Urban Area Growth Projection until 2035.

CSE 881: Project Intermediate Report
Pavan Yachamaneni
Sandeep Vemulapalli
Thanishq Tanmay

Difficulty level: Difficulty Level is the same as written in the proposal with medium in level. With components involving

a. Data Extraction
b. Data Engineering
c. Data Preparation
d. Forecasting Model Selection
e. Model Evaluation
f. Result in the form of Dashboard.

Abstract:

To predict urban growth, it's essential to choose a reliable metric that accurately represents the expansion of urban areas. In this study, we focus on the vertical growth of urban areas, which can be effectively monitored using satellite microwave backscatter techniques. Microwave backscatter, often referred to as "backtracking," is a method where microwaves are emitted towards the Earth's surface and the reflected signals are captured by satellites.

The strength and pattern of the reflected signals provide valuable insights into the characteristics of the surface. Buildings and other urban infrastructures have unique geometric properties that cause distinct reflection patterns, differentiating them from natural terrains. By analyzing these reflection patterns, which can be quantified into backscatter values ranging from 40 dB (indicating smooth surfaces) to 10 dB (representing rough surfaces like urban areas), we can effectively track and predict urban growth patterns.

Microwave backscatter is particularly sensitive to changes in surface roughness and dielectric properties. Urban regions, characterized by buildings and other structures, create pronounced corner reflections due to the interplay between built structures and the ground. This phenomenon leads to elevated backscatter values when compared to

natural or agricultural terrains. Consequently, a rise in backscatter values can be interpreted as an increase in urban structures.

Leveraging historical microwave backscatter data, this project aims to forecast urban growth in rapidly developing regions of South India. The goal is not only to quantify the growth of urban areas but also to determine the rate at which they are expanding.

Work Done Till Now: Till now Data Selection, Data Collection, Preliminary Data Exploration was made. Source of Data: NASA SEDAC

Part I Dataset Description:

The research paper titled "A global urban microwave backscatter time series data set for 1993–2020 using ERS, QuikSCAT, and ASCAT data" by Steve Frolking provides information about a dataset that captures urban microwave backscatter over time. Given a specific location in pixel, you can extract the backscatter values (σ°) in dB for that location across all the years (1993–2020) using the sensors mentioned. The data is available in both monthly and seasonal formats, allowing for a detailed analysis of changes over time.

Here are the relevant details about the features that can be extracted for a particular location in pixel across all the years:

1. Dataset Name: Global Monthly and Seasonal Urban and Land Backscatter Time Series (1993–2020).

2. Sensors and Eras:
- ERS (January 1993 – December 2000)
- QuikSCAT (July 1999 – November 2009)
- ASCAT (January 2007 – December 2020)

3. Spatial Metadata:
- Extent X: −180 to +180
- Extent Y: −58 to +64.5
- Resolution: 0.05 decimal degrees
- Coordinate reference system: longitude/latitude (WGS84 datum)
- Temporal resolution: monthly (12/yr) or seasonal (4/yr) JanMar, AprJun, JulSep, OctDec.

Part II Data Preprocessing:

1. Radiometric Calibration: Convert raw digital numbers (DN) from the satellite image to physical units, typically expressed in decibels (dB). This ensures that the data is consistent and comparable across different dates and sensors.
2. Maintain Data Uniformity (Not Done Completely): From the provided information, the sensors and their respective bands are:
   i. Cband Sensors: ERS ASCAT
   ii. Kuband Sensor: QuikSCAT

   To make the data uniform across all years, especially when combining data from different sensors and bands, we ran through the following steps:
   1. Radiometric Calibration
   2. Spatial Resampling
   3. Normalization

Part III Data Preparation:

The core aspect of the project is to have data prepared in such a way that we can extract maximum information from spatiotemporal data. Therefore our feature selection has 3 categories.

1. Statistical Features: Here given any location, we also select nearby data points which are around 380 sq km and add summarization features to capture spatial data and add values as we can particularly bank on single pixel location. The statistical feature we planned to add are:
   i. Mean
   ii. Standard Deviation
   iii. Skewness
   iv. Kurtosis

2. Texture Features:
   a. The GrayLevel Cooccurrence Matrix (GLCM) is a statistical method used in image processing and computer vision to analyze the texture and spatial relationships of pixels within an image. It is particularly valuable for characterizing the patterns and structures present in an image. GLCM is commonly employed in tasks like texture classification, segmentation, and feature extraction.

b. Texture Features are:
    i. Contrast
    ii. Dissimilarity
    iii. Homogeneity
    iv. Energy
    v. Correlation

3. Spatial Autocorrelation:
   Measures like Moran's I or Getis-Ord Gi* can be used to determine the spatial autocorrelation of backscatter values. High positive autocorrelation might indicate urban clusters.

4. Lagged Features:
   a. For each year, extract the features as before but also include the features from n previous years as lagged features.
   b. For now we consider n as a hyper-parameter.

Part IV Project Time-Line:

Oct 28 - Nov 07: Data Preparation
Nov 07 - Nov 20: Model Selection which involves ARIMA, LSTM, GRU
Nov 20 - Dec 02: Model Evaluation
Dec 02 - Dec 05: Dashboard Building

Part V Project Individual Contributions

1. Data Processing-Preparation and Dashboard building by Pavan.
2. Model Selection, Hyperparameter tuning by Sandeep.
3. Model Evaluation by Tanishq.