PECAN STREET

Digital Dirt: Regenerative Agriculture R&D Roadmap

Harnessing Open Data and AI to Measure Soil Carbon Sequestration

July 2021

pecanstreet.org

PECAN STREET

Digital Dirt: Regenerative Agriculture R&D Roadmap

Harnessing Open Data and AI to Measure Soil Carbon Sequestration

Summary

Globally, agriculture and animal production account for 11% of greenhouse gas emissions. To transform the sector from a net emitter of greenhouse gases to a global carbon sink requires significant technological innovations and farmland management changes. Further, the way that we manage our agricultural lands is at the nexus of two of the greatest challenges facing humanity: 1) reversing land and soil degradation while maintaining sustainable production of food and fiber to meet growing global demands, and 2) harnessing the potential of soil carbon sequestration for drawdown of atmospheric CO2 to help stabilize our climate.

Accomplishing these goals will require fundamental changes in agricultural land management. Central to this restructuring must be an emphasis on improving soil health, which has the critical co-benefit of sequestering carbon in soils, and managing agricultural systems as part of the broader ecosystem. Practices aimed at building soil health are collectively referred to as regenerative farming, which includes maintaining continuous plant or residue cover, minimizing soil tillage, and using more diverse crop rotations that increase soil organic matter stocks. Supporting the transition to, and maintenance of, regenerative farmland management requires monetizing the value of the various environmental services provided and ensuring that farmers retain a greater proportion of the value of both the crops and environmental services they cultivate. Capturing the value of these environmental services, particularly soil organic carbon (SOM) sequestration, is inhibited by a lack of cost-effective, reliable and rapid measurement and verification processes for changes in SOC content.

To help overcome this critical data and technology gap, Pecan Street's Digital Dirt Initiative convened an *AI for Soil Carbon Monitoring and Verification Working Group* in partnership with the Texas Advanced Computing Center (TACC). Leading soils scientists, advanced computing scientists, economists, and carbon offset market managers worked together through this initiative to identify specific data gaps and technology development opportunities that would unlock new markets and revenue opportunities for regenerative farming, and then identified AI approaches from other research domains that hold the potential to overcome these barriers. The working group determined that rather than focus on development of innovative in-situ soil sensors, near-term success in the creation of a cost-effective solution was more likely to be achieved through benchmarking and improvement of existing agricultural models.

A lack of high-quality data to calibrate models is a fundamental problem slowing development of monitoring and verification (M&V) methodologies that are needed to realize the potential of soil carbon sequestration at scale. The working group achieved consensus on a promising artificial intelligence (AI) and machine learning (ML) research and development (R&D) roadmap to close data gaps more quickly and cost-effectively than waiting for development and adoption of affordable, in-situ soil organic carbon (SOC) sensors. Models currently constrained by insufficient datasets can be fed AI generated synthetic data, based on existing measurements and verified against real-world data, to transform their market utility. Developing open, benchmark datasets and model outputs will accelerate development of reliable, low-cost

tools for estimation of soil carbon sequestration that can be deployed equitably across a wider array of farms. The result is a powerful climate solution that is good for our food and fiber production systems and our rural economy.

The Climate Opportunity

Soil contains three times as much carbon as the atmosphere while agriculture accounts for roughly 20% of the world's greenhouse gas (GHG) emissions. Cropland soils are believed to hold the potential to sequester between 0.90 and 1.85 Pg C/year, which could provide 10% of the world's Paris Climate Accord emission reduction commitments. At less than $10 per megagram of carbon dioxide equivalent, this is considered a low-cost pathway.

Fortunately, carbon and nitrogen — the two GHG emissions that result from crop production — are important soil nutrients. Increasing carbon and nitrogen content within the soil, at the right levels of balance, increases soil health and fertility and has the added potential benefit of reducing atmospheric GHGs.

Supporting farmers in the the transition to regenerative farming requires monetizing the value of the environmental services that it provides. Carbon offset markets could provide a viable economic pathway for farmers to be compensated for carbon sequestration. Additionally, there is significant, pent-up customer demand for 'climate-beneficial' food and fiber products, similar to organic, free-range, grass-fed product offerings. And there is an opportunity for movement by leading corporations to reduce Corporate Value Chain (Scope 3) emissions that could provide added revenue for implementing practices that sequester carbon.

To participate in carbon markets, a farm would need to be able to produce an auditable, verifiable account of the baseline level of SOC on the farm and the additive carbon sequestered over time in the soil. These practices would have to be codified in carbon marketplace standards to ensure compliance. Carbon credit standards for carbon sequestration on farms and ranches have begun to emerge over the past year, though at the time of writing this report, no carbon credits have been issued under these new standards. A report by Environmental Defense Fund (EDF) and the Woodwell Climate Research Center released in July 2021 analyzes where these codes can be standardized and improved to provide greater certainty to financial markets and to farmers.

Currently, most farmers seeking verifiable measurements of their soil carbon content get information the same way they've been doing it for decades. They take a one-cup sample of soil and mail it to a lab that sends them back a result. Collecting these samples manually is arduous and the cost per SOC test ranges from $12 to $32 per sample. Therefore, farmers limit their sampling and apply these point-source measurements to an entire field area. While these measurements provide a good baseline for farmers to utilize in their land management decisions around fertilizer and other soil inputs, it is not sufficient to provide market confidence or to show compliance with carbon market regulations.

Due to underinvestment in the soil sciences and the agricultural sector for decades, application of the data and low-cost sensor revolution that has transformed almost every other economic sector is only beginning to take shape in agriculture. To date, low-cost, in-situ soil sensors that could provide high-fidelity, time-series, geospatial data on soil carbon are in the early stages of development. The sensors that are available in the marketplace almost all rely on near-infrared

(NIR) spectroscopy, have variable results on accuracy and are not cost-effective to return a profit within the carbon marketplace, which currently sells carbon offsets for an average of $15 per metric ton.

Most carbon offset markets and certification programs rely on a combination of modeled estimates and ground-truth data from lab-tested soil samples. These models are typically referred to as ‘process models’ and were developed decades ago for agricultural yield and farm planning support. The most commonly used models are DayCENT and DSSAT , along with COMET-FARM, a model and web interface that builds on DayCENT with a specific focus on helping farmers understand how to modify their practices to reduce GHG emissions. These models are extended by soil scientists as needed when new discoveries about soil, plant innovations, or data become available. Because soil carbon and nitrogen are important soil health indicators, these models are able to estimate soil carbon increases or decreases that would result from changes to cropping and land management practices, and which also incorporate changes in temperature and moisture as model inputs. The accuracy of these estimates is only as good as the data used to build the models for each farm, and they often require a sophisticated understanding of the model to be correctly used. Most farms lack the adequate historical and soils data necessary for high-quality model calibration. And research is lacking on the impacts of missing data for each model input on the accuracy of soil carbon and nitrogen estimation.

Without widespread availability of affordable ground sensors, the industry must continue to rely on these models and improve their ease of use and certainty parameters in order to serve as an effective measurement and validation (M&V) tool for carbon offset markets and other forms of revenue for climate-benefical farming and ranching.

With its decade of experience in bridging data gaps that are preventing the development and scaling of climate solutions in the energy and water sector, Pecan Street convened a working group to develop innovative solutions to these agricultural data and technology challenges with a focus on near-term solutions that can catalyze regenerative farming as a powerful, global tool to fight climate change.

Collaborating on Design of Equitable and Cost-Effective M&V

Pecan Street launched our Digital Dirt initiative to jump start collaborations and assemble resources that will systematically remove data barriers from regenerative agriculture’s path to tackling climate change.

The soil sciences are severely under-invested in compared to the value they provide to civilization for food security, drinking water quality, and air quality. The discipline is currently undergoing paradigm shifts in the understanding of soil organic matter (SOM) dynamics (Lehmann & Kleber 2015), and in more extensive use of data–model integration to leverage ecological ‘big data’ for more extensive evaluation of soil models (and the underlying mechanistic hypotheses they embody) against a wider variety of observed data (Campbell, Field & Paustian 2018). Soil modeling has traditionally been viewed as a fundamentally data-limited endeavor since long-term, statistically rigorous measurements of SOM changes over time are slow and laborious to collect. Iterative ecological forecasting is increasingly recognized as a means for both providing additional information to stakeholders and making near-term model-based predictions that can be validated against observations (Dietze et al. 2018). SOM modeling may greatly benefit from the use of informal but voluminous soil health test data crowdsourced from local farmers, and from novel data streams of high spatial and/or temporal resolution—e.g., in-situ soil or water quality sensors—for improvement and validation.

The technical challenge is that SOM consists of a complex mixture of compounds that range in their half-life in soil from hours to millennia and thus lack a simple, uniform chemical signature for easy characterization and quantification in situ. Spatial SOM distribution is also highly variable both horizontally and with depth. Spectroscopic methods (e.g., visNIR, MIR, LIBS; Johns et al. 2015) for rapid, lower-cost measurement of such spatial variability exist, but have limited extrapolation value due to high site- and instrument-dependent calibration requirements (Ge et al. 2011). In contrast, model-based estimates provide a straightforward, independent estimate of SOM dynamics based only on climatic and edaphic factors and thus may improve accuracy, but they lack the high spatial resolution and precision of in situ sensors. Machine-learning approaches such as artificial neural networks may prove useful for extracting the best elements of each approach and providing synthesized estimates with improved accuracy and precision.

In 2020, Pecan Street collaborated with TACC to convene an interdisciplinary working group to address the critical need to map an open source R&D approach for using AI and ML to backfill synthetic data to produce reliable benchmark data sets and model outputs that can serve as a foundation for rapid innovation. Participants in this AI *for Soil Carbon Monitoring and Verification Working Group* included:

Name	Organization	Title	Expertise
Dr. Eric Slessarev	Lawrence Livermore National Lab	Research Fellow; ISCN's Soil Health Coordinator	Soil Scientist; Modeling Expert
Dr. Kathe Todd Brown	University of Florida	Assistant Professor, Environmental Engineering Sciences; ISCN's Data Coordinator	Soil Scientist; Modeling Expert
Dr. Dan Harburg	IndigoAG	Vice President of Innovation, Carbon	Carbon Markets Expert
Kenneth Walker	GSI Environmental Inc.	Lead on SOM measurement team for BCarbon	M&V Standards & Protocols for Carbon Offset Market
Dr. Paul Navratil	TACC	Director of Visualization	Data Management & AI; Project Lead
Stefan Jirka	Verra	Innovation Manager, Agriculture	Carbon Markets
Dr. Kenneth Medlock	Rice University	Professor, Energy & Resource Economics	Carbon Markets & Economics

Dr. Nithya Rajan	Texas A&M	Director, Rajan Lab; Associate Professor, Soil and Crop Sciences	Soil Scientist
Dr. Johannes Lehmann	Cornell University	Professor, Soil Sciences & Agroecology	Soil Scientist
Dr. Pramod Pokhrel	Texas A&M	Postdoctoral Research Associate, Rajan Lab	Soil Scientist
Dr. Matthew Smith	Agrimetrics	Chief Product Officer	Data Management; ML
Dr. Weijia Xu	TACC	Research Engineer; Manager, Data Mining and Statistics	AI / ML and data for public good
Dr. Zhao Zhang	TACC	Research Associate; Data Mining & Statistics	AI / ML and data for public good
Dr. Marc Boudria	Independent	Independent	AI / ML

The working group's mission was to identify the critical M&V challenges faced by carbon offset market managers related to agriculture offset projects and standards development, and to identify AI and ML applications that may be able to rapidly produce leapfrog solutions.

Over the past decade, AI & ML have grown in their sophistication and impact as the amount of data collected on people and industries grows. Other sectors (biomedical research, satellite management, and object detection) have proven that when fed with enough data, AI carries the potential to solve problems and detect connections that are beyond human capabilities. However, AI technologies behave very differently than traditional analytical and computational approaches. To be effectively applied in the agriculture sector, AI developers need to work together with soil scientists to marry the domain expertises.

The working group met monthly over a 6-month period to exchange perspectives on the following core questions and to discuss domain-specific challenges and best practices:

- Do AI advances have the potential to unlock verifiable, accurate proxy measurements for SOM content and/or accurate predictions of SOM changes over time due to specific farm management practices and/or other factors, such as weather and soil typology?
- Is there sufficient data on soils and farm practices from the USA to enable application of ML and AI? Where insufficient datasets are identified, can AI/ML applications produce useful synthetic data or analytics capabilities with existing measurements to produce new insights into previously impenetrable phenomena? Can the results be verified against real-world data and used to transform or benchmark the accuracy of existing SOM models?
- What data sources can be combined to fill known gaps, and what can we learn from them?

- What ML or AI approaches are best suited to developing solutions for proxy or remote SOM sensing? What ML or AI solutions are best suited to developing predictive algorithms for SOM changes over time?

As our progress has shown, collaboration across domains is the key to rapid innovation that solves big problems. To further facilitate collaboration between the various domains that need to work together to develop and build M&V solutions for SOM, Pecan Street built a database of soil scientists in the USA working on soil carbon sequestration or soil properties measurement and sensing. The database is hosted on Pecan Street's Dataport site along with the open soil data sets. We encourage data scientists, model developers and others interested in engaging in this space for research or technology development to utilize the datasets and engage with the soil scientists on collaborative work.

Consensus-Based R&D Roadmap

The working group achieved consensus on several conditions that will shape the solution to resolve the data gap:

- High-confidence, in-field baseline measurements through soil sampling or a proven sensor combined with predictive models based on existing open-source agricultural models, such as DSSAT and DayCent, are the best option to estimate carbon credit generation over the lifespan of a carbon credit offset project for regenerative agriculture.
- Ground-truth measurements of soil carbon content at five year intervals combined with market tools such as a "reserve fund" to accommodate known measurement and prediction model uncertainties is recommended for verification and offset production risk management.
- Combining models and some ground-truth data reduces the need for total ground-truth data from yet-to-be-developed sensors or proprietary M&V systems.
- SOM accumulation and M&V for the top 30 cm of soil testing is adequate because there is not a consensus among soil scientists down to one meter depths.

Further, the working group identified a three-step process to apply existing AI/ML approaches to tackle the soil carbon data gaps that are limiting the ability to benchmark and/or improve existing soil carbon sequestration models:

1: Develop a benchmark dataset for soils model improvement

  - Generate and validate synthetic data to fill gaps in existing, publicly-available datasets and
  - Convert existing heterogeneous datasets into homogeneous data that complies with an ontological schema

2: Develop a repeatable knowledge extraction/mapping process to move future heterogeneous data sets into a homogeneous dataset.

3: Develop and verify the reliability of models to enable coarse/remote sensing measurement to infer more granular values based on small samples.

Pecan Street and TACC are making rapid progress on step one, creation of a benchmark dataset, and have acquired several government and NGO sourced datasets from the following sources:

1) National Cooperative Soil Survey (NCSS) Soil Characterization Data

The NCSS dataset contains soil characterization data from the National Soil Survey Center (NSSC), Kellogg Soil Survey Laboratory (KSSL), and cooperating laboratories. The dataset contains soil samples taken from soil sites primarily based in the United Stated but includes samples from other countries. Sites are identified by unique site identifiers. A site may have one or more pedons of soil profiles. Soil samples analyzed at a NCSS laboratory include soil properties such as bulk density, particle size distribution, and pH.

2) Soil Survey Geographic Database (SSURGO)

The Soil Survey Geographic Database (SSURGO) contains data collected by the National Cooperative Soil Survey over the course of a century. This dataset includes data from the United States and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS. Examples of information available from this dataset includes water capacity, soil reaction, electrical conductivity, and frequency of flooding; yields for cropland, woodland, rangeland, and pastureland; and limitations affecting recreational development, building site development, and other engineering uses.

3) SOils DAta Harmonization (SoDaH) & Synthesis Database

The SoDaH database contains data from 215 locations and 186 unique study sites. Data for this dataset is contributed from Detritus Input and Removal Treatments (DIRT), the Nutrient Network (NutNet), Long-Term Ecological Research (LTER) Network, National Ecological Observatory Network (NEON), and Critical Zone Observatories (CZO) networks.

4) e-RA Rothamsed Archive -Broadbalk-Rothamsted Long-term Experiments

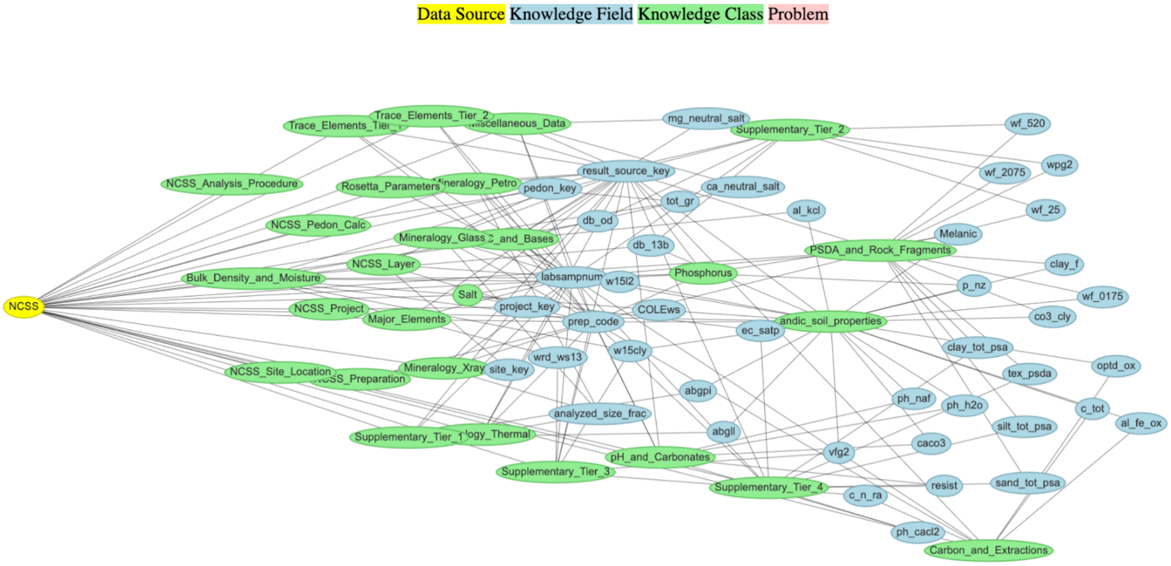
The Broadbalk Experiment Data from the e-RA Rothamsted Archive is one of the oldest continuous agricultural experiments in the world. The objective of the experiment is to test the effects of different organic manures and inorganic fertilizers (N, P, K, Na and Mg) on the winter wheat yield. The experiment site has a control strip that has not received organic manure of inorganic fertilizer since 1843.

5) Weather Database

The weather data is provided by Dark Sky. Data points including temperature, dew point, humidity, air pressure, wind speed, cloud cover and precipitation are provided. The dataset provided contains data for all NCSS site locations from 2016 – 2020. However, researchers interested in using data for additional site locations or other years can use the Jupyter Notebook that the team created to help pull weather data. The Jupyter Notebook was used to help the team narrow down the 285 unique locations where the weather data needed to be collected to cover thousands of unique NCSS site locations.

The individual datasets have been transformed for ease of comparison and integrated as individual files into Pecan Street's researcher database Dataport (www.pecanstreet.org/dataport). TACC has begun combining the datasets into a single cohesive database.

The first step in combining datasets is to map each dataset into common structures. In the below figure, the NCSS dataset is structured into categories of information - "Knowledge Class" and to specific data elements - and "Knowledge Field" that make up each category. Several Knowledge Fields are part of multiple Knowledge Classes, illustrating the complex, interconnected nature of these datasets.



Once this structure was established for the NCSS dataset, the team mapped which "Classes" and "Fields" were required for specific use cases such as running the DSSAT model. This provides a visual representation of the dataset that is relevant to the required use case. As expected, each dataset does not have sufficient content to support running models such as DSSAT, DayCent or COMET_Farm.

The additional datasets will be similarly mapped to enable easy comparison and selection of available data for specific objectives.

The subsequent open dataset will be made available by Pecan Street and TACC. The datasets are available on Pecan Street's Dataport as individual files at www.pecanstreet.org/dataport. The homogenized, integrated database of these datasets will be made available through TACC once additional funding is secured to complete development of the homogenization ML process.

This approach was recently validated in a report released by EDF and the Woodwell Climate Research Center, “Agricultural Soil Carbon Credits: Making sense of protocols for carbon sequestration and net greenhouse gas removals” (www.edf.org/soilcarbon). The report found that the best near-term option to advance and standardize M&V for soil carbon sequestration funded by carbon offset markets is benchmarking the accuracy of available process-based models:

> *“While a hybrid approach that combines in situ measurements with process-based models is likely a better solution than relying on process-based models alone, protocols need to improve the accuracy and scalability of the models by benchmarking them with independent and high-quality measurements from soil sampling…*
>
> *Most recommended models currently require a very high level of specialized knowledge and a deep dive into the scientific literature to understand their overall performance and quality.*
>
> *Development of an open-model registry with common performance metrics would greatly reduce the current opacity in soil carbon models. A set of sites that have long-term records of all required model inputs — including management records, soil properties, climate data and yield — and outputs — carbon fluxes including gross primary production and carbon dioxide respiration, nitrogen fluxes and long-term soil organic carbon (SOC) change — can serve as primary calibration and validation sites.*
>
> *In addition, a larger set of auxiliary sites located in important crop production zones could provide a reduced suite of measurements — crop yields, SOC, bulk density and N balance — to allow for true, out-of-sample model validation within each production zone in which the model will be applied.*
>
> *Existing research networks such as the USDA Agricultural Research Service’s Long-term Agroecosystem Research Network, some of the National Science Foundation’s Long-term Ecological Research Network and the NSF National Ecological Observatory Network, in combination with research stations at many land-grant institutions and other research centers, can provide the backbone for such a model benchmarking effort.*
>
> *Much of the data necessary already exists, but it will require a concerted effort to bring together these disparate data sources and ensure interoperability. Once existing data streams are identified, gaps in geographic data collection can be identified and targeted as new auxiliary sites in this benchmarking effort.”*

Pecan Street and TACC, under the guidance of its AI for Soil Carbon Sequestration working group, are well underway to achieving this recommended work, with enhancements through intentional open-platform design.

Model benchmarking: initial results

As part of its McGovern-funded groundwork, Pecan Street hired researchers at Texas A&M University's Rajan Lab led by Dr. Nithya Rajan to undertake an initial benchmarking assessment of the DSSAT model's soil organic carbon estimations. The assessment was undertaken by Dr. Pramod Pokhrel, a post-doctoral researcher.

To carry out this assessment, Dr. Pokhrel identified long-term farm experiments managed by Texas A&M from which he could gather the necessary data to build an accurate model of soil carbon and soil organic matter over time. The model results for estimates of SOC were then compared to real measurements to gain as assessment of the model's accuracy. In summary, the study - which will be published in full on Pecan Street's website - found that the measured SOC in no-tillage management were 1.26%, 0.79%, and 0.74% on top 5cm, 10cm, and 20cm, respectively. DSSAT estimated SOC was 0.99% in the top 20cm of soil. The comparative assessment found that the estimated SOC value produced by the DSSAT model (specifically, the CENTURY-Parton module of DSSAT) was 6% higher than the measured SOC. In the conventional tillage experiment, the measured SOC was 0.88%, 0.72%, and 0.63% at 5cm, 10cm, and 20cm depths, respectively. However, DSSAT estimated 0.99% SOC in the top 20cm of soil. The DSSAT estimated SOC in both tillage regimes was satisfactory.

Soil samples were also taken from the winter wheat – soybean rotation plots in 2002 and 2015. In 2002 with no-tillage management SOC was 1.68%, 0.89% and 0.8% at 5, 10 and 20 cm depths. The SOC content was lower in conventional till plots with the value of 1.12%, 0.87%, and 0.68% SCO at three depths. Overall SOC amount decreased slightly in 2015 in both tillage management and depths. In no-tillage management SOC was 1.48%, 1.12%, and 0.69% at 5, 10, and 20 cm depths whereas it was 1.01%, 0.96%, and 0.69% with conventional tillage. The DSSAT simulation experiment estimated 0.99% SOC in 2002 and 1.06% SOC in 2005 under no-tillage management. Similarly, the estimated SCO values were 1% in 2002 and 1.03% in 2015 with conventional tillage.

The overall result shows that the DSSAT SOC estimations had good agreement with the no-tillage management data but over-estimated SOC for the conventionally managed site. Nicoloso et al. (2020) also reported that DSSAT estimates trended to over-estimating SOC in conventional tillage management. Many factors that affect soil organic carbon such as initial soil organic carbon, soil type, amount of residue, and tillage. The difference in observed and estimated SOC may have been due to the factors that limit the amount of biomass produced and their incorporation. The research also found that mono-cropped production farms in the process of transitioning from conventional tillage to a no-till approach had a net decrease in SOC compared to farms that continued to practice conventional tillage within Texas or farms that transitioned to no-till in conjunction with integrated farming practices. This report points to the need for additional research on practice-SOC relationships and model variables and metrics.

R&D Next Steps for Soil Carbon M&V Development

Open Data and AI Sandbox Development

With a standardized aggregation of current, publicly available datasets, the next focus will be to create a benchmark dataset that will enable the Pecan Street team to benchmark the available models and enable users of the DSSAT, DayCent and COMET-Farm models to validate their outputs for specific projects. To accomplish the creation of this dataset, the team will:

1. Identify data gaps and use existing ground-truth measurements to train ML models to create synthetic data to backfill those gaps.
2. Characterize DSSAT, DayCent and COMET-Farm model output when using benchmark data against key usability metrics such as accuracy, precision, error ranges, etc. and confirm adequacy with carbon market participants.

To ensure scalability of the dataset and replication of the approach and schema for additional benchmark datasets, the research team build off existing ontological standardization efforts to establish an ontology for the benchmark dataset by:

1. Mapping existing data elements' relationships and naming conventions, and working with domain experts to resolve conflicts and propose a collective naming convention.
2. Accommodating nuances required by the DSSAT, DayCent and COMET-Farm models.
3. Circulating to various stakeholders to gain feedback and iterate the proposed ontology until sufficient agreement is reached.

The working group estimates this work can be accomplished in 18 months if adequately resourced. All work and outputs will be open platform and open source, helping to ignite innovation and market confidence in the outcomes.

Data & Ontology Standardization

In parallel with the creation of an open dataset and transparent soil carbon model benchmarking, the working group recommends continued collaboration with other industry leaders and researchers to:

- Establish a recommended data ontology standard;
- Create a data reporting and database architecture schema to standardize future datasets;
- Create a data documentation guide to support non-soil scientists in working with the soils data for AI, model, and soil carbon market innovations; and
- Establish a set of open API's that facilitate researcher integration of datasets to facilitate scientific discoveries.

The establishment of a standardized data ontology for soil data is under development by an international collective of researchers, co-led by working group member Dr. Kathe Todd-Brown. This ontology is expected to be complete and published by the end of 2021. In the meantime, several ontologies have been published through peer-reviewed processes, including by DEMETER (https://h2020-demeter.eu/wp-content/uploads/2020/10/DEMETER_D21_final.pdf), ETSI (https://saref.etsi.org/saref4agri/v1.1.2/), Leeds University (https://archive.researchdata.leeds.ac.uk/42/), and AgroRDF (http://data.igreen-services.com/agrordf). For the initial data standardization process, the TACC team selected the Leeds University developed ontology, which is more concise and focused on soil properties. Where measurement protocols differ or are not adequately described, deviations will be noted and explained in the data documentation to be produced by TACC and Pecan Street. The AgroRDF ontology may be utilized once more datasets are integrated as it is more comprehensive, but for the current needs it is less directly relevant.

Model & Sensor Innovation

In addition to the work undertaken by Pecan Street and TACC, the Digital Dirt collaboration identified three additional recommendations for activities that need to be undertaken to accelerate a high-confidence, standardized M&V approach for soil carbon sequestration:

1. Sensitivity analysis for model improvement

At the same time that sophisticated data analytics and synthetic datasets are developed to enable model benchmarking and improved modeling estimations for SOM, Pecan Street urges investment in innovation and improvement of soil, agricultural and biogeophyisical models as well as soil sensors.

To improve the existing models, users need to understand how inaccuracies or incompleteness in data for each model input variable impacts the accuracy of outputs. This research has yet to be comprehensively undertaken in relation to soil carbon and nitrogen content. To close this knowledge gap, a sensitivity analysis by variable for the different soil and climate zones for each model could be undertaken to identify where investments should be made to improve model accuracy and cost of data acquisition for valid estimates, predictions and verification. A sensitivity analysis is also an important step to simplifying model inputs for farmers to use in estimation of their potential for SOM sequestration in transitioning to regenerative farming practices.

2. Creation of modular benchmark datasets that can be applied to specific regions that account for variability in soil type, topography and climate.

One of the M&V challenges for SOC, as discussed above, is the cost to undertake soil samples and to invest in in-situ sensors. While improved and benchmarked models combined with a limited set of field measurements present an opportunity to rapidly and significantly decrease the cost and the time burden to undertake SOC M&V, in order for these modeled outputs to be acceptable to carbon markets and broader financial markets, verifiable and replicable validation of the modeled outputs needs to be provided. Verra Carbon Standard, for example, is currently requiring auditors to validate the performance of modeled estimates. These auditors review the underlying calculations and reasonable certainty of the modeled results for each project. Methods to reduce the overhead cost of both aggregators and auditors for participation in carbon markets would allow participating food producers to capture more of the value of the carbon offset purchase, which is critical for the economics of regenerative agriculture. An open, transparent benchmark dataset that can be tailored for model calibration to specific farms or could be run through a model and compared to the outputs of the project model as an accuracy comparison is a valuable capability in attesting to an offset credit's validity that could drive down auditor review time and costs.

3. Creation of an open platform AgTech performance validation center

AgTech investments are critical to attract innovators and entrepreneurs into this space. Innovation is needed right now to develop novel approaches to regenerative farmland management and soil carbon sequestration, and also to decrease the costs and time to participate in new carbon markets.

A similar trajectory of increased investments in CleanTech occurred over the past 20 years, driven in large part by policy signals from the federal government and international markets that pointed towards significant future investments in renewable power, smart grid and smart energy management systems. Important lessons can be learned from the CleanTech history and applied to the current injection of AgTech financing that can help to ensure investments are pointed towards scaling up solutions that will achieve the promised benefits and be attractive to customers. The most critical of these lessons is the creation of an open platform product validation and optimization center. The US Department of Energy has invested in several similar platforms for energy and water, including the PLATFORM for Product Launch at Pecan Street's lab in Austin, Texas.

A powerful and easily implementable model in the AgTech sector could build on the USDA's existing network of Long-Term Agricultural Research Centers (LTAR) or long-term experimental farm sites at agricultural research universities and extension services. By leveraging existing study sites that have well-documented land management history, results can be applied to similar regions. Pecan Street found in its product validation and testing program that results need to be confidential to the company otherwise innovators will feel disincentivized to have their products evaluated. However, by engaging investors to inform them of this service, they can require companies to secure a third-party performance and environmental impacts evaluation as part of the financing discussions. The result will help direct investments to the solutions that show the most promise.

Scaling Solutions from Farms to Regions

The challenges of rapidly scaling-up cooperation on regenerative farming to entire communities has heretofore been a barrier to achieving ecosystem benefits. We theorize that community-level coordination and decision-making for farmland management will be critical to ensuring sustainability and resiliency of our food systems and climate into the future. Providing incentives and reducing barriers to individual farmers for managing their croplands in a way that provides ecosystem benefits through regenerative farming practices is important; however, collaboration across farms is key to ensuring that ecosystem services are lasting and provide net benefits to an entire community.

Pecan Street's Digital Dirt initiative is working with researchers across the country, supported by a NSF award, to drive towards community and regional adoption of regenerative farming practices by first answering the following questions:

- Can technological innovations in data–model integration sufficiently improve quantification of environmental services to spur market development for compensation of these services?
- What are the psychological and SOMiological challenges that impede or promote collective land management decision-making among regional farmers?
- What are the economic incentives needed to nudge farmers toward collective decisions that produce more, and longer lasting, ecosystem services?
- What is the technology integration platform or model that fits with farmer decision-making to drive collective behavior for richer ecosystem benefits?
- What, if any, technologies need to be developed in order to create a 21st century farmer cooperative model?
- What, if any, policies and/or economic incentives need to be developed in order to gain farmer participation in a cooperative model that provides more co-benefits to the community?
- What is the best community to partner with on a full proposal that will be the most likely to result in a scalable and translatable solution for other agrarian communities?
- How can this model and new precision agriculture technologies be optimized with policies, economic incentive programs, and organizational structures to deliver better economic returns for farmers and America's rural agricultural communities?

While the process models described above are currently implemented as an individual field- or farm-scale modeling tool, a shift towards a landscape perspective would enable new applications for assessment of community-scale land management, while creating novel opportunities for model validation and improvement. Data sources for calibration and validation of the models at landscape scales—i.e., in response to local gradients in soil texture, land use history, soil health, topography, and micro-climate—have been extremely limited, and model calibration and validation efforts have typically focused on much smaller (i.e., individual agricultural experiment stations) or much larger (i.e., regional- or national-scale collections of experiment station data) spatial scales (Field et al. 2016).

Past work in the bioenergy space has demonstrated that community-scale assessment reveals new opportunities for landscape design and optimization, resulting in greater aggregate soil carbon sequestration and ecosystem service provisioning than would be possible through optimizing the management of individual farms (Field et al. 2018; Mishra et al. 2019).

Social analysis of collaborative, regenerative farmland management has been limited to date. While there is evidence of emerging interest from individual farmers in soil health enhancement (i.e., a "soil stewardship ethic"; Roesch-McNally, Arbuckle & Tyndall 2017) as an adaptive response to changing climate, there are few examples of coordinated landscape-scale farmland management based on shared socio-ecological goals (Jellinek et al. 2018). However, collaborative community management and monitoring of carbon storage has been shown to be viable and beneficial in the context of forest protection—where the carbon is readily visible as aboveground biomass (Larrazábal et al. 2012). We hypothesize that more transparent and convenient soil carbon monitoring tools and data that can support the development of markets that preference community-scale action and adequately compensate participants for the enhanced climate and environmental benefits achieved through community-scale action will help make environmental service provisioning more tangible to farmers and other stakeholders. Community-scale shifts towards regenerative agriculture also solve some of the scientific unknowns related to soil carbon sequestration such as transferability of SOM between geospatial carbon pools. A farm-level approach would not be able to account for how much carbon accrual in soil is occurring due to atmospheric drawdown or from homeostatic or other chemical processes in the soil that draw carbon from other regions of soil into the area under regenerative farming. A community-scale approach would resolve this issue for the most part if sufficient participation is achieved from neighboring farms.

Consideration of regenerative farming adoption and scaling needs to include a concerted focus on addressing systemic racism and resulting disparities in access to information and financing across racial lines.

Outreach & Adoption

The outputs of this R&D effort are useful only if the farming and agriculture industry accelerates the transition to regenerative farming and embraces open data and open platform innovation. Open platforms spur innovation at a faster rate, by breaking down the barriers to market entry and enabling rapid evaluation of ideas by investors, consumers and policymakers. With the climate crisis upon us, rapid innovation and adoption of proven solutions that convert agriculture from a source of net greenhouse emissions to a global carbon sink is critical.

Community-Based R&D

While significant investments are being made to solve the technical challenges and optimize the characteristics of soil that enable it to bind and sequester carbon, technical feasibility must be paired with an understanding of how regional groups of farmers make decisions regarding technology adoption and practice changes. Through this understanding, process optimization can be developed by extension services, universities, purpose-based NGO's, foundations and companies to shrink the gap between what becomes technologically achievable and what is actually achieved.

As noted by Finistere Ventures in their 2019 AgriFoodTech Investment Report, "[The digital transformation of agriculture has taken longer than many investors anticipated] due to challenges with farmers and channel partners integrating diverse data and insights into meaningful action. We believe that there will be a second wave of investment focusing around not only integration, but working with partners downstream who value the data sets being collated at farm level -- for example, in climate certification or consumer valued insights such as traceability. The age of big data on farms is likely to morph beyond agronomy to consumer and value chain benefits."

A deeper understanding of farmers' current attitudes towards technology adoption and practice changes, particularly as relates to sensors and data collection on their farms is critically important in the current economic climate. As reported by the Farm Bureau in October 2019, "Chapter 12 farm bankruptcies continued to increase as farmers and ranchers struggle with a prolonged downturn in the farm economy that's been made worse by unfair retaliatory tariffs on U.S. agriculture as well as two consecutive years of adverse planting, growing and harvesting conditions. Over the prior 12 months, Chapter 12 bankruptcies totaled 580 filings and were up 24% from the previous 12 months."

While a more challenging financial picture for farms is likely to make farmers more risk averse and less likely to make non-critical investments, the burgeoning crisis in farm bankruptcies combined with the digital revolution in farming is resulting in new financing and lending practices at banks that could offer more options to farms that have more data on soil conditions. As stated by Finistere in a pre-pandemic analysis: "The cloudy picture around trade over the last year has pushed farm bankruptcies above 20% in the US. However, the debt-fueled dynamics of farming have also encouraged the ongoing revision of strategies for lending against the value of farmland as a form of real estate. Enterprise resource planning systems can assist in assigning those valuations based on yields, soil quality and productivity, with multiple lenders, including Wells Fargo, trying to combine ERP and data collection tools to inform the underwriting process. In 2019, Farmers Business Network raised $175 million from Expanding Capital and Kleiner Perkins at a post-money valuation that minted a new agtech unicorn in part on this promise. More importantly, the outsized round illustrates the renewed interest in farm management systems that can connect operations to cost savings mechanisms based on data-driven financing options in an uncertain macro-environment. (pp. 10-12)"

At the same time, significant investments are underway into AgTech start-ups and innovations. Between 2014 and 2018, 282 agtech companies raised $5.5 billion of private capital across 481 transactions in the U.S. and Canada. These investments will flop and the social and environmental benefits will fail to be realized if farmers don't adopt the regenerative farming practices that many of these technologies are designed to facilitate. Designing solutions with farmers as an early and on-going part of in the input process with help to ensure these solutions achieve the intended outcomes.

To develop a better understanding of how to support farmers in changing their land management practices, Pecan Street undertook an initial survey of 79 farmers across New York and California. One of the questions asked farmers to rank the factors they considered most important or valuable in making decisions about farm management. The six factors they could rank were profitability, ease and time efficiency, soil health, farm worker and community health, government incentives, and novelty and/or personal interest. For both NY and TX farmers, profitability and ease and time efficiency rank as the most important factors while government incentives rank as least important. When asked how they evaluation a potential farm management practice change, the words time and cost are the most prevalent in the responses. The majority of survey respondents reported that their primary source of information about new technologies, research or funding came from neighbors and other farmers/ranchers in their communities, with about 1/3 of respondents also reporting that farmers associations and local extension agencies were also important sources of information. Given the large influence of social networks on a farmers access to information and resulting opportunities, gaining an understanding of how those social networks and information channels differ for farmers & ranchers of color is necessary to ensure equitability of opportunities and access to resources.

Pecan Street's farmer survey, and other similar sector analysis, also reveal that access to land and financing are more challenging for individuals of color. The history of racist lending and land ownership policies in the United States that prevented individuals of color from owning land and resulted in the theft of their land continues to systematically bar people of color from farming and ranching. Our economics of food and the high cost of productive land make it extraordinarily difficult for an individual without independent wealth to acquire land for food production. Nearly 88% of the respondents to Pecan Street's survey identified as Caucasian. Nearly half of the respondents to the survey reported that they grew up on a farm and 87% plan to leave their farm or ranch to family members. It is more difficult to be a farmer or rancher if you don't inherit your land. Addressing racial disparities in farming requires pursuing restorative justice for individuals of color to access farmland ownership. As farm lenders move towards incorporation of increasingly common soil health data in their risk assessment and property valuation, ensuring equitable access to information and land management technologies will be important to help undo system racism within land ownership and agriculture.

To support racial and climate justice efforts as pertains to agriculture and carbon offset markets for soil carbon sequestration, Pecan Street and its partners will undertake additional research to identify and resolve racial disparities in access to information, financing and resources related to regenerative agriculture and food production profitability. As a first step, Pecan Street is developing a research project that will focus on food producers of color and communities of color to map out information and resource access differences compared to white individuals and communities, and to document community-based solutions for information and resource access in relation to food production. We will also work to undo the theory of early adopter-based change that preferences individuals who have the time, resources and connections to identify and invest in new technologies by targeting R&D and access resources to those individuals, furthering a cycle of privilege and white supremacist culture.

Advancing open data approaches

As discussed above, the carbon offset market industry is aligning with the hybrid data-model integration approach to SOC M&V. However, open access to historic and future data is important to accelerate these markets and to offer farmers an economically-viable and easy pathway to invest in regenerative agriculture. As the carbon offset market and other revenue pathways grow for regenerative agriculture, private companies and investors will seek to capture the value of this market by pushing proprietary services and technologies. While development of a robust market for AgTech and low-cost soil sensors can help to further drive innovation around sustainable farming. It is important that behemoth companies are not able to capture and privatize the bulk of farming data, which would serve to inhibit innovation and market growth. It will be critical for policymakers, farmers and carbon market managers to understand the implications of data ownership around soil sensors and modeling outputs, and to balance the value of open data with privacy.

Education and engagement with agricultural funders and industry leaders on open data standards, privacy best practices, and carbon market development will be important to ensure that those most affected by these decisions are involved in making them. Pecan Street will additionally continue to pursue grants and other funding sources to support the outreach, education and collaboration required to develop, prove out and educate industry on validated, open soil carbon models for predictive assessments and validation of soil carbon sequestration.

Soil carbon sequestration market development

Finally, on the soil carbon M&V development roadmap intentional development of optimal market structures needs to be researched and undertaken. A core piece of this effort will be cultivation of the ecosystem and coalitions necessary to achieve a consensus-based market structure. A second core component of this process should be evaluation of carbon and environmental services markets in other industries. Examples from the agricultural and energy industries along with other relevant markets should be examined for best fit to valuing and incentivizing climate-beneficial food and fiber production.

While the carbon offset marketplace is an important and timely economic pathway to value and compensate farmers and ranchers for soil carbon sequestration services, the current market structure is limited in two ways. First, by requiring the separation of the value of soil carbon sequestration from the value of the food or fiber produced. Second, by layering in additional administrative and other soft costs into the qualification and validation of carbon offsets. For agricultural offset projects, aggregators and auditors will likely be required, taking a cut of the offset value along with the portion taken by the offset market manager.

Programs for regenerative agriculture or simply for verified soil carbon or zero emissions farming that enable the food commodity to garner a higher price to sellers should be explored and compared to the revenue and ease of compliance associated with participating in similar programs such as an organic certification. Several such certification programs for regenerative agriculture are currently in place and need to be standardized so consumers can properly value and understand the certification.

An interesting market approach is under development in the biofuels sector that combines the efficiency and market-validated approach to certification and valuation of biofuel mixes with the Renewable Energy Certificates (RECs) approach. RECs are a process whereby the origin and climate impact of energy are accounted for, tracked and assigned a tradable commodity value. RECS are the accepted legal instrument through which renewable energy generation claims are substantiated in the U.S. The EPA published a useful guide explaining the difference between offsets and RECs, available at https://www.epa.gov/greenpower/offsets-and-recs-whats-difference, that could be informative to market development for soil carbon sequestration on farms and ranches. A RECS-based approach may enable farmers and ranchers to capture more of the value of the sequestered carbon.

As soil carbon M&V standards are developed, it is worth simultaneously pursuing market research and development to ensure that these efforts will enable and support multiple market pathways.

Thank You

Thank you to the Patrick J. McGovern Foundation, the National Science Foundation and The Texas Advanced Computing Center for their support of Pecan Street's Digital Dirt Initiative. In addition to the working group members listed above, Pecan Street is grateful for the contributions of Bart Bohn, Dr. LaKisha Odom, Karl Rabago, Dr. Keith Paustian, Dr. Richard Stedman, Graham…, Dr. Ken Schlather, Dr. Nithya Rajan, Dr. Pramod Prahkol, and Dr. BCARON GUY.