

AgMIP

The Agricultural
Model Intercomparison
and Improvement Project

Building a Collaborative Multi-Model Framework for Simulating Soil Carbon and Nitrogen Dynamics in U.S. Agricultural Systems

Farm-Model Intercomparison Project (Farm-MIP)

WORKSHOP REPORT

JANUARY 12-14, 2026
THE FORUM
COLUMBIA UNIVERSITY
NEW YORK, NY



Table of Contents

| | |
|--|-----------|
| Day 1: Introduction, Project Goals, and Protocols | 1 |
| Modeling Group Presentations | 1 |
| APSIM Model | 1 |
| APEX & EPIC Models | 1 |
| DayCent Model | 1 |
| DNDC Model | 2 |
| DSSAT System | 2 |
| SALUS Model | 3 |
| SWAT+ Model | 3 |
| TASC Model | 4 |
| Guided Group Discussion 1: Farm-MIP Draft Protocols | 4 |
| Breakout Session 1: Defining Farm-MIP Consortium Goals and Scope | 6 |
| Day 2: Datasets and Model Inputs | 7 |
| Group Discussion 1: Data Sharing Agreements and Protocols | 7 |
| Group Discussion 2: Candidate Calibration Sites | 7 |
| Group Discussion 3: Datasets for Regional- and National-Scale Modeling | 8 |
| Breakout Session 2: Farm Practice Simulations | 9 |
| Guided Group Discussion 2: Input Data Quality Control and Gap-Filling | 10 |
| Group Discussion 4: Refine Phase 0 Model Input Files | 11 |
| Day 3: Wrap-Up and Next Steps | 12 |
| Updated Scientific Questions | 12 |
| Review of Progress on Data and Calibration | 13 |
| Group Discussion 5: Plan for Disseminating Results | 14 |
| Group Discussion 6: Strategies to Ensure Success | 15 |
| Appendix | 15 |

Day 1: Introductions, Project Goals, and Protocols

Modeling Group Presentations

APSIM Model

- APSIM is a modular agricultural system simulation framework coupling soil, crop, atmosphere, and management processes. The model started with simulating point-level soil X crop X management effects.
- APSIM evolved from a soil and crop model into a full agricultural systems model. It is process-rich and particularly strong in integrating crop growth, soil water balance, and residue-carbon processes. It is now flexible across scales: field scale (50% efforts), national scale (40% efforts), and global or sub-field (10% efforts).
- APSIM Classic is being phased out, and Next Gen is now the primary development platform. The model is open-source.
- Challenges and key interest in Farm-MIP: Simulate complicated farm practices; better quantification of uncertainty; parameterization at a sub-daily time step; new experimental datasets.
- Recent developments: biochar module (very unique); water logging and water table simulations; more detailed corn and soybean phenology.

APEX & EPIC Models

- EPIC is mainly developed as a field-scale agroecosystem model operating at a daily/ monthly time step.
- APEX is a “collection” of EPIC as it extends EPIC to watershed and landscape scales by routing outputs from multiple fields. Source codes are similarly formatted for APEX and EPIC, but the detailed soil processes are different.
- Both EPIC and APEX can simulate the impact of topography and primarily focus on hydrology and plant growth simulations.
- The core EPIC/APEX philosophy is scenario-based “what-if” analysis to evaluate alternative management options, predict environmental impacts, and support yield and production assessments for decision making.
- APEX has been used in USDA’s CEAP project to evaluate conservation management practices. The CEAP project leverages NRI points and has a timeline of every 10 years (5 years of survey + 5 years of modeling).
- Challenges and key interest in Farm-MIP: “what-if” scenario testing; evaluating productivity-environment tradeoffs.
- Recent developments: CH₄ module; adaptation of CENTURY for the soil module; enhanced freshwater modeling capacity; routines for biochar.

DayCent Model

- The model has been applied across different scales. Key applications are COMET-FARM at the field scale and the U.S. national inventory.
- The model structure follows a production → litter → surface SOC → active/slow/passive SOM framework with paralleled mineral N pools. Gas emissions are sensitive to WFPS.

- C14 data can be used to improve representation of soil carbon pools.
- The calibration framework has evolved from manual approaches to global sensitivity analysis (Sobol), Bayesian calibration, and Monte Carlo based uncertainty propagation, supported by long-term experimental and large-scale (global) datasets.
- The official version has been maintained by the CSU Soil Carbon Solution Center for ~3 years and is freely accessible for research, with fees only applied to commercial carbon credit uses.
- Total N₂O magnitude is captured reasonably well, but nitrification vs. denitrification contributions need to be further refined. The model has focused more on C sequestration than the magnitude of SOC.
- Challenges and key interest in Farm-MIP: Answer research questions related to regenerative agriculture.
- Recent developments: NASA-funded SIF/LAI modeling; incorporation of freeze thaw dynamics to better capture nutrient and gas flux cycling; related MEMS version can account for measurable C pools.

DNDC Model

- DNDC is a biogeochemical model originally developed to mainly focus on simulating greenhouse gas emissions. The model was later expanded to represent detailed carbon and nitrogen dynamics across ecosystem types.
- The model uses substrate availability, soil environmental conditions, and microbial activity to jointly regulate GHG emissions. Natural and management factors are used to simulate microbial activities.
- The model has been applied across different scales (field, regional, national)
- DNDC uses a relatively generic crop module rather than crop-specific parameterizations, which allows simulation across multiple crop types and vegetation systems.
- DNDC licensing has been transferred from UNH to a private company, which now manages development and distribution of the model.
- DNDC is currently sponsored by CARB to conduct inventory work for California.
- Challenges and key interest in Farm-MIP: Unique collaboration opportunities with other modeling groups; new datasets for model testing; uncertainty quantification and process-based understanding.
- Recent developments: nitrifier denitrification, manure DNDC; whole farm GHG accounting and decision support tools.

DSSAT System

- DSSAT is a crop model community system originated as a crop-driven modeling framework integrating multiple crop types. DSSAT continues to evolve through improved representation of soil and plant processes.
- DSSAT is an ecosystem of crop model users, trainers, and developers.
- DSSAT source codes are fully open source under a BSD license.
- A core DSSAT principle is that new crop models must be validated against multiple real field experiments before being added to the system. It is very important that the models are calibrated and evaluated for local varieties.

- The modeling system uses a modular architecture inspired by APSIM and integrates process representations from models such as EPIC and CENTURY. The plant modules are based on CERES crop growth concepts.
- The DSSAT group is working to organize the broader community to advance model development. There is also an annual training program.
- The input focus has shifted from standardizing model input formats to defining minimum data standards at the data collection stage to improve usability of experimental data for model evaluation and comparison.
- Challenges and key interest in Farm-MIP: model improvements; leverage recent updates from other models (e.g, DayCent soil algorithms); good quality data for SOC and GHG; initial conditions; stimulating discussions.
- Recent developments: hackathons and open forum discussions; tree and additional crops added to the system.

SALUS Model

- Started in the 90s with a goal of simulating tillage impacts. The original model included pool and decomposition components similar to DayCent and crop representation similar to DSSAT, but recent developments have caused it to diverge from both DayCent and DSSAT.
- SALUS is a process-based crop model designed to simulate interactive impacts of soil, crop genetics, weather, and management impacts.
- SALUS uniquely simulates changes in BD and effects of intermittent tillage.
- Key applications have been conducted across scales. At the regional scale, a multi-model ensemble was conducted to simulate cover crop and tillage impacts. At the national scale, SALUS is used to simulate CRP grasslands and marginal lands.
- Challenges and key interest in Farm-MIP: landscape-scale management impacts; downscaled results to support farm decisions; novel farm management.
- Recent developments: new applications (e.g., supporting farmer decision making, digital agriculture for large-scale management recommendations, and evaluation of bioenergy crops); crop-livestock system and grasslands; incorporation of remote sensing.

SWAT+ Model

- The model is fundamentally a hydrological ecosystem model, primarily used for large-scale water resource quantification, environmental assessment, and scenario analysis.
- Ongoing SWAT+ efforts aim to provide simulation results for the entire U.S. The modeling group currently collaborates with Field-to-Market and is expected to utilize datasets from the program.
- A global version of SWAT+ is already available as open-source (in Github) and is under continuous development. Community-based development is encouraged, and gatekeeper keeps track of newly implemented changes.
- SWAT+ can uniquely simulate impacts of heavy metal and pathogens.
- The carbon module is currently in testing and development; it can simulate terrestrial carbon cycling but will need to be improved for simulating cross-system carbon transportation.

- SWAT+ transitioned to an object-based modular architecture, which enables flexible simulation of interactions among soil, plant, and management components, as well as multi-plant competition and farm-scale field connectivity. The new structure also enables explicit spatial representation of Hydrologic Response Units (HRUs).
- Challenges and key interest in Farm-MIP: microbial and spatial processes; C module improvements.
- Recent developments: water allocation module; open-access API; broad-scale applications.

TASC Model

- TASC model (SWAT-Carbon) is developed upon the SWAT 2012 version and is an open-source model.
- TASC is a land-water coupled process-based model that integrates terrestrial and riverine carbon, hydrology, and biogeochemical processes. The model aims to provide a holistic understanding of landscape-scale processes and land-river hydrologic and biogeochemical linkages.
- The model maintains the SWAT temporal and core process framework while extending the system through additional physics and biogeochemical modules. Particularly, the terrestrial C algorithms leverage components from Century, DNDC, DSSAT, EPIC, and recent literatures.
- The model supports simulation from individual fields, to farms composed of interacting fields, and further to watershed-scale systems.
- An automated calibration tool (R-SWAT) is developed.
- Challenges and key interest in Farm-MIP: collaboration; identify gaps in SWAT-C; simulate regenerative practices.
- Recent developments: nitrogen module; incorporation of model components representing more complicated processes (e.g., microbial efficiency-type models); disease-related parameters and electron transport process representations; TASC-Photo (detailed photosynthesis processes).

Guided Group Discussion 1: Farm-MIP Draft Protocols

- **What can be standardized:** Input datasets, output variables and resolutions; management and weather simulations; model calibration guidance; model intercomparison, ensemble, and inter-model uncertainty analysis; required documentation of model structure, parameters, calibration, and improvement; open-source status.
- **What cannot be standardized:** Model spin-up length and algorithms; default params & params for adjustments; intra-model uncertainty analysis methods; computer platform and coding language; model calibration algorithms and methods; upscaling methods at the broader scale; specific aspects for model improvement.
- **Simulation procedure:**
 - Phase 0: internal testing;
 - Phase 1: uses site-level calibration and evaluation based on space (site-to-site), time (year-to-year), and treatment (management-to-management) splits;
 - Phase 2: complete site-level calibration and evaluation.

- **Consideration of potentially “turn-off” some model functions for intercomparison:**
 - Turning off model functions should be done cautiously to avoid distorting core processes, and suggested running the model in full mode first to evaluate impacts.
 - Example of turning off selected functions: N stress for DSSAT simulation. Other examples are associated with unique procedures that only some models have (e.g., pesticide use, river processes)
 - Another example related to accounting for soil erosion impacts was mentioned.
 - DSSAT has different versions and for intercomparison standard configurations should be applied.
 - Look into temperature-based sensitivity since models all have such representations, but Gerrit mentioned that the representation is slightly different from some other models.
 - Instead of proposing turning off some modules, we can consider this as a potential model sensitivity test where different modules can be experimented.
 - **Summary:** We could potentially consider this down the road, but it needs to be done with a lot of caution.
- **Calibration data representation:**
 - Verification is very different and is about code checking. Suggestion that we use "evaluation" rather than “validation.”
 - Need to assess dataset representativeness before dividing calibration and validation sites, which can ensure that major and combined management practices are covered in the calibration dataset so that independent evaluation runs are rigorous.
 - Testing model performance on new management practices not included in calibration sets is also useful to assess generalizability.
 - Project is unique in the sense that we involve soil, crop, and hydrological models for the first time, so we should test out the baseline performance of different models. Sotirios also mentioned that crop model calibration starts with full N rate (non-stressed treatment).
 - **Summary:** We will use “evaluation” rather than “validation” as our standard language for Farm-MIP. We should try out uncalibrated runs for all sites to assess baseline model performance. For splitting management into calibration and evaluation, the calibration set must have non-stressed levels.
- **Model output:**
 - Discussed feasibility of encouraging model groups to provide as many output variables as possible (e.g., mineralization, respiration, biomass, ET) to support outlier detection and broader model accuracy evaluation.
 - Importance of going beyond N₂O for the N cycling; instead, N leaching, plant N uptake, and soil N levels should also be compared.
 - Suggestion that the focus can be placed on calculating annual N budget, C balance, and water balance.
 - Consider leveraging ET and moisture datasets developed by AgMIP Soybean.

- Questions raised: (1) Do we have observations to evaluate all the processes? No but it's still valuable to see differences; (2) Can we use results to diagnose outliers and identify trends? Yes but we also need to be careful especially when there are no observations, as the general agreements from some models might still lead to bias.
- **Summary:** We should encourage modeling groups to generate as many outputs as possible that are related to the C, N, water, and crop components.
- **Required documentation:**
 - Modeling groups could try to maintain a stable model version throughout the project, but improved versions should also be allowed and encouraged given clear documentation.
 - Encouraged modeling groups to publish up-to-date handbooks with DOI, but this might be a high bar. Alternative options include key references, documentation, videos, and variable tables.
 - Scripts for data conversion are encouraged but not required.
 - Check out the Kimball publication about what was documented.
 - Outputs should be clearly defined (e.g., soil depth), and model implementation steps and process assumptions should be traceable and documented.

Breakout Session 1: Defining Farm-MIP Consortium Goals and Scope

- **Farm-MIP goals:** We should focus on producing high-quality, decision-relevant simulation outputs and evaluating the added value of ensemble approaches across scales. The main scientific value of model intercomparison is identifying process-level pathway differences that explain performance differences across models.
- **Consideration of standardized datasets and calibration:** Standardized data and protocols are prerequisites for intercomparison, ensemble analysis, and model improvement. Major model groups prefer to retain their own calibration approaches, while agreeing on harmonized data inputs and frameworks for splitting calibration and evaluation sets (e.g., by location and time period).
- **Uncertainties of model intercomparison:** Multi-model intercomparison will be conducted collaboratively, but multi-scale interactions remain a key uncertainty. At regional scales, watershed environmental differences may drive systematic differences in model outputs. As simulations move from point to larger spatial and temporal scales, models that do not represent lateral flows may diverge from those that do, leading to systematic differences in results at watershed and regional scales. Model performance may be domain-dependent (i.e., soil, crop, and water).
- **Benefit farmers and ranchers:** Near-term benefits are expected through strengthening existing model-to-stakeholder information pipelines, while longer-term outcomes may include enabling next-generation decision support tools. The ensemble tool may be further developed to help improve the assessment of regenerative agricultural practices for farmers and ranchers. We should organize meetings/ workshops with stakeholders down the road to better engage them for model co-improvement.

Day 2: Datasets and Model Inputs

Group Discussion 1: Data Sharing Agreements and Protocols

- **What principles should govern the use of open-access versus restricted datasets to ensure transparency, fairness, and appropriate use?**
 - Data sharing and accessibility (FAIR: findable, accessible, interoperable and reusable): The data are available under a free license for research use; the main limitations concern redistribution of the data and potential commercial use. Data access should require a simple formal application.
 - Authorship and collaboration: Data providers should be properly cited, and we also plan to offer co-authorship to encourage active engagement of data providers to help fill gaps in key site information (e.g., management history and experimental details). The consideration might be slightly different for sites that have already been published for a long term and widely used by the modeling community. The existing framework used in the CREDIT system can serve as a reference, in which contributions are transparently documented and authorship eligibility is determined based on defined participation thresholds. We should include data providers in a data paper related to Farm-MIP. For the rotation system in the protocol, we should change the language to include at least one from each modeling group as co-author for “external” publications. This also depends on the level of model granularity: if individual models are shown by name, model team involvement is required; if only aggregated outputs are used, coordinator-level involvement may be sufficient.
- **What protocols/safeguards are needed to protect data privacy during the project and after its completion?**
 - We may sign memorandums of agreement or non-disclosure agreements if we consider datasets that are not open access, but for now we should focus on open-access datasets.
- **Which platforms are suitable for hosting and sharing standardized datasets and model simulation results?**
 - GitHub; if a DOI is not required, public platforms such as DKRZ or ISMIP archive can be used.
- **AI policy:** Need to include an AI policy for Farm-MIP.

Group Discussion 2: Candidate Calibration Sites

- **Dataset size discussion (10 cropland sites and 5 grassland sites):**
 - Concern that the dataset is too small to support robust inference and serve both calibration and evaluation purposes. It would be preferable to focus on prioritized practices given the data size. For DayCent national inventory, 15 sites were used for calibration and another 65 were used for evaluation.

- Current dataset size is driven by three main factors: most modeling groups are comfortable working with 10-20 sites; site eligibility is constrained by the requirement to include both N₂O data and long-term SOC records. Ten sites with multiple measurements currently form the core dataset, with additional datasets to be prepared; some site PIs have never responded to emails so it is hard to confirm data availability.
- There are sites that meet the criteria but were not on our current list.
- Suggestion that ten sites should be sufficient and suggested that model recalibration may not be required given all 9 models have been well calibrated in the past. In this case, we can consider using these sites directly for model evaluation.
- For data acquisition, it was suggested that we should request each modeling group to continuously contribute their commonly used core datasets, and that we need to further expand the current dataset using long-term datasets from the published literature.
- May focus more on croplands given the data gap – suggested that we should keep at least one or two grass sites which ideally represents hay/pasture land use.
- Suggestion of voting for top three sites to get started with. Voting results of sites: KBS (7), Lincoln, NE ARS site (6), Purdue GraceNet site (3), Morrow plots (3).

Group Discussion 3: Datasets for Regional- and National-Scale Modeling

- **Point-based evaluation data sources:** A combination of unused site-level observations, network or survey-based datasets (USDA NASS, SOLUS/NASIS, RaCA, DSP4SH, CASH, NAPESHM, and others), meta-analysis datasets for SOC and N₂O, and other large-scale datasets (e.g., aboveground biomass, soil moisture, C fluxes). The USDA dataset connecting SOC with management might be helpful, but initial conditions are still unknown so the evaluation is more focused on assessing spatial representations.
- **DayCent's future dataset:** A dataset within the DayCent group is currently under review that provides quarter-degree gridded data from 1979 to 2023. The dataset includes cropland management information, such as fertilizer application, manure use, and tillage practices. It is open access with crop-specific information available. Future data updates are still under discussion with USDA, and they were previously planning on a two-year update cycle. Other groups are very excited about using the dataset to run the “counterfactual simulations” at the large scale.
- **Large-scale calibration:** There is likely not sufficient soil data to support calibration, but we can potentially use NASS yield data and AgMIP crop calendar for calibration associated with the crop modules.
- **Differences between site-level and point-based large-scale analysis:** Site-level calibration relies on long-term, rigorous datasets, while point-level evaluation at the national scale can use datasets with broader spatial coverage.

- **Resolution:** A 12-km grid resolution is considered most feasible based on capacity of all 9 models; grid resolution is often constrained by weather inputs (many models using PRISM data at a 4-km resolution). Pixel/grid-based simulations can be challenging for EPIC/APEX/SWAT+ for contiguous U.S. considering the computational power needed and that the models are focused on using watersheds as modeling units.

Breakout Session 2: Farm Practice Simulations

- **Which practices should we prioritize at the three project scales? No-till, cover crop, fertilizer, others?**
 - There is a list of 15 regenerative agricultural practices (cover crops, rotation, foraging, contour farming, mulching, etc.) at USDA. We should consider aligning our simulations with this list.
 - **Group 1:** Fertilizer management (amount, type, timing), cover crop (stratified by main region), and tillage should be priorities. The simulations should also be region-specific (e.g., LRR) based on farmer and NRCS feedback. We also need to determine how many of the ten sites can be used for site-level evaluation of the same practice.
 - **Group 2:** Prioritize a well-defined research question targeting problems relevant to specific regions and stakeholders. This can help us finalize the farm practice simulation list. Starting with simple scenario runs, sharing preliminary results for feedback, and then iteratively refining and expanding scenario analyses. The majority of detailed scenario analysis can be done at the field scale, and we should keep national scale simulation simpler.
 - **Group 3:** Simulations should be region specific and based on NRCS/farmer recommendations. For large scale simulation, begin with point-based simulations at representative locations within a region, focusing on a limited set of key management practices, and then progressively scale up to regional analyses based on the results. For upscaling simulation, it would be useful to create a bundle of practices.
- **How do we handle combinations of practices?**
 - **Group 1:** Farmers often think about a combination of management practices together rather than separately, so we should simulate these combinations. Our nine models should be able to handle a combination of management practice simulation pretty well. Outputs from such simulations can be huge, so we need to have a reasonable number of combinations and think about leveraging AI for output analysis.
 - **Group 2:** Some management practices (e.g., cover crops and no-tillage) are inherently embedded within crop rotations and cannot be fully isolated. We should clearly identify the outputs of interest (e.g., year-scale analysis, cover crops and no tillage with crop rotation) and ensure that all models are able to simulate these outputs consistently.
 - **Group 3:** Feasible and factorial designs could be applied at representative points and area to disentangle individual and combined effects.

- **Which weather scenarios should we prioritize? Dry vs. wet, hot vs. cold, others?**
 - **Group 1:** Run the 30-40 year historical baseline, analyze historical weather patterns (e.g., dry versus wet periods), and then apply clustering analyses to identify patterns. The weather scenario might be designed to be regional-specific (e.g., wet spring and dry summer in the Midwest). Consider applying a +1 °C temperature perturbation for model simulations. The weather generator could also be used.
 - **Group 2:** Focus on historical simulations (e.g, 30 years) to identify patterns and evaluate the interactions between weather and farm management.
 - **Group 3:** The primary focus should be retrospective analysis, followed by probability-based analyses (e.g., the likelihood of dry versus wet conditions in the coming year). Future climate change scenarios should largely be avoided in USDA projects, though one test scenario could be included.
- **Additional discussion:**
 - Consider computational limitations and research novelty. In addition to common practices, exploring high-efficiency fertilizers, biochar, and other emerging agricultural products. Biochar was repeatedly highlighted for its novelty, but not all groups are comfortable with their existing biochar module.
 - Suggestion that we should target one specific management practice for a paper within a year. Ideas related to avoiding repeating his paper is to focus on cover crop species, manure amendments, and pasture simulations. Meanwhile, some other PIs mentioned that they don't mind simulating major practices even though they are already done in *Basso et al.*
 - Before moving to regional-scale analyses, the project should focus on intensive point-based simulations in a few selected regions. For example, in some regions, irrigation may be more important for crop yield and soil carbon dynamics than other management practices.
 - The concern was raised that genetic modules may not be ready in many models. Therefore, we should not require all models to perform the same tasks related to crop simulation, but instead we should help identify which models are currently capable and where further development is needed.
 - Models should not be forced to simulate processes they were not designed to represent. A framework allowing different model types to contribute their respective strengths was proposed. In this case, we should focus more on how models diverge in their simulation results related to soil, water, and C cycles.

Guided Group Discussion 2: Input Data Quality Control and Gap-Filling

- **Weather data:**
 - Required: Min and max temperature, precipitation and solar radiation.
 - For gridded based data, most of these datasets are fairly well quality control and gap filled; for station data, some might be messy and we should be cautious.

- At the site level, station data are often the most commonly used source. Precipitation data are generally more reliable in station-based datasets. For minimum and maximum temperature and solar radiation, station data may require additional quality control.
- A good dataset called AgMERRA was mentioned for the weather data, especially for solar radiation.
- For the acceptable distance between field sites and weather stations, it was suggested to first review the existing calibration sites and confirm whether they have onsite weather measurements. At one site, the weather station is located more than 10 km away; at all other sites, it is generally within the 5-km range.
- Regarding weather data, several datasets presented earlier extend sufficiently far back in time, which should allow the spinup period to be extracted as needed.
- **Soil data and model initialization:**
 - **Model initialization**
 - Spin-up is ~10-20 years for most models. Models like DayCent require much longer time for spinup.
 - Many debates about whether initial conditions should be standardized: Model evaluation should focus more on carbon stock changes rather than absolute carbon stock values. In this case, starting from the same initial condition is less important; In an existing paper, the initial condition was set to be the same for different models; Large differences in initial conditions should be avoided to ensure comparability.
 - Temporary conclusion is that shared spinup datasets should be provided, but individual modeling groups should determine whether and how to conduct initialization.
 - **Bulk density**
 - Regarding the observed uncertainty, using a pedo-transfer function should be considered fine, but selecting a single bulk density value across different depths without a defensible range would not make sense.
- **Crop and management data:**
 - For site-level simulations, when using experimental sites reported in the literature, we should contact data providers to obtain additional management and site information. If gaps remain, national inventory databases can be used to infer representative management histories for the region.
 - It was also emphasized that crop processes should be represented beyond yield alone. At minimum, planting and harvest timing should be specified (for models like DayCent and DNDC), and detailed phenology (e.g., using crop calendar data) is often required for crop models.

Group Discussion 4: Refine Phase 0 Model Input Files

- At least 20 years of land use history should be collected for each site, assumptions should be made based on regional/national land use for sites lacking such information.

- A fixed and transparent methodology should be used for gap-filling and flux integration across missing observation periods. Clear definitions of variables and units are needed. The AgMIP-soybean template should be adapted for this project.
- We need to specify water management details (irrigation and drainage).
- Consider adding ET/PET data.
- Need to make the depth information clearer.
- Consider using .csv rather than .xlsx to meet FAIR standards.
- Check soil moisture data for potential unit error.
- Missing data should be confirmed with site managers where possible. A defined response time threshold should be established, after which a fallback plan should be carried out.
- For soil type and related variables, data should not be taken solely from databases (e.g., GRACEnet). Information in relevant literature should also be double checked. While here a known issue is that cited publications do not always correspond to the experiments that actually generated the data. Therefore, citations cannot be assumed to represent reliable data provenance.
- In addition to the minimal dataset, commonly used “good-to-have” variables across modeling groups should also be prepared.
- N₂O data should preferably include annual or seasonal total emissions and standard deviations to allow calibration.

Day 3: Wrap-Up and Next Steps

Updated Scientific Questions

1. To what extent do existing agroecosystem models consistently capture the influence of regenerative agricultural practices on soil health and N cycling?
2. When and where can regenerative agricultural practices sustain or increase agricultural productivity while improving water and N use efficiency?
3. How do we leverage AI, field observations, and model ensembling to improve simulation of weather-driven variability and its interaction with regenerative agricultural practices?

Scientific Question Refinement Discussion

- **Additional points to incorporate:**
 - Emphasize “regenerative agriculture” in framing scientific questions for this project.
 - Mention edaphic characteristics and specifically note differences among soil textures. We discussed that “environmental factors” can be used to represent both soil and weather factors.
 - Question 3 could be made more project-specific by highlighting approaches such as model-data fusion to improve model accuracy.
 - Questions should follow a hierarchical progression, with the third question perhaps extending the second one toward actionable guidance (e.g., improved monitoring) for farmers and ranchers.

- **Details to expand upon:**
 - Technical details should be expanded in the next phase of the project: For example, what management interactions and what kind of weather conditions would we simulate, how many years should we simulate, and whether these simulations will focus on evaluating yield stability over time.
 - We should expand the project to focus beyond SOC and N₂O. Consider focusing on broader nitrogen cycling processes and loss pathways (e.g., volatilization, NO_x and effects of nitrogen deposition, suggested by Limei).
 - About AI utilization, we should discuss more about the targeted methodologies, not only model ensembling (weigh outcomes) and big data analysis, but also the potential to develop a new AI technology and find the optimal solution for ranking sensitive variables, which can be helpful for stakeholders. We can also leverage AI for data gap-filling, model upscaling, and diagnosing where model outputs vary among groups. Keeping “AI” as a key word in the scientific questions and objectives are important for USDA.

Review of Progress on Data and Calibration

- **If given three sites, what would you do with model calibration? What about additional sites beyond the first three?**
 - DayCent plans to only conduct uncalibrated runs for the three. If the site number is sufficient in the end (a higher number than we currently have), then their group may consider tuning soil parameters at the regional or national scale.
 - SWAT family models can utilize the data to conduct site-level calibration. They may conduct regional-scale calibration, but it has not been decided.
 - The crop models would like to focus on site-level calibration. APSIM can also run calibration with all sites provided.
 - All modeling groups mentioned that they prefer to run ten sites together, rather than three first and then seven sites. However, for practical reasons (it will take a while to get data for the ten sites), we may adjust the strategy.
- **Representation:** Consider whether existing candidate sites can fulfill the intended farm practice simulation, since they may not represent livestock grazing and other regenerative practices well.
- **Views on calibration:**
 - Some modeling groups believe a unified calibration across multiple sites (i.e., opposite to tuning parameters for each site) is needed to ensure large-scale representativeness. This is particularly the case for DayCent as their models have already undergone regional calibration using a larger dataset, so they would like to skip “site-based calibration” in this case.
 - The crop and SWAT family models still require site-specific calibration for tuning parameters such as those related to crop genetics. As to whether they will also conduct “regional-scale” calibration, the answer is still uncertain – modeling groups would like to think through the procedure more when we scale up.

- For model calibration sequence, SWAT family models follow the water > soil > crop sequence while DayCent tune different parameters simultaneously using Bayesian calibration.
- Based on the agreement and disagreement so far, it was proposed to firstly run a blind simulation without any calibration to make sure all models are on the same page, and then we can leave it to the modeling groups to decide whether they will conduct site-based calibration and how they plan to do so.
- **Representativeness of calibration design:** Upscaling strategy should be considered alongside calibration design to ensure that site selection and calibration approaches can support future regional applications
- **Plot vs. field scale:** We should differentiate plot scale and field scale, for example soil survey data such as gSSURGO cannot be used to accurately represent plot-level spatial variability but may be used at the field scale.
- **Citation:** The importance of referencing data providers' papers was emphasized.
- **Data standardization:** Discussed the importance of aligning future data submissions with FAIR principles and existing data standards to improve data quality, sharing, and long-term usability. Suggested that we send out a model output template, which was agreed upon by other modeling groups.

Group Discussion 5: Plan for Disseminating Results

- **Tier-2 models and involvement:**
 - To keep the protocols and datasets open access for the community remains the long-term goal for Farm-MIP, but safeguard should be in place for participation of Tier-2 modeling groups in case they publish before our core groups. In this case, a staged data sharing strategy should be used, with coordinated input data access potentially restricted until agreed milestones are reached, while model outputs may require stricter controls.
 - Engaging new groups through Tier-2 participation should prioritize adding new models rather than duplicating existing models.
- **What stakeholder groups could be engaged to maximize reach and usefulness?**
 - Engage stakeholders early by clearly communicating project goals and data needs. We may also conduct a survey at some point.
 - Reach out to the community actively to identify existing high-quality datasets.
 - Focus external messaging on research and management outcomes rather than modeling details.
 - Build engagement in layers from scientific communities to agencies and producer groups, and design platforms and workflows to support long-term data sharing and feedback.
 - A nitrogen fertilizer application decision support tool (n-fact.ag/start) was presented as an example of a farmer-oriented decision tool and different scenario comparison.
 - Can perhaps modify one research question to go beyond model performance. Consider emphasizing tools and applications that will come out of this project.

- **Peer-reviewed papers: Literature review, major results**
 - Research questions should define the primary publication targets, with at least one paper expected for each question. Model-focused research questions will require simulation outputs and may lead to multiple papers across project phases.
 - Timeline planning will be an important next step. Literature review papers are expected earlier (e.g., within a year), while model application and regional prediction papers will likely follow as results mature.
 - A data paper should consider integrating metadata availability across field, regional, and national scales. It may be more appropriate to consider it as a meta data paper because we need permission to share the full dataset. We should invite data providers for this paper.
 - Protocol papers are expected later as methods evolve, followed by a potential final synthesis paper.
 - It was suggested that we send out a survey or shared table early on to collect potential paper topics as a next step.

Group Discussion 6: Strategies to Ensure Success

- Due to differences across model types, analysis of uncalibrated model outputs should be conducted first, followed by comparison after applying model-specific calibration approaches.
- Temporal resolution, simulation length, and output variables should be revisited in future meetings, with continued communications expected through emails and all group meetings.
- A key metric of success is to be able to obtain feedback from targeted stakeholders.

Appendix

Slides: [Introduction to the Farm-MIP Project](#)