

RR_WLD_2025_378 — Reproducibility Package

Paper: Sahu, Saslavsky & Schultz, “Reimagining Logistics and Connectivity from Space: Artificial Intelligence and Satellite Imagery Applications”

This package contains the code and the redistributable data behind the exhibits of the paper. Several datasets are commercial, confidential or of unconfirmed terms of use, and are **not included**. Section 3 explains, dataset by dataset, what is missing, why, and how to request access. Section 4 states which exhibits can be rerun from this package alone.

1. Contents

Item	Description
README.pdf	This file
LICENSE.txt, WB-IGO-RIDER.txt	Licence and rights statement
reproducibility_report_RR_WLD_2025_378.pdf	Reproducibility verification report
data_hash_report.csv	Location and SHA-256 hash of every data file used in the verification, including the files that are not in this package . Use it to confirm that data obtained from the authors/providers is identical to the data used in the verification.
variable_inventory.csv	Variable names of the datasets that are not included
Reproducibility package/	Everything needed to run the code: <code>main.py</code> (master script), <code>ORCHESTRATOR_README.md</code> , and one folder per exhibit (code, documentation, outputs and, where redistributable, data). Each exhibit folder has its own <code>README.md</code> , <code>Data_Availability_Statement.md</code> , <code>requirements.txt</code> and, if automated, <code>run_all.py</code> .

Data folders inside exhibit folders whose data is not redistributable are empty or absent. Code, documentation and the outputs generated during verification remain in every folder, so results can be compared against the published report.

2. How to run

1. Install Python 3 (verified with 3.13 and 3.14) and create a clean virtual environment:

```
python3 -m venv venv
./venv/bin/python -m pip install --upgrade pip
```

2. From inside the `Reproducibility package/` folder run `./venv/bin/python main.py`. The script locates the exhibit folders from its own location, so no directory has to be edited. Use `--skip-install` to reuse an existing environment and `--figures Figure4 Table4` to run selected exhibits.
3. `main.py` installs each folder's `requirements.txt` and runs its `run_all.py`. Logs are written to `Reproducibility package/run_logs/`.

4. Folders with no `run_all.py` are reported as SKIPPED. Exhibits whose input data is not included (Section 4) will be reported as FAILED until the data is obtained and placed at the paths listed in `data_hash_report.csv` (relative to Reproducibility package/).

Approximate run time of the full package with all data: 1 hour.

3. Data availability statement

Statements about provenance, custodians and permissions below were supplied by the authors (delivery of 26 September 2026) and have not been independently confirmed by the reviewer. For all requests contact the authors at ssahu8@worldbank.org.

Dataset	Used in	Access	In package?	How to obtain
Copernicus Sentinel-1 GRD SAR (project-processed occupancy estimates; Date, Occupancy)	Figs 1, 3, 14	Public	Yes (derivatives only; raw scenes are not included)	Raw scenes: https://dataspace.copernicus.eu . Exact AOI polygons and full product IDs were not recovered.
Georgian Revenue Service TIR-park truck counts (ground truth)	Figs 1, 14	Custom licence, no redistribution	No	Request the archived study extract and site/date mapping from the authors, who coordinate with the Georgian Revenue Service. Access is not guaranteed and the live service (https://www.rs.ge/TirPark-en) cannot recreate the historical observations.
PlanetScope imagery (Planet Labs PBC) and derived detections, scores, masks and annotations	Figs 2, 12, 13	Commercial licence, no redistribution; derivative release pending WBG-Planet terms	No	Requires a Planet licence. The Beitbridge scene list (15 Dec 2022 – 22 Apr 2023) is available from the authors. Original order IDs were not recovered.
Pléiades 1A/1B VHR scene of Incheon (Airbus, via SkyWatch) and derived truck points and Incheon building tables	Fig 6, Table 8	Commercial licence, no redistribution	No	Request licensed access from SkyWatch using scene SKYWATCH_PHR_PS_20231120T0215_ALL_6bff.tif (four tiles). No order ID was recovered.
Logistics warehouse information, Korea (MOLIT National Logistics Information Center) and derived warehouse labels	Tables 3, 7	Terms of redistribution not confirmed	No	Source: http://nlic.go.kr/nlic/WhsInfoWarehouseSch.action (snapshot of 14 June 2023). Derived labels are available from the authors once permission is confirmed.
Aktau terminal monthly throughput (confidential) and the indexed series derived from it	Fig 11	Confidential, no public access route	No	Requests can be addressed to the authors. Access is not guaranteed.
Figure 3 source workbook (mixed PlanetScope, GPS and Sentinel-1 data) and	Figs 2, 3, 8, 12	Contain restricted imagery and/or	No	Not available.

Dataset	Used in	Access	In package?	How to obtain
ArcGIS packages (Figure3_SAR.ppkx, Figure8.ppkx, Figure 2 and Figure 12 .ppkx)		Esri basemap content		
Microsoft Global ML Building Footprints, OpenStreetMap, WorldPop (1 km, 2020), World Settlement Footprint (WSF Evolution, WSF3D)	Figs 6, 7, Tables 3, 4, 7, 8	Public	Yes	Sources listed in Data_Availability_Statement.md of each exhibit folder.
U.S. Surface Transportation Board rail service data (Chicago Clearing)	Fig 10	Public	Yes	https://www.stb.gov/reports-data/rail-service-data/
Values transcribed or digitized from the published report	Figs 4, 5, 15, 16, Tables 1, 2, 5	No third-party data	Yes	Included.

Removed on the authors' request because no script depends on them: the two road/toll/parking directories of Figure 13 (Lebombo and Beitbridge) and `korea_logistic_assets_feature_updated.geojson`.

A historical `Figure4.xls` (482 Date/Occupancy observations, 2 March 2015 – 16 March 2024) was recovered by the authors. Its link to the published figure could not be verified, so it is not included and Figure 4 keeps its reconstruction qualification.

4. Exhibits and what can be rerun

Can be rerun with this package alone: Figure 4 (a reconstruction digitized from the published report, not a rerun from original Sentinel-1 inputs), Figures 7 and 10, and Table 4.

Conceptual or descriptive, nothing to rerun: Figures 5, 15, 16 and Tables 1, 2, 5.

Require data that is not included (verification used data provided directly by the authors):

Exhibit	Missing input
Figures 1, 14	Georgian Revenue Service ground truth
Figure 2	PlanetScope scenes, detections and scene manifest
Figure 3	Source workbook <code>Figure3.xlsx</code> (mixed-source)
Figure 6, Table 8	Pléiades scene, digitized truck points, Incheon tables
Figure 8	ArcGIS project (produced manually in ArcGIS Pro, not generated by <code>main.py</code>)
Figure 11	Aktau indexed series
Figure 12	PlanetScope-based ArcGIS packages

Exhibit	Missing input
Figure 13	PlanetScope detector scores, reference annotations and mask
Tables 3, 7	Korea warehouse labels (Korea results only)

5. Notes on exhibit status

- **Figures 4 and 13** are reconstructions from previously published output or manual digitization, not reruns from original raw source data.
- **Figures 3 and 13** include manual edits.
- **Figure 10:** the manuscript reports Pearson $r = 0.74$ and a mean relative error of 11%; the code reproduces $r = 0.7021$ and 9.9%.
- **Figure 12:** redistribution terms for the PlanetScope imagery are not confirmed.
- The data classification above rests on author-supplied provenance; open permission and lineage questions are listed in the verification report.
- **Figure 9 and Table 6** are referenced in author materials but have no folder in this package.

6. Computational environment

Verification was run on macOS (Apple silicon) with Python 3.13/3.14. ArcGIS Pro was used by the authors for Figure 8. Each exhibit's `requirements.txt` lists its Python dependencies.

`Table4_WBG_PR_closeout_package/requirements.txt` uses `numpy>=2.0` (the original `numpy==2.0.2` pin does not build on Python 3.13).

7. Rights

See `LICENSE.txt` and `WB-IG0-RIDER.txt`. Third-party data remain subject to their providers' terms.