Inferring Cuba's Hidden Economy: A Bayesian Latent Variable Model

Christopher M. Perez

May 6, 2025

What's Really Happening in Cuba?

Cuba forecasts a 1% GDP growth in 2025 despite the economic crisis

Cuban culture books

The regime claims that the 1% GDP growth expected in 2025 will rely on the recovery of tourism, exports, and energy improvements, despite an economy characterized by crises and setbacks.

CUBA

Cuba is a lot poorer than the government reports, a new study shows

By **Nora Gámez Torres**

January 22, 2018 5:56 PM

What's Really Happening in Cuba?

What we know about Cuba's economy

BY DREW DESILVER

Two-thirds of Americans favor an end to the decades-long U.S. trade embargo on Cuba, a January Pew Research Center study found, and the two nations reportedly are making progress on re-establishing diplomatic relations. As the communist government continues to slowly reform Cuba's economy, American businesses – from airlines to law firms – are exploring commercial opportunities on the island nation. But even if the embargo were to be lifted, it's not clear just what sort of Cuban economy those businesses would find.

Getting a handle on even basic information about Cuba's economy is difficult, for a number of reasons. The government still dominates economic activity on the island, both directly and through heavily subsidized state-owned enterprises. National statistics are not always complete or reliable. And Cuba's system of two parallel currencies – one peso for everyday transactions among ordinary Cubans, and a "convertible peso" for the tourism industry, foreign trade and the private sector – combined with multiple exchange rates complicates any international comparisons or discussions about the relative size of different parts of the economy.

Motivation

- Cuba's economic indicators are poorly reported or missing.
- International institutions lack consistent ground truth.
- Can satellite proxies and open data fill the gap?

Research Question

Goal

Estimate spatial economic activity using observable data and a principled statistical model.

Research Question

Goal

Estimate spatial economic activity using observable data and a principled statistical model.

- Leverage nightlights, NDVI, and road density.
- Use Bayesian inference to estimate a latent variable z_i per cell.

Data Sources

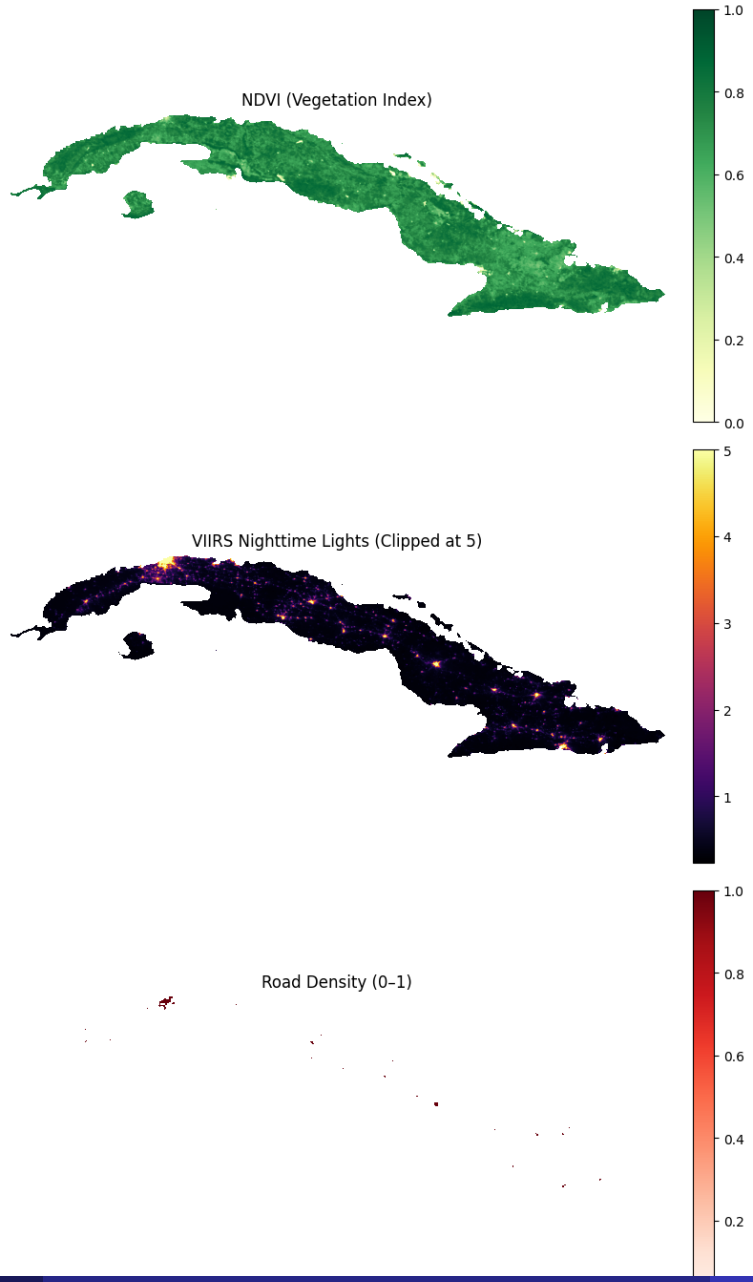
- **Nighttime Lights (VIIRS)**: Urbanization and infrastructure.
- **NDVI**: Vegetation intensity.
- **Road Density**: Accessiblity and development.

Data Sources

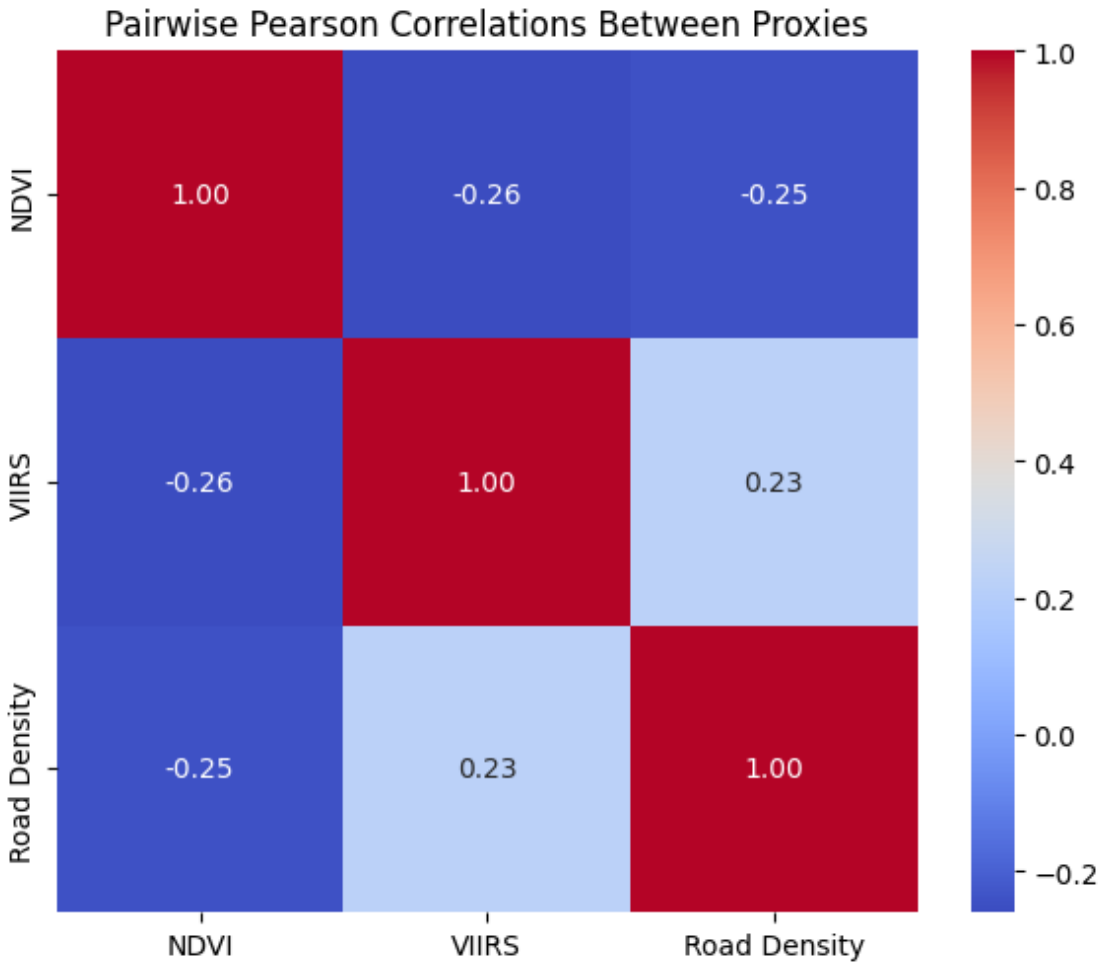
- **Nighttime Lights (VIIRS)**: Urbanization and infrastructure.
- **NDVI**: Vegetation intensity.
- **Road Density**: Accessiblity and development.

All proxies standardized and aligned to 500m grid. Water masked. 2024.

Preliminary EDA



Preliminary EDA



- Nightlights and roads weakly correlated ($\sim$0.23)
- NDVI negatively correlated with both roads and lights

Bayesian Latent Variable Model

- Latent index z_i per cell, measuring economic activity.
- Observation: $x_{i,k} \sim \mathcal{N}(\beta_k z_i, \sigma_k^2)$.
- Fixed $\beta_{\text{lights}} = 1$ for identifiability.
- Spatial smoothing via ICAR prior.

Bayesian Latent Variable Model

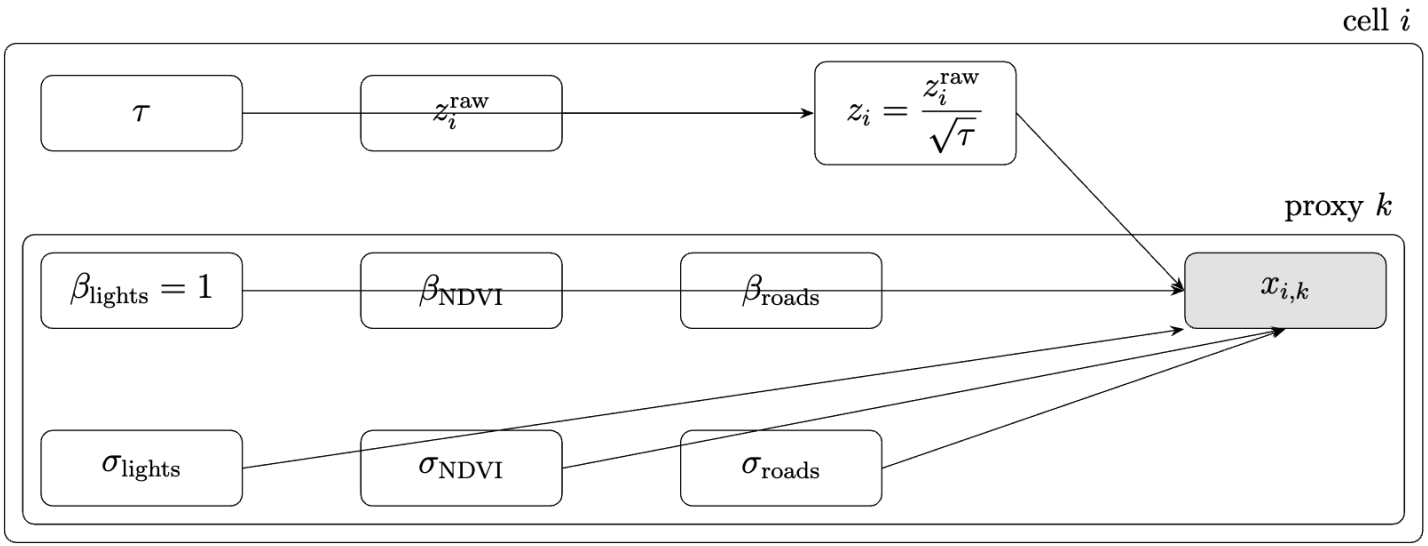
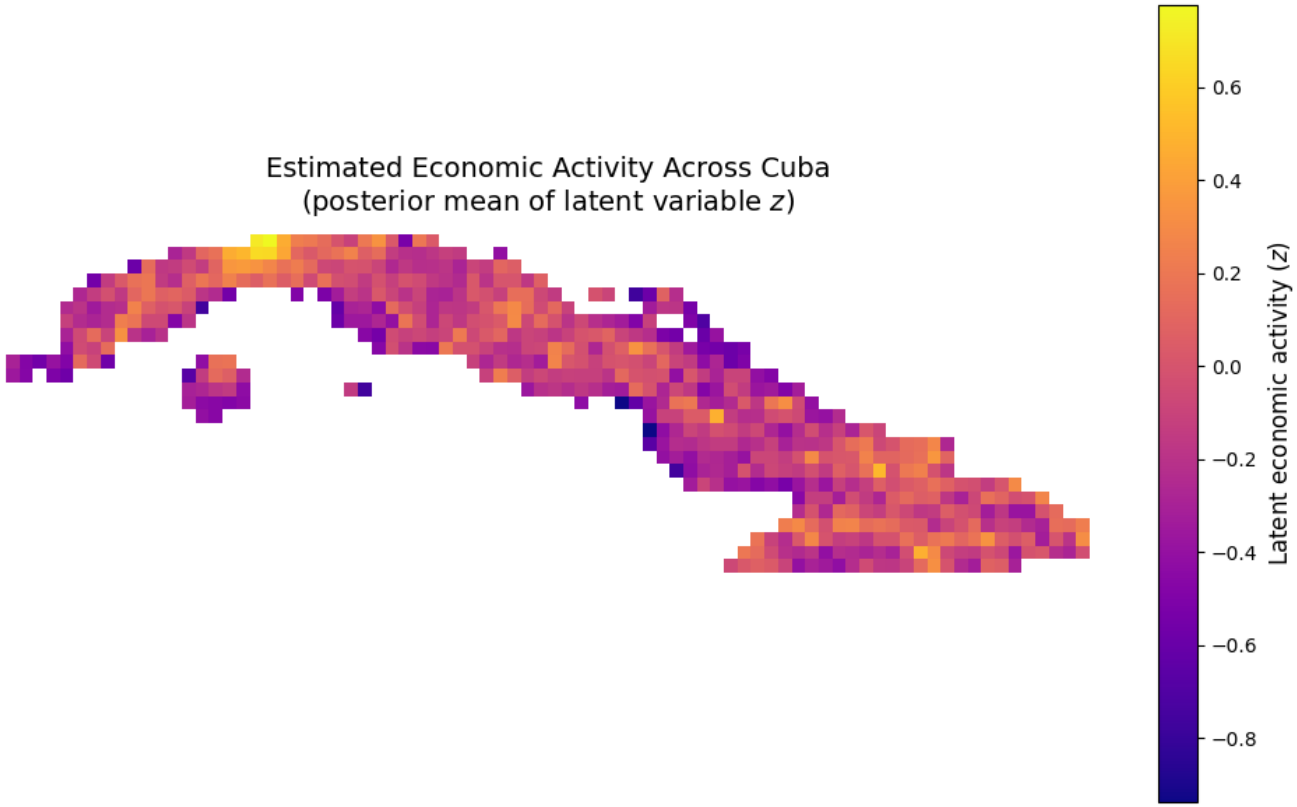


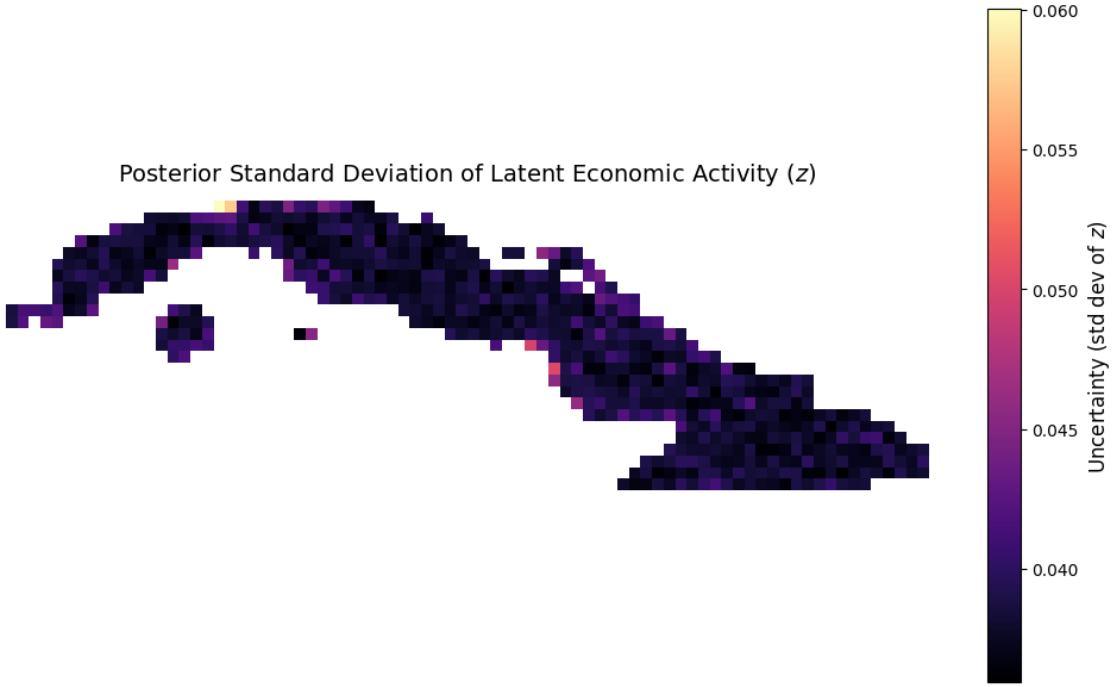
Figure 1: Graphical model of the Bayesian latent-variable hierarchy. Precision τ rescales raw latent draws z_i^{raw} to give spatial latent activity z_i. Observed proxies $x_{i,k}$ have Gaussian likelihood with mean $\beta_k z_i$ and noise σ_k. A fixed loading $\beta_{\text{lights}} = 1$ sets the latent scale.

Posterior Results: Cuba (15km Grid)



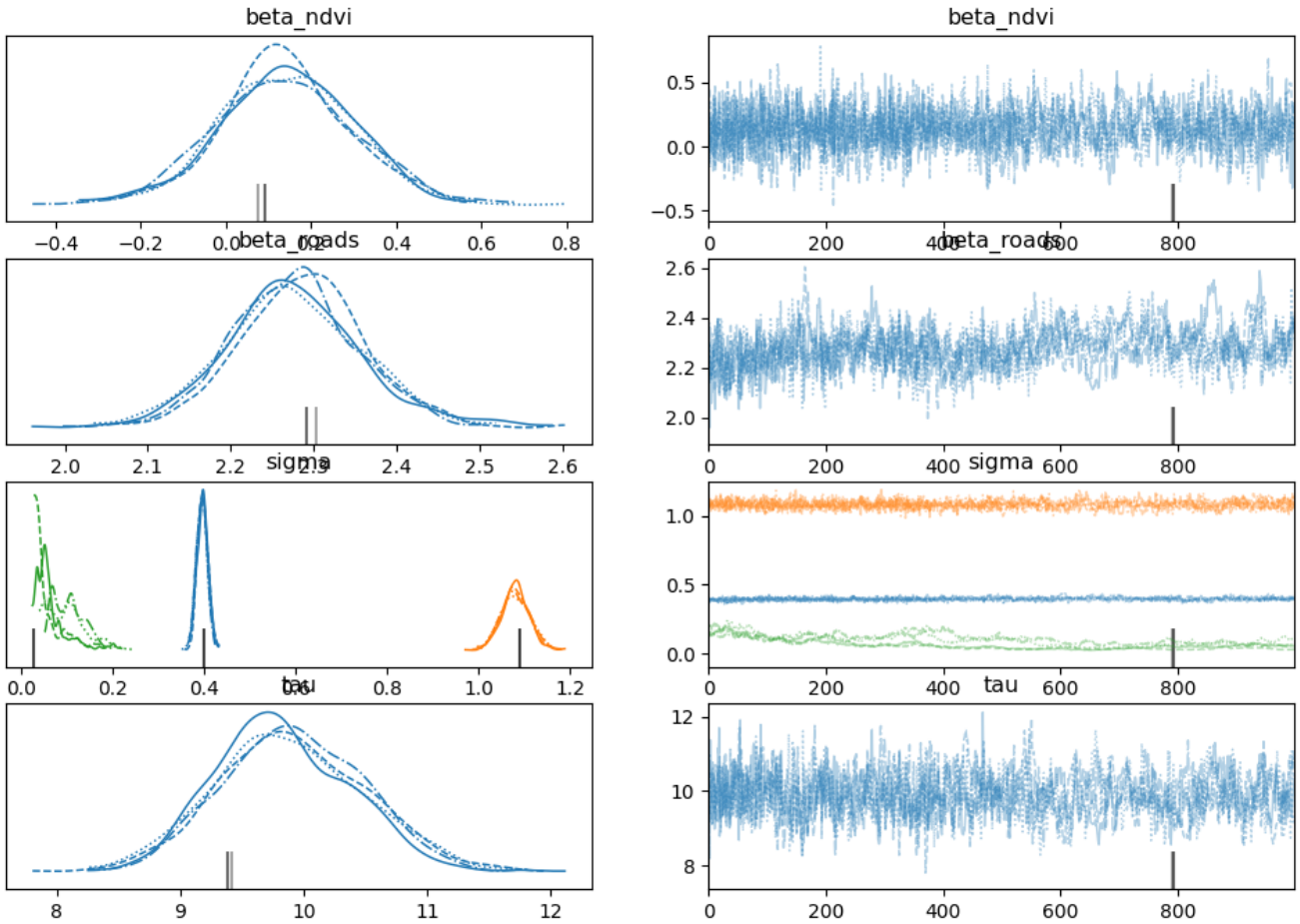
- High activity in Havana and other urban centers.
- Sparse regions have low z.

Posterior Uncertainty



- Posterior standard deviation is low ($\approx$ 0.04–0.06)
- Higher uncertainty in urban areas like Havana

Model Coefficients

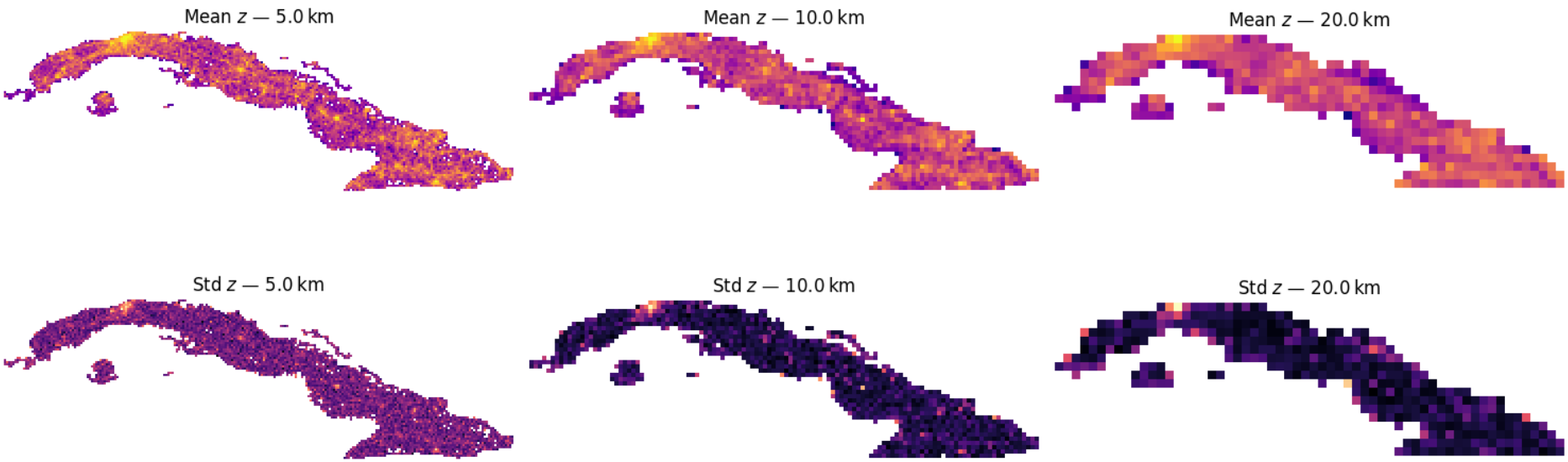


- $\beta_{\text{NDVI}} \approx 0.14$ (weakly positive).
- $\beta_{\text{roads}} \approx 2.29$ (strong positive).
- σ values show road is low-noise; NDVI is high-noise.

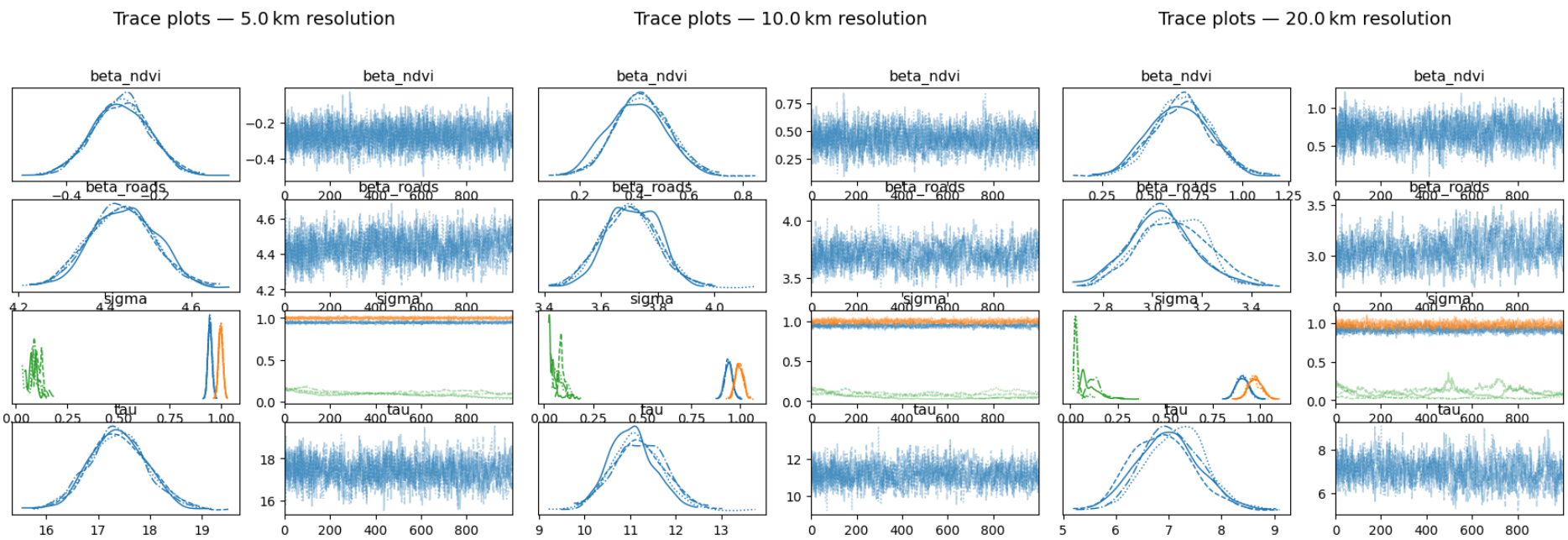
Multi-Resolution Overview

- Question: How does resolution affect our estimates?
- Fit model at 5km, 10km, 20km blocks.
- Compare variance, posterior uncertainty, and coefficients.

Resolution Comparison: Maps



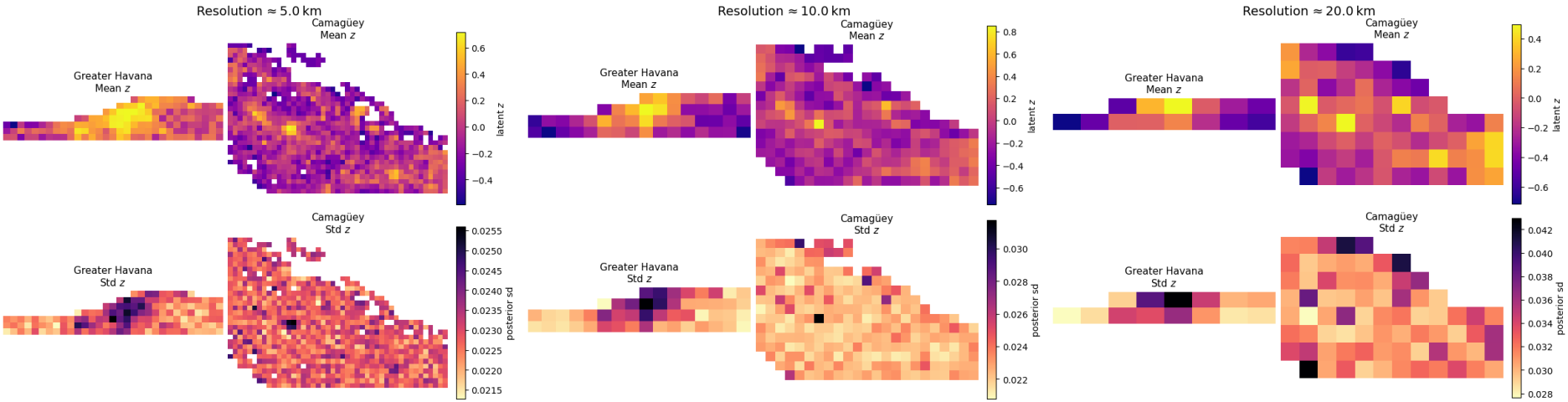
Resolution Comparison: Trace Plots



Multi-Resolution Results Summary

- β_{NDVI} flips from negative to positive as resolution increases
- β_{roads} declines: $4.44 \rightarrow 3.06$
- Posterior variance in z increases with block size
- Larger blocks allow more deviation between neighboring cells (smaller τ)

Zoom-In: Greater Havana and Camagüey



- Coarser resolution increases uncertainty per cell
- Urban vs. rural differences emerge clearly

Conclusion and Future Work

- Satellite proxies capture meaningful patterns in z_i.
- Bayesian model provides uncertainty quantification.
- Immediate Next Step: Cross-validate approach on Dominican Republic.
- Future: Add time dynamics and explore spatially varying β_k.

Thank You!

Thank you! Questions? Feedback?

References

- Liu, Bo et al. (2020). "An Economic Development Evaluation Based on the OpenStreetMap Road Network Density: The Case Study of 85 Cities in China." *ISPRS Int. J. Geo-Inf.*, 9(9), 517. https://www.mdpi.com/2220-9964/9/9/517
- Steele, Jessica E. et al. (2017). "Mapping Poverty Using Mobile Phone and Satellite Data." *J. Royal Soc. Interface*, 14(127), 20160690. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5332562/
- Besag, J., York, J., & Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43(1), 1-20. https://doi.org/10.1007/BF00116466