Website: The Tesseract Academy

# Testing Land Valuation Methodologies

**Lot 1: Market-based statistical valuation**

**Lot 2: Advanced algorithmic and machine-learning applications**

**Lot 3: Formula-based valuation by land area**

**Lot 4: Conventional valuation approaches**

**Lot 5: Innovative or experimental approaches**

# Executive Summary

## Introduction and background

This summary presents the findings from a comprehensive evaluation of five land-valuation methodologies applied to the Welsh context. This analysis utilises a large-scale land dataset comprising approximately one and a half million land-use classified parcels, of which 85% are residential. Residential stock includes Detached, Semi-Detached, Terraced, and Flats, which collectively provide a dense and methodologically tractable baseline for testing valuation methodologies.

Beyond residential land, the dataset contains a diverse set of non-residential land-use categories:

- Agricultural and natural land (13.2%), including:
  - Farmland – 114,459 parcels
  - Meadow – 72,163 parcels
  - Forest/Woodland – 2,870 parcels
  - Grassland – 1,467 parcels
  - Farmyard sites- 1,144
- Commercial and industrial land, including:
  - Retail – 6,737 parcels
  - Industrial – 2,277 parcels
  - Commercial (other) – 2,001 parcels

The methodological focus on the residential subset serves primarily as a baseline. Residential land is methodologically more tractable for valuation, with standardised attributes, dense market comparables, and relatively homogeneous characteristics, thereby providing a controlled environment for testing the relative performance, scalability, interpretability, and policy relevance of each approach. Crucially, the purpose of this analysis is not to limit valuation to residential land, but to assess how well these methodologies can generalise across the full spectrum of land types. The framework already incorporates satellite-derived parcel area data and land-use classification features, enabling preliminary decomposition of land values for commercial, industrial, agricultural, and mixed-use categories. The residential results therefore function as a methodological benchmark to understand where each method adds value, where it encounters limitations, and what adaptations would be required to extend the models across the diverse non-residential portfolio.

Land transactions with large parcels exceeding 0.5 hectares, present particular challenges for land-value decomposition, requiring specialised treatment of economies of scale, zoning constraints, and use-specific depreciation schedules. These dimensions are already partially captured through satellite-derived features but require further calibration for operational deployment across all land categories.

The study compares performance, scalability, interpretability, and policy relevance across a diverse set of methodological families:

- Traditional statistical models (Model 1)

    - Ridge Regression
- Machine Learning models (Model 2)
    - CatBoost
    - Gradient Boosting
- Expert-driven systems (Model 3)
    - K-Nearest Neighbours
    - “Fuzzy Matching”
- Formula-based economic methods (Model 4)
    - Depreciated Replacement Cost
- Advanced Agentic AI techniques (Model 5)
    - Claude Multi-Agent Valuation System

The overarching aim is to identify a valuation approach capable of supporting Wales-wide, repeatable, and defensible land-value estimation at granular geographic levels, that is, Lower Layer Super Output Area (LSOA) and below.

## Key Findings

Across all modelling approaches, performance was assessed using a test set of 9,606 residential land transactions drawn from nine Lower Layer Super Output Areas (LSOAs) across Wales. The five methodological families evaluated were: Model 1 (Stacked Ridge Regression ensemble), Model 2 (CatBoost Gradient Boosting), Model 3 (KNN comparable-sales automation / “fuzzy matching”), Model 4 (Depreciated Replacement Cost, DRC), and Model 5 (Multi-agent LLM ensemble). Among these, Model 2 (Gradient Boosting) achieved the strongest overall predictive results. However, its current accuracy remains below the level required for operational deployment in statutory or regulatory valuation settings.

Model 1 – Stacked Ridge Regression ensemble (market-based statistical approach) performed substantially better than a simple global Ridge baseline, demonstrating the value of a structured statistical ensemble even when predictive performance remained below operational requirements. Within the nine test LSOAs, within-LSOA $R^2$ ranged from 1.7% to 51.5% (average within-LSOA $R^2$ of 26.1%), with a mean absolute error of £69,646. The stacking architecture (5 property-type-specific models, 1 global model, and a meta-learner) showed clear benefits from model specialisation: the meta-learner assigned weights of approximately 1.05 to each property-type model while negatively weighting the global baseline (-0.025). Performance was strongest in rural and semi-rural markets (Powys: 51.5%, Bridgend: 46.3%, Monmouthshire: 42.8%) but much weaker in the large, heterogeneous urban market of Cardiff ($R^2$ = 1.7%). Model 1’s key strength remains interpretability and auditability, although its linear structure limits its ability to capture the non-linear interactions better handled by Model 2.

Model 2 – Gradient Boosting (CatBoost; machine-learning approach) delivered the best overall performance across the five model families. In its current configuration, it exhibited $R^2$ values ranging from 0.002 to 0.52 across the nine test LSOAs. In practical terms, this indicates moderate but still insufficient accuracy for operational valuation use, with substantial performance variation across different geographic contexts.

Model 3 – KNN comparable-sales automation ("fuzzy matching"; conventional comparables approach) did not meet the accuracy thresholds required for operational use. This model, which automates a nearest-neighbour comparables logic, performed substantially below the statistical and machine-learning approaches (Models 1 and 2). Prediction errors were typically large (often exceeding half of the land value), the method struggled to generalise across held-out LSOAs, and it produced unstable land-structure splits. These characteristics mean Model 3 is not currently reliable enough for valuation decision-making in its present form.

Model 4 – Depreciated Replacement Cost (DRC; formula-based approach) remained important from a methodological and policy perspective because it is the only fully transparent, rule-based approach in the study capable of explicitly decomposing total value into land and structure components. However, Model 4 (DRC) did not achieve predictive accuracy comparable to the data-driven models (Models 1 and 2). Across the test set, its explanatory power was negative (i.e., worse than a mean predictor), and its typical percentage error exceeded 100%. Although unsuitable as a standalone predictive valuation model, Model 4 still provides a useful conceptual and regulatory benchmark due to its interpretability and auditability.

Model 5 – Multi-agent LLM ensemble (innovative/experimental approach) produced encouraging but inconsistent results relative to the leading machine-learning model. Across the nine test LSOAs, Model 5 recorded $R^2$ values ranging from -0.02 to 0.46. This suggests the method has real capacity to interpret qualitative and contextual information, but its predictive accuracy was less stable and less reliable overall than Model 2 (Gradient Boosting).

When comparing performance patterns across geography and price levels, the same broad pattern appeared across all five models (Models 1–5). The strongest results tended to occur in areas with higher transaction density and more homogeneous land/property characteristics, while weaker performance was observed in areas with sparse data, greater heterogeneity, or atypical local market structure. High-value transactions and unique assets were consistently challenging for every modelling family, including the best-performing Model 2.

Taken together, the evaluation shows that Model 2 (Gradient Boosting / CatBoost) is the strongest candidate among the five approaches tested for empirical prediction, while Model 1 (Stacked Ridge) offers a valuable interpretable benchmark, Model 4 (DRC) remains the key transparent decomposition method, and Models 3 and 5 are currently better viewed as exploratory or supporting approaches rather than deployment-ready solutions. Even so, none of the models has yet reached the accuracy and consistency required for national deployment in formal valuation or taxation contexts. Improving performance in rural, high-value, and atypical market segments should be a priority in any future development phase.

**Land value as a proportion of total land value**

Using the Depreciated Replacement Cost (DRC) approach (Model 4) to separate total land value into land and structure components offers a transparent, rule-based framework that is fundamentally different from statistical and machine-learning techniques. DRC remains the only method in this study capable of producing an explicit, theoretically grounded land structure decomposition. However, when applied to the Welsh housing stock, its empirical performance demonstrates that it cannot be used as a standalone approach for estimating market values.

Across the nine test LSOAs, the DRC model exhibited $R^2$ values ranging from -0.79 to -0.01, indicating performance consistently below that of a simple mean predictor. The mean percentage error exceeded 100%, demonstrating that typical predictions deviated from actual prices by more than the land's full value. This pattern is consistent with the broader finding that the DRC formula does not adequately capture the heterogeneity of housing conditions, market demand, or localised price levels.

Because of these limitations, the results do not allow for a reliable or meaningful estimate of market values based on DRC alone. The overall predictive performance shows that the DRC method is not suited to estimating market-consistent land values in Wales.

Nevertheless, the method retains important value within a wider analytical framework. In particular, it provides a transparent conceptual basis for understanding how land and structures contribute to total land value, supporting auditability and interpretability in a way that complements empirical approaches. The DRC method can give a broad indication of relative land value patterns across different settlement types, particularly when examined alongside spatial variation in building age, condition, and market activity.

These strengths become clearer when considered alongside the empirical modelling results. Gradient Boosting (Model 2), with $R^2$ values ranging from 0.002 to 0.52 across the nine test LSOAs, provides the most reliable evidence for total market value in this study. While its accuracy remains insufficient for operational deployment, it offers a substantially more faithful reflection of market outcomes than the DRC formula. By combining the DRC method's transparent land structure decomposition with the empirical estimates generated by Gradient Boosting or other machine learning models, there is scope for developing hybrid approaches that balance interpretability with improved predictive accuracy.

For any potential policy use, the DRC method would therefore require careful handling, including reporting of both raw and adjusted statistics for transparency, and restricting its use to decomposition ratios rather than absolute valuations. Under these conditions, DRC can support policy development, internal modelling, and long-term monitoring, while total valuation exercises should continue to be grounded primarily in empirical models such as Gradient Boosting.

**National scalability and geographic coverage**

The modelling framework is able to generate land-level valuations for the land-use classified parcels contained in the dataset. Residential land transactions make up the majority of records, with 1,250,600 parcels (85.8%) classified as detached, semi-detached, terraced or flats. This extensive residential coverage provides the core empirical foundation for model development, reflecting both the composition of the Welsh housing market and the structure of the underlying transaction and spatial datasets. The density and geographic breadth of residential observations offer a robust basis for training, validating and comparing valuation methodologies

Beyond the residential sector, the dataset contains a diverse set of non-residential land-use categories which together account for the remaining share of Welsh land parcels. A substantial proportion of these fall within agricultural and natural land uses (13.2%), including 114,459 parcels of farmland, 72,163 parcels of meadow, 2,870 parcels of woodland, 1,467 grassland parcels, and 1,144 farmyard sites. These land types tend to exhibit markedly different spatial and economic characteristics from residential land, including substantially larger parcel sizes and distinctive market dynamics driven by agricultural suitability, environmental context and land management practices.

A smaller but economically significant segment, approximately 0.8% of all parcels, is composed of commercial and industrial land uses, incorporating 6,737 retail sites, 2,277 industrial premises, and 2,001 commercial land transactions. These assets are predominantly concentrated in towns and cities, where zoning, accessibility and local economic activity exert a greater influence on land values than physical characteristics alone.

Building floor area, derived from Energy Performance Certificate (EPC) records, provides an important source of information for understanding the relationship between building size and land value. The current framework therefore relies on floor area, which reflects the internal floor space of the building itself rather than the extent of the land parcel on which it sits.

Directly linked EPC floor area is 37.8% across the full dataset. Among land transactions for which floor area data are available, building sizes display remarkably similar distributions across land-use categories. Both residential and non-residential land transactions have a median floor area of around 86 square metres, with mean values of approximately 94 square metres. This lack of differentiation is a consequence of the underlying dataset rather than a reflection of the built environment itself. A large share of land transactions classified as “non-residential” by land-use context are, in fact, residential dwellings situated within agricultural, commercial or industrial zones. By contrast, genuinely large commercial or industrial buildings are extremely rare in the dataset: only 26 non-residential land transactions have a floor area greater than 500 square metres. Overall floor area ranges from very small structures of 1 square metre to a maximum of 8,412 square metres, although the distribution is highly concentrated around typical residential scales, with an interquartile range of roughly 71 to 107 square metres.

Within the modelling framework, floor area plays a significant role in estimating structure value, particularly for methods such as the Depreciated Replacement Cost (DRC) approach (Model 4) and for residual-based analyses that infer land value after accounting for building characteristics. However, the limited availability of floor area data constrains its use for comprehensive land value decomposition and highlights the need for improved data coverage if the framework is to support future land valuation work at national scale.

To assess the spatial consistency of model performance, the analysis makes extensive use of Lower Layer Super Output Areas (LSOAs). The dataset spans 1,912 LSOAs with at least one transaction in the final dataset, with a median of 724 land transactions per LSOA, allowing for robust spatial validation across urban, suburban and rural contexts. This geographic granularity is particularly important when assessing the applicability of the modelling framework to non-residential land uses, which are unevenly distributed and often concentrated in specific industrial, agricultural or commercial zones. LSOA-level metrics therefore offer critical insight into whether models trained primarily on residential data are capable of generalising to the distinct spatial and economic environments in which non-residential land values are formed.

A key component of the spatial analysis is the incorporation of Agricultural Land Classification (ALC) data, which intersects approximately 55.2% of land transactions in the dataset. However, 88.1% of matched records are classified as Grade U (Urban), leaving only 6.6% of all land transactions with agricultural grades one to five. These graded transactions remain important for understanding rural land values and land-use patterns and provide an evidence base for modelling non-urban land markets.

This multi-source pipeline is designed to be repeatable and updatable. As new transaction records, EPC certificates and spatial datasets are released, the framework can be refreshed automatically, ensuring that land and land valuations remain current and that the system can be scaled to support long-term monitoring, strategic planning and policy development across Wales.