FARMLAND IMPACT INDEX SUPPLEMENTARY METHODS, DATA, AND CODE

This appendix provides detailed information on data sources, variable construction, and analytical procedures to ensure the full reproducibility and transparency of the Farmland Impact Index (FII) methodology.

Data Sources and Acquisition

The primary data for this study were compiled from publicly available sources at the county level within the Southeastern United States. The specific datasets and their key characteristics are detailed below.

- **Farmland Impact Index (FII) Raw Data** (FIISE_with_FIPS.csv; available in GitHub repository)
    - **Source:** Compiled by the author from various public data sources, including the Census of Agriculture and American Community Survey. Full details are available in Table 1.
    - **Original Granularity:** County-level.
    - **Notes on Acquisition:** This proprietary file aggregates and pre-processes data from multiple public sources into a unified format for analysis. Specific original sources for individual metrics are detailed in the main manuscript's methods section and in Table 1 below.
- **U.S. County Boundary Shapefile** (cb_2024_us_county_500k.shp)
    - **Source:** U.S. Census Bureau, Cartographic Boundary Files.
    - **Access URL:** https://www2.census.gov/geo/tiger/TIGER2024/COUNTY/
    - **Original Granularity:** County-level polygons.
    - **Notes on Acquisition:** The shapefile was downloaded directly from the Census Bureau's website.
- **Rural-Urban Continuum Codes (RUCC)**
    - **Source:** United States Department of Agriculture, Economic Research Service (USDA-ERS).
    - **Access/URL:** https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

- **Original Granularity:** County-level codes.
- **Key Variables Used:** RUCC_2023.
- **Notes on Acquisition:** The RUCC data was downloaded from the USDA-ERS website and joined to the FII raw data using FIPS codes.

Table 1. Data sources and metrics used in the Farmland Impact Index.

INDEX DATA				
Environment				
Concept	**Metric**	**Publicly Available Data Sources**	**Citation**	**Notes**
Open space	% of county land in farms	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: DIVIDE (AG LAND - ACRES, LAND AREA, INCL NON-AG - ACRES)
Natural resource conservation	Acres enrolled in conservation reserve program	USDA FSA Historic CRP Data	USDA-FSA. (2025). CRP Enrollment and Rental Payments by County, 1986-2024 [Dataset]. https://www.fsa.usda.gov/tools/informational/reports/conservation-statistics/crp	Year Represented: 2022; Data Item: CONSERVATION RESERVE PROGRAM--CUMULATIVE ENROLLMENT BY FISCAL YEAR (ACRES) (Acres Under Contract at the End of Each FY)
Land stewardship	Acres of no till or conservation tillage	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: SUM (PRACTICES, LAND USE, CROPLAND, CONSERVATION TILLAGE, NO-TILL - ACRES, PRACTICES, LAND USE, CROPLAND, CONSERVATION TILLAGE, (EXCL NO-TILL) - ACRES)
Economic				
Concept	**Metric**	**Publicly Available Data Sources**	**Citation**	**Notes**
Farm viability	$ net farm income	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year represented: 2022; Data Item: INCOME, NET CASH FARM, OF OPERATIONS - NET INCOME, MEASURED IN $

Jobs	# agricultural jobs	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: LABOR, HIRED - NUMBER OF WORKERS
Taxes	$ taxes paid	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: TAXES, PROPERTY, REAL ESTATE & NON-REAL ESTATE, (EXCL PAID BY LANDLORD) - EXPENSE, MEASURED IN $
Social				
Concept	**Metric**	**Publicly Available Data Sources**	**Citation**	**Notes**
Food security	# of farms accepting SNAP	USDA FNS Snap Retailer Database	USDA-FNS. (2025). Where Can I Use SNAP EBT? | Food and Nutrition Service. https://www.fns.usda.gov/snap/retailer-locator	Year Represented: 2025
Cultural preservation	# farmers demographic diversity	USDA NASS Agriculture Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: SUM: (PRODUCERS, AMERICAN INDIAN OR ALASKA NATIVE - NUMBER OF PRODUCERS, PRODUCERS, ASIAN - NUMBER OF PRODUCERS, PRODUCERS, BLACK OR AFRICAN AMERICAN - NUMBER OF PRODUCERS, PRODUCERS, HISPANIC - NUMBER OF PRODUCERS, PRODUCERS, HAWAIIAN OR PACIFIC ISLANDER - NUMBER OF PRODUCERS)
Community engagement	# of direct to consumer operations	USDA NASS Agricultural Census	USDA-NASS. (2025). USDA - National Agricultural Statistics Service—Census of Agriculture. Census of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php	Year Represented: 2022; Data Item: COMMODITY TOTALS, INCL VALUE-ADDED, RETAIL, DIRECTLY MARKETED, HUMAN CONSUMPTION - OPERATIONS WITH SALES
Rural-Urban Continuum				
Concept	**Metric**	**Publicly Available Data Sources**	**Citation**	**Notes**

Urban typology	community code	USDA ERS Rural-Urban Continuum Codes	USDA-ERS. (2025). Rural-Urban Continuum Codes | Economic Research Service. Rural-Urban Continuum Codes. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes	Year Represented: 2023

Variable Construction and Standardization

All raw metrics were standardized to allow for comparison across different scales and units.

- **Z-score Standardization:** Individual metrics were converted into z-scores using the formula: $Zi=\sigma(Xi-\mu)$ Where:
    - is the z-score for observation i.
    - Xi is the raw value for observation i.
    - μ is the mean of the raw metric across all observations.
    - σ is the standard deviation of the raw metric across all observations. This standardization was applied to all quantitative metrics before aggregation.
- **Unweighted Farmland Impact Index (FII) Construction:**
    - **Composite Indicators:** For each dimension (Environmental, Economic, Social), the unweighted composite indicator was calculated by summing the z-scores of its constituent metrics. For example: Environmental_sum_Z=Zacres_enrolled_in_CRP+Zacres_in_conservation_or_no_till+Zcounty_in_agriculture
    - **Overall Unweighted FII Score:** The total unweighted FII score was derived by summing the unweighted composite indicator scores across all three dimensions.
- **Weighted Farmland Impact Index (FII) Construction (Principal Component Analysis - PCA):**
    - **Dimension-Specific PCA:** For each dimension (Environmental, Economic, Social), a separate Principal Component Analysis was conducted using the prcomp() function in R on the z-score standardized metrics. The first Principal Component (PC1) was extracted as the weighted score for that dimension. PC1 was selected as it consistently captured the largest proportion of variance within each dimension's set of indicators, reflecting the dominant underlying pattern.

    - **Overall Weighted FII Score:** The total weighted FII score was derived by summing the PC1 scores from each of the three dimensions.
- **Ranking:** All final composite indicator scores (unweighted and weighted) and overall FII scores (unweighted and weighted) were ranked in descending order using R's dense_rank(desc()) function. This method assigns a rank of 1 to the highest score, with subsequent ranks assigned sequentially without gaps, and tied scores receiving the same rank. A lower rank number therefore indicates a higher index score.

Urban Typology Grouping

The USDA Rural-Urban Continuum Codes (RUCC) were re-grouped into three broader urban typologies for this analysis. The specific mapping and county counts are as follows:

- **Urban:** RUCC codes 1, 2, and 3 (n=572 counties)
- **Suburban:** RUCC codes 4, 6, and 8 (n=510 counties)
- **Rural:** RUCC codes 5, 7, and 9 (n=301 counties)

Statistical Analysis Details

All data manipulation, statistical analysis, and visualization were performed using R version 4.4.2 (Posit Team, 2025) within RStudio version 2025.5.0.496 (Posit Team, 2025).

Key R packages utilized include:

- dplyr (for data manipulation)
- stringr (for string operations)
- sf (for handling spatial data)
- tmap (for thematic mapping)
- tidyverse (for general data science tools)
- psych (for Cronbach's Alpha)
- Base R functions for statistical tests (cor.test, prcomp, aov, TukeyHSD).

Differences in FII scores across urban typologies were assessed using one-way Analysis of Variance (ANOVA) via the aov() function in R. Statistical significance for ANOVA was set at p

< .001. Post-hoc pairwise comparisons were conducted using Tukey's Honestly Significant Difference (HSD) test via the TukeyHSD() function to identify specific group differences.

Reliability Analysis (Cronbach's Alpha): Cronbach's Alpha was calculated for the set of indicators within each FII dimension (Environmental, Economic, and Social) to assess their internal consistency and to evaluate which metrics to include in the composite indicators. This analysis was performed using the alpha() function from the psych package.

Correlation Analysis: Spearman's rank correlation was used to assess associations between individual indicators, calculated using the cor() function with method = "spearman". P-values for Spearman's correlations were derived based on t-distribution approximation. Kendall's tau correlation was used to test the association between the unweighted and weighted FII results, performed using the cor.test() function with method = "kendall".

Map Generation Details

Maps were generated for the Southeastern United States using the tmap package. A custom function, plot_fii_rank_map, was utilized to merge the FII data with county boundary shapefiles and generate choropleth maps. Data joining within this function was performed by matching county and state names (converted to uppercase for consistency) between the FII data and the shapefile. The tm_fill function was used for coloring based on the specified rank column. To ensure direct comparability across all maps, global 5-quantile breaks were calculated from the combined distribution of all rank scores and explicitly applied to the fill.scale argument using tm_scale_intervals(..., breaks = global_quantile_breaks). The viridis color palette was used for visualization.

R Code and Data Availability

All R code used for data processing, analysis, and visualization in this study is openly available in a public GitHub repository at: https://github.com/hdankbar/Farmland-Impact-Index-Analysis. This repository includes this supplemental methods description, the executable R code, along with the processed .csv data file (FIISE_with_FIPS.csv) used for the analysis. The analysis was conducted using the R and package versions specified in the notebook environment.