BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

DOE Bioenergy Technologies Office (BETO) 2019 Project Peer Review

ADO Session Presentation

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

March 7, 2019

Technology Session Area Review

Vi Rapp, Corinne Scown, and Kyle Niemeyer

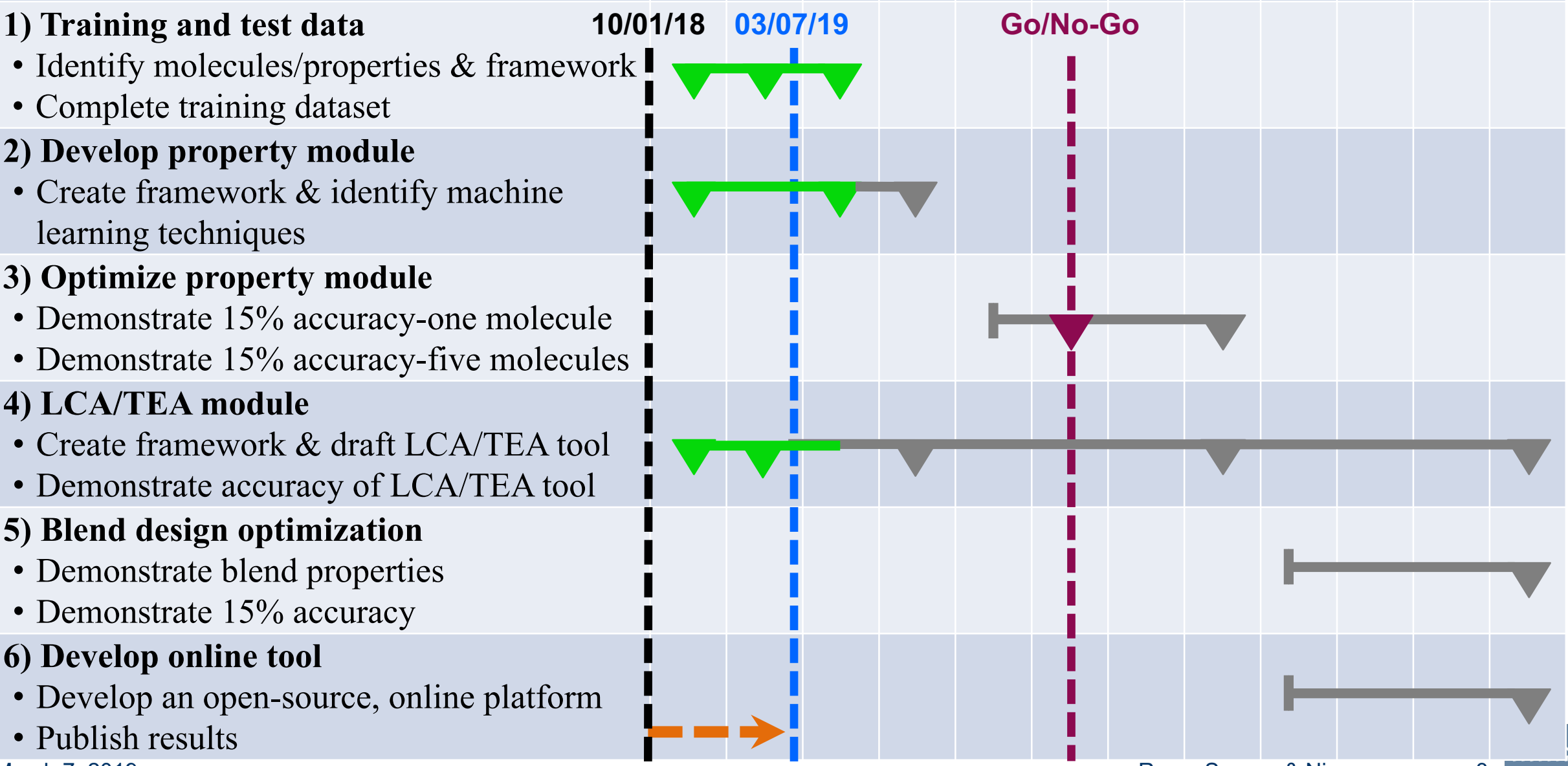
Lawrence Berkeley National Laboratory

Goal Statement

Increase domestic energy security and reducing U.S. reliance on foreign oil by developing a flexible, open-source tool that:

- Predicts properties of high-potential molecules (fuels, fuel co-products, and other bioproducts) derived from biomass using the power of machine learning
- Evaluates the cost, benefits, and risk of promising biobased molecules or biofuels to enable faster, less expensive bioprocess optimization, certification, and scale-up
- Aids industry with making progress towards the $3 per gallon goal

TASKS AND MILESTONE	FY 2019				FY 2020				FY 2021			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1) Training and test data • Identify molecules/properties & framework • Complete training dataset												
2) Develop property module • Create framework & identify machine learning techniques												
3) Optimize property module • Demonstrate 15% accuracy-one molecule • Demonstrate 15% accuracy-five molecules												
4) LCA/TEA module • Create framework & draft LCA/TEA tool • Demonstrate accuracy of LCA/TEA tool												
5) Blend design optimization • Demonstrate blend properties • Demonstrate 15% accuracy												
6) Develop online tool • Develop an open-source, online platform • Publish results												



Project Budget Table

	Original Project Cost (Estimated)		Project Spending and Balance		Final Project Costs
Budget Periods	**DOE Funding**	**Project Team Funding Allocation**	**Spending to Date**	**Remaining Balance**	**Funding needed to complete project**
FY 2019	**$350K** ***($325K received)***	**LBNL (78%) OSU (22%)**	**$14K (spent) $52K (encumb.)**	**$284K** ***($259K of received)***	
Task 1: Training and test datasets		LBNL (100%)			
Task 2: F2F property module		LBNL (53%) OSU (46%)			
Task 4: F2F LCA/TEA module		LBNL (100%)			

Project Budget Table (continued)

	Original Project Cost (Estimated)		Project Spending and Balance		Final Project Costs
Budget Periods	**DOE Funding**	**Project Team Funding Allocation**	**Spending to Date**	**Remaining Balance**	**Funding needed to complete project**
FY 2020	**$350K**	**LBNL (69%) OSU (31%)**			
Task 3: Optimize property module		LBNL (64%) OSU (36%)			
Task 4: F2F LCA/TEA module		LBNL (100%)			
FY 2021	**$350K**	**LBNL (81%) OSU (19%)**			
Task 4: F2F LCA/TEA module		LBNL (100%)			
Task 5: Blending Design		LBNL (54%) OSU (46%)			
Task 6: F2F online tool		LBNL (89%) OSU (11%)			

Quad Chart Overview

Timeline

- Project start date 10/1/2018
- Project end date 9/30/2021
- Percent complete ~14%

	FY 19 Costs	Planned FY 20 Costs	Planned FY 21 Costs	Total Planned Funding (FY 19-Project End Date)
DOE Funded	$350K	$350K	$350K	$1050K
Project Cost Share*	LBNL (78%) OSU (22%)	LBNL (69%) OSU (31%)	LBNL (81%) OSU (19%)	LBNL (76%) OSU (24%)

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

Partners:

Oregon State University (OSU)

Other Collaborations:

- Coordinating with NREL to obtain relevant data from Co-Optima database
- Collaborate with and support existing efforts for a future jet program: Sandia, NREL, PNNL, Dayton University, Georgia Tech

1 - Project Overview

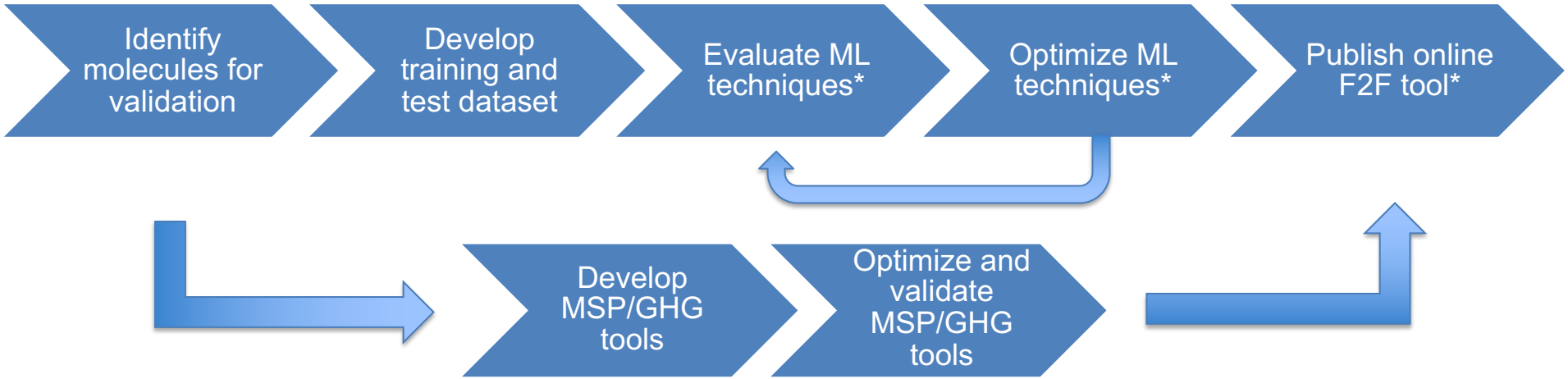
Advancements of bio-derived alternative jet fuels are limited by significant technical, social, and regulatory barriers

Standard property testing biofuel and biofuel blends is expensive, requires high fuel-volume for certification, and is conducted years after initial bench-scale experiments

High-Level Objectives:

- Develop and optimize machine learning (ML) techniques to predict properties of selected neat potential alternative jet molecules and blends
- Develop lightweight minimum selling price (MSP) and greenhouse gas (GHG) emissions estimation tool
- Develop and demonstrate an online tool that uses ML to predict desired properties and evaluate the cost, benefits, and risk of promising bio-based molecules or biofuels

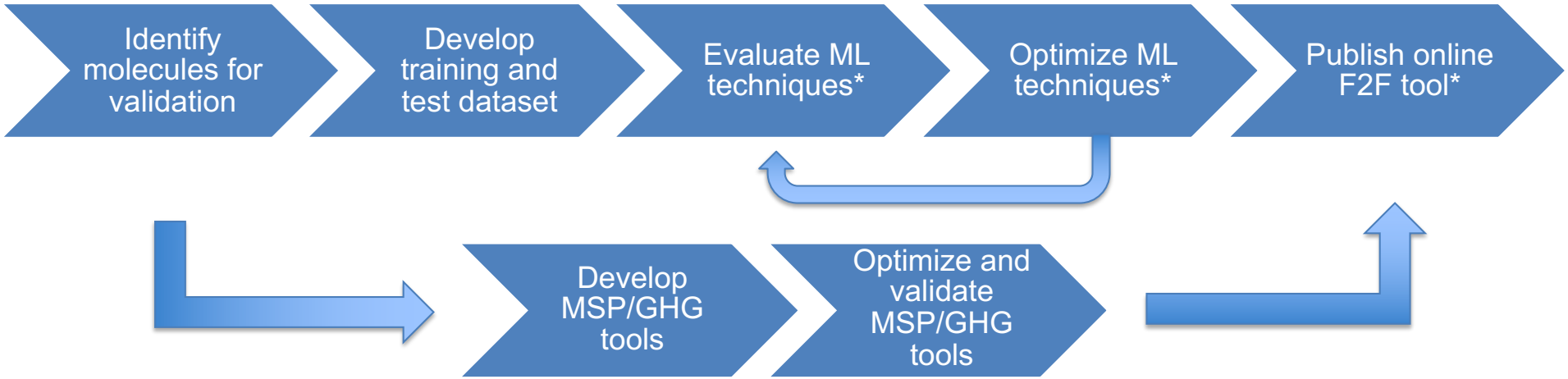
2 – Technical Approach



- **Develop dataset for machine learning (ML) based property prediction *[LBNL]***
  - Leverage 400+ molecules and blends from Co-Optima database, NIST, and literature
  - Include required properties from ASTM D7566 certified alternative jet fuels and validate properties with industry
- **Develop and optimize framework for property prediction *[LBNL & OSU]***
  - Automated machine learning tools in Python (e.g. TPOT, `scikit.learn`) to evaluate and rank supervised ML techniques for property prediction of neat molecules
  - Optimize algorithm(s) to improve predictive performance and inform Go/No-Go decision

*indicates joint LBNL/OSU effort

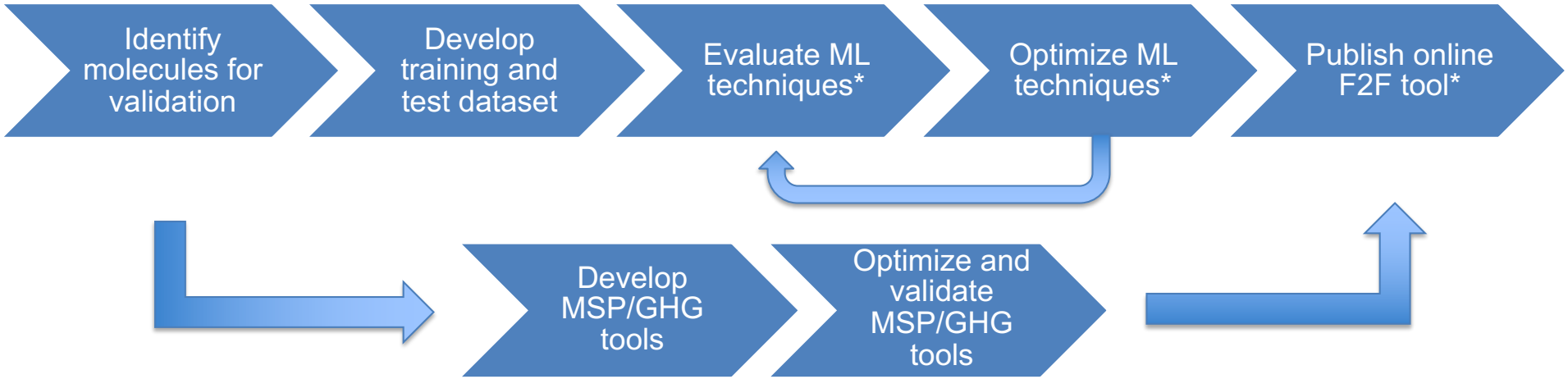
2 – Technical Approach (continued)



- **Develop minimum selling price (MSP) and greenhouse gas emissions (GHG) estimation tools *[LBNL]***
  - Develop a lightweight technoeconomic (TEA) and life-cycle assessment (LCA) model to estimate minimum selling price and life-cycle greenhouse gas emissions
  - Include potential tradeoffs between production-phase impacts and use-phase impacts
- **Publish online F2F webtool *[LBNL & OSU]***
  - Output predicted (using identified ML techniques) and measured neat and blend properties, cost analysis, and GHG emissions with sensitivity analysis

*indicates joint LBNL/OSU effort

2 – Technical Approach (continued)



- **Metrics of Success**
  - Demonstrate predicted properties (neat and blends) are within 15% of published values
  - Publish online F2F tool that predicts properties for potential alternative jet fuel molecules
- **Challenges**
  - Identifying and validating algorithm(s) to ensure accuracy, reduce computational resource demand, and reduce risk to schedule and scope
  - Collecting an abundance of reliable and accurate property data to accurately train tool
  - Validating lightweight TEA methods can predict mass/energy balances and costs

*indicates joint LBNL/OSU effort

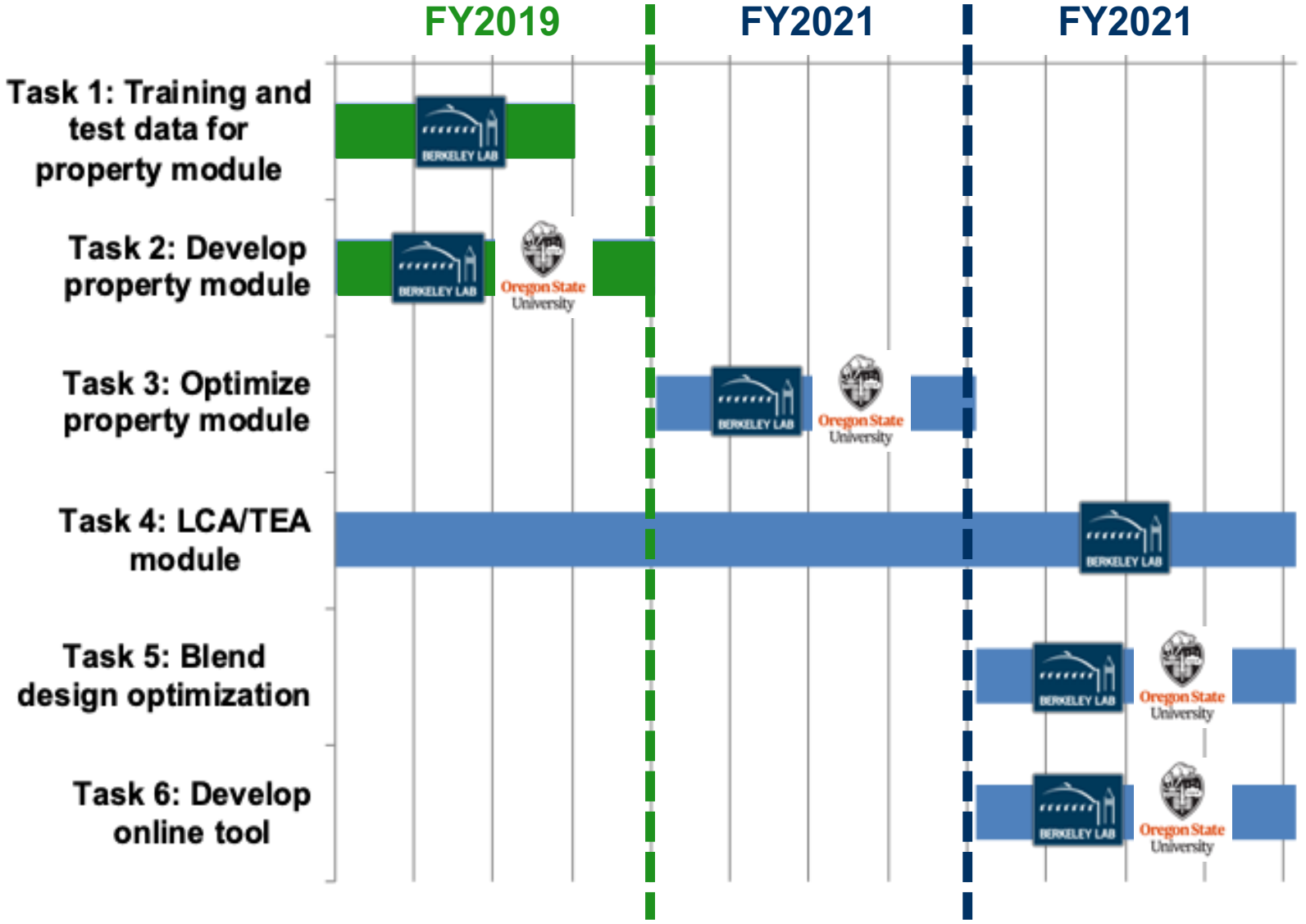
2 – Management Approach (Management)

OSU Milestones/Deliverables

- Weekly project updates
- Quarterly progress reports
- Software/source code development in support of the tasks

Other Potential Partnerships

- NREL - Co-Optima Program
- Existing efforts for jet program: Sandia, NREL, PNNL, Dayton University, Georgia Tech



3 – Technical Progress

TASKS AND MILESTONE	FY 2019 Q1	Q2	Q3	Q4	FY 2020 Q1	Q2	Q3	Q4
1) Training and test data • Identify molecules/properties & framework • Complete training dataset	Completed				Go/No-Go			
2) Develop property module • Create framework & identify machine learning techniques								
3) Optimize property module • Demonstrate 15% accuracy-one molecule • Demonstrate 15% accuracy-five molecules								
4) LCA/TEA module • Create framework & draft LCA/TEA tool • Demonstrate accuracy of LCA/TEA tool								

Key Accomplishments

- On-schedule for quarterly milestones
  - Database has 400+ different molecules and blends
  - Initiated evaluation of ML techniques for property prediction
  - Confirming with industry identified high-priority molecules and properties for database
- Ahead of schedule on future milestones
  - Predicted minimum selling price (MSP) and GHG emissions from limonene
  - Predicted cetane number of jet fuels
  - Created framework for online tool

3 – Technical Progress: Framework for F2F website

FEEDSTOCK TO FUNCTION

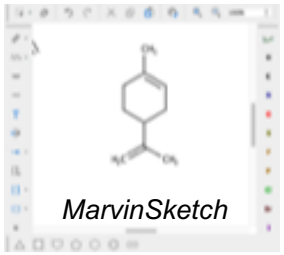
Home About Tools Data Sources Documentation References

Feedstock to Function Tool

The 'Feedstock to Function' tool harnesses the power of machine learning to predict properties of high-potential molecules derived from biomass and evaluates the cost, benefits, and risk of promising biobased molecules or biofuels to enable faster, less expensive bioprocess optimization and scale-up.

Search molecule property database using multiple platforms

Limonene



View measured and predicted properties

Names and Identifiers | Properties | Available Blends

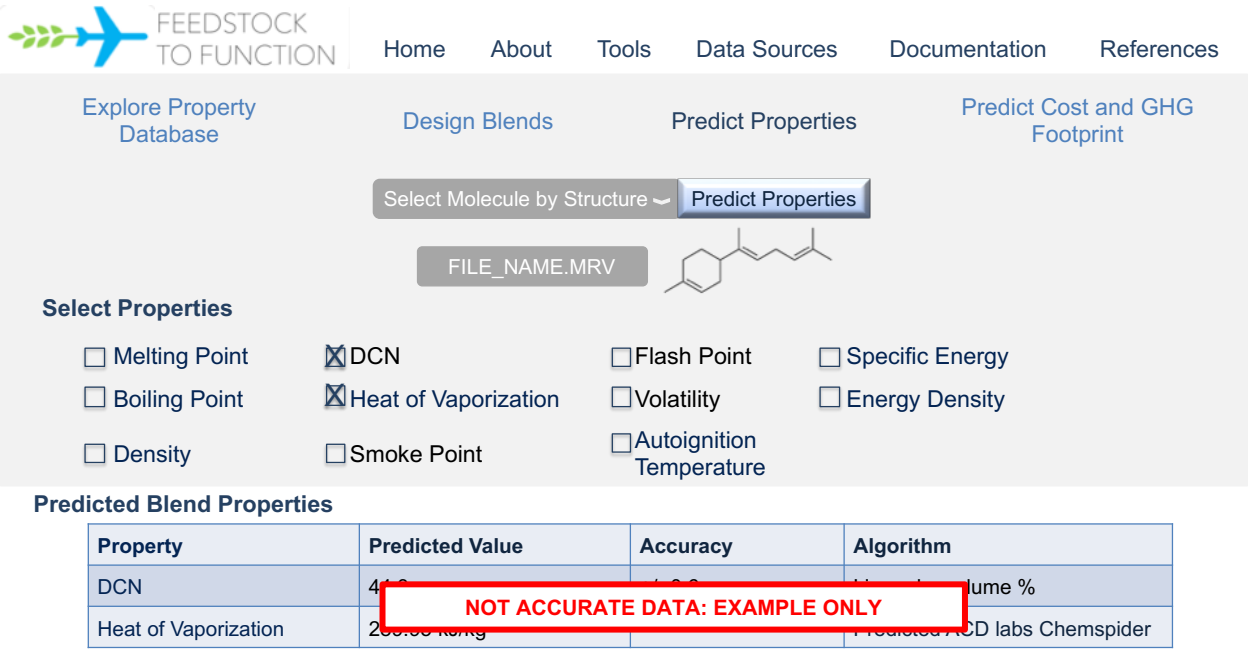
Property	Value	Source
Melting Point	-74.35 °C	Scifinder
Boiling Point	176.0 °C	Scifinder
Density	0.8402 g/cm³	Scifinder
RON	87.1	D2699, SNL
DCN	19.0	PNNL
Heat of Vaporization	289.95 kJ/kg	Predicted ACD labs Chemspider
MON	7	Predicted property using F2F on 03/2019

EXAMPLE ONLY

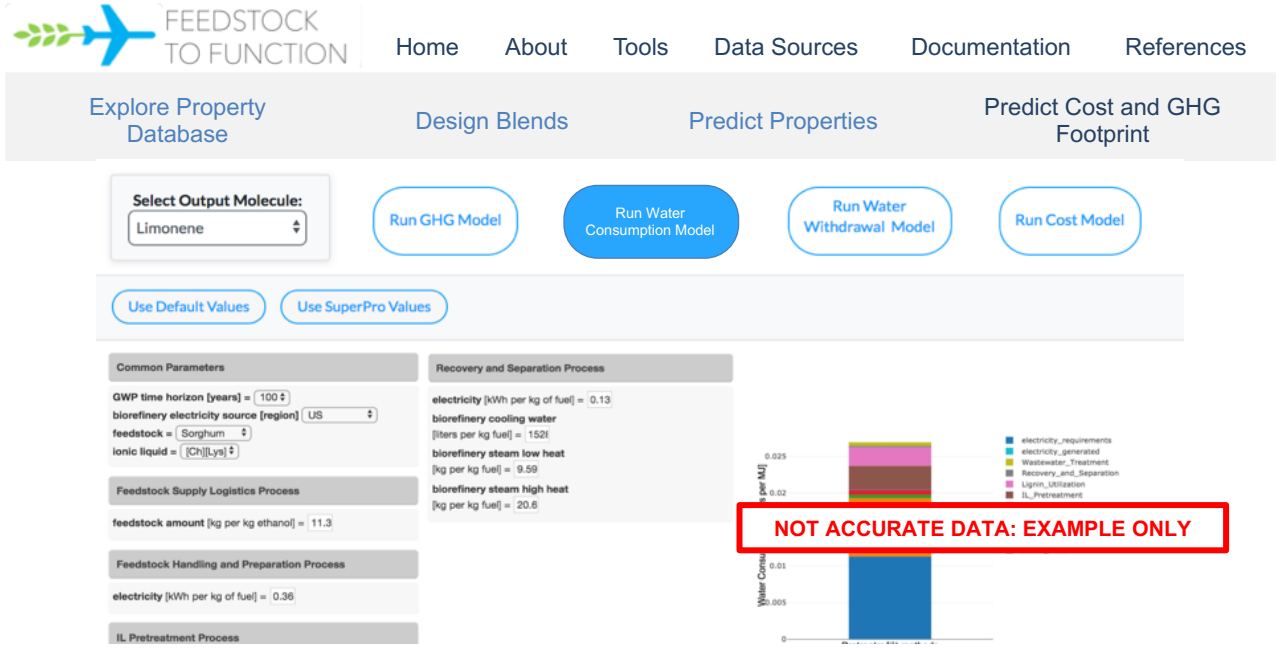
Explore Property Database | Design Blends | Predict Properties | Predict Cost and GHG Footprint

BERKELEY LAB

3 – Technical Progress: Framework for modules

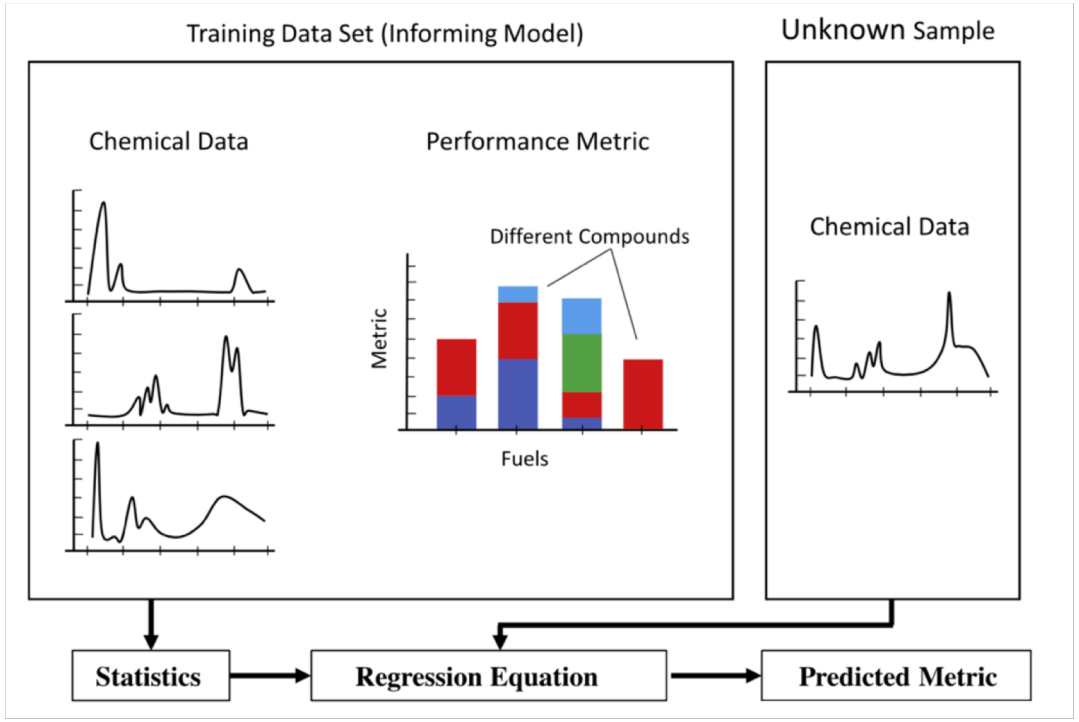


Predict Properties of New Molecules

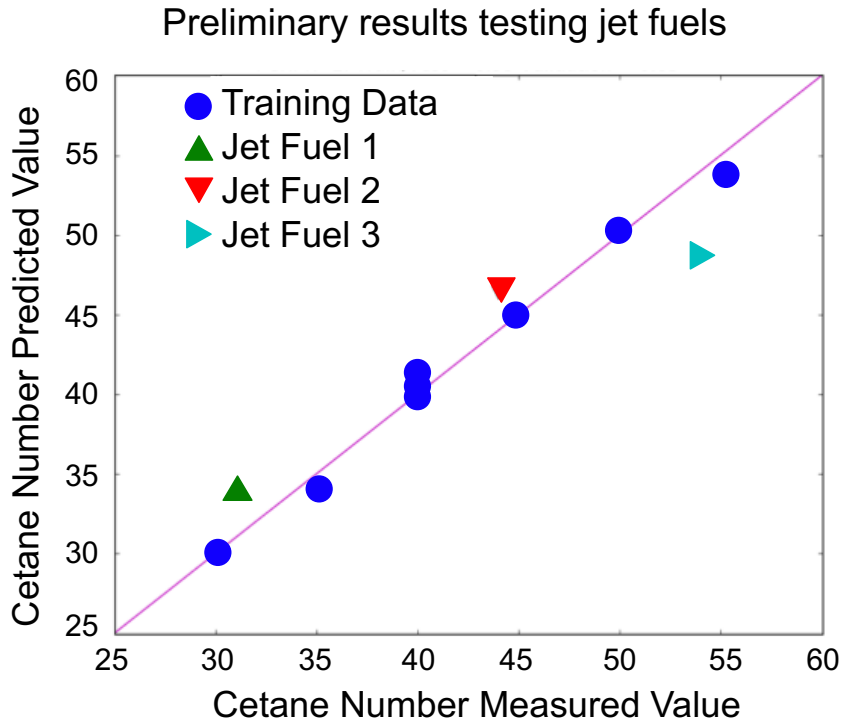


Predict Minimum Selling Price and Greenhouse Gas Emissions

3 – Technical Progress: Initial property module output

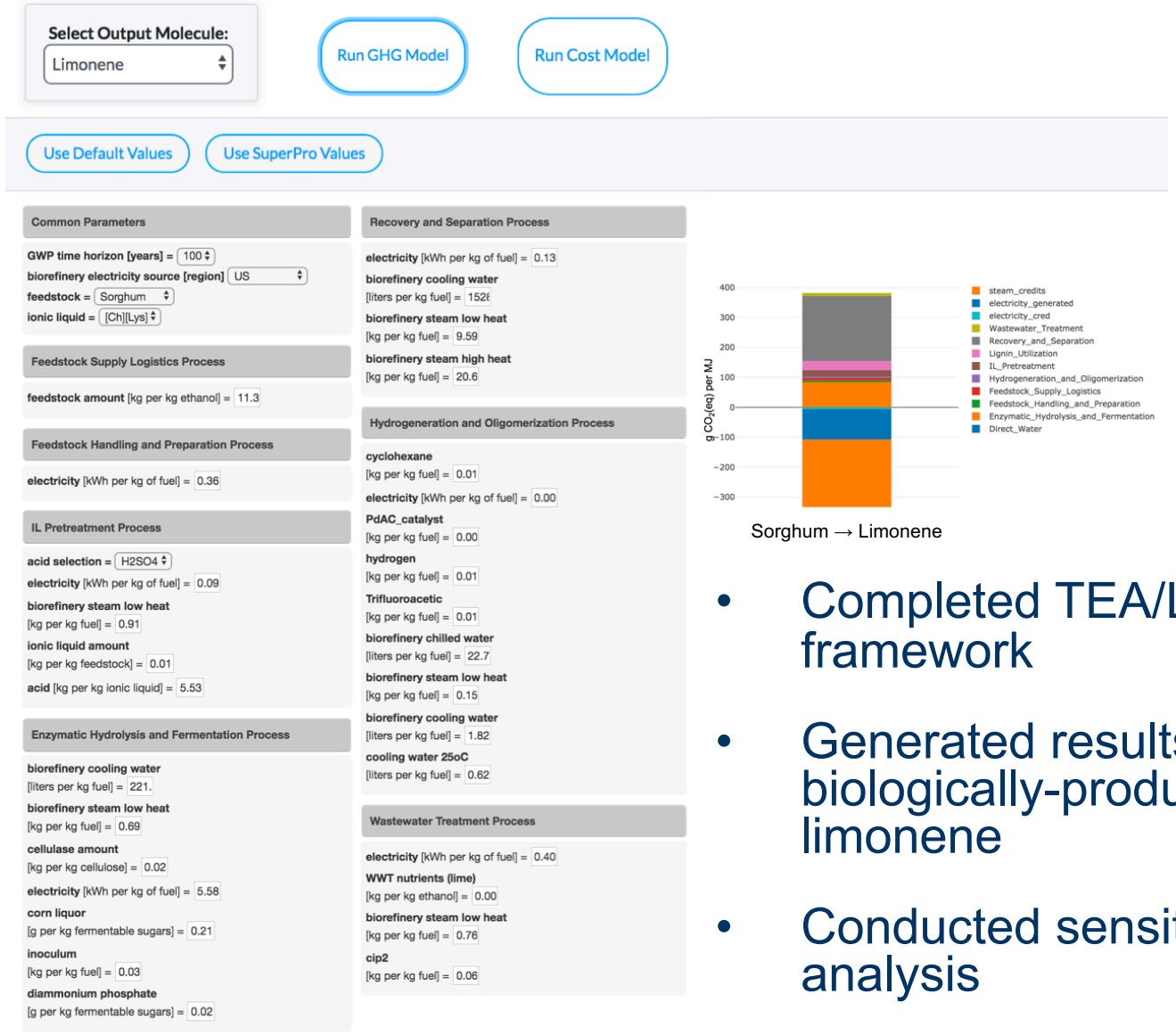


Flow chart from Daly et al. Fuel 183 (2016) 359–365



- Use Fourier-transform infrared spectroscopy (FTIR) to correlate spectra to cetane number (CN)
- Demonstrated principle component regression and support vector machine regression to correlate IR spectra to octane number and low temperature combustion index
- Initial results show good correlation with cetane number
- Results can be interpreted unlike artificial neural networks

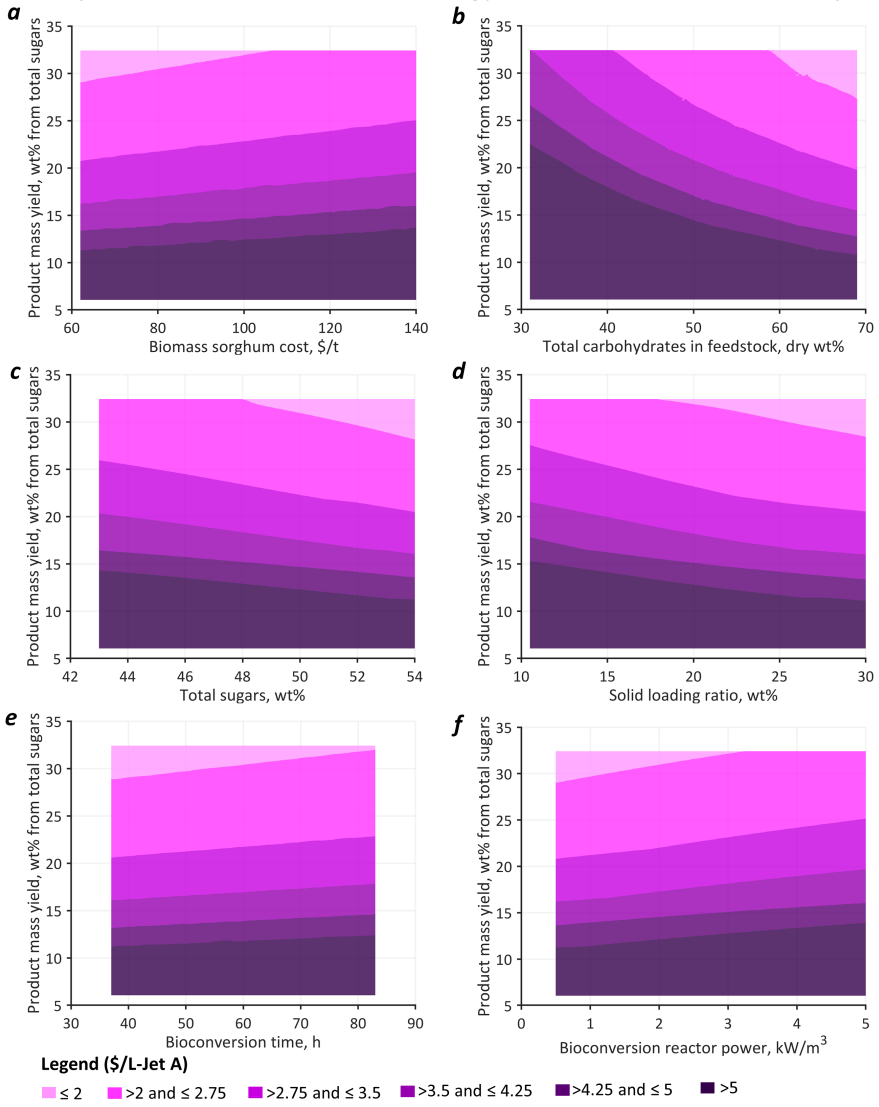
3 – Technical Progress: Initial LCA/TEA output



Sorghum → Limonene

- Completed TEA/LCA tool framework
- Generated results for biologically-produced limonene
- Conducted sensitivity analysis

Upgraded limonene cost sensitivity
(Source: Baral et al 2019 Energy & Environmental Science)



4 – Relevance

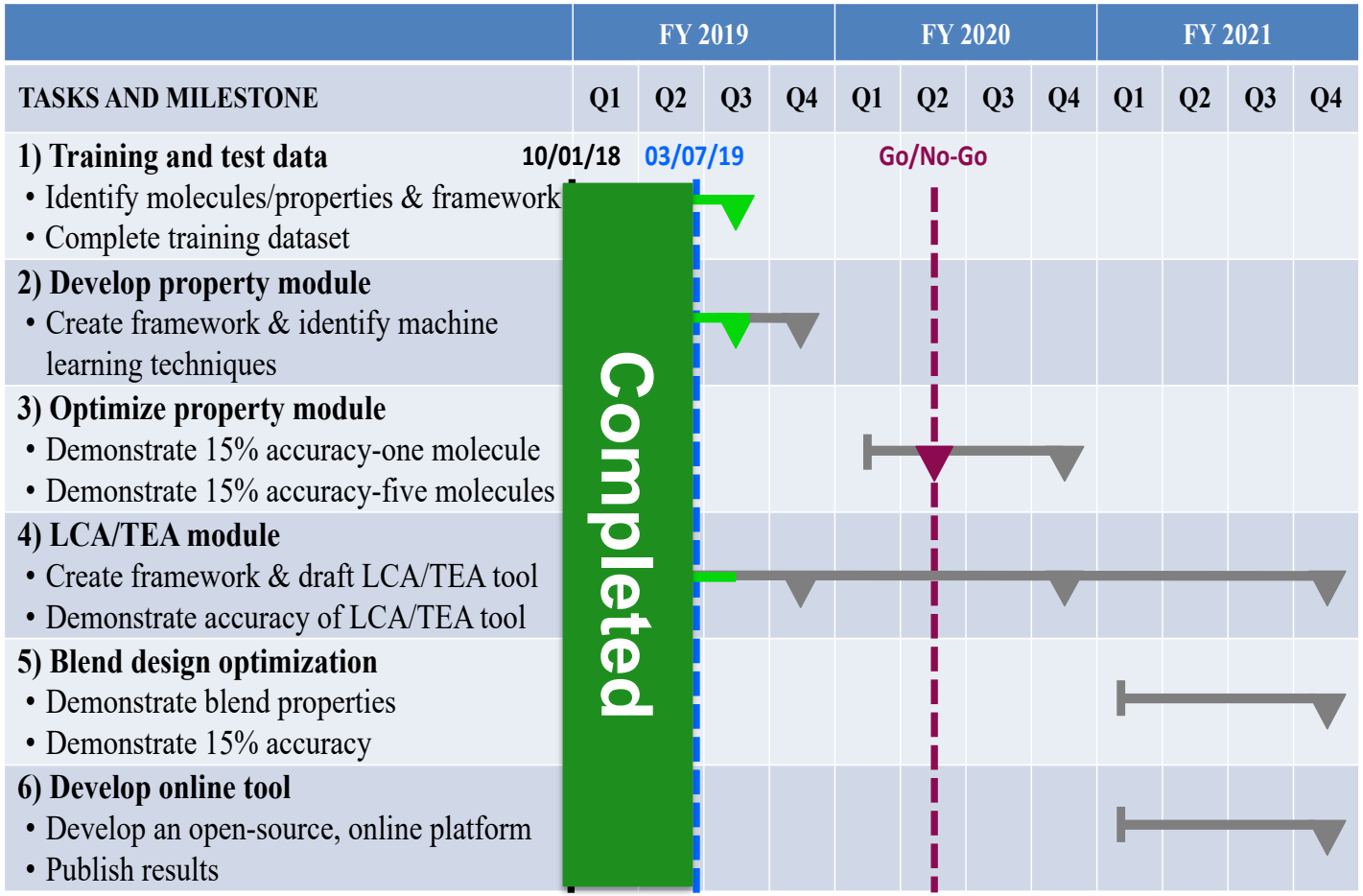
F2F is an open-source, online tool that will enable faster, less expensive bioprocess optimization, certification, and scale-up by predicting high value properties, MSP, and GHG emissions of new molecules and blends

- Supports DOE BETO Mission and Goals
  - "Develop and demonstrate transformative and revolutionary bioenergy technologies for a sustainable nation"
  - "Enable sustainable, nationwide production of biofuels that are compatible with today's transportation infrastructure, can reduce greenhouse gas emissions relative to petroleum-derived fuels, and can displace a share of petroleum-derived fuels to reduce U.S. dependence on foreign oil"
  - "Encourage the creation of a new domestic bioenergy and bioproduct industry"
- Supports industry with making progress towards DOE BETO performance goal: $3 per gallon gasoline equivalent (GGE) with GHG emissions reduction of 50% or more

4 – Relevance (continued)

- Supports ADO Mission and Goals
  - "De-risk bioenergy production technologies through validated proof of performance at the pilot, demonstration, and pioneer scales and to conduct activities that will transform the biofuels market by reducing or removing commercialization barriers."
  - Complementary to aviation fuel efforts
- Addresses Multiple BETO Barriers
  - 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
- Supports production of domestically produced biofuels that increase domestic energy security by reducing U.S. reliance on foreign oil, increase aviation resilience by introducing multiple fuel production pathways, and create jobs particularly in rural areas
- Project metrics and technical targets driven by industry targets

5 – Future Work

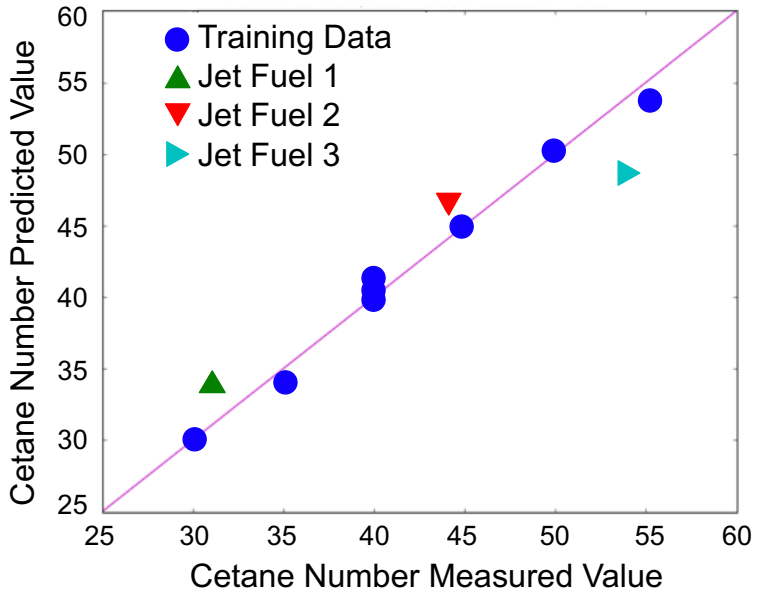


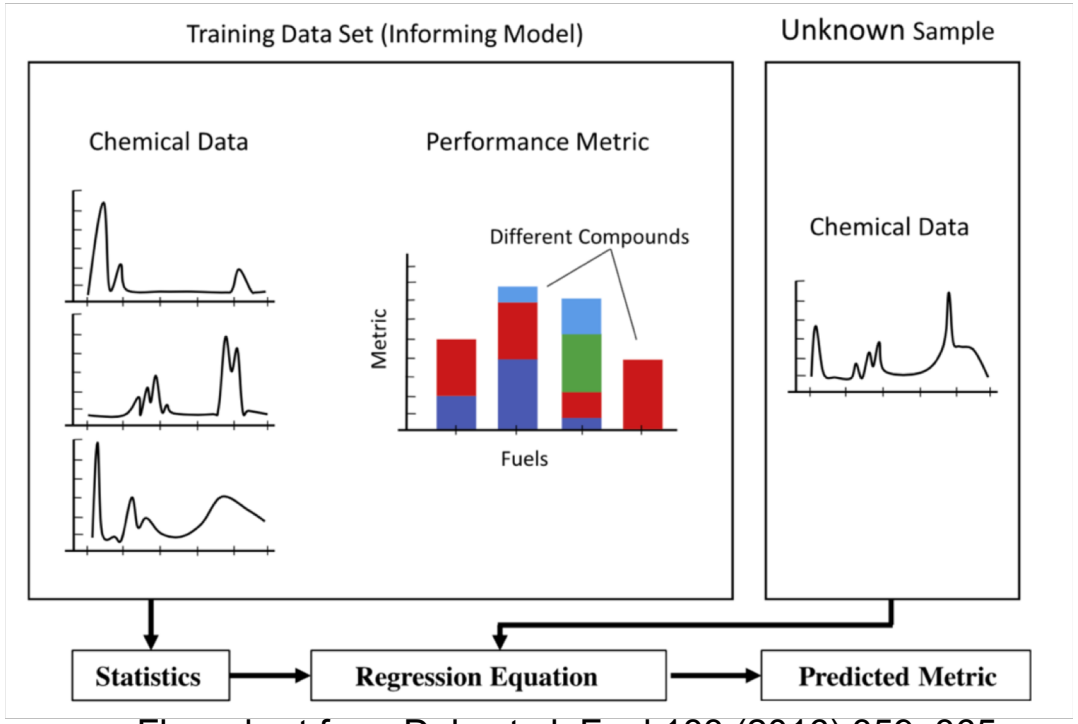
On schedule and within budget for remaining FY19 tasks

- Ahead of schedule on future milestones
  - Predicted minimum selling price (MSP) and GHG emissions from limonene
  - Precited CN of three jet fuels as proof of concept
  - Created framework for online tool
- Continue development of ML algorithms using TPOT and `scikit.learn` to predict high priority properties
- Optimize tool to demonstrate prediction capabilities for Go/No-Go decision

5 – Future Work: Expand property module

- Expanded properties to include:
  - Alternative jet fuel certification properties
  - High priority industry identified properties

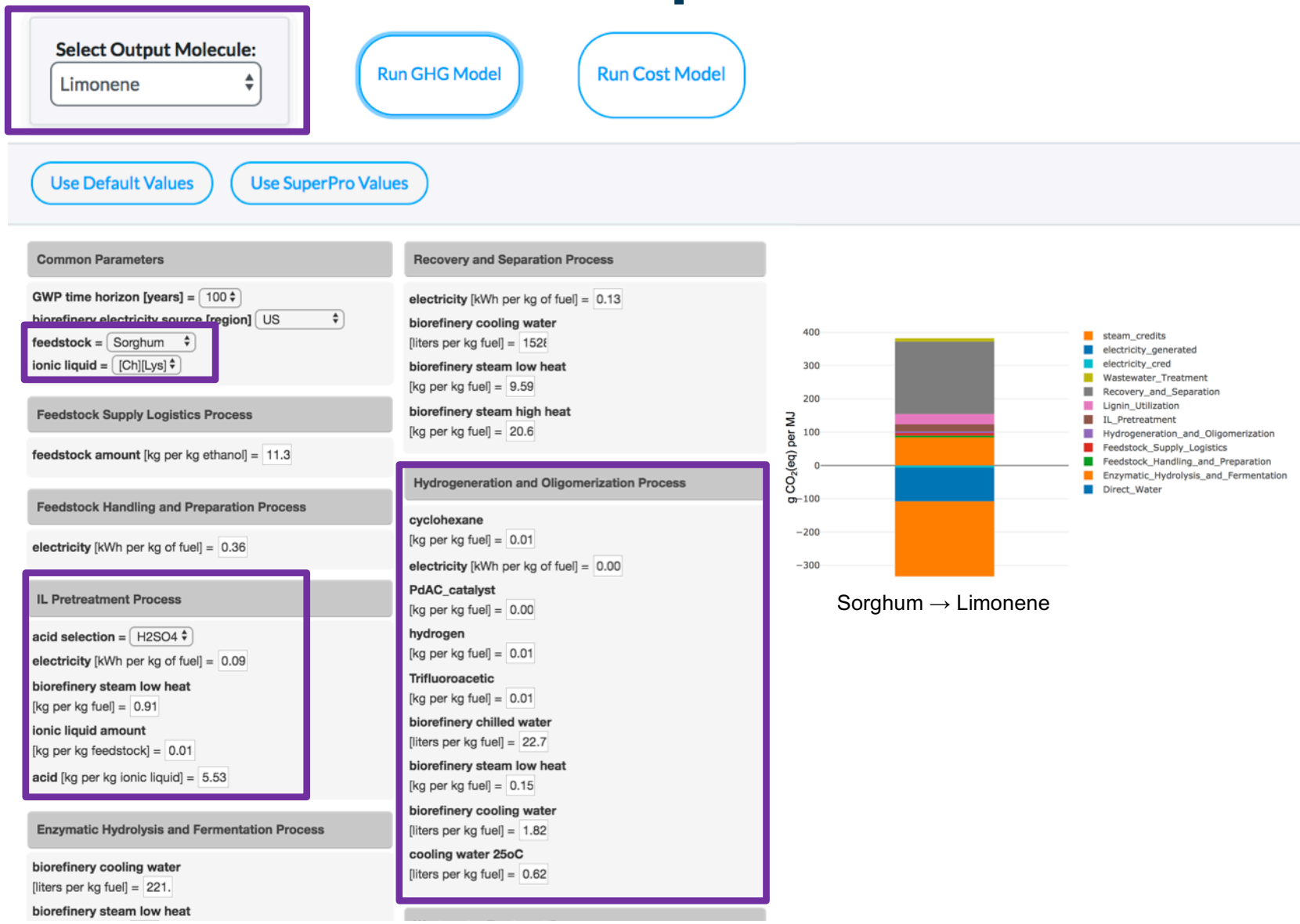




Flow chart from Daly et al. Fuel 183 (2016) 359–365

- Test and optimize using:
  - Previously ASTM certified alternative jet fuels
  - Industry molecules with sufficient experimental data

5 – Future Work: Expand and validate LCA/TEA module



Sorghum → Limonene

- Expand to include:
  - Previously ASTM certified alternative jet fuels
  - Add most relevant feedstocks for each fuel
  - At least 2-3 biomass deconstruction options
  - Relevant upgrading processes as needed
  - GHG and cost impact of use-phase property differences (e.g. density, LHV)

Summary

1. **Goal:** developing a flexible, open-source tool that predicts properties of high-potential molecules (fuels, fuel co-products, and other bioproducts) derived from biomass and evaluate the cost, benefits, and risk of promising biobased molecules or biofuels to enable faster, less expensive bioprocess optimization, certification, and scale-up

2. **Approach:** Leverage machine learning (ML) to predict desired properties and LBNL expertise to develop a lightweight TEA/LCA tool. Use alternative jet pathways and other well characterized molecules to validate ML algorithms

3. **Progress:** Predicted cetane number of a few jet fuels and estimated minimum selling price (MSP) and GHG emissions from limonene; developed framework for online tool; initiated database with 400+ different molecules and blends; initiated evaluation of ML techniques for property prediction; and confirming with industry high-priority molecules and properties for database.

4. **Relevance:** Aligns with DOE BETO goals and addresses multiple barriers (ADO-C, Ct-J, and Ct-N)

5. **Future work:** Expand, optimize, and validate property and LCA/TEA modules

Project Team

Lawrence Berkeley National Laboratory


Vi Rapp


Corinne Scown


Tyler Huntington


Ana Comesana

Oregon State University


Kyle Niemeyer


Bryony DuPont


Morgan Mayer

ADDITIONAL SLIDES

Risk Registry Table

	Risk Identified			Mitigation Strategy		Current Status
Risk ID	**Process Step**	**Risk Description**	**Severity (High/ Med/Low)**	**Mitigation Response**	**Planned Action Date**	**Active/ Closed**
Training Dataset Properties						
1	FY19Q3	Properties for training dataset are incomplete, contain multiple different entries, or molecules are too few to accurately train tool	Low	Select most common value (median) or average values for multiple entries; for missing entries, we will leave them open for the tool to predict or replace property with alternative desired property	6/30/2019	Active
Machine learning algorithm						
2	FY19Q4	Machine learning algorithm unable to predict properties to within 15%	Med	Will minimize risk by considering a wide breadth of algorithms and through optimization to achieve desired accuracy	9/30/2019	Active
High computational resources						
3	FY19Q4	Computational needs exceed resources	Low	Will optimize tool to reduce computational time for predicting properties and seek out additional computing resources	9/30/2019	Active
TEA Algorithm						
4	FY21Q4	Lightweight TEA methods prove unreliable in predicting mass/energy balances and costs	Low	Will use SuperPro Designer to provide reliable cost estimates and isolate key sources of non-linear effects that can be accounted for in other ways	9/30/2021	Active

Example: View Existing Database Properties

Home About Tools Data Sources Documentation References

Explore Property Database | Design Blends | Predict Properties | Predict Cost and GHG Footprint

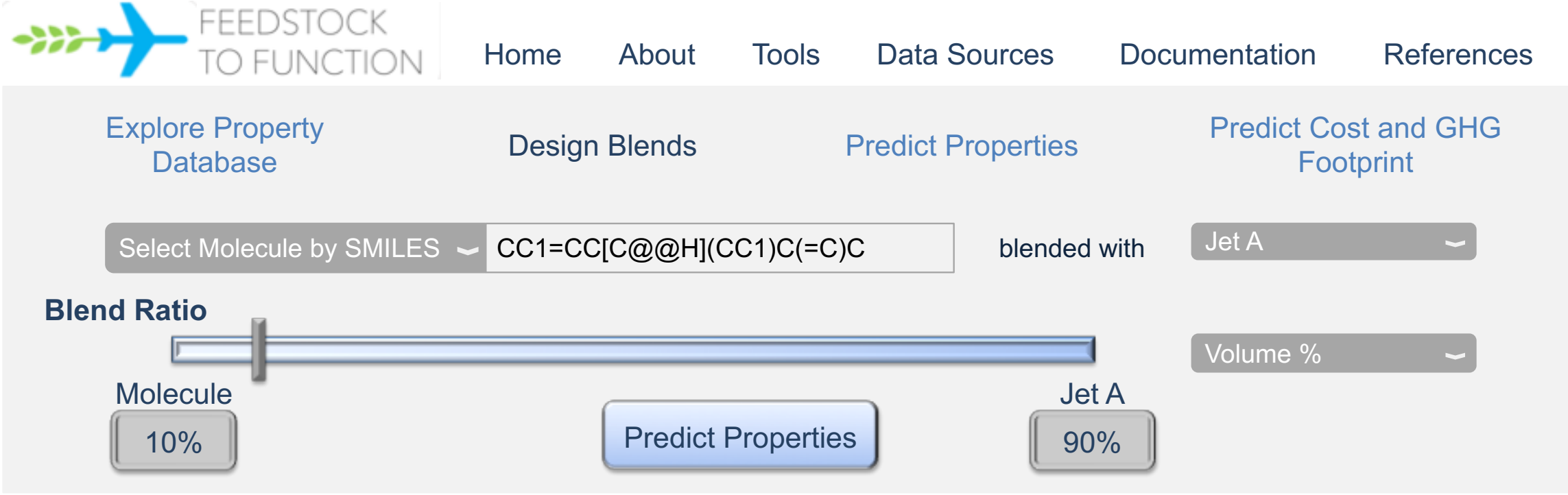
by SMILES | CC1=CC[C@@H](CC1)C(=C)C | Search

Names and Identifiers | **Properties** | Available Blends

Property	Value	Source
Mel	74.35 °C	Scifinder
Boiling Point		Scifinder
Density	0.8402 g/cm	Scifinder
RON	87.1	699, SNL
DCN	19.0	
Heat of Vaporization	289.95 kJ/kg	Predicted ACD
MON	76	Predicted property using F2F on 019

NOT ACCURATE DATA: EXAMPLE ONLY

Example: Predict properties of blends



Predicted Blend Properties

Property	Predicted Value	Accuracy	Algorithm
Boiling Point			
Density*	0.8402 g/cm³		
DCN	44.3	+/- 0.6	volume %
Heat of Vaporization	289.95 kJ/kg		D labs Chemspider

NOT ACCURATE DATA: EXAMPLE ONLY

* Predicted without measured neat properties