PREPRINT

*Author-formatted, not peer-reviewed document posted on 22/01/2026*

DOI: https://doi.org/10.3897/arphapreprints.e186006

# A Two-Phase Spatio-Temporal Interpolation Framework for Housing Values

**Shuang Tian, Fang Qiu**

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 22/01/2026. DOI: https://doi.org/10.3897/arphapreprints.e186006

# A Two-Phase Spatio-Temporal Interpolation Framework for Reconstructing Block Group Level Housing Values in California

**Shuang Tian**[1]* and **Fang Qiu**[1]

[1] Department of Geospatial Information Sciences,
University of Texas at Dallas,
800 West Campbell Road, MS GR31,
Richardson, TX 75080, USA

* **Corresponding author:** Shuang.Tian@utdallas.edu

Date: 1/20/ 2026

## Abstract

High-resolution historical housing value data are essential for urban analysis, socioeconomic research, and policy evaluation. However, housing values reported by the U.S. Census are only available at decennial intervals and are subject to changes in administrative boundaries, resulting in substantial spatial and temporal gaps. This study proposes a two-phase spatio-temporal interpolation framework to reconstruct annual block-group-level housing values in California by explicitly decoupling spatial and temporal dependence. In the first phase, ordinary kriging is applied independently to each census year (1990, 2000, 2010, and 2020) to reconstruct spatially continuous housing values on a unified 2020 block-group geometry, leveraging strong spatial autocorrelation in housing markets. In the second phase, temporal interpolation and extrapolation are performed independently at each block group using a locally constrained inverse time-distance weighting (ITDW) approach that exploits short-range temporal autocorrelation while avoiding unrealistic global temporal trends. Validation against American Community Survey data demonstrates robust predictive performance, with $R^2$ values ranging from 0.64 to 0.95 for interpolated years (2013–2019) and from 0.68 to 0.84 for the extrapolated years (2021–2023). The results indicate that housing values are more strongly influenced by nearest temporal neighbors than by long-range temporal trends, and that explicitly separating spatial and temporal processes enables stable and accurate reconstruction of fine-scale housing values from sparse historical observations.

## Key Word: Spatio-temporal interpolation; Ordinary kriging; Inverse time-distance weighting; Census block groups; Housing values

# 1. Introduction

Housing values exhibit pronounced spatial and temporal heterogeneity shaped by local market conditions, accessibility, demographic composition, and broader macroeconomic dynamics. Understanding the historical evolution of housing values at fine spatial scales is critical for urban planning, housing affordability assessment, and research on socioeconomic inequality [1–3]. Despite its importance, high-resolution historical housing data remain limited in the United States because census-based housing values are reported only at decennial intervals and are further complicated by frequent changes in census geographies, leading to substantial spatial and temporal gaps [4,5].

To address these limitations, previous studies have explored a range of spatial and spatio-temporal modeling approaches, including areal interpolation, spatial regression, space–time distance-weighted interpolation, and global trend fitting [6–10]. While these methods have proven useful in certain contexts, many implicitly assume smooth global temporal trends or continuous space–time dependence. Such assumptions are often violated in housing markets, where abrupt regime shifts—such as economic recessions, housing booms, and policy interventions—can produce non-linear and locally heterogeneous temporal dynamics [11,12]. Moreover, empirical evidence consistently shows that housing values exhibit strong spatial autocorrelation at fixed time points and strong short-range temporal autocorrelation at fixed locations, suggesting that spatial and temporal processes may operate at distinct scales rather than as a single coupled process [6,13].

Motivated by these observations, this study proposes a two-phase spatio-temporal interpolation framework that explicitly separates spatial and temporal dependency modeling. In the first phase, spatial structure is reconstructed independently at each census year using ordinary kriging to fully capture contemporaneous spatial autocorrelation. In the second phase, temporal interpolation and extrapolation are performed independently at fixed block-group locations using a locally constrained inverse time-distance weighting (ITDW) strategy that prioritizes nearest temporal neighbors. By design, this framework avoids unrealistic global temporal trends and reduces spatio-temporal leakage that can arise in fully coupled space–time interpolation methods.

Using California block-group housing values as a case study, we demonstrate that the proposed framework yields substantially improved reconstruction accuracy for both interpolated and extrapolated years when compared with traditional spatial residual-based and space–time distance-weighted approaches. The results highlight the importance of explicitly decoupling spatial and temporal processes when reconstructing long-term socioeconomic data from sparse historical observations.

# 2. Study Area and Data

This study focuses on the state of California, which exhibits pronounced spatial heterogeneity in housing markets, ranging from dense coastal metropolitan regions to suburban, exurban, and rural inland areas (***Fig. 1***). California's diverse economic structure, large population, and substantial regional housing disparities make it an ideal case for evaluating spatio-temporal reconstruction methods at fine spatial scales [1,11].

Median housing values at the census block-group level were obtained from U.S. Census sources [4] for four decennial years: 1990, 2000, 2010, and 2020 (***Fig. 2***). These data represent the most spatially detailed and consistently available housing value information over long historical periods in the United States [4,5]. However, differences in census block-group boundaries across decades introduce spatial inconsistencies that prevent direct temporal comparison.

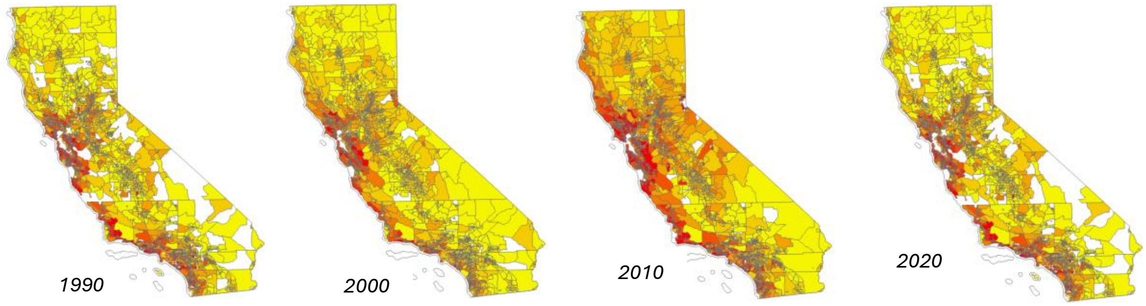


***Fig. 3.*** *Study area and spatial distribution of census block-group median housing values in California for the 1990, 2000, 2010, and 2020 decennial censuses. Blank areas indicate missing data.*

To ensure spatial consistency across time, all historical housing values were reconstructed onto a unified 2020 block-group geometry using areal harmonization techniques commonly adopted in small-area estimation and spatial interpolation studies [6,14]. Block-group centroids were used to represent spatial locations during interpolation, enabling point-based spatial modeling while preserving block-group-level interpretation. This approach eliminates inconsistencies arising from boundary changes and allows direct temporal comparison of reconstructed housing values across the full study period.

# 3. Methodology

## 3.1 Overview of the Two-Phase Framework

The proposed framework consists of two sequential and explicitly decoupled phases designed (**Fig. 2**) to separately model spatial and temporal dependence in census-based housing data:

1. **Spatial Reconstruction (Phase 1):** Ordinary kriging is applied independently to each census year to reconstruct spatially continuous housing values across a unified block-group geometry, explicitly capturing spatial autocorrelation.

2. **Temporal Interpolation (Phase 2):** At each fixed block group, housing values for target years are estimated using a local inverse time-distance weighting (ITDW) approach that exploits short-range temporal autocorrelation.

By separating spatial and temporal modeling, the framework avoids imposing strong assumptions of continuous space–time smoothness and allows each dependency structure to be modeled at its appropriate scale [7,8].

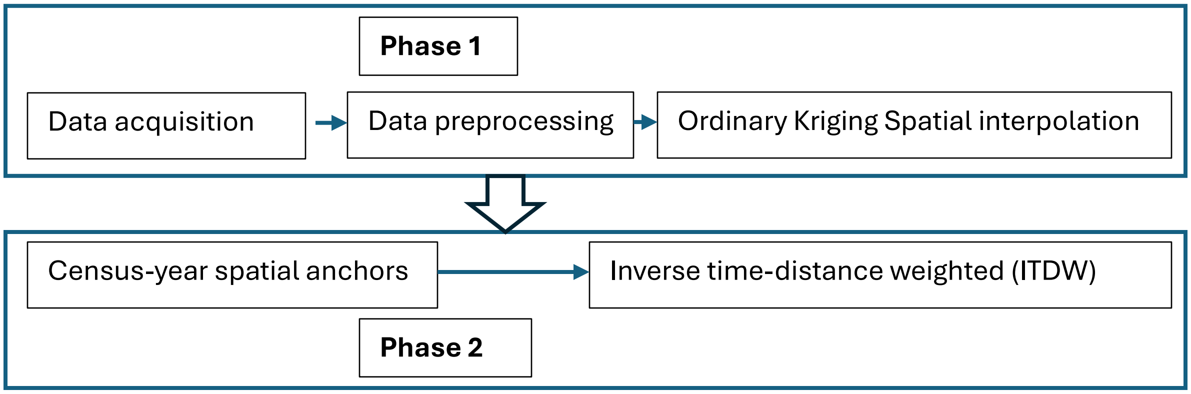


***Fig.2*** *Workflow of the proposed two-phase spatio-temporal interpolation framework, illustrating spatial reconstruction using ordinary kriging (Phase 1) and temporal interpolation using inverse time-distance weighting (Phase 2)*

## 3.2 Phase 1: Spatial Reconstruction Using Ordinary Kriging

In the first phase, spatial reconstruction of housing values is performed independently for each census year (1990, 2000, 2010, and 2020) using ordinary kriging (OK). Ordinary kriging is a geostatistical interpolation method that explicitly models spatial autocorrelation and provides best linear unbiased predictions under second-order stationarity assumptions [7,9]. Its application is particularly appropriate for housing markets, which are well known to exhibit strong spatial dependence consistent with Tobler's First Law of Geography [15].

For each census year, observed block-group median housing values are represented at block-group centroid locations. An empirical variogram is computed to quantify how variance between housing values changes as a function of spatial separation distance. The empirical variogram characterizes the spatial autocorrelation structure by relating semivariance to distance [7]. A theoretical variogram model is then fitted and used within the ordinary kriging system to derive optimal interpolation weights.

The kriging weights are obtained by minimizing the estimation variance subject to an unbiasedness constraint, ensuring that predicted housing values reflect the dominant influence of nearby block groups while accounting for spatial redundancy. Spatial reconstruction is conducted independently for each census year to avoid imposing assumptions about temporal continuity during spatial estimation. This design yields spatially continuous housing value surfaces on a unified 2020 block-group geometry while preserving year-specific spatial patterns.

## 3.3 Phase 2: Temporal Interpolation Using Inverse Time-Distance Weighting

Following spatial reconstruction, each block group possesses reconstructed housing values at four discrete census years. Temporal interpolation and extrapolation are then performed independently at each block group using a local inverse time-distance weighting (ITDW) approach.

Rather than fitting global temporal trends or regression-based curves, ITDW exploits short-range temporal autocorrelation by assigning greater influence to temporally proximate observations [10]. For a target year $t_0$, the estimated housing value $\widehat{H}(t_0)$is computed as:

$$\widehat{H}(t_0) = \frac{\sum_{i=1}^{n} \frac{H(t_i)}{| t_0 - t_i |^p}}{\sum_{i=1}^{n} \frac{1}{| t_0 - t_i |^p}}$$

where $H(t_i)$denotes the reconstructed housing value at census year $t_i$, $p$is a temporal decay parameter controlling the rate at which influence decreases with temporal distance, and $n$denotes the number of available census anchors.

Temporal interpolation is restricted to a local temporal neighborhood to ensure that estimates are dominated by nearest census years, thereby avoiding unrealistic influence from distant historical observations. This local temporal constraint allows abrupt market shifts and regime changes to be accommodated without enforcing smooth global trends. By decoupling spatial reconstruction and temporal interpolation, the framework preserves spatial and temporal autocorrelation structures independently and reduces spatio-temporal leakage commonly observed in fully coupled space–time interpolation methods [8].

## 3.4 Baseline Methods for Comparison

To evaluate the effectiveness of the proposed two-phase framework, its performance is compared against two commonly used baseline approaches:

1. **Spatial Residual-Based Interpolation:** Housing values are estimated by interpolating spatial residuals derived from a single reference year and applying them across time, implicitly assuming temporal stationarity of spatial patterns.
2. **Spatio-Temporal Distance-Weighted Interpolation (ST-IDW):** Spatial and temporal distances are combined into a unified distance metric, assuming continuous space–time smoothness and equal coupling between spatial and temporal dependence [10].

These baselines represent widely adopted approaches for reconstructing missing spatio-temporal socioeconomic data and provide appropriate benchmarks for evaluating reconstruction accuracy.

## 3.5 Spatial Autocorrelation Validation Using Moran’s I

To assess whether reconstructed housing value surfaces preserve realistic spatial dependence, spatial autocorrelation was evaluated using Global Moran’s I, a widely used statistic for diagnosing spatial structure in socioeconomic variables such as housing prices [6,15].

Global Moran’s I is defined as:

$$I = \frac{N}{W}\frac{\sum_i \sum_j w_{ij}\,(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where $N$ is the number of spatial units, $x_i$ and $x_j$ are housing values at locations $i$ and $j$, $\bar{x}$ is the mean housing value, $w_{ij}$ denotes the spatial weight between units $i$ and $j$, and $W = \sum_i \sum_j w_{ij}$.

Moran’s I values range from −1 to +1, with positive values indicating spatial clustering, negative values indicating spatial dispersion, and values near zero indicating spatial randomness. Statistical significance was assessed using permutation-based tests.

Spatial autocorrelation was computed for reconstructed housing value surfaces for **2015** (interpolated year) and **2023** (extrapolated year) using a row-standardized contiguity-based spatial weights matrix at the block-group level. These years were selected to evaluate whether spatial dependence is preserved both within and beyond the temporal range of observed census data. By validating spatial autocorrelation in reconstructed outputs, this analysis ensures that the proposed framework not only minimizes prediction error but also maintains the intrinsic spatial structure characteristic of housing markets.

# 4. Results

## 4.1 Reconstruction Performance Across Census and Non-Census Years

The performance of the proposed two-phase spatio-temporal interpolation framework was evaluated for both interpolated and extrapolated non-census years using observed American Community Survey (ACS) median house values as reference data. Validation was conducted at the census block-group level using three commonly adopted accuracy metrics: coefficient of determination ($R^2$), root mean square error (RMSE), and mean absolute error (MAE). Overall, the framework demonstrates consistently strong predictive performance across California, indicating its ability to effectively capture both spatial variability and temporal dynamics in housing values (***Figs.3-5***).

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 22/01/2026. DOI: https://doi.org/10.3897/arphapreprints.e186006

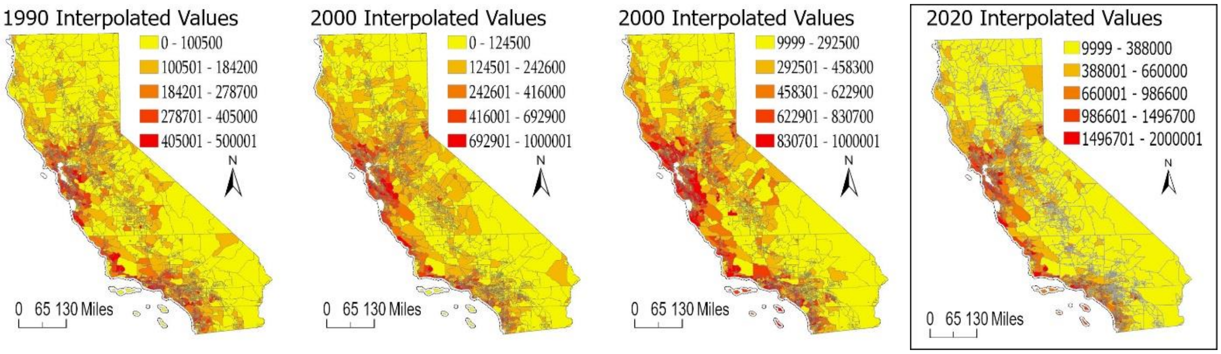


*Fig. 3.* ***Two-phase*** *Spatio-temporal interpolated census block-group housing values in California for census years: 1990, 2000, 2010, and 2020 decennial census years. The results illustrate pronounced spatial heterogeneity and long-term temporal evolution, with consistently higher housing values along the coastal regions. Major metropolitan areas, including San Francisco and Los Angeles, exhibit the highest housing values across all census years.*

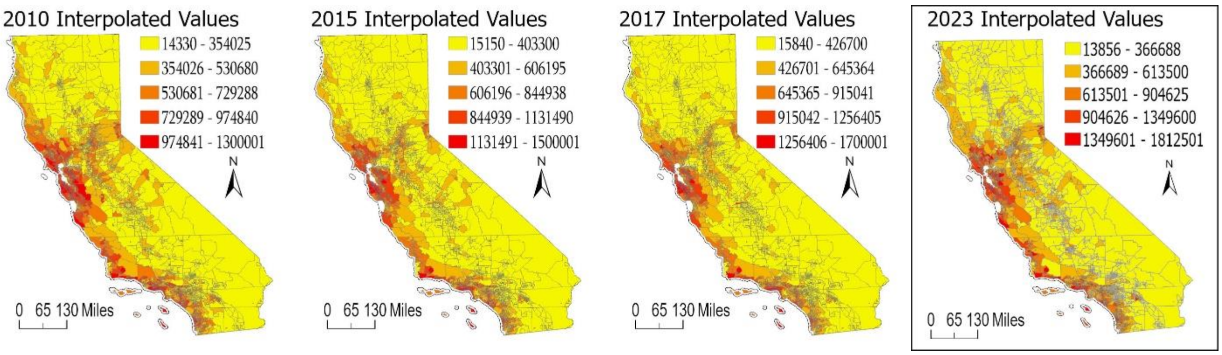


*Fig. 4.* ***Two-phase*** *Spatio-temporal interpolated census block-group housing values in California for non- census years: 2013, 2015,2017, and ,2023 interpolated and extrapolated housing values. The results illustrate pronounced spatial heterogeneity and long-term temporal evolution, with consistently higher housing values along the coastal regions. Major metropolitan areas, including San Francisco and Los Angeles, exhibit the highest housing values across all census years.*

## 4.2 Validation for Interpolated and Near-Term Extrapolated Years

For interpolated years, reconstructed housing values exhibit strong agreement with observed ACS data, confirming the effectiveness of the proposed framework in capturing short-range temporal dynamics. Among the interpolated years, **2017**, located near the midpoint between the 2010 and 2020 census anchors, achieves high validation accuracy with an **$R^2$ of 0.79** across **21,546 block groups** (RMSE = 166,628; MAE = 171,319). Performance improves further for **2019**, which lies closer to the 2020 census anchor, achieving an **$R^2$ of 0.84** across **21,535 block groups** (RMSE = 161,151; MAE = 96,550).

Earlier interpolated years exhibit slightly lower but still robust performance. Reconstructed housing values for **2015** yield an **$R^2$ of 0.704** across **21,593 block groups** (RMSE = 176,866; MAE = 139,015), while **2013**, which is more temporally distant from the nearest census anchor, achieves an **$R^2$ of 0.643** across **22,060 block groups** (RMSE = 147,055; MAE = 126,118). This progressive improvement in accuracy as target years approach census anchors highlights the importance of nearest temporal neighbors in housing value reconstruction.

.

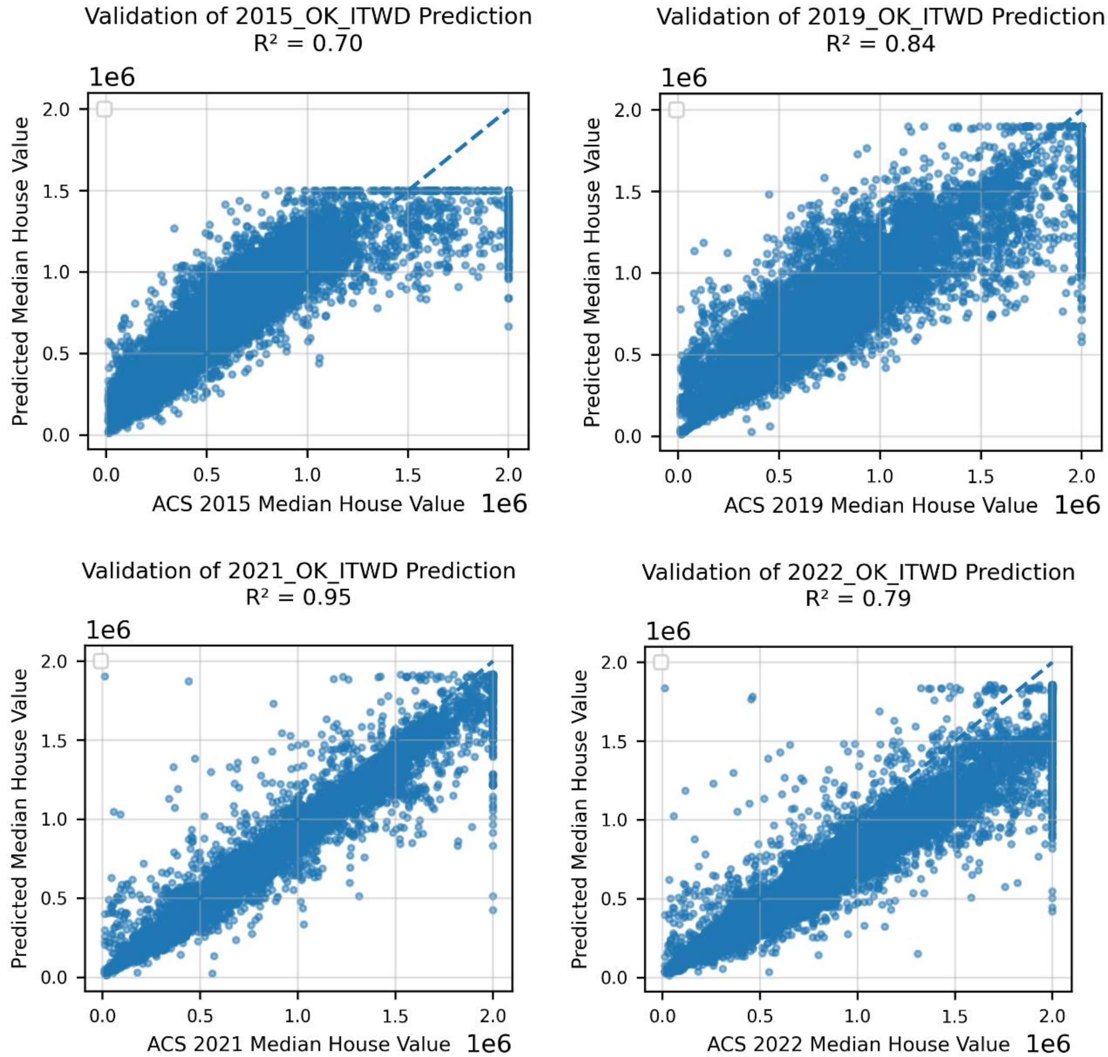


***Fig.5.*** *Scatter plots comparing reconstructed block-group-level median housing values against observed American Community Survey (ACS) data for non-census years 2015, 2019, 2021 (interpolated), and 2022 (extrapolated). Each point represents a census block group. The dashed line denotes the 1:1 reference line. Coefficients of determination ($R^2$) are reported for each year. The results demonstrate strong agreement between reconstructed and observed values across both interpolated and extrapolated years, with slightly increased dispersion for the extrapolated year reflecting greater temporal uncertainty.*

## 4.3 Validation for Extrapolated Year

To assess extrapolation performance, reconstructed housing values for **2021, 2022, and 2023** were compared with corresponding ACS observations. Despite the absence of a census anchor beyond 2020, the proposed framework maintains strong predictive capability for near-term extrapolation. **2021**, which is immediately adjacent to the 2020 census anchor, achieves an exceptionally high **$R^2$ of 0.95** across **23,089 block groups** (RMSE = 97,493; MAE = 59,130), reflecting minimal temporal drift.

As extrapolation distance increases, predictive performance declines in a controlled and expected manner. **2022** achieves an **$R^2$ of 0.79** across **23,096 block groups** (RMSE = 205,000; MAE = 158,421), while **2023** yields an **$R^2$ of 0.68** across **23,047 block groups** (RMSE = 260,638; MAE = 209,427). Despite increased uncertainty, reconstructed values for these years continue to preserve coherent spatial patterns and large-scale temporal trends (Fig. 5), underscoring the robustness of the nearest-time anchoring strategy.

## 4.4 Summary of Validation Results

Across all evaluated non-census years, **$R^2$ values range from 0.64 to 0.95** (***Table 1***), demonstrating stable and reliable performance under both interpolation and extrapolation scenarios. Interpolated years consistently achieve higher accuracy than more distant extrapolated years, while near-term extrapolation immediately following census anchors remains highly reliable. These findings indicate that resolving spatial autocorrelation first using ordinary kriging, followed by temporally constrained inverse time-distance weighting, provides a robust and interpretable strategy for reconstructing long-term block-group-level socioeconomic data from sparse census observations.

*Table 1. Validation metrics for reconstructed block-group housing values*

| Year | Validation Type | Block Groups (N) | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|
| 2013 | Interpolation | 22,060 | 0.643 | 147,055 | 126,118 |
| 2015 | Interpolation | 21,593 | 0.704 | 176,866 | 139,015 |
| 2017 | Interpolation | 21,546 | 0.79 | 166,628 | 171,319 |
| 2019 | Interpolation | 21535 | 0.84 | 161151 | 96550 |
| 2021 | Extrapolation | 23089 | 0.95 | 97493 | 59130 |
| 2022 | Extrapolation | 23096 | 0.79 | 205000 | 158421 |
| 2023 | Extrapolation | 23,047 | 0.680 | 260,638 | 209,427 |

***Note:*** *Validation is performed using observed ACS median house value data at the census block-group level. RMSE and MAE are reported in U.S. dollars.*

## 4.5 Spatial Autocorrelation Assessment

To evaluate whether reconstructed housing values preserve realistic spatial dependence, **Global Moran's I** statistics were computed for both census and non-census years. Moran's I provides a quantitative measure of spatial autocorrelation and is widely used to assess spatial clustering in socioeconomic variables.

For non-census years, Moran's I values remain strongly positive and statistically significant (**$p$ = 0.001**) for all evaluated years (***Table 2***). Moran's I values derived from ITDW-based reconstructions are consistently higher than those computed from temporally interpolated values alone, reflecting the re-imposition and preservation of spatial dependence through the spatial reconstruction phase. Spatial autocorrelation peaks for interpolated years between census anchors (2013–2019)

and declines modestly for extrapolated years (2021–2023), indicating increasing temporal uncertainty while maintaining coherent spatial structure.
For census years, Moran's I values computed from observed and reconstructed housing values are identical and statistically significant (***Table 3***), confirming that the spatial reconstruction phase preserves intrinsic spatial autocorrelation without introducing smoothing artifacts. Together, these results demonstrate that the proposed two-phase framework achieves not only strong predictive accuracy but also faithful preservation of the spatial organization characteristic of housing markets.

***Table** 2. Global Moran's I statistics for reconstructed housing values in non-census years*

| Year | Moran's I (Observed) | p-value | Moran's I (ITWD) | p-value |
|---|---|---|---|---|
| 2013 | 0.7956 | 0.001 | 0.8557 | 0.001 |
| 2015 | 0.7768 | 0.001 | 0.8968 | 0.001 |
| 2017 | 0.7921 | 0.001 | 0.9134 | 0.001 |
| 2019 | 0.8119 | 0.001 | 0.9144 | 0.001 |
| 2021 | 0.8025 | 0.001 | 0.8171 | 0.001 |
| 2022 | 0.8018 | 0.001 | 0.82420 | 0.001 |
| 2023 | 0.7981 | 0.001 | 0.8304 | 0.001 |

***Note:** Moran's I statistics were computed using a row-standardized contiguity-based spatial weights matrix at the census block-group level. "Observed" values correspond to housing values obtained through temporal interpolation or extrapolation using inverse time-distance weighting (ITDW) prior to spatial reconstruction, while "ITDW" values represent outputs after applying the ITDW-based temporal interpolation or extrapolation. Statistical significance was assessed using permutation-based tests (p = 0.001).*

***Table** 3. Global Moran's I statistics for reconstructed housing values in census years*

| Year | Moran's I (Observed) | p-value | Moran's I (Reconstructed) | p-value |
|---|---|---|---|---|
| 1990 | 0.6737 | 0.001 | 0.6737 | 0.001 |
| 2000 | 0.7374 | 0.001 | 0.7374 | 0.001 |
| 2010 | 0.7303 | 0.001 | 0.7303 | 0.001 |
| 2020 | 0.8006 | 0.001 | 0.8006 | 0.001 |

***Note:** Moran's I statistics were computed using a row-standardized contiguity-based spatial weights matrix at the census block-group level. Identical Moran's I values for observed and reconstructed data confirm that the spatial reconstruction phase preserves intrinsic spatial autocorrelation without introducing smoothing artifacts. Statistical significance was assessed using permutation-based tests (p = 0.001).*

# 5. Discussion

The results underscore the importance of aligning interpolation strategies with the intrinsic spatio-temporal dynamics of housing markets. Housing values exhibit strong spatial autocorrelation at fixed time points and pronounced short-range temporal autocorrelation at fixed locations. Methods that impose linear, smooth, or globally continuous temporal trends across extended periods are therefore poorly suited to housing data, as they fail to capture abrupt market shifts, cyclical behavior, and regime changes driven by macroeconomic conditions and policy interventions.

By explicitly decoupling spatial and temporal processes, the proposed two-phase framework mitigates spatio-temporal leakage and enhances reconstruction robustness. Spatial dependence is first resolved independently within each census year using ordinary kriging, ensuring realistic spatial structure is preserved. Temporal interpolation and extrapolation are then constrained to nearest temporal neighbors at fixed locations using inverse time-distance weighting. The strong predictive performance achieved using only four decennial observations demonstrates that local temporal weighting can effectively recover annual housing dynamics when spatial structure is properly reconstructed.

Beyond empirical accuracy, the proposed framework offers a transparent and interpretable baseline for more advanced spatio-temporal learning approaches. By explicitly separating spatial and temporal effects, the framework facilitates clearer diagnosis of model behavior and uncertainty, and provides a principled benchmark for evaluating graph-based and deep learning models. This separation is particularly valuable for hybrid modeling strategies that seek to integrate geostatistical methods with data-driven learning while maintaining interpretability.

# 6. Implications

The proposed two-phase spatio-temporal interpolation framework has several important implications for geospatial analysis, urban studies, and the reconstruction of long-term socioeconomic data.

**First**, from a methodological perspective, the results demonstrate that explicitly aligning interpolation strategies with the intrinsic dynamics of the target phenomenon is critical for reliable spatio-temporal reconstruction. Housing values exhibit strong spatial dependence at fixed time points and dominant short-range temporal dependence at fixed locations. The strong performance of the proposed framework highlights that separating spatial and temporal modeling—rather than enforcing fully coupled space–time smoothness—can substantially improve both accuracy and robustness when observations are temporally sparse.

**Second**, the framework provides a practical and scalable solution for reconstructing fine-scale historical housing data in contexts where annual observations are unavailable or inconsistent. By relying solely on decennial census data and a transparent interpolation strategy, the approach enables the generation of annual block-group-level housing estimates suitable for longitudinal urban analysis, housing affordability assessment, and studies of neighborhood change. The use of a unified spatial geometry further ensures temporal comparability, addressing a long-standing challenge in census-based socioeconomic research.

**Third**, the interpretability of the two-phase design has important implications for model transparency and reproducibility. Unlike black-box spatio-temporal learning approaches, the framework explicitly separates spatial and temporal effects, allowing researchers to diagnose sources of uncertainty, evaluate spatial realism using established metrics, and understand how temporal proximity influences reconstructed values. This transparency is particularly valuable for policy-oriented urban research and comparative regional studies, where methodological clarity and interpretability are essential.

**Fourth**, the framework demonstrates strong potential for near-term forecasting of housing values when recent census anchors are available. Anchored by the 2020 census, the proposed approach achieves high predictive accuracy for subsequent years, including 2021, 2022, and 2023. This result suggests that locally constrained temporal weighting, when combined with robust spatial reconstruction, can effectively support short-term housing market analysis without requiring complex forecasting models or extensive auxiliary data.

**Fifth**, the framework offers a powerful solution for spatial interpolation under conditions of severe data sparsity. Across multiple non-census years, a substantial number of block groups lack observed housing values (e.g., over 1,000–2,500 missing block groups per year between 2013 and 2023). The proposed framework successfully reconstructs housing values for these missing spatial units while preserving spatial autocorrelation and minimizing interpolation artifacts. This capability highlights the framework's utility for recovering massive spatial gaps in fine-scale socioeconomic datasets.

**Finally**, the proposed framework establishes a conceptual and computational baseline for future spatio-temporal modeling efforts. By resolving spatial structure prior to temporal interpolation, the approach can serve as a benchmark or initialization strategy for more advanced methods, including graph-based and hybrid spatio-temporal learning models. As such, the implications of this work extend beyond housing data reconstruction, offering a generalizable strategy for recovering fine-scale socioeconomic variables from sparse historical observations across diverse geographic and thematic contexts.

# 7. Limitations and Future Work

Despite its strong performance and interpretability, the proposed two-phase framework has several limitations that suggest directions for future research.

**First**, the temporal interpolation component relies on a local inverse time-distance weighting (ITDW) scheme with a fixed temporal decay parameter. While this design effectively captures short-range temporal autocorrelation and avoids imposing unrealistic global temporal trends, it does not explicitly account for heterogeneous temporal dynamics across different housing submarkets. As illustrated in ***Fig. 5***, predicted values for 2015 exhibit a saturation effect at the upper end of the value range, where extremely high housing prices are systematically underestimated and form an apparent ceiling relative to observed ACS values. This behavior reflects the averaging nature of fixed-parameter temporal weighting, which dampens extreme values when nearby temporal anchors do not exhibit comparable magnitudes. Future work could explore adaptive temporal weighting strategies, locally varying decay parameters, or data-driven temporal kernels to better accommodate region-specific housing dynamics, market segmentation, and nonlinear growth regimes.

.

**Second**, the framework assumes that spatial autocorrelation within each census year can be adequately characterized using block-group centroids and stationary variogram models. Although this assumption is common in geostatistical applications, housing markets may exhibit non-stationary spatial dependence due to urban–rural gradients, localized development patterns, or policy-driven effects. Extending the framework to incorporate non-stationary kriging approaches or multi-scale spatial modeling may further enhance spatial reconstruction accuracy.

**Third**, the current implementation focuses on univariate housing values and does not explicitly incorporate auxiliary socioeconomic or environmental covariates, such as median income, population density, or accessibility measures. Integrating such covariates through co-kriging or hybrid spatial–statistical models represents a promising extension that could improve reconstruction performance, particularly in regions with sparse or noisy observations.

**Finally**, while the proposed framework is intended as a transparent and interpretable baseline, it does not directly model complex nonlinear spatio-temporal interactions. Future research could integrate this two-phase design with advanced spatio-temporal learning approaches, such as graph neural networks, by using the reconstructed spatial and temporal structures as informed priors or initialization strategies. Such hybrid models may combine the interpretability and stability of geostatistical methods with the representational flexibility of deep learning, while maintaining robustness under sparse temporal observations.

# 8. Conclusion

This study proposes a two-phase spatio-temporal interpolation framework for reconstructing block-group-level housing values from sparse historical observations. By explicitly separating spatial reconstruction and temporal interpolation, the framework leverages spatial and temporal autocorrelation more effectively than conventional approaches that impose globally smooth space–time dependence. Spatial structure is first reconstructed independently for each census year using ordinary kriging, followed by locally constrained temporal interpolation at fixed block-group locations, resulting in stable and accurate annual housing value estimates.

Empirical validation using California housing data demonstrates that the proposed framework achieves strong predictive performance for both interpolated and extrapolated years while preserving realistic spatial dependence patterns. The results confirm that housing values are more strongly influenced by nearest temporal neighbors than by long-range temporal trends, and that decoupling spatial and temporal processes reduces spatio-temporal leakage and improves reconstruction robustness under temporally sparse observations.

Beyond empirical accuracy, the proposed framework offers a scalable and interpretable solution for reconstructing fine-scale historical socioeconomic data from census sources. It also establishes a transparent methodological baseline that can support subsequent integration with advanced spatio-temporal learning models, such as graph-based approaches, enabling systematic comparison and hybrid modeling in future research.

# 9. Data Availability Statement

The census-based socioeconomic data used in this study are publicly available from the IPUMS National Historical Geographic Information System (NHGIS) [4]. Derived datasets, graph structures, and model inputs generated during the study are available from the corresponding author upon reasonable request.

# 10. Ethics Statement

The authors confirm that they have read and complied with the ethical requirements for publication in the *Swiss Journal of Geosciences*. This study does not involve human participants, animal subjects, or data collected from social media platforms.

Generative artificial intelligence tools, specifically **ChatGPT (OpenAI)**, were used during manuscript preparation for language editing, text refinement, and clarity enhancement. All content generated with the assistance of this tool was carefully reviewed, verified, and edited by the authors, who take full responsibility for the accuracy, integrity, and originality of the final manuscript.

# 11. Author Contributions

**Shuang Tian** was responsible for data curation, data analysis, methodology, resources, visualization and writing original drafts.

**Dr. Fang Qiu** was responsible for conceptualization, investigation, project administration, supervision, validation, review and editing.

# 12. Competing Interests

The author declares no competing interests.

# 13. Funding Statement

This research did not receive funding.

# References


1. Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.
2. Glaeser, E. L., Gyourko, J., & Saks, R. (2005). Why is Manhattan so expensive? Regulation and the rise in housing prices. *Journal of Law and Economics*, 48(2), 331–369.
3. Bailey, N., van Gent, W., & Musterd, S. (2017). Remaking urban segregation. *Urban Studies*, 54(14), 3284–3302.
4. Jonathan Schroeder, David Van Riper, Steven Manson, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 20.0 [dataset]. Minneapolis, MN: IPUMS. 2025. **http://doi.org/10.18128/D050.V20.0**.
5. Logan, J. R., Xu, Z., & Stults, B. (2014). Interpolating U.S. decennial census tract data from as far back as 1970. *Geographical Analysis*, 46(4), 412–434.
6. Qiu, F., & Cromley, R. (2013). Introduction: Areal interpolation and dasymetric modeling. *Geographical Analysis*, 45(3), 213–215. https://doi.org/10.2747/1548-1603.49.5.644
7. Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical space–time models. *Mathematical Geology*, 31(6), 651–684.
8. Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression. *International Journal of Geographical Information Science*, 24(3), 383–401. https://doi.org/10.1080/13658810802672469
9. Li, S., Ye, X., Lee, J. *et al*. Spatiotemporal Analysis of Housing Prices in China: A Big Data Perspective. *Appl. Spatial Analysis* **10**, 421–433 (2017). https://doi.org/10.1007/s12061-016-9185-3
10. Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest. *Transactions in GIS*, 21(3), 446–467. https://doi.org/10.1111/tgis.12289
11. Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, 2, 299–362.
12. Gyourko, J., Mayer, C., & Sinai, T. (2013). Superstar cities. *American Economic Journal: Economic Policy*, 5(4), 167–199.
13. Pace, R. K., & Barry, R. (1997). Sparse spatial autoregressions. *Statistics & Probability Letters*, 33(3), 291–297. https://doi.org/10.1016/S0167-7152(96)00140-X
14. Qiu, F., Zhang, C., & Zhou, Y. (2012). The development of an areal interpolation ArcGIS extension and a comparative study. *GIScience & Remote Sensing*, 49(5), 644–663. https://doi.org/10.2747/1548-1603.49.5.644
15. Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(Suppl.), 234–240. https://doi.org/10.2307/143141