INDUSTRYLINE RESEARCH

Report:

Testing Land Valuation Methodologies:

Lot 3: Formula-based valuation by land area

Lot 5: Innovative or experimental approaches

Combined Report

March 2026

INDUSTRYLINE **RESEARCH**

Driving industries forward with market intelligence, Industryline Research is a specialist provider of business intelligence and market research within the housing and education sectors. Leveraging innovative data solutions, expert analysis and established methodologies, Industryline Research delivers innovative intelligence that drives growth and enables you to make informed strategic decisions backed by solid data.

CONTENTS

Table of Figures

Table of Tables

Glossary

ACCESS (Formula 3; Rung 3): The third step in the Lot 3 “Complexity Ladder”, introducing continuous penalties for accessibility and constraints to enable within-area differentiation.

Administrative comparisons: Aggregate “reasonableness” checks against known patterns or administrative evidence, used to assess plausibility without presenting the comparison as definitive validation.

Analytical parcel (analytic parcel): The polygon unit used for modelling. It is an agreed evidence unit used for attaching attributes and running formulae; it is not a legal title parcel.

Analysis lock window: A defined point at which data collection is treated as complete and the dataset is “locked” for analysis to preserve integrity and auditability.

Anchoring: A behavioural mechanism where judgements are shaped by an initial reference point (explicit or implicit), influencing value estimates and rankings.

ATAF (Automated Transactions and Analysis Framework): The programme’s structured workflow for transaction ingestion, standardisation, QA, spatial linkage, and analysis, enabling consistent and auditable outputs.

Attachment integrity: A QA check confirming that required attributes/features are successfully attached to each parcel (i.e., no silent loss of coverage through failed joins or missing attachments).

Audit boundary: The defined scope of what is included in audited outputs and therefore what can be claimed from the published artefacts (e.g., which parcels, which rungs, which inputs, and which extracts).

Auditable pipeline: A delivery approach where outputs are reproducible and defensible because execution is deterministic, logged, and traceable (e.g., run IDs, recorded parameters, and consistent staging).

BCIS (Building Cost Information Service): A source of construction cost indices referenced as potential evidence to support build-cost proxies within composite modelling where licensing allows.

Benchmarking hierarchy: A structured approach to comparing outputs against external evidence, with explicit coverage/definition statements to avoid over-claiming where evidence does not align to the analytic unit or scope.

BORIS / BORIS-D: German land value information systems/portals referenced as international examples of standardised land value reporting and evidence access.

Boundary-Line: A boundary dataset used for statistical and administrative geographies (e.g., LSOA boundaries) that define the study frame for analysis.

Cadastre (cadastral): A stable, authoritative parcel unit aligned to legal boundaries. It provides the strongest unit definition for defensible parcel-level valuation.

Calibration provenance: The documented lineage of how parameters were set (including sources, dates, and sensitivity work), enabling scrutiny and reproducibility.

Cap (capped display rate): A presentation control that limits mapped/displayed £/m² to avoid extreme values dominating the visual output; analytical reporting uses implied pre-cap rates and value totals.

Class attribution: Assigning each analytic parcel a land-use class used for scheduling and interpretation, including overlap-based assignment and fallback classification where required.

Clarity (Lot 5): How easy a method is to understand and apply, treated as distinct from perceived fairness/legitimacy.

Clipping: Restricting datasets to a boundary so analysis and reporting are confined to the study frame.

Code-Point: A postcode geocoding product/process used to locate records at postcode centroids for spatial linkage, noting known uncertainty introduced by centroid-based assignment.

Complexity Ladder: The staged hierarchy of formula “rungs” designed to make data dependency explicit and balance coverage with interpretability (FLAT, FLAT_RUR, ACCESS, COMP).

COMP (Formula 4; Rung 4): A prototype composite regression surface calibrated using transaction-linked evidence; treated as diagnostic/prototype rather than publication-ready in this programme.

Completion rules: Pre-defined criteria for what constitutes a complete response/record (and how partial completions are treated), supporting defensible interpretation.

Concentration diagnostics: Statistics describing how much total value is dominated by a small number of parcels (e.g., top parcel share, top ten share) alongside distribution metrics.

Confidence (Lot 5): A self-reported measure of how confident participants feel about their valuations, analysed separately from agreement/dispersion.

Coverage: The proportion of the study area (within the LSOA boundary) represented by the selected analytic parcel fabric and therefore eligible to receive model outputs; not synonymous with accuracy.

Coverage reconciliation layer: Reporting that quantifies, for each geography, the relationship between total area, valued parcel area, and explicit exclusions, including White Space.

CRS (Coordinate Reference System): The coordinate framework used to store and process spatial data. Consistent CRS use is required for valid area and distance calculations.

Data completeness gate (coverage gate): A stop/go QA control that blocks downstream valuation if required inputs do not fully cover the area being valued (for example, missing tiles in gridded fallback inputs).

Deduplication: Removing duplicate geometries/records in the parcel fabric to prevent double-counting and join artefacts.

Deterministic execution: Running the pipeline such that the same inputs and parameters produce the same outputs every time, supporting reproducibility and defensibility.

Dispersion (Lot 5): The spread of valuations/judgements across participants for the same parcel/task, treated as a standard output and diagnostic signal rather than "noise".

Distance-to-network metrics: Measures of proximity (in metres) from parcels to features such as primary roads or rail, computed in an appropriate CRS for distance work.

Driver decomposition (trigger fields): Parcel-level variables that would allow explanation of what drove a specific parcel's movement between rungs (e.g., which penalty/trigger applied).

EPSG:27700 (British National Grid): The CRS used for area and distance calculations in the analysis.

EPSG:4326 (WGS 84): A geographic CRS (latitude/longitude) used by some inputs, typically transformed for distance/area calculations.

ESA (European Space Agency): Organisation referenced as the provider of WorldCover.

Evidence lock protocol: Governance steps that define completion, dataset structure, and the analysis lock window, so outputs are auditable and defensible.

Exclusions (explicit): Surface types excluded by design, affecting interpretation of map “gaps” and coverage.

Fabric integrity: The soundness and consistency of parcelisation and feature attachment (valid geometries, controlled overlaps/gaps, stable joins).

Feasibility: The degree to which an approach can be implemented in practice under real-world constraints (data, operational burden, scalability, and governance).

FLAT (Formula 1; Rung 1): Baseline class schedule/rate card used as the controlled reference state.

FLAT_RUR (Formula 2; Rung 2): Discrete rural/low-access adjustment applied as stepped multipliers/flags, producing area-wide differentiation.

Flood exposure share (flood_share): The proportion of parcel area intersecting flood zones, used as a continuous constraint feature.

Gini coefficient: A summary measure of inequality/concentration in a distribution (here, indicating how uneven parcel values are within an LSOA).

GML (Geography Markup Language): A spatial data format used for distributing polygon datasets.

Gridded / tiled inputs: Raster datasets provided as tiles (e.g., land-cover GeoTIFF tiles); missing tiles can create structural gaps unless detected and blocked by QA.

HM Land Registry: The public body providing datasets used in the programme (including INSPIRE Index Polygons and Price Paid Data).

Implied pre-cap rate: Analytical rate computed as value (£) divided by parcel area (m²), used for reporting rather than the capped map display rate.

INSPIRE Index Polygons (INSPIRE parcel fabric): The polygon dataset used as the analytic parcel fabric for the report-facing runs.

Join completeness: QA checks confirming each parcel has required attributes attached (e.g., class, features, distances) with no “join leakage”.

Legitimacy (Lot 5): Whether a method is perceived as acceptable, fair, and contestable in practice; distinct from clarity.

LSOA (Lower layer Super Output Area): A small-area statistical geography used as the reporting unit for the pilot.

Median: The middle value of a distribution (50th percentile), used as a robust summary of parcel rates/values.

Missingness (Lot 5): Absence of responses or inability/unwillingness to provide precise valuations; treated as diagnostic evidence rather than administrative noise.

NaPTAN (National Public Transport Access Nodes): Dataset of public transport access points used for proximity measures.

NLUD-aligned five-class scheme: Simplified land-use classification used for consistent modelling (Residential; Commercial/Industrial; Agricultural; Community/Amenity; Undeveloped).

NULL (residual NULLs): Unclassified parcels remaining after primary attribution; treated as a completeness risk and addressed via fallback mapping.

Open datasets / open-source tooling: Delivery principle emphasising transparency and reproducibility through official/open inputs and open-source tools, within licensing constraints.

OS (Ordnance Survey): The UK mapping agency and provider of multiple datasets used in the analysis.

OS OpenMap Local: An OS product used in early baseline work as a proxy polygon base; required additional preprocessing and delivered lower coverage for this purpose than the INSPIRE parcel fabric.

OS Open Roads / OS Open Rivers / OS Open Greenspace: OS open datasets used for proximity and contextual feature engineering.

Parcel fabric: The polygon layer defining the units over which values are produced; treated as a programme decision (coverage, rules, exclusions), not an implementation detail.

Parameter governance: Version-controlled management of model parameters (including calibration provenance, run identifiers, and sensitivity testing), supporting defensibility.

P10 (10th percentile): The value at or below which the lowest 10% of observations fall.

P90 (90th percentile): The value at or below which 90% of observations fall (i.e., the threshold above which the top 10% sit).

Penalty logic: Rules/functions that reduce values based on accessibility and constraints measures (continuous adjustments rather than discrete steps).

Place-based reasoning: Tendency to infer value from area reputation and perceived status of place, which shapes interpretation and valuation.

PPD (Price Paid Data): Transaction dataset used as evidence for calibration/benchmarking and for constructing transaction-linked proxies.

Publishability: The standard for whether outputs can be released as evidence, requiring QA pass, coverage sufficiency, and a clear interpretation boundary.

QA (Quality Assurance): Checks and thresholds ensuring geometries, joins, coverage reconciliation, and pipeline execution meet publication standards.

Raster: A gridded spatial data format (e.g., GeoTIFF), used here in fallback classification only.

Readiness tiering (tiered readiness framework): Treating formula rungs as dependent on increasing inputs, so higher-complexity outputs are produced only where prerequisites are met.

Residual land value proxy: An estimated land value derived by subtracting a building value proxy from transaction price; not a direct observation of "land value".

Run ID (run identifier): A unique identifier for a pipeline execution instance, used to trace outputs to a specific input set and parameter configuration.

Salience effects: Where highly noticeable features disproportionately shape judgement, influencing value estimates and rankings.

Sensitivity testing: Structured testing of how outputs change under parameter or definitional variations to bound interpretation and support defensibility.

Stop conditions (stop/go checks): Non-negotiable QA gates (e.g., coverage thresholds, geometry validity, CRS discipline, join completeness) that prevent publication or further processing if unmet.

Structural exclusions: Land excluded by design from valuation eligibility, affecting interpretation of visible map gaps.

TAN 15 (Technical Advice Note 15): Planning guidance referenced as the basis for flood risk mapping and associated constraints measures.

Tile extents inventory: A maintained record of which raster tiles cover the study area, used to detect missing tiles and enforce completeness gates.

Topology harmonisation: Geometry processing to improve internal consistency when polygon sources are not parcel-designed (e.g., attachment/dissolve and overlap handling).

Transaction-linked evidence: External evidence (e.g., transaction records) used to calibrate or benchmark models.

Uncertainty capture (Lot 5): Treating expressed uncertainty (ranges, confidence, unknowns) as an essential output for interpretation and governance, rather than forcing false precision.

Unit definition: The rule-set and dataset determining what constitutes a “parcel” in the evidence base.

Validity repair (geometry validity): Fixing invalid polygon geometries so area and overlap operations behave correctly and reproducibly.

Valuation-eligible land: Land included within the analytic base and therefore eligible to receive model outputs; distinct from excluded land and distinct from “zero value”.

Value density: Value expressed per unit area (e.g., £/m^2), used to support comparability across parcels of different size.

value_gbp: Parcel value output in pounds, used for analytical totals and aggregation.

WIMD (Welsh Index of Multiple Deprivation): Used via the Access to Services domain quartile as an area-level access feature.

WorldCover (ESA WorldCover 2021 v200): Land-cover dataset used only as a fallback to fill residual unclassified parcels after overlap-based classification.

Executive Summary

Purpose, scope and audited evidence boundaries

This executive summary for the Welsh Government's Testing Land Valuation Methodologies programme summarises the audited evidence from two distinct methodological strands delivered by Industryline Research: Lot 3 (formula-based valuation) and Lot 5 (participatory valuation workshops).

Both lots in this programme are research and methodological testing exercises. The outputs should not be interpreted as professional valuation advice for individual sites or transactions, and the report does not claim that any tested method replaces established valuation approaches. Instead, it assesses how different approaches operate in practice, what inputs they depend on, how outputs can be audited and explained, and what limitations those dependencies imply for any potential future application.

Study geographies and bases

Lot 3 analysis is reported for a bounded pilot geography comprising nine Lower-layer Super Output Areas (LSOAs) in Wales:

- Gwynedd 009D (W01000114)
- Flintshire 015A (W01000255)
- Powys 011C (W01000449)
- Ceredigion 002D (W01000517)
- Pembrokeshire 002F (W01000617)
- Bridgend 019D (W01001045)
- Rhondda Cynon Taf 001F (W01001233)
- Monmouthshire 006F (W01001597)
- Cardiff 032H (W01002019)

Lot 5 evidence is drawn from ten participatory valuation workshops, comprising eight public workshops and two professional workshops (63 participants in total).

Methods tested

Lot 3: Formula-based valuation by land area

Lot 3 operationalises land value as an indicative land value density signal (expressed as £/m²) applied across an analytical parcel fabric. The method was implemented as a staged 'Complexity Ladder' designed to increase data dependency and within-area differentiation as additional variables are introduced.

In the audited run, the ladder comprises four rungs: Formula 1 (FLAT) establishes a uniform baseline; Formula 2 (FLAT_RUR) introduces stepped, area-wide adjustments; Formula 3 (ACCESS) incorporates continuous penalties for accessibility and constraints; and Formula 4 (COMP) is a prototype composite model anchored to transaction evidence and treated explicitly as diagnostic rather than policy-ready. All Lot 3 findings are reported for the audited export boundary and, unless explicitly stated otherwise, relate to the INSPIRE polygon base formula runs and associated outputs. The report prioritises value density and distribution (for example, median £/m² and percentile ranges) over absolute totals, because totals are model-implied aggregates and can be dominated by a small number of very large parcels.

‘White Space’ visible in maps is definitional and structural. It represents excluded surfaces and areas that are not parcelised or not valuation-eligible by design, and it must not be interpreted as missing data or ‘zero value’. In practice, White Space mainly reflects land deliberately excluded from the valuation-eligible fabric (for example, operational transport land and other out-of-scope surfaces) and areas not parcelised within the chosen fabric.

Lot 5: Participatory valuation workshops

Lot 5 provides a fieldwork-led, workshop-based evidence stream that captures how participants interpret valuation logic and what drives perceived clarity, perceived fairness, confidence, uncertainty, and trust under facilitated conditions.

Three methods are analysed in the audited dataset and are named consistently throughout the report: Comparable (Comp), Rule-of-Thumb Residual (RotR), and Point-Factor Index (PFI). Quantitative results are framed against audited denominators (workshops, participants, parcels, and response rows). The evidence supports describing how participants reasoned and responded under controlled, facilitated conditions, but it does not support treating the outputs as market-calibrated values or as transferable valuations for other sites. Method comparisons require a further guardrail: ‘method usage’ in this dataset is a design footprint (fixed exercise exposure) rather than preference or adoption, and differences between method-labelled value summaries must not be interpreted as causal ‘method effects’.

Key findings

Lot 3: What the audited evidence shows

Coverage and parcel fabric: The Lot 3 modelled land value surface amounts to 82.5% coverage of the total specified land area within the nine LSOA boundaries, using the INSPIRE Index Polygons as the primary parcel fabric. A core delivery learning is that parcel fabric choice is decisive. Alternative open polygon sources tested during scoping did not deliver sufficient, consistent coverage for report-facing outputs. The report therefore treats a consistent, high-coverage parcel fabric (and explicit, published exclusions) as a non-negotiable prerequisite for reproducible, publishable outputs.

Ladder behaviour: The staged ladder behaved as designed for the first three rungs: Formula 1 (FLAT) established a uniform baseline, Formula 2 (FLAT_RUR) introduced stepped, area-wide adjustments, and Formula 3 (ACCESS) produced a differentiated range of indicative values by applying continuous penalties for accessibility and constraints. Formula 4 (COMP) produced outputs orders of magnitude larger, confirming its role as a prototype/diagnostic tool rather than a policy-ready approach.

Between-area differentiation: Under Formula 3, median implied land values range from £75/m² in Bridgend 019D to £165/m² in several other LSOAs, consistent with the applied schedule and penalty structures.

Totals and rung-to-rung change are not monotonic: Across the nine LSOAs, model-implied totals are approximately £42.94bn (Formula 1), £46.87bn (Formula 2), and £44.51bn (Formula 3). In this audited run, Formula 2 uplifts totals by approximately 9.2% relative to Formula 1, and Formula 3 offsets part of that uplift (approximately 5.0% relative to Formula 2), resulting in a net change of approximately 3.7% from Formula 1 to Formula 3.

Governance-relevant signals in stepped schedules: Two LSOAs show a ‘silent step’ where Formula 2 does not uplift relative to the flat baseline (Flintshire 015A and Bridgend 019D). One LSOA reduces under Formula 2 (Monmouthshire 006F at −5.0%). These behaviours are governance-relevant signals that the underlying flag/threshold logic requires careful scrutiny before any operational use.

Distribution and concentration risk: Rate tails and value tails behave differently: Formula 2 median rates range from 142.5 to 165 £/m², while Formula 3 penalties introduce further dispersion, with P10 rates as low as 71.3 £/m² (Monmouthshire 006F) and medians as low as 75 £/m² (Bridgend 019D). Despite capped map rates, parcel values are highly skewed because parcel areas are highly unequal. Under Formula 3, the share of total value represented by the single highest-value parcel ranges from 2.32% (Pembrokeshire 002F) to 46.46% (Powys 011C), and the top ten parcels range from 17.44% (Pembrokeshire 002F) to 78.31% (Powys 011C). Gini coefficients range from 0.537 (Ceredigion 002D) to 0.971 (Powys 011C), evidencing structural concentration risk that constrains totals-only interpretation.

Class coverage and interpretive limits: The audited export includes four NLUD classes (Agricultural, Residential, Undeveloped, Community/Amenity). Commercial/Industrial does not appear in the audited export for the nine LSOAs due to the classification-transfer boundary of the redeployment and the fallback mapping used only to resolve residual unclassified parcels. Residential accounts for 58.6% of parcels but only 0.7% of parcelised area, whereas Agricultural accounts for 28.4% of parcels and 60.7% of parcelised area (with Undeveloped accounting for 35.1% of area and Community/Amenity 4.2%). The report therefore treats class coverage as an interpretive constraint. Outputs should be read within the audited class boundary and not extrapolated to omitted classes without further evidence and governance work.

Lot 5: What the audited evidence shows

Delivery, participation and completion: Across ten workshops (eight public and two professional), 63 participants took part. Overall, 61 of 63 participants are complete (96.8%), and 2 of 63 are partial (3.2%). Partial completion is treated as evidence relevant to toolkit accessibility and comprehension under real workshop conditions, with the logged reason being a language or terminology barrier.

Method deployment footprint (design exposure): The audited dataset contains 630 value responses across the three methods. The fixed deployment footprint is 40% RotR (252 responses), 30% Comp (189 responses), and 30% PFI (189 responses).

Clarity versus fairness is a central finding: Participants do not treat ‘clarity’ and ‘fairness’ as the same concept when evaluating valuation approaches. Across all participants, Comp is most frequently ranked as clearest (25 participants; 39.68%), while PFI is most frequently ranked as fairest (28 participants; 44.44%), with RotR ranked fairest by 23 participants (36.51%). This split is policy-relevant because it indicates that a method can be perceived as procedurally fair without necessarily being comfortable or easy for users to apply.

Indicative values show meaningful dispersion even within a bounded parcel set: Across all responses, the overall value range is £49,100 to £151,700, with a mean of £96,196 and a standard deviation of approximately £24,038. The report treats spread as a more defensible route to identifying ‘hard-to-value’ parcels than relying on headline highs and lows alone.

Confidence is moderate overall but differs by method: Across all responses, mean confidence is approximately 6.06 on a 1–10 scale (standard deviation approximately 1.28). The method most often perceived as fairest (PFI) has the lowest average confidence (mean 5.37; median 5), compared with RotR (mean 6.14; median 6) and Comp (mean 6.65; median 7).

Qualitative mechanisms explain the quantitative patterns: This indicates that explainability is not an optional communications layer; it is a core requirement of the valuation method if outputs are to be credible in public-facing settings. Across workshops, trust is repeatedly linked to whether participants can see how a number is produced and whether the underlying assumptions are visible and contestable; ‘black box’ concern is a recurring legitimacy gate. Uncertainty is dominated by below-ground unknowns and hidden constraints (for example, ground conditions, abnormal costs and remediation, utilities, and constraints not visible in surface-level evidence packs). The report concludes that assumptions are the hinge between perceived fairness, perceived clarity, and confidence, and that uncertainty capture should be an explicit feature rather than forcing false precision.

Comparison of approaches and joint implications

Lot 3 and Lot 5 operationalise land value differently and produce different types of evidence. Lot 3 generates parcel-level value density signals (£/m²) over an analytical parcel fabric within a defined audit boundary. Lot 5 generates facilitated human judgement evidence (rankings, confidence, qualitative rationales, and indicative numeric outputs) over a bounded parcel set within a separate fieldwork-and-analysis audit boundary. The appropriate basis for comparison is therefore feasibility, interpretability and usability, auditability and governance, scalability and coverage potential, and public intelligibility and legitimacy.

If the objective is repeatable coverage and consistent reporting surfaces, a Lot 3-type approach is necessary. Credible use requires strong governance: stable parcel fabric and exclusions; explicit explanation of White Space; controlled and versioned parameters so changes are traceable; readiness tiering so the method does not overreach where inputs are incomplete; and publication discipline that emphasises distributional statistics over totals where concentration risk is material.

If the objective is deliberation, scrutiny, trust-building and intelligibility testing, Lot 5 provides evidence that is necessary to design a credible public-facing approach. It cannot substitute for a scalable technical method where national coverage is required, but it reduces ‘translation risk’ by showing how people interpret value logic, what they regard as fair, and what undermines confidence.

The programme evidence supports a combined position in which technical outputs are paired with an explainability layer informed by participatory evidence. Bridge actions include making constraints and risks salient and comparable in public-facing engagement, using participatory testing to refine how technical maps and tables are explained (particularly where within-area differentiation is pronounced), and increasing overlap between participatory parcel examples and parcels in the technical fabric where feasible, without implying that the two approaches measure the same construct.

Conclusions and further considerations

Across both lots, the programme evidences that ‘land value’ is not a single observable quantity that different methods simply estimate with more or less precision. It is constructed through definitional boundaries (what is included and excluded, and the unit at which value is expressed), the drivers each method is able, or chooses, to operationalise, and the way outputs are interpreted by non-specialists and professionals. A technically auditable output can therefore still fail as policy evidence if it is not intelligible, perceived as legitimate, and accompanied by clear guardrails that prevent misinterpretation.

Scalability is conditional. Lot 3 is, in principle, scalable across Wales because its core value is a repeatable pipeline and parameter-governed execution; however, scale-up is conditional on governance and data readiness. The evidence indicates that scaling will be driven less by the formulae themselves than by stable fabric construction rules, consistent feature engineering, licensing and refresh cycles, Quality Assurance thresholds, and an auditable eligibility and coverage reconciliation that prevents misinterpretation of excluded land and White Space. Lot 5 is not scalable as a Wales-wide parcel valuation system, but it is scalable as a structured engagement and legitimacy layer.

Moving towards an operational, Wales-wide evidence product requires requirements to be stated explicitly, in a way that maps directly to risks surfaced across both lots. The consolidated report emphasises: (i) a published eligibility and coverage reconciliation layer that quantifies what has been valued, what has not, and why; (ii) publication-grade parameter governance and interpretability exports (including the fields required to answer common ‘why’ questions using published artefacts); and (iii) a public explanation framework, informed by participatory evidence, that anticipates anchoring and place-based reasoning, distinguishes procedural clarity from procedural fairness, and makes uncertainty explicit rather than hidden.

1 Introduction

1.1 Expertise and credentials

In 2025, Industryline Research was commissioned by the Welsh Government to undertake two strands of the multi-lot research programme, Testing Land Valuation Methodologies. The two stands of work undertaken fell under Lot 3 and Lot 5, related to testing land valuation formulas, and bidder-defined participatory workshops. The purpose of the work is methodological testing and evidence capture: what can be done with available inputs, how outputs can be generated and audited, and how different approaches may be interpreted and communicated responsibly.

Industryline Research is a specialist research and intelligence organisation delivering public-sector research in housing, education, and population and demographic research. The organisation undertakes mixed-method and quantitative programmes, including large-scale survey-based work in the UK education sector, and has delivered prior Welsh Government research in housing and the built environment, including evaluations and reporting related to construction, housebuilding, and programme delivery. Throughout the project, delivery has been governed through a structured assurance approach designed to support auditability and professional standards. This includes clear evidence boundaries, explicit assumptions, consistent documentation, quality checks embedded throughout the workflow, and a focus on reproducibility where feasible. These principles are particularly important for a programme that is explicitly testing feasibility and interpretability, rather than producing only numeric outputs.

1.2 Background and context of valuing land in Wales

Land valuation has relevance to a wide range of public functions in Wales, including planning and development decisions, infrastructure and land acquisition, the appraisal and prioritisation of public investment, and wider expectations of transparency in public decision-making. In practice, however, producing land value evidence that is both technically defensible and publicly intelligible is challenging. This programme sits within that practical context. It tests approaches that may support more consistent and explainable evidence, while recognising the limits imposed by data readiness, land heterogeneity, and the need for clear communication.

Across Wales, demand for housing, infrastructure and economic development places sustained pressure on a finite stock of developable land. Household formation, demographic change and patterns of internal and external migration contribute to demand for a range of housing tenures, including social and affordable housing, market-sale homes and private rented stock. At the same time, land is required to accommodate economic uses such as industry, logistics, retail and services, and to support essential infrastructure, including transport, utilities and community facilities.

Alongside these requirements, land is also needed to support environmental and climate objectives. This includes land for nature recovery, flood risk management, carbon sequestration, coastal and riverine adaptation, and the protection of biodiversity and ecosystem services. Agricultural land underpins rural livelihoods and food security, while landscapes and cultural assets contribute to well-being and tourism. In practice, this means that many locations are subject to competing claims and expectations, with different groups and interests attaching different values to the same parcels of land. Debates can therefore hinge on how land is valued, what uses are considered most appropriate, and how benefits and burdens are distributed between different groups and places.

Land-use decisions in Wales are taken within a structured planning system and legislative framework. Planning authorities are required to balance multiple objectives, including meeting housing requirements, supporting economic growth and local prosperity, enabling infrastructure delivery, promoting climate adaptation and nature recovery, and safeguarding cultural and heritage assets. The planning system is therefore a central arena in which competing demands on land are mediated and reconciled.

Within this system, land value plays a significant role. It underpins viability assessments for development proposals, informs judgements about the scale and type of development that can be supported in particular locations, and shapes negotiations about planning gain, infrastructure contributions and affordable housing delivery. How land value is understood and communicated can influence perceptions of fairness, including whether communities consider that they are receiving an appropriate share of the benefits associated with development, and whether costs and disbenefits are being borne disproportionately by particular groups or places. Planning services also operate under resource and capacity constraints, which can limit the time and resource available for detailed communication about technical issues such as land valuation and for additional modes of engagement beyond statutory processes.

Questions of fairness, transparency and trust recur in public and stakeholder debates about land and development. Communities may question whether particular developments are in the public interest, whether the distribution of benefits and impacts is equitable, or whether decisions are driven primarily by financial considerations. Businesses and landowners may be concerned about predictability and consistency, including how land value is assessed in relation to planning obligations or other requirements. Formal valuation processes can be perceived as opaque, in part because they rely on technical methods and data that are not routinely explained in accessible terms. From the perspective of residents or small businesses, the outcomes of land-related decisions can therefore appear difficult to reconcile with local perceptions of place and fairness, even where formal consultation requirements are met.

A central operational difficulty is that land is heterogeneous. Two parcels of the same area can differ materially in access, location, constraints, surrounding uses, service availability, permitted uses, and development viability. Value formation reflects both measurable factors (for example, connectivity, proximity to services, or constraints that affect use) and less tangible factors (for example, perceived desirability or market sentiment). These features make land valuation difficult to standardise and can create communication challenges. Even where a method is internally consistent, explaining why a value arises in a way that is accessible to non-specialists can be non-trivial.

Data completeness and readiness also vary. Different valuation approaches rely on different categories of inputs, and the feasibility of applying an approach consistently depends on whether those inputs are available at the required spatial granularity and with sufficient coverage. In a public-sector setting, this directly affects scalability and the strength of any claims that can be made about comparability across places.

In this context, the programme is deliberately experimental. It tests the applicability of different valuation approaches within a Welsh context, including the availability and usability of inputs, the robustness of outputs under practical constraints, and the conditions under which results can be interpreted responsibly. The intent is therefore broader than producing values alone: it is to understand strengths and limitations, the nature of the assumptions required, and the implications for transparency, auditability and communication. For the purposes of this report, the focus is on land valuation. The following sections set out the shared methodological principles applied across the research, followed by the lot-specific designs and data collection approaches.

1.2.1 What this report is not

Within this context, it is also important to note that due to the experimental nature of this programme of works, the evidence presented here is produced for methodological testing. It is not professional valuation advice for individual sites or transactions, and it should not be interpreted as a substitute for professional valuation practice. The report does not claim that any tested method replaces established valuation approaches; rather, it assesses how different approaches operate in practice, what inputs they depend on, how outputs can be audited and explained, and what limitations those dependencies imply for any potential future application.

The report is structured “by lot” because the two strands answer different questions and produce different forms of evidence. Lot 3 tests a formula-based approach that can generate consistent, auditable, parcel-level indicative outputs, subject to the availability and suitability of the required inputs. Lot 5 tests participatory approaches to examine how different audiences interpret land value problems, what they regard as clear or fair, where misunderstandings or heuristics emerge, and what this implies for the interpretability and perceived legitimacy of valuation outputs. Taken together, the strands are broadly complementary: technical reproducibility alone does not address public interpretability, and participatory evidence can inform how technical outputs are communicated and used without overstating precision.

1.2.2 Scope of the report

The programme evidence is presented for the geographies covered by the work (nine Lower Super Output Areas), and it is organised into six sections consistent with the required reporting structure: methods and data, findings, cross-lot comparison, conclusions, and further considerations. This structure is intended to make the differences between the two strands explicit, while enabling a coherent, consolidated narrative within a single report.

Section 2 sets out the methodology for each lot. Section 3 presents the findings for each lot, followed by sections on cross comparisons and conclusions. As this programme is exploratory and deliberately experimental, the report presents the evidence as it occurred in practice, including both successful elements and instances where an approach did not perform as intended or could not be applied reliably. In a delivery-focused valuation exercise, such issues would typically be handled through exception reporting and would not be set out in detail. Here, however, documenting limitations, breakdown points, and “non-working” cases is integral to the purpose of the work: to test feasibility, interpretability, data dependencies, and operational conditions for use at scale. Where shortcomings or failures are reported, they should therefore be read as findings in their own right, enabling clear lessons on why issues arose, what their practical consequences are, and what this implies for future application and further development.

1.3 International examples of comparative approaches

During the commissioning phase of the programme, reviewing international practice provided a clear view of how different land valuation approaches operate in practice and the conditions under which they are feasible. The Lot 3 and Lot 5 methods share key characteristics with three international precedents, which are summarised here as contextual background. These examples demonstrate that comparable approaches are used elsewhere, and they provide transferable lessons that may help inform future research and development in Wales. The three international examples below have been selected because they mirror the core “architectural” features being tested in this programme: (A) algorithmic value generation and the practical challenge of explainability; (B) published standard value maps and the management of boundaries (including “cliff-edge” effects); and (C) simplified, trust-forward valuation design that is intended to be understandable and contestable by non-specialists.

1.3.1 Denmark (model-based assessments)

Denmark provides a relevant comparator because its assessment regime explicitly separates overall property value from land value within the statutory assessment process, with those values then used as inputs to housing-related taxation (Danish Agency for Digital Government, 2025). In public guidance, the land-value element is framed in “undeveloped” terms (i.e., land assessed as if there were no buildings), which conceptually aligns with a land-value-per-surface focus rather than a full developed-property appraisal (Danish Tax Agency (Skattestyrelsen), n.d.).

Denmark is also relevant because the assessment process is explicitly model-based. Legal guidance describes assessments as produced using model calculations that are statistically and theoretically grounded, using recent “free sales” over a defined window and then correcting for the assessed property’s characteristics (Skatteforvaltningen, 2026). This is recognisably a mass-appraisal architecture: transaction evidence is translated into valuations through centralised modelling, rather than through bespoke, site-by-site manual valuation (Skatteforvaltningen, 2026).

In terms of alignment to Lot 3, the key parallel is the “auditable pipeline” logic: centrally specified rules and standardised inputs are applied consistently at scale, which increases auditability but also raises the bar for governance and explanation (particularly as the number of drivers and corrections increases) (Skatteforvaltningen, 2026). The Danish Property Assessment Agency’s own framing emphasises fairness, uniformity, and transparency as core objectives, reflecting that public confidence is treated as an operational requirement of the system, not a secondary benefit (Danish Property Assessment Agency (Vurderingsstyrelsen), n.d.).

A particularly transferable feature for Wales is Denmark's approach to "explainable fragments" of a complex model. In the Danish Parliamentary Ombudsman's review of the Agency's reasoning practices, assessment communications are described as providing (i) key factual datapoints, (ii) reference properties expected to inform the valuation, and (iii) supporting explanatory content that enables owners to understand the basis for the assessment and challenge incorrect inputs (The Danish Parliamentary Ombudsman, 2021). This is a pragmatic response to the "black box" problem: rather than attempting to expose full model internals, the system makes inputs and reference evidence contestable and provides a structured narrative of how the outcome is generated (The Danish Parliamentary Ombudsman, 2021).

Denmark also evidences a formal exception-handling route where model outputs appear implausible. An internal process note describes manual screening where computed land value is markedly higher than overall property value, distinguishing between (a) errors in data used by the model (including planning data), (b) cases where data appear correct but the model generates a disproportionate land value for other reasons, and (c) cases where the valuation is retained (Danish Property Assessment Agency (Vurderingsstyrelsen), 2023). This is closely analogous to the governance requirement in Lot 3 to separate defects in inputs from model behaviour, and to document how anomalies are triaged rather than silently corrected.

Denmark also illustrates the political fragility that can arise if explainability is not perceived to be adequate. The release of preliminary assessments in 2023 generated public controversy, including concerns where land values appeared unusually high relative to total property values (Schjødt, 2023). In 2025, the Danish Parliamentary Ombudsman explicitly framed the legitimacy issue as the need for citizens to obtain an explanation of computerised calculations in individual assessments, and opened an investigation focused on whether the Agency can explain and document the calculations behind assessments (including land value) (The Danish Parliamentary Ombudsman, 2025). The Agency's subsequent public communications acknowledge the trust challenge and describe additional quality assurance activity intended to improve the robustness of valuations and rebuild confidence (Danish Property Assessment Agency (Vurderingsstyrelsen), 2024).

For Welsh reporting purposes, the key learning is not that Denmark has "solved" explainability, but that even a mature, rules-based algorithmic system remains vulnerable to legitimacy challenge unless it can surface the key datapoints in a contestable form and provide a comprehensible pathway explanation (The Danish Parliamentary Ombudsman, 2021).

1.3.2 Germany (published standard land values, zone governance, and boundary management)

Germany's standard land values (Bodenrichtwerte) provide a close analogue for a "published schedule" approach. The core output is a zone-based standard value per square metre for a defined "standard plot" (an undeveloped, representative reference parcel for the zone), with values derived from transaction evidence held in purchase price collections (Bundesportal, n.d.). This design is deliberately zonal: expert committees collect notarised transaction information, maintain the purchase price collection, and publish standard land value information for public inspection (City of Bocholt, n.d.). Public access is supported through national and state portals, including BORIS-D for cross-state access (BORIS-D, n.d.).

Germany is also relevant because the zonal map can be used as a direct schedule in some administrative contexts. For example, Stuttgart's public guidance for property tax B explains that, for certain purposes, the tax base can be calculated by multiplying plot area by the applicable standard land value, with the building (and its characteristics) not forming part of that computation (City of Stuttgart, n.d.). Structurally, this is very similar to a "standard rate per m² by zone" valuation surface.

In terms of governance, Germany offers practical mechanisms for managing the "cliff-edge" problem at zone boundaries. A Brandenburg administrative instruction defines the standard land value as a zone-average location value per m² for a defined standard plot, and sets out expectations for zone construction (including that zones should not be spatially disconnected) and coordination at administrative boundaries to support coherence in value levels (State of Brandenburg, 2024). The same instruction also makes clear that boundary-setting does not replicate planning law and has no planning effect, which is important for public interpretation and for avoiding confusion about the purpose of zones (State of Brandenburg, 2024).

Germany's public-facing materials also address the limits of zonal averages and the implications for contestability. Stuttgart's guidance emphasises that standard land values are not market values for individual sites and that actual values may deviate because the published values relate to a reference date and standard plot assumptions (City of Stuttgart, n.d.). In addition, Stuttgart describes an "opening clause" mechanism in which owners can apply for a different value in certain tax contexts where a qualified appraisal evidences that the schedule-based value deviates by more than 30% from the site's actual value (City of Stuttgart, n.d.). Demonstrating the applicability of explicit standard-plot framing, boundary coordination expectations, and formal contestability mechanisms where zonal averaging produces perceived unfairness for a specific economic unit (State of Brandenburg, 2024).

1.3.3 Freetown, Sierra Leone (points-based, explainable design oriented to trust and compliance)

Freetown's points-based valuation reform is a strong real-world analogue for Lot 5's emphasis on intelligibility, fairness, and trust. The approach is designed to be explainable and verifiable by non-specialists: a base set of points is allocated (linked to measurable size), and points are then added or subtracted for observable characteristics (including proxies for location/service access and construction/finishing quality). The points are then calibrated into estimated values using rental market evidence (International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019).

The operational pipeline is also explicitly intended to work at scale. The reform model includes comprehensive identification and measurement using satellite imagery, enumerator field teams collecting standardised observable characteristics, calibration using rental data from a sub-sample, and an IT-enabled process spanning valuation, billing, payment, appeals, and enforcement (Kamara, 2023).

While Freetown is not a participatory workshop system, it is a relevant comparator because valuation design is explicitly framed around acceptability, transparency, and verifiability rather than around maximising market precision (International Centre for Tax and Development (ICTD), 2020). This is conceptually consistent with Lot 5's evidence that legitimacy is gated by whether people can understand and challenge the basis of an output. In practical terms, the "checklist" structure supports contestation at the level of recorded attributes (for example, whether characteristics have been captured correctly), rather than requiring citizens to interrogate a complex statistical model (International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019).

Freetown also highlights that “simplicity” does not remove governance risk: even where indicators are externally observable, there can be subjectivity and inconsistency between enumerators. The ICTD summary brief describes the need for calibration (using rental value data and econometric analysis), backchecking for consistency, and operational controls to reduce measurement variation (International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019). This point is directly relevant to any Wales-facing simplified schedule approach: transparency gains can be undermined if application of criteria is inconsistent across field teams. On legitimacy and acceptance, both the ICTD summary brief and the ATAF reform note emphasise the importance of political leadership and a credible “social contract” narrative linking taxation to visible service delivery (International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019). They also describe distributional effects consistent with a progressive reform narrative (higher relative liabilities for higher-value properties and reductions for lower-value properties), framed as improvements in fairness and compliance incentives (International Centre for Tax and Development (ICTD), 2020).

For Wales, the core transferable lesson is that simplified valuation can plausibly support acceptance where and that where the public can verify and contest recorded attributes, the logic can be explained without specialist translation, which results in decreased friction points during standarised land valuation (International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019).

1.3.4 Implications of international examples

Across the three examples, the most transferable evidence is how governance design responds to explainability, boundary effects, and contestability; mirroring the lenses used in this programme’s comparison section, and avoiding false equivalence between technical outputs and social legitimacy. This speaks directly to the Lot 5 finding that perceived legitimacy is shaped by whether people feel able to understand and challenge the basis of an output, not solely by whether an approach is technically sophisticated.

2 Methodology

Lot 3 Formula-based Valuation: Methodology

2.1 Scope and conventions

This section outlines the methodology used to generate standardised land value-per-area outputs for Lot 3 across the specified geographies. The unit of analysis is the analytic parcel: a constructed spatial unit designed to enable consistent feature attachment and formula execution. Analytic parcels are not legal title parcels (for example, Land Registry titles), and they are not buildings or building footprints. The outputs are standardised land value per unit area (typically expressed as £/m²), generated as evidence for methodological testing; they do not constitute professional valuation advice for individual sites or transactions.

The methodology sections are intended to explain both the underlying logic and the interpretive intent. Terminology is used consistently throughout to avoid downstream misinterpretation. “Valuation-eligible land” refers to parcels that are in-scope under the classification scheme and are eligible to receive £/m² outputs. “Undeveloped land” is a subset of valuation-eligible land: it remains in-scope, is valued, and should not be interpreted as “ignored” or assigned “zero”. “White Space” refers to structural exclusions arising from the parcelisation and valuation base construction; it is not an assertion that excluded areas have no value.

2.2 Research design and methodological approach

Lot 3 tests whether a transparent, reproducible, formula-based workflow can generate parcel-level land value-per-area outputs across diverse Welsh contexts using open and official datasets. The methodological intent is to enable scrutiny, re-runs, and structured refinement.

During the research design phase, Lot 3 was planned as a multi-phase approach. The project began by confirming and agreeing on the formula shortlist and the study frame, before moving into a data assembly and cleaning phase. This phase would build a harmonised database for nine selected LSOA areas. Once implemented, the four selected formula would be run across all parcels, followed by quantitative benchmarking and validation, including sensitivity analysis.

Consistent with its exploratory nature, the project sought full coverage within each LSOA across all OS OpenMap Local polygons, using five high-level categories aligned with national land use databases. Emerging findings, set out further in Section 3, showed that coverage across the nine LSOA areas was low when using OS OpenMap Local polygons, with only 23% coverage. Given that a key exploratory aspect of the project was to use open-source databases, the polygon base was therefore switched to HM Land Registry INSPIRE Index polygons spatial data, which delivered a substantial uplift in land coverage. Overall coverage increased from 23.33% in the baseline run to 82.49% in the INSPIRE redeployment, representing an uplift of +59.16 percentage points. [1]

Key design distinction is made between two output metrics: (i) the parcel value in GBP (value_gbp), which is the true, pre-cap output of the valuation formulae; and (ii) the displayed rate per square metre (rate_gbp_per_m2), which is capped at £150 for mapping and public-facing outputs to prevent extreme visual outliers. In practice, this ensures that maps and other public-facing materials remain legible and are not dominated by a small number of extreme values. However, for analytical reporting in this document (for example, Table 3.1), rates are expressed as the implied pre-cap rate (value_gbp / area_m2). Totals are always reported as the sum of value_gbp.

In turn, standardising outputs to a value per unit area supports comparability across heterogeneous parcels. Land varies materially in size, geometry and local context; therefore, expressing outputs in £/m² provides a consistent unit for comparing parcels and for summarising outputs at higher geographies (for example, by class or LSOA). To enable this consistently, analytic parcels are used because they provide a stable geography for attaching features, applying classification rules, and executing formula schedules, rather than relying on legal titles whose boundaries and availability are not designed for this analytic purpose.

Building on this, a deliberate “Complexity Ladder” was designed as part of the project methodology to make data dependency explicit and to balance coverage with interpretability where feature completeness varies. The ladder is implemented as a hierarchy of formula rungs, increasing in complexity and input dependence. This ladder makes up the four formulas within this project's scope and consisted of:

- FLAT: baseline class schedule (rate card)
- FLAT_RUR: discrete rural/low-access adjustment (stepped multiplier)
- ACCESS: continuous accessibility and constraint adjustments (distance and flood exposure)

[1] Coverage Calculation Note**: Overall study coverage (82.49%) is computed as Σ(parcel union area) ÷ Σ(LSOA area), a weighted aggregate that differs from the simple average of LSOA-level coverage percentages.

- COMP: composite regression model surface, calibrated using transaction-linked evidence

Accordingly, methodological learning retained in the approach is that a tiered readiness framework is not merely an implementation detail; it is required to balance coverage with methodological integrity in heterogeneous geographies, while also maintaining a clear audit trail of which inputs drive which outputs.

2.2.1 Technical implementation and data environment

Following the determination of the methodological design framework, the second phase of the project focused on establishing a host environment to support data ingestion and formula testing. To achieve this, the project utilised a spatial data environment established in Azure UK South, using a standard PostgreSQL database with the PostGIS extension enabled. This environment acts as the central warehouse for all geometry, attributes and valuation outputs. It is intentionally built around familiar, widely supported tools rather than bespoke or proprietary software, to enable reproducibility if required.

With the host environment in place, the next step is the construction of a full-coverage parcel fabric for each of the nine study LSOAs. The fabric is based on road-bounded blocks derived from OS OpenMap Local roads and Boundary-Line LSOA polygons, and is then refined by dissolving OpenMap polygons into an NLUD-aligned five-class scheme (for example, residential, commercial/industrial, agricultural, community/amenity, undeveloped). The result is a standardised core parcel list that covers valuation-eligible land in each LSOA and provides the main unit of analysis for the formulae.

Building on this parcel fabric, parcel-level features describing access and constraints are then engineered. These include measures based on WIMD Access to Services, distances to primary transport nodes (for example, main roads and rail stations derived from OS Open Roads and NaPTAN), and physical constraints such as flood risk derived from TAN 15 maps. These features are stored in dedicated feature tables and linked back to the parcel fabric, so that each parcel can be described not only by its land-use class and area, but also by its relative accessibility and exposure to constraints.

2.3 Steps to determining Land Values

The process is structured as a clear chain of operations, moving from spatial frame construction through to formula execution and verification.

Step 1 — Build the spatial frame and base parcel fabric (road-bounded blocks)

The spatial frame for Lot 3 comprises the nine specified LSOAs. Within this frame, a base “block scaffold” is constructed using road-bounded blocks, with road corridors treated as structural separators rather than valuation surfaces. As a result, road surfaces and associated movement corridors are removed during block construction and do not reappear as parcels later in the workflow. This design decision is therefore a primary, intentional contributor to White Space.

Once the road-bounded blocks are established, surface polygons from Ordnance Survey INSPIRE Index Polygons are attached and harmonised to form the base fabric. Polygons are first associated to blocks and then dissolved and standardised into analytic parcels. In turn, this ensures that each analytic parcel remains a stable unit for subsequent class attribution, feature attachment, and formula execution.

Step 2 — Operationalise valuation-eligible land, undeveloped land, and White Space

To ensure that key terms are applied consistently, definitions are first set out in plain English and then implemented via deterministic rules:

Valuation-eligible land: parcels classified into one of the five NLUD-aligned classes (Step 3) and therefore eligible to receive £/m² outputs.

Undeveloped land: a subset of valuation-eligible parcels that are not built but remain in-scope and are valued. Undeveloped is treated as an eligible class, not as an exclusion and not as “zero value”.

White Space: structural exclusions from the parcel fabric and valuation base arising from the fabric-building process and “never parcelise” rules. White Space indicates that the surface is excluded by design; it is not a claim about land value.

Building on these definitions, “never parcelise” / exclusions are then applied as follows:

- **Operational transport land:** treated as structural White Space; excluded from valuation parcelisation. This includes carriageways and road surfaces used solely for movement and associated operational corridors where represented in the source surfaces.

- **Water surfaces:** excluded by default. Foreshore and tidal extents are excluded by default and only included under an explicitly labelled sensitivity run (not part of the audited baseline).

Additional structural and infrastructure surfaces outside valuation scope: excluded to preserve the conceptual distinction between valuation-eligible land and White Space.

Step 3 — Apply the five-class NLUD-aligned classification scheme (final)

With the valuation base defined, a simplified, NLUD-aligned five-class scheme is applied to ensure consistency, transparency and fit-for-purpose modelling. This approach supports reproducible classification across diverse contexts, while avoiding overly granular categories that would require stronger, often unavailable, evidence for consistent assignment.

The five classes are:

- Residential
- Commercial/Industrial
- Agricultural
- Community/Amenity
- Undeveloped

Conceptually, each analytic parcel is assigned exactly one class for the purposes of rate scheduling, benchmarking summaries and interpretive comparison.

Step 4 — Granularity and unit-of-analysis under INSPIRE

During implementation, the team tested a wider range of open sources at the build stage, including sources that can support different parcel-fabric strategies. Under the original approach of utilising OS OpenMap Local polygons, this approach delivered only 23.33% overall coverage. An exploratory analysis of alternative polygon datasets was conducted, with the most suitable dataset identified as HM Land Registry INSPIRE Index polygons spatial data. The INSPIRE redeployment (*presented as the front-facing valuation run for the reported findings),* therefore*,* adopted INSPIRE polygons as the primary parcel fabric, which materially improved geographic completeness and reduced unclassified residuals.

Granularity management is primarily a data-governance issue. The audited INSPIRE redeployment utilises the INSPIRE parcel fabric directly and does not apply additional parcel-splitting logic; this simplifies interpretability but also means that parcel size effects and concentration diagnostics should be treated as a core part of interpretation.

Table 2.1: Granularity Quality Assurance checkpoint

Metric	INSPIRE redeployment (primary Lot 3 evidence)
Parcel count	10,415
Total area (m²)	275,444,158

Macro parcels	N/A (INSPIRE polygons)
Mega parcels	N/A (INSPIRE polygons)

This evidences a consistent unit-of-analysis definition under INSPIRE, with granularity controls handled at source rather than through downstream splitting.

Step 5 — Engineer access and constraint features (inputs to FLAT_RUR / ACCESS / COMP)

A concise feature inventory is used to ensure that feature engineering is auditable and interpretable. Features are engineered in a projected coordinate system suitable for distance and area calculation (EPSG:27700), and are attached to parcels via stable join keys or spatial operations.

Access features (audited)

1. WIMD Access to Services domain quartile (LSOA-level). The quartile is joined to parcels via lsoa_code. This provides a stepped, area-wide measure of access conditions.

2. Euclidean centroid distances in metres (EPSG:27700) to:
i) nearest rail station point (INSPIRE Index Polygons railway_station), and
ii) nearest primary road segment (Motorway + A-roads subset from INSPIRE Index Polygons roads).

3. Combined primary-node distance term. The combined distance is defined as the minimum of the rail-station distance and the primary-road distance, representing proximity to at least one principal access node.

4. Log transformation for scale stabilisation. A log transform is applied as ln(1 + distance) to reduce leverage from long-tail distances and preserve interpretability of marginal changes.

Constraint feature (audited)

Flood exposure share: flood_share = area(parcel ∩ flood_zones) / area(parcel), computed using Natural Resources Wales Flood Map for Planning polygons. flood_share is continuous on [0, 1] and is interpreted as proportional exposure rather than a binary “in flood zone” flag.

Dataset year governance (WIMD)

To prevent contradiction between planning inventories and the audited run, the WIMD Access to Services extract used in the audited outputs is identified by the publication identifier and extract date recorded in the Appendix. The method uses the Access to Services domain quartile as recorded for the audited run, rather than relying on an informal “current year” label.

Step 6 — Apply formulae (Complexity Ladder), including COMP calibration and benchmarking

Once the previous steps have allowed for the ingestion of relevant data, the next step in the approach was then to apply formulae. Each rung was described using a consistent template:

inputs → transformation → output → interpretive purpose → parameter publication location.

(1) FLAT — base class schedule

Inputs: parcel fabric v2; five-class NLUD-aligned scheme.

Transformation: assigns a baseline £/m² to each parcel according to its class using a fixed rate card (class schedule).

Output: baseline class-driven £/m² for each valuation-eligible parcel.

Interpretive purpose: provides a transparent starting surface that makes class assumptions explicit and enables straightforward auditing and sensitivity review.

Provenance: the baseline schedule is derived from calibration. Class-level £/m² values were calibrated from transaction-linked and residual evidence during the benchmarking phase, and then fixed as the auditable rate card for this research.

(2) FLAT_RUR — discrete rural/low-access adjustment

Inputs: FLAT baseline; WIMD Access to Services quartile (joined by lsoa_code).

Transformation: applies a deterministic stepped multiplier by WIMD quartile to the baseline £/m². The adjustment is discrete (quartile-based), not continuous.

Output: adjusted £/m² reflecting rurality/low-access conditions as a controlled, interpretable shift from the baseline.

Interpretive purpose: provides a low-complexity enhancement that uses a single, widely interpretable access indicator; it is designed to be robust when finer-grained features are incomplete.

(3) ACCESS — continuous accessibility and constraint adjustments

Inputs: FLAT baseline; WIMD quartile; ln(1 + combined primary-node distance); flood_share.

Transformation: applies multiplicative modifiers to baseline £/m²:

- a continuous accessibility term based on the log-distance metric,
- combined with the stepped WIMD quartile effect, and
- a continuous penalty term based on flood_share.

Output: £/m² that incorporates both access gradient effects and proportional constraint exposure.

Interpretive purpose: tests whether transparent, low-dimensional continuous features can improve face validity and spatial differentiation while remaining explainable and auditable.

Sliver handling and transparency: flood exposure is treated continuously and is not subject to ad hoc deletion. Geometry validity safeguards and area normalisation rules ensure re-runnability and consistent interpretation of small intersections.

(4) COMP — composite regression model

The COMP rung provides a calibrated demonstrator surface for the pilot areas, using transaction-linked evidence to fit a composite model. The audited published baseline is distinguished from internal assurance sensitivity work to ensure clarity and auditable publication.

Published (audited) COMP baseline

Evidence base: HM Land Registry Price Paid Data (PPD), 2025 year-to-date (1 January 2025 to 31 October 2025), clipped to the nine specified LSOAs.

Geocoding: Code-Point geocoding to postcode centroid; transactions are linked to analytic parcels using point-in-polygon assignment. This imposes a known granularity limit (postcode centroid rather than property point).

Sample size: 114 in-scope residential transactions after cleaning, geocoding and clipping.

Dependent variable: a residual land value proxy defined as transaction price minus an estimated building value.

Building value baseline (published): an auditable rebuild-cost proxy is used for the published baseline in contexts where BCIS indices cannot be redistributed.

Cleaning rules: exclusions for invalid or blank postcodes; a residual floor rule to prevent non-physical negatives; documented clipping/winsorisation thresholds where used; and explicit reason flags for excluded or clipped rows within the audit schema.

As this programme is exploratory and deliberately experimental, the report presents the evidence as it occurred in practice, including both successful elements and instances where an approach did not perform as intended or could not be applied reliably. Results presented for Formula 4 (COMP) are reported as feasibility and diagnostic evidence. However, the resulting values are orders of magnitude larger than those generated under Formulas 1–3 and should not be interpreted as a definitive, market-calibrated valuation model. Additional datasets (including raster terrain data, Price paid data and postcode data), and licensed rebuild cost indices, were used in an internal sensitivity run to test whether incorporating could help explain or constrain the order-of-magnitude divergence observed in Formula 4 relative to Formulas 1–3 and to inform whether their inclusion would be justified in any mid-project refinement. However, within the tested parameterisation, these additions did not materially reduce the scale divergence observed in Formula 4.

2.4 Data sources and analysis techniques

During initial implementation, the team trialled a more granular open polygon fabric sourced from multiple open portals to enable additional parcel refinement steps; however, overall polygon coverage across the nine pilot LSOAs was approximately 23% of LSOA area. Alternative polygon datasets were subsequently explored to prioritise higher surface coverage, which detrained the usage of the INSPIRE polygon spatial data.

During implementation, the work confirmed that the Lot 3 approach is practically deliverable with open datasets and open-source tooling, but only where input completeness is treated as a non-negotiable operational dependency. In particular, where fallback land-cover classification is used to assign parcel classes that cannot be inherited via overlap transfer, full geographic tile coverage for the study extent becomes a prerequisite. Missing tiles do not degrade outputs gracefully; they generate structurally incomplete classification (residual NULL classes) that must be treated as a hard stop for defensibility. The practical implication is that any production deployment needs an explicit “data coverage gate” and a maintained inventory of tile extents for any raster-derived fallback. For transparency, the datasets used to define the spatial frame, construct the parcel fabric, engineer features, and (where applicable) apply fallback classifications are grouped below to reflect their role in the pipeline.

Spatial frame and parcelisation

- LSOA boundaries for the nine specified geographies (frame definition and joins).

- HM Land Registry INSPIRE Index Polygons (GML by council) used as the parcel fabric.
- OS Open Roads (oproad_gb.gpkg) for road-based distance and access drivers.
- OS Open Rivers (oprvrs_gb.gpkg) for proximity/environmental features.

Classification and fallback

- ESA WorldCover 2021 v200 GeoTIFFs used only to fill residual NULL parcels after overlap-based classification (fallback).
    - Stored in EPSG:4326; transformed to EPSG:27700 during processing.

Within this framework, each dataset is described below in terms of its provider, role in the pipeline, and any key notes relevant to interpretation or reproduction.

- LSOA boundaries (pilot frame) | Provider: Official boundary source | Role in pipeline: Study frame; joins via lsoa_code | Key notes: Full metadata provided to Welsh Government at project closure.
- INSPIRE Index Polygons | Provider: HM Land Registry | Role in pipeline: Parcel fabric | Key notes: GML by council; EPSG:27700
- OS Open Roads | Provider: Ordnance Survey | Role in pipeline: Access / proximity drivers | Key notes: Open dataset; local-only use
- OS Open Rivers | Provider: Ordnance Survey | Role in pipeline: Environmental driver | Key notes: Open dataset; local-only use
- OS Open Greenspace | Provider: Ordnance Survey | Role in pipeline: Greenspace proximity | Key notes: Open dataset; local-only use
- WIMD Access quartile | Provider: Welsh Government | Role in pipeline: Stepped access feature | Key notes: LSOA-level feature
- ESA WorldCover 2021 v200 | Provider: ESA | Role in pipeline: Classification fallback | Key notes: Used only for residual NULLs

Baseline-only note

The early-stage baseline run (*c.23% coverage*) drew on additional sources such as OpenMap Local (roads/surfaces/points), NRW Flood Map for Planning, and NaPTAN. These are not used in the report facing valuation run figures and are listed here for continuity only to demonstrate the makeup of the first “low coverage”. This baseline required additional preprocessing because OS OpenMap Local polygons are not designed to represent land parcels. The extra work (attachment/dissolve and topology harmonisation) improved internal consistency and reduced obvious geometry artefacts, but it did not materially resolve the core limitation. The unit definition remained feature-led rather than parcel-led, and coverage remained low. In practical terms, this meant that outputs were less comparable across areas and more difficult to interpret as an “all-land” evidence surface, even where formula mechanics were functioning as intended.

For the report-facing redeployment, INSPIRE polygons were selected as the analytical parcel fabric because they provide a more coherent and stable unit definition, materially higher coverage, and a more defensible basis for aggregation and reporting. The baseline is therefore treated as an exploratory stage that demonstrates feasibility and highlights why unit choice is decisive, rather than as a candidate for publication-grade valuation outputs.

2.4.1 Analysis techniques (non-code description)

- INSPIRE parcel fabric clipped to the nine LSOA boundaries (EPSG:27700), with validity repair and deduplication.
- Deterministic class attribution using overlap transfer and WorldCover fallback mapping (parameterised, auditable).
- Distance-to-network metrics computed in EPSG:27700 (e.g., primary road and rail distances).
- Parameterised formula execution using fixed valuation parameters

The early-stage baseline run (approximately 23% surface coverage) used OS OpenMap Local polygon features as a proxy parcel fabric. As these polygons are not a dedicated cadastral parcel dataset, additional preprocessing and polygon harmonisation were required (e.g., attachment/dissolve and topology harmonisation). These steps were not required in the INSPIRE deployment because the parcel fabric is already polygonal; the redeploy instead uses deduplication, LSOA clipping, and validity repair. In the wider context, it is common to state that land valuation 'needs a cadastre'. The programme evidence supports the underlying principle that a stable, authoritative unit definition is a precondition for defensible parcel-level valuation. A cadastre is the strongest version of that unit definition because it aligns to legal parcel boundaries. However, for method testing and for producing an auditable evidence surface, an agreed analytical parcel fabric can be used provided the report makes the boundary of what is and is not being valued explicit (including coverage and exclusions), and avoids implying that outputs are legal, title-parcel valuations unless the unit definition supports that claim.

2.4.2 Verification and testing

Verification is presented as a three-part framework: (i) fabric integrity, (ii) behavioural checks by formula rung, and (iii) verification and review. These checks demonstrate integrity, reasonableness and interpretability under pilot constraints; they do not claim "ground truth" validation.

Fabric integrity and Quality Assurance

Fabric Quality Assurance ensures that parcelisation and feature attachment are geometrically sound, internally consistent, and operationally repeatable.

Table 2.2: Fabric Quality Assurance summary

Quality Assurance category	What is checked	Why it matters
Geometry validity and spatial reference	Valid geometries; consistent CRS (EPSG:27700)	Prevents artefacts in area and distance metrics
Attachment integrity	Unattached elements identified and resolved	Ensures no silent coverage loss
Area reconciliation	Study frame vs parcel totals; White Space quantified as structural exclusions	Supports correct interpretation of map gaps

Overlaps/gaps/micro-slivers	Managed via defined rules; no ad hoc edits	Preserves auditability and rerunnability
Operational readiness	Joins and runs execute consistently	Ensures a reproducible audit package

Behavioural checks and “technical signals” by rung

Behavioural checks confirm that each rung behaves as intended given its inputs, without asserting results. Checks include directionality, plausibility and sensitivity review.

External verification

External verification is designed as a structured reasonableness and interpretability review, not a claim of definitive validation. It comprises:

Review: Reviewers assess parameter reasonableness, definitional consistency, notable outliers, and whether outputs align with the declared interpretive logic of the rungs.

Administrative comparisons: aggregate checks are performed at LSOA and class level to assess internal consistency and directional plausibility against known spatial-economic patterns, while explicitly recognising that the programme is a exploratory research project and does not provide a national “ground truth” dataset for land values. Discrepancies are recorded and addressed through parameter sensitivity testing and definitional refinements, with the audit trail maintained via run identifiers and parameter sets.

Output handling discipline: outputs are treated as indicative, auditable surfaces for methodological testing. Verification establishes integrity and interpretability boundaries rather than asserting accuracy for any specific site.

A key lesson from the build is that auditability does not arise automatically from a formula; it requires explicit engineering guardrails. The implementation therefore benefited from formal gates that (i) verify coverage (union area) at LSOA and overall level, (ii) enforce classification completeness (no residual unclassified parcels after the defined fallback sequence), and (iii) confirm row-parity across formula outputs (one valuation row per parcel per rung, no join leakage). These checks convert the method from “a model that runs” into a method that produces outputs that are defensible as evidence, because failures are detected early and reported transparently rather than being carried forward into downstream tables.

2.5 Limitations and challenges

Scope and interpretive limits

The approach values analytic parcels, not title parcels, and it produces land value per unit area rather than building-level valuations or full development appraisal valuations. The work is also confined to the specified geographies and is intended to test methodological feasibility and communication readiness, rather than to provide determinative figures for individual sites.

Implication: outputs should be interpreted as standardised, evidence surfaces suitable for comparison and methodological learning, not as transaction-ready site valuations.

Structural exclusions and White Space

A number of surface types are excluded by design, including roads, operational transport land and water surfaces. Foreshore and tidal extents are excluded in the audited baseline and are only explored in explicitly labelled sensitivity work.

Implication: apparent “gaps” on maps must be interpreted as structural exclusions (White Space), not missing data and not “zero value”.

Granularity management and parcel size effects

In practice, granularity management is primarily a data-governance issue. The audited INSPIRE redeployment uses the INSPIRE parcel fabric directly and does not apply additional parcel-splitting logic. This simplifies interpretability; however, it also means that parcel size effects and concentration diagnostics should be treated as a core part of interpretation.

Evidence coverage constraints (particularly for COMP)

The COMP evidence base is limited to in-scope residential transactions within the pilot LSOAs and uses postcode-centroid geocoding. This introduces spatial uncertainty at fine scales and constrains inferential strength.

Licensing constraints and publication choices

Some cost indices used for building value estimation cannot be redistributed within a publishable audited package. For this reason, the published baseline uses an auditable proxy rebuild-cost approach, with licensed-index variants retained as internal sensitivity assurance evidence only.

Implication: wherever COMP is discussed, the report must clearly distinguish the published audited baseline from internal sensitivity assurance evidence.

Although the approach is intentionally open and replicable, open-source delivery introduces a specific class of operational risk that is distinct from methodological risk. In particular, library and driver behaviour (for example, how GML readers resolve schema references or attempt remote calls) can create unexpected latency or non-deterministic warnings unless a controlled execution environment is used. This does not undermine the conceptual validity of the Lot 3 approach; however, it does mean that a deployment-ready version should specify the runtime environment (tool versions, configuration flags, file discovery rules) and should treat the run log and lineage record as required artefacts rather than optional documentation. In effect, “open and low-cost” is achievable, but it depends on disciplined configuration management and repeatable execution to avoid fragility.

In the same way, scaling the approach beyond an exploratory investigation footprint is feasible, but it depends on data governance choices rather than on additional modelling sophistication. In practice, a production roll-out would require a maintained whitelist of permitted datasets and their versions, supported by a formal update cycle for input refresh (including land-cover tiles and reference boundaries). It would also require defined thresholds for stopping and rerunning where coverage or classification gates fail, alongside clear responsibilities for curating and validating the input inventory. Without these controls, scaling risks becoming a “repeat the run and hope” exercise; with them, scaling is primarily an operational exercise in ensuring consistent inputs and consistent execution.

2.6 Key Assumptions

Lot 3 outputs are “indicative value per square metre” signals produced for method testing. They are generated on an analytic parcel fabric designed for consistent computation and auditing, not on legal title boundaries. “White Space” reflects structural exclusions created by the fabric rules and scope decisions; it should not be read as missing data or “zero value”. The staged ladder is intentional: early rungs prioritise coverage and auditability with minimal inputs, while later rungs introduce additional features and therefore require stronger data completeness and governance to remain interpretable. Where engineered penalties or modifiers are applied (for example, accessibility and constraints), these are explicit modelling choices that must be explained and sensitivity-tested; they are not presented as discovered truths.

Lot 5 Participatory Approaches: Methodology

2.7 Scope and conventions

Running concurrently to Lot 3, Lot 5 represents a novel approach to land valuation through a fieldwork-led, workshop-based evidence stream. The appropriate audit boundary is therefore defined in two parts: (i) fieldwork completion date (final workshop delivered); and (ii) analysis lock window: transcription, quality checks, anonymisation, and thematic coding completed in the week following fieldwork. Evidence generated after the analysis lock window is treated as out-of-scope for the audited Lot 5 evidence package by design, although it may inform ongoing development activity.

Lot 3 and Lot 5 therefore use different operational audit anchors. Lot 3 is bounded by a database “last updated” cut-off for a locked run package, whereas Lot 5 is bounded by fieldwork completion and a defined analysis lock window for a locked qualitative–quantitative corpus. This report consolidates both evidence streams for comparison, but they remain distinct in their evidential basis and audit logic.

2.8 Research design and methodological approach

2.8.1 Why participatory valuation is being tested

Lot 5 tests participatory approaches as an innovative, experimental strand focused on public intelligibility, usability, perceived fairness, and trust or legitimacy, alongside technical valuation outputs produced elsewhere in the programme. The methodological aim is not to replace professional valuation practice; rather, it is to understand how different audiences interpret land value problems under structured information conditions, and what that implies for designing an approach that can be communicated responsibly and understood consistently.

The primary value of participatory evidence in this context is diagnostic learning. The workshops are designed to observe how participants reason, what cues they prioritise, where misunderstandings occur, and how structured prompts and materials affect consistency and confidence. These observations support the practical design question: what would a publicly intelligible toolkit need to manage (for example, information limits, heuristics, and differences in judgement) to be usable and defensible.

This is not an attempt to elicit “public prices” as a representative statistic for the population. Sampling and delivery are purposive and controlled: the purpose is to test feasibility and interpretive dynamics, not to generate population-level point estimates or to infer a market distribution from workshop outputs.

2.8.2 Why parcels were not anonymised

Parcels were intentionally not anonymised. This reflects likely real-world deployment conditions, in which parcels would be identifiable, and it enables observation of how local knowledge, place reputation, and location-based preconceptions shape valuation reasoning. This design choice intentionally makes potential location anchoring observable, enabling the toolkit to be stress-tested for bias risks and for how attribute information competes with area perceptions.

2.9 How land values were determined

Lot 5 is presented as an operational chain from recruitment through to auditable evidence capture. To maintain transparency and reproducibility, each step sets out its purpose, the participant's task, the outputs captured, and the associated Quality Assurance and traceability controls.

Step 1 — Recruitment and workshop segmentation (purposive; diagnostic)

The purpose of this step is to assemble a structured set of workshop environments that can test how valuation reasoning varies by audience type and context. Participants take part in a structured workshop session, with the activities set out below. Outputs captured at this stage include attendance records, participant profile categories at a level suitable for analysis and reporting, and workshop-level metadata (date, session type, location/area). For Quality Assurance and traceability, ten workshops were delivered in total and segmented into single-demographic sessions: eight public workshops and two professional workshops. Professional participation primarily included developers (including local authority developers, social housing developers, landlords/SLOs, and planners). Participation was offered bilingually (Welsh or English); in practice, all sessions were conducted in English by participant preference due to the discussion-led format. Workshop headcounts and profile summaries are recorded in the fieldwork count log used for the final programme overview table.

Recruitment was operationalised through two purposive pipelines aligned to the two workshop categories. Public participants were recruited through locally targeted, mixed-channel outreach to secure people with lived experience of the relevant places and practical engagement with local land use. This included promotion via local community groups and organisations, community venues (for example, community halls and similar local settings), local online community channels (including targeted social media outreach), and direct local promotion such as flyer distribution within the selected areas. Recruitment was monitored during fieldwork and adjusted where necessary to maintain viable group sizes within the delivery window and to secure a mix of public perspectives rather than a statistically representative sample.

Professional participants were recruited through targeted outreach to relevant organisations and roles using established sector contacts and stakeholder networks, supplemented by referral-based recruitment where initial contacts recommended further suitable participants. Where feasible, session composition was managed to avoid workshops being dominated by a single organisation and instead to include a mix of organisational types and functions (for example, planning, development, social housing practice, and related land and property roles), supporting rounded discussion and varied professional viewpoints.

Step 2 — Toolkit materials and standardised session delivery

With the workshop set established, the next step is to ensure comparability across sessions by standardising instruments, sequencing, and facilitation prompts. Participants work through method cards, structured worksheets and workbooks, calibration materials, and feedback instruments in a defined order. This produces completed worksheets, structured rankings and feedback, facilitator notes on comprehension and process, and deviation logs where applicable. To control facilitator-induced variability, delivery used standardised method cards, workbooks and worksheets, a calibration drill, and structured feedback instruments (including Forms exports), supported by a facilitator run-sheet to maintain consistent sequencing.

Step 3 — Parcel set design (8 core + 2 local; 10 parcels per workshop)

To enable cross-workshop comparability while also testing the role of local context and knowledge, each session uses a combined parcel set. Participants value a common set of parcels and a locally selected set of parcels within the same structured exercise. Outputs are valuation entries for ten parcels per session, recorded by method and accompanied by rationales and confidence ratings. For Quality Assurance and traceability, each workshop valued ten parcels in total. Eight parcels were common across all workshops to support comparability, while two parcels were locally selected to test the effect of local knowledge and contextual familiarity on reasoning. In some cases, workshops conducted in the same area used the same local parcels, resulting in occasional repetition of the full set of ten; however, the design principle remained “8 common + 2 local”.

Step 4 — Comprehension check and calibration drill (levelling; no priming)

Before the core valuation tasks, a comprehension check and calibration drill are used to establish a minimum shared understanding of the task, units, and materials, and to surface misconceptions early. Participants complete an initial comprehension check and then undertake a structured “toy parcel” calibration exercise using comparable logic. Captured outputs include comprehension check outcomes, facilitator notes on misunderstandings and questions, and calibration exercise work products retained as process evidence. In delivery, the comprehension check was used early in sessions to confirm that participants understood basic concepts, units, and task structure; where understanding fell below the assigned threshold, facilitators provided a recap and re-check. The calibration drill levelled understanding without providing currency anchors; calibration numbers were not entered into the analysis dataset. Facilitator notes taken during calibration formed part of the qualitative corpus for later interpretation of reasoning patterns and comprehension-related variance.

Step 5 — Core valuation exercise (independent first; then discussion)

The core valuation exercise is designed to capture initial individual judgements prior to group influence, and then to observe how reasoning evolves under structured discussion. Participants first complete valuations individually (silent/independent stage) using the simplified methods and parcel evidence packs, and then participate in facilitated discussion. Outputs include individual valuations recorded as £/m² (default), confidence scores on a 0–10 scale, and written rationales, alongside discussion contributions captured via audio recording and facilitator notes. The independent-first design reduces immediate group anchoring and creates a traceable separation between initial individual outputs and discussion-mediated reflections, supporting more defensible interpretation of convergence and divergence patterns.

Step 6 — Methods applied (core set) and professional elicitation add-on

Building on the core exercise, the next step tests a controlled set of simplified valuation methods for clarity and usability, while also capturing professional perspectives on method choice. Participants apply the core methods to the parcel set, and professional participants additionally describe what other methods they would use in practice. Outputs include valuations by parcel × method (Comp, RotR, PFI), confidence scores, and rationales, alongside professional discussion evidence on additional methods used in practice (elicitation prompt, not an additional structured valuation run). For Quality Assurance and traceability, the core set was applied consistently across sessions to maintain comparability, while the professional elicitation prompt was captured as qualitative evidence to inform toolkit refinement and interpretive context.

Step 7 — Structured feedback capture (clarity, fairness, trust; exit survey)

To complement valuation outputs with participant perceptions, structured feedback is captured on method clarity, perceived fairness, and trust in the process. Participants complete card-sort or ranking exercises on clarity and fairness, with reasons recorded, and complete an exit survey including trust and usability items. Outputs include ranked outputs with recorded reasons, Forms exports, and pre-session and exit measures of trust to observe whether exposure to the toolkit shifts expressed trust in the process (interpreted as diagnostic evidence, not as causal impact). Feedback instruments were standardised to support consistent coding of perceptions and to enable alignment between quantitative patterns and qualitative explanations.

Step 8 — Discussion capture and evidence trail (recording, transcription, Quality Control, anonymisation)

Finally, discussion capture is structured to preserve a rigorous evidence trail from workshop discussion through to an analysable, anonymised corpus. Participants take part in facilitated discussion, and outputs include audio recordings, transcripts, an anonymised transcript corpus, and facilitator Quality Control notes and anonymisation checks. Discussions were audio-recorded and transcribed using Whisper via Python. Facilitators quality-checked transcripts prior to anonymisation, because identifiable context assisted accuracy checking and speaker differentiation. Transcripts were then anonymised and entered into the qualitative analysis corpus, and at least one facilitator reviewed transcripts to confirm clarity and appropriateness of anonymisation prior to analysis use.

2.10 Data sources and analysis techniques

2.10.1 Evidence sources

Workshop inputs

- Parcel evidence packs (maps and attribute summaries; locally contextual information allowed by design).
- Method cards and calibration drill materials.
- Facilitator run-sheets to preserve sequencing consistency.

Workshop-generated primary data

- Worksheet/workbook valuations (values by parcel × method), confidence scores, and written rationales.
- Structured feedback outputs (card-sorts, rankings, and Forms exports).
- Facilitator notes and deviation logs (including comprehension issues and process observations).
- Audio recordings and anonymised transcripts.

2.10.2 Analysis approach (quantitative and qualitative; complementary)

Quantitative analysis summarises and compares participant outputs in a way that supports interpretability rather than asserting population estimates. To achieve this, analyses are structured to compare value distributions by method and by parcel, and to compare patterns by workshop type (public versus professional). Dispersion and convergence patterns are examined within each method and across methods, and participant outputs are compared against the surveyor benchmark ranges as an alignment and sense-check exercise (not as validation against a ground truth).

To complement this, qualitative analysis applies thematic analysis to participant rationales, facilitator notes, and anonymised transcripts. This is used to identify the dominant cues and heuristics that shape reasoning (including location anchoring), the ways in which participants trade off access/services versus constraints, how risk factors are recognised and incorporated, and how participants interpret clarity, fairness, and legitimacy or trust. In turn, these themes are also used to identify design improvements to the toolkit and to explain why divergence occurs (for example, comprehension limits, heuristics, information limits, or genuine judgement differences).

Taken together, the method is intentionally insight-forward. Quantitative patterns are interpreted alongside qualitative explanations, enabling a clearer account of why divergence occurs, which is critical for designing a publicly intelligible toolkit.

2.10.3 Standardisation, Data Quality Assurance and integrity controls

Workshop delivery integrity is supported through standardised instruments and sequencing. A consistent facilitator run-sheet was used alongside standardised method cards and worksheets, with a structured flow from orientation to comprehension check, calibration, independent valuation, discussion, and feedback capture. This reduces uncontrolled variation and strengthens comparability across sessions.

The evidence trail is explicitly controlled:

recorded audio → transcription → facilitator quality checks → anonymisation → inclusion in the analysis corpus.

This sequence ensures that the qualitative corpus is both accurate and ethically appropriate for analysis. Within this, missing and invalid handling follows a defined rule. Missing or incomplete valuation entries are excluded from numerical aggregation for that specific calculation, but retained as meaningful process evidence. Facilitators record reasons for non-completion or invalidity in workshop notes and deviation logs, and those reasons are incorporated into lessons-learned interpretation rather than being discarded.

Deviations were recorded contemporaneously (for example, comprehension issues attributable to technical phrasing and individual language fluency). Deviations are treated as relevant evidence: they inform interpretation of output variance and support targeted refinements to materials and facilitation guidance for future deployments.

2.11 Limitations and challenges

Purposive sampling and small group sizes

Workshops are designed for diagnostic learning rather than representativeness.

Implication: results are interpreted as feasibility and interpretability evidence, not as population estimates.

Recruitment and attendance constraints

Participation requires a time commitment (and, where relevant, travel), which can create practical barriers for both public participants and professionals attending in a work-related capacity. To reduce these barriers and support viable attendance public participants were offered a £50 gift voucher as a token of appreciation for their time (with a choice of voucher options). For professional workshops, it was agreed that it would be inappropriate for individuals to receive personal monetary benefit for participation linked to their professional role; instead, participating organisations were invited to nominate a charity to receive a donation in lieu of direct payment. Donations were set at £100 to £200 per participating organisation, depending on attendance levels and time commitment.

Implication: Incentives support delivery feasibility and reduce participation barriers, but they can also influence who is willing or able to attend. For transparency, an auditable record of voucher issuance and charitable donations (including nominated charities and payment confirmation) is retained within the project file and linked to the fieldwork count log.

Non-anonymised parcels

This improves real-world fidelity but elevates location anchoring risk and reputational cue effects.

Implication: potential anchoring is managed analytically (through thematic interpretation and cross-session comparison) rather than removed by design.

Group and facilitation effects

Discussion can influence convergence and the salience of certain cues. The independent-first design mitigates but does not eliminate such effects.

Implication: interpretation distinguishes initial independent outputs from discussion-stage reasoning evidence.

Simplified evidence packs

Inputs are intentionally constrained compared to what professionals might require for site appraisal.

Implication: divergence from benchmarks or across methods may reflect information limits and task design, not simply inconsistent judgement.

Benchmark constraints

Benchmarking draws on one surveyor, and benchmark outputs are ranges rather than point estimates.

Implication: benchmarking is used for context and sense-checking, not validation.

Timing boundary

Fieldwork completion and the analysis lock window define the audited evidence boundary.

Implication: later reflections and subsequent development work are treated as out-of-scope for the audited package.

2.12 Key assumptions

Lot 5 evidence is purposive and diagnostic rather than statistically representative. Its value is in identifying how people reason, what makes methods usable, what undermines confidence, and what is perceived as fair. Recognisable parcels are treated as an intentional design feature because place identity and local knowledge shape real-world interpretation. Non-completion and expressed uncertainty are treated as evidence about cognitive load and information sufficiency; incomplete entries are excluded from any specific quantitative calculation but retained in the qualitative synthesis. Recruitment routes and incentivisation are documented to make the sampling-to-recruitment route explicit for review.

3 Findings

Lot 3 Formula-based Valuation: Findings

This section reports the key findings for Lot 3, and provides an overview of the core land value outputs and observed behaviour of the Lot 3 formula “ladder” across 10,415 analytic parcels within nine Lower Super Output Areas (LSOAs). It sets out the value outputs generated, how those outputs change between rungs, and evidence-led interpretive lessons and diagnostic signals that are observable within the audited exports. Throughout, the analysis avoids attributing causal “drivers” where the required parcel-level trigger fields are not present in the audited extracts.

3.1 Audit boundary and how to read the outputs

All findings in this section refer to the audited export boundary for Lot 3.Unless explicitly defined as an alternative, the findings presented relate to the INSPIRE polygon base formula runs and associated outputs.

In reporting the figures generated by the formula runs, the report prioritises value density and distribution (for example, median £/m² and percentile ranges) over absolute totals. Totals are model-implied aggregates and can be dominated by a small number of very large parcels. This is not an error condition in itself; rather, it is a known interpretive risk when a per-m² rate is applied across heterogeneous parcel areas. Within this framing, the outputs support describing how values behave across the ladder (flat baseline, stepped between-area adjustment, then intended within-area differentiation). However, they do not fully support parcel-level attribution of “why” a given parcel is penalised (for example, which specific access trigger or constraint share drove a reduction), because the synthesis extracts do not include the parcel-level trigger columns required for driver decomposition. In parallel, “White Space” visible in maps is definitional and structural. It represents excluded surfaces and areas that are not parcelised or not valuation-eligible by design, and it must not be interpreted as missing data or “zero value”. Finally, given the experimental nature of this programme of work, the evidence presented is produced for methodological testing. It is not professional valuation advice for individual sites or transactions, and it should not be interpreted as a substitute for professional valuation practice. The report does not claim that any tested method replaces established valuation approaches; rather, it assesses how different approaches operate in practice, what inputs they depend on, how outputs can be audited and explained, and what limitations those dependencies imply for any potential future application.

3.2 Headline findings

Lot 3 modelled land value surface amounting to 82.5% coverage of the total specified land area within the nine LSOA boundaries, using the INSPIRE Index Polygons as the primary parcel fabric. The staged 'Complexity Ladder' of formulas behaved as designed for the first three rungs: Formula 1 (FLAT)[2] established a uniform baseline, Formula 2 (FLAT_RUR) introduced stepped, area-wide adjustments, and Formula 3 (ACCESS) incorporated continuous penalties for accessibility and constraints, producing a differentiated range of indicative values. In contrast, Formula 4 (COMP), a prototype composite model anchored to transaction evidence, produced outputs orders of magnitude larger, confirming its status as a diagnostic tool rather than a policy-ready approach.

The valuations reveal clear differentiation between LSOAs. For instance, under Formula 3, median implied land values range from £75/m² in Bridgend 019D to £165/m² in several other LSOAs, consistent with the applied schedule and penalty structures. A critical learning from constructing this evidence base is the foundational importance of data readiness and geometry governance. Initial tests using multiple open polygon sources (e.g., OS OpenMap Local) delivered only 23% coverage of the total pilot land area within the nine LSOA boundaries, underscoring that a consistent, high-coverage parcel fabric is a non-negotiable prerequisite. Furthermore, robust geometry processing—including harmonisation, validity repair, and controlled dissolving for feature attachment—proved essential for creating an auditable and reproducible analytic pipeline.

Table 3.1 provides the primary compliance summary: for each LSOA, it reports the number of parcels in the audited export, model-implied totals by rung, and the core distributional statistics that are interpretable from the audited extracts.

[2] Formula 1 (F1) is a flat baseline with parcel rates capped at £150/m² for display purposes. The median is therefore 150.0 for all LSOAs. Analytical tables report the implied pre-cap rate (value_gbp / area_m2).

Table 3.1: Summary land value outputs by LSOA and formula rung (Formulas 1–3).

LSOA (code)	Local authority (code)	Parcels (n)	F1 median £/m²[3]	F1 total (model-implied)	F2 median £/m²	F2 total (model-implied)	F2 vs F1	F3 £/m² (p10 / median / p90)	F3 total (model-implied)	F3 vs F2
Gwynedd 009D (W01000114)	Gwynedd (W06000002)	1,879	150.0	£18.29bn	165.0	£20.12bn	+10.0%	146.5 / 165.0 / 165.0	£19.10bn	-5.1%
Flintshire 015A (W01000255)	Flintshire (W06000005)	1,267	150.0	£2.28bn	150.0	£2.28bn	0.0%	107.8 / 150.0 / 150.0	£1.62bn	-28.8%
Powys 011C (W01000449)	Powys (W06000023)	984	150.0	£1.79bn	157.5	£1.88bn	+5.0%	78.8 / 157.5 / 157.5	£1.82bn	-3.4%
Ceredigion 002D (W01000517)	Ceredigion (W06000008)	426	150.0	£12.21m	157.5	£12.82m	+5.0%	154.4 / 157.5 / 157.5	£11.69m	-8.8%
Pembrokeshire 002F (W01000617)	Pembrokeshire (W06000009)	1,690	150.0	£9.19bn	165.0	£10.11bn	+10.0%	151.4 / 165.0 / 165.0	£9.81bn	-3.0%
Bridgend 019D (W01001045)	Bridgend (W06000013)	939	150.0	£106.59m	150.0	£106.59m	0.0%	75.0 / 75.0 / 75.0	£69.67m	-34.6%
Rhondda Cynon Taf 001F (W01001233)	Rhondda Cynon Taf (W06000016)	1,617	150.0	£10.86bn	165.0	£11.95bn	+10.0%	163.2 / 165.0 / 165.0	£11.76bn	-1.5%
Monmouthshire 006F (W01001597)	Monmouthshire (W06000019)	947	150.0	£191.40m	142.5	£181.83m	-5.0%	71.3 / 142.5 / 142.5	£162.12m	-10.8%
Cardiff 032H (W01002019)	Cardiff (W06000015)	666	150.0	£217.99m	157.5	£228.89m	+5.0%	78.8 / 157.5 / 157.5	£159.63m	-30.3%

[3] Formula 1 Cap Explanation: The displayed rate per square metre (rate_gbp_per_m2) is capped at £150 for mapping and public-facing outputs. For analytical reporting (e.g., Table 3.1), rates are expressed as the implied pre-cap rate (value_gbp / area_m2). This cap creates uniform median values of 150.0 across LSOAs in Formula 1 outputs.

The findings should be interpreted alongside a small set of practical risk signals identified during delivery. First, output sensitivity is not only conceptual, in the sense that different rungs produce materially different aggregate scales; it is also operational, because small issues in data completeness or classification can propagate quickly through valuation. Second, the approach remains strongly dependent on consistent geometry handling, including CRS discipline, clipping, validity repair and union-area coverage calculation. In practice, this means the method's transparency advantage is realised only where the pipeline is executed with consistent spatial rules and logged provenance, rather than through ad hoc, spreadsheet-style manipulation of intermediate figures.

Against this operational backdrop, model-implied totals across the nine-LSOA geographies are approximately £42.94bn for Formula 1, £46.87bn for Formula 2, and £44.51bn for Formula 3. This matters for interpretation, because the ladder should not be communicated as a monotonic "more complexity → higher value" sequence. In this audited run, Formula 2 uplifts totals by approximately 9.2% relative to Formula 1, and Formula 3 then offsets part of that uplift (approximately 5.0% relative to Formula 2), resulting in a net change of approximately 3.7% from Formula 1 to Formula 3.

To support a clear reading across rungs, Table 3.2 provides the rung-to-rung change metrics in a single view, including the derived net change from Formula 1 to Formula 3 (derived directly from model-implied totals).

Table 3.2: Change metrics across rungs (LSOA-level; derived from model-implied totals)

LSOA (code)	F2 vs F1	F3 vs F2	F3 vs F1 (net)
Gwynedd 009D (W01000114)	+10.0%	-5.1%	+4.4%
Flintshire 015A (W01000255)	0.0%	-28.8%	-28.8%
Powys 011C (W01000449)	+5.0%	-3.4%	+1.4%
Ceredigion 002D (W01000517)	+5.0%	-8.8%	-4.2%
Pembrokeshire 002F (W01000617)	+10.0%	-3.0%	+6.7%
Bridgend 019D (W01001045)	0.0%	-34.6%	-34.6%
Rhondda Cynon Taf 001F (W01001233)	+10.0%	-1.5%	+8.3%
Monmouthshire 006F (W01001597)	-5.0%	-10.8%	-15.3%
Cardiff 032H (W01002019)	+5.0%	-30.3%	-26.8%

Two diagnostic points are visible immediately and should be treated as governance-relevant signals rather than minor variation. First, two LSOAs show a “silent step” where Formula 2 does not uplift relative to the flat baseline (Flintshire 015A and Bridgend 019D). Second, one LSOA reduces under Formula 2 (Monmouthshire 006F at −5.0%). These behaviours matter because Formula 2 is the rung intended to represent a transparent, stepped schedule; silent or reversing steps imply that the underlying flag/threshold logic requires careful scrutiny before any operational use.

To ensure totals are not over-interpreted, Table 3.3 provides concentration diagnostics under Formula 3 at LSOA level. These show the extent to which a small number of parcels dominate model-implied totals within each area.

Table 3.3: Concentration diagnostics under Formula 3 (LSOA-level)

LSOA (code)	Largest parcel share of F3 total	Top 10 parcels share	Top 1% share	Gini (F3 parcel values)	Max parcel value (F3)
Gwynedd 009D (W01000114)	3.93%	25.44%	39.53%	0.933	£750.44m
Flintshire 015A (W01000255)	18.62%	63.13%	67.98%	0.950	£301.56m
Powys 011C (W01000449)	46.46%	78.31%	78.31%	0.971	£843.33m
Ceredigion 002D (W01000517)	7.74%	29.75%	23.85%	0.537	£0.90m
Pembrokeshire 002F (W01000617)	2.32%	17.44%	24.80%	0.865	£227.50m
Bridgend 019D (W01001045)	42.58%	72.18%	72.18%	0.854	£29.67m
Rhondda Cynon Taf 001F (W01001233)	6.97%	33.45%	42.53%	0.928	£819.68m
Monmouthshire 006F (W01001597)	12.10%	57.62%	57.62%	0.856	£19.62m
Cardiff 032H (W01002019)	14.47%	47.47%	40.75%	0.834	£23.09m

These diagnostics should be treated as the upper bound of value concentration that can arise under the audited polygon fabric and a binding £/m² cap.

3.3 Area profiles

The LSOA summaries below provide defensible cues on “how to read” the model outputs for each geographic area, drawing only on behaviours that are observable within the audit boundary. For consistency, each LSOA profile reports the same set of measures. ‘Coverage’ is the proportion of land area within the LSOA boundary that is represented by the valuation parcel fabric (i.e., the land area to which the model can assign values within this audited run). The rung-to-rung change metrics (F2 vs F1; F3 vs F2) show how totals change as additional rules are applied. Because map outputs are display-capped, each profile also reports the implied pre-cap £/m² distribution (P10 / median / P90), which shows how values are distributed within the area rather than relying on a single headline figure. Additionally, percentiles summarise the distribution in plain terms. P90 means the 90th percentile: 90% of parcels have values at or below this figure, and the top 10% of parcels are above it. Similarly, P10 is the point below which the bottom 10% of parcels sit. The concentration measures (largest-parcel share, top ten share, and the Gini coefficient) provide a defensible indication of whether total value in an area is dominated by a small number of very large parcels. The Gini coefficient ranges from 0 to 1, where 0 would indicate that value is evenly spread across parcels and 1 indicates extreme concentration. A Gini value such as 0.933 should therefore be read as very high concentration, meaning that a relatively small number of parcels account for a disproportionate share of the total, and totals should be interpreted alongside these distributional measures.

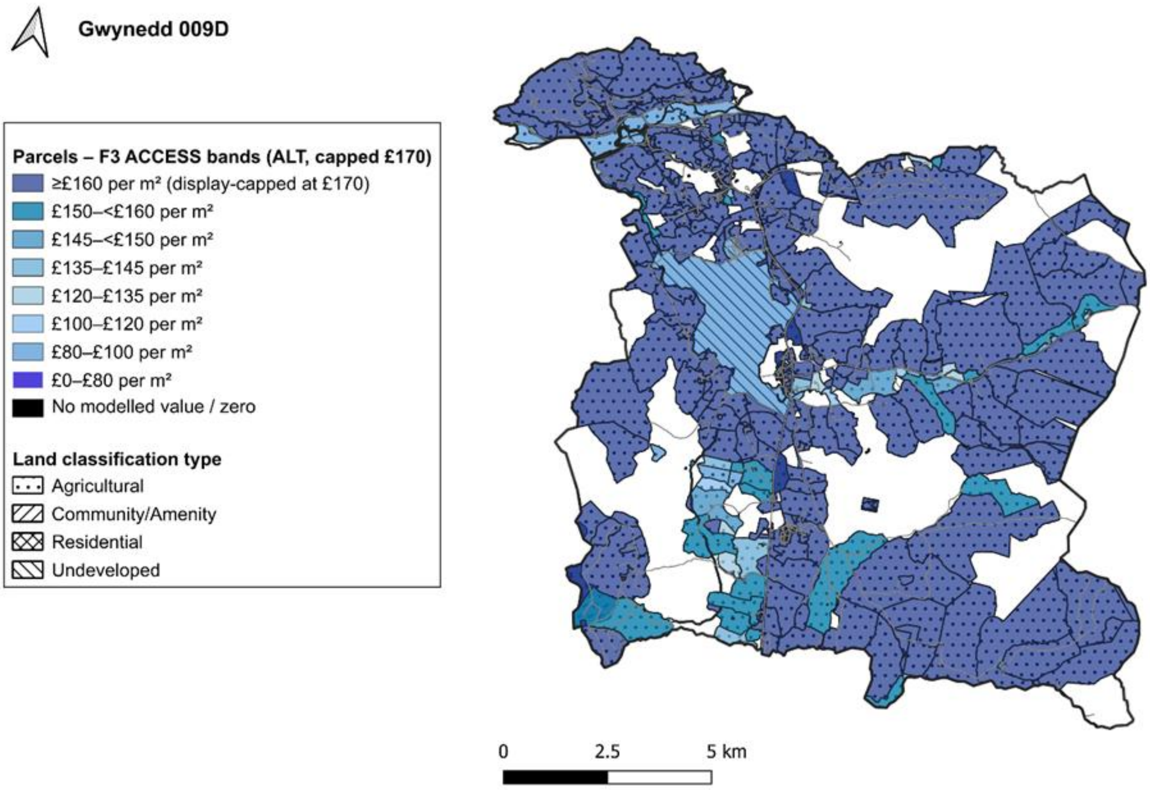
Gwynedd 009D

Coverage is 0.737. At LSOA level, applying Formula 2 increases total value by +10.0% compared with Formula 1, and Formula 3 then reduces total value by -5.1% compared with Formula 2. In cash terms this equates to F1 £18.29bn, F2 £20.12bn, and F3 £19.10bn, across 1,879 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 165.0 £/m². The bottom 10% of parcels are at or below 146.5 £/m², and the top 10% are at or above 165.0 £/m².

For the F3 parcel value distribution, the median parcel value is £67,482. The 90th percentile (P90) is £15,204,151, and the maximum parcel value is £750,440,433. Value is materially concentrated: the highest-value parcel accounts for 3.93% of total F3 value (max parcel value £750,440,433), the top ten parcels account for 25.44%, and the Gini coefficient is 0.933.

Figure 3.1: Bluescale rendered parcel map for Gwynedd 009D.



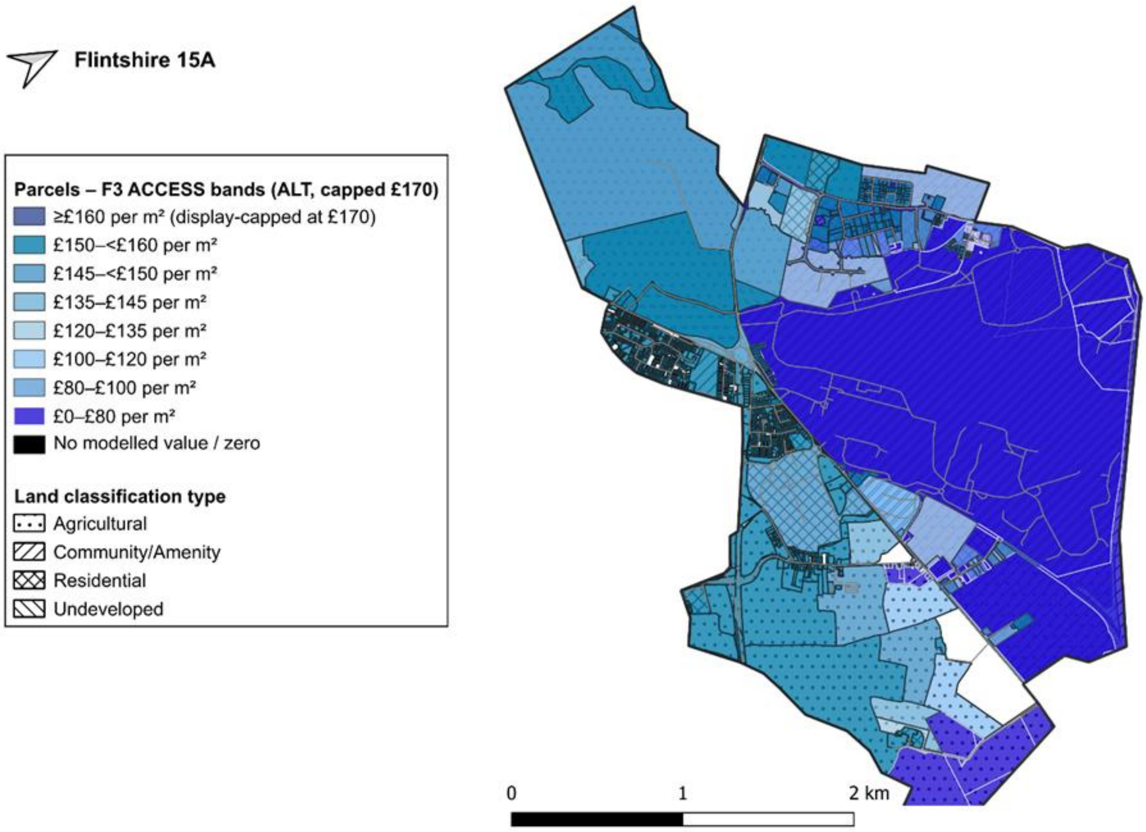
Flintshire 015A

Coverage is 0.967. At LSOA level, Formula 2 changes total value by 0.0% compared with Formula 1, and Formula 3 then reduces total value by -28.8% compared with Formula 2. In cash terms this equates to F1 £2.28bn, F2 £2.28bn, and F3 £1.62bn, across 1,267 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 150.0 £/m². The bottom 10% of parcels are at or below 107.8 £/m², and the top 10% are at or above 150.0 £/m².

For the F3 parcel value distribution, the median parcel value is £45,936. The 90th percentile (P90) is £529,789, and the maximum parcel value is £301,558,477. Value is materially concentrated: the highest-value parcel accounts for 18.62% of total F3 value (max parcel value £301,558,477), the top ten parcels account for 63.13%, and the Gini coefficient is 0.950.

Figure 3.2: Bluescale rendered parcel map for Flintshire 015A



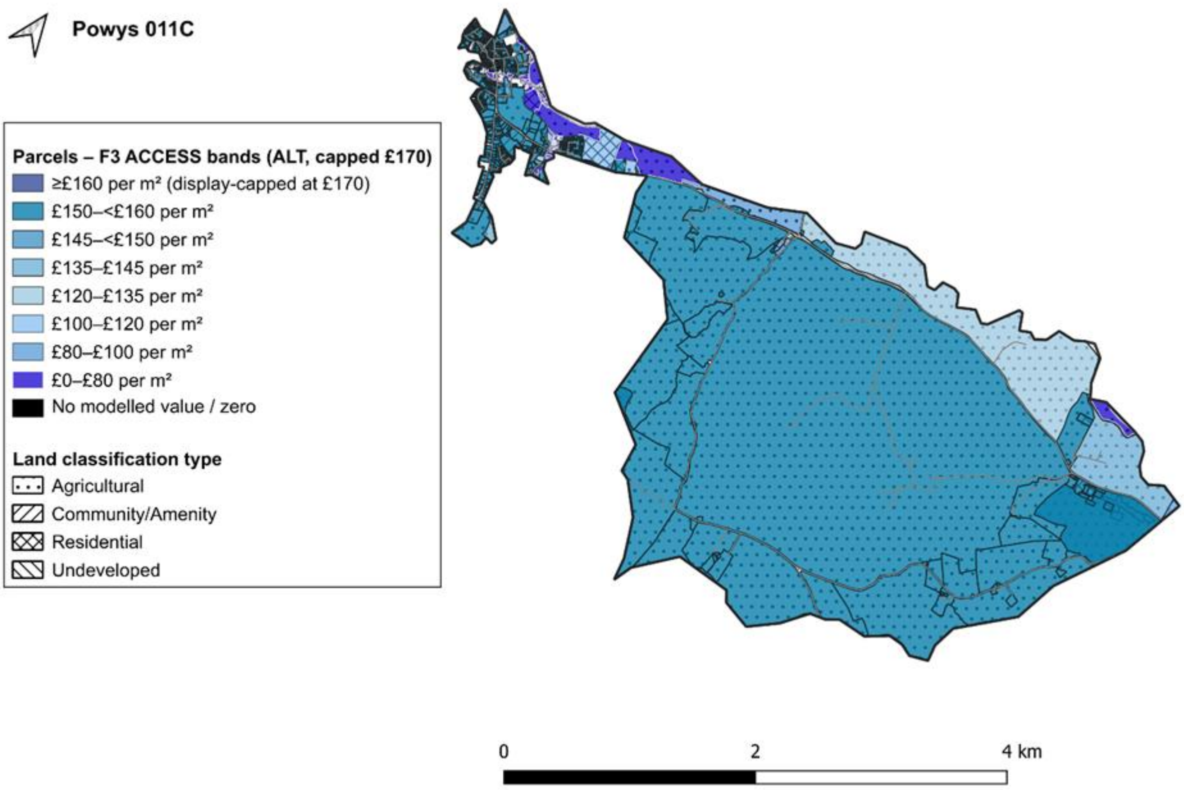
Powys 011C

Coverage is 0.976. At LSOA level, applying Formula 2 increases total value by +5.0% compared with Formula 1, and Formula 3 then reduces total value by -3.4% compared with Formula 2. In cash terms this equates to F1 £1.79bn, F2 £1.88bn, and F3 £1.82bn, across 984 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 157.5 £/m². The bottom 10% of parcels are at or below 78.8 £/m², and the top 10% are at or above 157.5 £/m².

For the F3 parcel value distribution, the median parcel value is £41,746. The 90th percentile (P90) is £355,760, and the maximum parcel value is £843,332,606. Value is materially concentrated: the highest-value parcel accounts for 46.46% of total F3 value (max parcel value £843,332,606), the top ten parcels account for 78.31%, and the Gini coefficient is 0.971.

Figure 3.3: Bluescale rendered parcel map for Powys 011C.



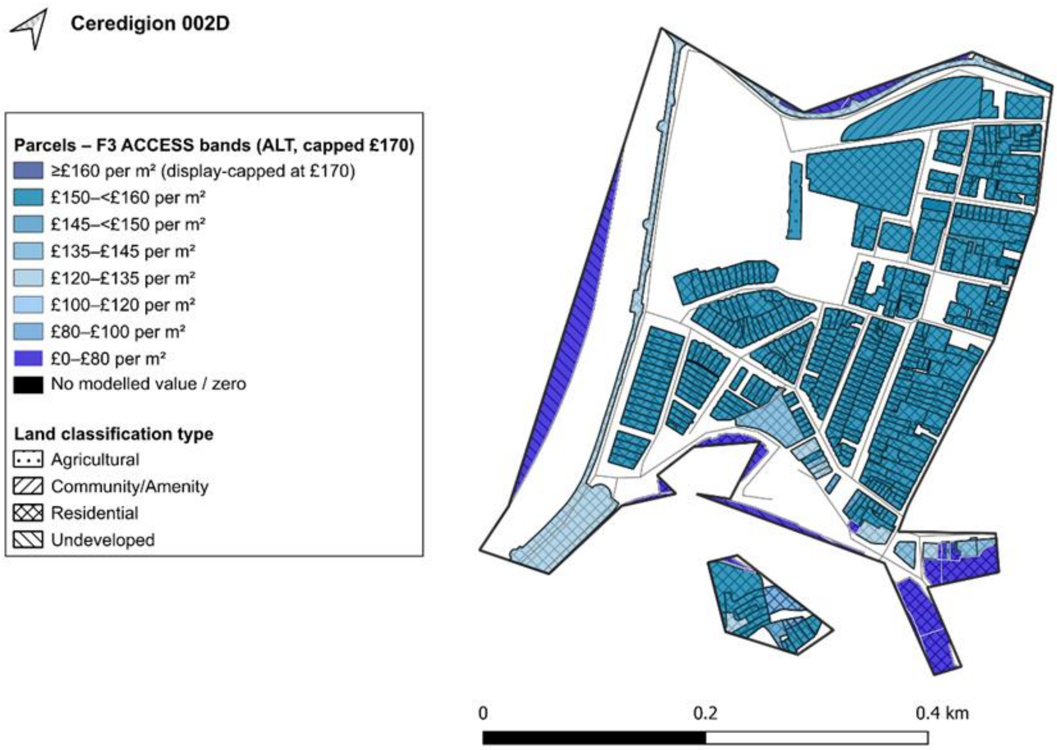
Ceredigion 002D

Coverage is 0.492. At LSOA level, applying Formula 2 increases total value by +5.0% compared with Formula 1, and Formula 3 then reduces total value by -8.8% compared with Formula 2. In cash terms this equates to F1 £12.21m, F2 £12.82m, and F3 £11.69m, across 426 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 157.5 £/m². The bottom 10% of parcels are at or below 154.4 £/m², and the top 10% are at or above 157.5 £/m².

For the F3 parcel value distribution, the median parcel value is £15,528. The 90th percentile (P90) is £42,278, and the maximum parcel value is £904,652. Value is materially concentrated: the highest-value parcel accounts for 7.74% of total F3 value (max parcel value £904,652), the top ten parcels account for 29.75%, and the Gini coefficient is 0.537.

Figure 3.4: Bluescale rendered parcel map for Ceredigion 002D.



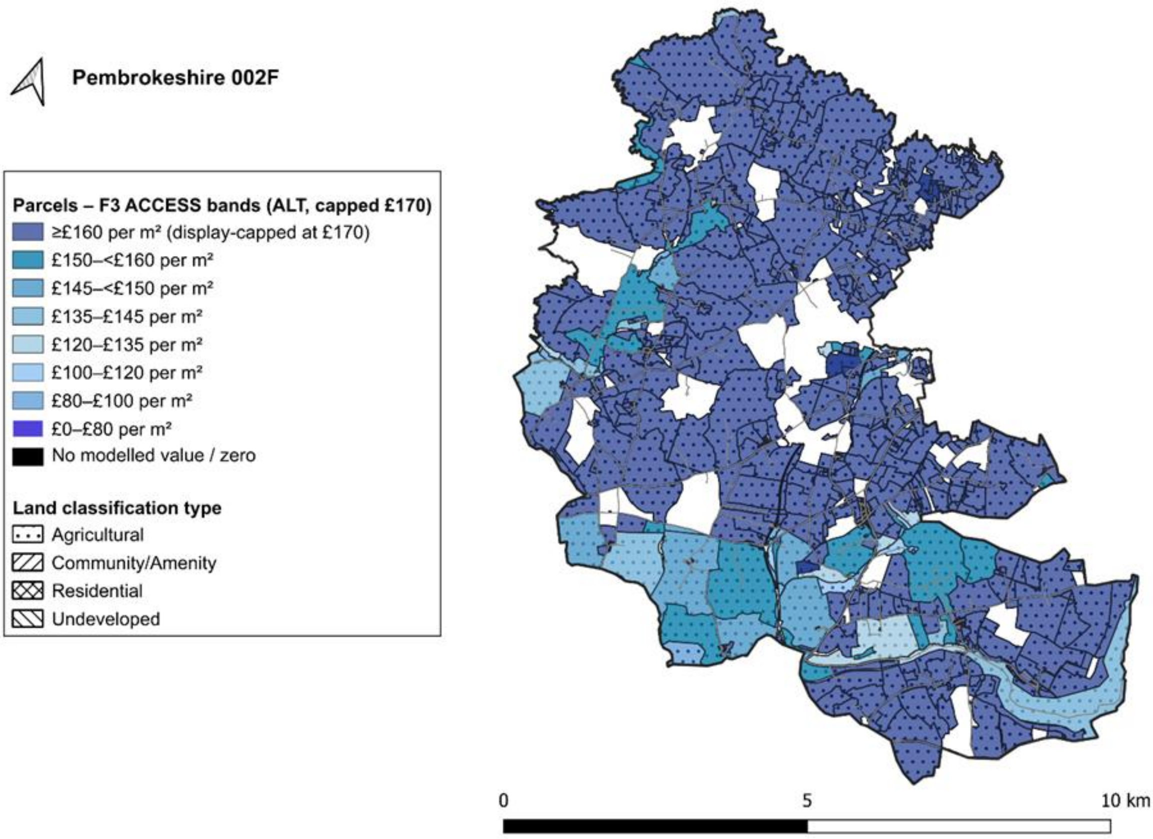
Pembrokeshire 002F

Coverage is 0.857. At LSOA level, applying Formula 2 increases total value by +10.0% compared with Formula 1, and Formula 3 then reduces total value by -3.0% compared with Formula 2. In cash terms this equates to F1 £9.19bn, F2 £10.11bn, and F3 £9.81bn, across 1,690 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 165.0 £/m². The bottom 10% of parcels are at or below 151.4 £/m², and the top 10% are at or above 165.0 £/m².

For the F3 parcel value distribution, the median parcel value is £240,349. The 90th percentile (P90) is £13,636,855, and the maximum parcel value is £227,499,055. Value is materially concentrated: the highest-value parcel accounts for 2.32% of total F3 value (max parcel value £227,499,055), the top ten parcels account for 17.44%, and the Gini coefficient is 0.865.

Figure 3.5: Bluescale rendered parcel map for Pembrokeshire 002F.



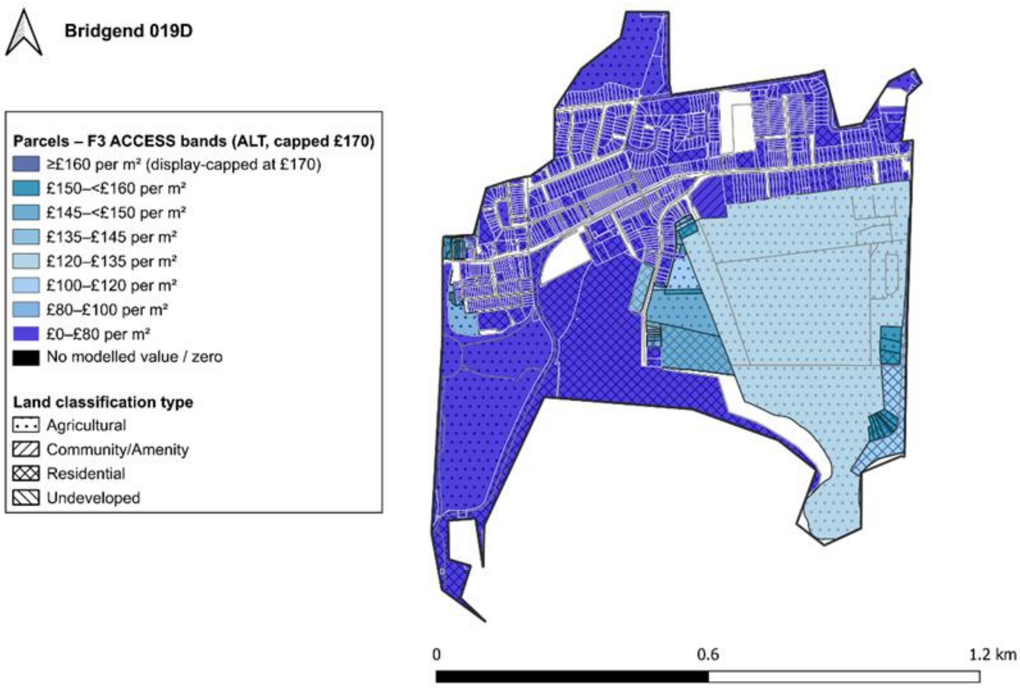
Bridgend 019D

Coverage is 0.890. At LSOA level, Formula 2 changes total value by 0.0% compared with Formula 1, and Formula 3 then reduces total value by -34.6% compared with Formula 2. In cash terms this equates to F1 £106.59m, F2 £106.59m, and F3 £69.67m, across 939 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 75.0 £/m². The bottom 10% of parcels are at or below 75.0 £/m², and the top 10% are at or above 75.0 £/m².

For the F3 parcel value distribution, the median parcel value is £13,825. The 90th percentile (P90) is £34,068, and the maximum parcel value is £29,666,844. Value is materially concentrated: the highest-value parcel accounts for 42.58% of total F3 value (max parcel value £29,666,844), the top ten parcels account for 72.18%, and the Gini coefficient is 0.854.

Figure 3.6: Bluescale rendered parcel map for Bridgend 019D.



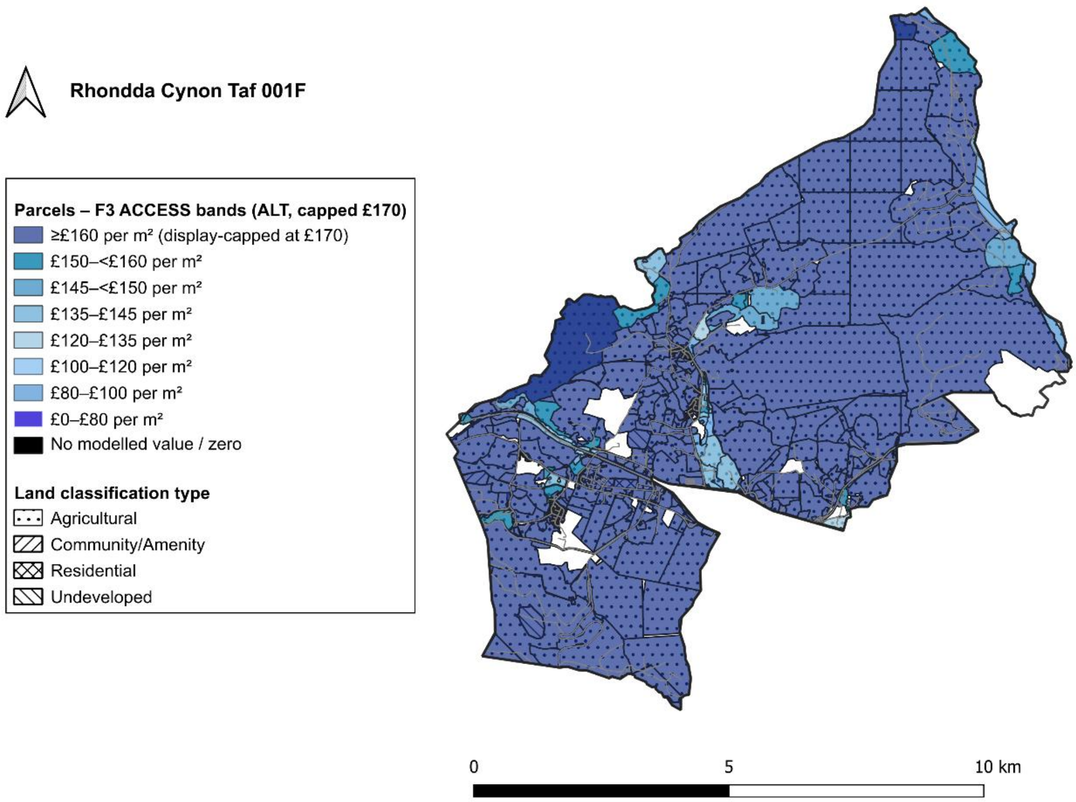
Rhondda Cynon Taf 001F

Coverage is 0.945. At LSOA level, applying Formula 2 increases total value by +10.0% compared with Formula 1, and Formula 3 then reduces total value by -1.5% compared with Formula 2. In cash terms this equates to F1 £10.86bn, F2 £11.95bn, and F3 £11.76bn, across 1,617 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 165.0 £/m². The bottom 10% of parcels are at or below 163.2 £/m², and the top 10% are at or above 165.0 £/m².

For the F3 parcel value distribution, the median parcel value is £84,244. The 90th percentile (P90) is £9,549,179, and the maximum parcel value is £819,684,840. Value is materially concentrated: the highest-value parcel accounts for 6.97% of total F3 value (max parcel value £819,684,840), the top ten parcels account for 33.45%, and the Gini coefficient is 0.928.

Figure 3.7: Bluescale rendered parcel map for Rhondda Cynon Taf 001F.



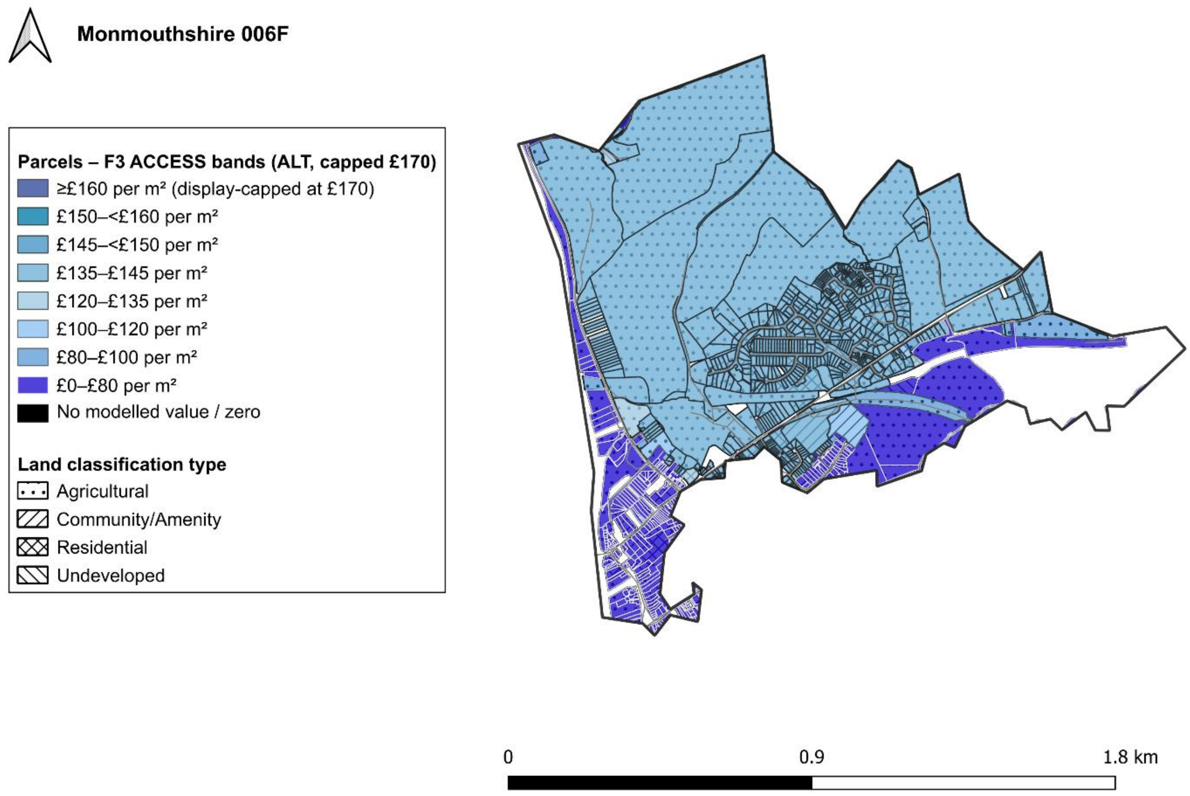
Monmouthshire 006F

Coverage is 0.862. At LSOA level, Formula 2 reduces total value by -5.0% compared with Formula 1, and Formula 3 then reduces total value by -10.8% compared with Formula 2. In cash terms this equates to F1 £191.40m, F2 £181.83m, and F3 £162.12m, across 947 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 142.5 £/m². The bottom 10% of parcels are at or below 71.3 £/m², and the top 10% are at or above 142.5 £/m².

For the F3 parcel value distribution, the median parcel value is £32,168. The 90th percentile (P90) is £125,832, and the maximum parcel value is £19,616,469. Value is materially concentrated: the highest-value parcel accounts for 12.10% of total F3 value (max parcel value £19,616,469), the top ten parcels account for 57.62%, and the Gini coefficient is 0.856.

Figure 3.8: Bluescale rendered parcel map for Monmouthshire 006F.



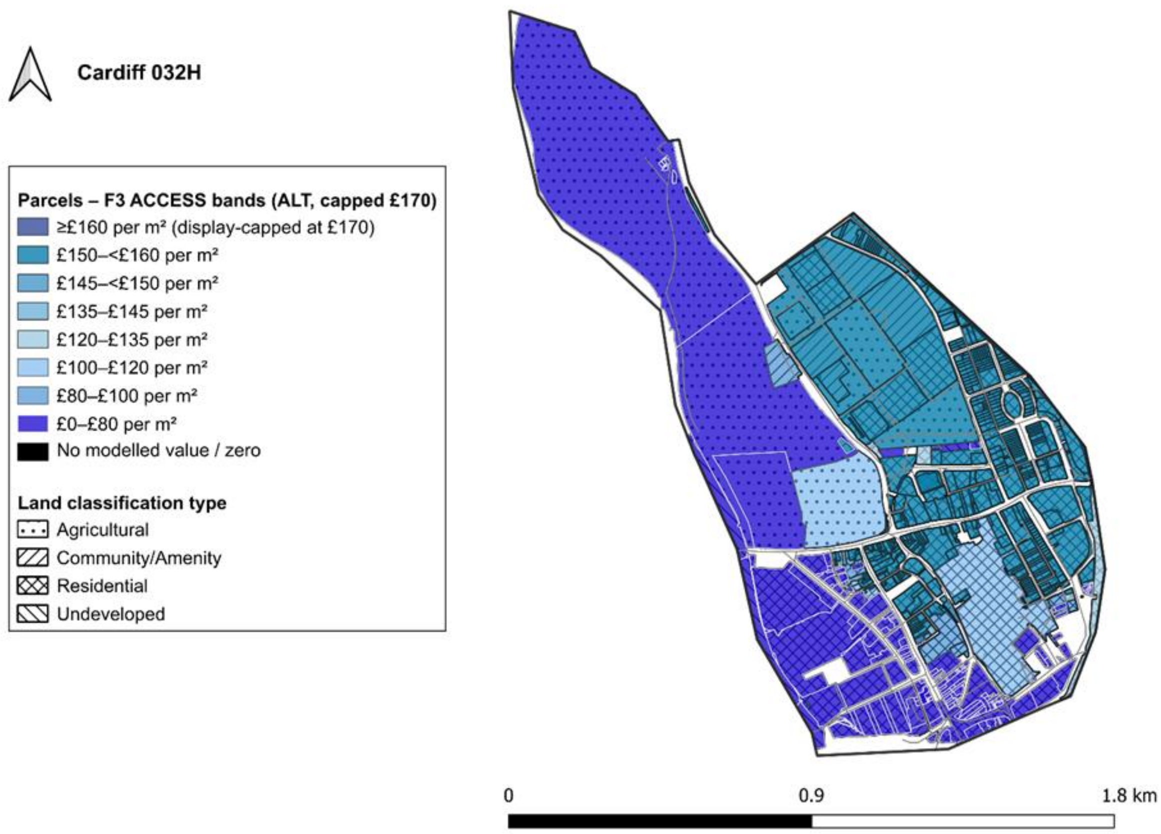
Cardiff 032H

Coverage is 0.840. At LSOA level, applying Formula 2 increases total value by +5.0% compared with Formula 1, and Formula 3 then reduces total value by -30.3% compared with Formula 2. In cash terms this equates to F1 £217.99m, F2 £228.89m, and F3 £159.63m, across 666 parcels.

For the implied pre-cap F3 rate (£/m²), the median is 157.5 £/m². The bottom 10% of parcels are at or below 78.8 £/m², and the top 10% are at or above 157.5 £/m².

For the F3 parcel value distribution, the median parcel value is £44,687. The 90th percentile (P90) is £296,261, and the maximum parcel value is £23,091,377. Value is materially concentrated: the highest-value parcel accounts for 14.47% of total F3 value (max parcel value £23,091,377), the top ten parcels account for 47.47%, and the Gini coefficient is 0.834.

Figure 3.9: Bluescale rendered parcel map for Cardiff 032H.



3.4 Findings narrative: what each rung adds

3.4.1 Coverage and variation of coverage across the nine pilot LSOAs

Coverage is a practical constraint on what can be valued and what can be claimed. In the INSPIRE redeployment, the Lot 3 parcel fabric covers 82.5% of the total land area inside the nine pilot LSOA boundaries. The remaining area is not “missing data”: it is largely explained by structural exclusions and definitional “White Space” created by how the analytic fabric is constructed (for example, operational transport surfaces and other excluded surfaces).

Coverage also varies materially between LSOAs. In this research, INSPIRE parcel-fabric coverage ranges from 0.492 (Ceredigion 002D) to 0.976 (Powys 011C). This variation matters because it affects how confidently readers can interpret totals and map “gaps” across areas. A lower coverage figure usually indicates that a larger share of the LSOA is outside the valuation surface by design (or by dataset boundary behaviour), not that the method has selectively “missed” parcels. The practical implication for Wales-wide scaling is therefore not only “high overall coverage”, but consistent and publishable coverage reconciliation (i.e., an explicit statement of what is in scope, what is excluded, and how that differs by geography).

3.4.2 Formula 1 (baseline): what it is and what it proves

Formula 1 is a deliberately flat baseline: it assigns a constant £150/m² across all valuation-eligible parcels in the audited export. Its purpose is not realism; it is control. A flat baseline establishes a transparent reference point that supports auditability and clear change attribution when subsequent rungs are applied. In practical terms, Formula 1 allows the programme to observe, and report, the direction and magnitude of the effects introduced by each additional layer of logic without conflating them with baseline differences.

3.4.3 Formula 2 (stepped between-LSOA differences): what it adds and what it does not add

Formula 2 is designed to introduce stepped differences between areas through a schedule-based adjustment, while remaining uniform within each LSOA. In the analytical outputs, this step behaviour is visible in the implied pre-cap median rates: under Formula 2, medians rise to 165.0 £/m² in some areas and to 157.5 £/m² in others, two LSOAs show no median uplift (Flintshire 015A and Bridgend 019D), and one shows a decrease (Monmouthshire 006F, to 142.5 £/m²). These patterns reflect the scheduled multipliers and provide the intended between-area differentiation.

However, the audited outputs also show that stepped transparency is highly sensitive to governance logic. Two areas display a silent step (Formula 2 equals Formula 1 in Flintshire 015A and Bridgend 019D), and one area reduces under Formula 2 (Monmouthshire 006F at −5.0% relative to the baseline). These are diagnostic signals because they imply that threshold, flag, or classification conditions are determining whether the rung “activates” at all. If this rung were to be used beyond methodological testing, the governing rules would require publication-grade clarity (including explicit thresholds, boundary conditions, and parameter control) to avoid misinterpretation and to support public intelligibility.

3.4.4 Formula 3 (access penalty): first within-LSOA differentiation and its implications

Formula 3 introduces within-LSOA differentiation by applying a downside-only penalty structure. Unlike the display-capped map outputs, this differentiation is now visible in the implied pre-cap rate distributions reported in Table 3.1 (p10 / median / p90), showing how penalties shift a share of parcels below the LSOA median while leaving upper tails unchanged where ceilings bind.

The magnitude of Formula 2 to Formula 3 reduction varies substantially between pilot areas, ranging from modest compression (for example, −1.5% in Rhondda Cynon Taf 001F; −3.0% in Pembrokeshire 002F; −3.4% in Powys 011C) to material compression (−28.8% in Flintshire 015A and −30.3% in Cardiff 032H) and very large compression in a diagnostic case (−34.6% in Bridgend 019D). This range is an interpretable behavioural result: it indicates that Formula 3 is the rung that meaningfully discriminates within areas, but that its effect size is highly variable by geography.

At the total level, the audited outputs show that Formula 2 uplifts totals by approximately 9.2% relative to Formula 1, and Formula 3 then offsets part of that uplift (approximately -5.0% relative to Formula 2). The net effect is a 3.7% change from Formula 1 to Formula 3 at the total level. This matters for how the ladder should be explained: it is not accurate to imply that additional complexity necessarily increases value in aggregate.

For mapping and visualisation, parcel-level £/m² rates are capped at 150.00, so map plates will not show the full differentiation produced by the model. Where differentiation between rungs is required for interpretation, it should be drawn from the implied pre-cap rate summaries (Table 3.1), the LSOA-level change metrics (Table 3.2), and the parcel-value distribution and concentration diagnostics (Table 3.3 and the LSOA profiles).

3.5 Distributional behaviour and concentration risk

A central empirical feature of the audited outputs is that rate tails and value tails behave differently.

The implied pre-cap rate schedule shows meaningful variation across LSOAs and rungs. F2 median rates range from 142.5 to 165 £/m², while F3 penalties introduce further dispersion, with P10 rates as low as 71.3 £/m² (Monmouthshire) and medians as low as 75 £/m² (Bridgend). Despite this, parcel values remain highly skewed due to extreme parcel area inequality, as evidenced by the Gini coefficients (0.537 to 0.971) and top-parcel concentration shares.

Parcel-level values under Formula 3 remain highly skewed (heavy-tailed) despite the uniform rate cap. Across LSOAs, the median parcel value ranges from £13,825 to £240,349, while the 90th percentile ranges from £34,068 to £15,204,151. The 99th percentile ranges from £199,367 to £253,138,659, and maximum parcel values range from £904,652 to £843,332,606.

Across LSOAs, concentration under Formula F3 varies widely. The share of total F3 value represented by the single highest-value parcel ranges from 2.32% (Pembrokeshire 002F) to 46.46% (Powys 011C). The top ten parcels' share ranges from 17.44% (Pembrokeshire 002F) to 78.31% (Powys 011C). Gini coefficients range from 0.537 (Ceredigion 002D) to 0.971 (Powys 011C). This reinforces that LSOA-level totals can be driven by a small number of parcels, and that distributional summaries should be read alongside totals when interpreting outputs.

Taken together, the distribution and concentration diagnostics indicate that LSOA-level totals can be sensitive to a small number of large parcels, even where the £/m² schedule is capped. This is a material interpretive consideration for policy communication: summary totals should be presented alongside distributional summaries and clear guardrails on how outputs should, and should not, be interpreted.

3.6 Class coverage and interpretive limits

The audited export includes four NLUD classes (Agricultural, Residential, Undeveloped, Community/Amenity). Commercial/Industrial does not appear in the audited export for the nine LSOAs. In this deployment, this reflects the classification-transfer boundary of the redeployment: the fallback WorldCover mapping used only to resolve residual unclassified parcels does not include a Commercial/Industrial category. Commercial/Industrial therefore cannot be populated by the fallback mapping and will remain absent where a direct transfer from source polygons is not available. This should be treated as an evidence boundary of the audited export, not as evidence that no commercial land exists in the selected LSOAs.

The audited class mix indicates that parcel counts and land area are distributed very differently by class. Residential accounts for 58.6% of parcels but only 0.7% of parcelised area, whereas Agricultural accounts for 28.4% of parcels and 60.7% of parcelised area (with Undeveloped contributing 35.1% of area and Community/Amenity 4.2%). This means that LSOA totals and concentration metrics are primarily shaped by large-area Agricultural and Undeveloped polygons rather than by the more numerous Residential parcels. This should be made explicit in interpretation to avoid conflating parcel counts with land surface coverage.

3.7 Formula 4 (COMP): prototype/diagnostic findings, explicitly not policy-ready

3.7.1 What was attempted

Formula 4 represents an attempted step towards a more evidence-linked, composite approach intended to generate a calibrated demonstrator surface. In programme terms, this rung is designed to explore feasibility and the additional governance requirements that arise when moving beyond transparent schedules towards calibration against transaction-linked evidence.

3.7.2 What the outputs look like (facts only)

In this audited run, Formula 4 outputs exhibit qualitatively different behaviour to Formulas 1–3 and are therefore treated as diagnostic. At LSOA level, Formula 4 totals fall in the billions to trillions (for example, £10,852,856,118,097.27 in Gwynedd 009D and £5,894,391,030,971 in Rhondda Cynon Taf 001F). A diagnostic check indicates the implied Formula 4 rates are orders of magnitude higher than Formula 3, consistent with a calibration scaling error in this run.

Parcel-level concentration under Formula 4 is not reported in the current artefact set. Any statements about dominance by single parcels or heavy-tail behaviour under Formula 4 should be avoided unless a dedicated parcel-level diagnostic extract is generated and Quality Assurance approved.

3.7.3 Why Formula 4 did not work

Formula 4 is a diagnostic rung that uses a different scaling regime from F1–F3. Unlike the core rungs, it does not align to the capped rate logic (e.g., the 150 £/m² presentation cap), and it applies a separate internal scaling to a composite score. This means the resulting totals are orders of magnitude larger than F1–F3 and must not be interpreted as market-comparable outputs. Its purpose is diagnostic sensitivity checking, not valuation. The implication is that F4 outputs are not directly comparable to F1–F3 totals, rates, or distributions and must not be used in cross-rung comparison or reporting.

3.8 What these findings imply

Within the context of these findings and results, there is scope to interpret what theses outputs imply for interpretability and for the practical fitness of the Lot 3 ladder as an experimental valuation design. Contextually, this is not to claim market “accuracy” in a professional valuation sense, but to identify where the outputs are behaviourally legible and decision-relevant (adequate), where interpretability is structurally constrained by the audited export boundary, and what risks and update constraints follow from the evidence.

3.8.1 Scale and granularity: where the method works well, and where it breaks down

The audited evidence indicates that the ladder is most interpretable at aggregated geographies (LSOA-level) and as a comparative “value density signal” rather than as a parcel-by-parcel explanation engine. At LSOA level, Table 3.1 provides stable medians and percentile ranges (p10/median/p90 under Formula 3) that support clear statements about between-area differences and within-area dispersion, while Table 3.2 provides rung-to-rung change metrics that show how schedules interact. Taken together, this supports defensible interpretation in terms of “how values behave across the ladder” and “how distributions shift” across areas.

By contrast, interpretability becomes materially weaker when attempting parcel-level attribution of “why” a parcel is penalised or uplifted. As already set out in the audit boundary note, the audited synthesis extracts do not include the parcel-level trigger fields required for driver decomposition. Consequently, parcel-level narratives should avoid causal explanations (for example, which access trigger or constraint share caused a reduction) unless the audited package is extended to include auditable trigger and decomposition layers.

Interpretability also reduces when totals are used as the primary lens. The totals in Table 3.1 are model-implied aggregates that can be dominated by a small number of very large parcels, and Table 3.3 shows that concentration is material in several LSOAs (for example, high top-share metrics and Gini values). This does not indicate an error condition in itself, but it does mean that totals are not a reliable proxy for typical parcel behaviour and are in sufficient for interpretation without accompanying distributional and concentration diagnostics.

3.8.2 Performance across areas: what the evidence supports saying

Across the nine-LSOAs, the evidence supports a clear statement that the ladder produces meaningful differentiation between areas and, under Formula 3, within areas. Under Formula 3, Table 3.1 shows medians spanning a wide range (for example, low medians in some LSOAs versus higher medians in others), while p10 values show that penalties can materially compress the lower tail even where the median remains relatively high. In parallel, Table 3.2 shows that the net effect of the ladder is not monotonic: Formula 2 uplifts can be partly or substantially offset by Formula 3 penalties, yielding net changes relative to Formula 1 that vary substantially by area. This is a key interpretability point, because it means the ladder must be described as interacting schedules rather than as a simple progression towards higher values.

The audited evidence also supports a defensible statement that distributional behaviour differs materially by area. Table 3.3 concentration diagnostics show that some LSOAs exhibit extreme skew (high Gini and high top-share metrics), implying that "headline totals" can remain stable while distributional equity and dominance effects vary materially. For interpretability, this means that area comparisons should routinely include both central tendency (median) and dispersion/concentration (percentiles, top-share, Gini), rather than presenting a single headline value.

3.8.3 Accuracy versus adequacy: what can and cannot be claimed from this evidence

The audited outputs do not support claims of market accuracy in a professional valuation sense, because the ladder is not presented as a transaction-calibrated valuation model for individual sites and because, within the audited package, the evidence base is structured for methodological testing rather than for market valuation verification. However, the evidence does support "adequacy" claims in a narrower and defensible sense: behavioural legibility of rung effects at aggregated scale, internal coherence of rung-to-rung changes as interacting schedules, and stability of interpretive signals when interpreted through medians, percentiles, and concentration diagnostics rather than solely through totals or cap-bound map products.

Accordingly, the report can defensibly frame Lot 3 as producing an indicative land value density signal that is suitable for testing explainability, governance-intelligibility, and operational feasibility. At the same time, it should avoid presenting the outputs as definitive absolute valuations or as substitutes for professional valuation practice, and it should avoid parcel-level causal attribution unless the audited extract boundary is expanded to include auditable trigger fields and decomposition outputs.

3.8.4 Key limitations and risks implied by the audited findings

Several limitations and risks follow directly from the evidence presented in Tables 3.1–3.3 and from the audited export boundary.

First, there is interpretive risk from concentration: in areas with high skew, a small number of parcels can dominate totals (Table 3.3). This creates a risk that totals or rank ordering by total value appear stable even when within-area distributions shift materially. If the programme intent is transparency and public intelligibility, totals must therefore be paired with distributional and concentration metrics as standard reporting outputs.

Second, the evidence shows governance sensitivity. Table 3.2 demonstrates that stepped schedules can generate silent steps and reversals in particular areas. This implies that intelligibility and defensibility depend on transparent publication of rule logic and parameters. Without explicit rule publication, readers may interpret stepped outcomes as arbitrary even when they are operating as designed.

Third, the audit boundary imposes a clear interpretability constraint. The absence of parcel-level trigger columns in the audited synthesis extracts prevents defensible driver decomposition. This is not a weakness of the concept, but it is a constraint on what can be claimed from this audited package. Where public-facing explanations require “why did this parcel change?”, the audited package must include trigger fields and auditable decomposition artefacts.

Fourth, there is presentation risk from map-capping. Where mapping outputs are capped for visualisation, map products will not show the full extent of differentiation. Interpretability must therefore rely on analytical values and implied pre-cap distributions (as reported in the tables) rather than on cap-bound visual inspection alone.

The evidence supports the view that the ladder can be updated and re-run on a repeatable basis, provided that operational prerequisites are treated as hard dependencies. In practice, updateability is shaped by three constraints evidenced in the build and reflected in the audit discipline adopted for this report: a maintained, high-coverage parcel fabric suitable for repeatable feature attachment; stable, versioned input datasets (and explicit inventory/coverage gates where tiled/raster sources are used); and parameter and run governance so that changes in schedules or inputs can be traced and explained.

Taken together, the audited evidence indicates that Lot 3 is most interpretable as an area-level value density and distribution signal, supported by standardised diagnostics, rather than as a parcel-level causal attribution model.

Lot 5 Participatory Approaches: Findings

This section reports the audited Lot 5 evidence from the participatory valuation workshops. It presents what participants did, what outputs were captured, and what can be inferred about intelligibility, perceived fairness, confidence, uncertainty, and trust under workshop conditions. The outputs are indicative values produced through facilitated exercises for a bounded parcel set; they are not market valuations and must not be interpreted as such.

3.9 Evidence boundary, dataset integrity, and what can now be claimed confidently

All findings in this section refer to the audited Lot 5 package.

Interpretation in this section requires two discipline points that mirror the Lot 3 approach: clear denominators and explicit evidence boundaries. First, all quantitative results are framed against the audited denominators (workshops, participants, parcels, and response rows). Second, the evidence supports describing how participants reasoned and responded under controlled, facilitated conditions, but it does not support treating the outputs as market-calibrated values, or as transferable valuations for other sites. A further structural boundary applies to method comparisons. “Method usage” in this dataset is a design footprint (fixed exercise exposure) rather than preference or adoption, and differences between method-labelled value summaries must not be interpreted as causal “method effects”. Method labels are structurally linked to the parcel tasks participants were assigned under the workshop design, so method-level differences may reflect parcel composition under each method rather than independent uplift/discount effects.

3.10 Headline findings

Lot 5 evidence shows that participants do not treat “clarity” and “fairness” as the same concept when evaluating valuation approaches. Across all participants, the method most often selected as clearest differs from the method most often selected as fairest: comparables (Comp) most frequently anchors clarity, while the point-factor index (PFI) most frequently anchors fairness. This clarity–fairness split is a policy-relevant signal, because it indicates that a method can be perceived as procedurally fair without necessarily being comfortable or easy for users to apply. It also indicates that explainability is treated as part of legitimacy, rather than as a post-hoc communication layer.

This distinction is reinforced by the numeric outputs, which show that meaningful dispersion persists even within a bounded parcel set under facilitation. In this context, value spread provides a more defensible route to identifying "hard-to-value" parcels than relying on headline highs and lows alone. Confidence provides a complementary lens: overall confidence is moderate, but the method most often perceived as fairest (PFI) also has the lowest mean confidence. The qualitative evidence helps explain this pattern, with trust repeatedly functioning as a gate. Where assumptions feel hidden or mechanisms feel "black box", trust becomes conditional. In particular, below-ground unknowns dominate uncertainty, implying that participatory tools must make uncertainty explicit rather than forcing false precision.

3.11 Delivery, participation and completion

This sub-section provides the denominators required to interpret every subsequent statistic.

Table 3.4. Workshop composition and participant bases

Workshop type	Workshops	Participants	Share of participants
Public	8	56	88.9%
Professional	2	7	11.1%
Total	**10**	**63**	**100.0%**

Overall, 61 of 63 participants are complete (96.8%), and 2 of 63 are partial (3.2%). The partial completion is not treated as a nuisance artefact. It is treated as evidence relevant to toolkit accessibility and comprehension under real workshop conditions.

The logged reason for partial completion is a language or terminology barrier, treated here as an accessibility and comprehension signal that informs toolkit design (including terminology choices, pacing, and how multi-step reasoning is supported in-session). For quantitative summary statistics, only completed entries contributing to that statistic are included. Partial entries are carried into the qualitative synthesis (facilitator notes and discussion evidence), and the specific workshop is explicitly flagged where it is relevant to interpretation.

3.12 Method deployment footprint (what "method usage" does and does not mean)

Three methods are analysed in the audited dataset and are named consistently throughout this section: Comparable (Comp), Rule-of-Thumb Residual (RotR), and Point-Factor Index (PFI).

Table 3.5. Method usage footprint (all responses, n = 630)

Method	Responses	Share
RotR	252	40%
Comp	189	30%
PFI	189	30%
Total	**630**	**100%**

The identical 40/30/30 pattern across public and professional workshops indicates fixed exposure. These proportions should therefore not be interpreted as adoption or preference rates. Method usage confirms that all participants encountered all methods across a fixed number of tasks. This strengthens the interpretability of ranking evidence because rankings were formed after exposure, not based on unfamiliar labels or unequal experience.

3.13 Method rankings: clarity versus fairness

This sub-section reports the most policy-relevant behavioural finding from Lot 5: participants distinguish between what feels clear and what feels fair, and they do not consistently choose the same method for both.

Across all participants (n = 63), the top-ranked method differs for clarity versus fairness.

Table 3.6. Overall top-choice distributions (n = 63)

Criterion	Comp	RotR	PFI
Clearest	25 (39.68%)	20 (31.75%)	18 (28.57%)
Fairest	12 (19.05%)	23 (36.51%)	28 (44.44%)

The ranking split reflects a meaningful conceptual distinction in how participants evaluated the methods:

1. **PFI as fairness**: fairness was associated with systematic treatment of factors and perceived consistency. Participants tended to treat structured factor application as reducing arbitrariness, even where the method felt demanding to use.

2. **Comp as clarity**: clarity was associated with being anchored in something that resembles market evidence, and with being easier to explain to others. Participants commonly framed comparables as intuitively legible: “this looks like how value is justified”.

3. **RotR as clarity (conditional)**: where RotR was experienced as clear, it was because it offered visible step logic. However, RotR’s clarity advantage is conditional on participants understanding the language and the implied assumptions. When terminology or concepts were barriers, RotR could become less accessible than methods that felt anchored (Comp) or factorised (PFI).

3.14 Indicative value outputs

The value outputs reported below are indicative workshop-generated values for a bounded parcel set. They are not market valuations. Differences observed by method label must not be read as independent method uplifts or discounts because method labels are structurally linked to the exercise design and the parcel tasks participants were assigned.

Table 3.7. Overall distribution summary (n = 630)

Statistic	Value
Minimum	£49,100
Maximum	£151,700
Mean	£96,196
Standard deviation	~£24,038

Table 3.8. Summary by method label (bounded; not causal)

Method	n	Mean	Median	Interquartile range	Range
Comp	189	£101,341	£101,500	£38,100	£62,500–£144,500
PFI	189	£94,754	£91,700	£38,200	£55,500–£147,000
RotR	252	£93,419	£92,600	£44,775	£49,100–£151,700

Because method exposure is structurally linked to parcel tasks (the fixed 40/30/30 deployment footprint), differences in these method-level summaries may reflect parcel composition under each method rather than an independent “method effect”. These tables are therefore reported as descriptive evidence about outputs under the implemented design, not as causal estimates of method uplift or discount.

3.14.1 Interpreting the value range

The overall range (£49,100 to £151,700) and standard deviation (~£24,038) indicate meaningful dispersion even within a controlled parcel set. This supports a core methodological observation: uncertainty and heuristic anchoring are present under facilitation and must be managed through tool design rather than assumed away.

3.14.2 Parcel-level differences: what is “easy” versus “hard” to value

This sub-section treats dispersion as the most defensible quantitative route to identifying “hard-to-value” parcels.

Across the parcel set, median values provide a stable anchor for describing central tendency. In the audited outputs:

1. **Lowest median**: Parcel_01 median £59,500; interquartile range £7,700; range £49,100–£71,100.
2. **Highest median**: Parcel_09 median £127,800; interquartile range £13,950; range £103,200–£144,500.
3. **High median**: Parcel_10 median £124,900; interquartile range £15,300; range £104,300–£151,700.

These values are reported as descriptive outputs for the bounded parcel set; they should not be interpreted as transferable values for other sites.

Hard-to-value parcels are operationalised as those with higher within-parcel spread (interquartile range) and/or larger ranges, not simply those with high values. In the audited outputs, the largest interquartile ranges occur for Parcel_08, Parcel_10, and Parcel_09, indicating that these parcels elicited greater disagreement under the bounded information conditions. The interpretive implication is that disagreement concentrates where participants face competing cues, higher uncertainty, or uncertainty about hidden constraints.

3.14.3 Workshop-level agreement and the Value_Spread classification

Workshop–parcel dispersion is summarised using the recomputed workshop–parcel aggregate dataset (100 rows). Value_Spread is derived from each workshop–parcel interquartile range and classified using dataset quartile thresholds:

1. Lower quartile (q1): £5,868.75
2. Upper quartile (q3): £12,775.00
3. Low spread: interquartile range ≤ q1
4. Moderate spread: q1 < interquartile range < q3
5. High spread: interquartile range ≥ q3

This classification is derived and reproducible; it is not a subjective label. In the corrected dataset, higher value spread does not mechanically imply lower confidence: the workshop–parcel correlation is near zero. In this context, this report does not claim that participants were less confident when they disagreed, as disagreement can reflect genuine judgement differences under uncertainty, rather than simple confusion.

3.14.4 Confidence: what it supports and what it complicates

Confidence provides a second lens on usability and legitimacy: a method can look procedurally attractive yet still leave users uncertain about their ability to apply it. Across all responses, mean confidence is approximately 6.06 on a 1–10 scale, with a standard deviation of approximately 1.28. This indicates moderate confidence overall, with meaningful variation.

Table 3.9. Confidence by method label

Method	Mean confidence	Median confidence
Comp	6.65	7
RotR	6.14	6
PFI	5.37	5

A key interpretive finding follows directly: **the method most often perceived as fairest (PFI) also has the lowest average confidence**. This is consistent with a separation between perceived procedural fairness and user comfort or competence. In practical terms, it implies that a method can satisfy fairness intuitions while still requiring design work to be usable at scale (language, guidance, worked examples, and explicit handling of uncertainty).

3.15 Qualitative themes: mechanisms that explain the numbers

This section provides the explanatory layer that makes the numeric outputs usable for programme learning. The analysis is intentionally insight-forward: values are interpreted alongside participants' explanations to identify why divergence occurs, whether due to comprehension barriers, heuristics, information limits, or genuine judgement differences. In that context, themes that recur across all workshops are treated as robust cross-context mechanisms rather than workshop-specific anomalies. The consistent presence of trust and comprehension drivers, alongside uncertainty drivers, indicates that these are structural determinants of whether a participatory valuation toolkit is intelligible and perceived as legitimate.

Across workshops, trust is repeatedly linked to whether participants can see how a number is produced and whether the underlying assumptions are visible and contestable. "Black box" concerns operate as a legitimacy gate: where the process feels opaque, trust becomes conditional. The programme implication is direct. Explainability cannot be treated as an optional communications layer applied after the calculation; in participatory contexts, it functions as part of the valuation method itself. Any future public-facing toolkit therefore needs to make assumptions explicit and challengeable, rather than merely outputting a value.

Uncertainty is the other persistent driver of dispersion, and it is dominated by below-ground unknowns and hidden constraints that participants view as capable of materially changing land value. Commonly cited uncertainties include ground conditions, abnormal costs and remediation, utilities, and constraints not visible in surface-level evidence packs. This explains why simplified and controlled information can still yield spread: participants are often not disputing what is presented; they are making different judgements about what is not presented. A practical learning point follows: participatory tools should include an explicit uncertainty capture mechanism that allows users to record information gaps and uncertainty, rather than forcing false precision.

Where confidence is higher, it is often associated with familiarity with local context or the presence of stronger anchors that reduce perceived arbitrariness. This requires careful handling. Confidence may reflect genuine clarity, but it can also reflect anchoring effects. Either way, reliance on local knowledge is not scalable, and a scalable approach should not assume that such knowledge will be present. Instead, it should substitute structured disclosure and worked examples that provide equivalent reassurance across settings.

Taken together, the qualitative evidence indicates that assumptions are the hinge between perceived fairness, perceived clarity, and confidence. RotR is often valued for transparent step logic, yet disagreement concentrates around whether the assumptions embedded in each step are reasonable. PFI is commonly experienced as systematic and therefore often fair, but can feel inaccessible, contributing to lower confidence. Comp is frequently experienced as anchored and explainable, but it prompts scrutiny over whether the selected comparables are genuinely comparable and whether local context has been adequately captured. The design implication is that a future toolkit should distinguish clearly between assumptions users are permitted to vary, assumptions fixed by guidance, and assumptions that require professional input.

3.15.1 Workshop patterns as diagnostic learning

Workshop-level dispersion is treated here as diagnostic learning rather than a basis for inference. Where workshops show high divergence, that divergence should be read as a cue to investigate what uncertainties were salient, which information gaps were most prominent, and whether the facilitation conditions influenced convergence. On that basis, workshop 2 stands out as having the highest high-spread share (60%), indicating a context in which parcel characteristics and/or discussion conditions produced greater dispersion. By contrast, workshop 7 shows 0% high spread, indicating tighter convergence. This should not be interpreted straightforwardly as “better” performance without triangulation, because convergence can reflect shared clarity, but it can also reflect shared anchoring or homogeneous assumptions. Workshop 3 provides a concrete design case study on accessibility. Its partial completion and distinct clarity pattern indicate that, where terminology comprehension is a barrier, methods that rely on multi-step residual logic (such as RotR) may become less accessible. In such conditions, participants may gravitate towards approaches that feel more anchored (Comp) or more clearly factorised (PFI). The programme learning is that accessibility and language should be treated as core methodological design variables rather than presentation details, because they materially shape which methods participants can use with confidence, and therefore the pattern of agreement and divergence observed in the outputs.

3.16 What Lot 5 findings contribute to the programme

Lot 5 adds evidence that is not accessible through technical modelling alone: it shows how people interpret valuation logic, what they treat as clear, what they treat as fair, and what drives confidence and trust under controlled information conditions. The clearest and fairest methods differ: Comp is most often experienced as clearest, while PFI is most often experienced as fairest (with professional comparisons treated as indicative due to small base). Numeric outputs show meaningful dispersion even within a bounded parcel set; dispersion is a more defensible indicator of “hard-to-value” parcels than headline highs and lows. Confidence and fairness are not equivalent: the method most often perceived as fair (PFI) has the lowest average confidence, indicating accessibility constraints that would need to be addressed for any scalable deployment. Trust is consistently gated by explainability; black-box concern is pervasive across workshops and functions as a legitimacy condition. Below-ground uncertainty dominates reasoning, indicating that any future participatory approach must make uncertainty explicit rather than hidden.

4 Comparison of approaches

This section explains how different methodologies produce different outputs, and why, and sets out comparative advantages and disadvantages in a structured, evidence-led manner. The purpose is not to identify a single “best” approach; rather, it is to describe which approach is best suited to which purpose, and the governance conditions required for each to be used credibly.

Within that framing, the scope of the comparison is deliberately not an “accuracy league table”. Lot 3 and Lot 5 operationalise land value differently and produce different types of evidence. Lot 3 generates parcel-level value density signals (£/m²) over an analytical parcel fabric within a defined audit boundary. Lot 5 generates facilitated human judgement evidence (rankings, confidence, qualitative rationales, and indicative numeric outputs) over a bounded parcel set within a separate fieldwork-and-analysis audit boundary. The appropriate basis for comparison is therefore feasibility, interpretability and usability, auditability and governance, scalability and coverage potential, and public intelligibility and legitimacy. Across the programme, a practical lesson is that public intelligibility and analytical defensibility both depend on the discipline of boundaries and guardrails. For Lot 3, those guardrails are technical (coverage union area, classification completeness, row-parity across formula outputs). For Lot 5, they are procedural (structured facilitation, controlled presentation of information, and clear limits on what workshop outputs can and cannot claim). In both cases, the strongest evidence is produced when the method is treated as a controlled system with explicit stop conditions, rather than as an “analysis exercise” where issues are handled informally midstream.

Lot 3 Formula-based Valuation: In comparison

Lot 3 produces parcel-level land value density signals (£/m²) applied to an analytical parcel fabric, staged through a stepped methodology (“complexity ladder”) that increases data dependency and within-area differentiation as additional variables are introduced. For the audited run, the primary reported artefacts are distributional summaries and maps of £/m² outputs at parcel level, alongside rung-to-rung behaviour that is mechanically attributable to the introduction of new input features and modifiers.

What Lot 3 can credibly claim now is primarily technical and governance-led: computational integrity within the audited run boundary; reproducibility of outputs given the same parcel fabric, parameters, and inputs; and an auditable rung-by-rung trace in which behaviour changes as additional variables are introduced. It can also claim interpretability in the specific sense that later rungs are engineered and parameterised, meaning that, in principle, the direction of effects is explainable even where the audited synthesis extracts do not carry all parcel-level trigger fields needed for full driver decomposition.

What Lot 3 cannot claim, by itself, is evidence of public intelligibility, perceived legitimacy, or acceptability of the outputs. It also cannot claim coverage of “all land” beyond the valuation-eligible analytical base defined by parcelisation and exclusions: White Space and other structural exclusions are outside the valued fabric by design, and audited extracts do not, in themselves, constitute a whole-landscape eligibility reconciliation.

4.1.1 Why Lot 3 outputs change across rungs (mechanistic explanation)

Differences across rungs are expected and are driven by model mechanics rather than by interpretive narrative. The main mechanisms are:

1. Newly introduced variables (for example, area-level modifiers, access terms, constraints) that were not present in earlier rungs.
2. Operationalisation choices (flags versus continuous terms; stepped schedules versus gradients; one-sided penalties versus two-sided adjustments).
3. Parcel fabric granularity and concentration effects, where very large parcels can dominate totals and can influence how map patterns are perceived, even when £/m² values are bounded.

Mechanism examples, expressed generically:

1. Baseline rung: uniformity within the scope base by design; map outputs are displayed with a £150/m² cap for visualisation, but the underlying implied pre-cap rates under later rungs do show the intended between-area and within-area differentiation.
2. Stepped rung: between-area differentiation by schedule; limited within-area differentiation by design.
3. Access/constraint rung: within-area differentiation emerges through continuous penalty terms; effects are uneven where feature readiness, spatial structure, and context differ.

4.1.2 Strengths and limitations

Strengths:

1. High auditability and re-runnability: parameter set control and reproducible execution support systematic re-testing, refinement, and consistent reporting cycles.
2. Scalability potential: the approach is structurally capable of producing wide coverage where input readiness is sufficient and governance conditions are met.

3. Mechanistic transparency: staged complexity makes data dependency explicit and supports disciplined interpretation.

Limitations:

1. Sensitivity to parcel fabric definition and to parameter governance; totals and area-weighted narratives can be distorted by concentration in very large parcels.

2. Interpretive risk for non-specialists: without explicit explanation of analytic parcels and White Space, technically defensible outputs can be misread as missing data, omitted land, or definitive “site values”.

Lot 5 Participatory Approaches: In comparison

Lot 5 produces evidence about intelligibility, usability, perceived fairness, confidence, trust, and reasoning mechanisms under facilitated workshop conditions. The output artefacts are structured rankings of methods (including the distinction between what is judged clearest and what is judged fairest), confidence ratings, indicative numeric values produced through structured tasks for a bounded parcel set, and qualitative evidence (transcripts, written rationales, facilitator notes) that explains why participants judged and behaved as they did. What Lot 5 can credibly claim now is legitimacy-and-design evidence: how people reason about land value when given standardised information; where misunderstandings and uncertainty concentrate; and what design choices appear necessary for a publicly intelligible toolkit. It can also claim evidence about usability barriers, including non-completion as a signal rather than an error to be hidden.

What Lot 5 cannot claim is representativeness of the numeric values as market-grade valuations, or population inference about public preferences. Workshop outputs are bounded by small group sizes, purposive recruitment, facilitation effects, and a deliberately limited parcel set. Quantitative reporting therefore requires strict denominators and missingness discipline, and interpretive claims must remain diagnostic.

4.1.3 Why participatory outputs diverge

Where participatory outputs diverge, the divergence is driven by behavioural mechanisms rather than by engineered model mechanics. In the audited evidence, the dominant mechanisms include:

1. Anchoring and adjustment: participants often anchor on place identity, perceived desirability, and local narratives, then adjust (to varying degrees) based on parcel attributes and constraints.

2. Salience effects: risks and constraints are used more consistently when they are made salient in materials; they are less reliably introduced unprompted.

3. Refusal of false precision: ranges, rounding, and non-completion are meaningful signals of cognitive load and uncertainty, especially where below-ground unknowns dominate.

4. Perceived legitimacy: trust is conditional on explainability; “black box” concern is a recurring gate on acceptance.

4.1.4 Strengths and risks

Strengths:

1. Produces evidence that computational approaches cannot generate: how outputs are understood, contested, trusted, and judged as fair or unfair.

2. Identifies translation risks early: where and why non-specialists will misinterpret, resist, or over-trust outputs.

3. Supports design refinement: workshop evidence is actionable for improving method cards, terminology, assumption visibility, and uncertainty capture.

Limitations:

1. Limited scalability as a coverage instrument: the approach is not designed to produce national parcel coverage.

2. Susceptible to anchoring, salience bias, and facilitation effects unless mitigated in toolkit design.

3. Numeric outputs can be misread as valuations if publication guardrails are not applied rigorously.

4.2 Why the two approaches produce different “results”

4.2.1 The fundamental reason for divergence: different constructs of “value”

Lot 3 encodes value as an auditable per-area signal applied to an analytical parcel fabric, with differences driven by staged data dependency and formula mechanics. Lot 5 encodes value as a human judgement under structured facilitation, in which anchoring, salience, uncertainty, and perceived legitimacy materially shape outputs. Different “results” are therefore expected because the two lots are measuring and testing different constructs: one prioritises repeatable computational signals; the other prioritises intelligibility and behavioural validity under public-facing conditions.

4.2.2 The translation gap: defensible versus accepted

A named cross-lot mechanism evidenced in the programme is the clarity–fairness split. Participatory evidence indicates that what participants experience as clearest is not necessarily what they experience as fairest. This matters because a technically defensible output can still fail in practice if it is not accepted as legitimate or if it cannot be explained in a way that aligns with people's fairness intuitions. This is not an argument against technical modelling; it is an evidence-led governance requirement. Lot 3 optimises technical consistency and auditability; Lot 5 evidences which aspects of explanation, assumption visibility, and procedural framing are necessary if technical outputs are to be credible to non-specialists.

4.2.3 Divergence

Place identity and reputation

Participatory evidence demonstrates strong place anchoring. A formula approach will only reflect place narratives to the extent that they are proxied by engineered variables. Divergence risk is therefore highest where place identity is a dominant cue and where proxies are limited or contested.

Risk and constraints

A formula approach can systematise penalties consistently (where constraints are measured and operationalised), while workshop evidence shows that constraints are less likely to be introduced unprompted unless made salient. Divergence risk is highest where hidden constraints and unknowns (especially below-ground) materially affect judgement. Additionally, Lot 3's analytic parcel fabric and definitional White Space exclusions are essential boundary-setting devices for computational feasibility, but they create interpretive risk if not communicated. Lot 5's use of recognisable parcels increases realism but can intensify anchoring. Divergence is most likely where unit boundaries are not understood by readers or where participants' mental model of "a parcel" differs from the analytical construct.

4.3 Comparative benefits and disadvantages table

Table 4.1 compares Lot 3 and Lot 5 across purpose, outputs, auditability, scalability, strengths, limitations, and best-use cases, and that it includes guardrails preventing misinterpretation.

Table 4.1: Comparative benefits and disadvantages of Lot 3 and Lot 5 approaches

Dimension	Lot 3: Formula-based valuation by land area	Lot 5: Participatory approaches
Primary purpose	Test feasibility of producing consistent, auditable parcel-level value density signals using engineered drivers and staged complexity.	Test public intelligibility, usability, perceived fairness, confidence and trust under controlled facilitation; identify reasoning mechanisms and design risks.
Output type and unit	Parcel-level £/m² signals on an analytical parcel fabric; maps and distributional summaries by rung.	Rankings (clarity andfairness), confidence scores, qualitative rationales; indicative numeric outputs for a bounded parcel set under facilitation.
What it can credibly claim now	High technical defensibility within audited boundary: reproducible, parameterised, rung-by-rung behaviour attributable to introduced mechanics.	Strong legitimacy and intelligibility evidence: how methods are understood, trusted, contested; where uncertainty and comprehension barriers concentrate.
What it cannot credibly claim now	Does not evidence acceptance or legitimacy; does not, by itself, reconcile whole-land coverage beyond the valuation-eligible analytic base.	Does not produce market valuations; numeric outputs are indicative and non-representative; small n constrains inference.
Auditability and QA traceability	Strong: rerunnable pipeline; parameter-set control; Quality Assurance artefacts; mechanical trace from rung changes.	Strong within fieldwork boundary: locked corpus; reconciled counts; recording→transcription→QC→anonymisation→analysis trail; deviations treated as evidence.
Coverage and scalability potential	Potentially scalable where input readiness and governance are sufficient; designed for repeat reporting.	Not designed for coverage scaling; designed for diagnostic learning and engagement/communication testing.
Data and resource dependency	Dependent on spatial fabric integrity, feature readiness, and parameter governance; sensitive to unit definition and concentration effects.	Dependent on facilitation quality, materials, and participant comprehension; sensitive to anchoring, salience, and information limits.
Strengths	Reproducible; auditable; supports systematic re-runs; staged complexity makes readiness explicit; suited to consistent surfaces and monitoring.	Produces evidence on acceptance and comprehension; identifies translation risks; captures uncertainty and reasoning mechanisms; informs explainability design.
Weaknesses and risks	Outputs can be misread without explanation of analytic parcels and White Space; totals can be dominated by large parcels; requires disciplined parameter governance.	Susceptible to anchoring and bias unless mitigated; small-n workshop patterns are indicative; numeric outputs easily misinterpreted without guardrails.
Best use-case in the programme	Technical feasibility and governance testing for parcel-level surfaces; identify readiness gaps and parameter sensitivities for future refinement.	Stress-test intelligibility and legitimacy; refine method explanations; identify where assumptions and uncertainty need explicit handling.

Public intelligibility and legitimacy	Requires an explainability layer; technical defensibility does not guarantee acceptance.	Direct evidence on legitimacy and trust; clarity–fairness split demonstrates that acceptance is multi-dimensional.
Likely divergence from the other approach	Where reputational and place cues dominate but are not proxied; where unit boundary and exclusions are misunderstood; where concentration distorts totals.	Where anchoring overrides attributes; where risks are not salient; where bounded information forces judgement under uncertainty.
Guardrail cell (anti-misread)	Outputs are bounded to the analytic parcel fabric and explicit exclusions; do not infer “all land” coverage without eligibility reconciliation.	Workshop values are indicative; interpret alongside denominators, confidence and qualitative justification; do not frame as market valuations.

4.4 Joint implications for use in practice

If the objective is repeatable coverage and consistent reporting surfaces, a Lot 3-type approach is necessary. However, credible use requires strong governance: stable parcel fabric and exclusions; explicit explanation of White Space; parameter versioning; and publication discipline that emphasises distributional statistics over totals where concentration risk is material. Without this, technically defensible outputs are vulnerable to misinterpretation by non-specialists.

If the objective is deliberation, scrutiny, trust-building and intelligibility testing, Lot 5 provides evidence that is necessary to design a credible public-facing approach. It cannot substitute for a scalable technical method where national coverage is required, but it materially reduces “translation risk” by showing how people interpret value logic, what they regard as fair, and what undermines confidence.

Minimum viability conditions for “good” use:

1. **For a Lot 3-type method:** a stable unit of analysis (a consistent parcel fabric); explicit, published exclusions (including a plain-English explanation of White Space); controlled and versioned parameters (so changes are traceable); readiness tiering (so the method does not overreach where inputs are incomplete); and published outputs that make clear what the figures represent, what they exclude, and what questions they cannot answer without additional published driver fields.

    - The research demonstrates that this is feasible for bounded areas using open and official datasets, provided the parcel fabric and input layers meet minimum completeness standards. Moving to Wales-wide operation is principally constrained by data and publication discipline (coverage reconciliation, consistent fabric governance, and completeness of any tiled and gridded inputs used for fallback classification), rather than by the existence of formulae in principle.

2. **For a Lot 5-type method**: clear denominators and missingness rules; explicit anchoring management; design features that make risks and constraints salient; allowance for uncertainty capture (confidence and ranges); and a traceable evidence trail including deviations.

Where each approach is inappropriate or high-risk:

1. Lot 3 is high-risk if presented as a single public-facing “truth” without explanation of analytic parcels, exclusions, and why adjacent parcels can differ materially under engineered penalties.
2. Lot 5 is high-risk if presented as a replacement for professional valuation or as producing definitive parcel values; it is also high-risk if place cues and anchoring are not explicitly managed in toolkit design.

Bridge actions that follow from the comparison (cross-lot design requirements):

1. Pair technical outputs with an explainability layer informed by participatory evidence, explicitly addressing both clarity and fairness narratives.
2. Make constraints and risks salient and comparable in any public-facing engagement, because they are not reliably introduced spontaneously.
3. Use participatory testing to refine how technical maps and tables are explained, especially where within-area differentiation is pronounced.
4. Where feasible, increase overlap between participatory parcel examples and parcels in the technical fabric to strengthen cross-reading, without implying the two approaches measure the same construct.

4.5 Comparative or complementary?

Taken together, Lot 3 and Lot 5 show that different valuation methodologies will produce different outputs because they encode different constructs of value: Lot 3 produces auditable per-area computational signals over an analytical parcel fabric, while Lot 5 produces structured judgement evidence shaped by intelligibility, salience, uncertainty and perceived legitimacy under facilitation. The comparative issue is therefore not which method is “right”, but what each method can credibly support and what governance conditions are required for its outputs to be interpreted responsibly. Lot 3 evidences feasibility, repeatability and parameter-governed transparency; Lot 5 evidences the legitimacy conditions under which value logic is accepted or contested, including the observed separation between clarity and fairness. In practical terms, credible future application requires alignment: technical repeatability must be coupled with public-facing explainability, explicit handling of uncertainty, and disciplined communication about unit boundaries and exclusions across Wales and beyond.

5 Conclusions

5.1 Summary of key findings, including lessons learnt through the process

Across both lots, the programme evidences that “land value” is not a single observable quantity that different methods simply estimate with more or less precision. It is constructed through definitional boundaries (what is included and excluded, and the unit at which value is expressed), the drivers each method is able, or chooses, to operationalise, and the way outputs are interpreted by non-specialists and professionals. This is a substantive finding in its own right, because it implies that a technically auditable output can still fail as policy evidence if it is not intelligible, perceived as legitimate, and accompanied by clear guardrails that prevent misinterpretation.

Lot 3 (formula-based valuation by land area) demonstrates high computational integrity within the audited analytical parcel fabric. Within that bounded base, the staged methodology behaves in a coherent, interpretable sequence: a baseline schedule provides a controlled reference point; a stepped adjustment introduces between-area differentiation; and an access/constraint stage introduces within-area differentiation through penalisation logic. The key lesson is therefore not a claim about “true values”, but an evidence-led account of what each rung adds, what it cannot add at this stage of maturity, and which governance requirements become binding as complexity increases (particularly parameter discipline, unit definition, and safeguards against concentration effects driven by parcel area).

Lot 5 (participatory approaches) shows that participants can engage meaningfully with structured valuation tasks, but that their reasoning is predictably shaped by behavioural mechanisms such as anchoring, place-based reasoning, salience effects, and the conversion of uncertainty into discounting. A central programme-level lesson is that missingness, refusal of precision, and requests for additional context are not administrative noise; they are diagnostic signals about what is hard to value and what will be hard to communicate publicly. The audited corpus is bounded to ten workshops (Workshops 1 to 10), 63 participants and a fixed set of ten parcels per workshop Within those bounds, the evidence supports robust learning about intelligibility and legitimacy conditions, including the observed separation between procedural clarity and perceived fairness.

Taken together, the two lots generate complementary learning. Lot 3 is strongest on repeatability, parameter governance, and the feasibility pathway to scale. Lot 5 is strongest on intelligibility, legitimacy, and the translation risks that will determine how any technical outputs are received and used. The programme’s combined value lies in treating these as distinct but interdependent requirements rather than as alternatives.

5.2 Cross-programme lessons and insights

Deliverability requires governed systems, not just workable methods

Across both lots, delivery evidence shows that the methodological concept can be technically sound yet still fail under scrutiny if governance is weak. Lot 3 is reproducible only when run as a logged, deterministic pipeline with disciplined geometry handling and an explicit audit boundary. Lot 5 is defensible only when treated as a fieldwork-and-analysis system with fixed denominators, completion rules, and a locked analysis window. The practical implication is that any future roll-out must specify “system requirements” as part of the method: audit boundary and cut-off discipline, versioning, Quality Assurance gates, naming conventions, and publication rules.

Interpretability is a first-order outcome; it must be designed and evidenced

The programme’s core question is not whether numbers can be produced, but whether they remain intelligible under public and professional scrutiny. Lot 3 shows that cap-bound maps can suppress visible differentiation even when underlying distributions and implied rates vary materially. Lot 5 shows that participants separate clarity from fairness and gate trust through explainability. The implementation implication is that a minimum publication pack should standardise interpretability artefacts—distributional reporting, concentration diagnostics, publishable rule logic and assumptions, and uncertainty capture—rather than relying on headline outputs alone.

Governance-sensitive logic is where reputational risk concentrates

Where outcomes appear non-intuitive, the failure mode is rarely a technical error; it is an absence of publishable explanation. Lot 3 demonstrates that small rule differences can generate silent steps, reversals, and variable compression magnitudes. Lot 5 shows that hidden assumptions are experienced as black-box behaviour and reduce trust. The implementation implication is that publishable rule logic and assumption visibility are not optional mitigations; they are the core controls that prevent legitimacy risk.

5.2.1 Lot 3 lessons learned

Operational deliverability and reproducibility are feasible, but only as an auditable pipeline

Lot 3 is implementable across the nine pilot LSOAs using open datasets and open-source tooling, provided it is run the same way each time and is fully traceable (for example, through a unique run ID, recorded parameter settings, and consistent staging of the ladder). Without that discipline, the same method becomes hard to defend because outputs cannot be reproduced cleanly against an agreed, audited input set. In practice, this means building in non-negotiable Quality Assurance “stop conditions” as the minimum standard for publication, including minimum coverage thresholds, geometry validity rules (ensuring polygons are valid), CRS discipline (using a consistent coordinate system for area and distance calculations), and join completeness checks (confirming each parcel has the required attributes attached).

Data readiness is a first-order dependency; the method does not fail gracefully when coverage breaks

Delivery evidence shows that partial coverage is not simply lower confidence; it can create structurally incomplete attribution, such as residual NULL classes or missing feature attachments. That is a publishability failure because it introduces systematic, non-random gaps that cannot be defended as “uncertain but usable”. In practice, this means introducing a simple “stop/go” quality check before any valuation is run or published. If a required dataset does not fully cover the area being valued (for example, if any map tiles are missing from a gridded land-cover dataset used as a fallback), the process should stop and report the gap, rather than producing partial figures that look complete but are not defensible.

Parcel fabric selection is decisive for coverage, comparability, and interpretability

This research demonstrates that “any polygon layer” is not an adequate proxy parcel fabric. Low-coverage or heterogeneous sources fragment the analysis surface, degrade aggregation comparability, and introduce artefacts that then require additional normalisation to achieve stable joins and totals. In practice, this means treating parcel fabric choice as a programme decision with clear acceptance criteria—coverage, contiguity, overlap rules, exclusion definition, and aggregation suitability—rather than as an implementation detail.

Geometry governance is a defensibility requirement, not an engineering preference

eometry processing choices directly affect what can be claimed, because they change area stability, join behaviour, and aggregation results. Silent coercion (undocumented “fixes” applied during processing) undermines auditability by obscuring where and why outputs differ between runs. In practice, this means publishing explicit geometry rules—validity criteria, overlap resolution, dissolve rules, clipping and boundary handling—and logging exceptions rather than allowing undocumented fixes.

Audit boundary discipline determines the claim boundary

The audited package supports reporting ladder behaviour, rung-to-rung change metrics, and distribution and concentration diagnostics for the exported parcel set. It does not support parcel-level causal explanations because the audited synthesis extracts do not include the "trigger" fields required to explain precisely what drove each parcel's movement. In practice, this means keeping the interpretive boundary explicit in any narrative that risks drifting into "drivers", and adding auditable trigger and decomposition artefacts if future use-cases require parcel-level explanations.

Separate analytical values from display rates to prevent presentation-driven misinterpretation

Capped visual outputs can create a false impression of uniformity, particularly where display limits suppress visible differentiation. Interpretive credibility improves when the presentation layer is explicitly separated and analytical reporting remains anchored to value_gbp totals and implied pre-cap distributions. The implementation implication is to standardise this separation in both pipeline outputs and publication narrative, including a clear rule wherever maps are shown.

Formula 1 provides governance control, not just a starting point

The baseline's value is auditability. It provides a controlled reference state that allows later rung effects to be measured cleanly and explained transparently. In practice, this means retaining Formula 1 in any operational stack, even if policy focus is on later rungs, because it anchors explanation and dispute resolution.

Formula 2 is intelligible, but legitimacy depends on publication-grade rule governance

Stepped schedules can be easy to communicate and function like visible policy levers. However, the evidence shows that they can generate silent steps and reversals depending on thresholds and multipliers. The implementation implication is to treat the schedule and activation logic as a publication-grade specification, with transparent parameters and boundary conditions, rather than as internal code logic.

Formula 3 enables within-area differentiation, but interpretation must be diagnostic-led

Within-area differentiation is feasible, but visual inspection of cap-bound maps is not a defensible interpretation route. Distributional evidence, change metrics, and concentration diagnostics are required because totals can remain correlated while within-area distributions shift materially. The implementation implication is to standardise percentile ranges and concentration metrics alongside totals in any policy-facing narrative.

Distributional skew and concentration are structural risks that persist across rungs

Highly unequal parcel areas create heavy-tailed value distributions, meaning totals and rank ordering can be dominated by a small number of parcels even when schedules behave as designed. This is an interpretive and communication risk, not inherently an error. The implementation implication is routine publication of concentration measures alongside medians and percentile ranges, with explicit caution against totals-only interpretation.

Formula 4's value is diagnostic: it evidences fragility and the need for QA gates

The composite prototype demonstrates how calibration can become fragile without strict unit governance, scaling discipline, and publishable evidence baselines, producing outputs that are numerically precise but substantively non-interpretable. The key takeaway within this formula 4 related findings is that if operationalising this formula based approach, then any composite method must define mandatory unit checks, scaling checkpoints, and evidence lineage artefacts before outputs can be presented responsibly.

5.2.2 Lot 5 lessons learned

Lot 5 is deliverable as an auditable evidence system, not as a loose engagement exercise

The credibility of participatory outputs depends on governance discipline: clear boundaries, fixed denominators (i.e., how many participants contributed to each result), completion rules, and controlled analysis windows. In practice, this means an explicit "evidence-lock" protocol that defines fieldwork completion, dataset structure, and the analysis lock window as part of deployment.

Denominator discipline is the foundation of defensible interpretation

Completeness and deviations are not peripheral; they are part of the evidence that determines what can be claimed. Treating partial completions as evidence strengthens credibility because it demonstrates performance under real conditions, including accessibility barriers. In practice, this means consistently reporting completion rates and deviations, and carrying those signals into design recommendations.

Method usage is design coverage, not an adoption signal

Fixed exposure strengthens the defensibility of rankings because participants rank after structured exposure, but it limits causal inference about method effects. The implementation implication is to keep this guardrail prominent wherever method-labelled value summaries are presented, so that differences are not misread as independent uplifts or discounts.

Clarity and fairness are separable constructs, and the split is policy-relevant

Participants distinguish procedural clarity from procedural fairness. Optimising for one does not guarantee the other, which means participatory legitimacy is multi-dimensional. The real-world implication of this finding is that people can and do understand that, although a particular approach may be simpler and easier to follow, it is not necessarily the fairest where fairness is understood as the approach that most accurately and comprehensively reflects real-world conditions by considering a wider range of relevant variables.

Fairness can coexist with low confidence, making usability the scalability gate

A method can satisfy fairness intuitions while remaining difficult to apply, and low confidence can constrain adoption even where fairness perceptions are positive. The implementation implication is targeted usability scaffolding—plain language, guidance, worked examples, and uncertainty prompts—particularly for the method perceived as fairest.

Dispersion persists even under facilitation, and must be managed rather than assumed away

Even with structured facilitation, participants do not converge on a single "correct" number because much of the valuation task is driven by unobservables and judgement calls (below-ground conditions, future use potential, local constraints, and incomplete evidence). Facilitation improves the quality of reasoning and makes assumptions visible, but it does not eliminate the underlying uncertainty. Dispersion therefore behaves like a diagnostic signal: it identifies where evidence is insufficient, where assumptions differ materially, and where a single-point valuation would be overstating precision.

Hard-to-value parcels are best identified through spread diagnostics, with a stable definition

Spread is a defensible proxy for contested valuations under bounded information and is actionable for tool refinement. The implementation implication is to lock the spread definition for publication and avoid mixing threshold schemes unless clearly labelled by aggregation level.

Confidence is not a proxy for convergence

Agreement does not imply correctness, and disagreement does not imply failure. Confidence and dispersion capture different constructs and should be interpreted separately. The implementation implication is to report and interpret them independently to avoid simplistic programme conclusions.

Accessibility and language are methodological variables

Methods that require people to hold several concepts in mind and derive value indirectly (for example, valuing land by subtracting components from an overall figure) are vulnerable to comprehension drop-off, even when the method is transparent on paper. In practice, comprehension is not merely a “communications issue”; it changes the valuation itself by increasing cognitive load, prompting shortcuts, and increasing non-completion or low-confidence responses. The evidence, therefore, implies that accessibility constraints can systematically bias outputs, not just reduce participation.

Trust is a gate, making explainability and contestable assumptions part of the method

Participants accept that assumptions are necessary, but they react strongly to assumptions that are not visible, not contestable, or appear to be “baked in” without justification. Trust is therefore contingent: people will use the output only if they can see what the method is assuming and whether those assumptions align with their understanding of place and context. This means “black box risk” is not a perception problem; it is a legitimacy risk that directly affects whether valuation outputs can be used in policy. A credible valuation method must embed assumption transparency and structured challenge as part of the method itself.

Below-ground uncertainty dominates, so tools must surface it rather than force false precision

Divergence frequently reflects what cannot be observed from the evidence pack. This makes uncertainty capture a minimum viable feature, not an enhancement. The implementation implication is explicit recording of unknowns and uncertainty as part of the value statement, and treating that information as a standard output for interpretation and governance.

5.3 Advantages and disadvantages

Lot 3 advantages are primarily technical and governance-led. Within the audited boundary, it provides re-runnable outputs with parameter-set control and a staged behavioural structure that can be explained and interrogated. This is directly relevant where policy needs require consistent surfaces, repeat reporting, and the ability to revise methods under explicit version control rather than informal recalculation.

Lot 3 disadvantages and risks relate less to the existence of formulae and more to the conditions of publishability. Outputs are sensitive to unit definition and parcel fabric granularity, and totals can be distorted by concentration effects where very large parcels dominate area-weighted results. In addition, where audited extracts omit full driver fields and eligibility reconciliation, the interpretive envelope becomes binding: the results remain valid within the valued analytical base, but the report must avoid overreach into whole-land coverage claims or parcel-level causal explanation.

The practical way these challenges were managed in the programme was through explicit audit boundaries, staged complexity (so that readiness constraints are made visible rather than hidden), and an emphasis on distributional interpretation (medians and percentiles) rather than on area-dominated totals. The most important operational lesson is that scale-up is constrained less by "formula logic" than by fabric governance, feature engineering standards, parameter governance, and audit extracts designed for interpretation rather than only for computation.

Lot 5 advantages are legitimacy and intelligibility evidence. It reveals how people construct value judgements in practice, what they experience as clear versus fair, and where reasoning mechanisms (anchoring, salience, uncertainty) can override technical attributes. This is practical evidence for designing public-facing explanations and engagement processes around land value, including how to present assumptions, risk, and uncertainty without encouraging false precision.

Lot 5 disadvantages and risks are intrinsic to the method. It cannot substitute for a Wales-wide valuation surface where coverage and repeatability are required. It is vulnerable to anchoring and place andstatus heuristics unless mitigations are built into materials and facilitation. It can also generate numeric outputs that appear precise in form while participants' own confidence and qualitative explanations indicate residual uncertainty. The practical lesson is that, if used operationally, the toolkit must be explicit about what it is (structured scrutiny and explanation) and what it is not (a replacement for professional valuation or a definitive valuation instrument).

5.4 What is driving the different land values across Wales?

This question is answered in two layers, because the two lots evidence different categories of drivers.

Layer 1: Attribute-driven drivers evidenced in Lot 3

In the evidence, this ladder behaviour is visible in implied pre-cap medians rising to 165 £/m² under Formula 2 in some areas and penalties reducing medians to 75 £/m² under Formula 3 in others.

Lot 3 outputs vary because the method explicitly encodes drivers and changes outputs when additional drivers are introduced rung-by-rung. At the audited stage, the baseline is structurally schedule-driven (value density by class); stepped modifiers introduce between-area effects; and engineered access/constraint measures introduce within-area differentiation through discounting logic. These are model mechanics derived from defined inputs and parameters. Their interpretability depends on the availability of driver extracts and coverage reconciliation that allow "why" questions to be answered transparently rather than inferred.

Layer 2: Cognition-driven drivers evidenced in Lot 5

Lot 5 shows that non-specialists and professionals do not begin from engineered variables; they typically begin from anchors (often place-based) and then adjust based on what is salient and intelligible in the evidence presented. Uncertainty is frequently converted into discounting or expressed as reluctance to provide false precision. This creates a predictable translation gap: adjacent parcels may differ in a model due to a penalty curve or engineered driver, while participants may explain differences in terms of area reputation, demand narratives, or local knowledge. The implication is not that one is "wrong", but that any policy communication must anticipate this gap and address it explicitly when presenting outputs as public evidence.

5.5 Confidence in robustness of values and of the approach taken

For Lot 3, confidence is highest in (a) reproducibility and computational correctness within the audited analytical parcel fabric and (b) the internal logic of the staged ladder as an interpretive frame. Confidence must be explicitly bounded where the evidence base does not support full transparency of coverage (a reconciled "all land" statement) or parcel-level causal explanation (why a specific parcel differs), particularly where those elements are not present in the audited synthesis extracts. The report should therefore distinguish robust execution of an audited run from robust claims about market-true values.

For Lot 5, confidence is highest in recurring behavioural patterns supported by triangulation across structured instruments, rationales, and qualitative evidence, and in operational learning about intelligibility, legitimacy, and cognitive load. Confidence is lower for generalising numeric distributions beyond the workshop evidence base, given purposive sampling and small group sizes, and for treating participant numeric outputs as comparable to market valuations without a completed benchmarking hierarchy that is explicitly stated and evidenced.

At programme level, robustness is strongest when the two approaches are treated as a learning system: one stream evidences what can be run, governed, and scaled (Lot 3), while the other evidences what can be understood, scrutinised, and accepted (Lot 5).

5.6 Land types and each approach

Lot 3

At the current maturity, land is easier to value where it is consistently parcelised under the fabric rules, is not dominated by extreme parcel size heterogeneity, and has the required engineered features available in auditable form. Land becomes harder to value where outputs become highly sensitive to definitional choices and unit effects, particularly where very large heterogeneous parcels dominate totals or where the value concept is weakly market-revealing without additional governance. In such cases, "hard to value" is often better described as "hard to publish responsibly without additional safeguards".

Lot 5

Land is easier to value where participants can form and justify an anchor using a small number of legible cues and where uncertainties are bounded. Land becomes harder to value where uncertainty is structural (particularly below-ground unknowns), where risk information is technical and not salient unless prompted, and where participants refuse false precision (ranges, rounding, non-completion). In conclusions, these should be presented as design-relevant signals: hard-to-value contexts are frequently hard-to-explain and hard-to-standardise in public-facing engagement.

5.7 Scalability for the whole of Wales

Lot 3 is, in principle, scalable across Wales because its core value is a repeatable pipeline and parameter-governed execution. However, scale-up is conditional on governance and data readiness. The evidence indicates that scaling will be driven less by the formulae themselves than by stable fabric construction rules, consistent feature engineering, licensing and refresh cycles, Quality Assurance thresholds, and an auditable coverage an eligibility reconciliation that prevents misinterpretation of excluded land and White Space. Lot 5 is not scalable as a Wales-wide parcel valuation system because workshop delivery cannot generate national parcel coverage. It is, however, scalable as a structured engagement and legitimacy layer. It can be expanded through stratified workshop programmes (by area type and participant group) and through toolkit refinements that reduce cognitive load and improve accessibility, while preserving auditability of materials, deviations, and learning outputs.

5.8 Wider requirements needed to value the whole of Wales

To move to an operational, Wales-wide evidence product requires the requirements to be stated in a way that maps directly to the risks surfaced across both lots. The programme evidence indicates that the key risks are not limited to whether results can be produced, but whether they can be reproduced, explained, and defended under scrutiny, and whether they remain intelligible and credible to non-specialist audiences.

5.8.1 Requirements for an auditable, scalable computational approach (Lot 3-led)

For a Wales-wide computational approach to be auditable and scalable, it must be possible to state clearly—by reporting geography—what land has been valued, what has not, and why. This requires a published eligibility and coverage reconciliation layer that quantifies the relationship between total area, valued parcel area, and explicit exclusions. In particular, it must quantify and explain White Space and any other structural exclusions, so that “gaps” are not misread as missing data or undervaluation.

In addition, published outputs need to be designed for interpretation, not only calculation. This means that the audit extracts made available for scrutiny should include the fields needed to answer common “why” questions using the published evidence pack, rather than relying on informal re-analysis. It also means routinely publishing concentration diagnostics alongside headline totals, so that readers can see when results are dominated by a small number of very large parcels and avoid totals-only interpretation.

Operational defensibility also depends on explicit, version-controlled parameter governance. Settings must be published in a way that is clear, versioned, and defensible, including calibration provenance, run identifiers linking outputs to the configuration used, and documented sensitivity testing where the mechanics of the model materially shape results. This is the practical route to ensuring that changes over time are explainable and that future iterations can be assessed against a stable baseline. In addition to these required factors, any Wales-wide deployment would always require a clear benchmarking hierarchy for any external comparisons and a transparent statement of coverage and comparability. This avoids over-claiming where administrative or professional evidence does not align cleanly with the analytical parcel fabric or with the bounded design of the research, and ensures that comparisons are used appropriately and consistently.

5.8.2 Requirements for public intelligibility, scrutiny, and acceptance (Lot 5)

In parallel, the programme evidence indicates that technical auditability is necessary but not sufficient: the evidence product must also be intelligible, contestable, and credible to non-specialists. This requires a public-facing explanation framework that anticipates anchoring and place-based reasoning, distinguishes procedural clarity from procedural fairness and legitimacy, and explains why adjacent parcels can reasonably differ. That explanation must explicitly cover definitional boundaries, including White Space, because these boundaries strongly influence how outputs are interpreted in practice.

The toolkit used to support participatory understanding and scrutiny also needs refinement so that it functions well under real-world conditions. This includes reducing cognitive load, enabling participants to record uncertainty (including ranges and confidence), and treating non-completion as an explicit signal about usability and evidence sufficiency rather than as silent missing data. These design features are required to prevent false precision and to ensure that the evidence captures both the value statements and the limits of what can be concluded from the information available.

Accessibility and bilingual delivery must be treated as core design variables rather than as presentation choices. Where deviations occur in delivery, they should be recorded *and* paired with mitigation strategies so that operational learning is preserved and can be scaled responsibly. This is necessary both for fairness of access and for the validity of the evidence collected.

5.8.3 Combined end-state implied by the programme

Taken together, the programme implies a combined end-state: a Wales-wide computational evidence base that is repeatable and auditable, paired with a participatory and communications layer that makes the evidence usable, contestable in a structured way, and publicly intelligible. This framing preserves a disciplined separation between coverage and legitimacy. It avoids implying that participatory processes substitute for national scale, while also avoiding the assumption that technical outputs alone are sufficient to confer public acceptance.

6 Further Considerations

6.1 Additional research to refine the evidence base

If the purpose of the programme evolves from feasibility testing towards an operational evidence product, the external verification layer needs to move from light-touch sense-checking to structured, repeatable benchmarking with clear hierarchies of evidence and explicit disagreement capture. A next-stage approach should include a structured panel review rather than reliance on a single professional comparator. This panel should be stratified by land class, geography type and value bands, with consistent prompts that ask reviewers to assess plausibility, assumption reasonableness, and sources of divergence. Disagreement should be recorded systematically (not treated as failure), because disagreement itself is evidence about sensitivity, definitional ambiguity and where further governance is required.

Benchmarking should also be formalised as a hierarchy, recognising that different evidence sources are valid at different scales and for different questions. Parcel-level comparators will not always align cleanly with analytical parcels; neighbourhood-scale signals may be more defensible for some comparisons; and administrative datasets may be informative but not directly commensurate with the analytical unit. A published hierarchy would reduce the risk of inappropriate “apples-to-oranges” comparisons and would provide readers with a clear basis for what counts as stronger or weaker verification evidence. Where residual-style evidence is used, the residual specification should be strengthened and the sensitivity logic should be published in annex form, even if licensing constraints prevent full disclosure of proprietary components. The priority is not to reveal restricted data, but to document how the programme stress-tested the residual assumptions and how sensitive outputs are to rebuild-cost baselines, cleaning rules and geocoding granularity.

6.1.1 Refining the parcel fabric and eligibility definition (Lot 3)

To reduce misinterpretation risk and improve transparency, future work should make “what is valued” and “what is excluded” numerically explicit and repeatable for every reported geography.

A core next-step is an eligibility and coverage reconciliation layer that quantifies, for each geography, the valued parcel area, the area excluded as White Space, and the area excluded for other definitional reasons. This should be published alongside maps and summary tables so that “gaps” cannot be misread as missing data or “zero value”. Where exclusions are policy-relevant (for example, transport corridors or water), the reconciliation should separate these categories rather than merging them into a single remainder.

Classification and mapping rules should be formalised as a published reference table with examples of common edge cases. This is important because explanation cannot depend on a technical audience reading annexes; non-specialist interpretability requires that the boundary-setting rules are available in plain language, with concrete examples that show how surfaces are mapped into classes and how exclusions are applied.

Further controlled refinements to the fabric may be warranted where large parcels dominate totals or distort interpretability, but these should only be introduced as parameterised, auditable transformations with clear before/after diagnostics. The guiding discipline is that refinements must remain re-runnable and version-controlled; manual patching would undermine auditability and weaken defensibility.

6.1.2 Improving driver completeness and feature readiness (Lot 3)

If the ladder is intended to represent increasing realism, higher rungs should add drivers that are both salient and explainable. Future work should therefore treat feature readiness as a publishability condition, not merely a modelling convenience.

Accessibility measures could be extended beyond straight-line distances through structured sensitivity testing (for example, network distances, service density proxies or multi-node accessibility composites), but only where added complexity demonstrably improves interpretability and remains auditable. Complexity should not be introduced solely because it is technically possible; it should be introduced because it clarifies rather than obscures the story of why outputs differ.

Constraint coverage could also be expanded beyond flood exposure where feasible and appropriate, including planning constraints, protected designations and contamination proxies. Each added driver should be documented consistently: its data source and refresh cycle, operational definition, why it matters, and how it affects value (flag, penalty, interaction). Without this documentation discipline, drivers become difficult to explain publicly and risk being perceived as arbitrary.

A specific cross-lot requirement is that any additional drivers should be tested for salience in participatory materials. If a driver is technically important but not publicly intelligible, its inclusion may improve computational realism while simultaneously harming perceived legitimacy unless communication design keeps pace.

6.1.3 Deepening participatory learning and reducing bias risk (Lot 5)

The next stage for participatory work is to move from describing how people reason under a given toolkit design to identifying how to design a public tool that produces stable, intelligible judgements with managed bias risks.

A practical development step is controlled A/B testing of workshop materials to isolate behavioural mechanisms: location salience, risk salience, and assumption visibility. For example, comparing anonymised versus recognisable parcels would quantify how much place anchoring changes reasoning patterns; presenting risk information using alternative formats (maps versus categorical flags versus narrative prompts) would test how salience affects incorporation of constraints.

The evidence base on “clarity versus fairness” should be deepened by testing alternative framings of the three methods and their explanations, and observing how rankings, confidence and rationales shift. The objective is not to engineer agreement, but to identify which explanation designs reduce misunderstanding, improve comfort, and preserve procedural legitimacy.

Targeted workshops with under-represented groups, including those with lower technical confidence, would strengthen accessibility learning and reduce the risk that toolkit refinement is inadvertently biased towards those most comfortable with technical or quasi-technical reasoning. Accessibility should be treated as a methodological variable with its own Quality Assurance discipline, not as a delivery afterthought.

6.2 How findings could have been more refined

6.2.1 Data quality and data processing discipline

For participatory datasets, the Quality Assurance pipeline should include automated integrity checks as standard deliverables: expected row counts by key, uniqueness checks, checksum and reconciliation totals, and automated exception reports. This reduces the risk of structural issues (such as misalignment or duplication) persisting into analysis and provides an auditable record that integrity gates were passed.

For formula outputs, each audited run should produce a standard interpretability export as a default artefact: parcel identifier, class, key driver fields, exclusion reason (where relevant), and rung outputs. This enables reporting teams to answer “why” questions from published artefacts without informal re-analysis or bespoke extraction work and reduces interpretive dependence on technical staff. A clearer separation in the data pipeline between analytical values (value_gbp) and display-capped rates (rate_gbp_per_m2) from the outset would have prevented initial presentational inconsistencies and improved interpretability.

6.2.2 Denominators, missingness and uncertainty capture

Uncertainty should be treated as a primary output in participatory exercises, not simply as commentary. Ranges, confidence statements, and explicit missingness reasons should be captured systematically and analysed as signals about cognitive load, information insufficiency and where public explanation will struggle. For model outputs, a parallel robustness indicator should be developed so readers do not assume all outputs are equally reliable. This indicator does not need to pretend to be a statistical confidence interval; it can be a structured rating informed by driver readiness, parameter sensitivity and fabric stability. The key point is to make “how robust is this signal?” visible rather than implicit.

6.2.3 Cross-lot alignment for stronger triangulation

Triangulation would be strengthened by increasing overlap between parcels discussed in workshops and parcels represented in the analytical fabric. This allows qualitative reasoning mechanisms to be mapped more directly onto technical outputs, improving the ability to explain divergence without implying equivalence. Participatory findings should also be used to prioritise technical drivers that matter for perceived legitimacy (for example, how risk is presented, what comparability narratives are accepted, how assumptions are explained). Where those drivers are then added or refined technically, their inclusion should be tested again participatorily to check whether technical improvements also improve intelligibility and acceptance.

6.3 Explaining the approach to members of the public

The programme evidences that legitimacy is gated by explainability, and that people distinguish between what is clear and what is fair. A credible public intelligibility strategy should therefore be operational and repeatable, not dependent on ad hoc narratives.

6.3.1 The minimum public explanation model

A public-facing explanation should answer five questions every time, in plain language:

1. What land is included and excluded, and why.
2. What “value” means in this context: an indicative signal, not a guaranteed sale price.
3. What factors drive the signal, with concrete examples.
4. How uncertainty and constraints are handled, including what happens when information is missing.
5. How a person can query, challenge or seek clarification on an output, including what evidence would be relevant.

These are not communications preferences; they are governance requirements if outputs are used in a public-facing setting.

6.3.2 Use the clarity versus fairness distinction explicitly

Communications should not assume that one explanation satisfies both clarity and fairness. A practical approach is two-layered:

A clarity layer that uses comparable, everyday framing: similar land in similar contexts tends to fall within a given range, with the explanation emphasising recognisable reference points and transparent summaries.

A fairness layer that emphasises procedural consistency: a consistent set of factors is applied so land is treated the same way across Wales, with visible rules and contestable assumptions.

This is not messaging spin. It aligns with the behavioural evidence that legitimacy depends on both comprehensibility and perceived procedural consistency.

6.3.3 Make assumptions visible and contestable

Assumptions should be published as part of the output, not hidden in annexes. A short “assumptions card” can list key assumptions and why they were used, alongside worked examples showing how changing an assumption changes the estimate (sensitivity demonstration, not false certainty). Black box concern should be treated as a primary risk: if users cannot see how factors shape results, trust will decline even if the underlying method is technically strong.

6.3.4 Explain why adjacent parcels can differ

Adjacent-parcel differences are a predictable flashpoint. Public explanation should therefore include a plain-language rationale for why neighbouring parcels can differ because constraints, access, classification and exclusions can differ, and because the analytical parcel fabric is not identical to legal title boundaries. Exclusions (including White Space) should be described explicitly as deliberate scope rules rather than missing data.

6.4 Future potential within a Welsh context

A Wales-wide computational evidence base is, in principle, achievable through a repeatable, audited pipeline, but only if coverage reconciliation, driver readiness and parameter governance are mature enough to sustain national roll-out without misinterpretation. Even if a national model exists, participatory evidence remains essential as the legitimacy layer: it supports scrutiny, tests intelligibility, and helps design the public explanation framework that determines whether outputs are accepted as credible. Without taking a policy position, the programme implies three requirements for any land value tax style instrument or related land policy tool. First, exceptionally clear scope definitions and a defensible treatment of exemptions and non-market land uses are non-negotiable, because definitional boundaries directly determine perceived fairness and dispute risk. Second, public acceptance will depend on perceived fairness, consistency, and the ability to understand and contest outcomes. Participatory evidence indicates that legitimacy is not secured by technical strength alone. Third, technical scalability is achievable only if governance and interpretability exports are treated as core deliverables rather than optional technical extras. If outputs are intended to support fiscal or quasi-fiscal instruments, the evidential standard for transparency, auditability and dispute handling must be designed in from the outset.

References

- Danish Agency for Digital Government, 2025. Property assessments and property taxes. Life in Denmark (borger.dk). Access: https://lifeindenmark.borger.dk/housing-and-moving/buying/property-assessments-and-property-taxes
- Danish Tax Agency (Skattestyrelsen), n.d. Introduction to property in Denmark. skat.dk. Access: https://skat.dk/en-us/individuals/property/introduction-to-property-in-denmark
- Skatteforvaltningen, 2026. The Legal Guidance 2026-1: H.A.5.2.1 General information on assessing land value for owner-occupied homes. info.skat.dk. Access: https://info.skat.dk/data.aspx?oid=2296671
- Danish Property Assessment Agency (Vurderingsstyrelsen), n.d. English. vurdst.dk. Access: https://vurdst.dk/english
- The Danish Parliamentary Ombudsman, 2025. The Ombudsman looks at the Property Assessment Agency's possibilities for explaining calculations behind property assessments. en.ombudsmanden.dk. Access: https://www.en.ombudsmanden.dk/news/2025/the-ombudsman-looks-at-the-property-assessment-agency%E2%80%99s-possibilities-for-explaining-calculations-behind-property-assessments
- The Danish Parliamentary Ombudsman, 2021. 20-3523 Vurderingsstyrelsens begrundelser i det nye ejendomsvurderingssystem. ombudsmanden.dk. Access: https://www.ombudsmanden.dk/Media/638436673186546531/20-3523_Vurderingsstyrelsens_begrundelser_i_det_nye_ejendomsvurderingssystem.pdf
- Danish Property Assessment Agency (Vurderingsstyrelsen), 2024. Vurderingsstyrelsen tager Folketingets Ombudsmands kritik til efterretning. vurdst.dk. Access: https://vurdst.dk/nyheder/vurderingsstyrelsen-tager-folketingets-ombudsmands-kritik-til-efterretning
- Danish Property Assessment Agency (Vurderingsstyrelsen), 2023. Proces for manuel håndtering af de ejendomme, hvor grundværdien er markant højere end ejendomsværdien. vurdst.dk. Access: https://vurdst.dk/media/foocrw50/bilag-f-proces-for-manuel-haandtering-af-de-ejendomme-hvor-grundvaerdien-er-markant-hoejere-end-ejendomsvaerdien.pdf
- Schjødt, 2023. New property assessments in Denmark and new rules on property taxation. schjodt.com. Access: https://schjodt.com/news/new-property-assessments-in-denmark-and-new-rules-on-property-taxation?utm_source=chatgpt.com
- Bundesportal, n.d. Standard land values: inspection of the standard land value map and information. verwaltung.bund.de. Access: https://verwaltung.bund.de/leistungsverzeichnis/EN/leistung/99012109023000/herausgeber/TH-731008/region/160000000000

- City of Bocholt, n.d. Expert committee for property values in the city of Bocholt. bocholt.de. Access: https://www.bocholt.de/en/gutachterausschuss
- BORIS-D, n.d. BORIS-D: Bodenrichtwertinformationssystem für Deutschland. bodenrichtwerte-boris.de. Access: https://www.bodenrichtwerte-boris.de/
- City of Stuttgart, n.d. Standard land values with information on property tax B. stuttgart.de. Access: https://www.stuttgart.de/en/leben/bauen/grundstueckswerte/bodenrichtwerte-fuer-die-grundsteuer/Bodenrichtwerte-mit-Hinweisen-zur-Grundsteuer-B
- State of Brandenburg, 2024. Verwaltungsvorschrift zur Ermittlung von Bodenrichtwerten im Land Brandenburg (Brandenburgische Bodenrichtwertvorschrift – VV-BRW). bravors.brandenburg.de. Access: https://bravors.brandenburg.de/verwaltungsvorschriften/vv_brw/2
- International Centre for Tax and Development (ICTD) and partners (Grieco, Kamara, Meriggi, Michel, Prichard, Stewart-Wilson), 2019. Simplifying Property Tax Administration in Africa: Piloting a Points-Based Valuation in Freetown, Sierra Leone. ICTD (Summary Brief No. 19), ictd.ac. Access: https://www.ictd.ac/wp-content/uploads/2019/11/ICTD_SumBrief19_Freetown_2.0.pdf
- International Centre for Tax and Development (ICTD), 2020. Freetown just implemented a new property tax system that could quintuple revenue. ictd.ac. Access: https://www.ictd.ac/blog/freetown-new-property-tax-system-quintuple-revenue/
- Kamara, A.B., 2023. Property Tax Reform: Lessons from Freetown. African Tax Administration Forum (ATAF), ataftax.org. Access: https://ataftax.org/wp-content/uploads/2023/01/Property_Tax_Reform_Lessons_from_Freetown_ENG.pdf

Appendix 1: Lot 3 Audit Register and Publication Criteria

1.1 Purpose and audited cut-off date

This appendix provides the minimum information needed to (a) understand what datasets and extracts underpin the reported Lot 3 outputs, and (b) evidence that the outputs were produced under controlled, publishable Quality Assurance conditions.

For the consolidated report, the audited reporting cut-off is set to **9th of January 2026**. Most underlying inputs and extracts were updated and/or finalised during the first week of January 2026 (between **5 January 2026 and 9 January 2026**).

1.2 Interpretation guardrails

These points are included here to prevent common misreadings when the report is reviewed without technical context:

- Lot 3 outputs are produced on an **analytical parcel fabric** (a modelling unit), not a legal title parcel base.
- **Coverage** describes the proportion of the reporting geography represented by the analytical parcel fabric and eligible for valuation outputs; it is not a measure of accuracy.
- **White Space** and other exclusions are **structural by design** (for example, operational transport land and water where excluded). Map "gaps" should be read as exclusions unless explicitly stated otherwise.
- Capped map display is a **visualisation control**; analytical interpretation should be anchored to uncapped distributions and parcel/value totals.

1.3 Dataset register and extract metadata

Dataset / Extract	Provider / Publisher	Product / Publication reference	Final update date/audited outputs	Key use in audited outputs
INSPIRE Index Polygons	HM Land Registry	INSPIRE Index Polygons	7 Jan 2026	Analytical parcel fabric
WIMD Access to Services (quartile)	Public Health Wales (as published)	WIMD domain extract / publication	7 Jan 2026	Area-level access signal (joined by LSOA)

Flood zones used for constraint	Natural Resources Wales	Flood Map for Planning (or equivalent)	10 DEC 2025	Flood exposure share (flood_share)
NaPTAN access nodes	Department for Transport	NaPTAN	7 Jan 2026	Distance-to-node features
Roads network used for distance metrics	Ordnance Survey	OS Roads	8 Jan 2026	Distance-to-road features
WorldCover (fallback only, if invoked)	European Space Agency	WorldCover 2021 v200	9 Jan 2026	Fallback mapping for residual unclassified parcels

1.4 Publication criteria (QA stop/go checks)
The following checks are treated as publication criteria. If any "stop" condition is triggered, outputs should not be published; instead, the failure should be reported and corrected upstream, then the run repeated under the same audited cut-off logic.

1.4.1 Geometry and CRS checks

- All parcel geometries are valid, or validity repairs are applied using declared rules (no silent coercion).
- All area and distance calculations are performed in a consistent projected CRS appropriate for measurement (e.g., British National Grid), with transformations documented.
- Deduplication and overlap handling are applied consistently, with exceptions logged.

1.4.2 Coverage and completeness checks (stop conditions)

- A coverage reconciliation statement exists for each reporting geography: total area, valued parcel area, and explicit exclusions (including White Space).
- Where any required input is incomplete for the reporting geography, the run stops. This includes missing tiles for any gridded/raster fallback inputs used to classify residual parcels.

1.4.3 Attribute attachment and row-integrity checks (stop conditions)

- Join completeness: each parcel has the required attributes attached for the rung being executed (class, engineered features, and any flags).
- Row parity: exactly one output record per parcel per rung; no join leakage or accidental duplication.

1.4.4 Classification completeness checks (stop conditions)

- No residual unclassified parcels remain in the publishable parcel set after the defined attribution and fallback sequence, unless explicitly labelled and excluded from publication totals.

1.5 Minimum audited artefacts (what must exist for reproducibility)
For each audited run, the following artefacts must be retained and attributable to the run identifier:

- Run metadata: run identifier, parameter-set identifier, and the dataset register reference (Appendix 1.3).
- Parcel extract: parcel identifiers, geometry, LSOA code, eligibility/exclusion flags, class, and area.
- Feature tables: engineered access/constraint variables with units and definitions.
- Rung outputs: one row per parcel per rung including rate and value outputs.
- Quality Assurance log: pass/fail results for the stop/go checks in Section 1.4, including any exceptions and resolutions.

Appendix 2: Lot 5 Governance Summary and Instruments Index

2.1 Purpose and evidence boundary

This appendix records the minimum governance information required to interpret Lot 5 as an auditable evidence system. It is intentionally concise. Full workshop materials (workshop guide, exercise guide, discussion guide, participant workbook templates, and digital capture templates) are supplied separately as part of the project resource handover.

Lot 5 outputs should be interpreted as evidence about intelligibility, legitimacy (including fairness perceptions), usability, and uncertainty under structured exposure, rather than as definitive valuations.

2.2 Denominators and completion rules

Lot 5 reporting must be anchored to explicit denominators and completion rules. The minimum requirements are:

- A clear statement of the eligible participant base for each workshop and instrument.
- A consistent definition of “complete”, “partial”, and “not completed” for each instrument.
- Treatment of partial completion and non-completion as evidence (for example, accessibility or cognitive burden), not as silent missingness.

2.3 Instruments index

This index lists the instruments used to capture structured evidence. Full instrument content is provided in the resource handover.

2.3.1 Structured method materials (method cards)

- Method card set (four methods): **[Insert method names as used in the workshop pack]**
- Bilingual versions used: **[Insert yes/no and language handling notes]**

2.3.2 Participant workbook / pack

- Participant workbook / pack: parcel briefs, worked examples, valuation capture prompts (including ranges and confidence), and reflection prompts.
- Accessibility features included (terminology support, pacing prompts, etc.)

2.3.3 Calibration drill materials (“toy parcel” exercise)

- Calibration exercise: simplified parcels used to baseline confidence and surface divergence drivers.
- Recording mechanism (paper/digital): **[Insert]**

2.3.4 Digital capture instruments (e.g., Forms)

- Comprehension check
- Clarity ranking
- Fairness/legitimacy ranking
- Exit survey (trust, assumption visibility/contestability, improvement needs)

2.4 Accessibility and bilingual delivery (validity conditions)

Accessibility and bilingual delivery are treated as methodological variables. Deviations (for example, where terminology comprehension created barriers) should be recorded alongside mitigation steps so the learning can be scaled responsibly.

Appendix 3: Digital Artefacts, GitHub Codebase, and Data Handover Note

3.1 GitHub codebase (Lot 3 pipeline and supporting documentation)
A GitHub has been provided to Welsh Government containing the reproducible Lot 3 codebase and supporting documentation required to re-run the pipeline as an auditable process. The repository is intended to include:

- Pipeline code for ingestion, processing, feature engineering, formula execution, and export.
- Configuration and parameter specifications (including versioning approach and run identifiers).
- Environment/dependency documentation (setup notes and runbook).
- Output schema documentation (tables/fields for parcel extracts, features, rung outputs, and Quality Assurance logs).
- Quality Assurance checks implemented as explicit pass/fail criteria consistent with Appendix 1.4.
- The repository is not intended to embed controlled datasets unless they are explicitly open and distributable under their licensing terms.

3.2 Data provision and secure transfer arrangements
Data required to re-run the pipeline is being handled via agreed Welsh Government arrangements. Where datasets are access-restricted, controlled, or otherwise unsuitable for embedding within a public code repository, they are transferred via the agreed secure route and governed by the relevant licensing and data-sharing terms.