedback

https://feedstock-to-function.lbl.gov/



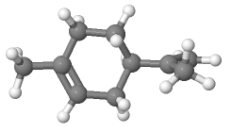
Search by molecule name or property parameters

4 – Progress and Outcomes

Displays predicted properties, cost, and emissions for molecules

Chemical Details

Structure



JSmol

SMILES ID	C=C(C)C1CC=C(C)CC1
Molecular Name	DIPENTENE
Molecular Formula	C10H16
IUPAC Name	1-methyl-4-prop-1-en-2-ylcyclohexene
InChiKey	XMGQYMWWDOXHJM-UHFFFAOYSA-N
Synonyms	DIPENTENE Kautschin DL-Limonene Cajeputene Cinene Dipentene LIMONENE

Properties

Property	Experimental Value	Source(s)	Predicted Value	Estimator
Boiling Point	175.2 °C	PubChem CoOptima	176.2 ± 8.3 °C	Boiling Point Prediction Algorithm
Melting Point	-95 °C	Bradley	-75.6 ± 5.8 °C	Melting Point Prediction Algorithm
Flash Point	44.9 °C	Saldana	46.2 ± 2.4 °C	Flash Point Prediction Algorithm
Cetane Number	18.9	ECNet	18.9 ± 4.0	Cetane Number Prediction Algorithm
Yield Sooting Index	136.7	Yale	137.0 ± 17.8	Yield Sooting Index Prediction Algorithm

Predict Life-cycle Costs & Emissions | Download Data as CSV

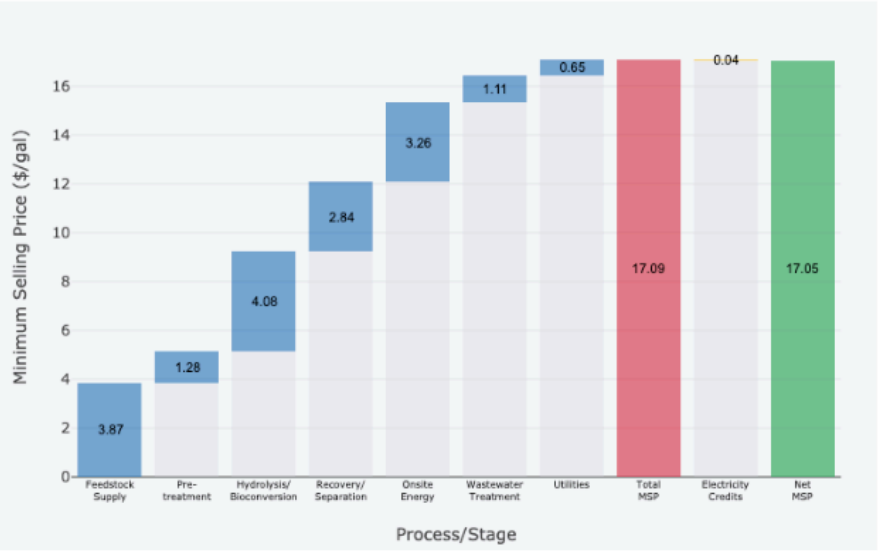
Output Molecule: Limonene | MSP Model | GHG Model | Run Model | Tech Specs

Set Parameters to Scenario Defaults:

State of Technology | Baseline | Optimal

Note: Modeling methods are still experimental in nature. Results should be interpreted with caution.

- Feedstock Supply
- Biomass Pretreatment & Neutralization
- Enzymatic Hydrolysis & Bioconversion
- Recovery & Separation
- Wastewater Treatment
- On-site Energy Generation
- Economic Evaluation Parameters
- Other Parameters

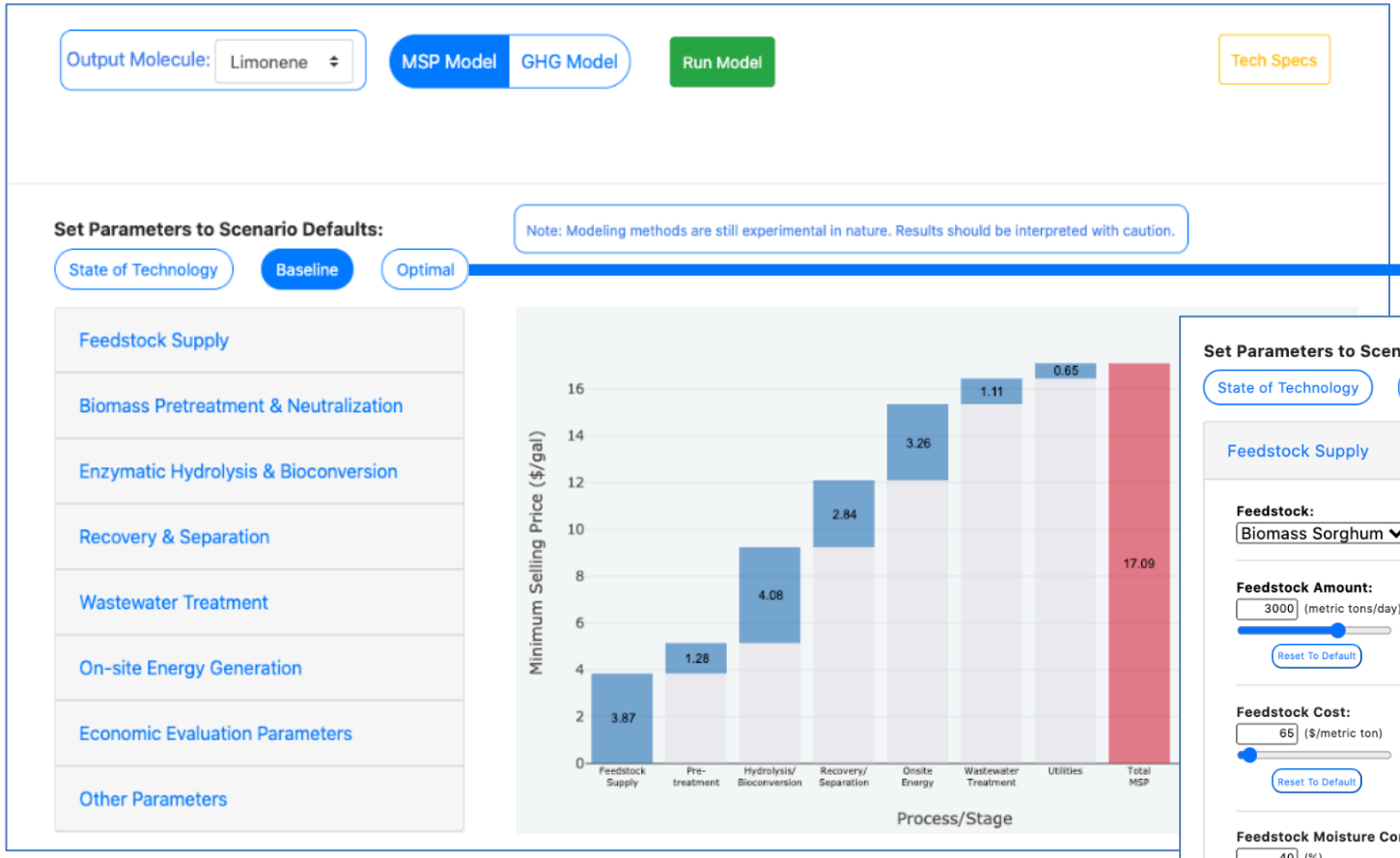


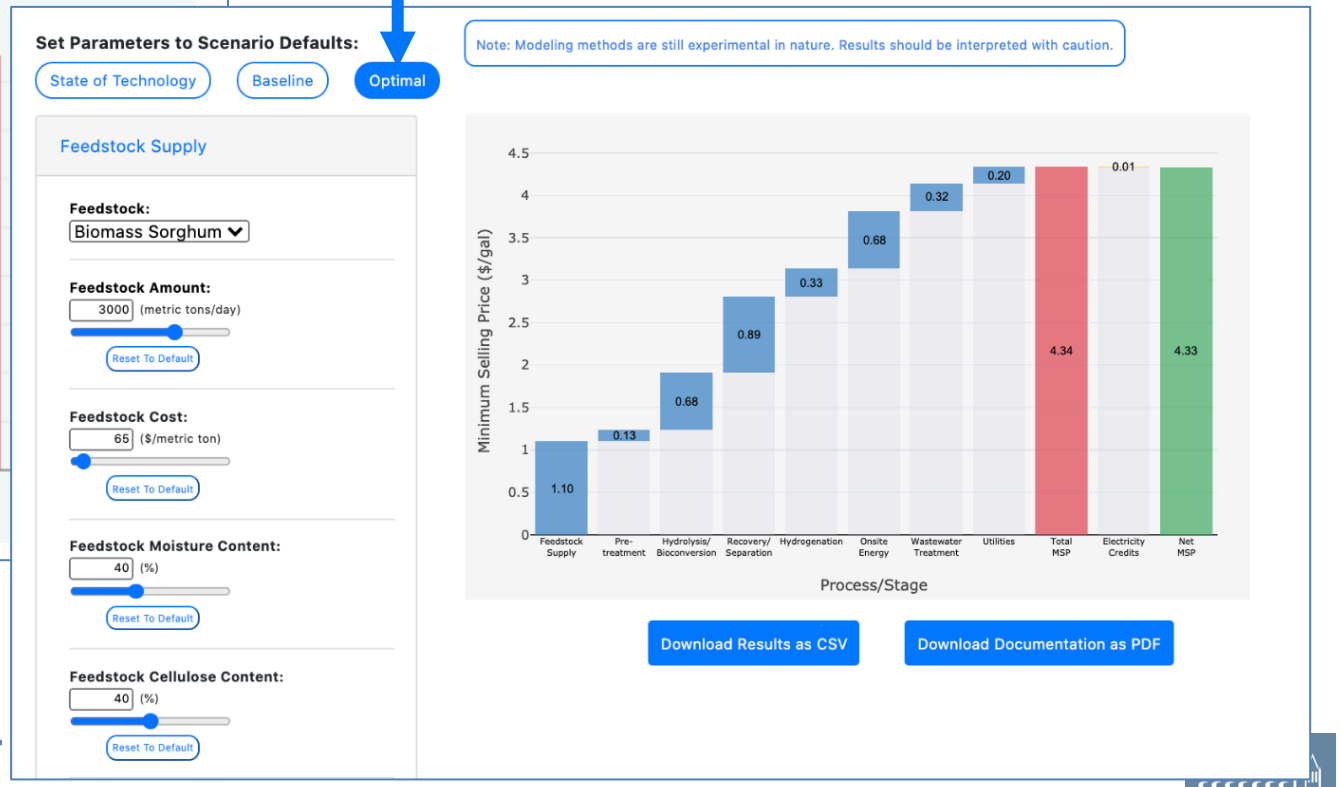
Download Results as CSV | Download Documentation as PDF

BERKELEY LAB

4 – Progress and Outcomes

Vary parameters to explore effect on cost and emissions





4 – Progress and Outcomes

Data collected and cleaned to train machine learning models

Fundamental property data

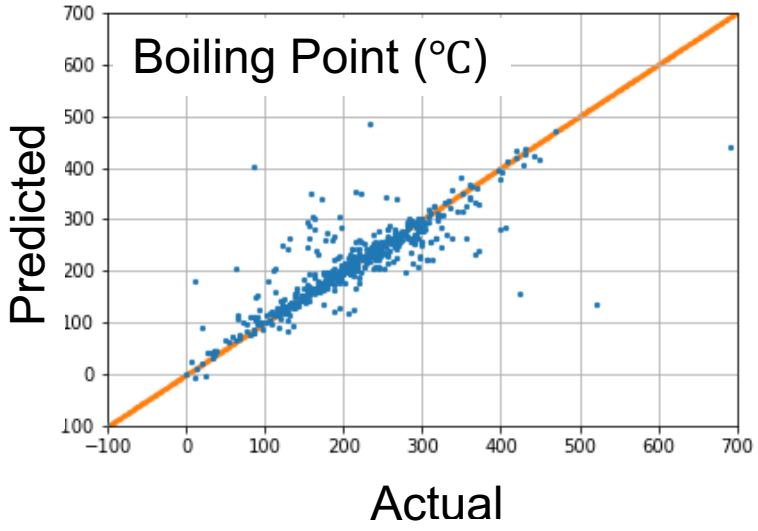
	Flash Point	Boiling Point	Melting Point	Cetane Number	Yield Sooting Index	Heat of Combustion
Number of neat molecules	1,221	2,604	11,327	474	485	349
Number of neat FTIR spectra	45	38	43	49	36	41
Number of blend FTIR spectra	62	59	62	50	0	52

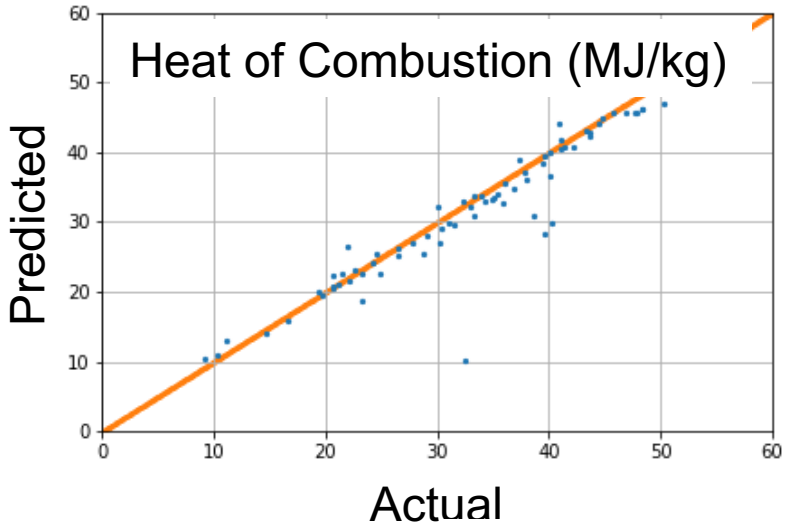
Cost and emissions data

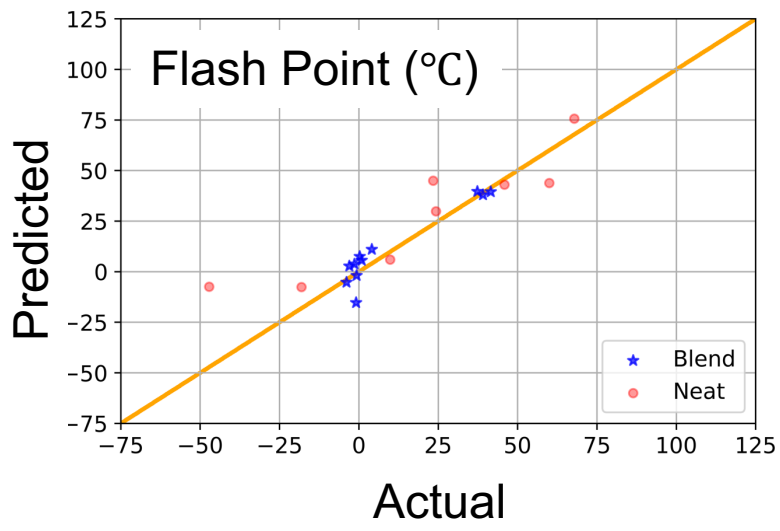
- Data from 8,000 trials of process simulation outputs using commercial software
- Molecules evaluated include limonane, limonene, bisabolane, bisabolene, and ethanol

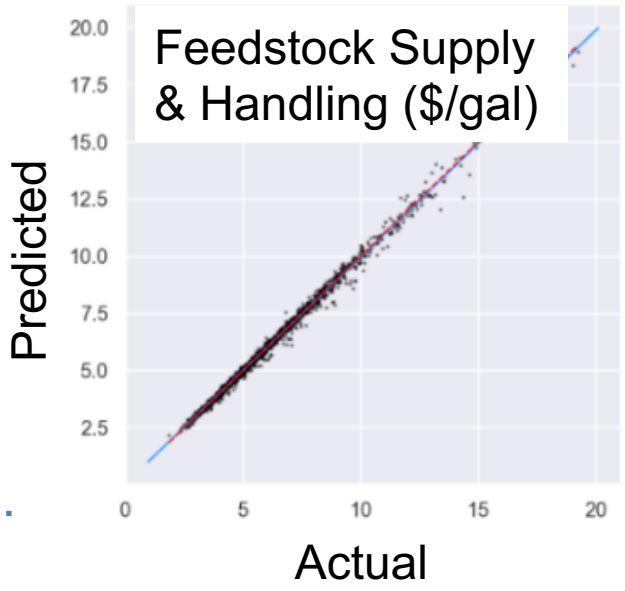
4 – Progress and Outcomes

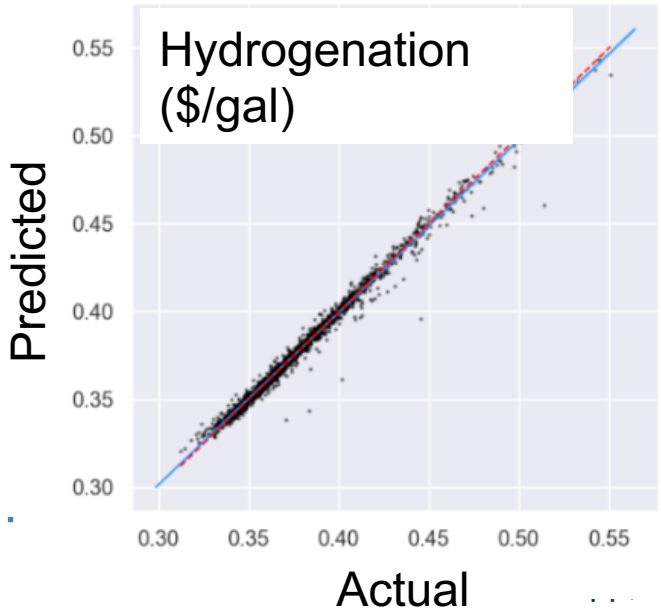
Machine learning models met performance targets











Summary

1. **Goal:** Develop a 'Feedstock to Function' tool that predicts biomass derived molecule performance and evaluates the cost, benefits, and risk of promising molecules early in the R&D cycle to enable faster, less expensive bioprocess optimization and scale-up

2. **Approach:** Demonstrate webtool concept with molecules for aviation fuels; leverage machine learning to predict fundamental aviation fuel properties, and cost and emissions of promising biobased molecules; validate models with experimental and published data

3. **Progress:** Published preliminary online webtool (feedstock-to-function.lbl.gov) and receiving feedback; met predictive performance targets for property predictions, minimum selling price, and GHG emissions; established database with thousands of molecules and blends

4. **Potential impact:** Accelerate R&D efforts, reduce expenditure, enable productive and successful experimentation, and identify new opportunities previously unexplored

5. **Future work:** Incorporate user feedback on webtool and complete fuel blend property predictions; expand webtool for other applications

Quad Chart Overview

Timeline

- Project start date: 10/01/2018
- Project end date: 9/30/2021

	FY20	Active Project
DOE Funding	$350K	$1050K

Project Partners

- Oregon State University (OSU)

Barriers addressed

- ADO-C. Codes, Standards, and Approval for Use
- Ct-J. Identification and Evaluation of Potential Bioproducts
- Ct-N. Multiscale Computational Framework toward Accelerating Technology Development

Project Goal

Develop a webtool that predicts biomass derived molecule performance and evaluates the cost, benefits, and risk of promising bio-based fuels and products. Allows researchers to ”fail quickly” with little cost.

End of Project Milestone

Demonstrate F2Ft predicts properties for up to three different blending ratios of well characterized or certified biobased molecule to within 15% of published experimental values. Develop an open-source, online platform tool for predicting molecule properties based on the predictive algorithm. Combine fuel module and LCA/TEA module. Publish results in peer-reviewed journal.

Funding Mechanism

BETO Lab Call, 2018

ADDITIONAL SLIDES

Responses to 2019 Reviewers' Comments

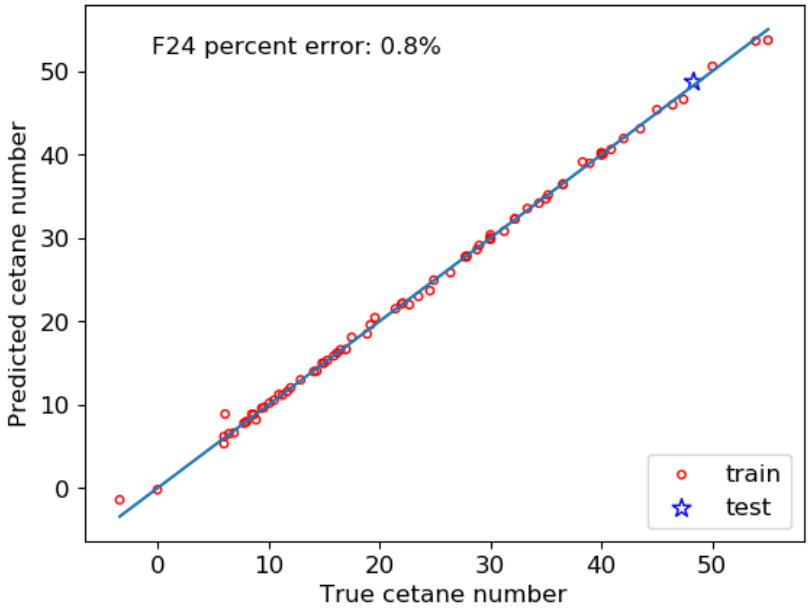
Comment 1: *"Approach does not seem to include a clear means of validating data that comes externally into the project; garbage in, garbage out is a great risk when pulling data from the internet and many other sources, and project should include some means of validating the molecule properties that it collects as it goes."*

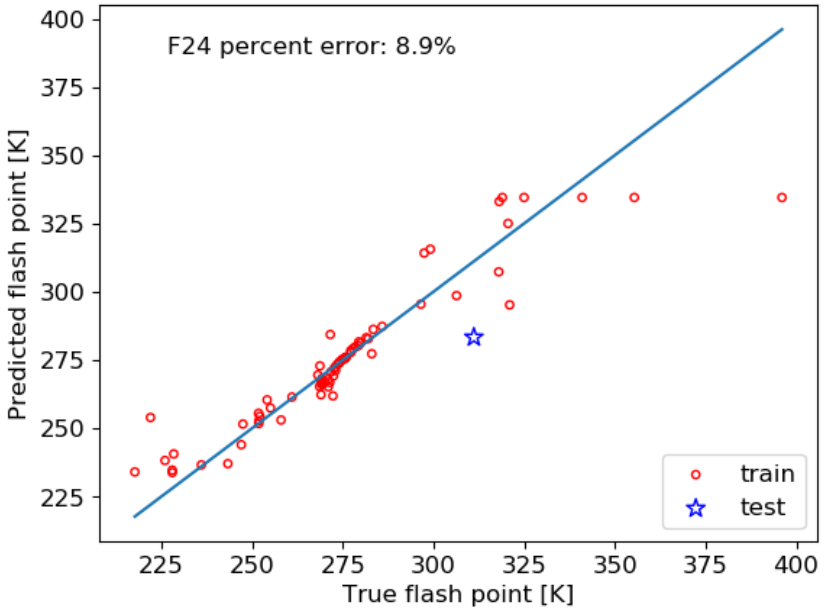
Response 1: We took several approaches to ensure the data for property training was of good quality. We collected data from reputable sources that are screened, we collected overlapping data for cross validation, we plotted data to visualize discrepancies, and we manually checked values to correct any potential errors. We were also able to correct data using the predictive model results. For the TEA/LCA tool, training data was generated using validated commercial software.

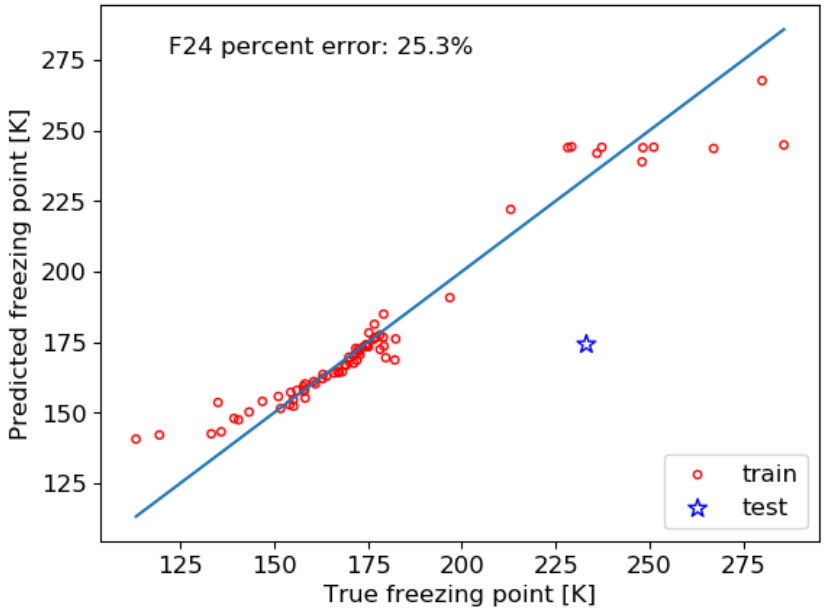
Comment 2: "Scientific basis supporting analysis rationale is not clearly identified. Model will only add value if it has reliable basis."

Response 2: To ensure model is supported by science, we have developed a feature selection method that ensures selected features have high statistical significance and scientific rationale. We have also validated our findings with published literature.

FY20 Go/No-Go Milestone: Property prediction of F-24 Blend within 15% for flash point and cetane number

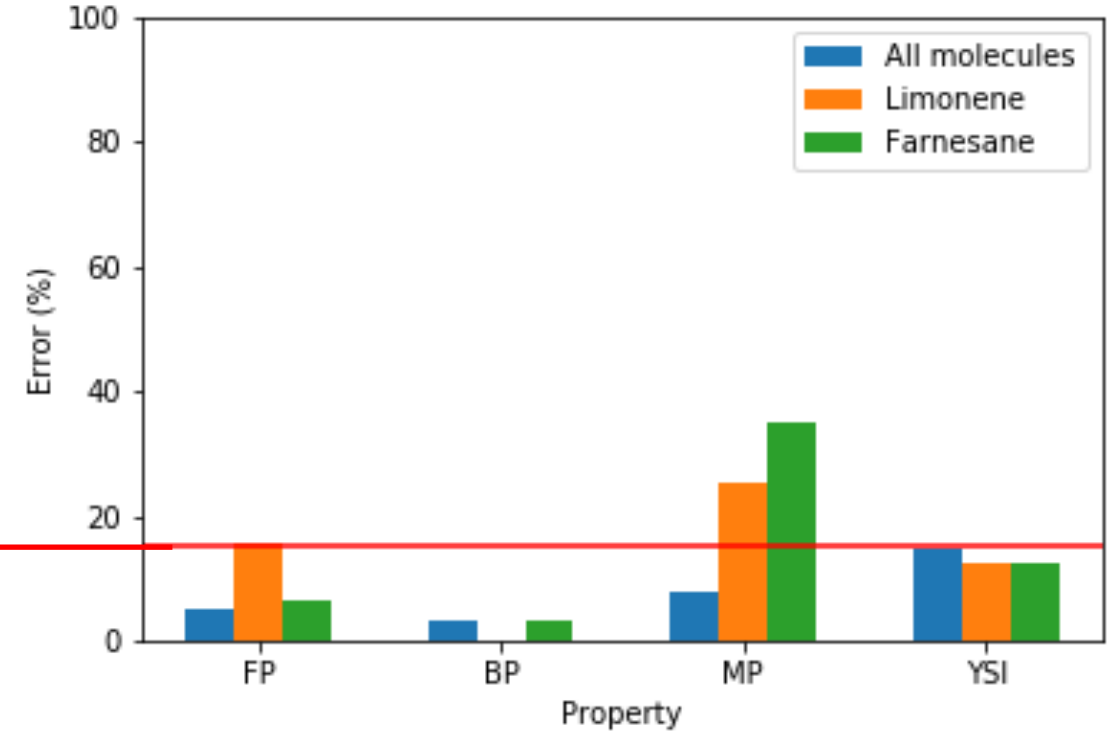






- Model correlates FTIR spectra with desired properties
- Dataset 72 fuels for freezing point, 74 fuels for flash point, and 69 fuels for cetane number (mostly gasoline-like molecules and blends)
- Additional jet fuel and jet-blend data would vastly improve the prediction for freezing point
- Approach shows promise for accurately predicting blend properties; recommend the blend design optimization task commence

FY20 Go/No-Go Milestone: Most property predictions within 15% of experimental value for neat molecules



Project goal:
<15% prediction error

MP	Melting Point
BP	Boiling Point
FP	Flash Point
YSI	Yield Sooting Index

% error for fuel test data (20% of dataset) is calculated as:

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{actual_i - prediction_i}{actual_i}\right|$$

Presentations and Webtool Release

Webtool Release

Feedstock to function website https://feedstock-to-function.lbl.gov/

Presentations

- Vi Rapp, “Machine learning for fuels, chemicals and food ingredients,” UC Congressional Briefing: The promise of Artificial Intelligence Research in Washington D.C. on Dec. 11, 2019.
- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Optimizing Humans and Machines to Advance Science” presented virtually at SciPy Conference in July 2020. (https://youtu.be/ENOf0IZDla8)
- Corinne Scown, "Using TEA and LCA to Answer Your Burning (Cost and Carbon-Related) Questions," at the Joint BioEnergy Institute in Berkeley, CA on Oct 23, 2019.
- Corinne Scown, “"Sustainability at the Joint BioEnergy Institute," AIChE Bioenergy Sustainability Conference in Nashville, TN on Oct. 21, 2019.
- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Innovative Biofuel Development through Machine Learning” poster presented at Bay Area Scientific Computing Day in Berkeley, CA on Dec. 16, 2019.
- Morgan Mayer, Tyler Huntington, Ana Comesana, Vi Rapp, Kyle Niemeyer, “Can machine learning predict fuel properties accurately?” Fall 2019 Western States Section of the Combustion Institute Meeting in Albuquerque, NM on Oct. 14, 2019.
- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Predicting Bio-jet Properties Using a Tree-Based Pipeline Optimization Tool” paper accepted for presentation at Spring Western States Section of the Combustion Institute in Stanford, CA on TBD (rescheduled due to COVID-19).
- Morgan Mayer, Ana Comesana , Tyler Huntington, Vi Rapp, Kyle Niemeyer, “Challenges in predicting fuel properties with machine learning” paper accepted for presentation at Spring Western States Section of the Combustion Institute in Stanford, CA on TBD (rescheduled due to COVID-19).