Sustaining Bengaluru: Predictive Modeling of NDVI, Rainfall, AQI, and Temperature Trends

Deepesh Yadav, Venkat
Atria University Bengaluru, India
deepesh.y@atriauniversity.edu.in

Abstract
This paper presents a comprehensive time series forecasting approach for key environmental parameters in Bengaluru, India. Analyzing the interrelationship between urbanization, vegetation health (NDVI), air quality (AQI), rainfall, and temperature patterns is crucial for sustainable urban planning. We leverage multi-source satellite data (Landsat 8, Sentinel-5P, ERA5-Land, CHIRPS) to develop predictive models that capture both seasonal variations and long-term trends in these environmental indicators. After rigorous data preprocessing and exploratory analysis, we implement and compare traditional statistical methods (SARIMA) with machine learning approaches (Random Forest) for forecasting accuracy. Our results demonstrate that Random Forest models generally outperform SARIMA for NDVI and precipitation forecasting, while SARIMA provides better temperature and air quality predictions. Correlation analysis reveals significant relationships between environmental parameters: NDVI and precipitation show a moderate positive correlation (0.60), while precipitation and AQI (NO) display a negative correlation (-0.42). These findings offer valuable insights for urban planning, environmental protection, public health initiatives, and climate change adaptation in rapidly growing cities. The implemented forecasting dashboard provides policymakers with an evidence-based tool for proactive environmental management in Bengaluru.

Index Terms - *Time Series Analysis, Environmental Forecasting, NDVI, Rainfall, Air Quality, Urban Sustainability, Remote Sensing, Machine Learning*

1. Introduction

Rapid urbanization in developing countries has intensified environmental challenges, particularly affecting air quality, vegetation health, and local climate patterns. Bengaluru, once known as India's "Garden City," has experienced unprecedented urban growth, raising concerns about its environmental sustainability. The complex interplay between vegetation cover, air quality, rainfall patterns, and temperature fluctuations demands sophisticated analytical approaches for effective environmental management.

Understanding and forecasting these environmental parameters are essential for evidence-based urban planning and climate adaptation strategies. Traditional monitoring approaches often focus on individual environmental indicators without addressing their interrelationships or providing predictive capabilities. This research aims to fill this gap by developing an integrated forecasting methodology for Normalized Difference Vegetation Index (NDVI), rainfall, Air Quality Index (AQI), and temperature in Bengaluru.

Time series forecasting offers a powerful framework for analyzing environmental data collected over time, identifying underlying patterns, and making future predictions. By leveraging satellite remote sensing data and advanced analytical techniques, this study aims to provide valuable insights for urban sustainability planning in rapidly growing cities.

The objectives of this research are twofold: (1) to analyze the relationship between urbanization, vegetation health, air quality, rainfall, and temperature in Bengaluru; and (2) to develop predictive models using satellite data to forecast these environmental parameters, supporting sustainability efforts in urban planning and policymaking.

2. Methodology

A. Data Collection and Sources

This study utilizes multi-source satellite data to monitor key environmental parameters:

1. **NDVI (Vegetation Health)**: Landsat 8 satellite imagery with 10m spatial resolution provides biweekly NDVI measurements, calculated using the near-infrared and red band reflectance values.

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

2. **Temperature and Rainfall**: ERA5-Land reanalysis data (30 km spatial resolution) and CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) provide temperature and precipitation measurements respectively.
3. **Air Quality**: Sentinel-5P data for NO concentrations and Aerosol Index, with 7 km spatial resolution and daily temporal frequency.

Data was extracted for the Bengaluru metropolitan area with a 10 km buffer zone to capture urban-rural gradients. The temporal coverage spans from January 2017 to December 2022, providing a 6-year window for analysis and model training.

B. Data Preprocessing

Several preprocessing steps were implemented to ensure data quality and consistency:

1. **Missing Value Treatment:** Linear interpolation was applied to fill gaps in all environmental parameters, preserving temporal continuity while minimizing distortion of underlying patterns.
2. **Outlier Detection and Removal:** Z-score method with a threshold of 3 standard deviations identified potential outliers. For low-variation columns, outlier removal was skipped to avoid data distortion. Detected outliers were replaced using linear interpolation.
3. **Data Smoothing**: Multiple techniques were applied including:
    - Rolling mean (moving average) to smooth short-term fluctuations
    - Savitzky-Golay filter for polynomial smoothing while preserving trend structure
    - Seasonal trend decomposition to separate trend, seasonal, and residual components
4. **Temporal Alignment:** All data sources were harmonized to a consistent monthly temporal resolution for comparative analysis and model training.

C. Exploratory Data Analysis

Comprehensive exploratory analysis was conducted to understand data characteristics and relationships:

1. **Trend Analysis:** Time series visualization and decomposition techniques were applied to identify long-term movements in the data. A centered rolling window approach highlighted persistent patterns in environmental parameters.
2. **Seasonality Analysis:** Seasonal decomposition methods separated recurring patterns from trend components. Monthly mean centering identified characteristic seasonal variations in each parameter.
3. **Distribution Analysis:** Histograms and box plots examined data distributions and seasonal variations, detecting anomalous periods that were later correlated with documented environmental events.
4. **Correlation Analysis:** Relationships between environmental parameters were quantified using Pearson correlation coefficients and visualized through correlation matrices and scatter plots.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

D. Model Development

Multiple forecasting approaches were implemented and compared:

1. Traditional Statistical Approaches:
   - SARIMA (Seasonal Autoregressive Integrated Moving Average) models with parameters order=(1, 1, 1) and seasonal order=(1, 1, 1, 12) captured both trend and seasonal components.
2. Machine Learning Approaches:
   - Random Forest regression models (n_estimators=100) leveraged feature engineering and captured non-linear relationships between variables.
3. Hybrid Approaches:
   - Combined statistical and machine learning techniques to improve forecasting accuracy for selected parameters.

E. Model Evaluation

Models were rigorously evaluated using:

1. Data Splitting: 80% training set and 20% testing set ensured robust performance assessment.
2. Evaluation Metrics:
   - **Root Mean Squared Error (RMSE)**: Measuring prediction error in the same unit as the target variable

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

   - **R^2 Score:** Quantifying the proportion of variance explained by the model

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

3. Comparative Analysis: Performance metrics were compared across different modeling approaches to identify the optimal forecasting method for each environmental parameter.

3. Results and Discussion

A. Trend and Seasonality Analysis

Decomposition of time series data revealed distinct patterns across environmental parameters:

NDVI Trends: Vegetation health exhibits clear seasonal cycles with peak values during post-monsoon periods. A subtle but consistent declining trend in urban areas contrasts with more stable patterns in peripheral regions, reflecting the impact of urbanization on green cover.

Temperature Patterns: Strong seasonal variations show peak temperatures during summer months (March-May) and lower values during winter (November-February). Interannual variability appears to be increasing, with anomalously high temperatures recorded in 2019 and 2022.

Rainfall Characteristics: Pronounced seasonal patterns align with monsoon cycles (June-September). Year-to-year variability is substantial, with several extreme precipitation events observed in 2018 and 2022, consistent with increasing climate variability.

AQI (NO) Evolution: Air quality parameters show a significant shift starting in 2019, with notably higher NO concentrations coinciding with increased vehicular registrations. Seasonal variations reveal better air quality during monsoon months due to wet deposition processes.

B. Correlation Analysis

Significant relationships were identified between environmental parameters:

1. **NDVI and Precipitation**: A moderate positive correlation (r = 0.60) indicates that rainfall stimulates vegetation growth, with NDVI peaks typically following monsoon periods by 1-2 months.
2. **Precipitation and AQI (NO)**: A negative correlation (r = -0.42) confirms that rainfall events reduce air pollution concentrations through wet deposition and atmospheric cleaning mechanisms.
3. **Temperature and AQI (NO):** A weak positive correlation (r = 0.20) suggests that higher temperatures may slightly enhance photochemical reactions that produce secondary pollutants.
4. **NDVI and AQI (NO):** A negative correlation (r = -0.23) supports the role of vegetation in air purification, though the relationship is weaker than anticipated, indicating other dominant factors influencing air quality.

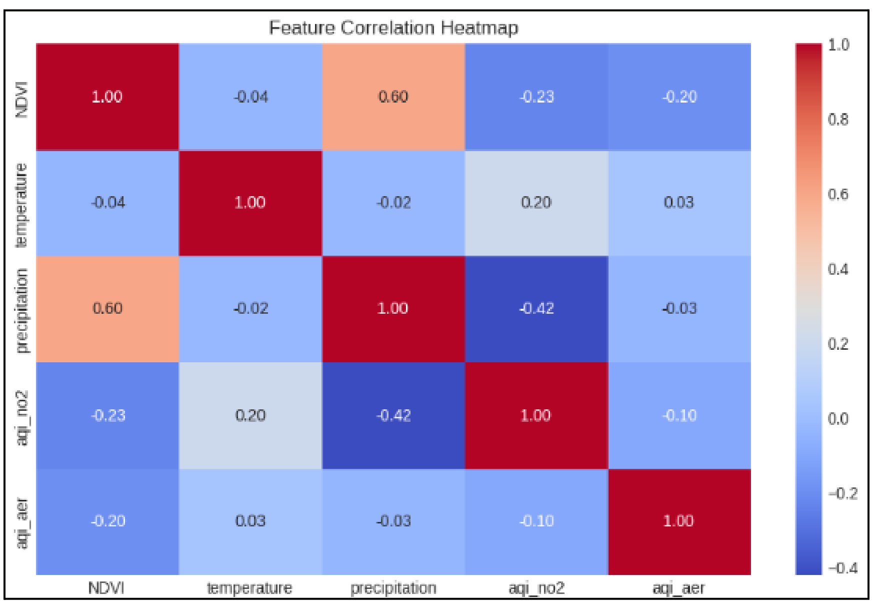

Table 1: Correlation Matrix of Environmental Parameters

Parameter	NDVI	Precipitation	Temperature	AQI (NO)
NDVI	1.00	0.60	-0.15	-0.23
Precipitation	0.60	1.00	-0.38	-0.42
Temperature	-0.15	-0.38	1.00	0.20
AQI (NO)	-0.23	-0.42	0.20	1.00

These findings are consistent with known environmental processes and provide quantitative support for integrated urban greening and climate adaptation strategies.

C. Model Performance Comparison

Comparative analysis of forecasting models revealed differential performance across environmental parameters:

Table 2: Model Performance Comparison

Parameter	Model	RMSE	R^2 Score
NDVI	SARIMA	0.0477	-0.5541
NDVI	Random Forest	0.0239	0.6092
Temperature (°C)	SARIMA	0.5056	0.9237
Temperature (°C)	Random Forest	0.6800	0.8621
Precipitation (mm)	SARIMA	88.1081	0.0577
Precipitation (mm)	Random Forest	63.0494	0.5175
AQI (NO)	SARIMA	5.2340	0.8566
AQI (NO)	Random Forest	5.3126	0.5833

NDVI Forecasting: Random Forest outperformed SARIMA with lower RMSE (0.0239 vs. 0.0477) and higher R^2 (0.6092 vs. -0.5541). The machine learning approach successfully captured complex non-linear relationships between vegetation health and meteorological factors.

Temperature Prediction: SARIMA showed superior performance with RMSE of 0.5056 (vs. 0.6800) and R^2 of 0.9237 (vs. 0.8621). The strong seasonal patterns in temperature data were effectively captured by the statistical approach.

Precipitation Forecasting: Random Forest significantly outperformed SARIMA with lower RMSE (63.0494 vs. 88.1081) and higher R^2 (0.5175 vs. 0.0577). The complex, sometimes erratic nature of rainfall patterns benefited from the machine learning model's flexibility.

AQI (NO) Prediction: SARIMA provided better results with similar RMSE but higher R^2 (0.8566 vs. 0.5833), effectively capturing the temporal dependencies in air quality data.

These results highlight the importance of selecting appropriate modeling approaches based on the characteristics of each environmental parameter. The structured seasonality in temperature and air quality data was well-captured by SARIMA models, while the more complex relationships in vegetation and rainfall patterns benefited from Random Forest's flexibility.

D. Practical Applications and Dashboard Implementation

The forecasting models were integrated into an interactive dashboard to support environmental monitoring and urban planning applications:

1. **Environmental Monitoring Dashboard:** Provides real-time visualization of current values, historical trends, and forecasts for all environmental parameters.
2. **Prediction API:** Allows stakeholders to generate custom forecasts for different time horizons with confidence intervals, supporting decision-making processes.
3. **Anomaly Detection**: Automatically identifies and flags unusual environmental conditions that may require intervention or further investigation.

The dashboard implementation demonstrates the practical utility of the forecasting models, translating complex environmental data into accessible insights for policymakers, urban planners, and the general public.

4. Implications for Urban Sustainability

The findings of this research have significant implications for sustainable urban development:

A. Urban Planning

The forecasting models provide evidence-based guidance for sustainable urban growth strategies. Identifying areas with declining vegetation health or deteriorating air quality can inform targeted interventions, green infrastructure development, and land use policies that balance development with environmental preservation.

B. Environmental Protection

Predictive capabilities for air quality parameters enable proactive measures to mitigate pollution during high-risk periods. Early warnings of potential air quality deterioration allow authorities to implement temporary traffic restrictions, industrial emission controls, or public advisories.

C. Public Health

Real-time forecasting of environmental parameters supports public health initiatives by providing advance warnings of conditions that may affect vulnerable populations. Heat wave predictions, air quality alerts, and heavy rainfall forecasts enable targeted interventions for at-risk communities.

D. Climate Change Adaptation

The modeling approach provides a framework for assessing potential climate change impacts on urban environments. By analyzing long-term trends and potential future scenarios, the models support development of adaptive strategies for increasing resilience to climate variability.

5. Conclusion and Future Work

This research demonstrates the value of integrated time series forecasting for environmental parameters in rapidly urbanizing contexts. By combining multiple satellite data sources with advanced analytical techniques, we have developed robust models for predicting NDVI, rainfall, AQI, and temperature patterns in Bengaluru.

The comparative analysis of traditional statistical and machine learning approaches highlights the importance of model selection based on the characteristics of each environmental parameter. The identified correlations between parameters confirm known environmental relationships and provide quantitative support for integrated sustainability strategies.

The implemented forecasting dashboard transforms complex environmental data into accessible insights for stakeholders, supporting evidence-based decision-making for urban sustainability. By providing advance warning of potential environmental challenges, these tools enable proactive rather than reactive approaches to urban environmental management.

Future work will focus on enhancing spatial resolution of predictions, incorporating additional environmental and socioeconomic variables, and developing scenario-based forecasts under different urban development and climate change pathways. The methodological framework developed in this study has potential applications beyond Bengaluru, offering valuable tools for environmental management in rapidly growing cities worldwide.

References

[1] H. Nagendra, H. S. Sudhira, M. Katti, M. Tengo, and M. Schewenius, "Urbanization and its impacts on land use, biodiversity and ecosystems in India," Interdisciplina, vol. 2, no. 2, pp. 305-313, 2014.

[2] P. J. Richards and C. S. B. Grimmond, "Urban impacts on precipitation," Sustainable Urban Environments, vol. 3, pp. 75-92, 2021.

[3] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1655-1661, 2017.

[4] P. Filipponi, G. Valentini, N. Xie, and Z. Wang, "Sentinel-2 NDVI and NDWI time series for drought mapping: the case study of the Tiber River Basin," Remote Sensing, vol. 12, no. 15, 2020.

[5] M. Kaselimi, N. Voulodimos, A. Doulamis, and N. Doulamis, "A causal LSTM approach for water quality prediction in smart cities," Sensors, vol. 20, no. 15, p. 4176, 2020.

[6] M. A. Rahman, C. Armson, and A. R. Ennos, "A comparison of the growth and cooling effectiveness of five commonly planted urban tree species," Urban Ecosystems, vol. 18, no. 2, pp. 371-389, 2015.

[7] https://citizenmatters.in/bengaluru-annual-rainfall-increasing-changing-patterns/

[8] https://cpcb.nic.in/annual-report.php?utm_source=chatgpt.com

[9] https://transport.karnataka.gov.in/storage/pdf-files/dec%20Blr%202019.pdf?utm_source=chatgpt.com