



Scanning for Soil Health: Assessing the Accuracy and Scalability of Innovations for In-Situ Soil Measurement

First Submission: RR_UGA_2026_753

Vaishnavi Dhas

reproducibility@worldbank.org

October 1, 2026

This review verifies the reproducibility of the exhibits included in the paper "*Scanning for Soil Health: Assessing the Accuracy and Scalability of Innovations for In-Situ Soil Measurement*".

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main section of the paper and the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. R: Opening the R project "Replication.Rproj". Restoring the environment by running `renv::restore()` and following the prompts. Running `o_packages.R`. Opening `3-analysis.R`, updating working directory in line 6, and running it.
 2. Stata: Reviewers did not run the Stata code. Instead, reviewers virtually verified the construction of the intermediate datasets from the raw data. Reviewers verified that this intermediate data matches what was provided by the authors; results of this verification are documented in the file *comparison_report.csv*.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 15 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by email on September 9, 2026.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 3 types of data: open, restricted and forthcoming.
 - Data folder structure and access policy:

Data

- |— Raw - Confidential: Forthcoming on the World Bank MDL (excluding restricted GPS data) following an embargo period.
- | |— Soil Data
- | |— Survey Microdata
- |— Raw - Public: Open data included in the public reproducibility package.
- |— Intermediate - Confidential: Intermediate data containing restricted GPS data, not deposited in the World Bank MDL.
- |— Processed – Public: Intermediate data included in the public reproducibility package.

- Reviewers used intermediate data provided directly by the authors to conduct the reproducibility verification, and this is the starting point of the public reproducibility package.
- Reviewers virtually verified the construction of the intermediate datas from the raw data.
- `data_hash_report.csv` lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire forthcoming data can use this file to verify that the data has not been altered.

- **Reproducibility Summary:**

- **Data:** Some data is restricted and has not been included in the reproducibility package. For more details, please refer to the README file. Intermediate data is included in the reproducibility package and can be used to fully reproduce the results in the paper.
- **Code:** All code files are included in the reproducibility package but operate from intermediate data, not raw data.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Dependencies environment:** The reviewers reproduced an existing environment for dependencies using dependency files or an environment metadata file provided by the authors.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1** Reproduced.
- **Figure 2** Reproduced.
- **Figure 3** Reproduced.
- **Figure 4** Reproduced.
- **Figure 5** Reproduced.
- **Figure 6** Reproduced.
- **Figure 7** Reproduced.
- **Figure 8** Reproduced.
- **Figure 9** Reproduced.
- **Figure 10** Reproduced.

- **Figure 11** Reproduced.
- **Table 1** Does not apply. The exhibit is not an analytical output.

Results in the Annex

- **Figure A1** Reproduced.
- **Figure A2** Reproduced.
- **Figure A3** Reproduced.
- **Figure A4** Reproduced.
- **Figure A5** Reproduced.
- **Figure A6** Reproduced.
- **Table A1** Does not apply. The exhibit is not an analytical output.
- **Table A2** Reproduced.
- **Table A3** Does not apply. The exhibit is not an analytical output.

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: Windows 11 Enterprise
- Processor: INTEL(R) XEON(R) PLATINUM 8562Y+ 2.80 GHz (2 processors)
- Memory available: 32.0 GB
- Software version: R 4.5.1, Stata 19.0 MP