SAND2017-7182C

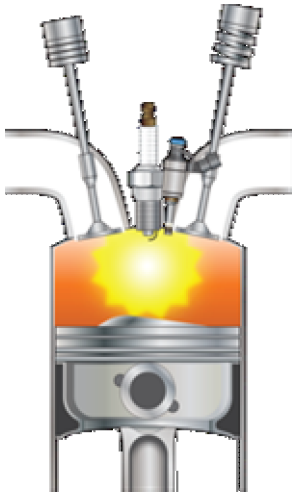
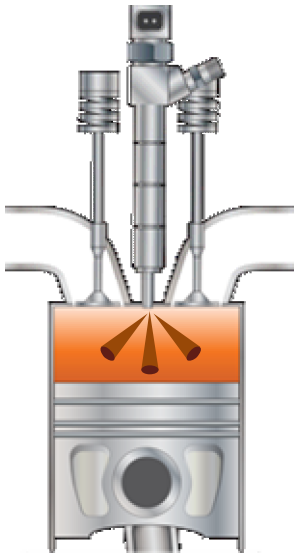
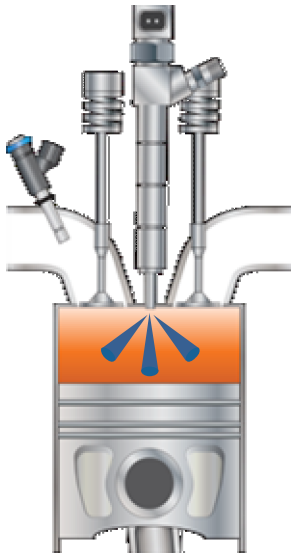
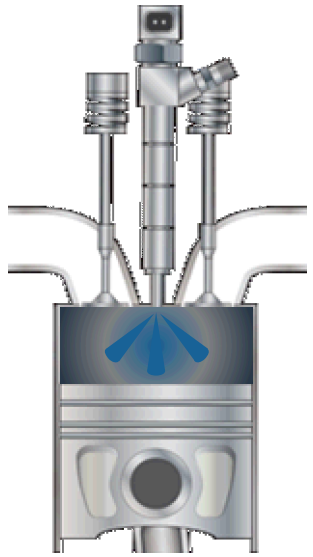
Developing a high- and low-cetane classifier for biologically produced chemicals using variable quality training data

Leanne S. Whitmore & Corey M. Hudson
Sandia National Laboratory
4/6/2017
ACS National Meeting SF/CA

- Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.

Modern Engine Types

Engine type and technology has an important relationship to the desired **fuel properties.**

Spark Ignition (SI)	Advanced Compression Ignition (ACI) kinetically-controlled and compression-ignition combustion	
	 	
Low reactivity fuel	Range of fuel properties – Principally Compression Controlled	High reactivity fuel

Co-Optima 2017

Cetane as a fundamental fuel property in diesel-like engines

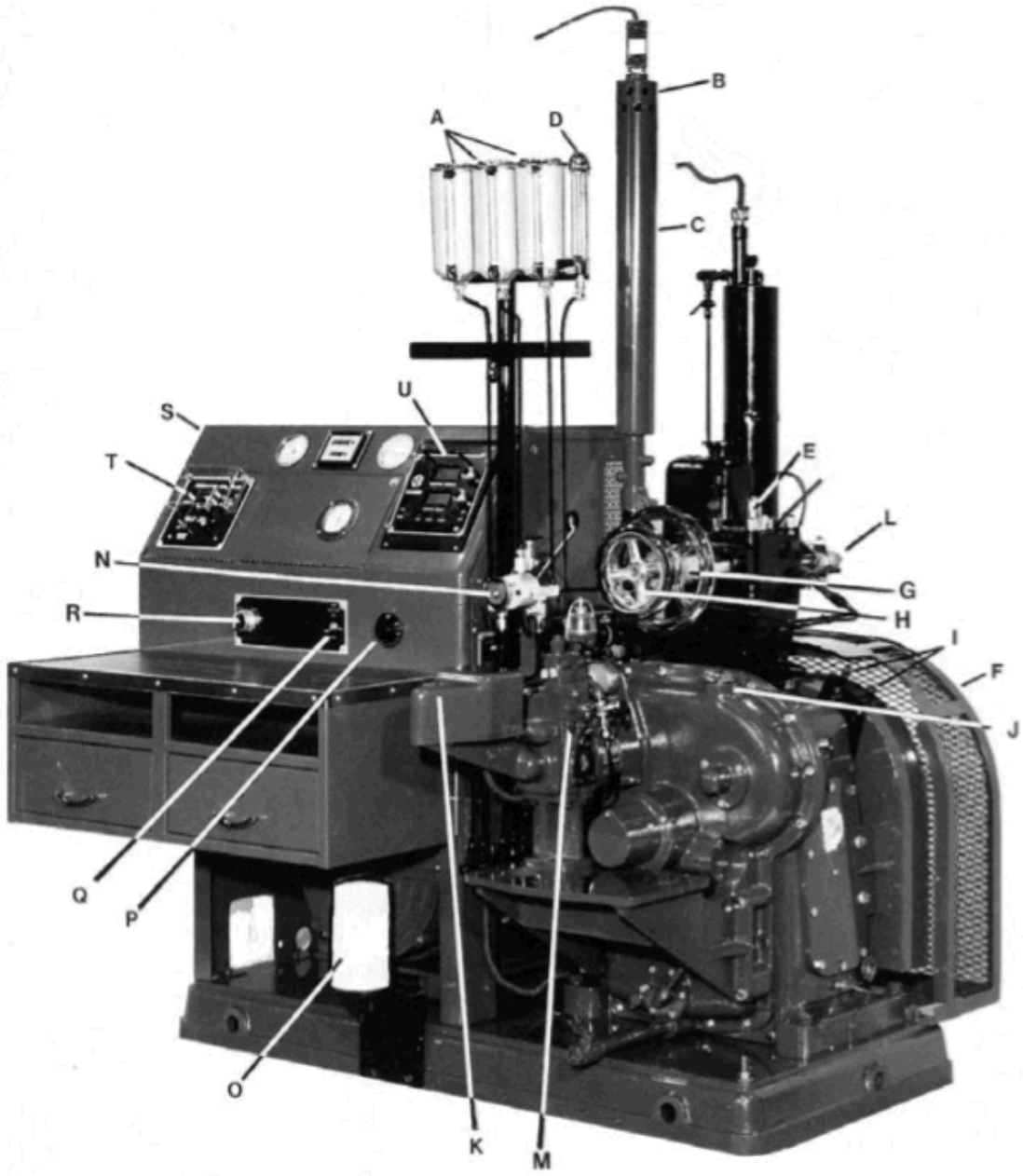
- Cetane is a parameter measuring the time delay between fuel injection and combustion.
- In general diesel-like fuels need to have cetane > 40 to be generally compliant with fuel property standards and for engines to run efficiently.
- Some states or countries may require even higher values (>45, 50s)

Methods for Measuring Cetane in Literature

- CFR: ISTM D613 – Cooperative Fuels Research: Gold standard for cetane measurement
- IQT: ISTM D6890 – Ignition Quality Tester: Correlative method – very high precision
- FIT: ISTM D7170 – Fuel Ignition Tester: Correlative method
- Blending Methods – Interpolative method
- Other Methods – Novel, often one-off methods
- Unknown Methods – Historical/literature based methods

CFR – ASTM D613

- Oldest method
- Gold standard
- Requires 400-500mL of pure compound



A—Fuel Tanks
B—Air Heater Housing
C—Air Intake Silencer
D—Fuel Flow Rate Buret
E—Combustion Pickup
F—Safety Guard
G—Variable Compression Plug Handwheel
H—V.C.P. Locking Handwheel
I—Flywheel Pickups
J—Oil Filler Cap
K—Injection Pump Safety Shut-Off Solenoi
L—Injector Assembly
M—Fuel Injection Pump
N—Fuel Selector-Valve
O—Oil Filter
P—Crankcase Oil Heater Control
Q—Air Heater Switch
R—Engine Start-Stop Switch
S—Instrument Panel
T—Intake Air Temperature Controller
U—Dual Digital Cetane Meter

FIG. 1 Cetane Method Test Engine Assembly

CFR – ASTM D6890

- High precision
- Large database of measurements
- Requires 100mL of pure compound

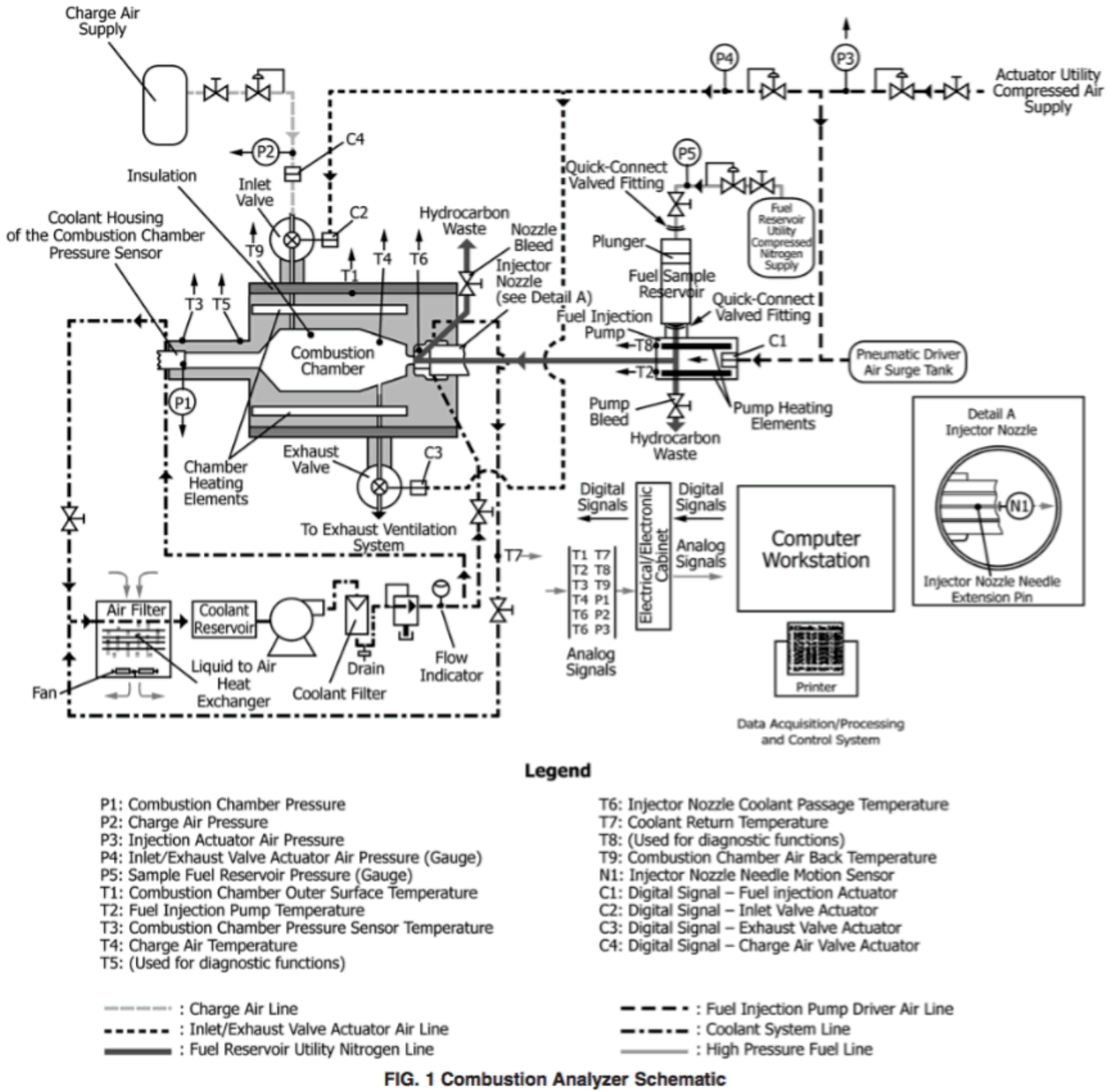
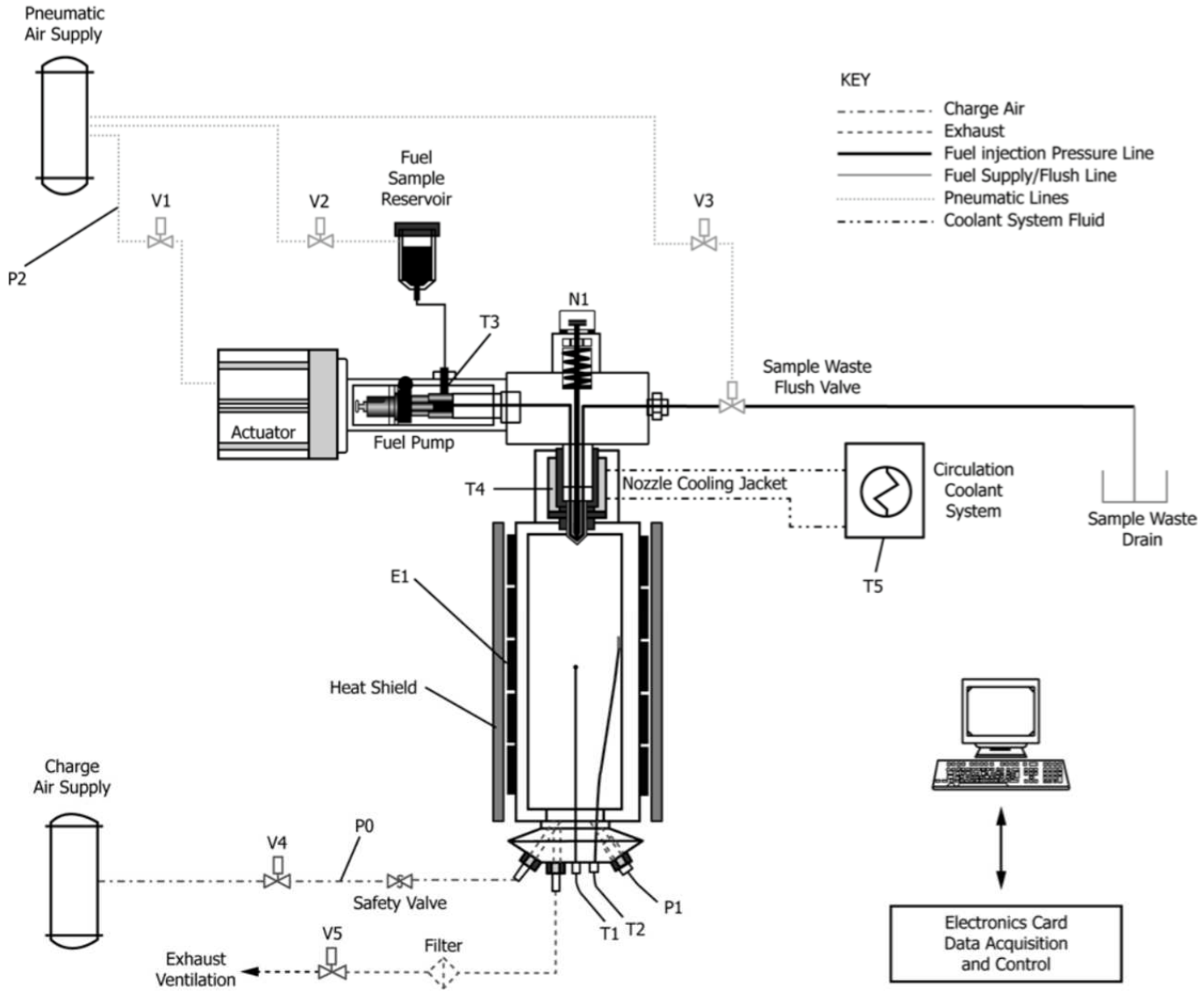


FIG. 1 Combustion Analyzer Schematic

FIT– ASTM D7170

- Few databased measurements
- 220 mL pure compound at sample time

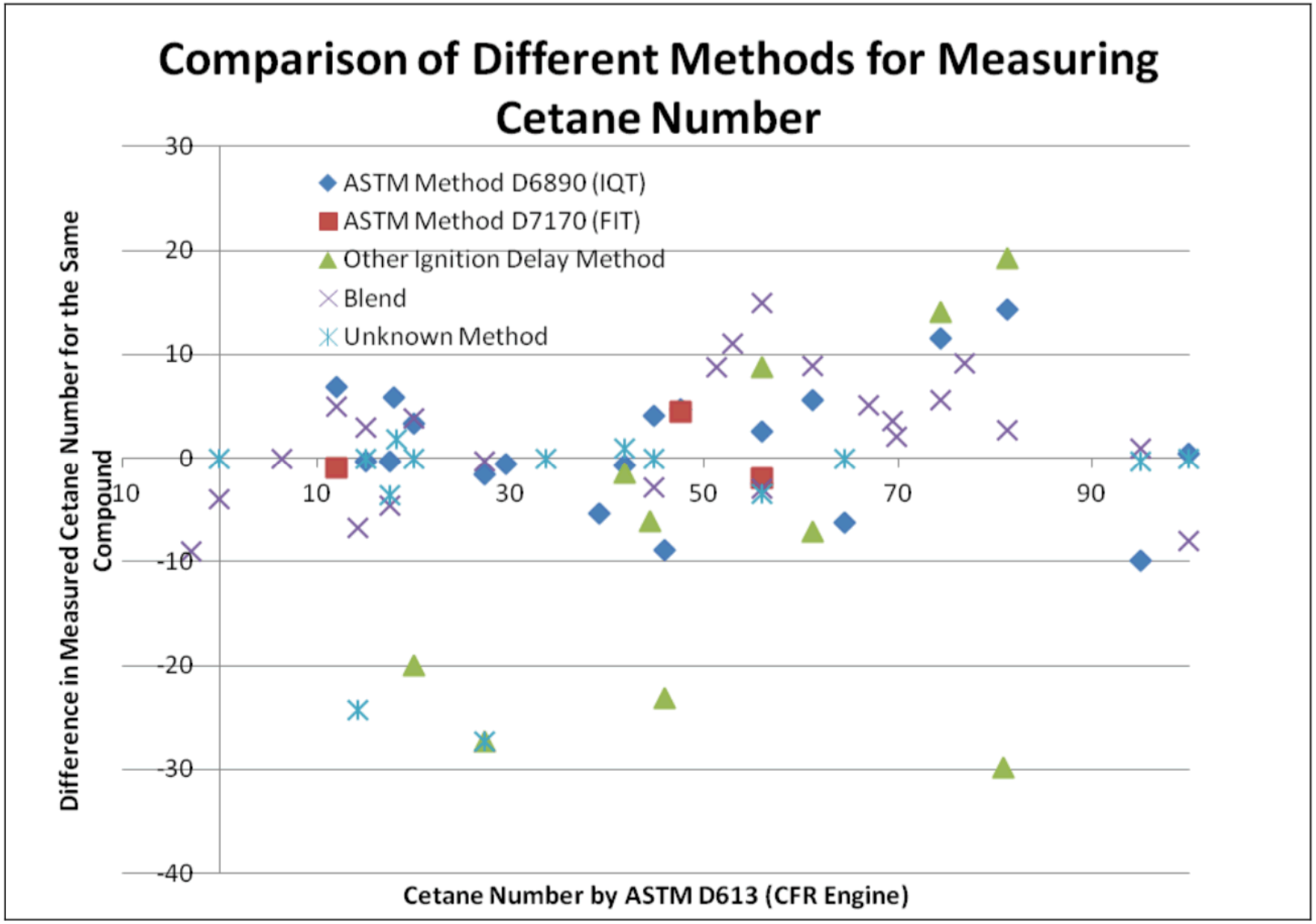


Digital Signals	Analog Signals
V1: Actuator Air Valve	T1: Chamber Charge Air Temperature
V2: Sample Fuel Reservoir Valve	T2: Chamber Inner Wall Temperature
V3: Sample Waste Flush Valve	T3: Fuel Injection Pump Temperature
V4: Charge Air Valve	T4: Injection Nozzle Cooling Jacket Temperature
V5: Exhaust Valve	T5: Circulation Coolant System Temperature (External)[A]
E1: Control Power to Chamber Heating	P0: Chamber Static Pressure Sensor
N1: Injection Nozzle Motion Sensor	P1: Chamber Dynamic Pressure Sensor
P2: Injection Actuator Air Pressure Switch Gauge	

[A] T5 is not located on the instrument. It is the temperature of the auxiliary Circulation Coolant System adjusted to maintain T4.

Non-standard methods

- Blending – interprolative: small amount of compound is blended and cetane of pure compound is interprolated: 85-135 reported measurements in the literature
- Other ignition delay methods – vary in methodology, uncertain correlations with gold standard methods: at least 70 such measurements
- Unknown methods – often older, less certain methods: 142-189 measurements in the literature


Comparison of Different Methods for Measuring Cetane Number
ASTM Method D6890 (IQT)
ASTM Method D7170 (FIT)
Other Ignition Delay Method
Blend
Unknown Method
Difference in Measured Cetane Number for the Same Compound
Cetane Number by ASTM D613 (CFR Engine)
30
20
10
0
-10
-20
-30
-40
10
10
30
50
70
90

NREL 2014

Existing database of cetane values

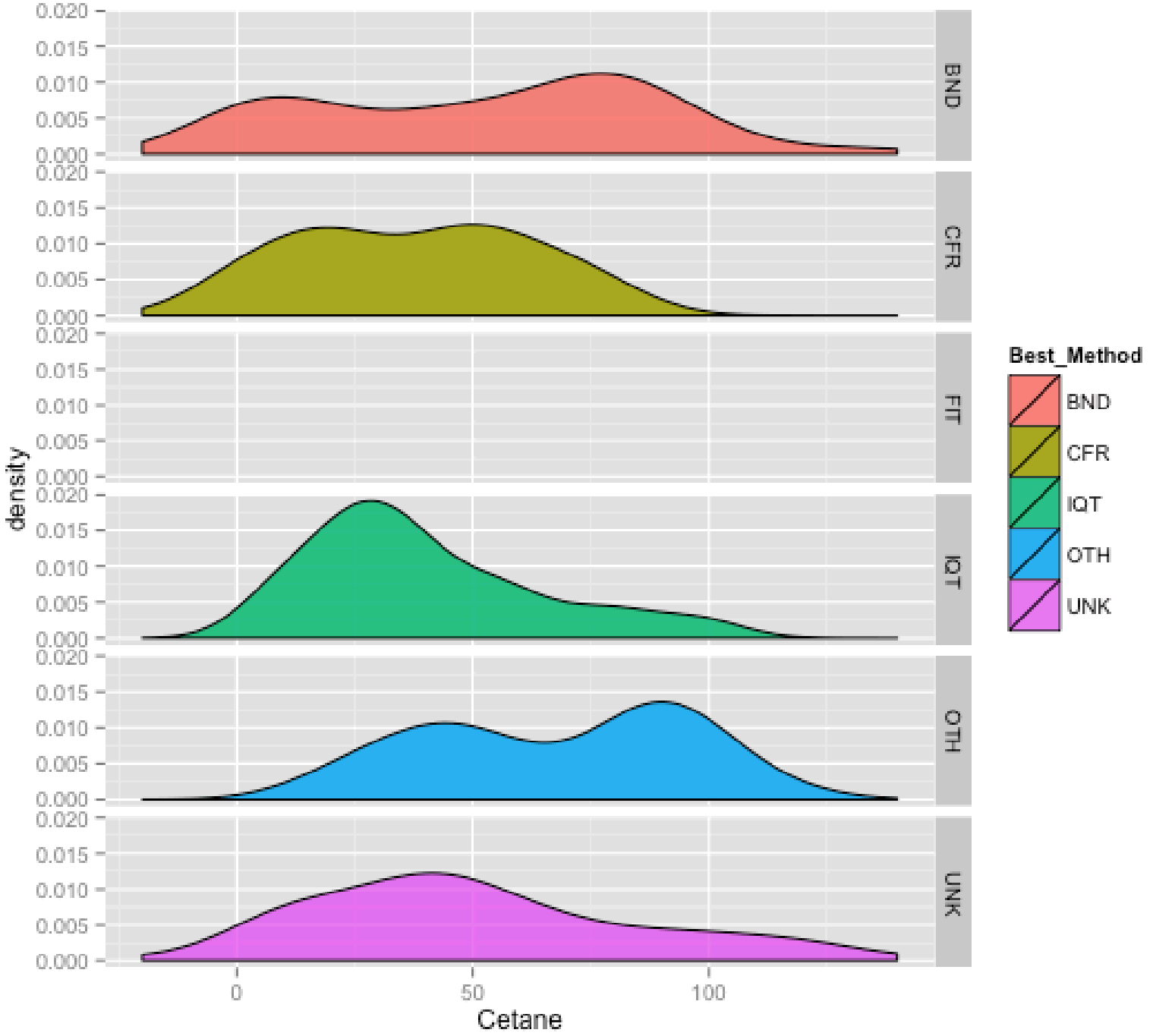
- Training database:
  - 2014 NREL Compendium of cetane values
  - 320 pure compounds with associated CAS numbers
    - Partially redundant
  - 6 methods of measurement

CFR	FIT	IQT	BLEND	OTHER	UNKNOWN
29	1	102	80	32	79

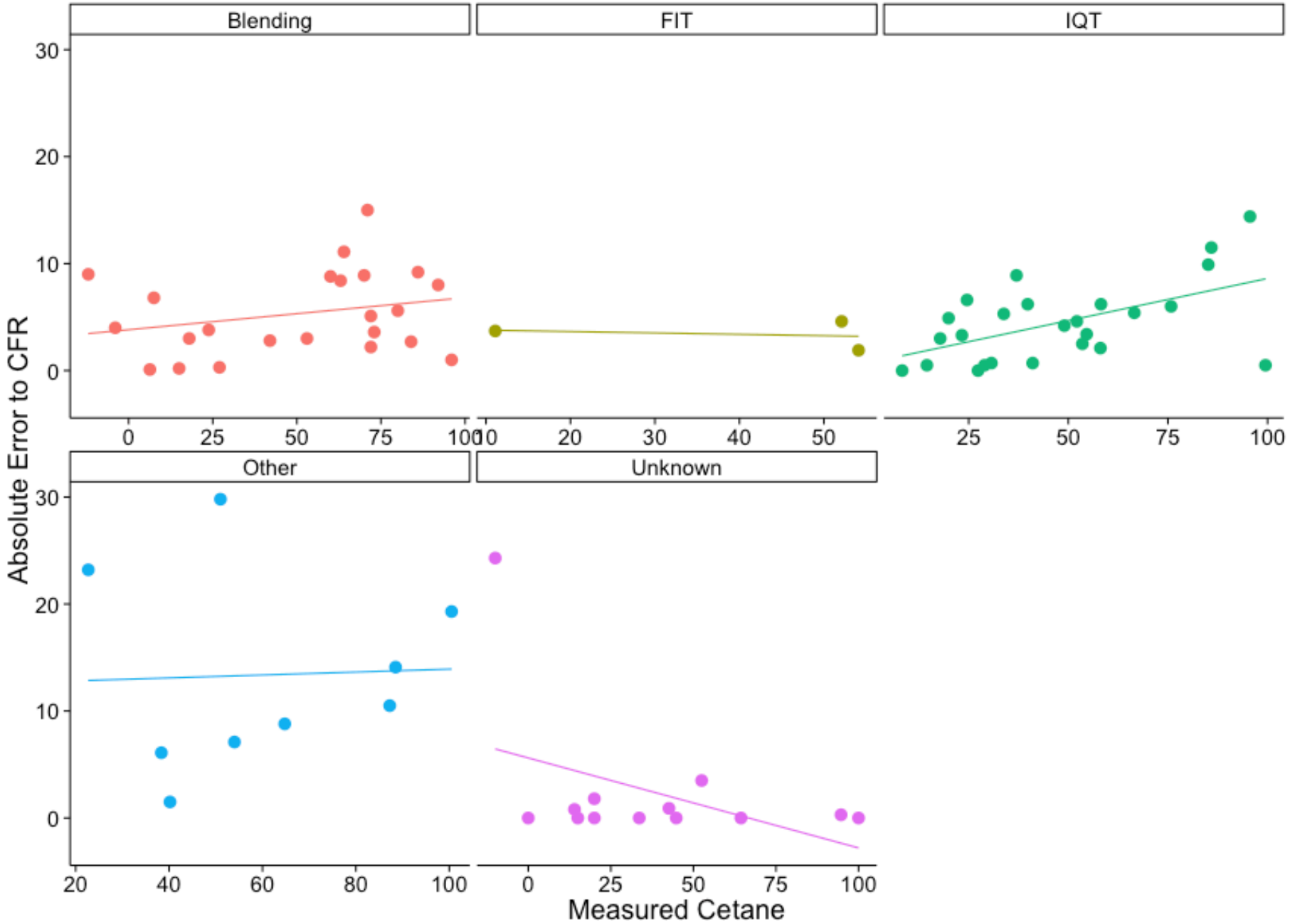
- Testing database:
  - 2017 Update of NREL Compendium of cetane values
  - 58 CFR measurements as confirmatory

Distributions of variable measures

- IQT and CFR methods reflect systematic study of the range of cetane values.
- Other ignition methods and blending methods are skewed toward high cetane



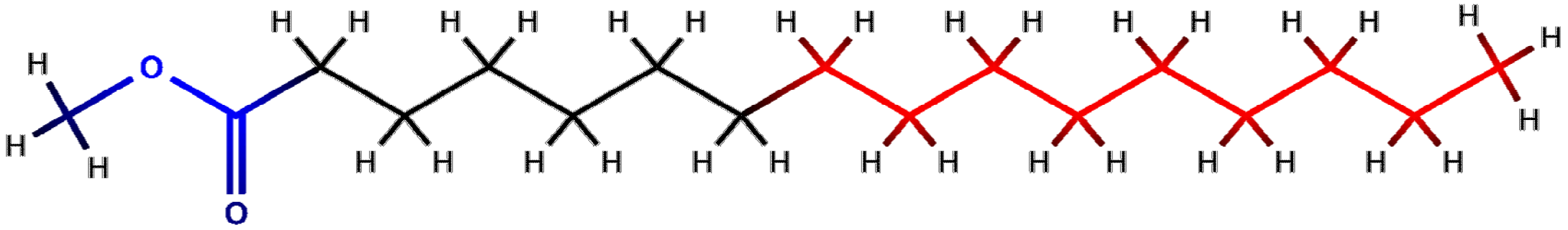
Errors by method



Exploring the compound variability

- Chemical similarity can be assessed by molecular fingerprints

'[!#1][CH2][CH2][CH2][CH2][CH2][CH2][CH2][CH3]'



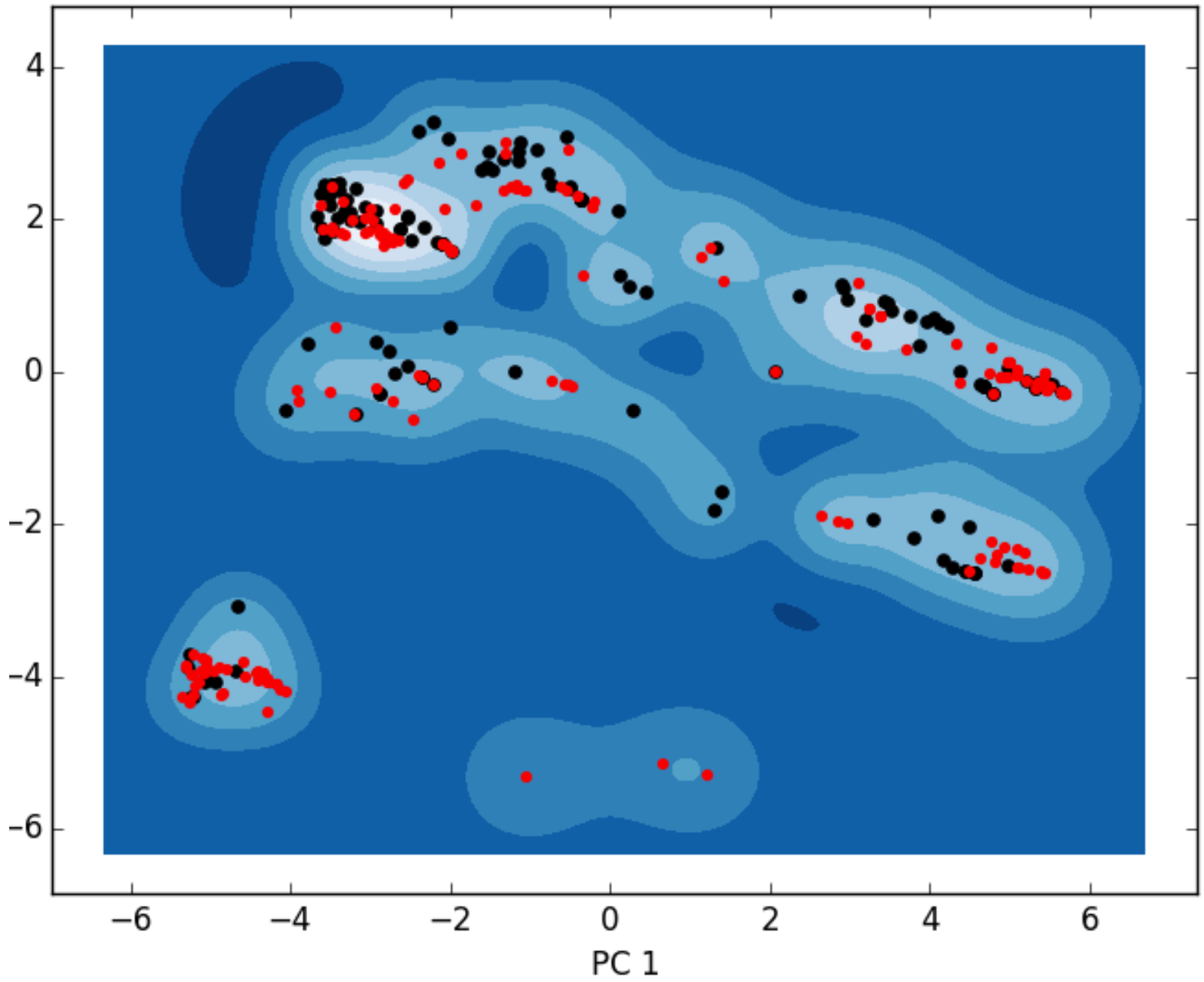
C~O

Methyl pamitate

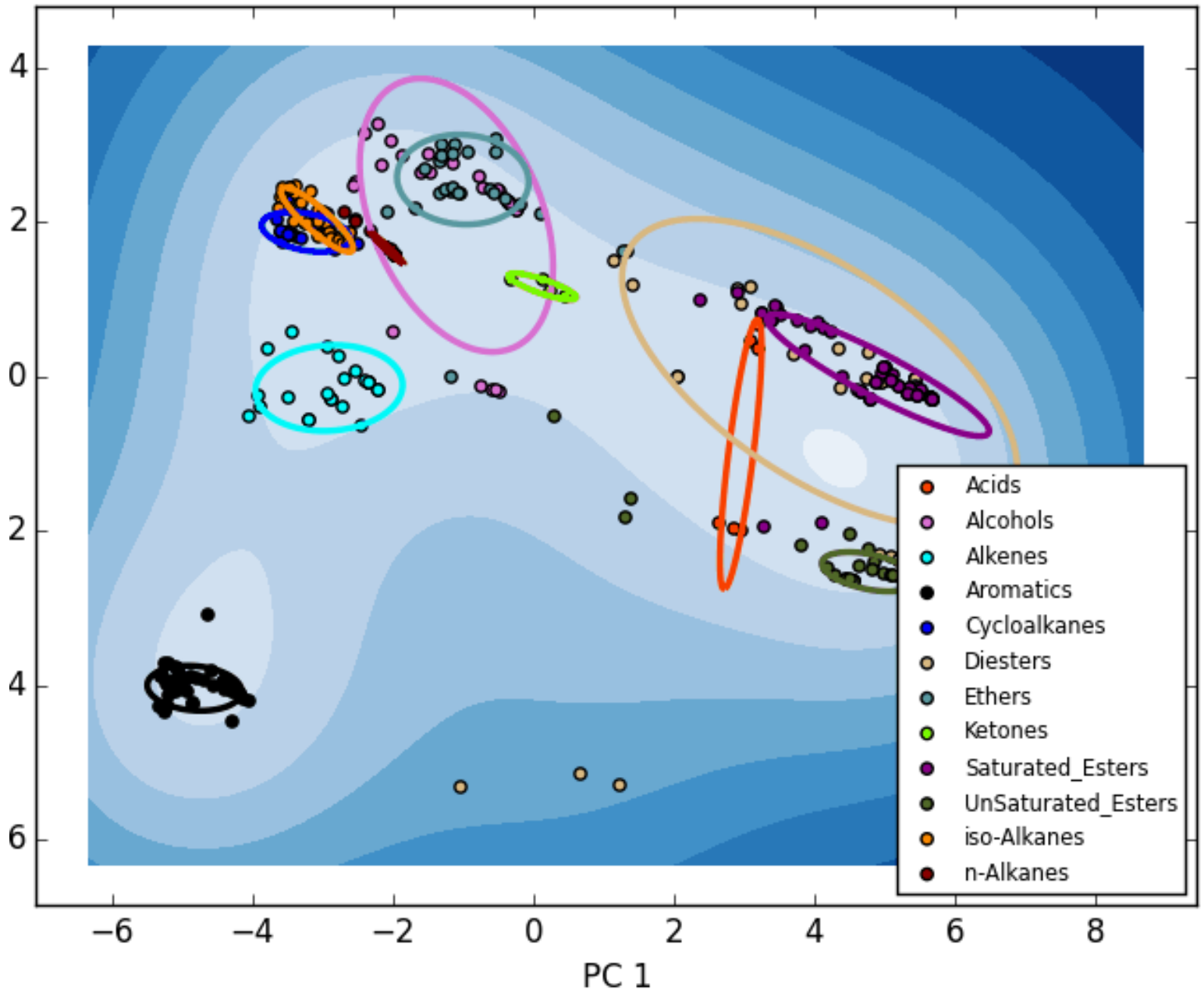
Data source

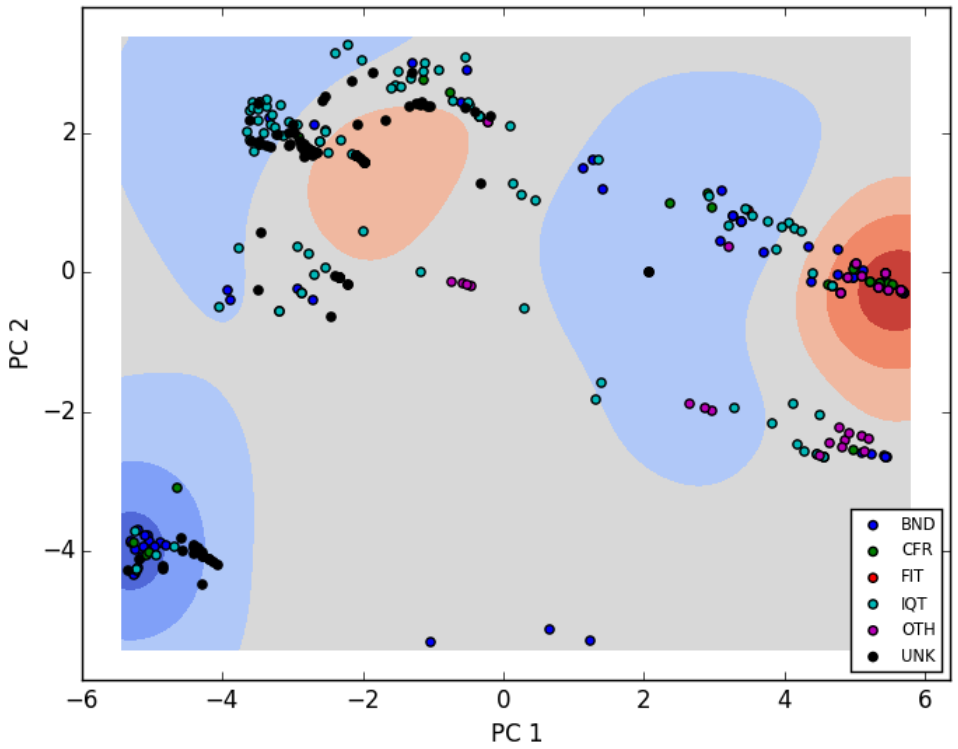
- Fingerprints: 6127
  - PubChem: 881
  - Estate: 1024
  - Klekota-Roth: 4860
  - SubStructure: 307
- Gold-standard data:
  - IQT: 101
  - FIT: 1
  - CFR: 29
- Other data:
  - Unknown, Other Ignition, Blending: 189

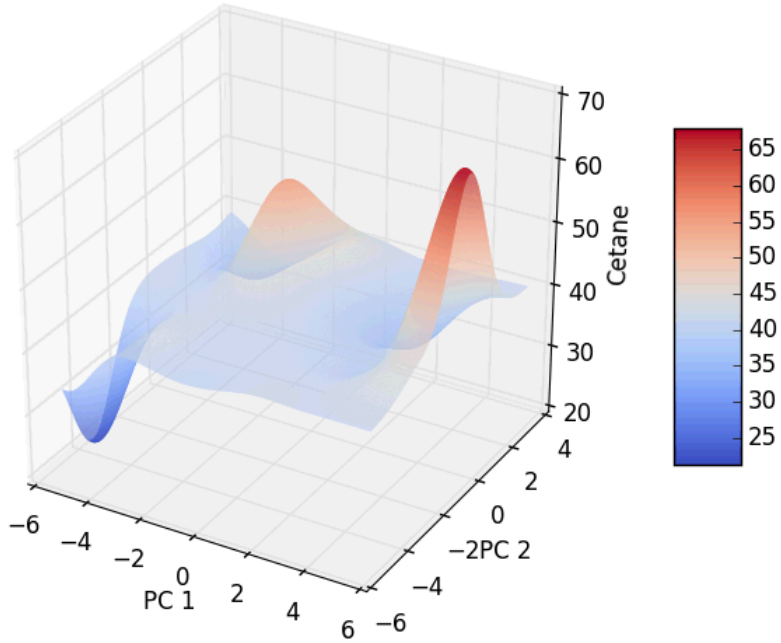
Total variability in chemical space



Total variability by class







Total variability by cetane

Comparisons of variability

features = 6127

compounds = 320

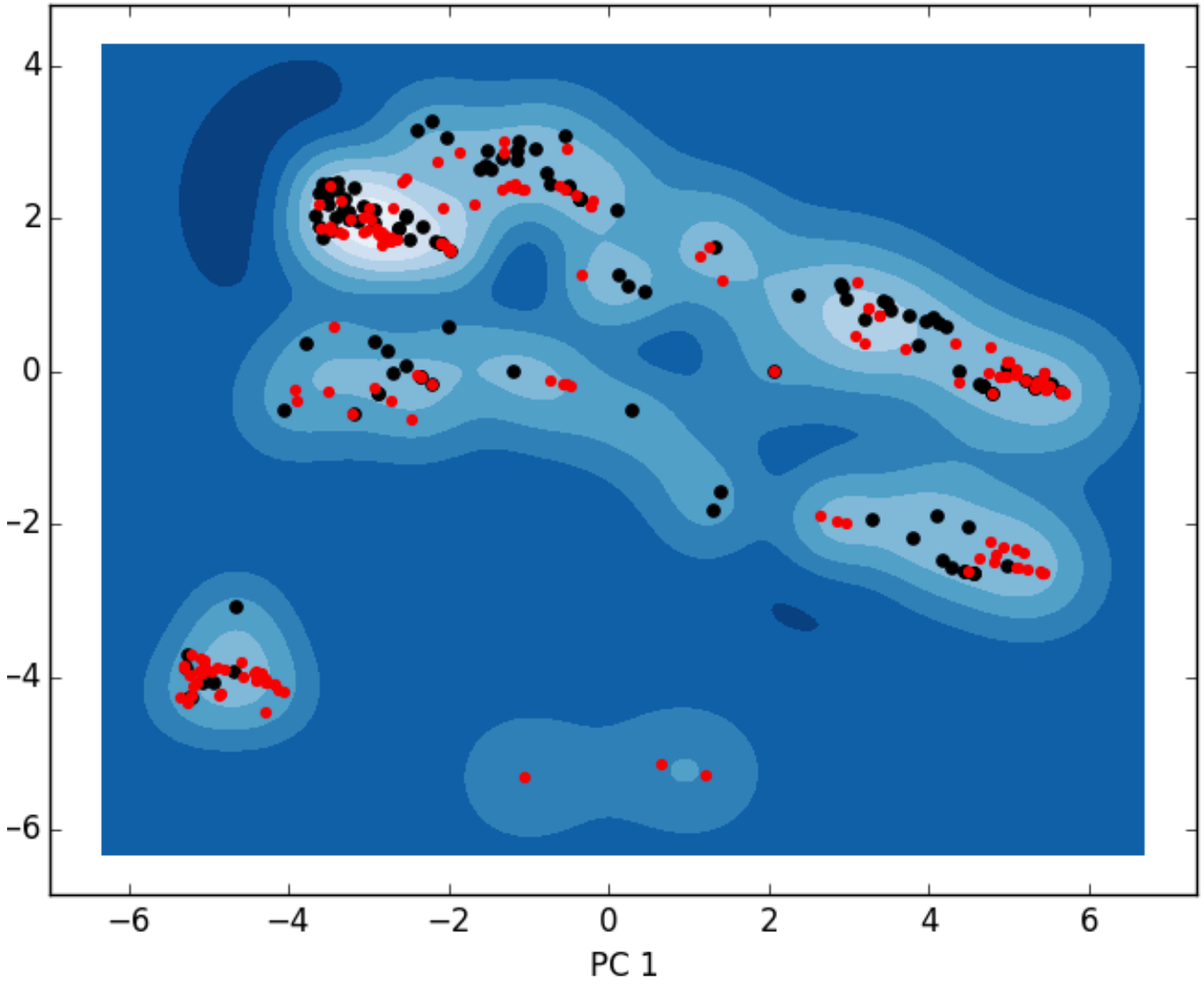
'Gold Standard' compounds = 131

Variance: $\frac{\sigma^2 \lambda_{gold}}{\sigma^2 \lambda_{total}}$

Gold Proportion: 40.9%

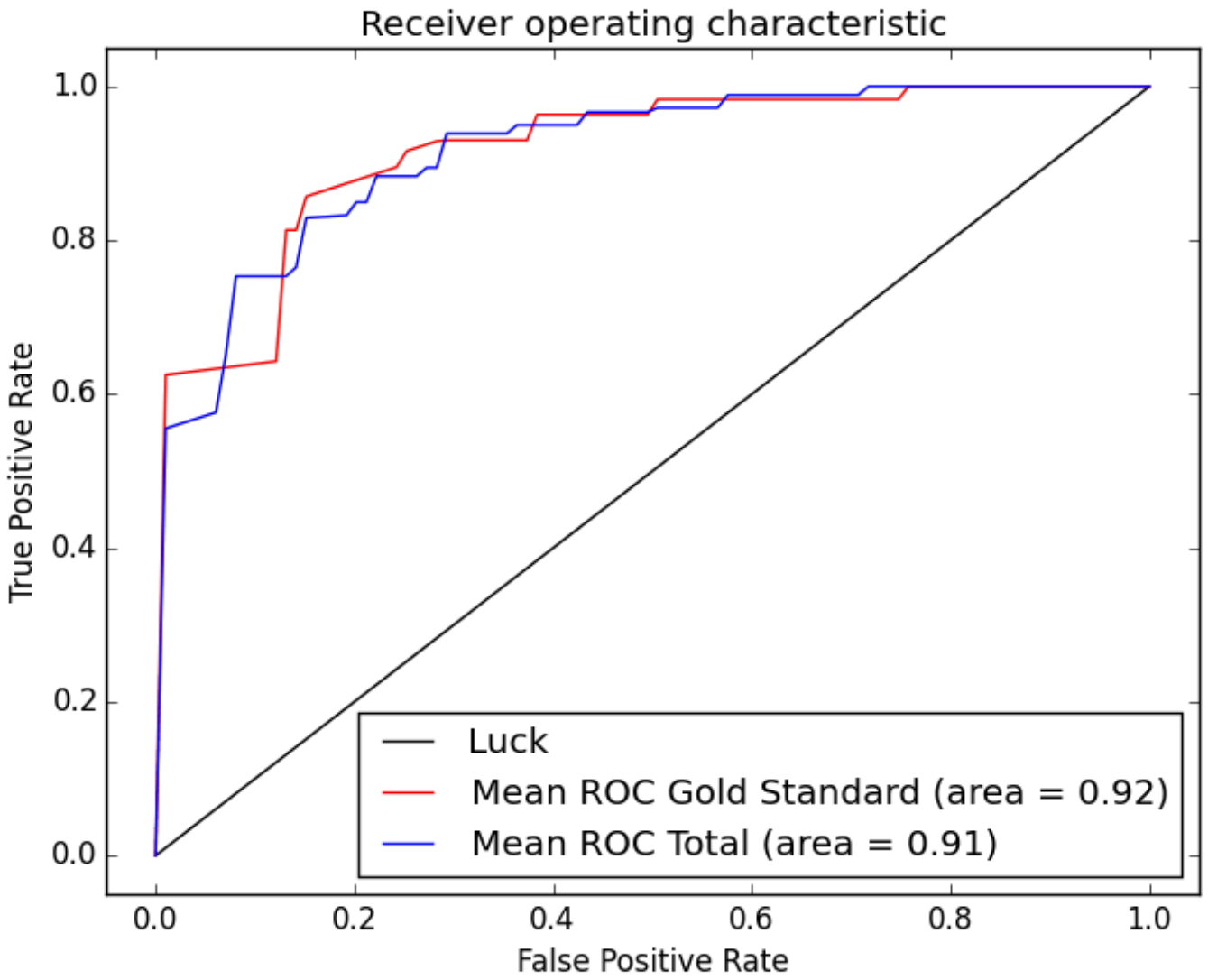
Variance Proportion: 85%

Levene Test (Gold vs. Nonstandard): $P < 0.05$



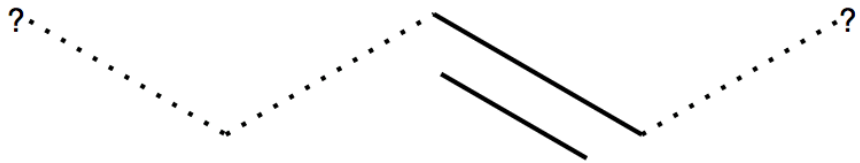
Comparison of models

- 10X Cross Validation
- Gradient Boosted Decision Trees
- 50 estimators, Max Depth = 2, Gamma = 0.29, Subsampling = 0.7, Column Sampling = 0.3, Learning Rate = 0.58
- Mean of CV for Gold Standard and Total

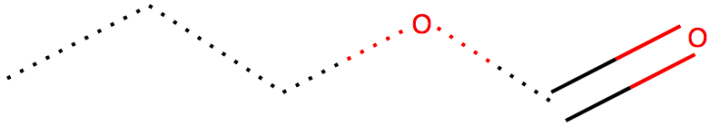


Differences in Feature Importances

[!#1][CH2][CH]=[CH][!#1]



CCCOC=O



[#6]=,:[#6]-,:[#6]=,:[#6]

[!#1][CH2][CH2][CH2][!#1]

[!#1][CH2][CH2][CH2][OH]

Performance of Gold Standard Model

- 50% cross validation
  - Accuracy: 0.8224 +/- 0.0783
  - Precision: 0.8114 +/- 0.1462
  - Recall: 0.7496 +/- 0.1809
  - Receiver Operator, AUC: 0.8112 +/- 0.0847
- 10X cross validation
  - Accuracy: 0.8757 +/- 0.1814
  - Precision: 0.8550 +/- 0.2363
  - Recall: 0.8733 +/- 0.3096
  - Receiver Operator, AUC: 0.8754 +/- 0.1885
- Confirmational dataset (44 compounds added to NREL Compendium)

Acknowledgements

- This research was conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Offices. (Optional): Co-Optima is a collaborative project of multiple national laboratories initiated to simultaneously accelerate the introduction of affordable, scalable, and sustainable biofuels and high-efficiency, low-emission vehicle engines.
- Thank you: Bob McCormick (NREL), Janet Yanowitz (BioEngineering, Inc.), Anthe George (SNL), Ryan Davis (SNL), Oliver Killian (SNL), John Gladden (SNL)