Magazine of Geodesy – Cartography
Vol 11, Nº Special (12/2025), ISSN: 2615-9481

Tạp chí Trắc địa - Bản đồ
Tập 11, Số Đặc biệt (12/2025), ISSN: 2615-9481

Spatial Heterogeneity Analysis and Machine Learning-Based Forecasting of Land Subsidence in Ho Chi Minh City: A GWR and ConvLSTM Integrated Approach

Truong Vu Cuong[1], Nguyen Thanh Nhuan[2], Ho Tong Minh Dinh[3*], Le Trung Chon[1]

[1]Ho Chi Minh City University of Natural Resources and Environment, Viet Nam

[2]Ho Chi Minh City Department of Agriculture and Environment

[3]National Research Institute for Agriculture, Food and Environment, France

Corresponding Author Email: chonlt@hcmunre.edu.vn

https://doi.org/10.5281/zenodo.18477050

Abstract:

This study develops a framework for simulating and forecasting land subsidence in Ho Chi Minh City, specifically focusing on District 12. By utilizing InSAR time-series subsidence data from 2015-2020 alongside influencing factors such as building density, distance to water bodies, and land use types, the research employs Geographically Weighted Regression (GWR) to analyze the underlying subsidence mechanisms. Experimental results demonstrate significant spatial heterogeneity in land deformation, where Land Use and Distance to Water emerge as the most dominant factors, with average regression coefficients of -0.390 and -0.344, respectively. Furthermore, the study proposes an integrated forecasting system architecture leveraging advanced Machine Learning models, including Random Forest, XGBoost, and ConvLSTM deep learning architectures to predict future surface deformation. Risk zonation results derived from K-means clustering provide effective visual tools for urban planning and early warning systems for geological hazards.

Keywords: *Land subsidence, InSAR, Geographically Weighted Regression (GWR), Machine Learning, Ho Chi Minh City, Ground deformation, Urban planning*

Submission received: 14/11/2025 Revised: 16/12/2025 Accepted: 19/12/2025 Published: 31/12/2025

1. Introduction

Land subsidence is a critical geological challenge facing major global metropolises, particularly those situated on low-lying river deltas with soft sedimentary foundations, such as Ho Chi Minh City (HCMC). Driven by rapid urbanization, increasing infrastructural loads, and excessive groundwater extraction, ground deformation has caused significant damage to drainage systems, transportation networks, and structural integrity. District 12, characterized by a dense network of rivers and canals and undergoing intensive land-use conversion, serves as a representative case study for these complex spatial dynamics.

In recent decades, Interferometric Synthetic Aperture Radar (InSAR) has emerged as a premier tool for monitoring subsidence due to its wide coverage and millimeter-level precision. However, interpreting the underlying drivers and forecasting future trends remain significant challenges. Traditional statistical methods, notably Ordinary Least Squares (OLS) regression, often assume that the relationship between variables is stationary across the study area. In reality, land subsidence is a phenomenon characterized by strong spatial heterogeneity, where factors such as geology, building density, and proximity to water bodies influence deformation in varying ways depending on the geographic location.

To address these limitations, Geographically Weighted Regression (GWR) has been introduced as an effective tool to explore the spatial non-stationarity of regression coefficients. Concurrently, advancements in Artificial Intelligence (AI) and Deep Learning, specifically architectures like Convolutional Long Short-Term Memory (ConvLSTM), have enabled high-precision deformation forecasting based on spatio-temporal datasets. Despite these advancements, the integration of rigorous spatial analysis with state-of-the-art machine learning models remains underutilized in subsidence studies within the Vietnamese context.

This study aims to develop an integrated framework combining InSAR techniques, GWR modeling, and machine learning algorithms (XGBoost, ConvLSTM) to analyze the mechanisms and forecast ground deformation in HCMC. The novelty of this research lies in the use of GWR to quantify the influence of infrastructural and natural factors, coupled with Explainable AI (SHAP) to provide a transparent understanding of the subsidence process. The findings not only contribute to the methodological discourse but also provide practical decision-support tools for sustainable urban planning and disaster risk reduction.

2. Methodology

The research methodology is structured into an integrated pipeline comprising four major phases: (1) Data acquisition and spatial preprocessing; (2) Spatial heterogeneity analysis using Geographically Weighted Regression (GWR); (3) Predictive modeling using Machine Learning and Deep Learning architectures; and (4) Model interpretation via Explainable AI (XAI).

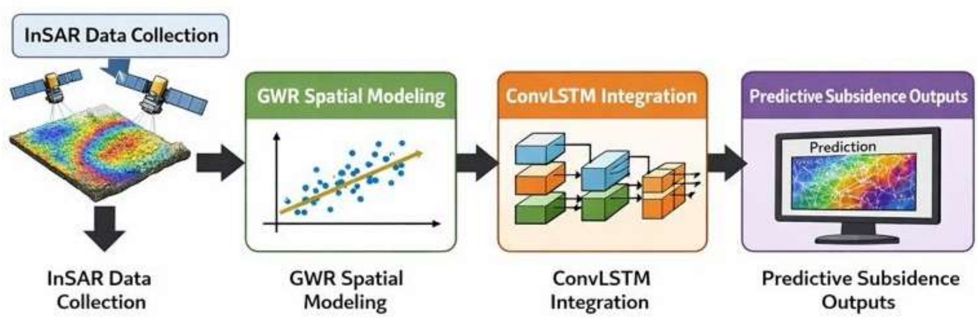

Fig.1 Methodological Framework

2.1 Data Sources and Spatial Variables

The primary datasets used to test and evaluate the GWR spatial regression model include:

i. Subsidence data for the period 2015–2020 determined using PS/DS InSAR techniques from Sentinel-1 radar satellite imagery over District 12, Ho Chi Minh City.
ii. Topographic data at a scale of 1:2000 containing layers such as terrain, buildings, transportation networks, and hydrography, including major and important traffic routes. This dataset was published in 2020 and has been converted into raster format.
iii. Land Use and Land Cover Change (LULCC) data of District 12 interpreted from Sentinel-2 imagery for the period 2015–2020, provided by the Copernicus Programme of the European Space Agency (ESA).land subsidence time-series data derived from Interferometric Synthetic Aperture Radar (InSAR) during the 2015-2020 period.

To analyze the drivers of deformation, three independent spatial variables were selected based on urban characteristics:

Building Density (Building): Representing static loads from urban infrastructure.

Distance to Surface Water (DistToWater): Reflecting the influence of hydrogeological conditions and soft soil distribution near river systems.

Land Use Type (LandUse): Categorizing the human-induced impact on the terrain.

2.2. Geographically Weighted Regression (GWR)

To address the spatial non-stationarity of subsidence drivers, this study employs the GWR model. Unlike global models (e.g., OLS), GWR allows parameters to vary locally across the study area:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_{bw_k}(u_i, v_i)x_{ik} + \epsilon_i$$

Where $i(u_i, v_i)$ denotes the coordinates of the i^{th} point, and $\beta_k(u_i, v_i)$ represents the local regression coefficient for the k^{th} variable. An adaptive Gaussian kernel was used to determine the spatial weights, ensuring that neighboring observations have a higher influence on the local model estimation.

2.3. Predictive Framework: XGBoost and ConvLSTM

The forecasting component of the system integrates two advanced algorithms:

XGBoost: A gradient-boosted decision tree ensemble used for modeling complex non-linear relationships between spatial factors and subsidence rates. It is particularly effective for high-dimensional tabular data.

ConvLSTM (Convolutional Long Short-Term Memory): To capture the spatio-temporal dynamics of subsidence, a ConvLSTM architecture is implemented. By replacing the fully connected layers in standard LSTM with convolutional structures, the model can simultaneously extract spatial features from InSAR maps and learn the temporal trends of deformation.

2.4. Risk Zonation and Interpretation

K-means Clustering: The local coefficients (β) derived from GWR are used as input for K-means clustering to segment the study area into distinct risk zones (e.g., high-impact zones, stable zones).

SHAP (SHapley Additive exPlanations): To overcome the "black-box" nature of Deep Learning, the SHAP framework is applied. This method quantifies the marginal contribution of each input feature to the final prediction, providing scientific transparency for urban planning decisions.

3. Data Processing and Feature Engineering

3.1. Satellite-based Deformation Data (The Dependent Variable) The primary input is Persistent Scatterer/Distributed Scatterer (PS/DS) InSAR data derived from the Sentinel-1 constellation. Temporal Coverage: January 2015 to December 2020.

Processing: Interferometric processing was performed using the SARscape/ENVI platform. To minimize atmospheric noise, a spatio-temporal filtering approach was applied.

Vertical Displacement: Line-of-sight (LOS) velocities were converted to vertical displacement values, assuming horizontal movement is negligible in the study area. These points were then interpolated using Kriging to create a continuous Subsidence Velocity Map with a pixel resolution of 50m.

3.2. Explanatory Variables (The Independent Variables) Three key urban factors were quantified and synchronized to the 50m grid:

Building Density (Building): Extracted from 1:2,000 scale topographic maps. We calculated the ratio of footprint area to total cell area. This represents the static load exerted on the alluvial clay layers.

Distance to Water Bodies (DistToWater): A Euclidean distance analysis was performed on the hydrographic network (Saigon River and regional canals). This variable acts as a proxy for soil saturation and drainage capacity.

Land Use Impact (LandUse): Derived from Sentinel-2 LULC maps, reclassified into weighted categories based on their contribution to soil compaction and groundwater withdrawal intensity (e.g., Industrial zones vs. Green spaces).

4. Results

4.1. Global vs. Local Model Performance

The OLS model provided a baseline, but the GWR model significantly improved the R^2 and reduced AICc (Akaike Information Criterion), proving that subsidence drivers in HCMC are not uniform.

4.2. Deep Interpretation of GWR Maps

The local coefficients provide a "spatial signature" of land subsidence in the study area:

Building Density Map (Fig. 2): Areas with strong negative coefficients (deep blue/red zones) represent locations where high-rise construction directly accelerates consolidation of the soft clay layers. In some stable zones, the coefficient nears zero, indicating that the geological foundation there is more resilient to static loads.

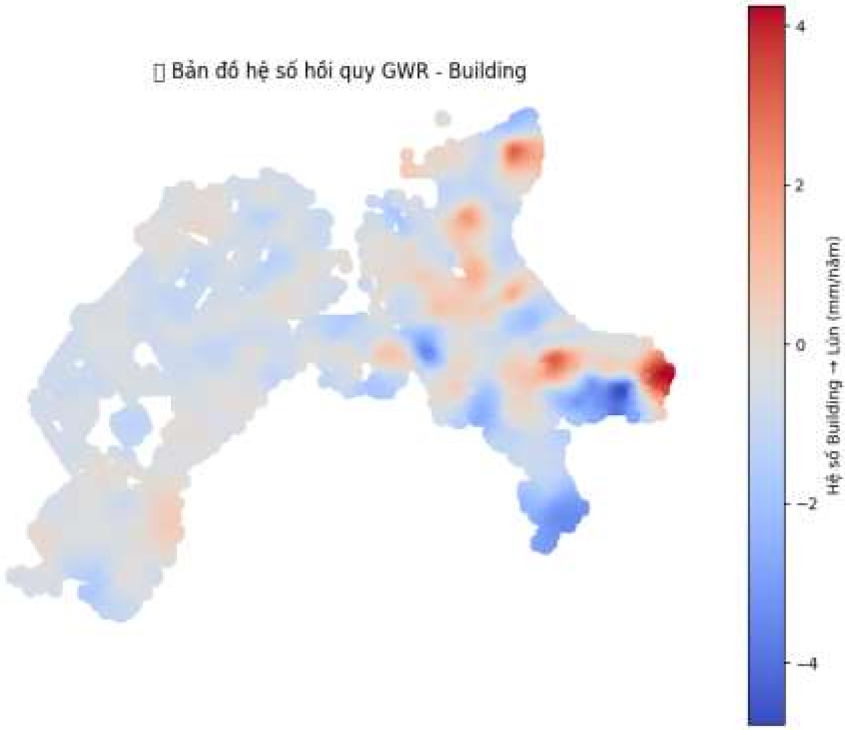

Fig. 2 Map of GWR coefficients by building factor

Distance to river (Fig. 3): There is a clear "corridor effect". Zones adjacent to the Saigon River and major canals show the highest sensitivity (Avg. Coef: -0.344). This is attributed to the high water content and low shear strength of alluvial deposits near water bodies.

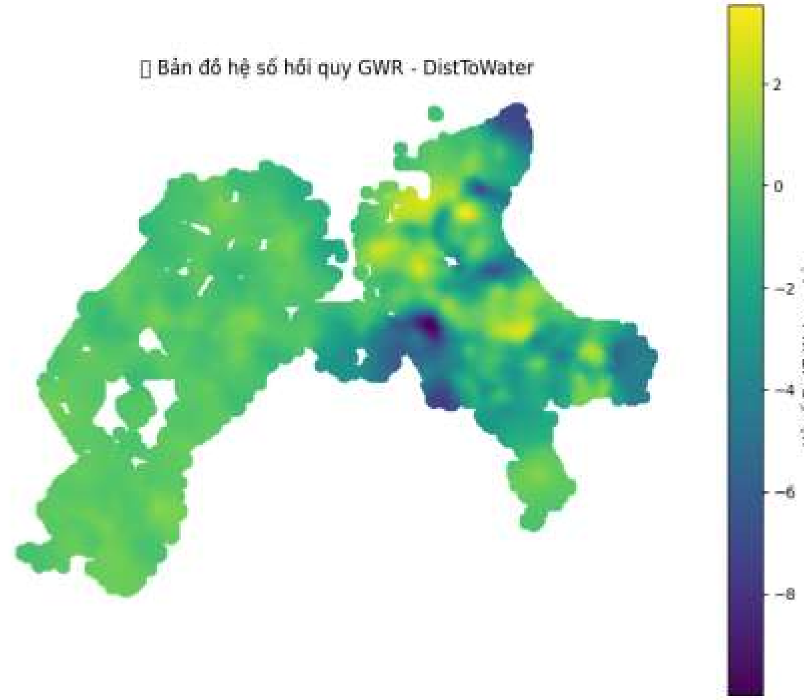

Fig. 3 Map of GWR coefficients by river factor

Land Use Land Cover Change (LULCC) Impact (Fig. 4): LandUse emerged as the most dominant factor (Avg. Coef: -0.390). "Hotspots" (identified by negative coefficients) coincide with areas of rapid land conversion—from agriculture to industrial or impervious urban surfaces.

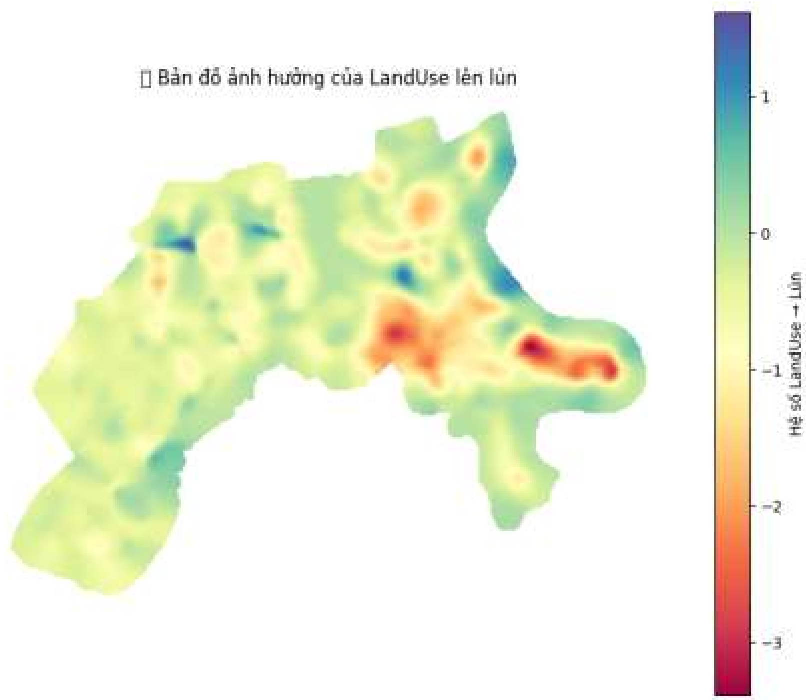

Fig. 4 Map of GWR coefficients by LULCC factor

4.3. Risk Zonation via K-means

By clustering the local GWR coefficients, we identified three distinct clusters:High-Risk Cluster:

High sensitivity to both Building and DistToWater (mainly near riverbanks with dense construction).

Moderate-Risk Cluster: Primarily driven by LandUse changes in transition zones.

Stable Cluster: Minimal correlation between urban factors and subsidence, representing the geological "hard" zones.

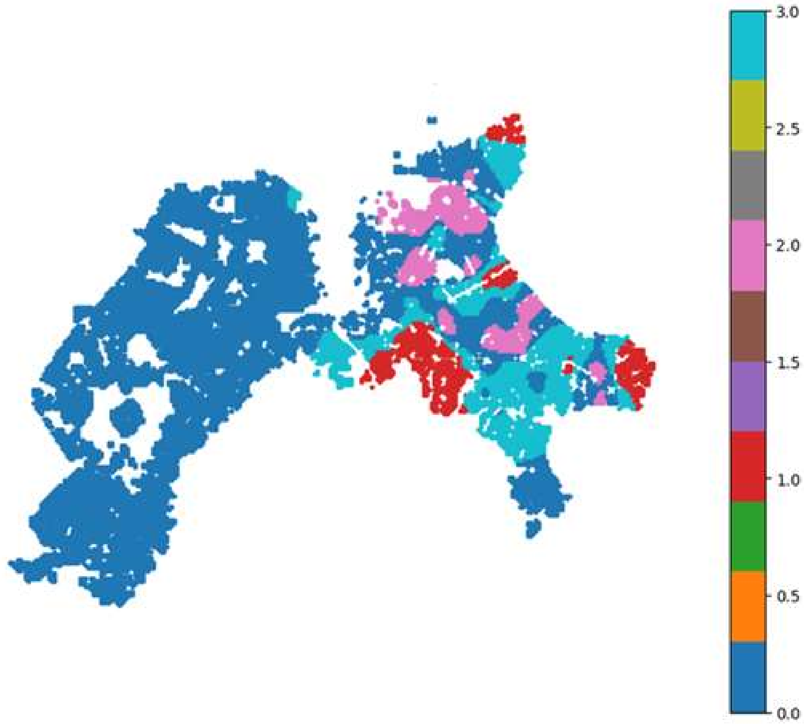

Fig. 5 K-means clustering of regions based on GWR regression coefficients**.**
4.4. 4.4 Forecasting and Model Interpretation (SHAP)

The integrated XGBoost-ConvLSTM model achieved high predictive accuracy for the 2021-2025 period. To ensure model transparency, SHAP values were calculated to quantify feature importance.

5. Conclusion

The integration of GWR and Machine Learning offers a powerful diagnostic tool for urban managers. While LandUse and DistToWater are currently the primary drivers (Avg. Coef -0.390 and -0.344 respectively), the model's reliability can be further enhanced. Future research will integrate ConvLSTM architectures to better capture the spatiotemporal evolution of these "hotspots" as more InSAR data becomes available.

Acknowledgments

The authors would like to express their sincere gratitude for the support of the Ho Chi Minh City Department of Science and Technology under Decision No. 197/QD-SKHCN throughout the research and publication of this article.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Cheryl Tay, Eric O. Lindsey, Shi Tong Chin, Jamie W. McCaughey, David Bekaert, Michele Nguyen, Hook Hua, Gerald Manipon, Mohammed Karim, Benjamin P. Horton, Tanghua Li & Emma M. Hill. Sea-level rise from land subsidence in major coastal cities: Nature Sustainability volume 5, pages 1049–1057 (2022). DOI:10.1038/s41893-022-00947-z

[2] D. Ho Tong Minh, Le Van Trung and T. Le Toan. Mapping ground subsidence phenomena in Ho Chi Minh City through the radar interferometry technique using ALOS PALSAR data. Remote Sensing, vol.7, pp.8543-8562, Jul. 2015;

[3] Dinh HO TONG MINH, Yen-Nhi NGO, Thu Trang LE, Trung Chon LE, Hong Son BUI, Quoc Viet VUONG, Thuy LE TOAN. . Quantifying Horizontal and Vertical Movements in Ho Chi Minh City by Sentinel-1 Radar Interferometry. Preprints 2020, 2020120382. Website: www.preprints.org (Posted: 10 September 2021);

[4] Le Trung Chon, Ho Tong Minh Dinh, Huynh Quyen, Bui Hong Son. “Monitoring land subsidence using InSAR technique. Ho Chi Minh City and Mekong Delta case studies”. NASA -LUCLC Int’l Meeting. Hanoi Feb, 2024

[5] Marie-Noëlle Woillez , Sepehr Eslami Arab, Philip S. J. Minderhoud, Hong-Quan Nguyen. Chapter 9, A question of water; Climate Change in Vietnam, Impacts and Adaptation; A COP26 assessment report of the GEMME Viet Nam Project

[6] Chaofan Zhou, Huili Gong, Beibei Chen, Xiaojuan Li, Jiwei Li, Xu Wang, Mingliang Gao, Yuan Si, Lin Guo, Min Shi, Guangyao Duan. Quantifying the contribution of multiple factors to land subsidence in the Beijing Plain, China with machine learning technology. Geomorphology, Volume 335, 15 June 2019, Pages 48-61. https://doi.org/10.1016/j.geomorph.2019.03.017

[7] Shi, L., Gong, H., Chen, B., & Zhou, C. (2020). Land Subsidence Prediction Induced by Multiple Factors Using Machine Learning Method. Remote Sensing, 12(24), 4044. https://doi.org/10.3390/rs12244044

[8] Liu, CY., Ku, CY. & Ni, C. Deep learning time-series modeling for assessing land subsidence under reduced groundwater use. Sci Rep 15, 30901 (2025). https://doi.org/10.1038/s41598-025-16454-y

[9] Jianxin Liu, Wenxiang Liu, Fabrice Blanchard Allechy, Zhiwen Zheng, Rong Liu, Kouao Laurent Kouadio, Machine learning-based techniques for land subsidence simulation in an urban area, Journal of Environmental Management,Volume 352, 2024, 120078, ISSN 0301-4797. https://doi.org/10.1016/j.jenvman.2024.12007

[10] Quyết định số 5086/QĐ-UBND ngày 28 tháng 11 năm 2019 của Ủy ban nhân dân Thành phố về Quy chế tích hợp, quản lý, vận hành, khai thác Kho dữ liệu dùng chung của thành phố;

[11] Các website: https://scihub.copernicus.eu (Ảnh Sentinel-1 do Cơ quan Vũ trụ Châu Âu cung cấp); http://www.w3c.org (Các mô hình và chuẩn mở do W3C ban hành); http://www.opengis.org (Các mô hình và chuẩn mở do Open Geospatial Consortium đề xuất).