



STATISTICS

Earth Observations & Machine Learning for Gridded Macroeconomic Data

**StatEO CONFERENCE
MAY 6, 2026**

Marco Marini · Jim Tebrake · Andinet Woldemichael
IMF Statistics Department · Big Data Center

Motivation: Everything Happens Somewhere — But Macro Data Hides It

THE GAP

Macro Statistics Aggregate Away Spatial Detail

- GDP is published at national or, at best, subnational level (state, province, NUTS-2).
- Most low- and middle-income countries publish no granular GDP at all.
- Localized risks — floods, wildfires, hurricanes — operate well below admin boundaries.
- G-20 Data Gaps Initiative identifies granular physical and transition risk indicators as priority data gap.

WHAT POLICY NEEDS

Spatially Explicit, Sector-Specific Economic Data

- Climate & disaster risk assessment at the sub-administrative level.
- Place-based policy: infrastructure gaps, regional disparities.
- Cross-country comparability where official subnational accounts are sparse.

Our Contribution: Three Advances Over Existing Gridded GDP Work

01

Industry Dimension

First gridded GDP dataset that disaggregates by 10 economic activities, not just total output. Each cell carries an interpretable industry composition.

02

Reconciled With National Accounts

Quadratic reconciliation (Di Fonzo & Marini, 2011) ensures grid-level estimates aggregate exactly to official totals — by region and by economic activity.

03

Policy-Ready Granularity

1 km² output, designed as input for IMF GeoPulse (forthcoming) — combinable with hazard, financial, and demographic layers for place-based analysis.

Two-Step Framework: Spatial Random Forest ▶ Quadratic Reconciliation to Official Accounts



PART I

Methodology

Data Sources · Two-Step Pipeline · Quadratic Reconciliation

Data: Official National Accounts × High-Resolution Geospatial Data

OFFICIAL TARGETS

National & Subnational GDP

- BEA — country, state, county
- StatCan — country, province
- GGDC 10-sector classification
- Anchored to WEO PPP\$ for international comparability

EARTH OBSERVATIONS

High-Resolution Covariates

- Nighttime lights (VIIRS)
- Population grids (WorldPop)
- Built-up area & buildings
- Land cover (ESA 11 classes)
- NO₂ emissions, temperature, precipitation
- Distance to highways, built surface, cropland
- Heavy industry plants within 50 km

KEY CORRELATIONS

Subnational GDP Correlates With

0.95

correlation with population

0.94

correlation with nighttime light radiance

...but predictive power varies sharply across the 10 sectors — services and government are the hardest to detect from space.

Method: A Two-Step Pipeline — Predict, Then Reconcile

1. ASSEMBLE

Stack ~70 geocovariates at WorldPop 100 m. Aggregate to admin level for training; keep at grid for prediction.

2. TRAIN

Spatial random forest (CAST package, R). Forward feature selection per response variable.

3. PREDICT

Disaggregate GDP and 10 sector shares to grid. Constraint: shares sum to 1; zero in water/desert cells.

4. RECONCILE

Quadratic adjustment to official subnational totals — by region and by sector — preserving ML spatial pattern.

OUTPUT

1 km² Gridded GDP × 10 Economic Activities

Fully consistent with official national accounts at all aggregation levels.

Quadratic Reconciliation — Formal Statement

SETUP (Di Fonzo & Marini, 2011)

• Indices:

- g = grid cells
- k = industries (10 GGDC sectors)
- r = subnational regions

• Inputs:

- \hat{y} = ML grid–industry predictions (preliminary)
- y^* = reconciled estimates (target)
- H = sparse aggregation matrix (grids → regions)

• Constraints (exact):

- Region × industry totals must match official
- Region totals across industries must match official

Drawback: Proportional adjustment creates ‘step’ problem at borders of subnational areas

SOLUTION

Generalized least squares reconciliation:

$$y^* = \hat{y} + W H' (H W H')^{-1} (T - H \hat{y})$$

with $W = \text{diag}(\hat{y}^2)$ — proportional quadratic loss

KEY PROPERTIES

- Exact match with official regional & industry totals
- Proportional loss → relative deviations penalized
- Correction depends only on $(T - H \hat{y})$ — discrepancy
- H is highly sparse → sparse Cholesky factorization
- Scales to 1 km² grids, applied per country

PART II

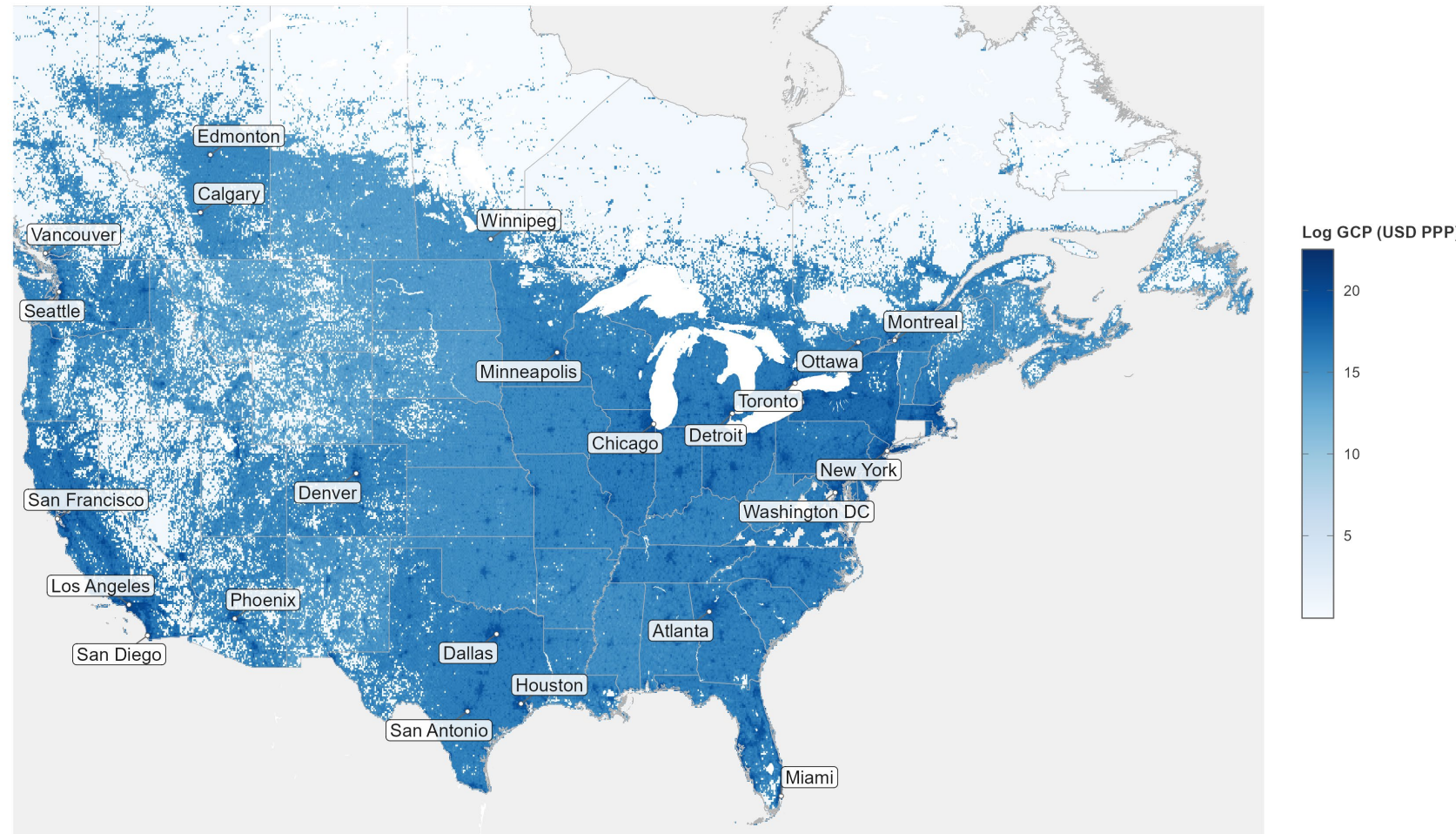
Application: Canada & United States

1 km² Gridded GDP × 10 Economic Activities · Reference Year 2021

Map — Total Economy: Gridded GDP, Canada & US, 2021

Gridded GDP, Total Economy, Canada and United States

Year: 2021 | Total Economy | Post-Benchmarking



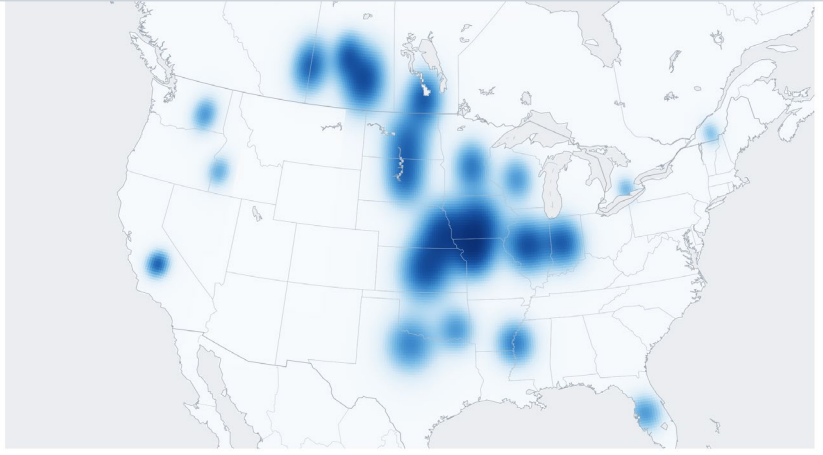
WHAT THE MAP SHOWS

Concentrated Corridors

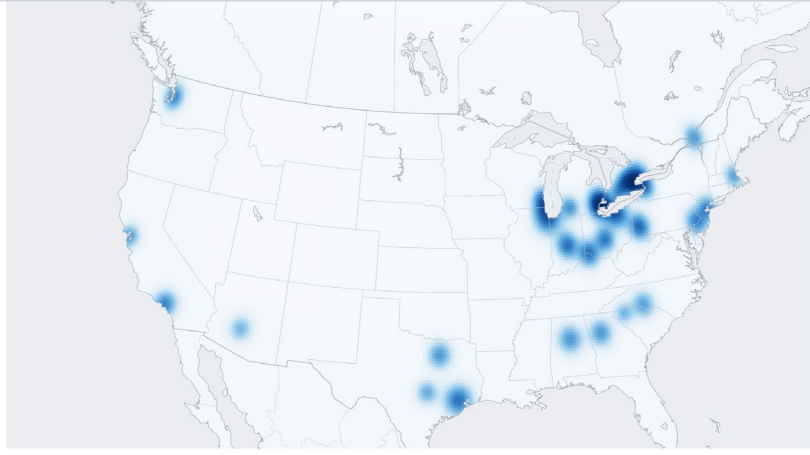
- **Boston–Washington Corridor**
the densest belt nationally
- **Great Lakes Belt**
Chicago → Detroit → Toronto
- **Windsor–Quebec Corridor**
Toronto · Ottawa · Montreal
- **Coastal Anchors**
LA, SF, Seattle, Vancouver
- *Near-zero in northern & rural areas*

Map — Industry Composition: Four Sectors, Canada & US, 2021

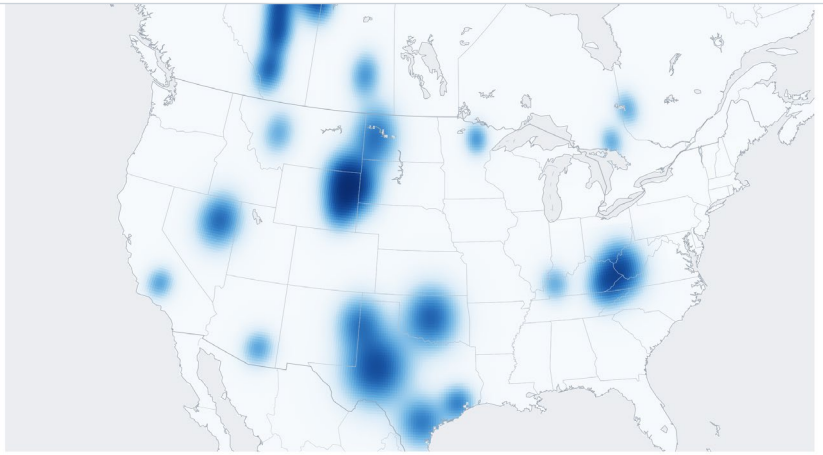
Agriculture



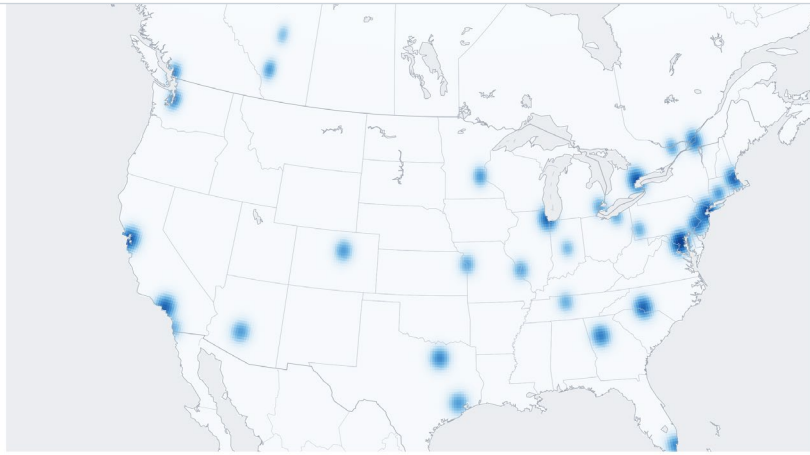
Manufacturing



Mining



Financial Intermediation



WHAT EACH MAP SHOWS

Distinct Spatial Patterns

Agriculture

Concentrated in the US Corn Belt (IA, IL, IN), Plains and the Canadian Prairies (SK, AB, MB).

Manufacturing

Great Lakes belt dominates: Detroit, Chicago, Cleveland, plus Ontario auto cluster.

Mining

TX/OK oil (Permian), Wyoming coal, Appalachian coal, Alberta oil sands, Northern Canadian hardrock.

Financial Intermediation

Boston–Washington corridor saturates dark; Toronto, SF Bay, Charlotte, Chicago strong.

Why It Matters: Industry dimension is what makes grid cell values interpretable — and what enables sector-specific risk overlays.

CAVEATS & LIMITATIONS

- Production interlinkages across space (mobile capital, intermediate goods, intangibles) are not directly observable at grid level.
- Service-sector activity leaves a weaker EO footprint than agriculture or manufacturing.
- Extractive industries in remote regions (Alaska, Nunavut) are systematically under-detected by nighttime lights.
- Reconciliation can only correct for aggregate consistency — within-region spatial pattern still rests on ML model.

GEOCOVARIATES — KEY GROUPS

- **Population & Demography**
WorldPop ~100m
- **Nighttime Lights**
VIIRS
- **Buildings & Built-Up Area**
Footprint count, density, area
- **Distances to Infrastructure**
Highway, built surface, cropland
- **Land Cover (11 Classes)**
ESA WorldCover
- **Climate**
Annual temperature, precipitation
- **Heavy Industry**
Iron, steel, cement plants <50 km
- **Air Quality**
NO₂ tropospheric column

OUR FUTURE WORK

1. Better Rural-Area Predictors

Add road-network density, agricultural productivity, distance-to-market.

2. Population Consistency Check

Avoid allocating GDP to grid cells with implausibly low population.

3. Geographic Extension

Central and South America, Europe, Sub-Saharan Africa, Asia.

4. Integration with GeoPulse

Feed gridded GDP into IMF's geospatial analytics platform for risk overlay.

Thank you · Marini (mmarini@imf.org) · Tebrake · Woldemichael · IMF Statistics Department

StatE0

5-7 May 2026 | ESA-ESRIN | Frascati (Rome), Italy



THANK YOU!