Journal of Science and Technology in Civil Engineering, HUCE, 2026, 20 (2): 44–54

CROSS-DWELLING VALIDATION OF INDOOR ENVIRONMENTAL MONITORING FOR OPERATIONAL RISK SCREENING IN SOCIAL HOUSING PORTFOLIOS

Thuy-Hien Thi Nguyen[a,b], Ruoyu Jin[b,*], Weifeng Chen[b], Lu Gan[b]

[a] *Research group on Industry 4.0 in Transportation (I4T group), University of Transport Technology, 54 Trieu Khuc road, Thanh Liet ward, Hanoi, Vietnam*
[b] *Brunel University of London, Kingston Lane, Uxbridge, UB8 3PH, United Kingdom*

Article history:
Received 20/3/2026, Revised 03/5/2026, Accepted 04/5/2026

Abstract

Social housing providers use indoor environmental monitoring within asset management systems. The extent to which these data can differentiate operational risk domains across independent dwellings has not been fully evaluated in operational deployment. Current predictive modelling frequently relies on random data partitioning, failing to reflect situations in which models are applied to previously unseen properties. This study examines the cross-dwelling explanatory capacity of environmental exposure indicators within a London-based housing portfolio. Five years of monitoring data from 93 UK social housing dwellings were linked with operational risk records, yielding 5,748 monthly dwelling-level observations. Indicators derived from temperature, relative humidity, and carbon dioxide were analysed using Ridge regression and Random Forest models under five-fold property-grouped cross-validation. Under grouped validation, indoor air quality and excess heat domains show positive explanatory power across dwellings. In contrast, envelope-related domains, including heat loss, draught, and cold home risks, produce near-zero or negative R^2 values, indicating limited cross-dwelling information in bulk indoor environmental measurements. Random Forest models do not consistently improve over regularised linear models. These findings identify the risk domains that can be informed by environmental screening at portfolio level and those which require further or direct structural assessment within asset management practice.

Keywords: indoor environmental quality (IEQ); social housing; asset management; internet of things (IOT); housing risk classification; portfolio management.

https://doi.org/10.31814/stce.huce2026-20(2)-04

1. Introduction

Indoor environmental conditions in social housing affect occupant health, regulatory compliance, and asset performance. Dampness and mould are associated with respiratory outcomes [1, 2], and inadequate ventilation increases indoor pollutant concentrations [3]. In housing asset management, these risks are commonly identified through tenant reporting and scheduled inspections. Such approaches may not capture intermittent ventilation deficiencies or gradual fabric-related deterioration [1].

Internet of Things (IoT) systems now enable continuous measurement of temperature, relative humidity, and carbon dioxide across housing stocks [4, 5]. These data are incorporated into asset management processes, including Net-Zero retrofit programmes in the United Kingdom. Environmental monitoring is being used alongside inspection-based methods.

The asset management system examined in this study covers 1,597 social housing properties. Monitoring devices were installed in 93 dwellings within this portfolio. Portfolio-wide deployment

*Corresponding author. *E-mail address:* ruoyu.jin@brunel.ac.uk (Jin, R.)

was not feasible because of installation, maintenance, and data management constraints. The monitored dwellings were selected by the housing provider to reflect the range of property types and construction characteristics within the portfolio. The 93-dwelling cohort represents the operational monitoring sample available for analysis.

The ability of routinely collected indoor environmental indicators to differentiate operational risk domains across independent dwellings has not been evaluated in operational deployment. Studies on overheating [6] and ventilation dynamics [3] are typically based on individual buildings or limited samples. Predictive analyses of housing datasets also frequently apply random data partitioning. When panel data include repeated observations from the same dwelling, random partitioning allows dwelling-specific patterns to appear in both training and testing subsets. This does not reflect deployment conditions, where models are applied to dwellings not used in model development.

For asset management, the key question is whether environmental exposure indicators explain variation between dwellings within a portfolio, rather than variation within a single dwelling over time. This requires validation procedures that preserve dwelling-level independence and assess cross-dwelling explanatory capacity.

This paper evaluates whether longitudinal indoor environmental exposure indicators explain variation in recorded operational risk domains across a UK social housing portfolio. Five years of monitoring data from 93 dwellings are linked with asset management risk records. Property-grouped cross-validation is applied so that model evaluation is conducted on dwellings excluded from model training. The analysis identifies domains for which environmental monitoring provides measurable cross-dwelling explanatory capacity and domains for which variation is not reflected in bulk indoor air measurements.

2. Literature review

2.1. Indoor environmental monitoring in residential assets

Research on indoor environmental quality in housing has largely examined individual dwellings or small samples. Studies of overheating [6] and ventilation dynamics [3] commonly analyse single buildings or limited groups of properties. These investigations describe environmental conditions in detail but do not assess whether exposure indicators differentiate risk across independent dwellings within large housing portfolios. Distributed sensor networks allow continuous measurement of temperature, relative humidity, and carbon dioxide across multiple residential properties [4]. In housing portfolios, monitoring systems generate longitudinal datasets with repeated observations recorded at dwelling level. This structure differs from cross-sectional building assessments and affects how model performance should be evaluated.

2.2. Validation in panel-structured housing data

Housing monitoring datasets typically contain multiple observations for each dwelling over time. In panel data of this type, model validation strategy determines what performance metrics represent. Random data partitioning is widely used in predictive analyses of housing datasets [7, 8]. When applied to panel data, observations are assigned to training and testing subsets without preserving dwelling identity. Measurements from the same dwelling may appear in both subsets, which often leads to data leakage and over-optimistic performance estimates [9]. Under this arrangement, performance reflects consistency associated with specific dwellings rather than differences between dwellings. Portfolio-level asset management concerns variation across properties [10]. Validation procedures that group observations by dwelling separate entire properties between training and testing subsets. This allows assessment of explanatory capacity across independent dwellings under conditions consistent with application to properties not used in model development.

2.3. Application in housing asset management

Ventilation deficiencies and fabric-related deterioration are commonly identified through tenant reporting and scheduled inspections [1]. These methods provide periodic assessment. Continuous monitoring produces longitudinal records of temperature, relative humidity, and carbon dioxide [5]. When aggregated at dwelling level, these indicators can be used for routine screening across a portfolio. Environmental screening may assist in identifying properties with sustained exposure associated with ventilation or overheating. Screening does not replace physical inspection. It may inform prioritisation of maintenance visits and retrofit planning where resources are limited.

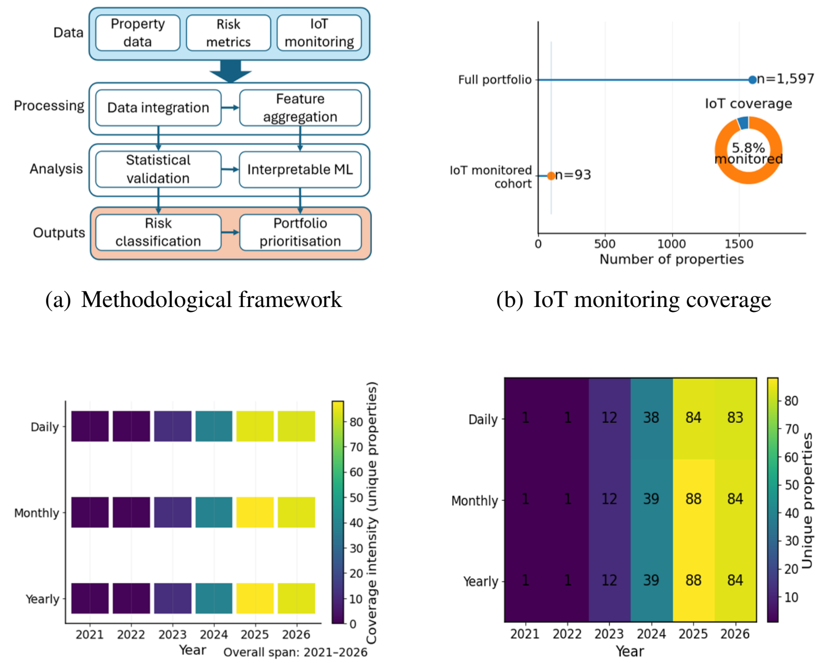
2.4. Measurement scope

Environmental sensors record indoor air conditions within the monitored spaces. Structural characteristics of the building envelope are not captured through these measurements. Thermal transmittance, air permeability, and junction detailing cannot be directly inferred from bulk air parameters alone. Interpretation of environmental indicators should remain consistent with this measurement boundary. Certain operational risk domains may correspond to measurable environmental exposure patterns. Other domains are primarily determined by structural conditions and require direct technical assessment within maintenance planning frameworks.

3. Methodology

3.1. Portfolio and data sources

The dataset links longitudinal indoor environmental measurements with operational housing risk records from a social housing provider. These operational risk records are hybrid, combining tenant-reported maintenance requests with subsequent technical inspections performed by professional surveyors. The overall methodological workflow is illustrated in Fig. 1(a), while the portfolio coverage and temporal data distribution are detailed in Figs. 1(b)–(d).



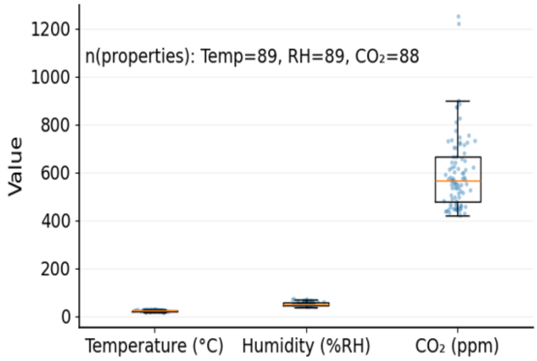
(a) Methodological framework

(b) IoT monitoring coverage

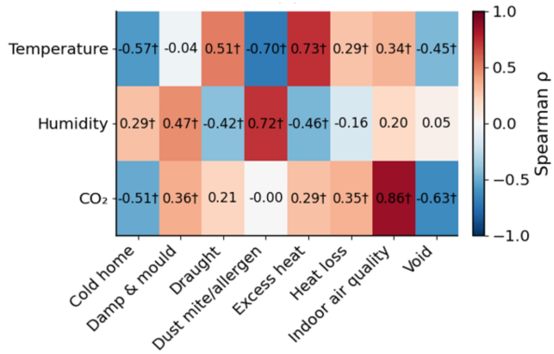
(c) Temporal coverage intensity

(d) Unique monitored properties

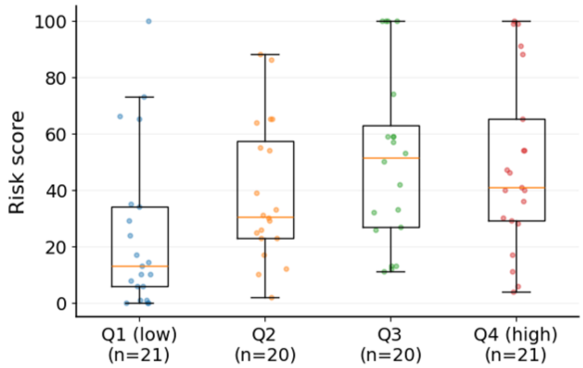
Figure 1. Methodological framework and dataset characteristics



(a) Environmental exposures



(b) Indicator-risk correlations



(c) Risk scores by quartile

Figure 2. Exploratory data analysis of the monitored cohort

The asset management system includes 1,597 properties. Environmental monitoring devices were deployed across a broader set of dwellings during the monitoring programme. Following linkage between sensor files and asset management records, and after applying completeness criteria to ensure alignment of environmental and risk information, 93 dwellings satisfied the requirements for inclusion in the analysis. These dwellings cover the principal property types within the stock. Exploratory data analysis of the monitored cohort, including environmental exposures and risk domain correlations, is presented in Fig. 2.

Indoor environmental conditions were recorded in primary living areas at approximately 1.1 m above finished floor level, consistent with established residential monitoring practice [11]. Air temperature, relative humidity, and carbon dioxide (CO_2) were measured at 15-minute intervals over five years.

Sensor data were filtered through an automated validation procedure based on instrument specifications and physical feasibility constraints. Records outside the operational measurement bounds were removed prior to aggregation. The specified bounds were 0 °C to 50 °C for air temperature, 0 to 100 percent for relative humidity, and 250 ppm to 10,000 ppm for CO_2 concentration. No interpolation or smoothing was applied during this step.

The cleaned measurements were aggregated to monthly dwelling-level indicators and linked with operational risk records. The analytical dataset comprises 5,748 dwelling-month observations. The data exhibit panel structure, with repeated monthly records for each dwelling. Sample size varies across risk domains due to eligibility criteria applied during model construction.

3.2. *Risk construct formulation and data structure*

Environmental exposure indicators were derived by aggregating 15-minute indoor measurements to monthly dwelling-level metrics. Carbon dioxide concentration was used as an indicator of ventilation adequacy. Sustained indoor CO_2 levels above 1,000 ppm are commonly interpreted as reflecting reduced ventilation effectiveness in residential settings [12]. Moisture exposure was defined by relative humidity exceeding 70 percent, which is associated with increased condensation and mould risk [1]. Excess heat exposure was defined using a 26 °C threshold based on residential overheating criteria [13]. Cold exposure was defined using an 18 °C threshold in line with minimum indoor temperature guidance for health protection [1].

Operational housing risk domains were extracted independently from the asset management system. Recorded domains included cold home, damp and mould, dust mite allergens, excess heat, heat loss, draught, indoor air quality, and void classification. After temporal alignment and linkage, each analytical observation corresponds to a dwelling-month unit combining environmental exposure indicators and the associated domain-specific risk status.

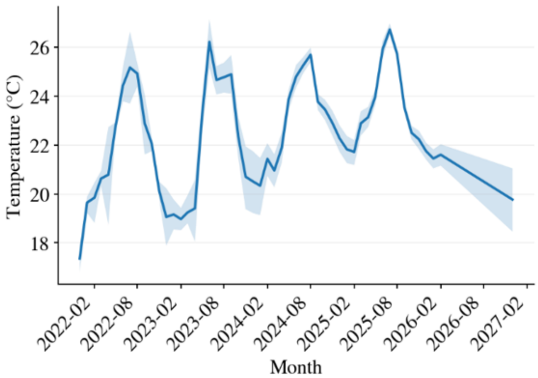
Descriptive statistics of environmental indicators and valid modelling records following linkage and eligibility filtering are summarised in Table 1. Monthly exposure indicators exhibit seasonal variation across the monitoring period. These temporal patterns are presented in Fig. 3.

Table 1. Definition and summary statistics of environmental exposure indicators

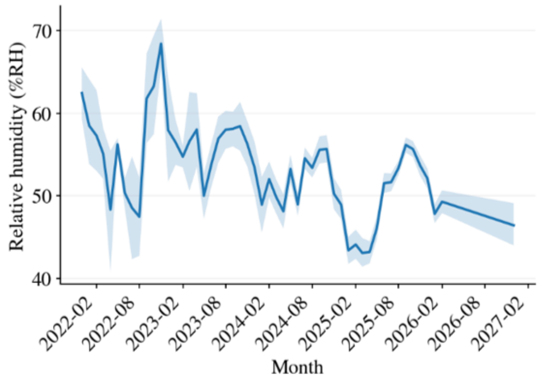
Panel A. Environmental exposure characteristics			
Variable	Mean ± SD	Range	Missing (%)
Temperature (°C)	23.18 ± 3.29	7.89–31.98	0
Relative humidity (%)	50.58 ± 10.34	18.95–82.89	0
CO_2 (ppm)	632.78 ± 248.42	392.04–2736.22	0
Panel B. Valid modelling records by domain			
Risk domain	Valid records (n)	Missing (n)	Missing (%)
Cold home	5.333	415	7.22
Damp and mould	5.333	415	7.22
Dust mite allergens	5.333	415	7.22
Excess heat	5.333	415	7.22
Heat loss	5.333	415	7.22
Draught	5.187	561	9.76
Indoor air quality (IAQ)	5.187	561	9.76
Void	5.187	561	9.76

Reported observation counts correspond to the property–time analytical unit following sensor–risk linkage and eligibility filtering.

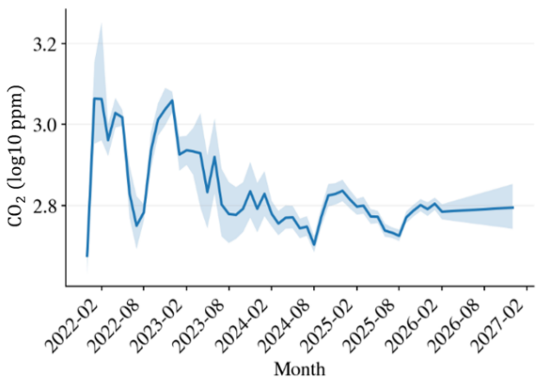
Bivariate relationships between environmental exposure indicators and operational risk domains were examined prior to regression modelling. Spearman correlation coefficients calculated at dwelling-month level are shown in Fig. 4. As expected, strong positive associations emerge between specific environmental indicators and their corresponding physical risks, notably CO_2 concentration with Indoor Air Quality ($\rho = 0.87$) and average temperature with Excess Heat ($\rho = 0.72$).



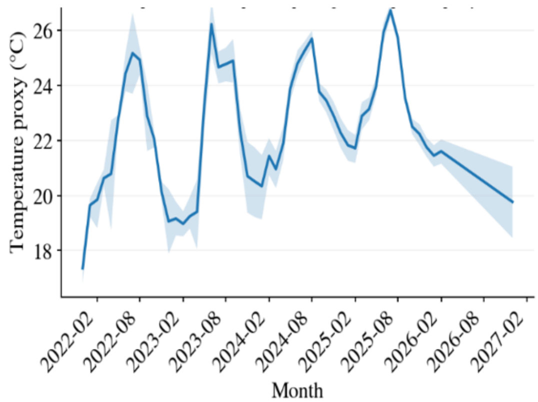
(a) Temperature



(b) Relative humidity



(c) CO_2 concentration



(d) Temperature proxy

Figure 3. Seasonal variations of environmental exposure indicators

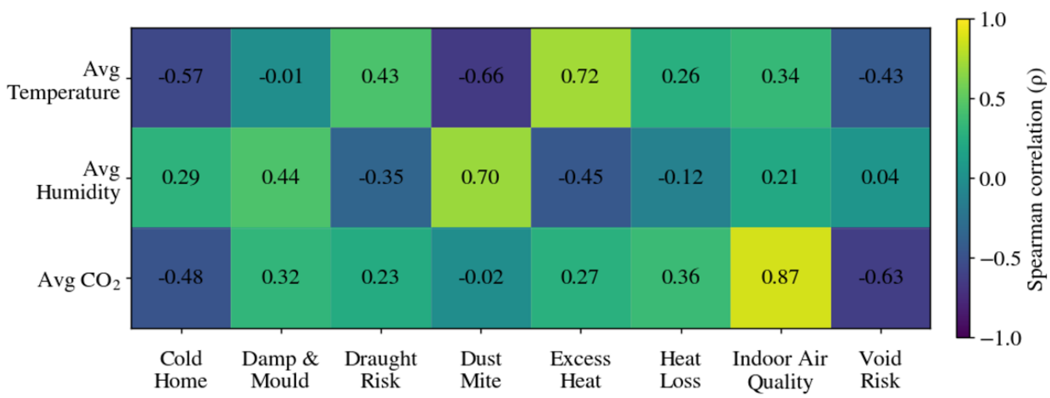


Figure 4. Spearman rank correlation coefficients (ρ) between monthly dwelling-level environmental exposure indicators and operational risk domains

3.3. Modelling and validation protocol

Regression models were estimated separately for each operational risk domain using dwelling-month observations as analytical units. Domain-specific risk status was treated as the dependent variable, and aggregated environmental exposure indicators were included as explanatory variables.

Ridge regression was implemented as a regularised linear baseline to mitigate coefficient instability under correlated predictors [14]. Random Forest regression was applied to evaluate whether non-linear relationships or feature interactions improved explanatory performance [15]. All predictors were standardised prior to model estimation.

Model evaluation employed five-fold property-grouped cross-validation. All longitudinal dwelling-level observations associated with a given property were assigned exclusively to either training or validation folds. This grouping preserves dwelling-level independence and prevents data leakage between training and validation subsets [16, 17]. Hyperparameter tuning for Random Forest models was conducted within training folds only.

Model performance was assessed using the coefficient of determination (R^2) and mean absolute error (MAE). Under grouped validation, R^2 represents the proportion of between-dwelling variance explained in properties not used for model estimation. Negative R^2 values indicate that the fitted model performs below an unconditional mean predictor under cross-dwelling evaluation. Such results reflect limited between-dwelling information contained in the environmental exposure indicators for specific domains rather than instability in the estimation procedure.

4. Results

4.1. Cross-dwelling explanatory performance

Table 2 reports five-fold property-grouped cross-validation results for each operational risk domain. Entire dwellings were held out from model estimation and used only for validation, so reported metrics reflect cross-dwelling evaluation on properties not used during training.

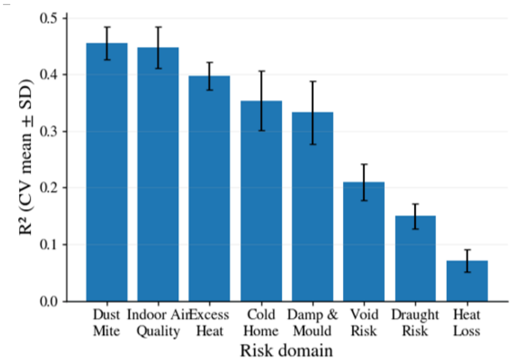
Table 2. Cross-validated explanatory performance of exposure indicators across risk domains

Domain	Model	n	Groups	R^2 (mean ± SD)	MAE (mean ± SD)
Cold home	Ridge	4974	79	−0.042 ± 0.626	11.63 ± 3.67
	RF	4974	79	−0.09 ± 0.692	10.04 ± 3.96
Damp and mould	Ridge	5053	81	−0.093 ± 0.444	12.33 ± 2.35
	RF	5053	81	−0.372 ± 1.086	12.15 ± 2.83
Draught	Ridge	5053	81	−0.05 ± 0.07	18.85 ± 1.45
	RF	5053	81	−0.258 ± 0.248	20.24 ± 0.56
Dust mite allergens	Ridge	4939	78	0.164 ± 0.33	10.62 ± 1.93
	RF	4939	78	0.124 ± 0.345	10.89 ± 1.36
Excess heat	Ridge	5051	80	0.191 ± 0.106	22.27 ± 3.89
	RF	5051	80	0.165 ± 0.152	21.48 ± 2.2
Heat loss	Ridge	5082	82	−0.052 ± 0.048	20.8 ± 3.13
	RF	5082	82	−0.295 ± 0.169	22.87 ± 2.96
Indoor air quality (IAQ)	Ridge	5051	80	0.273 ± 0.211	12.86 ± 2.51
	RF	5051	80	0.156 ± 0.225	13.65 ± 3.13
Void	Ridge	5051	80	−0.07 ± 0.233	17.93 ± 4.46
	RF	5051	80	−0.164 ± 0.401	14.54 ± 4.98

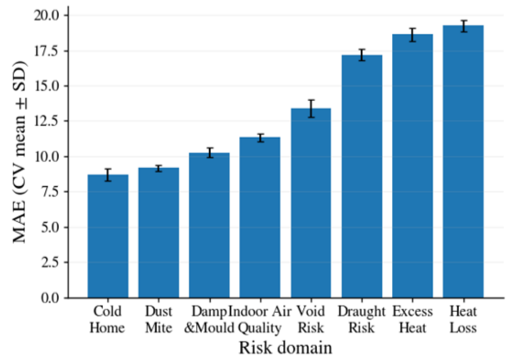
Indoor Air Quality and Excess Heat retain positive R^2 values under grouped validation [18]. This indicates that aggregated environmental exposure indicators explain a measurable proportion of between-dwelling variance for these domains. Dust mite allergens also show positive explanatory capacity, with lower R^2 values relative to ventilation- and temperature-driven risks.

Fabric-related domains, including Cold Home, Damp and Mould, Heat Loss, and Draught, yield near-zero or negative R^2 values under grouped evaluation. Void status also yields near-zero or negative R^2 under grouped validation. Negative R^2 indicates performance below that of an unconditional mean predictor when evaluated on unseen dwellings [19, 20]. Across these domains, differences between Ridge and Random Forest models remain limited [21–23], and non-linear model flexibility does not translate into consistent improvements in cross-dwelling explanatory capacity.

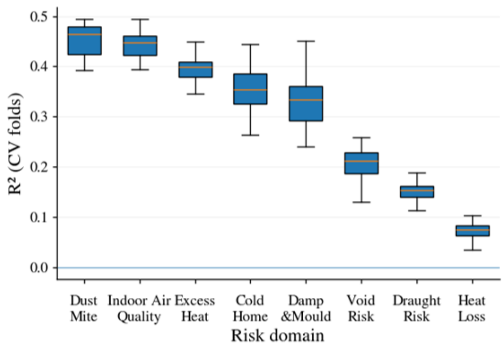
Cross-validated performance metrics and fold-level dispersion patterns are shown in Fig. 5(a)–(c). Diagnostic residual plots are presented in Fig. 5(d). Ventilation- and heat-related domains exhibit narrower interquartile ranges and more stable medians across folds. Fabric-related domains show wider dispersion, with longer whiskers and more frequent outlying fold results, consistent with higher variability in cross-dwelling performance under grouped validation [24].



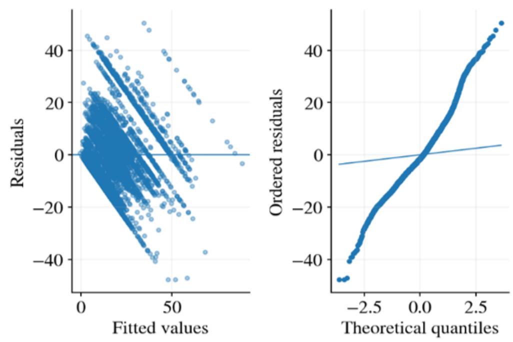
(a) Mean cross-validated R^2



(b) Mean cross-validated MAE



(c) Fold-level R^2 dispersion



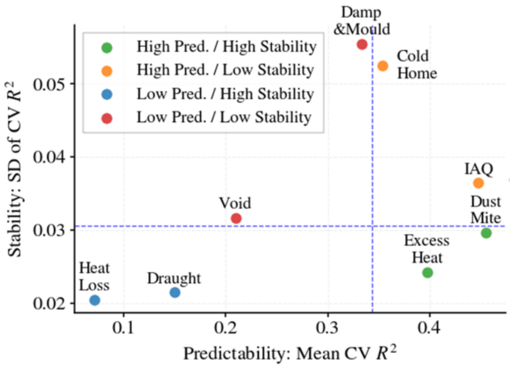
(d) Residual diagnostics

Figure 5. Cross-dwelling predictive performance and model diagnostics

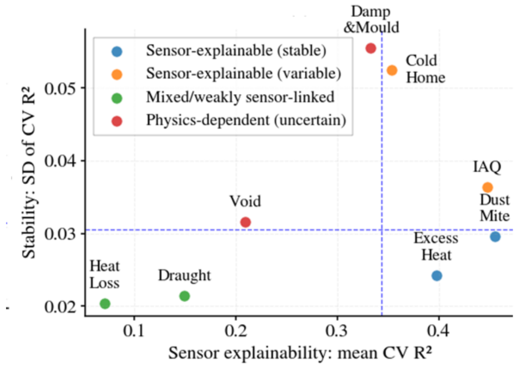
4.2. *Portfolio-level decision support implications*

Grouped cross-dwelling validation distinguishes the practical role of environmental monitoring across operational risk domains. Fig. 6 presents mean grouped CV R^2 alongside the standard deviation of CV R^2 across folds, allowing joint consideration of explanatory capacity and stability.

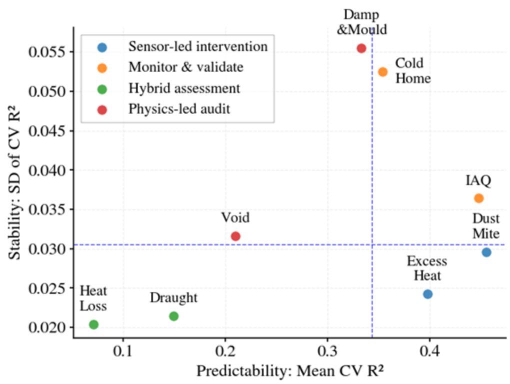
Indoor Air Quality and Dust Mite Allergens combine positive cross-dwelling R^2 with comparatively low fold dispersion. Aggregated exposure indicators differentiate dwellings consistently for these domains. Carbon dioxide concentration relates to ventilation adequacy under occupancy conditions [5, 12, 25], and the resulting exposure signal supports identification of dwellings with persistently elevated indoor pollutant levels.



(a) Predictability vs. stability



(b) Sensor explainability



(c) Operational strategies

Figure 6. Portfolio-level decision support and operational strategies

Excess Heat retains positive explanatory performance with greater fold-level variability. Temperature exposure corresponds to established overheating assessment approaches [6, 13], but cross-dwelling differentiation is less uniform. Monitoring data in this domain may guide prioritisation while requiring confirmatory inspection where necessary.

Cold Home and Damp & Mould yield negative mean R^2 values and elevated fold-level dispersion, indicating limited cross-dwelling explanatory capacity. The relationship between exposure indicators and recorded risks fluctuates among individual properties. Monitoring data in these domains offer contextual evidence rather than reliable predictive discrimination.

Heat Loss and Draught occupy the low-explainability region of Fig. 6. Under grouped evaluation, environmental indicators do not account for between-dwelling variance beyond a mean baseline. These risks depend on envelope transmittance, air permeability, and construction detailing, which are not directly captured through monthly indoor air measurements. Void status reflects operational occupancy conditions rather than exposure-related differences. For these domains, direct physical assessment remains necessary within maintenance and retrofit planning [25, 26].

Fig. 6(c) translates these statistical patterns into operational strategies. Risk domains exhibiting positive and stable cross-dwelling R^2 are well-suited for sensor-led screening. Those showing variable predictability necessitate a hybrid approach of monitoring supplemented by targeted inspections. Categories yielding near-zero or negative explanatory power remain strictly dependent on direct physical audits.

5. Discussions and conclusions

This paper analysed a linked dataset combining five years of high-frequency indoor environmental monitoring with operational housing risk records in a UK-based social housing portfolio. Monthly dwelling-level exposure indicators were evaluated against domain-specific risk status adopting Ridge and Random Forest regression. Property-grouped cross-validation was applied to assess cross-dwelling explanatory capacity and to prevent information leakage.

Environmental exposure indicators account for part of the between-dwelling variance in Indoor Air Quality and Excess Heat. In these domains, monitoring data allow comparison of relative exposure conditions and can inform inspection prioritisation at the portfolio level. For fabric-related domains, including Heat Loss and Draught, grouped evaluation yields near-zero cross-dwelling explanatory capacity. These risks depend on envelope properties not directly reflected in aggregated indoor air measurements. Monitoring data alone can not replace physical inspections in these cases.

Model performance differs under grouped and random validation. Random partitioning can retain dwelling-specific information in both training and test subsets, which inflates apparent performance. Grouped evaluation removes this effect and provides a more appropriate estimate of cross-dwelling applicability.

Hybrid operational risk records introduce misclassification label noise. False negatives mask unrecorded structural faults, and false positives reflect subjective discomfort rather than verified fabric failures. This subjective variance degrades cross-dwelling explanatory capacity in fabric-dependent domains. Property-grouped cross-validation directly mitigates this limitation. Isolating entire properties across data folds blocks dwelling-specific reporting bias from leaking into evaluation metrics. This strategy yields an objective assessment of environmental exposure indicators.

Monthly aggregation reduces short-term variability in exposure conditions, and operational risk records may vary in reporting consistency. The monitored dwellings represent a subset of the wider stock. Reported R^2 values reflect explanatory performance under the specific aggregation, linkage, and validation framework.

Further work will examine alternative temporal aggregation, inclusive of structural attributes such as construction age and retrofit history, and integration of environmental and fabric data. Application of the same grouped validation approach to other housing portfolios would clarify the stability of cross-dwelling explanatory performance across different stock compositions.

Acknowledgment

The authors gratefully acknowledge the support and facilities provided by Brunel University London and the University of Transport Technology (UTT) in conducting this research.

References

[1] Organization, W. H. (2009). *WHO guidelines for indoor air quality: dampness and mould*. Geneva, Switzerland: World Health Organization.

[2] Mendell, M. J., Mirer, A. G., Cheung, K., Tong, M., Douwes, J. (2011). Respiratory and Allergic Health Effects of Dampness, Mold, and Dampness-Related Agents: A Review of the Epidemiologic Evidence. *Environmental Health Perspectives*, 119(6):748–756.

[3] Fisk, W. J., Lei-Gomez, Q., Mendell, M. J. (2007). Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. *Indoor Air*, 17(4):284–296.

[4] Li, S., Xu, L. D., Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, 17(2): 243–259.

[5] Mélois, A., Rios, J., Legrée, M., Depoorter, J., Moury, A. M., Rebières, S., Guyot, G. (2025). Performance of two humidity-controlled ventilation systems after 15 years in service in French social housing buildings. *Building and Environment*, 288:113908.

[6] Taylor, J., Davies, M., Mavrogianni, A., Chalabi, Z., Biddulph, P., Oikonomou, E., Das, P., Jones, B. (2014). The relative importance of input weather data for indoor overheating risk assessment in dwellings. *Building and Environment*, 76:81–91.

[7] Zhao, H., Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6):3586–3592.

[8] Amasyali, K., El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205.

[9] Bourdeau, M., Zhai, X. Q., Nefzaoui, E., Guo, X., Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48: 101533.

[10] Too, E., Weaver, P. (2014). The management of project management: A conceptual framework for project governance. *International Journal of Project Management*, 32(8):1382–1394.

[11] ANSI/ASHRAE Standard 55-2023 (2023). *Thermal Environmental Conditions for Human Occupancy*. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

[12] ANSI/ASHRAE Standard 62.2-2022 (2022). *Ventilation and Acceptable Indoor Air Quality in Residential Buildings*. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

[13] CIBSE TM59 (2017). *Design methodology for the assessment of overheating risk in homes*. London, UK: Chartered Institution of Building Services Engineers.

[14] Hoerl, A. E., Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.

[15] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

[16] CIBSE TM52 (2013). *The limits of thermal comfort: avoiding overheating in European buildings*. London, UK: Chartered Institution of Building Services Engineers.

[17] Gelman, A., Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

[18] Kuhn, M., Johnson, K. (2013). *Applied Predictive Modeling*. Springer, New York, NY.

[19] Hastie, T., Tibshirani, R., Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2 edition, Springer, New York, NY.

[20] Niculescu-Mizil, A., Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 625–632.

[21] Fisher, A., Rudin, C., Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20(177):1–81.

[22] Vaicenavicius, J., Widmann, D., Andersson, C., Lindsten, F., Roll, J., Schön, T. B. (2019). Evaluating model calibration in classification. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 3459–3467.

[23] Guo, C., Pleiss, G., Sun, Y., Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 1321–1330.

[24] Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N. D. (2009). *Dataset Shift in Machine Learning*. MIT Press, Cambridge, MA.

[25] Bellido-Lopez, F. J., Sanz-Bobi, M. A., Muñoz, A., Gonzalez-Calvo, D., Alvarez-Tejedor, T. (2025). Maintenance-Aware Risk Curves: Correcting Degradation Models with Intervention Effectiveness. *Applied Sciences*, 15(20):10998.

[26] Gómez Fernández, J. F., Fernández, E. C., Márquez, A. C. (2025). A Structured Data Model for Asset Health Index Integration in Digital Twins of Energy Converters. *Energies*, 18(12):3148.