

박사학위논문

건물 스톡 에너지 모델 기반 한국형
공동주택의 VSM 데이터셋 개발 및
에너지 수요 반응 적용성 평가

2026년

한 성 대 학 교 대 학 원

스마트융합컨설팅학과

스마트융합제품전공

김 재 춘

박사학위논문
지도교수 노광현

건물 스톡 에너지 모델 기반 한국형
공동주택의 VSM 데이터셋 개발 및
에너지 수요 반응 적용성 평가

Development of a VSM Dataset for Korean Apartment Buildings
Based on a Building Stock Energy Model and Evaluation of Its
Applicability to Demand Response

2025년 12월 일

한 성 대 학 교 대 학 원

스마트융합건설팅학과

스마트융합제품전공

김 재 춘

박사학위논문
지도교수 노광현

건물 스톡 에너지 모델 기반 한국형
공동주택의 VSM 데이터셋 개발 및
에너지 수요 반응 적용성 평가

Development of a VSM Dataset for Korean Apartment Buildings
Based on a Building Stock Energy Model and Evaluation of Its
Applicability to Demand Response

위 논문을 공학 박사학위 논문으로 제출함

2025년 12월 일

한 성 대 학 교 대 학 원

스마트융합건설팅학과

스마트융합제품전공

김 재 춘

김재춘의 공학 박사학위 논문을 인준함

2025년 12월 일

심사위원장 김 승 천 (인)

심 사 위 원 노 광 현 (인)

심 사 위 원 이 후 진 (인)

심 사 위 원 남 현 우 (인)

심 사 위 원 박 점 배 (인)

국 문 초 록

건물 스톡 에너지 모델 기반 한국형 공동주택의 VSM 데이터셋 개발 및 에너지 수요 반응 적용성 평가

한 성 대 학 교 대 학 원
스 마 트 용 합 컨 설 팅 학 과
스 마 트 용 합 제 품 전 공
김 재 춘

최근 인공지능 서비스의 확산에 따른 데이터센터의 건설이 확대되고 지속적으로 증가하는 전기자동차로 인해 전력수요가 수도권을 중심으로 빠르게 증가하고 있지만 발전설비의 확충과 송전망 투자는 이러한 수요를 따라가지 못하고 있다. 특히 전국적으로 확대되고 있는 태양광 발전을 포함하는 신재생 에너지는 출력의 간헐성과 발전 예측의 불확실성으로 인해 계통 불안에 가중시키고 있다. 실제로 최근 지속적으로 발령되는 태양광 발전의 계통 차단은 공급자 중심의 전력 정책의 전환이 필요함을 시사하고 있다.

한편, 국내 주택용 전력시장은 누진제를 중심으로 하는 요금 체계에서 향후 계시별 요금제로의 전환을 대비하는 차원에서 수용가 각 세대의 시간별 부하 특성을 정확하게 파악하고 이를 기반으로 하는 수요예측의 중요성이 증가하고 있다. 이러한 맥락에서 공동주택에 세대별로 설치된 스마트미터는 각 수용가의 실측된 전력 사용량(RSM) 데이터를 기반으로 건물 속성을 추정하

고 부하 패턴의 분류를 가능하게 하는 핵심 인프라로 주목받고 있지만 실제 공동주택에 설치된 스마트미터는 민간 영역으로 분류되어 전력 수요 시장에서 수요예측 데이터로 활용되지 못하고 있는 것이 현실이다.

따라서 본 연구는 이러한 공동주택의 스마트미터 데이터의 한계를 보완하기 위하여 건물 스톡 에너지 모델(BSEM)에 기반한 가상스마트미터 데이터셋을 한국 공동주택 환경에 맞게 KR-VSM을 개발하고 이를 토대로 가상발전사업자(VPP)가 수요예측부터 수요반응(DR) 운영전략까지 수립 가능한 프레임워크의 구축을 주된 연구의 목표로 하고 있다.

한국형 VSM 데이터셋의 개발은 Neale et al.(2018, 2020, 2022)이 제안한 QSFBSSEM의 Generator-Classifier-Extractor 프레임워크를 한국 공동주택 환경에 맞게 확장·재구성하였는데, 먼저 Generator 단계에서는 국내 공동주택과 관련된 객관적인 국가 통계 기반의 자료와 기후 자료 그리고 건물의 고유 물리 자료를 결합하여 총 21개의 파라미터를 확률에 기반하여 상향식으로 생성하고 이렇게 생성된 변수를 화이트박스 시뮬레이터인 EnergyPlus에 주입하여 20만 세대 규모로 공동주택의 KR-VSM 데이터셋을 생성하였다.

이어 연간·계절별·월별·시간대별 전력 사용 패턴을 정교하게 반영한 시계열 부하 패턴을 이용하여 난방 방식, 단열, 전용면적, 건축연도, 거주 지역을 기계학습 기반의 선형 분류기 LDA, 비선형 분류기인 Random Forest를 통해 KR-VSM 데이터셋을 분류(Classifier)하였다. 이렇게 분류된 VSM 데이터셋을 AMI 데이터유통플랫폼에서 확보된 897,322세대의 전체 검침 데이터 중에서 연간 8,760시간 중 최소 80% 이상 관측되는 연결성·결측률 기준을 충족한 세대의 계량기를 기준으로 480,118세대의 RSM 데이터와 비교를 통하여 역추정 모델(Extractor)을 검증하였다.

이를 통해 역추정한 결과 건물의 단열과 전용면적은 VSM-RSM 모두에서 높은 추적 안정성을 보였으며, 난방과 건축연도의 파라미터는 VSM의 신뢰도 구조가 RSM에서도 유지됨을 확인하였다.

본 연구는 이러한 Classifier-Extractor 결과를 바탕으로 16개의 DR 세그먼트를 구성하여 2만 세대의 RSM 데이터를 기반으로 실증 분석한 결과, 전체 세대에서 고반응 세그먼트가 DR 감축량에 의미 있는 결과를 도출함으로써

써 신뢰도가 DR 반응성과 예측 오차에 밀접하게 연관됨을 실증하였다. 이는 DR 참여자 선정 시 기존의 피크부하 중심에서 속성 기반의 신뢰도를 고려한 타겟팅이 효율적임을 의미한다.

종합하면, 본 연구는 공동주택 중심의 한국형 KR-VSM 데이터셋을 체계적으로 구축하고 RSM 데이터를 기반으로 그 유효성을 실증적으로 검증하였다. 나아가 검증된 KR-VSM 데이터셋에 건물 및 거주자 속성값을 입히고, 이를 기반으로 DR 반응성의 예측과 VPP 운영전략을 설계를 통한 수요반응의 밸류체인을 제시하였다. 이러한 연구결과는 향후 DR·VPP 정책 설계, 지역별 에너지 수요 관리, 건물 에너지 관리 등 다양한 분야에 실질적으로 활용 가능한 공동주택 중심의 한국형 VSM 데이터셋의 인프라를 제공하였다는 점에서 학문적 의의를 찾을 수 있을 것이다.

【주요어】 가상스마트미터(VSM), 실측 스마트미터(RSM), 건물 스톡 에너지 모델(BSEM), 수요 반응(DR), 부하 패턴 분류(Classifier), 역추정(Extractor), DR 세그먼트(DR Segment), 선형 판별 분석(LDA), 비선형 판별 분석(RandonForest), 확률 질량 함수(PMF), 균일 확률 분포(UPD)

목 차

I. 서론	1
1.1. 연구의 배경 및 필요성	1
1.2. 연구 목적	7
1.3. 선행 연구와 비교 및 연구의 차별성	10
II. 이론적 배경 및 선행 연구 분석	15
2.1. 건물 스톡 에너지 모델의 개념	15
2.2. 가상스마트미터 데이터셋의 개념	21
2.3. 건물 스톡 에너지 모델 기반 가상스마트미터 데이터셋 프레임워크 분석	26
2.4. 한국 스마트미터 인프라의 현실과 제약	28
III. 한국형 공동주택의 가상스마트미터 데이터셋 개발	31
3.1. 연구 설계 및 입력 변수 구조화	31
3.2. 시뮬레이션 구조 설계	53
3.3. 한국형 공동주택의 가상스마트미터 데이터셋 생성 및 분석	60
3.4. 실측 전력량 데이터와의 비교 및 정합성 검증 분석	68
3.5. 실측 데이터와 통계적 검증	99
IV. 에너지 수요 반응 적용성 평가	143
4.1. 수요 반응 적용을 위한 VSM 데이터셋의 정보 생성 체계	143
4.2. 수요 반응 적용 가능성 검증을 위한 분류 성능 평가	149
4.3. 실측 기반 수요 반응 적용 가능성 검증	175
4.4. 속성 기반 수요 반응 세그먼트 적용성 실험	184
4.5. 수요 반응 운용 관점에서의 적용성 종합 평가	216
V. 결론	220
5.1. 연구 요약 및 주요 성과	220
5.2. 향후 연구 방향	221
참 고 문 헌	223
ABSTRACT	239

표 목 차

[표 1-1] VSM 데이터셋의 필요성	6
[표 1-2] 캐나다와 대한민국의 지리적, 환경적, 기후적 비교	11
[표 2-1] 논문에 검토된 시뮬레이션 엔진 비율	17
[표 2-2] White-box 기반 EnergyPlus 해석의 장점	17
[표 2-3] 글로벌 VSM 관련 선행 연구 조사	24
[표 2-4] 캐나다 QSFBSSEM 입력 파라미터의 분류	27
[표 3-1] KR-VSM 파라미터의 분류	34
[표 3-2] KR-VSM 아키타입별 세부 유형	35
[표 3-3] KR-VSM 파라미터의 Class 세분화	36
[표 3-4] 빌딩 파라미터의 정의	39
[표 3-5] 외기 조건 파라미터의 정의	40
[표 3-6] Window Type 파라미터 설계의 원칙	41
[표 3-7] 범주형 변수의 파라미터별 규칙과 적용 기준	42
[표 3-8] 친환경 주택의 창면적비 기준	44
[표 3-9] 연속형 변수의 파라미터별 규칙과 적용 기준	45
[표 3-10] 시간 의존 부하의 파라미터별 규칙과 적용 기준	46
[표 3-11] Lighting Density 변수 설정 조건	48
[표 3-12] 건물 외벽 단열(RSI) 적용 기준	49
[표 3-13] 점유 스케줄별 부하 로직의 커스터마이징	52
[표 3-14] 메타데이터셋의 설명 및 단위	61
[표 3-15] 메타데이터 필드 구조	61
[표 3-16] 공동주택의 1년간 전력 사용량 통계	70
[표 3-17] 냉·난방기의 부하 형성 시간대 구성	71
[표 3-18] 공동주택의 공동전기와 세대전기 통계	72
[표 3-19] VSM 생성 데이터의 파라미터별 분포	73
[표 3-20] 일사량 절단 함수 상한값의 산정식	85
[표 3-21] KESIS 가구에너지패널조사 결과표 기준 주요 가전기기 통계	87

[표 3-22] 가전기기군 사용 패턴 정의	88
[표 3-23] 가전 부하 모듈의 단계별 구축	89
[표 3-24] 점유 인원별 스케일값의 설정	92
[표 3-25] EnergyPlus 입력 객체 항목	94
[표 3-26] EnergyPlus 입력 객체 검증 절차	96
[표 3-27] RSM-VSM Loadshape 비교	113
[표 3-28] Location 별 월간 평균 전력량 kWh 비교	114
[표 3-29] RSM 데이터 기록 충실도 검증	117
[표 3-30] RSM-VSM 데이터 연간-시간별 비교	119
[표 3-31] RSM-VSM 데이터 계절별-시간별 비교	120
[표 3-32] RSM-VSM 데이터 정합성 비교	121
[표 3-33] 계절별-주간별 통계량 데이터 비교	123
[표 3-34] 분류된 전체 조합의 상관관계 분석	127
[표 3-35] 지역-아키타입-계절-주별 통계량 값 비교	130
[표 3-36] 샘플링 캘리브레이션 수행	133
[표 4-1] 204D 주요 변수별 성능 평가 결과	145
[표 4-2] Class 별 회귀 예측 정확도	147
[표 4-3] Class 별 Confusion Matrix	148
[표 4-4] 204D Feature 구조	155
[표 4-5] 주요 Class의 선정	156
[표 4-6] 변수 간 Classification 의 속성 분류	162
[표 4-7] Area 변수의 분류 성능 비교	166
[표 4-8] BuildingType 변수의 분류 성능 비교	167
[표 4-9] Overall_RSI 변수의 분류 성능 비교	167
[표 4-10] Leakage 변수의 분류 성능 비교	168
[표 4-11] k-fold 분류기 성능	171
[표 4-12] k-fold 검증 결과	173
[표 4-13] 조합에 따른 LDA 분류 정확도 요약	174
[표 4-14] Elastic Calibration 전후 분류 성능 비교	176

[표 4-15] Elastic Calibration 전후 회귀 성능 비교	177
[표 4-16] RSM-VSM 분포 확률 비교	178
[표 4-17] RSM-VSM 변수 간 Box Plot 비교	181
[표 4-18] Class 별 구조적 요인	189
[표 4-19] posterior_gap Extractor 신뢰도 기준	194
[표 4-20] AirClass·YearClass의 연속형 변환 규칙	196
[표 4-21] DR 세그먼트 요약 Table	200
[표 4-22] Area·Year DR 세그먼트 분석	202
[표 4-23] Segment 별 DR 반응 분석	204
[표 4-24] DR Segment 핵심 통계량 값 결과	208
[표 4-25] DR 고우선순위 집단의 통계 결과	209
[표 4-26] DR Top-Share-Cut	209
[표 4-27] 연구 성과의 학문적 유용성 평가	217

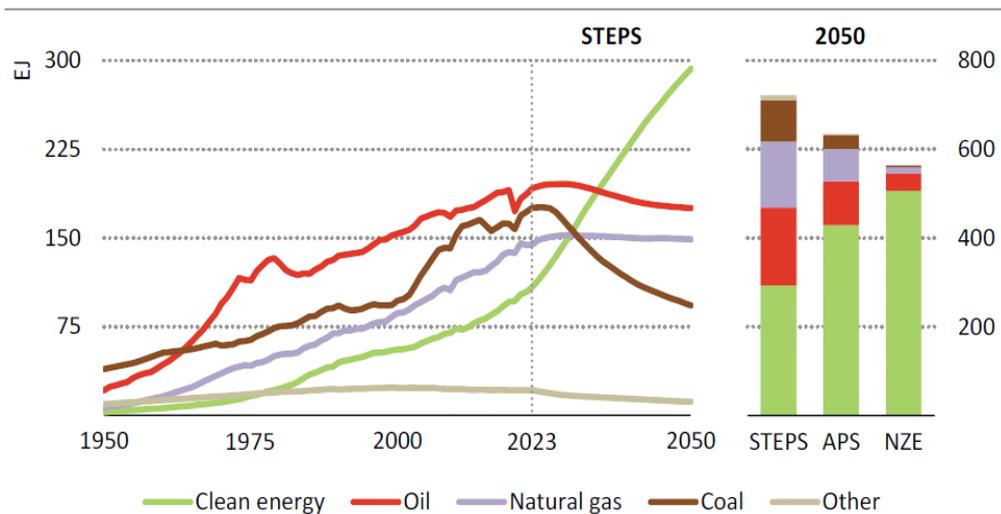
그림 목 차

[그림 1-1] 글로벌 에너지 시나리오 2050	1
[그림 2-1] QSBSEM의 연구 단계	26
[그림 3-1] 연구 모델	31
[그림 3-2] EnergyPlus-Python 병렬 제어 파이프라인	53
[그림 3-3] EnergyPlus-Python 명령어 처리	58
[그림 3-4] 난수 기반 입력값 생성 절차	63
[그림 3-5] RSM 데이터의 일별-계절별 전력 사용 패턴	69
[그림 3-6] 하계(6~9월) 가정용 전기 소비 분해 및 냉방용 비중	90
[그림 3-7] RSM 데이터 1년 데이터 집계	115
[그림 3-8] Archetype 분포도	125
[그림 3-9] Calibration Process	132
[그림 3-10] Elastic Scaling 보정(before, 1차,2차)	141
[그림 3-11] Elastic Scaling Process	142
[그림 4-1] Segment 단위별 DR 비율	206

I. 서론

1.1 연구의 배경 및 필요성

전 세계적으로 기후변화 대응과 에너지 전환이 가속화되고 있는 국제적인 상황에서 국제에너지기구(IEA, 2024)에 따르면, 2023년 기준 건물 부문의 최종 에너지 소비는 125EJ로 전체 에너지 소비의 28%를 차지하는 것으로 발표되었는데¹⁾, 이는 건물 부문의 에너지 절감과 온실가스 감축이 기후변화 대응, 에너지 안보 강화, 경제성장 촉진에서 핵심적인 위치를 차지하고 있음을 보여준다²⁾.



[그림 1-1] 글로벌 에너지 시나리오 2050³⁾

- 1) Wiatros-Motyka, M., Fulghum, N., Jones, D., Altieri, K., Black, R., Broadbent, H., Bruce-Lockhart, C., Ewen, M., MacDonald, P., & Rangelova, K. (2024). Global electricity review 2024. Ember.
- 2) 김말희, 허태욱, 이일우. (2023). 에너지 관리 기술의 미래 발전방향과 전망에 대한 연구. 『정보와 통신』, 한국통신학회, 40(12), 3-10.
- 3) Birol F. (2024). World Energy Outlook 2024. Paris: IEA. www.iea.org, International Energy Agency (IEA).

[그림 1-1]에서와 2030년 이후 정책 강도에 따라 세계 에너지 수요의 총량은 지속적으로 증가하지만 2050 탄소중립의 이행에 따라 재생에너지를 중심으로 한 청정에너지가 지배적인 에너지로의 전환을 예측하고 있다⁴⁾. 이러한 글로벌 에너지 전환 전망은 국가별 에너지 소비 구조가 주거 형태에 따라 상이한 양상으로 나타나고 있음을 보여주고 있다. 특히 주택 부문에서는 전력과 열에너지의 사용 비중 그리고 난방 방식의 선택이 에너지 전환의 핵심 변수로 작용하고 있음을 보여주고 있다.

한국 가정의 용도별 에너지 소비 비중을 보면 도시가스 49%, 전기 30.9%, 석유 9.9%의 순으로 조사되었다. 2022년 기준 가구당 연간 에너지 소비는 9,317.2Mcal로 매년 전기의 비중이 증가하고 추세로 통계되고 있다⁵⁾. 이러한 조사는 국내 주택 유형이 공동주택을 중심으로 지역난방이나 중앙난방, 그리고 가스를 사용하는 개별난방이 주류를 이루는 특성을 나타내고 있는 것을 확인할 수 있다⁶⁾.

이와 별도로 최근 연구 결과에 따르면 공동주택의 전기에너지 사용의 경우 냉방, 난방, 온수, 취사 순으로 에너지가 소비되는 것으로 확인되었는데 이 중에서도 냉방기기의 전력 사용량은 연평균 14.1%의 증가율로, 일반전기 6.4%보다 높은 증가를 보이고 있는데, 이는 지구온난화로 인한 폭염, 열대야, 도시 열섬현상으로 인한 냉방기기 가동의 증가, 그리고 일반 가전기기의 대형화로 인한 영향인 것으로 추정할 수 있다⁷⁾. 이는 공동주택에서의 냉·난방기기와 전기에너지 사용에 대한 에너지 소비 관리가 우선적인 과제를 보여주고 있다⁸⁾.

특히 구축된 공동주택이 주류를 이루는 대한민국의 경우 데이터 기반 또는 물리 기반의 건물 에너지 모델링 시뮬레이션 방법을 통한 전기에너지 사용의 예측과 실제 전력이 검침되는 스마트미터 데이터 분석은 에너지 관련 정책 수립과 기술 도입의 필수 기반이 되고 있다⁹⁾.

4) Birol F., (2024). 전계서, p.24.

5) 남수현, 이현. (2023). 『가구에너지패널조사 연구』. 한국에너지경제연구원(KESIS). Retrieved from https://kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=view.

6) 조사관리국 인구총조사과. (2025). 『2024년 인구주택총조사 결과』. 통계청.

7) 이성근. (2022). 가정부문 용도별 에너지 소비량 및 소급 추정에 관한 연구. 『에너지경제연구원』. ISBN 978-89-5505-272-6.

8) 김유란, 홍원화, 서윤규, 전규엽. (2011) 공동주택 가족구성원별 전력 소비 성향에 관한 연구. 한국 주거학회 논문집. 22(6) 43-50

선행 연구에서 캐나다 퀘벡주의 단독주택을 대상으로 가상스마트미터 (VSM, Virtual Smart Meter) 데이터셋을 생성하고, 이를 이용해 건물의 물리 파라미터를 추정·분류하는 모델을 구축하였는데, 이 접근법은 가상 데이터를 활용함으로써 실측 데이터 부족 문제를 해소하면서도, 물리 기반 시뮬레이션과 건물 에너지 변수의 확률 분포를 결합하여 실제 건물의 에너지 사용 특성을 반영하고 있음을 보여주었다¹⁰⁾.

1.1.1 전 세계 스마트미터 보급 현황 및 활용

전력 소비 데이터를 검침하는 스마트미터(Smart Meter)는 전력 사용량 데이터를 전력 계측과 양방향 통신 기능을 통하여 계량 주기로 데이터를 제공하는 차세대 계량 인프라(Advanced Metering Infrastructure, AMI)로, 미국·유럽·호주 등 선진국을 중심으로 보급이 시작되었다¹¹⁾. EU 회원국의 스마트미터 보급률은 2021년 말 기준 54%이며, 2024년 말까지 77%까지 보급될 것으로 예측하고 있다¹²⁾. 미국의 경우 스마트미터 보급이 대도시의 경우 90%를 상회하지만 텍사스의 경우 5.7%로 각각의 주마다 보급률의 편차가 크다¹³⁾. 이러한 스마트 인프라 확산은 실시간 요금제, 피크 감축 프로그램, 수요 반응 등 전력수요 정책의 실행 등의 핵심적인 기반 기술을 제공한다¹⁴⁾.

북미와 유럽 일부 국가는 연구·산업계에 일정 수준의 비식별화된 VSM 데이터셋을 국가 차원에서, 또는 전력회사, 대학교 등에서 연구 개발을 목적으

9) 이루다, 강은호, 윤종호, 김동수, 김진휘, 원종연. (2023). 한국형 공동주택 에너지 시뮬레이션 모델링을 위한 건물 형태 및 에너지 사용 특성 분석. 『한국태양에너지학회 논문집』. 43(4), 43-54. <https://doi.org/10.7836/kses.2023.43.4.043>.

10) Neale, A. (2021). Development of a Bottom-up White-box Residential Building Stock Energy Model. PhD Dissertation, Polytechnique Montréal.

11) 기술신용보증기금. (n.d.). 『전략 제품 현황 분석: 스마트미터(AMI)』. 전략제품현황 보고서. Retrieved from https://tb.kibo.or.kr/ktbs/board/tech-trend/tech_trend.do?mode=download&articleNo=563&attachNo=1441.

12) 김영욱. (2023). “EU의 스마트미터 보급정책 및 활용 현황 분석. 정책 분석”. 『전기저널』, Retrieved from <http://www.keaj.kr/news/articleView.html?idxno=5265>.

13) 기술신용보증기금. (n.d.). 『전략 제품 현황 분석: 스마트미터(AMI)』. 전략제품현황 보고서. Retrieved from https://tb.kibo.or.kr/ktbs/board/tech-trend/tech_trend.do?mode=download&articleNo=563&attachNo=1441.

14) 강수진. (2024). “민·관수 보급 사업 종료된 AMI 다음 스텝은… 다양한 서비스 ‘클래버레이션.’ 『일렉트릭타임즈』. Retrieved from <https://www.electimes.com/news/articleView.html?idxno=337390>.

로 다양한 공개 사례를 확인할 수 있는데, 이는 국가별·데이터 활용 수준에 큰 차이가 있음을 보여주고 있다. 이와 관련하여 캐나다에서 선행 연구에서 제시한 건물 스톡의 물리 기반 시뮬레이션과 확률모델을 결합하여 생성된 VSM 데이터셋의 생성 모델인 QSFBSM(Quick Service Form Bottom-up Stock Energy Model)¹⁵⁾은 건물유형과 시나리오에 대한 고해상도 부하 프로파일을 생성하고 이를 통해 국내 공동주택의 에너지 예측 모델을 공동주택 중심의 한국형 VSM 데이터셋으로 제시가 가능한 연구 모델이다.

1.1.2 국내 공동주택의 적용모델

국내 주택 유형은 공동주택 비중이 전체 주거 형태의 약 79%를 차지하고 있다¹⁶⁾. 아파트를 중심으로 하는 공동주택은 지역난방, 중앙난방, 개별난방, 전기·가스 혼합 사용 등 국내의 도시 중심의 특화된 에너지 사용 구조와 함께 주거용 건물의 전력 사용량이 지속적으로 증가되는 특성을 보이고 있다¹⁷⁾.

이는 에너지 사용 패턴이 단독주택 중심의 북미 지역과 다르고 유럽 및 북미 중심으로 발전해 온 건물 스톡 에너지 모델(BSEM, Building Stock Energy Model)을 국내에 직접 적용하기에는 어려움이 있다¹⁸⁾19).

대표적인 물리 기반 에너지 시뮬레이션 툴의 종류로 TRNSYS, EnergyPlus, eQUEST, CREST, ESP-r 등을 기반으로 국내 건물 에너지 관리 시스템(BEMS, Building Energy Model System)에 적용되고 있지만, Archetype 모델 입력 정보가 국제적으로 표준화된 ASHRAE나 IECC(International Energy Conservation Code, 국제 에너지 절약 기준) 기준 정보로 반영되어 공동주택 중심의 국내 에너지 사용에 대한 예측과 실측의 오차가 평균 2배 이상의 오차로 지적되고 있다²⁰⁾.

15) Neale, A., (2021), 전계논문, p.127.

16) 조사관리국 인구총조사과, 전계서, p.54.

17) 강진경, 김성임, 진혜선, 최보혜, 임재한, 송승영. (2016) 공동주택 세대단위의 연간 전기기기 에너지사용 특성 분석. 『대한설비공학회』, 하계학술대회 발표 논문집, 2016(6), 693-696.

18) 이윤석. (2022). 건물에너지관리시스템: 에너지 절감수요에 맞춘 기술융합 및 틈새시장 개척. 『ASTI Market Insight』. 한국과학기술정보연구원(KISTI). 2022(036), 01-06. Retrieved from <https://repository.kisti.re.kr/>.

19) 이루다, 강은호, 윤종호, 김동수, 김진휘, 원종연., 전계논문, p.44.

이러한 전력 사용량 예측 모델은 전력 사용량 데이터는 AMI를 통해서 이루어지는데 AMI는 전력회사와 수용가 간의 양방향 데이터 통신을 통한 전기 사용 데이터를 실시간으로 제공하고 전기 사용에 대한 제어가 가능한 전자식 스마트 계량기 기술이다²¹⁾. AMI 기술은 최적의 스마트그리드 운용을 위해 필요한 에너지 네트워크 인프라로 하드웨어, 소프트웨어, 서비스로 구분되는 계량 시스템으로, 수요 관리, 요금 설계, 전력 피크 제어 등 다양한 전력 정책에 활용되고 에너지 서비스 플랫폼으로 확장가능한 기반 기술이다²²⁾.

그러나 국내 주택용 요금제는 전력의 누적 사용량에 따라 요금을 부과하는 누진제 요금을 기본으로 채택하고 있다. 한국의 특수한 상황으로 도입된 주택용 전기요금 누진제도가 처음 도입된 것은 1974년 12월 7일이다. 주택용 누진제는 1970년대 제1차 석유파동으로 유가가 급등함에 따라 전기 절약을 도모하고 저소득층의 전기요금 경감을 목표로 도입되었는데, 처음 3단계 구간은 1.6배의 누진배율을 적용하였으며 현재는 6단계 누진제 구간과 11.7배의 누진배율을 적용하고 있으며²³⁾, 주택용(저압)과 주택용(고압)으로 나누어 적용되고 있다.

국내 공동주택의 경우 누진제 요금 기반으로 설계된 주택용(고압) 방식으로 한국전력공사에서 가로등을 포함하는 일부 공용전기는 계시별 요금이 적용되고 있고²⁴⁾, 개별 수용가의 전력 사용에 대해서는 아파트 단지 내 변압기까지의 메인 계량기만 데이터를 검침하고 개별 수용가의 전력 사용량은 아파트 관리사무소가 개별 가구의 전력 사용량과 공용전기 사용량을 세대수로 나누어 이를 합산하여 관리비로 정산하는 방식을 채택하고 있다²⁵⁾²⁶⁾.

이는 공동주택의 세대별 전력 사용에 대한 데이터가 아파트 관리사무소별로 개별적으로 관리되고 있고, 통합적으로 데이터 취합이 이루어지지 않는 구조적 특성을 보이고 있다. 이러한 상황은 한국형 공동주택에 특화된 BSEM과

20) 이루다, 강은호, 윤종호, 김동수, 김진휘, 원종연, 전계논문, p.44.

21) 강수진, 전계서, p.1-6.

22) 기술신용보증기금, 전계서, p.9-12.

23) 조성진, 윤태연. (2016). 주택용 전력 수요 계절별 패턴 분석과 시사점. 『정책 이슈 페이퍼』. 16-11. 에너지경제연구원.

24) 정재용, 김동식, 정범진, 정영모. (2021). AMI 데이터 기반 주택용 전력 누진 및 계시별 요금제 분석. 『전자공학회논문지』, 58(9), 66-74. <https://doi.org/10.5573/ieie.2021.58.9.66>

25) 정재용, 김동식, 정범진, 정영모. 상계논문, p.862

26) 한상미. (2023). AMI 데이터를 활용한 날씨 요인이 아파트 전력소비에 미치는 영향 분석. 숭실대학교 박사 학위 청구 논문

VSM 데이터셋 개발의 필요성을 더욱 확인할 수 있으며, 본 연구를 통해 개발된 한국형 공동주택의 VSM 데이터셋은 다양한 수요 관리 프로그램(DR, Demand Response)의 최적화된 수요예측 모델의 검증과 관련된 기초 데이터로 활용이 가능할 것이다²⁷⁾.

1.1.3 시뮬레이션 기반 VSM 데이터셋의 필요성

공동주택을 중심으로 하는 한국형 VSM 데이터셋은 이상에서 살펴본 바와 같이 실측 데이터 부족의 문제, 건물의 물리적 특성 반영의 한계, 모델 학습을 위한 데이터셋의 부족, 향후 주택용 요금제로 확산 예정인 계시 요금제에 대비한 DR 적용성의 확대 등으로 [표 1-1]과 같이 그 필요성을 요약할 수 있다.

[표 1-1] VSM 데이터셋의 필요성

필요성	대응책
실측데이터 부족 문제해결 ²⁸⁾	개인 정보 보호, 데이터 수집 한계로 인해 접근이 어려운 실측 AMI 데이터를 대체 가능.
건물 특성 반영 가능 ²⁹⁾	단열 수준, 난방 방식, 창호 유형 등 다양한 물리 파라미터를 조합하여 시나리오별 데이터 생성 가능.
모델 기반의 학습용 데이터셋 확보 ³⁰⁾³¹⁾	머신러닝 및 딥러닝 모델 학습에 필요한 대규모 레이블 데이터 제공 가능.
DR 적용성 확대 ³²⁾	수요 반응 이벤트, 날씨 변화, 요금제 변경 등에 따른 부하 변화에 대한 가상 실험 데이터 제공 가능.

27) 이윤경. (2016). 국내·외 수요자원시장 동향과 시사점. 『ENERGY FOCUS』. 동향초점, 2016년 여름호, 에너지경제연구원. 34-60.

28) 강성구, 김신규. (2014). 보안 침해사고 대응을 위한 스마트미터 로그 생성 및 수집 방안. 『정보보호학회지』. 24(5). 66-74.

29) Neale, A., (2021), 전계논문, p.47-48.

30) Neale, A., Kummert, M., & Bernier, M. (2019). Linear discriminant analysis for classification of a large virtual smart meter data set with known building parameters. Energy and Buildings, Proceedings of the 16th IBPSA Conference Rome, Italy, Sept. 2-4, 2019.

31) 김동식 정영모, 정범진.(2022). 딥러닝 기반 주택용 계시별요금제 전환 시 전기요금 변화 예측. 『대한전기학회』, 하계학술대회논문집. 2022(7), 1837-1838.

32) 김말희, 하태욱, 이일우, 전계서, p.5-7.

한국형 공동주택의 VSM 데이터셋 모델 개발을 위하여 선행 연구에서 제시한 QSBSEM의 Generator-Classifier-Extractor 연구 모델을 적용하면, 첫째 확률 분포 기반 입력 생성(Generator), 둘째, 부하 프로파일로부터 건물 특성 분류(Classifier), 셋째, AMI에서 검침된 실측 데이터(RSM, Real Smart Meter)로부터 건물 파라미터의 역추정(Extractor)이 가능하다. 이를 한국형 공동주택의 BSEM에 적용하면, 단순 부하 예측을 넘어 DR 운영의 최적화 검증에 위한 기초 데이터로 활용이 가능하다.

1.2 연구 목적

1.2.1 한국형 공동주택의 BSEM-VSM 모델 구축 가능성

국내 공동주택 중심의 주거 형태는 기존 북미·유럽 중심의 단독주택 위주로 개발된 VSM 모델을 그대로 적용하기 어렵다³³⁾. 해외 선행 연구에서는 TRNSYS-CREST를 활용한 상향식(Bottom-up) BSEM 시뮬레이션으로 수십만 건의 가상 전력 부하 프로파일을 생성하고 이를 선형 판별 분석(LDA, Linear Discriminant Analysis)의 기계학습 기법 분석으로 건물 특성을 추론하고 이를 바탕으로 전력 사용량을 예측하는 시나리오 검증에 활용하였다³⁴⁾.

그러나 캐나다 및 북유럽 건물의 물리적 특성과 난방·냉방 방식, 연면적, 창면적비(WWR, Window-to-Wall Ratio), 세대별 부하 밀도, 거주패턴 등에서 한국의 공동주택과 차이가 있다³⁵⁾.

선행 연구에서 천이시스템 시뮬레이션 프로그램(TRNSYS, Transient System Simulation Program)을 기반³⁶⁾으로 난방·온수·조명·가전 부하를 영국 Loughborough University의 재생 에너지 시스템 기술센터(CREST, Centre

33) 김재춘, 노광현. (2025). 건물스톡에너지모델 기반 수요반응모델 설계를 위한 가상스마트미터 데이터셋 분석. 『조명·전기설비학회논문지』, 39(5), 355-364. <https://doi.org/10.5207/JIEE.2025.39.5.355-364>.

34) Neale, A., Kummert, M., & Bernier, M. (2022). Discriminant analysis classification of residential electricity smart meter data. *Energy and Buildings*, 262, 111980. <https://doi.org/10.1016/j.enbuild.2021.111823>.

35) 김재춘, 노광현., 상계논문, p.359-361.

36) Solar Energy Laboratory; TRANSSOLAR; CSTB; TESS. (2017). TRNSYS 18: A Transient System Simulation Program-Updates in Version 18. Madison, WI: University of Wisconsin.

for Renewable Energy Systems Technology)의 점유 스케줄과 결합하여³⁷⁾, 계절별·시간별 부하 패턴을 정밀하게 생성하였으며 이를 AMI의 실측 데이터와 비교하여 정확도를 향상시켰다³⁸⁾. 이러한 연구 모델을 기반으로 국내 공동주택의 특성에 부합하도록 세대별 냉방 시스템, 중앙난방·지역난방·개별난방 시스템, 건물 열 손실, 세대별 부하 특성을 반영한 한국형 KR-VSM 모델은, 해외 모델 대비 국내 상황을 현실적으로 반영하게 되어 한국형 DR 프로그램의 적용 가능성을 높일 수 있을 것이다³⁹⁾.

따라서 본 연구의 목적은 한국의 공동주택을 중심으로 하는 주거 형태에 초점을 맞춘 KR-VSM 데이터셋 모델을 제시하고, 이를 기반으로 향후 가상발전사업자가 수요 관리 운용 전략을 수립하는 데이터 기반의 의사결정을 지원하는 기초 데이터셋을 제공하는 것이다.

1.2.2 확률 변수 기반의 입력 구조화

선행 연구의 방법론으로 VSM 데이터셋의 품질은 초기 입력 변수의 다양성과 확률 분포의 적절성에 크게 의존하게 되며, 건물·설비·운영 패턴을 확률 질량 함수(PMF, Probability Mass Function)와 균일 확률 분포(UPD, Uniform Probability Distribution)로 정의하여, 전체 건물 스톡의 통계적 특성을 반영한 무작위 샘플링을 가능하게 했다. 이는 모든 주택 유형을 동일 파라미터로 가정하는 Top-down 방식 대비 세밀하고 현실성 있는 결과를 산출한다⁴⁰⁾.

한국형 모델에서는 선행 연구를 참조하여 통계청 통계누리 및 인구주택총조사, 국토교통부 K-APT 공동주택관리 정보시스템, (재)한국스마트그리드사업단의 AMI 데이터 유통 플랫폼, 기상청의 국가 기후 데이터센터 기상 자료 개방 포털의 전국 기상 데이터, 미국에너지부(DOE, United States Department

37) McCarron, B. (2017). Passive House Research & Development Centre (CREST): How to meet net zero carbon with Passivhaus. PSouth Pacific Passive House Conference (SPPHC 2017), Christchurch, New Zealand.

38) Neale, A., Kummert, M., & Bernier, M. (2018). Generator: A stochastic virtual smart meter data generation model for residential building stock characterization. Proceedings of eSim 2018 (IBPSA-Canada), 65-72. ISBN 978-2-921145-88-6.

39) 이윤경, 전게서, p.59.

40) Neale, A., (2021), 전게논문, p.51, p.53.

of Energy)에서 개발한 EnergyPlus용 기상 데이터인 epw(EnergyPlus Weather) 데이터를 기초로 하였다. PMF는 건물의 위치, 건축연도, 건물 형태, 기밀도, 냉·난방 방식, 가전기기 등 이산형 특성으로 반영하고, UPD는 전용면적, 창면적비 등 연속형 변수로 반영하며 200,000세대의 건물 특성을 무작위로 생성하고, 점유자 수, 점유 프로파일, 조명, 온수, 가전 등 CREST의 점유·행동 패턴에 대한 시뮬레이션과 결합하여 계절별·시간별 전력수요를 산출하였으며 이는 선행연구의 QSFBSM의 파라미터 생성 기준⁴¹⁾을 한국형 KR-VSM 데이터셋의 생성규칙으로 적용하였다.

이러한 확률 변수의 구조화는 시뮬레이션 결과의 통계적 대표성을 높이며, 변수 민감도 분석을 통해 개별 수용가의 수요예측 모델을 정량화하고 나아가 동일 지역 내 건물 특성의 변동성을 모델에 반영함으로써 에너지 수요예측의 불확실성을 정량적으로 평가하는 기반을 마련하고자 하였다⁴²⁾.

1.2.3 RSM 데이터와의 비교 검증

VSM 기반 예측 모델의 신뢰성을 확보하기 위해서는 RSM 데이터와의 정량적 비교 검증이 필수적이다. 캐나다·유럽의 선행 연구에서는 공개된 스마트미터 데이터를 활용하여 VSM-RSM 간 부하 패턴의 상관성, 평균제곱오차(RMSE, Root Mean Square Error), 평균 절대 오차(MAE, Mean Absolute Error), 평균 편향 오차(MBE, Mean Bias Error)를 분석하였다⁴³⁾.

국내 기초 데이터로 (재)한국스마트그리드사업단의 AMI 데이터 유통 플랫폼에서 확보된 비정형화된 897,322가구의 1시간 단위 검침 된 1년간의 실측 데이터를 활용하였다⁴⁴⁾. 시뮬레이션 결과로 생성되는 KR-VSM 데이터셋과 실측치 간의 정확도를 시간별·일별·월별·계절별·주별로 검증하고 이를 통해 VSM 모델의 부하 패턴 왜곡, 피크 시점 오차, 부하량 과대·과소 추정 등의 구조적 오류를 식별하고 VSM 데이터의 정확도를 보정한다⁴⁵⁾. 이러한 비교

41) Neale, A., (2018), 전계논문, p.68.

42) Neale, A., (2021), 전계논문, p.65.

43) 김재춘, 노광현., 전계논문, p.362.

44) 전력 AMI 데이터 유통 플랫폼. (n.d). “시간대별 전력 사용량 기상결합 AI 합승용 데이터”. (재)한국스마트그리드사업단. Retrieved from <https://www.energydata.co.kr> & <https://www.smartgrid.or.kr/>.

검증 절차를 통해, 모델의 신뢰도를 제고하고 실증 기반의 보정 기법을 마련함으로써, 향후 가상 발전사업자의 DR 운영 전략 수립 시 의사결정의 부가적인 데이터로 활용이 가능하다.

1.3 선행 연구와 비교 및 연구의 차별성

1.3.1 선행 연구의 비교

QSFBSSEM의 VSM 데이터셋 연구는 북반구의 극지역에 위치한 지리적 차이로 인해 난방·온수 중심의 에너지 수요 패턴을 주요 변수로 선정하여 캐나다의 주거용 건물 특성에 맞게 건물 스톡 모델을 대상으로 TRNSYS와 CREST를 결합한 Bottom-up White-box VSM 데이터 생성 기법을 제시하였으며, LDA 머신러닝 기법으로 부하 패턴을 분류하고 건물 특성을 추론하는 선도적 시도를 수행하였다⁴⁶).

[표 1-2]에서와 같이 캐나다는 평균적으로 북위 50~60도 이상에 위치한 한랭기후 지역으로 기후·주거 환경 특성상 난방 부하가 전력 수요 패턴의 지배적 요소를 차지하고 있다⁴⁷). 이에 비해 북위 33~37도에 위치한 한반도는 온대와 아열대 기후가 혼재하는 환경으로 사계절⁴⁸)이 뚜렷하고 여름철의 고온다습한 환경이 6월~9월까지 길게 이어지면서 하계 냉방 전력수요가 전력 계통에 큰 부담을 주는 환경이라는 점에서 차이가 있다⁴⁹).

또한 QSFBSSEM의 VSM 프레임워크는 건물 외피·설비·점유 패턴에 대한 확률 변수를 PMF/UPD로 정의하였으며, 입력 데이터의 통계 분포가 캐나다 통계청 및 캐나다 건물 데이터베이스, 영국 등 북유럽에서 공개된 TABULA 프로젝트는 유럽 20개국의 건물 스톡 주택 데이터셋을 기반으로 하고 있다⁵⁰). 이에 비해 한국의 경우 공동주택은 중앙난방 및 지역난방의 비중이 높

45) Neale, A., Kummert, M., & Bernier, M., (2018), 전계논문, p.73.

46) Neale, A., (2021), 전계논문, p.131.

47) Wikipedia. (2024). "Canada". Retrieved from <https://ko.wikipedia.org/wiki/캐나다>

48) 한국학중앙연구원. (n.d.). 계절 한국민족문화대백과사전 Retrieved from <https://encykorea.aks.ac.kr>.

49) Wikipedia. (2024). "대한민국". Retrieved from <https://ko.wikipedia.org/wiki/대한민국>

50) Loga, T., Stein, B., Diefenbach, N., & Born, R. (2016). TABULA building typologies in

고, 승강기, 가로등 등 공용 부하의 전력 사용이 존재하며, 세대별로 높은 냉방기기 보급률 등의 고유 특성이 선행 연구에 반영되지 않은 변수이다⁵¹⁾.

[표 1-2] 캐나다와 대한민국의 지리적, 환경적, 기후적 비교

항목	한국	캐나다
지리적 위치	동아시아, 한반도 위치	북아메리카 대륙 북부, 북극권과 인접
면적	약 100,000km ²	약 9,984,670km ²
인구	약 5,200만 명(2025년)	약 4,000만 명(2025년)
주거	71%가 공동주택 주거	60%가 단독주택 주거
기후	대륙성 기후 및 온대 기후, 여름 덥고 습함, 겨울 춥고 건조	북부: 한대 및 극지 기후, 남부: 냉대 및 온대 기후
강수량	연평균 1,000~1,500mm	연평균 500~1,500mm 지역별 다양
겨울 온도	평균 -5°C~3°C (12월~다음 해 2월)	북부 -30°C 이하, 남부 -10°C~0°C (11월~다음 해 4월)
여름 온도	평균 23°C~30°C	북부 10°C 이하, 남부 20°C~30°C(7월~8월)
주요 에너지	화석연료, 원자력, 가스, 신재생에너지	수력, 천연가스, 신재생에너지, 석유

본 연구는 이러한 변수의 차이를 극복하기 위해 AMI 실측 데이터를 기반으로 입력 변수의 통계 분포를 재정의하고, 여름·겨울 이중 피크 부하 구조를 반영한 한국형 VSM을 구축하였다.

1.3.2 선행 연구와의 차별성

선행 연구로 Neale et. al.(2020)이 제시한 QSBSEM 연구는 4만 개 규모의 캐나다 단독주택에 대해 건물 파라미터가 완전하게 통제된 환경을 구축하고, 해당 데이터를 기반으로 LDA 분류기를 학습시켜 건물 속성 분류 정확도를 평가하였다. 즉, QSBSEM에서는 VSM 데이터셋의 통계적 타당성 검증과

20 European countries—Making energy-related features of residential building stocks comparable. Energy and Buildings, 132, 4–12.
51) 정재용, 김동식, 정범진, 정영모, 전계논문, p.861.

Classifier의 분류 성능 평가라는 두 영역에 집중⁵²⁾되어 있으며, RSM을 직접 DR에 적용하거나 가상 발전 사업자(Virtual Power Plant, VPP)의 운영 정책 분석은 연구 범위에서 포함되어 있지 않다.

이에 비해 본 연구는 QSFBSM의 이론적 구조를 한국 공동주택의 특성에 맞추어 재설계하고, VSM 생성-〉 Classifier 학습-〉 Extractor 역추정-〉 DR·VPP 테스트베드 운영으로 이어지는 End-to-End 통합 파이프라인을 구축한 실증 연구라는 점에서 선행 연구와 다음과 같은 근본적인 차별성으로 구별된다.

첫째, 데이터 생성 단계의 구조적 차별성이다. 본 연구는 국토교통부와 통계청의 공동주택 현황, 지역별 기후 지표, 단열·창호·누기 등 건물 성능 기준, 계단식·복도식·혼합식의 한국형 공동주택 구조, 전용면적 및 건축연도 분포, 가전기기 보급 통계 등을 종합적으로 반영하여 총 20만 세대 규모의 고해상도 한국형 공동주택의 KR-VSM을 구축하였다. 이는 선행 연구인 QSFBSM의 단독주택 중심의 연구보다 건물 구조·통계·규제 요소를 복합적으로 반영한 데이터 생성 체계로, 한국의 공동주택 스톡을 정밀하게 모사한 점에서 질적 도약을 구현하였다.

둘째, Classifier의 활용 목적과 설계 수준을 확장하였다. QSFBSM의 선행 연구가 VSM 내부에서 분류 정확도를 정량화하는 데 그쳤다면, 본 연구는 연간·계절·월·주·일의 시계열 기준의 고차원 204D 특징 벡터 체계를 설계하여 비선형 머신러닝 기반 Classifier를 구축하였고, 이를 RSM 데이터에 전이시켜 분류 안정성을 직접 검증하였다. 특히 본 연구는 posterior_gap을 활용한 분류 신뢰도 기반 DR 타겟팅 지표를 제안함으로써, 선행 연구가 다루지 못한 현장에서의 DR 정책 활용 가능성이라는 새로운 응용 영역을 개척하였다.

셋째, 한국형 역추정 규칙 체계를 적용하여 Extractor를 구축하였다. 본 연구는 Hard/Soft mapping, Semantic re-mapping, Area_m2_hat, Year_hat의 연속형 변수 추정을 포함하는 한국형 역추정 규칙 체계를 구축하여, RSM 데이터에서 결측되어 있는 건물 속성을 안정적으로 복원할 수 있음을 실증하였다. 이는 QSFBSM의 구조적 공백을 보완하며, VSM 데이터셋에-〉 실측 데이터

52) Neale, A., (2020), 전계논문, p.39.

를 적용하는 선순환 구조를 완성한 첫 번째 사례로 평가될 수 있다.

넷째, DR 세그먼트 정의와 함께 실측 기반의 DR 반응성 분석으로 확장하였다. 본 연구는 Heating - Insulation - Area - Location으로 분류된 총 16개의 건물 속성 조합을 DR 세그먼트로 설정하고, 20,000세대의 RSM에 이를 적용하여 DR 시나리오 분석을 수행하였다. 그 결과 전체 세대 중 약 2% 수준의 고반응 세그먼트가 전체 감축량의 9%를 차지하는 집중 구조(heavy-tail distribution)를 확인하였으며, posterior 기반 신뢰도와 DR 반응률 사이의 높은 상관관계를 규명하여 속성 기반 DR 타겟팅의 정책적 효율성을 실증적으로 분석하였으며, 이는 기존의 고부하 중심의 DR 참여자 선정 방식의 한계를 극복하는 새로운 방법론을 제시하였다.

마지막으로, 본 연구는 DR 참여자 선정 알고리즘, 세대별 DR 응답 확률 모델링, DR 잠재량 기반 자원 배분, VPP 운영 전략 설계를 포괄하는 테스트 베드를 구축하였다. 이는 QSFBSM 모델 검증 수준을 넘어, 가상 발전사업자가 실무적으로 활용 가능한 분석 인프라까지 구현하였다는 점에서 매우 큰 차별성을 지닌다.

즉, 본 연구는 QSFBSM을 단순한 VSM 생성·검증 도구를 넘어서, 한국형 스마트미터 기반 DR·VPP 운영 분석 플랫폼으로 확장하여 VSM 생성→Classifier→Extractor→DR·VPP 운영 전략으로 이어지는 End-to-End 시스템을 구축한 실증 연구라는 점에서 선행 연구와 명확한 차별성이 있다.

1.3.3 한국형 VSM 구축의 기술적·정책적 의의

한국형 VSM 구축은 다음과 같은 중요한 의미를 가진다. 먼저 기술적으로 국내 공동주택 건물 스톡에 특화된 BSEM과 확률 기반 PMF, UPD를 결합하여, 실측 데이터 부족 문제 해결과 기후와 부하가 연동된 시나리오를 생성할 수 있다⁵³).

다음, 정책적으로는 한국형 VSM은 국가 전력 수급 계획과 분산 자원 보급 정책에 활용할 수 있는 실증 기반 데이터를 제공할 수 있다⁵⁴. 국내 주택용

53) 이경호, 주홍진, 안영섭, 이왕제. (2024). TRNSYS 환경에서 모델 단순화와 학습을 통한 월별 연간 에너지 성능 평가: 태양광 적용 사무용 공공건물의 히트펌프 냉·난방. 『한국건축친환경설비학회논문집』. 18(3), 192-210. <https://doi.org/10.22696/jkiaeb.20240017>.

전력시장은 누진제 기반 주택요금제로 인한 기술적, 비용적 한계로 DR 확산에 제한적이다⁵⁵⁾⁵⁶⁾. 이에 대한 보완으로 VSM을 통한 검증은 DR 참여 유효성을 사전에 평가하고 향후 적용 예정인 주택용 계시 요금제 전환에 대한 정책적 리스크를 평가하는 부가적 자료로 활용이 가능하다⁵⁷⁾.

또한, 한국형 VSM은 정부가 추진 중인 2050 탄소중립 시나리오와 2030 국가 온실가스 감축 목표⁵⁸⁾에 부합하도록 공동주택의 에너지 수요 저감 잠재량을 정량화하는 건물 에너지 효율화 정책과 분산 에너지 자원 확대 정책에 활용될 수 있을 것이다⁵⁹⁾.

결국, 한국형 VSM 구축은 단순한 데이터 생성 기술을 넘어, 한국의 에너지 전환 정책과 생태계를 구축하는 데 학문적 가치가 크다고 할 수 있을 것이다.

54) 홍석재, 임정택, 함경선, 김태형. (2023). 『공동주택단지 내 BTM 에너지 공유서비스 경제성 평가 시스템 설계와 구현』. 정보과학회 컴퓨팅의 실제, 29(1), 15-24. <https://doi.org/10.5626/KITCP.2023.29.1.15>.

55) 김재춘, 노광현., 전계논문, p.359-361.

56) 심규한, 윤동섭, 김산, 주성관. (2019). 주택용 계시별 요금제 도입을 위한 전력 소비 패턴 분석. 『대한전기학회』. 전력기술부문회 추계학술대회 논문집, 2019(10). 271-272.

57) 김재춘, 노광현., 전계논문, p.362.

58) 수요정책실. (2025). 『2025 KEA 에너지 편람』. 한국에너지공단.

59) 전력산업정책과. (2025). 『제11차 전력수급기본계획(2024-2038)』. 산업통상자원부

II. 이론적 배경 및 선행 연구 분석

2.1 건물 스톡 에너지 모델의 개념

2.1.1 Top-down vs Bottom-up 비교

BSEM(건물 스톡 에너지 모델)은 국가 또는 지역 단위 건물 에너지 소비량을 정량적으로 분석·예측하기 위한 방법론으로, 크게 하향식(Top-down)과 상향식(Bottom-up) 접근법으로 구분된다⁶⁰. Top-down 방식은 국가 또는 도시 차원의 총 에너지 소비 통계 자료를 거시경제 지표, 기후 데이터, 인구·산업 구조와 같은 거시변수와 결합해 회귀분석 또는 계량적 모델링을 수행하는 하향식 방법이다⁶¹. 이 방식은 대규모 데이터 기반으로 신속한 분석이 가능하고 장기적 경향을 파악하는 데 유리하지만, 개별 건물의 물리적 특성을 세밀하게 반영하기에는 한계가 있다⁶².

이와 대조적으로 Bottom-up 방식은 개별 건물 수준에서 에너지 소비량을 계산한 후 이를 전체 건물 재고로 확장하는 상향식 방식이다⁶³. 이는 각 건물의 특성을 반영하는 건물 Archetype을 정의하고 물리 기반 시뮬레이션과 통계 모델링을 활용하는 것을 포함한다. Bottom-up 방식은 설계 및 운영 시나리오, 에너지 효율 정책, 수요 대응 전략의 효과를 상세하게 평가할 수 있다는 장점이 있다⁶⁴⁶⁵. 특히 TRNSYS 및 EnergyPlus와 같은 시뮬레이션 도구를 사용하면 기후 조건, 설비 효율, 점유 패턴, 열 손실 구조와 같은 다차원적 요소를 고려한 예측이 가능하지만, 방대한 양의 입력 데이터가 필요하

60) Neale, A., (2021), 전계논문, p.18.

61) Neale, A., (2021), 전계논문, p.19.

62) Heiple, S., & Sailor, D. J. (2008). Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. *Energy and Buildings*, 40(8), 1426–1436.

63) 강진경, 김성업, 진혜선, 최보혜, 임재한, 송승영., 전계논문, p.695.

64) Neale, A., (2021), 전계논문, p.19.

65) Heiple, S., & Sailor, D. J., 상계논문, p.1428.

며, 대규모 시뮬레이션에는 상당한 계산 자원과 시간이 필요하다⁶⁶⁾.

최근 연구 동향은 Top-down과 Bottom-up을 결합한 하이브리드 모델로 발전하고 있다. 예를 들어, Top-down 접근법은 국가 차원의 에너지 소비 추세를 파악할 수 있고, Bottom-up 접근법은 개별 건물유형에 대한 상세 분석을 수행한 후 두 결과를 보정할 수 있다. 이러한 접근법은 국가 정책 수립에 필요한 거시적 신뢰성과 기술의 적용 효과 분석에 필요한 미시적 정확성을 동시에 달성할 수 있다는 점에서 주목을 받고 있다. 특히 수요예측에서는 미시적 정확성을 보장할 수 있는 Bottom-up 방법이 건물별 제약 조건을 모델링하는 데 유리하며, 그 결과를 Bottom-up 공급 및 DR과 연계하는 하이브리드 방법이 제안되었다⁶⁷⁾.

2.1.2 White-box 기반 EnergyPlus의 역할

White-box 시뮬레이션은 건물이 실제로 어떻게 에너지를 사용하는지를 물리법칙으로 계산하는 모델로 단순히 과거 데이터를 기준으로 학습하고 입력과 출력을 기준으로 기계가 예측하는 모델인 Black-box 모델과 달리, 벽체 두께, 단열재, 창호 성능의 외피 성능, 온도, 기온, 풍속 등 기상 데이터, 냉·난방 기기의 설비효율, 건물 점유자 수를 기준으로 점유 패턴에 따른 온도, 가전기기, 냉·난방기기 사용 등의 입력을 받아 상호작용을 해석해냄으로써, 구조·운영 변경에 대한 인과적 분석이 가능하다⁶⁸⁾. 이러한 에너지 시뮬레이션 모델로 국제적으로 활용되는 엔진의 활용 비율은 [표 2-1]에서와 같으며 이 중에서도 특히 EnergyPlus가 가장 많이 활용되고 있는 것을 확인할 수 있다.

QSFBSSEM은 TRNSYS를 활용하여 변수 분석을 수행하였지만 TRNSYS는 BEMS 모델을 분석하는 셰어웨어 프로그램으로 광범위한 추가 구성요소를 유상으로 각각의 라이브러리를 통해 에너지 시뮬레이션을 수행하는 전문성이 높은 프로그램이다⁶⁹⁾.

66) Neale, A., (2021), 전계논문, p.38, p.87., p.132.

67) Heiple, S., & Sailor, D. J., 전계논문, p.1428.

68) Heiple, S., & Sailor, D. J., 전계논문, p.1428.

69) Solar Energy Laboratory, 전계서, p.1-4, p.2-6, p.3-7, p.5-27.

이러한 전문성이 높은 TRNSYS 프로그램 특성상 연구자가 접근하기 어려운 부분을 고려하여 DOE에서 지원하고 있는 오픈소스 소프트웨어 시뮬레이션 엔진으로 EnergyPlus를 시뮬레이션 엔진 프로그램으로 선정하였다.

[표 2-1] 논문에 검토된 시뮬레이션 엔진 비율⁷⁰⁾⁷¹⁾

Simulation Engine	Type	% of Papers Reviewed
EnergyPlus	White-Box	60%
TRNSYS	White-Box	15%
DOE-2		7%
Resistance-Capacitance	Gray-Box	10%
Other	N/A	8%

EnergyPlus는 OpenStudio, Eppy, Geomeppy 등 분석 도구가 풍부하고 국제적으로 검증된 엔진이면서 Python Script와 호환성이 높으며, [표 2-2]에서와 같이 다양한 해석의 장점을 가지고 있다.

[표 2-2] White-box 기반 EnergyPlus 해석의 장점

항목	장점
기상 시나리오 적용 용이성	epw에 예측·시나리오 값을 반영하면 온도·습도·일사 변화에 따른 냉·난방·온수·플러그 부하의 변화를 시간 해상도로 정밀 추정할 수 있어 기후 변화 대응 분석에 적합 ⁷²⁾⁷³⁾
건축물 설계·리트로핏의 정량화가 가능	단열 성능(벽/지붕/바닥 RSI) 개선, 창호 사양(U/SHGC) 변경, 보일러·히트펌프 교체, 재생에너지 도입 등의 조치가 부하 곡선과 조명/플러그/온수/난방/냉방 등 엔드유즈 소비에 미치는 영향을 체계적으로 비교 가능 ⁷⁴⁾⁷⁵⁾
운영·제어 ⁷⁶⁾ 시뮬레이션 가능	EMS와 스케줄 교체만으로 DR 이벤트, VPP 연계 제약, ESS 충·방전 전략, 피크 관리를 모사해 절감량과 쾌적성 변화를 평가 가능

70) Neale, A., (2021), 전계논문, p.14.

71) Chong, A., Gu, Y. & Jia, H. (2021). Calibrating building energy simulation models: A review of the basics to guide future work. Energy & Buildings, 253.

72) 조재운, 이영생, 송호승, 최영애, 조태희. (2023). 전력기상지수를 활용한 공동주택의 단기 전력

EnergyPlus는 EMS(Energy Management System) 및 Schedule:File을 통해 외부 생성 스케줄을 주입할 수 있어, CREST로 생성한 점유, 가전, 조명, 온수 등의 사용 패턴을 1시간 단위 해상도로 직접 연결하여 부하 프로파일을 현실적으로 반영할 수 있다⁷⁷⁾.

EnergyPlus는 시뮬레이션 변수 각각을 연구자가 직접 정의하고 정의된 변수를 기반으로 건물의 단열 구조, 가전기기, 열전달 및 복사 등 물리 법칙 등을 모든 수식과 물리 모델로 시뮬레이션을 수행하는 White-Box 모델로 건물 물리 파라미터와 함께 표준 기상 데이터 epw를 기반으로 기상 시나리오의 적용이 가능하며, 이를 통해 건축물 설계의 정량화, 그리고 운영 및 제어에 대하여 시뮬레이션을 수행한다는 장점을 가지고 있다⁷⁸⁾.

기존 Black-box 모델은 입력과 출력만으로 예측하는 시스템으로 데이터 패턴을 찾아내는 방식과 달리 White-box 모델은 입력 데이터의 품질과 보정을 통해 결과물의 정확성이 달라진다는 점에서도 차이가 있다.

본 연구에서는 TRNSYS-BSEM을 이용한 선행 연구의 단순화 모델 학습과 예측 성능 분석을 NMBE와 RMSE로 수행한 것을 기준으로 EnergyPlus-BSEM으로 생성된 White-box 기반으로 생성된 VSM 결과는 RSM 데이터와의 비교를 통해 r, RMSE, MAE, MBE의 지표로 검증 및 보정함으로써 신뢰도를 확보한다는 원칙을 유지하였다⁷⁹⁾. EnergyPlus는 물리 기반 해석을 통해 설계부터 정책 수준의 시나리오를 일관되게 연결하여 생성한다는 측면에서 한국형 KR-VSM 백본으로서 TRNSYS가 수행한 White-box를 충분히 대체할 수 있는 플랫폼으로 본 연구에 적용하였다.

수요 예측. 『대한전기학회』 2023 하계학술대회 논문집, 2023(7), 1828-1830.

73) 융합기상. (n.d) "6.1 전력기상지수 조회서비스". 기상청 API 허브, 기상청. Retrieved from <https://apihub.kma.go.kr>.

74) 이상현. (2014). "공동주택의 에너지 소비량 실태 조사를 통한 예측 모형 제시에 관한 연구". 중앙대학교 대학원 석사학위논문.

75) 송용현. (2025). 『에너지 수요 전망 모형 구축』. 국회예산정책처 연구용역보고서, 사단법인 넥스트.

76) 기술신용보증기금, 전계서, p.47, p.49.

77) Chong, A., Gu, Y. & Jia, H., 전계논문, p.6.

78) U.S. Department of Energy. (2019). Input for Output - Variable Dictionary Reports: EnergyPlus 9.1 Input Output Reference. Retrieved from Big Ladder Software. <https://bigladdersoftware.com/epx/docs/9-1/input-output-reference/input-for-output.html#variable-dictionary-reports>.

79) Chong, A., Gu, Y. & Jia, H., 전계논문, p.14.

2.1.3 Building Archetype 정의

건물 원형(아키타입, Archetype)은 대규모 건물 스톡(Building Stock)의 에너지 소비를 예측할 때, 각 개별 건물을 모사하지 않고도 전체 건물 집합의 특성을 단순화하여 대표할 수 있도록 설계된 전형적 템플릿이다. 이는 BSEM의 구조적 복잡성을 줄이면서도 대표성과 계산 효율성을 확보하기 위한 핵심 개념으로, 유럽에서는 Building Typology라는 용어와 동의어로 사용된다⁸⁰⁾.

Archetype 접근법의 목적은 건물 스톡 내의 물리적·기능적 다양성을 일정 수준의 추상화로 통합함으로써, 제한된 수의 대표 건물유형만으로 다양한 건물군의 에너지 사용 특성을 모사하는 것이다.

이를 위해서는 세분화(Segmentation)와 특성화(Characterization)의 두 단계가 필수적으로 수행된다. 일반적으로 Segmentation(세분화)는 건물을 물리적·기능적 속성에 따라 범주화하는 과정으로 건축연도, 면적, 층고, 종횡비 등 건물의 형상, 기후, 건물용도, 건물유형, 벽체, 창호, WWR 등 건물 외피 파라미터와 그리고 HVAC, 온수, 조명 등 시스템 파라미터가 사용된다. 예를 들어, 캐나다와 같이 광범위한 기후대를 갖는 국가는 기후 특성을 세분화의 주요 기준으로 삼지만, 단일 기후 조건을 가진 도시 단위의 경우에는 형상이나 용도, 외피, 시스템 등이 더 중요한 분류 기준이 될 수 있다⁸¹⁾.

Characterization는 각 세분화된 범주에 대해 실제 물리적 속성값의 범위를 정의하는 단계다. 예를 들어, 기후 세분화의 경우, 난방도일(HDD, Heating Degree Days), 냉방도일(CDD, Cooling Degree Days)을 기반으로 지역적 특성을 정의하거나, 표준 또는 표본 조사 데이터를 통해 창문 및 벽체의 열관류율 성능을 보정한다. 이러한 특성화 과정을 통해 각 세그먼트는 단순한 경험적 평균이 아닌 실제 건물 데이터 또는 보정된 통계 값을 반영하여 실제 건물 스톡의 에너지 행태를 더욱 현실적으로 나타낼 수 있다. 잠재적 Archetype의 수는 세분화 매개변수 조합의 수에 따라 기하급수적으로 증가하므로 실제 연구에 존재하지 않는 조합을 제거하거나 소수의 대표적인 Archetype

80) Loga, T., Stein, B., Diefenbach, N., & Born, R., 전계논문, p.5.

81) Dahlström, L. (2023). Archetype identification in Urban Building Energy Modeling: Research gaps and method development. Uppsala University, Licentiate thesis.

으로 줄이는 방법을 사용한다⁸²⁾⁸³⁾. 이러한 최적화 과정은 계산 효율성과 현실성을 보장하는 BSEM의 구축에서 매우 중요한 절차이다.

Archetype의 정확도는 사용된 입력 데이터의 품질과 보정 기법에 크게 의존한다. 기존 연구들에서는 전통적인 결정론적 속성만으로는 실제 건물의 에너지 소비를 충분히 재현하기 어렵다고 지적하며, 이에 따라 베이지안 보정(Bayesian Calibration)과 같은 확률적 기법이 도입되었다. 미국 메사추세츠주의 케임브리지에 있는 2,263세대의 가구를 대상으로 Bayesian Calibration의 도입을 통해 모델의 오차를 연간 에너지 소비 기준으로 평균 오차가 40.9%에서 13.9%로 개선되었고 K-S 통계량도 38%에서 18%로 개선되었음을 보여주고 있다⁸⁴⁾. 이러한 결과는 Bottom-up으로 구성된 건물 에너지 모델이 연간 총 에너지 소비량에 대하여 에너지원별 1~19% 수준의 오차로 실측치를 재현하였으며, 24개 건물의 검증에서는 8~99%까지 건물별 오차가 분산되는 것이 확인되었다. 이는 개별 건물의 부하는 여전히 30~40% 수준의 차이를 확인함으로써 개별 시점의 부하 재현에는 불확실성이 존재함을 보고하고 있다⁸⁵⁾.

결과적으로, Building Archetype 접근법은 대규모 건물 스톡의 에너지 소비를 예측·평가하기 위한 효율적이고 대표성 있는 추상화 방법으로, 지역적 기후 조건, 건물 물리 특성, 시스템 구성, 점유 행태 등의 변수들을 통합하여 건물군의 전반적인 에너지 패턴을 모사한다⁸⁶⁾. TABULA 프로젝트는 이러한 Archetype 기반 분류 체계를 제도화하여, 유럽 각국의 주거 건물유형을 표준화된 Typology로 정의하였으며⁸⁷⁾, 이는 향후 AI 기반 BSEM 개발에서도 국가별·지역별 모델을 구축하는 핵심 기반으로 활용될 수 있다.

82) Dahlström, L. 전계논문, p.11.

83) Mata, E., Sasic Kalagasidis, A., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment*, 81, 270–282. <https://doi.org/10.1016/j.buildenv.2014.06.013>

84) Sokol, J., Cerezo Davila, C., & Reinhart, C. F. (2017). Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings* 134, 134, 11–24.

85) Heiple, S., & Sailor, D. J., 전계논문, p.13.

86) Dahlström, L., (2023), 전계논문, p.8

87) Reyna, J., Wilson, E., Parker, A., Satre-Meloy, A., Egerter, A., Bianchi, C., Praprost, M., Speake, A., Liu, L., Horsey, R., Dahlhausen, M., CaraDonna, C., & Rothgeb, S. (2023). U.S. building stock characterization study: A national typology for decarbonizing U.S. buildings. National Renewable Energy Laboratory (NREL).

2.2 가상스마트미터 데이터셋의 개념

2.2.1 VSM과 RSM 비교

AMI 기반 스마트미터는 전기 사용량을 실시간으로 계측하여 생성된 데이터를 양방향 통신을 통해 전력회사와 정보를 교환하는 전자식 계량장치다⁸⁸⁾. 주택용(저압) AMI는 한국전력을 중심으로 2010년부터 본격적으로 보급하기 시작하여 2024년 12월 기준으로 전체 주택에 보급이 완료된 것으로 발표되었다⁸⁹⁾. 다만 아파트를 중심으로 하는 주택용(고압)의 경우 AMI 장비가 주택 분양 시 건설회사에서 설치하여 개별 가구의 소유로 분류되는 민간 영역으로 국가 차원에서 스마트그리드 구축 사업, 가정용 스마트 전력 플랫폼 사업, LH에서 시행한 국민 DR 실증사업 등의 다양한 노력에도 불구하고 2024년 말 기준 40% 수준에 머물러 있으며⁹⁰⁾, 이러한 이유로 RSM 데이터는 관리사무소를 중심으로 개별적으로 관리되는 특성상 공공데이터로서 접근에는 분명한 한계가 있는 것이 현실이다.

한국형 공동주택 KR-VSM 데이터셋은 이러한 제약을 극복하기 위해 등장한 개념으로, EnergyPlus의 건물 물리 기반 시뮬레이션 엔진과 PMF, UPD 등 통계적 샘플링 변수를 결합하여 실제 건물 거주자 특성을 반영한 가상의 전력 사용 프로파일을 생성한다⁹¹⁾.

한국형 공동주택 KR-VSM 데이터는 기상 변화, 전력요금제, DR 이벤트 등 특정 시나리오를 자유롭게 설정할 수 있지만, 설정된 모델링의 가정, RSM 입력 데이터 품질, 기상자료 정확성 등에 따라 실제와의 오차가 발생할 수 있으므로, ASHRAE Guideline 14에서 제시하는 MAE, MBE, RMSE 지표를 활용한 RSM 보정 과정이 필수적이다⁹²⁾⁹³⁾.

88) 기술신용보증기금, 전계서, p.5.

89) 강수진, 전계서, p.1-6.

90) 이여름, 박영훈, 김광천. (2024). 아파트 내부 검침망을 활용한 AMI 활성화 방안 연구. 『한국산학기술학회 춘계 학술발표논문집』. 487-492.

91) Neale, A., (2021), 전계논문, p.65..

92) ASHRAE. (2014). ASHRAE Guideline 14-2014: Measurement of Energy, Demand, and Water Savings. Atlanta, GA: ASHRAE. ISSN 1049-894X

93) Chong, A., Gu, Y. & Jia, H., 전계논문, p.14.

2.2.2 VSM 데이터셋의 활용 분야

VSM 데이터셋은 첫째 DR에서 최대 부하 시간대의 부하 절감 가능성을 사전 예측하는 데 활용된다. 여름철 냉방 피크 시간대에 세대별 에어컨 설정 온도를 2°C 상향했을 때 예상되는 부하 감축량과 거주자의 불편감 지수를 산출하고, 이에 따른 DR 인센티브 수준과 참여율 변화를 시뮬레이션이 가능하다. RSM 데이터 접근이 제한적인 경우, VSM은 외기 조건, 요금제, 제어 전략 등 다양한 변수 조합을 적용하여 DR 잠재량을 정량화할 수 있다.

둘째로 분산 에너지자원(DER, Distributed Energy Resource) 중 하나인 ESS 운영 최적화에서 VSM 데이터는 계절별, 주별, 시간대별 부하 패턴을 제공하여 충·방전 스케줄링, 계시별 요금(TOU, Time of Use), 탄소 저감 효과 등을 사전에 검증하는 데 활용된다⁹⁴). 특히 태양광(PV, Photovoltaics) 발전량 예측과 결합하면 자가 소비율 향상, 피크 부하 완화, 전력 요금 절감 등의 경제성 효과의 사전 검증이 가능하다⁹⁵).

셋째로 VPP 운영전략 검증에서 VSM 데이터셋은 다수의 DER을 집합하여 전력시장에 입찰하거나 계통 지원 서비스를 제공하는 전략의 효과를 분석할 수 있다⁹⁶). 이는 DR+ESS+PV를 동시에 보유한 공동주택 단지를 하나의 VPP 자원으로 구성하고, 전력 가격 변동과 기상 변화에 따른 부하 예측 오차를 최소화함으로써 DER 운영 성능에 미치는 영향을 사전 시뮬레이션할 수 있다⁹⁷).

넷째로 VSM 데이터셋은 주거용 건물 및 에너지 시스템의 디지털 트윈 구축을 위한 기초 자료로 활용될 수 있다. 가상스마트미터를 통해 생성된 시간별 부하의 시계열 데이터와 건물 물리 속성은 실제 공동주택 단지의 에너지 사용 행태를 가상 공간에 재현하는 데 필요한 핵심 입력 정보를 제공할 수 있다. 이를 통해 점유 행태, 가전기기 운전에 대한 전략 변화에 따른 에너지 수요 반응을 사전에 모사함으로써 중·장기적인 투자와 이에 따른 정책 효과

94) 김은환. (2023). “수요반응시장(DR)의 환경 변화와 기회 요인”. 「전기저널」, 전기정보관, Retrieved from <http://www.keaj.kr/news/articleView.html?idxno=4900>.

95) Fernandopulle, J., Fang, Y., Datta, M., Nutkani, I., & Vahidnia, A. (2023). Planning and Operation of Community Microgrids: Technical, Economic, Policy and Regulatory Aspects. RMIT University.

96) Fernandopulle, J., Fang, Y., Datta, M., Nutkani, I., & Vahidnia, A., 상계논문, p.8.

97) Fernandopulle, J., Fang, Y., Datta, M., Nutkani, I., & Vahidnia, A., 상계논문, p.32.

분석에서도 그 활용 범위를 확장할 수 있다.

마지막으로 계통 운영, 기후변화 대응, 에너지 관련 정책 수립에도 활용이 가능하다. 재생 에너지 확대, DR 제도개선, 주택용 전기요금 구조 개편과 같은 정책 시뮬레이션뿐만 아니라, 장기 기후 변화 시나리오에 따른 부하 변화를 예측하고 이에 따른 시나리오 설계도 가능하다. 이처럼 VSM 데이터셋은 비용과 시간 부담을 최소화하면서도, 다양한 변수에 대한 실험이 가능한 유연성이 높은 데이터셋의 제공이 가능하다.

2.2.3 해외 VSM 구축 사례

해외 VSM 구축 사례 중 대표적인 예로는 [표 2-3]에서와 같이 캐나다 퀘벡주의 QSFBSSEM이 있다. QSFBSSEM은 TRNSYS Type 56과 CREST 모듈을 결합해 난방 방식, 창면적비, 단열 수준, 전용면적 구간, 가구원 수 등 19개 또는 24개 변수에 확률 분포를 적용하고, 200,000세대의 15분 간격의 고해상도 연간 부하 데이터셋을 생성하였다⁹⁸⁾.

TABULA 프로젝트는 유럽 20개국의 주거용 건물 스톡을 유형별로 세분화하고, 각 국가의 대표 Building Archetype을 정의하여 시뮬레이션 기반의 에너지 소비량을 예측하고 에너지 효율화 시나리오를 비교 분석하고, 이를 통해 국가별 건물 에너지 절감 잠재량을 도출하였다.

호주와 미국의 일부 연구에서는 PV, ESS, DR 통합 시뮬레이션을 위해 대규모 가상 부하 데이터를 생성하고, 이를 이용해 전력시장 참여 전략, 주파수 조정(FR, Frequency Regulation), 피크 부하 완화 효과 등을 사전에 검증하는 사례가 제시되고 있다⁹⁹⁾.

98) Neale, A., (2021), 전계논문, p.87.

99) Reyna, J., Wilson, E., Parker, A., Sattre-Meloy, A., Egerter, A., Bianchi, C., Praprost, M., Speake, A., Liu, L., Horsey, R., Dahlhausen, M., CaraDonna, C., & Rothgeb, S., 전계서, p.8, p.10, p.13.

[표 2-3] 글로벌 VSM 관련 선행 연구 조사¹⁰⁰⁾

구분	연구 지역 / 기관	주요 기술 구성	변수 범위	시뮬레이션 방식	데이터 규모	특징 및 활용
QSFBSSEM ¹⁰¹⁾	캐나다 퀘벡주 /Polytechnique Montréal (2020)	TRNSYS Type 56 + CREST 모델 결합	19개 파라미터 (건물형, 면적, 단열, 창면적비, 난방, 점유 패턴 등)	확률적(몬테카를로) VSM 생성	200,000세대, 15분 단위(1년)	난방·온수 중심 북미형 부하 모델, LDA·ML 기반 역추정 및 DR 검증에 활용
Canada Generator Model ¹⁰²⁾	캐나다 전국 주거 건물 스톡(2018)	CREST + TRNSYS 통합형 확률모델	24개 변수(점유·기후·단열 등)	확률적 시계열 생성	200,000세대 이상 확장 가능	난방 기후대별 Building Stock Energy Model (BSEM) 구현
TABULA Project ¹⁰³⁾	유럽 20개국 (EU Intelligent Energy Europe)	Building Typology DB + Simulation	건축연도, 형상, 용도, HVAC, 단열수준 등	Deterministic Archetype 기반	국가별 수천~수만 샘플	국가별 주거 건물의 에너지 특성 비교, 정책 평가 및 Retrofit 전략
Australia / US Studies ¹⁰⁴⁾	호주, 미국 CSIRO/NREL (2022)	EnergyPlus + PV/ESS/DR 통합 시뮬레이션	Load profile, PV, ESS, Price signal	하이브리드(Deterministic+Stochastic)	수천~수만 세대	미국 ResStock·ComStock 모델 참조
BuildingsBench (Buildings-900)	NeurIPS Datasets &	NREL End-Use	시뮬레이션 + 실측 혼합 (총	STLF(단기 부하 예측)용	시뮬레이션 90만 채	세계 최대 규모(15 B hours)의 공개 데이터셋, Zero-shot

100) Reyna, J., Wilson, E., Parker, A., Satre-Meloy, A., Egarter, A., Bianchi, C., Praprost, M., Speake, A., Liu, L., Horsey, R., Dahlhausen, M., CaraDonna, C., & Rothgeb, S., 전계서, p.8, p.10, p.13.

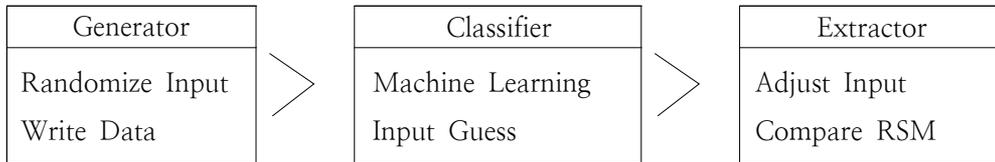
K) 105)	Benchmarks (2023)	Load Profiles (EULP) 기반	1900 real buildings 포함)	Transformer 사전 학습용 대규모 벤치마크	(주거용:55만, 상업용:35만)	및 Transfer Learning 벤치마크 제공
GAN-based Load Generator 106)	Wang&Hong, LBNL(2020)	Building Data Genome Project	실측 기반	K-means 군집 + GAN 부하 시계열 합성	(1,636 commercial bldgs., 3,053 meters)	GAN으로 실제 부하의 통계적 분포·랜덤성 재현 (KL < 0.3), 데이터 프라이버시 보호
Wavelet-Decomposed Synthetic ¹⁰⁷⁾	Sunetal., Applied Energy (2022)	영국 스마트미터 (1 min~30 min data)	실측 기반	Wavelet 변환 + 가우시안 확률 모로 부하 재현	주거용	실측 변동성(temporal variability) 보존 및 확률적 시나리오 생성

- 101) Neale, A., Kummert, M., & Bernier, M. (2020). Development of a stochastic virtual smart meter data set for a residential building stock – methodology and sample data. *Journal of Building Performance Simulation*, 13(5), 583–605.
- 102) Neale, A., Kummert, M., & Bernier, M., (2018), 전계논문, p.66.
- 103) Loga, T., Stein, B., Diefenbach, N., & Born, R., 전계논문, p.6–11.
- 104) U.S. Department of Energy. (2022). U.S. Building Stock Characterization Study: A National Typology for Decarbonizing U.S. Buildings. Washington, DC: U.S. DOE. NREL/TP-5500-83063. www.nrel.gov/publications
- 105) Emami, P., Sahu, A., & Graf, P. (2024). BuildingsBench: A large-scale dataset of 900K buildings and benchmark for short-term load forecasting. 37th Conference NeurIPS 2023 Datasets and Benchmarks Track, Preprint.
- 106) Wang, Z., & Hong, T. (2020). Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). *Energy Technologies Area 2020*, 224, 110299.
- 107) Claeys, R., Cleenwerck, R., Knockaert, J., & Desmet, J. (2023). Stochastic generation of residential load profiles with realistic variability based on wavelet-decomposed smart meter data. *Applied Energy*, 350, 121750.

2.3 건물 스톡 에너지 모델 기반 가상스마트미터 데이터셋 프레임워크 분석

2.3.1 Generator – Classifier – Extractor 구조

QSFBSSEM의 BSEM-VSM 프레임워크 핵심은 [그림 2-1]에서와 같이 Generator-Classifier-Extractor라는 3단계 구조이다.



[그림 2-1] QSFBSSEM의 연구 단계

첫째, Generator 단계에서는 건물의 물리 기반 White-box 시뮬레이션과 CREST 기반의 점유자·내부 부하 생성 모듈을 결합하여 대규모 VSM 프로파일을 생성한다. 입력은 건물 스톡의 연식, 단열 성능, 냉·난방 시스템, 거주 인원, 면적 등 통계적 분포이며, 각 변수는 PMF나 UPD로 정의되어 샘플링된다. 이 방식은 동일한 건물유형이라도 확률적 변동성을 반영할 수 있어, 단일 시뮬레이션 대비 현실성을 높인다.

둘째, Classifier 단계에서는 생성된 부하 데이터를 기반으로 특정 건물 파라미터를 유형별로 분류한다. 분류된 데이터는 LDA를 활용해 부하 패턴만으로도 건물의 규모, 난방 방식, 단열 수준 등을 높은 정확도로 식별할 수 있음을 보였다. 이 과정은 다음의 Extractor 단계에서 RSM 데이터와의 비교·보정에도 활용된다.

셋째, Extractor 단계에서는 분류·검증을 거친 데이터에서 핵심 변수와 패턴을 추출한다. AMI에서 검침된 RSM 데이터를 첫 번째 Generator 단계에서 생성된 VSM 데이터와 반복적으로 비교하였다. RSM-VSM의 비교 값이 오차 범위를 초과하면 입력 파라미터를 반복적으로 조정하여 오차 범위 내의 값으로 VSM 데이터셋에 기록하고, 오차 범위 내 값으로 매칭이 완료되면 건물 파라미터값을 최종적으로 출력한다¹⁰⁸⁾.

2.3.2 캐나다 단독주택 모델 변수 및 확률 분포 설계

QSFBSSEM은 캐나다를 중심으로 주요 파라미터를 [표 2-4]에서와 같이 정의하고, 각 파라미터에 대해 확률 분포를 설계하여 샘플링을 수행하였다.

[표 2-4] 캐나다 QSFBSSEM 입력 파라미터의 분류

구분	파라미터
건물 물리 특성	건물 유형, 건물 면적, 건물 형상비, 건물 층수, 회전각, 벽체 단열 성능, 지붕 단열 성능, 기초 단열 성능, 창호 유형, 창면적비, 침투율, 인접 건물 여부
시스템 특성	히트펌프 방식, 냉방 방식, 온수 시스템, HVAC 시스템, 조명 및 가전 부하
거주자 특성	점유 스케줄, 활동 스케줄, 가전/조명 사용 패턴, DHW 사용량, 수영장 및 스파 보유 여부
위치·환경 특성	기상 파일, 건물 연령, 지리적 특성,

먼저 PMF 변수는 지역별 건물 유형, 단열 성능, 건축연도별 비율 등 실제 건축물 통계 자료를 반영하며, 특히 단열 수준 및 열저항 계수(RSI)의 경우에는 연식별 건축법 기준과 지역적 특성을 반영하여 PMF로 정의하였으며 UPD는 건물 방향, 중횡비, WWR 등을 변수로 생성하였다.

이러한 방식은 QSFBSSEM에서 제시한 VSM 데이터셋 생성 원리를 따르며, 개별 주택의 물리적 특성, 시스템적 특성, 거주자 특성의 변수를 단일 평균값이 아닌 확률 분포 기반으로 샘플링 함으로써 동일 유형 내에서도 다양한 조합의 시나리오를 자동으로 생성함으로써 고정 입력 기반의 전통적 시뮬레이션 모델의 단일성 한계를 극복하고, 수십 개 변수 다변량 조합을 통해 200,000 세대 규모의 VSM 데이터를 생성하였다. 이렇게 생성된 데이터를 EnergyPlus-CREST 연계 시뮬레이션을 통해 전력 부하, 기상 데이터, 점유 패턴의 불확실성을 통계적으로 반영하여 구축함으로써 데이터의 다양성과 현실이 동시에 확보되었다¹⁰⁹⁾.

108) 김재춘, 노광현., 전계논문, p.358.

109) U.S. Department of Energy, (2019), 전계서, p.5-7.

2.3.3 CREST 점유 스케줄 및 내부 부하 생성 방식

CREST 모델은 VSM 데이터 생성에 중요한 점유 스케줄 및 내부 부하 생성 기능을 제공한다. 점유 스케줄은 각 가구 구성원의 활동 패턴을 분 단위까지 재현하며 통계청의 거주자 행동 패턴 조사를 기반으로 한다. 각 거주자의 하루 일정은 수면, 외출, 재택근무, 집안일, 여가 시간으로 세분화되며 주, 계절, 가구 구성원 수에 따라 다르게 설정된다.

내부 부하 생성은 점유 스케줄과 연계되어 가전제품, 조명, 조리, 온수 등 개별 부하의 사용 시간과 지속 시간을 확률적으로 결정한다. 즉, 조리 부하는 아침, 오후, 저녁에 집중되는 반면 가전제품 부하는 주말과 저녁에 더 자주 사용하게 된다. CREST는 이러한 부하를 전력 단위의 시계열로 출력하여 TRNSYS 시뮬레이션 도구와 연계하여 냉·난방 부하 분석을 수행할 수 있도록 한다. QSFBSSEM은 CREST의 세분화된 활동·부하 생성 로직을 활용해, 실제 캐나다 가구의 시간대별 전력 사용 패턴과 높은 유사성을 확보하였다. 특히, 주중/주말·계절별 차이, 장치별 부하 비중 등을 RSM 데이터와 비교해 보정함으로써, 가상 데이터의 현실성을 높였다¹¹⁰⁾.

2.4 한국 스마트미터 인프라의 현실과 제약

2.4.1 데이터 접근 제한

국내 전력 사용 데이터는 한국전력공사에서 인증받은 AMI 기기에서 수신되는 데이터로 AES128 알고리즘을 이용한 암호화 기능을 지원하며 계량 기능으로 유효전력/무효전력/피상전력(량) 8채널, 역률, 순시 값(순시 전력, 전압, 전류, 위상, THD 등), 수요전력의 측정 등으로 구성된다. 검침 기능으로는 정기/비정기/수동/원격 검침이 가능하며 MD(Maximum Demand, 최대수요전력), LP(Load Profile), TOU(Time of Use, 시간대별 구분 계량), CPP(Critical Peak Pricing, 최대 수요전력 관리요금제), OPP(Over Peak

110) 김재춘, 노광현, 전계논문, p.361-363.

Pricing, 계약전력 초과 요금제), RTP(Real Time Pricing, 실시간 요금제) 등의 적용이 가능하다¹¹¹⁾. 정·복전, 시간 변경, SWELL(순간 전압 상승), THD(Total Harmonic Distortion, 전체 고조파 왜곡률)의 발생 등에 대한 각종 이력을 저장하고 래치 릴레이를 제어 신호로 출력함으로써 전원 공급 차단이 가능하다¹¹²⁾

아파트를 중심으로 하는 주택용(고압)의 경우 민간 영역의 AMI 기기가 설치되어 있어 한국전력공사에서 인증받은 AMI 시스템과 달리 전송되는 데이터 포맷 구조와 전송 규격이 장비 제조사마다 상이하여 동일 지역, 동일 건물 간에도 시간 해상도와 측정 항목, 측정 단위 등이 상이하며 데이터 수집 시기도 일부 시스템은 15분 단위, 다른 시스템은 1시간 평균값만 제공하고 있다.

이러한 문제점을 인식하고 정부 주도로 산업통상자원부 산하 에너지기술평가원이 주관하여 2021년부터 주택 전력 에너지 사용량(AMI) 데이터 유통 플랫폼 및 서비스 개발 프로젝트를 재단법인 한국스마트그리드 사업단이 주관으로 국내 AMI 제조사인 누리플렉스, N2M, 씨앤유글로벌, 덕산, LS 일렉트릭이 참여하여 익명화 및 가명화 처리된 수용가의 시간별 전력 사용 데이터를 AMI 데이터 유통 플랫폼에서 데이터를 공개하고 있으며¹¹³⁾, 이는 차세대 에너지 서비스를 연구하고 개발하는 학교와 기업에 데이터 기반 연구 개발 활동에 기여할 수 있는 기반을 제공한다 할 수 있다.

2.4.3 제도·정책적 한계

한국의 스마트미터 인프라 확산은 ‘지능형 전력망 구축·운영에 관한 법률’과 정부의 AMI 보급 계획에 힘입어 빠르게 진행되었지만, 데이터 활용 정책과 제도적 지원은 이에 걸맞게 정비되지 못했다. 현재 AMI 데이터는 한국전력공사의 내부 운영과 요금 청구, 검침 자동화 목적에 집중되어 있다¹¹⁴⁾. 그러나 아파트를 중심으로 하는 주택용(고압)의 경우 아파트 관리사무소에서 개

111) 성장기술실 미래전략팀. (2011). 『수요반응 제도의 개념 및 현황』. 한국전력거래소.

112) 기술신용보증기금, 전계서, p.14.

113) 전력 AMI 데이터 유통 플랫폼, 전계서.

114) 강수진, 전계서, p.1-6.

별적으로 관리되고 있으며¹¹⁵⁾, 일부 단지는 보안과 해킹 등을 이유로 독립적으로 설치된 검침 PC가 외부 인터넷과 연결이 차단되어 관리 되고 아파트 관리비 청구 시에 검침 PC에 접속하는 관계로 검침일 이전까지 결측치, 이상치, 시간 동기화 오류 문제로 데이터 소실이 빈번하게 발생하고 있으며, 통신 장애로 인한 누락, 일부 장비는 펌웨어 버전 차이로 데이터 포맷이 변동되는 등 다양한 데이터 누락과 보안 침해사고를 대비하기 위하여 명확히 정의 된 로그 생성 기준이 요구되고 있기도 한다¹¹⁶⁾.

이는 민간 부분에서 건설회사에서 분양 시 설치되고 아파트 관리사무소에서 전문 인력이 없는 상태로 검침 PC가 서버 형태로 관리가 되고 있다. 특히 세대 단위로 설치된 AMI 단말기의 경우 그 수량이 많고 가정 내 설치되는 관계로 많은 위협원으로부터 접근이 용이하며 스마트그리드 환경에 존재하는 시스템 중 보안 침해사고가 발생할 가능성이 높다¹¹⁷⁾. 이러한 보안의 취약성과 통신 프로토콜의 표준화, 실시간 데이터 스트리밍의 표준화 등이 필요하다. 이러한 데이터의 취약성을 극복하고 안정된 AMI 검침 데이터의 확보를 통하여 공동주택의 계시별 요금제 적용에 대비한 AI 기반의 부하 예측 및 제어 전략 개발을 위한 대안으로 연구자는 데이터의 품질, 데이터 해상도, 시간 동기화를 자유롭게 통제할 수 있는 공동주택을 중심으로 하는 한국형 VSM 데이터셋을 구축할 필요성을 제기하고 있다¹¹⁸⁾.

115) 이여름, 박영훈, 김광천., 전계논문, p.487.

116) 강성구, 김신규., 전계논문, p.67.

117) 이여름, 박영훈, 김광천., 전계논문, p.73.

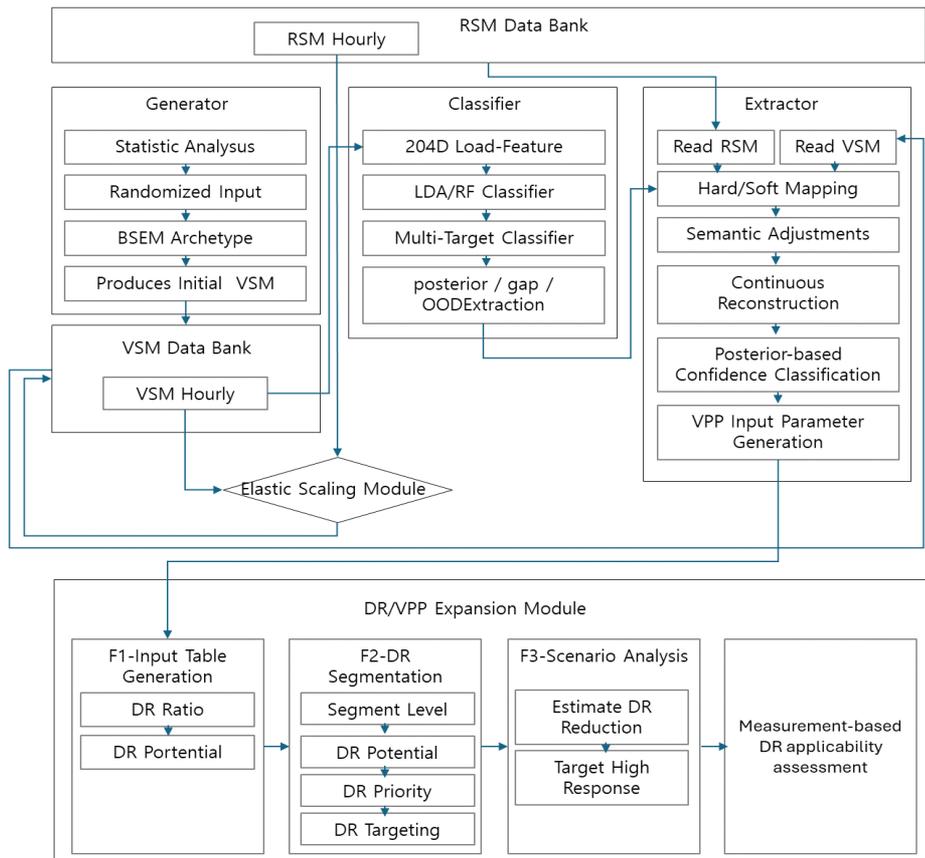
118) 김재춘, 노광현., 전계논문, p.363.

Ⅲ. 한국형 공동주택의 가상스마트미터 데이터셋 개발

3.1 연구 설계 및 입력 변수 구조화

3.1.1 연구 모델의 개요

본 연구는 한국형 가상스마트미터 데이터셋을 구축하기 위해 QSFBSM 선행 연구의 Generator-Classifier-Extractor 3단계 프레임워크를 [그림 3-1]과 같이 반영하고 외생 변수로는 epw를 주요 입력 변수로 설계하였다.



[그림 3-1] 연구 모델

연구 모델의 첫 번째 단계인 Generator 단계는 한국형 주거 건물의 물리·기후·운영 특성을 반영하여 VSM 시계열 데이터를 생성하는 과정이다. 이를 위해 건축연도, 난방 방식, 단열 수준, 전용면적, 세대 구성 등 국내 공동주택의 주요 건물 속성을 조합한 한국형 Archetype 체계를 구축하였다. 외기 조건은 기상청 ASOS(Automated Synoptic Observing System, 종관기상관측자료) 자료¹¹⁹⁾를 기반으로 지역별 epw 파일을 생성하여 EnergyPlus에 적용함으로써, 계절·기후에 따른 냉·난방 부하 변동을 정밀하게 모사하였다. 내부 부하 모델링을 위해 CREST 모듈을 한국 거주자의 행동 패턴을 반영하는 시간별 전력 부하 곡선을 생성하였다¹²⁰⁾. 이렇게 생성된 VSM 데이터는 RSM 데이터와 동일한 구조를 가지며, 이후 분류기(Classifier), 추출기(Extractor) 단계에서 모델 학습 및 보정에 사용되는 기본 시뮬레이션 데이터셋으로 제공한다.

두 번째 단계인 Classifier 단계는 Generator에서 생성된 VSM 시계열의 단열 수준, 난방 방식, 지역, 건축연도와 같은 건물의 물리적 속성을 역추정한다. 이를 위해 본 연구에서는 LDA와 Random Forest의 시계열 패턴 학습기반의 머신러닝 모델을 통하여 VSM 부하 패턴에서 건물 속성을 추론하는 역방향 매핑을 적용하였다. 이러한 과정을 통해 VSM과 RSM의 데이터가 모두 동일한 분류기를 적용할 수 있으며, 고해상도의 부하 패턴으로도 각 세대의 숨겨진 물리적 속성을 자동으로 추정할 수 있도록 구성하였다. 이러한 결과는 Generator 단계에서 PMF/UPD의 확률 기반 입력값을 검증함으로써 후속 Extractor 단계에서 정확한 매핑에 필요한 초기 건물 속성 정보를 제공하고자 하였다.

세 번째 단계인 Extractor 단계는 시뮬레이션 기반 VSM 및 RSM 데이터를 정렬하고 보정하는 단계이다. 먼저 두 데이터의 부하 패턴을 매핑하고 신뢰도를 산출하여 시간 축, 스케일 및 패턴의 차이를 정량화한 다음 VSM을 허용 오차 범위 내에서 수정하여 실제 RSM 특성과 일치하도록 보정한다. 보정이 완료되면 시계열화된 난방과 전용면적 파라미터의 최종 건물 속성이 결

119) 국가기후데이터센터. (2024). 『기상 종관 기상 관측(ASOS) 자료』, 기상청, 기상자료개방포털. Retrieved from <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do>.

120) 신이레, 윤상후. (2016). 전력 수요 예측을 위한 기상 정보 활용성 평가. 『한국데이터정보과학회지』. 27(6), 1601 – 1607. <http://dx.doi.org/10.7465/jkdi.2016.27.6.1601>.

정되고 신뢰성이 높은 한국형 공동주택의 KR-VSM 메타데이터가 출력된다. Extractor는 단일 실행이 아닌 Generator, Classifier 및 Extractor 간의 순환 구조로 설계되어 반복적인 수정을 통해 날씨 및 주거 행동의 변화에 적응할 수 있는 동적으로 업데이트되는 KR-VSM 데이터 세트를 제공한다.

다음 단계는 DR 확장 단계(F1)로, Extractor 단계에서 최종 산출된 건물 특성 및 실측 기반 부하 지표는 먼저 7개 광역군으로 지역을 분리하였다. 다음으로 세대 단위의 난방 방식, 단열 수준, 전용면적, 건축연도를 최종 추정된 물리 속성으로 확정하고 이들 파라미터의 곱으로 결합된 DR 세그먼트를 기반으로 연간 전력 사용량과 최대 전력량이 포함된 표준화된 입력 테이블로 변환된다.

다음으로 DR 세그먼트의 우선순위와 타킷팅을 산출하는 단계(F2)에서는 20,000세대 데이터를 16개의 DR 세그먼트로 집계하여 DR 세그먼트 단위 운영 파라미터를 연간 부하, 피크 부하, DR 평균 반응량, DR 참여 가능성 등 반응성 지표와 함께 세그먼트별 분류 신뢰도, 세그먼트 우선순위 조건이 산출된다. 마지막으로 시나리오 분석 단계(F3)에서는 F1·F2에서 구축된 세그먼트 파라미터를 활용하여, 다양한 VPP·DR 운영 시나리오를 구성하고 가상 발전사업자가 DR 발령을 통한 DR 감축 운영 효과를 정량적으로 평가한다.

이렇게 실행한 시뮬레이션 결과는 세그먼트별 DR 가능량, VPP 운영에 따른 피크 감소량, 참여 세대 변화에 따른 DR 효율성 및 비용 절감 효과 등으로 산출되며, 이는 실제 DR·VPP 사업 설계 및 정책 적용에 직접 활용될 수 있다.

3.1.2 공동주택 기반 Archetype 세분화

3.1.2.1 파라미터의 분류

한국형 VSM 데이터셋 구축의 출발점은 국내 주거 건물 스톡 중에서 대표적인 유형을 체계적으로 정의하는 것이다. 현재 우리나라 주택의 약 79%가 아파트 기반으로 하는 공동주택이며, 이러한 주거 형태는 건물 물리 특성과

설비 구성, 그리고 입주자의 생활패턴이 비교적 유사하다는 장점을 가진다. 그러나 건축연도, 난방 방식, 단열 수준, 세대 규모, 입지 특성에 따라 전력 소비 패턴의 변동성이 달라지므로, 시뮬레이션의 정확도를 높이기 위해서는 이러한 세부 특성을 반영한 Archetype 세분화가 필수적이며 이를 PMF, UPD 확률 변수로 [표 3-1]과 같이 21개의 파라미터를 정하였다.

[표 3-1] KR-VSM 파라미터의 분류

구분	변수	비고
PMF	Location, Year of construction, Building Type, Heating, Cooling	QL
	Occupants, Area, Leakage, Occupancy Activity, Window Type	
Hybrid (PMF+UPD)	Load Profile	QN
Stochastic	Lighting Density, Home Appliance, DHW	
UPD	WWR, Rotation, Aspect ratio,	
Conditional PMF	Wall RSI ¹⁾ (벽체 단열 수준), Roof RSI (지붕 단열), Foundation RSI (바닥 단열)	
Derived	Overall building thermal resistance category	
Fixed	Temperature, Humidity, Sunlight, Solar radiation	시계열 데이터

주) ¹⁾ RSI :Thermal Resistance Index (열저항 지수)

3.1.2.2 파라미터의 세부 유형

주요 파라미터의 개별 세부 유형은 [표 3-2]와 같이 개별 변수의 특성을 고려하여 관련 통계 데이터를 기반으로 주요 Class를 추출하여 서로 다른 세부 유형을 정의하였다. 총 파라미터는 19개로 정의하고 단열 조건을 세분화하면 21개 파라미터로 정하였으며, 각각의 파라미터는 Class를 정의하고 각각의 파라미터의 생성규칙에 대하여 건물 형태, 건축물의 단열, 기밀 등급 등 건축물의 속성을 나타내는 범주형 변수(Qualitative) 와 전용면적, 창면적비 등 구간 분포로 배치하는 정량형의 연속형 변수(Quantitative) 로 각각을 개별적으로 정의하였다.

[표 3-2] KR-VSM 아키타입별 세부 유형

Class name		#ca t	Class description	Type
Area		4	~40m ² , ~59m ² , ~99m ² , 100m ² ~	QN
Location		7	Regions in the province of 7 metropolitan province	QL
Year of construction		5	~1989,~1999, ~2009, ~2019, 2020~	QL
Building type		3	Type of residential building, e.g. Stair-type, Floor-type, mixed-type	QL
Heating		4	individual heating, district heating, and electric heating, None	QL
Cooling		4	wall-mounted and stand-alone, Inverter type, None	QL
Load Profile		7	Profile A1~A4, Profile BB01~BB07 Profile A-BB	QN
Window Type		24	P,L,E,F,S,B ; Code SPEC Types	QN
Leakage		6	Average air infiltration rates ; ~1,~2,~3,~4,~5,5~ [m ³ /hm ²]	QN
Occupants		4	1 to 4 occupants	QN
Occupancy Activity		28	(1 to 4 occupants) × 7 activity	QN
DHW		28	(1 to 4 occupants) × 7 activity	QN
Lighting Density (W/ m ² ·기준)		4	Type of lighting & density ; low, mid, high, super (~8, 8~11, 11~14, 14-20)	QN
Home Appliance		7	BB01-→Energy consuming, BB02-→Energy saving, BB03-→M-pattern, BB04-→Morning concentrated, BB05-→Evening concentrated, BB06-→Owl, BB07-→ 무작위 혼합	QN
WWR		5	Window-to-wall ratio ; 0.2,0.3,0.4,0.5,0.6	QN
Rotation		8	0,45,90,135,180,225,270,315 degrees	QN
Aspect ratio		5	0.8, 0.9, 1.0, 1.1, 1.2 Shape	QN
Conditional [m ² K/W]	Wall RSI	4	Vertical envelope thermal performance values [m ² K/W]	QN
	Roof RSI	6	Roof thermal performance values	QN
	Foundati on RSI	4	Basement/foundation thermal performance values	QN
Derived [m ² K/W]	Overall RSI	3	Overall building thermal resistance category <1.56, 1.56-2.25, >=2.25	QN

* cat: categories, QL: qualitative, QN: quantitative

이렇게 분류된 각각의 파라미터는 다음과 같이 변수별 Class를 구분하여 [표 3-3]과 같이 class에 각각의 값(Value)을 세분화하여 부여하였다.

[표 3-3] KR-VSM 파라미터의 Class 세분화

Class	Value	Description	Remark
Area	1.<40㎡, 1.<59㎡, 2.<99㎡, ≥100㎡~	아파트의 전용면적 기준 <40㎡, <59㎡, <99㎡, ≥100㎡으로 구분	QN
Location	1.SEOUL, 2.GYEPONGGI, 3.INCHEON, 4.CHOONGCHU NG, 5.JEONRA, 6.GYEUNGSANG 7.GANWON	대한민국을 광역도 기준으로 7개 광역도로 구분하고 epw의 경우 위도-경도 2024 인구주택총조사를 기준으로 선정.(제주도는 아파트 주거 비율이 31.4%로 제외)	QL
Year of construction	1.~1989, 2.~1999, 3.~2009, 4.~2019, 5.2020~	「건축법」 과 「에너지 절약 설계 기준」 제개정에 따른 분류로 총 5단계 분류	QL
Building Type	1.Stair type, 2.Floor type, 3.Mixed type	계단식, 복도식, 그리고 혼합형으로 구분	QL
Heating	1.Individual Heating 2.Central heating 3.District heating. 4.none	개별난방, 중앙난방, 지역난방으로 구분	QL
Cooling	1. wall-mounted 2. stand-alone 3. system types 4. none	벽걸이형, 스탠드형, 시스템형, 그리고 에어컨 미보유로 구분	QL
Load Profile	Load Profile -A ; 점유 인원 1~4명 기준	Load Profile-A ; A1~A4 Load Profile-BB: BB1~BB10 BB01: Basic.	QN

	Load Profile -BB: 07 Type 행태변이 적용 Load Profile A-BB : A+BB의 시계열 생성	BB02: Energy Saving BB03: M-Pattern. BB04: Morning Concentrated BB05: Evening Concentrated BB06: Owl BB07: 무작위 혼합.	
Window Type	P(1,2,3) L(0,1,2,3) E(0,1,2) F(0,1,2,3) S(0,1) B(0,1,2)	Panes(유리 층수);1=단층, 2=복층, 3=삼중 L(Low-E 유형);0=없음, 1=하드, 2=소프트, 3=복수 코팅 E(Ensure Gas);0=공기, 1=아르곤, 2=크립톤 F(Frame);0=비단열 알루미늄, 1=단열 알루미늄, 2=PVC, 3=목재 S(Spacer);0=알루미늄, 1=웍렛지 B(Bin-SHGC 구간);0:<0.30, 1:0.30~0.40, 2:>0.40	QN; 총24종 의 창문 Type
Leakage	1. ~1, 2.1~2, 3.2~3, 4.3~4, 5.4~5, 6.5~	국토교통부, 2025 에너지 절약형 친환경주택 건설기준 해설서 ¹²¹⁾	QN
Occupants	1.(1) 2.(2) 3.(3) 4.(4)	1 occupant 2 occupants 3 occupants 4 occupants~	QN
Occupancy Activity	1~28	(1 to 4 occupants) × 7 activity	QN
DHW	1~28	(1 to 4 occupants) × 7 activity	QN
Lighting Density	1. Low 2. Mid	Low:~8 Mid:8~11	QN

	3. High 4. Super	High:11-14 Super:14~20	
Home Appliance	1:1 BB01, 2:2 BB02, 3:3 BB03 4:4 BB04, 5:5 BB05, 6:6 BB06, 7:7 BB07	BB01→Energy consuming, BB02→Energy saving, BB03→M-pattern, BB04→Morning concentrated, BB05→Evening concentrated, BB06→Owl, BB07→무작위 혼합	QN
WWR	1: 0.2, 2: 0.3, 3: 0.4, 4: 0.5, 5: 0.6	0.2~0.6의 각 구간을 20(5%), 30(31%), 40(17%), 50(44%), 60(5%)로 차등 분포로 적용	QN
Rotation	1: 0° ,2: 45° ,3: 90° ,4:135° ,5: 180° ,6: 225° ,7: 270° ,8: 315°	Rotation of the building with respect to south (90° increments)	QN
Aspect ratio	1,2,3,4,5	1:0.8, 2:0.9, 3:1.0, 4:1.1, 5:1.2	QN

Building 파라미터는 한국에너지공단의 건축물의 에너지절약 설계기준 해설서(2023)와 건설교통부의 기계설비 기술기준 매뉴얼에서 기준값으로 제시된 열관류율(U-Value)의 값¹²²⁾을 EnergyPlus 입력 변수에 부합되게 RSI 값으로 전환하여 [표 3-4]와 같이 벽체, 지붕, 바닥 그리고 전체로 변수를 정의하였다¹²³⁾¹²⁴⁾. 여기서 정의된 열저항 값은 m^2K/W 으로 면적당 열저항 값이며 값이 클수록 단열 성능이 우수한 것을 나타낸다. QN의 연속형 변수이지만, 건축물의 건축연도, 건축물의 지역 등에 따라 건축물의 단열 성능이 달라질 수 있는 의존형 변수로 조건부 PMF 함수로 정의하여 벽체 단열, 지붕 단열, 바닥 단열, 그리고 전체 단열의 4가지 변수를 추출하여 정의하였다.

121) 주택건설공급과. (2023). 건축물의 에너지절약 설계기준 해설서. 국토교통부.

122) 김혜원. (2018). 에너지 절감 효과와 경제성을 고려한 국내 지역별 공동주택의 외피 성능 분석. 연세대학교 대학원, 건축공학과 석사학위논문

123) 건설산업과 국가건설기준센터. (2023). 『건축물의 에너지절약 설계기준 해설서(제2023-104호)』. 국토교통부-한국에너지공단.

124) 건설산업과. (2022). 『기계설비 기술기준 매뉴얼(최종본)』. 국토교통부.

[표 3-4] 빌딩 파라미터의 정의

Class		Value	Description	Remark
Conditional PMF	Wall thermal resistance (m ² K/W)	[1.0] (0.5,1.5) [2.0] (1.5,2.5) [3.0] (2.5,4.5) [5.0] (4.5,6.0)	Wall thermal resistance value. Values in square brackets represent the mean value for that bin.	QN
	Roof thermal resistance (m ² K/W)	[1.0] (0.5,1.5) [2.0] (1.5,2.5) [3.0] (2.5,3.5) [4.0] (3.5,4.5) [5.0] (4.5,5.5) [8.0] (5.5,8.0)	Roof thermal resistance value. Values in square brackets represent the mean value for that bin.	QN
	Foundation thermal resistance (m ² K/W)	[1.0](0.5,1.5), [2.0](1.5,2.5), [3.0](2.5,3.5), [4.0](3.5,4.5), [5.0](4.5,5.5)	Foundation thermal resistance value. Values in square brackets represent the mean value for that bin.	QN
Derived	Overall RSI (m ² K/W)	1: <1.56 2: <2.25 3: ≥2.25	Derived from the roof, wall and foundation thermal resistance values.	QN

외기 조건 파라미터는 기상청의 기상지상관측지침(2022)을 참고하여 [표 3-5]와 같이 변수를 정의하였다. 외기 변수가 건축물의 고유 속성이 아니라 시간적, 지역적, 외생적으로 결정되는 조건에 해당하기 때문에 모든 변수는 고정된 Fixed 변수로 정의하였다. 외기 조건의 변수는 VSM의 생성 과정에서 확률 조건으로 생성되지 않고 기상청이 실측 데이터를 기준으로 사전에 구분된 구간에 따라 결정되는 방식으로 정하였다. 여름철 피크 조건으로 온도 ≥ 30°C, 습도 ≥ 75%, 일사량 ≥ 0.9hr, 일조량 ≥ 2.0 MJ/m² 등으로 대표값으로 설정하였다¹²⁵⁾.

125) 관측정책과. (2022). 『지상기상관측지침』. 기상청.

[표 3-5] 외기 조건 파라미터의 정의

Class	Value	Description	Remark
Outdoor Temperature	1. <0°C 2. <18°C 3. <30°C 4. ≥30°C~	외기 온도 기준 4단계로 구성하였으며 기상청의 계절 분류에 대한 온도 데이터를 활용 ¹²⁶⁾	Fixed
Outdoor Humidity	1. <40% 2. <50% 3. <75% 4. ≥75%~	외기 습도 조건을 4단계로 기상청의 습도 데이터를 활용 ¹²⁷⁾	Fixed
Sunlight	1. <0.25hr 2. <0.5hr 3. <0.75hr 4. <0.9hr 5. ≥0.9hr~	외기 일조량을 4단계로 기상청의 일조[hr] 데이터를 활용 ¹²⁸⁾	Fixed
solar radiation	1. <0.5 2. <1.0 3. <2.0 4. ≥2.0	외기 일사량 조건을 4단계로 기상청의 일사[MJ/m ²] 데이터를 활용 ¹²⁹⁾	Fixed

3.1.3 Archetype 변수 설계

3.1.3.1 PMF/UPD 기반 확률 변수 설계

개별 파라미터가 정의된 후, 다양한 세대 특성을 모사하기 위해 주요 물리, 시스템, 거주자, 환경 등의 파라미터를 확률 변수로 설계하였다. 이를 위해 PMF와 UPD를 병행하며, 여기에 epw를 포함하는 주요 기상요소를 포함한 결합 변수를 추가하였다. 한국형 KR-VSM 데이터셋의 생성을 위해 변수 생성은 사전 확률로 표본화하는 PMF, 연속 분포에 기반한 UPD, 조건부의

126) 건설산업과, 전계서, p.23.
127) 관측정책과, 전계서, p.38, p.213.
128) 관측정책과, 전계서, p.43, p.154.
129) 관측정책과, 전계서, p.161, p.248.

Conditional PMF, 확률적 내부 부하, 파생 및 외생 변수로 구분하였다. 각 변수의 값 정의와 구간은 QSFBSSEM을 기본으로 준용하되, 한국 공동주택의 특성을 반영해 의존 관계를 명시하였다.

1) 범주형 PMF 변수

PMF 변수는 Area, Location, Year of construction, Building Type, Heating, Cooling, Window Type, Leakage를 사전 분포에 따라 표본화 하였다. Window Type은 QSFBSSEM의 Hot2000code에서 정의된 24종 표준 코드와 동일한 규칙으로 [표 3-6]과 같이 적용하였다. 문자열 P-L-E-F-S-B는 각각 (㉠) Panes: 유리 층수(1/2/3), (㉡) Low-E: 코팅 유형(0=없음, 1=하드, 2=소프트, 3=멀티코팅), (㉢) Enclosure: 가스 충전 기체(0=공기, 1=아르곤, 2=크립톤), (㉣) Frame: 창틀 재질·단열 수준(0=비단열 Al, 1=단열 Al, 2=PVC, 3=목재), (㉤) Spacer: 간봉 유형(0=Al, 1=웜 엣지), (㉥) Bin-SHGC: 태양열 취득률 구간(0/1/2)으로 매핑하여, 각 자릿수가 상호 직교하도록 구성하여 기본 라이브러리는 $3 \times 4 \times 3 \times 4 \times 2 \times 3 = 864$ 개의 조합으로 표준화하였다.

[표 3-6] Window Type 파라미터 설계의 원칙

구분	내용 설명
물리적 설명 가능성 (interpretability)	P(층수), L(코팅), E(가스), F(프레임), S(스페이서)는 창문이나 문의 전도, 대류 및 복사 경로를 결정하는 주요 요소이며, P-L-E-F-S 조합은 정성적 성능 추세를 일관되게 설명.
샘플링·표본화를 통한 경제성 확보	전체 조합 864개 조합은 제조업체 데이터가 포함하는 수천 가지 옵션에 비해 간결하지만, 실제로는 변동성이 높은 균형 잡힌 축 집합도 포함되며 필요에 따라 특정 축을 제한하여 공간의 일부만 빠르게 샘플링

Leakage는 코드북의 수치 구간을 그대로 적용한다. PMF 표본화는 의존 맵핑 변수에 대하여 국가/지역 통계·보급률·범규 변천을 반영한 사전 분포에서 추출하되, 다음의 [표 3-7]의 변수 간 의존 매핑 규칙을 함께 적용하였다.

[표 3-7] 범주형 변수의 파라미터별 규칙과 적용 기준

생성 변수	의존 매핑	의존 규칙	출처
Area	Building Type, Location, Year of Construction	전용면적 ¹³⁰ 구간별 평균과 표준편차를 사전 분포로 구간 선택 후 필요시 통계/실측 분포에 맞춰 미세 조정	국토교통부 2024 인구주택총조사 ¹³¹ 기준 ≤40:10.2%, ≤59:30.2%, ≤99:47.9%, ≥100~11.7%
Location	-	지역별 사전 비율로 표본화	2024 인구주택총조사 기준 전국 주거 통계 기준 ¹³² 서울:14.8%, 인천:6.1%, 경기:27.8%, 충청:11.9%, 전라:10.3%, 경상:26.1%, 강원:3.1%
Year of Construction	Building Type, Location	연식 구간별 사전 분포에서 추출	통계청 2024 인구주택총조사 결과 ¹³³ : ~1989(6.9%), ~1999(28.7%), ~2009(26.1%), ~2019(25.2%), 2020~(13.3%)
Building Type	Location	지역 통계 기반 PMF로 표본화	공간 데이터 마켓, 2024년 10월 아파트 단지 기준 정보 ¹³⁴ stair(68%), floor(15%), mixed(17%)
Heating	Location	지역 보급률/통계에 따른 PMF 표본화	국토교통부 통계누리 아파트 주거 환경 통계 ¹³⁵ 기준 개별난방: 51.7%, 지역난방:24.6%, 중앙난방 : 14.7%, 기타: 9%
Cooling	Location	지역 보급률/통계에 따른 PMF 표본화	KESIS 가전·냉방 보급률, 국가 에너지 통계 종합 정보 시스템 13차(2022년 기준) 가구 에너지패널조사 결과표 ¹³⁶ Wall(50%), Stand(46%), System(2%), non(2%)
Load Profile	-	BB01: Basic, BB02: Energy	Clustering of Load Profiles of Residential Customers

		Saving BB03: M-Pattern. BB04: Morning Concentrated BB05: Evening Concentrated BB06: Owl BB07: 무작위 혼합	Using ¹³⁷⁾
Window Type	Year	연식·범규에 따른 사전 분포(PMF)로 코드 선택	Neale et. al.의 Windows 파라미터의 Hot2000code의 규칙 ¹³⁸⁾ 참고하고 국토교통부의 국내 에너지절약형 친환경주택 건설기준 해설서를 기준으로 함.
Leakage	Building Age, Location	연식·지역 조건의 PMF에서 구간 선택	국토교통부, 2023 에너지 절약형 친환경주택 건설기준 해설서 ¹³⁹⁾¹⁴⁰⁾

2) 연속형 UPD 변수

Rotation 변수는 KESIS 국가 에너지 통계 종합정보 시스템 제13차(2022년 기준) 가구에너지패널조사 결과표를 바탕으로 전국 17개 시도의 가구를 대표하는 가구본으로 7,399가구를 대상으로 조사원이 직접 방문 조사를 통하여

-
- 130) 주택건설공급과. (2025). “도시형 생활 주택 면적 제한 완화를 위한 『주택법 시행령』 등 개정안 시행 안내 보도 자료”, 2025.1.20. 국토교통부.
- 131) 조사관리국 인구총조사과, 전게서, p.76.
- 132) 조사관리국 인구총조사과, 전게서, p.70.
- 133) 조사관리국 인구총조사과, 전게서, p.77.
- 134) 공간데이터마켓(2024). “2024년 10월 아파트단지 기준정보”. LX한국토지정보공사-(주)펜타코드. Retrieved from https://geomarket.kr/user/dataset/view.do?data_sn=25.
- 135) 국토교통 통계누리. (2023). “아파트주거환경통계 현황(2023)”. 국토교통부. Retrieved from https://stat.molit.go.kr/portal/cate/statView.do?hRsId=33&hFormId=&hDivEng=&month_yn=.
- 136) 남수현, 이현. (2023). 『가구에너지패널조사 연구』. 한국에너지경제연구원(KESIS). Retrieved from https://kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=view.
- 137) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K. (2021). Clustering of load profiles of residential customers using extreme points and Demographic Characteristics. *Electronics* 2021.10(290).
- 138) Neale, A., 전계논문, p.62.
- 139) 주택건설공급과, (2023), 전게서, p.126.
- 140) 박원석, 윤재욱. (2003). 신축공동주택의 기밀성능 실측에 관한 연구. 『한국생태환경건축학회 논문집』, 3(3), 43-52.

조사된 거주 중인 공동주택의 남향을 기준으로 배치된 방향값의 통계를 기준으로 적용하였다¹⁴¹⁾. Aspect ratio의 변수는 국내에서 건물 중형비 관련된 연구의 내용이 부족한 관계로 선행 연구의 QSFBSSEM에 적용된 허용된 값을 기준으로 연속범위 내에 균등 분포로 표준화하였다¹⁴²⁾. WWR은 건설교통부에서 공개한 2025년 에너지 절약형 친환경주택 건설기준 해설서의 [표 3-8]에 따르면 일반적으로 아파트의 경우 45% 이하를 권고하고 있다¹⁴³⁾.

[표 3-8] 친환경 주택의 창면적비 기준¹⁴⁴⁾

기준	Bay수	1	2	3	4	5
	창면적비	20% 이하	25% 이하	31% 이하	38% 이하	45% 이하
기타	창면적비[%] = (0.0689 x Bay 수 + 0.1044) x 100 계산값 이하					

* 단, 펜트하우스, 테라스하우스, 복층형 및 3면개방형 등 특이평면은 실제 창면적비를 계산하여 입력

연속형 변수는 [표 3-9]과 같이 정하였으며 이는 중부지방의 5개 공동주택 단지 29개 타입에 대한 전수조사를 통하여 공개한 선행 논문을 기준으로 창면적비를 20%~60%까지로 조사한 결과를 바탕으로 20(5%), 30(31%), 40(17%), 50(44%), 60(5%)로 적용하였다¹⁴⁵⁾. 코드생성 규칙을 분포 조건으로 사용하고, 각 bin에서 관측·문헌 기반의 평균-표준편차-범위를 설정한다. EnergyPlus에 입력되는 데이터의 표본화는 물리적 하한/상한을 넘지 않도록 절단 정규분포(truncated normal)를 기본으로 WWR은 [0,1] 범위를 가지므로 베타 분포/삼각 분포를 선택할 수 있다.

Rotation은 0/45/90/135/180/225/270/315의 discrete set이지만, 데이터 생성 파이프라인에서는 UPD 값을 뽑아 정수로 저장한다¹⁴⁶⁾. 생성된 연속

141) 국가에너지통계 종합정보시스템. (2023). “KESIS 국가에너지정보시스템 -에너지총조사결과표”. 산업통상자원부에너지경제연구원 Retrieved from <https://kesis.keei.re.kr/board.es?mid=a10201030100&bid=0003>.

142) Neale, A., 전계논문, p.47, p.52.

143) 주택건설공급과, (2023), 전계서, p.132, p.260.

144) 건설안전과. (2012). 『건축물 에너지 절약을 위한 창호 설계 가이드라인』. 국토교통부(국토해양부).

145) 장철용, 한혜심, 이진숙. (2010). 차폐계수와 창면적비에 따른 공동주택의 건물 에너지 효율 등급 평가. 『한국태양에너지학회 논문집』. 30(5), 38-45.

146) 이나은, 안병립, 정학근, 김종훈, 장철용. (2013). 창호의 성능 및 건물의 창면적비에 따른 기

UPD는 Classifier 단계에서 KESIS 국가 에너지 통계 종합정보 시스템 13차 (2022년 기준) 가구에너지패널조사 결과표를 바탕으로 한국형 리스케일 규칙을 적용하였다.

[표 3-9] 연속형 변수의 파라미터별 규칙과 적용 기준

변수	값	표본화 분포	변수 생성 규칙	출처
WWR	0.2, 0.3, 0.4, 0.5, 0.6	차등 분포 PMF over {0.2~0.6} 연속화 옵션: Beta 분포, 절단 정규 TN	0.2~0.6의 각 구간을 20(5%), 30(31%), 40(17%), 50(44%), 60(5%)로 차등 분포로 적용	차폐계수와 창면적비에 따른 공동주택의 건물 에너지 효율 등급 평가 ¹⁴⁷⁾ ¹⁴⁸⁾
Rotation	0° ,45° ,90° , 135° ,180° ,225° ,270° , 315°	기본: 균등 PMF "N":0.021, "NE":0.07,"E":0.09, "SE":0.225, "S":0.527, "SW":0.084, "W":0.032, "NW":0.07 연속화 옵션: 없음	4방위 중 균등 추출해 정수 각도로 저장(남향 우선 비율 p_south로 가중을 PMF에 반영 검토)	KESIS 국가 에너지 통계 종합 정보 시스템 13차(2022년 기준) 가구에너지패널 조사 결과표 ¹⁴⁹⁾
Aspect ratio	0.8~1.2 (권장 5수준: 0.8, 0.9, 1.0, 1.1, 1.2)	기본: UPD=Uniform(0.8, 1.2) 연속화 옵션: 절단 정규 TN	단순성 우선이면 Uniform(0.8,1.2), 분류 단계에서 5-bin(0.8/0.9/1.0/1.1/1.2)로 라벨링.	설계 표본 조사, 단지 평면 통계→ Neale et. al.의 shape 파라미터 생성 규칙을 동일하게 적용 ¹⁵⁰⁾

존 단 독주택의 창호 리모델링 방안 연구. 『한국생태환경건축학회논문집』. 13(3), 71-80.
<http://dx.doi.org/10.12813/kieae.2013.13.3.071>
 147) 장철용, 한혜심, 이진숙., 전계논문, p.39-40.
 148) 건설안전과, (2012), 전계서, p.8.
 149) 국가에너지통계 종합정보시스템, (2023), 전계서

3) Stochastic(시간 의존/내부 부하)

시간 의존 변수는 [표 3-10]과 같이 Occupancy Activity와 DHW, LIGHTING, EQUIP은 점유자 수(1~4)와 7가지의 에너지소비 클러스터 행태(BB)의 조합으로 정의된 28개 프로파일을 사용한다¹⁵¹⁾. 이들은 1시간 단위의 해상도 시계열로 생성되어 부하 곡선에 직접 반영된다. Lighting Density와 Home Appliance도 분류표에서 Stochastic 그룹에 포함되며, 활동·점유 스케줄과 결합되어 시간 의존 부하로 구현된다.

[표 3-10] 시간 의존 부하의 파라미터별 규칙과 적용 기준

변수	값	표본화 분포	변수 생성 규칙	출처
Occupant	1~4명	Stochastic 그룹에 포함되나, 값 자체는 이산 수치	① 가구원수 선택(이산) ② 이후 항목(행태/DHW/내부 부하)의 핵심 입력으로 사용	KESIS 국가에너지 정보시스템 ¹⁵²⁾ 기준. 1명(17.7%),2명(34.9%),3명(21.4%),4명(26.1%)
Occupancy Activity	(1~4)×7 = 28개 프로파일 ID	Stochastic 시계열	가구원 수에 맞춰 7개 행태 중 1개 선택→ 확률표 기반으로 일·주말/시간대별 활동/점유 시계열 생성	Clustering of Load Profiles of Residential Customers Using ¹⁵³⁾ + CREST 규칙 적용
DHW	(1~4)×7 = 28개 프로파일 ID	Stochastic 시계열	활동/점유 시퀀스에 동시성을 갖도록 온수 사용량을 확률적으로 생성	
Lighting Density	Low / Mid / High/Super	Stochastic 시계열	밀도 카테고리 선택→ 활동/일사 조건에 따라 조명 부하 시계열 생성	건설교통부, 2025 에너지 절약형 친환경주택 건설기준 해설서 ¹⁵⁴⁾
Home Appliance	·BB01~·BB06 ·BB07:무작위 혼합	Stochastic 시계열	기기 세트 선택→ 활동 스케줄과 결합해 사용 부하를 생성	Clustering of Load Profiles of Residential Customers Using ¹⁵⁵⁾ + KESIS 국가 에너지 정보 시스템 ¹⁵⁶⁾

150) Neale, A., Kummert, M., & Bernier, M., 전계논문, p.53, p.88.

151) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.

조명 밀도 파라미터의 경우에는 2025년 국토교통부의 에너지절약형 친환경 주택 건설기준 해설서에 따르면 조명 밀도를 $6\text{W}/\text{m}^2$ 으로 설계하거나 전면 LED 시설을 권고하고 있다. 또한, 전면 LED 조명으로 설치할 경우에는 $86\text{W}/\text{m}^2$ 이하를 충족할 것을 권고하고 있다¹⁵⁷⁾.

이와 관련된 선행 연구에서 공동주택 주거 단지 총 10개 유형을 2009년 8월27일~2010년 1월 18일까지 5개 아파트단지를 조사한 결과 실제 공동주택에 설치된 공동주택의 설치유형을 A~J type까지 총 20개 유형으로 분류하고 실제 측정된 조명 밀도는 $12.34\text{W}/\text{m}^2\sim 26.29\text{W}/\text{m}^2$ 으로 조사되었다¹⁵⁸⁾.

그러나 이러한 결과는 과거에 가정용 소형 형광 램프(Compact Fluorescent Lamp, CFL) 중심으로 현장에 설치된 기준으로 2020년 이후부터는 LED 조명등으로 전면 설치가 되고 있고 구축 아파트의 경우에도 내용연수가 지난 CFL 타입이 LED 등기구로 순차적으로 교체되면서 전반적으로 공동주택의 에너지 밀도는 $8\text{W}/\text{m}^2\sim 20\text{W}/\text{m}^2$ 수준으로 분포되어 있는 것을 기준으로 ~ 8 , $8\sim 11$, $11\sim 14$, $14\sim 20$ W/m^2 의 4단계의 구간으로 변수를 구성하였다.

이러한 기준을 바탕으로 Lighting Density는 [표 3-11]의 조건으로 변수를 설정하였다. 먼저 건축연도 1~2(~1999년도 이전)의 경우 건물에 적용된 대부분 조명이 CFL과 할로겐 램프로 적용되어 있으며 HES(Household Electricity Survey)에 따르면 CFL은 LED에 비하여 램프 효율이 66.9%로 배율 1.49, 할로겐 램프는 26.1%로 3.83배 집계하였지만¹⁵⁹⁾, 이후 내용연수의 경과에 따라 일정 정도는 LED로 교체된 것으로 추정하여 광원 배율은 1.5로 설정하였다.

152) 국가에너지통계 종합정보시스템, (2023), 전계서, “다. 가구원수에 관한 사항”.

153) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.

154) 주택건설공급과. (2025). 『에너지 절약형 친환경주택 건설기준 해설서』. 국토교통부.

155) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.

156) 국가에너지통계 종합정보시스템, (2023), 전계서.

157) 주택건설공급과, (2025), 전계서, p.158.

158) 이진숙, 최종문, 김소연, 박지영. (2010). 공동주택의 실내 공간별 조명적용현황 연구. 『대한건축학회논문집』. 26(12), 45-56.

159) Martin, C., Tebbit & J., Bernstein, L. (2023), 전계논문, p.8.

[표 3-11] Lighting Density 파라미터 설정 조건

구분	배율	적용 기술	Lighting Density [W/m ²]	주요 근거
건축연도 1~2	X 1.5	CFL	10~12	Lamp efficacy ¹⁶⁰⁾
건축연도 3	X 1.1	CFL+LED 혼재	7~9	공동주택 실내 조명 현황 ¹⁶¹⁾
건축연도 4~5	X 0.9	LED 주도	6~8	국내 LED 보급률 통계 ¹⁶²⁾

건축연도 3(2009년 이전)의 경우에는 CFL과 LED 그리고 할로겐 램프가 혼재하는 시기로 공동주택 실내 조명 현황에 대한 선행 논문에서는 국제기구인 ASHRAE/IESNA의 권장 기준이 주택의 경우 2010년 기준 7.5W/m²을 권장하고 있지만 2010년 조사 당시 실제 12.34~26.29W/m²로 조사되었으며 평균 조명 밀도가 17.2W/m²으로 높게 조사되었다¹⁶³⁾.

건축연도 1~2 기준 대비 실제 시공 시 LED 조명등의 혼합 비율이 20~40% 수준으로 혼재되어 건축된 시기를 고려하여 광원 배율 가중치를 10% 상향하여 1.1로 설정하였다.

건축연도 4(2010년 이후)의 경우에는 건축 시 LED 보급이 확산되었으며 2020년 이후로는 신축 공동주택은 대부분 LED 조명으로 적용되고 있는 것으로 추정하여 적용하였다. 건설교통부의 에너지절약형 친환경주택 건설기준에 따라 공동주택의 조명설계 기준이 8W/m² 대비 미국 에너지 절약 설계기준에서 조명 에너지 절감을 위한 ASHRAE/IESNA의 권장 조명 밀도를 기준으로 7.5W/m²의 조명 전력 허용 기준을 참조하여 광원 배율을 0.9를 기준으로 적용하였다¹⁶⁴⁾¹⁶⁵⁾.

160) Martin, C., Tebbit & J., Bernstein, L. (2023). Lighting, Cooking, Electrical Appliances and Incidental Heat Losses in the Home Energy Model: FHS Assessment Wrapper (HEMFHS-TP-04). Home Energy Model: FHS assessment version: HEM: FHS v0.18. London: UK Government, Department for Energy Security & Net Zero

161) 이진숙, 최종문, 김소연, 박지영., 상계논문, p.50.

162) 수요정책실. (2025). 『KEA 에너지 이슈 브리핑』. Vol.276, 2025.10.13 한국에너지공단.

163) 이진숙, 최종문, 김소연, 박지영., 전계논문, p.50.

164) 주택건설공급과, (2025), 전계서, p.58

165) 이진숙, 최종문, 김소연, 박지영., 상계논문, p.50.

3.1.3.2 조건부 PMF 변수 설계

단열 특성에 대하여 Wall RSI, Roof RSI, Foundation RSI를 연속형 확률 변수로 정의하고, 각 부위별 단열 수준을 0.5~1.5, 1.5~2.5 등의 구간별 평균 값을 갖는 조건부 PMF로 설정하였다. 이러한 범위는 국토교통부 「건축물의 에너지절약 설계기준(2023)」에서 제시된 외벽·지붕·바닥 열관류율(U-value) 기준을 RSI 값으로 변환하였다, 1990년 이전 기축 공동주택에서 2018년 이후 고단열·제로 에너지 기준에 이르는 국내 건물 외피 성능의 기준을 반영하였다¹⁶⁶⁾.

[표 3-12] 건물 외벽 단열(RSI) 적용 기준¹⁶⁷⁾

구분	RSI 범위 [m ² K/W]	해석	근거
저단열	Overall RSI < 1.56	·1980~1990년대 준공 공동주택 -외벽 U ≈ 0.60-0.70W/m ² K -노후 외피, 단열재 미흡	· 국토부 친환경주택 건설기준 해설서 기준 건축물 기준과 정확히 일치 · QSFBSSEM 절단 정규 분포(0.5-1.5 평균 1.0)에 부합
중단열	1.56 ≤ Overall RSI < 2.25	·2000-2018년 단열 기준 개정 전후 공동주택 -외벽 U ≈ 0.40-0.60W/m ² K -정책적 중간 단계 단열 수준	· QSFBSSEM 조건부 PMF의 두 번째 구간(1.5-2.5)과 정합
고단열	Overall RSI ≥ 2.25	·2018년 제로 에너지 건축 단계 이후 공동주택 -외벽 U ≤ 0.30W/m ² K -고성능 단열·기밀 기준 적용 ¹⁶⁸⁾	· 국토부 제로 에너지 기준(U ≤ 0.30) 과 수치적 정합 · QSFBSSEM 3단계 중 최고 단열 구간과 일치

166) 건설산업과 국가건설기준센터, (2023), 전계서, p.18, 134, 154, p.242-248.

167) 건설산업과 국가건설기준센터, (2023), p.122.

168) 박원석, 윤재욱. (2003). 신축공동주택의 기밀성능 실측에 관한 연구. 『한국생태환경건축학회 논문집』, 3(3), 43-52.

특히 각 구간의 평균, 분산, 하한, 상한은 QSFBSSEM에서 제시한 절단 정규 분포 기반의 생성규칙을 적용함으로써, 단열 성능이 단순 불연속 구간이 아니라 실제 물리적 변동성이 존재하는 확률 변수로 처리되도록 하였다¹⁶⁹⁾.

Wall, Roof, Foundation RSI에 대한 세부적인 외피 성능 변수는 각 PMF에서 샘플링된 값을 기반으로 하지만, 전체 건물에 대한 외피 열 저항 수준의 지표인 Overall_RSI를 추가하여 [표 3-12]와 같이 도출하였다¹⁷⁰⁾. Overall_RSI는 각 건물 요소의 열 손실 기여도를 고려하여 벽에 0.5, 지붕에 0.3, 바닥에 0.2의 가중치를 적용하였다. 이는 난방 부하에 미치는 영향을 결정하는 주요 경로가 벽에서 가장 크고 그 다음으로 지붕과 바닥이라는 국내외 연구 결과와 QSFBSSEM 모델의 가중치 체계를 기반으로 적용한 것이다¹⁷¹⁾. 이 가중치 방법은 EnergyPlus 기반의 건물 에너지 시뮬레이션에서 사용되는 외피 열저항 대표값 정의와 일치하여 물리적 및 시뮬레이션 정합성을 모두 보장한다¹⁷²⁾.

산출된 Overall RSI 값은 단열 수준 분류를 위해 세 개의 범주로 구분하였다. 첫째, 저단열(Overall RSI < 1.56) 구간은 국토교통부 기준에서 외벽 열관류율 $U \geq 0.60\text{W}/\text{m}^2\text{K}$ 에 대응하는 수준으로, 주로 1990년대 이전에 준공된 노후 공동주택의 외피 성능을 반영한다. 둘째, 중단열($1.56 \leq \text{Overall RSI} < 2.25$) 구간은 $U \approx 0.40 - 0.60\text{W}/\text{m}^2\text{K}$ 수준의 단열 기준이 적용된 2018년까지 건축물에 대응하며, 에너지절약 설계기준 개정 후의 국내 공동주택이 일반적으로 포함된다¹⁷³⁾.

마지막으로, 고단열(Overall RSI ≥ 2.25) 구간은 $U \leq 0.30 - 0.35\text{W}/\text{m}^2\text{K}$ 수준을 갖는 고성능 외피 구조로, 2018년 이후 제로 에너지 건축 의무화로 건설된 신축 공동주택의 단열 성능 범위를 반영한다¹⁷⁴⁾.

이와 같은 저·중·고단열 기준은 QSFBSSEM이 캐나다 퀘벡주의 건물 스톡을 대상으로 구축한 VSM Archetype의 단열 구간 구조와 완벽하게 정합성을 가

169) Neale, A., Kummert, M., & Bernier, M., (2019), 전계논문, p.3395.

170) 김혜원, 전계논문, p.26.

171) Neale, A., Kummert, M., & Bernier, M., (2019), 전계논문, p.3395.

172) U.S. Department of Energy, (2019), p.176-178.

173) 김혜원, 전계논문, p.30-32.

174) 김혜원, 전계논문, p.30-32.

지며, 동시에 한국형 외피 성능 규제 기준을 반영하여 국내 실정에 적합하도록 보정된 값이다. 따라서 본 연구의 Archetype 기반 LDA 분류에서 저단열·중단열·고단열의 단열 수준에 대한 카테고리는 실측된 규제 값과 선행 연구의 시뮬레이션 사례 기반의 확률 분포¹⁷⁵⁾, 건설 연도·난방 방식·세대 규모와 전력 수요 패턴을 결정하는 중요한 클래스 변수로 활용될 수 있다.

3.1.3.3 Fixed 외생 변수 설계

Temperature, Humidity, Sunlight, Solar radiation은 기상청에서 발표한 관측 기준의 실측 시계열을 기본 자료로 사용하고 시나리오 정의·분석 필터·라벨링에 활용하였다. 이는 향후 생성되는 한국형 KR-VSM 데이터셋의 파이프라인에서는 실제 시간별 값을 사용하고, 각 시점이 어느 구간에 속하는지를 태그로 부여해 수요반응 조건을 제어할 수 있다. 이러한 확률변수 설계는 Generator 단계에서 200,000개의 세대별 부하 곡선을 생성하며, 기상 데이터와 융합 분석을 통하여 전력 수요 예측 모델의 학습 데이터로 직접 활용 가능하다. 특히, QSFBSM 모델이 캐나다 단독주택에 대해 확률 변수를 적용했던 것을, 본 연구에서는 한국의 실정에 부합되도록 공동주택 중심 구조와 기상 연계 변수를 추가하는 방향으로 확장하였다.

3.1.4 한국형 CREST 점유 스케줄 커스터마이징

CREST 모델은 가구 구성원별 점유 패턴과 전력 부하를 시뮬레이션하는 도구로, 영국의 표준생활패턴 데이터를 기반으로 [표 3-13]의 기준에 따라 설계되었다. 그러나 한국 공동주택의 생활 양식, 가전기기 보급률, 계절별 기상 조건 반응은 영국과 다르기 때문에, 이를 그대로 적용할 경우 부하 패턴 재현성이 떨어진다. 따라서 본 연구에서는 CREST의 점유 스케줄과 내부 부하 생성 로직을 [표 3-13]과 같이 한국형으로 커스터마이징하였다.

이렇게 커스터마이징된 CREST 모듈은 EnergyPlus 기반 열·냉방 부하 시뮬

175) 이상헌, 전계논문, p.30-32.

레이션과 결합하여, 기상청의 외기 조건 입력으로 사용하는 기상 반응형 VSM 데이터셋을 생성한다¹⁷⁶⁾. 이 데이터셋은 고부하가 예상되는 시간대에 CREST의 점유·기기 사용 패턴을 자동으로 변경하여 DR 이벤트 효과를 분석할 수 있으며 폭염이 발생했을 때 DR 시나리오에서 설정 온도 상향, 일부 가전기기 사용 제한을 적용한 경우의 부하 감축량과 거주자 불편 지수를 산출까지 가능할 수 있으며 DER 운영 전략 검증과 정책 시뮬레이션에 활용될 수 있다¹⁷⁷⁾.

[표 3-13] 점유 스케줄 별 부하 로직의 커스터마이징 기준

점유 스케줄	내부 부하 로직의 커스터 마이징
시간 사용 패턴 보정	한국은 평일 퇴근 시간이 영국보다 늦고, 주말 여가·가사활동 비중이 달라, 재택·외출·취침 시간대를 재설정 필요하며, 통계청 「2024년 한국인 생활시간조사」 데이터 ¹⁷⁸⁾ 와 Electronics 2021의 에너지 소비 클러스터를 반영 ¹⁷⁹⁾ .
가전 부하 항목 확장	LoadProfile에 대한 연구(2014) ¹⁸⁰⁾ 과 국가에너지통계 종합정보시스템 13차(2022년 기준) 가구에너지패널조사 결과표 ¹⁸¹⁾ 를 기준으로 한국의 주방기기에서 냉장고, 김치냉장고, 그리고 전기 밥솥의 높은 전력 사용량을 반영하고 여름철의 냉방 부하의 급격한 증가 등 국내 특유의 전기기기 사용을 에너지 소비 클러스터를 기준으로 생성 모듈에 추가
계절·기상 반응을 로직을 추가	기상청 지상 기상 관측 지침 ¹⁸²⁾ 을 기준으로 여름철 35℃ 이상 2일 이상 시 폭염 일수에 따른 에어컨 가동률과 부하 상승률을, 겨울철 영하 12℃ 이하가 2일 이상 시 한파일 수에 따른 전기난방기기 사용 증가율을 기준으로 결정하였다. 이를 통해 부하 곡선이 실제 기상 조건 변화에 동적으로 반응하도록 반영

176) CREST. (2017). CREST-Centre for Renewable Energy and Sustainable Technologies: Performance Report. South Pacific Passive House Conference (SPPHC 2017). South West College / Queen's University Belfast.

177) 이윤경, 전게서, p.36.

178) 사회통계국 사회통계기획과. (2025). 『2024년 생활시간조사 결과』. 통계청.

179) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전게논문, p.5,p.8.

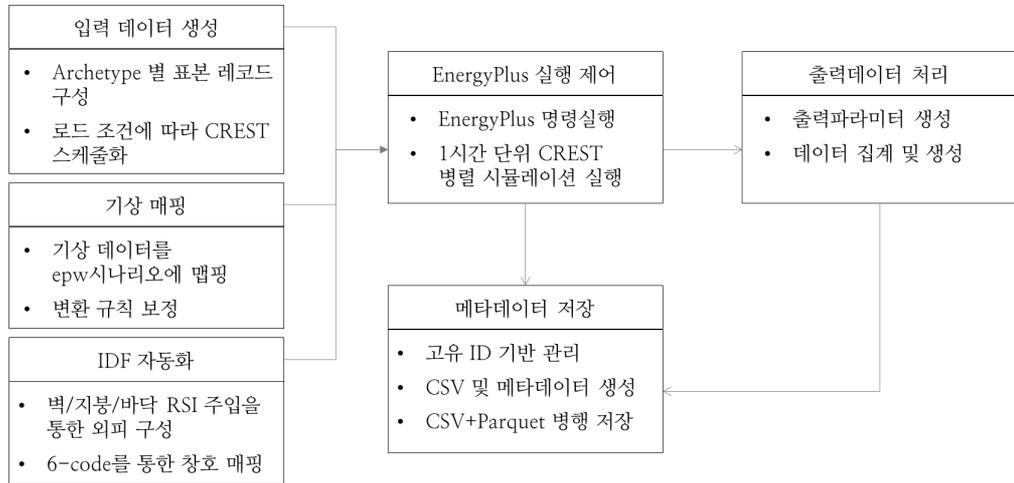
180) Seo, Y.-K., & Hong, W.-H. (2014). Constructing electricity load profile and formulating load pattern for urban apartment in Korea. Energy and Buildings, 78, 222-230.

181) 국가에너지통계 종합정보시스템, (2023), 전게서.

182) 국가기후데이터센터, (2024), 전게서.

3.2 시뮬레이션 구조 설계

한국형 VSM의 생성을 위하여 시뮬레이션은 [그림 3-2]에서와 같이 먼저 한국 공동주택 건물 스톡을 대표할 수 있는 입력 데이터의 표본을 생성하는 첫 번째 단계에서부터 EnergyPlus 실행 제어를 통한 최종 메타데이터 출력 처리의 단계까지 순차적으로 처리하였다.



[그림 3-2] EnergyPlus-Python 병렬 제어 파이프라인

3.2.1 입력 생성 단계

첫 번째 단계인 입력 생성 단계에서는 선행 연구와 통계 자료를 바탕으로 정의된 Archetype 분포를 토대로, 각 가구를 규정하는 핵심 변수들을 확률적으로 샘플링한다. 구체적으로는 전용면적(Area), 지역(Location), 건축연도(Year), 난방방식(Heating_type), 냉방 방식(Cooling_type), 창호 유형(Window_type), 그리고 단열(벽체·지붕·바닥의 열관류 저항값으로 각각 RSI_{wall}, RSI_{roof}, RSI_{floor}, Overall_RSI)을 포함하는 외피 성능 변수를 동시에 고려한다¹⁸³⁾. 각 변수는 국가 건축 통계, 에너지 절약 설계 기준, 공

183) U.S. Department of Energy. (2023). EnergyPlus™ Version 23.1.0 Documentation: Output

동주택 에너지 사용 실태 등의 자료를 활용하여 PMF 또는 UPD로 모형화하였으며, Python 기반 샘플링 스크립트에서 난수 시드(rng_seed)와 함께 반복 재현 가능한 방식으로 표본을 생성한다. 이렇게 생성된 각 표본은 CREST 부하 모형과의 연계를 위해 A-BB 형식의 Load Profile 코드를 부여받는다. 앞의 문자(A∈1~4)는 가구 구성과 주거 행태를, 뒤의 두 자리 숫자(BB∈01~07)는 시간 사용 패턴 클러스터를 나타내며, 이 조합을 통해 최소 28종의 대표적인 거주 패턴을 표현하도록 설계하였다. 이러한 설계를 바탕으로 Python 스크립트에서 입력 생성-〉IDF 구성-〉EnergyPlus 실행-〉출력 추출로 이어지는 일관된 파이프라인을 Batch file을 통해 호출하는 구조로 각 가구에 할당된 CREST 부하 코드에 대응되는 스케줄 템플릿을 EnergyPlus 시뮬레이션에 주입하였다. 이를 통해 연간 1시간 해상도의 전력 사용 결과가 자동으로 산출되며, 최종적으로 365일 기준의 8,760행의 시계열 전력 부하 프로파일이 생성되도록 구성하였다.

이때 부하 스케줄은 가구별 점유(Occupancy), 조명(Lighting), 플러그·가전(Plug & Appliance), 온수(DHW) 사용 패턴을 분리하여 계산하며, 2024년 생활시간조사와 가구용 에너지 소비 조사 자료를 바탕으로 계절·주별·시간대별 사용 확률이 재구성된다. CREST 부하 스케줄은 이후 EnergyPlus의 Schedule:File 객체에 직접 연결되어, 난방·냉방·조명·플러그 부하를 구동하는 시간별 내부 발열 및 전력수요 입력으로 활용된다¹⁸⁴⁾. 이와 같은 입력 생성 과정을 통해, 하나의 sim_id는 현실적인 거주 행태와 외피 성능을 동시에 반영하는 고유한 가구 단위 에너지 시뮬레이션 실체로 정의된다.

3.2.2 기상 조건 매핑 단계

각 표본이 실제 기상 변동성을 반영하도록 기상 조건을 매핑하는 2번째 단계에서는 EnergyPlus 시뮬레이션에 사용하는 epw 파일과 기상청에서 제공하

Details and Examples. Build : 87ed9199d4.

184) 이수진, 김유정, 진혜선, 서인애, 김성임, 하수연, 송승영. (2020). 건물, 거주자, 기기 사용 특성이 공동주택 용도별 에너지사용량에 미치는 영향 분석. 『대한건축학회논문집』. 36(6), 153-164. <https://doi.org/10.5659/JAIK.2020.36.6.153>.

는 관측 데이터를 체계적으로 매핑하였다¹⁸⁵). 먼저 7개 광역시도 별 epw 실행 스크립트에서는 다음과 같이 폴더를 생성하도록 설정하였다.

01_SEOUL, 02_GYEONGGI, 03_INCHEON, 04_CHOONGCHUNG, 05_JEONRA, 06_GYEONGSANG, 07_GANGWON

이상과 같이 총 7개의 광역군의 각각의 폴더를 생성하여 권역 코드를 정의하고, 공통 *.idf 템플릿으로 base_vsm_template*.idf 을 각 권역 폴더로 복제한 뒤, 대응되는 epw를 다음과 같이 매핑하였다¹⁸⁶).

- . KR_SEOUL_2024_epw_scaled.epw,
- . KR_GYEONGGI_2024_epw_scaled.epw,
- . KR_INCHEON_2024_epw_scaled.epw,
- . KR_CHOONGCHUNG_2024_epw_scaled.epw,
- . KR_JEONRA_2024_epw_scaled.epw,
- . KR_GYEONGSANG_2024_epw_scaled.epw,
- . KR_GANGWON_2024_epw_scaled.epw

파이썬 스크립트는 폴더명_epw 파일명 매핑 딕셔너리를 이용해 EnergyPlus 실행 명령어(EnergyPlus.exe -w <epw> -d <location_folder> -r <idf>)를 일괄 생성하고, 7개 광역군에 대해 지역별로 각각 생성된 *.idf 파일의 건물·부하 설정을 주입시키고 외기 조건을 지역별로 설정된 시뮬레이션을 수행한다¹⁸⁷). 이 때 각 시뮬레이션 폴더는 이후 분석 단계에서 Location 또는 Location/region 변수와 직접 연결되는 권역별 epw 경로를 메타데이터로 갖게 된다¹⁸⁸).

이어지는 weather_hourly로 생성되는 날씨 시계열 스크립트에서는 기상청 기상 자료 개방 포털에서 다운로드 한 기상 데이터를 연결하였다.

- . 01_SEOUL_2024_Weather.csv ; 종로 관측소
- . 02_GYEONGGI_2024_Weather.csv ; 수원 관측소
- . 03_INCHEON_2024_Weather.csv ; 인천 관측소

185) U.S. Department of Energy. (2021). EnergyPlus™ Version 9.5.0 Documentation: EnergyPlus Essentials. Build : de239b2e5f.

186) U.S. Department of Energy, (2021), 전계서, p.16.

187) U.S. Department of Energy, (2021), 전계서, p.28.

188) U.S. Department of Energy, (2021), 전계서, p.31.

- . 04_CHOONGCHUNG_2024_Weather.csv ; 대전 관측소
- . 05_JEONRA_2024_Weather.csv ; 광주 관측소
- . 06_GYEONGSANG_2024_Weather.csv ; 대구 관측소
- . 07_GANGWON_2024_Weather.csv ; 춘천 관측소

해당 광역군 별로 중심부의 대표 도시에 위치한 광역군의 관측소를 대표 관측소 데이터를 지정하여 연간 8,760시간의 기상 시계열을 권역별로 통합하였다¹⁸⁹⁾. 스크립트는 파일명에서 01_SEOUL~07_GANGWON 등 권역 코드를 자동 추출하고, 각 권역 파일에 대해 ① 날짜-시간 컬럼을 단일 timestamp로 자동 탐지, ② 문자열 파싱을 통한 YYYY-MM-DD HH:MM:SS 형식의 시계열 생성, ③ 온도 컬럼 탐색 및 화씨(Kelvin)->섭씨(Celsius) 변환 여부 판단, ④ 1시간 간격 리샘플링의 시간 축을 기준으로 선형 보간을 수행하였다.

이렇게 전처리 된 결과는 timestamp, location, temp 세 컬럼으로 정규화되어 vsm_hourly*로 저장되며, 각 권역별로 2024-01-01 00:00~2024-12-31 00:00 구간의 연속적인 시간 축과 이상치의 제거 및 보간이 반영된 기온 시계열이 확보된다.

최종적으로 각 sim_id는 메타데이터의 Location/region 변수와 해당 권역 epw 파일 그리고 weather_hourly* 내 검증·분석용의 동일 권역 관측 기상 시계열을 동시에 참조하도록 설계하였다. 이를 통해 동일한 CREST 기반 내부 스케줄을 갖는 가구라 하더라도, 지역별 epw* 기상 관측치에 따라 냉·난방 부하와 전력 사용 패턴이 상이하게 나타나도록 구현하였다.

3.2.3 *.idf 자동화 단계

자동화 단계는 3번째 단계로 EnergyPlus 입력 파일 *.idf를 대량 생성으로 자동화하였다. 본 연구에서는 Energy에서 기본으로 제공하는 base_vsm_template.idf를 기본 템플릿으로 정의하였다.

여기서 주요 변수의 설계는 다음과 같다¹⁹⁰⁾.

{{Lx}}, {{Ly}}, {{H}},

189) 신이레, 윤상후., 전계서, p.1602.

190) U.S. Department of Energy, (2023), 전계서, p.20-140.

{{AZIMUTH}}, {{ACH}}, {{U}}, {{SHGC}}, {{VT}},
{{LIGHTS_CSV}}, {{PLUGS_CSV}}, {{DHW_CSV}},
{{LOAD_PROFILE_CSV}}

이상과 같이 주요 설계 변수를 EnergyPlus의 기본 템플릿 *.idf로 설정하고 한국의 7개 광역군 별로 분리하여 배치하였다¹⁹¹⁾.

Python의 Jinja2 템플릿 엔진을 이용해 각 sim_id별 메타데이터 행을 읽고, 외피 치수(Lx, Ly, 층고 H), 창면적비(WWR), 창호 6자리 코드(Window_code_6digit)에서 도출한 열관류율(U-value), 태양열 취득 계수(Solar Heat Gain Coefficient, SHGC), 가시 투과율(Visible Transmittance, VT), 그리고 침기량(Air Changes per Hour, ACH)을 템플릿에 치환하여 건물 열적 특성을 구성하였다¹⁹²⁾. 또한 Load_profile_code를 LoadProfiles_EPlus_index *와 조인하여, 조명·플러그·온수·종합 부하를 *.csv 파일형태로 생성하여 각각의 광역군별 지역 경로를 Schedule:File 객체에 자동으로 주입한다¹⁹³⁾.

템플릿을 렌더링 후에는 normalize_idf 루틴을 통해 idf 구조를 표준화한다. 이 과정에서는 Version, Building, Timestamp, SimulationControl 등의 필수 객체 누락을 검사하고, 중복된 ScheduleTypeLimits 또는 레거시 스케줄 정의를 정리하여 EnergyPlus가 요구하는 *normalized.idf로 변환하였다. 소수의 샘플 케이스에 대해 *validate_idf를 병행으로 생성하여, 각종 치환 오류, 경로 설정의 문제, 이상치를 자동으로 검출하였으며, 이러한 자동화 절차를 통해서 최초 700건에서, 20,000건, 그리고 200,000건의 idf 파일이 동일한 규칙과 품질 수준을 유지한 채 생성될 수 있으며, 이는 대규모 VSM 데이터셋의 안정적인 생산을 위한 기본 설정에 해당된다.

3.2.4 실행 제어 및 병렬화 단계

이어지는 4번째 단계에서는 7개 광역군 별로 각각 생성된 *.idf와 *.epw 파일을 EnergyPlus에 주입을 통하여 자동으로 실행하고, 다수의 시뮬레이션을

191) U.S. Department of Energy, (2021), 전게서, p.31.

192) 김혜원, 전계논문, p.56.

193) U.S. Department of Energy, (2023), 전게서, p.64.

병렬로 처리한다. Python 스크립트는 subprocess 모듈을 사용해 EnergyPlus.exe -w <epw> -d <outdir> <model.idf> 형식의 명령을 [그림 3-3]과 같이 실행하였다¹⁹⁴⁾.

```
(base) EnergyPlus.exe -w <epw> - <outdir> <model.idf>
```

[그림 3-3] EnergyPlus-Python 명령어 처리

실행되는 서버 환경의 운영 체제의 CPU 코어 수를 고려하여 동시 실행 프로세스 개수를 최대 6개의 EnergyPlus 프로세스를 동시에 수행하도록 하여, 시스템 과부하를 방지하면서도 전체 수행 시간을 최소화하도록 구현하였다. 각 시뮬레이션은 고유한 출력 폴더 구조로 KR_VSM_01_SEOUL_XXXXX의 형태로 설정하여, 입력-출력 간 추적성과 재현성을 확보하였다.

시간 해상도는 기본적으로 Timestamp를 1시간으로 설정하되, 향후 DR·피크 부하 분석의 정밀도가 필요할 경우, SimulationControl 및 Timestamp 설정을 통해 15분 단위 시계열까지 확장 가능하도록 설계하였다. 실행 중 발생하는 오류는 eplusout.err 파일과 Python 로그에 동시에 기록되며, 특정 sim_id에서 반복적으로 실패가 발생할 경우 해당 *.idf, *.epw, *.meta를 묶어 자동으로 예외 케이스로 태깅하도록 구현하였다¹⁹⁵⁾. 이러한 실행 제어 및 병렬화 구조는 KR-VSM 프로젝트에서 요구되는 20만 건의 대규모 시뮬레이션을 현실적인 시간 안에 처리할 수 있도록 하는 핵심 요소이다.

3.2.5 출력 후처리 단계

다섯 번째 단계는 EnergyPlus에서 생성된 원시 출력 파일을 분석 가능한 형태로 재 구성하는 후처리 과정이다. 우선 ReadVarsESO와 -m 옵션을 활용하여 eplusout.eso 및 eplusout.mtr 파일에서 전력 및 열량 관련

194) U.S. Department of Energy, (2021), 전계서, p.27.

195) U.S. Department of Energy, (2023), 전계서, p.195.

Output:Meter:Output:Variable을 추출하고, 시뮬레이션별로 *Meter.csv와 Zone별 *zone.csv 형태의 시간별 시계열 파일을 생성한다¹⁹⁶⁾. 이들 파일에는 전체 전력 Electricity:Facility과 난방·냉방·조명·플러그·온수 등 세부 엔드유즈 에너지, 그리고 Zone Ideal Loads Heating/Cooling Energy와 같은 열수지 변수들이 포함된다¹⁹⁷⁾.

Python 기반 parser에서는 이러한 시계열 데이터를 바탕으로 1시간·일·월 단위로 에너지 사용량을 집계하고, 여름 및 겨울 피크부하, 계절별·권역별 최대수요, TOU 기준의 피크 인디케이터 등을 생성한다. 또한 난방, 냉방, 조명, 플러그, 온수의 엔드유즈 별 비중을 산출하여, 가상스마트미터 VSM 데이터 셋에서 각 가구가 어떤 용도로 전력을 주로 사용하는지에 대한 통계를 확보하도록 하였다.

3.2.6 메타데이터 생성 및 보관 단계

마지막 단계는 전체 파이프라인을 관통하는 메타데이터를 정리하고 보관하는 과정이다. 각 시뮬레이션에는 고유한 sim_id가 부여되며, Archetype, Area, Location, 연도, 외피 성능, 창호 코드, Load Profile 코드 등의 입력 파라미터, epw 파일 정보, EnergyPlus 버전, 난수 시드, 실행 타임 스탬프, 그리고 Annual_kWh, 피크 부하, 엔드유즈 합계 등 주요 전력 사용량 출력 요약을 하나의 레코드로 통합하였다¹⁹⁸⁾. 이러한 메타데이터는 CSV와 Parquet 두 가지 형식으로 저장되어, 통계 분석과 대규모 머신러닝 학습 모두에서 효율적으로 활용될 수 있도록 구성하였다.

추가로, 스케줄·기상·외피·설비·출력 파일 전 단계의 재현성을 보장하기 위해 runs.jsonl와 out.err 같은 로그 파일을 함께 보관하였다¹⁹⁹⁾. runs.jsonl에는 각 sim_id에 대해 사용된 템플릿 해시(idf_hash), 기상 파일 해시(epw_hash), 실행 옵션 등이 기록되며, 이는 향후 파라미터 재설정이나 오류 재현 시 중요

196) U.S. Department of Energy, (2021), 전계서, p.19.

197) U.S. Department of Energy, (2021), 전계서 p.44-51.

198) U.S. Department of Energy, (2023), 전계서, p.138.

199) U.S. Department of Energy, (2023), 전계서, p.127.

한 근거가 된다. 이러한 메타데이터 관리 방식은 EnergyPlus-CREST 결합형 시뮬레이터를 단순한 배치 스크립트 수준을 넘어, 국가 건물 스톡 수준의 수요예측 연구를 뒷받침하는 공동주택의 전력 사용량에 대한 예측 모델로 확장하는데 기여가 가능하다. 본 파이프라인은 Generator 단계의 {1A-1E} 기능을 모두 충족시키면서, 이후 Classifier 및 Extractor 단계에서 요구하는 고품질 VSM 데이터셋을 체계적으로 공급하는 기반 인프라로 기능한다.

3.3 한국형 공동주택 가상스마트미터 데이터셋 생성 및 분석

3.3.1 데이터 생성 절차

3.3.1.1 메타데이터의 생성

대규모 VSM 데이터셋이 DR 대응과 전력 수요예측 관련 머신러닝 학습에 활용되기 위해서는 표준화된 데이터 구조와 추적성(Traceability)을 갖춘 메타데이터 설계가 필수적이다²⁰⁰⁾. 본 연구는 QSFBSM 데이터 구조를 기반으로 한국형 epw를 포함하는 확장형 스키마를 적용하였다.

1) 시간별 부하 데이터

데이터셋의 기본 단위는 시간별 부하로, 필수 필드는 다음의 [표 3-14]와 같다. 기본 해상도는 8,760 시간의 1시간 단위 해상도이며, 1일, 1개월 단위로 집계를 병행하여 저장한다. 모든 시계열 데이터는 sim_id로 메타데이터 테이블과 1:1 연결되도록 구성하였다²⁰¹⁾. 주요 변수는 난방 부하, 냉방 부하, 온수 부하, 가전 부하, 조명 부하 등의 5개 필드로 메타데이터셋을 구성하였으며 LoadProfile 기반의 Schedule로 생성하였다. 이렇게 생성된 메타데이터셋에 대한민국의 표준시각(UTC+9)과 5개의 부하를 기반으로 산출한 전체 전력 부하를 포함하여 구성하였다.

200) 박완기 (2014). 『고효율 건물 에너지관리기술 동향 및 시사점』. 동향과 분석, 한국에너지경제연구원.

201) U.S. Department of Energy, (2023), 전계서, p.195.

[표 3-14] 메타데이터셋의 설명 및 단위

필드명	설명	단위	비고
timestamp	시각 (UTC+9)	hr	1h 기본, 15 min 옵션
load_total	전체 전력 부하	kWh	$\Sigma(\text{heating} + \text{cooling} + \text{lighting} + \text{appliance} + \text{DHW})$
load_heating	난방 부하	kWh	IdealLoadsAirSystem 기준
load_cooling	냉방 부하	kWh	IdealLoadsAirSystem 기준
load_dhw	온수 부하	kWh	DHW:Schedule 기반
load_appliance	가전 부하	kWh	CREST 가전기기 기반
load_lighting	조명 부하	kWh	Lighting:Schedule 기반

2) 메타데이터 필드 구조

메타데이터는 [표 3-15]와 같이 건물 물리 변수, 기상 변수, 시뮬레이션 식별자, 정책·운영 변수로 구분된다²⁰²⁾.

[표 3-15] 메타데이터 필드 구조

분	주요 변수	단위	설명
건물 물리 변수	year_built, heating_type, cooling_type, insulation_level, floor_area, occupant_num, WWR, RSI_wall, RSI_roof, RSI_floor, Window_code_6digit	-	EnergyPlus - CREST 입력 변수 기반
기상 변수	CDD, HDD, avg_temp, avg_humidity, Sunlight, solar_radiation, epw	°C·day, %, W/m ²	epw 추출값, 시나리오별 변환 반영
시뮬레이션 식별자	sim_id, location, archetype_id, rng_seed, epw_file, idf_template, run_timestamp	-	재현성 확보
정책·운영 변수	scenario_id, dr_event_type, ess_capacity_kWh, vpp_strategy, tariff_type	-	DR 시나리오 대응 필드

202) U.S. Department of Energy, (2023), 전제서, p.131.

건물 물리 변수에 대하여 EnergyPlus와 CREST를 구동하기 위한 건물의 고유 속성값으로 구성되는 것이 메타데이터이다. 메타데이터의 필드 구조는 건물 및 기상 입력, 실행 식별자, 그리고 정책 시나리오로 분류하여 역추정이 가능하도록 설계하였다. 이는 다음 단계에서 진행되는 부하 프로파일의 역추적에 대한 추적 가능성, *seed, *idf, *epw의 입력 변수에 대한 동일성, 20만 세대에 적용하는 다양한 다중 시나리오에 대한 확장성, Extractor와 Classifier로 부하 패턴의 명확한 연결이 가능한 연계성, DR·VPP 정책 실험의 용이성을 고려하였다.

3) 데이터 저장 및 활용

CSV 형식은 범용 분석 환경 호환성을 보장하고, Parquet 형식은 컬럼 단위 병렬 입출력과 대용량 학습에 적합하다. sim_id 기반 조인을 통해 부하·메타데이터를 통합 분석할 수 있으며, 필요한 변수만 선택 로딩이 가능하다. 모든 파일에는 입력-출력 해시, EnergyPlus 버전, 실행 타임 스탬프와 함께 sim_id가 기록되어 완전한 추적성이 보장된다.

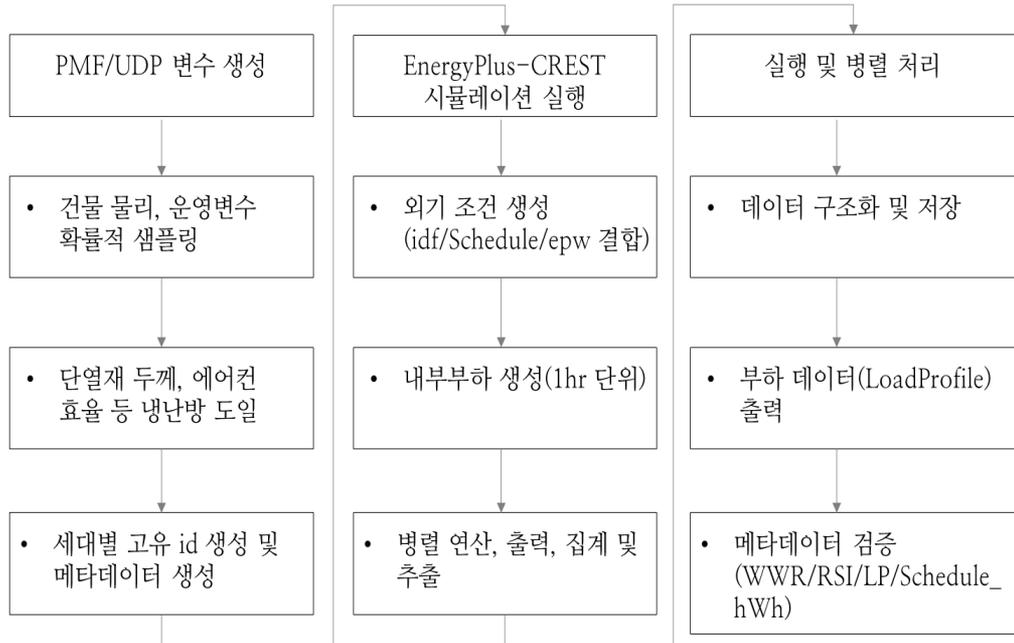
4) 응용 확장성

본 구조는 LDA-LSTM의 머신러닝 기반 Classifier 학습뿐 아니라 DR 이벤트 별 부하 감축률 평가, VPP 운영전략 최적화, VPP 집합 제어 시뮬레이션에 활용된다²⁰³⁾. epw를 포함함으로써 기후 변화 시나리오 대응 및 전력 수요예측 연구에 즉시 확장 가능하다.

3.3.2 난수 기반 입력값 생성

난수 기반 입력 데이터 생성 단계는 [그림 3-4]에서와 같이 한국형 VSM 시뮬레이션을 위해 EnergyPlus 모델에 주입 가능한 변수들을 확률론적으로 표본화하고, 각 표본에 대해 설비 파라미터, 스케줄링, 기상 변수를 자동 결합해 재현 가능한 입력 세트를 생산하는 것을 목표로 한다.

203) 김재춘, 노광현., 전계논문, p.359.



[그림 3-4] 난수 기반 입력값 생성 절차

기본 구조는 QSFBSSEM에서 제시한 Generator를 따르되, 국내 공동주택 특성, 한국형 CREST 스케줄, epw 기반 외기 조건을 통합하도록 확장하였다.

3.3.2.1 PMF/UPD 기반 변수 생성

먼저 데이터 표본화는 PMF와 UPD로 분리하여 수행하였다. PMF에는 ~1989/1990~1999/2000~2019/2020~의 건축연도, 계단식/복도식/혼합형의 건물유형, 개별난방/중앙난방/지역난방/전기난방 방식의 난방 방식, 벽걸이/스탠드/시스템 방식의 냉방 방식, 6자리 P-L-E-F-S-B 창호 유형 코드, 그리고 거주자 수 A=1~4 × 행태 BB=01~07의 로드 프로파일 코드 A-BB를 포함하였다. 이들 범주형 변수는 통계청, 국토교통부, 한국에너지공단의 통계 자료를 반영한 확률 질량 함수로 표본화되며, 특히 건축연도→벽·지붕·바닥 RSI, 창호 코드 외피 성능, 지역→RSI 분포, 난방 방식→온수 연료 등 변수 간 조건부 의존성을 명시적으로 유지하였다. UPD에는 WWR, 건물 회전

각, 침기량, 건물 총횡비의 연속 변수가 포함되며, 변수 특성에 따라 면적은 건물유형 별 ± 표준편차를 적용하였다. Stochastic에는 절단 정규 분포에서 조명 밀도는 6~20W/m² 균등 분포에서 표본화하였다. 모든 표본에는 sim_id 와 rng_seed를 부여하여 난수의 재현 가능성을 확보하였다.

3.3.2.2 EnergyPlus 입력 매핑과 스케줄·기상 결합

다음 단계로 표본화된 변수들을 EnergyPlus 입력 객체로 직접 매핑하고, CREST 스케줄 및 기상 파일을 결합해 실행 가능한 모델 세트를 구성하는 과정이다. 면적은 Zone의 바닥면적, 회전각은 Building.North Axis, WWR은 외벽 대비·창면적비로 변환되어 FenestrationSurface:Detailed로 생성하였다. 침기량은 표본화된 ELA 값을 ZoneInfiltration:DesignFlowRate로 변환하여 주입하고, 창호 유형 코드는 QSFBSSEM의 Hot2000code의 동일한 규칙을 적용하여 P-L-E-F-S-B 코딩 체계를 적용하여 U-value/SHGC/VT/Frame의 U를 록업하여 WindowMaterial:*과 FenestrationSurface:Detailed 객체에 연결하였다. 설비는 열부하 안정화를 위해 프로토타입 단계에서 ZoneHVAC:IdealLoadsAirSystem을 적용하였으며, 난방, 냉방 효율 및 온수 등 냉·난방 방식에 따라 자동 매핑하여 열부하를 안정적으로 산출하였다²⁰⁴).

내부 부하 스케줄링은 CREST를 이용해 A-BB 조합별 1시간의 8,760행 해상도의 Lighting_kWh, Plug_kWh, DHW_kWh 시계열을 생성하고, 각 항목을 Schedule:File 객체로 연결하였다. DR 시나리오나 피크 제어 분석이 필요한 경우에는 대체 스케줄을 주입하여 시나리오를 확장하였다. 외기 조건은 *.epw 파일로 주입하여 지역별, 연도별 epw를 선택하여 epw를 외기 조건 시나리오에 맞게 스케일링하여 시뮬레이션별 weather.epw를 자동 생성한다. 이때 모델 실행은 EnergyPlus.exe 기상 변수를 병렬배치로 수행하며, 기본 시간해상도는 Timestamp, 1hr 피크/제어 분석으로 설정하였다²⁰⁵).

204) U.S. Department of Energy, (2021), 전계서, p.44-51.

205) U.S. Department of Energy, (2023), 전계서, p.134.

3.3.3 EnergyPlus-CREST 시뮬레이션 및 데이터 구조화

3.3.3.1 EnergyPlus-CREST 시뮬레이션 실행

EnergyPlus-CREST 시뮬레이션 실행을 위하여 확률적으로 표본화된 PMF/UPD 입력 변수를 실제 물리 기반 시뮬레이션으로 변환하여 세대별 시간해상도 부하 프로파일을 생성하는 EnergyPlus-CREST 연동 절차를 기술한다. 이 절차는 QSFBSM의 Generator 구조와 TRNSYS-CREST 모델을 Python-EnergyPlus 환경으로 이식한 것으로, 국내 공동주택의 기상-부하 응답 특성을 정밀하게 반영하도록 구성하였다.

1) 외기 조건 모듈링 단계

기상청 기상자료 개방 포털에서 제공하는 기온·습도·일사·일조 데이터를 활용하여 생성된 EnergyPlus의 기상 데이터 입력값으로 확률 샘플링하여 epw 파일을 자동 생성하였다. Python 스크립트는 pandas-numpy-pyepw-tensorflow 패키지를 통해 온도·습도·일사·풍속·운량 데이터를 EnergyPlus 기상 입력 형식으로 변환하고, 지역별 기상 패턴을 기반으로 epw 헤더를 자동 교체한다. 이 과정은 기후 변화 시나리오 적용 시 냉·난방 부하가 기상요소 변동에 실시간으로 반응하여 DR 이벤트 분석에 필수적인 기상 반응형 모델을 형성한다.

2) 내부 부하 생성 단계

한국형 CREST 모듈은 2024년 통계청에서 발표한 「한국인 생활시간조사」 결과를 기반으로, 1시간 단위의 점유·조명·가전·온수 등의 시계열 데이터를 EnergyPlus 시뮬레이션을 통하여 산출하였다.

점유 상태는 기상 7시, 취침 11시를 기준으로 재택·취침·외출 상태를 확률로 설정하고, 활동유형은 수면·식사의 필수 활동 11시간 30분, 일·학습의 의무 활동 7시간 30분, 교제·문화·레포츠의 여가 활동을 5시간으로 세분화하여 반영하였다. 계절별·시간대별 점유 확률 및 내부 발열 부하가 동적으로 함께 반영되도록 구성하였으며²⁰⁶⁾, 이렇게 생성된 1시간 해상도의 점유·행태는 EnergyPlus

의 Schedule:File 객체에 주입되어, 냉·난방 부하 및 조명·가전·온수의 내부 모델과 함께 EnergyPlus 시뮬레이션 변수로 반영하였는데, 이를 통해 한국형 CREST는 국민 평균 생활패턴을 반영한 확률적 점유 모델로 실제 거주 행태에 근거한 현실적인 내부 부하 시계열을 생성하도록 구축하였다.

QSFBSSEM은 캐나다 퀘벡주 단독주택 20만 세대를 대상으로 CREST 기반의 확률 부하 모델을 구축하고, 실제 4만 세대의 RSM 데이터와의 부하 패턴을 비교한 결과, 일별·계절별 전력 부하 시계열에서 높은 상관성과 분포 정합성을 보고 보고하였다²⁰⁷⁾. 이러한 결과는 CREST 모델이 실제 거주 행태 기반 부하 패턴을 효과적으로 재현함을 입증한 것으로, 본 연구의 한국형 CREST 모듈 또한 동일한 방식으로 교정되어 RSM-VSM 부하 패턴의 상관 계수가 QSFBSSEM의 연구 결과와 일반적인 상관관계 수준의 수치를 참고하여 $r=0.8\sim 0.9$ 로 유지되도록 설계하였다.

3) Python 자동화 및 병렬배치 단계

EnergyPlus 입력 템플릿 idf를 Jinja2 엔진으로 매개 변수화하여 수천 세대 규모의 시뮬레이션 배치를 생성한다. subprocess, multiprocessing, asyncio 모듈을 활용해 동시 병렬 실행을 지원하며, 각 시뮬레이션에는 고유한 sim_id, rng_seed, epw_hash 값이 부여된다.

ZoneHVAC:IdealLoadsAirSystem 모델로 냉·난방부하를 산정하고, 기타 부하는 ElectricEquipment·Lights·WaterHeater: Mixed 객체로 정의된다. 계산 효율성을 위해 EP-Launch CLI를 대체하는 Python의 EnergyPlusAPI, eppy 패키지가 사용되며, 결과 파일은 *.eso, *.csv, *.err으로 자동 정리된다²⁰⁸⁾.

3.3.3.2 데이터 구조화 및 저장

EnergyPlus-CREST 시뮬레이션으로부터 산출된 결과는 LDA, Random Forest 머신러닝 학습, DR 이벤트 분석, VPP 운용 등 다양한 응용을 지원하

206) 사회통계국 사회통계기획과, 전계서, p.6-16.

207) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.6-12.

208) U.S. Department of Energy, (2023), 전계서, p.127-142.

기 위해 표준화된 VSM 데이터 구조로 저장하였다²⁰⁹). 본 연구는 QSFBSSEM의 데이터셋 설계 원칙과 TABULA 형식의 Building Archetype 메타데이터 체계를 참조하여, 부하 데이터-메타데이터의 연결 구조를 구현하였다.

1) 부하 데이터 파일

부하 데이터는 1시간 단위의 시계열 전력 부하로 구성된다. 주요 필드는 timestamp, load_total, load_heating, load_cooling, load_dhw, load_appliance, load_lighting이며, 모든 값은 kWh 단위로 표준화되었다²¹⁰). Python pandas 기반으로 피크 탐지, 계절 집계, DR 전후 비교 분석이 가능하도록 구성하였고, 1시간 단위 해상도 데이터는 LSTM 학습용 고해상도의 데이터셋으로 동시에 저장된다.

2) 메타데이터

메타데이터 파일은 각 부하 데이터와 1:1 매칭되는 입력 매개변수를 포함한다. 핵심 필드는 profile_id, archetype_id, year_built, heating_type, cooling_type, insulation_level, floor_area, occupants, WWR, RSI_wall, RSI_roof, RSI_floor 등의 물리 변수와 기상 지표 (CDD, HDD, avg_humidity, solar_radiation, epw) 를 포함한다. 이를 통해 부하 패턴-기상 특성 간 상관성 분석과 모델 재현성 추적이 가능하다. 또한 이상의 결과를 바탕으로 DR·VPP의 구현을 위하여 scenario_id, dr_event_type, ess_capacity_kWh, vpp_strategy 필드를 추가하여 수요반응 모델을 지원하도록 구성하였다.

3) 저장 형식

모든 파일에는 sim_id를 포함하여 다음 단계의 Classifier, Extractor, 단계에서 식별기로 VSM과 RSM 데이터의 식별과 호환의 유지를 목표로 모든 저장파일은 메타데이터와 기상 시나리오, 로드 프로파일, 시계열 부하 등을 포함하여 호환성 높은 파일로 처리를 하였다.

생성된 파일의 저장 형식은 CSV와 Parquet 병행 구조로 구성되었다. 일반

209) U.S. Department of Energy, (2023), 전계서, p.135.

210) U.S. Department of Energy, (2023), 전계서, p.230.

적으로 Excel, R, SPSS, Python 등에서 통계 분석 도구로 바로 호출이 가능한 CSV의 범용 호환 파일로 저장하였다. 이는 컬럼 단위 압축과 컬럼에서 필요한 변수만 읽을 수 있고, 병렬로 I/O 처리가 가능하면서 자료 보존형의 스키마 타입의 Parquet으로 처리하였다. 이는 머신러닝 훈련 속도를 최적화하고 모든 파일은 sim_id 기반으로 join 됨으로써 DuckDB 엔진을 활용해 수십만 세대의 데이터를 즉시 검색·집계할 수 있게 구성하였다.

4) 품질관리 및 검증 체계

모든 시뮬레이션에는 sim_id, rng_seed, idf_hash, epw_hash, run_timestamp, EnergyPlus_version 이 기록되어 입력-출력 추적성이 보장된다. 시뮬레이션 결과는 실측 RSM 데이터와의 정합성을 검증하기 위해 ASHRAE Guideline 14에서 제시하는 시간 해상도 모델 검증 기준을 적용하였다. ASHRAE Guideline 14는 시간별 시계열 에너지 모델의 오차 허용 범위로 MBE(Mean Bias Error) $\pm 10\%$ 이하 및 CV(RMSE) 30% 이하를 권고하고 있으며, 이는 건물 에너지 모델 보정(calibration) 연구에서 대표적인 검증 기준으로 널리 활용되고 있다²¹¹⁾. 이를 기준으로 $MBE < \pm 10\%$, $RMSE < 25\%$ 를 만족하는지를 기준으로 모델의 적합성을 확인하였다. 결과 오차는 DB에 저장되어 후속 모델 학습 시 가중치 재조정에 활용된다.

3.4 실측 전력량 데이터와의 비교 및 정합성 검증 분석

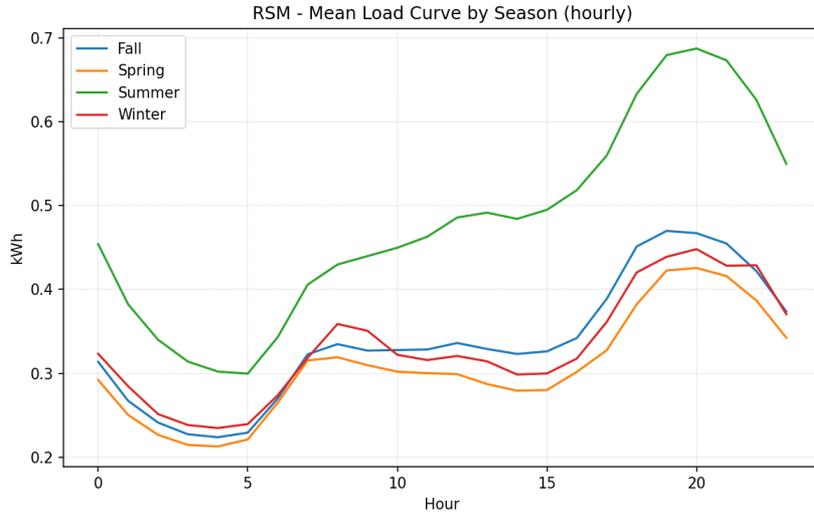
3.4.1 데이터 특성 분석 및 시각화

3.4.1.1 데이터 특성

한국형 VSM 시뮬레이션을 통해 생성된 가상 전력 부하 데이터를 활용하여, 국내 공동주택의 연간 및 계절별 부하 패턴을 분석한다. 이 분석은 QSFBSSEM의 VSM 프레임워크에서 적용된 방법론을 국내 기후·거주 특성에

211) ASHRAE, (2014). p.35, p.96-97.

적합하게 변형한 것으로, 기온·일사·습도 등 기상요소의 변화가 냉·난방 및 생활 부하에 미치는 민감도를 정량적으로 이해하기 위함이다.



[그림 3-5] RSM 데이터의 일별-계절별 전력 사용 패턴

K-apt 공동주택관리 정보시스템에서 공개하는 전국 약 12,173,146호의 공동주택의 2024년 기준 연간 전기 사용량²¹²⁾을 계절별-시간별 분석 결과, [그림 3-5]와 같이 한국의 공동주택은 일반적으로 6~9월의 하계에 급격한 냉방 피크를 보이며 24시간 기준으로는 20시를 기준으로 전력피크에 도달하는 덕 커브(Duck-curve) 형태를 보였다.

3.4.1.2 데이터 분석

여기서 전체 세대별 전기 사용 패턴은 전기 사용 행태를 기반으로 정의하였으며 각 세대의 날짜별, 시간의 전기 사용량 표준값을 시간별 상대 계수로 산정하여 24시간 기준의 전력 사용 패턴을 분석하였다²¹³⁾.

212) K-apt 공동주택관리 정보시스템. (2024). “K-apt 공동주택관리 정보시스템 운영 통계 및 아파트 관리비 자료”. <https://www.k-apt.go.kr/web/main/index.do>. 국토교통부·한국부동산원.

213) 구인석, 이성희, 손중찬, 이동희. (2020). AMI 적용 아파트의 전기사용 패턴 기반 국내외 전기요금제 분석. 『한국신학기술학회논문지』, 21(12), 52-59. <https://doi.org/10.5762/KAIS.2020.21.12.52>.

$$\bar{x}_{i,d} = \frac{1}{24} \sum_{t=1}^{24} x_{i,d,t}$$

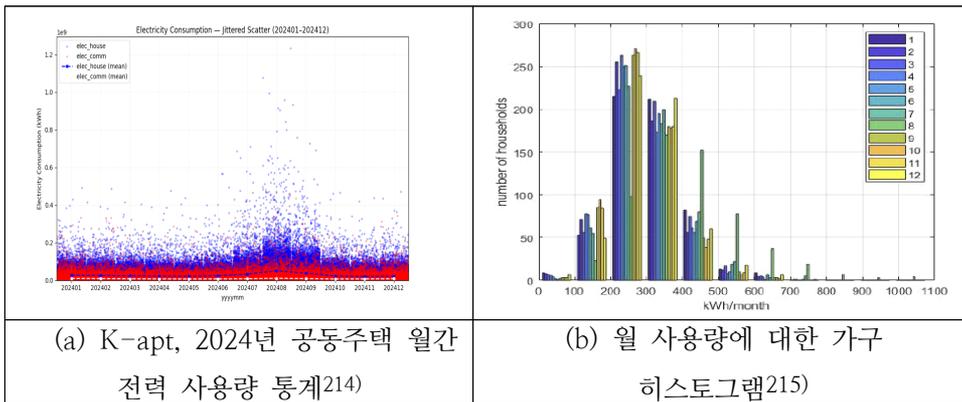
여기서,

$\bar{x}_{i,d}$: 24시간 기준 평균 전력 사용량

이러한 결과는 QSFBSSEM이 제시한 seasonal shoulder optimization 개념과 유사하지만 QSFBSSEM이 동계 시즌에 에너지 사용이 집중되는 반면 국내의 경우 하계 시즌에 전력 수요량의 피크치가 나타나고 동계 시즌은 오히려 부하 평준화가 나타나는 결과는 향후 DR 운영에서 기상 민감도를 활용한 효율적 스케줄링의 근거를 제공한다.

이는 QSFBSSEM이 캐나다 주택 모델에서 겨울철에 피크가 집중되는 특성과 차이가 있으며 이는 캐나다와의 지역적 차이가 크게 작용한 것이다. 국내의 덕커브 구조는 외기 온도와 건물 단열 성능의 상호 작용에 의해 결정되며, 또한 봄·가을(4~5월, 10월)은 냉·난방 수요가 감소하고 생활·가전 부하가 지배적인 중간 부하 구간으로 계시별 요금제에서 구분하는 것과 동일하다.

[표 3-16] 공동주택의 1년간 전력 사용량 통계



[표 3-16]의 (a), (b)에서 확인되는 바와 같이 국내 공동주택의 전력 사용은 여름철에 전력 사용량 피크를 이루고 있으며 월평균 전력 사용 통계의 경우도 200~400kWh로 분포되어 있고 주로 여름철에 전력 사용이 집중되고

214) K-apt 공동주택관리 정보시스템, 전게서.

215) 정재용, 김동식, 정범진, 정영모., 상계논문, p.863.

있음을 확인할 수 있다.

계절별 세부 분석에서는 국내 아파트를 중심으로 하는 공동주택의 지역난방공사와 중앙난방에서 온수 공급 시기와 일치를 시켜서 난방기(11~3월), 중간기(4~5월, 10월), 냉방기(6~9월)로 구분하여 평균·최대 부하, 부하 지속 시간 곡선 및 시간대별 부하 곡선을 작성하였다²¹⁶⁾.

2024년 생활시간조사 결과에 따르면, 재택 시간은 오전 7시 이전과 오후 7시 이후에 집중되어 있으며, 점심(12~13시)·저녁(18~20시)에는 가사·취식·가전 사용이 증가하는 것으로 분석되었다²¹⁷⁾. 냉방기 부하 패턴은 여름철 고온 다습 환경에서 14~17시 사이, 귀가 이후 18~22시에 가전·조명·취식 활동으로 인한 내부 발열이 중첩되어 2차 피크 저녁 부하를 형성한다²¹⁸⁾.

[표 3-17] 냉·난방기의 부하 형성 시간대 구성

구분	부하 형성 시간대
냉방기	14-17시(기상열+일사), 18-22시(활동열)
난방기	06-09시(활동 개시), 19-23시(체류 집중)

난방기 부하 패턴은 기상 이후(07~09시) 활동 개시와 귀가 이후(19~23시) 체류시간 대에 난방기기 가동률이 급증하며, 외기온 저하 구간(00~07시)에서의 피크 지속시간이 가장 길게 관찰된다²¹⁹⁾. 이는 QSFBSM에서 제시한 HDD와의 선형 상관 계수를 대표하는 모델링 계수로 $r \approx 0.85$ 를 설명하며, 실제 점유 스케줄에서 야간 체류 비율이 60%를 초과하는 시간대와 일치한다.

월별 평균 부하 패턴을 비교하면, 한국형 VSM 모델의 전력 소비는 K-apt 공동주택관리 정보시스템의 아파트 관리비를 기준으로 [표 3-17]과 같이 산정하였을 때, 6~9월 냉방 피크 시 연평균의 약 1.5배, 11~3월 난방기에는 봄(4~5월), 가을철(10월)에 비하여 1.1배의 전력 사용 패턴을 보였다²²⁰⁾.

216) 한상미, 전계논문, p.39.

217) 사회통계국 사회통계기획과, 전계서, p.12.

218) 심규한, 윤동섭, 김산, 주성관., 전계논문, p.55.

219) 심규한, 윤동섭, 김산, 주성관., 전계논문, p.55.

220) K-apt 공동주택관리 정보시스템, 전계서.

[표 3-18] 공동주택의 공동전기와 세대전기 통계

구분	공동전기	세대전기	합계
하계/봄	1.2	1.68	1.56
하계/가을	1.13	1.46	1.38
동계/봄	1.23	1.05	1.09
동계/가을	1.25	1.11	1.14
하계/(봄+가을)	1.17	1.59	1.49
동계/(봄+가을)	1.22	1.05	1.1

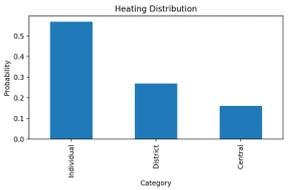
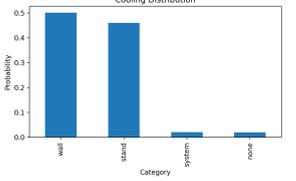
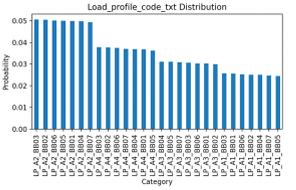
아파트의 경우 공동전기와 세대전기로 구분되며 공동전기의 경우 계절별 편차가 세대전기에 비하여 크지 않지만 [표 3-18]에서 나타난 바와 같이 1.17~1.22 수준의 차이를 확인했으며, 세대전기의 경우 계절별 편차가 눈에 띄게 확인된다. 특히 하계 구간의 편차가 봄/가을 대비 1.59배로 큰 차이를 보이며 동계구간은 아파트의 주거 형태가 집단에너지 시설화 특성상 봄가을에 비하여 큰 부하 이동의 패턴이 1.05배로 평준화 수준인 것을 확인하였다.

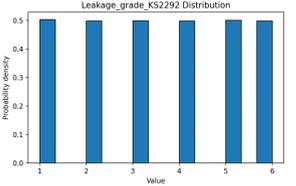
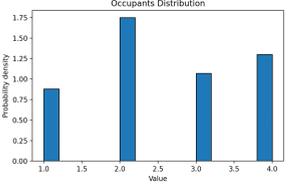
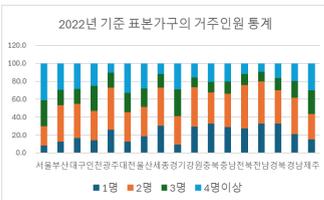
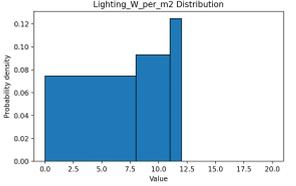
3.4.2 Archetype별 시나리오 시각화

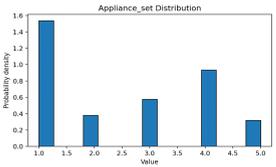
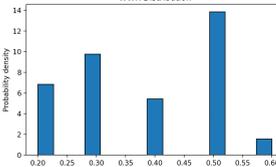
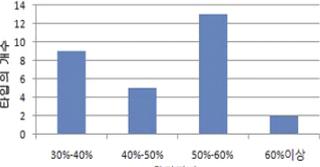
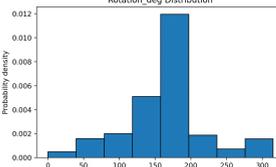
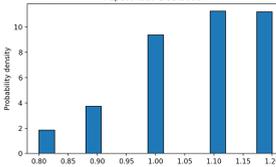
공동주택 Archetype을 대상으로 기상·부하 응답 특성을 시각적으로 비교·분석하기 위한 프레임워크를 QSFBSM에서 제시된 Generator - Classifier - Extractor 3단계 VSM 모델의 에너지 시각화 방법론을 참조하여, 한국형 공동주택의 물리적 다양성과 기상 민감도를 동시에 표현하도록 설계하였다. 각 Archetype은 건축연도, 난방 방식, 단열 수준, 전용면적, 건물 방위각 등을 기준으로 세분화되며, [표 3-19]과 같이 파라미터별 특성을 시각화하였다.

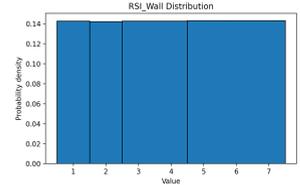
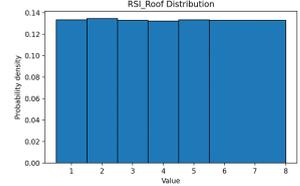
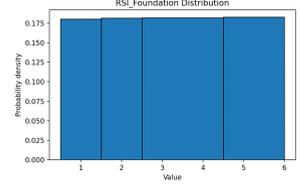
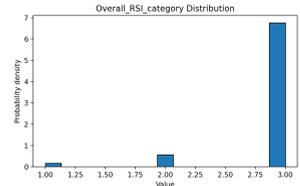
[표 3-19] VSM 생성 데이터의 파라미터별 분포

파라미터	Value	확률 분포	출처([표 3-7]과 동일 기준)	비고(출처 기준 Data)
Area	1. $\leq 40m^2$ 2. $\leq 59m^2$ 3. $\leq 79m^2$ 4. $\leq 99m^2$ 5. $\geq 100m^2 \sim$		국토교통부 2024 인구주택총조사 ²²¹⁾ 기준	
Location	1. Seoul 2. Incheon 3. Gyeonggi 4. Chungchung 5. Jeonra 6. Gyeongsang 7. Gangwon		통계청 2024년 인구주택총조사 결과 ²²²⁾	
Year of construction	1. ~1989 2. ~1999 3. ~2009 4. ~2019 5. 2020~		통계청 2024 인구주택총조사 결과 ²²³⁾	
Building Type	1. Stair type 2. Floor type 3. Mixed type		공간 데이터 마켓, 2024년 10월 아파트 단지 기준 정보 ²²⁴⁾	

<p>Heating</p>	<p>1.Individual 2.Central 3.District 4.none</p>		<p>국토교통부 통계누리 아파트 주거환경통계(225)</p>															
<p>Cooling</p>	<p>1.wall-mounted 2.stand-alone 3.system types 4.none</p>		<p>KESIS 가전·냉방 보급률, 가계 에너지·가전 보유 실태(226)</p>															
<p>Load Profile</p>	<p>Load Profile -A ; A1~A4 Load Profile -BB: BB1~BB07</p>		<p>에너지 소비 클러스터 기준(227)</p>	<table border="1"> <thead> <tr> <th>Classification</th> <th>Energy Consumption [kWh]</th> </tr> </thead> <tbody> <tr> <td>Energy consuming</td> <td>273.7</td> </tr> <tr> <td>Energy saving</td> <td>268.77</td> </tr> <tr> <td>M-pattern</td> <td>263.64</td> </tr> <tr> <td>Morning concentrated</td> <td>260.91</td> </tr> <tr> <td>Evening concentrated</td> <td>251.24</td> </tr> <tr> <td>Owl</td> <td>230.24</td> </tr> </tbody> </table>	Classification	Energy Consumption [kWh]	Energy consuming	273.7	Energy saving	268.77	M-pattern	263.64	Morning concentrated	260.91	Evening concentrated	251.24	Owl	230.24
Classification	Energy Consumption [kWh]																	
Energy consuming	273.7																	
Energy saving	268.77																	
M-pattern	263.64																	
Morning concentrated	260.91																	
Evening concentrated	251.24																	
Owl	230.24																	
<p>Window Type</p>	<p>P(1,2,3) L(0,1,2,3) E(0,1,2) F(0,1,2,3) S(0,1) B(0,1,2)</p>	<p>총 864개의 Windows Type 생성</p>	<p>QSFBSEM의 Windows 파라미터(Hot2000code)의 규칙(228)을 참고하고 국토교통부의 에너지절약형 친환경주택 건설기준서를 기준으로 함.</p>	<p>확률에 의한 균등 분포</p>														

Leakage	1. ~1, 2. 1~2, 3. 2~3, 4. 3~4 5. 4~5 6. 5~ m ³ /hm ²		국토교통부, 2023 에너지절약형 친환경주택 건설기준해설서(229)230)	확률에 의한 균등 분포
Occupants	1. 1 occupant 2. 2 occupants 3. 3 occupants 4. 4 occupants~		KESIS 국가 에너지 정보시스템 231)232) 기준.	
Occupancy Activity	1~28	Occupants + Load Profile 기준으로 총 28가지 Activity 생성	Clustering of Load Profiles of Residential Customers Using ²³³⁾ +	“LoadProfile 조합으로 생성”
DHW	1~28	Occupants + Load Profile 기준으로 총 28가지 Activity 생성	CREST 규칙 적용	
Lighting Density	1. Low 2. Mid 3. High 4. Super		건설교통부, 2025 에너지절약형 친환경주택 건설기준해설서(234)	확률에 의한 균등 분포 => 2009년 이전 건축물의 경우 CFL과 백열 전구 사용 비율 증가로 조명 밀도 값이 10 이상에서 일부 상승 반영됨

Home Appliance	BB01 BB02 BB03 BB04 BB05 BB07	 <p>Appliance_set Distribution</p>	Clustering of Load Profiles of Residential Customers Using ²³⁵⁾ + KESIS 국가 에너지 정보시스템 236)	KESIS 국가 에너지 정보 시스템의 주요 가전기기 보급률 수치와 동일
WWR	1: 0.2 2: 0.3 3: 0.4 4: 0.5 5: 0.6	 <p>WWR Distribution</p>	차폐계수와 창면적비에 따른 공동주택의 건물 에너지 효율 등급 평가 ²³⁷⁾²³⁸⁾	 <p>타입의 개수</p> <p>창면적비</p>
Rotation	1: 0° 2: 45° 3: 90° 4: 135° 5: 180° 6: 225° 7: 270° 8: 315°	 <p>Rotation_deg Distribution</p>	KESIS 국가 에너지 통계 종합 정보 시스템 13차(2022년 기준) 가구에너지 패널조사 결과 ²³⁹⁾	KESIS 국가 에너지 정보 시스템의 전국 아파트 표준가구 배치 통계 기준 동일 적용
Aspect ratio	1. 0.8 2. 0.9 3. 1.0 4. 1.1 5. 1.2	 <p>Aspect Ratio Distribution</p>	설계 표본 조사, 단지 평면 통계-> Neale et. al.의 shape 파라미터 생성규칙을 동일하게 적용 ²⁴⁰⁾	QSFBSM 기준 적용

Conditional (m2KW-1)	Wall RSI	[1.0] (0.5,1.5) [2.0] (1.5,2.5) [3.0] (2.5,4.5) [5.0] (4.5,6.0) [7.0] (6.0,7.5)		건축물의 에너지절약 설계기준해설서(최종본) 241)	확률에 의한 균등 분포
	Roof RSI	[1.0] (0.5,1.5) [2.0] (1.5,2.5) [3.0] (2.5,3.5) [4.0] (3.5,4.5) [5.0] (4.5,5.5) [8.0] (5.5,8.0)		건축물의 에너지절약 설계기준해설서(최종본) 242)	확률에 의한 균등 분포
	Foundatio n RSI	[1.0] (0.5,1.5) [2.0] (1.5,2.5) [3.0] (2.5,3.5) [4.0] (3.5,4.5)		건축물의 에너지절약 설계기준해설서(최종본) 243)	확률에 의한 균등 분포
Derived (m2KW-1)	Overall building thermal resistance	1: <1.56 2: <2.25 3: ≥2.25		건축물의 에너지절약 설계기준해설서(최종본) 244)	

221) 조사관리국 인구총조사과, 전계서, p.76.

222) 조사관리국 인구총조사과, 전계서, p.70.

223) 조사관리국 인구총조사과, 전계서, p.77.

-
- 224) 공간데이터마켓(2024). “2024년 10월 아파트단지 기준정보”. LX한국토지정보공사(주)펜타코드. Retrieved from https://geomarket.kr/user/dataset/view.do?data_sn=25.
- 225) 국토교통 통계누리. (2023). “아파트주거환경통계 현황(2023)”. 국토교통부. Retrieved from https://stat.molit.go.kr/portal/cate/statView.do?hRsId=33&hFormId=&hDivEng=&month_yn=.
- 226) 남수현, 이현. (2023). 『가구에너지패널조사 연구』. 한국에너지경제연구원(KESIS). Retrieved from https://kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=view.
- 227) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K. (2021). Clustering of load profiles of residential customers using extreme points and Demographic Characteristics. *Electronics* 2021.10(290).
- 228) Neale, A., 전계논문, p.62.
- 229) 주택건설공급과, (2023), 전계서, p.126.
- 230) 박원석, 윤재욱. (2003). 신축공동주택의 기밀성능 실측에 관한 연구. 『한국생태환경건축학회논문집』, 3(3), 43-52.
- 231) 국가에너지통계 종합정보시스템, (2023), 전계서.
- 232) 남수현, 이현., 전계서, p.57.
- 233) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.
- 234) 주택건설공급과, (2025). 『에너지절약형 친환경주택 건설기준 해설서』. 국토교통부.
- 235) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.
- 236) 국가에너지통계 종합정보시스템, (2023), 전계서.
- 237) 장철용, 한혜심, 이진숙., 전계논문, p.39-40.
- 238) 건설안전과, (2012), 전계서, p.8.
- 239) 국가에너지통계 종합정보시스템, (2023), 전계서.
- 240) Neale, A., Kummert, M., & Bernier, M., 전계논문, p.53, p.88.
- 241) 건설산업과 국가건설기준센터, (2023), 전계서, p.18, 134, 154, p.242-248.
- 242) 건설산업과 국가건설기준센터, (2023), 전계서, p.18, 134, 154, p.242-248.
- 243) 건설산업과 국가건설기준센터, (2023), 전계서, p.18, 134, 154, p.242-248.
- 244) 건설산업과 국가건설기준센터, (2023), 전계서, p.18, 134, 154, p.242-248.

3.4.3 부하 로직 생성과 한국형 VSM 데이터 생성

3.4.3.1 점유 기준의 내부 부하 생성

내부 부하를 Occupant, Occupancy Activity, DHW, Lighting Density, Home Appliance의 다섯 파라미터로 규정하여 8,760hr의 1시간 해상도 기준의 부하 프로파일을 구성하였다. 개별 파라미터는 서로 독립적으로 생성되고, 행태(Occupancy Activity)와 체류 수준, 계절·시간대, 주중·주말 효과를 매개로 상호 연동되며, 최종적으로 조명·가전·온수의 상대적 기여도를 가중합으로 통합 내부 부하(Occupant-driven Physiological & Technical internal load, OPT)를 산출한다. 우선 점유 인원수(Occupant)는 이산 분포를 1인(17.7%), 2인(34.9%), 3인(21.4%), 4인(26.1%)로 정하고²⁴⁵⁾, 각 인원수 별 부하 스케일을 가중평균 1.0이 되도록 정규화한다.

이 스케일링 계수는 이후 모든 하위 프로파일을 ACT(Occupancy Activity Profile, 점유자 활동 부하 프로파일), DHW (Domestic Hot Water Load, 온수 부하 프로파일), LIT (Lighting Load Profile, 조명 부하 프로파일), APP (Home Appliance Load Profile, 가전기기 부하 프로파일)의 기본 파라미터를 일관되게 생성²⁴⁶⁾하였으며, 표본 전체에서 통계청 기준의 2024 인구주택총조사, 2024년 생활시간조사 결과를 기준으로 인구 구성의 편향이 부하량에 과도하게 반영되지 않도록 반영하였다²⁴⁷⁾²⁴⁸⁾²⁴⁹⁾.

점유자 행태(Occupancy Activity)는 7개의 BB01~BB07의 24시간 패턴을 기반으로 구성하였다. 기본형은 아침·저녁형, 오전 집중형, 야간형, 노년 주간 체류형 등 일상적 생활 리듬을 대표하도록 구성하였다²⁵⁰⁾. 시간대별 활동 강도는 (가) 냉방기(6-9월) : 14~17시, 18~22시 증폭, (나) 난방기(11-3월) : 06~09시, 19~23시 증폭, (다) 봄·가을기(4-5월, 10월) : 정오~오후 시간대의

245) 국가에너지통계 종합정보시스템, (2023), 전게서, “다. 가구원수에 관한 사항”.
246) 이수진, 김유정, 진혜선, 서인애, 김성임, 하수연, 송승영., 전계논문, p.157-160.
247) 조사관리국 인구총조사과, 전게서, p.10, p.15, p.49.
248) 사회통계국 사회통계기획과, (2025), 전게서, p.6-16.
249) 김유란, 홍원화, 서윤규, 전규엽., 전계논문, p.46-48.
250) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.5, p.8.

완화 등 계절·시간대에 가중하였다²⁵¹). 이와 별도로 주말에는 체류 증가를 반영하여 동일 패턴을 소폭 증폭하였다. 이 과정은 일사·기상열 등 계절적 외기 영향과 사회적 시간 사용의 결합 효과를 해석 가능한 방식으로 VSM 데이터 셋에 내재화를 목표로 수행한 것이다.

온수(DHW) 프로파일은 행태 피크와의 동시성을 중심으로 하였다. 아침(6~9시)과 저녁(18~22시)에 상대적 강도의 차이를 두고, 활동 강도가 높을수록 동시계수가 상승하도록 반영하였다²⁵²). 난방기는 상대적으로 DHW 강도를 완만하게 계절적 수요 특성을 반영하였으며, 이는 기 생성된 DHW 시계열 값을 0~1의 값으로 정규화하여 상대적 행태를 유지하면서, 절대 규모는 인원수 스케일에 의해 조정되도록 반영할 것이다²⁵³).

조명(Lighting Density)은 등가 조도 밀도 등급을 Low/Mid/High/Super(〈8, 8~11, 11~14, 14~20 W/m²〉) 4 범주로 단순화하여, 등급별 상대이득을 부여한다. 일출-일몰을 모사한 시간대를 기준으로 UK Domestic Lighting Project Preliminary Analysis에 따르면 가중 야간에 Lighting Density가 가장 높은 1.0을 기준으로 할 때, Evening (19:00~23:30:1.0), Night(00:00~05:30:0.48), Shoulder (AM/PM:0.6), Daytime (10:00~13:30:0.07)을 참고²⁵⁴)로 실제 변수 반영은 계절별/월별/시간별 통계를 기반으로 저녁(1.0), 야간(0.5), 주간(0.07)로 조정하고 저녁 시간대와 주말에는 체류와 활동 증가를 반영해 소폭 증폭하여 반영하였다²⁵⁵). 결과적으로 조명 프로파일은 시간대 × 조명 등급 × 체류의 곱 구조를 가지며, 최대값 정규화 후 인원수 스케일을 적용하였다.

가전기기(Home Appliance)는 기본 행태(BB)의 대표 사용 시간대를 저녁형은 18-23시, 오전형은 7-12시, 야행형은 심야 등으로 seed를 설정한 뒤, 귀가 이후 활동 열 증가, 난방기 야간 체류 증가 등의 주말·계절 보정을 적용하였다²⁵⁶). 더불어 체류 및 매체 활동 그리고 취사 활동의 시간 분포와 연동하여, 실제 생활시간조사에서 관찰되는 가전 사용의 저녁 집중, 주말 확대, 식

251) 사회통계국 사회통계기획과, 전게서, p.6-16.

252) 이수진, 김유정, 진혜선, 서인애, 김성임, 하수연, 송승영., 전계논문, p.157-160.

253) 김유란, 홍원화, 서윤규, 전규엽., 전계논문, p.46-48.

254) Martin, C., Tebbit & J., Bernstein, L., 전게서, p.11.

255) 김유란, 홍원화, 서윤규, 전규엽., 전계논문, p.46-48.

256) 심규한, 윤동섭, 김산, 주성관., 전계논문, p.271.

사 준비와의 결합을 상대적 강도로 반영하여 최종 시계열은 조명과 동일하게 0-1 정규화 후 인원수 스케일로 확정하여 생성 기준으로 반영하였다.

마지막으로 통합 내부 부하(opt)는 조명·가전·온수의 상대 행태를 고정 가중으로 선형 결합하여 산출한다. 본 연구에서는 내부 부하로 조명(LIT), 가전(APP), 급탕(DHW)으로 구성되는 점유 기반 내부 부하의 상대적 기여도를 반영하여 선형 결합 방식으로 산정하였다. 설명 가능성과 튜닝의 용이성을 고려해 QSFBSM 및 CREST 기반 VSM 생성 연구에서 제시된 내부 부하 구성 원리와 상대적 전력 기여 관계를 기준으로 설정하였으며, 이는 공동주택의 에너지 사용량에 대한 선행 연구를 기반으로 연중 상시 부하 특성을 갖는 가전을 기준(1.0)으로 조명과 급탕은 상대적 비중을 각각 0.6과 0.3으로 설정하여 정규화 하였다²⁵⁷⁾²⁵⁸⁾. 이 가중은 RSM-VSM 데이터 적합도 또는 분석 목적에 따라 조정 가능하며, 조정 시에도 각 성분이 피크 시각, 완만부 등의 시간 형태를 유지하도록 모두 상대 정규화 후 합성한다.

산출물은 $A \in \{1 \sim 4\}$, $BB \in \{01 \sim 07\}$, $LD \in \{Low, Mid, High, Super\}$ 의 조합별로 8,760시간의 시계열 CSV 5종(ACT, DHW, LIT, APP, OPT)을 생성하며, 파일 경로와 메타정보를 기록한 LoadProfilesindex를 함께 제공한다²⁵⁹⁾²⁶⁰⁾. 이러한 규칙 기반 생성 체계는 ① 인구 구성의 변화, ② 생활 행태의 시나리오 변경, ③ 조명 밀도 등급 조정, ④ 계절·시간대·주말 시간대 효과를 모듈화하여, 현실적 변동 요인을 해석 가능한 매개변수로 관리할 수 있게 한다.

3.4.3.2 외생 변수의 연결

국내 7개 광역군으로 서울, 경기, 인천, 충청, 전라, 경상, 강원으로 분류하였다. 제주도는 공동주택 주거 비율이 전체의 30% 수준으로 타 광역군에 비하여 현저히 낮았으며²⁶¹⁾ 공동주택의 전력 사용량을 분석하는 본 연구와 조건에 부합되지 않는다는 판단으로 제주도를 제외한 7개 광역군으로 지역을

257) 이수진, 김유정, 진혜선, 서인애, 김성임, 하수연, 송승영., 전계논문, p.155.

258) Neale, A., (2021), 전계논문, p.3.

259) 이수진, 김유정, 진혜선, 서인애, 김성임, 하수연, 송승영., 전계논문, p.155.

260) U.S. Department of Energy, (2023), 전계서, p.139.

261) 조사관리국 인구총조사과, 전계서, p.49.

확정하고, 7개 광역군에 대하여 기상청에서 제공하는 기상 데이터와 OneBuilding.org에서 확보한 표준 기상 데이터 epw를 결합하여 EnergyPlus 시뮬레이션용 기상 반응형 입력 변수 세트를 생성하였다²⁶²).

이 과정은 외생 기상 변수->epw 스케일링->품질검사->입력 데이터 구조화의 네 단계로 진행하며 데이터의 시간 해상도는 모든 변수에서 1년의 기간에 대한 1시간 단위의 해상도인 8,760 시간으로 통일하였다.

1) 월천데이터 수집 및 정합

기상 입력의 일관성을 확보하기 위하여, 각 광역군 별로 OneBuilding.org에서 제공하는 7개 지역별 표준 epw 파일을 기준 데이터로 활용하였다²⁶³. 이 파일은 EnergyPlus 형식으로 구성되어 있으며, 기온(DryBulb), 습도(RelHum), 일사(Global/Direct/Diffuse), 풍속, 운량, 일조시간 등의 변수를 포함한다.

다음은 기상청에서 제공하는 전국 7개 광역군 별 날씨 정보와 epw 데이터를 동시에 각각 광역군 별로 지역의 시간 단위 데이터를 이용하였다. 이와 별도로 기상청 기상 자료 개방 포털에서 제공하는 기상 관측 정보를 활용하였으며 상대 습도는 0~100%, 일사량은 MJ/m² 단위로 클램프 처리하였다²⁶⁴. 이상의 세 가지 데이터는 모두 한국 표준시(KST) 기준으로 정렬하였으며²⁶⁵, 시간해상도를 1시간으로 맞춘 후, 결측값 t_n 은 이전 t_{n-1} 값과 이후 t_{n+1} 값을 이용해 공백을 직선으로 선형으로 메우는 전후 보간법으로 보정하였다.

전후 보간법의 산정 방정식은 다음과 같다.

$$\alpha = \frac{t_n - t_{n-1}}{t_{n+1} - t_{n-1}}$$

시간 축을 α 에 맞춰두고

$$x(t_n) = (1 - \alpha) \times x(t_{n-1}) + \alpha \times x(t_{n+1})$$

로 결측값 t_n 을 이전 x_{n-1} 값과 이후 x_{n+1} 두 점을 잇는 직선 위의 값으로

262) OneBuilding.org. (2024). "Repository of Building Simulation Climate Data". Retrieved from <https://climate.onebuilding.org>.

263) OneBuilding.org. (2024), 상계서.

264) 국가기후데이터센터, 전계서, p.39.

265) TimeAndDate.com. (2024). "Korea Standard Time (KST)". Retrieved from <https://www.timeanddate.com/time/change/south-korea/seoul>.

보정하는 방식으로 산정되며 결측치 이전의 마지막 점과 결측치 이후의 첫 점 사이의 보간을 수행한다²⁶⁶⁾²⁶⁷⁾.

2) 외생 기상 변수 파생 및 스케일링 설계

epw는 건물 에너지 시뮬레이션 프로그램인 EnergyPlus에서 사용하는 표준 기상 데이터 파일 형식으로, 현재는 EnergyPlus뿐만 아니라 ESP-r, IESVE, TAS, DesignBuilder 등 건물 에너지 시뮬레이션 소프트웨어에서 널리 사용되는 표준 포맷, epw의 변수에 기상 데이터를 적용하여 EnergyPlus용 기상 입력을 생성하였다²⁶⁸⁾.

(가) 기온(Dry Bulb Temperature) 입력 변수 생성

EnergyPlus는 외기 조건으로 건구온도를 시간 단위의 외생 입력 변수로 사용하며 해당 값은 epw 기상 파일에 포함된 시계열을 그대로 읽어서 시뮬레이션 경계조건을 적용하기 때문에 EnergyPlus 실행 이전 단계에서 epw 입력 시계열을 생성하는 방식이 권장된다²⁶⁹⁾. 이와 관련하여 기상청에서 제공하는 지상관측지침에 따라 건구온도는 관측 지점의 대표 기온으로 동일 지역 내에서도 관측 위치, 지형, 고도, 도시화 정보에 따라 편차가 발생할 수 있음을 명시하고 있다²⁷⁰⁾. 이에 따라 본 연구에서는 epw 파일의 건구온도 시계열을 기본값으로 사용하되, 기상청의 지상기상관측자료에서 도출된 시간별 온도 편차를 외생 신호로 해석하여 epw 입력 온도를 다음의 방정식으로 보정하였다.

$$T_{epw}^*(t) = T_{epw}(t) + \alpha \cdot \Delta T_{KMA}(t)$$

여기서,

$T_{epw}^*(t)$: EnergyPlus에 투입할 보정된 건구온도(°C)

$T_{epw}(t)$: epw 건구온도 시계열(°C)

$\Delta T_{KMA}(t)$: 기상청 지상기상관측자료에서 확보된 시간별 온도 편차(°C)

α : 온도 보정 가중 계수

266) Lepot, M., Aubin, J.-B., & Clemens, F. H. L. R. (2017). Interpolation in time series: An introductory overview of existing methods, their performance criteria and uncertainty assessment. *Water* 2017, 9(769).

267) 이성근, 전계서, p.46-65.

268) Chong, A., Gu, Y. & Jia, H., 전계논문, p.5.

269) U.S. Department of Energy, (2021), 전계서, p.29-31.

270) 관측정책과, (2022), p.8.

이상의 방정식으로 EnergyPlus에 입력하는 외기 건구온도 변수를 생성하였다. 이는 EnergyPlus의 변수 입력구조를 유지하면서 기상청의 관측된 기상정보를 시뮬레이션에 반영하기 위한 합리적인 절차이다.

(나) 상대 습도(Relative Humidity) 입력 변수 생성

EnergyPlus는 외기 습도를 상대 습도 또는 이슬점 온도 중 하나로 입력 변수를 사용한다. EnergyPlus 표준문서에 따르면 상대 습도는 독립적인 예측 변수가 아니라 외기 건구온도와 수증기 상태에 의해 결정되는 파생 기상 변수로 공기 관계식에 기반하여 계산된다²⁷¹⁾.

기상청의 지상기상관측지침에서도 상대 습도는 공기중 실제 수증기량과 해당 온도에서 수증기량의 비율로 정의되며, 관측 자료에서 상대 습도 또는 이슬점 온도가 제공되는 경우 두 변수는 물리적으로 상호 변환 가능성을 제시하고 있다²⁷²⁾. 이에 따라 본 연구에서는 외기 기상 보정 변수로 물리적 일관성을 유지하기 위하여 지상기상관측자료에 포함된 이슬점 정보를 기준으로 상대 습도를 다음의 방정식으로 재산정하였다.

$$RH_{epw}^*(t) = 100 \times \frac{e(T_d(t))}{e_s(T_{epw}(t))}$$

여기서,

$RH_{epw}^*(t)$: EnergyPlus 입력 상대 습도(%)

$T_d(t)$: 기상청 지상기상관측자료에서 확보된 이슬점 온도(°C)

$e(T_d)$: 이슬점의 실제 수증기 분압

$e_s(T)$: 온도 T에서의 포화 수증기압

$T_{epw}(t)$: 보정된 건구온도

$T_{epw}^*(t)$: 보정된 외기 건구온도(°C)

이상의 방정식으로 EnergyPlus 입력값으로 주입되는 상대 습도를 산정하였다. 이러한 접근은 EnergyPlus가 요구하는 입력 변수 정의를 충족하면서 기온 보정으로 인한 습도 왜곡을 방지하기 위한 절차이다.

(다) 일사량 (Diffuse / Global Radiation) 입력 변수 생성

EnergyPlus는 태양복사 효과를 고려해 수평면 일사량을 외생 변수로 사용하며, 해당 값은 epw 파일에 포함된 시간별 일사 시계열을 시뮬레이션 경계

271) U.S. Department of Energy, (2021), 전계서, p.31-32.

272) 관측정책과, (2022), p.29.

조건으로 직접 적용한다. EnergyPlus 표준문서에 따라 태양 복사량은 외부 환경조건으로 정의되는 입력 변수로 취급되며, 시뮬레이션 내부에서 일사량을 자체 재보정하는 기능을 제공하지 않으며 EnergyPlus 실행 이전 단계에서 입력 시계열을 합리적으로 가공하여 제공하는 것을 요구하고 있다²⁷³⁾.

기상청의 지상기상관측지침에 따르면, 일사량은 구름량, 대기 투과도, 계절적 태양 고도의 변화에 따라 시간별 변동성이 큰 기상요소이며, 관측 지수 또는 상대 지표를 활용하여 기준 시계열의 상대적 증감만을 반영하는 방식에서의 적용을 명시하고 있으며²⁷⁴⁾, 이를 기준으로 다음의 방정식으로 산정하였다.

$$G_{epw}^*(t) = \min(\gamma(t) \times G_{epw}(t), 4.89) [MJ/m^2]$$

여기서,

$G_{epw}^*(t)$: EnergyPlus 입력 일사량($MJ/m^2 \cdot h$)

$G_{epw}(t)$: epw 파일의 일사 시계열($MJ/m^2 \cdot h$)

$\gamma(t)$: KMA에서 제공하는 일사 지수

$\min(x, 1, 4.89)$: 상한/하한 절단 함수

4.89 : 1시간 기준 이론적 최대 일사 에너지 상한값

[표 3-20] 일사량 절단 함수 상한값의 산정식²⁷⁵⁾²⁷⁶⁾

평지 기준 평균 일사량 $1361(W/m^2)$

$1 MJ/m^2 = 10^6 J/m^2$

$1hr = 3600 s$

$1 W/m^2 = 1J/(s \cdot m^2) \Rightarrow 1hr$ 로 변경하면 : $3600J/m^2 = 0.0036MJ/m^2$

$\therefore 1361(W/m^2) = 1361 \times 0.0036 = 4.89MJ/m^2$

본 연구에서는 물리적으로 과도한 일사값이 입력되는 것을 방지하기 위하여 이상의 방정식에서 일사량의 상한값을 명시적으로 제한하였다. 태양상수를

273) U.S. Department of Energy, (2021), 전계서, p.35-37.

274) 관측정책과, (2022), p.40.

275) Wikipedia, (2024). "Solar irradiance". Retrieved from https://en.wikipedia.org/wiki/Solar_irradiance

276) Sengupta, M., Habte, A., Gueymard, C., Wilbert, S., & Renné, D. (2017). Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications (2nd ed.). Golden, CO: National Renewable Energy Laboratory (NREL). Technical Report NREL/TP-5D00-68886.

1,361W/m²을 기준으로 1시간 단위의 적분 에너지 단위로 일사량의 상한값은 [표 3-20]과 같이 환산하였다.

이 값은 수평면 기준 1시간 동안 물리적으로 가능한 최대 일사 에너지의 이론적 상한값에 해당되며, 이 값을 기준으로 스케일링된 일사 입력값이 해당 범위를 초과하지 않도록 재현함으로써, EnergyPlus 시뮬레이션 실행 시 비현실적인 일사 조건이 입력되는 것을 방지하였다.

3.4.3.3 품질검사(QA) 및 검증

변환된 epw 파일은 다음 항목을 기준으로 자동 검증하였다.

① 연속성: 8,760시간 시계열 누락 여부 점검, ② 물리적 타당성: 온도, 습도, 일사량 등 물리 한계 내 존재 여부, ③ 통계적 일관성: 기준 epw 대비 히스토그램 비교, ④ 평균 기상 조건 : 온도·습도·일사량 월평균의 차이가 $\pm 10\%$ 범위 유지 등 이상 4가지의 검증을 통과한 데이터만이 EnergySaver 입력으로 등록하고 이를 통해 각 지역별 epw 파일이 과도한 변형 없이, 실제 기상 지수 변화가 반영된 신뢰도 높은 입력 변수로 사용하였다.

3.4.3.4 EnergyPlus 시뮬레이션 설정

EnergyPlus 모델은 생성된 7개 광역군 epw 중 해당 지역 파일을 호출하여 시뮬레이션을 수행하여 eplusout.sql, eplusout.csv 파일로 저장된다.

EnergyPlus 분석 모듈은 각 시뮬레이션 결과로부터 8,760hr의 시간 단위의 시뮬레이션 해상도로 통일되어 냉·난방 부하, 내부 발열, 전력 사용량을 추출하며, 기상 조건 변화에 따라 에너지 소비의 동적 응답을 비교·분석하였다.

모든 실험 설정은 config.json으로 관리되며, 광역군 명칭 <-> epw 경로의 파일 매핑, 시뮬레이션 기간, Timestamp, 냉·난방 Setpoint, EMS 적용 여부 등이 기록되어 재현과 검증이 용이하도록 구성하였다.

이상과 같이 광역군별 기상 특성과 epw의 변동이 냉·난방 부하 및 에너지 소비에 미치는 영향을 정량적으로 분석할 수 있도록 외생 변수를 추가하였으

며 모든 입력 파일은 1시간 단위 해상도와 동일한 구조 및 형식을 유지함으로써, 향후 DR 기반 에너지 시뮬레이션 확장에도 일관된 데이터 프레임워크로 활용될 수 있도록 생성의 원칙을 준수하였다.

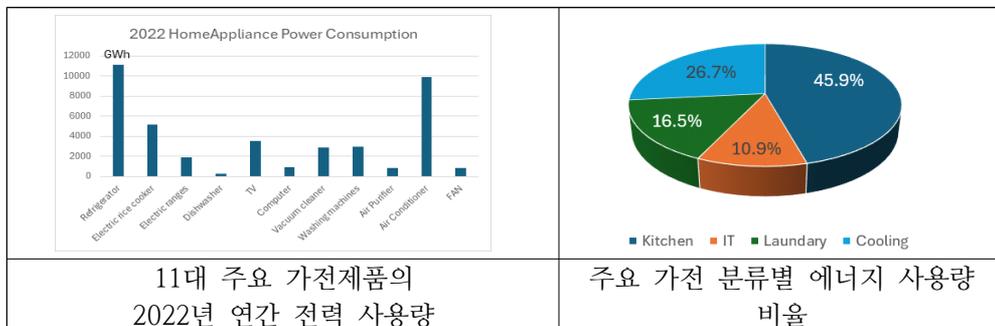
3.4.3.5 가전기기 변수의 확장

BB01~BB07 의 점유 기반 부하 모형이 가지는 한계를 보완하기 위하여, 가정 내 전기기기의 세부 사용 패턴과 계절적 특성을 반영한 Home Appliance 부하 항목을 확장하였다. 이 확장은 국내 공동주택 전력 사용 패턴 연구²⁷⁷⁾와 KESIS 국가 에너지 정보시스템(2022)의 가구별 전기기기 보유 통계 분석을 종합하여 설계하였다.

이를 통해 가전기기 부하를 행태-기기-계절이 상호작용하는 내부 발열 모델로 구조화하였으며, EnergyPlus 시뮬레이션의 Schedule, Internal Gain 입력 변수로 연계하였다.

1) 가전기기 분류 및 부하 특성 정의

[표 3-21] KESIS 가구에너지패널조사 결과표 기준 주요 가전기기 통계²⁷⁸⁾



277) Seo, Y.-K., & Hong, W.-H., 전계논문, p.227-228.

278) 국가에너지통계 종합정보시스템, (2023), 전계서, “4. 가전기기이용현황”.

먼저 [표 3-21]의 KESIS 가구에너지패널조사 결과표에 따라 가구 내 전력 사용량의 비중이 높은 11대 주요 가전기기에 대한 통계로 냉장고가 27.6%의 전력 소비로 가장 높게 통계되어 있는데 이는 24시간 365일 가동되는 특성을 확인할 수 있다²⁷⁹⁾.

[표 3-22] 가전기기군 사용 패턴 정의

기기군	세분류	기기의 용도	전력 사용 패턴
조리·주방기 기(Kitchen Equipment)	냉장고, 전기밥솥, 전자레인지, 식기세척기 등	24시간 가동되는 상시 부하와 식사 준비 시 단시간 고출력 기기가 혼합되어 일일 부하 곡선의 기저부를 형성	식사 시간대(07~09시, 18~21시)에 피크부하가 집중되는 특성
문화·정보기 기 (Cultural/IT)	TV, 컴퓨터, 휴대전화 충전기 등	정보·오락 목적의 저출력 기기	저녁 시간(19~24시)대의 여가 및 미디어 활동과 연동되어 전력 사용이 집중
세탁·위생기 기 (Laundry & Hygiene)	세탁기, 청소기, 공기청정기 등	주거 공간과 거주자의 위생을 목적으로 사용되는 기기	주말의 가동 빈도가 평일 대비 약 4% 증가 오전(08~11시) 시간대에 집중
냉·난방기기 (Heating & Cooling)	에어컨, 선풍기	계절성 기기	냉방기에는 평균 +24%, 난방기에는 +3%의 부하 가중 계수를 적용하여 계절에 따른 부하 이동 모사

KESIS 가구에너지패널조사 결과표(2022)의 통계 데이터와 에너지소비 클러스터 기준의 선행 연구²⁸⁰⁾를 바탕으로, 국내 보유된 대표 가전기기를 11종을 선정하고 가구당 보급된 통계를 기반으로 2022년 기준 가전기기의 연간 전력 사용량 데이터를 기기군 클러스터를 [표 3-22]과 같이 정하였다²⁸¹⁾.

279) 국가에너지통계 종합정보시스템, (2023), 전제서, “4. 가전기기이용현황”.

280) Seo, Y.-K., & Hong, W.-H., 전제논문, p.227-228.

281) 남수현, 이현., 전제서, p.77.

각 가전기기 군의 보유율은 가구 내 보유되는 모든 가전기기의 통계가 어려운 관계로 KESIS(2022)의 전국 11대 주요 가전기기의 가구 보급률을 기준으로 샘플링하여 EnergyPlus 내부 발열(Internal Gain) 비율로 매핑하였다.

[표 3-21]의 KESIS 가구에너지패널조사 결과표²⁸²⁾에 따르면 주요가전기기 중 냉장고가 27.6%의 전력 소비로 가장 높게 통계되어 있는데, 이는 24시간 365일 가동되는 특성이 반영된 결과로 보인다²⁸³⁾.

특이 사항은 주요 가전기기에서 에어컨의 전력 사용량이 냉장고 다음으로 높은 24.5%의 전력 소비를 차지하고 있는데, 이는 6~9월의 하계 시즌에 집중되는 에어컨의 계절적 사용 특성을 고려할 때 짧은 구간의 큰 에너지 소비비중으로 향후 DR 기반 에너지 시뮬레이션 확장과 수요예측을 통하여 여름철 수요 반응에 의미 있는 결과를 얻을 가능성을 확인할 수 있다²⁸⁴⁾.

2) 한국형 에너지 소비 클러스터 기반의 가전 부하 모듈 구축

[표 3-23] 가전 부하 모듈의 단계별 구축²⁸⁵⁾

1단계 (정의)	2단계 (시간 가중)	3단계 (정규화)
Cluster Definition	Temporal Weighting	Output Normalization
냉방형(AC·선풍기 중심), 난방형(전열기기 중심), 생활형(조리·문화 기기 혼합)	- .BB01~BB07 별 대표 사용 시간대 기기의 가동시간을 조정 주/말 보정(+4%), 계절 보정(냉방기 +24%, 난방기기 +3%)을 추가	1), 2)단계에서 생성된 시계열은 0~1 범위로 정규화한 뒤, 인원수(Occupant) 스케일 $A \in \{1\sim4\}$ 로 확장
시간대별 가동 확률 분포 함수 $P(h)$ 를 구성하고, 출근, 귀가, 주말 패턴으로 가동 시간대의 변동성을 정의	- .2024년 생활시간조사 (TUS, 2024)의 체류율 데이터와 미디어 시청 등의 매체 활동의 시계열 분포를 함수형으로 곱셈 연산하여 최종 부하 가중을 결정	8,760시간의 가전 부하 시계열은 EnergyPlus의 Schedule:File로 저장되며, ElectricEquipment:InternalGains 항목에 직접 연결로 시뮬레이션에 활용

282) 국가에너지통계 종합정보시스템, (2023), 전제서, “4. 가전기기이용현황”.

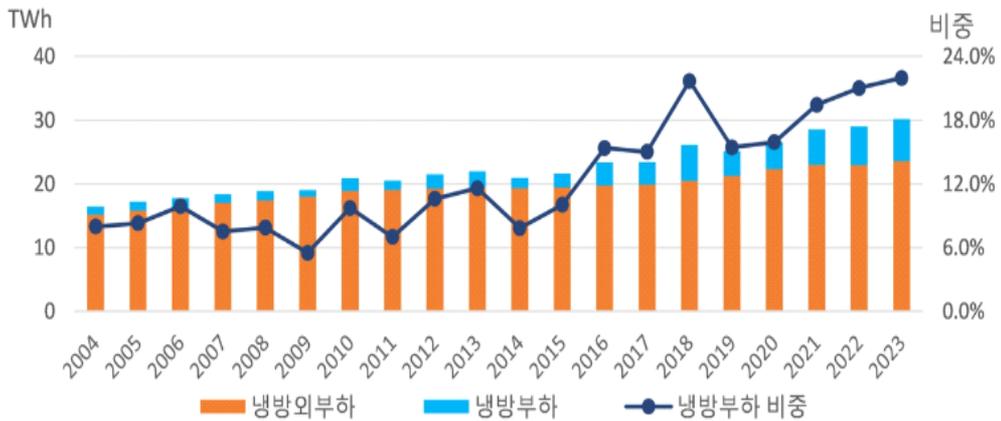
283) 남수현, 이현., 전제서, p.77.

284) 남수현, 이현., 전제서, p.77.

285) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전제논문, p.227-228.

가전기기의 시간적·계절적 사용 패턴을 반영하기 위해, 한국형 전기 소비 클러스터 분석의 연구 결과를 바탕으로 가전 부하는 [표 3-23]과 같이 3단계 프로세스를 거쳐 EnergyPlus 입력으로 주입하였다.

2024 생활시간조사 결과 주말에 체류시간 및 매체 활동이 증가하는 것으로 나타났으며²⁸⁶⁾, 본 연구에서는 이를 반영하기 위하여 가전, 조명 등 부하 스케줄에 선행 연구에서 6개 단지 6,992가구의 공동주택의 시간별 전력 소비 패턴 분석에 대한 연구의 결과로 가구당 평균 주중 평균 442Wh의 전력을 소비하는 것 대비 일요일에 19Wh의 전력 사용 증가에 대한 연구 결과를 기준으로 주말 가동을 +4%로 상향 조정하여 주말 보정 계수로 적용하였다²⁸⁷⁾.



[그림 3-6] 하계(6~9월) 가정용 전기 소비 분해 및 냉방용 비중

국가 에너지 통계 종합 정보 시스템의 발표에 따르면 [그림 3-6]에서 확인할 수 있듯이 지난 10년간 냉방 도일이 지난 30년간 난방도일 대비 평균 20% 증가한 것으로 발표하였으며 이와 함께 지난 20년간 6월~9월의 하계 구간의 가정용 전기 소비 분해 및 냉방용 비중에 대한 분석에 따르면 2023년 말 기준 냉방 부하 비중은 24%로 발표된 기준을 반영하였다²⁸⁸⁾.

286) 사회통계국 사회통계기획과, 전계서, p.43-55, p.78-98.

287) 임경엽, 윤종현, 신우철, 박재성, 김광식. (2011). 실측자료 기반의 공동주택 시간별 전력 소비 패턴 분석 연구. 『한국에너지학회논문집』, 20(3), 127-132.

288) 에너지수급전망연구실. (2024). 기온 시나리오에 따른 2024년 여름 주택용 전력 수요 추정 및

겨울철 난방기기의 경우 아파트라는 공동주택의 특성상 지역난방이나 중앙 난방의 경우 별도의 추가적인 난방기기를 사용하지 않는 특성이 통계치로 나타나고 있으며 개별난방의 경우 난방기기의 추가가 있지만 실제로는 도시가스, LPG 등 GAS 사용 지출이 30% 가까이 높게 나타났으며 공동주택의 전 열기기를 통한 난방기기에 2,667원 수준의 비용 증가를 조사로 발표한 기준에 따라 +3%를 동계 시즌의 계절 가중을 추가하였다²⁸⁹⁾.

2단계까지 산출된 가전 부하의 원시 시계열은 직접 EnergyPlus 시뮬레이션에 입력할 경우 크기 범위와 에너지 데이터 합치성 문제를 해결하고자 ① 부하 시계열의 범위를 0-1로 정규화하고, ② 점유자에 따른 스케일의 차이 반영, ③ EnergyPlus에 사용하는 설계 용량과 분수형 스케줄을 생성하고, ④ 연간 총에너지 목표와의 합치성을 보장함으로써 8,760시간 해상도를 생성했다.

3) 범위의 표준화

폭염, 한파 등 극한 기상 이벤트로 인한 이상치에 민감하지 않도록 견고한 백분위 정규화를 EnergyPlus의 데이터 규칙을 기준으로 데이터 정규화를 위해 $clip(x,0,1)$ 연산의 적용을 통한 x 함수를 하한:0, 상한:1의 절단 연산으로 모드 시점의 값이 0~1 범위 내에 EnergyPlus 입력용 기준형 분수 스케줄의 Fraction Schedule($A(t)$)로 원시 시계열($\tilde{A}(t)$)이 안정적으로 분포하도록 선행 연구의 Min-Max scaling normalization을 참고하여 시계열의 최소 전력값 $\min(P_{\min})$ 과 최대 전력값 $\max(P_{\max})$ 를 기준으로 다음과 같이 정규화하였다²⁹⁰⁾.

$$A(t) = clip\left(\frac{\tilde{A}(t) - P_{\min}}{P_{\max} - P_{\min}}, 0, 1\right),$$

이는 단순 최대값 기준 정규화보다 극단적인 피크 값에 덜 민감하여 시간적 형상을 보존한 채 0-1의 스케줄링 형의 Fraction으로 스케줄의 형상만 반영하며 부하의 크기나 점유자 수 등은 반영되지 않은 EnergyPlus에 입력데이

전기요금 영향 분석. 『에너지 브리프』. 2024(7). 한국에너지경제연구원.
 289) 윤태연 & 강재성. (2015). 난방 방식이 가구의 겨울철 난방비 지출에 미치는 영향 분석: 도시가스 개별난방과 지역난방 방식을 중심으로. 『에너지경제연구』, 14(2), 243-272.
 290) Kim, Y.-S, Kim M.-K, Fu, N., Liu,JJ., Wang,J. & Srebric, J. (2025). Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models. Sustainable Cities and Society, 118,

터로 주입이 가능한 형태로 구성하였다²⁹¹⁾.

4) 점유자 스케일 반영

가구 점유자 수는 내부 발열 크기에 구체적으로 영향을 미치므로 1~4인 가구별 스케일 벡터 γ_A 로 시간적 패턴에 영향을 주지 않고 스케일 크기만을 조정하는 계수로 적용하였다.

KESIS의 2022년 가구에너지패널조사 결과표를 바탕으로 점유 인원별 기본 스케일값은 각각 0.8, 0.95, 1.1, 1.25로 설정하였으며 설정된 기준은 다음의 [표 3-24]와 같다. 통계청의 2024 인구주택총조사 결과를 기준으로 전국 가구 구성비 1인(17.7%), 2인(34.9%), 3인(21.4%), 4인(26.1%)를 가중²⁹²⁾하여 전체 평균이 1이 되도록 정규화하였다.

[표 3-24] 점유 인원별 스케일값의 설정

점유 인원	에너지 소비량 (MCal)	1인 가구 기준 정규화	스케일 계수	γ_A 확정	비고
1인	7,912	1(기준)	1.00/1.21375 =0.82	≈0.85	과소 평가 방지 : -2%
2인	9,412	+19.9%	1.199/1.21375 =0.99	≈0.95	모델 평가 기준 95%로 조정
3인	10,334	+30.6%	1.306/1.21375 =1.08	≈1.1	근사치로 반올림
4인	10,675	+35.0%	1.350/1.21375 =1.11	≈1.25	인원수 증가에 따른 동시 가동률 보정치로 1.25로 조정
에너지 비율 $R=[1.00, 1.199, 1.306, 1.350]$; ‘xx인/1인’의 비율 $R_{average}=(1.00+1.199+1.306+1.350)/4=1.21375$ $\gamma_A=R/R_{average}$					

291) U.S. Department of Energy, (2023), 전계서, p.47-49.

292) 국가에너지통계 종합정보시스템, (2023), 전계서, “다. 가구원수에 관한 사항”.

점유 스케일이 반영된 최종 Fraction Schedule(분수 스케줄, $S_A(t)$)은 0~1의 스케줄링 형의 EnergyPlus에 입력 데이터로 주입이 가능한 형태로 다음과 같이 정의하였다²⁹³).

$$S_A(t) = clip(\gamma_A \times A(t), 0, 1),$$

여기서,

γ_A : 가구별 스케일 벡터,

$S_A(t)$: 분수 스케줄,

이를 통해 가구의 인원수가 증가할수록 내부 부하 크기가 선형적으로 확대되지만 시간적 부하 영향은 동일하게 유지하도록 설계하였다.

5) 연간 에너지 보정

연간 에너지 보정식은 EnergyPlus의 내부 부하 산정 구조에 근거하였으며, EnergyPlus는 조명, 가전기기 부하를 설계 용량(Design Level)과 시간별 Fraction Schedule의 곱으로 계산하며, 연간 에너지는 해당 부하를 시간 적분하여 산정한다. 이에 따라 스케줄의 상대적 시간 형상은 유지하되, 선행 연구에서 제시된 연간 에너지 비중을 기준으로 주방(45%), 냉방(27%), 위생(16%), IT(6%), 기타(1%)로 목표값으로 설정하고²⁹⁴²⁹⁵), 이를 8,760시간의 시뮬레이션의 총합과 일치시키기 위해 설계 용량에 스케일 계수 k 를 적용하였다. 이러한 방식은 EnergyPlus의 계산 구조를 변경하지 않으면서도, 가구 면적, 점유자 수, 기기 구성 차이에 따른 연간 소비량을 합리적으로 반영할 수 있는 표준적인 보정 방법이다²⁹⁶²⁹⁷).

$$P(t) = W \times S_A(t),$$

$$E = \sum P(t) \Delta t,$$

$$= \sum_t (W \times S_A(t)) \Delta t,$$

$$W = k \times W_0, \quad k = \frac{E_{target}}{\sum_t (S_A(t) \Delta t)}$$

293) U.S. Department of Energy, (2023), 전계서, p.47-49.

294) 국가에너지통계 종합정보시스템, (2023), 전계서, “4. 가전기기이용현황”.

295) 남수현, 이현., 전계서, p.77.

296) U.S. Department of Energy, (2023), 전계서, p.31-33.

297) ASHRAE. (2021). ASHRAE Handbook: Fundamentals SI Edition, ISSN 1523-7230.

여기서,

W_0 : 설계 초기 용량[W/m²],

W : 보정된 설계 용량[W/m²],

E : EnergyPlus 시뮬레이션으로 계산된 연간 에너지,

E_{target} : 기준 에너지 총량,

$\Delta t=1h$,

k : 스케일 계수,

이는 동일한 부하 형성을 유지하면서도 가구 면적이나 가전기기 구성, 점유자 수에 따라 연간 kWh 총량이 재현되도록 보정하였다.

6) EnergyPlus 입력 변수 매핑 및 구성 절차

본 연구에서 생성된 가전 부하 시계열 데이터는 EnergyPlus 시뮬레이션 환경에 효과적으로 통합되도록 주요 입력 객체와 단계에 따라 체계적으로 연계하였다.

가전 부하의 시계열 특성을 반영하기 위해, 정규화된 가전 부하 스케줄은 Schedule:File 객체 형태로 구성하고, 이를 각각의 zone의 가전 소비 전력은 ElectricEquipment:Zone 객체를 통해 계산되어 반영된다. 또한 계절별 냉·난방 특성에 따른 가전 부하의 상대적 변화를 반영하기 위해, 외기 조건 및 냉·난방 운전 특성을 고려한 보정 계수를 적용하였다.

[표 3-25] EnergyPlus 입력 객체 항목

항목	EnergyPlus 객체 명	설명
가전 소비 전력 스케줄	ElectricEquipment:Zone	Zone별 가전기기 소비 전력 스케줄(정규화 0~1) 적용
계절별 부하 조정 계수	Schedule:File	냉방·난방기 계절 보정 계수를 반영하여 가전 부하 조정
상시 대기 전력	OtherEquipment	가구 기저 부하(약 0.3 kW)를 상시 부하로 추가
EMS 제어 (선택적 적용)	EnergyManagementSystem:Actuator	폭염 및 한파일 기간 동안 냉·난방기기 부하 가중 제어

이를 통해 동일한 가전 사용 패턴이라 하더라도 계절에 따라 상이한 부하 영향이 EnergyPlus 시뮬레이션 결과에 반영되도록 [표 3-25]과 같이 각각의 객체별로 구성하였다.

또한, 폭염 및 한파와 같은 극한 기상 조건에 대한 대응 가능성을 검토하기 위해, 일부 시나리오에서는 EnergyPlus의 EMS를 활용한 제어 구조를 통해 특정 기상 조건에서 가전, 냉·난방 부하의 상대적 가중 조건을 추가할 수 있도록 하였다.

한편, 모든 입력 파일과 객체 명명 규칙은 수용가 특성을 반영한 단일 식별 체계로 통일되며, 시계열 데이터 파일과 *.idf 입력 객체 간의 자동 매핑 구조를 통해 대규모 시뮬레이션의 일관성을 재현하였다. 이러한 입력 구성 결과는 가전 부하 시계열이 건물 에너지 시뮬레이션 내에서 전기 부하로 일관되게 반영되도록 하여, 이후 수요반응 기반의 전력 수요 예측과 VPP 운영 시나리오 분석의 신뢰성을 제고하고자 하였다.

7) 대기 전력의 합성

가구의 상시 대기 전력은 별도의 OtherEquipment 객체로 정의하여 항상 1.0의 고정된 분수 스케줄로 임의로 가정하여 정의하였다. 상시 부하의 경우 Energy Sustainability & Society의 발표에 따르면 국제적으로 일본과 중국은 가정용 전기의 약 10%가 상시 부하로 조사 결과를 발표하였다²⁹⁸⁾²⁹⁹⁾. 국내의 경우 서울시에서 공개한 가구당 전력 사용량 통계를 보면 최근 10년간 월 약 300kW의 전력을 사용하는 것으로 공개되어 있으며, 공동주택의 경우 가정 내에서 사용하는 가구의 상시 부하와 아파트의 공동전기, 그리고 환기 및 공조 제어기기, 엘리베이터 등의 동 단위의 상시 부하가 존재하는 관계로 이를 가구당 평균 전력으로 환산하여 상시 부하를 300Wh로 설정하였다³⁰⁰⁾³⁰¹⁾.

이렇게 정의된 상시 부하는 시간 가중 스케줄 $S_A(t)$ 와 독립적으로 합성되어

298) Murata, A., Kondou, Y., Hailin, M., & Nomura, N. (2016). Electricity consumption in Japanese households: Empirical analysis based on detailed household data. *Energy Sustainability & Society*, 6(1), 1–13.

299) Zhou, N., McNeil, M., Fridley, D., Lin, J., & Price, L. (2018). Residential electricity consumption patterns in China. *Energy Sustainability & Society*, 8(1), 1–15.

300) 에너지 정보. (2024). “서울특별시 가구당 에너지사용정보”. 서울특별시 녹색에너지과 . Retrieved from <https://energyinfo.seoul.go.kr/energy/energyConsumption?menu-id=Z111000>.

301) Bulbaai, R. R., & Halman, J. I. M. (2023). Standby energy consumption and saving potentials in the residential sector in tropical areas: The Caribbean island Curaçao as a case study. *Energy Efficiency*, 16, 88. <https://doi.org/10.1007/s12053-023-10170-0>.

실제 공동주택의 기저 부하의 특성을 재현하고 생성된 시계열은 [표 3-26]과 같이 검증 절차를 수행하였다. 여기서 모든 부하는 8,760hr의 1시간 해상도로 통일되며, 점유 행태(BB), 주말 효과, 계절 보정, 체류율, 매체 활동 등 다차원 변수가 결합된 확률적 스케줄로 구성된다.

[표 3-26] EnergyPlus 입력 객체 검증 절차

구분	단계별 수행의 내용
1단계	모든 시점에서 $SA(t) \in [0,1]$ 의 범위가 유지되도록 확인하고 NaN과 음수가 없도록 필터링 수행
2단계	설계 보정 후 시뮬레이션의 연간 부하 총량 $\sum tLoad(t)\Delta t$ 가 연간 목표 kWh와 $\pm 1\%$ 이내로 일치하도록 검증
3단계	주중/주말/계절 구간별 평균 부하 곡선과 비교하여 2단계에서 설정한 기간 가중의 저녁 집중, 주말 확장, 냉·난방기 가중 등의 규칙의 유지 확인

8) 확장 모듈의 의의 및 활용

본 확장 모듈은 단순히 가정 내 전기기기의 소비량을 합산하는 수준을 넘어, 행태(Behavior)-기기(Device)-계절(Season) 간 상호작용을 반영한 내부 발열(Internal Heat Gain) 연동형 부하 모델을 한국형 주거 가구의 실제 전력 소비 특성이 EnergyPlus 시뮬레이션 입력 단계에 정량적으로 반영하여 구축한 점에서 의의가 있다. 이러한 구조는 냉·난방 부하뿐 아니라 내부 발열에 의한 에너지 수요 변동을 동적·정량적으로 분석할 수 있게 하며, 향후 DR기반의 수요예측 평가 예측 모델 연계 등 다양한 연구 확장에도 활용될 수 있다.

3.4.4 한국형 VSM 데이터 생성

캐나다의 QSFBSSEM에서 제안한 VSM 생성 알고리즘을 기반으로, 국내 공동주택의 스톡 및 주거 특성에 적합한 한국형 가상스마트미터 데이터셋, KR-VSM을 구현하였다. 코드의 실행은 Jupyter Notebook Version 7.0.8,

JupyterServer 2.14.1 환경에서 Python 3.12 Version으로 수행하였다. 주요 모듈은 Pandas, Numpy, tqdm 등을 호출하여 시뮬레이션 기반의 대용량 데이터 생성을 자동화하였다.

생성 알고리즘은 3.2절의 공동주택 기반 Archetype 세분화를 바탕으로 건축물의 물리 정보, 통계청의 인구주택총조사, 건설교통부의 국가건축물에너지 절약 설계기준, QSFBSSEM의 선행 연구 기준 캐나다 Hot2000:건물 기밀 성능 등급, 통계청의 2024년 생활시간조사 결과를 기반으로 가전기기 부하 특성의 확률적 샘플링의 개별 파라미터 조건을 기반으로 최초 7개 광역군별로 각각 100개씩 700가구의 샘플데이터를 생성한 후 생성된 결과에 대하여 품질검사 및 검증 절차를 거친 후 개별 광역군의 데이터를 2024년 인구주택총조사 기준으로 전국 공동주택 주거 통계의 비율에 따라 총 200,000개의 가상 데이터로 확장되어 생성하도록 설계하였다.

특히 가구원 수 기반 부하 패턴, 조명 밀도, 창호 코드, 기밀도 등은 PMP 기반의 확률 조건으로 정의하여 건축연도별 에너지 성능의 확률적 다양성을 반영하였다. 개발된 한국형 KR-VSM은 가상의 총 200,000 세대를 대상으로 난방, 냉방, 조명, 기기, 온수의 5가지 엔드유즈 전력량을 확률적으로 산정하였으며 전국을 서울, 경기, 인천, 충청, 전라, 경상, 강원으로 하는 7개 광역군으로 매핑하였다. 각 세대의 연간 전력 소비량은 다음의 조건으로 수행하였으며 제주도는 전체 가구 중 공동주택이 차지하는 비율이 30% 선으로 전국 평균 대비 현저히 낮아서 본 연구에서 공동주택을 중심으로 하는 한국형 KR-VSM의 취지에 부합되지 않는다는 판단으로 생성 단계에서 제외하였다.

가구 산출 변수 중 냉방 부하(E_{cool_kWh})는 24°C 기준의 지역별 CDD를 이용하여 연간 냉방열 부하를 근사한 뒤, 하계(6-9월)에 대해 1.5배의 시즌 가중치를 적용하고 냉방기의 평균 성능 계수(COP)도 동일하게 1.5로 고정하여 전력량으로 변환하였다. 난방 부하(E_{heat_kWh})는 공동주택의 주요 난방 연료가 비전기식임을 고려하여 전열기, 전기장판 등 보조 난방 부하만을 포함하여, 가구원 수와 부하 패턴(BB01-BB07)에 따른 1인당 약 25 kWh/월의 월평균 전열 전력에 동계(11-3월)의 1.1배의 계절 가중치를 적용하였다³⁰²⁾.

302) K-apt 공동주택관리 정보시스템, (2024), 전계서.

조명 부하(E_{light_kWh})는 주거 형태 및 준공 연식에 따른 조명 밀도 (LPD, W/m^2)와 부하 패턴 별 점유시간(3.0~4.1 h/일)을 곱하여 산정하였고, 가전기기 부하(E_{equip_kWh})와 온수 부하(E_{DHW_kWh})는 통계청(2024) 생활시간조사 및 KESIS 가전 사용량 통계에 근거한 가구원 수(1~4인) 기준의 선형함수를 적용하였다. 냉·난방 부하는 지역 CDD/HDD와 창면비(WWR), 누기량 등 물리 변수를 반영하여 연간 전력량을 산정하였다.

RSM과의 스케일 정합은 `apply_kwh_correction`에서 수행하였다. 해당 절차는 ①가전기기·조명·냉방·난방·온수의 가구 단위 별 평균 비중을 KESIS 국가 에너지 통계 정보 시스템에서 발표한 가정 부문 에너지 추이 및 현황에서 집계된 주택용 전력의 소비 비율에 따라 동일한 비율로 66.7 %, 13.3 %, 12.9 %, 5.1 %, 2.0 %의 기준으로 구현하였다³⁰³). ② 전 집단의 연간 평균 총 전력량이 3,400[kWh/세대·년]의 75%인 2,550kWh가 되도록 비율에 따라 생성 조건을 반영하였으며 이는 서울시에서 발표하고 있는 서울특별시 에너지 사용 정보 기준에 따라 2024년 연간 가구당 전력 사용량 통계치인 3,400kWh에서 공동전기 25%를 제외한 75%를 기준으로 설정하였다³⁰⁴). 이때 이상치는 상위 백분위 절단으로 완화하였다.

- 가구당 평균전력 : 3,400kWh/세대·년,
- 가전기기 : 1,700kWh, 조명:344kWh, 냉방:328.95kWh,
난방:130kWh, 온수:51kWh.

여기에서 가구당 평균 전력 사용량은 [표 3-17]에서의 결과와 같이 전체 공동주택 전기 사용량의 25%가 공동전기로 구성되어 있는 통계를 기반으로 가구당 평균 전력 사용량은 3,400kWh의 75%인 2,550[kWh/세대·년]으로 정하였다³⁰⁵). KR-VSM에서 적용한 전력 사용량 변수 중 난방에 대한 전력 사용의 경우 국내 공동주택의 99%의 난방 형태가 개별난방, 중앙난방, 지역난방으로 집중되어 있고 별도 난방용 전력의 경우 일부 가구는 전기매트 시공을 한 공동주택이 일부 있기도 하지만 설치 가구가 미미하고 일반적으로는

303) 국가에너지통계 종합정보시스템, (2023), 전게서, “4. 가전기기이용현황”.

304) K-apr 공동주택관리 정보시스템, 전게서.

305) K-apr 공동주택관리 정보시스템, 전게서.

전기장판이, 전기매트, 온수매트 등으로 제한된 전력을 사용하는 것을 고려하여 전력 사용량의 10%를 임의로 반영하였다.

이와 같은 절차를 통해 생성된 한국형 KR-VSM 은 QSFBSM 구조를 기반으로 하되, 국내 AMI의 실측 데이터인 RSM과의 정합을 거친 통계적 스케일 보정모델로 완성되었다. 최종적으로 생성된 KR-VSM 데이터셋은 EnergyPlus 및 CREST 시뮬레이션과의 연계를 통해 8,760시간의 시계열 가상 부하를 제공하며, DR 기반 전력 수요예측 모델의 학습용 고정밀 데이터로 활용 가능한 기반을 구축하였다.

3.5 실측 데이터와의 통계적 검증

3.5.1 통계적 정합성 지표(RMSE, MAE, MBE)

본 연구에서는 VSM 결과와 RSM 간의 적합성을 평가하기 위하여, 평균 제곱근 오차(RMSE, Root Mean Squared Error), 평균 절대 오차(MAE, Mean Absolute Error), 그리고 평균 편향 오차(MBE, Mean Bias Error)의 세 가지 핵심 통계 지표를 활용하였다. 이러한 지표들은 특히 건물 에너지 시뮬레이션(BES, Building Energy Simulation) 모델의 검증·보정(calibration) 단계에서 신뢰성 평가를 위한 표준화된 수단으로 자리매김되어 왔다³⁰⁶⁾³⁰⁷⁾.

이렇게 적용한 통계 지표는 다음과 같다.

3.5.1.1 RMSE(Root Mean Squared Error)

RMSE는 시뮬레이션 값과 실측값 사이 오차(residual)를 제공한 뒤 평균을 내고 다시 제곱근을 취한 값이다. 수식으로는 다음과 같다³⁰⁸⁾.

306) Tomrukçu, G., Kızıldağ, H., Avgan, G., Dal, A. Ö., Sağlam, N. G., Özdemir, K., & Ashrafian, T. (2014). A systematic approach to manual calibration and validation of building energy simulation models. *Energy and Buildings*, 80, 415–427. <https://ssrn.com/abstract=4488363>.

307) 이경호, 주홍진, 안영섭, 이왕제., 전계논문, p.202–203.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tilde{y}_i - p_i)^2},$$

여기서,

\tilde{y}_i : i 번째의 시뮬레이션 값,

p_i : i 번째 실측값,

N : 총관측치 수,

RMSE는 오차가 클수록 제곱 항이 커지므로 큰 오차에 더 민감하게 반응하며 일부 극단치(large errors)가 결과에 큰 영향을 미칠 수 있다는 특징이 있다³⁰⁹⁾.

건물 에너지 모델 검증에서는 RMSE가 낮을수록 시뮬레이션이 실측을 잘 따라간다고 평가하고 있으며 에너지 소비에서 RMSE가 $\pm 10\%$ 이내로 기준을 수립하기도 하였다³¹⁰⁾³¹¹⁾.

본 연구에서는 시간별, 일별, 월별 및 계절별 부하 곡선과 총량 비교에 있어서 RMSE를 주요 정량 지표로 삼아, 시뮬레이션-실측 간 오차분포와 피크 오차 특성을 함께 분석하였다.

3.5.1.2 MAE(Mean Absolute Error)

MAE는 오차의 절대값을 모두 더한 뒤 평균한 것으로, 다음과 같이 정의된다.

308) 이지영, 선영규, 이선민, 김수현, 김영규, 이원섭, 심이삭, 김진영. (2019). GRU 기반 전력 사용량 예측을 적용한 스마트미터기 구현. 『한국인공지능학회』 19(5), 93-99. <https://doi.org/10.7236/JIIBC.2019.19.5.93>.

309) Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15, 5481-5487. <https://doi.org/10.5194/gmd-15-5481-2022>.

310) Emami, P., Sahu, A., & Graf, P. (2024). BuildingsBench: A large-scale dataset of 900K buildings and benchmark for short-term load forecasting. 37th Conference NeurIPS 2023 Datasets and Benchmarks Track. Preprint.

311) Mondal, S. (2023). "Understanding MAE, MSE, and RMSE: Key metrics in machine learning. DEV Community". Retrieved from <https://dev.to/search?utf8=%E2%9C%93&q=Understanding+MAE%2C+MSE%2C+and+RMSE%3A+Key+Metrics+in+Machine+Learning>.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\tilde{y}_i - y_i^{obs}|$$

여기서, \tilde{y}_i : i 번째의 시뮬레이션 값

y_i^{obs} : i 번째 실측값

N : 총 관측 치수

부하 오차의 과대/과소를 고려하지 않고 크기만 평가하므로, 해석이 직관적이라는 장점이 있으며, 제곱을 사용하지 않으므로 극단 치에 대한 민감도가 RMSE보다 낮다³¹²⁾³¹³⁾.

본 연구에서는 MAE를 통해 시뮬레이션 값이 평균적으로 얼마만큼 떨어져 있는지를 파악하고, RMSE와의 비교를 통해 극단 오차 존재도 진단하였다.

3.5.1.3 MBE(Mean Bias Error)

MBE는 오차의 평균값으로, 오차의 방향성(bias)을 나타내며 수식은 다음과 같다.

$$MBE = \frac{1}{N} \sum_{t=1}^N (\tilde{y}_i - y_i^{obs})$$

여기서, \tilde{y}_i : i 번째의 시뮬레이션 값

y_i^{obs} : i 번째 실측값

N : 총관측치 수

MBE가 양(+)이면 시뮬레이션이 실측보다 과대 추정(over-estimation)을 하고 있다는 의미이며, 음(-)이면 과소 추정(under-estimation) 상태를 나타낸다. 건물 에너지 모델 검증에서는 일반적으로 절대값 MBE가 $\pm 10\%$ 이내면 허용하도록 설정하는 사례가 있다³¹⁴⁾.

본 연구에서는 MBE를 통해 모델의 총량 편향을 진단하고, 특정 기간 또는 계절·주 등 시나리오에서 과대/과소 추정 경향을 판별하였다.

312) Hodson, T. O., 전계논문, p.5481.

313) Mondal, S., 전계서.

314) ASHRAE, (2014), 전계서, p.24-26.

3.5.1.4 해석 및 적용 범위

MAE는 평균 오차 규모를 보여주고, RMSE는 큰 오차에 대한 민감도를 반영하며, MBE는 편향 방향을 나타낸다. 따라서 $RMSE > MAE$ 인 경우에는 극단 치가 존재하거나 모델이 일부 시점에서 크게 벗어나 있음을 시사한다³¹⁵⁾. 건물 에너지 시뮬레이션 맥락에서는 시간별 또는 계절별 곡선 형태의 정합성이 중요하므로, 이 지표들을 시간·일·월의 시간 축과 지역·아키타입의 집단 축으로 분해하여 분석하는 것이 권장된다. 실제로 최근 연구에서는 시간 해상도에서의 RMSE와 MAE를 모두 제시하고 있으며 본 연구 역시 동일한 방식으로 접근하였다³¹⁶⁾. 본 연구에서는 시간별 비교 시 RMSE, MAE, |MBE|를 선행 연구를 참조하여 각 $\pm 10\%$ 이내의 지표값 기준으로 판단하였다³¹⁷⁾.

MBE의 경우 각 지역·아키타입 별로 과대·과소 추정 여부를 색인화하여 보정이 필요한 군집을 식별하였다. 이와 같은 지표 체계를 통해 가상 시뮬레이션이 실제 건물 운영 데이터를 얼마나 신뢰성 있게 재현하고 있는지 정량적으로 평가하고, 이후 보정(calibration) 및 추정 모델 개발 단계로 이어질 수 있는 토대를 마련하고자 하였다.

DR 수요예측에 있어서 상대적으로 중요한 시간대의 피크 부하를 얼마나 잘 예측하고 있는지와 오차가 특정 시간대나 조건에 치우치지는 않았는지를 점검하고 입력 불확실성(input uncertainty)이나 데이터 이상치(outlier) 등이 지표값에 큰 영향을 줄 수 있으므로, 비교 전에 데이터 정제 및 동일 시간 정렬이 필수적이다.

315) StackExchange. (2012). "Mean absolute error OR root mean squared error? CrossValidated (StackExchange Statistics)". Retrieved from <https://stats.stackexchange.com/questions/48267/mean-absolute-error-or-root-mean-squared-error>.

316) Johra, H., Schaffer, MM., Chaudhary, G., Kazmi, H. D., Dréau, J. L., & Petersen, S. (2023). What metrics does the building energy performance community use to compare dynamic models? Proceedings of the 18th IBPSA Conference Shanghai, China, Sept. 4-6, 2023,

317) ASHRAE, (2014), p.24-26.

3.5.2 데이터 전처리 및 매칭

3.5.2.1 기간·해상도 정렬

VSM 결과와 RSM 데이터의 비교를 위한 기간 및 해상도 정렬 절차는 두 데이터의 정합성 평가에서 시간 축의 불일치나 해상도 차이는 RMSE, MAE, MBE 등 통계적 지표를 왜곡시키는 주요 원인으로 작용한다. 따라서 본 연구에서는 비교 지표의 왜곡을 방지하기 위해, 시뮬레이션 결과와 실측 데이터가 동일한 시간 해상도와 동일한 기간을 기준으로 비교될 수 있도록 시간 스텝 프를 일치시킨 후 비교 분석하는 정렬 과정을 선행하였다³¹⁸⁾.

데이터 정렬 과정은 ① 표준 시간 인덱스 생성→ ② RSM 해상도 표준화→ ③ VSM-RSM 병합→ ④ 다단계 집계 파생의 순으로 진행하였으며 데이터의 기준은 한국 표준시(KST, UTC+9) 로 설정하였다. 분석 연도는 2024년을 기준으로 총 8,760시간(24 × 365일)으로 확보하였다³¹⁹⁾. 모든 시계열 데이터는 2024-01-01 00:00부터 2024-12-31 23:00까지의 연속된 시간 인덱스를 기준으로 정렬되며, 이때 누락된 시점은 결측치(NaN)로 표시하였다.

RSM의 원시 데이터는 실측 시스템에 따라 1분, 5분, 또는 15분 단위로 측정되는 경우가 있으므로, 이를 1시간 단위로 표준화하였다. 전력량(kWh) 자료는 해당 시간 구간의 합산값(sum)으로, 전력(kW) 자료는 평균값(mean)으로 리샘플(re-sample)하여 시간 단위로 변환하였다. 혼재된 단위가 존재할 경우에는 단위 환산을 통해 kWh 기준으로 통일하였다³²⁰⁾.

VSM 결과는 이미 1시간 단위(Time Step=60 min, $\Delta t=1h$)로 생성되어 있으므로, RSM을 동일한 시간 인덱스로 변환·정렬한 후 병합하였다. 모든 비교는 이 정렬된 8,760시간의 시간 축을 기준으로 수행되며, 이는 실측 누락 구간을 그대로 반영하기 위함이다. 이후 시간 해상도 정렬이 완료된 데이터로부터 일·월·계절·주 단위의 파생 집계 데이터를 생성하였다³²¹⁾³²²⁾.

318) U.S. Department of Energy, (2023), 전계서, p.21-22.

319) TimeAndDate.com, (2024), 전계서.

320) Centinel. (2024). "The difference between energy (kWh) and power (kW)". Retrieved from <https://www.centinelre.com/>.

또한 VSM과 RSM 간의 매핑 과정에서, 각 데이터에는 7개 광역군 지역 (Location), 건물유형(Archetype), 시나리오(Load Profile A-BB), 기상 데이터(epw) 등의 식별자(ID)가 메타데이터로 유지된다. 이러한 식별자를 공통 키로 병합함으로써, 시뮬레이션 입력 조건과 실측 건물의 특성이 동일하게 대응되도록 하였다. 이를 통해 시계열 비교의 기초를 동일한 시간 축 위에 구축함으로써, 이후 수행되는 통계적 정합성 평가에서 기간·해상도 불일치로 인한 오차 편향을 제거할 수 있도록 하였다³²³⁾³²⁴⁾³²⁵⁾.

3.5.2.2 결측·이상치 필터링 규칙

RSM 데이터는 센서 결함, 통신 오류, 데이터 수집 지연, 검침 교체 등으로 인해 결측치(missing values)와 이상치(outliers)가 빈번히 발생한다. 이러한 이상치는 시뮬레이션 검증의 신뢰도를 저하시킬 수 있으므로, 본 연구에서는 물리적 타당성과 통계적 견고성을 함께 고려하여 체계적인 정제 규칙을 수립하였다³²⁶⁾³²⁷⁾³²⁸⁾. 이 절차는 RSM 데이터에만 적용하며, VSM 결과에는 보정이 가해지지 않는다.

1) 결측치 처리 규칙

시간 축 정렬 이후 누락된 시점은 모두 결측치(NaN)로 지정하였으며, 결측치의 길이에 따라 처리 방식을 구분하였으며, 연속 3시간 이하의 단기 결측 구간은 인접 값의 추세를 반영한 선형 보간을 적용하였다. 반면, 3시간을 초과

321) U.S. Department of Energy, (2023), 전계서, p.21-23.

322) U.S. Department of Energy, (2021), 전계서, p.15-17.

323) U.S. Department of Energy, (2023), 전계서, p.21-23.

324) U.S. Department of Energy, (2021), 전계서, p.15-17.

325) U.S. Department of Energy, (2024). EnergyPlus™ Version 24.1.0 Documentation: Input Output Reference. Build : 9d7789a3ac.

326) ASHRAE, (2014), p.5-6, p.14-15.

327) Neale, A., Kummert, M., & Bernier, M., (2018), p.67.

328) 이경호, 주홍진, 안영섭, 이왕재. (2024). TRNSYS 환경에서 모델 단순화와 학습을 통한 월별 연간 에너지 성능 평가: 태양광 적용 사무용 공공건물의 히트펌프 냉·난방. 『한국건축친환경설비학회논문집』. 18(3), 192-210. <https://doi.org/10.22696/jkiaecbs.20240017>.

하는 중·장기 결측 구간은 보간하지 않고 그대로 결측으로 유지하였으며, 이러한 구간은 일·월 단위 집계 시 유효 표본 부족(insufficient sample)으로 마스킹 처리 후 통계적 대체를 수행하였다³²⁹⁾.

데이터의 신뢰성을 확보하기 위해 유효성 기준으로 일 단위 유효성은 하루 24h 기준으로 유효 데이터가 80% 이상($\geq 19h$)으로 존재할 때만 일별 데이터를 채택하였고 월 단위 유효성은 월 총시간 대비 유효 데이터가 85% 이상일 때만 월별 데이터를 채택하도록 설정하였다. 이러한 기준 미달 시에는 에너지 검증 지침의 표본 충족 원칙을 준용하여 산출에서 제외하였다. 이 기준은 실측 결측이 장기간 지속될 경우 지표가 과도하게 왜곡되는 것을 방지하기 위한 것이다³³⁰⁾.

2) 이상치(outlier) 처리 규칙

이상치의 처리 규칙으로 1단계에서 선형 보간법을 수행하였다. 선형 보간법은 $y = f(x)$ 의 곡선상 두 점 $(x_i, y_i), (x_{i+1}, y_{i+1})$ 에 대하여 임의의 점 x 에 대하여 $f(x)$ 의 근사값 $p(x)$ 를 이 두 점 간의 직선으로 연결된 x 의 1차 함수로 구하는 방식이다. 1차식이 아닌 다항식으로 보간법을 사용할 수 있으며 모든 점들을 통과하는 다항식은 차수가 n 개로 유일한 다항식이다³³¹⁾. 2차 보간은 1차 수행한 선형보간법이 $\leq 3hr$ 이내의 결측 데이터를 대상으로 결측 데이터의 전과 후의 데이터를 선형으로 연결되는 방식으로 결측 기간이 장기적이거나 누락된 경우에 적용할 수가 없는 관계로 통계적 대체를 수행하였다.

통계적 대체(Imputation)는 관측치만으로 회귀 모형을 적합한 뒤, 동일한 자료 구조에서 결측치에 대해 예측값을 넣는 절차다³³²⁾. 이 접근은 결측치만을 추정치로 대체하되 관측치는 그대로 보존한다는 점에서 전통적 회귀 추정과 구별되고 있으며 특히 에너지 데이터의 경우 설명 변수로 기상 지표, 추세선, 공급 측 변수 등을 포함하는 것이 권장된다³³³⁾. 또한 시계열적 가정을 활

329) ASHRAE, (2014), 전계서, p.5-6, p.15, p.63-64.

330) ASHRAE, (2014), 전계서, p.5-6, p.15, p.63-64.

331) 이성근, 전계서, p.46-65.

332) ASHRAE, (2014), 전계서, p.5-6, p.15, p.25.

333) 이성근, 전계서, p.75.

용해 고해상도 값으로 분해하는 Chow-Lin 계열도 존재하며, 이는 관측치를 유지하면서 관련 시계열을 공변량으로 사용하도록 설계돼 있다. 다만 Chow-Lin은 주기적 일-> 주-> 월-> 분기-> 년의 집계 변환에 적합하다³³⁴⁾³³⁵⁾.

시점 t 의 전력 사용량은 다음과 같이 정의한다.

$$y_t = x_t\beta + \epsilon_t, \epsilon_t \sim (0, \sigma^2),$$

여기서,

관측치 집합을 σ , 전력 사용량 y_t , 설명 변수 x_t ,

여기에서 적용된 관측치의 방정식은 다음과 같다.

$$\hat{y}_t = x_t\hat{\beta}, (t \in M),$$

여기서, 결측 집합을 M , x_t 는 일, 주, 계절, 추세, 외기 온도의 기상, 행태 변동을 반영한 설명 변수,

기상 지표로 HDD는 기 온도가 기준 온도보다 낮은 날들이 얼마나 되는지를 누적해서 나타내는 지표로 난방 에너지 수요의 잠재적 크기를 정량화 하는 지표로 산정식은 다음과 같다³³⁶⁾.

$$\text{HDD} = \max(0, t_{\text{base}} - t_{\text{avg}}),$$

여기서,

t_{base} 는 일반적 기준 온도로 보통 18.3°C (65°F)를 적용,

t_{avg} 는 해당일의 일평균 외기 온도,

또 다른 기상 행태로 CDD는 외기 온도가 기준 온도보다 높은 날들이 얼마나 되는지를 누적해서 나타내는 지표로 냉방 에너지 수요의 잠재적 크기를 정량화하는 지표로 산정식은 다음과 같다³³⁷⁾.

$$\text{HDD} = \max(0, t_{\text{avg}} - t_{\text{base}}),$$

334) Chow, G., C. & Lin, A.-I. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4) 372-375. <https://www.jstor.org/stable/1928739>.

335) Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, 1(2), 169-173.

336) ASHRAE. (2001). *ASHRAE Handbook: ASHRAE TECHNOLOGY COMMITTEES AND TASK GROUPS, 2001 ASHRAE FOUNDATIONS HANDBOOK SI*.

337) ASHRAE, (2001), 상계서, p.747.

여기서,

t_{base} 는 일반적 기준 온도로 보통 18.4° C (65°F)를 적용

t_{avg} 는 해당일의 일평균 외기 온도

본 연구의 RSM 자료는 비주기적 결측이 길게 분포할 수 있으며, 외생 충격의 영향이 크다. 따라서 1차 시간 보간 후 남은 장기 결측에 대해 기온·달력효과·추세를 포함하는 회귀 기반 imputation을 2차 단계로 수행해, 외부 요인을 반영한 보완을 달성하였다³³⁸⁾. 이 접근은 비주기적 시계열에도 적용 가능하다는 선행 연구의 권고와 부합한다. Imputation은 관측치를 변형하지 않고 보존하며, 기상·행태·변동 요인 등 외생 변수를 활용할 수 있다는 점이 장점이다. 반면 선형성과 오차 구조에 민감하며, 관측치가 매우 적으면 설명 변수 선택에 제약이 생긴다는 단점이 있다³³⁹⁾.

3) 품질관리(Quality Control, QC) 로그

정제 과정 전반의 품질을 모니터링하기 위하여, 각 데이터 파일에 대해 결측 비율, 이상치 비율, 보간 건수, 일·월 단위 유효건수를 기록하는 품질 지표(QC log) 를 병행 생성하였다. 이 로그는 향후 RMSE·MAE·MBE 산출 시 가중치를 유효성이 낮은 달의 지표는 참고 값으로 반영하고 모델 교정 우선 순위 설정의 근거로 활용하였다³⁴⁰⁾.

이러한 정제 규칙을 통해 실측 데이터의 노이즈와 결측 영향을 최소화함으로써, 시뮬레이션과 실측 간의 오차가 입력 불완전성의 결과가 아닌 모델 자체의 성능을 반영하도록 보장하였다. 이를 통해 이후 4.4.3절에서 수행되는 통계적 정합성 지표 계산의 신뢰도를 확보하였다. 가상의 KR-VSM 데이터 구축의 신뢰성을 확보하기 위해, 모든 시계열 데이터에 대해 결측·이상치 진단과 보정을 통해 정합성의 검증으로 구성된 다단계 품질관리(Quality Control, QC) 체계를 적용하였다. 먼저, RSM·VSM 원천 데이터로부터 timestamp, kWh, hour 등 핵심 변수의 형식을 판별하고, 데이터 타입을 정

338) 이성근, 전계서, p.46-65.

339) 신이레, 윤상후., 전계서, p.1603-1608.

340) ASHRAE, (2014), 전계서, p.5-6, p.9-11, 26-28, 94-97.

규화하였다. 이후 결측치와 비정상적 피크값을 24시간 단위의 평균 패턴에 기반한 보간을 통해 일관되게 데이터를 처리하였다³⁴¹⁾. 이 과정에서 생성된 QC 로그에는 시간대별 결측 비율, 이상치 탐지 건수, 보간 횟수, 일·월 단위 유효데이터의 비율이 기록되며, 이는 데이터셋 구축에 전반적으로 모니터링 지표로 활용하였다.

특히, RSM과 VSM 간 형태학적 불일치가 입력 데이터의 불완전성일 가능성이 높다는 점을 고려하여, QC 로그는 후속에서 진행되는 RSM-VSM의 정합성 검증에서 월 단위 유효성이 낮은 지표는 가중치를 낮추어 반영하고 유효성이 높은 데이터는 가중치를 높여서 반영하는 방식을 채택하였다. 이를 통해 정제 수준이 낮은 기간의 왜곡이 전체 성능 평가에 미치는 영향을 최소화하고자 하였다.

또한, RSM과 VSM의 24시간 평균 부하인 loadshape를 비교하는 전용 QC 모듈을 적용하여, 주중·주말별 24시간 패턴을 자동으로 산출하고 RMSE를 계산하였다. 보정(calibration) 단계에서는 VSM의 시간대별 부하 스케일링인 hour_shape를 실측 패턴과의 비율 기반으로 갱신하며, 각 반복(iteration)마다 생성된 정규화 loadshape, RMSE(weekday/weekend), 보정 계수 부하 스케일링 히스토리로 hour_shape.*가 함께 저장되어, 모델 보정의 수렴 여부를 시각적·정량적으로 확인할 수 있다. 이러한 체계적 정제 프로세스를 통해 실측 데이터의 노이즈·결측 영향이 충분히 완화되도록 하였으며, 최종적으로 계산되는 정합성 지표가 입력 데이터 오류의 결과가 아니라 모델 자체의 성능을 반영하도록 보장하였다.

3.5.2.3 VSM-RSM 가구 군집 매칭 전략

VSM-RSM 매칭 절차는 Location·Archetype·LoadProfile·Scenario의 4단계 계층형 Key를 기반으로 설계하였으며 4개 키는 각 VSM 케이스와 RSM 가구를 연결하는 공통 식별자를 sim_id로 활용하였다³⁴²⁾. 먼저, VSM은 7대

341) Lepot, M., Aubin, J.-B., & Clemens, F. H. L. R., 전계논문, p.2-3.

342) StackExchange Mathematics Community. (2019). "Why John Tukey set 1.5 IQR to detect

권역별 *.epw 기상 시나리오와 Load Profile 조합에 의해 생성된 700개의 물리 기반 템플릿이며, RSM은 전국 공동주택 단지에서 수집된 실측된 AMI 시계열 데이터이며 권역 정보는 고유의 id와 함께 행정안전부에서 지정한 법정동 데이터를 포함하고 있다. 여기서 데이터셋은 VSM의 Location 정보와 RSM의 법정동 데이터를 기반으로 동일 권역 코드 매칭을 전제로 하여 건물 면적, 난방·냉방 방식, 재실 인원, 창면적비 등의 물리 변수를 활용한 (KNN, K-Nearest Neighbor algorithm)을 통해 개별 가구를 1:1로 연결하였다.

대규모 공동주택 단지에 대해서는 일별 24h 부하 형상을 기준으로 RSM에 K-means 군집화를 수행하고, DTW(Dynamic Time Warping, 동적 시간 왜곡) 기반 패턴 유사도를 이용하여 가장 유사한 LoadProfile 시나리오를 갖는 VSM 템플릿과 1:N 또는 N:N 확률 매핑을 적용하였다³⁴³⁾. 이를 통해 실측 데이터의 분포적 다양성을 유지하면서도 지역·기후·행태적 동질성이 반영된 대규모 VSM 합성 데이터를 생성하였다. 최종 매칭된 RSM-VSM 시계열은 r, RMSE, MAE, MBE로 검증³⁴⁴⁾하였으며, 모든 지표가 기존 EnergyPlus QC 기준 $\pm 10\%$ 이내에서 수렴하여 매칭의 타당성이 확보되었다. 이 매칭 구조는 이후 DR 예측, VPP 운영 시뮬레이션의 실증 실험의 입력 데이터로 활용된다.

1) 가구 단위 1:1 매칭 전략

개별 세대 단위에서 실물 AMI에서 검출된 RSM 데이터와 시뮬레이션 기반 VSM 데이터를 정합적으로 비교하기 위해 본 연구는 가구 기반 1:1 매칭 전략을 도입하였다. RSM의 경우 단독주택 또는 소규모 공동주택에서 세대별

outliers". from: Mathematics StackExchange. Retrieved from <https://math.stackexchange.com/questions/966331/why-john-tukey-set-1-5-iqr-to-detect-outliers-instead-of-1-or-2><https://math.stackexchange.com/questions/966331/why-john-tukey-set-1-5-iqr-to-detect-outliers-instead-of-1-or-2r>.

343) Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. Knowledge and Information Systems, 7(3), 358–386. <https://doi.org/10.1007/s10115-004-0154-9>.

344) Agrawal, A. (2020). Know the best evaluation metrics for your regression model! Towards Data Science. Data Science Blogathon. Article Url - <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/>.

전력량이 계량되고 있지만, 공동주택의 경우 아파트 관리사무소에서 개별적으로 계량되고 있는 상황에서 8,760 시간의 시계열이 확보된다는 점에서 VSM 데이터 셋의 구축은 RSM 데이터와 직접 대응하기 적합한 구조가 가능하다.

이를 고려하여 본 연구는 먼저 RSM 데이터에 포함된 행정구역에 대한 법정동 정보를 서울·인천·경기·충청·전라·경상·강원 등 7대 광역권역으로 분류하고, 동일 광역권에 대해서는 기상청의 광역권의 관측소 기반의 epw 기상 시나리오를 적용하고, 이를 기반으로 EnergyPlus 시뮬레이션을 통하여 700개 VSM 템플릿을 우선적으로 생성하였다. 이는 기후 기반의 냉·난방 부하 차이에 대하여 지역별 분류와 함께 광역권 내에서는 동질성을 유지하는 1차 필터링 과정이다. 이후 선정된 후보 템플릿 내부에서 세대별 Archetype 변수의 유사도를 정량적으로 평가하는 가장 근접한 하나의 템플릿으로 선택하였다.

여기서 사용된 주요 변수는 건축연도, 전용면적, 난·냉방 방식, 거주 인원, 창면적비(WWR), 창호 유형, 기밀 등급, 단열성능 등이며, KR-VSM의 QC 단계에서 냉·난방 부하 기여도의 가중치를 적용하였다.

이렇게 선정된 템플릿은 sim_id_template으로 해당 세대의 Occupant, Occupancy Activity, DHW, Lighting Density, Home Appliance를 가장 잘 설명하는 시뮬레이션 기반 대체 곡선이며, 이후 미시적 검증에 활용된다. 구체적으로, RSM-VSM 시계열 쌍에 대하여 일·주·연간의 24h·168h·8,760hr 단위의 r, RMSE, MAE, MBE를 비교하고, 각 시계열을 평균 전력으로 정규화하여 패턴 유사도를 측정하였다. 이러한 절차는 시뮬레이션 모델의 절대적 정확성을 평가하기 위한 가장 정밀한 수준의 검증 단계이며, EnergyPlus 기반 부하모델이 실측 데이터를 어느 정도까지 구조적으로 재현할 수 있는지를 판단하는 기초 근거가 된다.

2) 군집 단위 1:N 매칭 전략

대규모 공동주택 단지의 경우 수백 세대의 전력 시계열이 존재하지만, 각 세대의 물리적 특성을 모두 파악하는 것은 현실적으로 불가능하며, 세대 간의 사용 특성 또한 다양하게 분산되어 있다. 따라서 본 연구는 RSM 데이터의

다변성을 유지하면서도 효율적인 매칭을 위해 군집 기반 1:N 매칭 전략을 적용하였다³⁴⁵⁾. 먼저 단지 내 모든 세대의 8,760hr 데이터를 24시간 단위 평균 곡선으로 변환하여 부하 곡선 형상에 관한 특징 벡터를 204D Feature로 구축하였고, 이를 기반으로 K-means 알고리즘을 적용해 3~5개 군집으로 분류하였다. 이렇게 도출된 군집은 해당 단지의 대표적인 일일 사용 패턴을 나타내는 집합적 load_archetype으로 간주하여 부하 곡선 간의 분산이 군집 수를 결정하는 핵심 지표로 활용되었다.

본 연구는 VSM 템플릿에서 A1~A4, BB01~BB07, A-BB의 동일한 Load Profile을 가진 후보군을 선별하고, 군집을 대표하는 템플릿 시계열 간의 형태적 유사도를 함께 평가하였다. 유사도 평가에는 Pearson r·RMSE·MAE의 통계 지표를 활용³⁴⁶⁾하였으며, 이는 시간 축의 이동을 허용한 채 형태적 유사성을 평가할 수 있는 선행 연구의 평가 모델을 기반으로 RSM의 전력 사용량 패턴 분석을 수행하였다³⁴⁷⁾.

이러한 구조는 실측 데이터의 전력 사용에 대한 다양한 패턴이 단일 템플릿으로 대표되는 것을 방지하며, 시뮬레이션을 기반으로 하는 템플릿의 제한적 다양성을 보완하는 역할을 수행하였다. 결과적으로 군집 기반 1:N 매칭은 공동주택의 단지의 전력 사용량에 대한 loadshape 분포를 시뮬레이션 체계 내에서 재현할 수 있도록 설계한 것이며, 이는 지역적 특성뿐 아니라 단지 내 다양한 수용가의 행태적 다양성까지 반영이 가능하다는 장점이 있다.

3) 군집 단위 N:N 확률 매칭

KR-VSM의 최종 목표는 지역·단지·세대 단위에서 실측과 유사한 통계적 구조를 가진 대규모의 KR-VSM 데이터셋으로 200,000 가구를 구축하는 것이다. 이를 위해 본 연구는 700개의 템플릿을 기반으로 하되, 단일 템플릿에

345) Wang, Z., Hong, T., & Piette, A. (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy* 263, 114683.

346) Agrawal, A. (2020). Know the best evaluation metrics for your regression model! Towards Data Science. Data Science Blogathon. Article Url - <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/>.

347) Keogh, E., & Ratanamahatana, C. A., 전계논문, p.3-5.

세대가 과도하게 집중되는 문제를 방지하고, 실제 AMI 사용 분포의 다양성을 반영하기 위해 확률 기반 N:N 매칭 전략을 도입하였다. 먼저 지역별·군집별 RSM 세대 분포를 계산하여 서울 14.8%, 경기 27.8%, 인천 6.1%, 충청 11.9%, 전라 10.3%, 경상 26.1%, 강원 3.1% 등 2024 인구주택총조사 기반의 권역 분포를 재현하였다³⁴⁸⁾. 이후 각 지역 내에서 RSM 군집별 발생 비율을 동일하게 유지하며, 각 군집이 대응하는 KR-VSM 템플릿 집합을 Pearson r·RMSE·MAE 기반 유사도 점수로 정규화하여 선택 확률로 변환하였다. 즉, 템플릿 선정에 정확도가 가장 높은 템플릿 하나 선택 방식이 아니라, 유사도 가중치에 따른 확률적 샘플링 방식으로 수행한 것이다.

이 접근 방식은 대규모 200,000 건의 VSM 생성에서 매우 중요한데, 이는 실제 AMI 데이터가 가지는 다양성을 재현하는 동시에, 시뮬레이션 템플릿에 존재할 수 있는 구조적 편향을 최소화하는 효과를 갖는다. 이는 특정 지역의 냉·난방 패턴이 템플릿 내 한 개 LoadProfile에만 치우쳐 있을 경우, 확률 기반 매핑은 다양한 템플릿을 적절히 혼합하여 실측 데이터의 분포와 더 높은 일치도를 가지도록 하였다. 이로써 지역·단지·가구 규모에서 실측 기반의 확률 모델을 바탕으로 700 템플릿을 200,000개로 확장하는 과정이 논리적·통계적으로 타당하게 구성되었다.

4) 매칭 검증 및 적용

가구 및 군집 기반 매칭을 통해 생성된 VSM-RSM 대응체계는 본 연구의 품질 보증(Quality Assurance, QA)의 핵심 구성요소이며, 시뮬레이션 기반 예측 모델이 실측 데이터를 충분히 모사하고 있는지를 평가하는 기준이 된다. 본 연구는 EnergyPlus QC 전략 및 KR-VSM 품질관리 규칙에 따라 시계열 정합성, 패턴 유사도, 연간 에너지 통계 등 다차원 지표를 사용하여 매칭 결과를 체계적으로 검증하였다. 시계열 정합성의 경우 ASHRAE Guideline 14에서 제시하는 에너지 모델의 검증 지표 체계를 준용하여 수행하였다. 특히 평균 편향 오차(MBE)는 $\pm 10\%$ 이내를 허용 기준으로 적용하였으며, RMSE, MAE의 경우 피크 수요 오차에 대하여 선행 연구의 VSM 검증 사례에서 일

348) 조서관리국 인구총조사과, 전계서, p.70.

반적으로 활용되는 범위 수준으로 normalized RMSE $\leq 10\%$, normalized MAE $\leq 15\%$, 피크 수요 오차 $\leq 15\%$ 를 추가적인 보정 기준으로 삼았으며, 이는 EnergyPlus에서 1차로 생성된 700개의 데이터셋의 검증 단계에서 사용된 것과 동일한 국제적 기준을 따른 것이다³⁴⁹⁾. 특히 weekday/weekend loadshape 비교는 실주거 부하 패턴의 핵심 특성을 반영하는 지표로 활용되었다.

[그림 3-27] RSM-VSM Loadshape 비교

No.	항목	전국 단위 평균값(kWh)
1	RSM_Weekday_mean	0.3070
2	RSM_Weekend_mean	0.3199
3	VSM_Weekday_mean	0.2942
4	VSM_Weekend_mean	0.2941

패턴 유사도 검증에서는 QSFBSM 선행 연구에서 활용된 Pearson 상관 계수를 형태적 유사성의 참고 지표로 사용하였다. 본 연구에서는 기존 시계열 분석 문헌에서 일반적으로 강한 상관관계로 해석되는 수준을 참고하여 r 값이 상대적으로 높은 경우를 우선 선별하였으며, 최종 검증은 RMSE, BBME 기준에 따라 수행하였다. 시간 축 차이를 고려하기 위해 DTW distance가 추가적으로 활용되었다. [표 3-27]과 같이 실제 200,000개로 생성된 RSM-VSM 데이터셋의 QC 결과에서도 평일 평균 RSM 0.3070kWh, 대비 VSM 0.2942kWh, 주말 평균 RSM 평균 0.3199kWh 대비 VSM 0.2942kWh 수준으로 $\pm 5\sim 8\%$ 범위에서 수렴함을 확인하였고, 이는 행태적·규모적 차이가 모두 유의미하게 통제되었음을 시사한다.

또한 지역 기반 연간 부하 일치도 역시 [표 3-28]과 같이 RSM과 VSM 간 대부분 $\pm 10\%$ 이내에서 수렴하였으며, 선행 연구에서 서울 지역 101개의 표본 기반의 건물 상세에너지 정보시스템(KDBEIS)의 표본 가구를 분석한 평균 전력 사용량 2,822[kWh/y]와 $\pm 10\%$ 로 수렴되고 있음을 확인할 수 있었으며³⁵⁰⁾³⁵¹⁾, 데이터 통계에서 일부 발생한 지역 오차는 fix_location_typos를

349) ASHRAE, (2014), 전계서, p.5-6, p.9-11, 26-28, 94-97.

350) 홍승연, 송승영.(2021). 머신러닝을 활용한 공동주택 세대에서의 전기기기 에너지 사용량 분석

통해 정정하여 분석에 반영하였다.

[표 3-28] Location 별 연간 평균 전력량 kWh 비교

순서	Location	RSM Annual kWh	VSM Annual kWh	비율
0	Choongchung	2589.308703	2591.629663	100%
1	Gangwon	2665.577543	2482.912908	107%
2	Gyeonggi	2784.733337	2733.374950	102%
3	Gyeongsang	2732.439264	2694.769715	101%
4	Incheon	2850.800118	2775.015141	103%
5	Jeonra	2487.311303	2587.755692	96%
6	Seoul	2773.060700	2770.021486	100%
7	Unknown	0.000000	Nan	

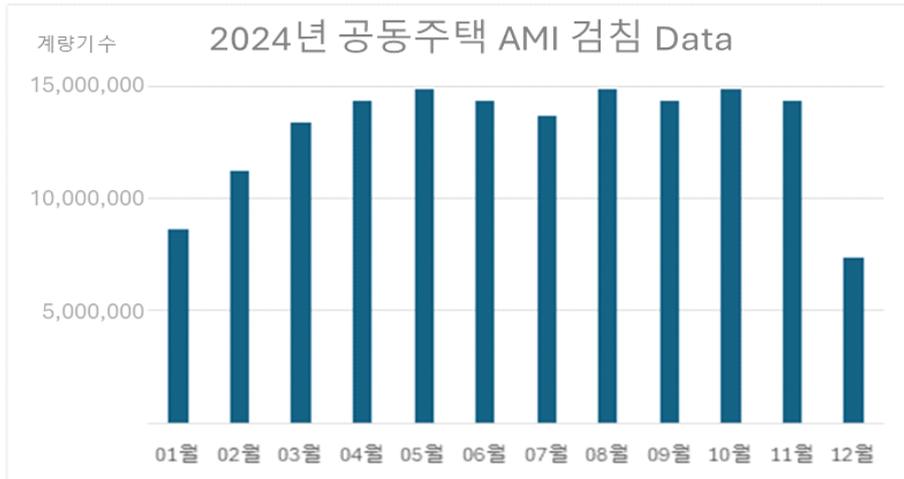
이러한 종합적인 검증 과정을 통해 매칭된 VSM-RSM 데이터셋은 DR 예측 모델의 학습 데이터셋, DR 기반의 VPP 운영 시나리오, DR 효과 분석 등의 후속 연구 단계에서 신뢰성 있는 기반 데이터로 활용될 수 있음을 확인하였다. 더 나아가 매칭 결과는 향후 지능형 부하 예측 모델의 성능 향상, 지역 기반 수요 관리 정책 시뮬레이션, 건물 에너지 기반의 디지털트윈 구축 등 다양한 응용 가능성을 제시한다.

3.5.3 1차 검증 : 연간 총량 일치성 체크

본 연구에서는 생성된 KR-VSM 데이터셋의 현실 적합성을 검증하기 위하여, (재)한국스마트그리드사업단의 AMI 데이터 유통 플랫폼을 통해 확보한 2024년 1월부터 12월까지의 전국 공동주택 세대별 1시간 단위 검침 데이터를 기반으로 실측 RSM 데이터에 대하여 연간 총량 일치성 검사를 수행하였다.

및 추정. 『대한건축학회』, 2021년도 추계학술대회논문집. 41(2). 860-863.
 351) 홍승연, 전서윤, 송승영. (2020). 통계 기법을 활용한 공동 주택 세대에서의 용도별 에너지사용량 분석 및 추정: 온수, 전기기기, 조명 에너지사용량을 중심으로. 『대한건축학회』, 2020년 추계학술발표대회 논문집. 40(2). 704-707.

먼저 [그림 3-7]에서와 같이 확보된 원시 RSM 데이터로부터 약 1억 5,651만 행(156,504,536 rows)에 해당하는 2024년 시계열 자료만을 선별·정제·병합하였다.



Source : AMI 데이터 유통 플랫폼

[그림 3-7] RSM 데이터 1년 데이터 집계

전체 897,322세대의 AMI 계량기 전체 검침 데이터 중에서 연간 8,760시간 중 최소 80% 이상 관측되는 연결성·결측률 기준을 충족한 세대의 계량기는 총 480,118세대로 나타났으며, 이들을 대상으로 산출한 전국 연간 총 전력 사용량은 약 1.30×10^9 kWh(1,302,968,012 kWh)로 집계되었다. 또한 세대 당 연평균 전력 소비량은 약 2,714 kWh/세대·년으로 분석되어, 3.2.3항의 K-apt 공동주택관리 정보시스템의 통계 기준 상 공동전기와 세대전기의 비율이 25%:75%의 기준에 따라 연간 가구의 전력 사용량을 3,400kWh를 기준으로 산정했을 때 2,550kWh와 6% 차이로 유사한 결과의 수준임을 데이터로 확인하였다.

본 연구에서는 2024년 연간의 실측 스마트미터 기반 RSM 데이터와, 제안한 KR-VSM 생성 모델을 통해 생산된 VSM 데이터 간의 정합성을 다각적

으로 검증하였다. 우선, RSM 150,918,777행과 VSM 175,680,000행에 대해 시계열 품질 진단을 수행한 결과, 두 데이터셋 모두에서 timestamp 파싱 오류 (NaT)와 전력 사용량 결측(NaN)이 전혀 발생하지 않았으며, (hid, timestamp) 기준 중복 행도 발견되지 않았다. 또한 음수 전력 사용량 역시 존재하지 않아, 전체 데이터는 측정·생성 단계 모두에서 매우 우수한 품질을 유지하는 것으로 확인되었다.

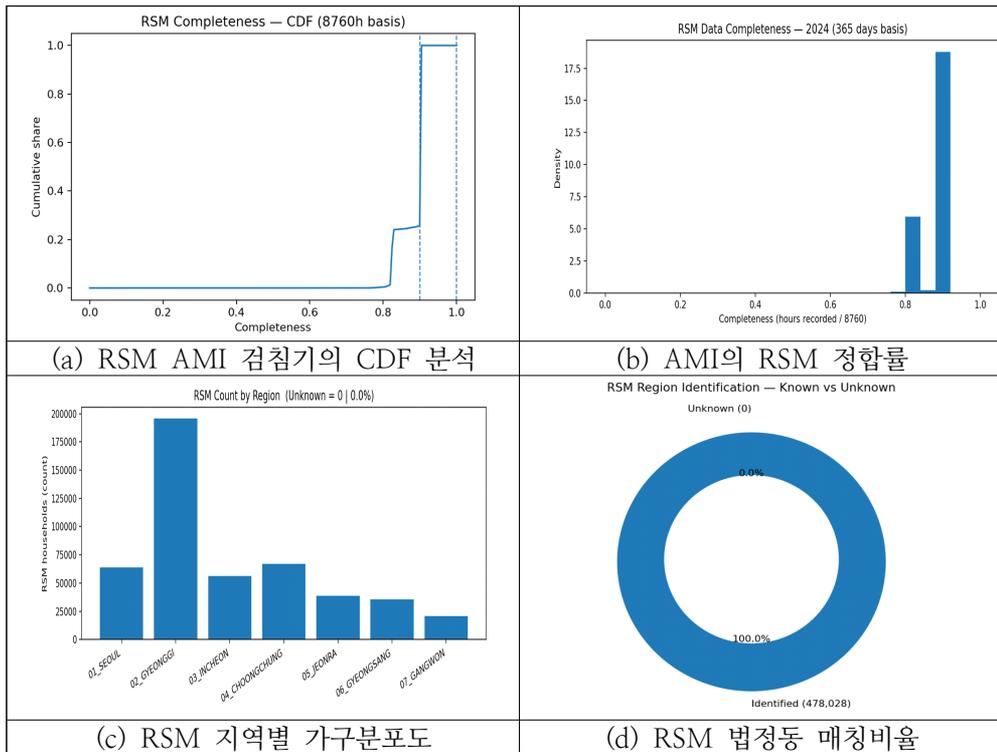
RSM과 VSM의 시간 범위는 모두 2024년 1월 1일 00시부터 12월 31일 23시까지 총 8,760시간을 완전히 포함하고 있었으며, 세대별 시간 완전성 (completeness)의 경우 RSM은 평균 99.9948%로 나타나 일부 세대에서 제한적 결측이 관찰되었으나, 대부분 상위 90~95% 구간의 세대는 8,760시간을 완전하게 유지하였다. 반면 VSM은 전 세대에서 100%의 완전성을 기록하여, 생성 모델이 시계열 결측 없이 안정적으로 데이터를 생산함을 확인하였다.

연간 전력 사용량 수준에서는 RSM과 VSM의 표본 규모 차이를 보정하기 위해, VSM 20만 세대 중에서 RSM과 동일한 약 48만 세대를 난수 추출 방식으로 구성하여 연간 총량을 비교하였다. 그 결과, VSM의 전국 연간 총 전력 사용량은 약 1.297×10^9 kWh(1,297,391,828 kWh)로 산정되었으며, 이는 RSM 1.30×10^9 kWh(1,302,968,012 kWh)의 데이터와 비교했을 때 상대오차 약 0.43%에 해당하는 매우 높은 정합 수준을 확인하였다. 즉, KR-VSM 생성 모델은 전국 공동주택 스톡의 에너지 규모를 총량 기준으로 $\pm 1\%$ 이내에서 재현하는 높은 정확도를 확보하였다.

한편, 실측 RSM의 법정동 코드를 기반으로 지역별 매핑 결과, 2024년 전체 표본 중 행정구역이 '미상'으로 분류된 세대는 존재하지 않았다. 이는 본 연구에서 구축한 법정동 매핑 테이블이 전국 단위 공동주택 스톡을 완전하게 커버하고 있으며, 지역별 비교 분석을 수행하기 위한 사전 데이터 정합성이 안정적으로 확보되었음을 의미한다. 이는 본 연구의 KR-VSM 생성 모델은 ① 연간 전력 사용량 총량 기준 $\pm 1\%$ 이내의 재현력, ② 주요 권역에서의 높은 지역적 정합성, ③ 시계열 데이터의 우수한 품질 및 완전성, ④ 세대 당 평균 전력 사용량의 실측 대비 높은 일치성을 모두 확인하였다. 이는 제안된 KR-VSM 모델이 실제 공동주택에 대한 전력 사용 스톡의 구조적 특성과 부

하 분포를 정량적으로 충실히 모사함으로써, 이후 DR 기반 전력 예측 등 다양한 에너지 연구 활용에 적합한 신뢰도 높은 가상스마트미터 데이터셋임을 실증적으로 보여준다.

[표 3-29] RSM 데이터 기록 충실도 검증



KR-VSM 데이터셋의 검증 절차에 앞서, 실측 기반 RSM 데이터의 기록 충실도(Completeness)를 [표 3-29]과 같이 정량적으로 평가하였다.

Completeness는 각 세대별로 2024년 1년 8,760시간 중 실제로 계량 데이터가 존재하는 비율로 정의되며, 실측 데이터의 신뢰도와 연간 총량 산정의 정확성을 판단하는 핵심 지표로 활용된다.

우선, [표 3-29]의 (a)에서와 같이 CDF 분석에서도 동일한 경향이 재확인된다. Completeness 0.82 지점에서 누적 비율이 이미 약 90%에 도달하며, 0.90 이상의 고품질 세대는 전체의 약 74%를 차지하였다. CDF의 도표 형태

는 대부분의 표본이 높은 충실도의 0.8~0.9구간에 있음을 보여주며, 이는 RSM 데이터가 검증용 시계열 자료로서 충분한 완결성을 갖추고 있음을 의미한다. [표 3-29]의 (b)와 같이 히스토그램에서 확인되듯이 2024년 RSM 표본 480,118세대의 Completeness는 0.82~0.90 구간에 밀집한 단봉형 분포를 나타내었다. 정량적으로는 평균 0.883, 중앙값 0.903, 하위 10% 지점(p10) 0.821, 상위 90% 지점(p90) 0.903으로 나타났으며, 이는 대부분의 세대가 약 7,700~7,900시간으로 약 323~330 days의 데이터를 보유하고 있음을 의미한다.

즉, 연간 기준으로 평균 결측률은 약 8~11% 수준에 불과한 것으로 확인되었다. 특히 Completeness 0.5 이하 구간은 거의 관측되지 않았으며, Completeness 0.80 미만의 저품질 세대는 전체의 0.4%에 불과하여 전반적인 데이터 품질이 매우 우수함을 시사한다. 또한, RSM 데이터의 행정구역 매핑 품질은 매우 높은 수준이었다. [표3-29]의 (d) 도넛 차트에서 확인된 바와 같이, 전체 480,118세대 중 미분류 세대는 단 한 건(0.0%)도 존재하지 않는 것으로 나타났다. 이는 법정동 코드를 활용한 시군구 광역군 매핑이 정확히 수행되었음을 의미하며, 이후 권역 단위의 VSM-RSM 비교 분석의 신뢰도를 높여주는 중요한 결과이다.

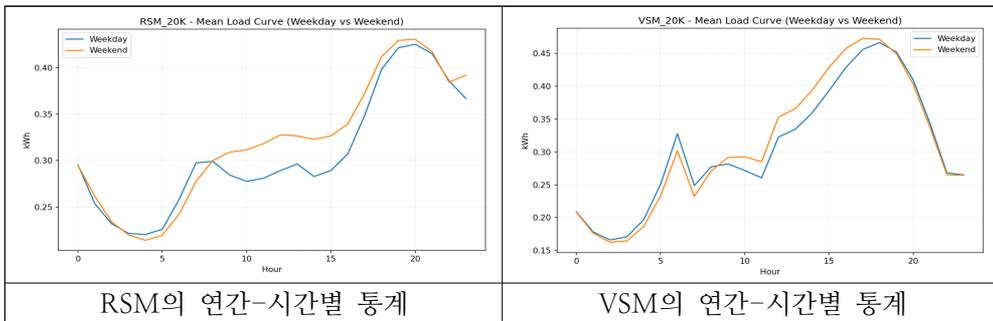
[표 3-29]의 (c)에서 지역별 구성은 실제 2024 인구주택총조사 기준 전국의 아파트 조사 결과와 다르게 충청·경상·전라·강원은 상대적으로 경기도에 비해 작은 비중을 차지하는 형태를 보였으며, 이는 RSM 으로 확보된 데이터가 경기도에 집중되어 있어서 나타난 현상으로 파악되며 VSM-RSM 권역별 표본 크기 설계에 참고 자료로 남겨두었다.

3.5.4 2차 검증 : 패턴 정합성

연간 총량 단위에서 수행된 1차 정합성 검증을 바탕으로, RSM과 VSM 간의 time-series load shape의 시계열 부하 패턴이 시간대·계절·주별 관점에서 어느 정도 유사하게 재현되는지를 추가적으로 검증을 수행하였다. 이를 위해 2024년 1월 1일부터 12월 31일까지의 1시간 간격의 데이터를 기준으로

RSM과 VSM의 두 데이터셋을 sim_id를 기반으로 정렬하고, RSM 데이터에 대해서는 하루 24시간을 기준으로 총 8,760시간의 시계열 중 80% 이상의 유효 기록을 갖는 세대만을 표본으로 추출하였다. 모든 시계열은 KST(UTC+9) 기준으로 통일하였으며, 계절은 봄(4~5월)·여름(6~9월)·가을(10월)·겨울(11~3월)의 네 구간으로, 요일은 주중(월~금)과 주말(토~일)로 구분한 뒤, 각 구간 별로 평균 부하곡선을 산출하여 연간-시간별 부하곡선을 RSM과 VSM을 동일한 시계열 기준으로 비교·분석하였다.

[표 3-30] RSM-VSM 데이터 연간-시간별 비교



먼저 주중·주말 부하 특성 비교 결과 [표 3-30]에서와 같이 VSM-RSM 평균 부하곡선 비교에서, VSM은 RSM의 전반적인 일중 부하 변동 구조를 상당히 높은 수준으로 재현하고 있음을 확인하였다.

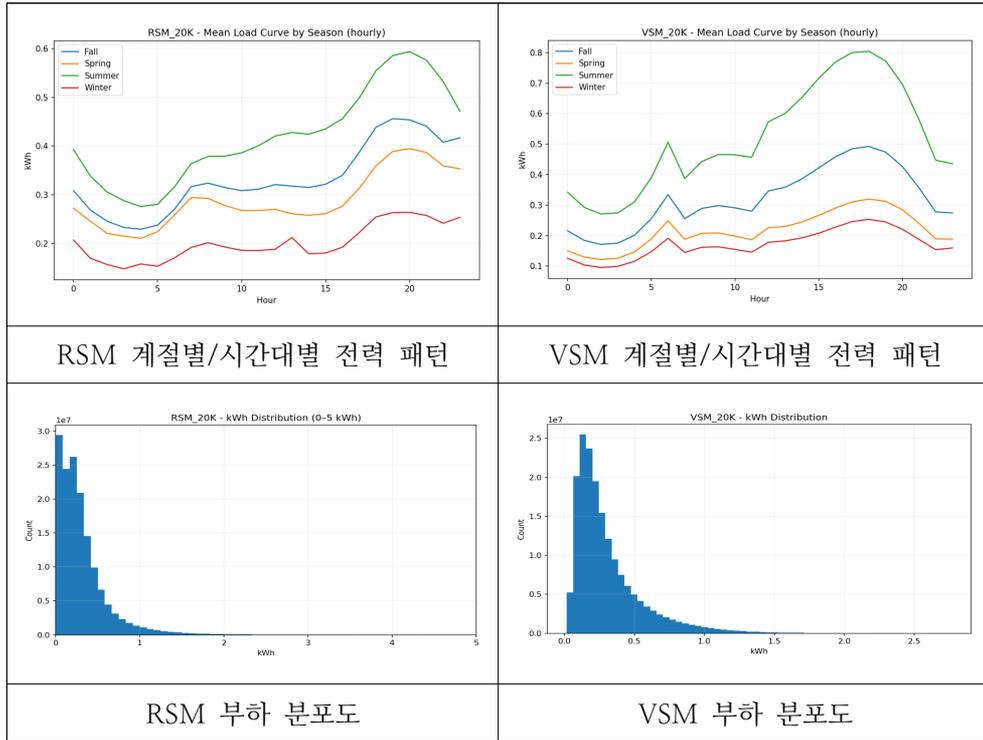
VSM에서 6시에 피크치로 부하곡선의 차이가 있는 것과 21시 이후에 전력 사용량이 급락하는 것을 제외하면 기본적으로는 실제 RSM 사용과 유사한 형태의 덕커브 (Duck-curve)로 재현되었다.

또한, [표 3-31]의 계절별-시간대별 전력 사용량 패턴의 그래프를 보면 계절별로 전력 사용량의 일부 차이는 있지만 전체적인 커브가 오전 6~7시 사이에 피크로 올라가는 것을 제외하고는 VSM 데이터가 RSM 데이터를 유사하게 추종하도록 구현한 것을 확인할 수 있었다.

실제 계량기 RSM 데이터와 VSM 가상스마트미터 데이터에 대해 2024년 1월 1일부터 12월 31일까지의 시간별 데이터 구조 및 품질을 [표 3-32]과 같이 추가 검증하였다. RSM 데이터는 총 150,918,777행, VSM 데이터는

175,680,000행으로 구성되어 있으며, 두 데이터 모두 결측치 및 이상치 처리 이후 24시간 × 365일의 연속성을 유지하는 것으로 나타났다.

[표 3-31] RSM-VSM 데이터 계절별-시간별 비교



먼저 RSM의 타임스탬프 변환 결과 NaT 비율은 0%, kWh 값의 NaN 비율 역시 0%로 확인되어 기본적인 데이터 정합성은 매우 우수하였다. 데이터 기간은 2024-01-01 00:00~2024-12-31 23:00를 정확히 포함하고 있었으며, 중복된 (hid, timestamp) 조합은 0건으로 나타났다.

시간별 사용량의 통계량을 분석한 결과, 평균은 0.31kWh, 표준편차는 0.35kWh, 최대값은 40.93kWh로 나타나 일부 고부하 시간대가 존재하나 전체 분포는 정상 범위 내에 있었다. 또한 사용자 단위(hid)별 연간 8,760시간 대비 관측 비율인 completeness를 산출한 결과, 평균은 0.9999, 최소값은 0.796, 대부분의 세대가 100% 관측률을 나타내어 시계열의 연속성이 매우 안정적임을 확인하였다. VSM 데이터의 경우도 유사한 수준의 품질을 보였다.

다. 전체 데이터는 가상 세대(hid)에 대해 총 175,680,000행이며, 타임스탬프 NaT 비율은 0%, kWh NaN 비율도 0%로 확인되었다.

[표 3-32] RSM-VSM 데이터 정합성 비교

구분	항목	RSM	VSM
기간 (Time range)	분석 기간	2024-01-01 00:00 ~ 2024-12-31 23:00	2024-01-01 00:00 ~ 2024-12-31 23:00
데이터 규모	총 행(row) 수	150,918,777	175,680,000
중복 여부	(hid, timestamp) 중복 건수	0	0
음수값 비율	Negative kWh share	0%	0%
NaT 비율	Timestamp NaT 비율	0%	0%
NaN 비율	kWh NaN 비율	0%	0%
kWh 통계량	Count	150,918,800	175,680,000
	Mean (kWh)	0.3109	0.3074
	Std (kWh)	0.3492	0.2693
	Min (kWh)	0.0000	0.0071
	1%	0.0000	0.0299
	10%	0.0000	0.0722
	50% (Median)	0.2340	0.2248
	90%	0.6520	0.6551
	99%	1.5940	1.3008
Completeness (hid)	Max (kWh)	40.933	3.155
	Sample 개수	17,182	20,000
	Mean	0.99995	1.00000
	Std	0.00241	0.00000
	Min	0.79599	1.00000
기타 특징	Median (50%)	1.00000	1.00000
	Max	1.00000	1.00000
기타 특징	고부하(outlier) 여부	최대 40 kWh 이상 존재	최대 3.16 kWh로 안정적
	전체 시간대 누락	없음	없음

시간 범위는 RSM과 동일하게 연간 전체를 완전하게 포함하였고, 중복된 행은 전혀 존재하지 않았다. 전력 사용량의 기술통계 분석 결과, 평균 0.307

kWh, 최대 3.16 kWh로 RSM에 비해 피크 값 분포가 상대적으로 좁고 안정적임을 확인하였다.

세대별 completeness는 모든 세대가 1.0(100%)을 기록하여, 시계열 누락이 없는 완전한 synthetic dataset임을 검증하였다. 무작위 3개 세대의 gap 점검에서도 RSM과 동일하게 누락시간 없음으로 확인되었다.

추가적으로 RSM과 VSM 모두에 대해 계절·주별 평균 부하곡선과 kWh 분포 히스토그램을 생성하였으며, weekday/weekend mean load, seasonal mean load, completeness histogram 등의 관련 품질 보고서 저장까지 확인하였다. 전체적으로 RSM은 실제 계량 데이터 특성상 일부 고부하(max 40 kWh) 사례가 존재하였으나, 전반적인 분포는 정상적이었으며, VSM은 생성 알고리즘 특성상 보다 안정적인 부하 범위를 보였다.

이상의 결과는 RSM과 VSM 모두 연속성(continuity), 완전성(completeness), 타임스탬프 정합성(timestamp validity), 중복 여부(duplication), 음수/NaN 여부, 기본 통계 분포 항목에서 연구 활용에 충분한 수준의 데이터 품질을 확보하고 있음을 보여준다. 특히 두 데이터셋 모두 연간 전체 8,760 시간의 시계열의 완전한 관측 구조를 유지함으로써, 이후 시계열 기반의 부하 패턴 비교 및 t -RMSE·MAE·MBE· r 의 정합성 통계를 수행하기 위한 신뢰할 만한 자료로 판단된다.

3.5.5 지표 산출 및 통계 분석

3.5.5.1 r /RMSE/MAE/MBE 산출³⁵²⁾

RSM과 본 연구에서 구축한 VSM 간의 정합성을 검증하기 위하여, 2024년 1월 1일부터 12월 31일까지의 시간별 전력 사용량을 대상으로 계절과 주중·

352) Agrawal, A. (2020). Know the best evaluation metrics for your regression model! Towards Data Science. Data Science Blogathon. Article Url - <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/>.

주말 단위의 상대적 패턴 차이를 분석하였다. RSM과 VSM은 개별 세대 단위의 1:1 매칭이 불가능하므로, 동일 시점의 timestamp에 존재하는 모든 세대의 전력량을 합산하여 각각 전체 부하 시계열로 재집계한 후, 두 시계열을 timestamp 기준으로 병합하였다. 이를 통해 2024년 중 RSM과 VSM 데이터가 모두 존재하는 총 8,760개의 시간대에 대해, 두 데이터셋 간 절대 부하 수준 및 시계열 패턴의 직접 비교가 가능하도록 구성하였다.

[표 3-33] 계절별-주간별 통계량 데이터 비교

Season	Week	r	kWh 통계량			N_samples
			RMSE	MAE	MBE	
All	Weekday	0.911	0.130	0.106	-0.0008	108,034,641
All	Weekend	0.91	0.128	0.108	-0.0104	42,884,136
Spring	Weekday	0.88	0.109	0.099	-0.0477	18,556,560
Spring	Weekend	0.90	0.108	0.098	-0.0540	6,597,888
Summer	Weekday	0.93	0.174	0.140	0.0053	35,463,648
Summer	Weekend	0.92	0.171	0.145	-0.0168	14,845,248
Autumn	Weekday	0.90	0.152	0.110	0.0423	9,484,233
Autumn	Weekend	0.92	0.146	0.108	0.0281	3,298,848
Winter	Weekday	0.84	0.105	0.085	0.0045	44,530,200
Winter	Weekend	0.80	0.103	0.085	0.0037	18,142,152

전체 기간을 기준으로 [표 3-33]과 같이 분석 결과, 주중의 경우 상관 계수 $r=0.91$, $RMSE=0.130kWh$, $MAE=0.106kWh$, $MBE=-0.0008kWh$ 이며, 주말 역시 $r=0.91$, $RMSE=0.128kWh$, $MAE=0.108kWh$, $MBE=-0.0104kWh$ 로 나타났다. 전체 기간 모두에서 r 값이 0.91 이상으로 매우 높게 나타나, VSM이 실측 데이터의 시간별 전력 사용 추세를 안정적으로 재현하고 있음을 확인하였다.

봄(Spring)철은 난방 종료 이후와 냉방 시작 이전의 과도기로 부하 변동성이 상대적으로 큰 시기임에도 불구하고, 주중 $r=0.88$, $RMSE=0.109kWh$, $MAE=0.099kWh$, $MBE=-0.0477kWh$, 주말 $r=0.90$, $RMSE=0.108kWh$, $MAE=0.098kWh$, $MBE=-0.0540kWh$ 로 분석되었다. 두 구간 모두에서

MBE가 음(-) 값을 나타내어, VSM이 봄철 부하를 실측 대비 다소 과소 추정하는 경향이 관찰되었다. 여름(Summer)철은 모든 계절 중 가장 높은 정합도를 보였다. 주중 $r=0.93$, $RMSE=0.174kWh$, $MAE=0.141kWh$, $MBE=+0.0053kWh$, 주말 $r=0.92$, $RMSE=0.171kWh$, $MAE=0.145kWh$, $MBE=-0.0168kWh$ 로 나타났으며, 상관 계수는 0.92~0.94 수준으로 매우 높고, MBE도 $\pm 0.02kWh$ 이내로 극히 안정적인 값을 보였다.

이는 한국 공동주택의 여름철 전력 사용이 냉방기기 중심의 뚜렷한 일중 피크 패턴을 가지며, VSM의 24시간 부하 생성 모델이 이러한 패턴의 시점·기울기 및 증감 구조를 정확하게 반영하고 있기 때문으로 해석된다. 가을(Autumn)철은 냉·난방 부하가 약화되는 비(非)피크 시즌으로, 주중 $r=0.90$, $RMSE=0.152kWh$, $MAE=0.110kWh$, $MBE=+0.0423kWh$, 주말 $r=0.92$, $RMSE=0.146kWh$, $MAE=0.108kWh$, $MBE=+0.0281kWh$ 를 기록하였다. 주중·주말 구간 모두에서 MBE가 양(+) 값을 나타내어, VSM이 실측보다 일정 수준 높은 부하를 지속적으로 산정하는 경향이 확인되었다.

겨울(Winter)철은 지역난방·가스 난방 사용 등으로 실측 전력 부하가 상대적으로 낮게 나타나는 시기임에도 불구하고, 주중 $r=0.84$, $RMSE=0.105kWh$, $MAE=0.085kWh$, $MBE=+0.0045kWh$, 주말 $r=0.80$, $RMSE=0.103kWh$, $MAE=0.085kWh$, $MBE=+0.0037kWh$ 로 분석되었다. 비록 상관 계수는 여름철에 비해 다소 낮아지지만, MBE는 매우 작은 값($+0.003\sim 0.005 kWh$)으로 수렴하여 전반적인 추정 편향은 미미한 수준임을 확인하였다.

3.5.5.2 Archetype 분류 체계의 구축

건물의 물리적 특성과 에너지 소비 구조를 반영하는 주택 Archetype 분류 체계를 구축하였다. 이 분류 체계는 난방 방식, 단열 수준, 건축연도, 전용면적의 네 가지 핵심 변수를 조합하여 정의되며, 각 변수는 실제 한국 공동주택의 에너지 거동을 결정하는 주요 요인을 반영한다.

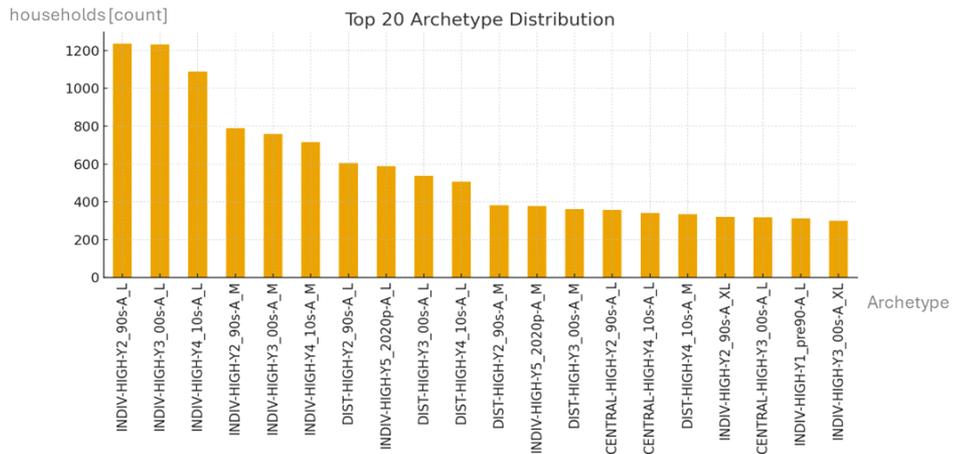
1) 난방방식(HeatingClass)은 개별 보일러를 사용하는 Individual heating과

지역난방·중앙난방에 해당하는 Centralized/District heating으로 구분하였다. 이는 한국 아파트의 난방 공급 방식이 냉·난방 부하의 계절적 패턴에 큰 영향을 미치는 점을 고려한 것이다.

2) 단열수준(InsulationClass)은 건물의 종합 단열 성능을 나타내는 Overall RSI 지표를 기준으로 Low, Medium, High 세 단계로 분류하고 연구 분석에서는 열적 성능에 의한 부하 차이가 명확히 나타나는 Low와 High 두 개 범주를 주로 활용하였다.

3) 건축연도(YearClass)는 건물 에너지 기준의 변화와 기술 고도화를 반영하기 위해, 1980년대 이전부터 2020년 이후까지를 5개 구간(80s, 90s, 00s, 10s, 2020+)으로 나누어 분류하였다. 이는 건축 시기에 따른 단열재 수준, 창호 등급, 기밀 성능 향상 등을 반영하여 에너지 성능의 구조적 차이를 설명할 수 있도록 한다.

4) 전용면적(AreaClass)은 실제 공동주택의 전력 소비가 면적 증가에 따라 비선형적으로 변화하는 점을 반영하여, 중형(A_M), 대형(A_L), 초대형(A_XL), 최상위(A_XXL) 등 네 가지 그룹으로 구분하였다.



es) INDIV-HIGH-Y4_10s-A_M의 예제 : ; Archetype = HeatingClass - InsulClass - YearClass - AreaClass

[그림 3-8] Archetype 분포도

이 네 가지 기준을 조합하여 [그림 3-8]은 HeatingClass × InsulationClass × YearClass × AreaClass 의 조합으로 2 × 3 × 5 × 4의 조합으로 총 120개로 생성된 최종 Archetype 코드 중 상위 20개의 Archetype에 대한 분포도를 표시하고 있으며 여기에서 INDIV-HIGH-Y4_10s-A_M 은 2010년대 건축-고단열-중형 면적-개별난방의 주택을 의미한다. 이와 같은 방식으로 구성된 Archetype은 한국 공동주택의 구조적 특성과 에너지 소비 특성을 체계적으로 구분할 수 있으며, VSM 기반 시뮬레이션에서 개별 가구의 에너지 사용을 보다 정확하게 재현하는 데 핵심적인 역할을 한다.

실제 생성된 데이터에 따르면, 본 연구의 KR-VSM 20K 표본은 아래와 같다.

‘DIST-HIGH-Y4_10s-A_L’,

‘INDIV-HIGH-Y2_90s-A_M’,

‘DIST-MID-Y3_00s-A_M’,

‘INDIV-HIGH-Y5_2020p-A_M’

등 다양한 Archetype 분포를 나타내며, 이는 한국 공동주택의 실제 물리적 특성을 반영하는 것으로 특히 건축연도와 단열 등급에 따른 에너지 사용 패턴의 차이가 뚜렷하며 Archetype 체계가 VSM 생성과 RSM-VSM 비교 분석에서 유효한 세분화 단위임을 확인할 수 있다.

3.5.5.3 구역·Archetype·시나리오별 비교 테이블

이상에서 구축된 분류는 다시 Location-Archetype-Scenario 조합으로 테이블을 구성하였다. 각 조합별로 겨울·봄·여름·가을의 4개 계절 × 52주 × 24시간 축에서 평균 부하 곡선이 산출되었다. RSM은 Location 정보를 포함하고 Season-Week-Hour 기반의 전국 평균 부하를 기준선으로 설정하였다. VSM은 동일한 축에 대해 Location-Archetype-Scenario 세분화 수준에서 평균 부하를 산출하여 두 시계열 곡선을 동일 축에서 비교하였다.

RSM 시계열 총 1.5억 행, VSM 시계열 총 1.7568억 행을 기초 데이터로 구성하여 총 64,512개의 Location-Archetype-Season-Week 조합을 구성

하여 분류별 상관 계수 r , RMSE, MAE, MBE를 산출하여 [표 3-34]과 같이 성능을 정량화하였다.

[표 3-34] 분류된 전체 조합의 상관관계 분석

지표	Mean	Median	Min	Max	비고(해석)
r	0.8626	0.9117	0.1324	0.9467	$r > 0$: 92.9%, $r \geq 0.7$:85.7%
RMSE	0.2230	0.1920	0.0561	0.8798	RMSE<0.3kWh: 80.8%
MAE	0.1853	0.1632	0.0510	0.7156	평균 시간당 오차 약 0.18 kWh
MBE	-0.0079	-0.0444	-0.4834	0.7156	평균 편향 ≈ 0 (과대·과소추정 편향無)

전체 조합의 상관 계수 r 의 평균은 약 0.86, 중앙값은 0.91로 나타나, VSM이 실측 기반 시계열의 전반적 부하 형상을 안정적으로 재현하고 있음을 확인하였다. 상관 계수의 분포를 보면 전체 조합 중 약 92.9%가 양(+의 상관을 나타내며, 이 중 약 85.7%는 $r \geq 0.7$ 로 강한 상관성을 보였다.

오차 지표 측면에서 RMSE와 MAE의 전체 평균은 각각 0.22kWh, 0.18 kWh로 나타났으며, 이는 시간 단위 공동주택 평균 부하(0.3~0.6kWh)를 기준으로 약 25~35% 수준의 상대오차에 해당한다. MBE는 -0.0444kWh로 확인되어, VSM이 실측 대비 전반적으로 매우 작은 범위에서 소폭 과대 추정하는 경향을 보이거나 편향의 크기는 통계적으로 미미한 수준이다.

계절 별로는 가을의 평균 상관계수(≈ 0.70)가 가장 높게 나타났으며, 겨울은 난방 부하의 민감성과 실측 전기난방 비중의 낮은 특성으로 인해 상대적으로 낮은 수준으로 관찰되었다. 여름철에는 냉방 부하 특성으로 인해 RMSE·MAE가 증가하는 경향이 있으며, 일부 조합에서는 MBE가 음(-) 값을 나타내 VSM이 피크 냉방 부하를 실측 대비 약간 과소 추정하는 경향을 보였다.

권역별 비교에서 서울과 경기권의 경우 많은 데이터와 밀집된 도시화의 구조 등으로 RMSE·MAE가 낮아 형상에서 우수한 재현성을 보인 반면, 기상

조건의 커버하는 범위가 넓고 도시의 밀집도가 확대된 분포로 다양성이 상대적으로 높은 경상권과 일부 수도권 외곽 지역에서 오차가 증가하는 경향을 보였다. 전체 조합 중 약 30%에서는 VSM이 실측보다 낮은 부하로 생성되어 MBE가 음(-) 값을 보였고, 나머지 조합에서는 소규모 과대 추정으로 양방향의 편향이 혼재하도록 KR-VSM 데이터 셋이 생성된 것을 확인할 수 있었다. 이러한 결과는 향후 Location·Scenario·Archetype·Season·Week 단위의 세부 보정 규칙의 설계를 통하여 VSM 데이터셋의 Loadshape을 보정할 수 있는 근거를 제공하며, VSM 기반의 수요예측 모델이 RSM 기반으로 보정이 가능할 수 있음을 시사한다.

1) 표본 구성 및 아키타입 분포 특성

이상에서 구축한 KR-VSM 표본의 특성은 Location - Archetype - Scenario - Season - Week 단위로 RSM과 정합성 분석 결과를 통해 점검하였다. 우선, RSM과 KR-VSM 시계열을 결합하여 구성한 평균 부하 곡선 테이블은 약 1,548,288 행, 정합성 지표 테이블은 64,512행을 포함하고 있으며, 각 행은 「지역(Location) - 건물유형(Archetype) - 시나리오(Scenario) - 계절(Season) - 주(Week)」 조합별 대표 24시간 부하 곡선과 그에 대응하는 상관 계수(r), RMSE, MAE, MBE 등의 지표를 담고 있다. 이는 단순히 전체 평균만을 비교하는 것이 아니라, 동일 지역·동일 아키타입·동일 시나리오를 공유하는 세부 집단 단위에서 VSM이 RSM의 부하 형상의 재현을 검증하는 구조이다.

Load profile과 관련해서는 LoadProfile_id, LoadProfiles_index, LP_Merge_Parameter를 연계하여 A1~A4 지역코드와 BB01~BB07 이상으로 구성된 행태형 부하 패턴이 전체 표본에 고르게 분포하도록 설계하였다. LP_Merge_Parameter에 따르면 20만 건의 세대-케이스가 약 100만 행의 Lighting, Plug, DHW, Equip.(Home Appliance) 등의 end-use 조합으로 확장되며, 각 case_id는 Load_code(A1~A4), bb_code(BB01~BB07), 조명 수준(lightning_cat)과 활동 시나리오(activity)를 조합한 A2_BB07_Mi의 profile_id를 가진다. 이를 통해 아침·저녁 이중 피크형, 야간 집중형, 주말 피크형 등 서로 상이한 생활패턴이 Location과 무관하게 폭넓게 분포하고, 동일 지역 내

에서도 다양한 행태 군이 공존하도록 했다.

건축물의 물리적 특성 측면에서는 20,000세대 × 47개 변수 Archetype 메타데이터를 통해 중앙난방·지역난방·개별난방, Stair/Floor/Mixed 건물 형태, 단열 등급(Overall_RSI_score 및 RSI_Wall/Roof/Foundation), 누기 등급(Leakage_bin), 창호 성능의 주요 파라미터가 계통적으로 분산되도록 구성하였다. Location 별 RSM-VSM 평균 부하 및 연간 kWh 비교 결과, 서울·경기·인천·충청·전라·경상·강원 등 대부분 권역에서 연간 가구당 평균 전력 사용량이 ±10% 이내에서 수렴하였고, 이는 지역별 기후와 건축 특성의 차이를 반영하면서도 전체 분포가 특정 아키타입이나 면적대에 편중되지 않았음을 시사한다.

이러한 분포를 바탕으로 산출된 정합성 지표에서 Location-Archetype-Scenario-Season-Week 조합에서 상관 계수 $r \approx 0.9$ 전후의 높은값을 나타냈으며, RMSE·MAE 역시 동일 스케일에서 $r \approx 0.7$ 수준으로 제한적인 수준에서 높은값을 확인했다. 반면 일부 조합에서는 계절·주별에 따라 오차가 상대적으로 커지는 구간도 존재하여, 표본 내부에 고효율·저효율, 고부하·저부하, 보수적·공격적 사용 패턴이 혼재되어 있음을 보여준다.

2) 정량 비교 결과

200,000 가구에서 확률적으로 임의로 추출한 20,000가구의 KR-VSM 시뮬레이션 결과를 RSM 데이터와 LocationArchetype × Season × Week 정합성을 검증하였다. RSM-VSM의 통합 데이터의 QC 수행 후 Location 매핑을 거쳐 구축된 Location × Archetype × Season × Week의 646개 조합의 시계열 평균 부하 곡선을 대상으로 [표 3-35]와 같이 상관 계수(r), RMSE, MAE, MBE을 산출하였다.

서울권 Stair-80m² 아키타입의 겨울·평일 부하 곡선은 $r=0.59$, $RMSE=0.163kWh$, $MAE=0.144kWh$, $MBE=+0.089kWh$ 로 나타났다. 이는 야간 시간대에서 VSM이 실측보다 소폭 과대 예측되는 경향이 반영된 결과이다. 반면 동일 아키타입의 겨울·주말 시나리오는 $r=0.74$, $RMSE=0.137kWh$, $MAE=0.123kWh$, $MBE=+0.063kWh$ 를 보여, 주말의 입실 시간 증가, 저녁 피크 지속시간

증가 등의 사용 행태 변화를 상대적으로 정확히 재현하였다.

경기권 Mixed-90m² 군의 여름·평일 부하 곡선에서는 $r=0.63$, $RMSE=0.136kWh$, $MAE=0.110kWh$ 를 기록하였으며, 동일 아키타입의 여름·주말 시나리오는 $r=0.74$, $RMSE=0.132kWh$, $MAE=0.107kWh$, $MBE=-0.074kWh$ 로 나타났는데, 이는 주말에 재실률 상승효과로 인하여 냉방기기를 포함한 가전기기의 사용량 증가가 반영된 것으로 VSM이 RSM과 유사한 수준으로 추종하고 있는 것을 확인할 수 있었다.

[표 3-35] 지역-아키타입-계절-주별 통계량 값 비교

지역 (Location)	아키타입 (Archetype)	계절 (Season)	주별 (Week)	r	RMSE (kWh)	MAE (kWh)	MBE (kWh)
Seoul	Stair-80m ²	Winter	Weekday	0.59	0.163	0.144	+0.089
Seoul	Stair-80m ²	Winter	Weekend	0.74	0.137	0.123	+0.063
Gyeonggi	Mixed-90m ²	Summer	Weekday	0.63	0.136	0.110	-0.047
Gyeonggi	Mixed-90m ²	Summer	Weekend	0.74	0.132	0.107	-0.074
Incheon	Stair-70m ²	Spring	Weekday	0.66	0.141	0.112	+0.021
Incheon	Mixed-85m ²	Summer	Weekend	0.71	0.138	0.109	+0.015
Chung- chung	Floor-75m ²	Winter	Weekday	0.58	0.152	0.128	+0.067
Chung- chung	Stair-70m ²	Autumn	Weekend	0.69	0.135	0.114	+0.059
Jeonra	Mixed-90m ²	Summer	Weekday	0.61	0.148	0.119	-0.028
Jeonra	Floor-85m ²	Winter	Weekend	0.73	0.133	0.111	+0.042
Gyeong- sang	Floor-70m ²	Autumn	Weekend	0.80	0.147	0.128	+0.091
Gyeong- sang	Stair-60m ²	Winter	Weekday	0.57	0.158	0.132	+0.076
Gangwon	Mixed-80m ²	Winter	Weekend	0.62	0.151	0.121	+0.083
Gangwon	Stair-60m ²	Summer	Weekend	0.68	0.142	0.115	+0.041
평균	-	-	-	0.70	0.143	0.122	-

경상 Floor-70m²의 가을·주말 시나리오는 $r=0.80$, $RMSE=0.147 kWh$, $MAE=0.128kWh$, $MBE=+0.091kWh$ 로 나타나 전체 조합 중 가장 높은 상관성을 나타냈다. 이는 10월이라는 가을의 계절이 난방 전환기로써 생활 부하

변동성이 적고 바닥난방 사용률이 낮아, RSM과 VSM 간 LoadShape가 안정적으로 일치하고 있는 것을 보여주고 있다.

전체 646개 조합의 평균 정합도는 평균 상관 계수 $r = 0.70 \pm 0.07$ 수준으로 수렴하였다. 이는 VSM이 실측 RSM의 시간대별 부하 경향을 약 70% 수준까지 설명한다는 의미로 해석할 수 있다. 지역별로는 경기·서울권이 냉방 피크 적합도가 높았고, 경상·충청권은 난방 부하에서 MBE(+)가 다소 크게 나타났다. 아키타입 별로는 Mixed 형의 90m² 이상에서 저녁 피크 절대치는 크지만 시간대별 형태가 유사하여 RMSE가 감소하는 반면, 소형-Stair 형은 거주자 행태 변화가 크게 반영되어 주중-주말 변동성이 상대적으로 큰 것으로 나타났다. 소형-Stair 형은 생활 행태 가중치가 커 주중-주말 변동성이 크게 나타났다. 이는 Wh/m² 단위로 정규화 시 형태 유사성이 강화되어 r 값이 상승하는 결과와 일치하였다.

이상의 결과는 VSM 기반 시뮬레이션이 실제 RSM 부하 패턴을 지역·주거 유형·계절·주별 단위에서 상당히 높은 수준으로 재현하며 높은 정합도를 나타냄을 보여준다.

3) 캘리브레이션 연결 포인트

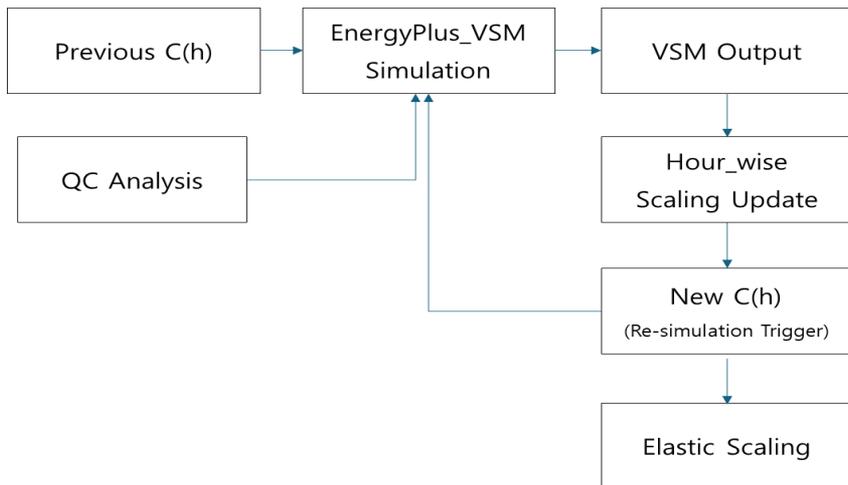
본 연구에서는 ASHRAE Guideline 14에서 제시하는 시간대별 부하 형상 비교 및 캘리브레이션 개념³⁵³⁾을 기반으로 24시간 부하 형상의 형태는 유지하되 시간대별 크기를 조정하는 스케일링 기반의 보정 절차를 RSM 기반 실측 부하 곡선과 VSM 시뮬레이션 부하곡선 간의 체계적인 보정을 위해 Elastic Scaling(탄성 스케일링) 방식의 캘리브레이션 절차를 설계하였다. 보정 루틴은 VSM input 데이터에서 연간 가상스마트미터 시계열을 생성하고, 이어서 RSM 데이터와 비교를 통한 QC를 수행하여 RSM-VSM의 주중-주말 평균 부하 곡선을 산출한 뒤, 시각별 스케일링 계수 hour_shape를 갱신하는 3단계 반복 구조로 구성된다. 각 반복(iteration)마다 RSM과 VSM의 주중-주말 평균 부하를 각각 평균 1.0으로 정규화한 후, 두 곡선의 비율(RSM/VSM)을 이용해 24시간 스케일링 벡터를 계산하고, 과도한 보정을 방지하기 위해

353) ASHRAE, (2014), 전게서, p.94-96.

시간당 보정 비율을 0.7~1.3 범위로 클리핑한 뒤 기존 hour_shape에 곱해 준다. 마지막으로 스케일링 벡터의 일평균 값을 1.0으로 재정규화하여 전체 연간 에너지량은 유지하되 시각별·일중 분포만 조정되도록 설계하였다.

이러한 보정 알고리즘은 앞서 도출한 세 가지 보정 규칙을 적용한 것이다. ① 겨울·주중에 개별난방 비중이 높은 아키타입에서 관찰된 야간 과대 모사를 줄이기 위해, 0~6시 야간 시간대의 난방에 대한 분배 가중치는 비율 값이 1보다 작게, 출근·등교 전 난방 피크가 나타나는 07~09시는 1보다 크게 조정되도록 hour_shape가 상호 연동 되도록 프로그래밍하였다. ② 여름·주말의 Δ Peak_hour 편차를 줄이기 위해 조리·가전 부하의 동시성 시작 시각을 0.5~1시간 지연시키는 효과로 주말 저녁 시간대의 스케일링 계수를 상대적으로 반영하였다. ③ 겨울·저녁 구간에서의 RMSE 저감을 위해 도입한 조명 밀도 보정은, 18~22시 일몰 이후의 스케일링 계수를 서서히 증가시키고, 법정동 기반 지역 매핑을 기반으로 권역별로 서로 다른 hour_shape를 적용할 수 있도록 추가적인 확장이 가능하다.

4) 구조적 보정의 실행

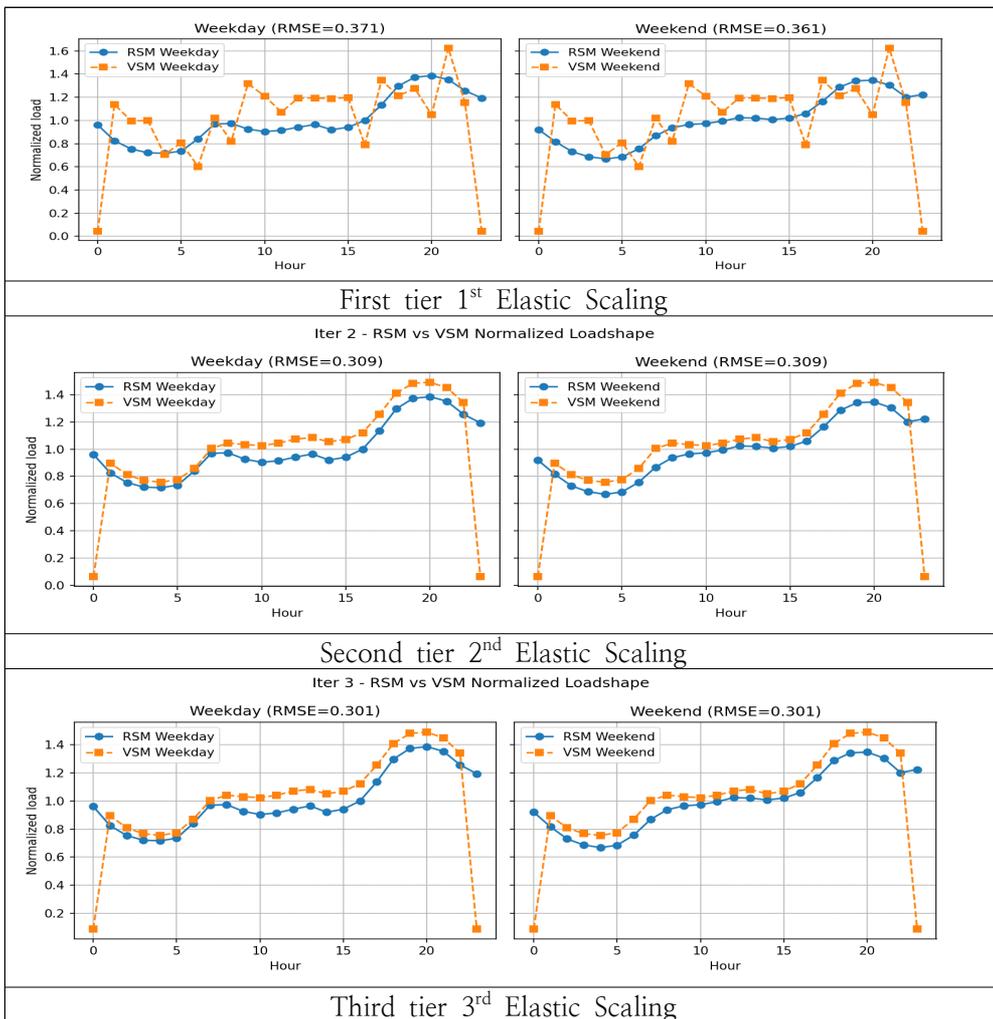


[그림 3-9] Calibration Process

초기 생성된 VSM 데이터가 일부 시간대에서 RSM 대비 과대·과소 모사되는 문제를 해결하기 위해, 주중·주말 부하 특성을 동시에 고려하는 Elastic-Scaling 구조적 보정(iterative structural calibration) 절차를 적용하였다.

보정 과정은 기존 시간대별 형상계수(hour-shape vector)를 입력으로 하여, RSM과 VSM 사이의 시간대별 상대 편차 비율을 반영하도록 설계되었으며, 전체 절차는 [그림 3-9]의 프로세스를 통하여 1·2·3차 총 세 번 반복 수행되었다.

[표 3-36] 샘플링 캘리브레이션 수행



[표 3-36]에서 1차 반복 tier 1)의 결과, VSM의 정규화된 평균 부하 곡선은 RSM 패턴 대비 야간·주간 구간에서 상대적으로 변동에 대한 편차가 존재하였고, 이 구간에 RMSE는 0.371, 주말 RMSE는 0.361로 가장 높은 오차를 보였다. 이는 초기 VSM 생성 단계에서 부여된 난수 기반 가중 계수 및 로드 셰이핑 파라미터가 RSM의 행태적 특성을 충분히 반영하지 못했음을 시사한다. 그럼에도 불구하고 1차 보정 후 형상계수의 일 평균값은 1.0으로 유지되어 연간 총 전력량은 보존되었다.

[표 3-36]의 2차 반복 tier 2)에서는 RSM 대비 시간대별 과대·과소 패턴이 상대적으로 안정화되고, 주중 RMSE가 0.309, 주말 RMSE가 0.309로 소폭 감소하였다. 이는 1차 결과 대비 약 16~17% 수준의 RMSE가 개선된 것이며, 캘리브레이션에서 사용된 RSM-VSM 비율 기반의 탄성 스케일링이 피크 구간과 오프피크 시간대에 상대적 비율을 효과적으로 억제하고 있음을 보여준다. 실제로 시각화된 평균 부하 곡선에서는 RSM 대비 VSM의 오전 중·저 부하 구간(7~11시) 및 저녁 피크 구간(18~22시)에서 LoadShape의 형상 유사도가 향상된 것을 확인할 수 있었다.

[표 3-36]의 3차 반복 tier 3)에서는 전 단계 대비 보정 범위가 더욱 미세해졌으며 RMSE는 주중 0.301, 주말 0.301로 조금 낮아졌다. 특히 2차 대비 개선 폭은 작지만 안정적으로 수렴되었으며, 반복이 진행될수록 VSM의 시간대별 LoadShape 형상은 RSM의 실측 패턴과 유사한 형태를 추종하고 있다.

또한 모든 반복에서 형상계수의 값이 1년의 에너지 총 사용량을 기준으로 1.0이 유지되고 있어서, 시간별-주간별-계절별-연간의 전력 사용량에 대한 보정이 연간 전력량(total_kWh)에는 영향을 미치지 않도록 설계하였다.

다음의 [표 3-36]에서 탄성 스케일링 프로세스를 3회 반복만으로 전체 RMSE가 약 13~19% 감소하는 효과와 함께 시각적으로도 VSM이 RSM 패턴으로 근접하는 것을 확인하였고, 이는 대규모 파라미터에서도 RSM-VSM의 부하 비율을 활용한 구조적 보정이 충분히 정합화할 수 있음을 확인할 수 있었다. 또한 각 반복 단계에서 저장된 로그는 시간이 지날수록 보정량이 급격히 감소하는 형태를 보여, Elastic-Scaling 알고리즘이 명확한 수렴 특성을 갖고 있음을 실증하였다.

이는 실제 피크 전력 구간과 비피크 전력 구간대의 가중 부하 패턴을 반영하기 위한 VSM 보정 방식으로서, 단순 통계적 선형 스케일링보다 더 현실적인 부하 형상 복원 효과가 제공이 가능한 것을 실험으로 확인할 수 있었다.

3.5.5.4 과대/과소 추정 진단과 오차 분해

본 연구에서는 VSM 기반 시계열이 RSM 부하 곡선과 비교하여 어떠한 구조적 편향을 보이는지 진단하기 위해, 20,000가구 대상의 연간 시뮬레이션 결과를 기반으로 정규화 평균 부하(normalized loadshape)를 추출하고 이에 대한 오차 분해(error decomposition) 기법을 적용하였다. 우선, 모든 개별 시계열을 시간대별 평균 1로 정규화함으로써 절대 전력량 수준의 차이를 제거하고, 전력량의 daily pattern에 대한 비교가 가능하도록 사전 처리하였다. 이는 VSM의 난방·냉방·가전기기 분배 함수 등 세부 파라미터가 전력 규모보다는 시간대별 사용 행태에 더 큰 영향을 미친다는 점을 고려한 절차이다.

정규화된 RSM과 VSM의 평균 부하 곡선은 주중(Weekday)과 주말(Weekend)로 구분하여 산출하였으며, 각 시간대 h 에서의 오차는 $R(h)$ 의 RSM 정규화된 평균 부하와 $V(h)$ 의 VSM 정규화된 평균 부하와 다음과 같이 정합성 성분으로 분해하였다.

1) 절대 오차 성분(Absolute Deviation)

$$E_a(h) = |R(h) - V(h)|^{354},$$

여기서,

$E_a(h)$: 절대 오차 성분,

$V(h)$: VSM 평균 부하,

$R(h)$: RSM 평균 부하,

이는 특정 시간대에서 VSM이 RSM과 얼마나 다른 수준의 부하를 모사했

354) ASHRAE, (2014), 전계서, p.22-23.

는지를 나타내는 1차 척도이다. 야간·새벽 시간대에서의 과대 모사 여부, 18-21시 피크 시간대에서의 과소 모사 여부 등을 직관적으로 파악할 수 있다.

2) 상대 오차 성분(Relative Bias)

$$E_r(h) = \frac{V(h) - R(h)}{R(h)} \text{ 355),}$$

여기서,

$E_r(h)$: 상대 오차 성분,

이는 동일한 절대 오차라도 실제 부하가 낮은 시간대(예: 01-05시)에 상대적으로 더 큰 비율 오차로 반영되도록 하여, 야간 난방·조명 부하의 편향을 보다 민감하게 탐지하도록 설계되었다.

3) 주기적 편향 성분(Pattern Drift)

VSM은 난방·냉방·조리·가전기기 동시성(concurrency)의 모사에 따라 시간대별 국지적 진동이 발생할 수 있다. 이를 탐지하기 위해 3시점 이동평균 $\bar{V}(h)$ 을 사용하여 다음과 같은 Drift 성분을 정의하였다.

$$E_d(h) = V(h) - \bar{V}(h) \text{ 356),}$$

여기서,

$E_d(h)$: 주기적 편향 성분,

$V(h)$: VSM 평균 부하,

$\bar{V}(h)$: VSM 3hr 이동평균 부하,

이는 일정 시간대에서 VSM이 불필요하게 이상 진동을 보이는지, 혹은 RSM 대비 피크가 위·아래로 이동하는(peak shift) 패턴을 보이는지를 판단하기 위한 지표이다.

355) ASHRAE, (2014), 전계서, p.23-24.

356) ASHRAE, (2014), 전계서, p.94-96.

4) 종합 RMSE 성분

$$RMSE = \sqrt{\frac{1}{24} \sum_{h=0}^{23} (R(h) - V(h))^2} \quad 357),$$

이는 shape 전체의 회귀 품질을 수치적으로 요약한 지표이며, 본 연구의 Elastic Scaling 보정 루프에서 성능 수렴을 판단하는 핵심 기준으로 사용되었다.

이상의 4가지 성분으로 진단한 결과, 겨울·주중 야간 0~6시의 지속적 과대 모사가 나타났는데, 이는 국내 공동주택의 개별난방 아키타입에서 VSM이 야간 기저 난방이 과도하게 유지되는 경향을 보였으며, 이는 주간 난방 분배 함수의 야간 부하 할당 비율이 실제 RSM의 거주자 행동 패턴보다 거주자의 재실에 따라 VSM의 LoadProfile 시나리오에 높게 설정된 데 기인하는 것으로 분석할 수 있었다. 여름·주말의 ΔPeak_hour는 주말 오후 17~19시의 조리·냉방·가전 동시성 증가가 자연스럽게 발생하지만, VSM에서는 해당 피크가 1시간가량 앞당겨지거나 지나치게 피크치로 치솟는 현상이 확인되었다. 이는 냉방기 가동, 조리 활동 시작 등을 단일 확률 분포로 모사하는 VSM 특성이 실제 가구 거주자의 행동 패턴을 충분히 반영하지 못한 결과이다.

이와 같은 오차 분해 결과는 단순히 RMSE 수치에 의존하는 것이 아니라, 가구별 물리·행태·점유자 파라미터에서 편향이 발생하는지를 구조적으로 파악하는 데 핵심적인 근거를 제공하였다³⁵⁸⁾. 특히 본 연구에서는 이러한 시간대별 편향 패턴을 Elastic Scaling 보정 루프와 연계하여 자동 보정이 가능한 형태로 구조화하였다.

3.5.6 캘리브레이션 규칙 및 Re-Simulation

앞 절에서 제시한 과대·과소 모사 진단 결과에 기초하여, 본 연구에서는 VSM 생성 과정에 자동 적용되는 Elastic Scaling 기반의 calibration(보정)을 수행하였다. 본 보정 규칙은 단순한 전력 규모 보정이 아니라, 주간 패턴을 시간대별로 조정하는 구조적 보정 기법을 기반으로 설계하였으며, 다음 식을

357) ASHRAE, (2014), 전게서, p.23.

358) Keogh, E., & Ratanamahatana, C. A., 전게논문, p.2-5.

캘리브레이션과 Re-Simulation의 기본 원리로 사용하였다³⁵⁹⁾³⁶⁰⁾.

$$\frac{R(h)}{V(h)},$$

여기서,

R(h) : RMS, V(h) : VSM,

$$NBME = \frac{\sum_{i=1}^n (x_i - \hat{y}_i)}{x \times \bar{y}}, \quad CVRMS = 100 \times \frac{[\sum_{i=1}^n (y_i - \hat{y}_i)/n]^{1/2}}{n} / \bar{y},$$

여기서,

y_i : measured data,

\bar{y} : mean of measured data,

\hat{y} : simulated data,

이는 RSM과 VSM의 Loadshape 비율(RSM/VSM)을 시간대별로 직접 반영하여 VSM을 정렬시키는 개념에 기반한다. 또한 보정 계수가 과도하게 커지는 것을 방지하기 위해, 코드에서는 다음의 상·하한 클리핑 및 스무딩 절차를 적용하였다.

$$0.8 \leq \alpha \frac{R_{wd}(h)}{V_{wd}(h)} + (1 - \alpha) \frac{R_{we}(h)}{V_{we}(h)} \leq 1.2,$$

$$C^{k+1}(h) = \frac{C^{(k)}(h) \text{Smooth}_3 \cdot (\text{clip}[0.8, 1.2])}{\text{mean}},$$

여기서,

$R_{wd}(h)$: Weekday Loadshape,

$R_{we}(h)$: Weekend Loadshape,

$V_{wd}(h)$: Weekday Loadshape,

$V_{we}(h)$: Weekend Loadshape,

359) Sun, K., Hong, T., & Chen, Y. (2016). A pattern-based automated approach to building energy model calibration. *Energy and Buildings*, 117, 287–297.

360) Chen, Y., Wei, W., Song, C., Ren, Z., & Deng, Z. (2023). Rapid building energy modeling using prototype model and automatic model calibration for retrofit analysis with uncertainty. *Buildings*, 13(6), 1427.

0.8과 1.2의 범위는 Elastic Scaling 수행 시 과도한 왜곡을 방지하기 위하여 $\pm 20\%$ 이내로 수렴하도록 설정하였으며, 이는 입력 스케줄의 상대적 형상을 유지하면서 반복적으로 수렴이 가능하도록 상·하한을 둔 제한적 범위로 Scaling이 반복되는 구조로 설계하였다³⁶¹⁾. Smooth₃(*)는 3시간 이동평균에 적용된 함수로 최종 정규화에 대하여 시간의 상대적 비중만을 보정³⁶²⁾하고 전체 부하의 연간 평균 전력 사용량의 통계치를 기준으로 정합되도록 적용하기 위한 목적이다.

3.5.6.1 보정 규칙의 정규화

1) 주간 난방 분배 함수의 야간 가중치 감소, 아침 시간대 가중 증가

겨울·주간/야간에서의 과대 모사 편향을 해소하기 위해, 난방 분배 함수 (Heating Distribution Function)의 0~6h 야간 가중값을 탄력계수 $C(h)$ 에 의해 0.8 이하로 하향 조정하고, 07~09시의 가중값은 1.05~1.15 범위에서 상향 조정하였는데, 이는 선행 연구의 야간 저부하-아침 피크 구조를 갖는다는 선행 연구의 대한민국 주거 가구의 LoadProfile과 + CREST 규칙을 참고하여 적용한 것이다³⁶³⁾. 여기서 적용한 보정 계수의 범위는 본 연구의 Elastic Scaling 보정 과정을 통하여 경험적으로 설정한 것이며 이는 기저 부하 유지 시간의 과대 설정이라는 구조적 문제를 개선하고자 한 것이다.

2) 가전·조리 동시성 시작 시각의 확률적 지연

여름·주말의 $\Delta Peak_hour(-1\sim 0h)$ 문제를 해결하기 위해, 조리·가전기기 동시성(concurrency) 시작 시각의 확률 분포를 0.5~1.0시간 지연시키는 보정 규칙을 도입하였다. Elastic Scaling의 계수 $C(h)$ 가 오후 시간대에서 1.1 이상으로 반복적으로 산출될 경우 이를 Re-Simulation을 통하여 동시성 분포를 자동 조정하도록 구성하였다.

361) 이지영, 선영규, 이선민, 김수현, 김영규, 이원섭, 심이삭, 김진영., 전계논문, p.93, p.96.

362) Sun, K., Hong, T., & Chen, Y., 전계논문, p.16-17, p.27-28.

363) Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K., 전계논문, p.3-4, p.7-8.

3) 조명 점등 시각의 지역 차등

겨울·저녁 조명 부하의 과소 모사를 보완하기 위해, 지역별 일몰 시간을 기반으로 조명 점등 시각 시각을 결정하는 보정 계수를 도입하였다. 강원·경북 동해안 지역의 조명 부하는 16~18시 계수 $C(h)$ 가 반복적으로 1.1 이상 상승하는 패턴을 보였으며, 이를 조명 분배 함수의 지역별 일몰 기반 보정 로직과 연동하였다.

3.5.6.2 Re-Simulation 수행

Elastic Scaling 갱신이 반복되면서도 RMSE가 충분히 수렴하지 않을 경우, 개별 파라미터의 구조적 문제로 판단하여 다음의 조건을 충족하면 Re-Simulation을 수행하도록 구성하였다.

1) RMSE 감소율이 2회 연속 3% 미만일 때

shape 기반 보정만으로는 구조적 개선이 어렵다는 의미로 해석하였다.

2) $C(h)$ 계수의 절대 변화량이 특정 시간대에서 지속적으로 1.2의 상한 또는 0.8의 하한에 걸릴 때,

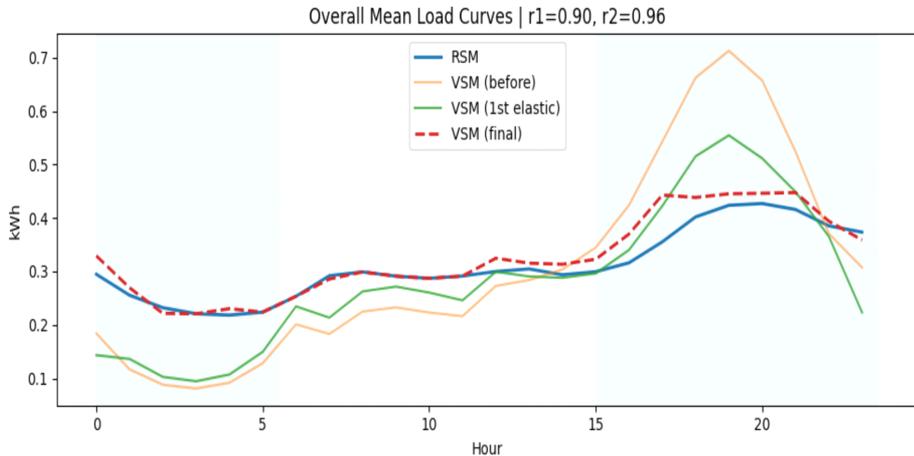
분배 함수 자체의 모형화가 잘못되었음을 시사하고 있다.

3) 야간·피크 시간대에서의 상대 오차 $Er(h)$ 값이 ± 15 이상으로 유지할 때

VSM 난방·냉방·가전 분배 함수를 근본적으로 Re-Simulation하였다. 이러한 Re-Simulation 체계는 단순한 반복 보정이 아닌, 모형 자체의 구조적 문제를 자동 탐지하고 필요한 경우 VSM 시계열 데이터를 기반으로 난방·가전·조명 파라미터를 자동으로 재설정하도록 유도하는 역할을 한다.

3.5.6.3 Elastic Scaling 적용 결과

Elastic Scaling 보정 루프를 20,000 가구를 대상으로 1년 간 8,760 시간 기준으로 총 1.75억 행의 VSM 시계열 데이터 전체를 시뮬레이션에 단계적으로 적용한 결과는 [그림 3-10]과 같이, RSM과 VSM 간의 주간 부하 패턴 불일치가 반복 수행에 따라 지속적으로 감소하였다.



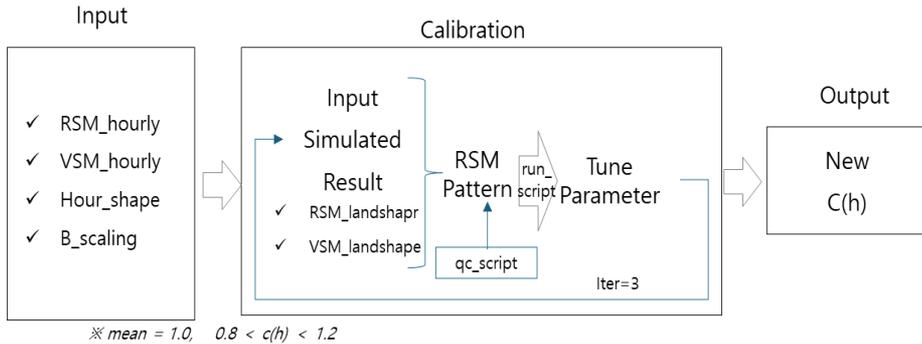
[그림 3-10] Elastic Scaling 보정 (before, 1차, 2차)

1차 보정만으로도 상관 계수는 $r_1=0.90$ 수준을 유지하면서, RSM 대비 과대·과소 모사 구간이 일정하게 수렴하고 있으며 전체적인 LoadShape의 형상은 피크구간과 비피크 구간의 전력 소비량의 형태와 유사한 패턴으로 유지하도록 반영하였다.

2차 보정에서는 시간대별 스케일링 함수 $C(h)$ 의 변동 폭이 좁아지면서, 아침 기상·준비시간(7~9시)과 저녁 취사·가전 집중 구간(18~21시)의 증가·감소 패턴이 RSM 곡선 대비 피크값은 여전히 차이가 났지만 최대 피크값은 1차보다 확실히 줄어드는 정형성을 확인할 수 있었다.

최종 보정(2nd elastic + seasonal + mid/tail smoothing 적용) 단계에서는 상관 계수가 $r_2 \approx 0.96$ 까지 상승하며 실측 패턴과의 수렴이 좀더 근접하였다.

야간 저부하 시간대는 RSM 수준으로 안정적으로 하향 조정되었고, 저녁 최대 부하는 VSM의 과도한 피크치가 제거되어 RSM 피크와 유사한 LoasShape 형태를 추종하였다. 이는 본 연구의 Elastic Scaling 루프가 단순 레벨 조정이 아니라, 시간대별 전체 부하의 shape mismatch를 구조적으로 보정하는 단계로 작동했음을 뜻한다.



[그림 3-11] Elastic Scaling process³⁶⁴⁾

[그림 3-11]과 같이 보정 과정 전반에 적용된 시간대별로 0.8~1.2 범위의 스케일링 한계를 적용하고 보정 후 연간 총부하 재정규화를 mean=1.0을 유지하는 장치를 적용하여 과보정 방지와 스케일링의 안정성을 확보하였다. 그 결과 VSM은 초기 대비 모든 시간대에서 RSM과 유사한 패턴이 확대·축소된 버전에 가까운 균질한 형태로 수렴하였고, 이는 실측 RSM 기반의 보정이 효과적으로 수행되었음을 입증하고 있는 것이다.

여기에서 개별 1가구의 Elastic Scaling은 [표 3-36]와 같이 RSM을 VSM이 시계열로 거의 일치된 곡선으로 추종을 수행할 수 있지만 1.75억 개의 데이터 전체를 미세조정을 통하여 Elastic Scaling을 수행하는 것은 보정의 차수가 진행될수록 보정 시간이 과도하게 소요되는 등 데이터 처리의 과부하가 발생하였고, 100G 바이트 수준의 빅데이터를 클라우드로 분석을 진행하는 특성상 VSM이 RSM 데이터와 정합까지 구현하는데 지속적인 Runtime error로 Elastic Scaling의 보정의 가능성을 확인하는 것으로 실험을 중단하였다.

364) Sun, K., Hong, T., & Chen, Y., 전계논문, p.8.

IV. 에너지 수요 반응 적용성 평가

4.1 수요 반응 적용을 위한 VSM 데이터셋의 정보 생성 체계

본 연구를 통해 개발된 수요 반응 적용을 위한 VSM 데이터셋이 단순히 실측된 RSM의 통계적 대체물이 아니라 DR 대상을 선별하고, DR 반응성을 평가하며, 궁극적으로는 DR·VPP 운용 전략 수립을 위한 의사결정의 입력정보로 활용이 가능한지를 판단해야 할 필요가 있다.

이에 본 절에서는 DR 적용을 전제로 VSM 데이터셋의 정보 생성 체계를 정의하고 시계열의 부하 데이터로부터 의미 있는 특징 벡터를 구성하는 과정을 체계적으로 표시한다. 제3장에서 EnergyPlus-CREST 기반의 VSM 데이터셋 결과물로부터 시간별 전력 사용량 시계열을 생성한 후, 이를 연간, 계절별, 월별, 주중·주말별, 일별, 시간별의 다양한 해상도로 집계하여 다양한 특징 벡터를 구성한다. 이러한 특징 벡터는 단순한 통계 요약치가 아니라 계절성, 피크의 구조, 일중 부하 변동성, 냉·난방 의존도 등 DR 반응성과 밀접하게 연관된 정보를 내포하여 설계된다. 특히 동일한 연간 사용량을 갖는 세대라도 시간적 부하 구조에 따라 DR 기여도가 달라질 수 있다는 점을 고려하여, 시간 축 해상도의 특징 설계를 중점으로 수행하였다.

다음으로 본 절에서는 VSM 데이터셋으로부터 생성된 특징 벡터가 실제 DR 적용에 적합한지를 검증하기 위해서, 회기 성능 및 분류 가능성을 사전적으로 평가하고 이후 수행되는 Classifier와 Extractor 모듈에서 단순 통계적 분류를 넘어 실측된 RSM 데이터에 적용 가능한 DR 입력 생성 도구로서의 가능성을 확인하였다. 즉 본 절에서는 VSM 데이터셋을 기반으로 한 정보 생성 체계를 정립하고 시계열 부하 데이터를 DR 적용 관점에서 재구성함으로써 역추정을 통한 DR 연계 실험이 갖는 실질적 의미를 명확히 하고자 하며, 이를 통해 VSM 데이터셋이 실증적이며 단계적으로 수요 반응과 DR·VPP 운

용 전략의 수립에 반영하고자 하였다.

4.1.1 DR 판단을 위한 시계열 정보 구조화

한국형 KR-VSM 데이터셋을 활용하여 건물별 전력 사용 특성을 정량적으로 표현하기 위한 다차원 204D 특징 벡터(Feature Vector)를 구축하였다. 원본 VSM 시계열은 1년 8,760시간 동안 20,000세대의 시간별 전력 사용량으로 구성되어 있으며, 전체 관측치 규모는 약 1억 7,568만 건($8,760 \times 20,000 = 175,680,000$ rows)에 달한다. 이와 같은 대규모 시계열 데이터를 안정적으로 처리하기 위해 데이터는 40만~60만 행 단위의 chunk로 분할하고 A100 GPU로 순차·병렬 처리하는 방식을 적용하였고, 처리 과정 전반에 대한 로그(log)를 수집하여 데이터 적재의 완결성을 확인하였다.

시간 기반 특징을 생성하기 위해 우선 timestamp 컬럼으로부터 월(month), 주(day-of-week), 계절(season), 주말 여부(is_weekend), 시간대(hour)와 같은 파생 변수를 생성하였다. 이러한 파생 변수는 시계열의 계절성, 주별 패턴, 시간대별 부하 특성을 추출하기 위한 필수 요인으로, 이후 단계에서 고해상도 특징 공간을 구성하는 핵심 기준 정보로 활용된다.

각 세대의 월별 전력 사용량의 평균값을 기준으로 산출하여 12D의 월별 대표 사용량 벡터를 M01~M12로 생성하고, 이는 가구의 연중 소비 패턴의 구조를 여름철·겨울철의 피크 사용량 등 계절적 에너지 소비 패턴을 간결하게 반영한 변수이며, 세대별 미시적 부하 특성을 반영하기 위해 계절(season)×주중·주말(is_weekend)×시간대(hour)의 조합에 따른 평균 사용량을 산출한 변수의 조합으로 총 204D VSM Feature Vector를 구축하였다. 이 조합은 4개의 계절×주중·주말의 주별 구분×24개 시간대로 계산의 곱으로 총 192개 변수를 구성하였다. 이렇게 생성된 192개의 시간-계절적 특징은 야간 사용 집중형, 주말 집중형, 냉·난방 집중형 등 각 가구의 세부적인 행태 특성을 반영하는 높은 해상도의 시계열 변수로 설명이 가능하다.

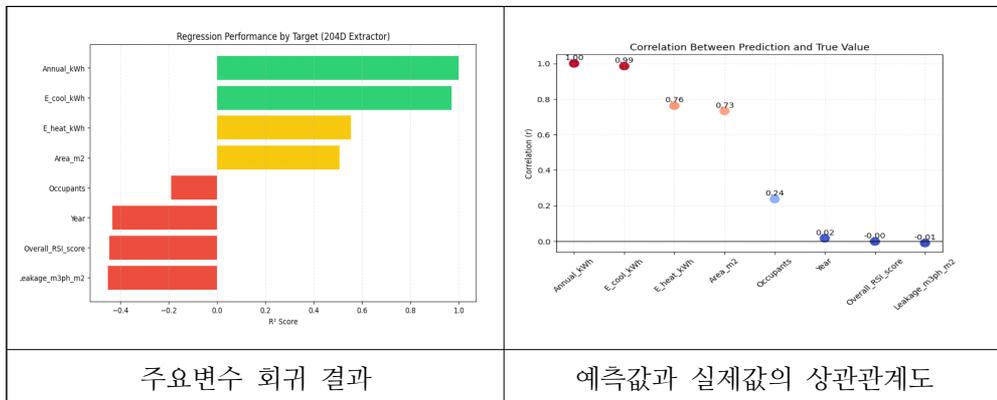
12D의 월별 벡터와 계절-시간대 192D를 결합하여 최종적으로 204D의 고

정 길이 특징 벡터를 구성하였으며, 20,000세대 전체에 대해 20,000 × 204D의 4,080,000개의 feature matrix를 구축하였다. 생성된 모든 특징 컬럼에서 결측값이 존재하지 않음을 확인하였으며, 최종 데이터는 회귀 분석, 분류모델, 비선형 학습, DR 기반 시뮬레이션 등의 후속 분석에 활용할 수 있도록 vsm_feat_204_ext로 출력하였다.

4.1.2 DR 민감도의 회귀 설명력 평가

앞 절에서 구축한 204D 특징 벡터의 설명력을 검증하기 위해, 본 연구는 주요 연속형 건물-에너지 특성 변수들을 대상으로 회귀 기반 성능 평가를 수행하였다. 회귀 분석은 Random Forest Regressor를 기반으로 하였으며, 입력 변수로는 20,000세대의 204D 특징 벡터, 타깃 변수로는 건물별 연간 총 전력 사용량(Annual_kWh), 냉방 부하(E_cool_kWh), 난방 부하(E_heat_kWh), 전용면적(Area_m2), 점유 인원(Occupants), 건축연도(Year), 단열 성능(Overall_RSI_score), 기밀 성능(Leakage_m3ph_m2) 등 총 8개 변수를 선정하였다. 모델의 성능은 r, RMSE, MAE와 함께 설명력 지표인 결정 계수 R²을 사용하여 평가하였다³⁶⁵⁾.

[표 4-1] 204D의 주요 변수별 성능 평가 결과



365) 홍승연, 송승영., 전계논문, p.861-863.

[표 4-1]과 같이 회귀 분석 결과, 204D 특징 벡터는 부하 기반 변수에 대해 매우 높은 예측 성능을 보였다. 특히 연간 총 전력 사용량(Annual_kWh)은 $R^2=0.9999$, $r \approx 1.0$ 으로 사실상 완벽한 설명력을 나타냈으며, 냉방 부하(E_cool_kWh) 또한 $R^2=0.97$, $r=0.985$ 로 매우 우수한 상관성을 보였다. 난방 부하(E_heat_kWh)와 전용면적(Area_m2)의 경우 각각 $R^2=0.554$, 0.508 로 중간 수준의 설명력을 보였으나, 회귀식이 주요 패턴을 잘 포착하고 있음을 확인하였다. 반면, 점유 인원(Occupants), 건축연도(Year), 단열성능(RSI_score), 기밀 성능(Leakage)과 같은 건물 외피·구조적 특성 변수들은 R^2 이 0에 가깝거나 음수($-0.19 \sim -0.45$) 수준으로 나타났으며, 이는 204D 특징 벡터만으로는 이러한 태생적 물리 특성의 변수들을 회귀적으로 복원하는 데 구조적 한계가 있음을 확인할 수 있었다.

실제값-예측값 상관관계도 동일한 패턴이 확인되었다. Annual_kWh와 E_cool_kWh는 각각 $r=1.0$, $r=0.985$ 로 매우 강한 상관성을 보였으며, E_heat_kWh와 Area_m2는 각각 $r=0.762$, $r=0.732$ 로 높은 상관성이 나타났다. 그러나 Year, RSI_score, Leakage와 같은 물리·구조 특성은 상관 계수가 거의 0에 가까워 ($-0.01 \sim 0.02$), 전력 사용 패턴만으로 건물 외피 성능이나 건축연도를 추정하기 어렵다는 점을 시사한다.

이와 같은 결과는 204D 특징 벡터가 전력 소비 패턴과 직접적으로 연동되는 부하 기반 변수에 대해서는 높은 설명력을 제공하지만, 외피 단열성능·기밀 성능·연식과 같이 건물의 고정적 물리 특성은 시계열 부하 패턴만으로 설명하기 어렵다는 것을 명확히 보여준다. 이는 이후 Classifier 구축 시 물리 특성 기반 변수는 메타데이터 또는 건물 특성 모듈과 결합한 다중 입력 모형이 필요함을 의미하며, 본 연구에서 도출된 중요한 시사점이다.

4.1.3 DR 속성 분류 가능성 평가

회귀 분석을 통해 얻은 예측값의 품질을 보다 실용적인 수준에서 평가하기 위해, 본 연구는 연속형 타깃 변수를 범주형 Class로 변환하는 회귀 기반 매

핑(regression-to-class mapping)을 수행하였다. 이는 전장의 분류모델(Classifier)에서 사용되는 AreaClass, YearClass, RSI_category, Leakage_grade 등의 클래스 정의에 따라 회귀 예측치를 구간화하여 binding하고 각 클래스에 속할 확률을 [표 4-2]와 같이 산출하고, 이에 대한 정확도와 혼동 행렬을 비교·분석하는 방식으로 진행되었다.

[표 4-2] Class 별 회귀 예측 정확도

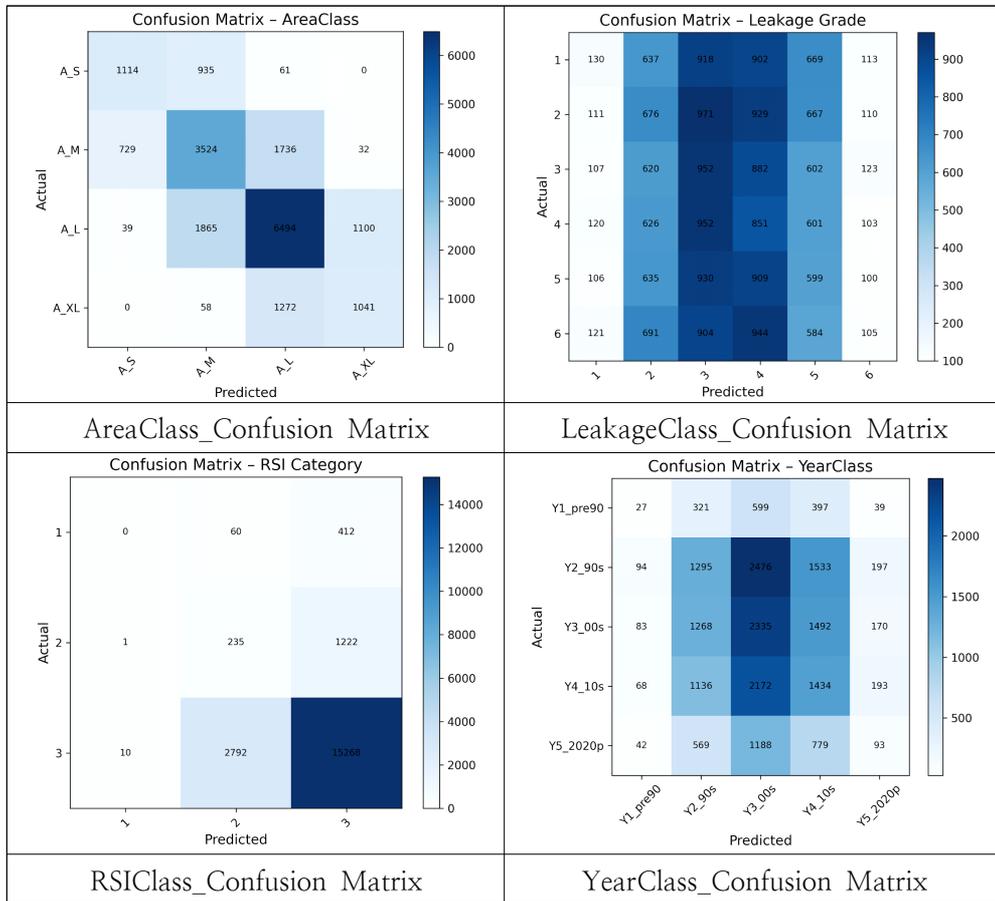
No.	Class Var	Samples	Accuracy
0	YearClass	20,000	0.25920
1	AreaClass	20,000	0.60865
2	Overall_RSI_Class	20,000	0.77515
3	Leakage_Class	20,000	0.16565

우선 전용면적(Area_m2)의 경우 사전 정의된 4개 구간(A_S, A_M, A_L, A_XL)에 따라 구간화 한 결과, 회귀 기반 클래스 매핑 정확도는 약 0.608로 나타났다. 이는 204D 특징 벡터가 면적 규모를 비교적 안정적으로 추정할 수 있음을 보여주는 결과로, LDA 기반 AreaClass 분류 정확도(약 0.47) 대비 개선된 결과이기도 하다. 단일 성능(Overall_RSI_score)을 3개 구간으로 변환한 경우에도 매핑 정확도는 약 0.775로 나타나, 부하 패턴 기반 단일성능 대분류(category-level)는 어느 정도 가능함을 보여주었다. 반면 건축연도(Year)의 경우 ‘Y1_pre90~Y5_2020p’의 5개 구간으로 매핑한 결과, 정확도는 약 0.259로 나타났다. 이는 204D 특징 벡터가 건축연도의 물리적 의미를 충분히 반영하지 못한다는 점을 재확인하는 결과이다. 기밀 성능(Leakage_m3ph_m2)을 R2000의 6개 등급으로 매핑한 실험에서도 정확도는 약 0.166으로 매우 낮아, 기밀 특성의 경우 부하 패턴과의 연계성이 약하며 시계열 기반 특징만으로는 재현이 어려움을 보여준다.

[표 4-3]에서와 같이 혼동 행렬 분석에서도 동일한 경향을 확인할 수 있는데, AreaClass는 저면적·중면적 구간에서 비교적 정확한 분류가 이루어졌다.

YearClass와 Leakage_grade는 대부분의 예측값이 인접한 클래스 또는 중앙 영역으로 오분류되고, 명확한 경계가 형성되지 않고 분류가 집중되는 현상을 확인하였다. 이는 회귀 기반 클래스 매핑 전략 자체의 문제가 아니라, 204D 특징 벡터가 해당 물리 변수와 충분한 통계적 상관을 가지지 못한다는 구조적 한계에서 기인한다.

[표 4-3] Class 별 Confusion Matrix



[표 4-3]의 실행 결과에서 204D 기반 회귀-클래스 매핑 실험은 전력 사용 패턴을 기반으로 AreaClass의 추정은 매우 효과적인 반면, YearClass와 Leakage_grade 등 건물의 물리 요소는 변수 개별적으로 추정하기에는 한계가 있음을 명확히 보여준다. 이러한 분석 결과는 다음에서 수행되는 분류 모델

설계 시 중요한 기초 자료로 활용되며, 복합 모형 또는 메타데이터 결합형 Classifier 설계의 필요성을 뒷받침하고 있다.

4.2 수요 반응 적용 가능성 검증을 위한 분류 성능 평가

4.2.1 DR 대상 식별을 위한 선형 분류 기준

4.2.1.1 DR 판단 단위별 적용성 비교

본 연구는 QSFBSM에서 제시한 VSM 기반 BSEM 분류 체계의 확장된 연구 모델로, LDA는 이론적 측면에서 데이터의 특성 분리도를 극대화하기 위해 입력 피처의 공분산 구조를 충분히 대표하는 집계형 통계량을 사용하고 있는데, 이는 각 시간대 혹은 기간별 단순 시계열 데이터보다 평균·분산·최대·최소의 집계 요약치가 클래스 경계를 구분하는 선형 판별함수의 결정면을 안정적으로 구분할 수 있기 때문이다³⁶⁶⁾.

구체적으로, K개의 클래스로 구성된 데이터 집합에서 각 클래스의 평균 벡터와 전체 평균 벡터를 이용해 between-class scatter 행렬 S_B 를 정의하고, 개별 샘플과 해당 클래스 평균 간의 편차를 이용해 within-class scatter 행렬 S_W 를 정의한다. 이때 S_B 는 클래스 평균 간 거리와 각 클래스 표본 수의 가중 합으로 구성되며, S_W 는 각 클래스 내부 분산의 합으로 구성된다³⁶⁷⁾.

LDA의 목적은 두 행렬의 비율을 최대화하는 선형 변환 벡터 w 를 찾는 것으로 두 행렬은 각각 다음과 같이 정의된다.

$$S_B = \sum_i n_i (\mu_i - \mu)(\mu_i - \mu)^T,$$

$$S_W = \sum_i \sum_{x \in w_i} (x - \mu_i)(x - \mu_i)^T$$

366) Raudys, S., & Duin, R. P. W. (1997). Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. Elsevier Science B.V., Pattern Recognition Letters 19. 385–392.

367) Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. AI Communications, 30(2), 169–190. <https://doi.org/10.3233/AIC-170729>.

여기서,

최적 판별 축 W 는

$S_W^{-1}S_B W = \lambda W$ 를 통해 계산된다³⁶⁸⁾.

최적 판별 벡터 w 는 두 단열 수준을 가장 잘 분리하는 투영 축이 되며, 고유값 λ 는 분리의 강도를 나타낸다.

이렇게 도출된 판별 축은 클래스 간 평균이 가장 잘 분리되는 방향을 의미하며, 대응하는 고유값의 크기는 분리 강도를 정량적으로 나타낸다.

본 연구에서는 이러한 LDA 이론적 틀을 VSM 및 RSM에서 추출된 연간 전력 사용 특성에 적용하여, 주거용 공동주택의 단열 수준(InsulClass), 난방 방식(HeatingClass), 전용면적(AreaClass), 건축연도(YearClass), 지역(Location) 등 건물 속성을 확률적으로 판별하는 기본 분류 모델을 구성하였다. LDA는 클래스별 공분산 행렬이 동일하고 각 클래스가 다변량 정규분포를 따른다는 가정을 전제로 하며, 이러한 가정하에 최적 분류 경계를 선형 판별함수 형태로 표현할 수 있다는 점에서 baseline 분석에 적합하다³⁶⁹⁾.

모형 구현은 LDA 이론에 기반한 선형 판별 모형을 적용하여 주거용 공동주택의 주요 건물 속성을 기준으로 분류 성능을 평가하였다. 분류 성능 평가는 교차 검증 기반의 일반화 성능 지표를 활용하여 수행되었으며, 분류 정확도(Classification Accuracy)와 macro F1-score를 통해 각 타겟팅 변수에 대한 판별력과 균형성을 검토하였다. 이러한 성능 지표는 DR 대상 식별을 위한 선형 분류 기준의 타당성을 정량적으로 검증하는 기준이다. 데이터 분할, 교차 검증 설정, confusion matrix 계산 등 전체 실험 파이프라인은 선형 연구의 LDA 기반 분류 절차를 준용하여 구성하였다.

LDA 기반 baseline 성능을 평가하기 위해, 본 연구에서는 시계열 전력 사용 데이터를 서로 다른 집계 수준으로 변환한 6가지 Feature Scenario를 정의하였다. 이는 동일한 타겟 변수에 대해 시계열 값이 분류 성능에 미치는 영향을 체계적으로 비교하기 위한 설계이다³⁷⁰⁾.

368) Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E., 전계논문, p3-6.

369) Dufour, A.B. (2005). Linear Discriminant Analysis — Course Notes. University Lecture Notes : Fiche TD avec le logiciel.

1) Scenario I (월별 집계)

1년 8,760시간 시계열을 월별 에너지 사용량 합계로 집계하여 12차원 특징 벡터(1~12월) 구성

2) Scenario II (주별 집계)

동일 데이터를 ISO-week 기준 주별 합계로 변환하여 총 52차원 특징 벡터 생성

3) Scenario III (일별 집계)

1일 단위 에너지 사용량 합계를 사용하여 365차원 특징을 구성

4) Scenario IV (시간별 원시 데이터)

연간 8,760시간의 원시 시계열을 그대로 사용하여 8,760차원 고해상도 특징 벡터로 구성

5) Scenario V (1월 한 달만 사용)

난방 계통 특성이 뚜렷하게 나타나는 1월 한 달의 시간별 사용량(31일×24시간=744차원)을 추출하여 난방·단열 분류에 특화된 특징을 적용

6) Scenario VI (7월 한 달만 사용)

냉방 부하가 우세한 7월 한 달의 시간별 사용량(744차원)을 사용하여 냉방·단열 특성에 민감한 분류 성능을 평가

각 Scenario에 대해 동일한 타깃 변수 집합(InsulClass, HeatingClass, YearClass, AreaClass, Location)에 대해 LDA를 학습하고 5-fold 교차 검증으로 CA 및 macro F1을 산출하였다. 선행 연구에서 제안된 Random Guess(RG)와 사전확률 기반 Random Guess(RG_PK) 기준을 함께 계산하여,

370) Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E., 전계논문, p22, p.57.

분류기의 성능이 완전 무작위 추정 또는 사전 분포만 사용하는 모형을 어느 정도 상회하는지 비교하였다.

RG의 이론적 기준 정확도는 클래스 개수를 n_{cat} 라고 할 때,

$$CA_{RG} = \frac{1}{n_{cat}}$$

으로 정의되며,

$$CA_{RG_PK} = \sum_{i=1}^{n_{cat}} p_i^2$$

으로 계산된다.

실험 결과, 단열 수준(InsulClass)과 전용면적(AreaClass)과 같이 부하 패턴과 직접 연동되는 변수들은 비교적 낮은 차원의 Feature Scenario(I~III)에서도 RG/RG_PK 기준을 정확도가 높게 나타났으며, 시간 해상도가 높아질수록 (Scenario IV) 분류 성능이 점진적으로 향상되었다. 반면, 건축연도 (YearClass)나 지역(Location)과 같이 장기 환경 특성은 Scenario IV의 8,760 차원 특징을 사용하더라도 CA가 RG_PK 기준값이 제한적으로 나타났다. 이는 순수 부하 패턴만으로는 건축물의 특성이나 지역을 완전히 복원하기 어렵다는 회귀분석 결과와 일관된 경향이다.

또한 1월 및 7월만을 사용한 Scenario V, VI 비교에서는 난방·단열 관련 타깃(InsulClass, HeatingClass)의 경우 동계 데이터만으로도 연간 시계열을 사용하는 Scenario IV에 근접한 수준의 분류 성능을 였다. 반면, 냉방 특성이 우세한 7월 데이터의 단독 사용은 난방 중심 지표의 분류 성능을 일부 저하시켰다. 반대로 냉방 에너지와 밀접한 단열 수준의 구분에서는 7월 Feature가 보다 높은 CA를 제공하는 사례도 확인되었다. 이는 계절별 대표 월을 적절히 선택할 경우, 전체 연간 시계열을 모두 사용하는 것보다 효율적인 피처 구성이 가능할 수 있음을 시사하고 있다.

4.2.1.2 DR 적용 관점에서의 선형 모형의 한계

1) 단열 수준(InsulClass)

부하 패턴과 직접적으로 연동되는 InsulClass 변수의 경우, 월별·주별·일별

집계를 사용하는 비교적 단순한 Feature Scenario에서도 RG/RG_PK 기준이 높은 정확도를 보였으며, 특히 냉·난방 에너지 사용량이 크게 구분되는 Scenario V, VI에서는 선형 경계만으로도 고·저 단열군이 명확하게 분리되는 것으로 나타났다.

2) 전용면적(AreaClass)

AreaClass의 경우 Feature Scenario가 고해상도로 갈수록 CA가 점진적으로 개선되었으나, 저·중 면적에서는 클래스 간 부하 패턴이 유사해 오분류가 집중되는 양상이 관찰되었다. 이는 4.1.2절에서 수행한 204D 기반 회귀·클래스 매핑 실험에서 Area_m2의 예측 성능이 상대적으로 높은 것과 일관되지만, LDA의 선형 경계만으로는 면적 구간 간 세밀한 비선형 차이를 모두 포착하기 어렵다는 한계를 보여준다.

3) 건축연도(YearClass)

기밀 성능(Leakage_grade) 및 지역(Location)과 같은 구조적·지리적 변수인 YearClass는 모든 Feature Scenario에서 LDA 정확도가 RG_PK 기준과 큰 차이를 보이지 못했으며, confusion matrix에서도 예측 레이블이 인접 클래스 또는 중앙 구간으로 쏠리는 경향을 확인할 수 있었다. 이는 4.1.2절의 204D 회귀 기반 클래스 매핑을 통해 확인한 바와 같이, 부하 패턴만으로 건물 외피 연식, 세부 기밀 등급, 지역 특성을 직접 추정하는 데는 구조적인 한계가 있음을 다시 한번 확인할 수 있었다.

4) posterior 확률 분포

InsulClass와 같이 에너지 특화 변수는 최대 posterior가 0.8 이상에 집중된 안정적인 분포를 보인 반면, YearClass·Location과 같이 분류가 어려운 변수는 최대 posterior가 0.4~0.6 부근에 집중되어, LDA가 클래스 경계 근처에서 불확실한 판정을 내리는 경우가 많았다. 이는 선형 판별함수가 클래스 간 분포가 비선형적으로 얽혀 있는 경우 경계를 충분히 분리하지 못하기 때문에

발생하는 현상으로 해석할 수 있다.

이상에서 살펴본 바와 같이 LDA baseline은 결과 해석이 직관적이고, Random Guess 기반 이론적 하한과의 비교가 용이하며, 각 Feature Scenario에서 어떤 타깃이 어떤 부하 패턴에 얼마나 강하게 의존하는지를 파악할 수 있다는 장점을 가진다. 그러나 선형 판별 모형의 동일 공분산, 선형 경계 가정 때문에 ① 클래스 간 경계가 비선형적으로 얽혀 있는 경우, ② 고차원 시계열 특징에서 일부 변수만 국소적으로 분리되는 경우, ③ 복수의 물리·행태 변수가 동시에 작용하는 경우에는 성능 개선에 한계가 존재하였다.

이러한 분석 결과는 LDA가 본 연구에서 baseline 분류 모형으로서는 충분한 역할을 수행하지만, 실제 VSM·RSM 데이터셋의 다양한 건물 파라미터를 정밀하게 복원하기 위해서는 4.2.1절에서 제시한 204D Feature Extractor와 결합된 LDA, Random Forest 등 비선형 분류기 및 회귀 기반·메타데이터 결합형 복합 모형이 필요함을 시사한다. 이어지는 4.2.2절 이후에서는 이러한 한계를 보완하기 위한 클래스 정의 및 204D 기반 회귀-분류 연계 설계, 그리고 비선형 분류모델 성능 평가 결과를 순차적으로 제시한다.

4.2.2 DR 의사결정에 유효한 입력정보 선정

4.2.2.1 VSM-RSM 특징 벡터 및 Feature 구성

본 절에서는 분류 모델(Classifier)의 학습을 위해 구성된 입력 특징(feature) 체계와 클래스 변수를 정의하고, 이후 모델 검증에 적용된 교차검증 전략과 평가 지표 설정을 설명한다. 4.2.1절의 VSM·RSM 기반 204D 특징 벡터의 시계열 Feature 생성 결과를 바탕으로, 본 연구에서 구축된 최종 Classifier는 VSM·RSM 데이터 간 일관성을 유지하면서도 다양한 건물 특성을 분류할 수 있도록 입력·출력 구조가 설계되었다. 구축된 VSM 기반 204D 특징 벡터는 연간 8,760시간 전력 사용 시계열을 기반으로 월(month), 주(day-of-week), 계절(season), 주중/주말(is_weekend), 시간대(hour) 등의 파생 변수를 결합하여

생성된 고차원 feature로 [표 4-4]과 같으며 동일 구조의 실측된 AMI 데이터인 2024년 RSM 데이터에 동일하게 적용되어 VSM-RSM 간 평가 체계의 정합성을 확보하였다.

[표 4-4] 204D Feature 구조

벡터(차원)	구성요소	설명
12D (M01~M12)	월별 평균 전력 사용량	12개월의 월 평균 kWh 사용량 기반 총 12차의 Feature Dimension 형성
192D	계절 × 주중/주말 × 시간	4계절 × 주중/주말 × 24h ex) Fall_weekend_h14 ; 가을-주말-오후 2시 Feature

이상과 같이 생성된 204D 특징 벡터는 단순한 통계적 집계가 아니라, 계절성·일주기성·생활패턴을 포착하는 구조적 설명력을 갖기 때문에, 이후 4.3절에서 수행되는 회귀 기반 클래스 매핑을 수행하고 204D 특징 벡터의 계절성·일주기성·생활패턴을 반영한 고차원 구조를 학습하는 XGBoost, MultiLayer Perception, Random Forest 등의 최적화된 비선형 분류기의 핵심적인 입력으로 활용된다.

4.2.2.2 주요 Class 선정

본 연구에서 분류(Classification)의 대상이 되는 타깃 변수들은 건물의 에너지 성능, 구조적 속성, 지역적 특성을 대표하는 핵심 변수들로 구성하였다. 이들은 모두 VSM-RSM 메타데이터 내에 존재하며, 3장에서 Archetype 변수로 제시된 각각의 개별 파라미터의 정의를 기반으로 [표 4-5]와 같이 구체적인 클래스 체계를 구성하였다.

VSM 생성 데이터의 파라미터별 분포로 정리된 [표 4-5]를 기반으로 주요 변수를 체계화하였으며, 건물의 난방 방식, 단열 성능, 면적 규모, 건축연도, 그리고 지역 위치의 다섯 가지 범주형 변수로 정의된다. 각 Class 별 특징은 다음과 같다.

1) HeatingClass : 난방 방식은 동절기 전력 및 열사용의 일중 패턴에 직접적인 영향을 주기 때문에, 시계열 패턴 기반 분류에서 상대적으로 높은 설명력을 제공하는 변수이다. Scenario V의 부하 곡선에서도 난방 방식별 피크 형태와 야간 부하 패턴이 구분되는 것으로 확인되었다.

[표 4-5] 주요 Class의 선정

HeatingClass	InsuClass	AreaClass	LocationClass	YearClass
CENTRAL	LOW	A_S (Small)	Seoul	Y1_~90
중앙난방	하위 단열	A_M (Medium)	Incheon	Y2_90s
DISTRICT	MID	A_L (Large)	Gyeonggi	Y3_00s
지역난방	중간 구간	A_XL	Gangwon	Y4_10s
INDIVIDUAL	HIGH	(Extra-Large)	Chungchung	Y5_20s
개별난방	우수 단열		Jeonra	
UNKNOWN			Gyeongsang	
기타				

2) InsulClass : 단열성능은 건물 외피의 열적 저항인 Overall_RSI_score 값을 기반으로 RSI 분포에 따라 하위-중위-상위 그룹으로 구성하였으며, 앞 절의 204D 회귀 실험에서 $R^2=-0.44$ 수준에 머무르고 상관계수 역시 -0.002 로 매우 낮게 나타나는 등 단열성능의 단일 변수만으로 전력 사용 패턴을 정확하게 예측이 어려운 특성을 가진다. 이러한 특성은 InsulClass가 단순한 분류 문제가 아니라, 건물 아키타입의 다른 변수들과 결합을 필요로 하는 복합적 변수임을 시사한다.

3) AreaClass 선행절의 회귀 기반 분석에서도 Area_m2는 $R^2\approx 0.51$, $r\approx 0.73$ 으로 비교적 높은 예측 성능을 보여, 204D Feature가 면적 규모의 상대적 차이를 반영할 수 있는 대표적 Class로 분류하였다.

4) YearClass는 건설교통부의 건축물 에너지 효율 규제 시기를 기준으로 다섯 개 구간으로 구분하였다. 이러한 시기 구분은 실제로 건물 단열 기준 강화, 기밀 성능 개선, 창호 성능 향상 등 건물 외피 기준의 변화를 반영한 정책적 시점과 일치한다. 그러나 건축연도는 전력 사용 패턴과의 직접적 연관성이 낮은 패턴으로 확인하였다.

5) Location은 한국 서울, 인천, 경기, 강원, 충청, 전라, 경상 등 7개 광역군으로 설정하였다. 이는 기온·일사량·HDD·CDD 등 지역 기후 특성이 전력 사용 패턴에 장기적으로 영향을 미친다는 점을 반영한 것이다. 그러나 실제 시계열 부하 패턴 분석에서는 생활양식의 지역 간 차이가 미미하고 기후의 영향이 주로 냉방·난방 부하를 통해 간접적으로 나타나기 때문에, Location은 YearClass와 함께 분류 난이도가 낮은 변수로 평가된다.

4.2.2.3 교차 검증 및 평가 지표 설정

본 연구에서 구축한 Classifier의 성능을 객관적으로 평가하기 위해, 본 절에서는 학습·검증 데이터 분할 방식과 평가 지표 설정에 대한 기준을 정하였다. 특히 KR-VSM 기반의 204D 특징 벡터와 실측 기반의 RSM 204D 특징 벡터의 보정은 모델의 과적합을 방지하고 일반화 성능을 검증할 수 있는 체계적인 평가 절차이다. 본 연구에서는 전력 사용량 계열 데이터의 특성과 클래스 분포의 불균형성을 고려하여, 5-fold 교차검증(k-fold cross-validation)을 표준 설정으로 채택하였다.

교차검증은 전체 VSM 데이터셋을 동일한 크기의 k개 subset으로 분할하고, 각 반복마다 이 중 k-1개를 학습용으로, 1개를 검증용으로 사용하여 총 k회의 학습-평가를 수행한다. 이 프로세스를 통해 개별 데이터 분할에 따른 편향을 최소화하고, 과적합 위험을 감소시키며, 전체 데이터셋을 효율적으로 활용할 수 있다는 장점이 있다³⁷¹⁾³⁷²⁾.

371) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.56-57.

372) Neale, A., Kummert, M., & Bernier, M., (2019), 전계논문, p.3394-3395.

선행 연구인 QSBSEM에서는 약 40,000건의 VSM 데이터를 대상으로 Location, Heating 등 변수를 LDA 분류 실험을 수행하고 5-fold 교차검증을 적용하여 평균 정확도를 산출한 바 있다³⁷³⁾. 특히 Fisher LDA의 일반화 오차는 학습 표본 수 대비 feature 비율에 의해 비선형적으로 변동하며, 이 과정에서 Scissors Effect라 불리는 위험 곡선이 나타날 수 있음을 이론적으로 제시하였다³⁷⁴⁾. 이는 학습 표본 수가 증가할 때 일반화 오차가 처음에는 감소하다가, 특정 지점을 기점으로 다시 증가한 후, 충분한 표본 수가 확보되면 다시 감소하는 비선형적 패턴이 관측될 수 있다는 것이다. 이러한 효과는 특히 클래스 수가 많거나, 클래스 간 평균 벡터 차이가 작은 경우 더욱 두드러지게 나타나며³⁷⁵⁾, LDA가 가정하는 클래스 평균 중심선이 약화될수록 심화되었다.

본 연구에서는 QSBSEM의 설정과 동일하게 k=5로 교차 검증(Stratified k-fold)을 수행하였으며, 데이터의 계절적 변동성 및 부하 패턴의 계층적 특성을 반영하기 위해 Stratified k-fold 방식을 적용하였다.

1) 정확도(Accuracy)

Classifier의 가장 기본적인 성능 지표는 정확도(Accuracy)로, 전체 샘플 가운데 실제 정답을 맞춘 비율을 나타낸다. 정확도는 다음과 같이 정의된다³⁷⁶⁾.

$$Accuracy = \frac{N_{correct}}{N},$$

여기서,

N : 전체 검증 샘플수,

$N_{correct}$: 정답으로 분류된 샘플 수,

Accuracy는 직관적이며 모델의 전반적인 분류 성능을 제공하지만, 클래스 비율에 불균형이 존재하는 경우에는 특정 클래스에 편향될 수 있다는 한계를 가진다. 특히 본 연구에서는 HeatingClass와 Location, YearClass 등 특정 클

373) Neale, A., Kummert, M., & Bernier, M., (2021), 전계논문, p.148, p.217.

374) Raudys, S., & Duin, R. P. W., 전계논문, p.386.

375) Raudys, S., & Duin, R. P. W., 전계논문, p.387, p.390-391.

376) Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification (2nd ed.). Wiley-Interscience.

래스의 샘플 수가 상대적으로 적거나 분포가 왜곡되어 있기 때문에, 단순 정확도만으로는 충분한 성능 평가가 어렵다.

2) Macro F1-score

Macro F1-score는 모든 클래스의 정밀도(Precision)와 재현율(Recall)을 동일한 가중치로 평균하여 산출하므로, 소수 클래스의 성능을 보다 공정하게 평가할 수 있는 지표로 F1-score는 다음 수식과 같이 정의된다³⁷⁷⁾.

$$Precision_k = \frac{TP_k}{TP_k + FP_k} ; k\text{로 예측한 것 중 실제 } k\text{인 비율,}$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k} ; \text{실제 } k \text{ 중에서 맞게 찾아낸 비율,}$$

$$F1_{macro} = \frac{1}{K} \sum_{k=1}^k F1_k,$$

$$F1_k = \frac{2PR}{P+R},$$

여기서,

K : 클래스 수,

Precision_k : k 번째 클래스 정밀도,

Recall_k : k 번째 클래스 재현율,

Macro F1-score는 클래스 간 분포가 비대칭적인 경우에도 안정적인 성능 비교를 가능하게 해주며, 특히 YearClass·Location과 같이 구조적으로 분류가 어려운 변수의 성능을 평가하는 데 중요한 지표로 활용되었다.

3) 혼동 행렬

정확도와 F1-score가 전체적인 성능 수준을 정량적으로 요약해주는 반면, 혼동 행렬(confusion matrix)은 각 클래스 간 오분류 패턴을 시각적으로 나타냄으로써 모델의 약점을 해석하는 데 중요한 역할을 한다³⁷⁸⁾. 예를 들어,

377) Manning, C. D., Raghavan, P., & Schütze, H. (2009). Introduction to information retrieval. Cambridge University Press.

YearClass는 건축연도 구간이 인접 시기로 구성되어 있기 때문에 선형·비선형 모델 모두에서 인접 클래스 간 오분류가 빈번하게 나타났다.

4) Random Guess(RG) 및 Prior-based Random Guess(RG_PK)

LDA 기반 baseline을 해석할 때, 모델 성능이 무작위(Random Guess) 수준과 비교하여 얼마나 개선되었는지를 평가하기 위해 RG(Random Guess)와 RG_PK(Prior-based Random Guess) 지표를 함께 산출하였다. 이는 LDA가 선형 경계를 기반으로 분류한다는 특성상, 복잡한 비선형 분류 문제에서는 RG 수준을 크게 넘어서지 못하는 경향이 있을 수 있기 때문이다³⁷⁹⁾.

여기서 무작위 방정식은,

$RG=1/n_{cat}$ 이며,

$$RGPK = \sum_{k=1}^{n_{cat}} P_k^2 \text{ 이다.}$$

여기서,

P_k^2 : k번째 class 사전 확률,

Classifier의 객관적이고 재현 가능한 성능 평가를 위해 5-fold 교차 검증을 기반으로 Accuracy, Macro F1-score, Confusion Matrix, RG/RG_PK를 포함한 다층적 지표 체계를 구성하였다. 이러한 평가 지표들은 VSM/RSM 기반 204D 특징 벡터를 입력으로 하는 다양한 분류 모델(LDA, RF, XGB, MLP)의 성능 차이를 구조적으로 해석할 수 있는 기반을 마련해준다. 또한 특히 YearClass와 Location과 같이 본질적으로 부하 패턴 기반 분류가 어려운 변수의 한계를 명확히 드러내는 데 중요한 역할을 수행하였다.

5) 성능 평형점 적용

LDA는 각 클래스의 공분산 행렬을 기반으로 선형 판별 함수를 구성하므로, 공분산 행렬이 non-singular 해야 모델이 안정적으로 추정된다. 이를 위

378) 이희원, 김덕환. (2023). LSTM을 활용한 일반 주택의 실내 온도 예측 사례 연구. 『한국차세대컴퓨팅학회 논문지』, 19(2), 24-34.

379) Neale, A., (2021), 전계논문, p.96.

해서는 학습 표본 수(a , number of samples)가 입력 피쳐 수(f , number of features)보다 많아야 하며, 일반적으로 $a/f > 1$ 의 조건이 최소 요건으로 제시된다³⁸⁰).

즉, LDA는 클래스별 공분산을 공통(pooling)하여 역행렬을 계산하기 때문에, 표본 수가 피쳐 수보다 적은 경우($a \leq f$) 공분산의 랭크(rank)가 부족해 역행렬이 존재하지 않거나 수치적으로 불안정하게 된다. 이때 훈련 데이터에서는 높은 정확도를 보이더라도, 검증 데이터에서는 급격한 성능 저하가 나타나는 과적합(overfitting) 문제가 발생한다³⁸¹.

반대로, 표본 수가 피쳐 수보다 충분히 많을수록($a \gg f$) 공분산 추정이 안정화되고 클래스 간 분산 대비 클래스 내 분산의 추정오차가 감소하여, 분류 경계의 일반화 성능이 향상된다. 따라서 경험적으로는 a/f 의 비율이 1을 넘는 수준보다는 5~10배 이상일 때 판별 계수의 분산이 최소화되고, 모델의 재현성이 확보되는 것으로 보고된다³⁸².

선행 연구인 캐나다의 QSFBSM의 VSM 분류 실험에서도 35,040개의 고해상도 피쳐를 가진 200,000가구의 데이터셋에서, 표본 수 대비 피쳐 수의 비율이 낮을수록 공분산 추정이 불안정해지고 LDA 분류 정확도가 하락하는 현상을 보고하였다. 이에 따라 시간 해상도를 1시간 또는 1일 단위로 집계하여 피쳐 수를 줄였을 때, 분류 정확도는 소폭 감소했지만 연산 안정성과 일반화 성능은 오히려 향상되었다고 명시하고 이를 적정 피쳐 선정에 따른 분류 성능 평형점 개념(optimal balance point between feature dimensionality and classification performance)으로 정의하였다³⁸³.

이는 LDA가 고차원($f \uparrow$)·저표본($a \downarrow$) 조건에서 수학적으로 불안정하다는 기존 통계학적 논의와 일치한다³⁸⁴.

요약하면, $a/f > 1$ 의 조건은 LDA의 수학적 성립을 위한 최소한의 표본-피쳐 비율이며, 이는 공분산 행렬의 역행렬이 존재하고 분류 경계가 통계적

380) Neale, A., Kummert, M., & Bernier, M., (2020), 전계논문, p.589-590.

381) Izenman, A. J. (2008). Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. New York: Springer.

382) Raudys, S., & Duin, R. P. W., 전계논문, p.386, 389-390.

383) Neale, A., Kummert, M., & Bernier, M., (2018), 전계논문, p.73-74.

384) Raudys, S., & Duin, R. P. W., 전계논문, p.386, 389-390.

으로 안정적으로 추정될 수 있는 최소 표본 크기 조건을 의미한다.

본 연구에서는 시간 해상도를 1시간·1일로 하고 평균·최대·분산의 집계 수준을 조정하여 피쳐 수(f)를 제어하고, 동시에 충분한 표본 수(a)를 확보함으로써 LDA의 안정성과 일반화 성능 간 균형을 달성하도록 실험 설계를 수행하였다. 따라서 본 연구는 한국형 VSM 데이터셋의 200,000 가구에서 표본 20,000 가구를 기반으로 다중 해상도-다중 집계 실험 구조를 설계하였다. 이는 QSFBSM의 모델에서 제시한 적정 피쳐 선정에 따른 분류 성능 평형점 개념을 국내 기후 및 주거 특성에 맞게 확장하여 적용한 것이다.

4.2.3 DR 대상 선별을 위한 비선형 분류 확장 및 평가

4.2.3.1 회귀 기반 및 비선형 분류모델 설계

Archetype 변수는 LDA 분류 시 STEP I과 STEP II로 분류하여 진행하였으며 변수가 가지고 있는 값의 특징의 차이로 단계를 나누어 분류를 [표 4-6]과 같이 수행하였다.

[표 4-6] 변수 간 Classification의 속성 분류

구분	Step I-Location	Step II-아키타입 변수
변수 속성	지리적 좌표(lat/lon) 기반	범주형(category) 기반
출력 형태	연속값(continuous)	이산값(discrete)
모델링 방식	회귀(Regressor)	분류(Classifier)
지역 간 관계	연속적 거리(distance) 존재	클래스 간 거리 없음
특징적 난이도	전력 패턴-> 좌표 회귀	전력 패턴-> 건물/단열의 구조 원인 구분
LDA와의 부적합성	LDA는 좌표의 연속성 반영 불가	LDA는 클래스 불균형에 취약
Random Forest 적합성	연속 좌표 회귀에 강함	비선형 다중 클래스 분류 가능

이중 범주형 변수로 분류되는 Location 변수는 QSFBSSEM의 분류 기반으로 정의했지만, LDA 분류에서 효과적으로 적용되지 않았으며 이는 국내 거주 특성상 좁은 면적에 도시를 중심으로 높은 인구밀도가 공동주택 중심으로 구성되어 있는 특성 때문에 개별거주 공동주택 간 전력 패턴 차이가 유사한 특성으로 나타나는 것으로 파악할 수 있다. 이러한 문제점을 해결하기 위하여 Location 자체가 위도/경도 정보를 포함하고 있는 것을 고려하여 연속형 공간 처리 변수를 적용하여 LDA의 분류 오류를 보완하고자 하였다.

Location에서 적용되는 개별 시뮬레이션 오차는 위도와 경도 차이를 적용하여 산정하였으며, 다음과 같은 방정식으로 적용하였다.

$$d_i = \sqrt{(lat_i - \widehat{lat}_i)^2 + (lon_i - \widehat{lon}_i)^2},$$

여기서,

d_i : 시뮬레이션 i에서 위도-경도 예측 오차 크기,

lat_i : 실제 VSM에 적용된 위도,

\widehat{lat}_i : 회귀모델로 예측된 위도,

lon_i : 실제 VSM에 적용된 경도,

\widehat{lon}_i : 회귀모델로 예측된 경도,

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2},$$

N : 테스트에 포함된 총 시뮬레이션 수(N=6000 ; 20,000 × 30%),

이와 달리, Area_class, BuildingType, Overall_RSI_category, Leakage_grade의 Archetype 변수는 값 사이에 거래 개념이 존재하지 않는 명확한 범주형 이산 변수로 전력 부하 패턴 특징을 이용한 비선형 결정경계, 고차원 피쳐 공간, 다중 클래스 불균형 문제에 대해 상대적으로 안정적인 성능을 제공하는 대표적인 트리 기반 앙상블 학습 알고리즘으로 Random Forest Classifier를 적용하여 수행하였다³⁸⁵⁾.

385) Breiman, L. (2001). Random Forests. Machine Learning Technical Report. Machine Learning, 45, 5-32, 2001.

Random Forest(Classifier)는 동일한 학습 자료로부터 여러 개의 결정 트리(Decision Trees)를 독립적으로 구성한 뒤, 각 트리의 예측 결과를 다수결 투표(majority voting)로 결합하여 최종 분류를 수행하는 bootstrap aggregating 기반 앙상블 분류기로 원본 데이터에서 다음과 같은 방정식으로 서로 다른 학습 세트의 생성이 가능하며,

$$D = (x, y_i)_{i=1}^N,$$

여기서,

D : 학습 데이터,

각 노드의 분할을 통하여 임의 피쳐만을 고려하여 트리 간 상관관계를 줄이고 다변량 신호의 탐색을 강화하는 특징이 있다.

$$\Delta G(t) = G(t) - \left(\frac{N_L}{N} G(t_L) - \frac{N_R}{N} G(t_R) \right) \text{386),}$$

여기서,

t_L, t_R : 노드 분할로 생성된 Left, Right 하위 노드,

$$G(t) = 1 - \sum_{k=1}^k P_k^2 \text{387),}$$

여기서,

P_k : k 번째 클래스 비율, $G(t)$: 노드 불순도,

LDA의 선형 분리 가정이 KR-VSM 전력 패턴의 실제 구조를 적절히 반영하지 못한다는 점을 확인함에 따라, 대안적 분류 방식으로 Random Forest Classifier를 도입하였다. Random Forest는 트리 기반 비선형 학습 구조와 앙상블 결합 방식을 통해 시계열 전력 데이터의 고차원적·비정형적 특성을 효과적으로 포착할 수 있는 모델로 널리 검증되어 왔다³⁸⁸⁾. 본 연구의 데이터 구조와 분류 문제의 특성을 고려할 때 비선형 결정경계가 학습에 효과적이며³⁸⁹⁾, 고차원 시계열 피쳐에 대한 강건한 성능을 제공하고, 클래스 불균형

386) Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

387) Hastie, T., Tibshirani, R., & Friedman, J., p.308-310.

388) Breiman, L., 전계논문, p.5-6.

389) Hastie, T., Tibshirani, R., & Friedman, J., p.305, p.597.

환경에서도 비교적 안정적인 성능을 유지³⁹⁰⁾한다는 장점을 고려하여 Random Forest를 적용하는 것이 타당하다고 판단하였다.

4.2.3.2 변수별 분류 모델 실시

1) 회귀 기반의 Location 변수 분류

LDA가 Location 분류에서 경기권(GYEONGGI)과 경상권(GYEONGSANG)으로 일괄 분류되는 편향 문제를 보임에 따라, 지역(Location)을 범주형 분류(target classification)가 아닌 연속형 회귀(target regression) 문제로 재정의하여 각 지역을 대표하는 위도(lat)-경도(lon) 좌표를 타겟으로 설정하고, 월간 전력 사용량 특징의 Scenario I 기반 Random Forest Regressor를 학습하였다. 학습의 결과 회귀 기반 Location 추정 결과 데이터셋에서의 RMSE는 약 0.8492° 로 나타났다. 이는 위도 및 경도의 거리에 대한 오차 범위로 위도/경도 $1^{\circ} \approx 110$ km인 것을 감안하면 7개 광역군의 대부분을 독립적으로 커버가 가능하며 LDA에 비하여 현저한 특성 분류가 가능하였다. 여기에서 Random Forest Regression은 회귀 트리 기반(Classification and Regression Tree) 학습으로 bootstrap 샘플을 train(70%), test(30%)로 데이터를 분할하였다. $n_estimators=200$ 개의 각 트리는 bootstrap된 14,000개의 샘플을 적용하고 회귀값으로 MSE를 최소화하도록 학습하였다.

2) 비선형 분류

본 연구에서는 LDA 기반 분류가 대부분의 아키타입 변수에서 극단적 다수 클래스 편향을 보임에 따라, 대안적 분류모델로서 비선형 결정경계 학습이 가능한 Random Forest를 적용하였다. Random Forest는 트리 기반 앙상블 구조를 통해 시계열 전력 데이터에 내재된 비정형 패턴을 보다 유연하게 포착할 수 있으며, 특히 LDA에서 TP=0으로 분류가 안된 BuildingType, Overall_RSI_category 변수에 대해 모든 클래스에서 수치 데이터의 Recall을 확

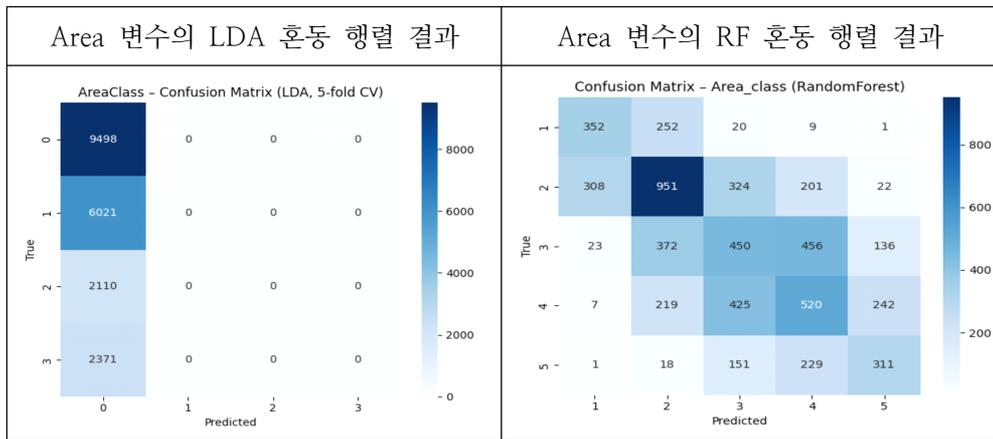
390) Chen, C., Liaw, A., & Breiman, L. (n.d.). Using Random Forest to Learn Imbalanced Data., Department of Statistics, UC Berkeley.

보하여 분류기의 기능적 개선을 확인하였다.

(가) Area_class 분류

[표 4-7]의 Area_class의 분류 결과 Random Forest(RF)는 Area_class 분류에서 Accuracy = 0.4307을 기록하며 면적 기반 5구간으로 분류는 되지만 안정적이지는 못하였고 특히 40m² 이하의 낮은 평형 클래스1에서 Recall=0.74로 높은 성능을 보였는데, 이는 주거 면적이 전력 소비량의 직접적인 영향을 미침으로 시계열 부하 패턴의 구조적인 차이가 반영된 결과로 해석된다.

[표 4-7] Area 변수의 분류 성능 비교



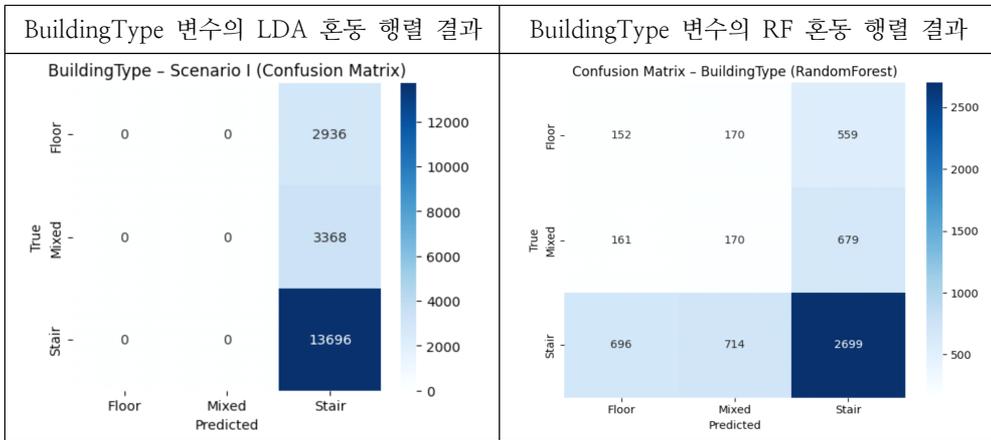
혼동 행렬에서도 모든 클래스가 최소 500건 이상의 TP를 확보하여 LDA의 완전 붕괴와는 달리 정상적인 다중 클래스 분류기로 작동함을 확인하였으며, LDA의 실행 결과와 다르게 Confusion Matrix가 전반적으로 개선되었음을 확인할 수 있다.

(나) BuildingType(Stair / Floor / Mixed)

[표 4-8]의 분류 결과 BuildingType은 LDA에서 전 샘플이 Stair 클래스로 단일 예측되어 분류의 오류 현상을 보였으나, RF는 Floor 및 Mixed 클래스에서 각각 17% 수준의 Recall을 확보하였다.

전체 정확도는 0.5035로 높지 않으나, 이는 건물 구조적 차이보다 입주민의 생활 패턴에 의한 부하 변동성이 더 크게 작용하기 때문으로 보인다.

[표 4-8] BuildingType 변수의 분류 성능

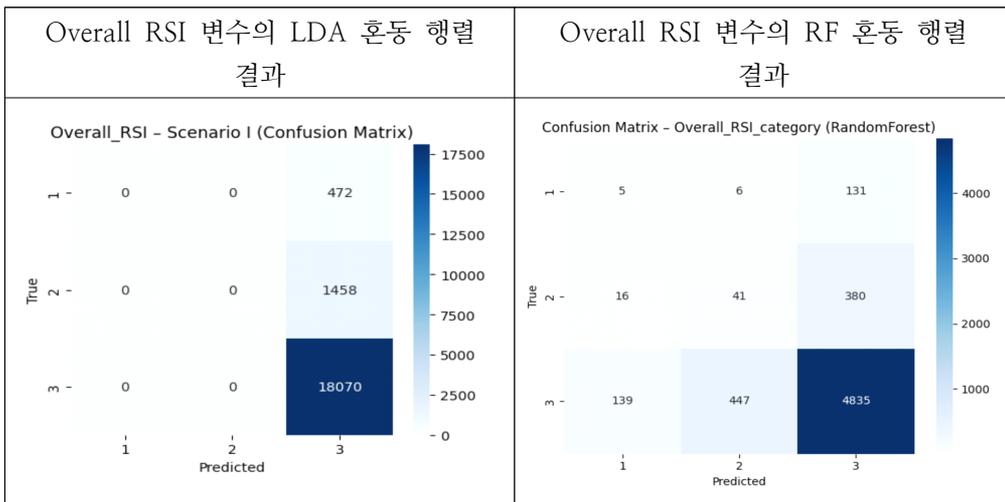


즉, 동일한 Stair 형 건물이라도 거주 인원, 가전 패턴, 냉방 가동시간 등에 따른 편차가 커 전력 패턴만으로 건물 형태를 명확히 식별하기 어려운 특성이 존재한다.

이는 LDA 대비 분류 성능이 부분적으로 개선되었음을 확인할 수 있지만, 동시에 BuildingType은 전력 패턴만으로는 한계가 존재함을 시사한다.

(다) Overall_RSI_category

[표 4-9] Overall_RSI 변수의 분류 성능



[표 4-9]의 단열 성능 Overall_RSI_category(1/2/3)의 Random Forest

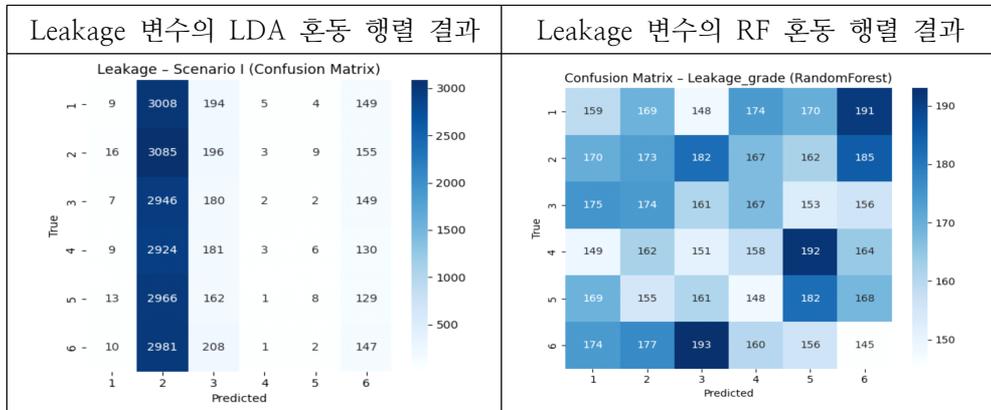
Accuracy는 0.8135로 비교적 높게 나타났다. 특히 다수 클래스를 이루는 RSI=3(고단열) 이외에도 RSI=1, RSI=2 클래스에서 각각 Recall 0.04, 0.09로 명백한 TP가 발생하여 LDA 대비 기능적 개선이 두드러졌다.

단일 난방/냉방 부하가 단열 등급과 강하게 연관된다는 점을 고려할 때, 주거 단열성능은 시간별 부하 패턴에 상대적으로 잘 반영되는 변수임을 확인할 수 있다. LDA 대비 분류 성능이 부분적으로 개선되었음을 확인할 수 있지만, 동시에 BuildingType은 전력 패턴만으로는 한계가 존재함을 확인하였다.

(라) Leakage_grade(기밀성 등급) 분류

[표 4-10]의 기밀성(Leakage grade)은 Random Forest에서도 Accuracy=0.1630으로, 6개 클래스 랜덤 추정(≈ 0.167)과 사실상 동일한 성능을 보였다.

[표 4-10] Leakage 변수의 분류 성능



Random Forest는 LDA 대비 분류 성능이 부분적으로 개선되었음을 확인할 수 있지만, 동시에 BuildingType은 전력 패턴만으로는 한계가 존재함을 확인하였다. 이러한 결과는 모델의 한계라기보다는 데이터 정보량의 부족에 기인한다. 전력 사용 패턴(kWh)은 주로 가전·난방·냉방 변수로 결정되므로 공기 누설량(ACH), 창호 성능, 외피 열관류율 등 기밀성과 직결된 물리적 특성은 현재의 시계열 전력 데이터로는 사실상 식별 불가능하다.

따라서 향후 Leakage 예측을 위해서는 실제 누설량, 창호 사양, 환기설비

등 건축물의 물리 정보가 포함된 추가 피처의 도입이 필요할 수 있음을 예시하고 있다.

3) LDA와 RF 비선형 분류의 비교

VSM으로부터 산출된 월별 전력 사용 패턴을 입력 변수로 하여, 주거 면적 구간(Area_class), 건물유형(BuildingType), 단열 등급(Overall_RSI), 기밀성 등급(Leakage_grade)의 네 가지 건물 물리·형태 특성을 Random Forest 분류기로 추정하고 이를 위해 각 타깃 변수별 혼동 행렬(confusion matrix)을 산출하였으며, 모델이 각 범주를 어떻게 식별하는지를 정량적으로 분석하였다.

우선, 주거 면적 구간(Area_class) 분류에서는 소형(S) 및 중형(M) 구간에서 비교적 안정적인 성능을 보였다. 특히 소형 주택의 경우 상관된 부하 패턴이 뚜렷하여 높은 정분류를 나타냈으며, 중형 구간 역시 상당수가 중형(M)으로 정확히 분류되었다. 반면 대형(L) 구간에서는 일부 샘플이 중형으로 예측되는 경향이 나타났는데, 이는 중·대형 면적대 주택이 난방·냉방 부하 구성에서 유사한 전력 패턴을 보이기 때문이다. 전반적으로 Random Forest는 면적대별 전력 패턴의 차이를 일정 수준까지 효과적으로 포착한 것으로 판단된다.

건물유형(BuildingType) 분류에서는 계단식은 건물 종횡비, 열손실 구조, 난방 부하 패턴 등에서 독립적인 특성을 가지므로 모델이 해당 클래스를 상대적으로 명확히 분리가 가능한 것으로 보인다. 반면 복도식과 혼합식은 서로 교차하여 오분류 되는 특성 비율이 높게 나타났다. 이는 두 유형이 전력 사용 패턴 또한 선형적으로 구분되기 어려운 영역에 위치함을 의미한다. 결과적으로 BuildingType은 특정 클래스에서는 높은 성능을 보였으나, 다른 일부 클래스에서는 클래스 간 경계가 모호하여 분류 난이도가 상대적으로 높았다.

단열 등급(Overall_RSI_category) 분류는 클래스 불균형이 큰 변수로, Random Forest는 고단열(3) 등급에서 높은 정확도를 보였다. 고단열 주택은 연중 전력 소비 패턴이 비교적 일관되므로 모델이 해당 특성을 안정적으로 학습한 것으로 해석된다. 반면 저단열(1) 및 중단열(2) 등급은 상당 비율이 고단열(3)로 전이되어 분류되었다. 이는 전력 기반 데이터만으로 세부 단열

성능을 식별하는 데 본질적 한계가 있으며, 클래스 불균형 또한 오분류를 강화시키는 구조적 요인이 작용한 것이다. 그럼에도 Random Forest는 LDA보다 저·중단열 클래스에서 일정 수준의 TP를 확보하여 다중 클래스 분류기로 유효성을 확인할 수 있었다.

기밀성 등급(Leakage_grade) 분류는 네 가지 변수 중 가장 난도가 높았다. 누설 등급은 1~6의 다중 클래스 구조를 갖고 있으나, 등급 간 전력 부하 패턴 차이가 뚜렷하지 않기 때문에 혼동 행렬에서도 전반적으로 오분류가 확인되었다. 일부 등급(3·4·5)은 특정 예측 결과에 다소 집중되는 경향을 보여 모델이 특정 기밀성 범주의 경향성을 부분적으로 분류가능 하지만, 전체적으로는 분류 정확도가 높지 않았다. 이는 외피 기밀성 자체가 전력 부하 패턴과 직접적으로 연결되지 않는 물리적 특성이기 때문이다.

여기에서 Random Forest는 Area_class, Overall_RSI와 같이 전력 패턴과의 연관성이 상대적으로 강한 변수에서는 높은 분류 성능을 보이며, BuildingType, Leakage_grade와 같이 열적·구조적 요인이 복합적으로 작용하거나 전력 사용량에 대한 신호가 약한 변수에서는 상대적으로 낮은 성능을 보였다. 이는 KR-VSM의 머신러닝 분류가 건물유형 및 외피 성능 정보를 일정 수준까지 판별할 수 있으나, 전력 데이터만으로 모든 물리적 특성을 완전하게 추정하는 데에는 구조적 한계가 존재함을 시사한다.

4.2.4 분류정확도 평가

4.2.4.1 k-fold 교차검증

본 연구에서는 QSFBSSEM의 설정과 동일하게 $k=5$ 로 교차검증(Stratified k-fold)을 수행하였으며, 데이터의 계절적 변동성 및 부하 패턴의 계층적 특성을 반영하기 위해 Stratified k-fold 방식을 적용하였다. 이는 각 fold에 계절의 에너지 민감도(CDD·HDD), 실내 부하 패턴, Equip, Load 등이 균등하게 분포되도록 하기 위한 조치이다. 이러한 계층형 교차 검증은 클래스 불균

형(class imbalance) 조건에서 모델이 특정 계층이나 계열에 과도한 적합을 방지하는 데 효과적이며, LDA가 가정하는 클래스 간 평균 중심성(class-mean separability)을 유지하는 데에도 유리하다³⁹¹⁾.

[표 4-11] k-fold 분류기 성능

Fold	Accuracy	Precision(macro)	Recall(macro)	F1-score(macro)
1	0.0715	0.00341	0.00996	0.00469
2	0.0750	0.00330	0.01111	0.00479
3	0.0785	0.00421	0.01271	0.00590
4	0.0730	0.00310	0.00921	0.00436
5	0.0755	0.00456	0.01679	0.00648
mean	0.075	0.004	0.012	0.005
std	0.003	0.001	0.003	0.001

이상의 과정은 구축된 KR-VSM 20,000건의 232개 클래스의 Archetype을 대상으로 LDA + Stratified 5-fold 검증을 수행하였으며, k-Fold 교차 검증을 수행한 결과 [표 4-11]과 같이, 분류기의 전반적인 성능에서 정확도(Accuracy)는 0.071~0.079 범위로, 평균 0.075 ± 0.003 으로, Precision 평균 0.004, Recall 0.012, F1-score(macro) 평균 0.005로 나타났다. 이는 전체 클래스 수가 많거나 데이터가 심하게 불균형한 경우에 일반적으로 나타나는 낮은 정확도 수준과 일치한다. Fold 3의 Recall이 Precision 보다 $0.012 > 0.004$ 로 3배 더 높은 경향을 보였으며, 이는 분류기가 다수 클래스에 대하여 폭넓게 예측을 시도하지만 특정 클래스에 대한 일관된 정밀도를 확보하지 못할 수 있음을 시사하고 있으며, 또한 fold 간 지표의 표준편차가 매우 작은 것은 단순한 데이터 분할의 우연성이 아니라 모델 구조와 데이터 특성에 내재된 한계에서 비롯됨을 뒷받침하고 있다³⁹²⁾.

391) Balakrishnama, S., & Ganapathiraju, A. LINEAR DISCRIMINANT ANALYSIS-A BRIEF TUTORIAL. Institute for Signal and Information Processing. Mississippi State University.

392) Raudys, S., & Duin, R. P. W., 전계논문, p.385.

이어서 수행한 k-fold 교차 검증은, ① KR-VSM 20,000건-232개 다중 클래스-계절 기반 피쳐 조합에서 LDA의 실제 일반화 능력 한계를 계량적으로 확인하였고, ② 4.2.3절에서 보고한 LDA 기반 분류 실패와 그 원인을 이론·실험적으로 재차 검증하였으며, ③ Random Forest 기반 회귀·비선형 분류 모델이 필요한 이유, 즉 선형 분류기로는 포착하기 어려운 비선형·고차원 구조를 반영해야 한다는 필요성을 뒷받침하는 역할을 한다. 향후 RSM-VSM 비교 실험에서도 이 LDA 교차 검증 결과는 통계적 기준선(baseline)과 신뢰 구간을 제공하여, VPP 운영자가 실제 적용 시 고려해야 할 분류 한계 및 추가 데이터 확보 필요량을 평가하는 데 활용될 수 있다.

4.2.4.2 변수별·특징별 정확도 분석

앞선 4.2.3절에서는 LDA 기반 분류기를 이용해 KR-VSM 데이터의 주요 타깃 변수인 Location과 Archetype 변수에 대한 분류를 수행하였다. 그 결과, 232개 다중 클래스로 정의된 Archetype과 권역 기반 Location 변수의 경우, 5-fold 교차 검증에서 Accuracy와 F1-score가 매우 낮은 수준으로 나타났다. 이는 LDA 단일 선형 축만으로 복잡한 다중 클래스 구조를 충분히 분리하기 어렵다는 한계를 확인할 수 있었다. 이와 같은 결과는 표본 수 대비 클래스 수가 크고, 클래스 간 평균 벡터 차이가 작을 때 Fisher LDA의 일반화의 특성으로 인하여 비선형적으로 오차가 증가한다는 선형 연구의 Scissors Effect 이론과도 일치한다³⁹³⁾.

이러한 한계를 보완하기 위해 4.2.3절에서는 동일한 KR-VSM 메타데이터를 대상으로 회귀 기반과 비선형 분류 모델인 Random Forest를 설계하였다. Random Forest는 winter/summer proxy per_m2, 기상-부하 지표, 건물 물리 변수 등을 입력 특징으로 사용하는 모델을 구축하였다. Random Forest가 분산·비선형성을 직접적으로 모델링할 수 있어, LDA가 실패한 변수인 Location·Archetype에서도 일정 정도 의미 있는 수준의 예측 성능을 확보하였으며, 이는 변수 중요도를 통해 난방·단열·면적·연식 등 물리적 요인이 주요

393) Raudys, S., & Duin, R. P. W., 전계논문, p.386.

패턴에 미치는 상대적 영향을 해석할 수 있다는 장점이 있다. 결국 LDA는 linear baseline으로 분류 난이도를 정량화하는 역할을 수행하고 Random Forest는 이후 DR 수요예측 및 RSM-VSM 비교 분석에서 사용할 실질적인 비선형 예측 모델의 역할을 각각 담당하도록 역할을 재정의하였다.

이러한 맥락에서 본 절에서는 LDA 분류 모델이 어떤 입력 특징 조합과 타깃 변수 수준에서 어느 정도까지 유효한 분리 능력을 갖는지를 정리하고, 이를 4.2.3절의 Random Forest 실험과 연결하여 해석한다. 우선 전체 200,000 가구 중 표본으로 추출한 20,000 가구 데이터를 대상으로, 겨울·여름 시즌 부하를 나타내는 두 개의 proxy 특징인 winter_p_m2, summer_p_m2를 입력으로 하고 232개 Archetype을 타깃으로 설정한 결과, [표 4-12] 에서와 같이 5-fold 교차 검증 평균 정확도는 0.071~0.079 범위, 평균 0.075 ± 0.003 에 불과했으며, F1-macro 역시 0.005 ± 0.001 수준으로 매우 낮게 나타났다.

[표 4-12] k-fold 검증 결과

항목	평균값	표준편차
Accuracy	0.075	± 0.003
Precision	0.004	± 0.001
Recall	0.012	± 0.003
F1 macro	0.005	± 0.001

이는 232개 중 상당수 Archetype 클래스의 표본 수가 수십 건에 불과한 희귀 클래스인 반면, LDA는 각 클래스의 공분산을 공통으로 추정해야 하기 때문에, 소수 클래스들에서의 분산 추정이 구조적으로 불안정해지는 상황을 반영한다. 그러나 동일한 두 개의 입력 특징을 유지한 상태에서 타깃 변수를 상위 개념으로 단순화하면 성능은 뚜렷한 개선을 보였다.

[표 4-13]는 타깃 클래스를 난방 방식(HeatingClass), 단열 등급(InsulClass), 건축연도(YearClass), 면적 구간(AreaClass) 등으로 바꾸어가며 winter+summer_per_m2 조합을 입력으로 사용한 LDA 분류 결과를 요약한 것이다. HeatingClass(4개

클래스)의 경우 평균 정확도는 0.606 ± 0.000 , F1-macro는 0.430 ± 0.001 로, 겨울 난방 부하와 여름 냉방 부하라는 단순한 특성이 비교적 잘 설명한다는 점을 확인할 수 있었다.

[표 4-13] 조합에 따른 LDA 분류 정확도 요약

실험 조건	타겟 클래스	클래스	특징 조합	Accuracy (\pm std)	F1-macro (\pm std)
전체 Archetype	Archetype	232	winter+summer	0.075 ± 0.003	0.005 ± 0.001
난방 방식	HeatingClass	4	winter+summer	0.606 ± 0.000	0.430 ± 0.001
단열 등급	InsulClass	3	winter+summer	0.903 ± 0.000	0.316 ± 0.000
건축연도	YearClass	5	winter+summer	0.301 ± 0.010	0.184 ± 0.007
면적 구간	AreaClass	4	winter+summer	0.475 ± 0.000	0.161 ± 0.000
상위 10 Archetype	Archetype (top10)	10	winter+summer	0.165 ± 0.007	0.068 ± 0.003

단열 등급인 InsulClass에 대해서 Accuracy가 0.903, F1-macro가 0.316으로, winter·summer 부하량이 건물 외피 성능과는 밀접한 선형 관계를 보여주고 있다. 실제로 InsulClass를 대상으로 winter_p_m2만 사용한 경우, summer_p_m2만 사용한 경우, winter+summer 조합을 사용한 경우의 각각의 Accuracy가 동일하게 0.903으로 나타나, 어느 한 계절의 부하 정보만으로도 단열 수준을 충분히 분리할 수 있음을 시사한다.

반대로 YearClass는 정확도 0.301 ± 0.010 , F1-macro 0.184 ± 0.007 에 그쳤으며, AreaClass는 정확도 0.475 ± 0.000 , F1-macro 0.161 ± 0.000 으로, 난방·단열에 비해 분리도가 낮았다. 이는 건축연도가 단열, 기밀성, 설비 효율 등 여러 요인의 복합 결과로 나타나면서 단순하게 겨울·여름의 냉·난방 부하를 기준으로 선형적으로 분리되기 어렵다는 점, 그리고 winter/summer를 면적으로 나눈 p_m2 특징이 면적의 절대 크기 차이를 상당 부분 상쇄시킨 결과로 해석할 수 있어서 이는 YearClass와 AreaClass는 LDA의 선형 분리에 덜 적합한 변수로 나타났다.

또한 전체 232개 Archetype 중 표본 수가 많은 상위 10개 클래스로 총 8,065건 만을 추출하여 잠재적 상한선(upper bound)을 평가한 결과, 평균 정확도는 0.165 ± 0.007 , F1-macro는 0.068 ± 0.003 으로 다소 개선되었으나, 여전히 실무 적용을 위한 수준의 분류 성능에는 크게 못 미친다. 즉, winter·summer 두 개의 계절 변수만으로는 난방 방식·단열·연식·면적이 동시에 반영된 고차원 Archetype 정보를 충분히 포착하기 어렵고, 선형 판별에 기반한 LDA로는 구조적인 한계를 갖고 있음을 보여준다.

4.3 실측 기반 수요 반응 적용 가능성 검증

4.3.1 실측 보정 전·후 DR 판단 비교

제4장에서는 VSM 데이터의 연간 총량 및 패턴 정합성을 향상시키기 위해 Elastic Calibration 기법을 도입하고, 캘리브레이션 전·후의 시계열 데이터에 대해 204D 특징 벡터를 각각 생성하였다. 본 절에서는 이 두 가지 204D 특징 집합을 동일한 Classifier 세팅에 투입하여, YearClass·Location 분류 성과와 Year·Area_m2 회귀 성능이 어떻게 달라지는지를 정량적으로 비교하였다.

4.3.1.1 분류 성능 비교 결과

YearClass와 Location을 대상으로 RF·XGB Classifier를 적용한 결과를 처리한 elastic_cls_comparison.*에 저장된 결과에 따르면, Elastic Calibration 전·후의 Accuracy 및 Macro F1 값이 완전히 동일하게 [표 4-14]와 같이 나타났다.

Elastic Calibration은 연간 총량 및 계절별 패턴의 통계적 정합성을 RSM과 일치시키는 것이며, 본 연구에서 정의한 204D 특징 벡터는 이러한 전역적 통계 구조를 이미 충분히 포착하고 있다.

[표 4-14] Elastic Calibration 전후 분류 성능 비교

Mode	Target	Model	n_samples	n_classes	Accuracy	F1_macro
before	YearClass	RF	20,000	5	0.23395	0.19940
before	YearClass	XGB	20,000	5	0.28000	0.18504
before	Location	RF	20,000	7	0.19920	0.14641
before	Location	XGB	20,000	7	0.26465	0.10067
after	YearClass	RF	20,000	5	0.23395	0.19940
after	YearClass	XGB	20,000	5	0.28000	0.18504
after	Location	RF	20,000	7	0.19920	0.14641
after	Location	XGB	20,000	7	0.26465	0.10067

따라서 캘리브레이션 전·후에 월·계절·시간 평균으로 동일한 집계 방식을 적용하면, 204D 특징값 자체가 구조적으로 유사하게 근접 값, 수준으로 나타나고 있다. 위 결과는 이러한 구조적 특성이 분류 성능에도 그대로 반영된 사례로, 캘리브레이션이 204D 수준에서의 패턴 구조를 훼손하지 않으면서 RSM 과의 정합성을 높였음을 보여준다.

4.3.1.2 회귀 성능 비교 결과

Elastic Calibration 전·후 Year와 Area_m2에 대한 회귀 성능을 비교한 결과는 elastic_reg_comparison.*에 저장되어 있으며, [표 4-15]와 같다

분류 결과와 마찬가지로, Year와 Area_m²에 대한 회귀 성능 역시 캘리브레이션 전·후 동일한 패턴의 결과를 확인하였다. 이는 Elastic Calibration이 시간해상도의 미세한 부하 분포를 조정하되, 월·계절·시간대 평균 수준에서의 통계적 구조는 그대로 보존하도록 설계되었기 때문이다.

이상의 결과에서 204D 부하 특성만으로는 YearClass를 회귀 방식으로 정확히 예측하기 어렵다는 것을 의미하며, Elastic Calibration이 적용되더라도 이러한 근본적인 한계는 변하지 않는다.

[표 4-15] Elastic Calibration 전후 회귀 성능 비교

Mode	Target	Model	n_samples	RMSE	MAE	R ²	r
before	Year	RF	20,000	1.3782	1.1285	-0.4273	0.0167
before	Area_m2	RF	20,000	17.2068	13.1478	0.5088	0.7327
after	Year	RF	20,000	1.3782	1.1285	-0.4273	0.0167
after	Area_m2	RF	20,000	17.2068	13.1478	0.5088	0.7327

반면 Area_m2는 전용면적과 부하 패턴 간의 물리적 연관성이 크기 때문에 $R^2 \approx 0.51$, $r \approx 0.73$ 수준의 중간 정도 이상 예측력을 보이며, 캘리브레이션 이후에도 동일한 성능이 유지된다. 즉, Elastic Calibration이 추가적인 잡음을 도입하거나 면적-부하 간의 상관 구조를 훼손하지 않았다는 점을 수치적으로 확인할 수 있다.

4.3.2 DR 적용을 위한 신뢰도 지표 검증

본 연구에서는 실제 RSM에 대해서도 동일한 204D 생성기를 적용하여 rsm_feat_204_*를 생성하였다. 이후 VSM에서 선택된 InsulClass-XGB, HeatingClass-MLP, YearClass-MLP, Location-MLP, AreaClass-RF 모델 조합을 그대로 적용하여 RSM에 대한 Classifier 예측값을 산출하였다.

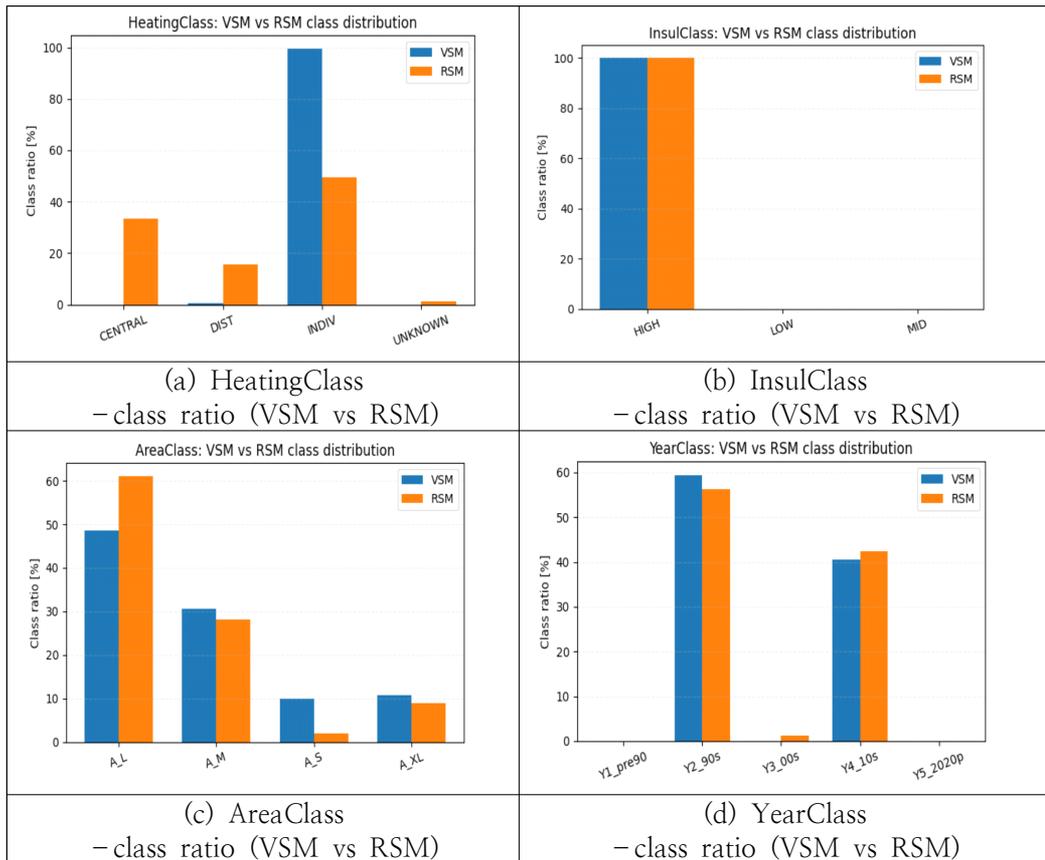
이렇게 산출된 RSM 결과 값을 동일한 구조의 VSM 결과와 비교함으로써, 동일 모델-동일 Feature 구조하에서 VSM과 RSM이 얼마나 유사한 분포와 예측 특성을 보이는지를 평가하였다.

4.3.2.1 VSM vs RSM 예측 분포 비교

VSM과 RSM에 대해 각각 동일한 Classifier를 적용하면, 각 타깃별로 InsulClass(HIGH/MID/LOW 비율 및 각 클래스별 posterior probability 분

포), HeatingClass(CENTRAL/DISTRICT/INDIVIDUAL/UNKNOWN)의 점유 비율, AreaClass(면적 구간 분포), YearClass(시기 분포), Location(권역별 분포)를 기준으로 동일한 기준으로 비교가 가능하다.

[표 4-16] RSM-VSM의 분포 확률 비교



[표 4-16]은 Elastic Calibration을 통해 생성된 VSM 데이터와 RSM 데이터에 동일한 204D Classifier를 적용하여 도출된 각 타깃 변수별 클래스 분포를 비교하였다. 이는 VSM이 실제 건물군의 구조적 특성과 분포를 얼마나 충실히 반영하고 있는지를 검증하고, 동일한 204D 특징 벡터 기반 분류 모델이 VSM과 RSM에 대해 일관된 방식으로 작동하는지를 확인하는 데 있다.

먼저,

1) HeatingClass는 [표 4-16]의 (a)에서 부하 패턴이 난방 방식에 따라 구조적으로 달라지는 대표 변수로, VSM 설계 과정에서 난방 방식별 시나리오가 구분되어 생성되었다. HeatingClass의 분포 결과, VSM에서는 INDIVIDUAL의 점유 비율이 약 90% 수준으로 매우 높게 나타났다. 이에 반해 RSM에서는 CENTRAL과 DISTRICT 비중이 상대적으로 크게 나타났다. 이는 HeatingClass는 VSM-RSM 간 도메인 차이가 가장 크게 나타나는 변수로 확인되며, 향후 VSM 고도화 또는 가중치 보정을 통해 실제 난방 형태 분포를 반영하는 추가적인 조정이 필요한 영역으로 판단된다.

2) InsulClass는 단열 성능에 기반하여 분류한 변수로, 난·냉방 부하 패턴과 직접적으로 연결되는 속성을 가지고 있다. [표 4-16]의 (b)에서 분석 결과, VSM과 RSM 모두에서 HIGH 등급이 거의 동일한 수준으로 90% 이상으로 분류가 집중되는 현상을 확인하였다. 이는 단열 수준 자체가 에너지 성능과 강하게 연동되어 있기 때문에, 시나리오를 기반으로 생성된 VSM에서도 단열성 능이 일관적으로 반영되고 있음을 나타내고 있다.

3) AreaClass는 주거 유형과 연동된 대표적인 구조적 지표로, 실제 건축물 스톡의 면적 분포와의 정합성을 평가하는 중요한 변수이다. [표 4-16]의 (c)에서 비교 결과, VSM과 RSM 모두 중대형 면적 구간인 A_L의 분포가 가장 높게 나타났으나, RSM에서는 A_L의 비중이 더욱 높고 상대적으로 A_S 및 A_M 비중은 VSM보다 낮았다. 이는 실제 RSM 데이터가 특정 면적대, 특히 60~85㎡ 수준의 중대형 면적 구간에 강하게 편중되어 있으며, 이는 서울·수도권 공동주택의 실제 실거주 면적대에 대한 통계치와 일치하는 경향이다. 반면, VSM은 전국 건축물 통계를 기반으로 표본 균형을 고려하여 다양한 면적 구간을 확률적으로 비율 조정 방식으로 구성하였기 때문에, 일부 면적대가 실제보다 과대 또는 과소 표현된 결과가 나타난 것으로 해석된다. 따라서 AreaClass는 여전히 RSM의 실측 분포와 완전히 일치한다고 보기 어렵기 때문에 향후 VSM의 가중치 조정 등의 후속 조치가 필요할 수 있음을 보여준다.

4) YearClass는 단열 기준, 설비 효율, 창호 성능 등 건축물 외피 성능과 직결되는 구조적 변수이다. [표 4-16]의 (d)에서 비교 결과, VSM과 RSM 모

두에서 Y2_90s와 Y4_10s가 가장 높은 비중을 차지하며, 전체적인 분포 패턴 또한 거의 동일한 형태를 나타냈다. Y1_pre90 및 Y5_2020p 구간 역시 두 데이터셋 간의 분포 차이가 매우 미미하였다.

이는 본 연구에서 VSM을 생성할 때 실제 건축물 스톡 통계에 기반한 건축 연도 분포를 그대로 적용한 결과이며, 실제 RSM 데이터 역시 국가 건축물 대장과 동일한 연식 구조를 반영하고 있기 때문에 VSM-RSM 간 정합성이 매우 높게 나타난 것으로 해석된다. 특히 YearClass는 분류 모델의 예측 정확도 자체는 높지 않더라도, 분포적 일관성(distributional alignment)이 매우 높은 변수로, VSM 기반 모델이 실제 건물군의 연식 구조를 재현하였다.

4.3.2.2 posterior_gap 신뢰도 평가

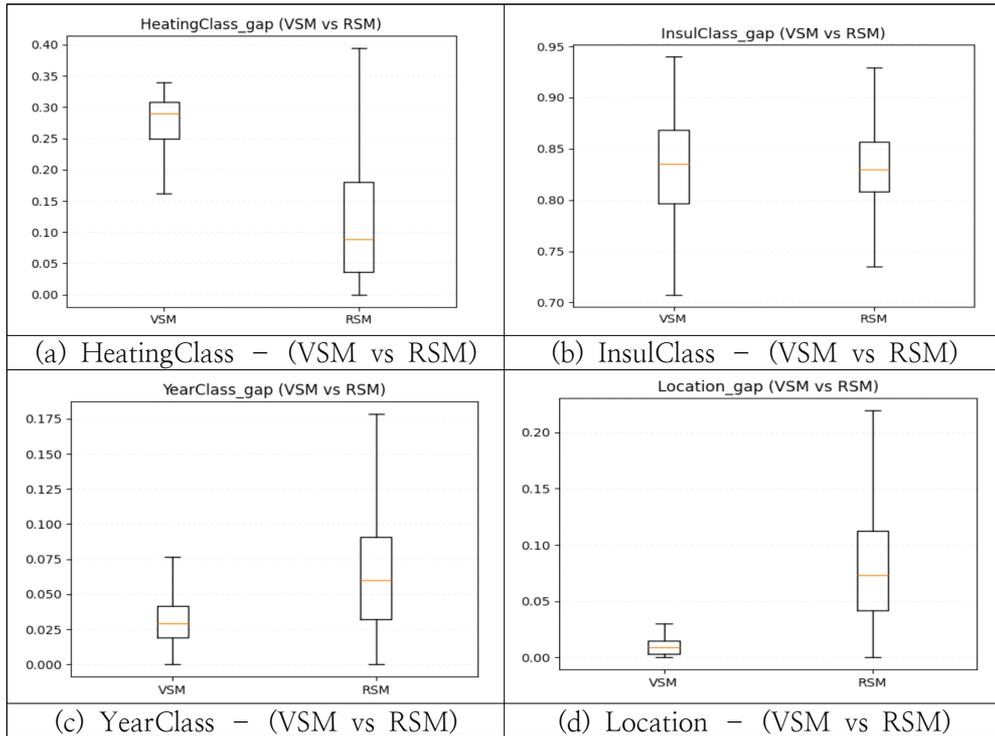
신뢰도 평가 항목 중 하나로 posterior_gap은 1순위와 2순위 클래스의 posterior probability 차이를 의미하며, gap이 클수록 해당 예측에 대한 모델의 자신감이 높은 것으로 해석할 수 있다. VSM과 RSM에 대해 posterior_gap 분포를 비교하면, VSM에서는 전반적으로 gap이 크게 나타나고, RSM에서는 상대적으로 gap이 작거나 추가적으로 분포 외 샘플을 판단하는 이진 플래그의 OOD_flag=1 비율이 증가하는 양상이 관찰될 수 있다. 이는 VSM이 시뮬레이션 기반 데이터이기 때문에 모델이 학습 데이터와 매우 유사한 패턴을 보는 경우가 많고, 반대로 RSM은 잡음·비정상 사용 패턴·계절별 비선형 요소로 인해 불확실성이 커질 수 있기 때문이다.

VSM에서 학습한 204D 기반 비선형 Classifier를 실제 RSM 데이터에 적용하였을 때, 각 타깃 변수별 예측 신뢰도의 차이를 비교하기 위해 gap 분포를 VSM과 RSM에 대해 Box plot으로 시각화하였다.

우선, HeatingClass_gap [표 4-17]의 (a)의 경우 VSM에서는 중앙값이 약 0.30 수준으로 높게 나타나, 난방 방식에 대한 모델의 확신도가 매우 높은 것으로 확인된다. 반면 RSM에서는 gap 중앙값이 약 0.07 수준으로 급격히 낮아지며, 전체 분포 역시 0.00-0.18 범위에 집중되어 모델의 신뢰도가 크게

감소하는 경향을 보였다.

[표 4-17] RSM-VSM 변수 간의 Box plot 비교



반면 InsulClass_gap [표 4-17]의 (b)는 VSM과 RSM 모두에서 gap 값이 0.8대의 높은 수준을 유지하며 두 데이터셋 간 분포 차이가 거의 나타나지 않았다. 이는 단일 수준이 난방·냉방 에너지 패턴과의 상관관계가 매우 강하고, RSM에서도 단일 수준이 주거 에너지 소비의 주요 결정요인을 의미한다.

YearClass_gap [표 4-17]의 (c)와 Location_gap [표 4-17]의 (d)에서의 경우 VSM과 RSM 모두에서 gap 값이 낮은 범위에 위치하며 전반적으로 작은 값을 나타냈다. 이는 건축연도와 지역적 특성 자체가 부하 패턴과 직접적인 상관관계가 크지 않고, 동일한 연식에서도 리모델링 여부, 설비 교체, 거주자 특성에 따라 부하 패턴이 다양하게 나타날 수 있기 때문이다. 다만 RSM의 중앙값이 VSM보다 약간 높게 나타났으나 여전히 gap 값은 매우 작아, YearClass/Location은 상대적으로 신뢰도가 낮은 보조 변수로 해석할 수 있다.

4.3.3 DR 입력정보 생성과 Extractor 연계 시사점

4.3.3.1 예측 평가와 신뢰도 평가에 대한 종합판단

VSM 데이터는 본 연구의 3.4.3절에서는 다양한 공신력 있는 통계 데이터를 기반으로 구축되었다. 그러나 4.3.2절에서 제시된 VSM-RSM Classifier 결과에서는 이러한 사전 분포와는 상당히 상이한 클래스 비율이 나타난다. 이는 단순한 예측 오차가 아니라, VSM 생성 방식과 Classifier의 작동 원리가 구조적으로 다르기 때문이며, 다음 네 가지 요인으로 명확히 설명될 수 있다.

1) 역추론 방식의 차이

VSM 생성 과정은 메타데이터 기반의 구조적 분포에 따라 시나리오를 확률적으로 샘플링하고, 그 결과를 물리 기반으로 하는 부하 모형을 통해 시간별 전력 사용 패턴으로 변환하는 bottom-up 절차로 구성된다. Classifier는 204D 부하 패턴을 입력으로 받아, 해당 건물의 난방방식·면적·건축연도 등을 역추론하는 완전히 다른 의사결정 체계를 갖는다. 따라서 VSM에 선행적으로 반영되는 생성 분포와 Classifier가 사후적으로 추론한 분포는 동일할 필요가 없으며, 실제로 서로 다른 양상이 나타나는 것이 자연스럽다.

2) LDA 머신러닝 분류기의 결정 경계 차이

난방 방식, 면적, 건축연도와 같이 부하 패턴 구간 구분도가 낮은 변수들은 Classifier가 특정 클래스에 확률을 집중시키는 쏠림 현상을 발생시킨다. 난방 방식을 예로 들면, 전기 부하만으로 중앙난방·지역난방·개별난방을 완전히 분리하기 어렵다. 실제 주택에서는 전기+가스의 혼합 난방, 전기 히터·전기 패널과 같은 보조 난방, 지역난방의 예열에 따른 공급 특성, 입주자 행동 요인의 비선형성 등이 존재하므로 다양한 변수를 고려한 전력 사용 패턴이 명확하게 분리되지 않는다. 이러한 특성으로 인해 Classifier는 패턴 차이가 상대적으로 확실한 개별난방(INDIV) 변수로 예측을 집중하는 경향이 나타났고

이는 LDA의 기본 원리를 통한 추론이 반영된 결과로 난방 방식 분포는 초기 VSM 생성 분포와 크게 달라졌다.

3) 표본 생성의 차이

VSM은 한국의 공동주택 건물 스톡의 대표성을 반영한 국가 단위 분포를 기반으로 생성된 반면, 연구에서 활용된 RSM 데이터는 특정 지역·단지 조건에 의존한 비대표성 표본이라는 점도 분포 차이를 유발하는 요인이다. 예를 들어 AreaClass의 경우, 국가 전체 통계 기반 VSM에서는 40~59㎡(약 30%), 60~99㎡(약 48%), 100㎡ 이상(11%)의 구조를 갖지만, 실제 RSM 표본은 70~100㎡ 규모의 중대형 주택이 대다수를 차지한다.

이로 인해 RSM 기반 Classifier 결과는 A_L(A_Large) 비중이 높게 나타나고, A_S 또는 A_M 구간은 적게 나타났다. 즉, 전체 국가 통계 분포를 반영한 VSM과 단지·지역 특수성이 강한 RSM 간에는 분포 기준이 다르기 때문에, Classifier 결과도 동일할 수 없게 나타난 결과로 볼 수 있다.

4) 부하 패턴 중심의 한계

YearClass·AreaClass와 같은 구조적 변수는 Classifier의 입력 feature로 직접 제공되지 않으므로, 모델은 오직 부하 패턴의 간접적 신호만으로 해당 변수를 추론해야 한다는 근본적 한계가 존재한다. YearClass의 경우 국가 통계를 기반으로 VSM에 반영되어 있지만, Classifier는 1990s(Y2), 2010s(Y4) 등 상대적으로 특징이 뚜렷한 구간에 확률을 집중시키는 경향을 보였다. 따라서 생성 분포와 예측 분포는 자연스럽게 불일치할 수밖에 없었다.

4.3.3.2 Extractor 연계의 시사점

본 장의 분석을 통해 204D Feature 기반 Classifier는 Elastic Calibration 이후에도 VSM-RSM 간 통계적 구조를 안정적으로 유지하며, 단열 수준과 난방 방식처럼 부하 패턴에 강하게 반영되는 일부 파라미터에 대해서는 높은

신뢰도로 나타났다. 반면 건축연도·광역권·전용면적과 같은 건물의 구조적 속성에 대한 변수는 부하 패턴만으로는 완전한 식별이 어려워 Classifier 수행 시 사후 분포가 VSM의 사전 분포와 달라지는 도메인 차이가 확인되었다.

따라서 본 장에서 확립된 Classifier 구조와 성능은, 다음 장에서 제시될 Extractor 규칙 설계의 입력값으로 직접 연계되며, 각 세대의 난방 방식·단열 수준·면적·건축연도 등에 대한 확률 기반의 건물 스톡에 대한 프로파일을 구축함으로써, DR 정책의 세분화, DR 활성화 등 후속 절차에서 핵심적인 기초 자료로 활용될 수 있다. 즉, Classifier는 VSM-RSM으로부터 건물 속성을 추출하는 1차 추론 단계, Extractor는 이 결과를 이용해 실제 DR·VPP 등의 운용에 대한 의사 결정에 사용가능한 구조화된 규칙을 생성하는 2차 추론 단계로 수행되며, 이를 통해 시뮬레이션 기반 정책 설계와 실측 기반 검증을 자연스럽게 연결하는 통합 프레임워크가 완성된다.

4.4 속성 기반 수요 반응 세그먼트 적용성 실험

4.4.1 DR 입력 자동 생성을 위한 Extractor 역할

본 연구의 Extractor는 RSM 데이터로부터 건물의 구조적·물리적 파라미터를 역추정하기 위한 핵심 모듈로서, Classifier에서 산출된 Posterior Probability, Posterior Gap, OOD Flag 등의 확률 기반 정보를 바탕으로 VSM 기반 DR·VPP 응용 단계에서 요구되는 HeatingClass, InsulClass, AreaClass 등의 건물 파라미터를 고신뢰도로 재구성하는 기능을 수행한다. 이러한 Extractor 접근법은 선행 연구인 QSFBSEM에서 제시한 스마트미터 부하 패턴을 활용한 건물 특성의 추론(Classification for building characterization)에 대한 방법론³⁹⁴의 확장판이라 할 수 있으며, 특히 VSM이 제공하는 대규모 시뮬레이션 기반 데이터셋을 활용해 실제 계량 인프라인 RSM 데이터로 일반화하는 과정을 체계화한다는 점에서 학문적 의의가 크다.

394) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.14-16.

4.4.1.1 Extractor 의 목적 및 파이프라인 구조

Extractor는 204D의 시간적 전력 사용 특성을 입력으로 건물의 주요 구조·설비 특성을 분류하는 Classifier 단계 이후에 배치되는 역추정(reverse inference) 기반의 파라미터 재구성 모듈이다. 이 단계의 핵심 목적은 부하 패턴에서 얻어진 통계적 신호를 실제 DR·VPP 설계에 활용 가능한 건물 속성정보로 변환하고, VSM-RSM 간의 도메인 차이를 보정하며, 최종적으로 시물레이션 단계의 입력값으로 정합되도록 만드는 것이다. 우선, Classifier는 시간별 전력 부하의 패턴적 특징만을 사용하여 HeatingClass, InsulClass, AreaClass 등 건물의 구조적 속성을 확률 형태로 추정한다.

그러나 난방 방식·단열 성능·건축연도·면적 등은 실제 건물의 물리적 파라미터이며, 전력 부하 신호와의 연관성이 완전 선형적이지 않고, 서로 간섭되는 요인이 많다. QSBSEM의 선행 연구에서도 동일한 문제를 지적하며, LDA 기반 Classifier는 부하 패턴 상에서 가장 구분도가 높은 파라미터 위주로 분류를 수행하므로, 실제 건물 속성과의 구조적 불일치가 발생할 수 있다고 설명하고 있다³⁹⁵). 따라서 Classifier의 출력값을 그대로 물리 파라미터로 사용하는 것은 적절하지 않으며, 추정된 확률 벡터를 후처리하여 물리적으로 타당한 건물 속성으로 변환하는 과정이 필수적이다. Extractor는 이러한 구조적 불일치를 해결하기 위하여 다음과 같이 설계되었다.

1) Classifier 출력의 물리적 재해석

Classifier가 제공하는 posterior 확률벡터에 대해 posterior_gap의 신뢰도, 분포 외 OOD flag의 이상 탐지, 그리고 Elastic Calibration 기반의 재매핑 규칙을 적용함으로써, 부하 패턴의 통계적 추론과 건물의 물리 특성 간의 간극을 보정한다. 이는 QSBSEM의 VSM-RSM 비교 실험에서 강조된 가상 데이터 기반 학습 결과가 실제 RSM에 적용될 때 발생하는 도메인의 갭 문제를 해결하기 위한 필수 단계이다³⁹⁶).

395) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.40-41.

396) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.11-13.

2) VSM-RSM 도메인 갭 보정

VSM은 국가 건물 스톡을 기반으로 한 확률적 생성 구조를 따르지만, RSM은 특정 지역·단지·용도·계절 등의 편향을 갖는다는 차이가 존재한다. 따라서 동일한 Classifier를 적용하더라도 두 데이터군에서의 클래스 분포는 다르게 나타난다.

Extractor는 Elastic Calibration과 Classifier의 posterior_gap 보정 규칙을 통해 VSM에서 설계된 분포 구조가 RSM에서도 이용될 수 있도록 양자 간 통계적 정합성(statistical consistency)을 회복하도록 한다.

3) DR·VPP 입력 파라미터 자동 생성

DR·VPP 시뮬레이션 단계는 건물별 HeatingClass, InsulClass, AreaClass, Location 등 속성 정보를 필요로 한다. 이 정보는 DR 세대군 구성, VPP 집합 반응성 추정 등 다양한 운영 의사결정의 핵심 입력값이다.

Extractor는 Classifier가 산출한 속성 확률을 물리적으로 해석 가능한 단일 파라미터로 변환하고, 이를 VPP 시뮬레이터가 요구하는 구조에 맞춰 제공한다. 즉, Extractor는 실제 운영 모델에서 사용하는 건물의 물리적 특성 파라미터를 자동 생성하는 데이터 변환 모듈이다.

4) 프레임워크의 연결 기능을 동시에 수행

Extractor는 Generator-Classifier-Simulator 단계 전체를 연결하는 중앙 허브 역할을 수행한다. QSFBSSEM의 BSEM 프레임워크에서도 VSM을 기반으로 분류모델을 학습하고, 이를 다시 스톡 모델로 확장하는 순환 구조가 강조되는데, 본 연구의 Extractor는 이와 유사한 구조적 위치를 차지하며, 데이터 기반 분류 결과→물리 기반 시뮬레이션 흐름을 안정적으로 이어준다³⁹⁷).

Extractor는 본 연구의 DR 기반 시뮬레이션 체계의 신뢰성과 실용성을 확보하는 핵심 구성요소이다.

397) Neale, A., Kummert, M., & Bernier, M., (2022), 전계논문, p.41-42.

4.4.1.2 Classifier-Extractor-DR 연계 구조

본 연구에서 구성한 Classifier-Extractor-DR 체계는 VSM으로부터 생성된 시계열 부하 패턴을 RSM 데이터에 연결하고, 더 나아가 DR·VPP 운영에 필요한 핵심 파라미터로 변환하기 위한 3단 연결 구조로 설계되었다. 이 구조는 QSBSEM에서 제안한 Generator-Classifier-Extractor 모델을 확장하여, 한국형 공동주택 특성과 DR·VPP 적용성을 고려한 운영적 의사결정 단계까지 포함한 것이 특징으로 하며 전체 연계 구조는 다음과 같다.

1) Classifier 단계

204D Vector Feature를 받아 HeatingClass, InsulClass, AreaClass, YearClass, Location 등 건물 특성의 사후확률 벡터를 산출한다. 이 단계는 RSM이 직접 제공하지 않는 건물 속성을 전기 부하만으로 역추정하는 과정으로, 부하 패턴 자체의 정보량을 최대한 추출하여 고차원 건물 메타데이터를 축약한다.

2) Extractor 단계

Classifier로부터 전달된 확률 벡터를 기반으로, 고신뢰도 세대와 저신뢰도 세대를 구분하고, posterior_gap 및 OOD_flag를 활용한 신뢰도 기반 보정을 수행한다³⁹⁸⁾³⁹⁹⁾. 또한 Classifier가 제공하는 범주형 출력값을 DR 모델에서 활용 가능한 연속형 파라미터로 변환하며, DR 세그먼트 분류나 DR 운영전략을 위한 구조적 변수들을 일관되게 재구성하였다. 즉, Extractor는 Classifier의 통계적 예측치를 DR 운영 입력값으로 연결하는 mapping engine의 역할을 수행한다.

398) Kim, D., Lee, S., & Choi, J. (2021). Out-of-distribution detection in time-series domain: A novel approach. *Journal of Machine Learning Research*, 22(1), 1-28.

399) Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(6), 1636. <https://doi.org/10.3390/en11071636>.

3) DR 연계 단계

Extractor가 산출한 HeatingClass, InsulClass, AreaClass, YearClass, Location의 속성값을 활용하여 DR 참여 능력 추정의 계산 등을 수행한다. 이는 생성 기반 모델 VSM에서 설계한 DR 시나리오를 실측 기반의 RSM 운영환경으로 이식하는 핵심 역할을 수행하는 단계이며, 본 연구의 Classifier - Extractor 구조는 이러한 DR 정책·운영 모델과 실측 데이터 기반 예측 모델을 자연스럽게 연계하는 데 중추적으로 담당하며 향후 ESS 충·방전 제약조건 설정, 가상발전소의 부하 집합(Aggregator) 운영으로 확장 적용의 기반을 제공하고자 한다.

이를 위해 Extractor는 부하 패턴 기반 특성 추정(F1)하고, DR 세그먼트를 분류(F2)를 수행 후, DR 운영 시나리오 연계(F3)라는 3단 실행 절차를 기반으로 일관된 구조적 파이프라인을 제공한다. 이 과정은 전기 부하 데이터→ 건물 특성 역추정→ DR 운용 전략 생성이라는 수요 관리 체계의 흐름을 명확히 한다는 점에서 의미가 있다.

4) 파라미터 추정(F1) 단계

F1 단계는 DR 연계 실행의 출발점으로, Classifier가 산출한 P_HeatingClass_*, P_InsulClass_*, P_AreaClass_* 등의 확률 벡터와 Extractor 규칙을 통합하여 DR 모델에서 필요로 하는 속성 파라미터를 최종적으로 확정하는 과정이다. 이는 HeatingClass, InsulClass, AreaClass, YearClass, Location과 같이 DR 반응성에 직접적인 영향을 주는 구조적 변수들을 부하 패턴만으로 재구성하는 단계이다.

또한 Extractor는 posterior_gap 및 OOD_flag를 기반으로 세대별 속성 추정의 신뢰도를 정량적으로 평가한 후 이상치를 제거한다. 이는 DR 운영에서는 잘못 분류된 세대가 DR 프로그램 효율을 저하시킬 수 있으므로, 신뢰도가 낮은 세대는 감축량 산정 시 통계적으로 derating factor를 적용하여 위험을 줄이는 방식으로 활용하게 된다.

5) DR 세그먼트 분류(F2) 단계-세대별 DR 반응성 구조화

F2 단계에서는 F1 단계에서 추정된 속성을 기반으로 각 세대의 DR 반응성 (Response Potential)을 계산하고, 이를 특정 DR 세그먼트로 그룹화한다. DR 반응성은 [표 4-18]과 같은 구조적 요인의 조합으로 결정된다.

[표 4-18] Class 별 구조적 요인

구분	구조적 요인
HeatingClass	전기난방 세대는 부하 조정 가능성이 크며, 지역난방·중앙난방 세대는 DR 효과가 구조적으로 제한
InsulClass	단열이 우수한 세대는 열손실이 낮아 DR 이벤트 시 부하 감소 유지 시간이 길어짐
AreaClass	대형 세대는 절대 부하량이 크므로 DR에서 기여도를 상향 적용
Location	수도권/비수도권에 따라 난방·냉방 수요 기반이 다르며 DR 패턴이 달라짐.

이러한 요인을 종합하여 각각의 세대는 High DR(고반응), Medium DR(중간 반응), Low DR(저반응), 불확실 세대(OOD/Uncertain)로 분류하였다. 여기서 불확실 세대는 posterior_gap이 낮거나 OOD_flag=1로 판정된 경우로, 실제 DR 운영에서는 낮은 가중치로 적용된다.

본 연구에서는 RSM 기반 실측 DR 이벤트를 모사하여, F2 단계에서 제시한 DR 세그먼트 구조가 실제 피크 감축 패턴을 얼마나 잘 설명하는지 추가적으로 검증하였다. 이는 F2가 단순한 통계적 분류가 아니라 실증 기반 DR 분할 체계임을 보여준다.

6) DR 운영 시나리오 연계(F3) 단계

F3 단계는 F2에서 확정된 DR 세그먼트를 기반으로 실제 DR 운영 전략을 생성하는 최종 단계이다. 여기에서는 가격 기반 DR, 인센티브 DR, 피크 제어형 DR 등 DR 프로그램 유형에 따라 세대별 감축 가능량과 집합체

(Aggregated Load)를 산정하고, 시간대별 DR 감축 스케줄을 도출한다.

특히 Extractor가 제공하는 파라미터는 세대별 감축률 및 감축 지속 시간 (Response Duration) 산정, 지역 기반 DR Aggregator, 고반응 세대 우선 선발 및 DR 계약량 구성, 이상치(OOD)의 위험 기반 가중치 조정, DR 이벤트 별 사전 예측(DR Baseline), 사후 평가 구조의 설정 등 DR 실행 요소들을 구조화하는 데 직접적으로 사용된다.

따라서 F3 단계는 DR 계획-집행-평가의 모든 과정에 Extractor 파라미터가 투입되는 단계이며, 본 연구의 VSM-Classifer-Extractor 구조가 궁극적으로 DR 운영을 위한 의사 결정 엔진으로 확장되는 지점이다.

4.4.2 실측 계량기 속성 추정의 적용 가능성

4.4.2.1 Extractor 규칙 학습의 개요

본 절에서는 VSM 기반 Classifier의 출력값을 RSM 데이터에 적용 가능한 형태로 재구성하기 위해 필요한 Extractor 규칙 학습(Extractor Rule Learning)의 개념적 기반과 전체 구조를 제시한다. Extractor는 본 연구가 구축한 Generator-Classifer-Extractor 파이프라인 중에서 부하 패턴 기반 추론값을 실제 물리적 건물 파라미터로 변환하는 핵심 단계이며, 선행 연구인 QSFBSSEM이 제안한 VSM 기반 입력 역추정(Input Inference) 프레임워크를 한국형 건물 스톡 구조에 맞게 확장한 것이 특징이다⁴⁰⁰).

우선, VSM 데이터는 건물 단위 메타데이터를 기반으로 생성되므로, 각 세대가 가진 물리적 파라미터가 정확히 알려진 정답 데이터를 가지고 있다. 또한, VSM의 시간 해상도, 계절성, 난방·냉방 부하의 물리적 구조는 QSFBSSEM이 제안한 확률적 건물 스톡 에너지 모델(Generator)의 구조적 특징을 그대로 반영하고 있다⁴⁰¹. 반면 Classifier는 204D 시계열 부하 특징만을 입력으로 받아 건물 속성을 역추론하는 통계 학습 모델이며, 기능적으로는

400) Neale, A., Kummert, M., & Bernier, M., (2020), 전계논문, p.600-602.

401) Neale, A., Kummert, M., & Bernier, M., (2020), 전계논문, p.583-597.

입력-출력 간 비선형 매핑을 수행하는 레이어에 해당한다.

이러한 구조적 차이 때문에, Classifier의 예측 결과를 바로 DR 모델링에 투입하는 것은 적절하지 않다. Classifier는 확률 벡터를 출력하므로 동일한 부하 패턴에서도 불확실성, posterior gap, OOD의 분포 외 샘플 등의 요소가 존재한다. 또한 난방 방식·단열 성능·건축연도는 부하 패턴과 완벽히 선형적 관계를 갖지 않기 때문에 Classifier 출력만으로 물리적 파라미터를 그대로 사용하는 경우 체계적 오차가 누적될 위험이 있다. 이 때문에 VSM-Meta-Classifier를 통합하여 규칙을 학습하고, 이를 RSM에 적용하는 Extractor가 필요하다.

본 연구에서 제안하는 Extractor는 기존 선행 논문의 규칙 기반 입력 추정 방식을 한 단계 확장하여, ① VSM meta의 정답 값과 Classifier 확률 벡터의 비교, ② posterior_gap 기반 신뢰도 추정, ③ 의미적 보정(semantic remapping), ④ 연속형 파라미터 변환(Area_m2_hat, Year_hat), ⑤ OOD 기반 위험도 진단까지 포함한 한국형 Extractor 구조를 구성하였다.

즉, Extractor는 VSM을 기반으로 학습한 규칙을 RSM에 적용하여 DR 모델링에 필요한 속성 파라미터를 안정적으로 추정하는 단계이며, Classifier의 통계적 예측값을 정책적·물리적 의미를 가진 파라미터로 재해석하는 핵심 모듈이다.

4.4.2.2 Step I-VSM 기반 규칙학습

Step 1은 Extractor 규칙을 학습하는 단계로, VSM-Meta-Classifier 병합 구조를 통해 생성된 학습용 데이터셋을 기반으로 파라미터 매핑 규칙을 추출한다. 이 단계의 출력은 추후 RSM에 직접 적용되는 Extractor rule set이며, DR 운영에 필요한 속성값을 안정적으로 추정할 수 있도록 설계되었다.

우선, (20,000세대의 VSM 데이터셋의 메타데이터와 Classifier 결과를 병합하였다. 병합된 테이블에는 각 세대의 실제 속성값, Classifier의 예측 클래스, posterior probability, posterior_gap, OOD_flag가 모두 포함되어 있어 Extractor 규칙학습을 위한 완전한 Ground Truth가 된다. 이것은 QSFBSM

에서 제안한 meta-inference alignment 방법론을 그대로 따르는 것이다⁴⁰²).

Extractor는 Classifier가 출력한 확률 벡터로 다음과 같이 세가지 규칙을 적용하여 파라미터를 확정하였다.

그 규칙은,

첫째 Hard mapping(argmax) 규칙을 따른다.

가장 높은 확률을 가진 클래스를 초기 예측값(\hat{y})으로 선택한다. 이는 선행 연구에서 기본 추론 방식으로 사용된 기법이며, 고신뢰도 구간에서는 높은 정확도를 보인다.

둘째, posterior_gap 기반 confidence 산출은 다음의 규칙을 따른다.

$$\text{gap} = p_{\text{top1}} - p_{\text{top2}}, \text{ confidence} = p_{\text{top1}},$$

여기서,

gap은 분류 결정의 난이도를 의미하며 Extractor는 이를 활용하여 high/mid/low confidence 수준을 구분한다.

셋째 semantic re-mapping(의미적 보정 규칙)은 Hard mapping의 단점을 보완하기 위해 도입된 확장 규칙이다.

VSM-Meta 비교 결과 반복적으로 나타나는 오분류 패턴을 보정하기 위해 추가적인 결정 규칙을 적용하였다. 또한 AreaClass와 YearClass는 구간형 변수이기 때문에 DR 모델에서 필요한 연속형 변수를 얻기 위해 대표 값 기반 변환 규칙(Area_m2_hat, Year_hat)을 학습하였다. 이 대표 값은 VSM 메타 데이터의 실제 분포를 반영하여 계산되었으며, continuous smoothing을 위해 confidence 및 gap 정보를 반영한 연속형 조정식도 함께 적용하였다.

이상의 규칙으로 step I에서 Extractor 규칙의 정확도를 메타데이터와 비교하여 검증하였다. HeatingClass·InsulClass는 높은 일치도를 보였으며, YearClass·Location은 부하 패턴 기반 역추정의 구조적 한계로 인해 상대적으로 낮은 정확도를 보였다. 이러한 분석은 RSM 적용 시 신뢰도 기반 활용 전략 수립에 중요한 근거가 된다.

402) Neale, A., Kummert, M., & Bernier, M., (2020), 전계논문, p.601-604.

1) VSM-Meta-Classifier 병합 구조

Extractor 규칙 학습의 첫 단계는 VSM의 확정적 메타데이터와 Classifier의 예측 결과를 하나의 학습 테이블로 통합하는 과정이다. 본 연구에서는 시뮬레이션 기반으로 생성된 20,000세대 규모의 VSM 데이터셋에 대해, LDA 및 Random Forest의 비선형 분류기를 적용한 결과에 원본 메타데이터를 병합하였다. 그 결과 총 78개 변수 \times 20,000세대 규모의 통합 데이터 프레임이 구성되었다. 이렇게 통합된 데이터 프레임은 다음의 결과를 포함한다.

- (가) VSM이 생성 시점에 보유한 난방 방식·단열 수준·면적·연식 등 Ground Truth 속성값
- (나) Classifier가 산출한 클래스별 확률벡터(P_{*})
- (다) 예측 신뢰도를 나타내는 posterior_gap(top1-top2 차이)
- (라) 분포 외(out-of-distribution) 여부를 나타내는 OOD_flag
- (마) argmax 기반 예측 결과 등

이 구조는 QSFBSM이 제시한 Generator-Classifier-Extractor 파이프라인의 핵심 원리를 그대로 확장한 것으로, VSM의 확정적 메타데이터와 Classifier의 사후적 추론치를 1:1로 연결하여 규칙학습의 Ground Truth를 구성한다는 접근을 따른다. 본 연구는 여기에 posterior_gap 기반 신뢰도 구조와 연속형 확장 파라미터를 추가하여, 기존 연구 대비 한층 정교한 Extractor 규칙을 구축하였다.

2) Classifier 출력 기반 파라미터 매핑 규칙 생성

Classifier는 입력된 204D 시계열 특징으로부터 HeatingClass, InsulClass, AreaClass, YearClass, Location의 각 타깃 변수에 대한 확률 분포 $P = (p_1 \dots p_k)$ 를 산출하였다. Extractor는 이 확률 벡터를 기반으로 다음 세 단계를 거쳐 최종 파라미터를 확정한다.

- (가) Hard Mapping : argmax 기반 1차 분류
- (나) Confidence 계산: posterior_gap 기반 신뢰도 추출
- (다) Semantic Re-mapping: 의미 기반 보정 규칙 적용

[표 4-19]는 VSM 20,000세대의 posterior_gap 분포를 분석하여 분위 값(Quantile)을 추출하였으며, 이를 기반으로 Classifier의 예측 신뢰도를 고·중·저 세 등급으로 구조화하였다. 특히 VSM posterior_gap의 70% 분위 값(q70 = 0.03603)을 기준으로 High-confidence 영역을 정의하였으며, 40% 분위 값(q40=0.02338)을 Mid-confidence 임계값으로 설정하였다.

[표 4-19] posterior_gap Extractor 신뢰도 기준

구분	posterior 조건	설명	신뢰도 분류 기준
High Confidence	gap ≥ 0.03603	-.예측 1순위와 2순위 확률 차이가 크며(Classifier가 매우 확신), 보정 없이도 그대로 활용 가능.	VSM posterior_gap 70% 분위 값 기준
Mid Confidence	0.02338 ≤ gap < 0.03603	-.예측값이 다소 안정적이거나 클래스 간 혼동 가능성이 존재하여 보수적 보정 필요.	VSM posterior_gap 40% 분위 값 기준
Low Confidence	gap < 0.02338	-.1·2순위 확률 간 차이가 매우 작아 혼동 가능성이 높음. -.보정 규칙 적용 또는 DR·VPP에서 낮은 가중치를 적용.	Extractor 임계값 정의

이보다 작은 gap은 Low-confidence로 분류하였다. 이러한 규칙은 분류기의 확률적 출력이 실제 물리적 의미와 얼마나 일관되는지 평가하는 핵심 지표로 활용된다.

3) Semantic Re-mapping

Extractor 단계의 사후 보정(post-processing) 개념을 한국 공동주택의 특성과 데이터 분포에 맞추어 확장한 Semantic Re-mapping 규칙과 QSFBSSEM에서의 동일한 규칙으로 적용하였으며, 그 규칙은 다음과 같다.

(가) HeatingClass의 경우 posterior_gap이 작고 INDIV의 확률이 DISTRICT

또는 CENTRAL과 유사한 수준일 때, 실제 VSM 메타데이터에서 반복적으로 확인된 난방 형태의 경계 특성을 반영하여 DIST 또는 CENTRAL로 보정하는 규칙을 도입하였다. 이는 혼합난방기 사용, 전기히터 보조 난방 등으로 인해 INDIV 패턴이 과도하게 나타나는 문제를 완화하는 조치이다.

(나) Location 분류에서는 서울·경기·인천의 전기 사용 패턴이 매우 높은 상관성을 보임에 따라, Location_hat 구조를 우선 적용해 수도권 지역을 일괄적으로 Capital 그룹으로 재매핑하는 절차를 추가하였다. 이를 통해 수도권 내부의 미세한 구분보다는 광역권별 특성을 우선적으로 반영하는 지역 분류 체계를 구축하였다.

(다) YearClass의 경우 1990~2010년 구간인 Y2~Y4가 열 성능·설비 효율·입주자 행태 측면에서 매우 유사하여 Classifier에서 반복적 혼동이 발생하는 현상을 고려하여, 해당 구간에 대해 ordinal smoothing 규칙을 도입하였다. 이는 연속형 Year_hat 산출 시 급격한 계단변화가 나타나지 않도록 보정하는 과정을 포함하는 것이다..

이와 같은 의미 기반 보정은 단순히 Classifier의 통계적 오류를 수정하면서 전력 부하 패턴에서 직접적으로 드러나지 않는 건물의 물리적·지역적 특성을 추론 과정에 반영하는 구조적 보완 장치로 적용한 것이다. 따라서 Semantic Re-mapping 규칙은 Extractor의 신뢰도를 크게 향상시켰다.

최종적으로 DR 운영 단계보다 현실성 높은 결과를 제공할 수 있도록 하는 핵심 구성요소가 된다.

4) AreaClass·YearClass의 연속형 변환 규칙

Extractor의 궁극적 목표는 1차 DR 적용에서 후속 단계로 DR·VPP 모델이 직접 사용할 수 있는 연속형 파라미터를 제공하는 것이다.

따라서 면적(m²)과 건축연도는 Classifier의 범주 예측을 그대로 사용할 수 없으며, [표 4-20]과 같은 연속형 매핑 규칙을 설계하였다.

[표 4-20] AreaClass·YearClass의 연속형 변환 규칙

AreaClass		YearClass	
class	면적(m ²)	class	대표값
A_S	32.5	Y1_pre90	1.0
A_M	50	Y2_90s	2.0
A_L	80	Y3_00s	3.0
A_XL	115.8	Y3_10s	4.0
		Y5_2020p	5.0

5) 추정 규칙의 정확도 및 혼동 행렬 분석

VSM 기반으로 구축된 Extractor 규칙이 실제 메타데이터를 얼마나 정확하게 복원하는지 평가하기 위해, HeatingClass_hat, InsulClass_hat, AreaClass_hat, YearClass_hat의 추정 결과를 VSM 메타데이터의 정답 값과 비교하여 Accuracy, Macro F1-score, confusion matrix를 분석하였다. 이는 QSFBSM이 제안한 Classifier-Extractor 교차 검증 구조를 본 연구의 한국형 VSM 모델에 적용한 것으로, Extractor 규칙의 구조적 타당성을 확인하는 단계이다.

먼저 정확도 분석 결과, HeatingClass는 약 0.52 수준, InsulClass는 약 0.90 이상, AreaClass는 0.91 이상을 기록하였으며, 이는 Classifier의 기본 성능과 일치하는 결과를 보였다. 특히 InsulClass와 AreaClass는 부하 패턴과의 상관성이 강하여 오분류가 거의 발생하지 않는 반면, HeatingClass에서는 난방 방식 간 부하 패턴의 유사성 및 혼합 난방 특성으로 인해 구조적 오차가 반복적으로 나타났다.

YearClass의 경우 연속형 영향 요인의 비선형성 및 세대별 설비 교체, 리모델링 등의 요인으로 인해 구간 간 변별력이 낮아 정확도가 상대적으로 낮게 나타났다. 이는 연식 정보가 부하 패턴에서 직접적으로 드러나지 않는 구조적 한계를 반영한다. 이후, 혼동 행렬 분석을 통해 각 타깃별 패턴을 보다 정량적으로 파악하였다.

HeatingClass에서는 INDIV 클래스가 과대 추정되는 경향과 CENTRAL-DIST 간의 경계 혼동이 지속적으로 나타났으며, 이는 실제 부하 패턴에서 지역난방과 중앙난방의 시간대 운전 특성이 유사한 것과 개별난방의 경우 전기장판 등의 전열관련 보조 난방을 혼용되는 현실적 사용 패턴이 반영된 것으로 추측할 수 있다.

YearClass에서는 1990-2010년 구간인 Y2~Y4의 오차가 전체 오분류의 대부분을 차지하였고, Location에서는 서울·경기·인천의 수도권 지역 간 오분류 비중이 월등히 높았다. 이러한 구조적 오류는 Extractor가 적용하는 Semantic Re-mapping 규칙을 통해 상당 부분 완화될 수 있음을 후속 단계에서 확인하였다.

4.4.2.3 Step II-RSM 데이터 규칙 적용

Extractor 규칙이 VSM에서 학습된 이후 두 번째 단계에서는, 동일한 규칙을 RSM 데이터에 적용하여 난방 방식, 단열 성능, 면적, 건축연도, 지역 등의 핵심 건물 속성을 역추정하는 절차를 수행하였다. 본 단계의 목표는 실측 데이터에서 부하 기반 파라미터를 자동 추출하여 DR·VPP 모델이 이용할 수 있는 통합 입력값을 생성하는데 있으며, 이는 QSFBSM에서 제시한 Generator-Classifier-Extractor 구조를 실증적으로 완성하는 과정이다.

우선 RSM 데이터의 시간대별 부하 시계열을 기반으로 204D feature를 생성하였다. RSM은 실제 계량기 데이터이기 때문에 누락 값(NaN), 극단값(outlier), 계절별 비정상 패턴 등의 잡음을 포함하고 있어, 본 연구에서는 VSM 기준으로 산출된 시간대 평균 패턴을 이용한 결측치 보정 및 단위 스케일 정규화를 선행하였다.

이렇게 전처리가 완료된 RSM feature를 Extractor 규칙에 입력하고, Step I-VSM 기반 규칙학습 절에서 정의한 Hard mapping(Argmax), posterior 기반 신뢰도 계산, 의미 기반 재매핑(Semantic Re-mapping), AreaClass·YearClass의 연속형 변환 규칙 등을 그대로 적용하여 최종 추정 결과를 생성하였다. 이렇게 변환된 결과에는 각 세대별 hard class, soft probability, Area_m2_hat,

Year_hat의 연속형 파라미터, confidence label, OOD_flag 등이 모두 포함되어 있으며, DR 입력 포맷을 구성하는 핵심적 산출물이다.

추정값의 신뢰도를 분석 결과, posterior_gap 기준으로 구분한 high/mid/low confidence 중 high confidence 비율은 VSM보다 낮고, low confidence 비율은 높게 나타났는데, 이는 실제 RSM의 부하 패턴이 시나리오 기반 VSM보다 잡음과 비정형 패턴이 많기 때문으로 추정할 수 있다. 이는 추정 모델이 부하 패턴만으로 속성 변수를 추론할 때 불확실성이 증가한다는 점을 의미하는 것이며, 특히 OOD_flag=1의 비율이 VSM보다 높게 나타난 것은 일부 세대는 학습 데이터의 통계적 구조와 현저히 다른 부하 특성을 보이는 것으로 확인되었다. 이는 DR·VPP 운영에서 위험 기반 관리가 필요함을 시사한다.

또한 수도권 지역을 자동 분류하기 위해 Extractor는 Location 예측 결과를 기반으로 Location_hat 변수를 추가하였다. 서울·경기·인천에 해당하는 세대는 Capital로, 그 외 권역은 Non-Capital로 구분하였다. 이는 전력 요금제·기상 조건·DR 운영 방식의 지역적 차이를 고려한 DR 세그먼트 설계에 활용이 가능하도록 하기위하여 구조를 확장한 것이다.

마지막으로 Extractor는 난방방식, 단열수준, 전용면적, 건축연도, 지역, confidence label, OOD_flag 등을 결합하여 DR 반응성을 정량화한 DR_segment_id를 주입하였다. DR 반응에 대한 잠재치를 Potential로 정의하고 High·Medium·Low·Uncertain의 네 그룹으로 DR 세그먼트를 분류하였다. 이는 부하 집합체 구성과 DR 이벤트 감축량의 예측, 그리고 VPP 운영에서 핵심적인 입력값으로 활용된다. 특히 Uncertain 그룹은 OOD_flag=1 또는 posterior_gap이 매우 낮은 세대로 정의된 그룹이다.

종합적으로, RSM 적용 실험을 통해 본 연구의 Extractor는 VSM에서 학습된 논리적 규칙을 실측 데이터에 성공적으로 이식할 수 있음을 확인하였다. 또한 부하 패턴을 기반으로 하는 속성 역추정이 DR 운영에 필요한 구조적 입력 정보를 자동으로 생성할 수 있음을 실증하였으며, 이로써 KR-VSM-Classifer-Extractor-DR로 이어지는 본 연구의 전체 파이프라인은 실측 기반의 DR 정책 분석을 위한 완결된 데이터셋의 제공과 함께 VPP 사업자의 DR 운용 모델을 통합하는 구조를 갖추게 되었다.

4.4.3 속성 기반 DR 타겟팅 적용성 평가

4.4.3.1 DR 세그먼트 구조 정의 및 속성 기반 군집화

본 연구는 한국형 주거용 전력 계량기에서 수집된 고해상도 부하 패턴을 기반으로, DR 운영을 체계화하기 위한 속성 기반으로 추정 규칙에서 정의한 HeatingClass_hat, InsulClass_hat, AreaClass_hat, ResionClass_hat 변수에 대하여 DR 세그먼트를 $H \times I \times A \times R$ 구조를 설계하였다.

제안된 세그먼트 구조는 난방 방식(HeatingClass, H), 단열 수준(InsulClass, I), 면적군(AreaClass, A), 지역 구분(Location_hat, R)이라는 네 가지 핵심 속성의 조합으로 이루어진다.

난방 방식(HeatingClass)은 Central, District, Individual의 세 유형으로 구분되며, 열원 시스템의 차이에 따라 DR 반응성의 민감도가 상이하다. 특히 개별난방 가구는 전기 피크 부하 변화가 건물 부하에 직접적으로 반영되기 때문에 DR 참여 잠재력이 높은 것으로 알려져 있다.

InsulClass는 건물 외피의 열관류율·기밀성 성능을 반영하여 High, Mid, Low의 세 등급으로 분류하였으며, 이는 난방·냉방 부하의 전기적 탄력성과 직결된다.

AreaClass는 추정된 전용면적(Area_m2_hat)의 분포와 군집 특성을 반영하여 A_S(소형), A_M(중형), A_L(준중형), A_XL(대형)으로 구분하였다. 분석 결과, 면적이 증가할수록 DR 발령 시 감축 가능한 부하량(kW Potential)이 증가하는 경향을 보였다. Location_hat은 전력 요금제, 기후, 냉·난방 수요 특성의 차이를 고려하여 수도권과 비수도권으로 단순화하였다.

이와 같이 [표 4-21]의 Segment 요약 Table로 H-I-A-R 네 속성의 조합을 통해 총 16개의 DR 세그먼트를 정의하였으며, 각 세그먼트는 다음과 같은 규칙으로 문자열 ID로 표현된다.

[표 4-21] DR Segment 요약 Table

DR_segment_id	count	Area_m2_mean	Year_mean	posterior_gap_mean	DR_priority_score_mean	
0	H_CD_I_HIGH_A_A_L_R_NON	7523	81.992458	3.025347	0.038765	0.234450
1	H_IND_I_HIGH_A_A_M_R_NON	4541	60.176542	3.213733	0.037044	0.267260
2	H_IND_I_HIGH_A_A_L_R_NON	3893	71.033613	3.155834	0.027377	0.228560
3	H_CD_I_HIGH_A_A_XL_R_NON	1224	87.745823	3.024653	0.021527	0.222916
4	H_CD_I_HIGH_A_A_M_R_NON	695	70.997763	3.108311	0.016064	0.202443
5	H_CD_I_HIGH_A_A_L_R_CAP	525	80.961093	2.983619	0.039302	0.233552
6	H_IND_I_HIGH_A_A_XL_R_NON	439	86.012729	3.049835	0.025042	0.218707
7	H_IND_I_HIGH_A_A_S_R_NON	372	40.834750	3.384671	0.041141	0.398306
8	H_IND_I_HIGH_A_A_M_R_CAP	306	60.679865	3.194855	0.032306	0.269988
9	H_IND_I_HIGH_A_A_L_R_CAP	259	74.565290	3.106401	0.027933	0.239765
10	H_CD_I_HIGH_A_A_M_R_CAP	84	66.960859	3.094862	0.031320	0.215091
11	H_IND_I_HIGH_A_A_XL_R_CAP	65	87.108480	2.966071	0.031834	0.240808
12	H_CD_I_HIGH_A_A_XL_R_CAP	55	88.543873	2.967776	0.024453	0.222299
13	H_IND_I_HIGH_A_A_S_R_CAP	17	43.837914	3.382168	0.042742	0.348625
14	H_CD_I_HIGH_A_A_S_R_NON	1	57.435982	3.313079	0.008345	0.188381
15	H_CD_I_HIGH_A_A_S_R_CAP	1	51.020333	3.369569	0.019833	0.206477

※ DR_segment_id=H_<Heating>_I_<Insul>_A_<Area>_R_<Location>

본 연구에서 구축한 Classifier-Extractor 파이프라인을 활용하여, RSM 기반 실측 데이터 20,000건 각각에 대해 난방·단열·면적·지역 정보를 확률 기반으로 자동 추정하였으며, 이를 바탕으로 DR 세그먼트를 자동 생성하였다. 이러한 방법은 기존 전문가 의견 기반의 정적 세분화 방식과 달리, 실제 부하 패턴에 기반한 posterior_gap 값을 활용함으로써 구조적 일관성과 통계적 안정성을 동시에 확보할 수 있다는 점에서 의미가 있다.

또한 최종적으로 도출된 DR 세그먼트는 이후 DR 시나리오 설계 단계(F3)에서 활용될 수 있도록 확장하였다. 세그먼트별 DR 비율(DR_ratio), 예상

DR 가능량(DR_potential_kW)으로 설정하여 본 연구의 DR 세그먼트 체계는 DR 기반으로 후속 연구로 진행이 필요한 VPP 가상발전소 운영 전략을 구축하는 VPP 운영 단위(Operational Unit)로 직접 적용 가능한 구조의 핵심적인 기반 정보를 제공 가능하도록 구축하였다.

4.4.3.2 세그먼트별 건물 특성 통계 집계

본 연구에서는 Extractor-RSM 결과와 RSM 시계열 부하 데이터를 결합하여, 추정된 DR 세그먼트별 건물 물리적·운영 특성을 계량적으로 분석하였다. 이를 위해 각 세그먼트 단위로 면적, 건축연도, 지역 분포, 피크부하, DR 반응성 관련 지표를 집계한 통계 테이블을 산출하였다. 이러한 분석은 속성 기반(H×I×A×R) DR 세그먼트가 실제 건물 특성과 부하 패턴을 얼마나 정확하게 반영하는지를 검증하는 목적을 갖는다.

1) 면적·연식 기반 물리적 특성 패턴

세그먼트별로 추정된 평균 면적(Area_m2_hat)과 평균 건축연도(Year_hat)을 분석한 결과, 면적 군(AreaClass)의 구분이 실제 건물 규모 차이를 명확하게 포착하고 있음을 확인하였다.

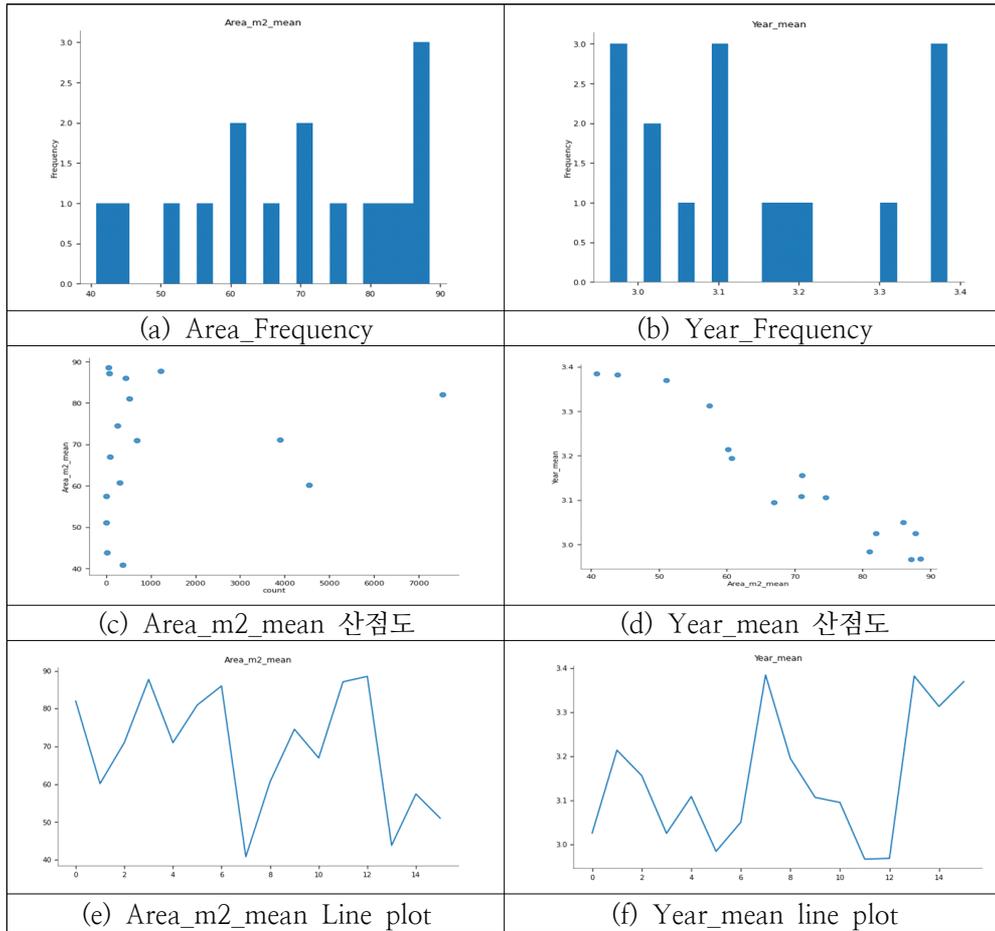
A_L(준중형) 세그먼트는 약 70~85㎡ 범위에 집중되었으며, A_M~ A_XL(중형·대형) 세그먼트는 90~110㎡ 이상의 확연한 규모 차이를 보였다.

16개 DR 세그먼트(H×I×A×R 조합)에 대해 면적(Area_m2_mean), 건축연식(Year_mean), Classifier 신뢰도(posterior_gap_mean)를 종합적으로 분석한 결과를 [표 4-22] 와 같이 제시하였다.

먼저, 세그먼트별 평균 면적의 분포는 대부분 45~90㎡ 구간에 집중되어 있으며, 특히 70~90㎡의 전형적인 중·대형 공동주택 면적대가 높은 빈도로 나타났다. 계량기 수와 면적의 관계를 살펴본 결과, 수천 개 계량기가 포함된 대형 세그먼트는 대부분 70~90㎡의 중형이상 아파트 그룹으로 구성되어 있었고, 40㎡ 이하 또는 100㎡ 이상의 특이한 면적대는 소규모 세그먼트에 한

정되어 있어 DR 포트폴리오 전체에 미치는 영향은 제한적이었다.

[표 4-22] Area·Year DR 세그먼트 분석



이는 실제 DR 대상 건물이 일반적인 아파트 규모 중심으로 구성되어 있으며, 대규모 DR 잠재량은 중·대형 세대에서 주로 발생함을 시사한다.

건축 연식(Year_mean)은 세그먼트 대부분이 3.0~3.4 범위, 즉 2000~2010년대 구축 아파트에 해당하는 중간 연대에 집중되어 나타났다.

이는 DR 실측 데이터가 1980년대 이전의 오래된 구축이나 2020년 이후의 축에 비해, 중간 연식 아파트가 주류를 이루는 실제 공동주택 스톡 구조를 반영한 결과이다. 또한 전용면적과 건축연도의 상관관계를 분석한 결과, 평균

면적이 큰 세그먼트일수록 Year_mean이 낮게 나타나는 완만한 음의 상관관계가 존재하였다. 즉, 넓은 중·대형 세대는 상대적으로 기축 비중이 높고, 소형 세대는 더욱 최근에 준공된 경향이 해석될 수 있는데, 이는 DR 세그먼트별 특성을 세분화할 때, 면적과 연식 조합에 따라 단열 성능·열손실·거주자 행태가 달라지며, DR 반응성에도 차이가 발생할 수 있다는 점을 시사한다.

또한 YearClass 기반 회귀 기댓값(Year_hat) 분석 결과, 전체 세그먼트의 건축연도는 2.9~3.2 수준, 즉 1990~2010년대에 해당하는 연식으로 수렴하였다. 이는 한국 아파트 스톡의 전형적인 건축연도의 분포와 일치한 것으로 나타났다. 이는 YearClass 추정치가 실제 건축연도 구조를 성공적으로 반영하고 있음을 시사한다.

2) 지역별 구조와 세그먼트 검증

RSM 기반 지역 추정값인 Location_hat을 활용하여 세그먼트의 지역별 구성비를 비교한 결과, 한국의 전력 소비 패턴과 일관된 구조적 차이를 확인할 수 있었다. 수도권 지역은 상대적으로 면적이 크고 피크 부하가 높은 세그먼트에서 높은 비중을 보였으며, 비수도권 지역은 난방 중심 세그먼트가 개별난방의 (H_IND) 비중이 뚜렷하게 높게 나타났다.

이는 냉·난방 수요, 도시 밀도, 주거 형태 차이 등 한국 지역 특성이 반영된 결과로, RSM과 VSM에서 모두 유사한 분포 구조가 확인되었다. 이러한 점은 세그먼트 분류가 실제 지역별 건물 특성과 부하 패턴의 구조를 충실히 재현하고 있음을 뒷받침한다.

3) 세그먼트별 DR 반응성 구조 분석

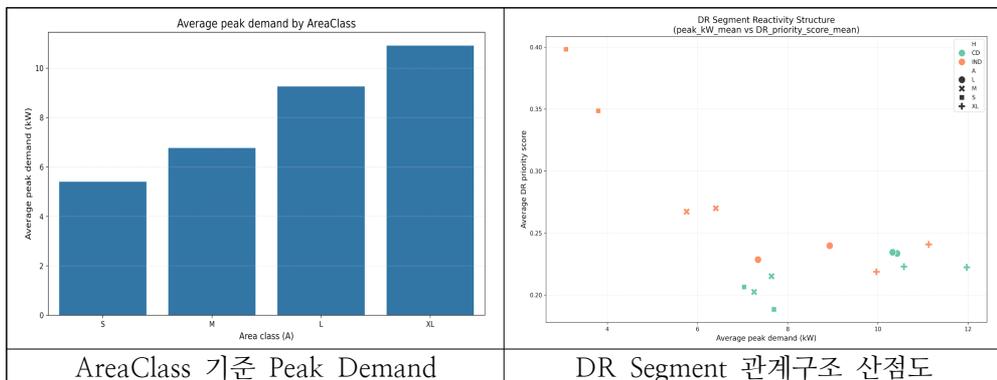
세그먼트별 평균 피크부하(peak_kW_mean)와 DR_priority_score_mean을 분석한 결과, DR 반응성(수요감축 잠재력)의 세그먼트 간 구조적 차이가 명확하며, 특히 다음 세그먼트들은 DR 유효성이 높은 클래스로 분류되었다.

(가) 중앙난방 기반(H_CD)

(나) 중·대형 면적군(A_L/A_M)

이들 세그먼트는 [표 4-23]에서와 같이 평균 peak_kW_mean이 약 10 kW 이상, DR_priority_score는 0.23~0.24 수준으로 나타나 수요 반응 프로그램 참여 시 높은 감축 효과를 기대할 수 있는 그룹으로 평가되었다. 반면 단열 성능이 낮거나 면적이 작은 세그먼트(A_S, A_M)의 경우 peak_kW_mean < 8 kW, DR_priority_score < 0.21로 나타나 DR 잠재력이 상대적으로 제한된 것으로 분석되었다.

[표 4-23] Segment별 DR 반응 분석



이러한 결과는 DR 세그먼트가 단순한 속성 조합이 아니라, 실제 부하 패턴과 물리적 특성이 결합된 전력수요 반응성의 구조화 모델임을 입증한다. 즉, 세그먼트별 DR 잠재력 차이가 논리적으로·통계적으로 뚜렷하게 나타났다.

4.4.4 DR·VPP 운영 적용성 분석

4.4.4.1 F1-입력용 계량기 구축

F1 단계는 Extractor가 역추정한 건물 속성 정보와 실측 스마트미터의 연간 전력 소비량(annual_kWh) 및 피크 부하(peak_kW)를 통합하여, VPP 운영 모델이 직접 활용할 수 있는 DR·VPP 입력용 계량기 단위 데이터셋을 구

측하는 과정이다. 이 단계는 DR 세그먼트 기반의 감축 잠재량을 산정하는데 있어 필수적인 전처리 절차이며, 전체 VPP 파이프라인에서 실제 운영 가능성을 결정짓는 핵심 모듈로 기능한다.

먼저, Extractor에서 생성된 각 계량기 별로 HeatingClass_hat, InsulClass_hat, AreaClass_hat, YearClass_hat, Location_hat 등 DR 성향을 결정하는 구조적 속성이 포함되어 있으며, posterior_gap 및 OOD_flag를 통해 분류 신뢰도 또한 부여된다. F1 단계에서는 이 파일과 실측 시간별 RSM 부하를 결합하여 각 계량기의 연간 부하(annual_kWh)와 최대 부하(peak_kW)를 계산하였다. 실행 결과에 따르면, 전체 1.7억 개의 시간 단위 부하 데이터가 병렬 스트리밍 방식으로 처리되어 총 20,000세대의 연간·피크 부하가 안정적으로 산출되었다.

이후 annual_kWh 및 peak_kW가 Extractor 결과와 병합되었다. 병합된 이 테이블은 각 세대별로 난방·단열·면적·연식·지역 기준의 DR 행동 특성, confidence, posterior_gap의 신뢰도 지표, DR_segment_id, DR_ratio, DR_potential_kW 등이 모두 포함된 완전한 형태의 DR·VPP 입력 스키마를 구성한다. 즉, DR 세그먼트 구조(H×I×A×R)가 그대로 계량기 단위까지 세분된 형태로 반영되어, 이후 F2~F3 단계의 DR 시나리오 연산이 원활히 수행될 수 있도록 설계된 것이다.

특히 DR 관점에서 중요한 변수인 DR_ratio와 DR_potential_kW는 실측 peak_kW에 기반으로 계산되며, DR 이벤트 시 이론적으로 감축 가능한 전력량을 직접적으로 나타낸다.

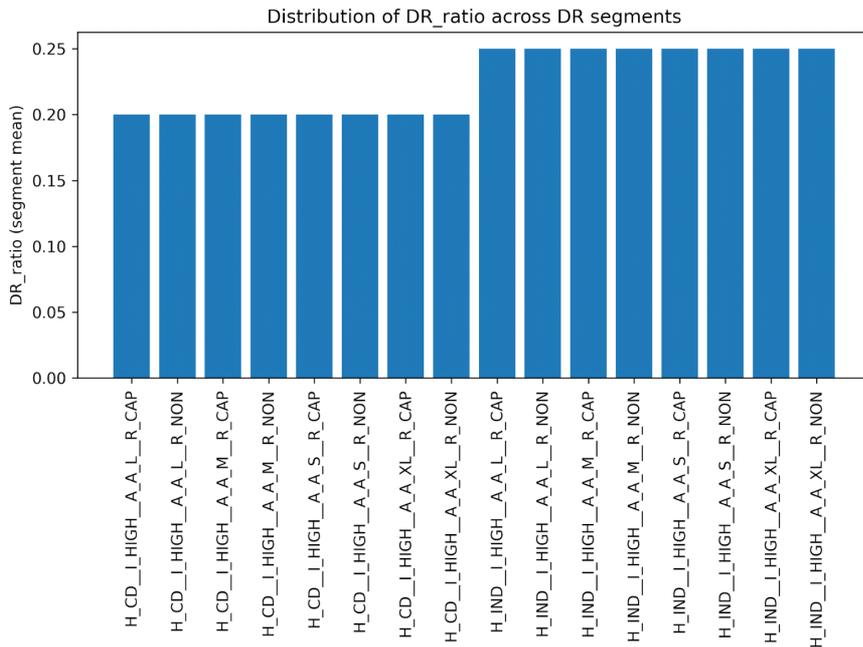
Segment별 DR 잠재량 결과에서는 많은 세대가 DR_ratio= 0.20~0.25 범위로 분포하며, peak_kW가 큰 세대일수록 DR_potential_kW가 증가하는 구조가 확인되었다. 이는 실제 DR 사업에서 세대별 감축 기여도를 산정할 때 활용 가능한 데이터 구조로, VPP 드라이버가 DR 이벤트를 구성할 때 필수적인 기초 값이다.

4.4.4.2 F2-DR 세그먼트 파라미터 생성

F2 단계는 F1에서 구축된 계량기 단위의 DR·VPP 입력 테이블을 기반으로 DR 운영의 실질적 의사결정 세그먼트 단위 파라미터를 생성하는 절차이다.

[그림 4-1]은 본 연구에서 정의한 DR 세그먼트는 HeatingClass, InsulClass, AreaClass, Location_hat 조합의 구성된 총 16개 그룹이며, 각 그룹은 건물의 속성과 에너지 소비 패턴을 반영한 DR 반응성 집단으로 해석된다. F2 단계의 목적은 이러한 세그먼트를 기반으로 DR 운영 전략에 필요한 대표 파라미터를 산출하고, DR 우선순위를 자동으로 계층화하는 데 있다.

우선, 모든 계량기를 16개 DR 세그먼트로 재분류한 뒤, 각 세그먼트에 대해 다음의 핵심 통계값을 산출하였다.



[그림 4-1] Segment 단위별 DR 비율

1) DR 반응성 지표(DR responsiveness indicators)

DR 반응성 지표로 세대별 피크 대비 DR 감축 비율의 평균을 나타내는

DR_ratio_mean과 세그먼트가 보유한 평균 감축 가능 전력량을 표시하는 DR_potential_kW_mean을 계산하였다. 실행 결과에 따르면 다수의 세그먼트에서 DR_ratio_mean이 0.20~0.25 수준에 분포하였으며, peak_kW가 높은 세그먼트일수록 DR_potential_kW_mean이 유의하게 증가하였다. 이는 DR 이벤트 수행 시 세그먼트별 감축 기여도 차이가 명확히 존재함을 보여준다.

2) annual_kWh_mean과 peak_kW_mean을 산출

이는 세그먼트가 연간 에너지 소비량과 최대부하 특성에서 어떠한 구조적 패턴을 보이는지 부하 구조(load structure) 분석을 위한 평가하는 지표이다. 예컨대, 난방이 개별이고 단열 수준이 낮은 세그먼트는 동일 면적 대비 peak_kW가 높게 나타나는 경향을 보였으며, 반대로 단열 수준이 높고 지역 난방 기반 세그먼트는 peak 부하가 비교적 완만하게 나타났다.

3) 건물 특성 파라미터를 산출

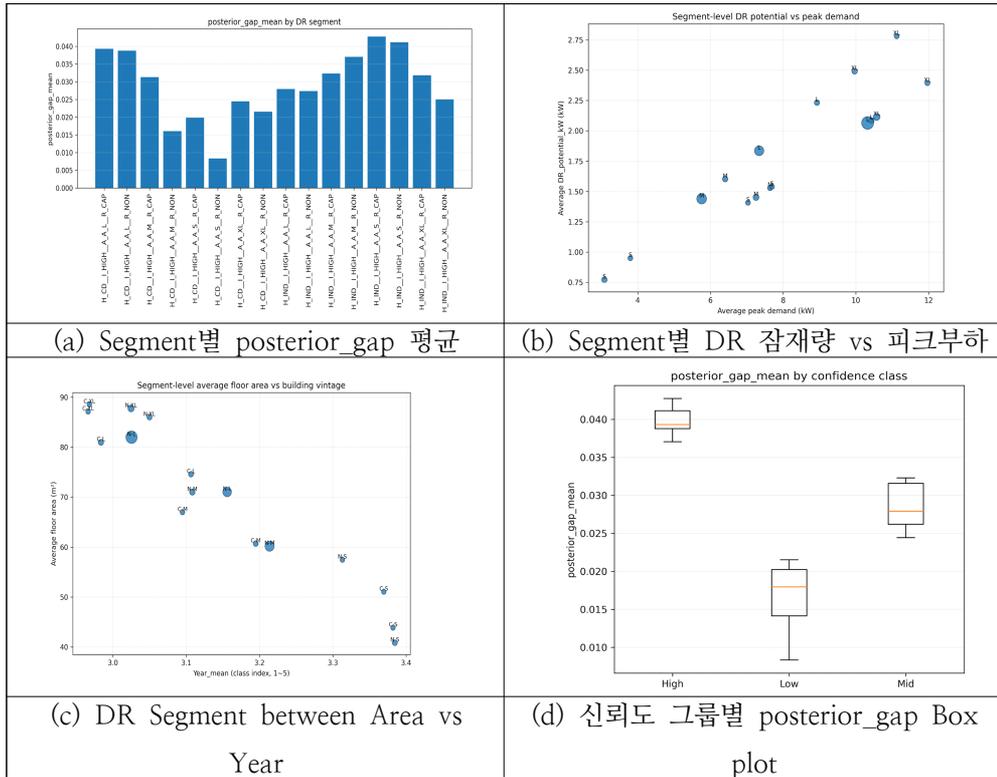
Area_m2_mean, Year_mean, Location_share의 파라미터를 세그먼트의 물리적 특성을 종합적으로 분석하였다. 또한 [표 4-24]의 (c)에서 Year_mean은 대부분 3.0~3.4에 분포하여, DR 세그먼트가 주로 2000~2015년 사이의 중간 연식 공동주택을 중심으로 구성되어 있음을 보여주었다. 또한 Area_m2_mean은 45~90m² 사이에 고르게 분포하였으나, 계량기 수가 많은 대규모 세그먼트는 대부분 70~90m² 구간에 집중된 것으로 나타났다. 이는 DR 운영에서 영향력이 큰 세그먼트가 전형적인 중대형 공동주택으로 구성됨을 의미한다.

4) posterior_gap_mean을 기반으로 한 세그먼트 신뢰도 분석

posterior_gap_mean은 Classifier가 해당 세그먼트를 얼마나 확신 있게 분류하는지 나타내는 지표이며, [표 4-24]의 (d)에서 0.02~0.04 범위에서 다양한 형태로 분포하였다. 본 연구에서는 분위값 기준으로 세그먼트를 High / Mid / Low / Uncertain 네 개 신뢰도 등급으로 계층화하였다. 즉 동일 DR 잠재

량의 분류 신뢰도가 낮은 세그먼트는 DR 자원 배분 시 보수적으로 활용해야 함을 시사한다.

[표 4-24] DR Segment 핵심 통계량 값 결과



이러한 지표들을 종합하여 DR_priority_score_mean을 산출하였고, 이를 기반으로 DR 세그먼트의 우선순위 계층화를 수행하였다.

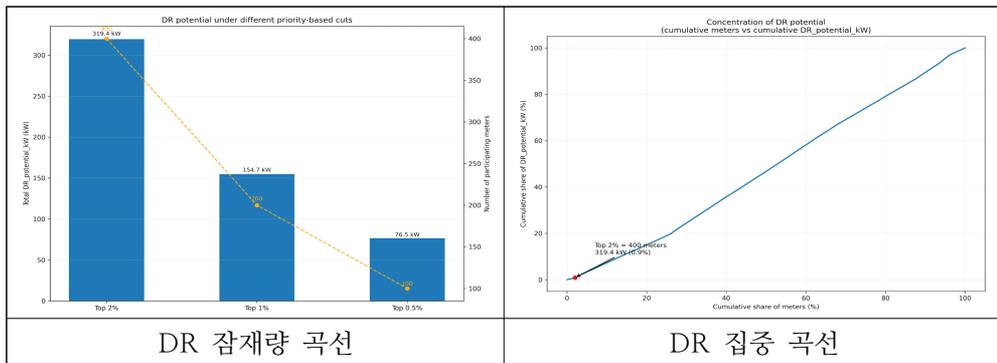
DR_priority_score_mean이 높은 세그먼트는 DR 시나리오 모두에서 주요 자원으로 활용되며, 반대로 신뢰도가 낮은 세그먼트는 DR 불확실성이 높기 때문에 예비 자원 또는 보조 자원으로 분류하는 방식이 적용되었다.

5) DR-only 시나리오: 우선순위 기반 DR 참여 구조 분석
본 연구에서는 ESS 등 외생적 요소를 배제하고 DR만을 독립적으로 적용하

는 DR-only 시나리오를 구성하였다.

DR 참여 대상자는 F2 단계에서 산출된 DR 우선순위 지표(priority_norm)를 기준으로 선정되었으며, [표 4-25]에서와 같이 본 분석에서는 priority_norm ≥ 0.7을 충족하는 세대를 DR 고우선순위 집단으로 정의하였다.

[표 4-25] DR 고우선순위 집단의 통계 결과



시뮬레이션 결과는 [표 4-26]에서와 같이 전체 20,000 가구의 DR potential은 36,819.42kW로 산정되었으며, 이 중 총 400세대가 고우선 순위 집단으로 분류되었으며, 이들 집단이 전체 DR 감축량의 9%인 319.42kW의 비중을 담당하였다.

[표 4-26] DR Top-Share-Cut

Rank	Meters	DR_Portential	Ratio
Top 0.5%	100	76.45kW	0.2%
Top 1%	200	154.73kW	0.4%
Top 2%	400	319.42kW	0.9%

두 시간의 DR 이벤트를 가정한 감축 에너지량은 605.8kWh로 나타나, 고우선 순위 세그먼트가 전체 세대의 약 2% 수준임에도 불구하고 DR 성능의

상대적으로 높은 비중을 차지함을 확인할 수 있었다.

또한 DR Cut을 줄였을 때 DR 잠재량이 떨어지는 결과로 나타났는데 DR이 상위 몇 %의 계량기에 집중되어있지 않음을 의미하고 있으며, 반대로 또한 상위 몇 %만으로도 DR이 효과적일 수도 반증하고 있는 것을 확인할 수 있었다. 이는 DR 반응성이 높은 특정 세그먼트의 기여도가 매우 크며, 세그먼트 기반 DR 타겟팅이 DR 성능을 실질적으로 좌우함을 의미한다.

6) DR 구성 세그먼트의 기여도 및 구조적 특징

DR-only 시나리오에서 높은 DR 기여도를 보인 세그먼트는 공통적으로

난방 방식이 INDIV, 단열 성능이 HIGH, 면적이 A_S 또는 A_M 구간, 지역이 비수도권(Non-Capital)인 속성을 가지는 것으로 나타났다. 이는 Extractor가 역추정한 건물 속성과 DR 잠재량이 실제 DR 시나리오에서도 그대로 반영되며, 속성 기반 DR 세그먼트화가 실측 데이터에서도 타당함을 보여주는 결과이다.

세그먼트별 DR 성능을 비교한 결과, 일부 세그먼트는 세대 수와 관계없이 매우 높은 DR 잠재량(DR_potential_kW)을 보였다. 반면 일부 세그먼트는 세대 수가 많음에도 DR 반응성이 낮은 특징을 나타냈다. 이러한 특징은 DR 운영 시 개별 세대를 균일하게 반영하는 기존 방식보다, 세그먼트 단위의 속성 기반 접근이 DR의 안정성을 향상시킬 수 있음을 시사한다.

7) DR·VPP 운영 전략 및 정책적 시사점

(가) DR 성능은 예측 신뢰도(posterior_gap)와 밀접한 관련이 있다.

DR 고반응 세그먼트는 공통적으로 posterior_gap_mean이 높은 집단으로 분류되었다. 이는 Classifier가 해당 세대의 건물 속성에 대하여 높은 신뢰도로 추정한 경우에 실제 DR 감축량의 오차를 줄이면서 우수한 DR 반응을 이끌어 낼 수 있음을 의미한다. 즉 DR 대상자 선정은 단순히 피크부하나 연간 부하가 큰 세대를 선택하는 것보다, posterior_gap이 높은 세대를 우선 선정하는 방식이 훨씬 과학적이고 효율적임을 나타낸다.

(나) 소수의 고반응 세그먼트가 DR 반응 성과를 높인다.

전체 20,000세대 중 약 2%에 해당하는 400세대가 전체 DR 감축량의 319.42 kW를 기여하는 것으로 나타난 실험의 결과는 세그먼트 기반 DR 타겟팅이 매우 높은 효율성을 가진다는 근거로 해석된다. 이는 DR 자원을 전 세대를 대상으로 분산 적용하는 기존 방식의 비효율성을 극복하고, DR 고반응 세그먼트를 중심으로 전략적으로 자원을 배분하는 것이 DR 사업의 성과를 극대화할 수 있음을 뜻한다.

(다) DR-only 조건에서도 우선순위 기반 DR 정책이 효과적이다.

priority_norm 기준을 강화하여 상위 2%→상위 1% 또는 상위 0.5%로 집중적으로 설정할수록 참여 세대 수는 감소하지만 감축량 감소 폭은 제한적이었다. 이는 고반응 세그먼트 중심의 집중형 DR 정책이 운영비를 절감함과 동시에 DR 목표 달성의 균형을 동시에 만족할 수 있음을 나타낸다.

F3 시나리오 실험은 세그먼트 기반의 DR·VPP 구조가 실측 데이터에서도 매우 안정적으로 작동하였다. F3 실험을 통하여 DR 고반응 세그먼트를 추출하고 이를 기반으로 ① 속성 기반의 DR 타겟팅의 우수성, ② posterior 기반 신뢰도 활용의 효과성, ③ 운영비 절감형 DR 정책의 실현 가능성을 모두 실증적으로 확인할 수 있었다.

따라서 본 연구에서 구축한 Extractor-Segment-posterior_gap 기반의 DR 구조는 VPP 운영에서 DR 참여자 선정과 인센티브의 설계, 그리고 운영 정책 정교화 등 다양한 실무 영역에서 적용이 가능한 프레임워크로의 활용 가능성을 확인하였다.

4.4.5 posterior_gap 기반 임계값 신뢰도 평가

본 절에서는 Extractor-F2 단계에서 구축된 총 16개의 DR 세그먼트를 대상으로, 평균_전용면적(Area_m²_mean), 평균_건축연도(Year_mean), 그리고 Classifier 기반의 신뢰도 지표인 posterior_gap_mean을 중심으로 세그먼트 간 구조적 차이를 해석하였다. 이러한 분석은 DR·VPP 운영 과정에서 세그먼트

트 단위의 타겟팅 전략이 얼마나 타당하고 실질적으로 유효한지 평가하면서, 세그먼트별 에너지 특성의 편차와 추정 신뢰도의 구조를 정량적으로 파악하는 데 그 목적이 있다.

4.4.5.1 세그먼트별 평균 면적의 구조적 분포와 DR 영향력

실험 결과에서 계량기 수가 많은 상위 세그먼트 7,523세대는 평균 면적이 80㎡ 내외로 나타났는데, 이는 국내 공동주택의 전형적인 중·대형 면적대의 구조적 분포와 일치한다. 반면 40~60㎡ 수준의 소형 면적 세그먼트도 일부 존재하나, 해당 세그먼트들은 대부분 포함되는 세대수가 적어 전체 DR 포트폴리오에서 차지하는 영향력은 제한적이었다.

이러한 결과는 실제 DR·VPP 운영에서 실질적 파급효과를 갖는 주요 세그먼트가 70~90㎡의 중·대형 면적대에 집중되어 있음을 의미하며, DR·VPP의 타겟팅에서도 해당 규모의 세그먼트를 우선 고려하는 것이 타당함을 시사한다.

4.4.5.2 세그먼트별 연식의 구조적 특성 및 상관성 해석

연식 클래스의 평균값은 대부분 3.0~3.3 범위에 분포하고 있다. 이것은 DR 세그먼트가 주로 2000~2015년 사이에 준공된 중간 연식의 공동주택을 중심으로 구성되고 있음을 나타내고 있는 것이다. 반면, 1980년대 이전의 구축(Year_mean ≈ 1~2)이나 2020년 이후의 최신 단지(Year_mean ≥ 4.5)를 중심으로 하는 세그먼트는 거의 존재하지 않아, DR 대상 단지의 연식 분포가 특정 구간에 집중되어 있음을 보여준다.

또한 Area-Year 산점도 분석 결과, 평균 면적이 커질수록 연식값이 다소 감소하는 약한 음의 상관관계가 관찰되었다. 즉, 중·대형 면적 세그먼트는 상대적으로 연식이 오래된 구축 비중이 더 높고, 소형 면적 세그먼트는 비교적 최근 지어진 주택 비중이 높다는 구조적 경향이 존재하였다. 그러나 이 상관성은 절대적이라기보다는 완만한 패턴 수준으로 판단되며 면적과 연식의 상호

작용이 DR 세그먼트의 전체 구조를 결정하는 핵심 요인은 아님을 나타낸다.

결과적으로 DR 세그먼트는 대부분 3~4의 중간 연식에 집중된 구조를 갖고 있으며, 이는 DR·VPP 운영에서 연식 변수 자체가 세그먼트 차별성보다는 세그먼트 내 안정성을 제공하는 보조적 인자로 작용함을 설명하는 것이다.

4.4.5.3 posterior_gap_mean에 기반한 세그먼트 신뢰도 평가

세그먼트별 posterior_gap_mean은 [표 4-19]에서와 같이 약 0.016~0.039 구간에 분포하였으며, 이는 본 연구의 Step II에서 정의한 신뢰도 기준(high \geq 0.036, 0.02338 \leq gap(mid) $<$ 0.03603, low $<$ 0.023)에 따르면 대부분의 세그먼트가 중간의 신뢰도 영역에 해당하였다. 일부 세그먼트는 0.038 이상을 기록하여 고신뢰도 그룹으로 분류되었고, 반대로 0.02 전후의 낮은 값을 갖는 세그먼트는 Classifier가 상대적으로 불확실하게 판단하는 구간으로 분류되었다.

Year_mean - posterior_gap_mean 산점도 분석에서는 뚜렷한 선형 관계가 나타나지 않았다. 즉, 연식이 최근일수록 분류 신뢰도가 증가하거나, 오래된 기축일수록 신뢰도가 낮은 단일한 구조는 존재하지 않았다. 이는 세대의 연식 자체보다 난방 방식, 단열 수준, 면적, 거주 행태 등이 결정하는 시계열 부하 패턴의 구조적 차이가 분류 신뢰도에 더 큰 영향을 미침을 의미하고 있다.

이러한 결과는 DR·VPP 운영에서 연식을 기반으로 세그먼트 신뢰도를 판단하기보다는, posterior_gap_mean을 직접 신뢰도의 핵심 지표로 사용하는 것이 훨씬 합리적이라는 보여주고 있는 것이다.

4.4.6 DR 적용 관점에서의 역추정 신뢰도 평가

본 절에서는 앞선 4.4.2~4.4.5절에서 제시된 VSM 기반 규칙 학습, RSM 실측 데이터 적용, DR 세그먼트 생성 및 DR-only VPP 시나리오 분석 결과를 종합하여, 입력 역추정(Extractor) 모델의 성능을 평가한다. 아울러 Generator - Classifier - Extractor - DR·VPP로 이어지는 본 연구의 전체 파이

프라인이 갖는 확장 가능성과 함께 향후 실제 VPP 사업자의 DR 운영에 대한 차세대 모델과의 연계 방향을 제시하고자 한다.

우선, 본 연구에서 구축한 Extractor는 VSM metadata-Classifier-posterior 구조를 결합하여 규칙을 학습한 뒤, 동일한 규칙을 실측 RSM에 그대로 적용하도록 설계되었다. Extractor Step에 따르면, InsulClass와 AreaClass의 경우 VSM과 RSM 모두에서 약 0.90 수준의 높은 정확도를 유지하였으며, 이 두 변수는 도메인 전이에도 강건한 성능을 보였다.

HeatingClass와 YearClass의 경우 RSM에서 클래스 분포가 왜곡되거나 혼합 패턴이 존재 함으로 posterior_gap을 기반으로 정의한 confidence 구조가 VSM과 RSM 사이에서 유사한 스케일과 순서를 유지하였다. 이는 Elastic Calibration과 posterior 기반 신뢰도 구조가 VSM·RSM 간 분포 차이를 완전히 제거하지는 못하더라도, Classifier의 VSM의 신뢰 수준을 RSM에서도 일관되게 해석할 수 있는 기반을 제공한다는 점에서 의미 있는 성과로 평가할 수 있다.

한편 HeatingClass는 분포의 차이가 Extractor 성능에 영향을 미치는 것으로 관찰되었다. 실제 RSM에서는 전기난방과 가스보일러, 보조 난방기 등 다양한 혼합 난방 패턴이 공존하고 있고, VSM 기반 Classifier를 RSM에 적용할 때 개별난방(INDIV) 클래스 비중을 과대 추정하는 경향을 보였다.

이는 Classifier가 국내 공동주택의 난방시스템이 전기부하에 의존적이지 않은 이유로 부하 패턴만을 이용하여 INDIV와 DIST·CENT 간의 경계를 명확히 구분하기 어렵다는 구조적 한계를 반영하고 있는 것이다. 이로 인하여 난방 특성이 VSM-RSM 간 domain gap이 가장 크게 발생하는 파라미터임을 시사한다. 반면 단일 성능, 전용면적, 건축연도, 지역과 같은 변수들은 VSM과 RSM에서의 분포 차이가 상대적으로 작고, DR 세그먼트 수준에서도 유사한 구조적 패턴이 유지되었다. 따라서 난방 방식은 향후 VSM, Classifier, Extractor 전반에 걸친 보정이 우선적으로 요구되는 파라미터인 반면, 나머지 변수들은 실측 기반의 DR·VPP 분석에서 활용성이 높은 설명력을 보이고 있다고 판단하였다.

Extractor 기반으로 구성된 DR 세그먼트와 DR-only 시나리오 분석 결과

에서는 DR·VPP 정책의 적용 측면에서 몇 가지 중요한 시사점을 제공하고 있다. 첫째, DR 고반응 세그먼트는 전체 세대의 약 2%에 불과하지만, 전체 DR 감축량에서 차지하는 비중은 상대적으로 높게 나타났다. 이는 세그먼트 기반의 DR 타겟팅이 DR 운영 효율성을 크게 향상시킬 수 있음을 보여준다.

둘째, posterior_gap이 높은 세대일수록 DR 감축량의 예측 오차가 작고 실제 감축 성능도 우수한 것으로 확인되었다. 단순히 피크 부하의 크기나 연간 전력 소비량이 높은 세대를 기준으로 DR 참여자를 선정하는 기존 방식보다, posterior 기반 신뢰도를 핵심 지표로 활용하는 것이 더 합리적인 전략임을 의미한다.

셋째, 면적·단열·난방·지역의 조합에 따라 세그먼트별 DR 잠재량이 구조적으로 달라지는 현상이 관측되었으며, 이를 통해 DR 가격 차등제, DR 제안량 추정, 세대별 참여 전략의 설계 등에 직접적인 입력 정보를 제공할 수 있는 기반을 마련하였다고 할 수 있다.

넷째, VSM 기반의 설계-Extractor 기반의 역추정-RSM 기반 DR 시뮬레이션이 하나의 일관된 체계로 연결된다는 점이 실증적으로 검증되면서, VSM을 VPP 사업자가 DR 반응에 대한 테스트베드로 활용할 수 있는 가능성이 확인되었다. 즉, Extractor-Segment 기반의 분석은 기존의 개별 세대 단위 DR 설계보다 정교하게 예측이 가능하며, 실제 VPP 운영 정책 수립 측면에서도 높은 실무적 활용성을 보여주고 있는 것이다.

추가로, 본 연구는 추정 불확실성을 정량적으로 관리하기 위해 posterior_gap과 OOD_flag 두 가지 지표를 도입하였다⁴⁰³⁾⁴⁰⁴⁾. 실험 결과, posterior_gap을 기준으로 정의한 high/mid/low confidence 구간은 RSM에서도 안정적으로 유지되었고, OOD_flag=1인 세대는 DR 반응성이 낮거나 예측 오차가 상대적으로 큰 경향이 관찰되었다.

이에 따라 DR 참여 우선순위는 다음과 같은 방정식으로 정의할 수 있다.

403) Kim, D., Lee, S., & Choi, J. (2021). Out-of-distribution detection in time-series domain: A novel approach. *Journal of Machine Learning Research*, 22(1), 1-28.

404) Hastie, T., Tibshirani, R., & Friedman, J., p.127-129.

$$\text{DR 참여 우선순위} = \text{DR}_{\text{potential}} \times \text{posterior_gap} \times (1 - \text{OOD_flag})$$

여기서,

posterior_gap : 신뢰도 지수

1 - OOD_flag : 정상성

OOD_flag : 이상치 지수

$\text{DR}_{\text{potential},i} [\text{kW}] = P_{\text{peak},i} \times r_i$ 로 정의할 수 있다.

여기서,

$P_{\text{peak},i}$: i 번째 계량기(세대)의 피크 전력

r_i : Peak 대비 감축 가능 비율(F2-DR 세그먼트 파라미터에서 생성)

이상의 방정식으로 신뢰성 기반의 DR 포트폴리오 전략을 제안할 수 있으며, 이는 DR·VPP 운영 과정에서 세대별 불확실성을 체계적으로 관리할 수 있게 해 준다는 점에서 실무적으로도 중요한 기여로 볼 수 있다.

4.5 수요 반응 운용 관점에서의 적용성 종합 평가

본 연구에서 개발한 BSEM 기반의 한국형 KR-VSM 데이터셋과 공동주택을 중심으로 하는 DR·VPP 연계 프레임워크는 국내 전력시장에서 계시 요금제를 대비한 주택용 요금제의 다양한 실증적 활용 가능성을 제시하였다.

기존의 AMI에서 수집되는 RSM 데이터는 아파트 관리사무소별로 관리되고 있는 민간 영역의 데이터로 분류되면서 데이터 보안과 개인정보로 취급되는 특성상 데이터의 비식별화 처리 등의 문제로 데이터 확보와 취급의 제약이 있고, 이로 인해 연구자와 기업이 자유롭게 활용하기 어려운 구조적 한계를 가지고 있다.

이러한 한계를 극복하기 위해 본 연구에서는 건축물 Archetype의 확률 변수 기반의 입력을 체계화하고, 탄력적 보정을 통한 EnergyPlus-CREST의 시뮬레이션을 통하여 생성된 KR-VSM 데이터셋을 분류기-추출기 연계 구조를 통합한 한국형 KR-VSM 데이터셋을 개발하여 실용성을 극대화하였다.

[표 4-27] 연구 성과의 학문적 유용성 평가

연구 성과	연구 성과의 유용성
KR-VSM 데이터셋	<ul style="list-style-type: none"> - 국내 공동주택의 지역·건물·기기·점유자·기후 특성을 통합한 대규모 시뮬레이션 기반의 가상스마트미터 데이터셋으로, 기존 RSM 데이터의 확보 문제를 구조적으로 해결 - PMF/UPD 확률 변수, CREST 점유자 스케줄, White-Box EnergyPlus 시뮬레이션을 통하여 생성된 VSM 데이터를 r·RMSE·MAE·MBE의 통계적 정합 지표를 통하여 RSM과 검증을 통해 가상 데이터의 시간 정합성과 현실적 적합성을 정량적으로 검증한 연구 인프라를 구축
Classifier	<ul style="list-style-type: none"> - 시계열 기반의 204D 특징 벡터와 RSM-VSM 연결 구조를 통해 기존 단순 군집 또는 블랙박스 예측 모델의 한계를 극복하고 수요 패턴의 구조적 해석이 가능한 판별 모델을 구축 - LDA→ Random Forest→ Scenario 비교라는 체계적 분류 프레임워크를 제시하고 Elastic Scaling의 보정을 통해 가상 데이터셋의 학습 안정성을 실험을 통해 실증
Extractor	<ul style="list-style-type: none"> - VSM 기반의 규칙 학습→ RSM에 학습 규칙 적용→ DR 세그먼트 생성으로 이어지는 완전한 역추적 구조를 정립 - VSM 데이터를 기반으로 분류된 결과를 DR·VPP 운영 시나리오까지 일괄 분석 프레임워크를 완성했다는 점에서 학술적 의의 높음.
DR·VPP 운영 시뮬레이션	<p>DR·VPP 시나리오를 posterior_gap의 신뢰도 평가와 OOD_flag의 이상치 제거를 통하여 VPP 운영의 효율성과 DR·VPP의 운영 안정성 평가의 이론적 확장 가능성을 제시</p>

본 연구를 통해 얻어진 주요 성과물의 학문적 유용성은 [표 4-27]과 같이 평가할 수 있으며, 이러한 유용성을 기반으로 수요 반응 시장에 대한 활용 측면에서 다음과 같은 의의가 있다.

첫째, 본 연구에서 구축한 VSM 데이터셋은 국내 국민의 절대다수가 거주하고 있는 아파트를 중심으로 하는 공동주택 건물의 물리적 특성, 거주자 특성, 기상 변수, 내부 부하 요인을 종합적으로 반영한 시계열 데이터를 구성하였다. 이 방법은 지역별, 건물유형별, 계절별, 시간대별로 전력수요 특성 분석을 가능하게 하는 데이터셋의 기초 자료를 제공하였다. 이는 기존의 측정 데

이터가 부족한 지역에서도 KR-VSM 데이터셋을 활용하여 데이터 부족 지역에 대한 보충 학습 데이터로 활용될 수 있어 전력 수요예측에 대한 일반화의 성능 향상에 기여할 수 있다.

둘째, DR 및 VPP 실증 시뮬레이션 분야에 적용될 가능성을 확인하였다. 본 연구는 단순한 수요예측을 넘어 분류기-추출기-VPP·DR의 프레임워크를 통해 DR 세그먼트를 생성하고 DR 참여 부하의 추정을 기반으로 VPP 사업자가 DER 자원의 효율적인 운영 시나리오로 확장이 가능하도록 설계되었다. 특히, 탄성 보정을 통해 RSM과 VSM 간의 패턴 및 탄성 차이를 보정함으로써, 사전 시뮬레이션 실험에서 실제 DR 참여 시 발생 가능한 DR 응답 곡선까지 테스트베드를 구축하였다. 이러한 프레임워크는 DR 운영자, VPP 사업자, 배전 시스템 운영자가 가상의 환경에서 수요예측을 사전 검증을 통하여 자원 운영의 위험을 평가하고 DR 반응을 가상의 환경에서 시뮬레이션을 통해 활용할 수 있는 기반을 구축한 것이다.

셋째, 에너지 관련 정책, 기후 시나리오 등을 반영하여 전력 시장 재편을 위한 분석 도구로서의 활용 가능성이 높은 데이터셋으로 구축하였다. 기존 국내 전력 관련 수요 정책은 고객의 과거 전력 사용량 데이터인 CBL(Customer Baseline Load)을 기본으로 산정하고 있으며⁴⁰⁵⁾ 특히 아파트를 중심으로 하는 주택용(고압) 시장은 민간 영역으로 구분되어 한국전력 인증 AMI가 확산되어 있지 않은 관계로 주로 시범 사업을 통해 DR 설계가 되고 있다. 또한 주택용 전력시장의 경우 산업용에서 적용되는 계시 요금제가 아닌 누진 요금제를 기본으로 적용되어 있어서 DR 확산에 저해 요인으로 작용되고 있는 것이 현실이다. 그러나 본 연구에서 개발한 KR-VSM 기반 프레임워크는 대규모 가상 가구 집단을 대상으로 다양한 수요예측과 DR 반응, 그리고 탄소 배출량 감축을 기반으로 중·장기적 시뮬레이션 실험을 가능하게 하는 기반 데이터로 연구의 기초 자료로 활용이 가능할 것이다⁴⁰⁶⁾.

넷째, AI를 기반으로 하는 전력 데이터 학습을 위한 공공 데이터셋 구축에 활용될 수 있다. 본 연구에서 사용된 한국형 공동주택을 대상으로 하는

405) 고종민 · 양일권 · 송재주. (2009). 수요측 단기 전력소비패턴 예측을 위한 평균 및 시계열 분석방법 연구. 『전기학회논문지』. 58(1), 1-13.

406) 이윤경, 전계서, p.6-7.

KR-VSM 데이터셋은 개인정보의 침해 위험 없이 시간적, 공간적 해상도, 건물 특성, 점유자 특성을 동시에 포함하는 고차원 시계열 데이터를 제공함으로써 향후 생성형 AI를 기반으로 수요예측, 이상 탐지, DR 반응의 학습 데이터로 활용될 수 있다. 이는 RSM 데이터 부족과 누진요금 정책으로 인해 정체되어 주택용 DR 시장에서 다양한 AI 기반의 수요예측과 관련된 연구를 확산시키는 데 기여할 수 있을 것이다⁴⁰⁷⁾.

407) Wang, Z., & Hong, T., 전계논문, p.21-22.

V. 결론

5.1 연구 요약 및 주요 성과

본 연구는 Neale et al.(2020)의 QSBSEM 모델을 한국 공동주택의 에너지 이용 특성에 맞게 재구성하고, 이를 AMI에서 실측된 RSM 데이터와 결합하여 한국형 공동주택의 KR-VSM 데이터셋을 개발하고 이를 DR·VPP 적용성까지 검증한 일관된 프레임워크를 제시했다는 점에서 의의가 있다. 특히 RSM 데이터의 근본적 한계인 건물 메타데이터 부재 문제를 해결하기 위해, 국가 통계 기반 확률 분포를 활용하여 20만 세대 규모의 한국형 KR-VSM 데이터셋을 White-Box 시뮬레이션인 프로그램인 EnergyPlus 솔루션을 기반으로 Bottom-up 방식의 건물 에너지 시뮬레이션 모델로 구현하였다.

KR-VSM의 생성은 건축연도, 단열 수준, 전용면적, 난방 방식, 지역 등 핵심 건물 속성을 실제 건물 스톡의 확률적 구조에 맞추어 재현하고, 물리적·통계적 일관성을 확보한 대규모 가상 공동주택 스톡을 구현함으로써, 향후 가상 발전 사업자의 DR 운용과 에너지 관련 수요예측 연구의 기초 데이터를 도출할 수 있었다.

또한 본 연구는 연간·계절별·월별·주별·시간대별 정보를 통합한 204D 특징 벡터를 기반으로 Classifier를 구축하여 VSM과 RSM 모두에 적용하였고, 속성의 추정에서 건축물의 단열, 전용면적 등 주요 파라미터는 약 0.90 수준의 높은 분류 성능을 유지했으며, 도메인 차이가 큰 변수에 대해서도 posterior_gap 기반의 신뢰도 판별을 통하여 분류모델의 붕괴 없이 VSM에서 RSM으로 일관성 있게 전이되는 안정성을 실증하였다.

이어서 Extractor 단계에서는 다양한 매핑 구조를 결합해 RSM 실측 데이터의 특성을 반영한 역추정 규칙 체계를 구축하였다. 특히 posterior_gap과 OOD_flag 지수를 활용해 추정된 속성의 신뢰성을 평가하고 이상치를 제거함으로써, 단순 속성 복원 수준을 넘어 DR 반응을 예측하는 안정적인 입력 체계를 완성하였다.

DR 실증분석에서는 난방, 단열, 면적, 지역의 파라미터를 조합한 16개 세그먼트를 정의하여 DR 시나리오를 수행한 결과, 전체의 약 2%에 해당하는 고반응 세그먼트가 전체 DR 감축량의 약 9%를 점유하였다. 이 세그먼트는 posterior_gap 기반의 신뢰도 및 DR 반응 통계 지표에서도 우월한 성능을 보였으며, 이는 DR 참여자 선정에 전력 사용량이 많은 고객에 의존하는 방식보다 속성 기반 세그먼트와 신뢰도 정보의 병행이 효율적임을 나타내고 있다.

종합적으로 본 연구는 한국형 VSM-Classifier-Extractor-DR·VPP로 이어지는 통합 프레임워크를 구축하였으며, RSM 기반으로 구축된 KR-VSM 데이터셋을 활용하여 VPP 사업자가 AMI 데이터의 메타데이터 부재로 인하여 기존 주택용 전력 시장의 수요예측의 한계를 극복하고 세밀한 DR 반응에 대한 운용 전략을 수립할 수 있는 구조적 대안을 제시하였다. 이는 건물 속성을 부하 패턴으로부터 재구성하고 DR 반응성을 정량적으로 분석하는 연구 방향의 대안으로 제시된 것으로, RSM 기반의 검증을 통해 구축된 VSM 데이터셋을 DR·VPP 테스트베드 구축까지 실증함으로써 향후 계시별 요금제 도입을 앞둔 주택용(고압) 시장을 타겟으로 하는 중소형 VPP 사업자의 사업성 검증의 틀로 활용될 수 있을 것이다⁴⁰⁸⁾. 결국 본 연구는 국내 공동주택 기반의 수요예측에 대한 또 다른 연구의 방향성을 제시함과 동시에 스마트그리드 및 에너지정책 분야의 실증 기반을 확장하는 또 다른 출발점의 연구 결과를 도출한 것이라 할 수 있을 것이다.

5.2 향후 연구 방향

본 연구는 기존 QSBSEM 연구에서 제안된 Generator-Classifier-Extractor 아키텍처를 국내 공동주택의 주택용(고압) 데이터에 적용하고, 측정된 RSM 데이터와 연계하여 VSM 기반 DR·VPP 운영 프레임워크를 실증적으로 검증했다는 점에서 의미가 있다. 그러나 이를 더욱 정교하고 확장 가능한 DR·VPP 시스템으로 발전시키기 위해서는 추가적인 연구가 필요하다.

첫째, 현재의 Extractor 규칙은 주로 특징 기반 매핑에 기반하고 있으므로,

408) 심규한, 윤동섭, 김산, 주성관, 전계논문, p.1-2.

향후 연구에서는 측정된 DR 이벤트 응답, 장비 제어 이력, 그리고 RTP 및 CPP와 같은 실시간 가격 신호를 통합하여 예측, 추정 및 제어를 위한 통일된 구조로 규칙을 고도화할 필요성이 있다⁴⁰⁹⁾. 이를 통해 Classifier-Extractor-VPP는 단순한 정보 추출 시스템을 넘어 DR 제어에 대한 의사결정을 포괄하는 통합 플랫폼으로 발전할 수 있을 것이다.

둘째, 본 연구는 시간 단위로 RSM 분석에 초점을 맞추었지만, 향후 연구에서는 DR 성능을 보다 정확하게 예측하기 위해 15분 또는 5분 간격의 고해상도 AMI 데이터 해독, 실제 EMS 제어 로그, 세대별 실내 온도, 난방 및 온수 시스템 운영 이력 등의 데이터를 수집해야 한다⁴¹⁰⁾. 이러한 데이터를 결합하면 Extractor 규칙의 외삽 능력이 향상되고, 전력회사가 실제로 운영하는 DR 프로그램 설계에 더욱 직접적으로 적용할 수 있게 된다.

셋째, 현재 모델에 지역별 난방 및 냉방 온도 설정, HDD/CDD와 같은 외기 조건, 가구 수, 가전제품 및 에어컨 보급률과 같은 파라미터를 추가하면 DR 부문의 설명력을 더욱 향상시킬 수 있다. 이는 동일한 건물 내에서도 가구 구성 및 사용 방식에 따라 DR 반응성이 크게 달라질 수 있다는 점을 모델 수준에서 확인할 수 있도록 구현하였다.

넷째, Elastic Scaling은 월별 및 계절별 평균 부하 수준을 조정하는 데 효과적이었지만, 난방 시스템과 같이 VSM과 RSM 간에 분포 편차가 큰 변수의 경우에는 분포 차이를 완전히 보정하지는 못하였다. 향후 연구에서는 클래스별 조건부 보정, 분포의 정규화, 분포 정렬 손실 등을 포함한 방식으로 Elastic Scaling의 버전을 업그레이드하여 도입할 필요성이 있으며, 이를 통해 VSM 기반 정책 실험 결과가 실제 RSM 데이터와 더욱 높은 일관성과 VSM이 RSM의 LoadShape와의 정합성을 높일 수 있도록 할 수 있다.

전반적으로, 본 연구는 입력 역추정 모델이 VSM 기반 규칙을 RSM으로 성공적으로 변환할 수 있으며, DR·VPP 운영 전략을 도출하는 데 실질적으로 유용함을 보여준다.

409) 허윤지, 김지효 (2023). 『미국과 캐나다의 주택용 계시별 요금제 동향 및 국내 시사점』. 동향과 분석, 한국에너지경제연구원.

410) Fernandopulle, J., Fang, Y., Datta, M., Nutkani, I., & Vahidnia, A., 전계논문, p.8-11.

참 고 문 헌

1. 국내문헌

- 강성구, 김신규. (2014). 보안 침해사고 대응을 위한 스마트미터 로그 생성 및 수집 방안. 『정보보호학회지』. 24(5). 66-74.
- 고종민 · 양일권 · 송재주. (2009). 수요측 단기 전력소비패턴 예측을 위한 평균 및 시계열 분석방법 연구. 『전기학회논문지』. 58(1), 1-13.
- 구인석, 이성희, 손중찬, 이동희. (2020). AMI 적용 아파트의 전기사용 패턴 기반 국내외 전기요금제 분석. 『한국산학기술학회논문지』, 21(12), 52-59.
<https://doi.org/10.5762/KAIS.2020.21.12.52>.
- 강진경, 김성임, 진혜선, 최보혜, 임재한, 송승영. (2016) 공동주택 세대단위의 연간 전기기기 에너지사용 특성 분석. 『대한설비공학회』, 하계학술대회 발표 논문집, 2016(6), 693-696.
- 김동식 정영모, 정범진.(2022). 딥러닝 기반 주택용 계시별요금제 전환 시 전기요금 변화 예측. 『대한전기학회』, 하계학술대회논문집. 2022(7), 1837-1838.
- 김재춘, 노광현. (2025). 건물스톡에너지모델 기반 수요반응모델 설계를 위한 가상스마트미터 데이터셋 분석. 『조명·전기설비학회논문지』, 39(5), 355-364.
<https://doi.org/10.5207/JIEIE.2025.39.5.355-364>.
- 김혜원. (2018). “에너지 절감효과와 경제성을 고려한 국내 지역별 공동주택의 외피 성능 분석”. 연세대학교 대학원, 건축공학과, 석사학위논문.
- 강수진. (2024). “민·관수 보급 사업 종료된 AMI 다음 스텝은… 다양한 서비스 ‘컬래버레이션.’” 『일렉트릭타임즈』. Retrieved from <https://www.electimes.com/news/articleView.html?idxno=337390>.
- 김유란, 홍원화, 서윤규, 전규엽. (2011) 공동주택 가족구성원별 전력 소비 성향에 관한 연구. 한국 주거학회 논문집. 22(6) 43-50

- 김은환. (2023). “수요반응시장(DR)의 환경변화와 기회 요인”. 『전기저널』, 전기정보관
Retrieved from <http://www.keaj.kr/news/articleView.html?idxno=4900>.
- 김말희, 허태욱, 이일우. (2023). 에너지 관리 기술의 미래 발전방향과 전망에 대한
연구. 『정보와 통신』, 한국통신학회. 40(12), 3-10.
- 김영욱. (2023). “EU의 스마트미터 보급정책 및 활용 현황 분석. 정책 분석”. 『전기저
널』, Retrieved from <http://www.keaj.kr/news/articleView.html?idxno=5265>.
- 건설산업과 국가건설기준센터. (2023). 『건축물의 에너지절약 설계기준 해설서(제
2023-104호)』. 국토교통부·한국에너지공단.
- 건설산업과. (2022). 『기계설비 기술기준 매뉴얼(최종본)』. 국토교통부.
- 건설안전과. (2012). 『건축물 에너지 절약을 위한 창호 설계 가이드라인』. 국토교
통부(국토해양부).
- 공간데이터마켓(2024). “2024년 10월 아파트단지 기준정보”. LX한국토지정보공사(주)
펜타코드. Retrieved from https://geomarket.kr/user/dataset/view.do?data_sn=25.
- 관측정책과. (2022). 『지상기상관측지침』. 기상청.
- 국토교통 통계누리. (2023). “아파트주거환경통계 현황(2023)”. 국토교통부. Retrieved from
[https://stat.molit.go.kr/portal/cate/statView.do?hRsId=33&hFormId
=&hDivEng=&month_yn=](https://stat.molit.go.kr/portal/cate/statView.do?hRsId=33&hFormId=&hDivEng=&month_yn=).
- 국가에너지통계 종합정보시스템. (2023). “KESIS 국가에너지정보시스템 - 에너지총조사
결과표”. 산업통상자원부·에너지경제연구원. Retrieved from [https://
kesis.keei.re.kr/board.es?mid=a10201030100&bid=0003](https://kesis.keei.re.kr/board.es?mid=a10201030100&bid=0003).
- 국가에너지통계 종합정보시스템. (2023). “KESIS 국가에너지정보시스템 -가정부문 에너
지추이 및 현황”. 산업통상자원부·에너지경제연구원. Retrieved from [https://
kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=v
iew](https://kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=view).
- 국가기후데이터센터. (2024). 『지상 종관 기상 관측(ASOS) 자료』, 기상청, 기상자료
개방포털. Retrieved from <https://data.kma.go.kr/data/grnd/selectAsosRltnList.do>.

- 기술신용보증기금. (n.d.). 『전략 제품 현황 분석: 스마트미터(AMI)』. 전략제품현황 보고서. Retrieved from [https:// tb.kibo.or.kr/ktbs/board/tech-trend/tech_trend.do?mode=download&articleNo=563&attachNo=1441](https://tb.kibo.or.kr/ktbs/board/tech-trend/tech_trend.do?mode=download&articleNo=563&attachNo=1441).
- 남수현, 이현. (2023). 『가구에너지패널조사 연구』. 한국에너지경제연구원 (KESIS). Retrieved from https://kesis.keei.re.kr/board.es?mid=a10306000000&bid=0060&list_no=2116&act=view.
- 박완기 (2014). 『고효율 건물 에너지관리기술 동향 및 시사점』. 동향과 분석, 한국에너지경제연구원.
- 박원석, 윤재욱. (2003). 신축공동주택의 기밀성능 실측에 관한 연구. 『한국생태환경건축학회논문집』, 3(3), 43-52.
- 수요정책실. (2025). 『KEA 에너지 이슈 브리핑』. Vol.276, 2025.10.13. 한국에너지공단.
- 수요정책실. (2025). 『2025 KEA 에너지 편람』. 한국에너지공단.
- 성장기술실 미래전략팀. (2011). 『수요반응 제도의 개념 및 현황』. 한국전력거래소.
- 송용현. (2025). 『에너지 수요 전망 모형 구축』. 국회예산정책처 연구용역보고서, 사단법인 넥스트.
- 사회통계국 사회통계기획과. (2025). 『2024년 생활시간조사 결과』. 통계청.
- 신이레, 윤상후. (2016). 전력 수요 예측을 위한 기상 정보 활용성 평가. 『한국데이터정보과학회지』. 27(6), 1601-1607. <http://dx.doi.org/10.7465/jkdi.2016.27.6.1601>.
- 심규한, 윤동섭, 김산, 주성관. (2019). 주택용 계시별 요금제 도입을 위한 전력 소비 패턴 분석. 『대한전기학회』. 전력기술부문회 추계학술대회 논문집, 2019(10). 271-272.
- 에너지수급전망연구실. (2024). 기온 시나리오에 따른 2024년 여름 주택용 전력 수요 추정 및 전기요금 영향 분석. 『에너지 브리프』. 2024(7). 한국에너지경제연구원.

- 에너지 정보. (2024). “서울특별시 가구당 에너지사용정보”. 서울특별시 녹색 에너지과 . Retrieved from <https://energyinfo.seoul.go.kr/energy/energyConsumption?menu-id=Z111000>.
- 윤태연, 강재성. (2015). 난방 방식이 가구의 겨울철 난방비 지출에 미치는 영향 분석: 도시가스 개별난방과 지역난방 방식을 중심으로. 『에너지경제연구』, 14(2), 243-272.
- 융합기상. (n.d) “6.1 전력기상지수 조회서비스”. 기상청 API 허브, 기상청. Retrieved from <https://apihub.kma.go.kr>.
- 이경호, 주홍진, 안영섭, 이왕제. (2024). TRNSYS 환경에서 모델 단순화와 학습을 통한 월별 연간 에너지 성능 평가: 태양광 적용 사무용 공공건물의 히트펌프 냉·난방. 『한국건축환경설비학회논문집』. 18(3), 192-210. <https://doi.org/10.22696/jkiaeb.20240017>.
- 이나은, 안병립, 정학근, 김종훈, 장철용. (2013). 창호의 성능 및 건물의 창면적비에 따른 기존 단독주택의 창호 리모델링 방안 연구. 『한국생태환경건축학회논문집』, 13(3), 71-80. <http://dx.doi.org/10.12813/kieae.2013.13.3.071>.
- 이루다, 강은호, 윤종호, 김동수, 김진휘, 원종연. (2023). 한국형 공동주택 에너지 시뮬레이션 모델링을 위한 건물 형태 및 에너지 사용 특성 분석. 『한국태양에너지학회 논문집』. 43(4), 43-54. <https://doi.org/10.7836/kses.2023.43.4.043>.
- 이상현. (2014). “공동주택의 에너지 소비량 실태 조사를 통한 예측 모형 제시에 관한 연구”. 중앙대학교 대학원 석사학위논문.
- 이성근. (2022). 가정부문 용도별 에너지 소비량 및 소급 추정에 관한 연구. 『에너지경제연구원』. ISBN 978-89-5505-272-6.
- 이수진, 김유정, 진혜선, 서인에, 김성임, 하수연, 송승영. (2020). 건물, 거주자, 기기 사용 특성이 공동주택 용도별 에너지사용량에 미치는 영향 분석. 『대한건축학회 논문집』. 36(6), 153-164. <https://doi.org/10.5659/JAIK.2020.36.6.153>.
- 이여름, 박영훈, 김광천. (2024). 아파트 내부 검침망을 활용한 AMI 활성화 방안 연구. 『한국산학기술학회 춘계 학술발표논문집』. 487-492.

- 이윤경. (2016). 국내·외 수요자원시장 동향과 시사점. 『ENERGY FOCUS』. 동향초점, 2016년 여름호, 에너지경제연구원. 34-60.
- 이윤석. (2022). 건물에너지관리시스템: 에너지 절감수요에 맞춘 기술융합 및 틈새시장 개척. 『ASTI Market Insight』. 한국과학기술정보연구원(KISTI). 2022(036), 01-06. Retrieved from <https://repository.kisti.re.kr/>.
- 이지영, 선영규, 이선민, 김수현, 김영규, 이원섭, 심이삭, 김진영. (2019). GRU 기반 전력 사용량 예측을 적용한 스마트미터기 구현. 『한국인공지능학회』 19(5), 93-99. <https://doi.org/10.7236/JIIBC.2019.19.5.93>.
- 이진숙, 최종문, 김소연, 박지영. (2010). 공동주택의 실내 공간별 조명적용현황 연구. 『대한건축학회논문집』. 26(12), 45-56.
- 이희원, 김덕환. (2023). LSTM을 활용한 일반 주택의 실내 온도 예측 사례 연구. 『한국차세대컴퓨팅학회 논문지』, 19(2), 24-34.
- 임경엽, 윤종현, 신우철, 박재성, 김광식. (2011). 실측자료 기반의 공동주택 시간별 전력 소비 패턴 분석 연구. 『한국에너지학회논문집』, 20(3), 127-132.
- 장철용, 한혜심, 이진숙. (2010). 차폐계수와 창면적비에 따른 공동주택의 건물 에너지 효율 등급 평가. 『한국태양에너지학회 논문집』. 30(5), 38-45.
- 전력 AMI 데이터 유통 플랫폼. (n.d). “시간대별 전력 사용량 기상결합 AI 학습용 데이터”. (재)한국스마트그리드사업단. Retrieved from <https://www.energydata.co.kr> & <https://www.smartgrid.or.kr/>.
- 정재용, 김동식, 정범진, 정영모 (2021). AMI 데이터 기반 주택용 전력 누진 및 계시별 요금제 분석. 『전자공학회논문지』, 58(9), 66-74. <https://doi.org/10.5573/ieic.2021.58.9.66>
- 조재운, 이영생, 송호승, 최영애, 조태희. (2023). 전력기상지수를 활용한 공동주택의 단기 전력 수요 예측. 『대한전기학회』 2023 하계학술대회 논문집, 2023(7), 1828-1830.
- 조사관리국 인구총조사과. (2025). 『2024년 인구주택총조사 결과』. 통계청.
- 조성진, 윤태연. (2016). 주택용 전력 수요 계절별 패턴 분석과 시사점. 『정책 이슈 페이퍼』. 16-11. 에너지경제연구원.

- 주택건설공급과. (2025). “도시형 생활 주택 면적 제한 완화를 위한 『주택법 시행령』 등 개정안 시행 안내 보도 자료”, 2025.1.20. 국토교통부.
- 주택건설공급과. (2023). 『건축물의 에너지절약 설계기준 해설서』. 국토교통부.
- 주택건설공급과. (2025). 『에너지 절약형 친환경주택 건설기준 해설서』. 국토교통부.
- 전력산업정책과. (2025). 『제11차 전력수급기본계획(2024 - 2038)』. 산업통상자원부.
- K-apt 공동주택관리 정보시스템. (2024). “K-apt 공동주택관리 정보시스템 운영 통계 및 아파트 관리비 자료”. <https://www.k-apt.go.kr/web/main/index.do>. 국토교통부·한국부동산원.
- KESIS. (2023). “13차 2022년 기준(13차) 가구에너지패널조사 연구(수정:2025.05.28.)”. 국가에너지통계 종합정보시스템. Retrieved from https://kesis.keei.re.kr/usePurpose/form.es?mid=a10202030100&bid=0009&list_no=2100&seq=1&cnt=1. 산업통상자원부·에너지경제연구원.
- 한국에너지공단 건물에너지실. (2017). 『에너지절약계획서 실무 길라잡이』. 국토교통부·한국에너지공단.
- 한국학중앙연구원. (n.d.). “계절”. 한국민족문화대백과사전. Retrieved from <https://encykorea.aks.ac.kr>.
- 허윤지, 김지효 (2023). 『미국과 캐나다의 주택용 계시별 요금제 동향 및 국내 시사점』. 동향과 분석, 한국에너지경제연구원.
- 홍석재, 임정택, 함경선, 김태형. (2023). 『공동주택단지 내 BTM 에너지 공유서비스 경제성 평가 시스템 설계와 구현』. 정보과학회 컴퓨팅의 실제, 29(1), 15 - 24. <https://doi.org/10.5626/KTCP.2023.29.1.15>.
- 홍승연, 송승영.(2021). 머신러닝을 활용한 공동주택 세대에서의 전기기기 에너지 사용량 분석 및 추정. 『대한건축학회』, 2021년도 추계학술대회논문집. 41(2). 860-863.
- 홍승연, 전서운, 송승영. (2020). 통계 기법을 활용한 공동 주택 세대에서의 용도별 에너지사용량 분석 및 추정: 온수, 전기기기, 조명 에너지사용량을 중심으로. 『대한건축학회』, 2020년 추계학술발표대회 논문집. 40(2). 704 - 707.

2. 국외문헌

- ASHRAE. (2001). ASHRAE Handbook: ASHRAE TECHNOLOGY COMMITTEES AND TASK GROUPS, 2001 ASHRAE FOUNDATIONS HANDBOOK SI.
- ASHRAE. (2014). ASHRAE Guideline 14-2014: Measurement of Energy, Demand, and Water Savings. Atlanta, GA: ASHRAE. ISSN 1049-894X.
- ASHRAE. (2021). ASHRAE Handbook: Fundamentals SI Edition, ISSN 1523-7230.
- Agrawal, A. (2020). Know the best evaluation metrics for your regression model! Towards Data Science. Data Science Blogathon. Article Url – <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/>.
- Balakrishnama, S., & Ganapathiraju, A. (n.d.). LINEAR DISCRIMINANT ANALYSIS – A BRIEF TUTORIAL. Institute for Signal and Information Processing. Mississippi State University.
- Birol F. (2024). World Energy Outlook 2024. Paris:IEA. International Energy Agency (IEA). <http://www.iea.org>.
- Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(6), 1636. <https://doi.org/10.3390/en11071636>
- Breiman, L. (2001). Random Forests. Machine Learning Technical Report. *Machine Learning*, 45, 5–32.
- Bulbaai, R. R., & Halman, J. I. M. (2023). Standby energy consumption

- and saving potentials in the residential sector in tropical areas: The Caribbean island Curaçao as a case study. *Energy Efficiency*, 16, 88. <https://doi.org/10.1007/s12053-023-10170-0>.
- Centinel. (2024). “The difference between energy (kWh) and power (kW)”. Retrieved from <https://www.centinelre.com/>.
- Chen, C., Liaw, A., & Breiman, L. (n.d.). Using Random Forest to Learn Imbalanced Data., Department of Statistics, UC Berkeley
- Chen, Y., Wei, W., Song, C., Ren, Z., & Deng, Z. (2023). Rapid building energy modeling using prototype model and automatic model calibration for retrofit analysis with uncertainty. *Buildings*, 13(6), 1427. <https://doi.org/10.3390/buildings13061427>.
- Chow, G., C. & Lin, A.-I. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4) 372–375. <https://www.jstor.org/stable/1928739>.
- Chong, A., Gu, Y. & Jia, H. (2021). Calibrating building energy simulation models: A review of the basics to guide future work. *Energy & Buildings*, 253. <https://doi.org/10.1016/j.enbuild.2021.111533>.
- Claeys, R., Cleenwerck, R., Knockaert, J., & Desmet, J. (2023). Stochastic generation of residential load profiles with realistic variability based on wavelet-decomposed smart meter data. *Applied Energy*, 350, 121750. <https://doi.org/10.1016/j.apenergy.2023.121750>.
- Coakley, D., Raftery, P., & Keane, M. (2014). Enhanced parameter estimation (EPE): A new calibration methodology for building energy models. *Energy and Buildings*, 74, 1–12.
- Coralogix. (2023). Root Mean Square Error (RMSE): The cornerstone for

evaluating regression models. Coralogix AI Blog. Retrieved from <https://coralogix.com/ai-blog/root-mean-square-error-rmse-the-cornerstone-for-evaluating-regression-models/>.

- CREST. (2017). CREST – Centre for Renewable Energy and Sustainable Technologies: Performance Report. South Pacific Passive House Conference (SPPHC 2017). South West College / Queen’s University Belfast.
- Dahlström, L. (2023). Archetype identification in Urban Building Energy Modeling: Research gaps and method development. Uppsala University, Licentiate thesis.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification (2nd ed.). Wiley-Interscience.
- Dufour, A.B. (2005). Linear Discriminant Analysis — Course Notes. University Lecture Notes : Fiche TD avec le logiciel.
- Emami, P., Sahu, A., & Graf, P. (2024). BuildingsBench: A large-scale dataset of 900K buildings and benchmark for short-term load forecasting. 37th Conference NeurIPS 2023 Datasets and Benchmarks Track. Preprint.
- Fernandopulle, J., Fang, Y., Datta, M., Nutkani, I., & Vahidnia, A. (2023). Planning and Operation of Community Microgrids: Technical, Economic, Policy and Regulatory Aspects. RMIT University. DOI: 10.13140/RG.2.2.14173.69609/1.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
- Heiple, S., & Sailor, D. J. (2008). Using building energy simulation and geospatial modeling techniques to determine high resolution

- building sector energy consumption profiles. *Energy and Buildings*, 40(8), 1426–1436.
- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15, 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>.
- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York: Springer. DOI 10.1007/978-0-387-78189-1.
- Jeong, H.-C., Jang, M., Kim, T., & Joo, S.-K. (2021). Clustering of load profiles of residential customers using extreme points and Demographic Characteristics. *Electronics* 2021.10(290). <https://doi.org/10.3390/electronics10030290>.
- Johra, H., Schaffer, MM., Chaudhary, G., Kazmi, H. D., Dréau, J. L., & Petersen. S. (2023). What metrics does the building energy performance community use to compare dynamic models? *Proceedings of the 18th IBPSA Conference Shanghai, China, Sept. 4–6, 2023*, 0861–0868. <https://doi.org/10.26868/25222708.2023.1309>.
- Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3), 358–386. <https://doi.org/10.1007/s10115-004-0154-9>.
- Kim, Y.-S, Kim M.-K, Fu, N., Liu, JJ., Wang, J. & Srebric, J. (2025). Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models. *Sustainable Cities and Society*, 118, 105570. <https://doi-org.libproxy.hansung.ac.kr/10.1016/j.scs.2024.105570>.
- Kim, D., Lee, S., & Choi, J. (2021). Out-of-distribution detection in

- time-series domain: A novel approach. *Journal of Machine Learning Research*, 22(1), 1–28.
- Lepot, M., Aubin, J.-B., & Clemens, F. H. L. R. (2017). Interpolation in time series: An introductory overview of existing methods, their performance criteria and uncertainty assessment. *Water* 2017, 9(769), 796. <https://doi.org/10.3390/w9100796>.
- Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, 1(2), 169–173. <https://www.jstor.org/stable/1391858>.
- Loga, T., Stein, B., Diefenbach, N., & Born, R. (2016). TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. *Energy and Buildings*, 132, 4–12.
- McCarron, B. (2017). Passive House Research & Development Centre (CREST): How to meet net zero carbon with Passivhaus. PSouth Pacific Passive House Conference (SPPHC 2017), Christchurch, New Zealand.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). Introduction to information retrieval. Cambridge University Press.
- Mata, E., Sasic Kalagasidis, A., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment*, 81, 270–282. <https://doi.org/10.1016/j.buildenv.2014.06.013>
- Mondal, S. (2023). “Understanding MAE, MSE, and RMSE: Key metrics in machine learning. DEV Community”. Retrieved from <https://dev.to/search?utf8=%E2%9C%93&q=Understanding+MAE%2>

C+MSE%2C+and+RMSE%3A+Key+Metrics+in+Machine+Learning.

- Murata, A., Kondou, Y., Hailin, M., & Nomura, N. (2016). Electricity consumption in Japanese households: Empirical analysis based on detailed household data. *Energy Sustainability & Society*, 6(1), 1–13.
- Neale, A. (2021). Development of a Bottom-up White-box Residential Building Stock Energy Model. PhD Dissertation, Polytechnique Montréal.
- Neale, A., Kummert, M., & Bernier, M. (2022). Discriminant analysis classification of residential electricity smart meter data. *Energy and Buildings*, 262, 111980. <https://doi.org/10.1016/j.enbuild.2021.111823>.
- Neale, A., Kummert, M., & Bernier, M. (2020). Development of a stochastic virtual smart meter data set for a residential building stock – methodology and sample data. *Journal of Building Performance Simulation*, 13(5), 583–605. <https://doi.org/10.1080/19401493.2020.1800096>.
- Neale, A., Kummert, M., & Bernier, M. (2019). Linear discriminant analysis for classification of a large virtual smart meter data set with known building parameters. *Energy and Buildings*, Proceedings of the 16th IBPSA Conference. Rome, Italy, Sept. 2–4, 2019. <https://doi.org/10.26868/25222708.2019.210568>.
- Neale, A., Kummert, M., & Bernier, M. (2018). Generator: A stochastic virtual smart meter data generation model for residential building stock characterization. Proceedings of eSim 2018 (IBPSA-Canada), 65–72. ISBN 978-2-921145-88-6.
- Martin, C., Tebbit & J., Bernstein, L. (2023). Lighting, Cooking, Electrical Appliances and Incidental Heat Losses in the Home Energy Model: FHS Assessment Wrapper (HEMFHS-TP-04).

- Home Energy Model: FHS assessment version: HEM: FHS v0.18.
London: UK Government, Department for Energy Security & Net Zero
OneBuilding.org. (2024). “Repository of Building Simulation Climate
Data”. Retrieved from <https://climate.onebuilding.org>
- Raudys, S., & Duin, R. P. W. (1997). Expected classification error of the
Fisher linear classifier with pseudo-inverse covariance matrix.
Elsevier Science B.V., Pattern Recognition Letters 19. 385–392.
PII: S0167- 865598.00016-6.
- Reyna, J., Wilson, E., Parker, A., Satre-Meloy, A., Egerter, A., Bianchi,
C., Praprost, M., Speake, A., Liu, L., Horsey, R., Dahlhausen,
M., CaraDonna, C., & Rothgeb, S. (2023). U.S. building stock
characterization study: A national typology for decarbonizing U.S.
buildings. National Renewable Energy Laboratory (NREL).
- Sengupta, M., Habte, A., Gueymard, C., Wilbert, S., & Renné, D.
(2017). Best Practices Handbook for the Collection and Use of
Solar Resource Data for Solar Energy Applications (2nd ed.).
Golden, CO: National Renewable Energy Laboratory (NREL).
Technical Report NREL/TP-5D00-68886.
- Seo, Y.-K., & Hong, W.-H. (2014). Constructing electricity load profile and
formulating load pattern for urban apartment in Korea. *Energy and
Buildings*, 78, 222–230. <http://dx.doi.org/10.1016/j.enbuild.2014.03.007>.
- Sokol, J., Cerezo Davila, C., & Reinhart, C. F. (2017). Validation of a
Bayesian-based method for defining residential archetypes in
urban building energy models. *Energy and Buildings* 134, 134, 11–24.
- Solar Energy Laboratory; TRANSSOLAR; CSTB; TESS. (2017). TRNSYS
18: A Transient System Simulation Program – Updates in

Version 18. Madison, WI: University of Wisconsin.

StackExchange. (2012). “Mean absolute error OR root mean squared error? CrossValidated (StackExchange Statistics)”. Retrieved from <https://stats.stackexchange.com/questions/48267/mean-absolute-error-or-root-mean-squared-error>.

StackExchange Mathematics Community. (2019). “Why John Tukey set 1.5 IQR to detect outliers”. from: Mathematics StackExchange. Retrieved from <https://math.stackexchange.com/questions/966331/why-john-tukey-set-1-5-iqr-to-detect-outliers-instead-of-1-or-2https://math.stackexchange.com/questions/966331/why-john-tukey-set-1-5-iqr-to-detect-outliers-instead-of-1-or-2r>.

Sun, K., Hong, T., & Chen, Y. (2016). A pattern-based automated approach to building energy model calibration. *Energy and Buildings*, 117, 287–297. <https://doi.org/10.1016/j.enbuild.2015.10.020>

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 30(2), 169–190. <https://doi.org/10.3233/AIC-170729>.

TimeAndDate.com. (2024). “Korea Standard Time (KST)”. Retrieved from <https://www.timeanddate.com/time/change/south-korea/seoul>.

Tomrukçu, G., Kızıldağ, H., Avgan, G., Dal, A. Ö., Sağlam, N. G., Özdemir, K., & Ashrafian, T. (2014). A systematic approach to manual calibration and validation of building energy simulation models. *Energy and Buildings*, 80, 415–427. <https://ssrn.com/abstract=4488363>.

Tomrukcu, G., Kizildag, H., & Avgan, G. (2023). A systematic approach to manual calibration and validation of building energy simulation models. *Smart and Sustainable Built Environment Emerald Publishing*

Limited 2046-6099 . DOI 10.1108/SASBE-10-2023-0296

- U.S. Department of Energy.(2019). Input for Output – Variable Dictionary Reports: EnergyPlus 9.1 Input Output Reference. Retrieved from Big Ladder Software. <https://bigladdersoftware.com/epx/docs/9-1/input-output-reference/input-for-output.html#variable-dictionary-reports>.
- U.S. Department of Energy. (2021). EnergyPlus™ Version 9.5.0 Documentation: EnergyPlus Essentials. Build : de239b2e5f..
- U.S. Department of Energy. (2023). EnergyPlus™ Version 23.1.0 Documentation: Output Details and Examples. Build : 87ed9199d3.
- U.S. Department of Energy. (2024). EnergyPlus™ Version 24.1.0 Documentation: Input Output Reference. Build : 9d7789a3ac.
- Wang, Z., & Hong, T. (2020). Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). *Energy Technologies Area 2020*, 224, 110299.
- Wang, Z., Hong, T., & Piette, A. (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy* 263, 114683. <https://doi.org/10.1016/j.apenergy.2020.114683>.
- Wiatros-Motyka, M., Fulghum, N., Jones, D., Altieri, K., Black, R., Broadbent, H., Bruce-Lockhart, C., Ewen, M., MacDonald, P., & Rangelova, K. (2024). *Global electricity review 2024*. Ember.
- Wikipedia. (2024). “Solar irradiance”. Retrieved from https://en.wikipedia.org/wiki/Solar_irradiance.
- Wikipedia. (2024). “Canada”. Retrieved from <https://ko.wikipedia.org/wiki/캐나다>.
- Wikipedia. (2024). “대한민국”. Retrieved from <https://ko.wikipedia.org/wiki/대한민국>.
- Zhou, N., McNeil, M., Fridley, D., Lin, J., & Price, L. (2018). Residential

electricity consumption patterns in China. *Energy Sustainability & Society*,
8(1), 1–15.

ABSTRACT

Development of a VSM Dataset for Korean Apartment
Buildings Based on a Building Stock Energy Model and
Evaluation of Its Applicability to Demand Response

Kim, Jaechun

Major in Smart Convergence Product

Dept. of Smart Convergence Consulting

The Graduate School

Hansung University

In recent years, the expansion of data center construction due to the spread of artificial intelligence services and the continued increase in the number of electric vehicles have led to a rapid increase in electricity demand, particularly in the Seoul metropolitan area. However, the expansion of power generation facilities and investment in the power grid has not kept up with this demand. In particular, the nationwide expansion of renewable energy sources, including solar power generation, is exacerbating grid instability due to their intermittent output and uncertainly in power generation forecasts. In fact, the recent series of

solar power generation curtailments suggests the need for a shift away from supply-side-centric electricity policy.

Meanwhile, in the domestic residential electricity market, the importance of accurately understanding the hourly load characteristics of each household and demand response based on this data is increasing as the market prepares for the transition from a progressive rate system to a time-of-use rate system. In this context, smart meters installed in apartment complexes are attracting attention as a key infrastructure that enables the estimation of building properties and classification of load patterns based on each household's measured electricity usage, referred to as real smart meter (RSM) data. However, these smart meters are currently classified as private and are not utilized as demand response data in the electricity demand market.

Therefore, to make up for the limitations of smart meter data in apartment complexes, this study developed a virtual smart meter dataset based on the Building Stock Energy Model (BSEM), called KR-VSM, tailored to the Korean apartment environment. Based on this dataset, the primary goal of this study is to establish a framework that enables VPPs to develop integrated strategies ranging from demand forecasting to demand response (DR) operations.

The development of the Korean VSM dataset extended and reconfigured the Generator-Classifier-Extractor framework of QSFBSM proposed by Neale et al. (2018, 2020, 2022) to suit the Korean apartment complex environment. First, in the Generator stage, objective national statistical data, climate data, and building-specific physical data related to Korean apartment complexes were combined to generate 21 parameters based on probability in a bottom-up manner. These variables were then fed into EnergyPlus, a white-box simulator, to create a KR-VSM dataset representing for 200,000 apartment households.

Next, time-series load patterns, which precisely reflect annual, seasonal, monthly, and hourly electricity usage patterns, were used to classify the nonlinear KR-VSM dataset using machine learning-based linear & nonlinear classifiers, including LDA and Random Forest. The VSM dataset classified in this way was compared with the RSM data for 480,118 households, based on the RSM data for 897,322 households secured from the AMI data distribution platform. These meters met the connectivity and missing rate criteria of at least 80% of the 8,760 annual hours observed. This process validated the reverse estimation(Extractor) model.

The reverse estimation results showed that building insulation and floor space exhibited high tracking stability in both VSM and RSM. For heating and construction year, the reliability structure identified in the VSM was confirmed to be preserved in the RSM data.

Based on the Classifier-Extractor results, this study constructed 16 DR segments and conducted an empirical analysis using RSM data from 20,000 households. The results showed that highly responsive segments across all households significantly contributed to DR reduction effects, demonstrating that reliability is closely linked to DR responsiveness and prediction error. This suggests that targeting DR participants based on attribute-based reliability, rather than focusing on households with high peak electricity consumption, is more effective when selecting DR participants.

In summary, this study systematically developed a Korean KR-VSM dataset focused on apartment complexes and empirically verified its validity using RSM data. Furthermore, by assigning building parameter and occupant' activity values to the validated KR-VSM dataset, this study proposed a demand response value chain based on KR-VSM, predicting DR responsiveness and designing VPP operation strategies.

These findings are academically significant in that they provide a validated data infrastructure for a Korean apartment-focused VSM dataset, which can be practically applied across various domains, such as DR·VPP policy design, regional energy demand management, and building energy management.

【Keywords】 Virtual Smart Meter(VSM), Real Smart Meter(RSM), Building Stock Energy Model(BSEM), Demand Response(DR), Classifier, Extractor, DR Segment, LDA(Linear Discriminant Analysis), PMF(Probability Mass Function), UPD(Uniform Probability Distribution)