Estimating the Magnitude and Spatial Distribution of Informal Trade in Central Asia

An EO and Economic Data Framework for Estimating the Magnitude of Bazaars in Kyrgyz Republic and Kazakhstan

THE PROBLEM

In many emerging economies, particularly within Central Asia, substantial informal commerce segments are systematically undercounted by official records. Traditional data collection through local surveys often remains incomplete, as it is frequently limited by the transparency and reliability of local government structures. Consequently, official statistics fail to capture the true magnitude of the informal trade sector, which serves as a primary source of regional employment.

To bridge these informational gaps, this project was developed in partnership with the World Bank Group to support the "Central Asia: Regional Trade Connectivity Linkages" initiative, which focuses on bolstering regional trade, investment, and connectivity.

THE EO APPROACH

EO data provides an independent and easily cross-checkable information source for sampling commerce activities, bypassing the potential lack of transparency or capacity within local government structures. By utilizing simple, robust indicators derived from Very High Resolution (VHR) satellite imagery, this approach remains effective across various sensors and different spatial or spectral resolutions. Such a flexible methodology ensures the process is both highly scalable and replicable for monitoring economic activity in diverse contexts

ACKNOWLEDGEMENTS

The work has been conducted in the framework of the EO Clinic (Earth Observation Clinic), which is an ESA (European Space Agency) initiative to create a rapid-response mechanism for small-scale and exploratory uses of satellite EO information in support of a wide range of International Development projects and activities.

REFERENCES

"Bazaars and Trade Integration in CAREC Countries," World Bank, May 2009 · UN COMTRADE / World Bank WDI (2006) · "Logistics and Transport Competitiveness in Kazakhstan," UN 2019 (ISBN: 978-92-1-117205-8) · Transport in Kazakhstan 2013–2017, stat.gov.kz · knoema.com (vehicle traffic data)

WEBSITE

VIDEO DESCRIPTION (AI generated)

STUDY SITES

Two of the largest bazaars in Central Asia were selected for analysis. Both are clearly identifiable in satellite imagery as major commerce centres, consisting of warehouses, container buildings, parking facilities and sub-bazaars, and both were growing and intermingled with surrounding residential areas at the time of study.

Dordoi Bazaar (Bishkek, Kyrgyz Republic)

~40,000 sales outlets. Part of the modern Silk Road; re-export hub. Open daily 05:00–16:00. Key events: EEU membership (2015), Belt & Road Initiative (2019), COVID-19 partial closure (2020). Founded on a former manufacturing plant, squeezed between a cemetery and residential areas a near-monolithic landscape of container malls.

Kyrgyz Republic

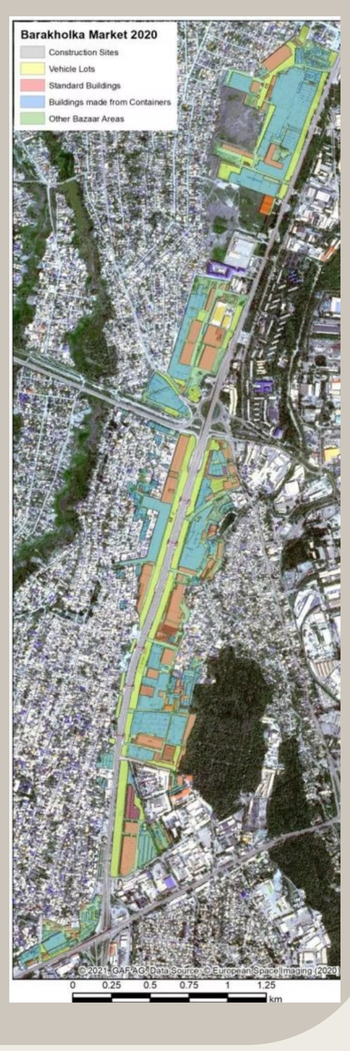
Bazaar goods = 54.7% of total imports; 65.4% of GNI

Barakholka Bazaar (Almaty, Kazakhstan)

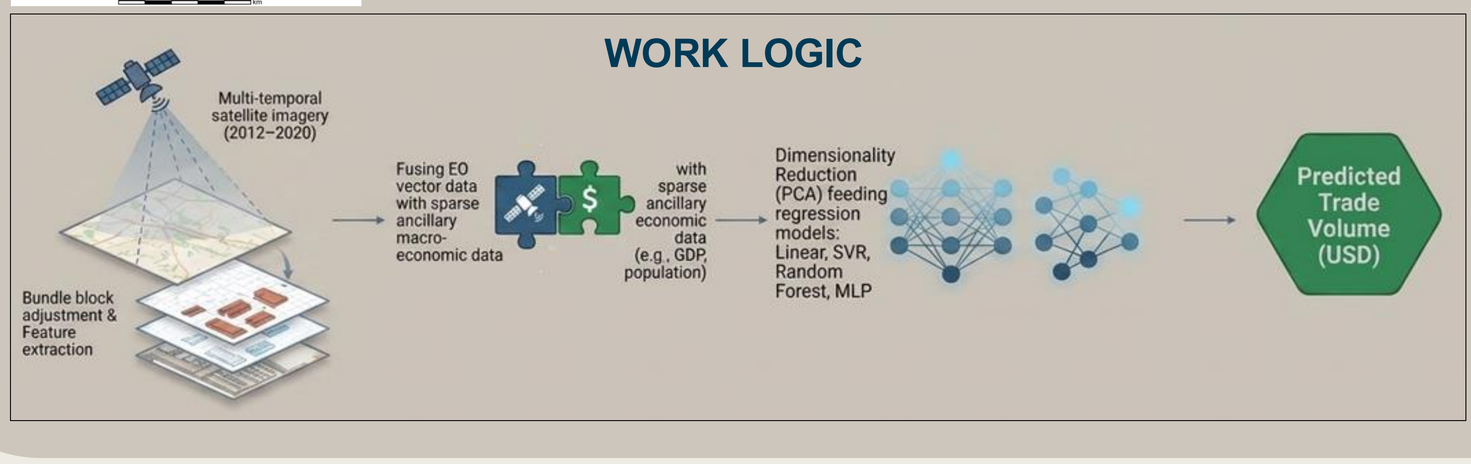
~15,000 sales outlets. Goods from China transferred onward to Russia, Turkey and Korea. Open Tue–Sun 10:00–17:00. Key events: "Initiative to eliminate chaotic circumstances" (2011), transition from containers to fixed buildings (from 2014), partial COVID-19 closure (2020). Located along motorway A-350; comprises several separated sub-bazaars.

Kazakhstan

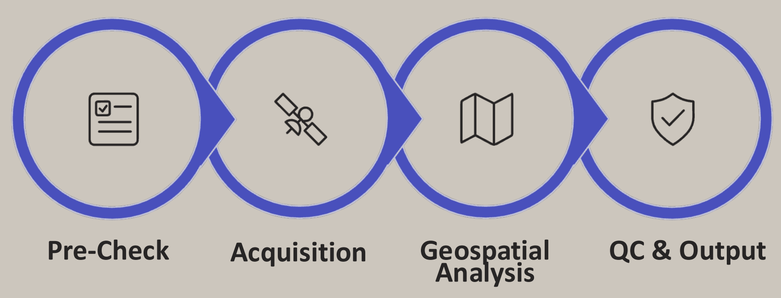
Bazaar goods = 20.1% of total imports; 65.3% share of CA bazaar imports



WORK LOGIC



EO GEOSPATIAL ANALYSIS METHODOLOGY



The workflow begins with an analysis using open-source information to determine meaningful observation dates, followed by ordered VHR imagery, expert visual interpretation, and a two-stage quality control process (QC1: semi-automatic; QC2: 100% independent senior analyst review).

Satellite Data

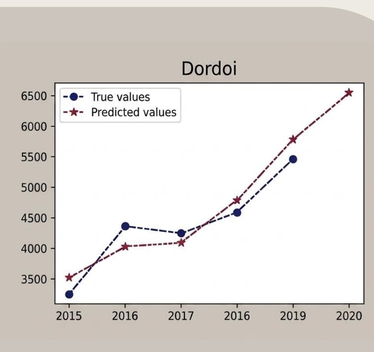
VHR imagery from WorldView-2, Pleiades, and GeoEye sensors was acquired for both bazaars, covering spring/summer and autumn seasons to detect construction activity and avoid seasonal effects. Barakholka: 12 scenes (2012–2020). Dordoi: 9 scenes (2012–2020). All scenes underwent bundle block adjustment for geometric alignment.

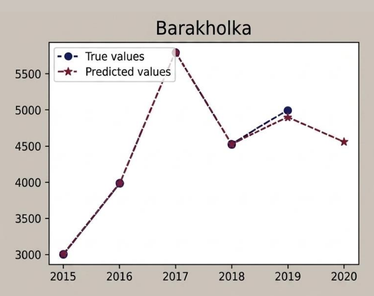
Extracted Indicators (per scene, in m²)

- Bazaar Area: Total bazaar area including roads and open spaces
- Building Area: Permanent fixed buildings
- Container Building Area: Container-based buildings including covered aisles
- Vehicle Parking Aera: Permanent, well-marked parking lots only
- Construction Site Area: Clearly visible, larger active construction sites
- Total Roofed Area: Sum of Building Area and Container Building Area

PREDICTIVE ANALYSIS RESULTS

Multiple regression and machine learning techniques were evaluated under different data scenarios. Models were trained on historical data (2015–2018), validated against 2019 benchmarks, and used to generate estimates for 2020. Results yield encouraging results but do not support definitive conclusions due to the limited availability and short temporal coverage of input data, which constrains the effectiveness of predictive techniques. Initial modelling, achieved prediction errors of approximately 10% for Dordoi Bazaar and 5% for Barakholka Bazaar. The inclusion of ancillary economic and mobility data substantially improved model performance for Barakholka, reducing the validation error to around 2–3%.





TEST 1 — Dordoi (EO only)

Best model: MLPR (abs. error 2019: 319 M USD). LR and SVR overfit (accuracy ~1.0 in test, ~1.0 in test, high error in validation). MLPR selected. **2020 Prediction:** ~6,000 M USD — an USD — an increase of **~+10%** vs. 2019. Score fit: 89.49%.

TEST 2 — Barakholka (EO only)

Best model: MLPR (abs. error 2019: 125 M USD, score fit >99%). Linear dependencies dependencies present. **2020 Prediction:** ~4,316 M USD — a decrease of **~−13.7%** vs. 2019. **~−13.7%** vs. 2019. Confirmed by all methods.

TEST 3 — Barakholka (EO + non-EO)

Best model: MLPR (abs. error 2019: 99 M USD). Adding ancillary data reduces validation validation error for LR and MLPR. **2020 Prediction:** ~4,683 M USD — a decrease of **~−6.85%** of **~−6.85%** vs. 2019. Score fit >99%.

TEST 4 — Barakholka (EO + non-EO + PCA)

Best model: LR (abs. error 2019: 99 M USD). PCA reduces 23 variables to 5 components with components with ~100% explained variance. **2020 Prediction:** ~4,672 M USD — a decrease a decrease of **~−6.64%** vs. 2019. Confirms data are largely linearly correlated.

CONCLUSION

The results of the analysis are considered cautiously encouraging, though the lack of a large volume of input data prevents the derivation of solid, definitive conclusions. The integration of ancillary data was shown to significantly improve predictive quality; for example, the prediction error for the Barakholka bazaar decreased to approximately 2–3% when such data was included. Overall, while there is coherence among various tests, the lack of comprehensive data results in high general uncertainty regarding the final predictions.

The most significant limitation of the study is the short length of the time series, as five or six data points are insufficient to accurately represent entire behaviors or reach high accuracy in validation. This limited data volume often leads to overfitting problems. Additionally, the low availability and quality of non-EO data required the use of extrapolation methods, which inherently introduce errors into the datasets and subsequently affect predictive capability.

Authors

Vito De Pasquale
depasquale@planetek.it

Giulio Ceriola
ceriola@planetek.it

Planetek Italia s.r.l.
http://www.planetek.it