Full paper

International Journal of Real Estate Studies

INTREST

Residential Price Forecasting in Shaoxing using Gaussian Process Regression with Bayesian Optimisation

Bingzi Jin[1], Xiaojie Xu[2]*

[1]Advanced Micro Devices (China) Co., Ltd., Shanghai, China
[2]North Carolina State University, Raleigh, NC 27695, US

*Corresponding author's email: xxu6@alumni.ncsu.edu; Xiaojie-Xu@hotmail.com

Article history: Received: *25 August 2025* Received in revised form: *1 June 2026*
Accepted: *2 June 2026* Published online: *30 June 2026*
DOI: https://doi.org/10.11113/intrest.v20n1.436

Abstract

The Chinese residential property market has undergone pronounced cyclical shifts over the last decade, moving from sustained expansion to a sharp downturn after late 2021, which poses considerable challenges for investors, planners, and regulators seeking to anticipate urban housing price trends. While machine learning methods are increasingly applied to real estate forecasting, Gaussian process regression (GPR) remains underutilised in China's spatially heterogeneous markets, and no prior study has deployed it to model price dynamics in Shaoxing—a representative, rapidly transitioning urban centre. To address this gap, the present study constructs a GPR forecasting framework with multiple kernel structures and adaptive basis functions, optimized through Bayesian inference and cross-validation, using monthly residential price data from January 2013 to July 2024. The model is trained on the period up to April 2022 and evaluated out-of-sample over May 2022–July 2024. For comparative benchmarking, identical test conditions are applied to a long short-term memory (LSTM) network, support vector regression, a regression tree, and a simple autoregressive model. The GPR model achieves a root mean square error of 30.98, markedly lower than LSTM (48.58), support vector regression (60.98), regression tree (73.34), and the autoregressive benchmark (87.28), corresponding to a relative root mean square error of 0.1774%. These results confirm the framework's ability to capture the nonlinear temporal dependencies that characterise Shaoxing's housing market. Beyond its methodological contribution, the study offers practical value for valuation practice, land administration, and housing policy by providing a transparent, scalable, and data-driven tool for price monitoring. The proposed system can be deployed independently or integrated with conventional econometric models, and its parsimonious input design facilitates replication in other urban contexts, supporting evidence-based decision-making across diverse real estate markets.

Keywords: Residential property price, Price forecasting, Gaussian process regression, Bayesian optimization, Cross validation, Shaoxing

■1.0 INTRODUCTION

After an extended period of sustained expansion over several decades, China's property sector has entered a structural adjustment phase marked by sustained depreciation patterns beginning in late 2021. Disentangling the multifaceted drivers of residential price formation—encompassing cyclical market oscillations and systemic instabilities—retains paramount importance due to its cascading effects on household consumption patterns, investment prioritization, and institutional governance mechanisms. In response, analysts and forecasting institutions now emphasize integrating advanced computational frameworks to enhance the precision of real estate appraisal systems. Amidst evolving macroeconomic pressures and heterogeneous economic indicators, the development of robust forecasting architectures for property value trajectories has become a strategic imperative for policymakers and institutional investors. These systems must account for nonlinear interdependencies between regional demographic shifts, fiscal policy adaptations, and speculative market behaviors, which collectively underscore the urgency of deploying adaptive modeling techniques. Furthermore, the integration of high-dimensional datasets reflecting urban planning initiatives, land-use regulations, and consumer sentiment indices offers novel pathways to mitigate forecasting uncertainties while aligning predictive outputs with contemporary socioeconomic realities.

The precision and dependability of temporal prediction systems have attracted significant attention across economic and financial analytical disciplines. Conventional econometric methodologies, including vector autoregressive (VAR), autoregressive (AR), and vector error correction (VEC) frameworks, have been subjected to rigorous scrutiny and progressive enhancement to improve predictive performance. Simultaneously, computational techniques like deep neural architectures, ensemble learning strategies, and artificial neural systems (ANNs) have demonstrated practical efficacy in resolving heterogeneous property market forecasting complexities. Neural network structures have risen to prominence as primary tools for real estate valuation, consistent with findings that validate algorithmic approaches within econometric applications, despite enduring challenges related to interpretative transparency and adaptability across diverse contexts. However, the application of Gaussian process regression (GPR) to analyze temporal patterns in property valuation trajectories remains relatively underexplored, highlighting a critical lacuna in methodological innovation at the intersection of computational economics and urban market analysis. This gap underscores the necessity for hybridized modeling paradigms capable of synthesizing probabilistic inference with domain-specific economic variables, thereby advancing both theoretical and applied dimensions of real estate forecasting.

Gaussian process regression (GPR) constitutes a probabilistic Bayesian methodology, building upon seminal work byNeal (Neal, 2012), which established theoretical linkages between Gaussian stochastic processes and neural architectures featuring infinite hidden layers via Bayesian functional priors. The GPR framework intrinsically supports hybrid observational paradigms, encompassing stochastic (Williams and Rasmussen, 1995) and deterministic data-generating mechanisms (Neal, 1997), utilizing its probabilistic modeling foundation to rigorously quantify uncertainty in prediction intervals. Empirical validation by Brahim-Belhouari and Vesin (Brahim-Belhouari and Vesin, 2001) established GPR's enhanced performance relative to radial basis neural architectures for stationary time-series modeling, a technical innovation subsequently advanced by Bermak and Brahim-Belhouari (Brahim-Belhouari and Bermak, 2004) through the integration of heterogeneous kernel configurations. The technique demonstrates adaptability in addressing non-stationary temporal systems while maintaining computational efficiency through closed-form matrix operations that systematically unify prior distributions with stochastic disturbance parameters (Brahim-Belhouari and Bermak, 2004). Informed by Brahim-Belhouari and Bermak's multi-model paradigm (Brahim-Belhouari and Bermak, 2004), this investigation employs ensemble-averaged GPR specifications to optimize predictive stability. Contemporary implementations in commodity valuation analyses and agroeconomic forecasting further validate GPR's methodological flexibility, solidifying its relevance for sophisticated econometric inquiries into temporal dependencies.

Research focused on predicting residential real estate prices in Shaoxing City remains limited, especially regarding the application of computational intelligence methodologies. D'Arcy, McGough and Tsolacos (1999) employed classical econometric models to explore commercial property valuation fluctuations, constructing a singular-equation framework to evaluate office rental volatility in Dublin. Their findings highlighted supply-demand disequilibrium, lagged GDP variations, and delayed inventory corrections as primary determinants of lease instability, with the model outperforming competing statistical techniques in predictive accuracy. Tonelli, Cowley and Boyd (2004) innovated a system dynamics model combining stakeholder perspectives with macroeconomic time-series data to forecast office rental trajectories, merging qualitative inputs with quantitative rigor for enhanced forecast stability. Li, Shen and Love (2005) developed a sequential regression framework for estimating initial construction costs of office structures, methodically integrating multivariate predictors to refine estimation reliability. Mohd, Jamil and Masrom (2020b) systematically compared computational algorithms—including Ridge/LASSO regression, random forests, decision trees, and linear regression—for appraising sustainable office buildings, with results identifying random forests as the most accurate predictor. Jamil, Mohd, Masrom and Ab Rahim (2020) determined decision-tree architectures as optimal for green construction cost estimation via comparative benchmarking of predictive models. Kim, Shin, Kim and Shin (2013) assessed regression, neural networks, and support vector machines for construction expenditure prediction, validating artificial neural networks (ANNs) as superior in precision. Tatari and Kucukvar (2011) engineered an ANN based system to quantify cost premiums in LEED-certified buildings, using sustainability metrics as inputs, with site selection and energy efficiency ratings identified as statistically significant cost determinants. Silver and Goode (1990) applied ordinary least squares (OLS) regression to formulate seemingly unrelated regression (SUR) equations, mitigating multicollinearity in UK retail rent estimation and proposing alternatives to classical econometric approaches. McGough and Tsolacos (1995) utilized autoregressive (AR), moving average (MA), and ARIMA models to project short-term UK commercial leasing costs, confirming time-series methods' utility in property analytics. Eppli, Shilling and Vandell (1998) analyzed metropolitan retail property performance using raw datasets, identifying macroeconomic announcement frequency as a critical driver of retail real estate volatility. Quigley (1999) executed a comprehensive investigation into historical variables, macroeconomic fundamentals, and housing price correlations, seeking to elucidate connections between economic conditions and valuation frameworks. Brooks and Tsolacos (2000) implemented a vector autoregressive (VAR) model to predict UK retail rents, identifying temporal forecasting windows as critical for precision. Brooks, Tsolacos and Lee (2000) extended this framework to explore interactions among macroeconomic, financial, and property variables, showing that macroeconomic and financial factors explain cyclical UK property value shifts, offering insights into historical returns and future asset volatility. Jackson (2001) applied multivariate linear regression to analyze premium retail rent determinants, revealing shopping center size and demographic attributes as significant predictors of leasing cost disparities across UK urban areas. Kim (2004) conducted a longitudinal analysis of excess returns in Singaporean retail and office markets, proposing macroeconomic shifts as predictors of returns, with results underscoring economic instability's role in commercial real estate forecasting. West and Worthington (2006) adopted a GARCH-M framework to assess macroeconomic impacts on Australian office, retail, and industrial properties, demonstrating enhanced predictive accuracy for office returns and heterogeneous sectoral responses to economic perturbations. Ghysels, Plazzi, Valkanov and Torous (2013) validated real estate return forecasting via integrated macroeconomic and localized indicators, emphasizing multi-scale analysis. Their approach was adapted to U.S. markets, differentiating residential and commercial segments to reveal divergent predictive trends. Panagiotidis and Printzis (2016) deployed a VECM to examine macroeconomic linkages with Greek housing indices, identifying mortgage credit availability and retail sales as significant price drivers, reinforcing credit access and consumption trends' centrality in housing dynamics. Seo, Salon, Kuby and Golub (2019) utilized spatial error regression to evaluate transportation infrastructure externalities on Phoenix commercial values, identifying road connectivity and light rail proximity as key determinants of intra-urban price differentials. Al-Shayea Al-Shayea (2012) championed cascade-forward backpropagation ANNs for retail property appraisal, emphasizing their ability to model non-linear urban market interactions. Yan, Wei, Hui, Yang, ZHANG, Hong and WANG (2007) proposed a hybrid regression-grey prediction system for seasonal commercial price forecasting, synthesizing outputs via wavelet neural networks (WNNs) to optimize error correction. Results indicated China's 2005 macroeconomic policies moderated real estate investment and residential inflation, illustrating regulatory-market interdependencies. Valier (2020) demonstrated machine learning driven AVMs surpass traditional regression in price prediction, underscoring computational intelligence's potential to enhance valuation consistency. Abidoye, Chan, Abidoye and Oshodi (2019) compared ANNs, ARIMA, and SVMs for Hong Kong's property index forecasting, establishing ANN dominance in resolving non-linear market relationships. Crosby, Davis, Damoulas and Jarvis (2016) empirically affirmed GPR's superiority over regression-kriging, random forests, and M5P decision trees in valuation, highlighting its ability to model spatiotemporal property data patterns. Mohd, Jamil, Johari, Abdullah and Masrom (2020a) systematically reviewed quantitative modeling innovations in real estate appraisal, emphasizing hybrid and machine learning advancements while evaluating scalability and accuracy limitations.

Gençay and Yang (1996) conducted a comprehensive evaluation of parametric and semi-parametric approaches rooted in econometric theory to predict residential real estate prices. Their analysis demonstrated that semi-parametric techniques exhibit enhanced adaptability in modeling non-linear market dynamics, yielding superior forecast precision and valuation consistency across housing systems. Glennon, Kiefer and Mayock (2018) explored hybrid methodological strategies for developing residential price indices, revealing that composite methodological architectures surpass conventional single-technique approaches in identifying nuanced market fluctuations and enhancing measurement reliability. Clapp and Giaccotto (1992) analyzed the efficacy of valuation techniques, advocating for the superiority of assessed-value methodologies over repeat-sales frameworks. They further introduced systematic calibration mechanisms to mitigate inherent biases in valuation processes, thereby improving the operational utility of real estate indices. Kaboudan and Sarkar (2007) contend that spatially aggregated pricing systems outperform disaggregated models in urban settings, positing that spatially consolidated methodologies more effectively capture localized market interconnections, bolstering valuation dependability in metropolitan contexts. Mei and Fang (2017) developed a dynamic statespace framework to project average transaction prices by integrating trend decomposition with multivariate regression, synthesizing cyclical temporal oscillations and exogenous variable impacts to examine price volatility mechanisms in unstable property markets. Levesque (1994) employed a geospatial case study to assess airport noise externalities on residential prices, quantifying correlations between acoustic exposure and value depreciation while elucidating urban infrastructure externalities' role in housing market dynamics. Hepşen andVatansever (2011) conducted a rigorous evaluation of ARIMA models, applying stringent performance metrics to assess forecasting accuracy across heterogeneous temporal datasets. Their findings emphasize the framework's ability to detect linear temporal dependencies while exposing limitations in non-stationary contexts. Baroni, Barthélémy and Mokrane (2005) applied principal component analysis (PCA) to formulate a repeat-sales index, illustrating its methodological precision in forecasting apartment price trends. Their technique resolves multicollinearity and dimensionality challenges in transactional data, advancing a refined paradigm for trend identification in segmented residential markets. Guo (2020) contrasted linear and non-linear regression frameworks for predicting price stability, evaluating their capacity to simulate intricate market interactions and enhance reliability in volatile economies. The research underscores non-linear methods' dominance, particularly their integration of asymmetrical relationships and threshold phenomena, refining stability prediction accuracy. Paris (2008) deployed artificial neural networks (ANNs) and computational methodologies to forecast fluctuations in UK residential indices across national and regional scales. This work underscores computational methodologies' potential to improve rental price prediction while offering insights into spatiotemporal heterogeneities across geographic zones. Chi (2017) proposed a spatial back-propagation neural network (SBPNN) framework to advance residential price estimation, emphasizing its proficiency in addressing spatial dependencies and non-linear interactions neglected by traditional approaches. The model integrates spatial autocorrelation and spatiotemporal variability, constructing a holistic architecture to resolve locational complexities in valuation. Bee-Hua (2000) introduced a hybrid technique combining evolutionary computational strategies with ANNs to simulate demand in residential construction sectors, merging adaptive optimization and non-linear pattern recognition to transcend classical econometric constraints. This methodology improves demand forecasting by capturing intricate interactions in heterogeneous markets. Štubňová, Urbaníková, Hudáková and Papcunová (2020) demonstrated neural networks' outperformance of regression-based approaches in residential appraisal, attributing this to their capacity to model non-linear correlations and complex variable interdependencies. Their critique of linear regression's inadequacy in capturing spatiotemporal valuation dynamics positions computational intelligence as indispensable for advancing automated valuation systems (AVMs). As evidenced by Seya and Shiroi (2021), incorporating Gaussian nearest-neighbor techniques into deep neural architectures enhances residential rental price prediction accuracy relative to traditional frameworks. This integrative strategy leverages spatial interdependencies and non-linear interactions to better approximate real estate market behaviors. Rafiei and Adeli (2018) introduced a novel architecture combining softmax layers with unsupervised deep Boltzmann machines (DBMs) to identify critical input features, refining construction cost forecasting through economic indicator analysis. The framework merges probabilistic feature extraction with discriminative classification, enabling robust interpretation of non-linear economic structures. Zhang, Hu, Li, Zhang, Yang and Qu (2021) devised a hybrid system integrating support vector regression (SVR) with radial basis functions and extra-trees regression to simulate granular spatiotemporal residential land price distributions. This approach facilitates high-resolution predictions of urban land valuation mechanisms. Yoo, Im and Wagner (2012) evaluated alternative hedonic pricing frameworks against OLS regression for residential appraisal, establishing random-forest and Cubist techniques as superior predictive capabilities. Empirical investigations by Dimopoulos, Tyralis, Bakas and Hadjimitsis (2018), Hong, Choi and Kim (2020), and Dimopoulos and Bakas (2019) corroborate machine learning-driven mass appraisal systems' efficacy. Collectively, these studies highlight advanced computational frameworks' transformative potential—including predictive algorithms and data-driven analytical frameworks—in augmenting the precision, consistency, and scalability of large-scale valuation infrastructures. Additionally, Ai, Liu, Jiang and He (2020) introduced a machine learning architecture to optimize residential land valuation accuracy and efficiency, emphasizing computational spatial analytics' role in urban planning. To resolve spatial heterogeneity in geospatial hedonic models, Picchetti (2017) adopted gradient-boosted regression trees (GBRT) to strengthen estimation robustness, reducing sampling bias through ensemble-based algorithms.

Shaoxing, a prefecture-level city in the southern Yangtze River Delta, occupies a distinctive position in China's urban hierarchy that makes it a compelling setting for residential price forecasting research. Unlike the country's first-tier megacities, which attract disproportionate attention, Shaoxing represents the large and economically consequential stratum of second- and third-tier urban centres that collectively underpin much of China's real estate wealth. The city's economy rests on a diversified manufacturing base—particularly in textiles, chemical fibres, and machinery—and has benefited from deep integration into the Hangzhou-Ningbo-Shanghai economic corridor. This integration has fueled sustained population inflows and rising household incomes over the past decade, which, together with municipal infrastructure expansion and the development of new urban districts, drove a prolonged housing price appreciation between 2013 and mid-2021. Yet, like many similar cities, Shaoxing was not immune to the nationwide property sector correction that took hold in late 2021. The combined effects of tightened credit access, the "three red lines" regulatory framework, and weakened buyer confidence abruptly reversed price momentum, producing a pattern of nonlinear boom-to-correction dynamics that is difficult to reconcile with simple trend-based forecasting models. This volatile trajectory makes Shaoxing a microcosm of the structural adjustments now reshaping regional housing markets across China, and it therefore offers a highly relevant context for testing advanced forecasting methodologies. The selection of Shaoxing for this study is further justified on practical and analytical grounds. First, the city's housing market is dominated by owner-occupiers and local investors rather than by the highly speculative capital flows that characterize a few top-tier cities. Consequently, price movements in Shaoxing are likely to reflect fundamental drivers—such as local economic performance, demographic change, and land supply policies—more faithfully than in markets where speculative noise can obscure underlying trends. This makes Shaoxing an instructive testbed for a model that seeks to capture systematic temporal patterns. Second, the availability of a consistent, long-run monthly price series from the Anjuke platform allows for rigorous temporal modelling without the complications of variable-definition changes or large data gaps that often hamper longer-span studies of smaller cities. Third, despite its economic significance and its representativeness of a large class of Chinese cities, Shaoxing has received limited attention in the computational real estate forecasting literature. Almost all existing machine-learning-based housing studies focused on China have concentrated on first-tier metropolitan areas such as Beijing, Shanghai, Shenzhen, and Guangzhou, or on a few well-known provincial capitals. This neglect leaves a significant empirical gap: the forecasting tools that have been developed for the most liquid and data-rich urban markets may not transfer straightforwardly to cities with thinner trading volumes, more pronounced policy sensitivity, and distinct local economic drivers. Existing modelling approaches, moreover, are not well suited to the unique dynamics present in Shaoxing's housing market. Classical econometric specifications assume linear dependence structures and are known to break down in the presence of structural breaks and regime shifts of the kind observed after late 2021. The non-normality, negative skewness, and heteroskedasticity documented in the present dataset, along with the rejection of linearity by the BDS test, signal that such linear models are unlikely to deliver adequate predictive performance in this context. On the other hand, while more flexible machine-learning methods, including long short-term memory (LSTM) networks, support vector regression, and regression tree techniques, can accommodate nonlinearities, their successful deployment often requires rather large training samples and extensive manual architecture design. In medium-sized cities where long, high-frequency records are scarce, these approaches risk overfitting or providing opaque forecasts that are difficult for policy-makers and valuation professionals to scrutinise. Furthermore, many of these algorithms were initially developed and benchmarked on national or first-tier-city datasets, and their behavioural characteristics under the administrative interventions and supply-side constraints typical of Shaoxing's market remain poorly understood. The confluence of these factors—rapid urbanisation, sharp policy-driven turning points, limited local modelling precedents, and the failure of standard linear and purely data-hungry nonlinear techniques to offer a balanced solution—creates a clear demand for a forecasting framework that can flexibly capture nonlinear temporal dependencies while maintaining statistical robustness and interpretability on a moderately sized dataset. Gaussian process regression, with its probabilistic nonparametric foundation and capacity to encode smoothness assumptions through kernel choice, emerges as a natural candidate to meet this demand. By situating the study in Shaoxing, the present research thus addresses not only a geographic gap but also a methodological one, providing evidence on the practical effectiveness of a modelling strategy that can be replicated in other second and third-tier cities facing similar structural transitions.

The research problem addressed in this study is the accurate forecasting of residential property prices in China's regionally heterogeneous and increasingly volatile housing markets, where nonlinear dynamics, policy shifts, and rapid urbanisation pose significant challenges to traditional econometric models. Although Gaussian process regression (GPR) has been extensively validated for housing valuation in contexts such as Boston (Algikar and Mili, 2023), its application remains markedly underdeveloped in China's spatially diverse real estate systems. In particular, no prior study has employed GPR to model residential price trajectories in Shaoxing City, Zhejiang Province—a representative, rapidly transitioning urban market—leaving a clear methodological gap. To fill this gap, the objective of this study is twofold: to deploy GPR for the first time on a longitudinal Shaoxing dataset spanning January 2013 to July 2024, and to evaluate its predictive performance under rigorous out-of-sample conditions. Methodologically, the GPR architecture integrates adaptive basis functions and multiple kernel specifications, with hyperparameters optimized via Bayesian inference and regularised through cross-validation to balance accuracy and generalisability. The study makes three principal contributions. First, it delivers the inaugural GPR-based residential price forecasting model for Shaoxing, extending nonparametric probabilistic methods to an under-researched Chinese urban context. Second, it demonstrates that a streamlined GPR configuration can attain exceptional forecast accuracy—achieving a relative root mean square error of 0.1774% over the May 2022–July 2024 testing period—thereby setting a new benchmark for machine-learning-driven appraisal in dynamic markets. Third, the proposed framework can function autonomously or in synergy with complementary econometric tools, offering a scalable, interpretable, and policy-ready instrument for evidence-based urban planning, regulatory decision-making, and future inquiry into the spatiotemporal determinants of residential valuations.

The remainder of this paper is organised as follows. Section 2 reviews prior computational and econometric approaches to property price forecasting, situating the present study within the broader literature. Section 3 describes the Shaoxing residential price dataset and its distributional properties. Section 4 details the Gaussian process regression framework, kernel and basis function specifications, Bayesian optimisation protocol, and performance metrics. Section 5 presents the empirical results, including model selection, out-of-sample forecasting performance, and residual diagnostics. Section 6 discusses the practical and policy implications of the findings, and Section 7 concludes with a summary of contributions and directions for future research.

■2.0 LITERATURE REVIEW

2.1 Geographic and Methodological Scope of Computational Property Valuation Studies

Computational methodologies for property valuation estimation have been rigorously investigated across diverse international contexts. Research spans multiple regions, including the United Kingdom (Wilson, Paris,Ware and Jenkins, 2002), New Zealand (Limsombunchai, 2004), Malawi (Embaye, Zereyesus and Chen, 2021), Nigeria (Abidoye and Chan, 2018), and the United States (Huang, 2019), illustrating their adaptability to heterogeneous economic and spatial configurations. In Asian contexts, academic focus has predominantly targeted South Korea (Kang, Lee, Jeong, Lee and Oh, 2020), China (Ho, Tang and Wong, 2021), Malaysia (Rahman, Maimun, Razali and Ismail, 2019), Singapore (Wang, Chan, Wang and Chang, 2016), and Iran (Rafiei and Adeli, 2016). European contributions encompass Spain (Rico-Juan and de La Paz, 2021), Italy (Morano, Tajani and Torre, 2015), and transnational studies (Ćetković, Lakić, Lazarevska, Žarković, Vujošević, Cvijović and Gogić, 2018), alongside methodological innovations from Turkey (Terregrossa and Ibadi, 2021) and Russia (Yasnitsky, Yasnitsky and Alekseev, 2021). African empirical analyses extend to Tanzania and Uganda (Embaye et al., 2021), supplemented by Australian inquiries (Milunovich, 2020). This global diffusion underscores the pervasive adoption of computational frameworks in real estate systems, reflecting their versatility across disparate economic and geographic landscapes. Machine learning techniques exhibit substantial potential for property appraisal through multivariate temporal analysis, with the current investigation concentrating on Shaoxing's residential sector. Existing work demonstrates remarkable geographic diversity, with implementations across varied urban agglomerations. Turkish studies have concentrated on Istanbul (Terregrossa and Ibadi, 2021) and Ankara (Kitapci, Tosun, Tuna and Turk, 2017), while Malaysian research examines Petaling, Kuala Lumpur (Sarip, Hafez and Daud, 2016) and Mukim Pulai, Johor Bahru (Rahman et al., 2019). Spanish analyses prioritize Alicante (Rico-Juan and de La Paz, 2021), Nigerian investigations focus on Benin (Igbinosa, 2011) and Lagos (Abidoye and Chan, 2018), Iranian efforts target Tehran (Rafiei and Adeli, 2016), and Russian inquiries explore Moscow (Yasnitsky et al., 2021). South Korean scholarship emphasizes Seoul (Kang et al., 2020), Italian work spans Milan (Morano and Tajani, 2013), Taranto (Chiarazzo, Caggiani, Marinelli and Ottomanelli, 2014), and Bari (Morano et al., 2015), while New Zealand's focus centers on Christchurch (Limsombunchai, 2004). U.S.-based studies encompass regions such as Wake County, North Carolina (Peterson and Flanagan, 2009); Rutherford County, Tennessee (Nghiep and Al, 2001); New York City (Ge,Wang, Xie, Liu and Zhou, 2019); Boston and Ames (Shahhosseini, Hu and Pham, 2019); Fairfax County, Virginia (Park and Bae, 2015); and California (Huang, 2019). Chinese research addresses metropolitan areas including Tangshan (Gu, Zhu and Jiang, 2011), Chongqing (Wang,Wen, Zhang and Wang, 2014), Shenzhen (Liu and Liu, 2019), Beijing (Li, Xiang and Xiong, 2020), Shanghai (Fu, 2018), Hong Kong (Ho et al., 2021), Kunming, Xuzhou, Changchun, and Handan (Liu and Wu, 2020), Dalian (Piao, Chen and Shang, 2019), and Guangzhou, Shenzhen, Shanghai, and Beijing (Xu and Li, 2021). The broad geographic distribution underscores computational frameworks' ability to tackle valuation intricacies across varied metropolitan settings. This research endeavors to offer novel empirical insights into the methodological deployment of machine learning in property price prediction while establishing a structured evaluative paradigm for residential valuation mechanisms. Concentrating on Shaoxing's residential sector as a case study, the analysis evaluates the operational effectiveness of computational techniques in producing precise valuation forecasts. By prioritizing this localized context, the outcomes aim to fortify data-driven policymaking frameworks for investors and municipal stakeholders, while assessing the scalability and transferability of machine learning models across heterogeneous real estate markets. These insights may further catalyze advancements in predictive analytics and promote their strategic integration within both nascent and mature property sectors.

2.2 Algorithmic Innovations and Comparative Approaches in Real Estate Forecasting

Prior scholarly inquiries into computational methodologies for property valuation have pursued multifaceted objectives across interdisciplinary domains. A central emphasis has involved advancing algorithmic innovations that synthesize heterogeneous real estate attributes (Xu and Li, 2021), alongside parallel efforts integrating macroeconomic indicators into predictive architectures (Kang et al., 2020). Technical forecasting paradigms have systematically incorporated macroeconomic variables such as GDP fluctuations, unemployment dynamics, and interest rate shifts (Yasnitsky et al., 2021), property-specific metrics encompassing parcel dimensions, locational amenities, and structural age (Terregrossa and Ibadi, 2021), and historical price trajectories (Liu and Wu, 2020). This investigation aligns with endeavors to improve computational frameworks for temporal lag analysis (Liu and Wu, 2020), deploying Gaussian process regression enhanced via Bayesian hyperparameter optimization and cross-validation protocols. Contemporary work reflects methodological heterogeneity, with some studies conducting comprehensive evaluations of diverse machine learning strategies (Embaye et al., 2021), while others prioritize singular methodological exploration (Yasnitsky et al., 2021). Earlier work underscores the enhanced efficacy of hybridized computational architectures (Terregrossa and Ibadi, 2021) and systematic juxtapositions between machine learning and classical econometric techniques (Milunovich, 2020). Recent empirical analyses suggest neural networks achieve superior performance relative to alternative methods (Kang et al., 2020), whereas support vector regression (SVR) and ensemble methods demonstrate competitive precision (Li et al., 2020). Comparative evaluations reveal context-dependent algorithmic performance disparities in appraisal tasks. For example, SVR frameworks exhibit heightened accuracy compared to back-propagation neural networks (Li, Xu, Zhao and Chen, 2009), and hybrid SVR models augmented with genetic algorithms outperform fuzzy neural architectures (Wu, Li, Fang, Hsu, Lin and Wu, 2009). Random forest and gradient boosting techniques surpass SVR, AdaBoost, Naive Bayesian networks, and RIPPER algorithms (Ho et al., 2021), while fuzzy-based approaches excel in niche applications (Li, Fong and Chong, 2017). LSTM models dominate convolutional neural networks and exceed evolutionary algorithms, back-propagation derivatives, and SVR benchmarks (Ge et al., 2019). Integrative systems merging SVR with econometric methods outperform boosting, random forest, decision tree, and linear regression frameworks (Huang, 2019), while XGBoost models achieve superior accuracy relative to random forest, LASSO, and bagging techniques (Yan and Zong, 2020). Non-linear methodologies consistently exceed linear regression in precision (Embaye et al., 2021). SVR frameworks outperform Bayesian vector autoregression and random walk approaches in empirical mode decomposition (Plakandaras, Gupta, Gogas and Papadimitriou, 2015), and multilayer perceptron models exceed ARIMA frameworks (Lim, Wang, Wang and Chang, 2016). Divergent empirical outcomes underscore the context-specific utility of algorithms, with no universally optimal methodology identified between deep learning and temporal econometric systems (Milunovich, 2020). Inconsistent findings further emphasize the critical role of dataset characteristics in shaping algorithmic efficacy. This study advances specialized discourse by examining Gaussian process regression applications in property valuation (Yasnitsky et al., 2021). Integrative approaches fusing neural networks with dynamic model averaging (Wei and Cao, 2017), multivariate regression (Terregrossa and Ibadi, 2021), and case-based reasoning (Taffese, 2007) address nonlinear pricing mechanisms inadequately captured by linear econometric models, thereby expanding the methodological frontier in real estate analytics.

2.3 Reported Predictive Accuracy across Computational Techniques

Empirical analyses reveal significant variability in predictive accuracy among computational methodologies applied to real estate appraisal tasks. Research documents mean absolute percentage error (MAPE) measurements spanning multiple intervals: 8%–9% (Kang et al., 2020), 7%–8% (Pai and Wang, 2020), 6%–7% (Liu and Wu, 2020), 5%–6% (Kang et al., 2020), 4%–5% (Kang et al., 2020), and 3%–4% (Liu and Wu, 2020). Lower error thresholds include 2%–3% (Pai and Wang, 2020), 1%–2% (Rico-Juan and de La Paz, 2021), and sub-1% benchmarks (Ho et al., 2021), reflecting methodological diversity and context-driven performance in computational frameworks. Sub-1% MAPE has been attained via neural architectures (Ma, Chen and Zhang, 2015), random forests (Ho et al., 2021), support vector regressions (Ho et al., 2021), augmented Holt's exponential smoothing (Liu andWu, 2020), and gradient-boosted systems (Ho et al., 2021). Algorithmic refinements in XGBoost (Shahhosseini et al., 2019), support vector regressions (Pai and Wang, 2020), neural networks (Lim et al., 2016), and modified Holt's techniques (Liu and Wu, 2020) achieve MAPE within 1%–2%. Error margins of 2%–3% are reported for support vector regressions (Liu and Liu, 2019), neural networks (Liu and Liu, 2019), and regression trees (Pai and Wang, 2020). Error rates of 3%–4% are observed in support vector regressions (Wang et al., 2014), deep restricted Boltzmann machines (Rafiei and Adeli, 2016), adjusted Holt's methods (Liu andWu, 2020), and neural networks (Shahhosseini et al., 2019). Methodologies yielding 4%–5% MAPE include support vector regressions (Wang et al., 2014), neural networks (Rahman et al., 2019), and genetic algorithms (Kang et al., 2020). Higher discrepancy ranges (5%–6%) are noted for neural networks (Liu and Wu, 2020), genetic algorithms (Kang et al., 2020), and support vector regressions (Li et al., 2020). Neural networks (Liu and Wu, 2020) and genetic algorithms (Kang et al., 2020) achieve 6%–7% MAPE, while genetic algorithms address 7%–8% thresholds (Kang et al., 2020). Neural networks (Pai and Wang, 2020) and genetic algorithms (Kang et al., 2020) also report 8%–9% error ranges. These results underscore the contextual dependency of algorithmic efficacy, where forecasting complexity aligns with temporal data attributes, and elevated error margins may remain operationally acceptable under heterogeneous data conditions, emphasizing the necessity for context-specific model selection.

2.4 Deployment of Gaussian Process Regression

A primary obstacle in implementing computational frameworks, particularly Gaussian process regression (GPR), stems from their vulnerability to overfitting and underfitting—challenges equally pervasive in classical econometric approaches. To address this, the study prioritized establishing an optimal balance between predictive precision and model resilience during hyperparameter optimization within the GPR-driven residential appraisal system. While machine learning models frequently demand operational intricacy surpassing traditional econometric counterparts, streamlined configurations remain strategically essential to improve accessibility for practitioners with limited technical expertise. The developed GPR architectures demonstrate computational efficiency on par with classical benchmarks while preserving structural clarity and interpretative accessibility. By eschewing excessive complexity, these designs reduce adoption barriers for users unfamiliar with advanced predictive analytics. This methodological approach integrates theoretical rigor with pragmatic functionality, ensuring reliable predictive performance alongside procedural transparency—an innovation pivotal for advancing the practical deployment of valuation systems across academic research and industrial applications. The framework's emphasis on equilibrium between sophistication and usability fosters interdisciplinary collaboration, enabling stakeholders to leverage predictive insights without necessitating specialized computational training. Moreover, its modular design accommodates iterative refinement, allowing adaptation to evolving market dynamics while maintaining interpretability—a critical consideration for institutional adoption and policy-relevant analytics.

■3.0 DATA

This study employs monthly residential real estate valuation data sourced from the Anjuke platform in Shaoxing, China, covering the period from January 2013 to July 2024. The Anjuke platform is one of China's largest online real estate information providers, aggregating extensive listing and transaction data from verified brokerage sources, and its price series are widely employed in both academic research and industry analysis. Figure 1 presents the dataset's distributional properties through quantile-quantile (Q-Q) plots, kernel density estimates (KDEs), and histograms, applied to both raw price indices and their first-differenced transformations. Summary statistics for the residential price series are detailed in Table 1, including outcomes from the Anderson-Darling normality test. The test yields a $p-$value rejecting the null hypothesis of normality at the 5% significance level, with the series displaying negative skewness and a platykurtic distribution. This non-normality aligns with empirical patterns frequently documented in financial and economic time-series data. The observed skewness and kurtosis metrics further corroborate the presence of asymmetric volatility clustering, a phenomenon recurrent in dynamic urban property markets characterized by episodic demandsupply imbalances and regulatory interventions. Such deviations underscore the necessity of robust nonparametric methodologies capable of accommodating heavy-tailed distributions and nonlinear temporal dependencies inherent to real estate valuation analytics.

Previous research has systematically examined nonlinear interdependencies within the intricate statistical properties of financial and economic time-series data. To empirically validate nonlinear dynamics in the analyzed residential price series, this investigation adopts the Brock-Dechert-Scheinkman (BDS) statistical framework (Brock, Scheinkman, Dechert and LeBaron, 1996). The BDS test is executed across varied embedding dimensions and distance parameters (ϵ), leveraging its capacity to rigorously assess temporal dependencies while mitigating assumptions of linearity. Empirical outcomes demonstrate statistically significant nonlinear structural dependencies, with computed $p-$values converging toward zero, unequivocally rejecting linearity assumptions and affirming complex nonlinear patterns within the valuation trajectory. Guided by these insights, the study emphasizes Gaussian process regression (GPR) as a theoretically congruent methodology for capturing nonlinear pricing mechanisms, seeking to optimize predictive fidelity while explicitly addressing the dataset's inherent spatiotemporal complexity. This methodological alignment ensures robustness against misspecification risks endemic to linear parametric models, while accommodating the volatility and regime shifts characteristic of urban real estate markets. The integration of GPR further enables probabilistic uncertainty quantification, a critical advancement for risk-sensitive decision-making in volatile economic environments.

Figure 1 Shaoxing City, Zhejiang Province, China: Monthly residential property prices, January 2013-July 2024

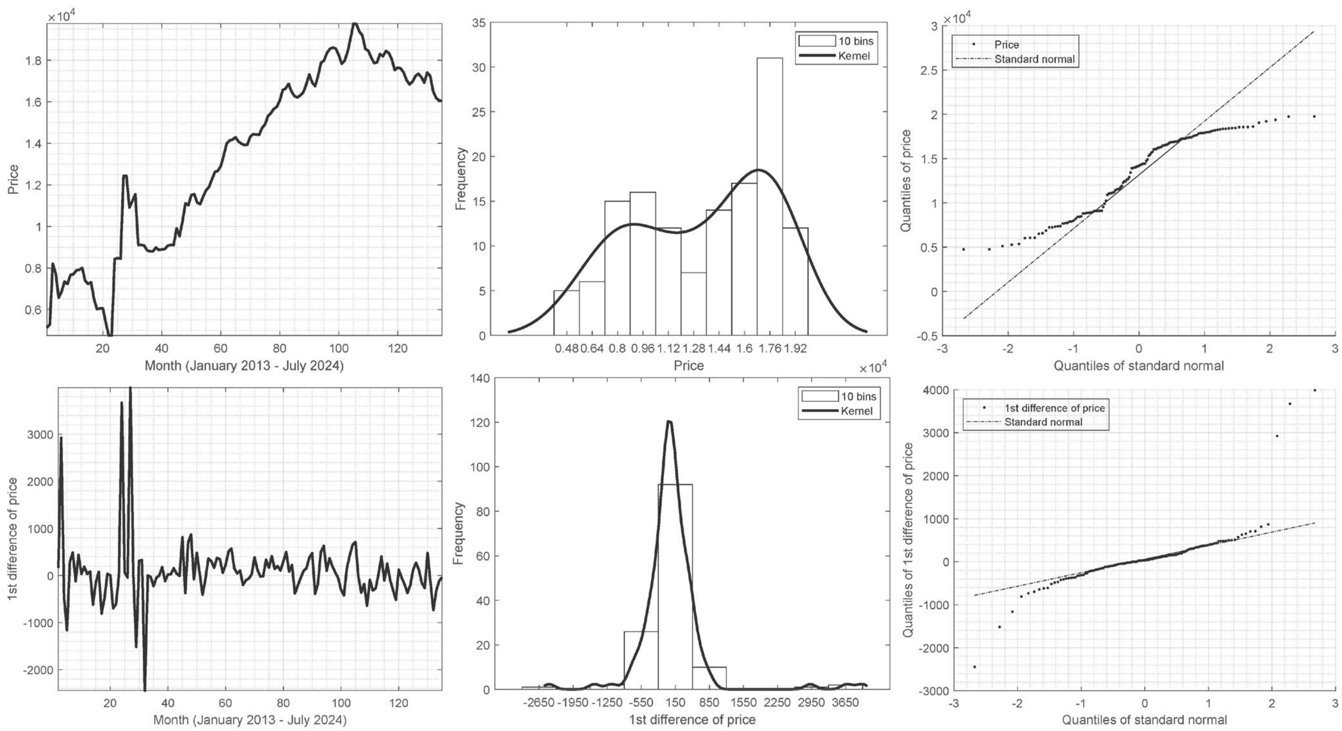

Table 1 Data summary of monthly residential property prices for Shaoxing City, Zhejiang Province, China, from January 2013 to July 2024

Series	Minimum	1st percentile	5th percentile	Mean	Median	Standard deviation	95th percentile	99th percentile	Maximum	Skewness	Kurtosis	Jarque-Bera	Anderson-Darling
Prices	4750	4750	6051.50	13314.74	14182	4340.91	18557.75	19744.15	19762	-0.3371	1.7345	0.0114	<0.0005
First differences	-2444	-1665.32	-638.40	81.41	35	673.60	693.20	3722.40	3987	2.5868	19.0726	<0.001	<0.0005

4.0 METHOD

This study investigates Gaussian process regression (GPR), a probabilistic nonparametric framework grounded in kernel-based learning, which has demonstrated empirical efficacy in modeling nonlinear dependencies across diverse scientific and engineering domains. The underlying mathematical framework formalizes training data through the structured set $\{(x_i, y_i); i = 1,2, \dots, T\}$, where $y_i \in \mathbb{R}$ represents a scalar response variable and $x_i \in \mathbb{R}^d$ corresponds to $d-$dimensional vector of covariates, collectively defining the stochastic structure characterizing the training dataset. In residential price prediction applications, predictors are operationalized as twelve-month historical price lag variables—a design strategy that utilizes sequential monthly observations to forecast thirteenth-month price dynamics. Conventional linear regression paradigms formulate this association as $y = x^T \beta + \epsilon$, operating under the assumption of Gaussian error distribution $\epsilon \sim N(0, \sigma^2)$. Diverging from this approach, GPR postulates a generative model where target variables arise from latent functions $l(x_i)$, modeled as realizations of multivariate Gaussian processes, combined with basis expansions b to enhance representational flexibility. The continuity and smoothness of response surfaces are governed by covariance kernel functions associated with $l(x_i)$, while b facilitates nonlinear transformations of predictors into higher-dimensional feature spaces. Such a formulation facilitates the systematic capture of intricate nonlinear dependencies inherent in housing market dynamics, offering a robust analytical tool for real estate valuation challenges.

This investigation delineates the mathematical foundations of Gaussian processes (GPs), formally defined by their mean function and covariance kernel function, which jointly specify the stochastic properties of these probabilistic models. The covariance kernel is mathematically formalized as $k(x, x') = \mathrm{Cov}[l(x), l(x')]$, while the mean function is defined by $m(x) = E(l(x))$. Within Gaussian process regression (GPR), the predictive framework is structured through the relationship $y = b(x)^T\beta + l(x)$, where $l(x) \sim GP(0, k(x, x'))$ represents a zero-mean Gaussian process and $b(x) \in \mathbb{R}^p$ denotes basis functions. The kernel $k(x, x'|\theta)$ is parameterized by hyperparameters θ, which are iteratively optimized during model training alongside parameters σ^2, β, and basis function specifications to enhance predictive accuracy. The analysis systematically examines two fundamental kernel typologies—isotropic and nonisotropic (automated relevance determination, ARD)—detailed in Equations (1)–(10), with five distinct formulations empirically validated within each category. Key hyperparameters include the scale-mixture coefficient $\alpha > 0$, signal variance σ_f, isotropic length-scale σl, and Euclidean distance metric $r = \sqrt{(x_i - x_j)'(x_i - x_j)}$. To ensure parameter positivity, logarithmic transformations are applied to σ_l and σ_f, yielding transformed hyperparameters $\theta = (\theta_1, \theta_2) = (\log \sigma_l, \log \sigma_f)$. For ARD configurations, distinct length scales σ_m ($m = 1, 2, ..., d$) are assigned to individual predictors, extending θ to a multidimensional hyperparameter space $\theta = (\theta_1, \theta_2, ..., \theta_d, \theta_{d+1}) = (\log \sigma_1, \log \sigma_2, ..., \log \sigma_d, \log \sigma_f)$. This adaptive parameterization enhances model flexibility in capturing heterogeneous predictor effects, thereby improving its capacity to identify variable-specific nonlinear patterns within highdimensional real estate datasets.

$$\text{Isotropic Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{r}{\sigma_l}} \tag{1}$$

$$\text{Isotropic Squared Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{1}{2}\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma_l^2}} \tag{2}$$

$$\text{Isotropic Matern 5/2: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)e^{-\frac{\sqrt{5}r}{\sigma_l}} \tag{3}$$

$$\text{Isotropic Rational Quadratic: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha} \tag{4}$$

$$\text{Isotropic Matern 3/2: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \frac{\sqrt{3}r}{\sigma_l}\right)e^{-\frac{\sqrt{3}r}{\sigma_l}} \tag{5}$$

$$\text{Nonisotropic Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\sqrt{\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \tag{6}$$

$$\text{Nonisotropic Squared Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{1}{2}\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} \tag{7}$$

$$\text{Nonisotropic Matern 5/2: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \sqrt{5\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} + \frac{5}{3}\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right)e^{-\sqrt{5\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \tag{8}$$

$$\text{Nonisotropic Rational Quadratic: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \frac{1}{2\alpha}\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right)^{-\alpha} \tag{9}$$

$$\text{Nonisotropic Matern 3/2: } k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \sqrt{3\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}\right)e^{-\sqrt{3\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \tag{10}$$

This investigation conducts a systematic evaluation of four candidate basis function formulations, formally defined through mathematical expressions in Equations (11)–(14), to assess their theoretical soundness and empirical effectiveness within the established modeling architecture. The analytical framework employs three core mathematical constructs:

.

$$X = (x_1,\ x_2,\ \dots, x_n)',\ X^2 = \begin{pmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1d}^2 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2d}^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1}^2 & x_{T2}^2 & \cdots & x_{Td}^2 \end{pmatrix}, \text{and } B = (b(x_1),\ b(x_2),\ \dots, b(x_n))'$$

$$\text{Empty: } B = Empty\ Matrix \tag{11}$$

$$\text{Constant: } B = I_{n\times 1} \tag{12}$$

$$\text{Linear: } B = [1, X] \tag{13}$$

$$\text{Pure Quadratic: } B = \left[1, X, X^2\right] \tag{14}$$

The consideration of these ten kernel families is motivated by their capacity to span a wide spectrum of covariance structures with fundamentally different smoothness and differentiability properties. The Exponential kernel produces rough, non-differentiable functions; the Squared Exponential kernel generates infinitely differentiable, very smooth trajectories; the Matern 3/2 and 5/2 kernels offer intermediate degrees of smoothness that often suit real-world temporal data; and the Rational Quadratic kernel, as a scale mixture of Squared Exponential kernels, can capture multi-scale variation through the additional parameter α. For each of these five isotropic formulations, a corresponding nonisotropic (automatic relevance determination, ARD) version is also evaluated. By assigning a distinct length-scale σm to each input predictor, the ARD kernels allow the model to learn the predictive relevance of individual lagged price variables and to automatically down-weight uninformative inputs—an important consideration given the twelve-month lag structure used here. The four basis function specifications are likewise chosen to represent a hierarchy of global trend complexity: the empty basis assumes no deterministic component and models the response purely as a zero-mean Gaussian process; the constant basis captures a global mean level; the linear basis accommodates monotonic long-term movements; and the pure quadratic basis allows for curvature in the price trajectory. Simultaneously optimizing over both the kernel type and the basis function therefore provides the model with sufficient flexibility to capture the stochastic local dependencies and any systematic deterministic trends present in Shaoxing's residential price series.

This research deploys a computational framework combining ten-fold cross-validation with Bayesian hyperparameter tuning, employing the Expected Improvement Per Second Plus ($EIPSP$) acquisition strategy for parameter refinement. The analytical methodology centers on a Gaussian process (GP) surrogate model approximating the latent function $f(x)$. Approximation of $y_i = f(x_i)$ utilizes Bayesian inference to probabilistically sample N_s data points x_i within specified bounds. During initial exploratory stages, N_s is set to four observational units. The framework persistently assimilates data, overcoming transient anomalies until achieving Ns valid evaluations. Following this, the algorithm iteratively performs its operational cycle. The workflow initiates with structural modifications to $f(x)$, succeeded by calculating posterior distributions across set Q, formalized as $Q(f|x_i,\ y_i$ for $i = 1, \dots, T)$. Subsequent phases determine an optimal sampling point x and define the minimization objective $a(x)$ governing the acquisition criterion. Function $a(x)$ evaluates the potential utility of candidate x within the probabilistic surrogate Q. Expected improvement approaches prioritize probabilistic benefits rather than incremental gains, x_{best} represents the posterior mean minimizer, with $\mu_Q(x_{best})$ indicating its minimal expected value. The improvement metric is mathematically expressed as $EI(x,Q) = EQ[\max(0,\ \mu_Q(x_{\text{best}}) - f(x))]$. Incorporating temporal weighting enhances optimization efficiency by modeling spatiotemporal relationships. An auxiliary surrogate model estimates evaluation latency $\mu_S(x)$, facilitating $\text{EIPS}(x) = \frac{EI_Q(x)}{\mu_S(x)}$, with $\mu_S(x)$ derived from the temporal GP model's posterior mean for computation time. Adaptive strategies are embedded to prevent local optima convergence and mitigate regional overexploitation. Here, $\sigma_F(x)$ captures posterior target variability, with σ_{PN} denoting additive noise variance, adhering to $\sigma^2{}_Q(x) = \sigma^2{}_F(x) + \sigma^2{}_{PN}$. A positive threshold $t_{\sigma PN} > 0$ governs algorithmic behavior. Following each iteration, the $EIPSP$ protocol assesses if $\sigma_F(x) < t_{\sigma PN}\sigma_{PN}$. Detection of overexploitation triggers kernel reparameterization and doubling of θ (Bull, 2011). Simultaneously, σ_Q is increased at intermediate points to expand exploration. A revised candidate x is generated using the adjusted kernel. Sustained exploitation prompts a tenfold increase in θ, with a five-iteration limit to identify viable candidates. The updated x dynamically adjusts exploration ratios via $EIPSP$'s adaptive mechanisms. This iterative approach equilibrates exploration of novel regions and exploitation of adjacent areas, optimizing predictive performance.

This research employs Bayesian optimization techniques integrating basis function specifications, kernel parameter configurations, σ hyperparameter tuning, and normalized covariate transformations to improve predictive accuracy. Model performance assessment leverages the relative root mean squared error ($RRMSE$), a standardized metric enabling comparative assessments across different modeling frameworks (Despotovic, Nedic, Despotovic and Cvetanovic, 2016). The $RRMSE$ is mathematically defined by: $RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{for})^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i^{obs}}$, where y^{for} and y^{obs} denote predicted and empirical values, respectively, with n representing the sample count. Complementary evaluation metrics include the mean absolute error (MAE) and root mean squared error ($RMSE$), preserving the dimensional units of the response variable: $MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i^{obs} - y_i^{for}|$ and $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{for})^2}$. Predictive robustness is further quantified using the correlation coefficient (CC), which evaluates linear association between observed and forecasted outcomes: $CC = \frac{\sum_{i=1}^{n}\left(y_i^{obs} - \overline{y^{obs}}\right)\left(y_i^{for} - \overline{y^{for}}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i^{obs} - \overline{y^{obs}}\right)^2 \sum_{i=1}^{n}\left(y_i^{for} - \overline{y^{for}}\right)^2}}$, where $\overline{y^{\text{obs}}}$ and $\overline{y^{\text{for}}}$ signify the arithmetic means of empirical and predicted values. These metrics collectively constitute a multidimensional validation framework for rigorous algorithmic evaluation, ensuring comprehensive scrutiny of model fidelity and generalizability.

■5.0 RESULTS

This analytical framework employs monthly residential real estate valuation records covering January 2013 through April 2022 for model calibration, succeeded by sequential single-step forecasting and rigorous empirical testing conducted across a 27-month horizon spanning May 2022 to July 2024. For cross validation, we segment the training phase into ten equal folds following the ordering of time. As visualized in Figure 2, the implementation of the $EIPSP$ optimization protocol on training data identifies normalized covariates, an isotropic exponential covariance structure (Equation 1), and invariant basis formulation (Equation 12) as the superior configuration. Comprehensive hyperparameter estimation outcomes for ten Gaussian process regression implementations, generated through ten-fold cross-validation, are systematically presented in Table 2. The tabular organization stratifies analytical outcomes across ten mutually exclusive validation subsets (designated CV1 through CV10), enabling critical assessment of parameter consistency among segmented training cohorts. This partitioned validation strategy permits thorough examination of predictive transferability while mitigating overfitting tendencies through cyclical parameter refinement across distinct data subdivisions, thereby enhancing algorithmic robustness against dataset-specific biases.

Figure 2 Results of Bayesian optimisations

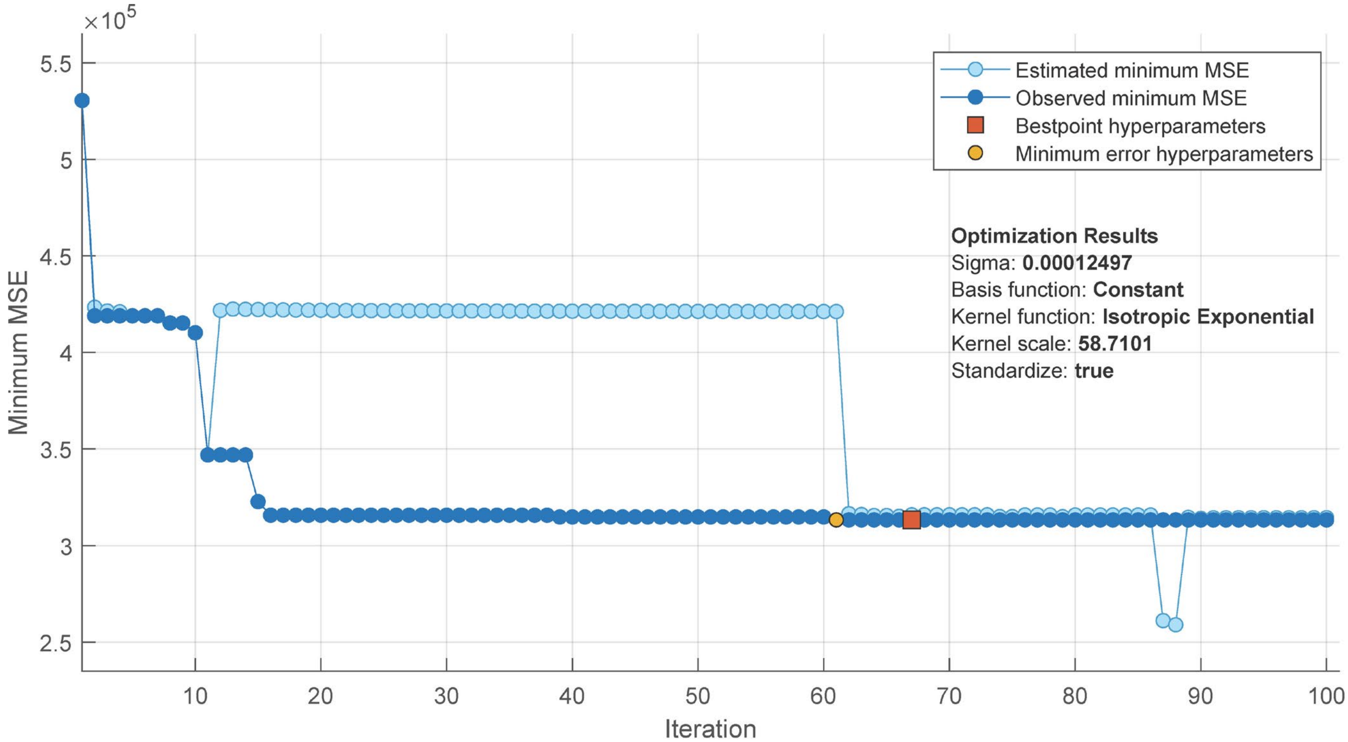

Table 2 Results of model parameter estimations

	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9	CV10
σ	40.9780	40.3291	39.7477	40.0577	41.2640	41.0860	40.5724	40.6965	39.8096	41.2363
β	13749.7617	13689.6160	13112.9790	13773.5780	13688.5990	13807.6757	13569.4144	13810.8344	13929.8609	13803.7033
σ_l	48.7125	47.0620	93.4218	49.2913	44.5015	49.0989	58.6190	46.7936	57.5353	48.2694
σ_f	6097.0052	5974.1848	7314.6934	6011.5289	5871.0753	6059.0477	6203.2897	6022.8422	6072.0219	6009.5952

This analysis encompasses a 27-month temporal scope from May 2022 to July 2024, employing ten Gaussian process regression variants (denoted CV1–CV10 in Table 2) to generate computational approximations of housing market valuations. Within each monthly operational cycle throughout the temporal window, ten independent predictive outputs are computed, with consolidated forecasts derived via an ensemble aggregation strategy utilizing arithmetic mean integration. This methodological framework reduces variance from outlier predictions among constituent models, enhancing result reproducibility and statistical robustness. Graphical juxtapositions of historical price trajectories and simulated outputs are presented in Figure 3, supplemented by percentage-based error distributions quantifying forecast deviations in Figure 4. Visual analysis demonstrates close correspondence between observed market dynamics and algorithmic projections. Quantitative validation employs standardized performance indices—root mean square error ($RMSE$), relative root mean square error ($RRMSE$), mean absolute error (MAE), and correlation coefficient (CC)—systematically cataloged in Table 3. The derived $RRMSE$ of 0.1774% aligns with established accuracy classification schemas (Despotovic et al., 2016), which stratify predictive capability into hierarchical tiers: exemplary ($RRMSE$ < 10%), competent (10% < $RRMSE$ < 20%), moderate (20% < $RRMSE$ < 30%), or inadequate ($RRMSE \geq$ 30%). These empirical outcomes validate the operational efficacy of the GPR-based framework in producing high-precision real estate forecasts, demonstrating congruence between theoretical architecture and empirical performance. The methodological rigor is evidenced through systematic cross-validation, ensemble error mitigation, and alignment with standardized evaluation taxonomies, collectively affirming the model's capacity to capture complex market dynamics while maintaining generalizability across temporal partitions.

Figure 3 Comparisons between predicted and actual prices of residential real estate during the May 2022-July 2024 out-of-sample period

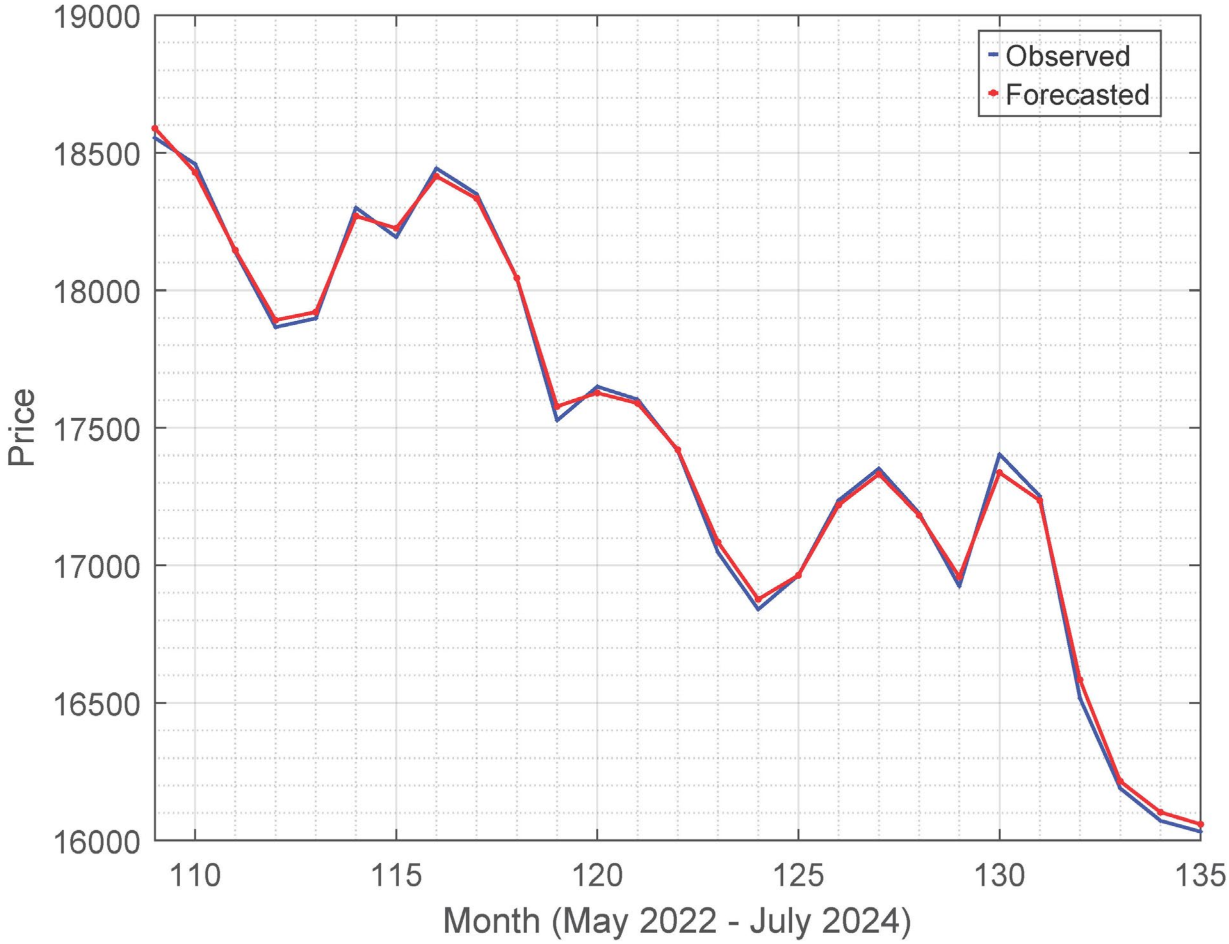

Table 3 Prediction performance for prices of residential real estate over the May 2022-July 2024 out-of-sample period

RRMSE	RMSE	MAE	CC
0.1774%	30.9836	26.1506	99.923%

This analytical inquiry conducted a systematic diagnostic assessment of autocorrelated residuals within forecasting errors to evaluate the structural validity of the computational framework, with analytical results graphically depicted in Figure 5. The investigation methodically analyzed autocorrelation behavior over twenty temporal lags, employing standardized autocorrelation metrics to maintain procedural transparency and analytical coherence. Experimental results revealed statistically insignificant autocorrelation values across consecutive lags, confirming the robustness of the formulated models. Additionally, although temporal discrepancies in empirical coherence were noted, identified autoregressive conditional heteroskedasticity (ARCH) phenomena could enhance predictive reliability within sequential forecasting frameworks. This observation underscores the potential value of incorporating volatility clustering patterns into accuracy evaluations, proposing that such approaches may optimize error modeling and parameter adjustment in time-series data contexts. The absence of significant residual autocorrelation aligns with theoretical expectations for well-specified models, suggesting minimal systematic bias in forecast errors. Conversely, the detection of ARCH effects implies latent volatility persistence that, when explicitly modeled, could refine uncertainty quantification in dynamic market environments. These dual findings—statistical independence in residuals alongside conditional variance clustering—highlight the nuanced interplay between mean and variance structures in temporal forecasting systems, advocating for integrated diagnostic frameworks that address both aspects to strengthen predictive validity.

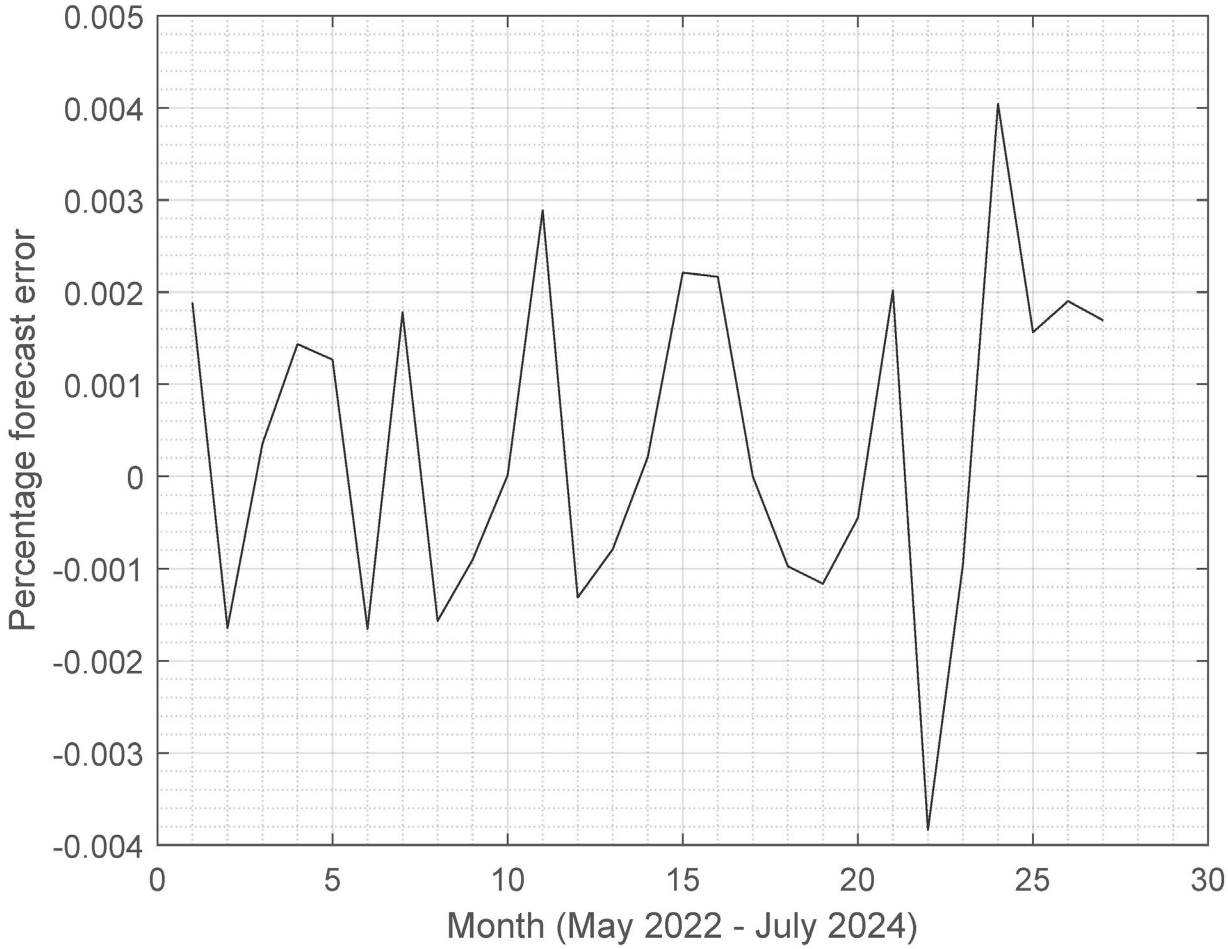

Figure 4 Percentage prediction errors for prices of residential real estate during the May 2022-July 2024 out-of-sample period

We performed a systematic evaluation to benchmark Gaussian Process Regression (GPR) against four competing methods, aiming to set a clear performance standard. The study included three established machine learning techniques—an LSTM (Long Short-Term Memory) network, support vector regression (SVR), a regression tree (RT)—plus a traditional econometric time-series method, the Autoregressive (AR) model, frequently used for economic forecasting. To guarantee a fair and rigorously controlled evaluation of the fundamental forecasting architectures, we intentionally configured the competing LSTM, SVR, RT, and AR models to receive exactly the same input features as the GPR model. This design guarantees that any disparities in predictive accuracy arise exclusively from each method's distinct algorithmic characteristics. Subsequently, the empirical investigation carried out an out-of-sample test covering the period from May 2022 to July 2024. During this evaluation window, the LSTM, SVR, RT, and AR approaches yielded RMSE values: LSTM 48.582, SVR 60.9757, RT 73.3413, and AR 87.2838. Notably, all these alternative methods produced substantially higher error metrics compared to the 30.9836 RMSE achieved by the GPR model. This considerable margin strongly indicates that the GPR modeling framework delivers superior predictive accuracy for Shaoxing real estate price indices. This clear performance advantage shows that GPR's inherent capacity to model complex, nonlinear temporal relationships is particularly well-suited for this forecasting task. Being a strictly linear method, the AR model is unable to capture the complex underlying patterns in the data. While LSTM, SVR, and RT are non-parametric machine learning methods that can model nonlinearity, their underlying structures turned out to be less effective than GPR for this task. Hence, the experimental findings solidly position GPR as a more powerful modeling tool for this financial forecasting domain, outperforming both simple linear econometric models and other flexible non-parametric machine learning methods.

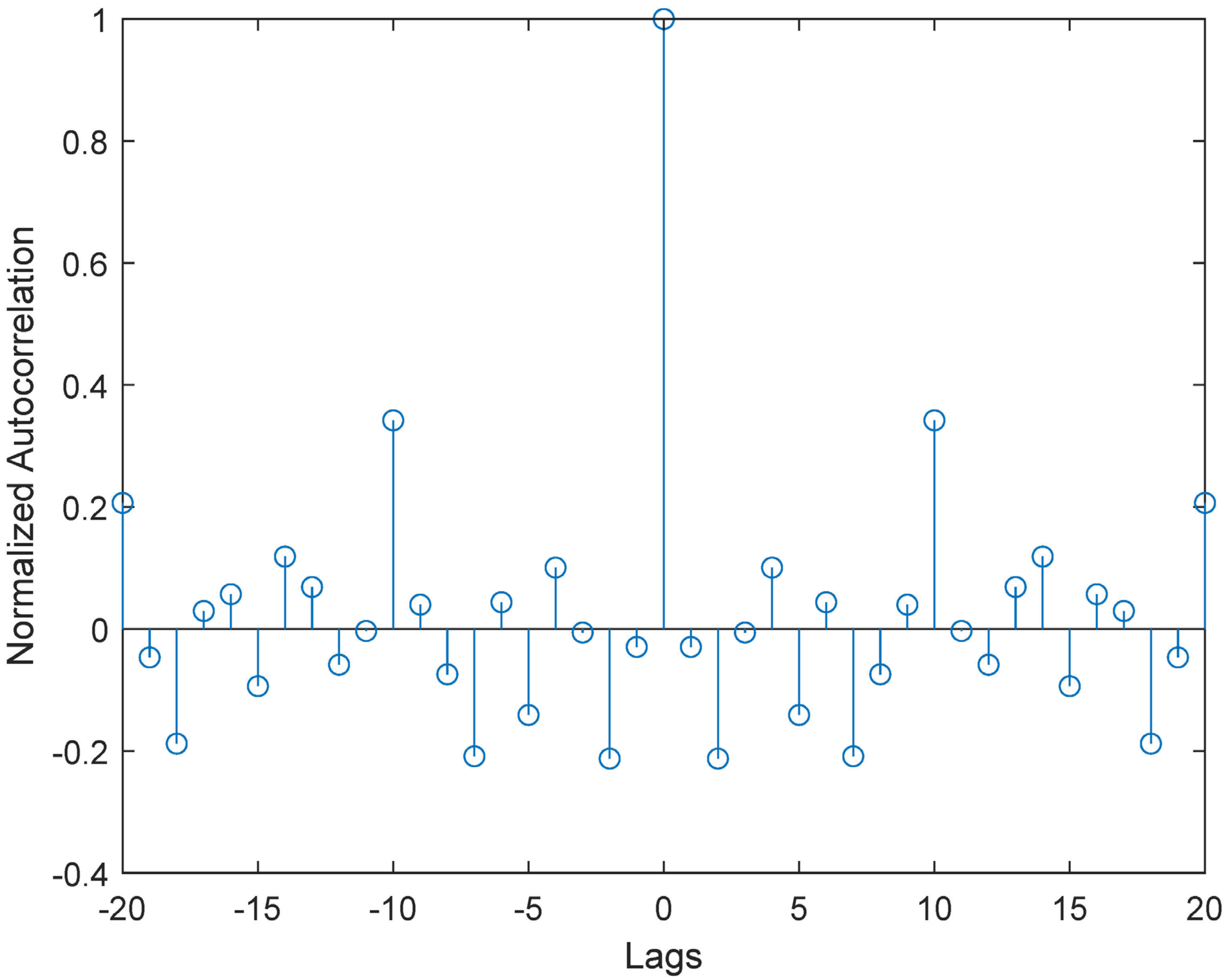

Figure 5 Results of error autocorrelation analysis

■6.0 IMPLICATION

This study demonstrates that a Gaussian process regression model trained on twelve-month price histories can generate accurate out-of-sample residential price forecasts for Shaoxing, achieving a RRMSE of 0.1774% and outperforming several common benchmarks. For valuation practitioners and mortgage lenders, the model offers a timely, data-driven benchmark that can be updated monthly using available transaction prices, supporting loan-to-value assessments, collateral valuation, and the reconciliation of comparable-sales estimates. Local government agencies tasked with property tax assessment can incorporate the forecasts into mass appraisal systems to reduce manual updates. From a market interpretation perspective, the model captures nonlinear boom-to-correction dynamics, helping planners and policymakers identify turning points and evaluate whether recent price movements align with long-term trends. However, several important limitations condition these contributions. The use of historical prices means the model does not exploit land-supply signals, demographic shifts, or policy variables. Its performance has been demonstrated for a single city over a specific historical window, so external validity across different urban contexts remains to be tested. Practically, implementation requires a clean, consistent price time series, which may be challenging in markets with thin trading or data gaps. Nevertheless, the model's parsimonious input structure makes replication feasible in other cities that maintain reliable price indices, allowing systematic benchmarking and comparison across institutional settings. Used as a complement to—not a replacement for—traditional valuation methods and qualitative market knowledge, the framework offers a transparent and interpretable forecasting tool that can strengthen evidence-based decision-making in both the public and private sectors.

Beyond these general contributions, the empirical findings carry specific and actionable implications for valuation practice, land administration, housing policy formulation, and the analysis of real estate markets beyond Shaoxing. In valuation practice, the demonstrated capacity to generate accurate point forecasts—attaining a relative root mean square error of 0.1774% over a 27-month out-of-sample horizon—positions the Gaussian process regression framework as a powerful complement to conventional appraisal techniques. Professional valuers, mortgage lenders, and financial institutions routinely rely on comparable sales methods, hedonic pricing models, and income capitalization approaches that, while well established, can struggle to keep pace with rapidly shifting market conditions or to furnish timely benchmarks in thinly traded segments. A systematically calibrated, data-driven forecasting engine that operates on publicly available historical transaction prices offers a transparent and repeatable numerical anchor. It can support the reconciliation of value estimates, provide early indications of turning points for loan-to-value ratio assessments, and strengthen the objectivity of collateral valuation in mortgage underwriting. Furthermore, local government bodies tasked with property tax assessment and mass appraisal may incorporate such point forecasts into computer-assisted mass appraisal systems, using the model's monthly projections to update cadastral values with reduced manual intervention and enhanced across-the-board consistency. From a land administration perspective, accurate residential price forecasts constitute an indispensable input for strategic land supply decisions, infrastructure investment planning, and zoning policy. Municipal authorities in Shaoxing and comparable expanding urban centres must decide when to release land parcels, how to price long-term land leases, and where to direct transport and utility investments. A forecasting framework that reliably captures the nonlinear propagation of past price dynamics into future months allows planners to align land release programmes with anticipated market demand, thereby mitigating both supply gluts and artificial scarcity. The model's performance during the post-2021 adjustment phase—a period marked by pronounced policy interventions and demand-supply disequilibrium—attests to its suitability for environments in which regulatory shifts and demographic pressures interact in complex ways. Urban planners can link forecast trajectories to land value capture mechanisms, estimating the likely appreciation attributable to new transit corridors or public amenities and calibrating developer contributions accordingly. While the present study relies on historical prices, the modular architecture readily accommodates the inclusion of spatially referenced variables such as distances to employment centres, schools, and metro stations, offering a natural pathway toward spatially explicit land value modelling. For housing policy, the availability of dependable forecasts enhances capacity for diagnostic monitoring and the design of market-stabilising interventions. Regulatory agencies seeking to dampen speculative bubbles or to cushion sharp corrections require forward-looking indicators that are both accurate and quickly updatable. The forecasting procedure validated in this research, which can be refreshed each month as new price observations become available, enables near-real-time tracking of price momentum. Policy-makers can compare short-term forecasts against long-term fundamental benchmarks to identify overheating phases, triggering graduated measures such as adjustments to down-payment requirements, purchase eligibility restrictions, or stamp duty rates. Conversely, during downturns, downward price projections can guide the timing and scale of fiscal stimulus, tax relief, or social housing procurement. Because the methodology focuses on price estimates, its outputs integrate directly into the types of deterministic scenario analyses that are common in policy briefings and legislative impact assessments, without requiring stakeholders to engage with over-complicated reasoning. In addition, city-level housing affordability programmes—including shared-ownership schemes, rental subsidy calibrations, and inclusionary zoning mandates—benefit from precise anticipations of market price evolution, which inform income thresholds and subsidy ceilings in a manner that keeps public resources aligned with market realities. Finally, the study's implications extend substantially to comparative real estate market analysis beyond Shaoxing. The methodological core—a Gaussian process regression with Bayesian optimisation of kernel and basis functions, regularised through cross-validation—is not tied to any local idiosyncrasy. Its input structure, which parsimoniously employs price histories, minimises data collection burdens and facilitates replication in cities where price time series are readily available. Researchers and government statistical agencies can deploy the same framework in other Chinese cities, or in international settings, to produce city-specific forecasting models whose predictive accuracy can be directly compared using the same relative error metrics. Such comparative benchmarking would shed light on how institutional factors—such as land auction systems, mortgage regulation, and urban growth boundaries—modulate the predictability of residential markets. Moreover, the ensemble approach that aggregates cross-validated model variants can serve as a robust baseline against which more complex, data-intensive architectures (e.g., those incorporating demographic flows, satellite imagery, or natural language sentiment indices) are evaluated. By establishing a high-accuracy floor with a transparent and transferable design, the Shaoxing case study encourages the adoption of advanced computational forecasting tools in emerging real estate research communities, fostering cumulative knowledge and evidence-based spolicy well beyond a single municipality.

■7.0 CONCLUSION

This study applied Gaussian process regression to forecast monthly residential property prices in Shaoxing, Zhejiang Province, using a longitudinal dataset from January 2013 to July 2024. The modelling framework considered ten covariance kernel families, four basis function specifications, and Bayesian hyperparameter optimisation through the Expected Improvement Per Second Plus ($EIPSP$) protocol, with model selection regularised via cross-validation. The inputs consisted of twelve-month historical price lags; no geospatial, demographic, or macroeconomic covariates were used. In out-of-sample testing over May 2022–July 2024, the ensemble of cross-validated GPR variants achieved a relative root mean square error of 0.1774% and a root mean square error of 30.98, substantially outperforming LSTM (48.58), support vector regression (60.98), regression tree (73.34), and a simple autoregressive model (87.28) under identical input conditions. These results demonstrate that a carefully regularised GPR configuration can capture the nonlinear temporal dependencies present in Shaoxing's housing market using lagged price information. Because the model relies on historical prices, its contributions should be stated with corresponding caution. The study provides robust evidence that GPR, when combined with Bayesian optimisation and cross-validation, can serve as an accurate and interpretable forecasting tool for a medium-sized Chinese city undergoing rapid structural adjustment. This has direct relevance for valuation practitioners, mortgage lenders, and local government agencies that require timely, data-driven price benchmarks but might lack access to extensive auxiliary datasets. The parsimonious input structure also makes the framework readily transferable to other cities where only transaction price series are available, facilitating comparative benchmarking and the development of city-specific forecasting models. At the same time, several limitations qualify the findings. First, the reliance on lagged prices means the model does not exploit potentially informative variables such as land supply, infrastructure investment, population flows, or regulatory changes. Second, the analysis covers a single city over a specific historical period and, while the out-of-sample results are strong, the generalizability to other urban contexts has not been empirically tested. Future research can address these limitations in several directions. Obvious extensions include incorporating macroeconomic indicators, land-use variables, or spatial accessibility measures as additional predictors to test whether they improve forecast accuracy beyond what lags can achieve. Hybrid architectures that combine GPR with econometric models, such as spatial lag or time-varying parameter frameworks, may help disentangle local from systemic price drivers. Replicating the methodology across multiple Chinese cities—and in international settings—would provide evidence on its broader applicability and reveal how institutional factors modulate forecasting performance. Finally, while the present work concentrates on the *EIPSP* algorithm for implementing Bayesian optimization, future work could explore Bayesian optimization methods together with different machine learning models for forecasting real estate prices.

Acknowledgement

We would like to acknowledge the anonymous reviewers for their critical and constructive comments that helped improve the quality of this paper.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Abidoye, R.B., Chan, A.P., 2018. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal* 24, 71–83. doi:10.1080/14445921.2018.1436306.

Abidoye, R.B., Chan, A.P., Abidoye, F.A., Oshodi, O.S., 2019. Predicting property price index using artificial intelligence techniques: evidence from hong kong. *International Journal of Housing Markets and Analysis* doi:10.1108/IJHMA-11-2018-0095.

Ai, H., Liu, Q., Jiang, Y., He, J., 2020. Urban residential land price appraisal via quantifying impact factors based on deep belief networks, in: *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, pp. 29–33. doi:10.1145/3383972.3384017.

Al-Shayea, Q.K., 2012. Estate market forecast using artificial neural networks .

Algikar, P., Mili, L., 2023. Robust gaussian process regression with huber likelihood doi:10.48550/arXiv.2301.07858.

Baroni, M., Barthélémy, F., Mokrane, M., 2005. A PCA Factor Repeat Sales Index (1973-2001) To Forecast Apartment Prices in Paris (France). ESSEC.

Bee-Hua, G., 2000. Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: the case of the singapore residential sector. *Construction Management & Economics* 18, 209–217. doi:10.1080/014461900370834.

Brahim-Belhouari, S., Bermak, A., 2004. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis* 47, 705–712. doi:10.1016/j.csda.2004.02.006.

Brahim-Belhouari, S., Vesin, J.M., 2001. Bayesian learning using gaussian process for time series prediction, in: *Proceedings of the 11th IEEE Signal ProcessingWorkshop on Statistical Signal Processing (Cat. No. 01TH8563), IEEE*. pp. 433–436. doi:10.1109/SSP.2001.955315.

Brock,W.A., Scheinkman, J.A., Dechert,W.D., LeBaron, B., 1996. A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235. doi:10.1080/07474939608800353.

Brooks, C., Tsolacos, S., 2000. Forecasting models of retail rents. *Environment and Planning A* 32, 1825–1839. doi:10.1068/a3332.

Brooks, C., Tsolacos, S., Lee, S., 2000. The cyclical relations between traded property stock prices and aggregate time-series. *Journal of Property Investment&Finance* doi:10.1108/14635780010357532.

Bull, A.D., 2011. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research* 12.

Ćetković, J., Lakić, S., Lazarevska, M., Žarković, M., Vujošević, S., Cvijović, J., Gogić, M., 2018. Assessment of the real estate market value in the european market by artificial neural networks application. *Complexity* 2018. doi:10.1155/2018/1472957.

Chi, J., 2017. Spatial bp neural network model in evaluation of residential real estate price. *Geospatial Information* 2, 86–90.

Chiarazzo, V., Caggiani, L., Marinelli, M., Ottomanelli, M., 2014. A neural network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia* 3, 810–817. doi:10.1016/j.trpro.2014.10.067.

Clapp, J.M., Giaccotto, C., 1992. Estimating price indices for residential property: a comparison of repeat sales and assessed value methods. *Journal of the American Statistical Association* 87, 300–306. doi:10.1080/01621459.1992.10475209.

Crosby, H., Davis, P., Damoulas, T., Jarvis, S.A., 2016. A spatio-temporal, gaussian process regression, real-estate price predictor, in: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4. doi:10.1145/2996913.2996960.

D'Arcy, E., McGough, T., Tsolacos, S., 1999. An econometric analysis and forecasts of the office rental cycle in the dublin area. *Journal of Property Research* 16, 309–321. doi:10.1080/095999199368067.

Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S., 2016. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews* 56, 246–260. doi:10.1016/j.rser.2015.11.058.

Dimopoulos, T., Bakas, N., 2019. Sensitivity analysis of machine learning models for the mass appraisal of real estate. case study of residential units in nicosia, cyprus. *Remote Sensing* 11, 3047. doi:10.3390/rs11243047.

Dimopoulos, T., Tyralis, H., Bakas, N.P., Hadjimitsis, D., 2018. Accuracy measurement of random forests and linear regression for mass appraisal models that estimate the prices of residential apartments in nicosia, cyprus. *Advances in Geosciences* 45, 377–382. doi:10.5194/adgeo-45-377-2018.

Embaye,W.T., Zereyesus, Y.A., Chen, B., 2021. Predicting the rental value of houses in household surveys in tanzania, uganda and malawi: Evaluations of hedonic pricing and machine learning approaches. *Plos One* 16, e0244953. doi:10.1371/journal.pone.0244953.

Eppli, M.J., Shilling, J.D., Vandell, K.D., 1998. What moves retail property returns at the metropolitan level? *The Journal of Real Estate Finance and Economics* 16, 317–342. doi:10.1023/A:1007783900583.

Fu, T., 2018. Forecasting second-hand housing price using artificial intelligence and machine learning techniques, in: *2018 8th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2018), Atlantis Press*. pp. 269–273. doi:10.2991/mcei-18.2018.54.

Ge, C., Wang, Y., Xie, X., Liu, H., Zhou, Z., 2019. An integrated model for urban subregion house price forecasting: A multi-source data perspective, in: *2019 IEEE International Conference on Data Mining (ICDM), IEEE*. pp. 1054–1059. doi:10.1109/ICDM.2019.00123.

Gençay, R., Yang, X., 1996. A forecast comparison of residential housing prices by parametric versus semiparametric conditional mean estimators. *Economics letters* 52, 129–135. doi:10.1016/S0165-1765(96)00851-8.

Ghysels, E., Plazzi, A., Valkanov, R., Torous, W., 2013. Forecasting real estate prices. *Handbook of Economic Forecasting* 2, 509–580. doi:10.1016/B978-0-444-53683-9.00009-8.

Glennon, D., Kiefer, H., Mayock, T., 2018. Measurement error in residential property valuation: An application of forecast combination. *Journal of Housing Economics* 41, 1–29. doi:10.1016/j.jhe.2018.02.002.

Gu, J., Zhu, M., Jiang, L., 2011. Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications* 38, 3383–3386. doi:10.1016/j.eswa.2010.08.123.

Guo, Z., 2020. The forecast of housing price in xi'an based on big data analysis. *Insight-Statistics* 3, 7–14.

Hepşen, A., Vatansever, M., 2011. Forecasting future trends in dubai housing market by using box-jenkins autoregressive integrated moving average. *International Journal of Housing Markets and Analysis* doi:10.1108/17538271111153004.

Ho, W.K., Tang, B.S., Wong, S.W., 2021. Predicting property prices with machine learning algorithms. *Journal of Property Research* 38, 48–70. doi:10.1080/09599916.2020.1832558.

Hong, J., Choi, H., Kim, W.s., 2020. A house price valuation based on the random forest approach: the mass appraisal of residential property in south korea. *International Journal of Strategic Property Management* 24, 140–152. doi:10.3846/ijspm.2020.11544.

Huang, Y., 2019. Predicting home value in california, united states via machine learning modeling. Statistics, *Optimization & Information Computing* 7, 66–74. doi:10.19139/soic.v7i1.435.

Igbinosa, S.O., 2011. Determinants of residential property value in nigeria–a neural network approach. *African Research Review* 5, 152–168. doi:10.4314/afrrev.v5i5.13.

Jackson, C., 2001. A model of spatial patterns across local retail property markets in great britain. *Urban Studies* 38, 1445–1471. doi:10.1080/00420980120076740.

Jamil, S., Mohd, T., Masrom, S., Ab Rahim, N., 2020. Machine learning price prediction on green building prices, in: *2020 IEEE Symposium on Industrial Electronics & Applications (ISIEA), IEEE*. pp. 1–6. doi:10.1109/ISIEA49364.2020.9188114.

Kaboudan, M., Sarkar, A., 2007. A gis framework to forecast residential home prices, in: *23rd Annual Meeting of the American Real Estate Society*.

Kang, J., Lee, H.J., Jeong, S.H., Lee, H.S., Oh, K.J., 2020. Developing a forecasting model for real estate auction prices using artificial intelligence. *Sustainability* 12, 2899. doi:10.3390/su12072899.

Kim, G.H., Shin, J.M., Kim, S., Shin, Y., 2013. Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine doi:10.4236/jbcpr.2013.11001.

Kim, L., 2004. Time-varying macroeconomic risk and commercial real estate: an asset pricing perspective. *Journal of Real Estate Portfolio Management* 10, 47–57. doi:10.1080/10835547.2004.12089695.

Kitapci, O., Tosun, Ö., Tuna, M.F., Turk, T., 2017. The use of artifi cial neural networks (ann) in forecasting housing prices in ankara, turkey. *Journal of Marketing and Consumer Behaviour in Emerging Markets* , 4–14.

Levesque, T.J., 1994. Modelling the effects of airport noise on residential housing markets: a case study of winnipeg international airport. *Journal of transport economics and policy* , 199–210.

Li, D.Y., Xu,W., Zhao, H., Chen, R.Q., 2009. A svr based forecasting approach for real estate price prediction, in: *2009 International Conference on Machine Learning and Cybernetics, IEEE*. pp. 970–974. doi:10.1109/ICMLC.2009.5212389.

Li, H., Shen, Q., Love, P.E., 2005. Cost modelling of office buildings in hong kong: an exploratory study. *Facilities* doi:10.1108/02632770510602379.

Li, R.Y.M., Fong, S., Chong, K.W.S., 2017. Forecasting the reits and stock indices: group method of data handling neural network approach. *Pacific Rim Property Research Journal* 23, 123–160. doi:10.1080/14445921.2016.1225149.

Li, Y., Xiang, Z., Xiong, T., 2020. The behavioral mechanism and forecasting of beijing housing prices from a multiscale perspective. *Discrete Dynamics in Nature and Society* 2020. doi:10.1155/2020/5375206.

Lim,W.T.,Wang, L.,Wang, Y., Chang, Q., 2016. Housing price prediction using neural networks, in: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE*. pp. 518–522. doi:10.1109/FSKD.2016.7603227.

Limsombunchai, V., 2004. House price prediction: hedonic price model vs. artificial neural network, in: *New Zealand agricultural and resource economics society conference*, pp. 25–26.

Liu, L., Wu, L., 2020. Predicting housing prices in china based on modified holt's exponential smoothing incorporating whale optimization algorithm. *Socio-Economic Planning Sciences* 72, 100916. doi:10.1016/j.seps.2020.100916.

Liu, R., Liu, L., 2019. Predicting housing price in china based on long short-term memory incorporating modified genetic algorithm. *Soft Computing* 23, 11829–11838. doi:10.1007/s00500-018-03739-w.

Ma, H., Chen, M., Zhang, J., 2015. Study on the prediction of real estate price index based on hhga-rbf neural network algorithm. *International Journal of u-and e-Service, Science and Technology* 8, 109–118. doi:10.14257/ijunesst.2015.8.7.11.

McGough, T., Tsolacos, S., 1995. Forecasting commercial rental values using arima models. *Journal of Property Valuation and Investment* doi:10.1108/14635789510147801.

Mei, H., Fang, H., 2017. A study on the real estate price forecast model in the midwest of china–based on provincial panel data analysis, in: *Proceedings of the Tenth International Conference on Management Science and Engineering Management, Springer*. pp. 525–536. doi:10.1007/978-981-10-1837-4_45.

Milunovich, G., 2020. Forecasting australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting* 39, 1098–1118. doi:10.1002/for.2678.

Mohd, T., Jamil, N.S., Johari, N., Abdullah, L., Masrom, S., 2020a. An overview of real estate modelling techniques for house price prediction. *Charting a Sustainable Future of ASEAN in Business and Social Sciences* , 321–338doi:10.1007/978-981-15-3859-9_28.

Mohd, T., Jamil, S., Masrom, S., 2020b. Machine learning building price prediction with green building determinant. *IAES International Journal of Artificial Intelligence* 9, 379. doi:10.11591/ijai.v9.i3.pp379-386.

Morano, P., Tajani, F., 2013. Bare ownership evaluation. hedonic price model vs. artificial neural network. *International Journal of Business Intelligence and Data Mining* 8, 340–362. doi:10.1504/IJBIDM.2013.059263.

Morano, P., Tajani, F., Torre, C.M., 2015. Artificial intelligence in property valuations: an application of artificial neural networks to housing appraisal. *Advances in Environmental Science and Energy, Planning* , 23–29.

Neal, R.M., 1997. Monte carlo implementation of gaussian process models for bayesian regression and classification. arXiv preprint physics/9701026 .

Neal, R.M., 2012. Bayesian learning for neural networks. volume 118. *Springer Science & Business Media*.

Nghiep, N., Al, C., 2001. Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research* 22, 313–336. doi:10.1080/10835547.2001.12091068.

Pai, P.F., Wang, W.C., 2020. Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences* 10, 5832. doi:10.3390/app10175832.

Panagiotidis, T., Printzis, P., 2016. On the macroeconomic determinants of the housing market in greece: A vecm approach. *International Economics and Economic Policy* 13, 387–409. doi:10.1007/s10368-016-0345-3.

Paris, S.D., 2008. Using artificial neural networks to forecast changes in national and regional price indices for the UK residential property market. *University of South Wales (United Kingdom)*.

Park, B., Bae, J.K., 2015. Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications* 42, 2928–2934. doi:10.1016/j.eswa.2014.11.040.

Peterson, S., Flanagan, A., 2009. Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research* 31, 147–164. doi:10.1080/10835547.2009.12091245.

Piao, Y., Chen, A., Shang, Z., 2019. Housing price prediction based on cnn, in: *2019 9th International Conference on Information Science and Technology (ICIST)*, IEEE. pp. 491–495. doi:10.1109/ICIST.2019.8836731.

Picchetti, P., 2017. Hedonic residential property price estimation using geospatial data: a machine-learning approach. *Instituto Brasileiro de Economia* 4.

Plakandaras, V., Gupta, R., Gogas, P., Papadimitriou, T., 2015. Forecasting the us real house price index. *Economic Modelling* 45, 259–267. doi:10.1016/j.econmod.2014.10.050.

Quigley, J.M., 1999. Real estate prices and economic cycles. *International Real Estate Review* 2, 1–20.

Rafiei, M.H., Adeli, H., 2016. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management* 142, 04015066. doi:10.1061/(ASCE)CO.1943-7862.0001047.

Rafiei, M.H., Adeli, H., 2018. Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of construction engineering and management* 144, 04018106. doi:10.1061/(ASCE)CO.1943-7862.0001570.

Rahman, S.N.A., Maimun, N.H.A., Razali, M.N.M., Ismail, S., 2019. The artificial neural network model (ann) for malaysian housing market analysis. *PLANNING MALAYSIA* 17. doi:10.21837/pm.v17i9.581.

Rico-Juan, J.R., de La Paz, P.T., 2021. Machine learning with explainability or spatial hedonics tools? an analysis of the asking prices in the housing market in alicante, spain. *Expert Systems with Applications* 171, 114590. doi:10.1016/j.eswa.2021.114590.

Sarip, A.G., Hafez, M.B., Daud, M.N., 2016. Application of fuzzy regression model for real estate price prediction. *Malaysian Journal of Computer Science* 29, 15–27. doi:10.22452/mjcs.vol29no1.2.

Seo, K., Salon, D., Kuby, M., Golub, A., 2019. Hedonic modeling of commercial property values: distance decay from the links and nodes of rail and highway infrastructure. *Transportation* 46, 859–882. doi:10.1007/s11116-018-9861-z.

Seya, H., Shiroi, D., 2021. A comparison of residential apartment rent price predictions using a large data set: Kriging versus deep neural network. *Geographical Analysis* doi:10.1111/gean.12283.

Shahhosseini, M., Hu, G., Pham, H., 2019. Optimizing ensemble weights for machine learning models: a case study for housing price prediction, in: *INFORMS International Conference on Service Science, Springer*. pp. 87–97. doi:10.1007/978-3-030-30967-1_9.

Silver, M., Goode, M., 1990. Econometric forecasting model for rents in the british retail property market. *Omega* 18, 529–539. doi:10.1016/0305-0483(90)90009-X.

Štubňová, M., Urbaníková, M., Hudáková, J., Papcunová, V., 2020. Estimation of residential property market price: Comparison of artificial neural networks and hedonic pricing model. *Emerging Science Journal* 4, 530–538. doi:10.28991/esj-2020-01250.

Taffese, W.Z., 2007. Case-based reasoning and neural networks for real estate valuation, in: *Artificial Intelligence and Applications*, pp. 98–104.

Tatari, O., Kucukvar, M., 2011. Cost premium prediction of certified green buildings: A neural network approach. *Building and Environment* 46, 1081–1086. doi:10.1016/j.buildenv.2010.11.009.

Terregrossa, S.J., Ibadi, M.H., 2021. Combining housing price forecasts generated separately by hedonic and artificial neural network models. *Asian Journal of Economics, Business and Accounting* , 130–148doi:10.9734/ajeba/2021/v21i130345.

Tonelli, M., Cowley, M., Boyd, T., 2004. Forecasting office building rental growth using a dynamic approach. *Pacific Rim Property Research Journal* 10, 283–304.

Valier, A., 2020. Who performs better? avms vs hedonic models. *Journal of Property Investment & Finance* doi:10.1108/JPIF-12-2019-0157.

Wang, L., Chan, F.F., Wang, Y., Chang, Q., 2016. Predicting public housing prices using delayed neural networks, in: *2016 IEEE Region 10 Conference (TENCON), IEEE*. pp. 3589–3592. doi:10.1109/TENCON.2016.7848726.

Wang, X., Wen, J., Zhang, Y., Wang, Y., 2014. Real estate price forecasting based on svm optimized by pso. *Optik* 125, 1439–1443. doi:10.1016/j.ijleo.2013.09.017.

Wei, Y., Cao, Y., 2017. Forecasting house prices using dynamic model averaging approach: Evidence from china. *Economic Modelling* 61, 147–155. doi:10.1016/j.econmod.2016.12.002.

West, T., Worthington, A.C., 2006. Macroeconomic risk factors in australian commercial real estate, listed property trust and property sector stock returns. *Journal of Financial Management of Property and Construction* doi:10.1108/13664380680001083.

Williams, C., Rasmussen, C., 1995. Gaussian processes for regression. *Advances in neural information processing systems* 8.

Wilson, I.D., Paris, S.D., Ware, J.A., Jenkins, D.H., 2002. Residential property price time series forecasting with neural networks, in: *Applications and Innovations in Intelligent Systems IX. Springer*, pp. 17–28. doi:10.1007/978-1-4471-0149-9_2.

Wu, C.H., Li, C.H., Fang, I.C., Hsu, C.C., Lin, W.T., Wu, C.H., 2009. Hybrid genetic-based support vector regression with feng shui theory for appraising real estate price, in: *2009 First Asian Conference on Intelligent Information and Database Systems, IEEE*. pp. 295–300. doi:10.1109/ACIIDS.2009.41.

Xu, L., Li, Z., 2021. A new appraisal model of second-hand housing prices in china's first-tier cities based on machine learning algorithms. *Computational Economics* 57, 617–637. doi:10.1007/s10614-020-09973-5.

Yan, Y., Wei, X., Hui, B., Yang, S., ZHANG, W., Hong, Y., WANG, S.y., 2007. Method for housing price forecasting based on tei@ i methodology. *Systems Engineering-Theory & Practice* 27, 1–9. doi:10.1016/S1874-8651(08)60047-2.

Yan, Z., Zong, L., 2020. Spatial prediction of housing prices in beijing using machine learning algorithms, in: *Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence*, pp. 64–71. doi:10.1145/3409501.3409543.

Yasnitsky, L.N., Yasnitsky, V.L., Alekseev, A.O., 2021. The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time. *Complexity* 2021. doi:10.1155/2021/5392170.

Yoo, S., Im, J., Wagner, J.E., 2012. Variable selection for hedonic model using machine learning approaches: A case study in onondaga county, ny. *Landscape and Urban Planning* 107, 293–306. doi:10.1016/j.landurbplan.2012.06.009.

Zhang, P., Hu, S., Li, W., Zhang, C., Yang, S., Qu, S., 2021. Modeling fine-scale residential land price distribution: An experimental study using open data and machine learning. *Applied Geography* 129, 102442. doi:10.1016/j.apgeog.2021.102442.