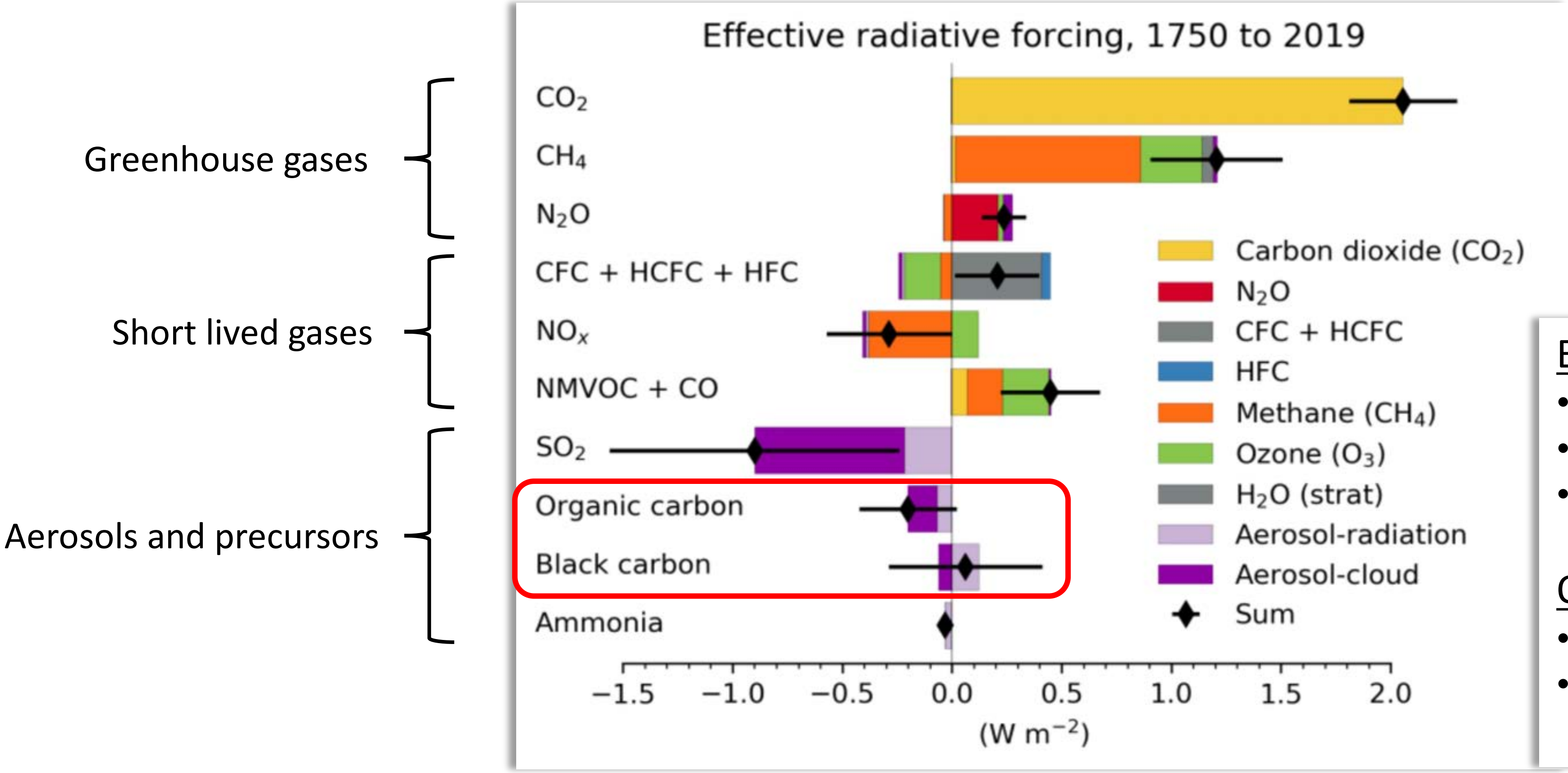
Application of machine learning to correct for biases in ARM filter-based aerosol light absorption datasets

Joshin Kumar[1] (j.kumar@wustl.edu)
Rajan K. Chakrabarty[1] (chakrabarty@wustl.edu)

Co-authors: Nishit J. Shetty[1], Patrick Sheridan[2], Allison Aiken[3], Manvendra Dubey[3], Connor Flynn[4]
[1]Washington University in St. Louis, [2]NOAA Global Monitoring Laboratory, [3]Los Alamos National Laboratory, [4]University of Oklahoma

.

Challenge: High uncertainty in radiative impact of aerosols on climate



Naik et al., 2021, IPCC

Black Carbon (BC):
- An optical definition for soot
- Most absorbing component
- Large uncertainty in forcing

Organic Carbon (OC):
- Low volatility organic material
- Low/No absorption at long visible wavelengths

Solution: Improvements in the measurements of light absorption by aerosols.

Various instruments are used to measure light absorption by aerosols

- **In-situ Instruments** – Photoacoustic Soot Spectrometer (PASS)
  - Principle: Light absorbing particles heat up the surrounding air which emits pressure waves that can be detected with a microphone
  - Pros: Accurate
  - Cons: Expensive

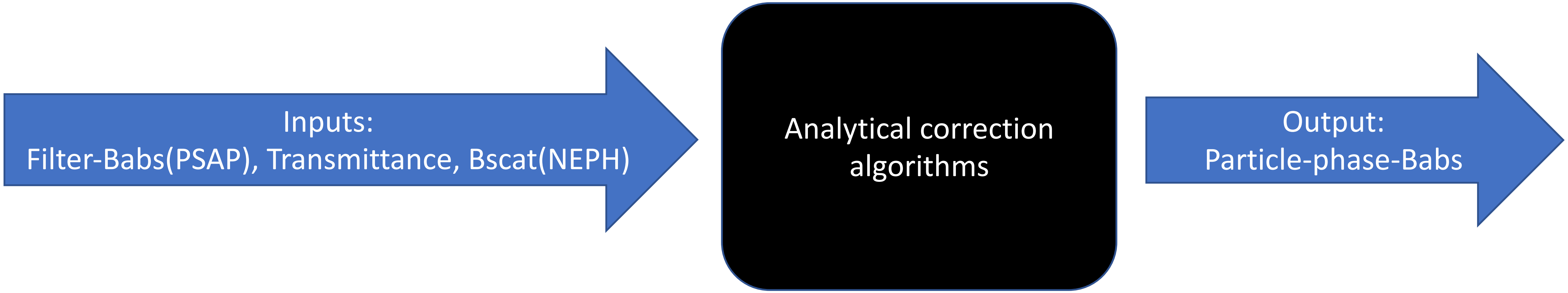
- **Filter-based Instruments** – Particle Soot Absorption Photometer (PSAP), Tricolor Absorption Photometer (TAP), Aethalometer
  - Principle: Change in filter transmission or reflection (or increased attenuation) after sampling is due to particle light absorption
  - Pros: Economic
  - Cons: subject to unquantifiable artifacts; Multiple scattering enhances the absorption measurement; Filter loading dependent

Hence, we need correction algorithms to correct for the biases in filter-based instruments.

Correction algorithms are used to correct filter-based aerosol light absorption measurements

- Bond et al., 1999
- Virkkula et al., 2005, 2010
- Muller et al., 2014
- Hanyang Li et al., 2020

$$\sigma_{AP}(PSAP) = (k_0 + k_1(h_0 + h_1\omega_0)\ln(Tr))\sigma_0 - s\sigma_{SP}.$$



Research Questions

1. How does the ARM's PSAP (filter-based) and PASS (particle-phase) compare?

2. How applicable are the traditionally used PSAP correction algorithms for SGP and for the TRACER main site?

3. Could Machine Learning be used to correct the filter-based absorption measurements more accurately than previously developed filter-correction algorithms?

4. What are different factors affecting accuracy of correcting PSAP data?

Contents:

1. Correcting for biases in filter-based aerosol light absorption measurements:
    - At SGP site.
    - At Laboratory from burn experiments.

2. Determining the factors affecting inaccuracies in filter-based aerosol light absorption measurements at TRACER main site.

Instrument data obtained from Southern Great Plain (SGP) Site

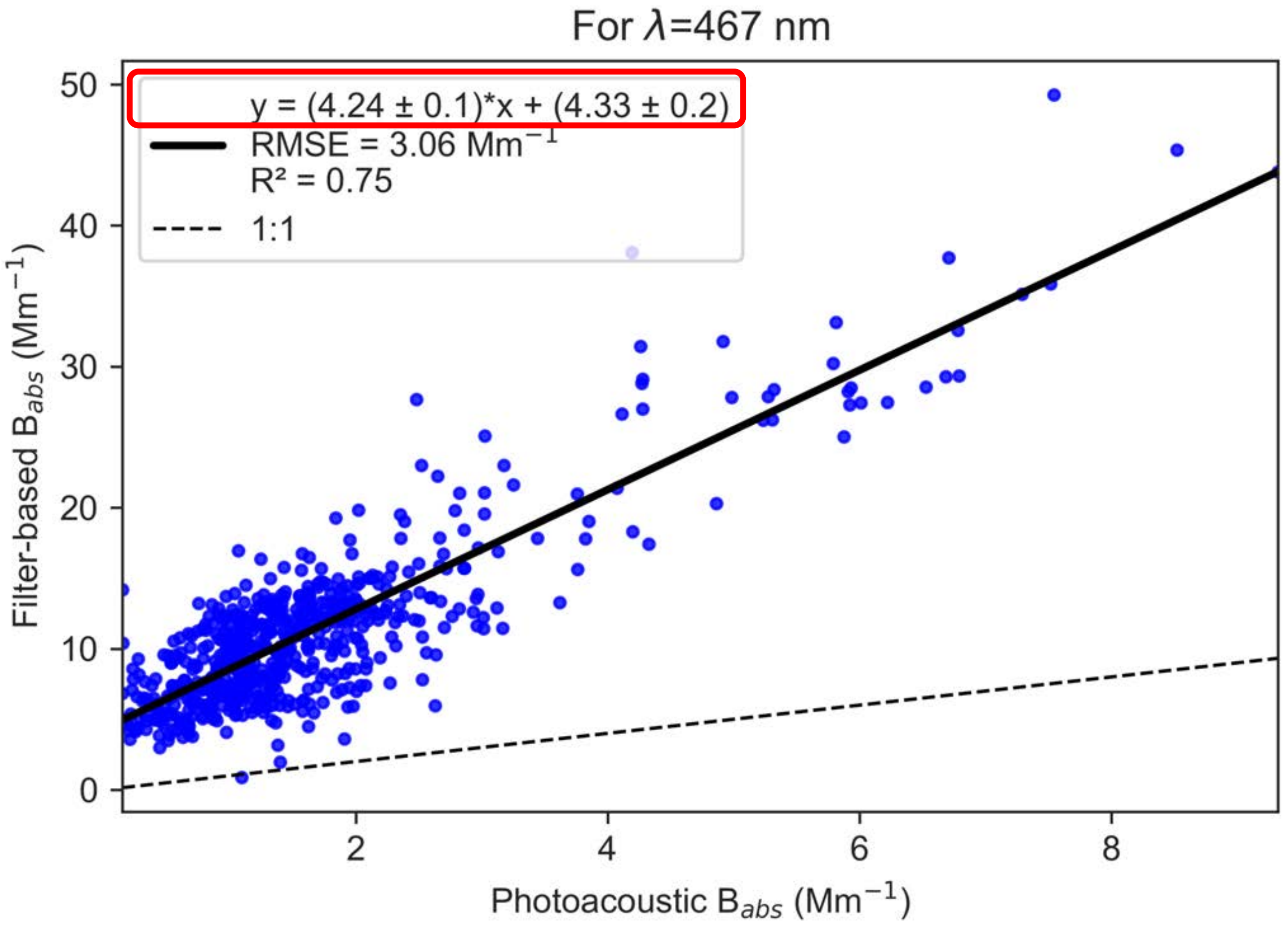
- Ambient ground-based data used from Atmospheric Radiation Measurement (ARM) user facility at Southern Great Plains (SGP)

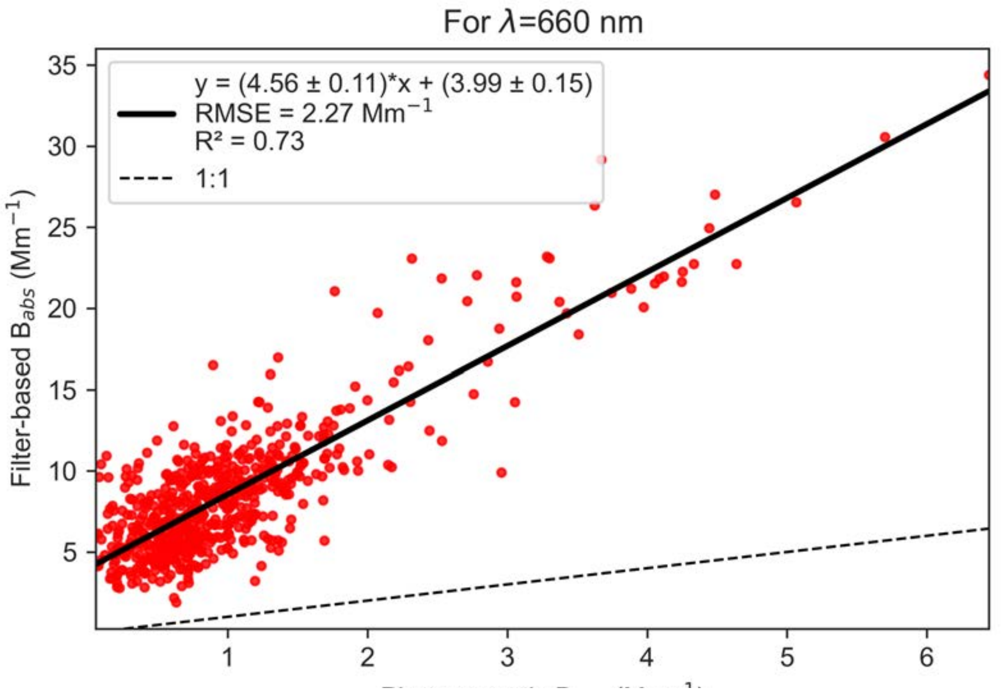
- SGP is a typical rural, mid-continental site and world's largest and most extensive climate research facility.

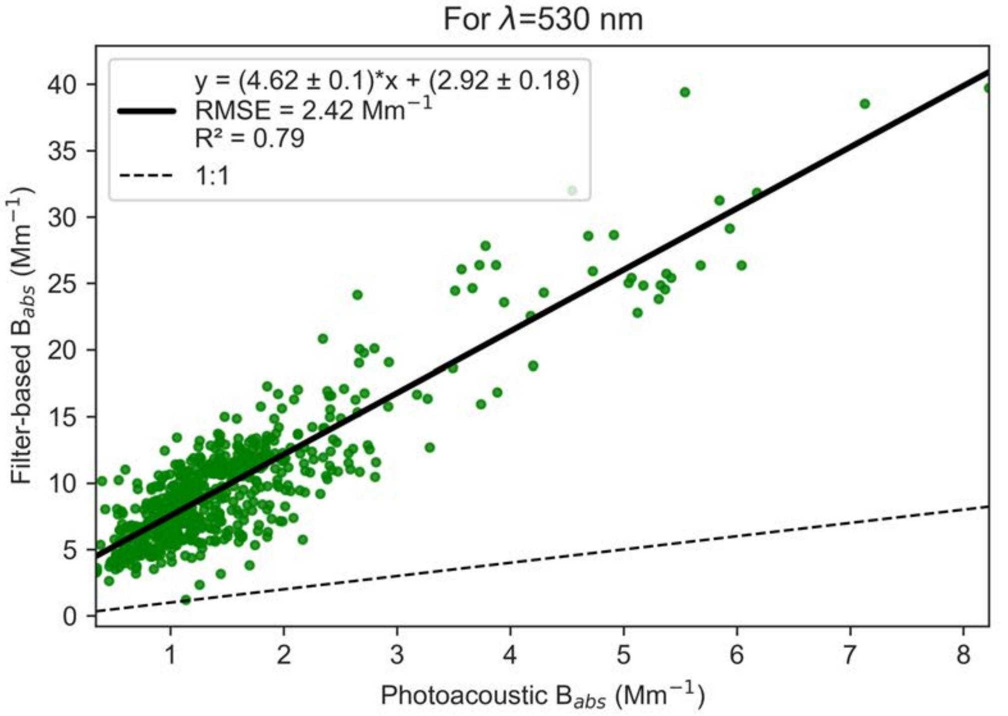
- High resolution timeseries data used from 27th Jun to 25th Sept, 2015
  - Photoacoustic Soot Spectrometer (PASS-3λ)
  - Particle Soot Absorption Photometer (PSAP-3λ)
  - Nephelometer (NEPH-3λ)
  - Aerodyne Aerosol Chemical Speciation Monitor (ACSM)

SGP site: PSAP(Filter) vs PASS(In-situ)

PASS adjusted to PSAP Wavelengths using inferred AAE values







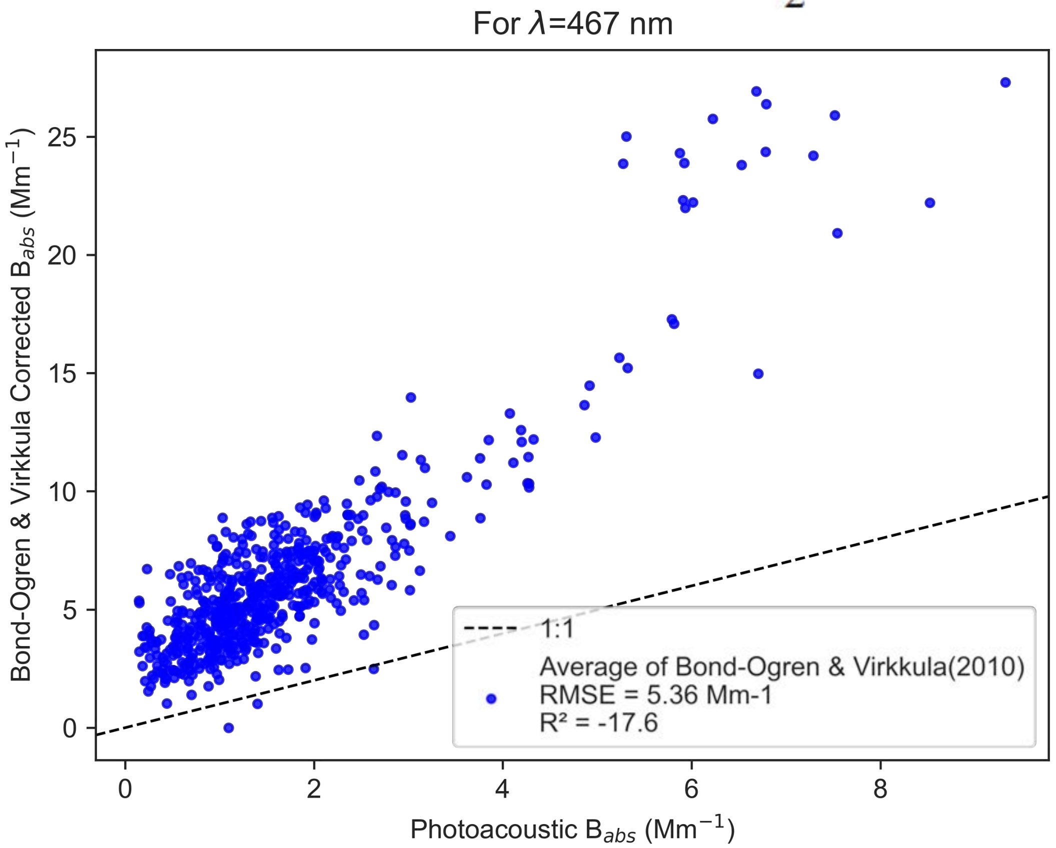
Key-take-away: 4x Overestimation of PSAP

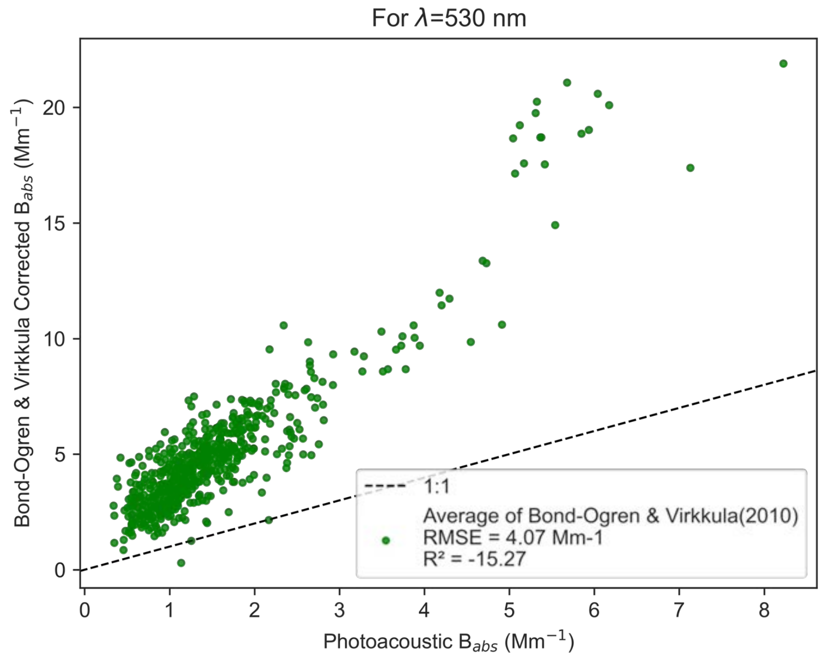
=> Need for filter-correction algorithms

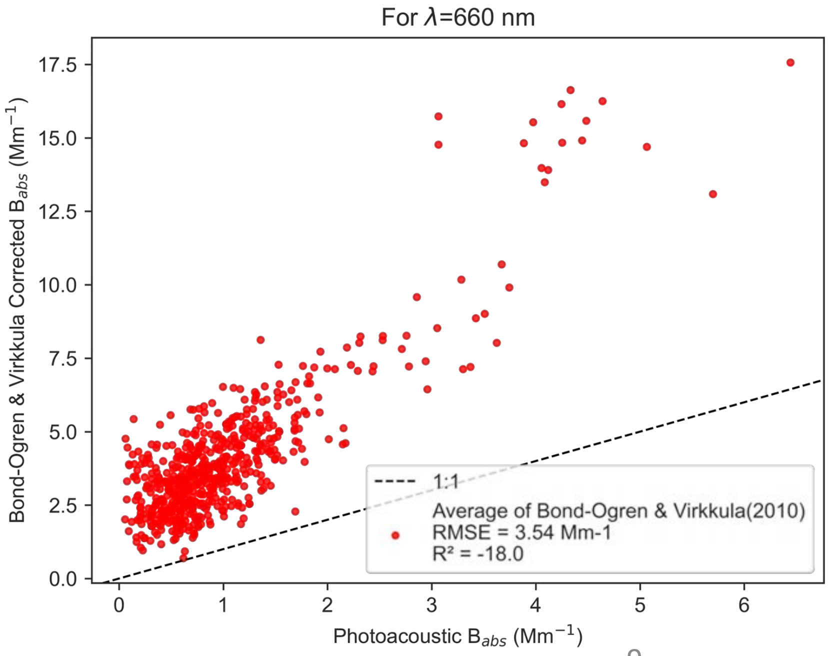
SGP: ARM's current correction algorithm: Average of Bond-Ogren and Virkkula (2010)

$$B_{abs}(\text{Bond–Ogren-corrected}) = B_{PSAP} \times \left(\frac{1}{1.5557 \times \text{Tr} + 1.0227} \right) - 0.0164 \times B_{scat}$$

$$B_{abs}(\text{average-corrected}) = \frac{\left[\begin{array}{c} B_{abs}(\text{unrevised Virkkula-corrected}) \\ + B_{abs}(\text{Bond–Ogren-corrected}) \end{array} \right]}{2}$$





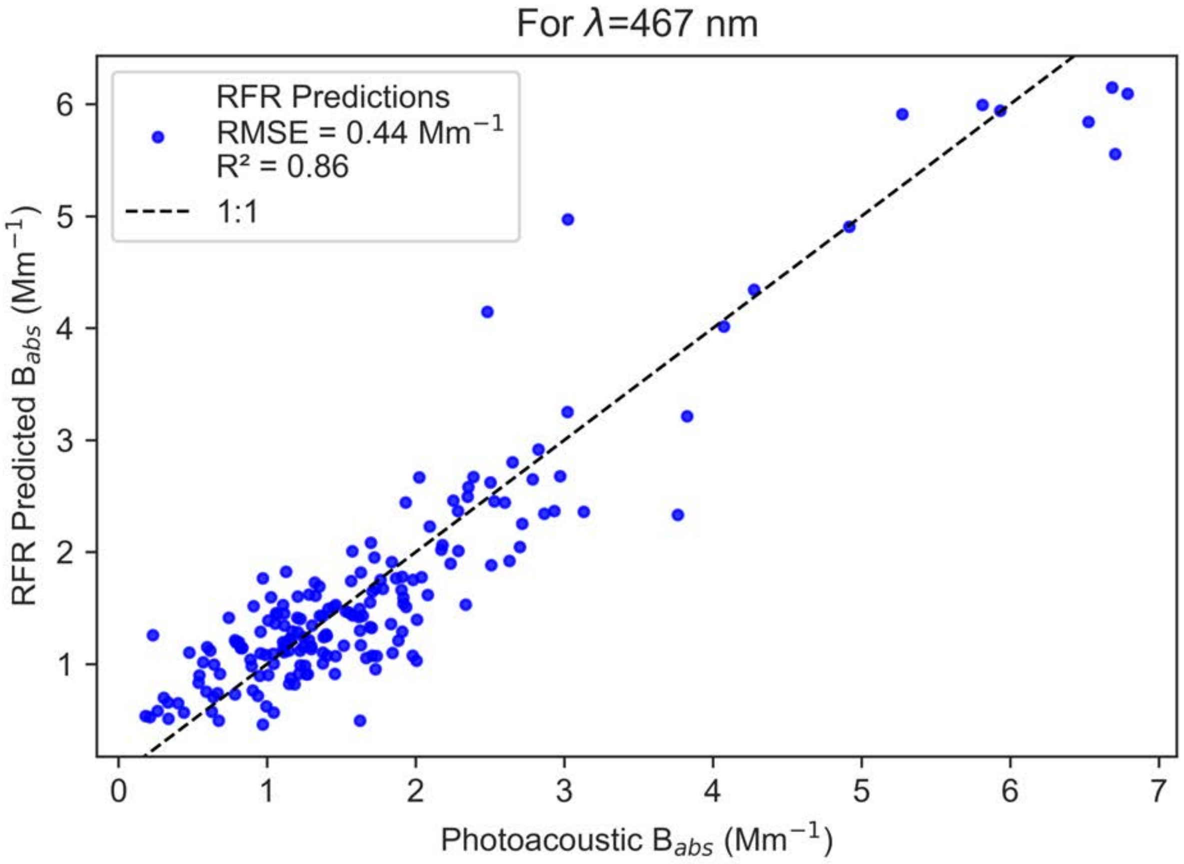


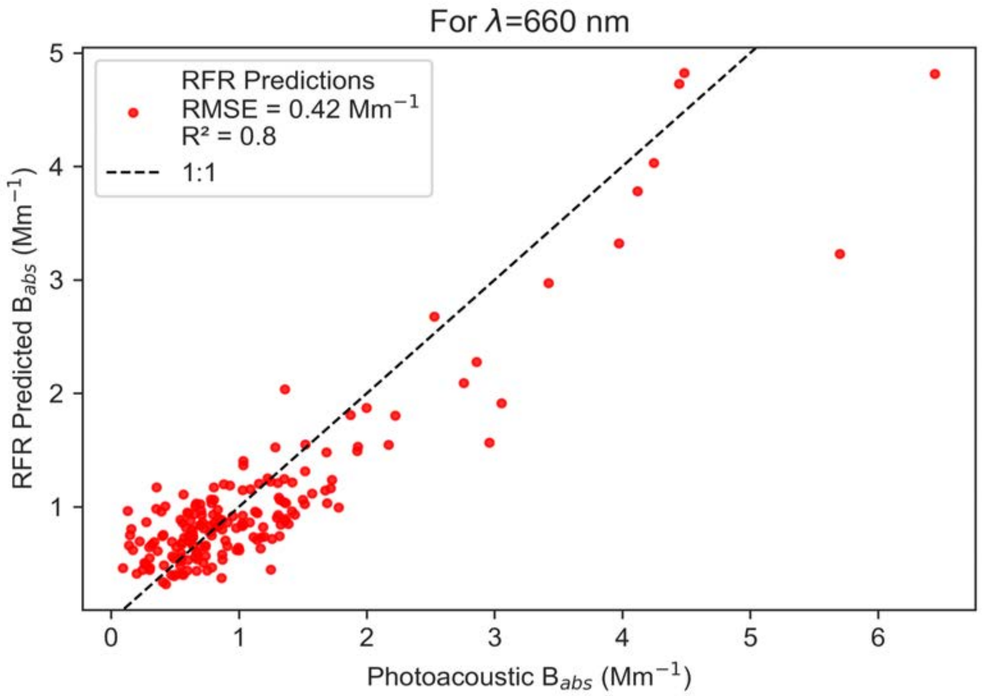
SGP Site: Random Forest Regression (RFR)

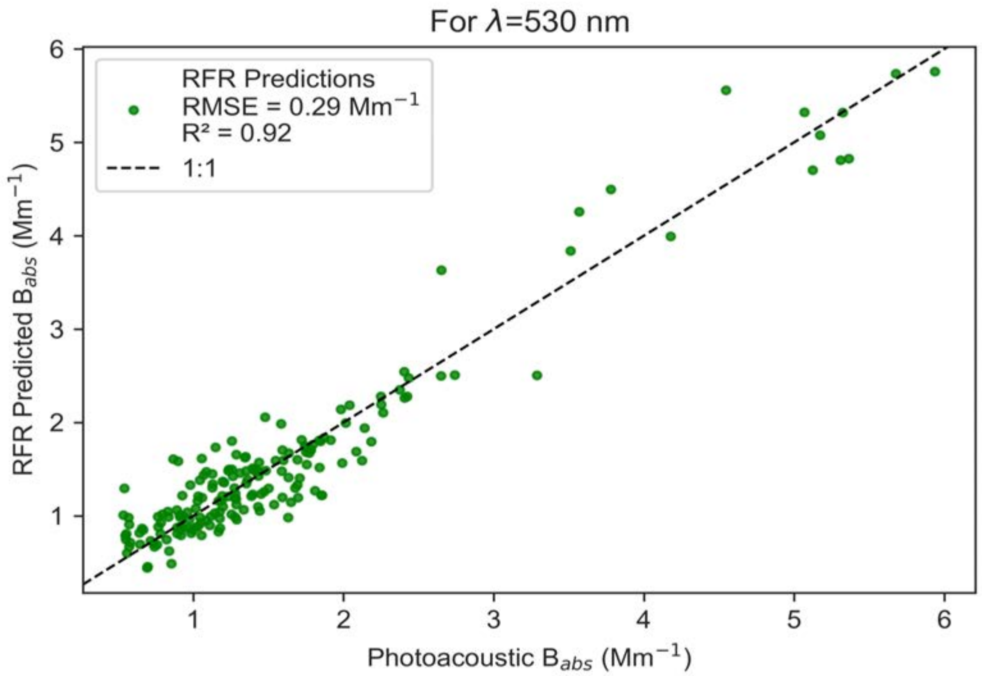
A supervised Machine Learning algorithm

Inputs: B_{abs} (PSAP), Transmittance (PSAP), B_{scatt} (NEPH), C_{Mass} (ACSM)

Output: Corrected particle-phase B_{abs}

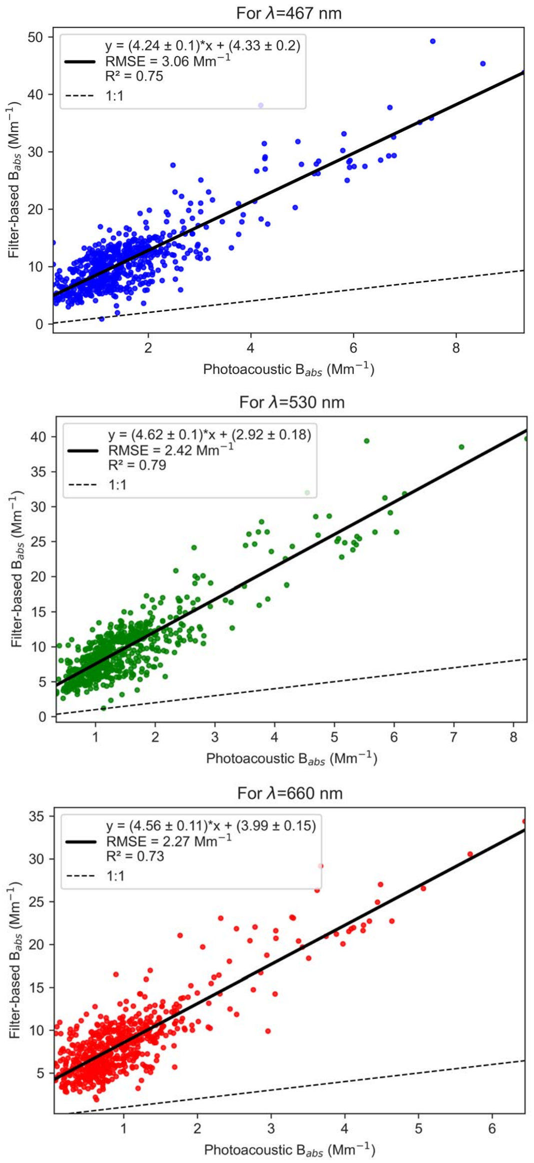






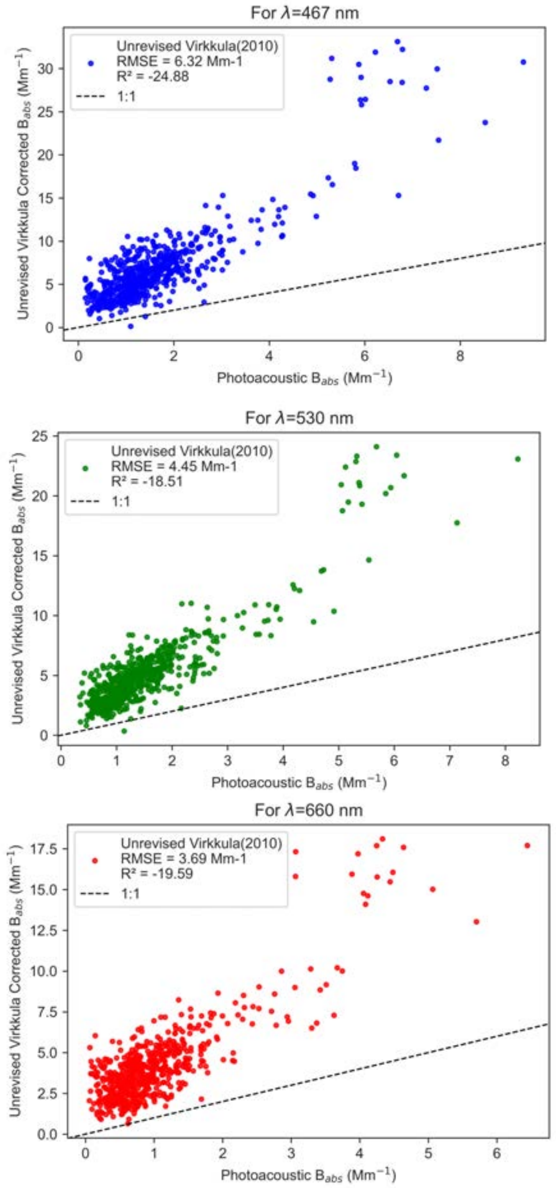
SGP Site: Raw Data: PSAP vs PASS

(PSAP =~ 4 x PASS)



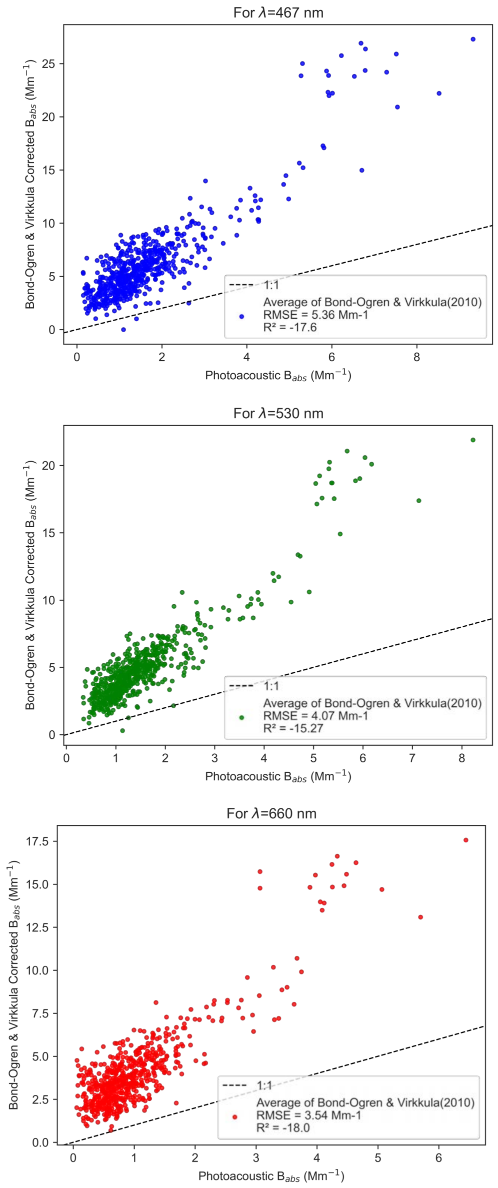
Virkkula (2010) correction

(Erroneous but extensively used)



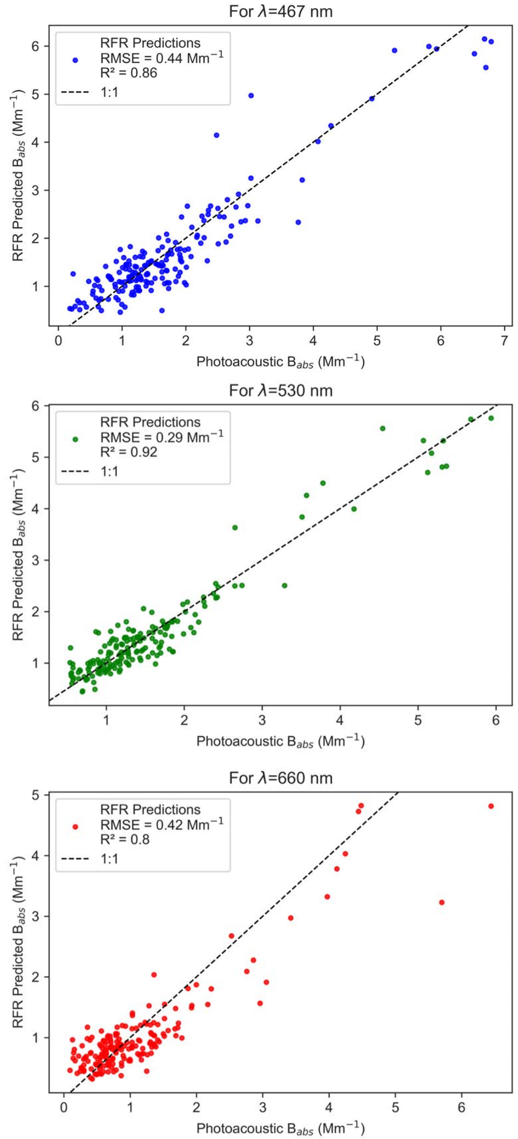
Current ARM correction: Bond-Ogren-Virkkula

(Better than just Virkkula)



Machine Learning: Random Forest Regression

(Best Accuracy)

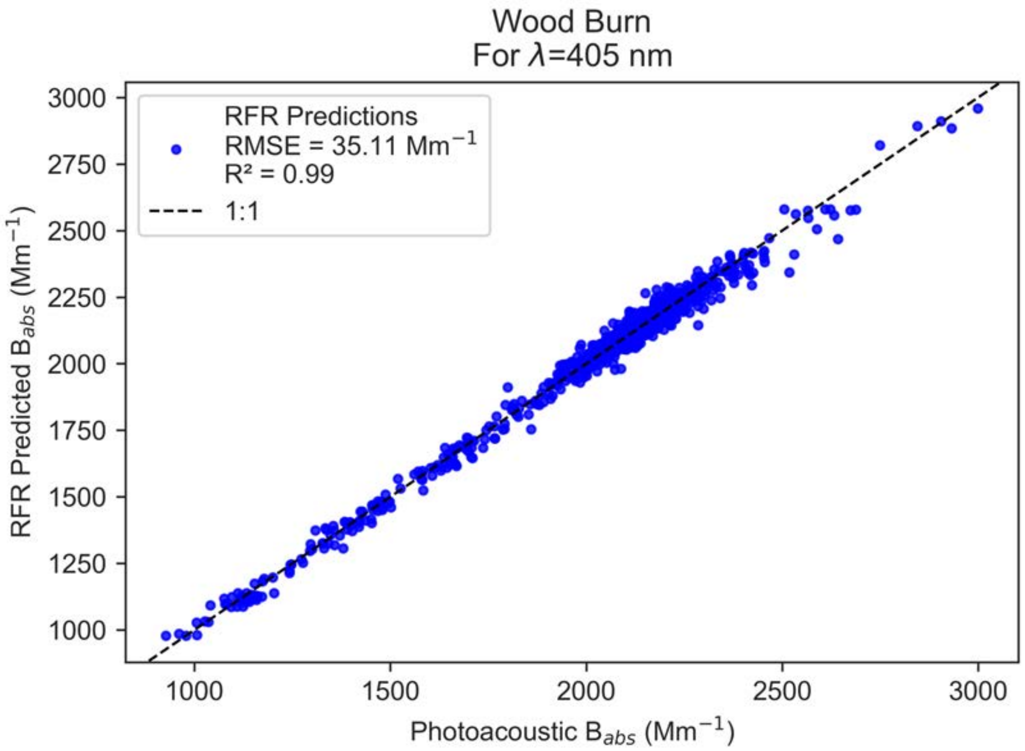


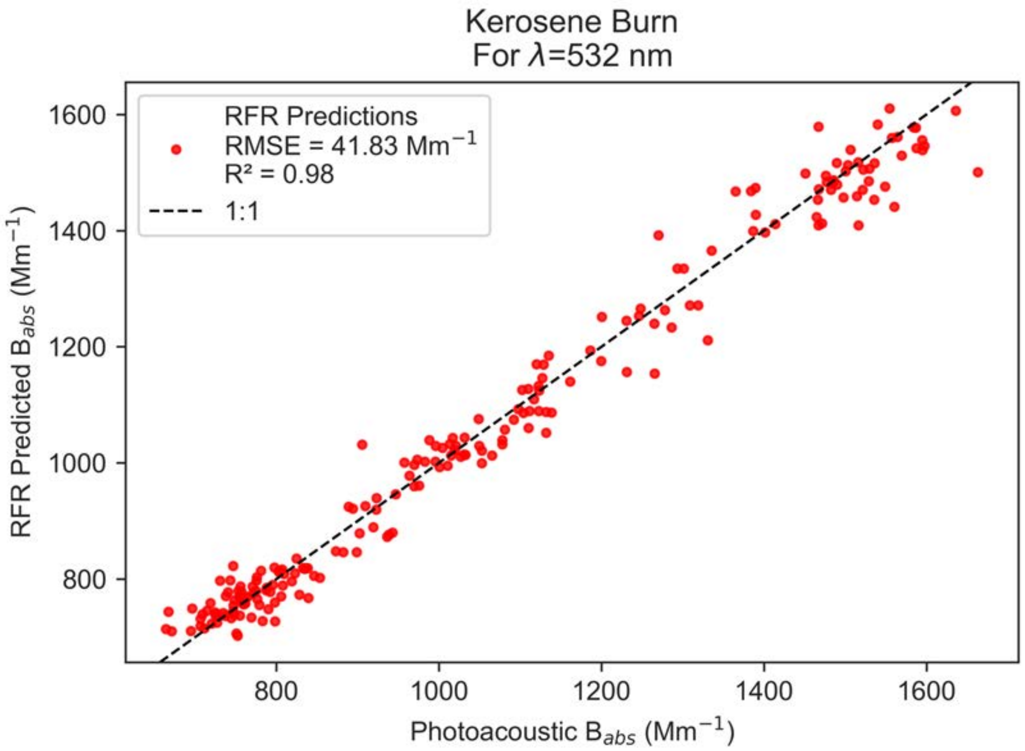
Exploring the potential of RFR in laboratory datasets:

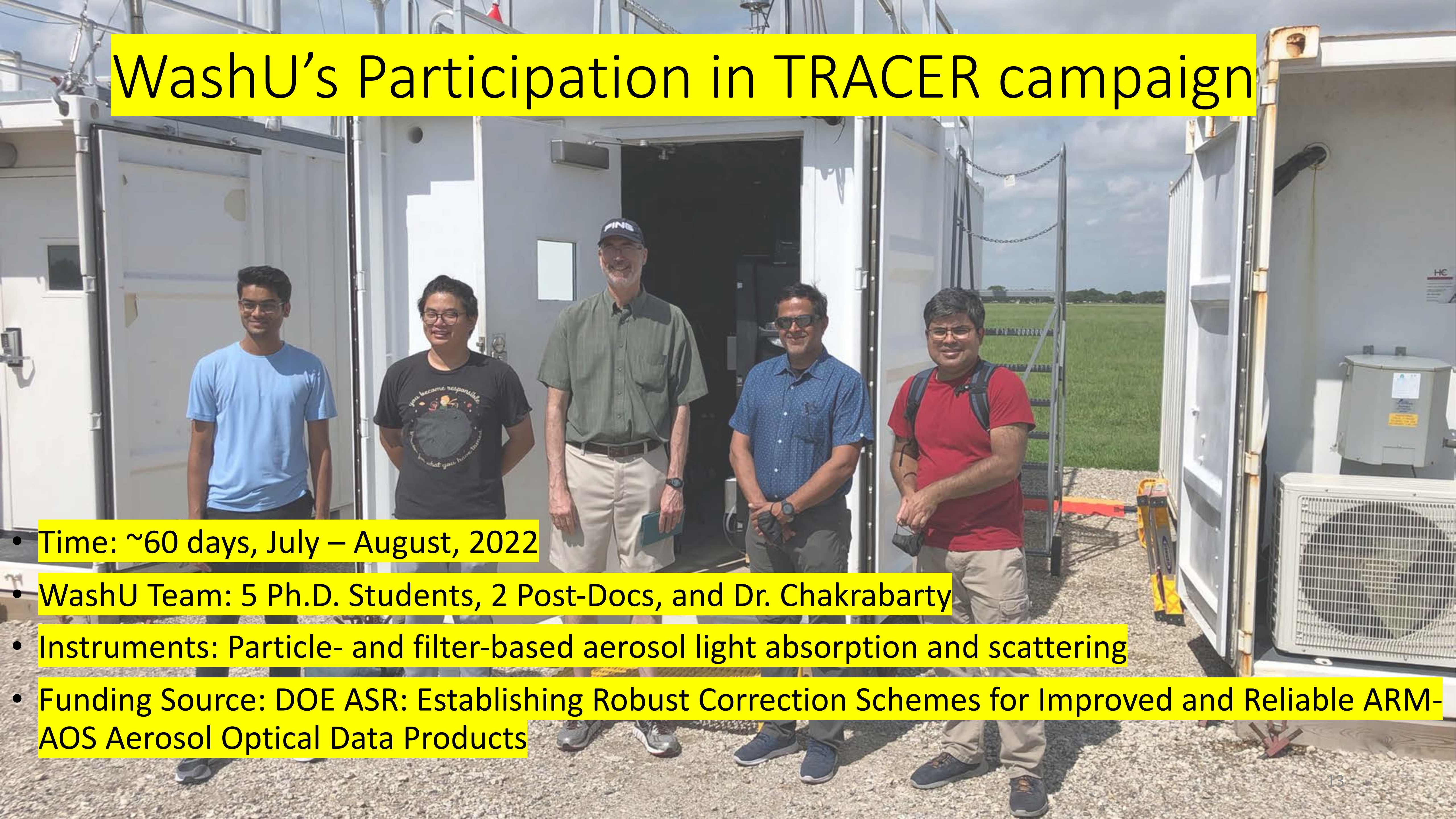
Applying RFR Algorithm on Lab generated Burn data

Inputs: B_{abs} (TAP), B_{scatt} (NEPH), Number size distribution parameters (N, μ_g, σ_g)

Output: Corrected particle-phase B_{abs}



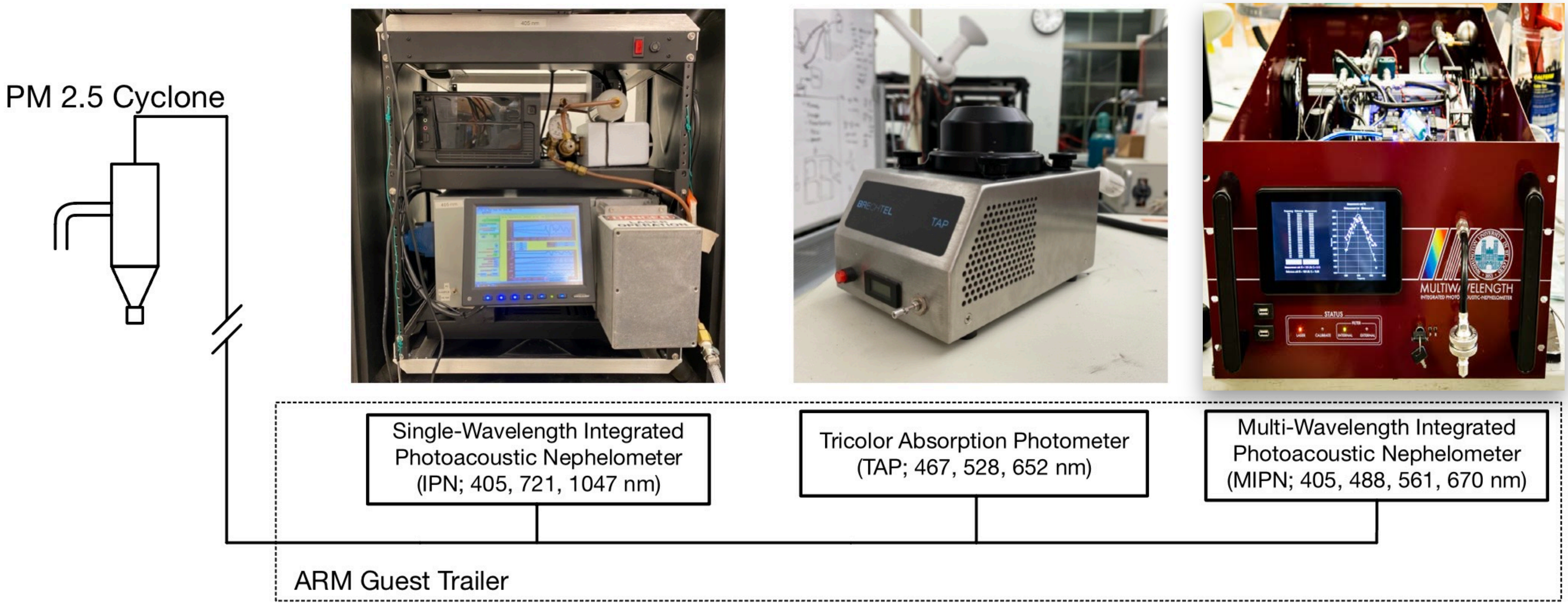




WashU’s Participation in TRACER campaign

- Time: ~60 days, July – August, 2022
- WashU Team: 5 Ph.D. Students, 2 Post-Docs, and Dr. Chakrabarty
- Instruments: Particle- and filter-based aerosol light absorption and scattering
- Funding Source: DOE ASR: Establishing Robust Correction Schemes for Improved and Reliable ARM-AOS Aerosol Optical Data Products

Schematic of the WashU's instrument setup in TRACER field campaign



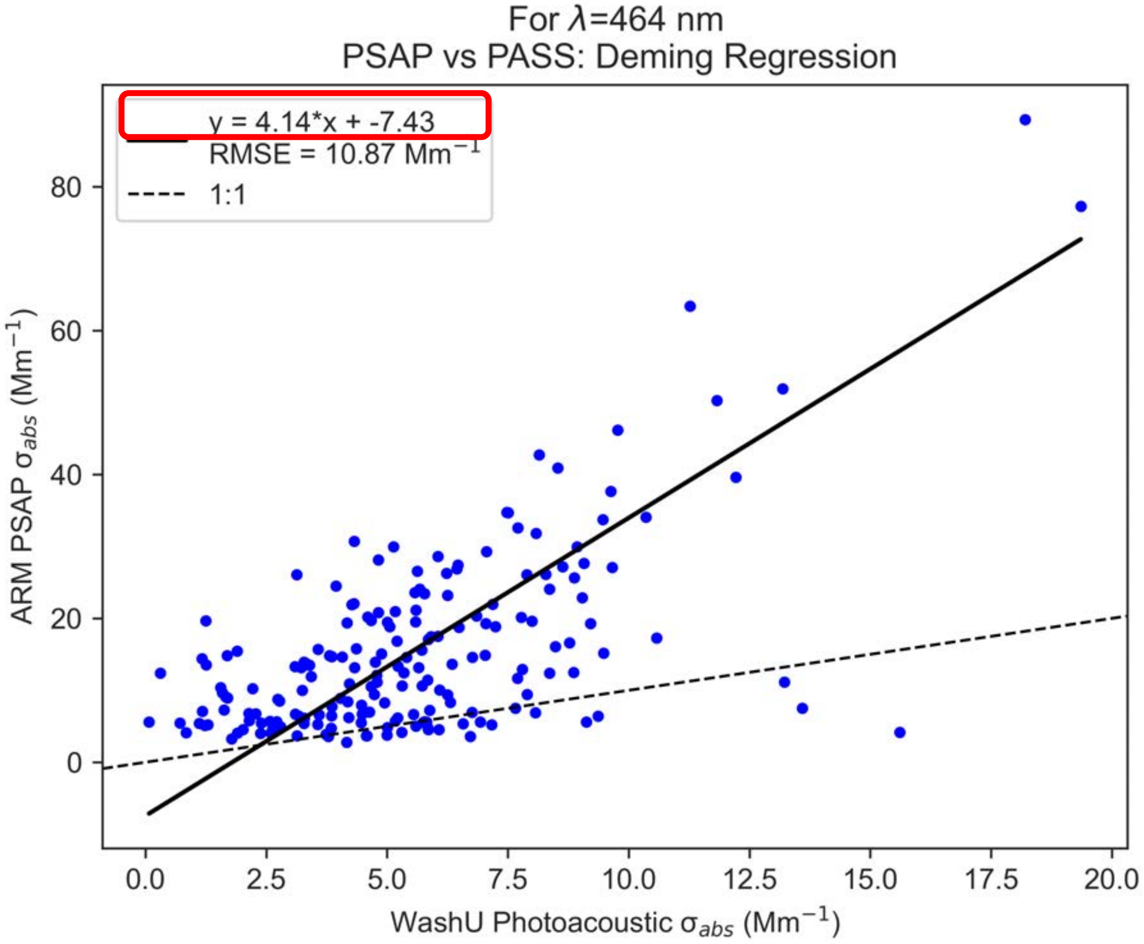
Note: Single-wavelengths of PASS were choosen such that they can be used to correct for filter-based measurements.

Instrument data obtained from TRACER campaign's La Porte Site

- Ambient ground-based data used from TRACER's La Porte, TX site.
- La Porte is a coastal industrial site with flaring events from factories near the site.
- High resolution timeseries data used from 1st – 29th August, 2022:
  - WashU's Photoacoustic Soot Spectrometer (PASS)
  - ARM-AMF1 Particle Soot Absorption Photometer (PSAP-3λ)
  - ARM-AMF1 Nephelometer (NEPH-3λ)
  - ARM-AMF1 Aerodyne Aerosol Chemical Speciation Monitor (ACSM)
  - ARM-AMF1 Single Particle Soot Photometer (SP2)

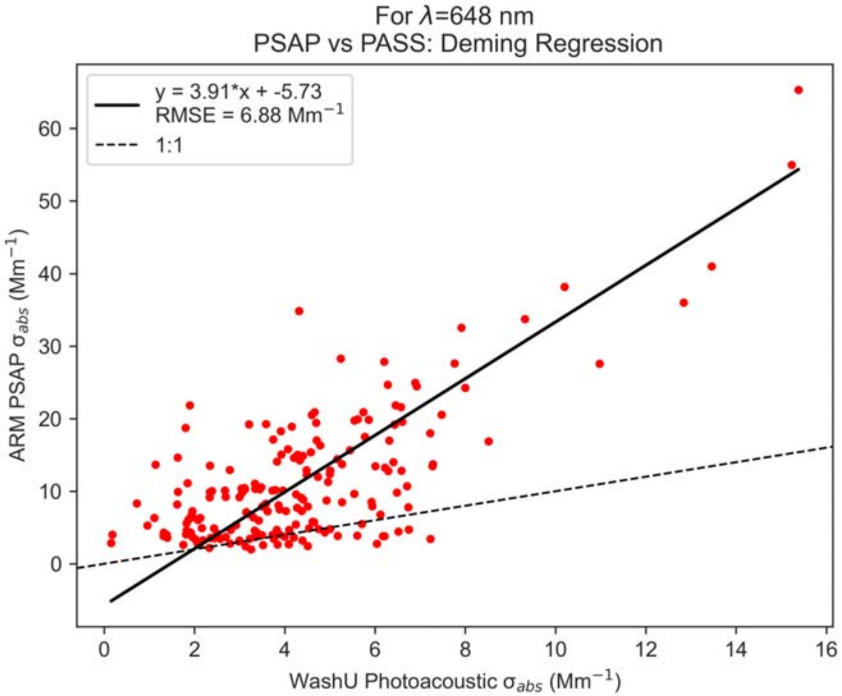
TRACER site: PSAP(Filter) vs PASS(In-situ)

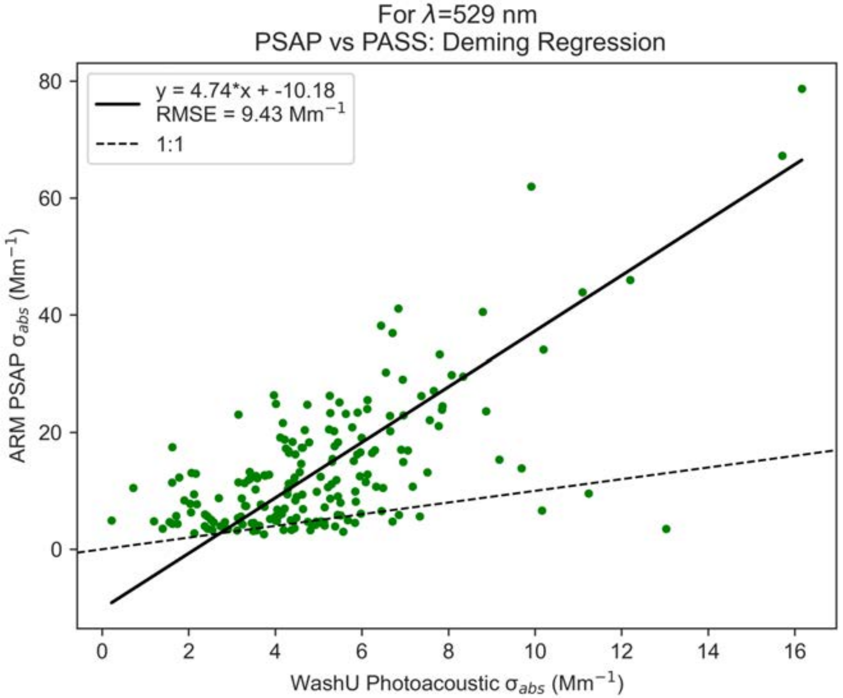
PASS adjusted to PSAP Wavelengths using inferred AAE values



Key-take-away: 4x Overestimation of PSAP

=> Need for filter-correction algorithms



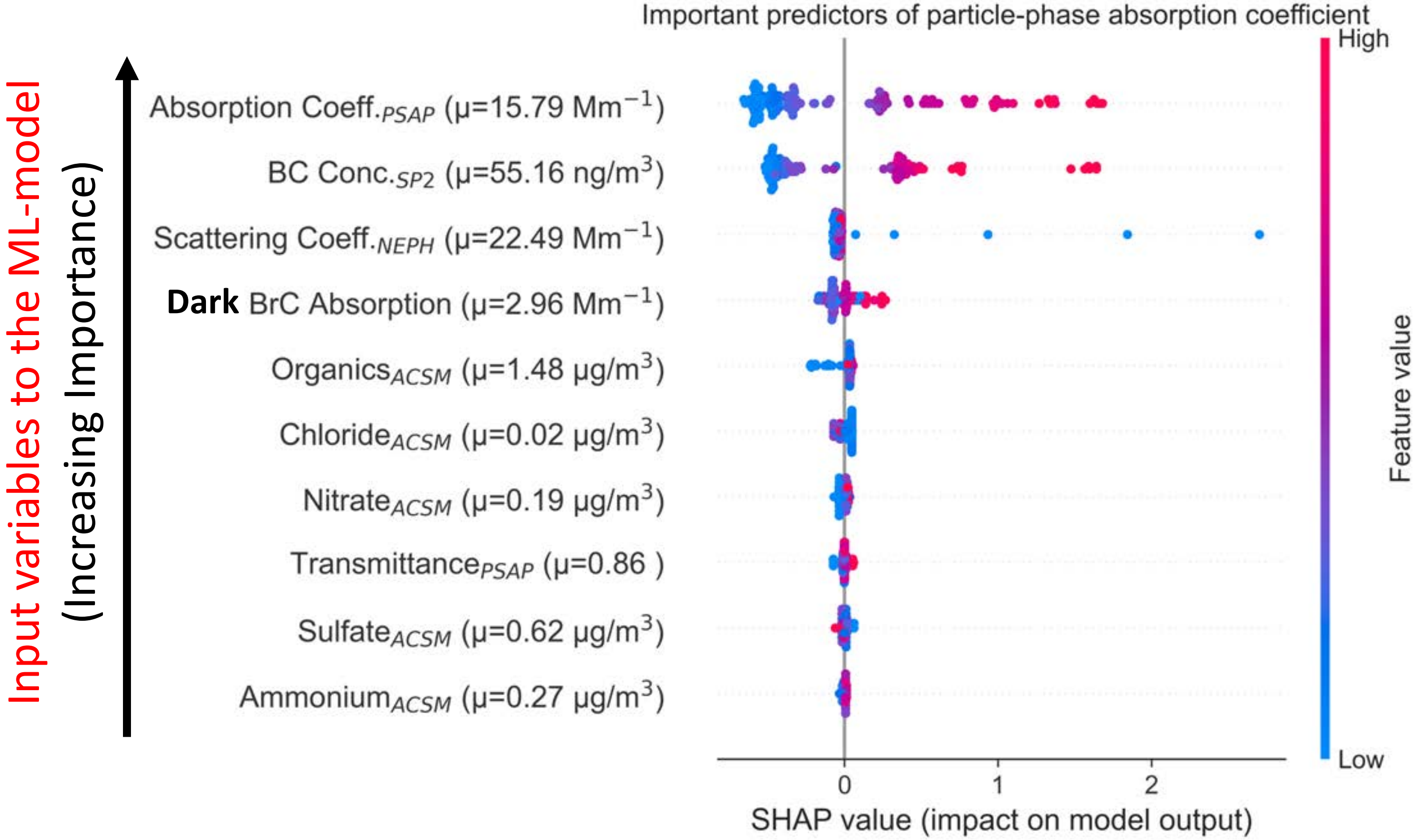


Parameters affecting PSAP correction at TRACER main site

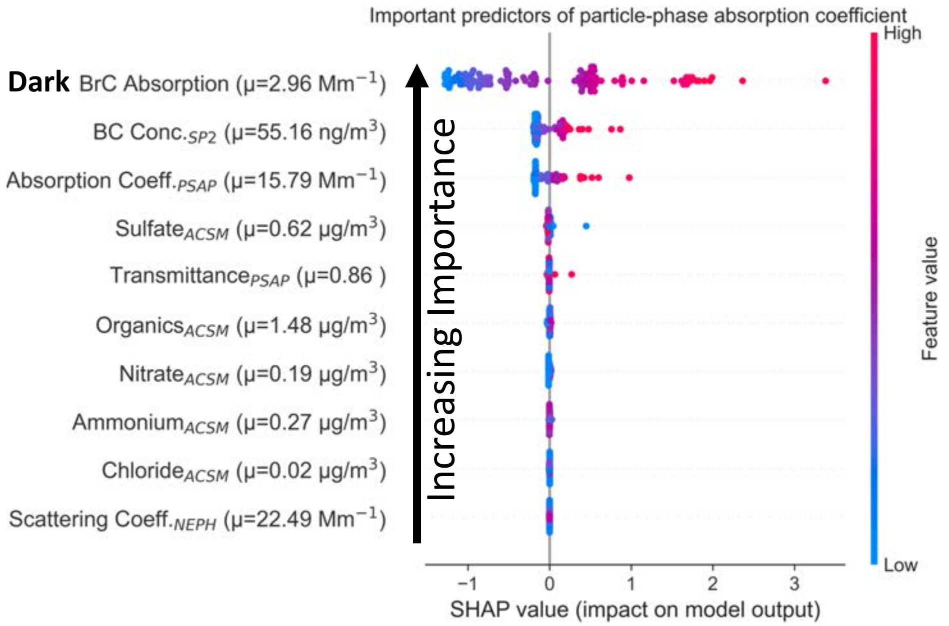
Inputs: B_{abs} (PSAP), Transmittance (PSAP), B_{scatt} (NEPH), C_{Mass} (ACSM and SP2)

Output: Corrected particle-phase B_{abs}

λ = 464 nm



λ = 648 nm



Key-take-away: BC, Dark BrC, and Organics affect PSAP correction

=> Reinforces need for ML-based PSAP correction algorithm

Conclusion

- The PSAP (filter-based) light absorption measurements overestimate by around **4x** compared to Photoacoustic (particle-phase) measurements.

- Filter-correction algorithms/models are site specific.

- The order based on accuracy of the correction algorithms for filter-based absorption measurements:

  ML > ARM’s current correction (Bond-Ogren-Virkkula) > Virkkula (2010)

- Filter-based inaccuracies are strongly influenced by:
  - BC concentration
  - Dark brown carbon absorption

Thank you!

Questions?

- Manuscript published at AMT

  Kumar, J., Paik, T., Shetty, N. J., Sheridan, P., Aiken, A. C., Dubey, M. K., and Chakrabarty, R. K.: Correcting for filter-based aerosol light absorption biases at the Atmospheric Radiation Measurement program's Southern Great Plains site using photoacoustic measurements and machine learning, Atmos. Meas. Tech., 15, 4569-4583, 10.5194/amt-15-4569-2022, 2022.

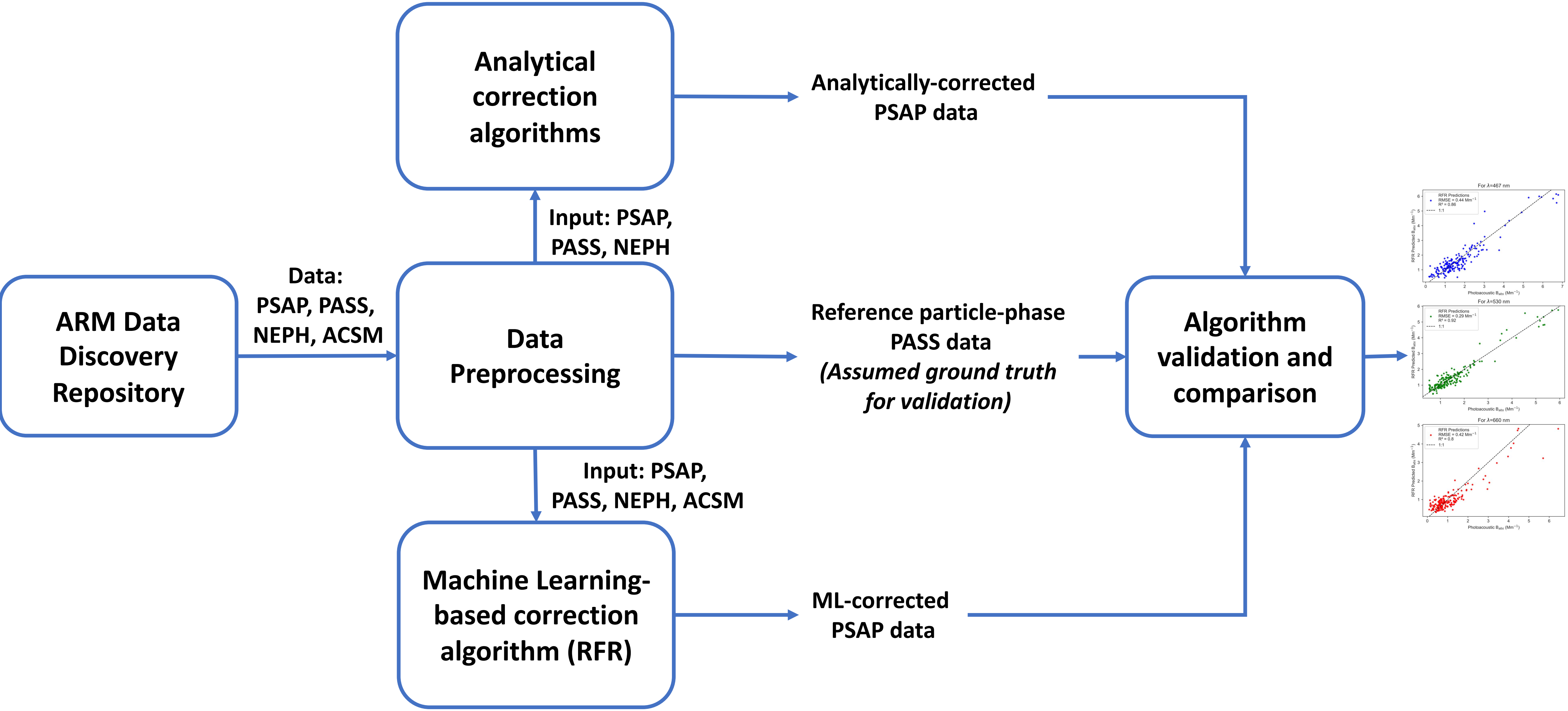
- Python codes for RFR correction are openly available on GitHub:

  https://github.com/joshinkumar/Filter-correction-ML-code

Contact us at:

Rajan K. Chakrabarty (chakrabarty@wustl.edu)

Joshin Kumar (j.kumar@wustl.edu)



Analytical correction algorithms
Analytically-corrected PSAP data
Input: PSAP, PASS, NEPH
ARM Data Discovery Repository
Data: PSAP, PASS, NEPH, ACSM
Data Preprocessing
Reference particle-phase PASS data
(Assumed ground truth for validation)
Algorithm validation and comparison
Input: PSAP, PASS, NEPH, ACSM
Machine Learning-based correction algorithm (RFR)
ML-corrected PSAP data
For λ=467 nm
RFR Predictions
RMSE = 0.44 Mm⁻¹
R² = 0.86
1:1
RFR Predicted B_abs (Mm⁻¹)
Photoacoustic B_abs (Mm⁻¹)
For λ=530 nm
RFR Predictions
RMSE = 0.29 Mm⁻¹
R² = 0.92
1:1
RFR Predicted B_abs (Mm⁻¹)
Photoacoustic B_abs (Mm⁻¹)
For λ=660 nm
RFR Predictions
RMSE = 0.42 Mm⁻¹
R² = 0.8
1:1
RFR Predicted B_abs (Mm⁻¹)
Photoacoustic B_abs (Mm⁻¹)

What is Random Forest?

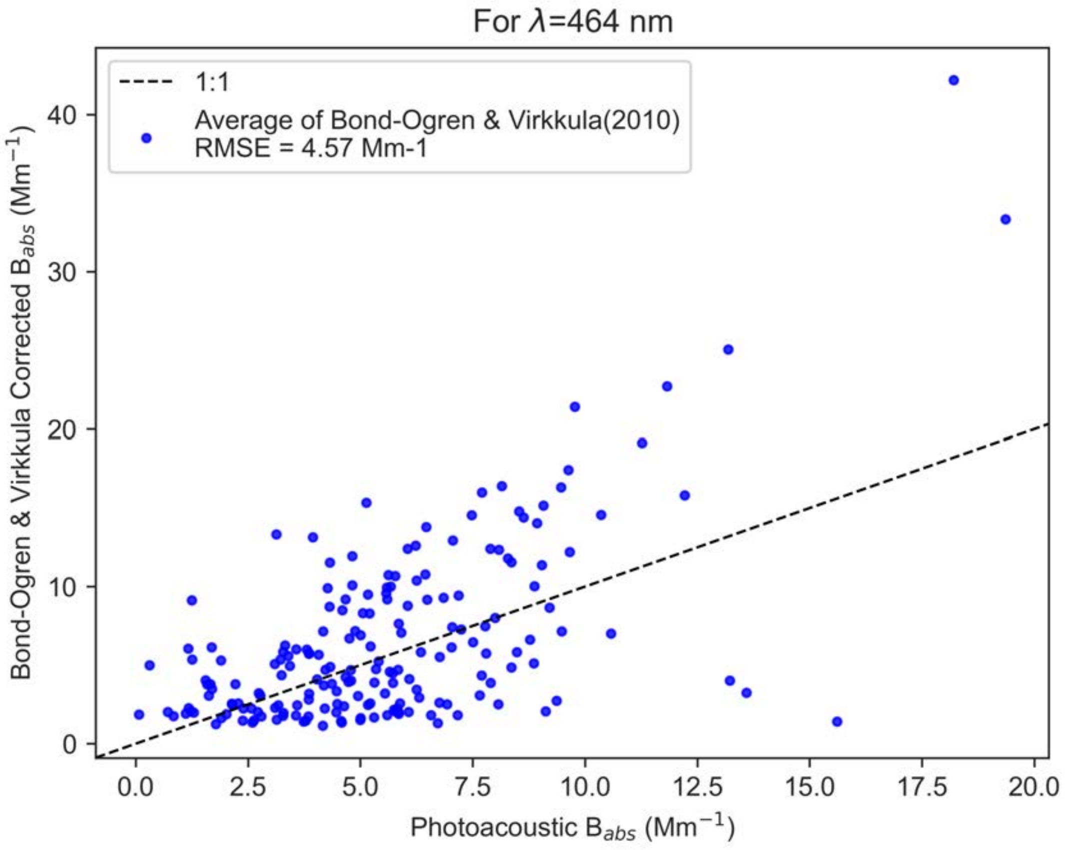
- Random Forest is a supervised machine learning algorithm made of “Decision Trees”
- Each Decision Tree is made up of “Nodes”
- Each Node divides the training data using inequalities on feature(s) variables
- The Leaf at the end(after series of nodes and data sample division) of the tree contain similar data samples and are used to make a decision once the tree is trained; i.e. also known as prediction of the Tree.
- Predictions from all the Trees are used to make a final optimized prediction for a sample data of Input.
- Hence, Random Forest is ensemble algo. based on the “Wisdom of Crowd” approach.

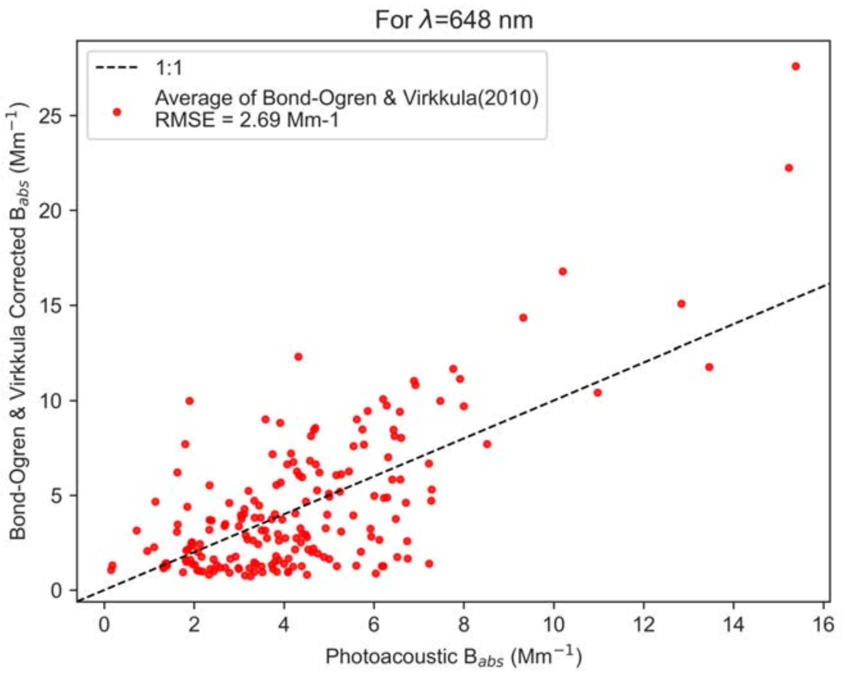
algorithm:

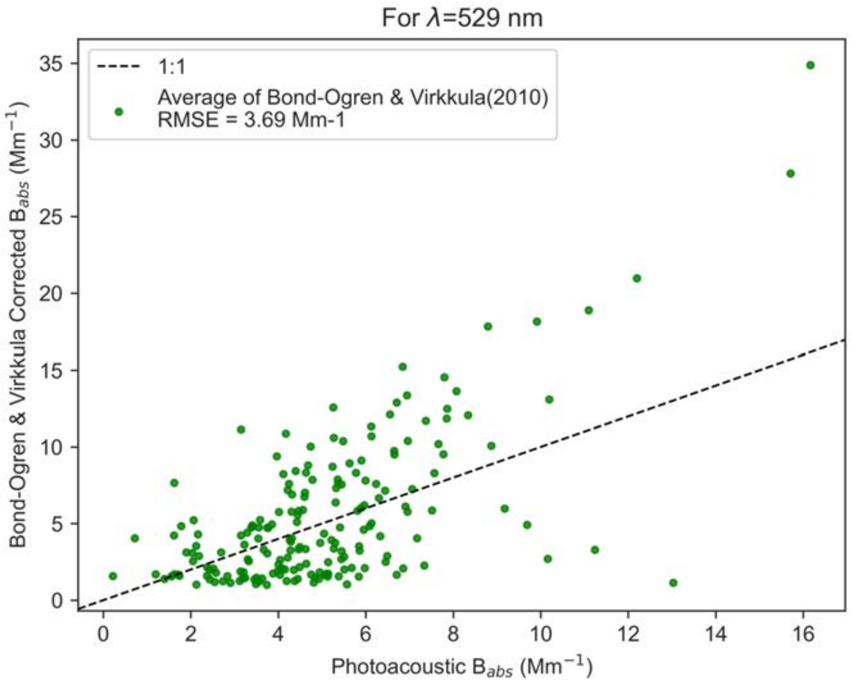
Average of Bond-Ogren and Virkkula (2010)

$$B_{abs}(\text{Bond–Ogren-corrected}) = B_{PSAP} \times \left(\frac{1}{1.5557 \times Tr + 1.0227} \right) - 0.0164 \times B_{scat}$$

$$B_{abs}(\text{average-corrected}) = \frac{\left[B_{abs}(\text{unrevised Virkkula-corrected}) + B_{abs}(\text{Bond–Ogren-corrected}) \right]}{2}$$







TRACER Site: Random Forest Regression (RFR)

A supervised Machine Learning algorithm

Inputs: B_{abs} (PSAP), Transmittance (PSAP), B_{scatt} (NEPH), C_{Mass} (ACSM and SP2)

Output: Corrected particle-phase B_{abs}