Linking Manure Management to Emissions: A Satellite-GeoAI Traceability Pipeline for U.S. Cattle Operations

Ashutosh Pawar[1], Vasit Sagan[1]*, Derek Tesser[1], Edward Boyda[2], Aaron Davitt[3]

[1] Department of Earth, Environmental and Geospatial Science, Saint Louis University, St. Louis, United States
[2] Earth Genome, Denver, United States
[3] WattTime, United States
* vasit.sagan@slu.edu

Keywords: bovine slurry index, cattle operations, GeoAI, manure ponds, remote sensing

Introduction

National livestock greenhouse-gas inventories estimate methane emissions from manure management using animal populations, climate, and the fraction of manure assigned to different animal waste management systems (AWMS). However, operation-level information on liquid manure storage remains limited, causing AWMS fractions to rely largely on national statistics, surveys, and regional assumptions (Dittmer et al., 2023; IPCC, 2019). This limits the ability of inventories to represent spatial variation in manure-management infrastructure.

Recent GeoAI approaches provide cattle-operation footprints at large spatial scales, creating a geographic framework for operation-level analysis. Yet these footprints identify the extent of an operation rather than the infrastructure within it and therefore cannot distinguish manure ponds from barns, pens, vegetation, soil, or other surface features. Satellite-based detection can provide this additional information, but spectral detections must be converted from raster pixels into geographically traceable objects and operation-level records before they can support spatial analysis or methane activity data.

Here, we use the Bovine Slurry Index (BSI-E), a Sentinel-2 spectral index developed for manure-pond detection within cattle-operation footprints (Pawar et al., 2026), as the upstream detection method. Its development and validation are reported separately and are not re-evaluated here. Instead, this study develops an operation-level satellite-GeoAI traceability pipeline that converts BSI-E detections into spatially explicit, auditable records and investigates an important spatial-support issue: whether the area of the operation footprint itself influences the probability of detecting a candidate pond. Using 4,525 cattle operations across California, Colorado, Kansas, and Texas, we aim to (1) derive traceable operation-level candidate-pond records, (2) characterize state and county spatial patterns, and (3) quantify how interstate differences in pond-positive prevalence change after accounting for operation-footprint area.

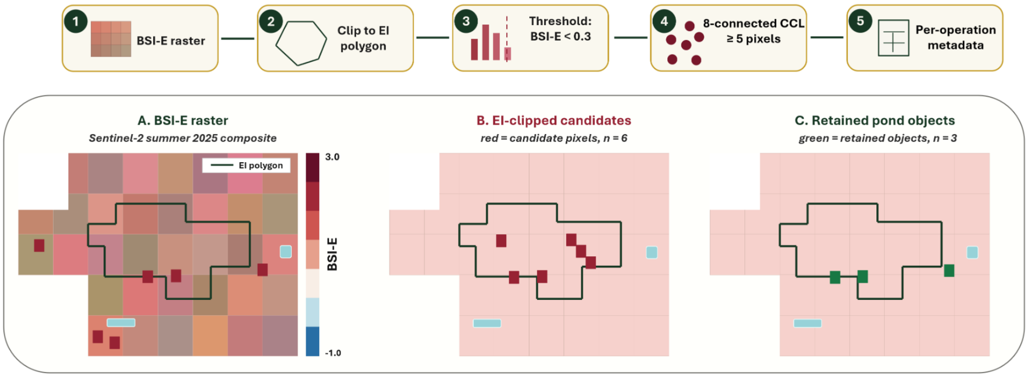

Figure 1: Satellite-GeoAI traceability pipeline for operation-level manure-pond mapping. Sentinel-2 BSI-E detections are clipped to cattle-operation footprints, thresholded, grouped using connected-component labeling, filtered, vectorized, and summarized into operation-level records. The raster grid is illustrative.

Method

We analyzed 4,525 Climate TRACE cattle-operation footprints across California (n = 1,100), Colorado (n = 827), Kansas (n = 1,103), and Texas (n = 1,495) (Davitt et al., 2025). Earth Index-derived footprints were treated as fixed spatial masks. For each operation, a warm-season Sentinel-2 surface-reflectance composite was clipped to the footprint, and BSI-E was calculated from B8A and B11 following Pawar et al. (2026).

The satellite-GeoAI traceability pipeline converted raster detections into operation-linked candidate objects **(Figures 1-2)**. Pixels with $BSI - E \; < \tau$ were grouped using eight-neighbor connected-component labeling; components smaller than five analysis-grid cells were removed and the remainder vectorized. Operations containing at least one retained component were classified as pond-positive. Here, pond-positive prevalence refers to the share of cattle operations within a given geographic unit that contain at least one retained candidate pond. For each operation, the resulting operation-level attributes included pond-positive status, retained object count, total candidate area, footprint coverage, BSI-E statistics, source footprint, imagery period, and geometry.

To evaluate whether operation-footprint area influenced pond-positive prevalence, we divided the pooled footprint-area distribution into quintiles and directly standardized each state's prevalence to the pooled quintile distribution. This allowed interstate comparisons under a common footprint-area structure rather than each state's observed footprint distribution. Raw prevalence was reported with 95% Wilson confidence intervals, while standardized prevalence and median footprint coverage among pond-positive operations were estimated using stratified bootstrap intervals.

We additionally fit a pooled logistic regression model relating pond-positive status to $log_{10}(footprint\ area + 1)$ to characterize the continuous association between footprint size and detection probability. Operation-level predicted probabilities from this model were aggregated to derive county-level expected prevalence and observed-minus-expected prevalence was mapped for counties containing at least 10 processed operations. These analyses were used to distinguish spatial patterns associated with footprint area from remaining geographic variation and were interpreted at the county level.

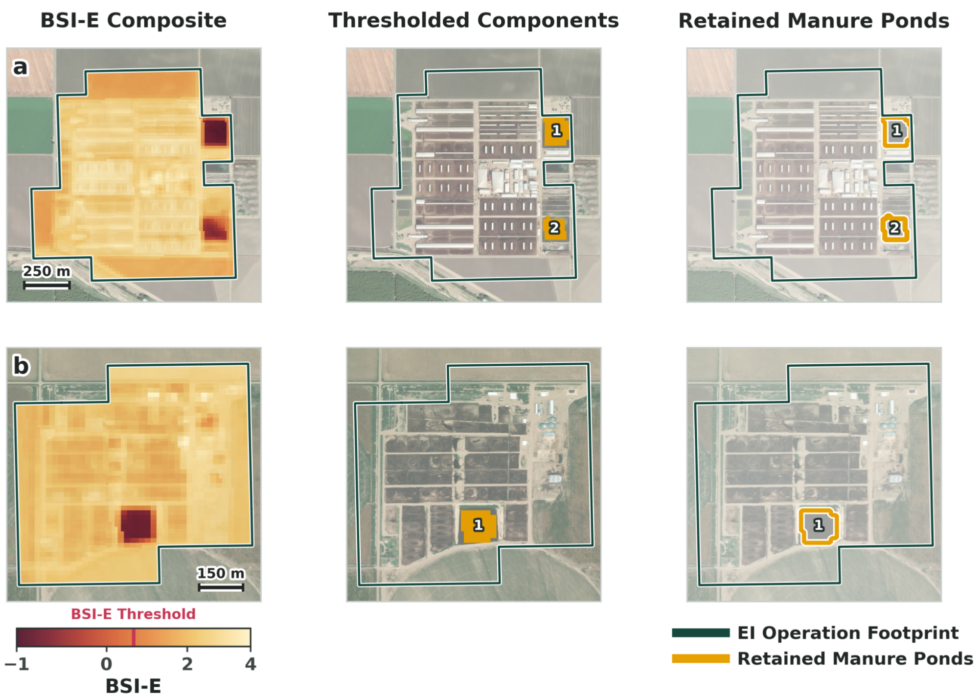

Figure 2: Representative operations showing the BSI-E composite, thresholded components, and retained objects after footprint clipping and size filtering. High-resolution imagery provides visual context.

Results

Across the 4,525 cattle operations, 1,576 operations (34.8%) were pond-positive, with 3,256 retained candidate objects totaling 2,105.1 ha. Raw pond-positive prevalence varied substantially, from 26.6% in Texas and 29.0% in Colorado to 31.7% in Kansas and 53.5% in California. Aggregate retained candidate area showed a different pattern: Texas contained the largest detected area (776.4 ha), followed by California (622.4 ha), Kansas (401.5 ha), and Colorado (304.8 ha). Thus, the frequency of pond-positive operations and the total mapped candidate area captured distinct dimensions of the operation-level result.

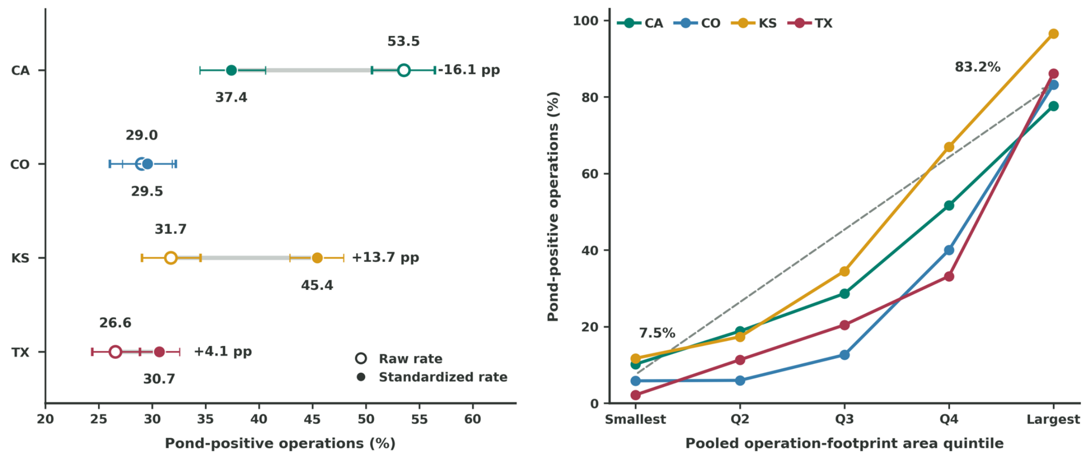

Figure 3: Influence of operation-footprint area on interstate pond-positive prevalence. The left panel compares raw prevalence with prevalence directly standardized to the pooled footprint-area distribution; the right panel shows pond-positive frequency across pooled footprint-area quintiles.

Operation-footprint area strongly structured these interstate differences **(Figure 3).** Across the pooled footprint-area distribution, pond-positive frequency increased from 7.5% in the smallest quintile to 83.2% in the largest quintile. California had a substantially larger median operation footprint (43.5 ha) than the other three states (15.3-20.5 ha). After direct standardization to a common footprint-area distribution, California's prevalence decreased from 53.5% to 37.4%, whereas Kansas increased from 31.7% to 45.4%. Colorado changed from 29.0% to 29.5%, and Texas from 26.6% to 30.7%. Consequently, the raw interstate ranking was not preserved after accounting for differences in operation-footprint area.

County-level patterns showed a similar spatial-support effect **(Figure 4)**. Counties with high raw pond-positive prevalence frequently coincided with larger mean operation footprints. However, the observed-minus-expected prevalence map contained both positive and negative differences across all four states after accounting for footprint area. Operation-footprint area therefore explained an important component of the county-level spatial pattern but did not account for all geographic variation.

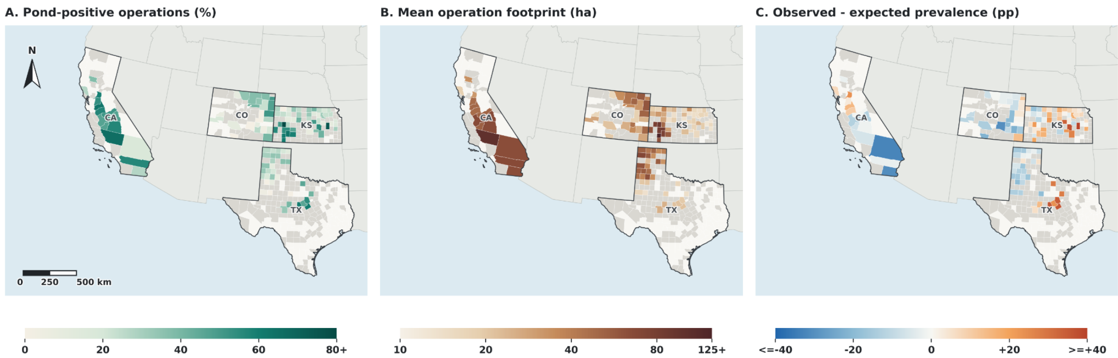

Figure 4: *County-level spatial patterns of BSI-E candidate detections.* Panels show raw pond-positive prevalence, mean operation-footprint area, and observed-minus-expected prevalence from the pooled footprint-area logistic model. Only counties with at least 10 processed operations are shown.

Discussion and Conclusion

The central GIScience finding is that the cattle-operation footprint functions as both a contextual mask and the spatial support of the measurement. Larger footprints may represent physically larger or more complex facilities, but they also create a larger search area in which candidate components can occur. Both mechanisms can increase pond-positive frequency. Consequently, raw differences in state- or county-level prevalence should not be interpreted solely as differences in manure-management practice. The substantial change in interstate prevalence after footprint-area standardization demonstrates that the geometry of the upstream GeoAI product can materially affect downstream spatial inference.

At the same time, footprint area does not explain all of the observed geographic variation. Positive and negative county-level differences remained after accounting for footprint size, indicating that additional factors may influence the spatial distribution of detected candidate infrastructure. Herd size, production system, facility configuration, climate, and independently verified manure-management information will be needed to distinguish true management-system differences from effects associated with facility scale and search area.

The retained objects should therefore be interpreted as candidate manure-management infrastructure rather than confirmed AWMS categories. BSI-E identifies spectral conditions associated with moisture-rich and turbid surfaces, and visually similar features such as open water, runoff, shadow, or transient wet surfaces may also be retained. In addition, the four-state analysis has not yet undergone independent, state-stratified validation. Candidate detections should not be directly translated into anaerobic lagoons or manure-routing fractions without additional interpretation or ancillary observations.

Within these limits, the satellite-GeoAI traceability pipeline converts raster detections into spatially explicit, operation-linked records while preserving processing and geographic provenance. These records provide a reproducible basis for targeted validation, temporal persistence analysis, and future integration with herd, climate, and management information for IPCC Tier 2 inventory development. The contribution is therefore not a new emission estimation method, but a traceable framework for converting satellite-derived manure-infrastructure detections into spatially explicit activity data.

Acknowledgements:

The authors thank the Climate TRACE coalition and its funders for organizational and financial support, Earth Genome for technical collaboration, and the Remote Sensing Laboratory at Saint Louis University for research infrastructure.

References

Davitt, A., Tesser, D., Boyda, E., Pawar, A., Al Akkad, O., Tulloch, P., Tulloch, A., Schiller, S., Sridhar, L., Betz, L., & Sagan, V. (2025). *Agriculture sector: Cattle operations asset-level enteric fermentation and manure management emissions*. Climate TRACE methodology documentation.

Dittmer, K. M., Wollenberg, E., Cohen, M., & Egler, C. (2023). How good is the data for tracking countries' agricultural greenhouse gas emissions? Making use of multiple national greenhouse gas inventories. *Frontiers in Sustainable Food Systems, 7*, 1156822. https://doi.org/10.3389/fsufs.2023.1156822

IPCC. (2019). *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories*. Intergovernmental Panel on Climate Change. https://www.ipcc-nggip.iges.or.jp/public/2019rf/

Pawar, A., Sagan, V., Tesser, D., Farbo, A., Boyda, E., & Davitt, A. (2026). *The Bovine Slurry Index (BSI): An integrated framework for manure pond mapping using GeoFM-derived cattle-operation detection and Sentinel-2 spectral formulation*. Manuscript submitted for publication.