

지역난방 열 수요와 날씨 빅데이터를 융합한 열 수요 예측

원자력발전소

1. 분석 배경 및 목표

최근 공동주택의 공급과 거주 비중이 증가하면서, 이에 따라 지역난방 수요 역시 지속적으로 확대되고 있다. 특히 공동주택의 열 수요는 외부 기상 요인의 영향을 크게 받기 때문에, 기온, 습도, 풍속 등 다양한 기상 변수에 기반한 정밀한 열수요 예측이 필수적이다. 이러한 예측은 에너지의 효율적인 생산과 공급, 그리고 피크 수요 시의 안정적인 대응에 핵심적인 역할을 한다. 기상청과 한국지역난방공사에서 제공하는 기상 데이터 및 열수요 빅데이터를 융합 분석함으로써, 열 수요의 패턴을 정량적으로 파악하고, 이에 기반한 예측 모델을 수립할 수 있다. 이는 냉난방 에너지의 과잉 공급이나 공급 부족을 사전에 방지하고, 장기적으로는 에너지 비용 절감 및 탄소 배출 감소에도 기여할 수 있다.

본 보고서에서는 2024년 한 해 동안의 열 수요를 예측하고, 그 결과를 바탕으로 향후 기간에 대한 수요 전망의 기초 자료로 활용하고자 한다. 이를 통해 지역난방 에너지 운영의 효율성과 예측 가능성을 높이는 데 기여하는 것을 목표로 한다.

2. 분석 데이터 및 전처리

2.1 데이터 정의

변수명	설명	변수명	설명
tm	날짜 및 시각	m_hr1	1시간 강수량
branch_ID	지사명	hm	정시 상대 습도
ta	정시기온	si	ASOS 일사량
wd	정시 10분 평균 풍향	ta_chi	체감온도
ws	정시 10분 평균 풍속	heat_demand	시간당 지사별 열 공급량
m_day	해당 시간까지의 일 강수량		

< 표 1 > 지점별 기상 관측 데이터 항목별 구성

2.2 데이터 전처리 - 결측치 보간

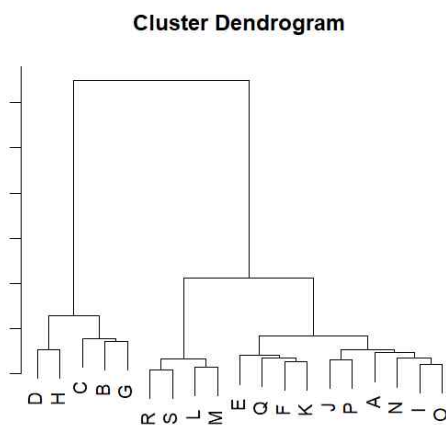
데이터전처리 단계에서 기온(ta, 12 997개), 풍향(wd, 18 815개), 풍속(ws, 18 815개), 일강수량(rn_day, 18 626개), 시간 강수량(rn_hr1, 19 154개), 상대습도(hm, 39 717개), 일사량(si, 232 922개), 체감온도(ta_chi, 20개), 열수요(heat_demand, 23개) 변수에 -9.9 또는 -99.9로 표시된 결측치가 확인되었다. m_hr1 변수에 이미 날씨 정보가 포함되어 있으므로 중복되는 m_day는 분석에서 제외하였다. 모든 결측 코드는 NaN으로 변환한 뒤, 시계열 특성을 유지하도록 보간을 수행하였다.

1 ~ 2시간 연속 결측인 경우 결측 구간의 직전·직후 실측치를 이용해 선형 보간(linear interpolation)을 적용하였다. 일사량(si)의 구조적 결측(MNAR) 경우 일조가 없는 08:00 이전과 18:00 이후의 결측은 물리적 특성을 반영해 0으로 대체하였다. 3시간 이상 연속 결측은 다음 문단에서 설명하는 클러스터링 기반 보간 절차를 따랐다.

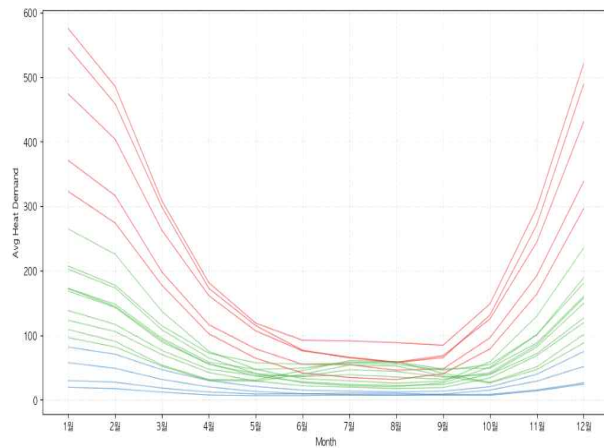
3 시간 이상 연속으로 발생한 결측 구간을, 동일-패턴 지점들의 정보를 활용해 보다 현실적으로 채운다. 지점 간 유사도 산출을 위해 DTW 거리 행렬을 이용하였다. 19개 branch_id

별 열수요(heat_demand) 시계열에 대해 Dynamic Time Warping(DTW)으로 모든 쌍의 거리를 계산하였다. DTW는 패턴이 비슷하지만 타이밍·속도가 어긋난 경우에도 작은 거리를 부여하므로, 형태적 유사성을 정확히 반영할 수 있기 때문이다. 이후 계층적 클러스터링 및 최적 군집 수 결정을 하였다. DTW 거리 행렬을 입력으로 계층적 군집화(average linkage)를 수행하고 덴드로그램을 분석한 결과, 3개 클러스터가 전체 변동성을 가장 잘 설명하였다.

개별 branch_id 단위로 보간을 수행하지 않은 이유는, 결측 구간이 동일한 값으로만 채워지는 문제가 있었기 때문이다. 따라서 유사한 수요 패턴을 가지는 지점을 군집화하고, 이를 바탕으로 각 클러스터별 대표적인 패턴을 활용하여 보다 신뢰성 있는 보간을 수행하고자 하였다. 보간에 대한 구체적인 식은 $0.5 \times (\text{결측치 기준 바로 이전 값}) + 0.5 \times (\text{같은 클러스터 내 해당 시점 평균})$ 으로 설정한 후 보간을 수행하였다.



< 그림 1> DTW 기반 클러스터링



< 그림 2 > 지사별 평균 heat_demand 그래프

2.3 파생 변수

열수요 예측의 설명력을 높이기 위해 다양한 파생 변수를 체계적으로 설계·도입하였다.

- 주말/휴일 : is_weekend, is_holiday 변수를 생성하여 주말과 법정 공휴일 여부
- 시간 주기성 : 하루 24시간을 기준으로 1차 및 2차 고조파를 적용한 삼각 함수 변환 및 월간 일수 차이를 고려한 일 주기 및 12개월 주기도 포함한다. 이러한 주기 변수는 회귀·트리 계열 모델에서 시간의 순환성을 부드럽게 학습하도록 돕는다.
- 불쾌지수(Discomfort Index, DI) : 기상·체감 요인을 포착하기 위해 도입한다.

$$DI = 1.8ta - 0.55(1 - hm)(1.8ta - 26) + 32$$

불쾌지수를 식으로 산출한 뒤, 68 미만(전원 쾌적)부터 83 이상(매우 불쾌)까지 6단계(0-5)로 등급화하였다.

- 시간별 난방도일(Heating Degree Hour, HDD) :

$$HDD = \max(0, 18 - ta)$$

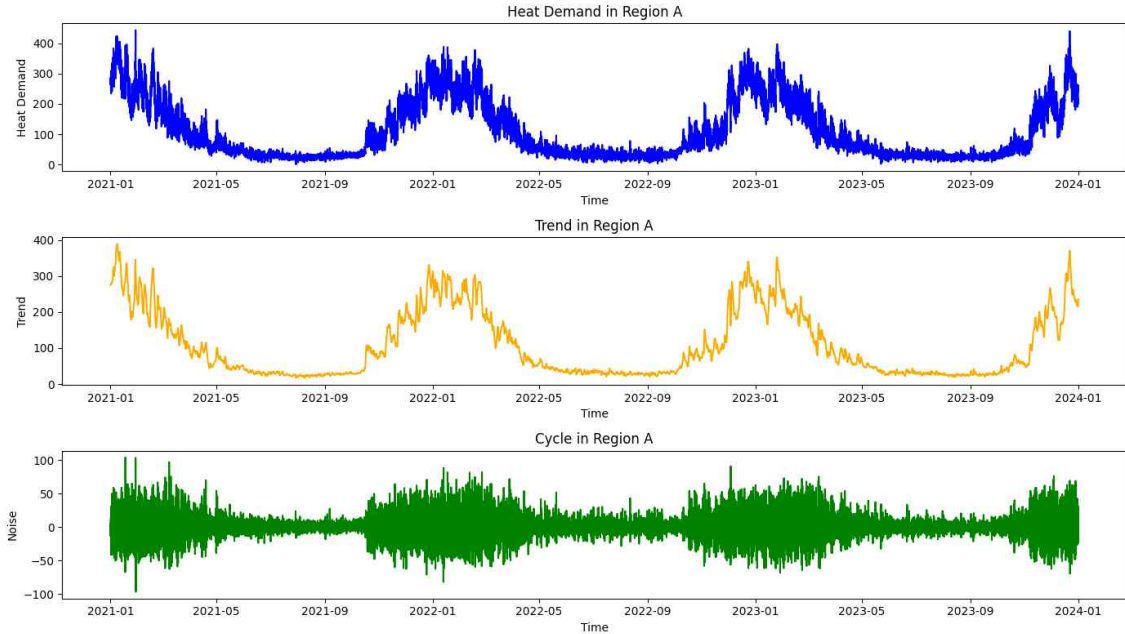
난방 수요 및 에너지 소비 예측을 위한 기후 지표이다. 일반적으로는 일별 평균 기온을 기준으로 계산하지만, 시간 단위의 효과를 반영하기 위해 시간별 난방도일을 계산하였다.

- 지연 변수 : ta(기온)는 1, 3, 6, 12시간 및 하루 전(24 시간)까지 초단기·단기·반일·일주기 영향을 모두 모델에 전달한다. hm(상대습도)은 변동 속도가 비교적 완만하므로 1-3 시간 범위의 단기 기억만 활용했고, ws(풍속)는 급격한 스파이크 특성이 강해 1 시간 직전값과 3 시간 지연값으로 충분히 설명하도록 하였다. 이러한 설계는 변수별 물리적 응답 시간을 고

려해 필요 이상의 공선성을 피하면서도, 열 수요에 영향을 미치는 자기상관 구조와 주기 패턴을 효과적으로 학습할 수 있게 해 준다.

3. 분석 방법

3.1 HP filter를 이용한 추세와 노이즈의 분해



< 그림 3 > HP filter를 이용한 분해 예시 : A 지사

본 연구에서는 열수요 시계열을 Hodrick-Prescott(HP) 필터로 분해해 추세(τ_t)와 노이즈 ($c_t = y_t - \tau_t$)으로 나누었다. HP 필터는

$$\min_{\tau_t} \sum_t (y_t - \tau_t)^2 + \lambda \sum_t (\Delta^2 \tau_t)^2$$

와 같이 관측치와 추세 간 오차(첫 항)와 추세의 곡률(둘째 항)을 동시에 최소화하므로, 부드러운 장기 흐름과 변동을 자동으로 분리할 수 있다. 모델의 복잡도를 의미하는 람다는 $\lambda = 1600$ 을 사용하였다. 이렇게 얻은 노이즈 c_t 는 평균이 0에 가까운 준정상 시계열이 되어 트리 기반 모델 및 부스팅 기반 모델 등에 그대로 투입할 수 있으며, HP 분해를 선행함으로써 모델이 계절·주기적 패턴 학습에 집중해 예측 정확도가 한층 향상된다.

3.2 XGBoost를 이용한 지사별 추세-잔차 이중 모형 구축 근거

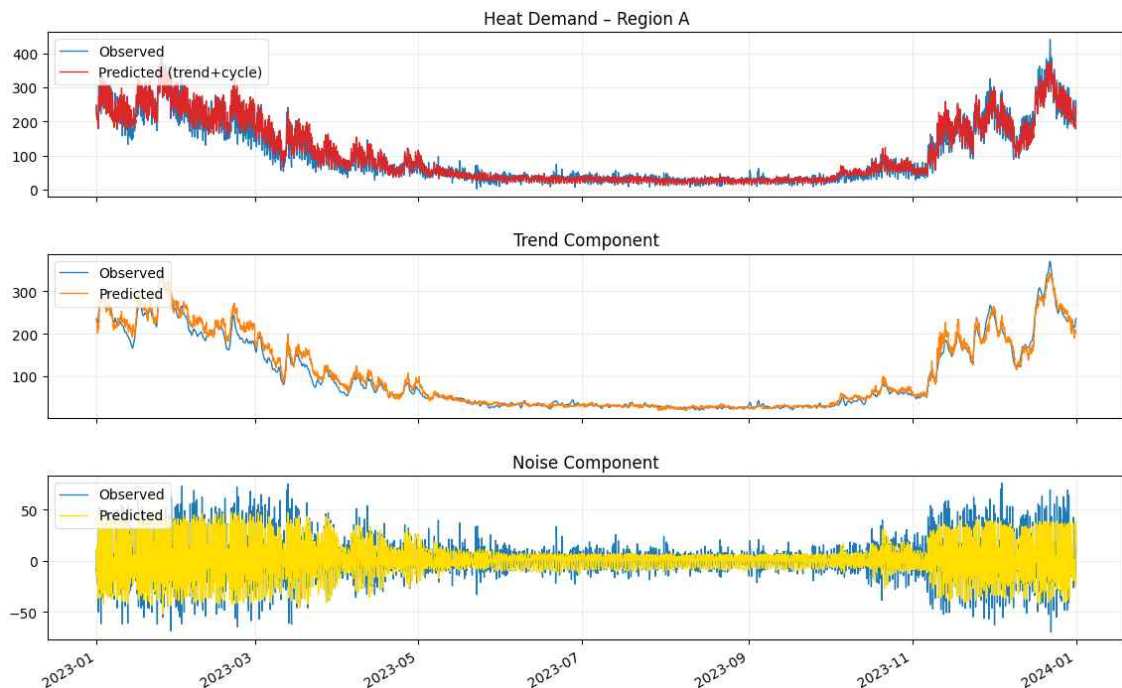
본 프로젝트의 핵심 목표는 각 지사(branch)의 시간당 열수요(heat demand)를 최대한 정확히 예측하는 것이다. 지점마다 난방 설비 용량·배관 손실·수요 패턴이 상이하므로, 지사별 모델링이 필수적이다. 지사별 열 수요 시계열을 HP 필터로 분해하면 장기 추세(τ_t)와 노이즈(c_t) 두 성분이 도출된다. 지사별 두 성분 모두를 XGBoost 회귀기로 별도 학습한 뒤 최종 예측을 $\hat{y}_t = \hat{\tau}_t + \hat{c}_t$ 로 합산하는 이중 모형 전략을 이용한다. 이를 XGBoost로 구현한 이유는 다음과 같다.

- 스케일·동역학이 다른 두 성분을 각각 최적화 : 장기 추세(τ_t)는 year, month_sin/cos, 누적 HDD 등 장기·계절 기반 변수를 중심으로 학습한다. XGBoost는 트리 분할을 통해 비선형 계절 효과나 기후 변화 추세(예: 온난화로 인한 베이스라인 하락)를 유연하게 포착한다. 노이즈 성분(c_t)는 ta_lag, ws_lag, hour_sin/cos, is_weekend 같은 단기·고주파 변수가 핵심이다. 부스팅은 복잡한 상호작용(예: “주말+갑작스런 한파” 조건)도 단계적으로 반영해 급등·급락을 정확히 추적한다.
- 정상성·추세 분리 요건 충족 : 시계열 트리 모델은 강한 추세·계절성이 그대로 남아 있으면 과적합·데이터 누수에 취약해진다. HP 필터로 장기 추세(τ_t)를 제거하고 잔차만 학습시키면, XGBoost는 상대적으로 i.i.d. 가정에 근접한 입력을 받게 되어 일반화 성능이 높아진다. 최종 예측은 $\hat{y}_t = \hat{\tau}_t + \hat{c}_t$ 로 복원하므로 추세 정보도 보존된다.
- 풍부한 파생 변수 활용 : Gradient Boosting은 수백 개의 지연(lag)·주기·체감·난방 부하 변수를 투입해도 스케일 조정이나 다중공선성 걱정 없이 학습한다.

지사별 HP-분해 + XGBoost 이중-모형은 지점 고유의 장·단기 수요 특성을 정확히 반영하면서도, 트리 부스팅의 비선형·상호작용 학습 능력을 극대화해 열수요 예측 정확도와 운영 해석력을 동시에 향상시킨다.

4. 결과 및 검증

4.1 미세조정을 통해 산출한 validation(2023년 예시 : 지사 A)



< 그림 4 > HP filter를 이용한 분해 후 예측 : A 지사

2023년 best rmse : 17.02

먼저 상단 그림에서 보이듯, 예측치(빨간 선)는 계절별 수요 변동을 전반적으로 정확히 따라가며 난방 종료·시작 시점의 급격한 하락과 상승을 모두 포착한다. 특히 봄·여름철 저부하 구간에서는 과대·과소 예측이 거의 없어 기본적인 수준 이동(level shift)을 안정적으로 재현합니다. 다만 겨울철 극한 부하가 발생한 며칠간은 실측 피크보다 약간 낮게 형성되었다.

중간 그림은 장기 추세만을 비교한 것으로, 실측과 예측이 겹쳐 보일 만큼 오차가 작다. 이는 커널 대역폭이나 로컬 회귀 창(window)이 적절히 선택돼 추세 성분을 과소·과대평가 없이 학습했다는 증거이다. 반면 하단 그림의 노이즈를 보면 분산이 계절에 따라 달라지는 이분산(heteroscedastic) 구조를 예측치(노란 영역)가 대체로 따라가긴 하지만, 실측 노이즈 스파이크는 일부 놓치는 모습이 남아 있다.

현재 모델은 계절 흐름과 장기 추세를 매우 잘 포착해 운영 계획 수립에는 충분히 유용하지만, 혹한기 극값·초단기 변동을 더 정확히 예측하려면 기상·행동 패턴 같은 외생 변수 추가, 이분산을 명시적으로 모델링, 노이즈의 자기상관 구조 분석을 통한 모델 고도화가 필요해 보인다.

4.2 검증 - 최종 test rmse = 15.53

본 연구는 지사별 수요 특성을 반영하기 위해 branch 단위로 개별 모델을 구축하고, 각 모델의 하이퍼파라미터를 Optuna-TPE 탐색으로 최적화한다. 학습 자료는 2021-2022년(Training), 검증 자료는 2023년(Validation)으로 구분하고, 학습 구간 내에서는 5-분할 TimeSeriesSplit을 적용해 시계열 누적 구조를 유지한 채 교차검증을 수행하였다.

구분	값
탐색 알고리즘	Optuna TPE sampler (seed = 42)
평가 지표	RMSE (5-분할 CV 평균)
시도 횟수	50 trials / branch
학습·검증 비율	2021-2022(학습) / 2023(검증)

< 표 2 > 탐색 설정

파라미터	범위(분포)
n_estimators	500 - 5 000 (step 100)
max_depth	3 - 12 (정수)
learning_rate	0.01 - 0.30 (log-uniform)
subsample, colsample_bytree, colsample_bylevel	0.70 - 1.00 (연속)
reg_alpha, reg_lambda	1e-8 - 1.0 (log-uniform)
min_child_weight	1 - 7
gamma	1e-8 - 0.5 (log-uniform)

< 표 3 > 주요 탐색 파라미터 범위

Trial별로 도출된 파라미터 조합으로 5회 CV를 수행한 뒤 평균 RMSE가 최소인 조합을 각 지사의 최종 파라미터로 채택하였다. 이후 2021-2022 전 구간으로 재학습해 2023년 검증 세트에 적용했으며, 지사별 최종 RMSE는 보고서 결과 표에 정리하였다. 이러한 전략은 시간 순서를 보존하면서도 과적합을 방지하고, 지사별 수요 패턴의 이질성을 반영할 수 있는 장점이 있다.

5. 활용 방안

추세-노이즈 분해 모델은 열수요 관리의 전략적·운영상 의사결정을 모두 지원할 수 있다. 먼저 장기적으로는 추세 예측을 바탕으로 연료 계약 규모나 보일러 증설 시점을 정교하게 계획함으로써 과투자 및 설비 부족 위험을 동시에 줄일 수 있다. 또한 연도별 탄소 배출량을 전망해 재생열원이나 열저장 설비 도입 효과를 시뮬레이션함으로써 중·장기 탈탄소 로드맵 수립에도 기여한다. 한편 단기 운용 측면에서는 노이즈 예측으로 하루 전·시간 전 부하 피크를 사전에 감지해 보일러 기동·정지 순서를 최적화할 수 있어 연료비와 기동 비용을 절감할 뿐 아니라 예비율 확보에도 유리하다. 예측치와 실측의 편차가 일정 한계를 넘을 때 즉시 경보를 발령하도록 하면 설비 고장, 센서 오류, 돌발 기상 등 이상 상황을 조기에 포착해 대응 시간을 벌 수 있다. 나아가 계절·시간대별 추세 패턴을 이용해 수요반응 프로그램의 인센티브 구간을 설계하고, 노이즈 분석으로 변동비와 고정비를 분리해 요금 체계를 투명하게 조정하면 사용자 피크 억제를 유도하면서도 비용 회수 구조를 안정화할 수 있다. 결국 추세는 중·장기 전략 수립에, 노이즈는 단기 운영과 리스크 관리에 각각 최적화되어 있어 두 성분을 함께 활용하면 에너지 비용, 탄소 배출, 서비스 신뢰도를 동시에 개선할 수 있다.