Pacific Northwest
NATIONAL LABORATORY

PNNL-38409

Integration and Quantitative Comparison of Up-Scaled Molecular Observation Network Data with Existing Soil Databases

September 2025

1 Kendalynn A Morris
2 Nathan J Wiens

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38409

Integration and Quantitative Comparison of Up-Scaled Molecular Observation Network Data with Existing Soil Databases

September 2025

1 Kendalynn A Morris
2 Nathan J Wiens

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

MONet provides novel soil molecular data to the research community for understanding biogeochemical processes and complementing other soil datasets. This study assesses MONet's ability to replicate known soil patterns via a comparative analysis of soil respiration (Rs), pH, and clay content against benchmark datasets. Results show moderate agreement for pH and clay content, highlighting MONet's strengths in capturing soil biogeochemical variation in underrepresented regions like urban areas. Rs data are marked by the appropriate trends relative to other datasets, but direct comparison is impractical due to methodological differences in underlying data. Strategic sampling is recommended to improve MONet's coverage and eventual utility in bridging molecular observations with global datasets.

Summary

Herein we evaluate MONet soil properties in comparison with SoilGrids and SRDB to explore its capacity for representing known soil properties. Our key findings include:

- Moderate agreement in pH and clay content comparisons, with R^2 values of 0.48–0.53 for pH and 0.62–0.66 for clay content.

- Strengths of MONet include its reliance on physical samples to capture unique trends (e.g., urban soils) and its potential to illuminate gaps in global coverage.

- Limitations include known methodological challenges in comparing soil respiration measurements, reflecting the challenges of comparing controlled lab-based metrics (MONet) with broad field-based observations (SRDB).

We emphasize the importance of expanding MONet sampling in precipitation transition zones (mean annual precipitation of 500–1000 mm) and enhancing methodological coherence for Rs measurements, potentially through collaboration with established site networks such as NEON. Ultimately, MONet is positioned as a valuable evolving tool for scaling molecular processes to global assessments

Acknowledgments

This research was supported by the Earth and Biological Sciences Directorate (EBSD) Mission Seed, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

The Molecular Observatory Network (MONet) is an effort led by the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the DOE's Office of Biological and Environmental Research and located at PNNL.

Acronyms and Abbreviations

EMSL	Environmental Molecular Science Laboratory
FAO	Food and Agriculture Organization of the United Nations
IIASA	International Institute for Applied Systems Analysis
MAP	mean annual precipitation
MAT	mean annual temperature
MONet	Molecular Observatory Network
Rs	soil respiration
SRDB	soil respiration database

Contents

Figures

1.0 Introduction

The Molecular Observatory Network (hereafter, MONet) is an initiative from the Environmental Molecular Science Laboratory (EMSL), one of the Department of Energy's Office of Science user facilities (Bowman et al. 2023). MONet addresses the inherent challenges of consistently collecting molecular-level soil data across diverse ecosystems and large geographic scales, providing researchers a transformative tool to potentially uncover patterns that were previously inaccessible. With its focus on the hidden molecular dynamics within soil, MONet seeks to address urgent environmental challenges, including climate change, ecosystem resilience, and sustainable land management. As all research builds on previous work, achieving this goal requires integrating and validating MONet data with and against existing databases of soil characteristics.

Soil has been a focal point for scientific study for as long as recorded history. Modern databases range from parameter-specific and individually curated efforts, such as the Soil Respiration Database (SRDB(Jian et al. 2021)), to machine learning-driven data products like SoilGrids (Hengl et al. 2017), which provide gridded global values across a suite of soil parameters. Other notable sources include international organizations, such as the Food and Agriculture Organization (FAO) with its Global Soil Information System and Harmonized World Soil Database (FAO and IIASA 2023), as well as government agencies like the USDA with SSURGO and gSSURGO databases (Soil Survey Staff 2014)) and USGS soil datasets, such as CONUS-wide mineralogy (Smith et al. 2014). Due to historical disciplinary silos, each of these databases reflects the specific needs and priorities of distinct user communities, leaving gaps in interoperability and synthesis across research efforts.

Despite MONet's potential to transform soil research with molecular data, it remains relatively unknown outside the EMSL and National Lab community. Expanding its impact will require public-facing tools and resources that emphasize open science principles, making MONet accessible and usable for a broad range of soil data user communities. To support this effort, we set out to assess MONet's current dataset through response functions and upscaling with globally gridded soil data and present this analysis in a publically available Rmarkdown. This work will lay the foundation for tools that enhance MONet's usability and foster the large-scale Earth system science and ecosystem ecology research necessary to improve molecular-to-global understanding. As MONet is still a relatively small dataset, validating its ability to replicate known patterns in soil properties can not only bolster confidence in its utility but also help identify gaps where additional data collection efforts should be prioritized.

2.0 Materials and Methods

2.1 Data

The current MONet dataset is compact and manageable in size, making it easy to work with and integrate into data-driven analyses. To evaluate MONet's ability to replicate known patterns in soil properties, we compared aspects of it with two globally recognized soil datasets: SRDB v5 and SoilGrids 2.0. These datasets offer complementary insights into soil parameters at different scales and resolutions, serving as benchmarks for testing MONet's observations.

SRDB v5 (Jian et al. 2021) and an associated gridded product of global 1-km resolution estimates of total soil respiration and its heterotrophic component (2021), see also (Hashimoto, Ito, and Nishina 2023). The gridded data is generated by combining SRDB observations with mean annual temperature, seasonal precipitation, vegetation cover data, and modelled via a quantile regression forest approach. This dataset's straightforward structure makes it relatively easy to integrate with MONet data (Figure 1).

SoilGrids 2.0 (Poggio et al. 2021) provides global maps of soil properties, including pH, clay content, organic carbon, and more, using machine learning models. These models are driven by soil profile observations sourced from the WoSIS database and other Earth observation data products (e.g., climate and vegetation covariates). With a spatial resolution of 250 meters, SoilGrids is one of the highest-resolution global datasets available. However, its large size makes processing computationally intensive, even for regional scale analyses.

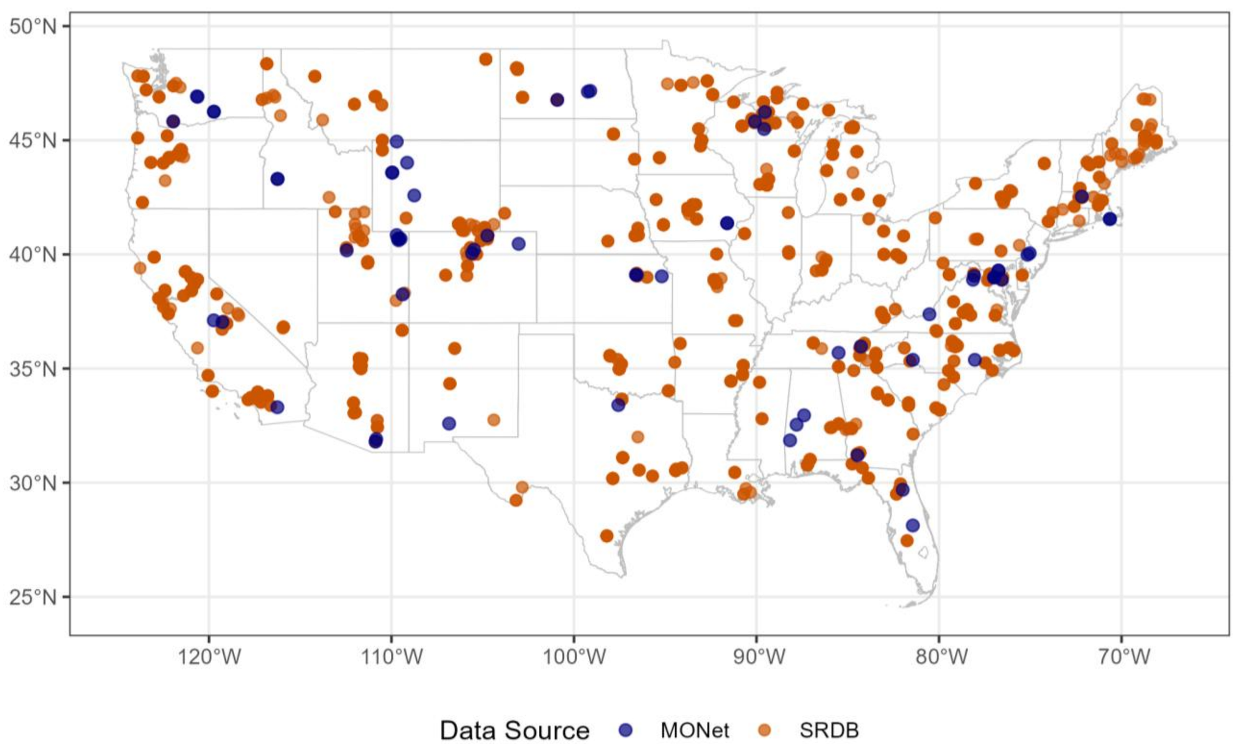


Figure 1. Location of MONet and SRDB sample sites in CONUS

2.2 Parameters

To capture representative features of soil without being exhaustive, soil respiration, pH, and clay content were selected for analysis. Soil respiration (Rs) captures aspects of biological activity in soil over variable time scales. It serves as a critical indicator for microbial processes and carbon fluxes influenced by both environmental gradients and molecular-scale interactions. Soil pH, acts as a proxy for biochemical interactions, particularly with abiotic factors such as soil mineralogy. As a fundamental property of soil chemistry, it provides valuable context for interpreting molecular observations. Clay content is included to represent soil's geological weathering and its impact on structure, water retention, and molecular adsorption.

Quantitative comparisons of theses parameters were as follows: (1) examining soil respiration relative to mean annual precipitation (MAP) and mean annual temperature (MAT), (2) assessing soil pH within Köppen climatic zones (Figure 2), and (3) evaluating the spatial distribution of soil clay content. We also completed direct comparisons between the data sources using linear regression. The original intent was to compare spatially upscaled MONet clay content to SoilGrids data, however detailed reading into existing spatial upscaling tools revealed that tasks to be beyond the resources currently available. Instead, we provide a visual comparison of the two datasets as well as an overlay of MONet clay content relative to a potential use-case, and current downstream data product of SoilGrids, the clay content of CONUS land-use regions of the Global Change Assessment Model (Bond-Lamberty et al. 2023).

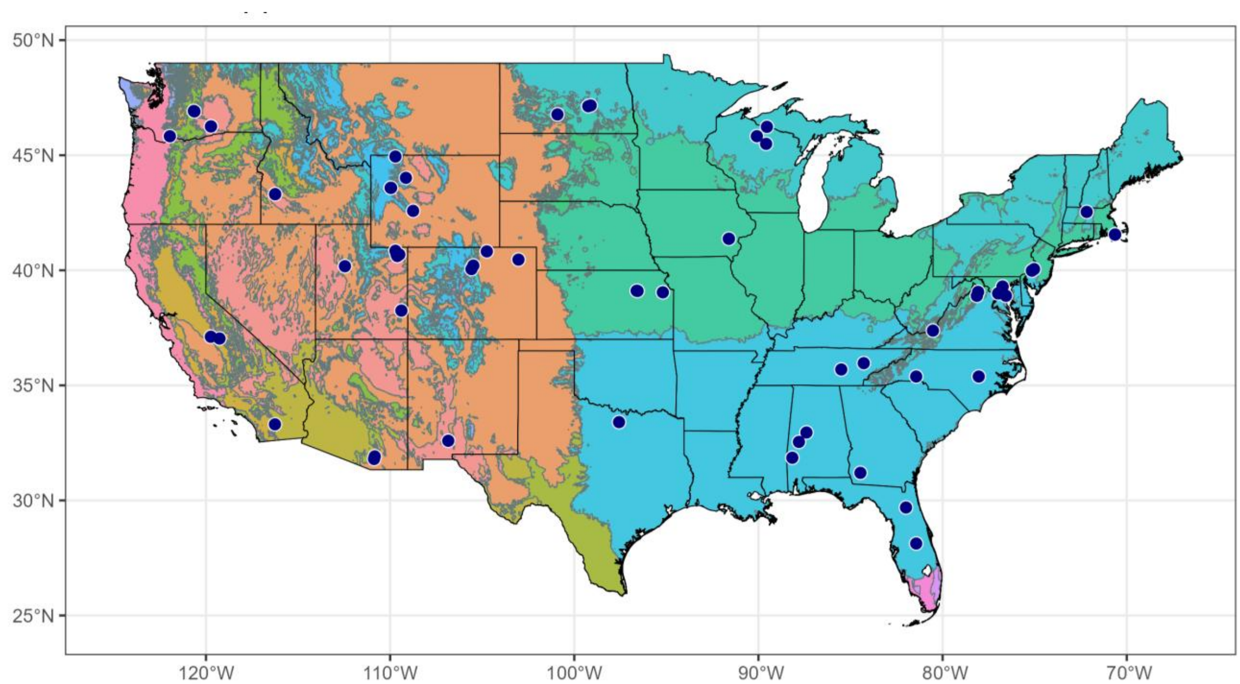


Figure 2. Köppen Climate Zones in CONUS with MONet sample locations

SRDB and derived data products report respiration fluxes on a spatial scale, while MONet reports this flux on the basis of soil mass. Additionally, MONet contains soil respiration data for the top 10 cm of a 30 cm deep soil core, as well as the bottom 10 cm. We calculated an index to integrate measurements of these sections as well as across two time points (24 hr and 96 hr), weighting the top core more heavily (65%) due to higher biological activity, and assigning greater importance to longer-term respiration rates (96 hours weighted at 66.7%) compared to shorter-term measurements (24 hours weighted at 33.3%). This approach provides a single representative value for each core, capturing dynamic microbial activity over time. For SRDB data, spatially aggregated soil respiration rates were converted into mass-based fluxes (mg CO2-C per gram of soil per day) using fixed assumptions about bulk density (1.33 g/cm^3), a soil depth of 10 cm, and appropriate area and temporal scaling factors. This allowed for consistent comparisons between MONet's experimental values and SRDB-derived data across climatic gradient analyses (FIGURE X). For comparisons with SoilGrids data, we used the 0 to 5 and 15 to 30 cm data products for comparison to the top and bottom core sections respectively.

2.3 User Orientation and R Markdown

MONet data was manually downloaded from two sources: EMSL's web data portal and the One Thousand Soils pilot project's Zenodo repository, links for both below. Shapefiles for down selecting data to the continuous United States, the spatial extent of MONet data, were sourced from the National Oceanic and Atmospheric Administration (National Weather Service 2018) and the Commission for Environmental Cooperation (Beck et al. 2018). All other data are downloaded using R code for maximum reproducibility.

MONet data:
https://shinyproxy.emsl.pnnl.gov/app/1000soils
https://zenodo.org/records/15328215

Shapefiles:
https://www.cec.org/north-american-environmental-atlas/climate-zones-of-north-america
https://www.weather.gov/gis/USStates

All analyses are conducted in the programming language R (R Core Team 2022) and fully reproducible using publicly available code. Users should refer to github documentation and visual aids (Figure 3) to get started.

```
.
├── MONetSynthesis.Rproj
├── Morris-Wiens_MONetSynthesis
├── R/
│   └── MONetDataPreprocessing.R
├── data/
│   ├── GCAM/
│   │   └── mapped_clay_KN.csv
│   ├── MONet/
│   │   ├── 1000S_processed_L2_summary.csv
│   │   ├── 1000Soils_Metadata_Site_Mastersheet_v1.csv
│   │   ├── pH/
│   │   │   └── processed_data/
│   │   │       ├── Column_Descriptions.xlsx
│   │   │       ├── Coordinates.csv
│   │   │       └── Soil_BioChemical_properties.csv
│   │   └── clay/
│   │       └── processed_data/
│   │           ├── Column_Descroptions.xlsx
│   │           ├── Coordinates.csv
│   │           └── Soil_BioChemical_properties.csv
│   ├── shapefiles/
│   │   ├── gcam_boundaries_moirai_3p1_0p5arcmin_wgs84/
│   │   │   └── main_outputs/
│   │   │       └── glu_boundaries_moirai_landcells_3p1_0p5arcmin.shp
│   │   ├── s_18_mr25/
│   │   │   └── s_18mr25.shp
│   │   └── na_climatezones_shapefile/
│   │       └── climatezones_shapefile/
│   │           └── NA_ClimateZones/
│   │               └── data/
│   │                   └── NorthAmerica_Climate_Zones.shp
│   ├── soilgrids*/
│   │   ├── crop_roi_igh_clay_0-5cm.tif
│   │   ├── crop_roi_igh_clay_15-30cm.tif
│   │   ├── crop_roi_igh_ph_0-5cm.tif
│   │   └── crop_roi_igh_ph_15-30cm.tif
│   ├── SoilResp_HeterotrophicResp_1928/
│   │   └── data/
│   │       └── soil_Rh_mean.tif
│   ├── srdb*/
│   │   └── srdb-20250503a/
│   │       └── srdb-data.csv
│   └── worldclim_data*/
│       └── climate/
│           └── wc2.1_10m/
│               └── wc2.1_10m_prec_xx.tif
├── R_data*/
│   ├── processed_Rs.RData
│   ├── processed_pH.RData
│   └── processed_clay.RData
├── figures
└── README.md
```

Figure 3. R project file tree

3.0 Results and Discussion

The comparative analysis of MONet data with SoilGrids and SRDB gridded data highlights notable strengths and limitations in capturing soil properties across datasets. Rs comparisons proved impractical due to fundamental methodological differences between MONet's laboratory measurements and field-based soil respiration estimates from SRDB. When converted to the same units (see markdown and Appendix A for more detail), values for the two datasets differed by 4 to 5 orders of magnitude. However, when both datasets were standardized using Z—scores, the overall pattern in relation to MAP and MAT were generally similar (Figure 4). Though there was better agreement between the relationship between MAP and Rs than for MAT.

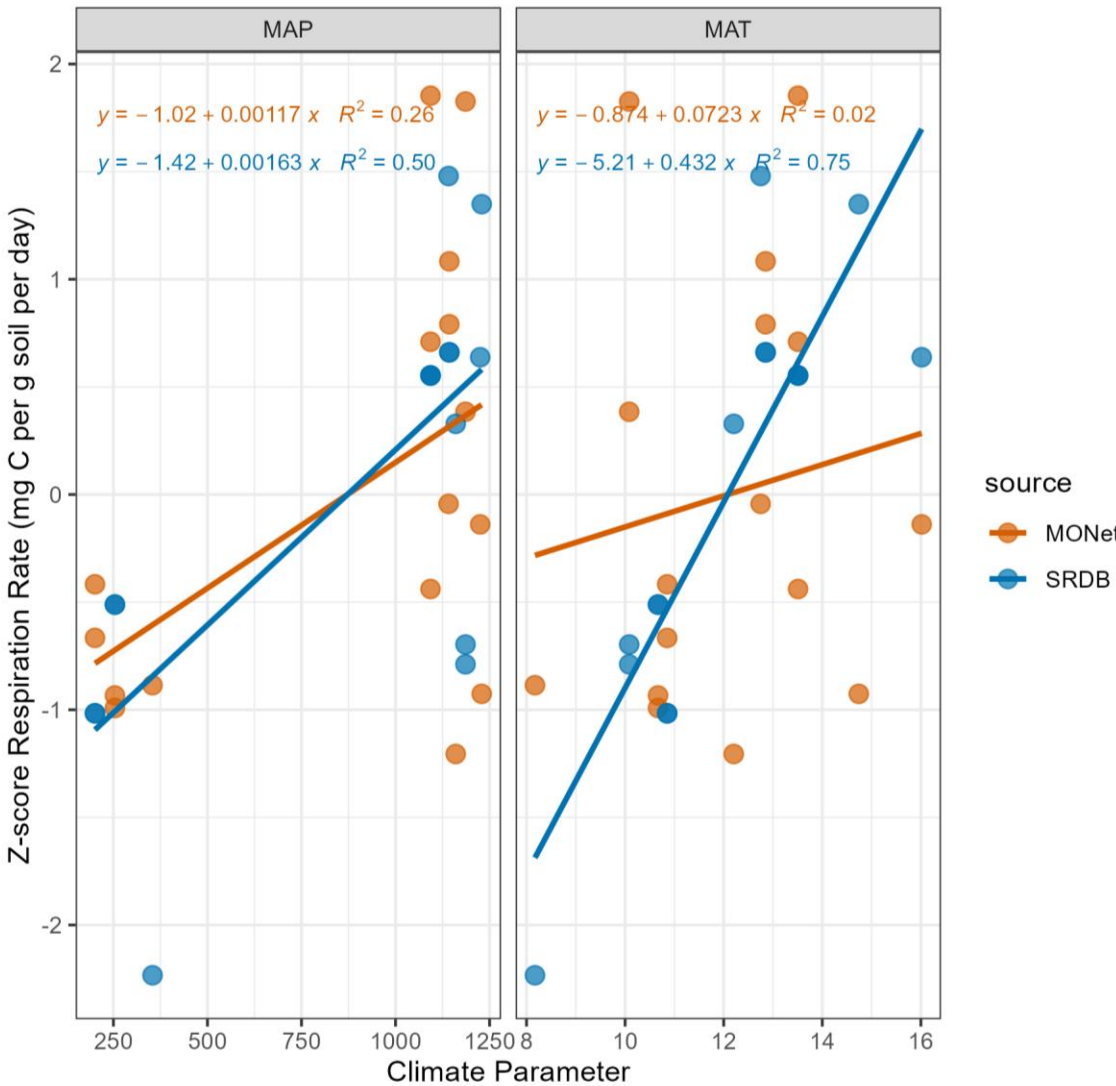


Figure 4. Z-score of respiration rate of MONet and SRDB samples in relation to MAP and MAT

For pH, the coefficient of determination (Figure 5, R^2) was 0.48 for bottom cores in our one to one comparison and 0.53 for top, suggesting moderate agreement with SoilGrids. However, MONet physical samples tend towards higher pH values at the lower, more acidic, end of the range and towards higher values at the alkaline extreme, reflecting challenges in aligning observations with predictions from broader-scale datasets. For the one to one comparison with clay content (Figure 6), R^2 values were stronger at 0.62 for bottom cores and 0.66 for top cores, but similar patterns of overestimation at low values and underestimation at higher values were observed, particularly for bottom layers.

Overall, the comparison underscores the variability in agreement across parameters, with clay content showing slightly stronger alignment than pH. Despite its strengths, MONet's relatively small sample size limits its ability to perfectly capture spatial heterogeneity, emphasizing the need for cautious interpretation of broader trends.

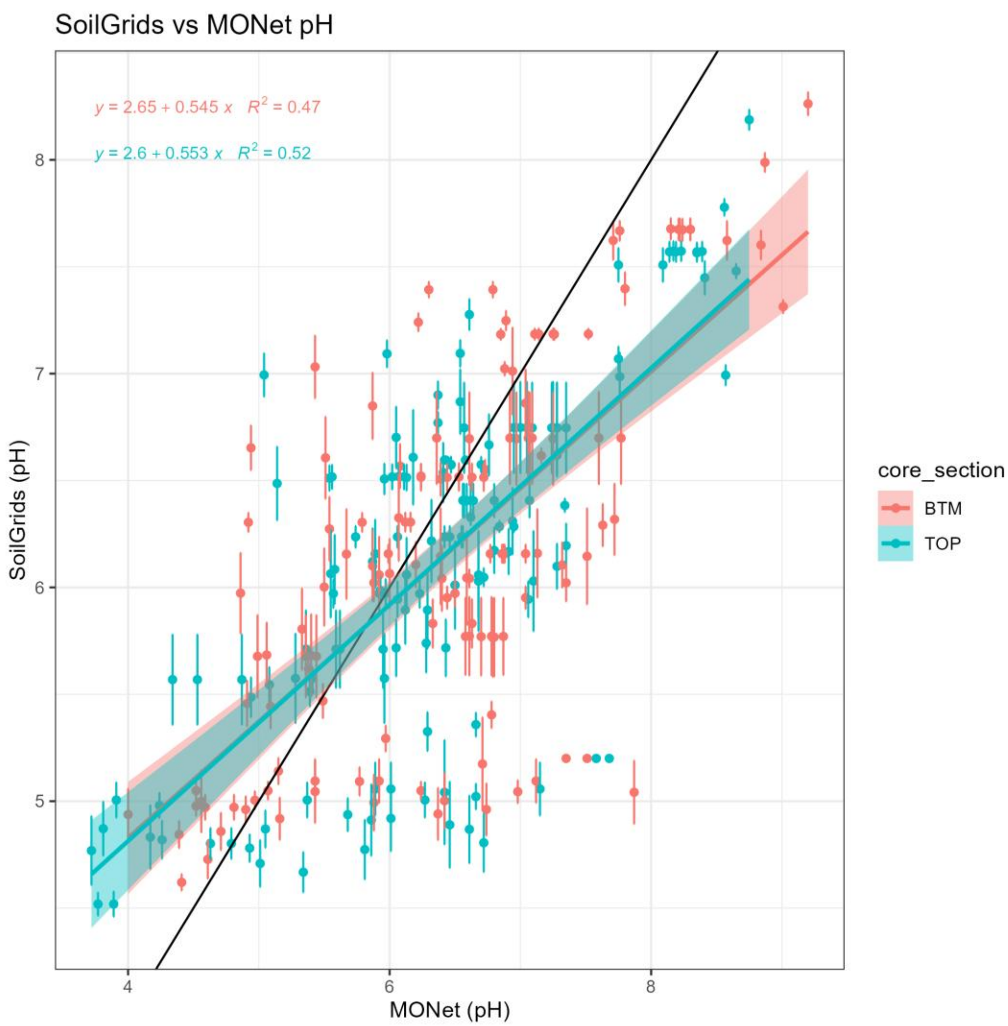


Figure 5. One-to-one comparison of pH with 1000m buffer confidence intervals

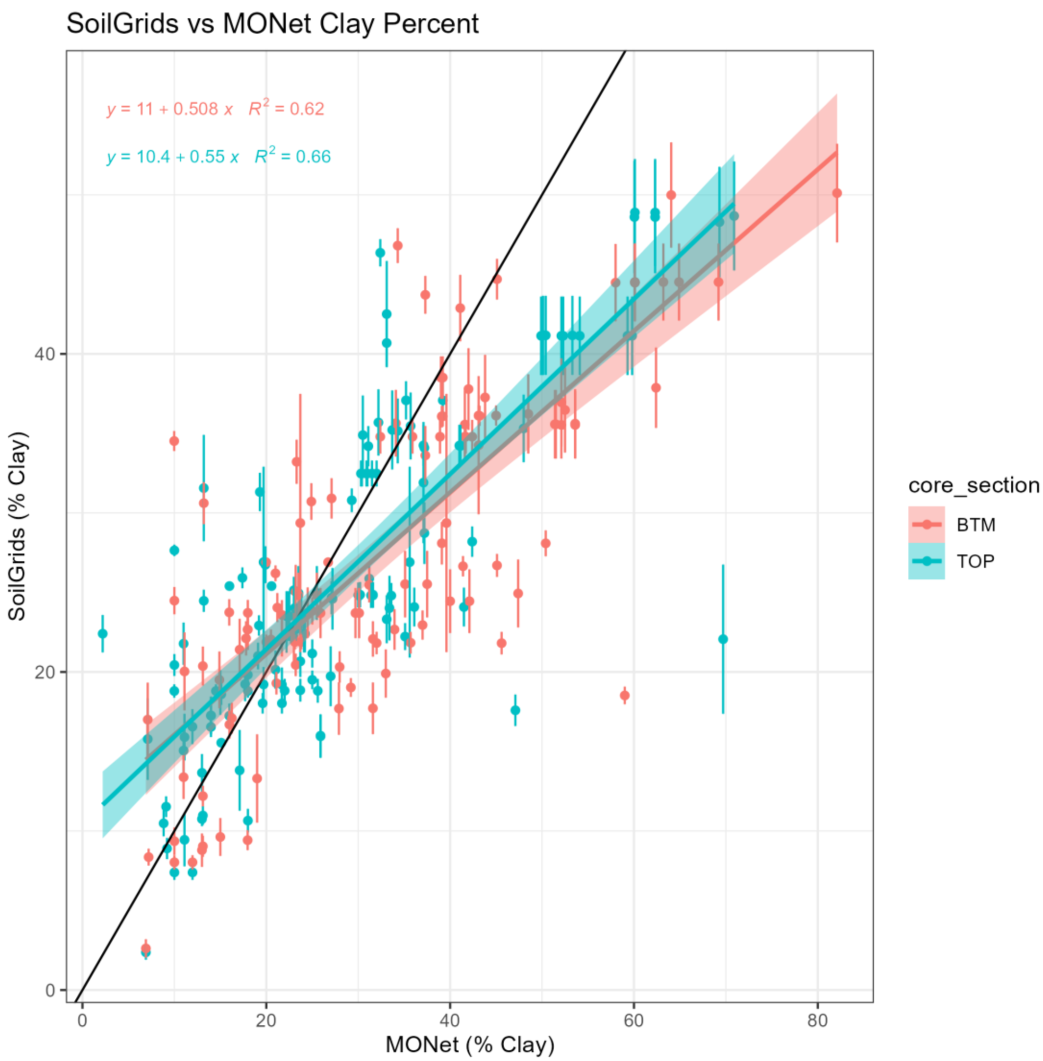


Figure 6. One-to-one comparison of clay with 1000m buffer confidence intervals

3.1 Strengths of MONet

While MONet pH data generally agree with that from SoilGrids, there were specific climate zones, generally with low sample volume, where pH values were somewhat divergent (Figure 7). Humid continental climate with year-round precipitation, which make of most of the land area of CONUS, tended to have lower pH values in SoilGrids, than what was represented in MONet by physical samples, but mean values were within a single pH unit.

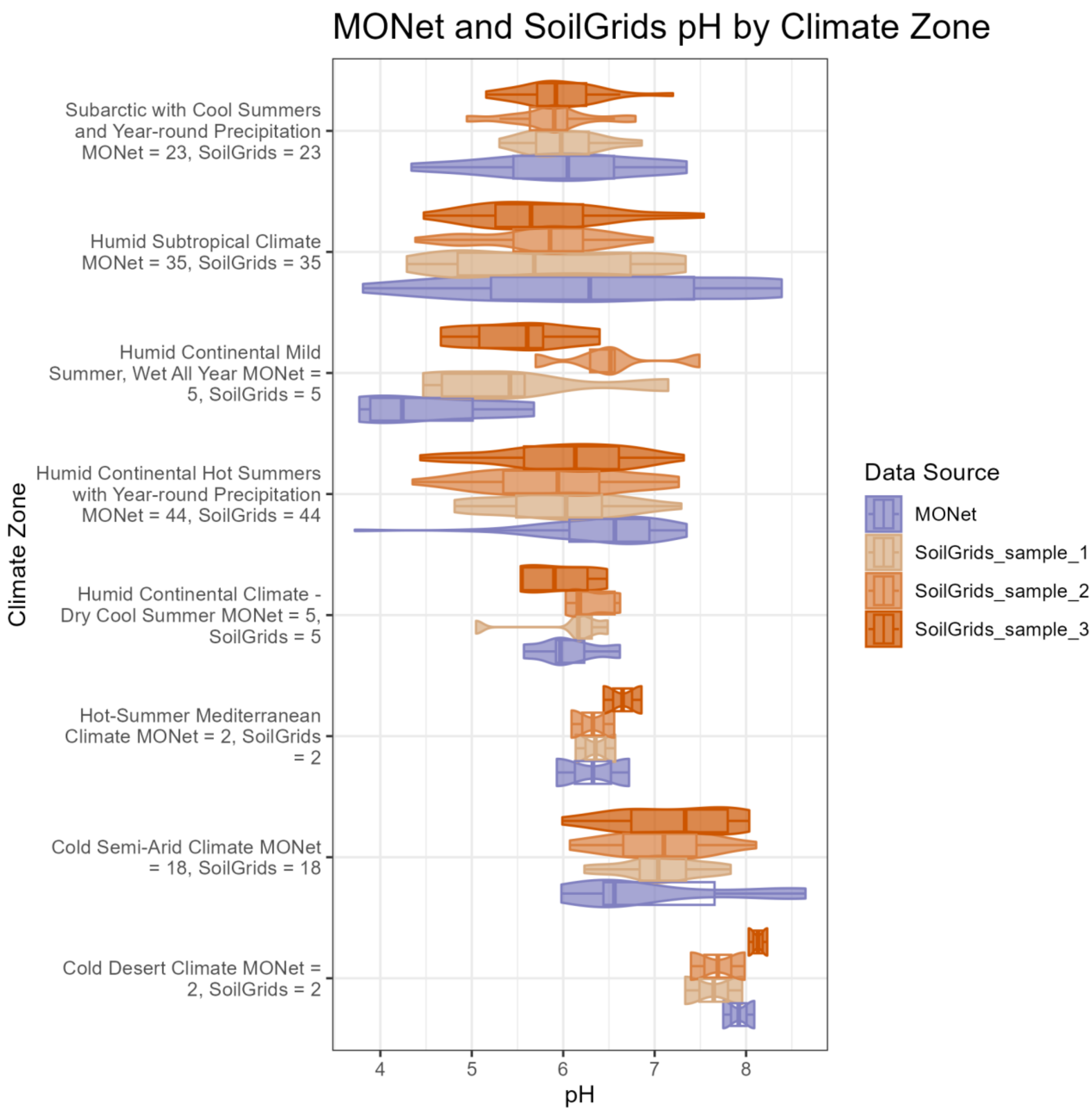


Figure 7. pH distribution by climate zone compared to random samples of SoilGrids

Its foundation on physical soil samples is a distinct advantage of MONet, which allows for sampling in regions excluded from SoilGrids, such as urban areas (Figure 8). This was exemplified by the identification of six MONet sample sites—three in College Park, MD, two in Baltimore, MD, and one in Philadelphia, PA—which can provide valuable insights into urban soil properties. These areas, excluded from SoilGrids' global models, highlight MONet's ability to expand the soil biogeochemistry dataset into underrepresented regions.

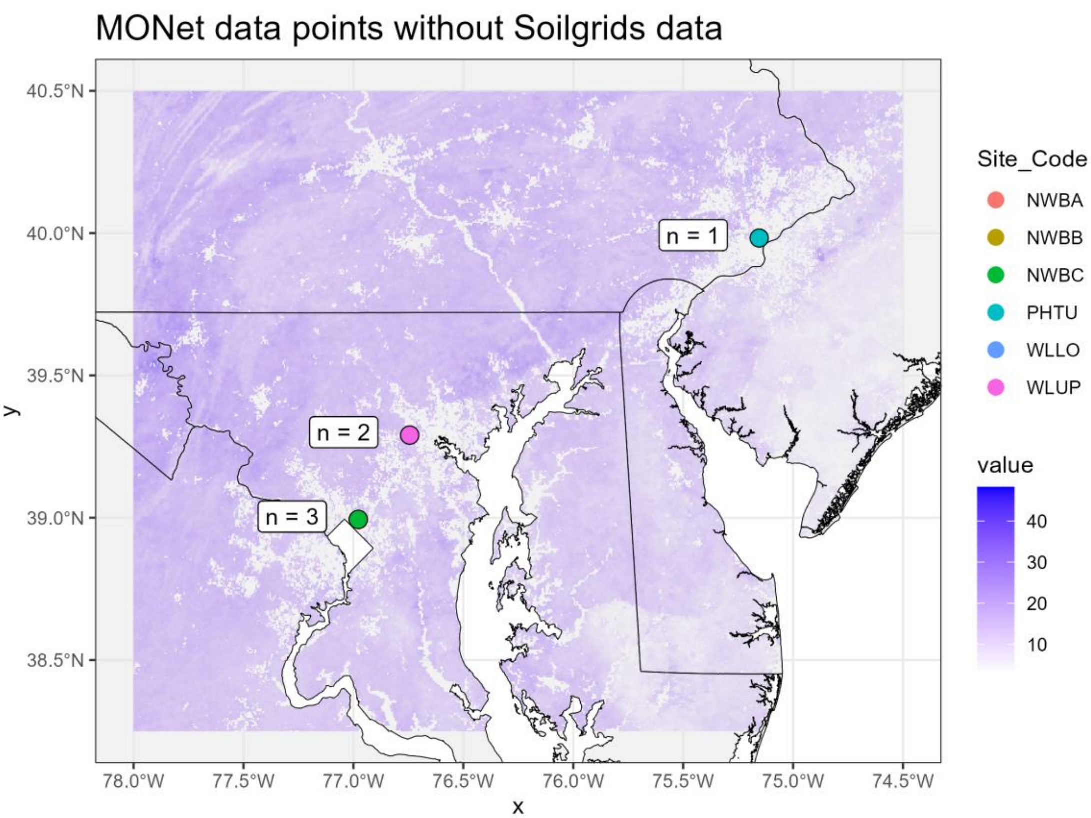


Figure 8. MONet sample locations without a corresponding value in SoilGrids

3.2 Areas for Improvement

For an applied-science comparison of MONet's ability to capture known spatial patterns in clay distribution, we pulled clay content from GCAM's CONUS land-use basins, values that are derived from spatial aggregation of SoilGrids (Figure 9). This spatial comparison shows that the low-end overestimation and high-end under estimation of SoilGrids' predictions relative to the physical samples of MONet, seen in the direct comparison (Figure 6) is exaggerated in this downstream data product (Figure 10). These inconsistencies highlight the need for careful consideration in cross-dataset interpretations and the importance of targeted validation for improving soil property accuracy across regions.

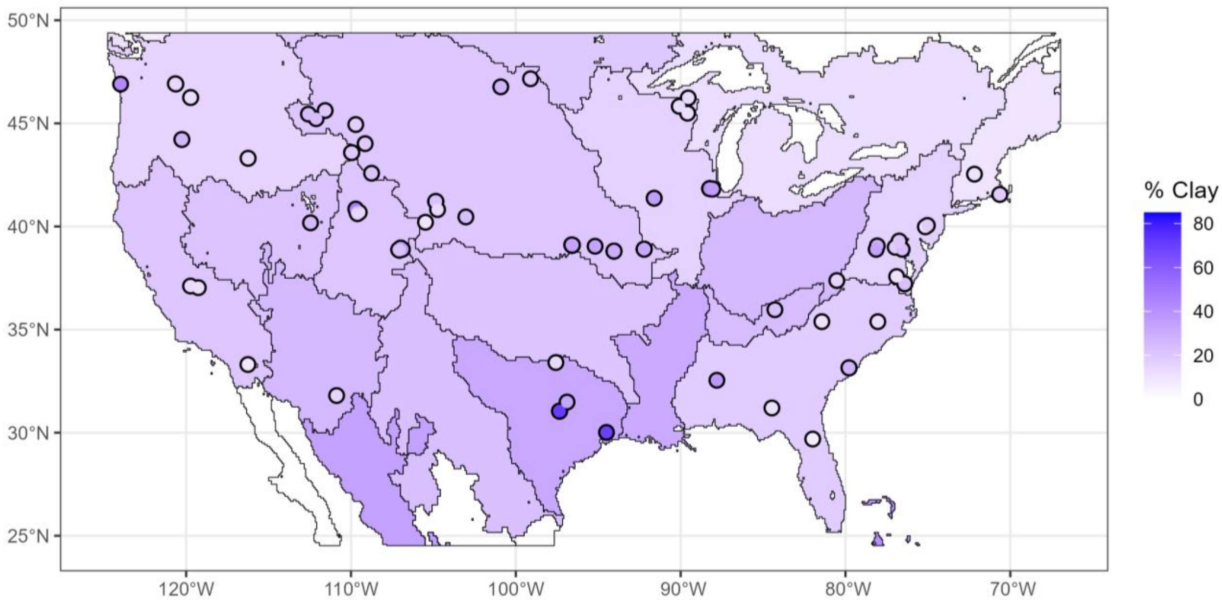


Figure 9. Spatial comparison of clay content at MONet site to GCAM basin clay content

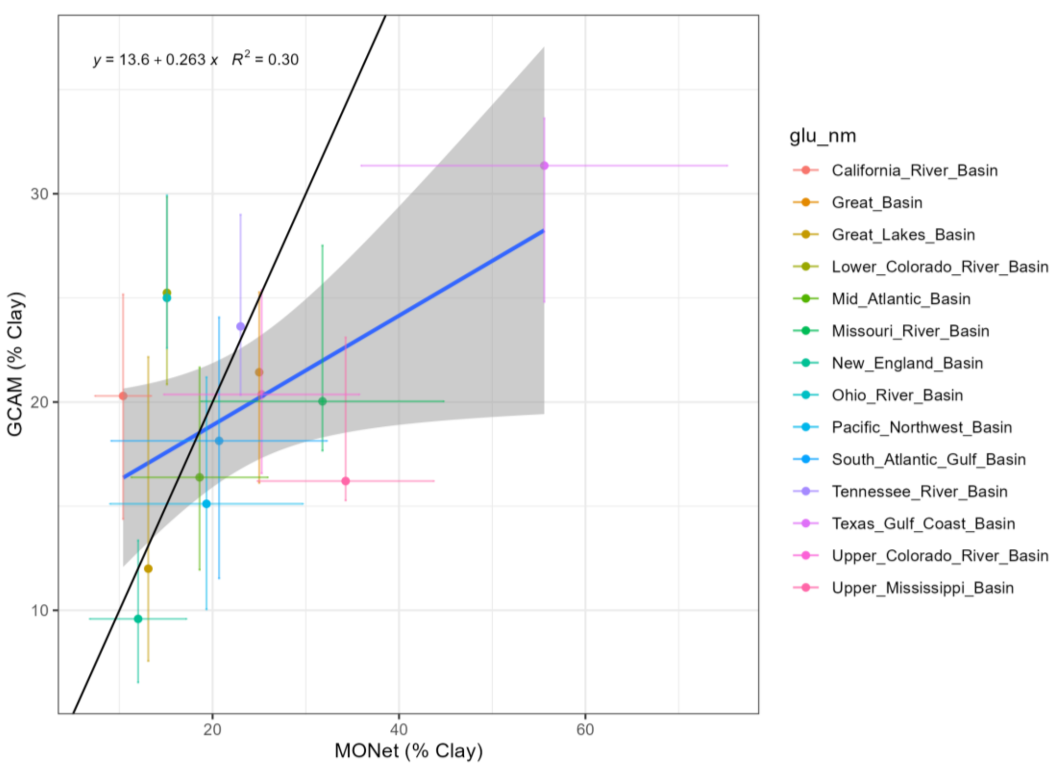


Figure 10. One-to-one comparison of GCAM CONUS basins (glu-nm) vs MONet clay content with confidence intervals

Current gaps within specific Köppen climate zones, such as hot deserts and temperate oceanic climates, highlight the need for expanded sampling efforts to improve representation across diverse environmental gradients. Notably, MONet Rs values show a substantial gap in regions with moderate rainfall (MAP 500–1000 mm/year), particularly transitional areas like the Great Plains or Intermountain West. Targeted sampling in these regions, characterized by temperate grasslands and semi-arid ecosystems, could close critical data gaps in soil biogeochemistry.

Laboratory-based MONet respiration measurements isolate heterotrophic components under control conditions, whereas field-based measurements are influenced by real-world environmental gradients. These methodological differences limit direct comparisons but underscore complementary value (Patel et al. 2022): MONet data excels in controlled mechanistic studies, while SRDB informs large-scale ecological trends. Future efforts should systematically compare the biases and strengths of each method to better align data integration.

3.3 Conclusion

MONet has shown promise in capturing key soil properties, including pH and clay content, with moderate agreement to global datasets like SoilGrids. Its ability to provide insights into underrepresented regions (e.g., urban soils) reinforces its unique contribution to soil biogeochemistry. However, key areas for improvement remain, including expanding sampling coverage within climate zones and addressing scale mismatches in soil respiration measurements. With strategic enhancements and interdisciplinary collaborations, MONet can evolve into a more robust tool, bridging molecular-scale observations with global datasets to inform Earth system science and ecosystem research.

4.0 References

Beck, Hylke E., Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg, and Eric F. Wood. 2018. “Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution.” *Scientific Data* 5 (1): 180214.

Bond-Lamberty, Ben, Pralit Patel, Joshua Lurz, pkyle, kvcalvin, Steve Smith, abigailsnyder, et al. 2023. *JGCRI/Gcam-Core: GCAM 7.0*. https://doi.org/10.5281/zenodo.8010145.

Bowman, Maggie M., Alexis E. Heath, Tamas Varga, Anil K. Battu, Rosalie K. Chu, Jason Toyoda, Tanya E. Cheeke, et al. 2023. “One Thousand Soils for Molecular Understanding of Belowground Carbon Cycling.” *Frontiers in Soil Science* 3 (April): 1120425.

FAO, and IIASA. 2023. *Harmonized World Soil Database Version 2.0*. Rome and Laxenburg: Food and Agriculture Organization of the United Nations (FAO) ; International Institute for Applied Systems Analysis (IIASA).

Hashimoto, Shoji, Akihiko Ito, and Kazuya Nishina. 2023. “Divergent Data-Driven Estimates of Global Soil Respiration.” *Communications Earth & Environment* 4 (1): 1–8.

Hengl, Tomislav, Jorge Mendes de Jesus, Gerard B. M. Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, et al. 2017. “SoilGrids250m: Global Gridded Soil Information Based on Machine Learning.” *PloS One* 12 (2): e0169748.

Jian, Jinshi, Rodrigo Vargas, Kristina Anderson-Teixeira, Emma Stell, Valentine Herrmann, Mercedes Horn, Nazar Kholod, et al. 2021. “A Restructured and Updated Global Soil Respiration Database (SRDB-V5).” *Earth System Science Data* 13 (2): 255–67.

National Weather Service. 2018. “State Metadata.” NOAA’s National Weather Service. May 16, 2018. https://www.weather.gov/gis/StateMetadata.

Patel, Kaizad F., B. Bond-Lamberty, J. Jian, Kendalynn A. Morris, Sophia A. McKever, Cooper G. Norris, Jianqiu Zheng, and V. Bailey. 2022. “Carbon Flux Estimates Are Sensitive to Data Source: A Comparison of Field and Lab Temperature Sensitivity Data.” *Environmental Research Letters* 17 (11): 113003.

Poggio, L., L. D. de Sousa, N. Batjes, G. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. 2021. “SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty.” *The Soil*, June. https://doi.org/10.5194/SOIL-7-217-2021.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing* (version 4.2.1). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Smith, David B., William F. Cannon, Laurel G. Woodruff, Federico Solano, and Karl J. Ellefsen. 2014. “Geochemical and Mineralogical Maps for Soils of the Conterminous United States.” *Open-File Report*. US Geological Survey. https://doi.org/10.3133/ofr20141082.

Soil Survey Staff. 2014. “Gridded Soil Survey Geographic (GSSURGO) Database for the United States of America and the Territories, Commonwealths, and Island Nations Served by the USDA-NRCS.” United States Department of Agriculture, Natural Resources Conservation Service. 2014. https://gdg.sc.egov.usda.gov/.

Stell, E. D., L. Warner, J. Jian, B. P. Bond-Lamberty, and R. Vargas. 2021. “Global Gridded 1-Km Soil and Soil Heterotrophic Respiration Derived from SRDB V5.” *NASA Earthdata*. Oak Ridge National Laboratory DAAC (ORNL DAAC). https://www.earthdata.nasa.gov/data/catalog/ornl-cloud-soilresp-heterotrophicresp-1928-1.

Appendix A – Static Markdown

The following images are a static version of our final R Markdown for illustration only. To generate an interactive markdown see: https://github.com/JGCRI/monet_synthesis

Code

Morris&Wiens_MONetSynthesis

Kendalynn A. Morris & Nathan J. Wiens

2025-09-12

Introduction

This R markdown is designed to be a guide for comparing and integrating MONet soil data with publicly available datasets such as SoilGrids (Hengl et al., 2017) and the Soil Respiration Database (SRDB, Jian et al., 2021). This script executes a series of three comparative and statistical analyses involving 1) soil respiration relative to mean annual precipitation (MAP) and mean annual temperature (MAT), 2) assessments of soil pH within Köppen climatic zones and directly to SoilGrids, 3) and the spatial distribution of soil clay content. Due to differences in the spatial scale of global data products verses a nationally focused effort such as MONet, the response of soil properties to known drivers will be our metric of comparison (Collier et al., 2018; Shao et al., 2013).

Prep our R Environment

Load Packages

General data manipulation: `dplyr` and `tidyr`

Plotting: `ggplot2`, `ggpubr`, `ggpmisc`

Geographic data handling: `sf`, `terra`, `tidyterra`, `raster`, `tiff`

Precipitation and Temperature data: `geodata`

Show

Show

Load Pre-processed Data 1 of 3 (this can take a few minutes)

Show

```
## [1] "Rs loaded, cheers!"
```

Load Pre-processed Data 2 of 3 (ditto)

Show

```
## [1] "Clay data loaded, cheers!"
```

Load Pre-processed Data 3 of 3 (ditto)

Show

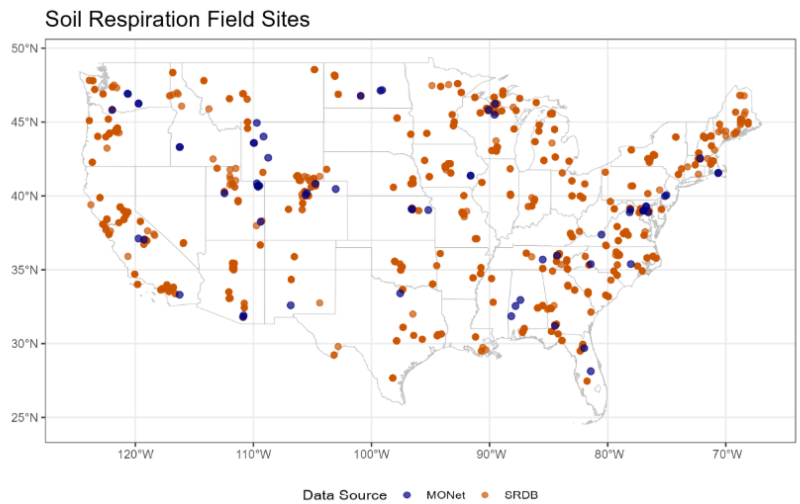
1. Examining Soil Respiration relative to Mean Annual Precipitation and Mean Annual Temperature

Soil respiration (Rs) captures aspects of biological activity in soil over variable time scales. It serves as a critical indicator for microbial processes and carbon fluxes influenced by both environmental gradients and molecular-scale interactions.

This section compares soil respiration measurements from MONet to the Soil Respiration Database (SRDB v5; Jian et al., 2021) and its associated gridded global ~1km^2 resolution product (Stell et al., 2021).

Sample Sites

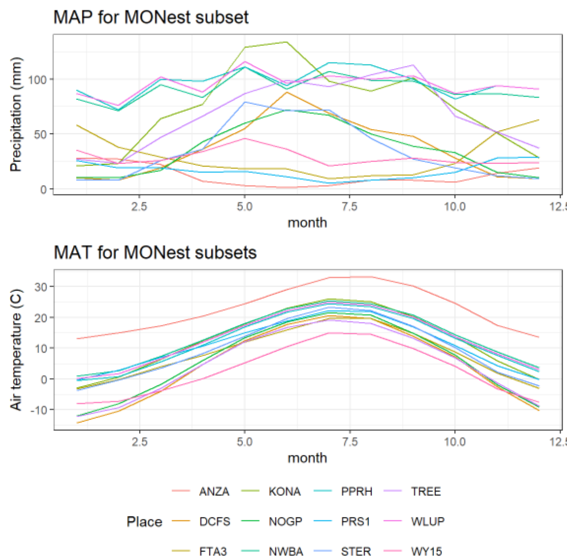
SRDB currently includes a significantly larger number of study sites across the United States compared to MONet. However, MONet is ongoing and this analysis can act as a guide for future sampling calls in order to increase sample representation across regions and soil characteristics.

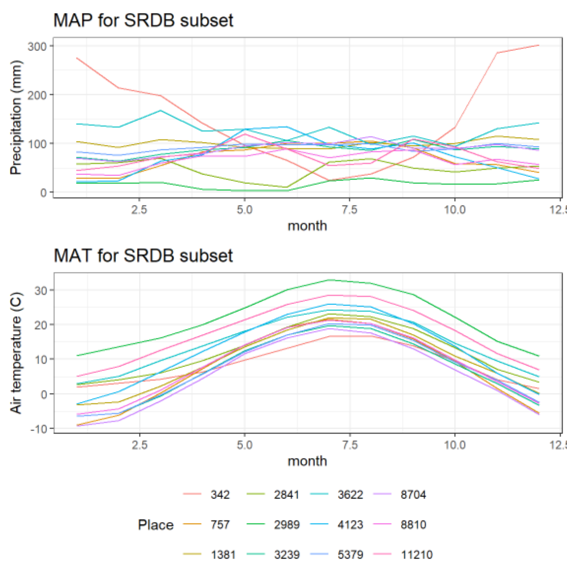


SRDB and MONet Rs Sample Sites

Extract Mean Annual Precipitation (MAP) and Mean Annual Temperature (MAT) for sample coordinates

Using the `geodata` package, ~1km^2 gridded 10 minute precipitation and average temperature data from 1970-2000 is downloaded and and extracted at the MONet and SRDB sample sites. These data are then used to create timeseries of MAP and MAP for each sample location. A selection of sites are plotted for illustration.





SRDB Conversion

SRDB soil respiration is converted from annual soil respiration to mg CO2-C / g soil / day using the following assumptions and parameters:

- **Bulk density** (ρ): 1.33 g/cm³ (mass of soil per unit volume)
- **Effective soil depth** (d): 10 cm as a common depth for soil respiration measurements
- **Area conversion**: 1 m² = 10,000 cm²

Using the equation:

$$\text{mg C g}^{-1}\text{ soil day}^{-1} = \left(\frac{\text{g C m}^{2}\text{ yr}^{-1}}{1.33 \times 10 \times 10,000}\right) \times \frac{1}{365} \times 1,000$$

MONet Rs Index

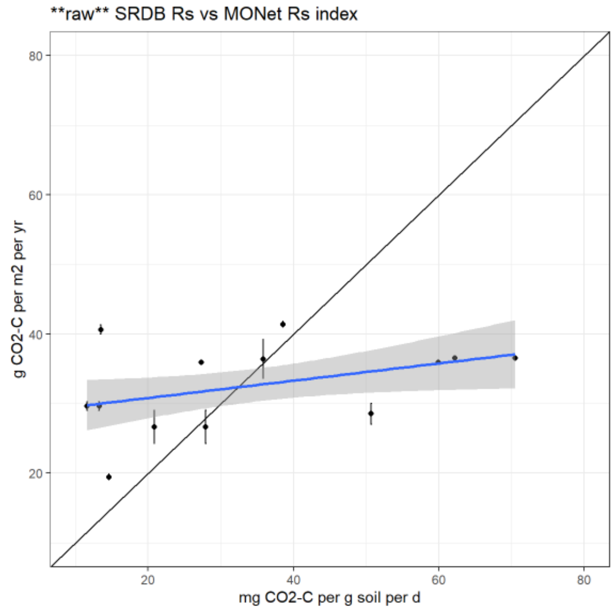
Each MONet sample core is made up of a top (0cm-10cm) and bottom (20cm-30cm) core section. To represent a single soil respiration data point for each MONet sample core, an index was developed which takes the top and bottom core sample into account as well as the two time points. Given that the majority of biological activity takes place in the top layers of soil, the top sample is weighted greater (65%) than the bottom (35%). After combining top and bottom data for each core, the one and four day respiration rates are given 1/3 and 2/3 weight respectively.

$$R_s\text{ Index} = 0.65 \cdot \left(\frac{1}{3} \cdot R_{s,\text{top, 24 hr}} + \frac{2}{3} \cdot R_{s,\text{top, 96 hr}}\right) + 0.35 \cdot \left(\frac{1}{3} \cdot R_{s,\text{bottom, 24 hr}} + \frac{2}{3} \cdot R_{s,\text{bottom, 96 hr}}\right)$$

One to One for Rs

Using the gridded version of SRDB, we can directly compare SRDB Rs to MONet Rs. To do so, a 1000m buffer zone was extracted from the gridded SRDB dataset and the median and standard deviation was calculated in order to create a confidence interval.

In the figure below, the vertical bars represent the calculated confidence interval of the SRDB Rs and the solid line represents y=x.



Summary of Respiration Rate by Source

Source	Min Respiration Rate	Max Respiration Rate	Mean Respiration Rate	Units
MONet	5.5488	92.3958	39.7783	g CO2-C per m2 per yr
SRDB	0.0004	0.0009	0.0007	g CO2-C per m2 per yr

Impact of MAP and MAT on Soil Respiration by source

The following figure show the relationship of MONet data points and median values of the gridded SRDB from a 1000m buffer around MONet sites. SRDB has a stronger correlation with MAP and MAT, but both MONet and SRDB show a positive relationship. These figures reveal that the current MONet dataset lacks representation of soil samples with MAP of 300-1000 mm/year

Show

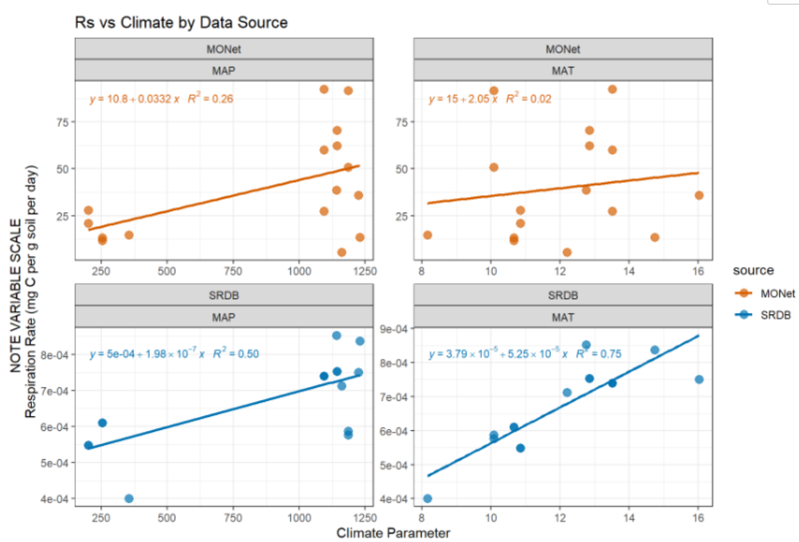
```
##
## Call:
## lm(formula = respiration_z ~ source + MAT + MAP, data = cd4m)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5983 -0.2258  0.1601  0.3166  1.6929
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.233e+00  9.587e-01  -2.330   0.0273 *
## sourceSRDB  -2.763e-16  2.841e-01   0.000   1.0000
## MAT          1.056e-01  9.277e-02   1.139   0.2645
## MAP          1.097e-03  4.292e-04   2.556   0.0163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8037 on 28 degrees of freedom
## Multiple R-squared:  0.3972, Adjusted R-squared:  0.3326
## F-statistic: 6.149 on 3 and 28 DF,  p-value: 0.002398
```

Show

Show

```
##
## Call:
## lm(formula = respiration_rate ~ source + MAT + MAP, data = cd4m)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.344  -7.534  -2.460  10.932  50.478
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  43.67653   22.83537   1.913   0.0661 .
## sourceSRDB  -39.77766    6.76833  -5.877 2.55e-06 ***
## MAT          -1.91560    2.20986  -0.867   0.3934
## MAP           0.02206    0.01022   2.158   0.0396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.14 on 28 degrees of freedom
## Multiple R-squared:  0.5855, Adjusted R-squared:  0.5411
## F-statistic: 13.18 on 3 and 28 DF,  p-value: 1.501e-05
```

Show

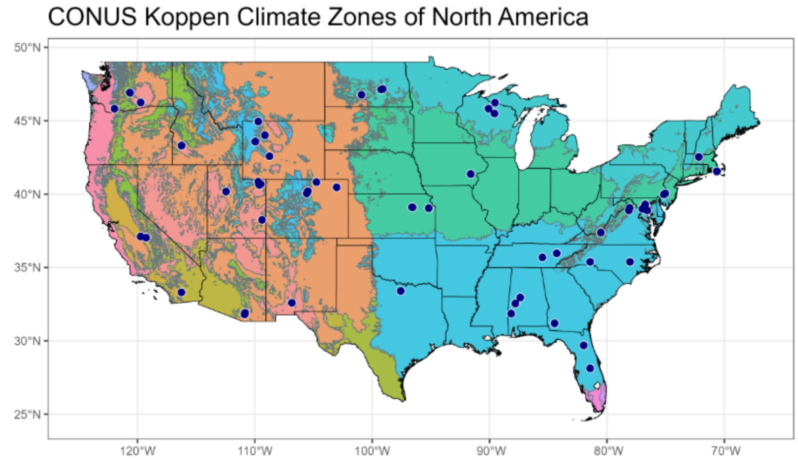


2. pH assessemnt within Köppen Climate Zones and directly to SoilGrids

Soil pH, acts as a proxy for biochemical interactions, particularly with abiotic factors such as soil mineralogy and atmospheric chemical deposition. As a fundamental property of soil chemistry, it provides valuable context for interpreting molecular observations. In this section, MONet soil pH measurements are compared to pH estimates from SoilGrids at regional scale by Köppen Climate Zone and by the location of each sample.

Soil pH, acts as a proxy for biochemical interactions, particularly with abiotic factors such as soil mineralogy and atmospheric chemical deposition. As a fundamental property of soil chemistry, it provides valuable context for interpreting molecular observations. In this section, MONet soil pH measurements are compared to pH estimates from SoilGrids at regional scale by Köppen Climate Zone and by the location of each sample.

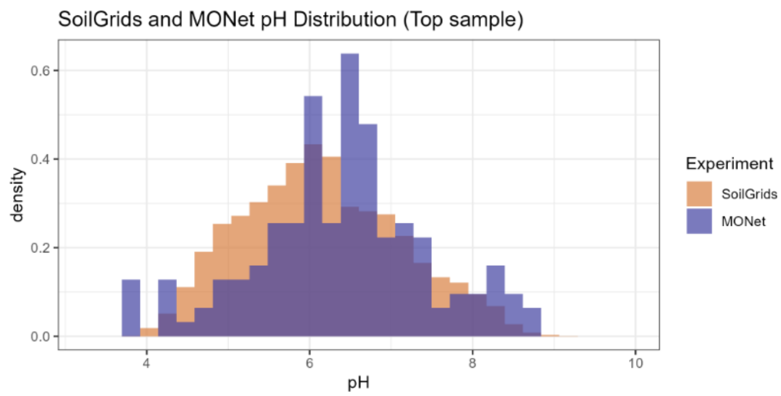
Extract Köppen Climate Zones for MONet data

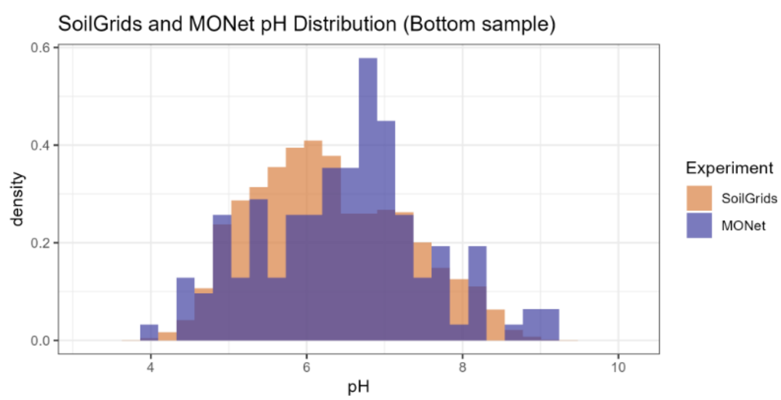


CONUS Köppen Climate Zones of North America

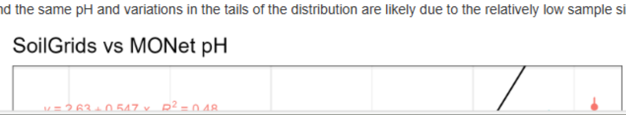
MONet vs SoilGrids density distributions

SoilGrids values are extracted at MONet sampling coordinates to enable direct point-to-point comparisons of pH. The distribution of these points are compared to evaluated ability of direct geographic comparisons.

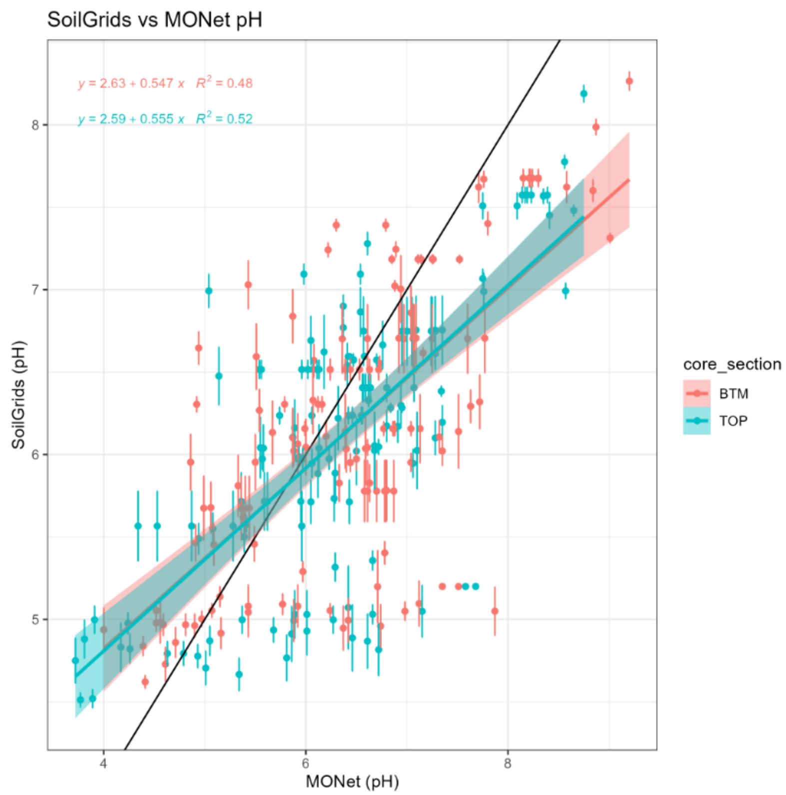




Density distribution of pH data.
In general, MONet data does moderately well at capturing the distribution of pH seen in SoilGrids. Both distributions peak at around the same pH and variations in the tails of the distribution are likely due to the relatively low sample size of MONet.



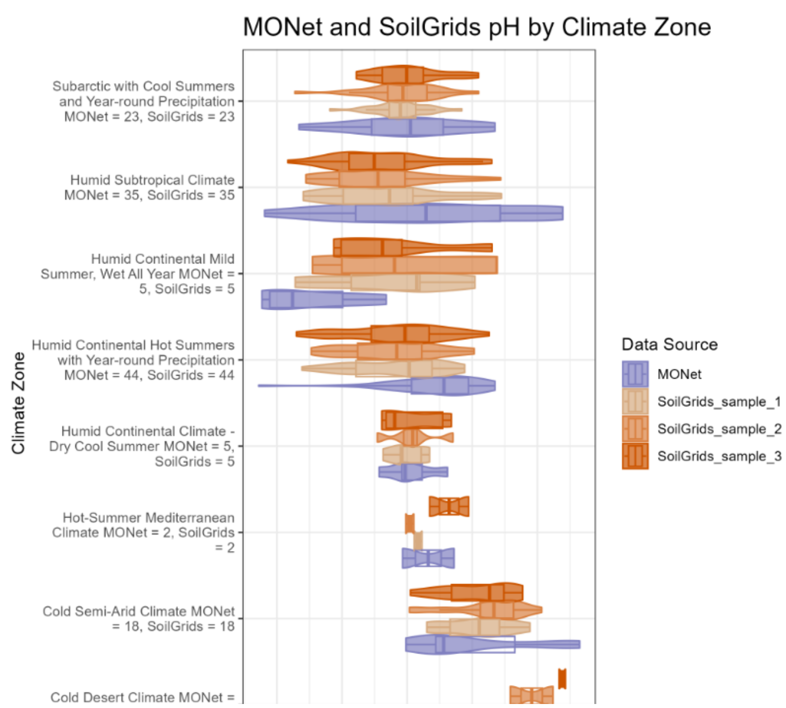
In general, MONet data does moderately well at capturing the distribution of pH seen in SoilGrids. Both distributions peak at around the same pH and variations in the tails of the distribution are likely due to the relatively low sample size of MONet.



As before, the vertical lines on each point represent a confidence interval calculated from SoilGrids pH values within a 1000m buffer around each MONet sample site. When comparing the data sources by MONet sample location, we found that compared to SoilGrids, MONet samples tend to underestimate values at the lower end of the pH range and over estimate at the higher end.

Distribution of pH in each region

We compared the distribution of MONet pH measurements by climate zone to a random sample of points from each climate zone in SoilGrids where n is the number of MONet sites per region. The median pH in the MONet samples was generally within the spread of the of the 25th and 75th confidence intervals of the SoilGrids samples.

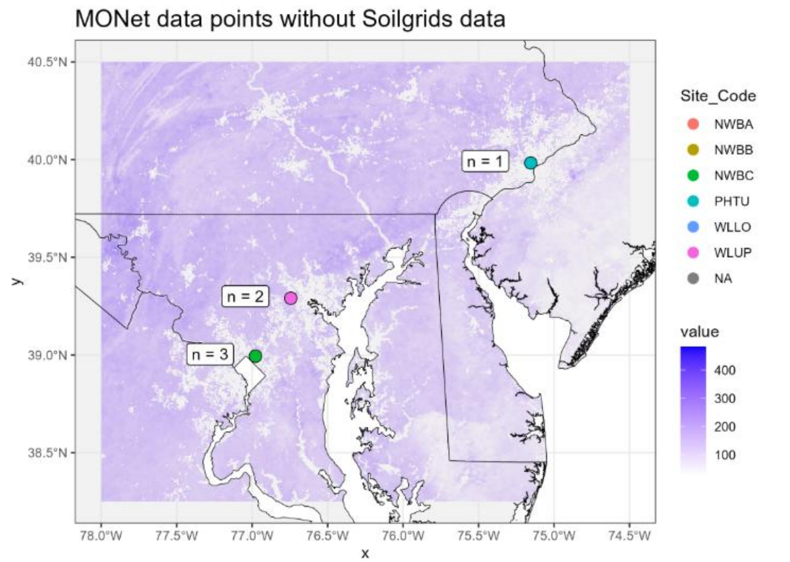


3. Spatial Comparison of Clay Content

MONet datapoints not in soilgrids

This analysis identified 6 MONet sample sites in locations with incomplete data in the soil grids dataset. Even though these locations are considered as urban sites and not recorded in soil grids, having data on the soil biogeochemistry composition can offer insights into soil health and processes in urban environments. The 6 locations were made up of 3 site in College Park, MD, 2 in Baltimore, MD, and 1 in Philadelphia, PA:

- Northwest Baltimore Base of Slope Adelphi, MD
- Northwest Baltimore Mid Slope Adelphi, MD
- Northwest Baltimore Top of Slope Adelphi, MD
- Winters Lane Lower Soil Pit (Baltimore, MD)
- Winters Lane Upper Soil Pit (Baltimore, MD)
- Urban CZO (Philadelphia, PA)

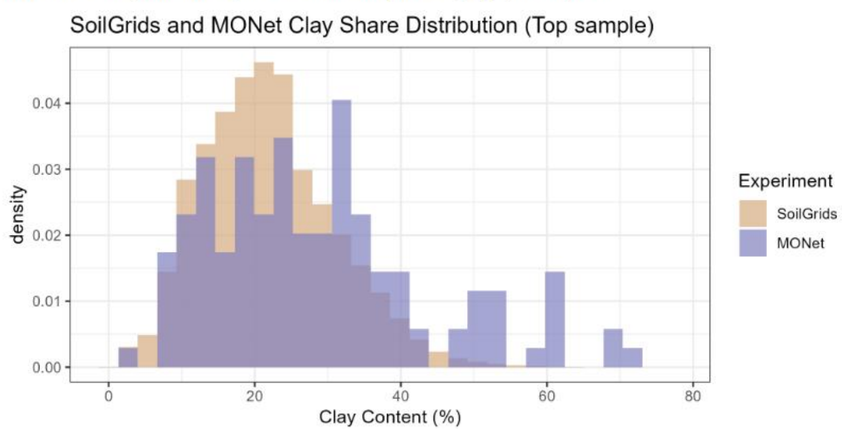


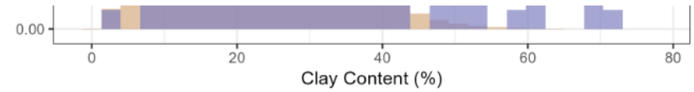
Gaps in SoilGrids Data

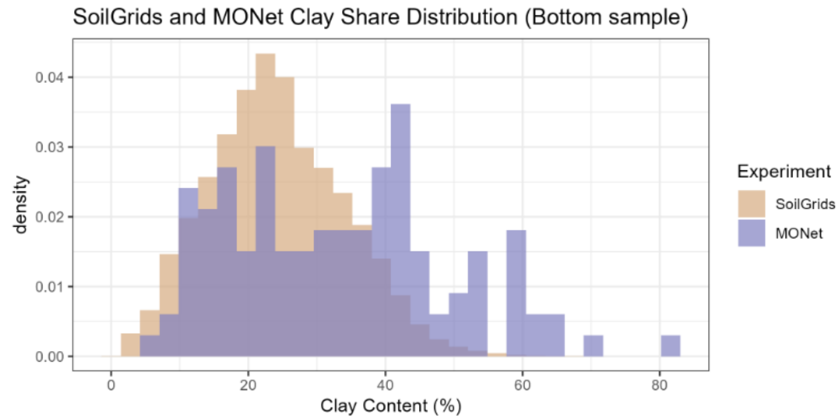
Clay 1:1

MONet vs SoilGrids density distributions

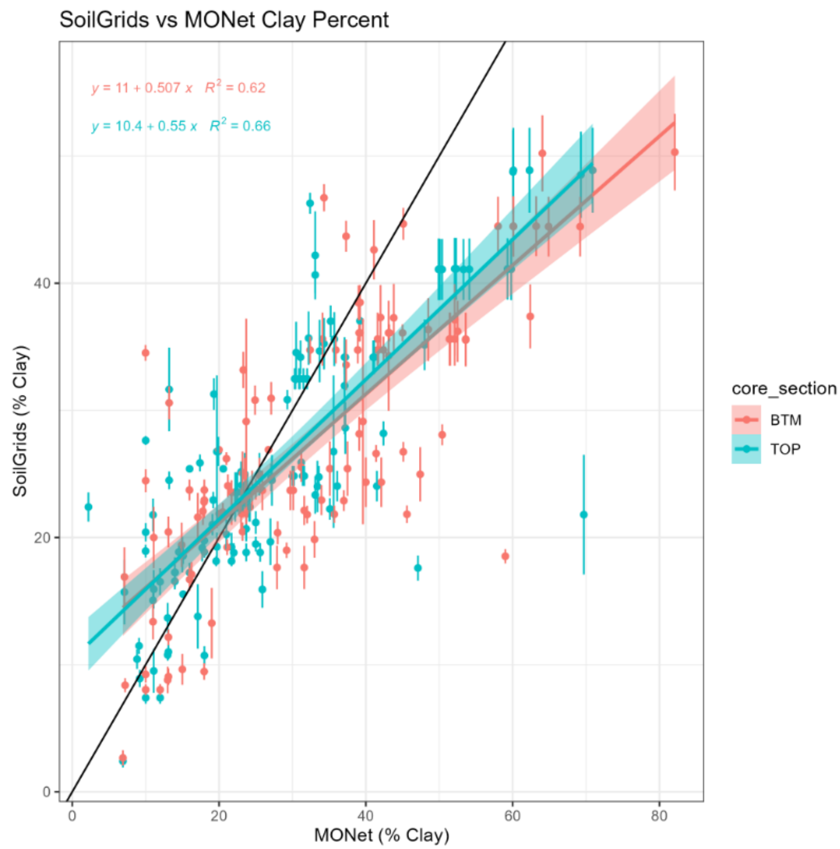
SoilGrids values are extracted at MONet sampling coordinates to enable direct point-to-point comparisons of clay content. The distribution of these points are compared to evaluated ability of direct geographic comparisons.





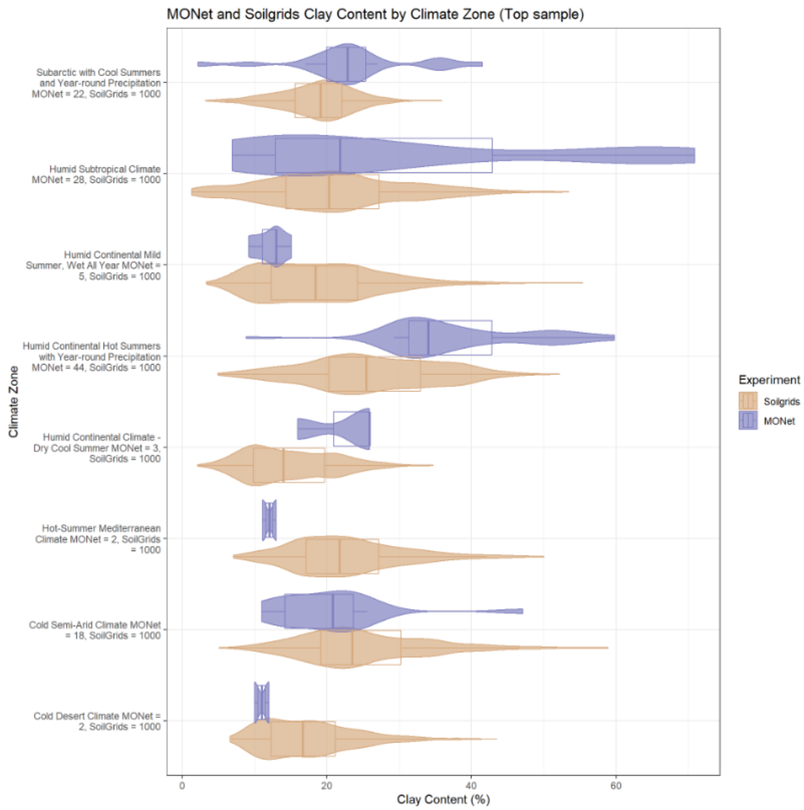


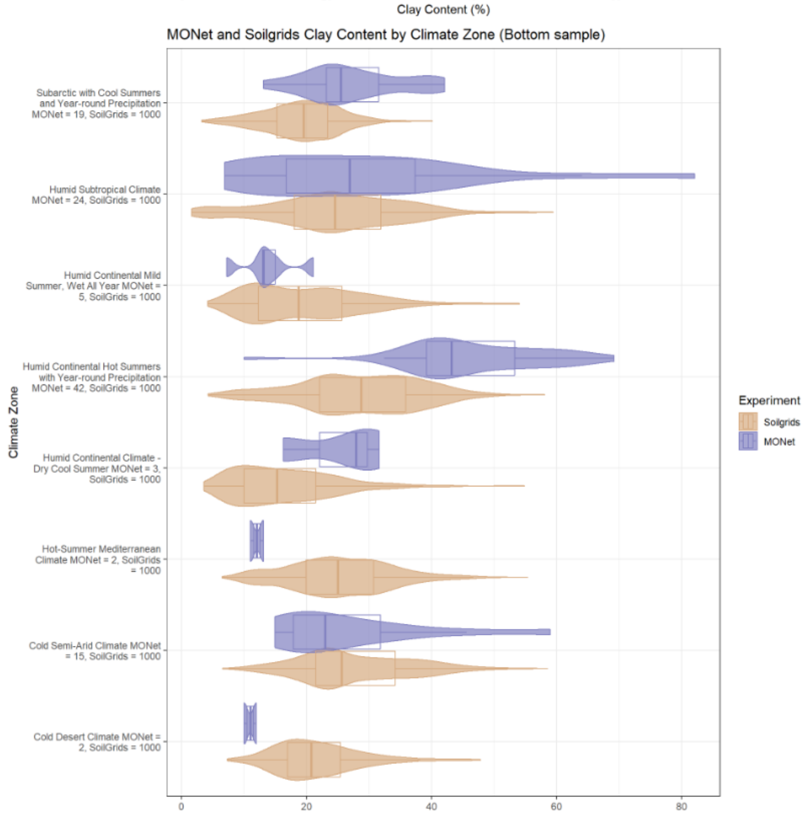
Density distribution of clay data.



One to One SoilGrids vs MONet Clay.
When comparing the data sources by MONet sample location, we found that compared to SoilGrids, MONet samples tend to underestimate values at the lower end of the clay content range and over estimate at the higher end.

Clay Content Violins

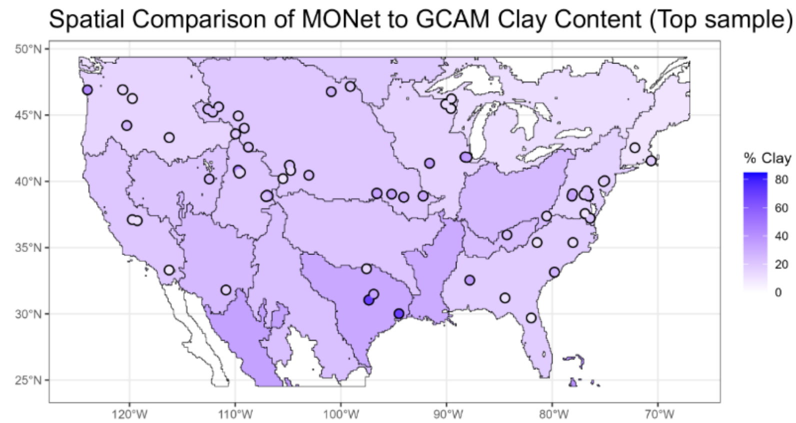




In climate regions with more MONet sample locations the distribution of clay content values tended to better match the distribution of SoilGrids values. Regions with fewer MONet samples tended to skew to one side of the interval.

In climate regions with more MONet sample locations the distribution of clay content values tended to better match the distribution of SoilGrids values. Regions with fewer MONet samples tended to skew to one side of the interval.

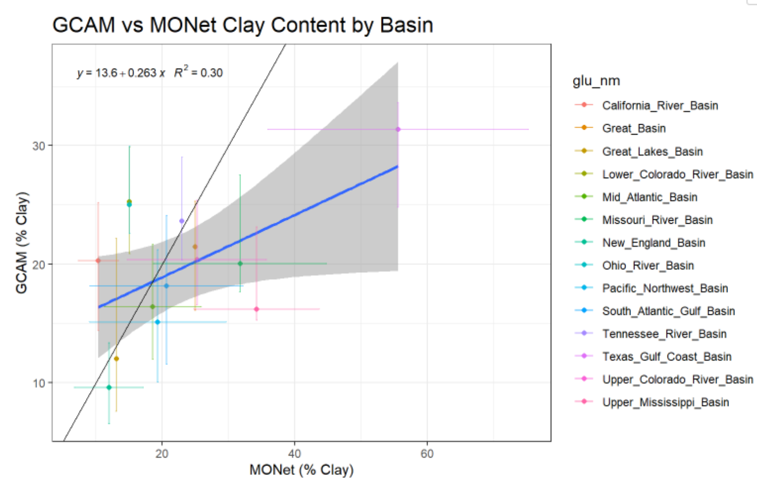
Applied Science Example: GCAM



Spatial Comparison of MONet and GCAM Clay.
GCAM (Global Change Analysis Model) is a dynamic-recursive model with technology-rich representations of the economy, energy sector, land use and water linked to a climate model that can be used to explore climate change mitigation policies including carbon taxes, carbon trading, regulations and accelerated deployment of energy technology. GCAM land use is represented in 232 basin which have clay content derived from SoilGrids. This figure shows a comparison of MONet values to the clay content of the 18 basins that make up CONUS.

GCAM Clay Comparisons

Show



Error bars for each point suggest variability and uncertainty in clay content measurements within each basin. The 1:1 line (black) shows where the datasets would perfectly agree, but most points fall below it, indicating that GCAM estimates tend to be higher than MONet values at lower clay contents and lower at higher clay contents.

Conclusion

This markdown has walked through some high-level analysis and comparison of MONet soil data and closely related databases. The code of making any figures that are called here can be found in MONetDataPrepocessing.R. There was so much more to explore than could be easily included in this compact format!

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov