SPATIAL-MORPHOLOGICAL MODELING

FOR MULTI-ATTRIBUTE IMPUTATION OF URBAN BLOCKS

Vasilii Starikov, Ruslan Kozlyak, Georgii Kontsevik, Sergey Mityagin. ITMO University, St. Petersburg, Russia. mityagin@itmo.ru

Abstract

Accurate reconstruction of missing morphological indicators of a city is crucial for urban planning and data-driven analysis. This study presents the spatial-morphological (SM) imputer tool, which combines data-driven morphological clustering with neighborhood-based methods to reconstruct missing values of the floor space index (FSI) and ground space index (GSI) at the city block level, inspired by the SpaceMatrix framework. This approach integrates global morphological priors with local spatial information for context-dependent interpolation. Hybrid SM models combined with inverse distance weighting (IDW) and spatial k-nearest neighbors (sKNN) outperform state-of-the-art baselines.

Problem Formulation: $X_i = \{x^{(r)}, x^{(rec)}, x^{(bus)}, x^{(ind)}, x^{(tr)}, x^{(sp)}, x^{(agr)}, s_i\},\ Y_i = \{FSI_i, GSI_i\}.$

Related Work

Existing approaches to urban data imputation include image-based morphology reconstruction, graph neural networks for spatial dependency modeling, and matrix-factorization-based multi-view imputation methods. Morphological clustering frameworks such as SpaceMatrix provide interpretable global priors for built-form analysis.

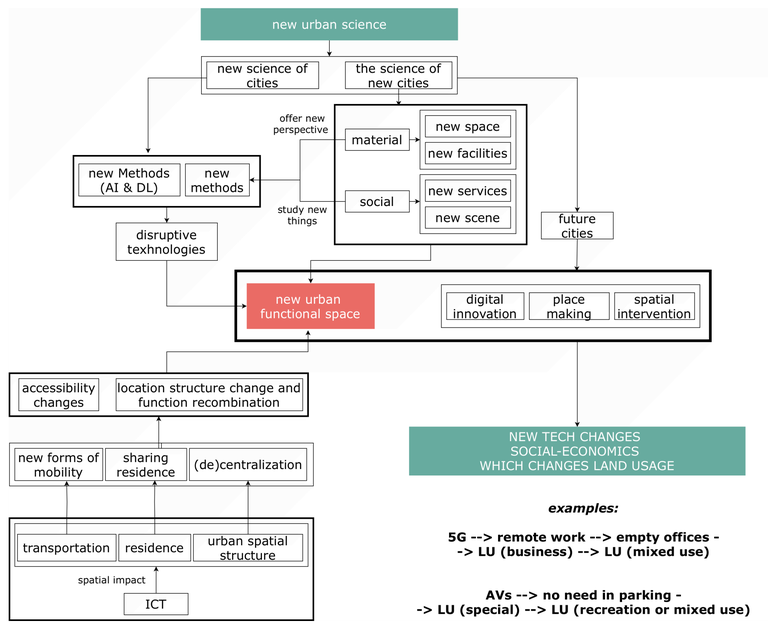
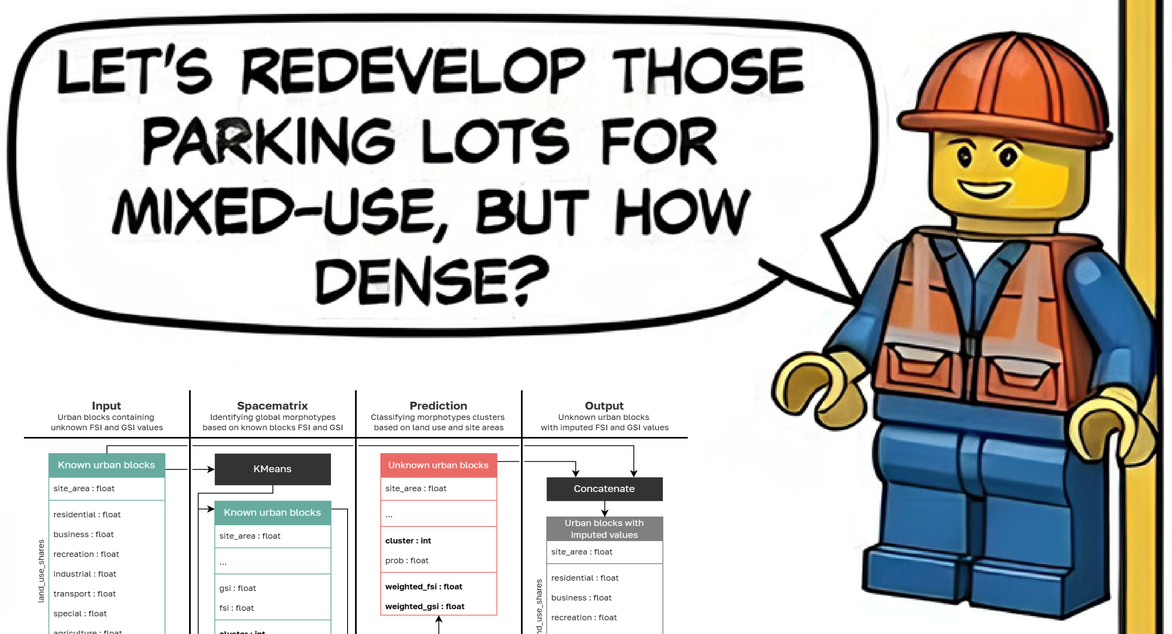


Figure 1: Place of the developed method in New Urban Science.

Methodology: SpaceMatrix and Imputation Model

SpaceMatrix classifies urban blocks into morphological types based on built-form indices such as FSI and GSI using unsupervised clustering. Each cluster represents a characteristic urban form, and its centroid defines typical built-form intensity values used for imputation. The proposed workflow consists of block-level input data, morphological clustering, CatBoost-based prediction, and reconstruction of missing built-form attributes.



Model: Catboost classifier

A CatBoost classifier predicts the probability distribution over morphological clusters for each urban block based on land-use composition and site area. Final FSI and GSI values are computed as probability-weighted averages of cluster medians, capturing city-wide morphological regularities while preserving local interpretability. Urban block geometries and topological relationships were generated using the BlocksNet framework.

Figure 2: Workflow of the Spatial-morphological (SM) imputation model.

Results

We evaluate the SM-Imputer on the Saint Petersburg dataset with synthetic missing values ranging from 10% to 70%. Baseline methods include inverse distance weighting (IDW), spatial k-nearest neighbors (sKNN), and SMV-NMF. Results show that SM+IDW and SM+sKNN achieve the lowest MAE and RMSE, demonstrating the benefit of combining global morphological priors with local spatial interpolation.

Acknowledgments This work supported by the Ministry of Economic Development of the Russian Federation (IGK 000000C313925P4C0002), agreement No139-15-2025-010.