T.R. Miles Technical Consultants Inc.

5475 SW Arrow Wood Lane, Portland OR 97225 Phone: (503) 780-8185

www.trmiles.com email: tmiles@trmiles.com

July 9, 2026

Search notes: AI in Soil Science and Biochar

How can AI be used to optimize the formulation and use of biochars and biochar amended products for soils, soil carbon, soil microbiomes? What types of AI tools can be deployed by a biochar producer to make biochars and biochar-based products that are suitable for purpose in crops, remediation of contaminated soils, and water quality? Provide a use-case oriented summary for integrating this paradigm into biochar/carbon removal project design.

AI can be used end-to-end in a biochar value chain: to design feedstock/pyrolysis conditions for targeted products, predict field performance (SOC, yield, microbiomes), and optimize application strategies and MRV for carbon removal projects.[1][2][3]

Below is a use-case oriented summary, focusing on tools and workflows a biochar producer or project developer could realistically deploy.

1. AI for product formulation (designing "fit-for-purpose" biochars)

a. Predicting biochar properties from feedstock and process

Multi-task machine learning can predict multiple key biochar properties (pH, ash, surface area, C/N, aromaticity, nutrient content) from biomass composition and pyrolysis settings. In one recent study, a multi-task XGBoost model trained on 666 experiments achieved $R^2 \approx 0.90$ across nine soil-relevant properties, with final temperature the dominant driver and heating rate relatively minor.[1]

Relevant tools:

- Gradient boosting models (XGBoost, LightGBM, CatBoost) for:
  - Forward prediction: given feedstock + process, predict resulting biochar properties.
  - Reverse optimization: given desired soil functions (liming value, permanence, nutrient profile), solve for feasible combinations of feedstock blends and process parameters.[1]

- SHAP and partial dependence plots for interpretability:
  - Identify which feedstock traits (proximate/ultimate analysis) and process parameters most strongly control agronomic and carbon permanence properties.[1]

Use case:
A producer wants a manure-derived biochar that:

- Raises pH in acidic soils,
- Has high fixed C and low H/C for permanence,
- Provides moderate P and K.

They can:

1. Build or adapt a multi-task ML model using their own proximate/ultimate analyses plus public datasets.[3][1]
2. Run "forward" simulations on candidate feedstocks (e.g., dairy solids + straw) and temperature ranges (450–600 °C).
3. Use "reverse" optimization to identify a narrow operating window that maximizes fixed C while keeping ash and liming capacity in a target band, then validate experimentally.[1]

2. AI for predicting soil organic carbon outcomes

a. Ensemble ML for SOC response to biochar

An ensemble ML framework (ExtraTrees + LightGBM + CatBoost) trained on 800 field observations from 104 studies predicts biochar-induced changes in SOC (log response ratio) across soils, climates, and cropping systems with $R^2 \approx 0.86$. SHAP analysis shows biochar rate, crop type, soil type, and pH as dominant factors, with strongest SOC gains in degraded, low-SOC soils and within neutral pH ranges.[2]

Relevant tools:

- Voting/stacked ensembles (tree-based regressors):
  - Input: biochar properties, soil properties (SOC, pH, texture), climate (MAT, MAP), crop and management.[2]
  - Output: expected SOC response over a given time frame.

- Explainable AI (SHAP, permutation importance, PDPs):
  - Quantify marginal impacts and thresholds (e.g., SOC response saturates around 120 t/ha biochar; largest marginal gain around 40 t/ha).[2]

Use case:
For a carbon removal project, you need ex-ante estimates of SOC gains for a CDR crediting methodology, and ex-post comparisons as MRV.

You can:

1. Integrate an ensemble SOC response model into your project design stack to simulate SOC outcomes under different:
   - Biochar types and rates,
   - Cropping systems (legumes vs cereals),
   - Soil textures and initial SOC.[2]
2. Use SHAP/PDPs to identify:
   - Sites where marginal SOC gains per tonne biochar are highest (low SOC, fine-textured, near-neutral pH).[2]
   - Application rates beyond which returns diminish, to avoid over-application from economic and sustainability perspectives.[2]
3. Wrap the model in a simple GUI for agronomist and farmer use (similar to the Streamlit interface developed in the SOC study).[2]

3. AI for soil microbiomes and biological responses

Current models focus mostly on SOC and yields, but AI frameworks are increasingly proposed for microbiome-linked functions.[4][3]

Relevant tools:

- ML models linking biochar properties and soil conditions to:
  - Microbial biomass and diversity,
  - Enzyme activities,
  - Soil respiration and N cycling.[3][2]

- Deep learning and sequence-based models:
    - Predict protein structure/function from metagenomic data to infer microbial traits impacted by biochar (e.g., nutrient cycling under drought).[4]
- Multi-agent AI systems:
    - Agents for literature mining, hypothesis generation, and experimental design around biochar–microbiome interactions (e.g., how specific biochar formulations alter microbial networks in degraded soils).[4]

Use case:
A vineyard remediation project wants to leverage biochar to improve microbial resilience and disease suppression in metal-contaminated soils.

You can:

1. Use an ML model to correlate existing biochar trials (texture, functional groups, metals, DOC) with microbial diversity and enzyme data from your sites or the literature.[3]
2. Let a multi-agent AI system:
    - Mine microbiome and metagenomics literature related to similar soils and contaminants,
    - Generate hypotheses about which pore size distributions, pH, and surface chemistries favour beneficial microbial consortia,
    - Propose sampling and experimental designs (e.g., factorial trials for manure vs woody biochars in contaminated rows).[4]

4. AI for contaminant remediation and water quality

a. Predicting sorption and remediation performance

Dedicated ML models have been used to optimize biochar for heavy metal adsorption and broader pollutant removal. These models typically take biochar properties (surface area, pore structure, pH, functional groups) and solution/soil chemistry as inputs and output removal efficiencies.[5][3]

Relevant tools:

- Supervised ML (Random Forests, SVMs, Gradient Boosting):
    - Predict adsorption capacity for metals and organics from feedstock, process parameters, and post-treatments.[5][3]

- Gamma testing and other data quality checks:
  - Evaluate nonlinear relationships and suitability of ML for sorption prediction.[5]
- Deep learning architectures (CNNs, LSTMs, hybrid models):
  - Capture more complex interactions in multi-pollutant systems or time-series water quality data.[3]

Use case:
A biochar producer targets storm-water filters and mine-tailings remediation:

1. Build a sorption performance library for different biochars (including your products), with standardized data on:
   - Metals (Cu, Zn, Pb, Cd),
   - Anions (nitrate, phosphate),
   - Organics (PAHs).
2. Train ML models to:
   - Predict removal at given contact times, pH, and ionic strength,
   - Identify optimal activation/post-treatment (e.g., Fe/N doping for Cr removal, steam activation for organics).[3]
3. Use the model to design product lines:
   - “Agronomic char” for fields,
   - “Filter char” for water,
   - “Remediation char” for hotspot soils, each with documented performance curves output by the model.[5][3]

5. AI across the biochar project lifecycle (soil amendment + CDR)

Below is a project-oriented view combining the above elements.

a. Design and scale biochar production

- Use multi-task ML to define operating envelopes for different product families (manure chars, crop residue chars, mixed feedstocks) tuned for specific soil functions and permanence.[6][1]

- Integrate production optimization:
  - Yield vs energy use vs emissions using ML/AI to support continuous improvement and lower cost/tonne C sequestered.[6][3]

b. Site selection and agronomic planning

- Use ensemble ML SOC models and digital soil mapping tools to:
  - Identify sites with highest SOC and productivity response per tonne biochar.[4][2]
  - Prioritize degraded, low-SOC, fine-textured soils in climates where biochar's water-retention and liming effects are most valuable.[3][2]
- Employ multi-agent AI to decompose broader questions:
  - E.g., "How will biochar + cover crops + manure affect SOC and N2O at region scale under projected climate?" into climate, soil, crop, and management sub-tasks.[4]

c. Application strategy and rate optimization

- Use SHAP/PDP outputs to:
  - Optimize rate (e.g., around 40 t/ha where marginal SOC gains are highest, avoiding saturation around 120 t/ha).[2]
  - Tailor product choice and rate by crop and soil type (e.g., stronger SOC response in legumes and fruits; fine-textured soils).[2]
- Combine ML outputs with economic and logistics models:
  - AI frameworks can help with cost-benefit and life-cycle assessment, including transport, spreading, and revenue from credits and yield uplift.[6][3]

d. Monitoring, reporting, verification (MRV)

- Build or adopt AI-assisted MRV pipelines:
  - Integrate sensor networks, remote sensing, yield data, and periodic soil sampling.[3][4]
  - Use ML models to interpolate/extrapolate SOC, moisture, N2O, and contaminant dynamics between sampling points.

- Use explainable AI outputs (SHAP, uncertainty quantification) to:
    - Provide transparent MRV for carbon registries,
    - Quantify uncertainty ranges and attribute variance to climate, management, and product differences.[3][2]

e. Knowledge generation and continuous improvement

- Multi-agent AI systems can:
    - Continuously mine new literature and databases for biochar outcomes, update models, and identify knowledge gaps.[4][3]
    - Generate and prioritize hypotheses (e.g., about mineral-associated carbon saturation, microbiome-mediated stabilization) and propose experiments to refine your product line and agronomic recommendations.[4][2]

6. Practical tool stack for a biochar producer

A realistic architecture for you (TCI) or similar producers could look like:

Layer	AI tools	Purpose
Product design	Multi-task XGBoost / LightGBM; SHAP	Map feedstock + pyrolysis to key properties; optimize formulations.[1][3]
SOC/agronomy	Ensemble ML (ExtraTrees, CatBoost, LightGBM)	Predict SOC response, yield trends, rate optimization.[2]
Remediation/water	RF/SVM/boosting; sorption ML models	Predict pollutant removal; design remediation and filter products.[3][5]
Microbiome	ML models + metagenomics DL	Explore biochar–microbiome impacts; guide pH and surface chemistry choices.[4][3]
Project & MRV	Multi-agent AI + digital soil twins	Integrate climate, soil, management; design experiments and MRV workflows.[4][2][3]

If you were to start implementing this in one flagship project, what do you see as the single most constraining factor right now—data availability/quality, in-house ML capacity, or integration into existing agronomic and MRV workflows?

1. Multi-task-learning-assisted-in-predict.pdf
2. GCB-Bioenergy-2026-Ray-Predicting-Biochar-Induced-Changes-in-Soil-Organic-Carbon-With-Ensemble-M.pdf
3. Recent-advances-in-machine_deep-learning-frameworks-for-biochar-application-in-soil-amendment-an.pdf
4. fsci-4-1721295.pdf
5. Optimizing-biochar-for-heavy-metal-adsorption_-an-AI-driven-predictive-modeling-approach-with-Gamma-testing_-Biofuels_-Vol-0-No-0-Get-Access.pdf
6. Leveraging-machine-learning-for-sustainable-biochar-and-hydrochar-production_-Advances-challenges-and-perspectives-in-thermochemical-processes-Scie.pdf
7. 1-s2.0-S2950155526000613-main.pdf
8. sustainability-18-05203-v2.pdf

Please feel free to contact me at +1 503-780-8185 should you have any questions.

Kind regards,

Thomas R. Miles

Thomas R Miles

President
Director International Biochar Initiative 2015-2024
Founder and director United States Biochar Initiative 2010-2024 (Now American Biochar Institute biochar.org)
Sponsor and host Biochar Discussion Group 2006-2026
https://biochar.groups.io/g/main/topics