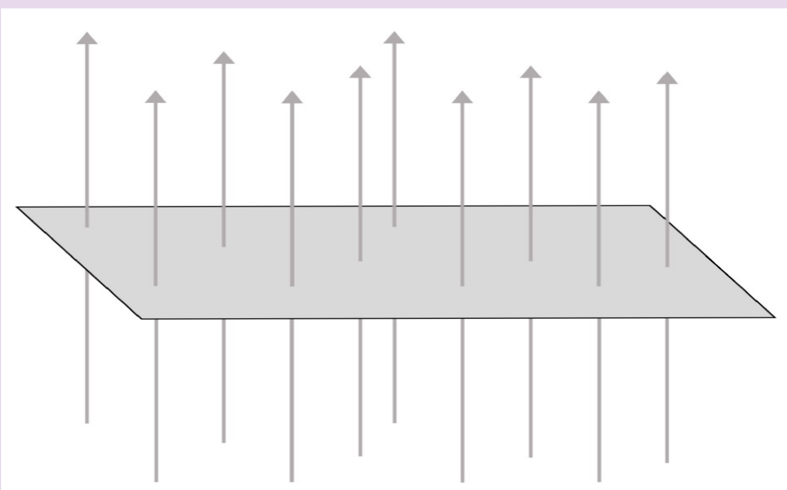
Exploring Data-Driven Equation Discovery to Model Moisture Flux (MF)

Rebecca Porter[1, 2], Yu Huang[3], Candace Agonafir[3]

1. University Corporation for Atmospheric Research (UCAR), Boulder, CO, 2. University of Saint Mary, Leavenworth, KS, 3. Columbia University, New York City, NY

1. Introduction

Visualization of MF (Fig. 1)



- Current empirical equations within climate models rely on overgeneralizations of sub-grid scale processes like MF.

MF Equation (Eq. 1)

$$w'q' = k\frac{dq}{dz}$$

eddy diffusivity m^2/s

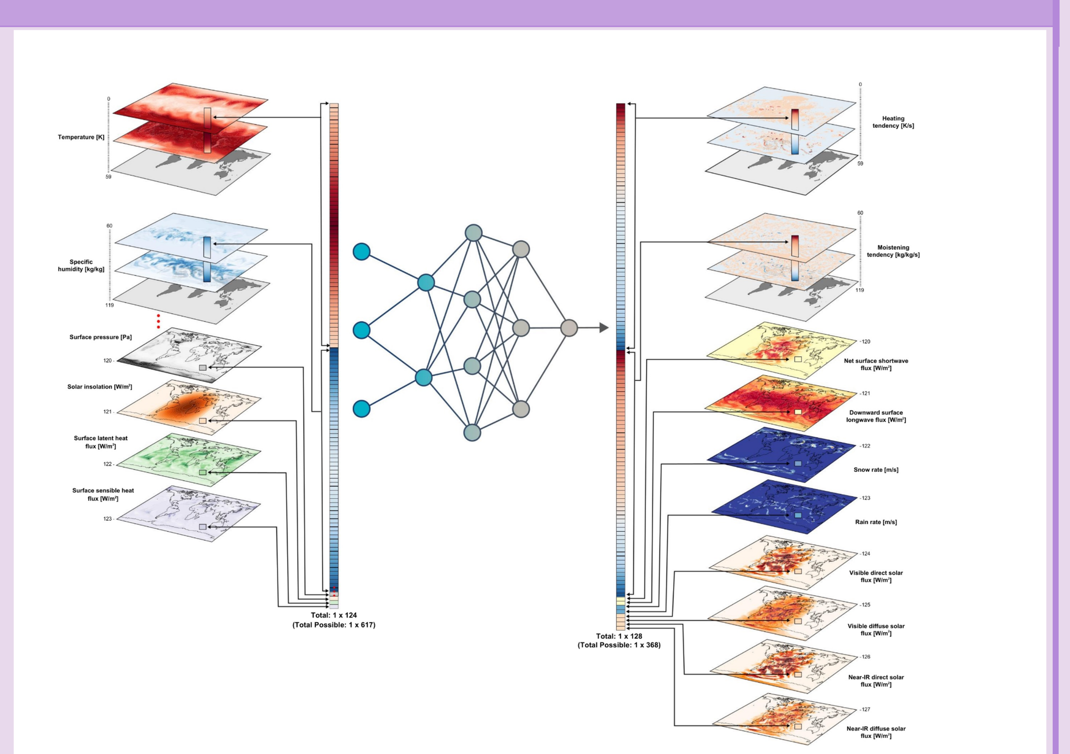
vertical gradient of moisture kg/kg/m

moisture flux m/s (or m/s * kg/kg)

- Improvement in **Eq. 1** aims to accurately capture this process without compromising computational efficiency.
- Objective: Apply equation discovery methods to improve calculation of moisture flux

2. Data and Methods

- Dataset: E3SM Climate Model Simulations
- Low-resolution, real-geography data over the equatorial Atlantic Ocean near the Caribbean Island.
- Lowest 3 levels near surface (723 meters and below)



ClimSim Dataset (Fig. 2)

- Time: July, year one

Eq. 2

$$w'q'(i) = -\frac{dq}{dt}(i)dz(i) - \left(\frac{dq}{dt}\right)(i-1)dz(i-1)\ldots - \left(\frac{dq}{dt}\right)(0)dz(0) + LHF$$

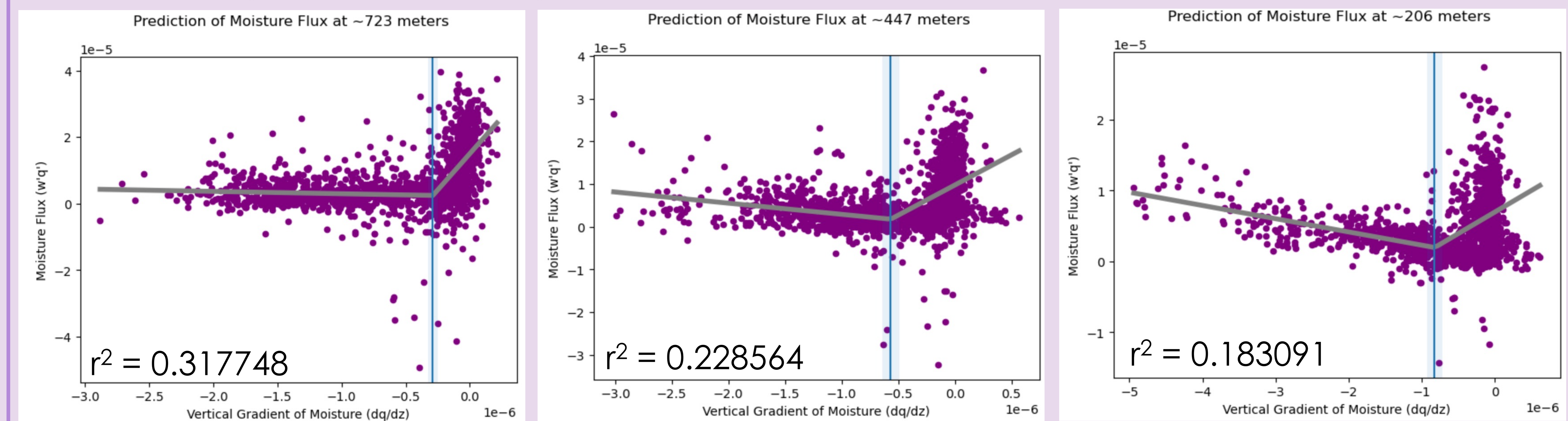
Eq. 3

$$dz = -\frac{R_d T_v}{g}dlnp$$

- Change in height dz and MF $w'q'$ are calculated using **Eq. 2-3**
- Methods: Symbolic Regression and Piecewise Regression
- Input: Vertical gradient of moisture $\frac{dq}{dz}$; Output: MF $w'q'$

3. Results

Piecewise Regression Model (Fig. 3-5)



Vertical line represents breaking point. Slanted lines represent model prediction. The model could not converge as accurately on a breaking point in the lower levels of the atmosphere.

PR Equation at 723 meters (Eq. 4)

$$y = \alpha x + c + \beta(x-\psi)H(x-\psi) + \varsigma$$

α = slope of seg. 1: -0.703693

c = y-int of seg.1: 2.21097×10^{-6}

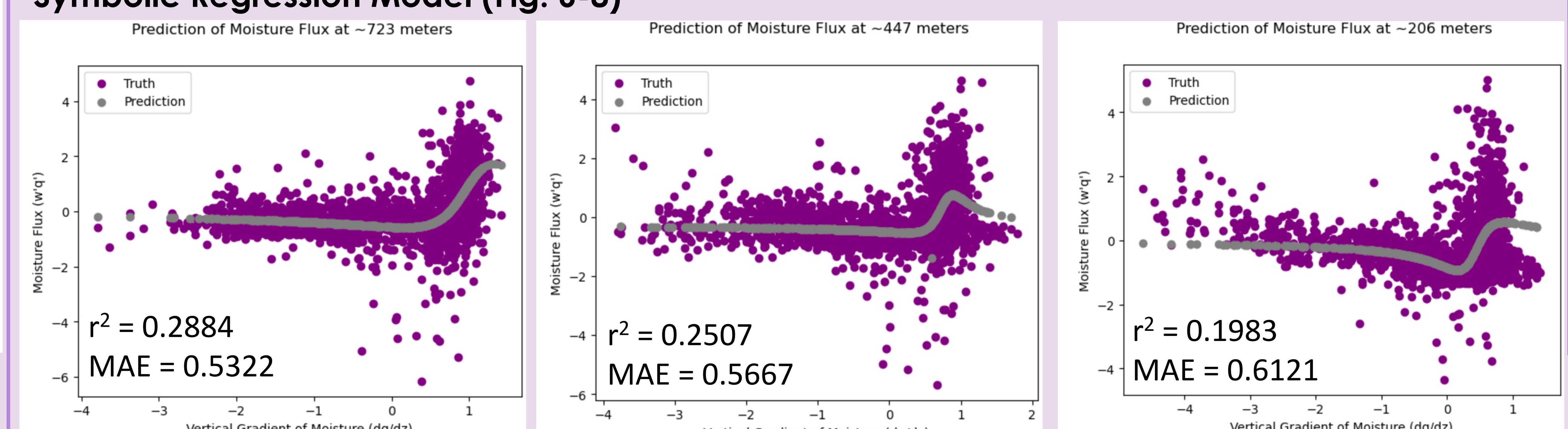
β = change in slope between segments: 44.3341

ψ = breakpoint value: 2.87898×10^{-7}

$H(x-\psi)$= heaviside step function: $H\left(\frac{dq}{dz} + 2.87898\times10^{-7}\right)$

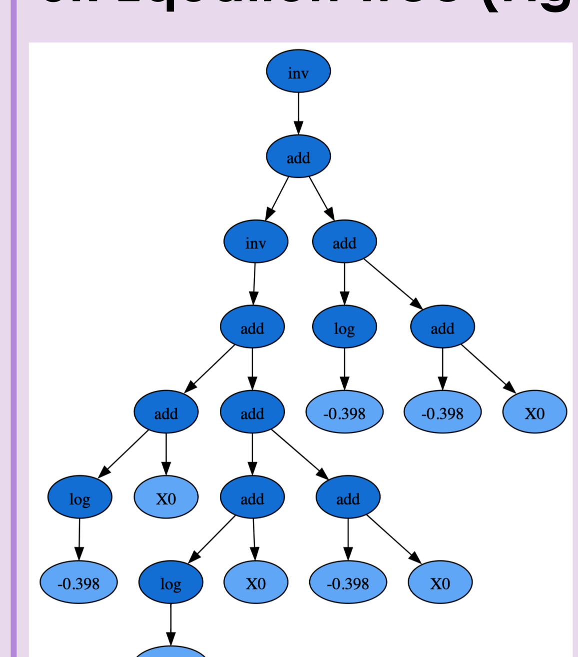
ς = noise

Symbolic Regression Model (Fig. 6-8)



Ran model for 45 generations with tournament size of 20. Mathematical operators were limited to addition, subtraction, multiplication, log, and inverse.

SR Equation Tree (Fig. 9)



SR Equation at 723 meters (Eq. 5)

$$w'q' = ((2\log(-0.398) - 0.398 + 3\frac{dq}{dz})^{-1} + \log(-0.398) - 0.398 + \frac{dq}{dz})^{-1}$$

Eq. 5 is the interpretation of the equation tree that is output by the SR model in **Fig. 9.**

4. Discussion

- Two equations for 723 meters high (**Eq. 4**, **5**) were discovered using the vertical gradient of moisture as input to PR and SR models.
- Due to the linear nature of the model, PR is unable to capture the shape of the data as well as SR.
- Both model predictions worsen as the height above the surface decreases.
- **Eq. 4** and **5** do not hold for other levels in the atmosphere.
- More accurate and interpretable equations than **Eq. 4** and **5** are necessary to make this study impactful.
- Hyperparameter tuning, adding more features such as temperature or pressure, and adding more function possibilities are all future steps that can be taken to improve model predictions.

Acknowledgements

Research Mentors:
Yu Huang (Columbia University)
Dave Lawrence (NCAR CGD)
Katie Dagon (NCAR CGD)
Linnia Hawkins (Columbia University)
Wayne Chuang (Columbia University)
Candace Agonafir (Columbia University)
Tian Zheng (Columbia University)
Writing Mentor:
Sharon Zhong (Michigan State University)

PIs on LEAP side:
Pierre Gentine
Mike Pritchard
Stephan Mandt
Sungduk Yu
Honorary Mentions~Jihae Moon, Sara Shamekh

This material is based upon work provided by the NSF STC: Center for Learning the Earth with Artificial Intelligence and Physics (LEAP) (Award #2019625) This material is based upon work supported by the National Science Foundation under Grant No. AGS-1641177 (SOARS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.