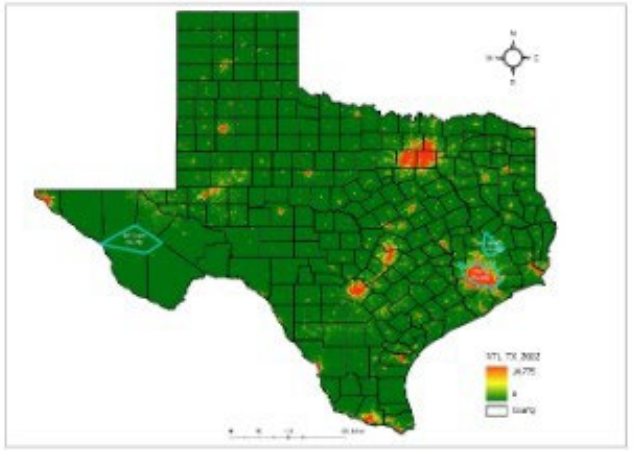


Applicability of the Nighttime Light Data as an Ancillary Tool to Estimate the Population at the County and Place Level of Texas

Muhammad Salaha Uddin, Ph.D., Special Research Associate, IDSER, UT San Antonio
Jeffrey A. Jordan, Ph.D., Associate Director, IDSER, UT San Antonio

Texas GIS Forum, Austin, Texas
Nov 06, 2025

IDSER
Institute for Demographic & Socioeconomic Research

Outline

1. Introduction
2. Datasets
3. Methods
4. Results
5. Conclusion and Future Work
6. Questions and Discussion

1. Introduction

- Texas Demographic Center (TDC) Population Estimates Program
- Different national, state, and local agencies utilize statistical methods and tools to estimate the population
- Conventional methods (Housing Unit, Ratio Correlation, and Component II) of population estimation and relative criticism
- Data availability, limitations, and time constraints
- Conduct robustness check of estimates from conventional methods by using alternative approaches and readily available datasets
- Utilize GIS and Remote Sensing data to supplement/enhance the population estimates
- Relative radiance of Nighttime Light (NTL) is a good proxy measure of economic activity

2. **Datasets**

US Census Bureau

- 2020 Decennial Census
- American Community Survey (ACS) 5-Year County Estimates
- Vintage 2022 County and Place Estimates (2021 and 2022)

REGRID - Matched Secondary Address Data

NASA Black Marble Product Suite - Nighttime Light (NTL)

- VNP46A4 for 2015 to 2022
- Yearly moonlight-adjusted snow-free time NTL's relative radiance
- Spatial resolution 15 arc-second (~450 meters)

2. **Datasets: US Census Bureau**

2020 Decennial Census

- 100% Count data (universe)
- Administered every 10 years
- Population, housing unit, and group quarters data
- Total Population used as the Dependent Variable for Base Model Development
- Occupied Housing Units used for calculating residential share of nighttime light radiance

American Community Survey (ACS) 5-Year Estimates

- Sample survey data (~1 in 38 households)
- Released every 5 years
- County-level data used (geographies down to Block Group)
- County estimates are used for trend analysis and exploring the relationship between NTL and the total county population, and calculating error metrics
- Occupied Housing Units used for calculating residential share of nighttime light radiance

Vintage 2022 County and Place Estimates (2021 and 2022)

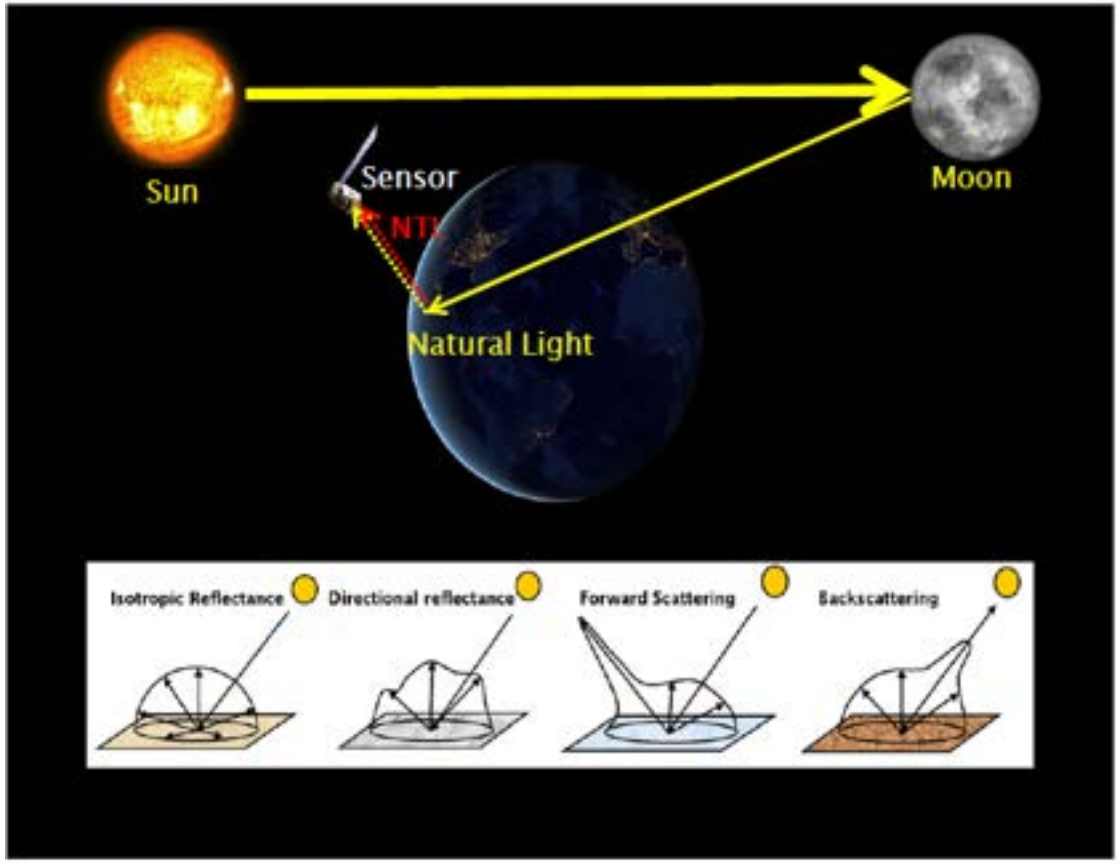
- Released annually
- Estimates after the Decennial Census to the current year may be revised
- County estimates are used for trend analysis and exploring the relationship between NTL and the total county population
- County and Place Estimates are used to calculate error metrics

2. **Datasets: Regrid**

REGRID - Matched Secondary Address Data

- Monthly updates
- Parcel records and primary situs addresses are matched to known secondary addresses
- Unit-level addresses for multi-unit buildings
- Geocoded address points
- All address points utilized in the study

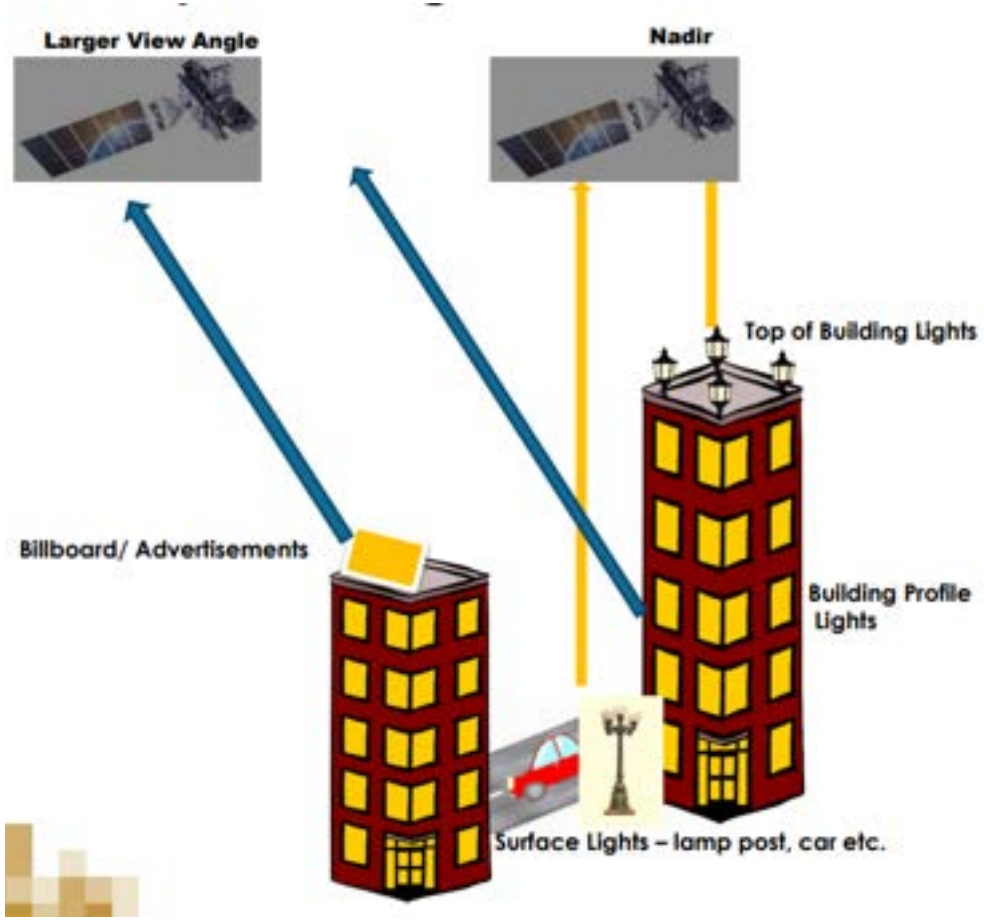
2. Datasets: NTL



Principles of Nighttime Remote Sensing

- Unlike in daytime remote sensing, there are multiple light sources
- Observations include moonlight, light directly emitted by a source (e.g., building and transport), and light scattered by the ground.
- Snow (both under moon-illuminated and moon-free conditions) can also increase the signal during winter months.
- Land features (such as buildings and trees) can also block the light source during different time periods.

2. Datasets: NTL



Principles of Nighttime Remote Sensing

- Light sources have different angular emission and reflection profiles
- Different satellite viewing angles may change light sources captured.
- Angular differences are more prominent in city centers (downtown areas) with tall buildings.

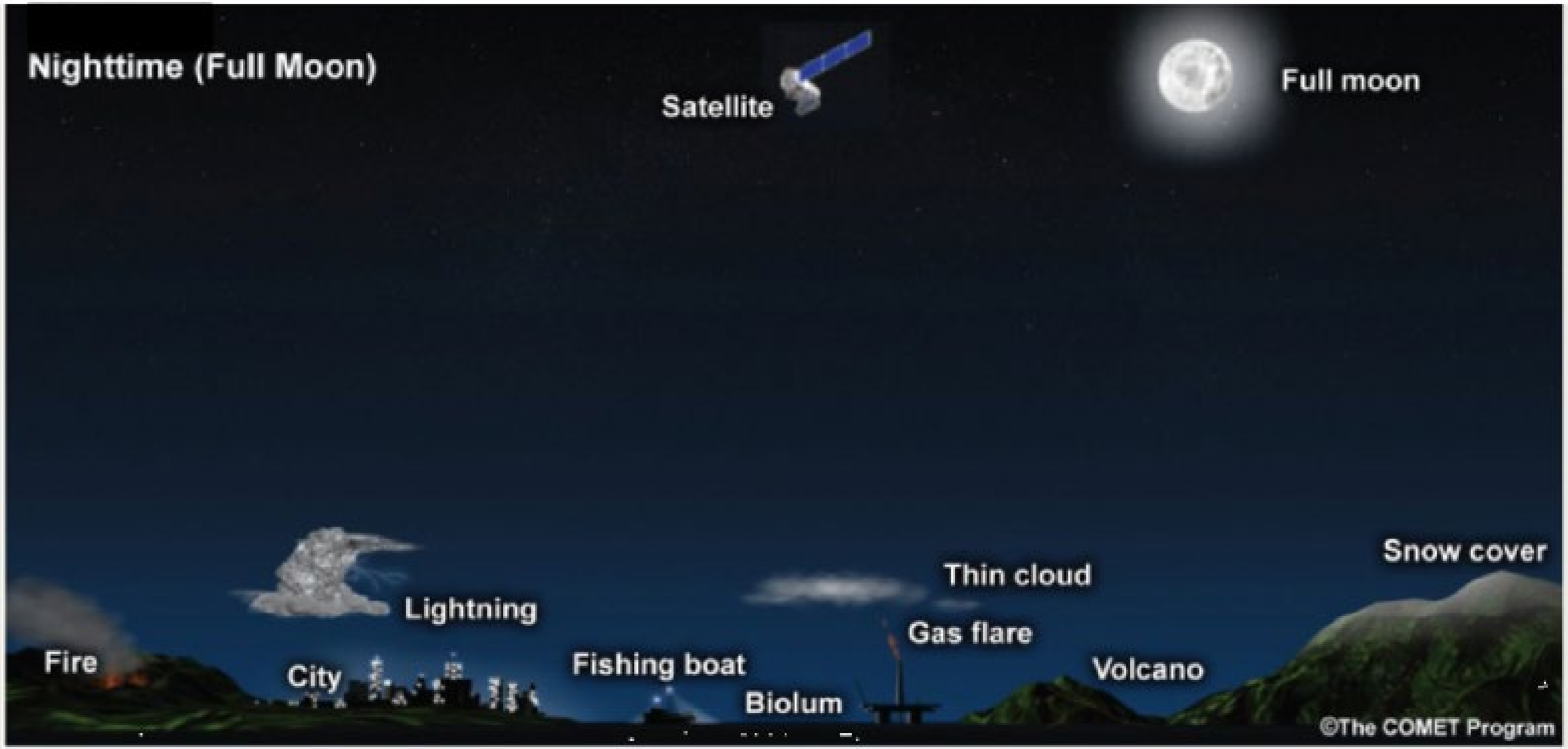
2. Datasets: NTL

Datasets: NTL Scientific Data Set (SDS) HDF Layers

AllAngle_Composite_Snow_Covered	Land_Water_Mask	OffNadir_Composite_Snow_Covered
AllAngle_Composite_Snow_Covered_Num	NearNadir_Composite_Snow_Covered	OffNadir_Composite_Snow_Covered_Num
AllAngle_Composite_Snow_Covered_Quality	NearNadir_Composite_Snow_Covered_Num	OffNadir_Composite_Snow_Covered_Quality
AllAngle_Composite_Snow_Covered_Std	NearNadir_Composite_Snow_Covered_Quality	OffNadir_Composite_Snow_Covered_Std
AllAngle_Composite_Snow_Free	NearNadir_Composite_Snow_Covered_Std	OffNadir_Composite_Snow_Free
AllAngle_Composite_Snow_Free_Num	NearNadir_Composite_Snow_Free	OffNadir_Composite_Snow_Free_Num
AllAngle_Composite_Snow_Free_Quality	NearNadir_Composite_Snow_Free_Num	OffNadir_Composite_Snow_Free_Quality
AllAngle_Composite_Snow_Free_Std	NearNadir_Composite_Snow_Free_Quality	OffNadir_Composite_Snow_Free_Std
DNB_Platform	NearNadir_Composite_Snow_Free_Std	

SDS HDF Layer	Units	Description	Bit Types	Fill Value	Valid Range
AllAngle_Composite_Snow_Free	nWatts・cm−2 .sr−1	Temporal Radiance Composite Using All Observations During Snow-free Period	16-bit unsigned integer	65,535	0 – 65,534

What can we study using Nighttime Light data?

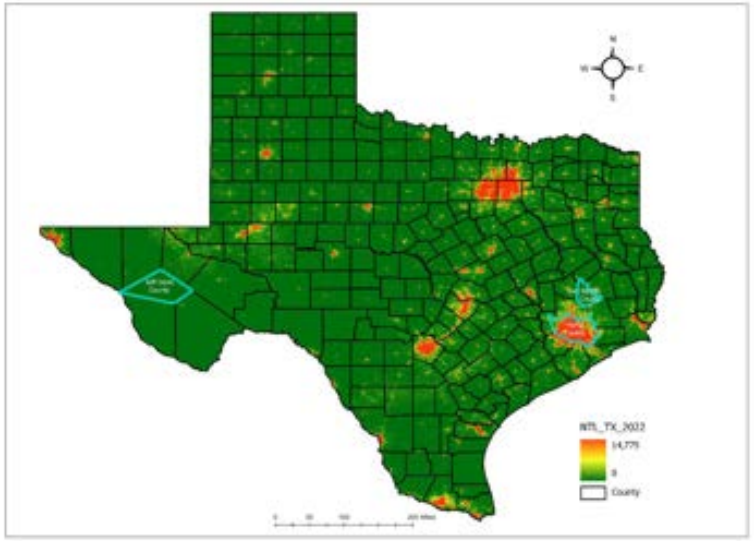


3. Methods

Methods: Data Processing

- Aggregated the cell values (relative radiance) at the County and Census Tract Level for 2015 to 2022
- Partial cell values are scaled by the overlapped area to the total area of the cell.

$$\textbf{County's Yearly NTL Radiance} = \sum_{i}^{n} Fully\ Contained\ Cell\ Value + \sum_{i}^{m} Partially\ Contained\ Cell\ Value$$



- At CT, yearly NTL's relative radiance scaled by the occupied housing unit to total number of address within the census tracts.

$$\textbf{Census Tract's Yearly NTL Radiance} = \left(\begin{matrix}\sum_{i}^{n} Fully\ Contained\ Cell\ Value \\ +\sum_{i}^{m} Partially\ Contained\ Cell\ Value\end{matrix}\right) * \frac{Total\ Occupied\ Housing\ Units\ in\ a\ Year}{Total\ Number\ of\ Address\ within\ a\ census\ Tract}$$

3. Methods

Methods: Model Development

- Global (Linear and Spatial Lag) vs. Local Regression (Geographically Weighted Poisson Regression: GWPR) Model
- County Level Development

$$E(Y_i) = e^{(\beta_{0i}\,(u_i) + \beta_{1i}(u_i) + \varepsilon)}$$

Where: $E(Y_i)$ represents the expected count of population for a County

β_{0i} is the local constant (in this case, the County) associated with the explanatory variable of NTL radiance for the ith location (County)

β_{1i} is the local coefficient associated with the explanatory variable of NTL radiance for the ith location (County)

- Census Tract Level Model

3. Methods

Methods: County Pop Estimation

Pop estimation from base model

- County pop estimation using the input data (NTL) in the base model
- Error Metrics Evaluation: APE, PE (over- and underestimation) by population density, population size, development type (rurality, urbanity)

Defined based on rural and urban population proportions as of the 2020 census

- Completely Rural (with a 100% rural population)
- Predominately Rural (Rural population 80% to 99.9%)
- Mostly Rural (Rural population 60% to 79.9%)
- Partially Rural (Rural population 40% to 59.9%)
- Semi-Urban (Rural population 20% to 39.9%)
- Urban (less than 20% Rural Population)

3. Methods

Methods: Place Pop Estimation

- Identify (Evaluation the error metrics) the County type and select a county and place for population estimation.
- Model estimated total populations at the CT level are adjusted in two steps.

$$ADJ\_TPOP\_CT_i = MD\_EST\_TPOP\_CT_i * \frac{CB\_EST\_TPOP\_CNT}{\sum_i^n MD\_EST\_TPOP\_CT_i} \ldots\ldots\ldots\ldots \text{(Adjustment 1)}$$

Ratio between the Census-estimated total population of a county to the total estimated population of all CTs in a County.

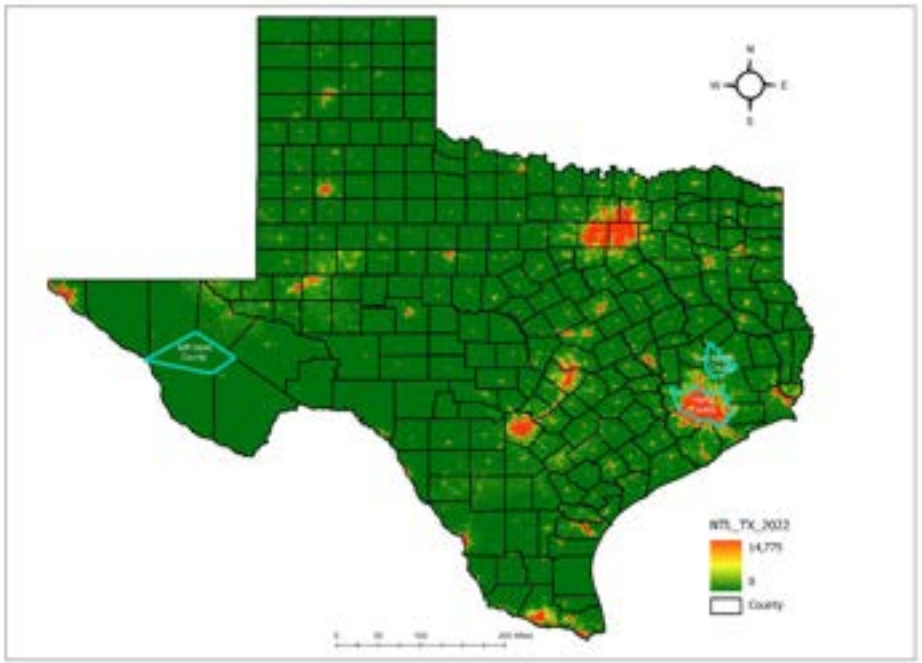
$$ADJ\_POP\_PCT_j = MD\_EST\_TPOP\_PCT_j * \frac{Res\_Add\_Wth\_PL\_PCT_j}{Tot\_Res\_Add\_Wth\_PCT_j} \ldots\ldots\ldots \text{(Adjustment 2)}$$

Ratio between the residential addresses located within the Place of a CT to the total residential addresses within a CT

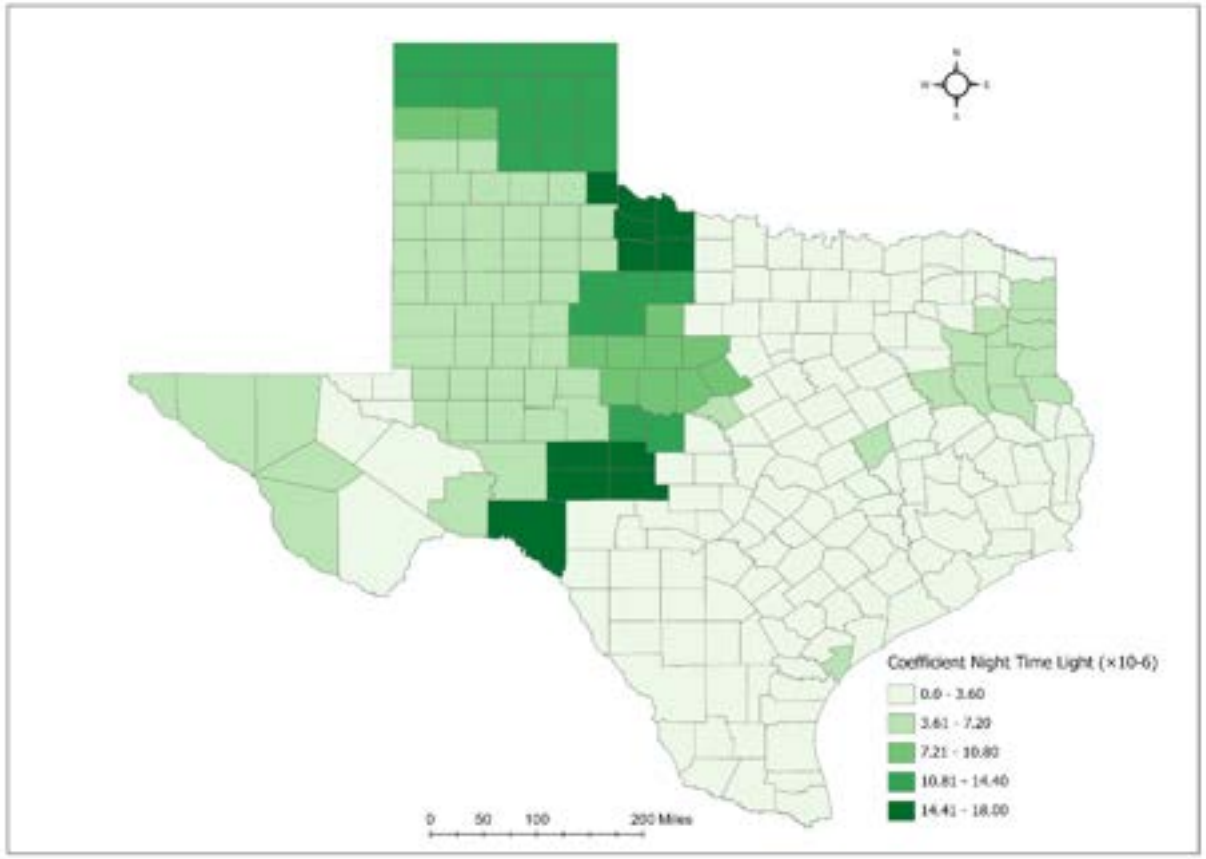
- $Total\ Place\ Population = \sum_i^n ADJ\_TPOP\_FCT_i + \sum_j^m Adj\_TPOP\_PCT_j$

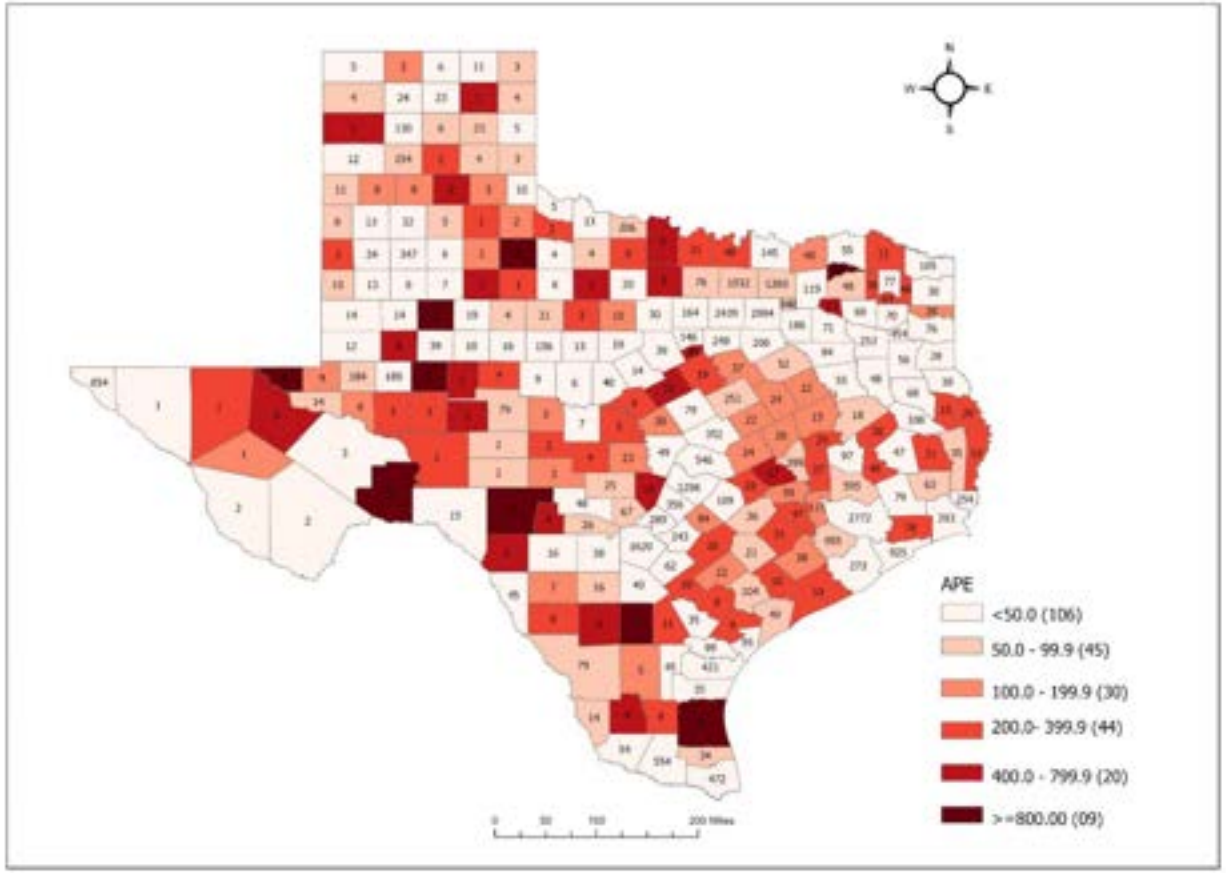
4. Results: Distribution of NTL Radiance

Year	Each Cell's NTL Relative Radiance Range	Min Value of Aggregated NTL's Relative Radiance at County Level	Max Value of Aggregated NTL's Relative Radiance at County Level	Median Value of Aggregated NTL's Relative Radiance at County Level
2020	0 - 8,652	400 (Jeff Davis)	9,326,682 (Harris)	47,782 (Deaf Smith)
2021	0 - 6,471	376 (Jeff Davis)	9,450,038 (Harris)	47,102 (Robertson)
2022	0 – 14,775	567 (Jeff Davis)	10,588,687 (Harris)	65,609 (Sa Jacinto)
Source: NASA's VNP46A4 data product and Author's calculation				



4. Results: County Model





➔ R^2 in Global Spatial Lag Model 0.97 overall Deviance explained by GWPR is 0.92.

➔ NTL coefficient's heterogeneity is observed

➔ Coefficient varies from 0-18*10^{-6}

➔ NTL estimation provides better results for County with high number of population and higher population density

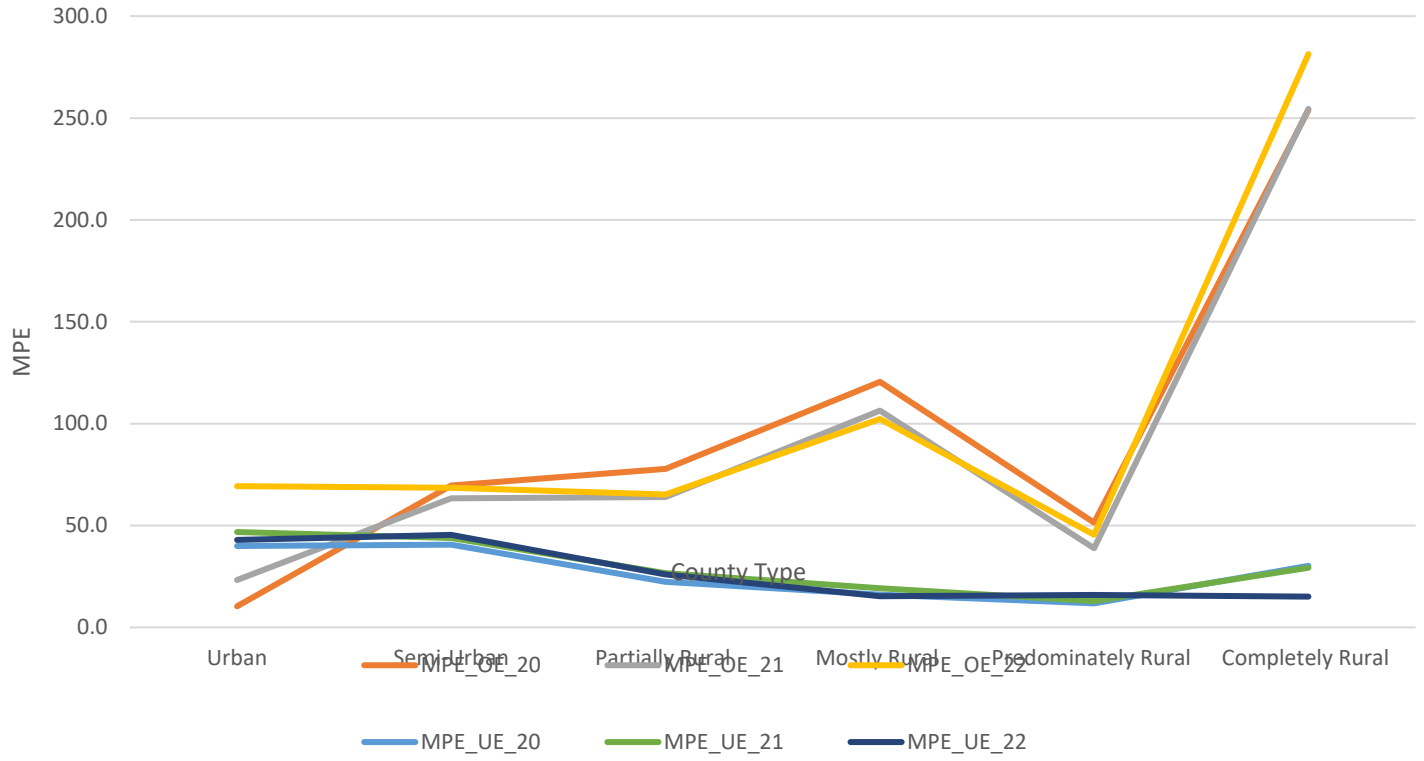
➔ NTL based model estimation provides better results for low density county (0-99) if population size belongs to 0.32 to1.2 million

4. Results: County Estimation

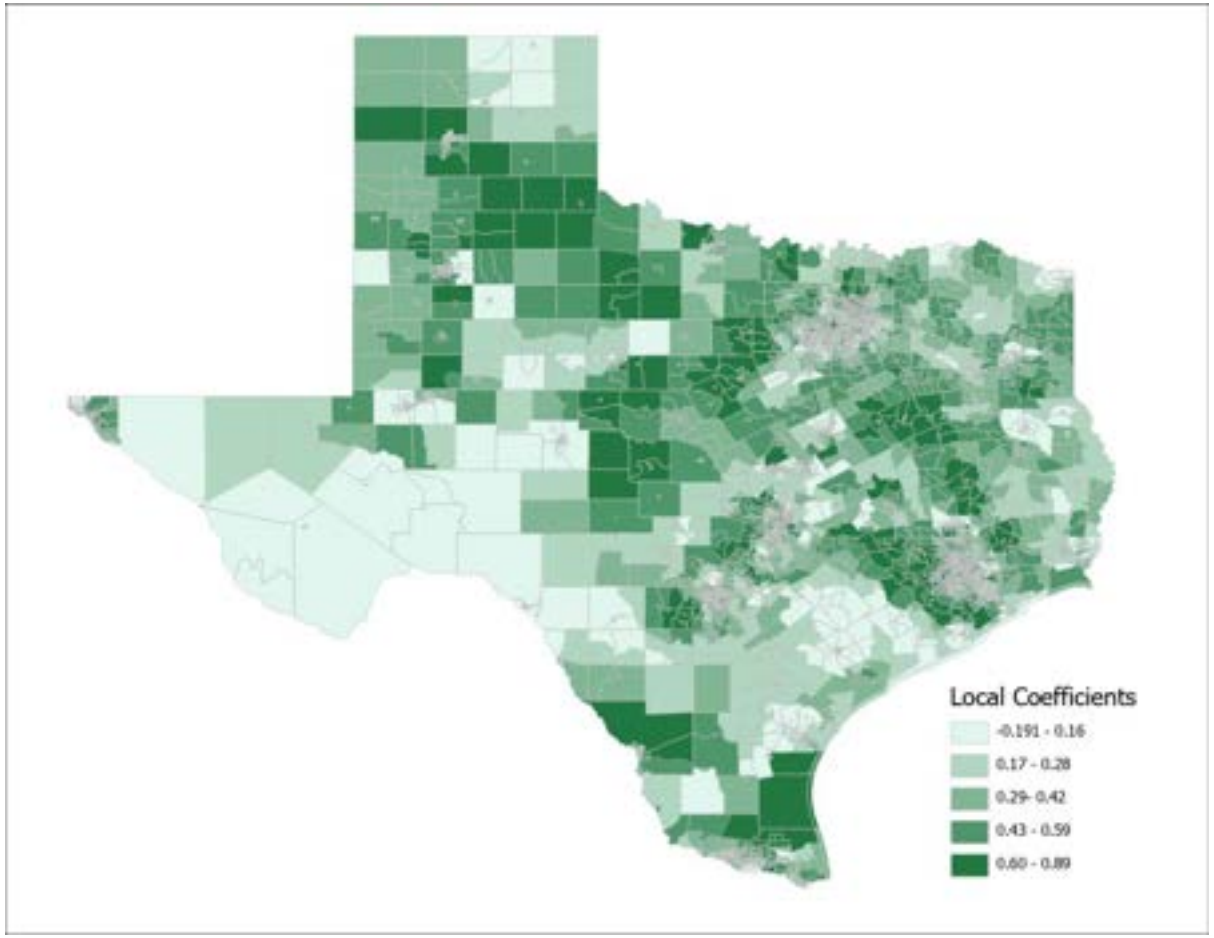
➔ Majority of underestimations are mostly (63.2) observed in urban Counties while overestimations are predominately (97.8) observed in completely rural Counties.

➔MPE for under estimation decreases with increase rurality of the counties.

➔MPE for over estimation increase with increase rurality of the counties.



4. Results: CT Model



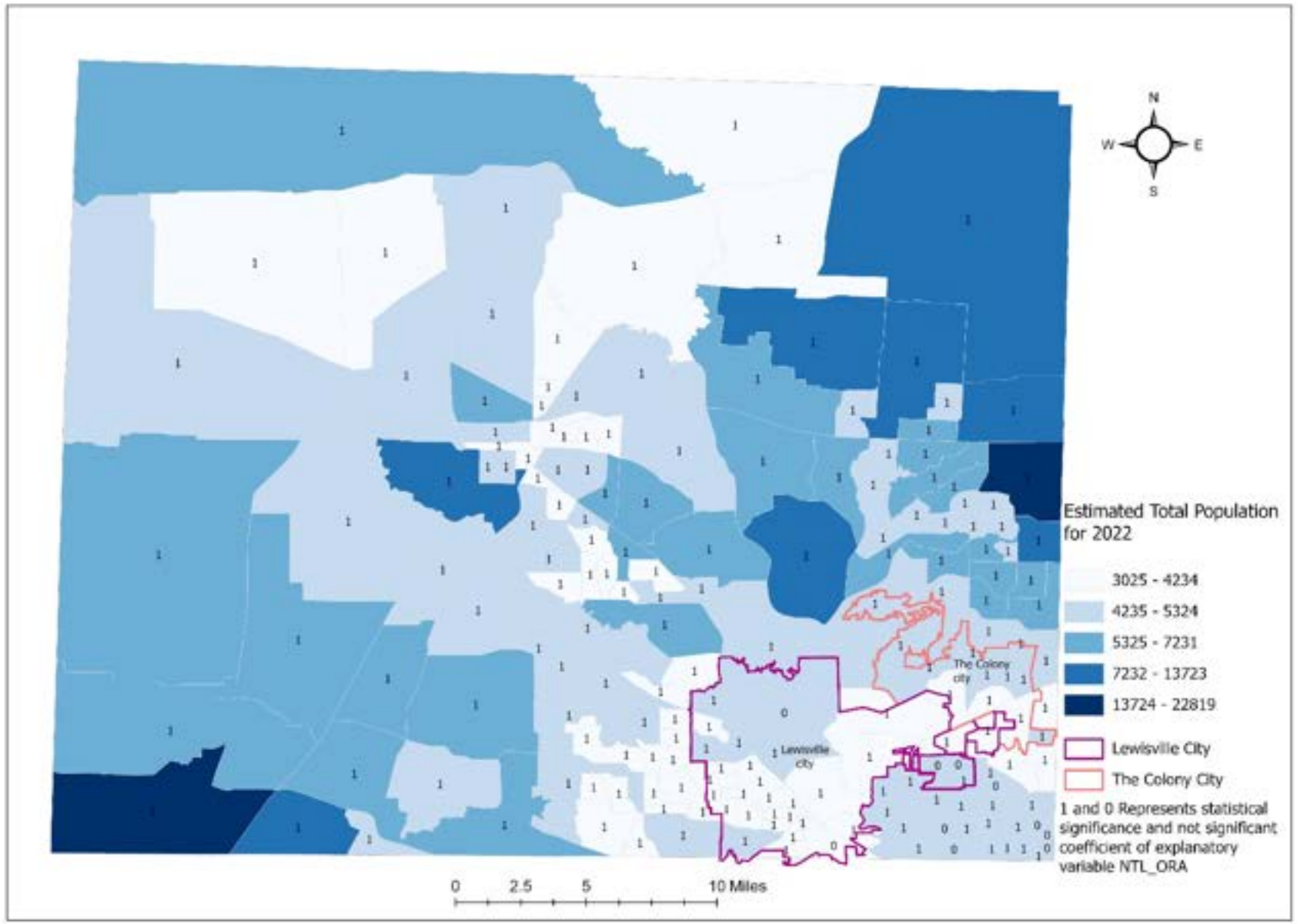
➔Overall deviance explained by GWPR at CT level is 0.57.

➔NTL coefficient's heterogeneity is observed

➔ Coefficient varies from -0.19-0.81

➔ Scaled NTL estimation provides better results for CTs in the County with high number of population and higher population density

Results: Place Population Estimation



4. Results: Place Population Estimation

	Denton County	Lewisville	Colony
Number of Census Tracts (CT)	193	29	12
CB Estimated total population for 2022	977,281	131,215	45,367
NTL Model-based estimation at the County level	407,013	-	-
APE	58.4	-	-
Aggregated estimated population of CTs (Without Adjustment)	954,012	146,299	53,621
APE	2.4	11.5	18.2
Adjustment Factor	1.02	-	-
Aggregated estimated population of CTs (After Adjustment)	977,281	149,869	54,919
APE	0	14.2	21.1
Adjustment due to partial CT with settlements within the city boundary		132,139	46,171
APE	0	0.7	1.8

5. Conclusion and Future Work

- NTL proves to be a reliable estimator of the total population at specific geographic levels
- At the individual County level, NTL estimates are comparable to Census Bureau (CB) estimates, particularly in counties classified as urban and semi-urban.
- NTL is applicable for estimation of Places that are located within Urban and Semi-Urban counties
- The COVID-19 pandemic has limited human activity this year, potentially affecting the precision of the local coefficients obtained from the 2020 database
- However, observed magnitudes of estimated errors are consistent by population size and County type for pre and post pandemic years, hence negating the influence of COVID-19 on the base year model.
- The application of NTL data for population estimation holds significant potential, especially when traditional data sources are unavailable.
- NTL data has substantial potential for estimating populations in smaller areas of urban and semi-urban counties.
- Improve the ratio correlation method of population estimation incorporating NTL as a predictive variable.

Thank You!

Q & A

Muhammad Salaha Uddin, Ph.D.

muhammad.uddin@utsa.edu

Jeffrey A. Jordan, Ph.D.

jeff.jordan@utsa.edu

IDSER
Institute for Demographic & Socioeconomic Research

UT San Antonio™
The University of Texas at San Antonio