TESTING LAND VALUATION METHODOLOGIES

FINAL REPORT

Lot 1: Market-based statistical valuation
Lot 2: Advanced algorithmic and machine-learning applications
Lot 3: Formula-based valuation by land area
Lot 5: Innovative or experimental approaches

Dr Rhys ap Gwilym, Prof Adrian Gepp, Dr Sadeque Hamdan, Dr Edward Jones, Pretty Karibo, Dr Graeme Pearce, Temidayo Popoola, Dr Xiaoxi Qu

Bangor University

March 2026

EXECUTIVE SUMMARY

The Welsh Government commissioned this project to test, compare and understand a range of methodologies for estimating land values in Wales. The work does not attempt to identify a single "true" value for land. Instead, it assembles the most comprehensive Wales-wide land, property and amenity datasets yet brought together, applies multiple modelling approaches at national and local levels, and examines how members of the public respond to different valuation methods. The findings show what can be achieved with today's data, what cannot, and where future effort would most usefully be directed.

WHAT WE DID

We developed and assessed four complementary strands:

- **Lot 1 – Hedonic regression.** This approach uses large numbers of past property sales to estimate how different characteristics — such as location, plot size, property type, proximity to services and local amenities — contribute to observed sale prices. Hedonic models are well-established and widely used in economic analysis, and they provide a structured way to separate the value of land from the value of buildings**.**

- **Lot 2 – Machine-learning modelling.** This strand applies more flexible and modern modelling techniques capable of capturing complex, non-linear relationships that traditional models may miss. These methods generally deliver the highest predictive accuracy for overall property, but their internal workings can be less intuitive to interpret.

- **Lot 3 – Formula-based approaches.** This strand reviews how other countries develop and maintain land-value systems, and tests simplified formula-based approaches inspired by international practice. These systems emphasise predictability, stability and ease of communication. They typically assign values using area-based factors or zoning rules rather than detailed parcel-level modelling. They are best viewed as simplified frameworks, not stand-alone valuation tools.

- **Lot 5 – Behavioural evidence.** Lot 5 explores how members of the public interpret and respond to different valuation methods. Using an interactive dashboard and small financial incentives, participants compared parcel-level outputs, examined underlying data, and expressed preferences among modelling approaches. The findings suggest that participants perceive a trade-off between accuracy and interpretability of land valuation methods. This strand highlights the importance of public acceptability and transparency if a valuation system were ever to be used in practice.

All lots draw on a unified data backbone created for the project: a Transactions Database (1.28 million Wales transactions, scaled to 2025 values), a National Land Parcel Database (1.41 million polygons with spatial attributes), and an LSOA Land Parcel Database for nine diverse case-study LSOAs. Together they enable like-for-like comparisons across methods and geographies.

KEY FINDINGS

Data is the binding constraint; modelling is secondary

- The project successfully fuses multiple data sources and constructs new spatial attributes. Nevertheless, critical gaps and inconsistencies remain – parcel geometries are problematic, coverage is incomplete, matching properties across datasets is challenging and there is a lack of data on key variables such as the planning system and the quality of amenities.

- These issues limit all modelling approaches. They explain why methods that should, in principle, be highly accurate still fall short of "policy-grade" precision for many uses. Improving the core data spine will yield larger gains than switching modelling techniques.

Three modelling approaches; three distinct strengths

- **Hedonic (Lot 1).** Provides an explicit, economically coherent decomposition of land and structure. Predictive performance is respectable for residential transactions, with coefficients that behave sensibly and a statistically insignificant intercept – an important signal that the assembled land-related variables are doing real work rather than relying on a baseline constant.

- **Machine learning (Lot 2).** Delivers the lowest prediction errors for total property values and captures important non-linearities and interactions that linear models cannot. It demonstrates what best-possible accuracy looks like given today's data, and produces spatial results that closely mirror Lot 1 at aggregate scales.

- **Formula-based (Lot 3).** Intentionally simple and communicable. Produces plausible relative spatial patterns using MSOA fixed effects, but requires external calibration to estimate absolute land values.

Across Lots 1–3, the broad geography of land values is consistent: the lowest values are concentrated in the south Wales valleys and other post-industrial areas; the highest values are found in parts of Cardiff, Monmouthshire, Swansea and selected coastal or amenity-rich pockets. This convergence suggests the underlying spatial structure of land value in Wales is robust to modelling choices, even if levels differ.

Decomposition is feasible in all models, but never incontrovertible
All three modelling approaches decompose land and structure by design. However, because "pure land value" is rarely and unreliably observed in the market, no decomposition can be uniquely validated. The philosophy matters: land and structures are complementary; separability is a modelling assumption, not an empirical truth. Lot 2's strongest performers – allowing richer interactions – reinforce this point.

Public preferences emphasise understandability
Lot 5 demonstrates that participants consistently preferred the hedonic model, followed by machine learning and then formula-based approaches. Crucially, these preferences do not change under different financial incentives. This reinforces the importance of approaches that people feel they can understand and that appear grounded in recognisable logic, even if those approaches are not the most statistically accurate.

WHAT THE MAP SHOWS (LOT 2)

Figure 1 presents average estimated land values per parcel in the National Land Parcel Database by LSOA using our best-performing Lot 2 configuration. It is included because machine-learning methods achieved the strongest predictive accuracy for property values and produce spatial variation that is very similar to the hedonic model. The map should be read as a plausible, data-constrained portrait of Wales' land-value geography – not a definitive measure of the "true" value of land.

CROSS-CUTTING LESSONS

1. **Land value is a construct, not an observed quantity.** Without systematic sales of unimproved land, all land values are model-based inferences. Different policy purposes imply different conceptual definitions (residual for insurance; social/location value for fiscal instruments; development potential for planning).

2. **Modelling discretion is unavoidable – its location differs.** Hedonic models require explicit choices about functional form and variables; machine-learning models automate many such choices inside algorithms; formula-based approaches place discretion with the formula's architects and calibration choices. No approach "removes" judgement.

3. **Data quality dominates method choice.** With better core data (parcel–property linkages, planning designations, amenity quality, non-residential structures), all approaches improve—especially their usability for policy.

4. **Spatial patterns are robust, levels are not.** Methods agree on where values are high or low; they differ more on how high or how low.

5. **Public legitimacy depends on intelligibility.** Understandable methods, transparent assumptions and visible data quality improve acceptability.

WHAT THIS MEANS FOR POLICY AND IMPLEMENTATION

- **Feasibility today.** All three modelling approaches are operationally feasible (we implemented each), but none delivers accuracy that would be acceptable for high-stakes statutory uses across Wales with current data.

- **Role of models.** Methods can be combined:
    - Hedonic models offer explicit control and a coherent decomposition.
    - Machine-learning models are well-suited to benchmarking, diagnostics and revealing interactions; they may be central if the chosen definition of land value requires non-linear complementarities to be modelled explicitly.
    - Formula-based outputs can support communication and stability when calibrated by richer models.

- **Institutional data spine.** The primary constraint is not modelling but data governance. Fragmented responsibilities and inconsistent identifiers make reliable parcel-to-property-to-attribute linkage difficult. Any durable valuation system will require a more coherent Welsh land information infrastructure.

- **Public engagement as a data asset.** The Lot 5 dashboard shows people can and will help validate and correct local parcel information. A controlled feedback mechanism could provide a distributed quality-assurance layer for Welsh land and property data.

SUMMARY

This project shows that land valuation in Wales is both possible and inherently constrained. The modelling work demonstrates clear progress on previous studies and strong internal coherence, but the accuracy of estimated land values is limited by data that is incomplete, inconsistent and not designed for the purpose.

At the same time, the project reveals stable spatial patterns across methods and provides new insight into how people perceive the legitimacy of valuation approaches.

Further progress will depend on clarifying the policy purpose of "land value", improving the core data infrastructure, and governing modelling choices transparently – ideally with citizens helping to validate the data that describe their own places.

Figure 1: Lot 2 average land values per parcel in the National Land Parcel Database by LSOA
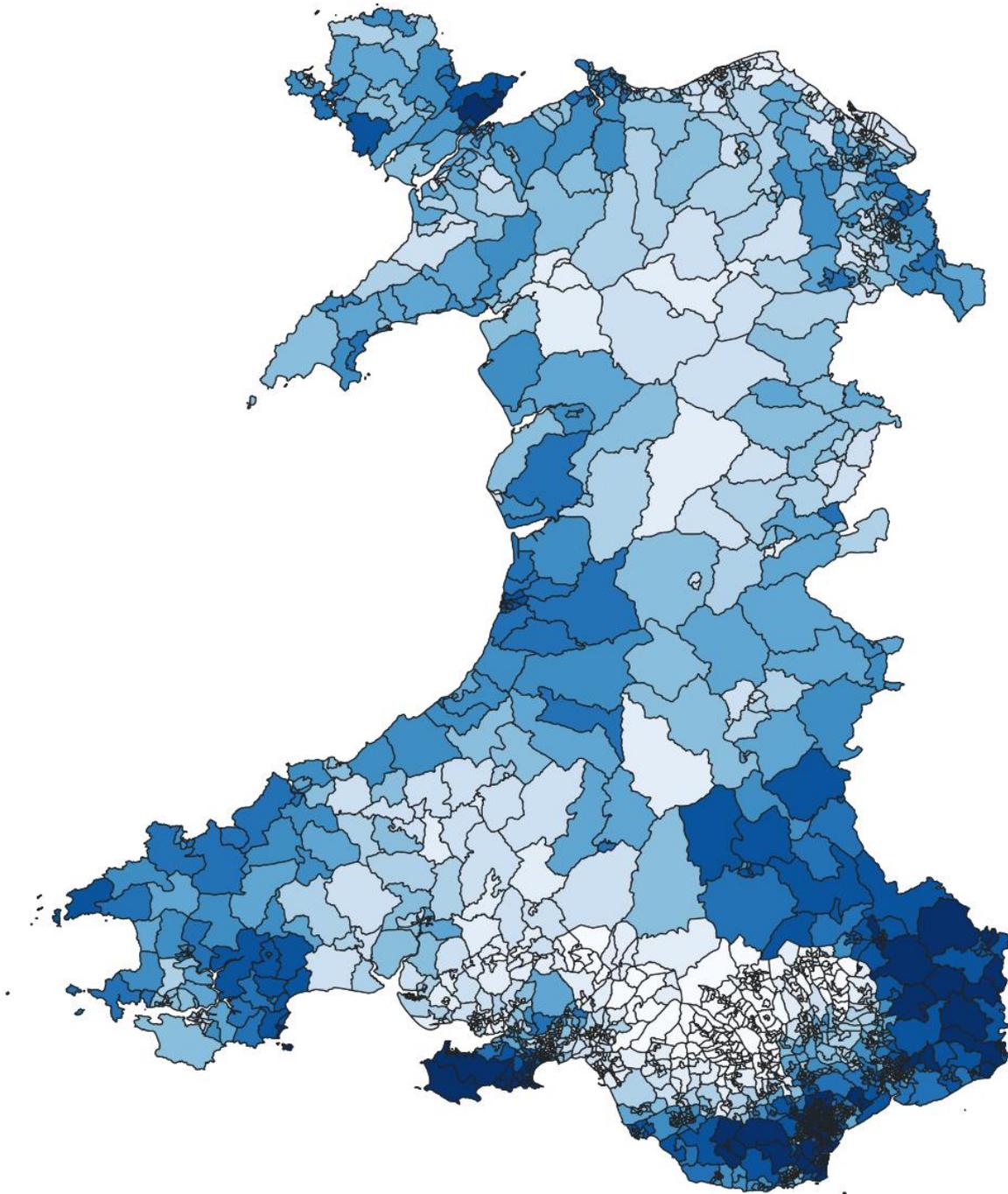
Darker shades of blue represent higher land values per parcel.

(Reproduces Figure 3.2.2 in the main report.)

Links to interactive, scrollable maps: Lot 1, Lot 2, Lot 3.

Links to maps at the parcel-level for the nine identified LSOAs: Lot 1, Lot 2, Lot 3.