CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Assessing the impact of energy retrofits on indoor climate conditions using mixed effects models: methodology and R implementation
Authors	Mishra, Asit Kumar
Publication date	2026-05-29
Original Citation	Mishra, A K 2026, 'Assessing the impact of energy retrofits on indoor climate conditions using mixed effects models: methodology and R implementation', Atmosphere, vol. 17, no. 6, 560, pp. 1-26. https://doi.org/10.3390/atmos17060560
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.scopus.com/pages/publications/105042724562 - 10.3390/atmos17060560
Rights	© 2026, the Author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-17 04:24:25
Item downloaded from	https://hdl.handle.net/10468/19256

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

atmosphere

MDPI

Article

Assessing the Impact of Energy Retrofits on Indoor Climate Conditions Using Mixed Effects Models: Methodology and R Implementation

Asit Kumar Mishra [1,2]

[1] School of Public Health, University College Cork, T12 XF62 Cork, Ireland; akumarmishra@ucc.ie
[2] DTU Sustain, Technical University of Denmark, 2800 Kogens Lyngby, Denmark

Abstract

Energy retrofit interventions have become increasingly critical as building sectors worldwide pursue decarbonization targets and improved energy efficiency. However, establishing robust causal inference about retrofit impacts on indoor climate conditions remains challenging due to confounding variables including outdoor climate fluctuations and occupant behavior. This paper presents a methodological framework for analyzing pre- and post-retrofit indoor climate data using linear mixed effects (LME) models, which explicitly account for building-level variability while controlling for environmental and behavioral factors. The approach is demonstrated using a case study analyzing partial pressure of water vapor in Irish residential homes before and after energy retrofit interventions. The analysis incorporates standardized coefficients to assess relative importance of predictive factors and employs model parsimony through stepwise removal of non-significant terms. Complete R code is provided to facilitate adaptation by other researchers. Our results demonstrate that LME models provide unbiased estimates of retrofit effects while avoiding aggregation bias that plague simpler analyses. This paper serves as both methodological reference and practical guide for practitioners seeking to rigorously evaluate building retrofit effectiveness across diverse indoor climate parameters.

Keywords: energy retrofit; indoor climate; mixed effects models; building performance; open-source coding; statistical models

1. Introduction

Growing concern over the building sector energy consumption and its contribution to greenhouse gas emissions has driven substantial investment in energy retrofit programs across the EU [1], with explicit focus in certain cases on the residential sector [2]. These interventions target thermal envelope improvements through enhanced insulation, window-door upgrades, airtightness measures, and replacing fossil fuel-based heating systems, with the primary objective of reducing heating energy demand [3]. These measures, in addition to affecting the energy use of a building, also affect the indoor environment and consequently the health and wellbeing of occupants [4,5].

Despite this massive policy and financial commitment to retrofitting [1]—from the EU's Renovation Wave, which aims to renovate 35 million buildings by 2030 [6], to national schemes targeting hundreds of thousands of homes [2]—empirical evidence on how retrofits affect indoor environmental quality and health remains uneven and context-dependent [4,7]. While many studies have reported improved thermal comfort and reduced

Check for updates

Academic Editor: Ferdinando Salata

Received: 2 April 2026
Revised: 22 May 2026
Accepted: 26 May 2026
Published: 29 May 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

dampness following envelope upgrades [7], indoor air quality outcomes are mixed, with some investigations finding increased concentrations of indoor pollutants, elevated radon levels, or greater summertime overheating in highly airtight homes that lack adequate ventilation [7–9]. This inconsistency in findings means that retrofit programs cannot be evaluated on energy performance alone. To inform future climate action plans and retrofit support schemes, there is a pressing need for robust, building-scale evidence on how specific retrofit strategies influence indoor humidity, temperature, and air quality under real operating conditions [7,10]. In practice, this requires methodological approaches that can disentangle the effect of the retrofit itself from background variability in weather, building stock, and occupant behavior [11].

1.1. Confounding in Retrofit Evaluation

Evaluating the impact of energy retrofits on indoor environmental conditions is complicated by multiple, overlapping sources of confounding. Indoor temperature and humidity respond not only to the retrofit measures themselves, but also to outdoor weather conditions, which drive heat and moisture transfer through the building envelope [7,12]. Occupant behavior—including window/door opening, thermostat set-points, and moisture-generating activities such as cooking, showering, and drying clothes—can substantially modify both indoor conditions and energy use [7,10]. These practices are, in turn, affected by the season, household composition, and socio-economic demographics [10]. Building characteristics such as thermal mass, ventilation system type, airtightness level, age, and construction standard further shape how a dwelling buffers or amplifies outdoor climate fluctuations [7]. Superimposed on these structural and behavioral factors are temporal patterns: heating versus non-heating seasons, day–night cycles and inter-annual variability in weather [13].

Traditional evaluation approaches, such as simple pre/post comparisons or cross-sectional regressions with limited covariate control, rarely account for this full set of confounders and are therefore prone to biased retrofit effect estimates [7,14]. For example, if pre-retrofit monitoring happens to coincide with an unusually dry winter while post-retrofit data are collected during a wetter season, before/after comparisons of indoor humidity may wrongly attribute climate-driven changes to the retrofit. Similarly, differences in occupancy patterns or window-opening behavior between homes that do and do not receive retrofits can be misinterpreted as retrofit effects if not explicitly modelled [10]. Robust retrofit evaluation therefore requires statistical methods that can separate intervention effects from background variability arising from weather, behavior, building stock and time [11].

1.2. Mixed Effects Modelling Framework

Linear mixed effects (LME) models provide a principled framework for addressing these challenges [11,15]. By combining fixed effects and random effects, LMEs allow analysts to adjust explicitly for measured confounders, e.g., outdoor climate, season, building characteristics and proxy indicators of occupancy, while simultaneously accounting for the hierarchical structure of the data (repeated measurements nested within rooms and homes). Fixed effects capture population-average relationships between predictors and outcomes, enabling direct estimation of the retrofit effect after controlling for climate and behavioral covariates. Random intercepts and, where justified, random slopes at the building level, accommodate between-home differences in baseline indoor conditions and in sensitivity to outdoor drivers (for example, some homes being more responsive to external weather than others) [11,15].

This hierarchical formulation has several advantages over more conventional regression or aggregation-based approaches. First, it makes efficient use of all available

observations without pre-averaging by home or time-period, thereby preserving within-home variability and avoiding information loss that can distort variance estimates and *p*-values [11]. Second, mixed effects models handle unbalanced longitudinal data naturally: homes with shorter monitoring periods or incomplete pre/post records still contribute information without requiring listwise deletion, an important property for real-world field studies where drop-out and sensor gaps are common. Third, the random effects themselves provide useful summaries of building-level heterogeneity in retrofit response, which can be related back to physical characteristics in subsequent analyses.

LME methods are now standard in biomedical and social sciences for precisely this combination of confounding control and hierarchical data handling [11]. But they have been applied less consistently in building performance and indoor environmental quality research [7], where simple before/after comparisons or ordinary least squares regression still predominate.

In building performance and indoor environmental quality research, the need for such models is increasingly recognized because monitoring datasets are typically structured as repeated observations nested within rooms, homes, or buildings, and are influenced simultaneously by weather, occupancy, and building-specific characteristics. The empirical retrofit and IEQ literature remains dominated by case-specific analyses focused on outcomes such as ventilation adequacy, pollutant concentrations, thermal comfort, or post-retrofit air quality. For example, studies have shown the importance of ventilation performance and occupant exposure patterns in deep energy-retrofitted dwellings, the broader indoor air quality implications of retrofit as part of a just transition, and the mixed effects of deep energy renovations on indoor air quality and thermal comfort [8,10,16].

However, these studies do not provide a general, openly documented modelling workflow that other researchers can readily adapt for pre/post retrofit evaluation across outcomes and contexts. The contribution of the present paper is therefore not another outcome-specific retrofit analysis, but a transferable and reproducible LME framework, with accompanying open-source R code (R version 4.4.3 was used for all analysis reported in this work), for estimating retrofit effects while accounting for confounding, clustering, and building-level heterogeneity.

1.3. Objectives and Presentaiton Structure

There is thus both the need for robust retrofit evaluation and the methodological challenges that routine analytical approaches struggle to address. This paper responds to that gap by providing a complete, transferable methodological framework—grounded in LME modelling—that other researchers can apply to their own indoor climate datasets. Indoor humidity, expressed as the partial pressure of water vapor, measured in 23 Irish residential homes across pre- and post-retrofit periods, is used throughout as a worked example to ground the exposition [17,18]. The primary contribution, however, is the framework and its open-source implementation rather than the findings of that specific case study.

The work pays particular attention to confounding control, hierarchical data structure, and unbalanced longitudinal designs. Step-by-step R code covering data processing, model construction, fixed and random effect selection, model diagnostics, and result visualization have been provided. The complete analytical pipeline is demonstrated using the case study. Open-source, documented codebase that practitioners and researchers can readily adapt to their data is made freely available [19].

The paper is structured as follows. Section 2 presents the methodological framework, describing the general approach to LME model construction, selection of fixed and random effects, model parsimony criteria, and interpretation of standardized coefficients. Section 3

presents results from the partial pressure of water vapor case study. Section 4 discusses the advantages and limitations of the approach and offers guidance for adapting it to future applications.

2. Methods

2.1. Study Design and Data Structure

2.1.1. Overview of Data Collection

The case study utilizes synthetic data formed based on the data gathered from 23 Irish residential homes monitored continuously before and after energy retrofit interventions [20]. Pre-retrofit monitoring periods occurred during early 2015 (February–June), while post-retrofit periods covered September 2015 through February 2017. This timing captures both heating seasons (October–May equivalent) and non-heating seasons (June–September equivalent) in both periods, allowing seasonal effects to be analyzed.

Indoor climate measurements were collected in multiple rooms including kitchen, living room, and multiple bedrooms. External climate conditions (temperature and relative humidity) were obtained from the nearest Met Eireann weather station (Dublin Airport) [17]. From the temperature and relative humidity data, the following variables were derived [21]:

Partial pressure of water vapor (Pw, kPa): derived using Magnus formula, quantifies the absolute amount of water vapor present;

Absolute humidity (AH, kg/kg dry air): mixing ratio representing the mass of water vapor per unit mass of dry air;

Dew point temperature (DPT, °C): temperature at which air becomes saturated with moisture.

A critical choice in our analysis workflow was to employ partial pressure of water vapor (Pw) as the primary outcome variable rather than relative humidity (RH), despite RH being more commonly reported in building science literature. This choice reflects important methodological principles that all practitioners should consider when selecting their response metrics. While both metrics describe the same physical phenomenon (water vapor in air), they differ fundamentally in their temperature dependence, which has profound implications for pre/post intervention studies.

The same absolute amount of water vapor in indoor air will produce different RH readings, depending on the air temperature [21]. This temperature dependence creates a serious confounding problem in retrofit analysis. Energy retrofit interventions typically alter heating patterns and thus indoor temperature distributions. This temperature change alone will lower RH readings by 5–10 percentage points regardless of whether actual moisture conditions have changed. Simple before/after comparison of RH values becomes impossible to interpret without detailed temperature accounting.

Partial pressure of water vapor (Pw, measured in kPa), by contrast, represents the absolute pressure exerted by water molecules independent of temperature or the presence of other gases. Pw is thus temperature-independent and directly comparable across seasons and heating regimes. Pw is also highly relevant to human thermal comfort evaluation [22] (p. 55). By selecting Pw as the response variable and explicitly including external temperature as a separate fixed effect in the mixed effects model (Section 2.2), simultaneous control of temperature effects is achieved while capturing genuine moisture-related retrofit impacts.

For practitioners implementing this methodology with different outcome variables, the selection principle should be to choose metrics that are mechanistically relevant and independent of potential confounders. For example, if studying thermal retrofits, prefer metrics independent of seasonal variation; if studying ventilation interventions, prefer metrics sensitive to occupant behavior but robust to outdoor conditions.

A sample, synthetic dataset comprising of ~28,000 hourly observations (which represents about 5% of the original dataset) has been used for this demonstration. The sample was constructed by randomly selecting a stratified subset of pre-retrofit and post-retrofit periods across all homes, maintaining temporal distribution to ensure representativeness, followed by addition of small perturbations to the measured indoor climate data. The R implementation shared accommodates any continuous response variable.

2.1.2. Tidy Data Structure for Pre/Post Retrofit Studies

For analysis to be reproducible and code to be readily adaptable across datasets, it is strongly recommend to organize pre/post retrofit data according to "tidy data" principles [23]. A tidy dataset should have the following structure: each row represents a single observation (e.g., one hourly temperature/humidity measurement from one room in one home at one time point). Each column represents a single variable. The recommended data structure is provided in Table 1. Users can expand upon this based on the parameters of interest in their study.

Table 1. Recommended tidy data structure for pre/post retrofit studies.

Column	Data Type	Description	Example
HomeID	factor	Unique home identifier	H01
RoomType	factor	Living Room or Other Rooms (averaged)	Living Room
DateTime	POSIXct	Date and time of observation	1 February 2015 09:00
RetrofitStatus	factor	Pre-Retrofit/Post Retrofit	Pre-Retrofit
RoomT	numeric	Indoor air temperature (°C)	19.5
RoomRH	numeric	Indoor relative humidity (%)	52
RoomPw	numeric	Indoor partial pressure of water vapor (kPa)	0.85
RoomDPT	numeric	Indoor dew point temperature (°C)	8.2
ExternalT	numeric	External air temperature (°C)	5.2
ExternalRH	numeric	External relative humidity (%)	78
ExternalPw	numeric	External partial pressure of water vapor (kPa)	0.43
Season	factor	Heating/Non-Heating season (DEAP definition)	Heating
Month	integer	Month of year (1–12)	2
Hour	integer	Hour of day (0–23)	9
BuiltType	factor	Construction type (End Terrace, Mid Terrace, Detached)	End Terrace
BuiltYear	integer	Year of original construction	1994
Occupants	integer	Number of occupants	3
Income	numeric	Annual household income (optional). Be mindful of privacy and GDPR requirements	35,000

2.1.3. Data Hierarchy and Nesting Structure

The data possesses a hierarchical structure critical for model specification. At Level 3, the highest level, there are the homes. At Level 2, there are the measurement occasions (hourly aggregate data) and they may be further grouped by aspects such as room type or season. At Level 1, the lowest level, there are individual measurements.

This nesting structure is inherent to building performance (energy and/or indoor environment) data collection. It violates the independence assumption of ordinary least squares (OLS) regression [11,15]. Multiple measurements from the same home are necessarily correlated, and ignoring this correlation produces artificially narrow standard errors and inflated Type I error rates [11,24]. Random effects at the home level are therefore required to properly account for intra-home correlation [11,15].

2.1.4. Data Cleaning and Handling Missing Values

Rigorous handling of missing data and outliers is essential before applying mixed effects models. Real-world building datasets inevitably contain gaps resulting from sensor failures, maintenance periods, and data transmission errors.

The most defensible approach is to retain complete observations where possible and allow the mixed effects modelling framework (via maximum likelihood estimation) to handle missing data appropriately under the Missing At Random (MAR) assumption [25]. In our workflow, missing data were converted to explicit NAs early in processing. This approach is preferable to listwise deletion (which discards entire observations with any missing values) because it retains observations where only isolated variables are missing. The lme4 package (ver. 1.1.36) in R [15] implements maximum likelihood estimation that naturally accommodates missing values in outcome and predictor variables, allowing use of the full effective sample size.

Extended gaps (>1 day) warrant investigation. They may reflect systematic patterns (e.g., sensor maintenance during winter maintenance periods) that could introduce bias if not accounted for. The proportion of missing data by home and by time-period must be documented during the analysis process.

For datasets with <5% missing values in any predictor, multiple imputation is often unnecessary [26]. However, for higher missing proportions or for secondary exploratory analyses, multiple imputation can be employed if three conditions are met: (1) missing data are plausibly MAR given measured variables; (2) enough observed data exist to generate reasonable imputations; and (3) the analysis model is specified before imputation.

Imputation does introduce additional uncertainty. If the retrofit effect remains significant across multiple imputation models but inference is sensitive to imputation method choice, this should be reported as a limitation. For the retrofit studies specifically, retrofit status or temporal classifiers (which define pre/post periods) are design variables that should be complete by definition and should not require imputation.

As the case study data were generated synthetically and form a large dataset, no missingness arose except for a small percentage in the date-time column due to a date-time format issue. In such field studies, missing data is most likely to arise from sensor communication dropouts and scheduled maintenance periods. Little's MCAR test of the sample data used ($p < 0.0001$) indicated missingness was unlikely to be completely at random. Variable-specific missing proportions are reported in Appendix A.

To support reproducibility for users applying the code to other datasets, we provide an accompanying missing-data diagnostic workflow in the shared R script, including variable-wise missingness summaries and an optional assessment of whether missingness is associated with observed covariates.

2.1.5. Outlier Handling

Two classes of outliers require consideration: physiologically implausible values and genuine extreme events. Physiologically implausible values include relative humidity > 100% or <0%, absolute humidity < 0, or recorded temperatures outside the physical range of installed sensors. These should be converted to NA as they represent data transmission errors or sensor faults.

Extreme but plausible values represent genuine extreme conditions (e.g., brief periods of very high indoor humidity from shower activity, or unseasonably cold outdoor temperatures). These should not be removed as they represent valid variations in building performance under real-world conditions. The mixed effects framework, with its explicit accounting of within-home variation through random intercepts and slopes, is particularly robust to isolated extreme values because it does not assume homogeneous variance across homes.

To evaluate robustness to extreme values, a sensitivity analyses can be conducted. In this demonstration, separate models were fitted, with and without observations beyond the 5th and 95th percentile of each continuous predictor's distribution. This was done to verify that inferences about the retrofit effect (the primary coefficient of interest) remain stable. If the retrofit coefficient changes substantially when extreme observations are excluded, this suggests either genuine interactions (e.g., retrofit effectiveness differs under extreme conditions) or problematic outliers requiring deeper investigation. Any sensitivity findings need to be documented transparently.

A data quality log should be maintained documenting the following: (1) proportion of missing values by variable and time period; (2) outlier decisions and any values removed or imputed; (3) any sensor replacements, data collection issues, or recalibrations during the study period; and (4) any other data manipulations applied. This transparency enables reviewers and future researchers to assess data quality and reproduce analyses faithfully. Version control of raw and processed datasets further enhances reproducibility.

2.2. Mixed Effects Model Framework

2.2.1. General Model Specification

A linear mixed effects model can be expressed as [11,15]

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ijk} + u_0 j + u_1 j x_{ij} + \epsilon_{ij} \quad (1)$$

where

y_{ij} is the response variable (e.g., partial pressure of water vapour) for observation i in home j

β_0 is the fixed intercept (population average baseline level)

β_k are fixed slopes (population average effects of predictors)

x_{ijk} are predictor variables

u_{0j} is the random intercept for home j (deviations of home-specific baseline from population average)

u_{1j} is the random slope for home j (deviations of home-specific response to a key predictor)

ϵ_{ij} is the residual error term

Random effects are assumed to follow normal distributions: $u_{0j} \sim N(0, \sigma_u^2)$ and $u_{1j} \sim N(0, \sigma_u^2)$, with residuals $\epsilon_{ij} \sim N(0, \sigma^2)$.

2.2.2. Application to Retrofit Analysis

For retrofit impact evaluation, the model takes the specific form [10]

$$y_{ij} = \beta_0 + \beta_R \text{RetrofitStatus}_j + \beta_C \text{Climate}_{ij} + \beta_B \text{Behaviour}_{ij} + \beta_S \text{Season}_{ij} + u_{0j} + u_{1j}\text{Climate}_{ij} + \epsilon_{ij} \quad (2)$$

where

- β_R is the primary effect of interest: the impact of retrofit on the indoor climate parameter
- Climate terms (e.g., external temperature, external humidity) capture the effects of outdoor conditions that drive indoor conditions
- Behavior terms (occupancy, window opening behavior, use of additional heating systems—where information is available) account for occupant-related factors
- Season terms capture seasonal variation in climate drivers and occupant patterns
- Random intercept (u_{0j}) allows baseline indoor climate to vary by home
- Random slope (u_{1j}) for the climate predictor allows homes to differ in their sensitivity to outdoor conditions (e.g., some homes may be better thermally insulated than others)

This specification directly addresses confounding by explicitly including measured confounders as fixed effects while allowing home-level heterogeneity (often covering unmeasured and unknown confounders) through random effects.

2.3. Model Building and Selection Strategy

2.3.1. Fixed Effects Selection

Model building followed a step-down approach beginning with the most complex model that includes all theoretically relevant predictors. To start, all measured variables with clear mechanistic relationship to the outcome were included. For indoor humidity (Pw), clear drivers include external Pw (direct entry of moist outdoor air), room temperature (affects saturation vapor pressure), occupancy (moisture generation), and season (affects occupancy and heating patterns). For all fixed effects, the variance inflation factors (VIFs) were calculated to assess multicollinearity [27]. VIF over 10 suggests problematic multicollinearity while values over 5 warrant investigation. Correlated predictors would be removed, retaining the more mechanistically relevant variables to the next step.

Model building followed a structured, theory-driven approach rather than automated variable selection. All candidate predictors were pre-specified based on mechanistic relevance to indoor humidity. The retrofit term was retained in all models regardless of *p*-value (see Section 2.3.3). Nested model comparisons used likelihood ratio tests and AIC/BIC to evaluate whether the added complexity of additional predictors improved model fit. Non-significant predictors without strong mechanistic justification were considered for removal only after confirming via LRT that their exclusion did not worsen model fit.

This approach prioritizes model interpretability while maintaining explanatory power. To facilitate comparison of relative importance across predictors with different scales, standardized coefficients were calculated using the sjPlot package (version 2.8.17) in R [28].

2.3.2. Random Effects Specification

Random effect specification involves decisions about random intercepts and random slopes. In analysis of retrofit data, it is needed to include the random intercepts at the grouping level (home). This allows baseline levels to differ across buildings. Given the acknowledged heterogeneity in building characteristics, this is an essential requirement [11].

Random slopes may be included for one or more of the key predictors. This is recommended when a predictor shows substantial variation across the grouping level (e.g., homes differ in how they respond to external conditions). In retrofit analysis, random

slopes for climate variables (e.g., external temperature) are often justified, as they represent building-level differences in thermal sensitivity.

Including random slopes increases the model complexity but if model comparison (via likelihood ratio tests [29]) indicates improved fit, then the more complex model is preferred.

Systematic random-effects selection was carried out to identify the most appropriate hierarchical structure for the retrofit data. A random-intercept-only model was first fitted to allow baseline indoor vapor pressure to vary across homes. This was then extended by adding a random slope for Retrofit Status, reflecting the expectation that homes may respond differently to the retrofit intervention depending on their envelope characteristics, ventilation configuration, and occupant behavior. A further random slope for external humidity was then considered. Homes may also differ in their sensitivity to outdoor moisture conditions due to differences in airtightness, thermal mass, and moisture buffering capacity.

Model comparison using likelihood ratio tests and information criteria (Appendix A) showed that each added random slope improved fit substantially. The random-intercept-plus-two-slope structure provided the best balance of statistical fit and interpretability. The selected model therefore included random intercepts and random slopes for RetrofitStatus and ExternalPw at the home level. This captured both heterogeneity in retrofit response and heterogeneity in coupling to outdoor climatic conditions.

2.3.3. Retaining the Retrofit Effect

A critical principle in retrofit analysis is that the primary effect of interest—the retrofit impact—must be retained in the model regardless of statistical significance [30]. This principle reflects the hypothesis-driven nature of such research. In this situation, retrofit impact is the research question, not a secondary inquiry. There is a risk of Type II error from premature removal of small but meaningful retrofit effects. Even when the retrofit impacts are not statistically significant, it is important to quantify the retrofit impacts, which may be of practical importance [31].

2.4. Diagnostics and Model Validation

2.4.1. Checking Model Assumptions

Before interpreting results, model assumptions must be verified. Normality of residuals were verified using Q-Q plots that compare observed residuals to normal quantiles. Substantial deviations would have suggested the need for transforming the data or alternative models. Where the data are large (as in the demonstrated case), formal tests (Shapiro–Wilk, Anderson–Darling) may be overly sensitive [32] and hence are not recommended.

Residual plots (predicted vs. residual values) were used to check the homogeneity of variance. Patterns such as funnel shapes indicate non-constant variance (heteroscedasticity). If such patterns were observed, variance-stabilizing transformations or weighted regression approaches would have to be considered. Residuals were examined for temporal autocorrelation. This is particularly important for time-series data, as in our case hourly logged data. Autocorrelation functions (ACF plots) and Durbin–Watson tests [33] can be used to identify this issue. If an issue if detected, autoregressive or GEE models [34] may be preferable to the linear models. Q-Q plots of random effects estimates were used to check normality assumptions. In this case, minor deviations are acceptable; serious violations may indicate model misspecification.

Residual autocorrelation was further assessed using autocorrelation and partial autocorrelation plots of model residuals (Appendix A), together with the Durbin–Watson statistic. Although the primary mixed-effects specification accounted for clustering at the

home level, the residual diagnostics indicated that some short-range temporal dependence remained in the hourly data. As a sensitivity analysis, an alternative model incorporating an AR(1) correlation structure within homes was fitted and compared with the original specification using Akaike's information criterion.

2.4.2. Model Comparison

Model selection proceeded using three complementary approaches. For nested models—where one model is a restricted version of another—likelihood ratio tests (LRTs) were used, in which the test statistic $-2(\ell_1 - \ell_2)$ follows a χ^2 distribution with degrees of freedom equal to the difference in the number of estimated parameters between the two models. This provides a formal significance test for the additional complexity introduced by a more elaborate random effects structure. For example, when evaluating whether a random slope improves upon a random-intercept-only specification, this is the suitable approach.

Where models are non-nested and cannot be compared via LRTs, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used [35,36]. Both reward goodness of fit while penalizing model complexity, with lower values indicating a preferable model. AIC applies a lighter complexity penalty and tends to favor models with greater explanatory flexibility. BIC's stronger penalty for the number of parameters makes it more conservative and is generally preferred when parsimony is the primary objective.

To quantify the variance explained by the model, marginal and conditional pseudo-R^2 values were calculated [28,37]. The marginal R^2 reflects the proportion of variance attributable to the fixed effects alone. The conditional R^2 captures variance explained by both fixed and random effects combined. Together, these two quantities allow assessment of how much of the outcome variability is accounted for by the population-level predictors versus building-level heterogeneity.

2.5. Interpretation of Standardized Coefficients

Standardizing predictors prior to model fitting places all fixed effect coefficients on a common scale, expressed in units of standard deviations of the respective predictor. This facilitates direct comparison of effect sizes across variables [38] that would otherwise be incommensurable due to differences in units and measurement scale: outdoor temperature in degrees Celsius, for example, cannot be compared directly to a binary retrofit indicator without standardization. In this demonstration, the standardized coefficients of predictors were calculated using the sjPlot package [28].

A standardized coefficient $\hat{\beta}_k$ is interpreted as: *"a one standard deviation increase in predictor k is associated with a $\hat{\beta}_k$ unit change in the outcome, holding all other predictors constant."* The relative magnitudes of standardized coefficients therefore indicate which predictors contribute most to explaining variability in the indoor climate outcome, independent of their natural scales.

For the retrofit indicator, which is coded as a binary variable (0 = pre-retrofit, 1 = post-retrofit), the standardized coefficient captures the estimated population-average shift in the outcome attributable to the retrofit after adjusting for all covariates. In the context of indoor humidity expressed as partial pressure of water vapor, a negative coefficient indicates drier indoor conditions post-retrofit, while a positive coefficient indicates increased humidity—a distinction with direct practical relevance for assessing moisture-related risks in retrofitted dwellings.

3. Results

To illustrate the methodological workflow described in Section 2, this section presents results from a case study. As discussed, this case study analyzes synthetic, representative data of 23 Irish residential homes. This section should be understood as a worked example demonstrating the analytical steps, from raw data characteristics through model specification, diagnostics, and interpretation. Researchers applying this methodology to different indoor climate outcomes (temperature, air quality), in different climates, or different intervention types can follow an identical workflow, applied to their parameter of interest.

3.1. Sample Characteristics

The case study dataset comprised hourly-aggregated measurements from 23 Irish residential homes (sample size: 28,822 observations total). The homes included a mix of building types (end-terrace, mid-terrace, semi-detached) constructed in 1994 or 2000. Pre-retrofit monitoring periods encompassed February–June 2015, while post-retrofit periods extended from September 2015 through February 2017.

Table 2 presents summary statistics for key variables, grouped by retrofit status. Data have been aggregated into two room types, living room and other rooms, based on the distinction used in the Irish Dwelling Energy Assessment Procedure (DEAP) [18]. Pre-retrofit indoor average partial pressure of water vapor (Pw) ranged from 0.31–0.98 kPa (mean 0.62 $\pm$ 0.18 kPa), with slightly higher values during non-heating seasons (mean 0.68 $\pm$ 0.19 kPa) compared to heating seasons (mean 0.57 $\pm$ 0.16 kPa). Post-retrofit values showed comparable ranges and distributions, though with slightly more humid conditions.

Table 2. Summary statistics of key predictors, grouped by retrofit status.

Variable	Level	Pre-Retrofit	Post-Retrofit
n (count)		7018	21,804
Room Type (% of readings)	Living Room	3509 (50.0)	10,902 (50.0)
	Other Rooms	3509 (50.0)	10,902 (50.0)
Season (% of readings)	Heating	5356 (76.3)	14,798 (67.9)
	Non-Heating	1662 (23.7)	7006 (32.1)
Built Type (% of readings)	End Terrace	3988 (56.8)	11,484 (52.7)
	Mid Terrace	2126 (30.3)	7966 (36.5)
	Semi Detached	904 (12.9)	2354 (10.8)
Built Year (% of readings)	1994	3502 (49.9)	11,256 (51.6)
	2000	3516 (50.1)	10,548 (48.4)
External Temperature—mean (SD)		8.9 (4.6)	10.0 (4.8)
External RH—mean (SD)		79 (13)	82 (11)
External Pw—mean (SD)		0.91 (0.24)	1.03 (0.30)
Room Temperature—mean (SD)		19.7 (2.3)	20.6 (2.1)
Room RH—mean (SD)		51 (8)	56 (8)
Room Pw—mean (SD)		1.16 (0.20)	1.36 (0.26)
Room DPT—mean (SD)		9.0 (2.6)	11.3 (2.9)

3.2. Mixed Effects Model

Model selection proceeded in two stages, using likelihood ratio tests (LRTs) to compare nested model pairs and AIC/BIC as supporting criteria. In Stage 1, an expanded model was first fitted, augmenting the candidate fixed effects (Equation (2)) with built type, construction year, and household income. Comparison with the more parsimonious model, without these added predictors, showed that both AIC (−44,380 vs. −44,426) and BIC (−44,223 vs. −44,302) favored the simpler model.

The expanded model returned a lower log-likelihood despite the additional parameters (logLik = 22,209 vs. 22,228; $\chi^2(4) = 0$, $p = 1$). This is consistent with a convergence failure likely attributable to the varying scales of the predictors and the limited number of homes (n = 23) relative to the added building-level predictors. The expanded model was discarded in favor of the simpler model.

A reduced model was then fitted by removing Season and Occupants from the fixed effects. The LRT strongly favored the model retaining these predictors ($\chi^2(2) = 16.34$, $p < 0.001$), and AIC confirmed this (−44,426 vs. −44,413), while BIC was marginally closer between the two models (−44,302 vs. −44,306). Given the LRT result and the theoretical basis for including seasonal patterns and occupancy in indoor humidity modelling, the model was not further simplified.

The selected final model was therefore as follows:

RoomPw ~ RetrofitStatus + ExternalPw + ExternalT + RoomT + RoomType
+ Season + Occupants
+ (1 + RetrofitStatus + ExternalPw | HomeID)

AIC = −44,426, Marginal R^2/Conditional R^2 → 0.75/0.82

The random effects structure allows each home to have its own baseline humidity level (random intercept) as well as home-specific responses to retrofit status and external vapor pressure (random slopes), capturing building-level heterogeneity in both pre/post retrofit shift and sensitivity to outdoor moisture conditions.

3.3. Relative Importance of Predictors

The standardized coefficients and their 95% confidence intervals are presented in Table 3. External vapor pressure was the strongest predictor of indoor humidity ($\hat{\beta} = 0.63$, 95% CI [0.59, 0.66]). Building envelopes, even post-retrofit, did not fully decouple indoor from outdoor humidity.

Table 3. Details for the predictors of the final selected model.

	Indoor Pw (Pa)					
Predictors	**Estimates**	**Std. Beta**	**CI**	**Standardized CI**	***p***	**Std. *p***
(Intercept)	−0.46	−0.48	−0.55–−0.38	−0.63–−0.34	<0.001	<0.001
Retrofit Status Post Retrofit	0.09	0.35	0.07–0.11	0.25–0.44	<0.001	<0.001
External Pw	0.56	0.63	0.53–0.59	0.59–0.66	<0.001	<0.001
External temperature	−0.01	−0.18	−0.01–−0.01	−0.19–−0.17	<0.001	<0.001
Room temperature	0.06	0.46	0.05–0.06	0.46–0.47	<0.001	<0.001
Room Type Other Rooms	0.11	0.41	0.10–0.11	0.40–0.42	<0.001	<0.001
Season Non-Heating	0.01	0.04	0.01–0.01	0.03–0.06	<0.001	<0.001
Occupants	0.01	0.08	−0.00–0.03	−0.00–0.16	0.051	0.051

Room temperature was the second strongest predictor ($\hat{\beta}$ = 0.46, 95% CI [0.46, 0.47]), with higher indoor temperatures associated with greater vapor pressure, likely reflecting combined effects of occupant-driven moisture generation and the temperature dependence of absolute humidity. External temperature had a moderate negative effect ($\hat{\beta}$ = −0.18, 95% CI [−0.19, −0.17]), consistent with cold outdoor air having lower absolute moisture content.

Room type showed a notable positive effect ($\hat{\beta}$ = 0.41, 95% CI [0.40, 0.42]), indicating that Other rooms carry systematically higher vapor pressure than Living rooms. This likely reflects differences in moisture source proximity (e.g., kitchens, bathrooms). The effect of season was small but statistically significant ($\hat{\beta}$ = 0.04, 95% CI [0.03, 0.06]), with slightly elevated vapor pressure during the non-heating season. Occupant count had the smallest effect among retained predictors ($\hat{\beta}$ = 0.08, 95% CI [0.00, 0.16]), with a p value just bordering upon significance, 0.051.

Importantly, the retrofit effect was positive and clearly significant ($\hat{\beta}$ = 0.35, 95% CI [0.25, 0.44]). This indicated higher indoor vapor pressure in the post-retrofit period relative to pre-retrofit, after adjusting for all covariates.

3.4. Model Specification and Diagnostics

Variance inflation factors for fixed effects were all < 1.5, indicating negligible multicollinearity—Table 4.

Table 4. VIF value and 95% confidence interval for fixed effect coefficients.

Term	VIF [95% CI]
RetrofitStatus	1.01 [1.00, 1.03]
ExternalPw	1.09 [1.08, 1.10]
ExternalT	1.15 [1.13, 1.16]
RoomT	1.14 [1.12, 1.15]
RoomType	1.00 [1.00, 1.45]
Season	1.14 [1.13, 1.16]
Occupants	1.00 [1.00, ~]

The residual diagnostics plot has been included in Appendix A (Figure A1). The Q-Q plot showed minor deviations at the tails but overall adherence to normality. Residuals vs. fitted values plot showed relatively even scatter around zero, though slight heteroscedasticity at extreme predicted values. The Q-Q plot of the random intercepts confirmed approximate normality. The Durbin–Watson statistic of 1.45 indicated minimal temporal autocorrelation despite hourly aggregation.

Diagnostic plots (Appendix A) for the original mixed-effects model showed positive autocorrelation at short lags in the residual auto correlation function (ACF), with a gradual decay in the partial autocorrelation function (PACF), consistent with low-order temporal dependence. Refitting the model with an AR(1) correlation structure yielded $\phi = 0.109$ and improved the model fit, with AIC decreasing from −44,426 to −44,763, suggesting that modest hour-to-hour dependence remained after adjustment for fixed and random effects. For this worked example, the simpler LME specification was retained to prioritize interpretability and transferability.

For the selected final model, marginal R^2 (variance explained by fixed effects) was 0.75 and conditional R^2 (variance explained by fixed and random effects) was 0.82. These values indicate that the model captures substantial variation in indoor humidity, with building-level heterogeneity accounting for an additional 7% of variance.

As the expanded model (Section 3.2, Stage 1 model testing) had a convergence gradient warning, a convergence stability check was run for the coefficients of the final model against a stable optimizer. The results showed minimal difference between the coefficients (Figure 1) indicating that the next step could be taken, despite the warning.

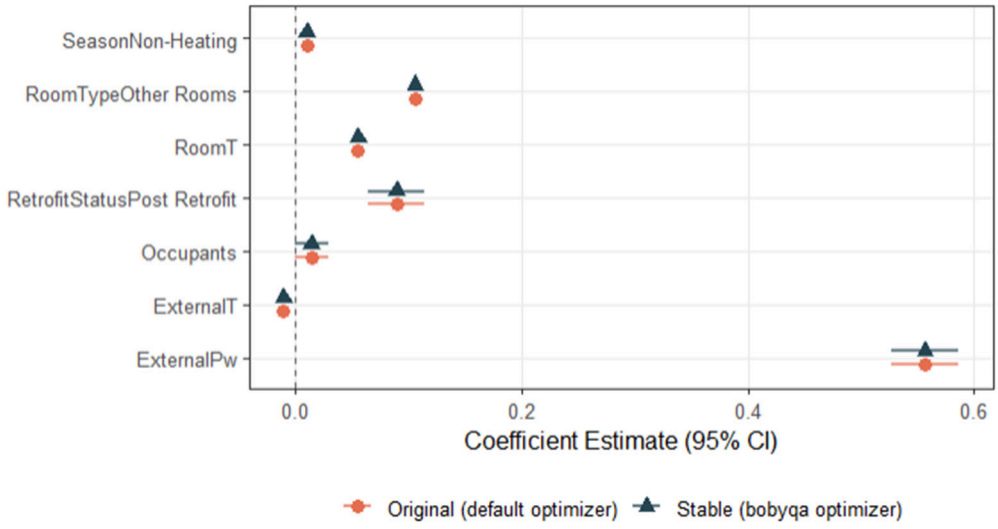

Figure 1. Convergence stability checks for model coefficients—default optimizer compared with Bobyqa optimizer results.

To assess sensitivity of the model to extreme observations, all observations falling outside the 5th–95th percentile range of each continuous predictor (External Pw, External T, Room T, Room Pw) were excluded, retaining 20,820 of the original 28,822 observations (72.2%). The final model was re-fitted on this trimmed dataset and fixed effect coefficients compared with those estimated from the full sample, referred to as the "Original" results in Table 5.

Table 5. Fixed effect coefficients from the full sample versus the 5th–95th percentile trimmed sample (72.2% of observations retained).

Coefficient	Original	Trimmed	Δ (kPa, %)
Intercept	−0.464	−0.470	−0.0061 (1.3)
Retrofit status—Post-retrofit	0.090	0.079	0.0109 (12.1)
External Pw	0.556	0.496	−0.0600 (10.8)
External T	−0.010	−0.010	0.0004 (4.3)
Room T	0.056	0.060	0.0042 (7.6)
Room Type—Other rooms	0.107	0.088	−0.0181 (17.0)
Season—Non-heating	0.011	0.011	0.0006 (5.7)
Occupants	0.015	0.012	−0.0022 (15.4)

Coefficients for external temperature, room temperature, and season showed negligible change across both datasets (Δ < 8%), confirming that these relationships are not driven by observations at the tails of the predictor distributions.

The retrofit effect attenuated modestly from 0.090 to 0.079 kPa (12.1%) but remained directionally consistent and of comparable magnitude. The key inference regarding retrofit impact on indoor humidity was therefore unchanged. External vapor pressure showed the largest absolute shift (Δ = −0.060 kPa, 10.8%), which is expected given that it has the widest dynamic range among the predictors. Its coefficient is consequently most sensitive

to tail exclusion. Room type and occupant count exhibited proportional changes of 17.0% and 15.4%, respectively. However, their absolute magnitudes remain small and do not alter the interpretation of either predictor.

3.5. Random Effects and Building Heterogeneity

The intraclass correlation coefficient (ICC) of 0.29 indicated that 29% of the total variance in indoor vapor pressure was attributable to systematic differences between homes, independent of the measured predictors. This underlines the importance of the random effects structure. A fixed-effects-only model would treat these between-home differences as unexplained noise, inflating residual variance and producing overly narrow standard errors for the population-level estimates.

The random intercept variance ($\tau_{00} = 0.02$) exceeded the residual variance ($\sigma^2 = 0.01$), confirming that baseline differences in indoor humidity across homes were a larger source of variability than within-home measurement noise. The random slope variances for both retrofit status and external vapor pressure were negligible ($\tau_{11} \approx 0.00$ for both), suggesting that while homes differed substantially in their baseline humidity levels, they responded to the retrofit intervention and to outdoor moisture conditions with broadly similar sensitivity once baseline level was accounted.

The strong negative correlation between random intercepts and the external vapor pressure slope ($\rho_{01} = -0.79$) is clearly visible in Figure 2. This figure plots each home's estimated baseline humidity level against its sensitivity to outdoor moisture conditions, separately for pre- and post-retrofit periods. Homes with higher baseline indoor humidity (positive random intercept, e.g., H09, H21) consistently exhibited weaker coupling to external moisture conditions, while homes with lower baselines (e.g., H02) showed the strongest outdoor–indoor moisture linkage. This pattern is physically plausible. Dwellings with persistently elevated indoor humidity may have characteristics, such as higher occupant moisture generation or less effective ventilation, that dominate the indoor moisture balance and reduce the relative influence of outdoor conditions. Notably, the pre- and post-retrofit panels of Figure 2 are near-identical. The near-identical structure across pre- and post-retrofit panels is itself informative: it demonstrates that the negative intercept–slope relationship is a stable property of the home sample rather than an effect of the intervention. Both panels are shown to make this stability directly visible.

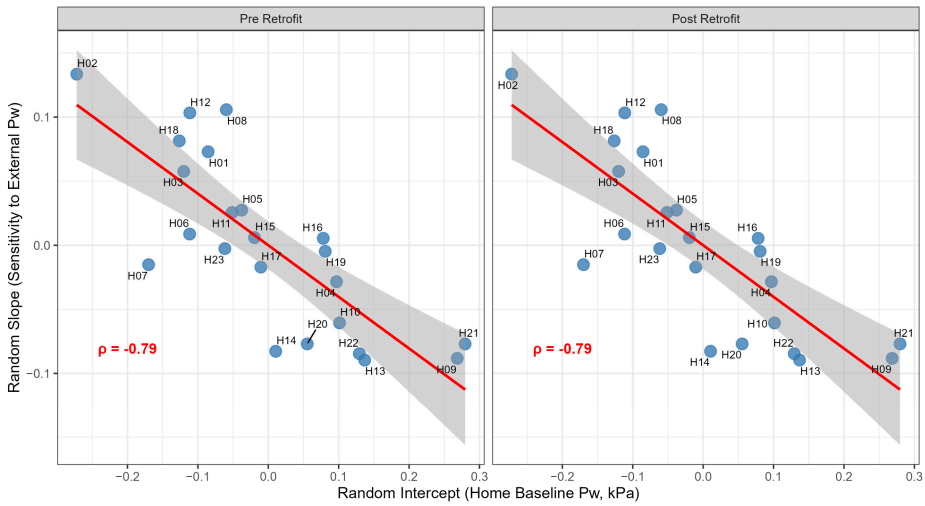

Figure 2. Random effects correlation structure: home-level random intercepts (baseline indoor vapor pressure) versus random slopes (sensitivity to external vapor pressure), shown separately for pre- and post-retrofit periods. Each point represents one home (n = 23). The red line and shaded band show the fitted linear trend and 95% confidence interval. ρ = −0.79 in both periods.

3.6. Room Type and Seasonal Effects

Figure 3 illustrates the distribution of indoor vapor pressure by room type and season, separately for pre- and post-retrofit periods. Indoor humidity was consistently higher in the non-heating season than in the heating season across both room types and both periods. Other rooms, i.e., not the living rooms, exhibited higher vapor pressure than living rooms throughout. Post-retrofit, both patterns were preserved but at a uniformly elevated level, relative to pre-retrofit conditions.

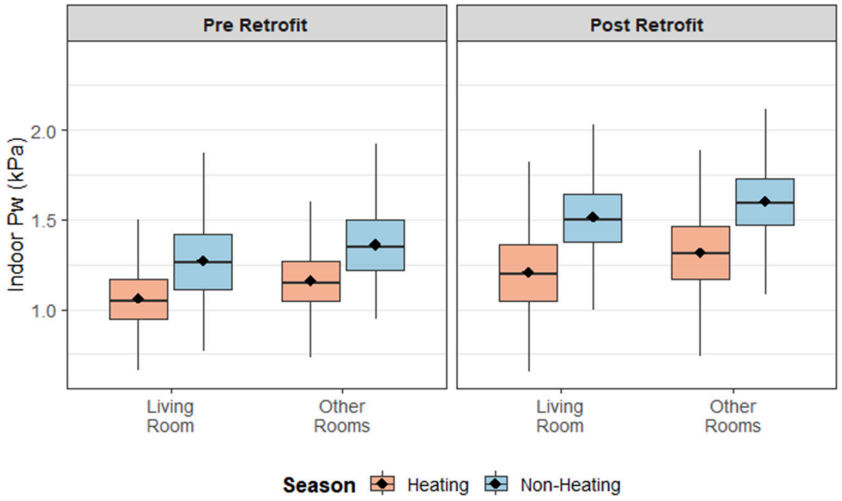

Figure 3. Indoor vapor pressure (kPa) by room type and season, pre- and post-retrofit (n = 23 homes, 28,822 observations). Boxes show interquartile range; horizontal line and diamond denote median and mean respectively. Heating season: October–May; non-heating season: June–September.

Although the seasonal separation is visually apparent in Figure 3, the standardized coefficient for season in the final model was small ($\hat{\beta}$ = 0.04, 95% CI [0.03, 0.06]). This indicated a modest contribution of seasons to indoor humidity variation relative to the dominant predictors of external vapor pressure and room temperature. Formal pairwise contrasts for season were therefore not pursued.

Tukey-adjusted pairwise comparisons from the estimated marginal means of the final model confirmed that these visual differences were statistically distinguishable after controlling for outdoor climate and occupancy, Table 6. Other rooms had significantly higher indoor vapor pressure than living rooms both pre-retrofit (Δ = 0.107 kPa, 95% CI [0.103, 0.110], $p < 0.001$) and post-retrofit (Δ = 0.107 kPa, 95% CI [0.103, 0.110], $p < 0.001$). The identical magnitude of this contrast in both periods indicates that the retrofit did not alter the relative humidity differential between room types. The structural moisture environment of other rooms (kitchens, bathrooms, bedrooms) remained persistently more humid than living rooms regardless of retrofit status. Elevated vapor pressure in rooms proximal to moisture sources such as kitchens and bathrooms is consistent with localized moisture generation, though the mediating roles of ventilation frequency and room usage pattern were not directly measured in this study.

Table 6. Pairwise comparisons of estimated marginal means for indoor Pw (kPa) by RoomType × RetrofitStatus. Estimates are Tukey-adjusted differences. Positive values indicate the first group has higher Pw than the second.

Contrast	Estimate, kPa [95% CI]	Std. Error	*p*-Value
Living Room Pre-Retrofit vs. Other Rooms Pre-Retrofit	−0.107 [−0.110, −0.103]	0.001	<0.001

Table 6. *Cont.*

Contrast	Estimate, kPa [95% CI]	Std. Error	*p*-Value
Living Room Pre-Retrofit vs. Living Room Post-Retrofit	−0.090 [−0.122, −0.058]	0.013	<0.001
Other Rooms Pre-Retrofit vs. Other Rooms Post Retrofit	−0.090 [−0.122, −0.058]	0.013	<0.001
Living Room Post-Retrofit vs. Other Rooms Post-Retrofit	−0.107 [−0.110, −0.103]	0.001	<0.001

The retrofit effect was also consistent across room types: both living rooms and other rooms showed a statistically significant increase in indoor vapor pressure post-retrofit of identical magnitude (Δ = 0.090 kPa, 95% CI [0.058, 0.122], $p < 0.001$ for both contrasts). The symmetry of these contrasts—both the room type difference and the retrofit effect—suggests that the retrofit acted as a uniform shift in the moisture baseline of the homes rather than differentially affecting rooms with distinct moisture source profiles.

3.7. Holdout Temporal Validation

To assess temporal generalizability of the fitted model, a 70/30 holdout assessment was performed. Observations were partitioned by stratified random sampling within homes, ensuring all 23 homes contributed to both training and test sets. This design evaluates whether the model generalizes to unseen time points from the monitored homes; it does not test transferability to an independent cohort of buildings. The model was re-fitted on the training set alone followed by testing the model created on the training set using the test data set.

On the held-out test set, the model achieved R^2 = 0.817 and RMSE = 0.11 kPa, closely matching the conditional R^2 of 0.824 obtained on the full dataset. This consistency between in-sample and out-of-sample performance indicates that the model is not overfitted to the training data and that the fixed effects structure generalizes well to unseen observations. Figure 4 shows the predicted versus observed vapor pressure values for the test set. Points cluster tightly along the 1:1 line across the full range of observed values (approximately 0.7–2.2 kPa), with no evidence of systematic bias at either the lower or upper extremes of the distribution.

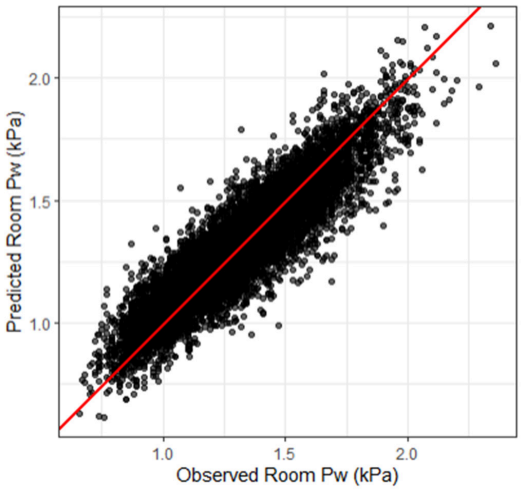

Figure 4. Predicted versus observed indoor vapor pressure (kPa) on the held-out test set (30% of observations, n ≈ 8647). The model was fitted on the training set (70%) and applied to unseen observations using fixed effect predictions with home-level random effects estimated from training data. The red line denotes perfect agreement (slope = 1, intercept = 0). Test R^2 = 0.817; RMSE = 0.110 kPa.

The training set-test set validation was performed to ensure the model was not overfitted. As the validation set comprises held-out observations from the same 23 homes rather than entirely new buildings, this assessment reflects temporal generalizability within the study sample. This is not about the transferability of the model to an independent cohort.

4. Discussion

This study demonstrates the rigorous framework provided by linear mixed effects models for evaluating energy retrofit impacts on indoor climate parameters while appropriately controlling for confounding variables. Using indoor vapor pressure in 23 Irish homes as a case study, the model identified a statistically significant post-retrofit increase of 0.090 kPa (approximately 6.6% relative to the pre-retrofit mean), after adjusting for outdoor climate, season, room type, and occupancy. This finding is consistent with studies documenting increased indoor moisture levels in homes where retrofit-driven reductions in air infiltration were not accompanied by adequate mechanical ventilation [7,9], and with Irish-specific evidence showing that post-retrofit improvements in airtightness can compromise indoor air quality when occupant ventilation behavior and installed ventilation systems are insufficient [8,10]. Whether post-retrofit humidity increases or decreases appears to depend critically on the ventilation strategy adopted alongside envelope improvements—a finding that reinforces the need for the kind of controlled, covariate-adjusted analysis presented here rather than simple before/after comparisons.

The dominant role of external vapor pressure and temperature in predicting indoor humidity reflects fundamental hygrothermal physics: interior moisture conditions arise primarily from the infiltration and diffusion of outdoor air moisture, modified by occupant-driven moisture generation and buffered by building thermal and hygric mass [12,39]. Even following envelope upgrades, outdoor climate remains the primary driver of indoor humidity variability, with external vapor pressure alone accounting for the largest standardized effect in the model ($\hat{\beta}$ = 0.63). This finding underscores that retrofit effectiveness in humidity management cannot be evaluated in isolation from the outdoor climate context—a design principle with direct implications for cross-study comparisons that pool results from different climatic zones.

The relative contribution of each pathway depends on building-specific properties such as airtightness, ventilation provision, and envelope permeability. This is partly reflected in the home-level variability captured by the random slope for ExternalPw in the selected model. Although the present work does not embed a first-principles moisture balance, the LME framework is designed to be outcome-agnostic and transferable across indoor climate parameters. A formal physical decomposition of the outdoor–indoor coupling mechanism will be pursued in a subsequent study using the complete dataset and focused specifically on the predictors of indoor humidity before and after retrofit. The regression coefficients can be contextualized against physical frameworks [12,39] in domain specific studies.

In practical terms, the detected 0.090 kPa increase in indoor vapor pressure, under typical temperate indoor temperature conditions, would correspond to ~4–5 percentage increase in RH. The precise value would be temperature dependent. This magnitude carries meaningful consequences for indoor environmental quality: sustained relative humidity above 60% is associated with conditions conducive to mold germination and dust mite proliferation, both of which are linked to respiratory health outcomes including asthma exacerbation and allergic rhinitis [40]. Where post-retrofit homes already operate near this threshold—as may be the case in Ireland's cool, humid Atlantic climate—even a modest humidity increase of this magnitude may shift occupancy periods from below to above the biological risk threshold. This finding highlights the importance of integrating

humidity monitoring and adequate ventilation specification into retrofit programs rather than focusing solely on thermal and energy performance metrics.

While this paper focuses on residential energy retrofits, the analytical framework and R code presented here is domain-agnostic. It has direct applicability to before/after intervention studies across a range of scientific and engineering disciplines. Any study design involving repeated measurements nested within experimental units, a clearly defined pre/post intervention structure, and multiple confounding variables is a candidate for this approach. The open-source code provided with this paper requires only that users substitute their outcome variable, redefine their grouping structure, and specify contextually appropriate fixed effects, making cross-disciplinary adoption straightforward.

4.1. Advantages of the Mixed Effects Approach

The mixed effects framework offers several methodological advantages over simpler approaches to retrofit evaluation that are worth making explicit, particularly for researchers considering its adoption in their own field studies.

Simple before/after comparisons attribute all observed change in an indoor climate parameter to the retrofit, regardless of whether contemporaneous changes in outdoor climate, season, or occupant behavior may equally explain the difference [7]. By including fixed effects for external vapor pressure, temperature, and season, the LME model partitions these contributions explicitly, isolating the retrofit effect from systematic variation in background conditions. This is not merely a statistical nicety: as demonstrated in Section 3.2, the selected model with seasonal and occupancy controls was significantly preferred over the reduced model by likelihood ratio test ($\chi^2(2) = 16.34$, $p < 0.001$), confirming that omitting these terms would have produced a materially biased retrofit effect estimate.

A related advantage concerns uncertainty quantification. Analyses that treat repeated measurements from the same home as independent observations—as ordinary least squares regression does—produce artificially narrow standard errors and confidence intervals, inflating Type I error rates [11,15]. By modelling both fixed effects and random variation across homes simultaneously, the LME framework ensures that confidence intervals for the retrofit effect properly reflect between-home heterogeneity. The ICC of 0.29 observed in our case study illustrates the practical importance of this: nearly one-third of total variance in indoor vapor pressure is attributable to home-level differences, a quantity that would otherwise be absorbed into the residual and misrepresent precision.

Beyond correcting for dependence, the random effects structure actively quantifies heterogeneity in retrofit response—a dimension of practical value that purely fixed-effects approaches cannot provide [15]. The home-level random slopes estimated here reveal that buildings differ in their sensitivity to outdoor moisture conditions, and the negative intercept–slope correlation ($\rho = -0.79$) identifies a systematic pattern in which homes with higher baseline humidity exhibit weaker outdoor–indoor coupling. Such information is directly actionable for practitioners: it points toward building-level characteristics that may moderate retrofit effectiveness and that warrant investigation in larger studies.

The framework also preserves the full information content of the dataset. Aggregation-based approaches—such as computing monthly or seasonal means per home prior to analysis—discard within-period variability and can introduce ecological fallacy-type artefacts where group-level associations diverge from individual-level relationships [41]. By retaining hourly observations and modelling their dependence structure directly, the LME approach avoids information loss while maintaining valid inference. This is particularly important when, as here, within-home variability across time is itself of scientific interest. Furthermore, the model handles unbalanced data—homes with different monitoring dura-

tions or missing observations in one period—without requiring listwise deletion, making it well-suited to the realities of longitudinal field studies [15].

Finally, the provision of fully reproducible, open-source R code directly addresses a broader credibility challenge in building performance research [19]. Publishing complete analysis code alongside results allows readers to verify methodological choices, extend the analysis to new outcomes or datasets, and adapt the framework to their own retrofit monitoring programs. In a field where study designs, data structures, and outcome definitions vary considerably, transferable and transparent analytical tools are arguably as valuable as any single empirical finding.

4.2. Limitations and Caveats

While the model controls for the primary measured confounders like outdoor climate, season, room type, and occupancy, several unmeasured sources of confounding remain. Occupants who have invested in a retrofit may alter their ventilation or heating behavior independently of the physical changes to the building envelope, a form of awareness-driven behavioral change that the model cannot distinguish from the structural retrofit effect. Sensor drift or physical relocation of monitoring equipment between pre- and post-retrofit periods could also introduce systematic measurement bias. Addressing these sources of residual confounding fully would require either randomized controlled trial designs, impractical at scale for residential retrofit programs, or instrumental variable approaches that exploit exogenous variation in retrofit uptake, both substantially more resource-intensive than observational field monitoring [7,14].

Indoor humidity exhibits daily and weekly cycles driven by occupancy patterns, ventilation behavior, and diurnal outdoor climate variation. While the Durbin–Watson statistic (1.45) suggests minimal first-order autocorrelation in the model residuals, it is just a preliminary check.

Our approach of modelling hourly observations with random effects captures much of the between-home temporal structure. But it does not explicitly model within-home autocorrelation at finer time scales. More sophisticated temporal models like autoregressive specifications, seasonal ARIMA, or state-space models, may provide improved fit for applications where the temporal dynamics of indoor climate parameters are themselves of primary interest, though typically at the cost of interpretability and computational tractability [13].

The presence of residual autocorrelation is not unexpected in hourly indoor climate data, where adjacent observations are influenced by persistent occupancy patterns, ventilation behavior, and thermal inertia. In this study, the autocorrelation was modest but detectable, and an AR(1) sensitivity model fits the data better than the simpler independence-based specification. Importantly, the substantive conclusions regarding the retrofit effect were unchanged, indicating that the main findings are robust to reasonable alternative assumptions about the temporal error structure. Future applications of the framework to similarly dense monitoring data may benefit from explicitly modelling serial dependence when it is clearly present.

The case study results are specific to the Irish residential context. The direction and magnitude of the retrofit effect on indoor humidity may differ substantially in other climate zones or in housing stocks with different construction standards, airtightness levels, or thermal mass. Retrofit strategies also vary widely in intensity and scope, and findings from envelope-focused interventions may not generalize to deep retrofits that include mechanical ventilation with heat recovery. Practitioners applying this framework in other contexts should treat the fixed effect coefficients as case study-specific and reassess model structure using locally relevant predictor sets and climate data.

The illustrative findings reported here, including the estimated post-retrofit increase in indoor vapor pressure and the observed room-type differences, are intended to demonstrate the interpretive capacity of the LME framework rather than to constitute a definitive humidity risk assessment. A formal assessment of moisture-related health risk, including computation of mold growth indices and integration of location-specific climatic exceedance thresholds, is outside the scope of this methodological paper. Similarly, the attribution of elevated humidity in certain room types to proximity to moisture sources is offered as a plausible interpretation consistent with the model output, not as a validated causal claim. Mediating effects of ventilation frequency and room functional usage were not directly measured and warrant investigation in dedicated future work.

With 23 homes and approximately 29,000 observations, the dataset provides substantial power for estimating population-average fixed effects, including the retrofit effect. However, power for detecting home-level differences in random slopes (for example, identifying which specific buildings exhibit unusually strong or weak retrofit responses) is more limited at this cluster count [24]. Researchers whose primary interest lies in building-specific effect heterogeneity rather than population-average inference should plan monitoring programs accordingly.

4.3. Practical Implementation Guidance

For researchers seeking to apply this methodology to their own datasets, the following guidance is available:

Understand your data structure and identify all levels of clustering (homes, rooms, time periods, geographic regions). Random effects should be specified at each relevant level. Select all mechanistically relevant predictors for the model. Include variables with clear theoretical/physical justification for the outcome. Check assumptions of the linear model and if violations are noted or warnings come up during execution, conduct sensitivity analyses. If needed, consider employing alternative link functions or variance structures [11].

To compare the contribution across predictors, you may either adjust the predictors (standardize the data, convert to z-scores) or compare the standardized β coefficients. In this work, the latter approach is used. Whichever path you take, document the process for reproducibility. Reporting should specify both fixed and random effects. The advantage of LMEs is that we are not limited to population-average effects. We can report the heterogeneity across grouping units. This information is crucial for practitioners seeking to tailor interventions. Choose the response metrics analyzed strategically, based on research questions and study objectives, not based on what is convenient to analyze. Practitioners should also consider outcome metrics aligned with specific standards, for example, EN16798 or Standard 55 with respect to thermal comfort.

To facilitate verification and adaptation, works must consider providing minimal reproducible code alongside the publication. Open-source code with detailed comments accelerates methodological progress and builds community trust [19]. To complement this, consider using tidy data standards, as described in Section 2.1.2. This ensures easy machine readability and the ability to apply existing code without restructuring. Providing representative sample data, to enable code testing without disclosing proprietary full datasets, is also good practice.

4.4. Future Directions

Future research would need to extend to multivariate models, incorporating multiple indoor climate parameters simultaneously (co-modelling temperature, humidity, and air quality) to assess coupling effects. Similarly, time series analysis can be added in to explicitly

analyze the temporal impacts of interventions, potentially using state-space or dynamic linear models [13]. The strong negative correlation between random intercepts and external humidity needs to be verified in studies with larger and more diverse home samples for generalizing conclusions.

The methodology needs to be applied to datasets spanning diverse climates and building stocks to assess generalizability. A deeper analysis of interaction effects between retrofit strategy and occupant behavior is also called for to identify optimal retrofit approaches for specific populations. It is envisaged that the open-source solutions provided through this work will accelerate these steps.

5. Conclusions

Energy retrofits represent a critical intervention for improving building sector energy efficiency and achieving decarbonization targets. However, rigorously evaluating their impacts requires appropriate statistical methodology that accounts for confounding, clustering, and heterogeneity. Linear mixed effects models provide such a framework, enabling unbiased estimation of retrofit effects while quantifying building-level variability in retrofit response.

This paper has presented complete methodological guidance and reproducible R code for implementing this approach. The substantial heterogeneity in retrofit response across buildings underscores the value of building-specific analysis. Researchers are encouraged to adopt and adapt this methodology for their own retrofit evaluation studies. The provided R code, with detailed comments, substantially reduces barriers to implementation. By lowering the technical threshold for rigorous retrofit evaluation, it is hoped to accelerate progress in building-climate science and support evidence-based retrofit policy and practice.

Funding: A.K.M. is a Dorothy and Marie Skłodowska-Curie fellow, partly funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101034345. The work was partly supported by the Building Climate-Resilient Education Systems project, which is funded through the BAOBAB project at ASCEND, which is jointly funded by UK aid through the Foreign, Commonwealth and Development Office (FCDO), the International Development Research Centre (IDRC), Canada and by the Ministry of Foreign Affairs of the Netherlands as part of the Climate Adaptation and Resilience (CLARE) research programme and Step Change initiative.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Complete R code and a representative 5% public sample of data are available in the GitHub repository: https://github.com/asitkm76/MixedEffectsModel_Retrofit_IndoorClimate (Last assessed 22 May 2026). All code is provided under Creative Commons license, CC BY-SA, encouraging community contribution and extension.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

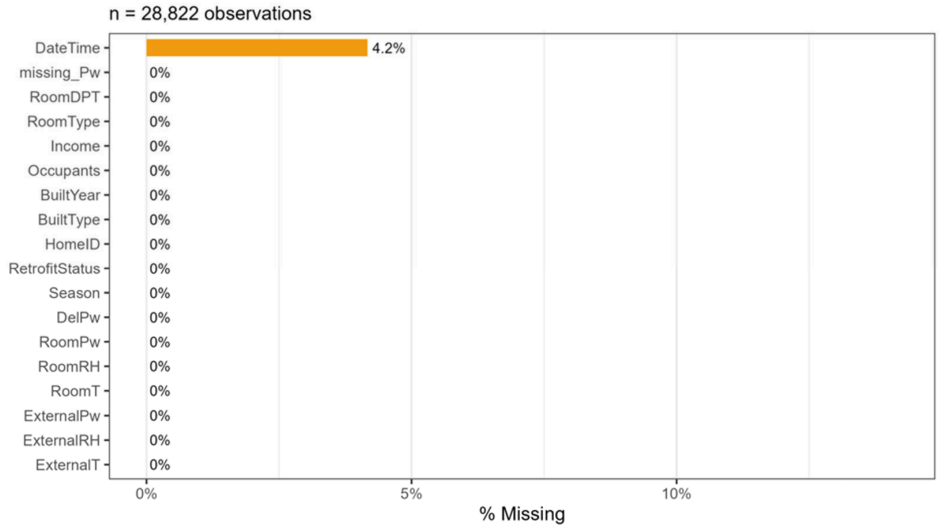

Figure A1. Variable specific missing proportions (Section 2.1.3).

Table A1. Systematic random slope selection—Likelihood Ratio table.

Random Slop Structure	No. of Parameters	AIC	Convergence—Gradient
(1 | HomeID)	10	−42,312	<0.00001
(1 + RetrofitStatus | HomeID)	12	−43,726	<0.00001
(1 + RetrofitStatus + ExternalPw | HomeID)	15	−44,426	0.00003

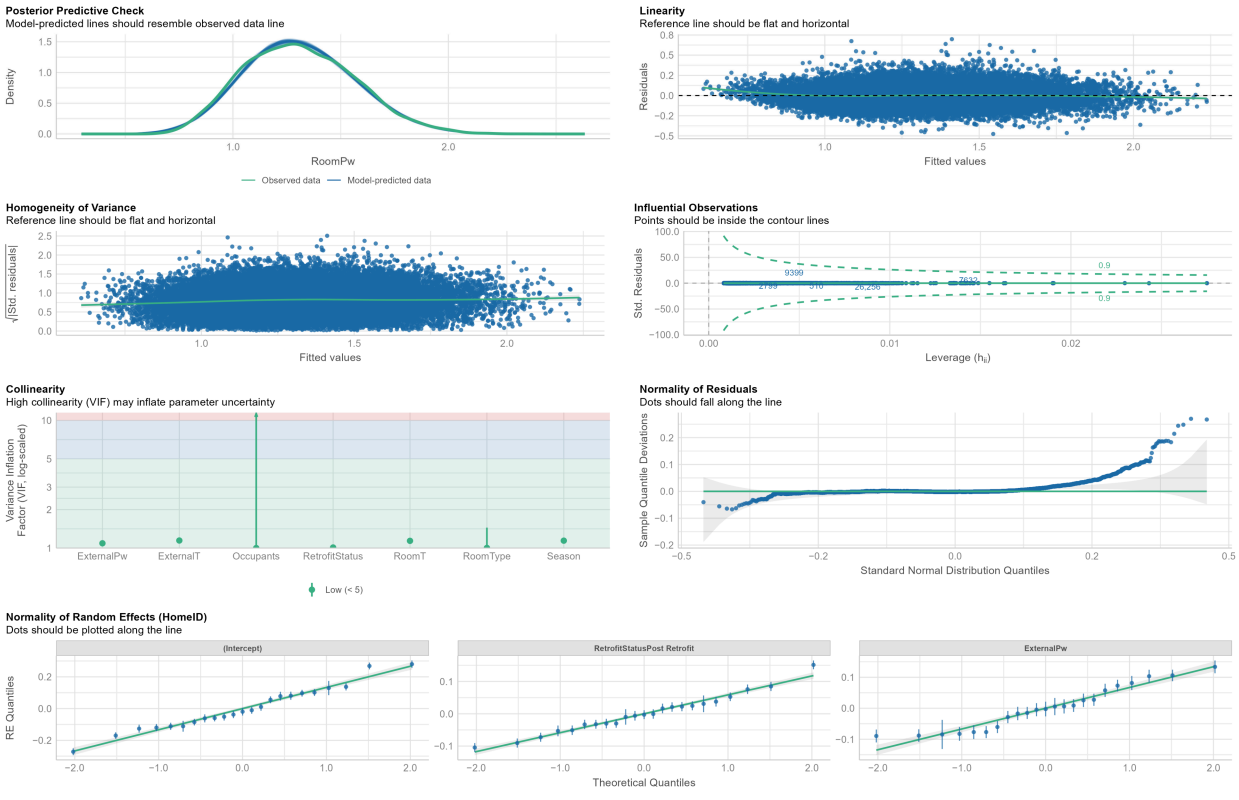

Figure A2. Regression diagnostic plots for the final model (Section 3.4).

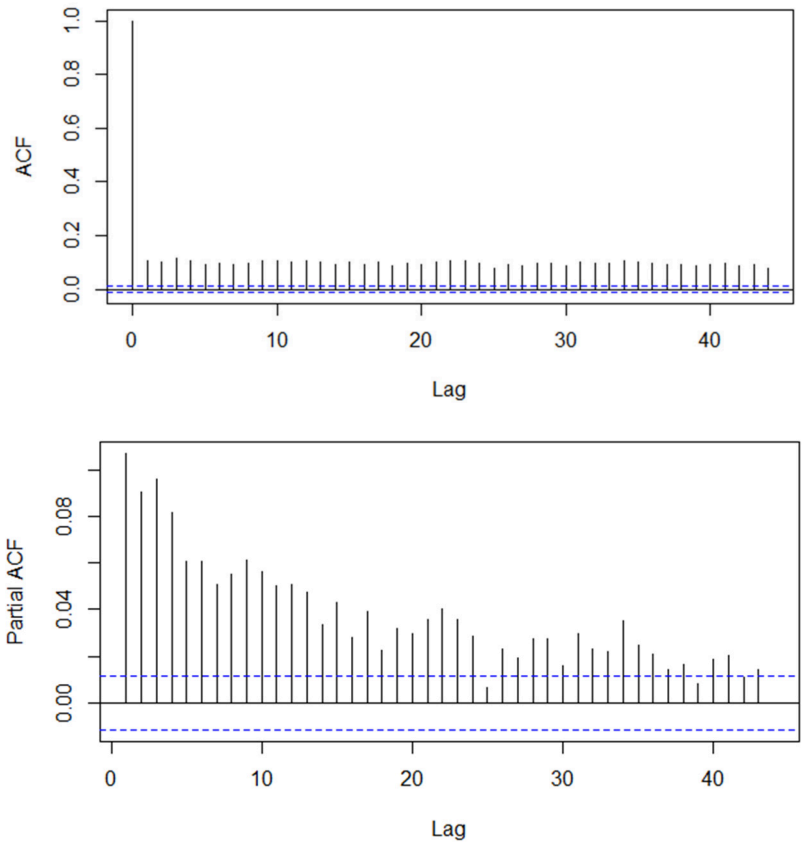

Figure A3. Auto-correlation (dashed lines indicate the approximate 95% confidence limits for zero autocorrelation; spikes outside these limits suggest statistically notable residual autocorrelation at the corresponding lag) and partial auto-correlation plots (dashed lines indicate the approximate 95% confidence limits for zero partial autocorrelation; spikes outside these limits suggest statistically notable residual dependence at the corresponding lag) for residuals.

References

1. European Commission. Fit for 55: Delivering on the Proposals. Available online: https://commission.europa.eu/topics/climate-action/delivering-european-green-deal/fit-55-delivering-proposals_en (accessed on 17 March 2026).
2. Department of Climate, Energy and the Environment. Climate Action Plan 2023. 2022. Available online: https://www.gov.ie/en/department-of-climate-energy-and-the-environment/publications/climate-action-plan-2023/ (accessed on 19 March 2026).
3. Dall'O', G.; Galante, A.; Pasetti, G. A methodology for evaluating the potential energy savings of retrofitting residential building stocks. *Sustain. Cities Soc.* **2012**, *4*, 12–21. [CrossRef]
4. Wang, C.; Wang, J.; Norbäck, D. A Systematic Review of Associations between Energy Use, Fuel Poverty, Energy Efficiency Improvements and Health. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7393. [CrossRef] [PubMed]
5. Dorizas, V.; Düvier, C.; Elnagar, E.; Zuhaib, S. Healthy Buildings Barometer 2024. 2024. Available online: https://press.velux.ch/healthy-buildings-barometer-2024/ (accessed on 17 March 2026).
6. European Commission. A Renovation Wave for Europe—Greening Our Buildings, Creating Jobs, Improving Lives. Available online: https://ec.europa.eu/newsroom/clima/items/690287/ (accessed on 26 March 2026).
7. Fisk, W.J.; Singer, B.C.; Chan, W.R. Association of residential energy efficiency retrofits with indoor environmental quality, comfort, and health: A review of empirical data. *Build. Environ.* **2020**, *180*, 107067. [CrossRef]
8. Hassan, H.; Mishra, A.K.; Wemken, N.; O'Dea, P.; Cowie, H.; McIntyre, B.; Coggins, A.M. Deep energy renovations' impact on indoor air quality and thermal comfort of residential dwellings in Ireland–ARDEN project. *Build. Environ.* **2024**, *259*, 111637. [CrossRef]
9. Shrubsole, C.; Macmillan, A.; Davies, M.; May, N. 100 Unintended consequences of policies to improve the energy efficiency of the UK housing stock. *Indoor Built Environ.* **2014**, *23*, 340–352. [CrossRef]
10. Coggins, A.M.; Hogan, V.; Mishra, A.K.; Norton, D.; Foster, D.; Wemken, N.; Cowie, H.; Doherty, E. Energy retrofits: Factors affecting a just transition to better indoor air quality. *Indoor Environ.* **2024**, *1*, 100058. [CrossRef]

11. Fitzmaurice, G.M.; Laird, N.M.; Ware, J.H. *Applied Longitudinal Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012. Available online: https://books.google.com/books?hl=en&lr=&id=0exUN1yFBHEC&oi=fnd&pg=PR17&dq=+Applied+longitudinal+analysis+(2nd+ed.)&ots=BfpROXtJ0C&sig=x5-zVr1fx8cTV-Ay4DSiktdVtPA (accessed on 19 March 2026).
12. Kunzel, H.M. *Simultaneous Heat and Moisture Transport in Building Components*; Fraunhofer Institute of Building Physics: Stuttgart, Germany, 1995.
13. Cowpertwait, P.S.; Metcalfe, A.V. *Introductory Time Series with R*; Springer Science & Business Media: Berlin, Germany, 2009. Available online: https://books.google.com/books?hl=en&lr=&id=QFiZGQmvRUQC&oi=fnd&pg=PR7&dq=Introductory+time+series+with+R.+&ots=p1qZpO_TYD&sig=-7Sn1_IT94hMxnNfl4QhCQHsvEA (accessed on 19 March 2026).
14. Kurnik, C.W.; Agnew, K.; Goldberg, M. *Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2017. Available online: https://docs.nlr.gov/docs/fy17osti/68564.pdf (accessed on 26 March 2026).
15. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
16. Coggins, A.M.; Wemken, N.; Mishra, A.K.; Sharkey, M.; Horgan, L.; Cowie, H.; Bourdin, E.; McIntyre, B. Indoor air quality, thermal comfort and ventilation in deep energy retrofitted Irish dwellings. *Build. Environ.* **2022**, *219*, 109236. [CrossRef]
17. Met Éireann. Historical Data—Met Éireann—The Irish Meteorological Service. Available online: https://www.met.ie/climate/available-data/historical-data (accessed on 18 March 2026).
18. SEAI. DEAP Software & Methodology. Available online: https://www.seai.ie/ber/support-for-ber-assessors/deap (accessed on 21 March 2026).
19. Peng, R.D. Reproducible Research in Computational Science. *Science* **2011**, *334*, 1226–1227. [CrossRef] [PubMed]
20. Mishra, A.K.; Moran, P.; Goggins, J. *Influential Determinants of Indoor Humidity in Homes and the Impact of Retrofits*; CERI: Dublin, Ireland, 2022.
21. Alduchov, O.A.; Eskridge, R.E. Improved Magnus form approximation of saturation vapor pressure. *J. Appl. Meteorol. (1988–2005)* **1996**, *35*, 601–609. [CrossRef]
22. ANSI/ASHRAE. *Standard 55. Thermal Environmental Conditions for Human Occupancy*; ANSI/ASHRAE: Atlanta, GA, USA, 2023. Available online: https://www.ashrae.org/technical-resources/bookstore/standard-55-thermal-environmental-conditions-for-human-occupancy (accessed on 11 August 2024).
23. Wickham, H. Tidy data. *J. Stat. Softw.* **2014**, *59*, 1–23. [CrossRef]
24. McNeish, D.M.; Stapleton, L.M. The effect of small sample size on two-level model estimates: A review and illustration. *Educ. Psychol. Rev.* **2016**, *28*, 295–314. [CrossRef]
25. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019. Available online: https://books.google.com/books?hl=en&lr=&id=BemMDwAAQBAJ&oi=fnd&pg=PR11&dq=+Statistical+Analysis+with+Missing+DataStatistical+Analysis+with+Missing+Data&ots=FCFS71GYVX&sig=aBS4CNvnpr4gBzJHTZx_dtNybIQ (accessed on 27 March 2026).
26. Buuren, S. *van Flexible Imputation of Missing Data*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2018; Available online: https://stefvanbuuren.name/fimd/ (accessed on 27 March 2026).
27. O'brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
28. Lüdecke, D. sjPlot-Data Visualization for Statistics in Social Science. Zenodo. 2021. Available online: https://ui.adsabs.harvard.edu/abs/2018zndo...2400856L/abstract (accessed on 21 March 2026).
29. Self, S.G.; Liang, K.-Y. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. *J. Am. Stat. Assoc.* **1987**, *82*, 605–610. [CrossRef]
30. Harrell, F.E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2015. [CrossRef]
31. Wasserstein, R.L.; Lazar, N.A. The ASA Statement on *p*-Values: Context, Process, and Purpose. *Am. Stat.* **2016**, *70*, 129–133. [CrossRef]
32. Royston, P. Approximating the Shapiro-Wilk W-test for non-normality. *Stat. Comput.* **1992**, *2*, 117–119. [CrossRef]
33. Durbin, J.; Watson, G.S. Testing for serial correlation in least squares regression: I. *Biometrika* **1950**, *37*, 409–428.
34. Liang, K.-Y.; Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, *73*, 13–22. [CrossRef]
35. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **2003**, *19*, 716–723. [CrossRef]
37. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [CrossRef]
38. Gelman, A. Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* **2008**, *27*, 2865–2873. [CrossRef] [PubMed]

39. Hens, H.S. *Building Physics-Heat, Air and Moisture: Fundamentals and Engineering Methods with Examples and Exercises*; John Wiley & Sons: Hoboken, NJ, USA, 2017. Available online: https://books.google.com/books?hl=en&lr=&id=q-Q5DwAAQBAJ&oi=fnd&pg=PA33&dq=Building+Physics+%E2%80%90+Heat,+Air+and+Moisture:+Fundamentals+and+Engineering+Methods+with+Examples+and+Exercises&ots=nhv39UaGtY&sig=PlKhWwrlFlJuQZGPZfi3gq0w1Vg (accessed on 21 May 2026).
40. Heseltine, E.; Rosen, J. *WHO Guidelines for Indoor Air Quality: Dampness and Mould*; World Health Organization (WHO): Geneva, Switzerland, 2009. Available online: https://books.google.com/books?hl=en&lr=&id=PxB8UUHihWgC&oi=fnd&pg=PP2&dq=WHO+Guidelines+for+Indoor+Air+Quality:+Dampness+and+Mould&ots=9COTOVQ1HK&sig=ozDC91V-W2cOZA3iHWK3mqd_wPc (accessed on 26 March 2026).
41. Robinson, W.S. Ecological correlations and the behavior of individuals. *Int. J. Epidemiol.* **2009**, *38*, 337–341. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.