PROcess-guided deep learning and DAta driven modelling (PRODA)
From realistic representation to mechanistic understanding of global soil carbon

Feng Tao, Yuanyuan Huang, Bruce A. Hungate, Xingjie Lu, Toby D. Hocking, Umakant Mishra, Gustaf Hugelius, Xiaomeng Huang, and Yiqi Luo

Feng Tao
Department of Earth System Science, Tsinghua University
taof18@mails.tsinghua.edu.cn, @phxtao

Thursday, 25 February 2021
Background
SOC in the Earth System

• Biggest terrestrial C pool

Atmosphere: ~750Pg C

Vegetation: ~560Pg C

Surface Soil (0-1m): ~1,400Pg C

Deeper Soil (0-2m): ~2,500Pg C

Disolved OC: ~700Pg C

Disolved IC: ~38,000Pg C

Time Period: 2009 - 2018  Imbalance: 0.4Pg
(Data: Global Carbon Budget 2019)
Background
SOC in the Earth System

- Biggest terrestrial C pool
- Potential nature-based resolution to rising CO₂
Background

From Local Observations to Global Mechanistic Understanding

Mechanisms of SOC Storage

I: Photosynthesis & Input Allocation
II: Organic C Decomposition
III: Environmental Modifiers
IV: Vertical Transportation
V: Microbial Activities & Mineral Protection

(WoSIS 2019 Snapshot, Mishra & Hugelius 2020, Invalid)
Background

From Local Observations to Global Mechanistic Understanding

I: Photosynthesis & Input Allocation

II: Organic C Decomposition

III: Environmental Modifiers

IV: Vertical Transportation

V: Microbial Activities & Mineral Protection

Mechanisms of SOC Storage

Global Process Model

\[
\frac{dX(t)}{dt} = Bu(t) - A\xi(t)KX(t) - V(t)X(t)
\]

Local Observations
Background

From Local Observations to Global Mechanistic Understanding

Mechanisms of SOC Storage

I: Photosynthesis & Input Allocation
II: Organic C Decomposition
III: Environmental Modifiers
IV: Vertical Transportation
V: Microbial Activities & Mineral Protection

Global Process Model

\[
\frac{dX(t)}{dt} = Bu(t) - A\xi(t)KX(t) - V(t)X(t)
\]

Local Observations

Global SOC Understanding

Background

From Local Observations to Global Mechanistic Understanding

Mechanisms of SOC Storage

I: Photosynthesis & Input Allocation
II: Organic C Decomposition
III: Environmental Modifiers
IV: Vertical Transportation
V: Microbial Activities & Mineral Protection

Global Process Model

\[
\frac{dX(t)}{dt} = Bu(t) - A\xi(t)KX(t) - V(t)X(t)
\]

Local Observations

Global SOC Understanding

- WoSIS 2019 Snapshot
- Mishra & Hugelius 2020
- Invalid


Background

From Local Observations to Global Mechanistic Understanding

I: Photosynthesis & Input Allocation

II: Organic C Decomposition

III: Environmental Modifiers

IV: Vertical Transportation

V: Microbial Activities & Mineral Protection

(Jackson et al. 2017 AR)

Mechanisms of SOC Storage

Global Process Model

$$ \frac{dX(t)}{dt} = Bu(t) - A \xi(t) KX(t) - V(t)X(t) $$

Absent Direct Connection

Local Observations

Global SOC Understanding

kg C m$^{-2}$

10

30

100

WoSIS 2019 Snapshot
Mishra & Hugelius 2020
Invalid

Tsinghua University
Background
From Local Observations to Global Mechanistic Understanding

Current: highly biased representation

Future: highly uncertain projection

(Luo et al. GCB, 2015)

IPCC AR5, 2013
Background
From Local Observations to Global Mechanistic Understanding

Current: highly biased representation
Future: highly uncertain projection

How to incorporate large local observations and complex ESMs to best understand global soil carbon cycle?

(Luo et al. GCB, 2015)
PRODA: PROcess-guided deep learning and DAta-driven modelling

1. Process Model
   - Meteorological Data
   - Big Observational Data

2. Deep Learning Algorithm

3. Prediction

(Tao et al. 2020 Frontiers in Big Data)
PRODA: PROcess-guided deep learning and DAta-driven modelling

Process-guided Model: CLM5

Meteorological Data

\[
\frac{dX(t)}{dt} = I(t) - A\xi(t)KX(t) - V(t)X(t)
\]

Five categories of processes
- Input allocation
- Decomposition rate
- Microbial CUE
- Vertical transportation
- Environmental modifiers

(Huang et al. GCB, 2017)

(Tao et al. 2020 Frontiers in Big Data)
PRODA: PROcess-guided deep learning and DAta-driven modelling

Big Observational Data

\[
\frac{dX(t)}{dt} = I(t) - A\xi(t)KX(t) - V(t)X(t)
\]

Meteorological Data

> 50,000 Vertical Soil Profiles

Site level data assimilation

Optimised parameter values

(Tao et al. 2020 Frontiers in Big Data)
PRODA: PROcess-guided deep learning and DAta-driven modelling

Deep Learning Algorithm

\[ \frac{dX(t)}{dt} = I(t) - A\xi(t)KX(t) - V(t)X(t) \]

Meteorological Data → Site level data assimilation → Optimised parameter values → Deep Learning → Prediction

> 50,000 Vertical Soil Profiles

Process model

Optimisation method

Optimisation target

60 Env. Covariates

(Tao et al. 2020 Frontiers in Big Data)
PRODA Performance
Improved SOC Representation

(Tao et al. 2021 in prep.)
**Big Data Retrievals**

**SOC Stock, Distributions and Turnover**

**Stock Total**

- 56±2% spatial variation of observations (11% by default setting)
- Whole depth (0 - 840cm): 2404±105 Pg C

**Turnover total**

- 0 - 100cm: 1322±75 Pg C

**Global SOC Stock and Distribution**

- Mean: 112±10 years
- Median: 44±2 years

(Tao et al. 2021 in prep.)
Big Data Retrievals
Key Mechanisms underlying SOC

Pervasive spatial variation of mechanisms

- Roots reach deeper soil in tropics than in boreal regions
- Higher baseline decomposition in boreal regions than tropics
- Env. modifiers increases from boreal regions to tropics
- CUE in boreal regions is higher than tropics

(Tao et al. 2021 in prep.)
Which Mechanism Determines SOC Storage?
CUE as the Main Regulator

How will global SOC change with different mechanisms?

Global SOC stock is most sensitive to CUE instead of carbon input or its allocation processes

(Tao et al. 2021 in prep.)
Which Mechanism Determines SOC Storage?

CUE as the Main Regulator

How will global SOC deviates from reality when ignoring mechanism’s spatial variability?

Model simulation deviates from observations most when ignoring CUE’s spatial variability EVEN IF all others’ spatial dependences are counted

(Tao et al. 2021 in prep.)
Environmental Depended Mechanisms

Importance of Environmental Variables

- **Soil texture** (i.e., clay, silt, sand content, etc.) is the most important feature in regulating processes.

- **Climate variables** (i.e., temperature, precipitation) are equally important with texture for environmental scalars and vertical transportation.

(Tao et al. 2021 in prep.)
PRODA strikes a balance between process modelling & data mining to best represent AND understand global SOC.

**Take-Home Message**

PRODA strikes a balance between process modelling & data mining to best represent AND understand global SOC.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Method</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Modeling</td>
<td>Met Data → Model Given Structure &amp; Parameters</td>
<td>![Graph 1]</td>
</tr>
<tr>
<td>Data-driven Modeling</td>
<td>Met Data → Model → Data Assimilation → Selected Structure &amp; Trained Parameters</td>
<td>![Graph 2]</td>
</tr>
<tr>
<td>PRODA</td>
<td>Met Data → Model → Input Variables → Neural Network → Optimised Parameters</td>
<td>![Graph 3]</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>Input Variables → Neural Network</td>
<td>![Graph 4]</td>
</tr>
</tbody>
</table>

**Processes Understanding**

**Data Mining**

**Processes**

**Understanding**
Take-Home Message

Environment dependent mechanisms & Microbial centred SOC stabilisation

Carbon Input

- Mechanism value (constant)
- Mechanism value (varied)
- Impact of mechanisms
- Information
- Carbon Flow

Input Allocation

Litter C

- Physical Env.

Decomposition

Carbon Flow

CO₂

Stabilised SOC

Carbon Input

- Input Allocation

Litter C

- Physical Env.

Vegetation
- Climate
- Geography

Soil Texture
- Soil Chemical properties

CUE

Decomposition

CO₂

Stabilised SOC
THANKS!

QUESTIONS TIME