Climate Impact on Residential Load: the Tale of 100K GAMs

Mike Ludkovski

UC Santa Barbara

Suggested Citation

M. Ludkovski, "Climate Impact on Residential Load: the Tale of 100K GAMs," in *Proceedings of the PowerUp Conference*, 2026.

BibTeX Entry

```
@inproceedings{ludkovski2026climateimpact,
  author = {Ludkovski, Mike},
  title = {Climate Impact on Residential Load: the Tale of 100K GAMs},
  booktitle = {Proceedings of the PowerUp Conference},
  year = {2026},
}
```

Climate Impact on Residential Load: the Tale of 100K GAMs

Mike Ludkovski†

***Abstract*—We design statistical methodology to project the impact of climate change on residential load at the level of substations. To achieve high-resolution load projections, we combine the outputs of downscaled global climate models (GCMs) with a statistical additive regression model. To this end we design a bespoke Generalized Additive Model (GAM) that links cooling/heating loads at given substation and hour of day to daily meteorological covariates. Our models are trained on the ResStock outputs for a realistic building mix and illustrated on the entire 1898 substations in the CATS synthetic grid. The results visualize feature relevance and aggregate climate impacts on California residential loads.**

***Index Terms*—climate impact, load modeling, residential power demand, generalized additive models**

I. Introduction

Ensuring grid reliability in a warming world requires granular modeling of climate impacts on future electricity demand. Traditional grid-wide projections mask potential vulnerability as climate impact on load is highly heterogeneous. On the one hand, climate effects vary spatially due to microclimates and geographic features. For example, temperature differentials between coastal and inland areas are expected to rise. On the other hand, different building and population mixes (such as urban vs rural), imply that load profiles will shift differently due to, say, increased summer heat. The resulting disparate demand changes will cause uneven impact on distribution and transmission across the grid.

In this applied study we design statistical methodology to incorporate the effects of climate change into the modeling of future residential load at a *substation* level. Our motivating application is capturing the shifts in CAISO—California Independent System Operator—over the next 10-15 years. Specifically, we aim to capture the impact of the anticipated climate shift between 2020 and 2035. Given the numerous and complex microclimates present in California (the marine layer, tule fog, etc.) and substantial intra-state heterogeneity in building stock and household characteristics, we seek to capture spatially resolved climate impacts that are further disaggregated by major housing types and load categories.

To achieve high-resolution load projections, we combine the outputs of downscaled global climate models (GCMs) with a statistical regression model that links weather and residential load. Our approach learns a nonlinear hourly-scale dependence of specific appliance loads on a collection of daily weather covariates, enabling predictions under unseen weather states. This methodology supercedes simple temperature-adjusted historical weather scenarios by (i) taking into account the role of other essential weather variables, such as precipitation, irradiance, wind, etc.; (ii) leveraging GCMs to lean into the upcoming distributional weather changes (more frequent and more prolonged heat waves, more extreme precipitation events, etc.) that cannot be captured by "bumping" the temperature profile uniformly across the year.

While engineering-based approaches to energy consumption of individual buildings are certainly the gold standard, they are computationally intensive and still require making lots of assumptions. Our machine-learning-flavored approach thus marries the highly accurate physical simulations from ResStock with a fast-to-evaluate map that connects weather inputs to hourly load profiles. A related challenge that our approach sidesteps is the resolution mismatch between GCM weather outputs that are at the daily scale, and the hourly inputs expected by ResStock.

Related Literature: There is voluminous literature on statistical approaches to load modeling and forecasting [1]–[4]. The first paper to apply GAMs to this task was [5] who targeted short-term load forecasting and primarily considered recent loads and temperature for the covariates. Similar GAM-based short-term forecasts were built by [6] for 2260 French substations. More recently, [7], [8] considered GAM models for jointly simulating daily net-load of 14 regions in UK by utilizing a collection of weather variables similar to ours. [9] perform medium-term (annual 8760 scale) load forecasting using a P-splines GAM model. A key distinguishing feature of our analysis is the different time-scales of our meteorological covariates (which are all daily) relative to the load profile, and the disaggregation into load categories. Other approaches include functional and multiple linear regression methods, for example group MCP or group LASSO [10]. An emergent strand proposes deep learning methods, e.g., using Convolutional Neural Networks or LSTM architectures [11].

Context: Our overall emphasis is applied research that weaves in statistical tools as part of a larger team pipeline. Companion in-progress works describe (i) residential electrification adoption pathways matched to residential housing stock configurations in California and mapped to electrical substations; (ii) updates to the CATS [12] synthetic grid to align its generation mix and transmission network with granular residential and commercial building mix, and realistic DERs; (iii) use cases of our climate- and adoption-adjusted load scenarios for grid planning and stress-testing purposes [13]. In aggregate, this yields a new state of the art toolbox for highly resolved projections of future grid conditions in CA.

Section II presents our data pipeline and the overall approach. We then present our empirical results about climate impacts across California residential substations in Section III before diving into the modeling details in Section IV. Section V discusses further in-progress work and the limitations of the present analysis.

Mike Ludkovski is with the Department of Statistics and Applied Probability University of California Santa Barbara, CA, USA. Email: ludkovski@pstat.ucsb.edu

II. Data Pipeline

We adopt a bottom-up approach to construct load profiles at the substation level by integrating multiple datasets. We begin by obtaining housing characteristics and energy demand at the census tract level using NREL's ResStock [14], an energy consumption model of the US housing stock, combined with the American Community Survey Housing Characteristics data. Next, we characterize buildings served by each substation using datasets from Zillow, Microsoft Building Footprints, and utility-provided spatial data on substation and feeder locations. A crosswalk between census tract-level data and substation-specific information is the final labor-intensive piece. Our analysis aims to reproduce the CAISO grid; to this end we use the public-domain CATS [12] synthetic transmission network that defines about 8800 buses, of which 1898 correspond to substations with some residential load.

To impute the weather-load relationships we employ the physical building models embedded in the 2024-vintage ResStock 3.3 tool, developed by the National Renewable Energy Laboratory (NREL) [15]. ResStock simulates loads broken out by 28 appliances and 5 building types: Single Detached, Single Attached, Multi-family 2-4, Multi-Family 5+ (aka apartment buildings) and Mobile Homes. Most of the appliances, such as dryer, dishwasher, exterior lighting, pool pump, etc, are either deterministic or (practically) weather-independent. For our pipeline we focus on the Cooling and Heating components. For a given building subtype, ResStock generates year-long 15-min energy demand profiles (35040 time points) based on AMY (Actual Meteorological Year) weather referenced from 2019 NOAA dataset.

For our climate forecasts, we seek typical California weather conditions at the year 2035 horizon. To this end, we rely on four GCMs that have been shown to be among the best performing for Western US: EC-Earth3, MIROC6, MPI-ESM1 and TaiESM1 [16]. All utilize the SSP-3.7 climate scenario which projects about 1.5 degrees of warming above pre-industrial levels by 2035. The respective GCMs are downscaled to 3×3km spatial resolution [17] and available via the Cal-Adapt tool [18]. Our Typical Meteorological Year (TMY) construction methodology is from [19].

ResStock-simulated loads are not directly applicable for assessing climate impacts because ResStock utilizes PUMA-wide weather (so all substations within a given PUMA are assigned same weather inputs), creating a spatial mismatch between a specific substation and the weather that was coupled to it. This mismatch can be quite severe, e.g., a coastal Mendocino substation has an average temperature gradient of nearly 6° versus the NOAA weather station that is 10+ miles inland. As a result, we do not have any historical benchmark; instead when reporting climate impacts we compare our predicted loads based on a given GCM to ones based on the reanalysized ERA 2019 data (downscaled same way as the GCMs).

For meteorological predictors, we use:

- `Tmax`/`Tmin`: daily maximum/minimum temperature at 2m altitude in $^\circ C$;
- `TmaxLag`/`TminLag`: yesterday's max/min temperature (lagged by one day);
- `day`: day of the year, normalized to $[0, 1]$ and treated as a cyclic covariate.
- `wind`: average daily wind speed at 10m in m/s;
- `irrad`: daily surface downward direct normal irradiance in W/m^2 ;
- `humid`: average daily relative humidity $\in [0, 100]$;
- `prcp`: total daily precipitation in mm.

The inclusion of lagged temperature is to account for thermal inertia, i.e., buildings retaining heat, and the fact that yesterday's max/min temperature directly impacts the early morning hours. In addition, we also employ two flags: `weekend` to distinguish weekdays (Mo-Fr) and weekend (Sa, Su) days, and `wet` which is an indicator of $\texttt{prcp} > 0$.

A. Empirical Features of CA Residential Load Profiles

Figure 1 displays a few sample load profiles. As illustration, we use the **parlier** substation, covering a small city in Fresno county with about 2200 single-family detached buildings. The left panel shows cooling load at four sets of two consecutive days. On a cold January day, cooling demand is minimal (though non-zero); it is an order of magnitude higher in the summer, especially when daily highs get well above $30^\circ C$. Peak demand is around 6pm, however the daily profile varies, for instance compare the broader peak in October relative to the sharp evening fall-off in April. Moreover, due to the diurnal seasonality, peak cooling demand is later in early summer than in Fall as respective sunsets occur later. The right panel shows heating demand at the same substation and same days. During the summer, heating demand is zero for all 24 hours. In colder conditions, the highest load is in the morning (hours 6-9); afternoon may be zero demand or not depending on how cold it gets. The loads zig-zag due to the intra-hourly cycling of the heaters.

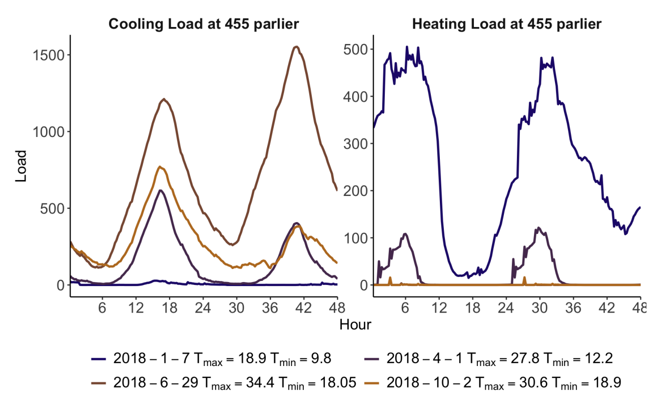

Fig. 1. Sample ResStock load profiles at 4 representative pairs of days at the parlier substation in Fresno county. Each day has $24 \times 4 = 96$ observations.

Several empirical features are essential when modeling weather-load dependence. First, the impact of most covariates is *nonlinear*. For instance, it is well known that temperature has a hockey-stick shape impact on AC cooling usage, with muted impact for moderate temperature and a much stronger sensitivity for temperatures at or above $25^\circ C$. Second, there are strong *interaction* effects. For example, sunlight duration drives the energy absorbed by buildings, creating an interaction between day and $T_{\max}$, with impact of temperature amplified in the summer. Third, many areas of California have zero heating demand during some days, creating a distinct mixed distribution to train on. Fourth, while many of the weather variables are highly correlated (e.g., irradiance and humidity), because those covariates are on the daily scale while our load profiles are hourly, they still contribute distinct effects. Hence, an a priori variable selection is treacherous.

As illustrated in Figure 1, the daily load profile can take on vastly different shapes. This is evidence of the deeper fact that the dependence of load on weather is qualitatively different during various times of day. As a consequence, we choose to separately model a collection of specific hours and then "stitch" the respective point-in-time projections into a daily load shape. Similar stitching reconstructs the full 8760 annual profile. The specific methodology we employ are Generalized Additive Models (GAM), discussed in Section IV. As a typical setup, we build GAMs for 12 daily timepoints, spaced every 2 hours, and then interpolate for other hours using cubic splines.

III. Aggregate results across California

In this section we present the aggregate results from our models across the 1898 residential substations in CATS [12]. This entails building $1898 \cdot 5 \cdot 2 \cdot 12 \simeq 220K$ models across substations/building types/load categories/hours of the day. Figure 2 visualizes the results, with a further summary in Table I. As expected, global warming causes cooling loads to rise, on average about 9-11% between 2020 and 2035. The largest impact (15%+ increase) is in the Sacramento Valley and the SF Peninsula in Northern CA. Higher temperatures will curb heating demand by about 18-25%. This reflects the mild climate in California, where heating is only used occasionally and will become even less needed in the future. Largest decreases are around Los Angeles and San Diego, where heating needs will drop by more than 30%; conversely, minimal heating change will occur in the Central Valley.

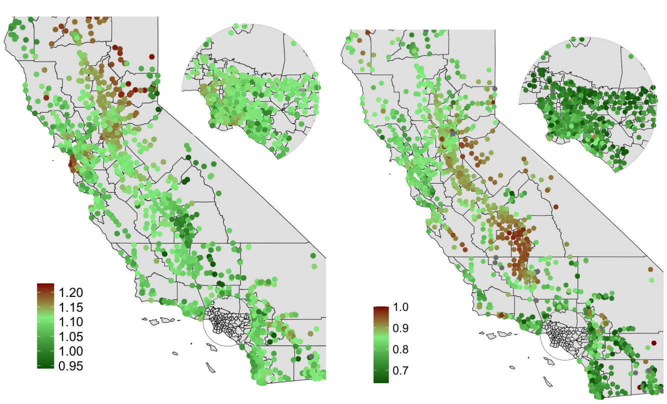

Fig. 2. Climate impact on total annual substation load. We report the ratio of projected load based on the 2035 Earth-EC3 downscaled GCM TMY relative to the 2019 ERA baseline. *Left panel*: Cooling; *Right:* Heating across the 1898 CATS substations. The colors and ranges are different in each plot.

Turning to the comparative effects across the 4 GCMs, we observe that MIROC6 gives the strongest climate shifts (its TMY also has the highest average temperature) and also the strongest variability, while MPI-ESM1 is projecting essentially no change in cooling demand. Notably, the MIROC6 GCM forecasts longer tails of heating load distribution, reflecting the phenomenon of "hotter hots and colder colds" and increased weather volatility. This happens especially in Riverside county and the Inland Empire.

In total, climate impact will increase total residential demand on average by 2–4%, see Figure A.7. This reflects the offsetting effects on cooling and heating and the presence of many weather-independent appliances that we hold constant. While this overall effect is muted, we emphasize that it is only about the summed annual demand and does not reflect daily variability. Notably, there is little correlation between climate impacts on heating vs. cooling (Fig. A.7 left).

In Table I we report climate impact by building types. Beyond the average annual load percentage change, we also look at the 98% quantile (corresponding to looking at the hottest 175 hours or about 1 week's worth) and the 99.8% (the hottest 18 hours). Mobile Homes experience the largest climate impact and Apartment Buildings the smallest. While the variability is not large, it is material: e.g. a 5% differential in heating load decline between SFD and SFA building types.

TABLE I
Mean, 98% and 99.8% Quantile Changes in Annual Hourly Residential Load across CATS substations by building type.

	Cooling			**Heating**		
	EC-Earth3 GCM TMY Input					
Building Type	Ave	$q_{0.98}$	$q_{0.998}$	Ave	$q_{0.98}$	$q_{0.998}$
Sngl Fmly Attached	9.8%	4.6%	9.5%	-23.8%	-20.3%	-18.3%
Sngl Fmly Detached	10.5%	4.8%	9.4%	-18.6%	-14.1%	-14.8%
Multi-Family 2-4	10.2%	4.4%	9.2%	-21.5%	-17.7%	-15.3%
Apartment Bldg 5+	9.5%	4.7%	10.3%	-21.8%	-18.8%	-15.7%
Mobile Homes	12.3%	6.0%	11.6%	-20.8%	-16.5%	-11.2%
Total	10.5%	4.9%	9.5%	-19.1%	-14.9%	-14.7%
Downscaled Global Climate Model (SFD building type)						
MPI ESM 1	-1.0%	0.7%	7.0%	-8.1%	-9.7%	11.1%
Tai ESM 1	10.8%	7.0%	10.8%	-19.5%	-19.5%	-20.2%
EC-Earth3	10.5%	4.8%	9.4%	-18.6%	-14.1%	-14.8%
MIROC 6	17.0%	9.6%	15.8%	-22.9%	-23.2%	-9.5%

IV. Statistical Modeling of Load Profiles

The statistical task is to model the 96-dim. daily load profile $\mathbf{f} = \{f(t) : t \in \{0.25, 0.5, \dots, 24\}\}$ as a function of input covariates. The above considerations imply that we seek a semi-parametric flexible model that can capture multiple nonlinear relationships, which naturally suggests Generalized Additive Models (GAM) . GAMs offer a high degree of flexibility and interpretability, transparently showing the impact of each covariate.

Denote by $\mathbf{X}$ the vector of d input covariates. GAM represents the output Y as a sum of several *smooths* f_j, which are in turn modeled as linear combinations of basis functions:

$$g(\mathbb{E}[Y|\mathbf{X}]) = \beta_0 + \sum_{j=1}^{J} f_j(\mathbf{X}), \tag{1}$$

where g is the link, and $f_j(\mathbf{X}) = \sum_{k=1}^{K_j} \beta_{j,k}\phi_{j,k}(X_{\alpha_j})$ depends on a (small) subset index α_j of the d covariates. In our setting, these smooths are of mixed type and include univariate smooths $|\alpha_j| = 1$; bivariate smooths $|\alpha_j| = 2$; and univariate and bivariate smooths modulated by categorical predictors.

We use cubic splines for the basis functions $\phi_{j,\cdot}$ and cyclic splines for day. We then build 12 models for the odd hours of the day $hr \in \{1, 3, \ldots, 23\}$. Using the R mgcv library [20], our model is

```
gam(load ~ wet + weekend +            # GAM-cool
  s(day,  k=12, bs="cc", by=weekend) +
  s(Tmax, k=12, bs="bs", m=c(3,1), by=weekend)+
  s(irrad, k=8, bs="bs", m=c(3,1), by=wet) +
  s(wind,  k=5, bs="bs", m=c(3,1)) +
  s(humid, k=5, bs="bs", m=c(3,1)) +
  s(prcp,  k=5, bs="bs", m=c(3,1)) +
  s(Tmin,  k=8, bs="bs", m=c(3,1)) +
  s(TmaxLag, k=8, bs="bs", m=c(3,1)) +
  s(TminLag, k=5, bs="bs", m=c(3,1)) +
  ti(Tmax, day, bs=c("cs","cc"), by=weekend)+
  ti(Tmax, Tmin, k=c(4,4), bs=c("cs","cs")),
  data = trainDF, select=TRUE)
```

Above k refers to the number of spline knots K_j for the respective smooth, and m refers to penalty order. Model (GAM-cool) features 6 univariate smooths (lines 4-9), 6 smooth-factor interactions (the first 3 terms with by= clauses that differentiate between weekdays and weekends and between dry and wet days) where a different univariate smooth is defined for each level of the factor variable, and 3 bivariate tensor-product smooths. The latter capture the interactions between $T_{\max}$ and day, further conditioned on the weekday flag, and between $T_{\max}$ and $T_{\min}$. Since we already model the respective univariate effects, we only retain the interaction terms $f^{4,4}(T_{\max}, T_{\min}) + f^{4,4}(T_{\max}, day, 1_{wet=T}) + f^{4,4}(T_{\max}, day, 1_{wet=F})$, where the superscripts indicate the dimension of each marginal basis. The above model has 122 coefficients to be estimated. To model heating loads, we swap the roles of the $T_{\max}$ and $T_{\min}$ variables, see Appendix.

The standard Restricted Maximum Likelihood (REML) [21] is employed for smoothness selection. To address the potential collinearity and unknown covariate relevance, we utilize built-in shrinkage tools via the `select=TRUE` option. This modifies the penalty matrix of the splines to penalize the null space (e.g., the linear and intercept components) of the smooth functions, allowing the model to shrink the effect of a non-informative covariate to zero. Hence the fitted model only retains covariates with significant explanatory power and mitigates spurious complexity.

The paragraphs below explain and discuss the specific form of the GAM model above and the choices for each term.

Zero-inflation: For heating loads, we face a mixed distribution with a large point mass at zero (representing periods with no active heating, especially during the summer) and a positive, right-skewed continuous distribution for non-zero demand. To handle this situation, we utilize the Tweedie distribution (`family=tw()`) which takes the variance of the response to be a power function of its mean, $\mathrm{Var}(Y|\mathbf{X}) = c \cdot \mathbb{E}[Y|\mathbf{X}]^p$, where the hyperparameter $1 < p < 2$ distinguishes between Poisson-like ($p \to 1$) and Gamma-like ($p \to 2$) behavior. This is combined with a log link function g to jointly estimate the probability of zero load and the continuous positive response within a single likelihood function. An alternative that we considered are two-part zero-inflated (or Hurdle) models, such as `ziplss` or Zero-Adjusted Gamma, that decouple the generative process and separately build a logistic regression for the probability of non-zero load and a GAM for the strictly positive component:

$$\mathbb{E}[Y \mid \mathbf{X}] = \mathbb{P}(Y > 0 \mid \mathbf{X}) \cdot \mathbb{E}[Y \mid Y > 0, \mathbf{X}],$$

where each of the two terms on the right is modeled with a distinct GAM model. While allowing flexibility through distinct smooths for zero-probability and magnitude, this approach is slower and more difficult to finetune.

Extrapolative Behavior: A key challenge to address is extrapolation to yet-unseen conditions, for instance extreme heat waves. Our training set is based on historical weather (AMY) in 2019, while our models are evaluated on 2035 climate-based TMY. This shift is exacerbated due to the mismatch between the ResStock input weather and the actual weather at a given substation, with the result that we often have substantial deviations, such as trying to predict load under a $T_{\max}$ that is 10°C higher than anything in the training set. Such setting is prone to major extrapolation errors where predictions become nonsensical, especially if using a log-link. For this reason, we employ cubic splines with a first-order difference penalty (`bs="bs", m=c(3,1)`). Standard second-order penalties extrapolate linear trends on the predictor scale, implying physically unrealistic exponential growth of demand on the response scale during extreme weather events. In contrast, first-order penalization shrinks the response towards a constant (zero slope in the linear predictor) beyond the observed data range, effectively modeling a gradual saturation of load beyond the training temperature range. Similarly, to prevent artifacts from the tensor product interaction terms during extrapolation, we utilize cubic regression splines with shrinkage penalties (`bs="cs"`) for the non-cyclic covariates. This ensures that beyond training regions the interaction effects decay to zero, allowing the model predictions to converge robustly to the sum of the additive main effects.

Observation Likelihood: There is no canonical way to handle the observation likelihood. The default Gaussian likelihood in (GAM-cool) corresponds to putting equal weight on capturing very high and very low loads. A Gamma-type likelihood would recognize that loads are non-negative and that larger swings are likely during heavy-load days, but would amplify extrapolation difficulties. As given, projected load based on (1) may end up (slightly) negative. This is caused by the Gaussian likelihood, as well as by the spline basis, which makes the smooths to have low curvature and therefore unable to capture a hockey-stick shape that is often present. We zero out all negative predictions.

In our setup, we focus on capturing the "ground truth" load, hence the predictive mean. One may bootstrap the training residuals to output predictive simulations, which may be more appropriate for grid stress testing, as it would recognize that realized loads can end up being even higher than projected due to "unmodeled" residuals.

Additional Terms: We considered incorporating hour of day via another GAM term, namely `s(hr,bs="cc")`. However the hourly load shape is not constant across seasons, so such additive structure is insufficient, and interaction terms like `ti(hr,day)` are needed. Because such interactions add a lot of extra GAM coefficients, we opt for the direct modeling of different hours.

Similarly, we have considered building a 3-term smooth to jointly model hour and season dependence, e.g., according to `ti(hr,day,Tmax)`; however, that produces a very slow model that does not perform noticeable better. One underlying challenge is that we only have 1 year's worth of training data, making capturing some of the interaction terms nontrivial.

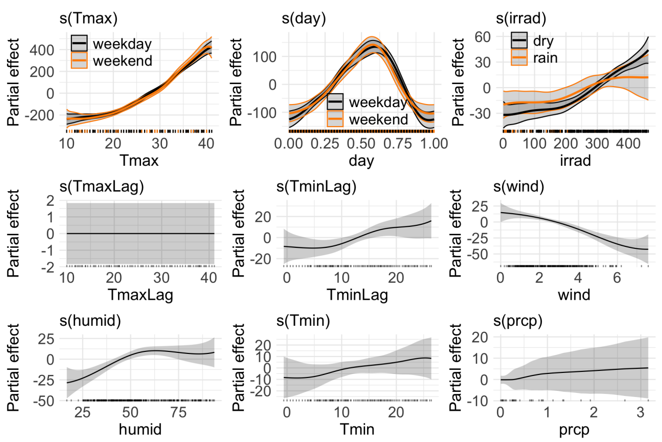

Fig. 3. Partial dependence of cooling loads at 6pm at **parlier** substation on univariate covariates in (GAM-cool). See Figure A.8 for the bivariate smooths.

A. Partial Dependence and Influential Covariates

Figure 3 visualizes the fitted GAM model for 6pm cooling load for SFD buildings at the **parlier** substation in Fresno county. We show the partial effect—conditional on keeping all other inputs constant—of all 1d smooths; see Figure A.8 in the Appendix for the three bivariate smooths. Note that the y-axes are all different and their range provides a ballpark estimate of term relevance. The major effects are due to $T_{\max}$, which exhibits an inflection point around 18-20$°C$, so that cooling loads increase in a convex fashion as daily max rises. For temperatures above 40$°C$ the slope lessens a bit, indicating partial saturation.

The other major effect is time of year day, with much higher cooling loads in the summer compared to winter. There is a complex annual pattern, indicating that peak loads are in August-September and lowest cooling is in January-February; this pattern is a bit flatter on weekends. Ceteris paribus (i.e., even for the same temperature), cooling is more prevalent in the summer, both because daylight is longer (so the buildings absorb more heat) and because many households keep their thermostats permanently on during the hot months, while turning them on only intermittently in winter.

Cooling demand falls when humidity is low, and increases when yesterday's $T_{\min}$ was high; note the respective hockey-stick shapes. There is a hockey-stick influence of $irrad$, (more cooling on sunny days), and a concave dependence on $wind$ indicating that calm weather implies more cooling. Finally, $T_{\max}(d-1)$ and $prcp$ are deemed irrelevant for this specific substation and their influence is shrunk to zero. Shrinkage penalties are essential to achieve the essentially linear relationships (see e.g., $wind$) and to drop irrelevant covariates, both of which are safeguards against overfitting.

For the bivariate terms (cf. Fig A.8), there is a strong interaction between $T_{\max}$ and $T_{\min}$: days with hot afternoons and hot nights have much higher cooling demand; there is also some interaction between $T_{\max}$ and day.

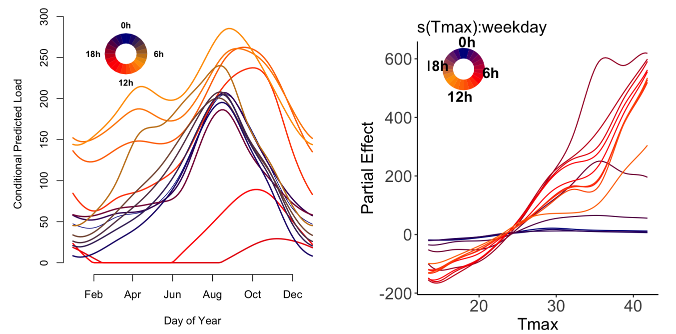

Fig. 4. Partial dependence plots across hours of the day for cooling load at **oceanview** substation in Orange county. *Left panel*: partial dependence on day of the year day; *Right:* partial dependence on $T_{\max}$ during weekdays.

To provide further insight, Figures 4-5 show accumulated local effects (ALEs) [22] to quantify the effect of a couple of key covariates at a different substation. ALEs are more indicative of the total impact of a given covariate and also mitigate extrapolation errors when the covariates are correlated. We observe that temperature effects are muted during the nighttime, and that seasonal effects peak at different hours of the day. Fig 5 shows that while there is the universal hinged dependence of cooling load on $T_{\max}$, there are variations across different substations, including some partial saturation (declining slope at extreme right of the plot) at some locations and inflection points occurring over the broad range of 18-22$°C$, rather than the fixed 18$°C$ Cooling Degree Day standard.

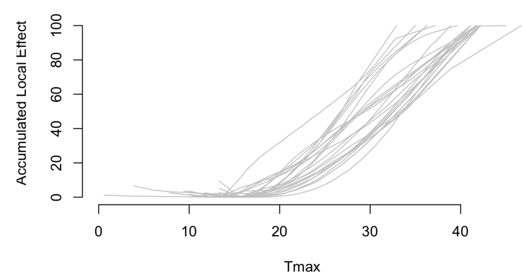

Fig. 5. Normalized accumulated local effects of maximum daily temperature $T_{\max}$ on cooling loads at $hr = $ 6pm across 50 randomly selected substations. The y-axis is standardized to $[0, 100]$ range.

B. Covariate Relevance

A powerful summary statistic to assess variable importance is the effective degrees of freedom (EDF) available for each term of the GAM. An EDF less than 1 indicates a variable that was shrunk to (almost) zero, and hence plays no importance in the fitted model; an EDF $\in [0.9, 1.3]$ indicates essentially a linear fit with minimal curvature. Finally, EDF > 1.3 corresponds to nonlinear dependence, with higher EDF indicating more complex fits.

Figure 6 visualizes the EDFs of a representative bus across hours of the day. We observe dramatic variations. For instance, in the morning and early afternoon, there is essentially no dependence on the interaction `ti(Tmax,day)`, but this becomes significant after 4pm. Similarly, $T_{\min}$ matters in the morning, but not in the afternoon. As might be expected, irradiance is only significant during the daytime hours. In the early morning lagged $T_{\max}$ plays at least as important a role as today's temperature, then progressively loses relevance.

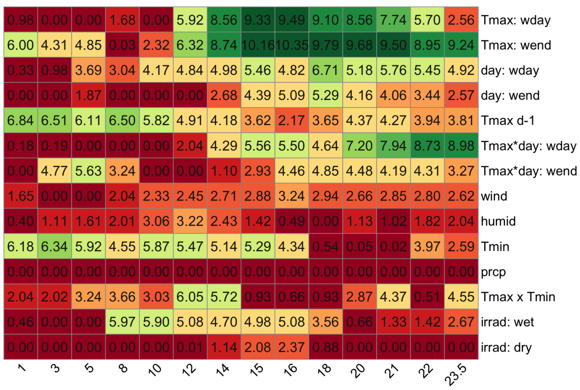

Fig. 6. Effective number of degrees of freedom across hours and each term in (GAM-cool) for cooling loads at **russell** substation in Alameda county. See Figure A.9 for corresponding heating loads.

Comparator Approaches: As a potential comparator, we investigated the use of Function on Scalar Regression (FOSR) models with variable selection, available in the `grpreg` package [23]. FOSR models subsume the hourly dependence on hr into a given basis (e.g., cubic splines ϕ), reducing to the estimation of the coefficients $B \equiv (b_{k,\ell})$ to the multivariate linear regression $\mathbf{Y} = \mathbf{X}B\Phi^\top + E$ where $\mathbf{X}$ is the $n \times (d+1)$ design matrix, Φ is the $h \times K$ basis matrix, and E are the $n \times h$ observation noises. The number of coefficients is $K \times d$ which ends up being several hundred and hence shrinkage techniques are prevalent. For instance, the grMCP strategy [24] groups coefficients corresponding to each covariate to allow their joint shrinkage to zero if the covariate is deemed irrelevant. While common in other contexts, FOSR has two major disadvantages compared to GAM. First, the results are quite sensitive to the shrinkage parameter which is difficult (i.e., expensive) to tune. Second, FOSR tends to use the same bases for all covariates, which makes it much less flexible (without a lot of customization) compared to GAM where it is straightforward to assign different degrees of freedom to different smooths. Finally, FOSR is oriented to linear dependence on covariates and hence is not suited to our setting where we would like to capture nonlinear interactions. The Appendix discusses validation of our GAMs and statistical metrics for comparing different load profile projections.

V. Conclusion

This study contributes to bridging climate science and power systems, by refining the spatial and temporal resolutions to gain more nuanced insights into how the California grid will react to future climate. Our 200K+ GAM models provide a large statistical corpus that contrasts sharply with monolithic grid-wide analyses.

Limitations: load projections have many facets and it is not possible to address all of them simultaneously. The presented study is geared towards grid-wide assessment and brings the following caveats. First, our models are trained on a single year of ResStock output, so may miss out on understanding the impact of some rare weather states; additional training data would better validate our results. Second, the reported projections are based on a single TMY, to be interpreted as the "average" weather conditional on the global climate state. There is substantial inter-annual variability, to the extent that a "hot" 2020 year and a "colder-than-usual" 2035 weather-year look almost identical. Consequently, actual realized loads in 2035 may well be multiple percent above or below our projections. In a separate study, we will address this issue, in particular to test the hypothesis that global warming is also increasing the inter-annual variability. In a related vein, since our focus is on *typical* load profiles in 2035, our results should not be used to study potential extreme weather/load behavior. The latter necessitates a projection methodology that is geared to fitting the tails, additional GCM outputs (extreme years, rather than TMY), and residuals bootstrapping. Third, we have focused on the larger features of load profiles, sidestepping fine-scale temporal modeling. For instance, ResStock output often features sub-hourly patterns witness the rough 15-min training profiles in Figure 1. By construction, our forecasts smooth out those effects.

Our granular appliance- and substation-level analysis lends itself to integrated downstream assessment of grid impacts. In an ongoing collaborative work, we inject our climate impact module into a larger framework that considers socioeconomic scenarios of DER adoption, electrification and EVs, together with overall shift of California electricity demand and generation to provide a more holistic view of future grid evolution. For this purpose, the residential load forecasts are augmented with similar models for commercial and industrial loads, as well as renewable generation capacity factors, to enable the running of fully-realistic grid OPF and dispatch algorithms. A parallel project combines our statistical models with several high-fidelity electrification models [25] (such as adoption of heat pumps) which shift the base loads we use for training and allows to capture the cumulative effect of electrifying residential energy demand and climate change. A further aspect of in-progress work is capacity expansion planning [13] to analyze additional grid resources needed to maintain a reliable grid in the two decades ahead.

VI. AI Usage Disclosure & Acknowledgements

AI chatbots were used to refine figure formatting and to iterate on the GAM model construction. AI assistants were not used for conceptualizing, writing or editing this article.

I thank three anonymous referees for their feedback on the early version. I am grateful to Yuting Ma, Patricia Hidalgo-Gonzalez, Qi Geng, Paul Serna Torre and Sanjay Poudel for discussions that informed model construction; Adam Sedlak was invaluable for generating the GCM inputs and mapping them to CATS substations. This research was partially supported by funds from the Climate Action Research Grants of the University of California, Grant Number R02CP7302.

Appendix

A. GAM Model for Heating Loads

For completeness, the heating GAM-Tweedie specification we employ is:

```
gam(load ~ wet+weekend +                # (GAM-heat)
   s(Tmax, k=10, bs="bs", m=c(3,1), by=weekend)+
   s(Tmin, k=12, bs="bs", m=c(3,1), by=weekend)+
   s(day, k=12, bs="cc", by=weekend) +
   s(irrad, k=8, bs="bs", m=c(3,1), by=wet),
   s(prcp,    k=5, bs="bs", m=c(3,1)) +
   s(TmaxLag, k=8, bs="bs", m=c(3,1)) +
   s(TminLag, k=8, bs="bs", m=c(3,1)) +
   ti(Tmin, day, bs=c("cs","cc"),
                 m=c(1,2), by=weekend) +
   ti(Tmax, Tmin, m=c(1,1)) +
   data = trainDF, select=TRUE,
   family=tw(), method="REML")
```

B. Additional Plots and Tables

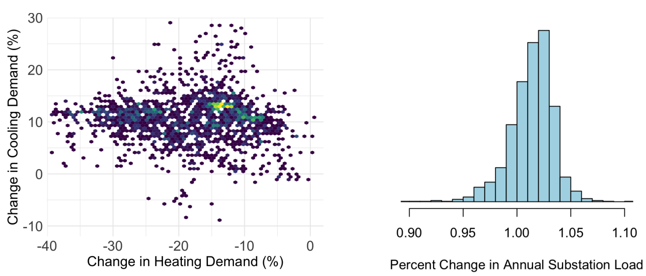

Fig. A.7. Distributions of annual substation load percentage change between 2019 AMY ERA and 2035 TMY EC-Earth3 climate scenarios. *Left:* Heating and Cooling substation loads. *Right*: Total annual substation load.

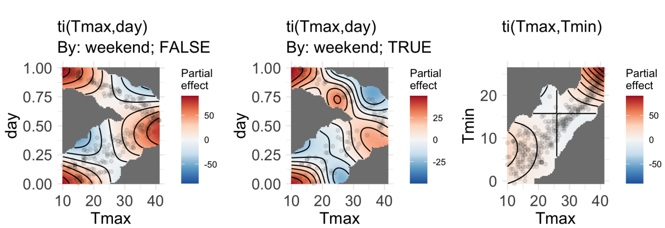

Fig. A.8. Bivariate smooths for SFD cooling loads at 6pm at **parlier** substation.

C. Statistical Model Validation

Our primary loss function for validating model fit is mean absolute deviation (MAD), with RMSE being inappropriate due to larger errors during high-load hours, and percentage errors being inappropriate due to presence of many zero-load hours. Moreover, since we only have a single year of training data, extensive testing is not possible since it would require reserving a substantial chunk of available data. Consequently, as default we utilize a random split using 300 days for training the model and 65 days held out for out-of-sample testing, all cross-validated 5 times.

Since we are interested in feature importance, we seek an interpretable model with shrinkage, so that predictive loss minimization is not the sole goal. As such, visual validation of daily load profiles predicted by the model, and analysis of partial dependence plots are essential components of overall goodness-of-fit checking. A further check is to make sure there is no seasonality (either day-of-year or hour-of-day) among the predictive errors, and a consistent model behavior across substations and building types.

While full validation is beyond the scope of this short version and will be treated in a separate article targeting statistics audience, in Figure A.10 below we present an illustrative residuals plot, and a plot comparing ground truth load versus model prediction for a few in-sample and out-of-sample days. We observe that the largest discrepancies are at the peak hour of the day, and the overall behavior is essentially identical for

TABLE A.2
Average Effective Degrees of Freedom by Weather Covariate across all substations and all hours. The second column reports the percentage of models where this covariate was deemed important (EDF > 1.3).

Covariate	Cooling		Heating	
`Tmax: wday`	5.62	76.0%	3.69	67.7%
`Tmax: wend`	4.94	78.7%	2.95	58.3%
`Tmin: wday`	4.25	89.5%	3.24	67.1%
`Tmin: wend`	–	–	2.83	61.5%
`Tmax` Lag 1	3.28	78.0%	3.01	71.1%
`Tmin` Lag 1	2.67	63.3%	2.13	70.2%
`day` × `wday`	6.71	97.0%	6.28	88.5%
`day` × `wend`	5.46	95.7%	4.27	82.1%
`ti(Tmax,day): wday`	2.97	79.1%	2.68	70.9%
`ti(Tmax,day): wend`	2.60	80.2%	2.24	68.5%
`ti(Tmax,Tmin)`	3.34	87.7%	4.67	90.4%
`wind`	1.37	66.8%	–	–
`humid`	1.35	64.9%	–	–
`prcp`	0.15	10.1%	1.96	74.6%
`irrad: dry`	3.14	80.8%	3.12	84.2%
`irrad: wet`	0.78	38.5%	3.14	77.9%

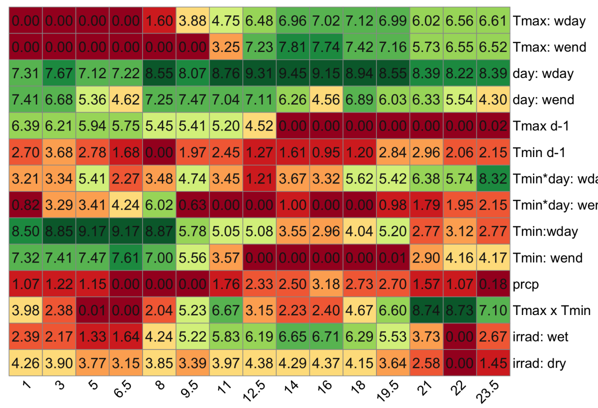

Fig. A.9. Effective number of degrees of freedom across hours and each term in (GAM-cool) for *heating loads* at **russell** substation.

both training and testing samples.

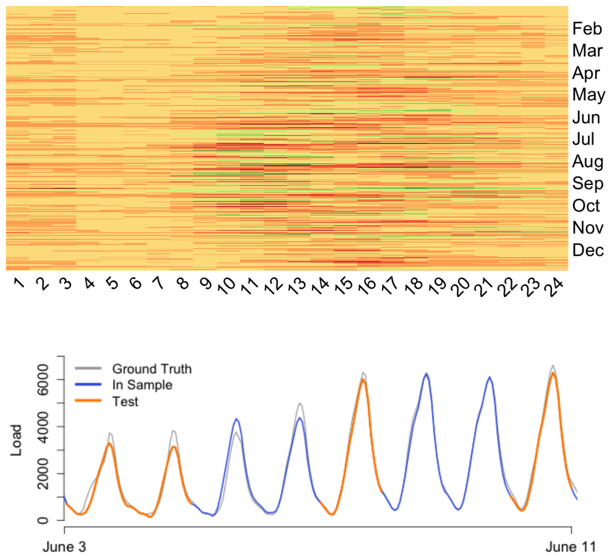

Fig. A.10. *Top:* Heatmap of GAM residuals (actual minus predicted) across hours (x-axis) and days (y-axis). *Bottom*: Actual and model-predicted load for a representative week with both in-sample and out-of-sample days. Cooling SFD loads at **oceanview** substation; 300 training days and 65 test days.

References

[1] A. Antoniadis, J. Cugliari, M. Fasiolo, Y. Goude, and J.-M. Poggi, *Statistical Learning Tools for Electricity Load Forecasting*. Springer, 2024.

[2] C. Capezza, B. Palumbo, Y. Goude, S. N. Wood, and M. Fasiolo, "Additive stacking for disaggregate electricity demand forecasting," *The Annals of Applied Statistics*, vol. 15, no. 2, pp. 727–746, 2021.

[3] J. de Vilmarest, J. Browell, M. Fasiolo, Y. Goude, and O. Wintenberger, "Adaptive probabilistic forecasting of electricity (net-) load," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4154–4163, 2023.

[4] V. Satgé, S. Gabriel, and C. Clastres, "Sensitivity analysis of load profiles: Implications for resource adequacy in future power system," *Energy*, p. 138272, 2025.

[5] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 134–141, 2011.

[6] Y. Goude, R. Nedellec, and N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 440–446, 2013.

[7] J. Browell and M. Fasiolo, "Probabilistic forecasting of regional net-load with conditional extremes and gridded NWP," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5011–5019, 2021.

[8] V. Gioia, M. Fasiolo, J. Browell, and R. Bellio, "Additive covariance matrix models: modeling regional electricity net-demand in Great Britain," *Journal of the American Statistical Association*, vol. 120, no. 549, pp. 107–119, 2025.

[9] M. Zimmermann and F. Ziel, "Efficient mid-term forecasting of hourly electricity load using generalized additive models," *Applied Energy*, vol. 388, p. 125444, 2025.

[10] P. H. T. Sousa, C. P. de Souza, and R. Dias, "Bayesian variable selection for function-on-scalar regression models: a comparative analysis," *arXiv preprint arXiv:2303.03521*, 2023.

[11] J. Chung and B. Jang, "Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data," *PLoS one*, vol. 17, no. 11, p. e0278071, 2022.

[12] S. Taylor, A. Rangarajan, N. Rhodes, J. Snodgrass, B. C. Lesieutre, and L. A. Roald, "California test system (CATS): A geographically accurate test system based on the California grid," *IEEE Transactions on Energy Markets, Policy and Regulation*, vol. 2, no. 1, pp. 107–118, 2023.

[13] P. Serna-Torre, A. Sedlak, Q. Geng, S. Poudel, Y. Min, G. De Roche, M. Singh, M. Ludkovski, C. Fernandes, D. Callaway, J. Kleissl, M. Mildenberger, G. Wu, R. Deshmukh, and P. Hidalgo-Gonzalez, "Impact of socioeconomic residential propensity of adoption of distributed energy resources on the climate resilient grid expansion," *in preparation*, 2026.

[14] E. Present, P. R. White, C. Harris, R. Adhikari, Y. Lou, L. Liu, A. Fontanini, C. Moreno, J. Robertson, and J. Maguire, "Resstock dataset 2024.1 documentation," tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2024.

[15] J. Reyna, A. Fontanini, E. Present, L. Liu, R. Adhikari, C. Bianchi, J. Brossman, R. Chintala, K. Clark, C. Harris, *et al.*, "Resstock technical reference documentation (v.3.3.0)," tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO (United States), 07 2025.

[16] W. Krantz, D. Pierce, N. Goldenson, and D. Cayan, "Memorandum on evaluating global climate models for studying regional climate change in California," tech. rep., UC San Diego's Scripps Institution of Oceanography, 2021.

[17] D. B. Walton, F. Sun, A. Hall, and S. Capps, "A hybrid dynamical–statistical downscaling technique. Part I: Development and validation of the technique," *Journal of Climate*, vol. 28, no. 12, pp. 4597–4617, 2015.

[18] Cal-Adapt: Analytics Engine, "Climakitae: A powerful Python toolkit for climate data analysis and retrieval with data from the Cal-Adapt analytics engine.," 2026.

[19] S. Wilcox and W. Marion, "Users manual for TMY3 data sets," tech. rep., National Renewable Energy Laboratory, 2008. Technical Report NREL/TP-581-43156.

[20] S. N. Wood, "Generalized additive models," *Annual Review of Statistics and Its Application*, vol. 12, no. 1, pp. 497–526, 2025.

[21] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society (B)*, vol. 73, no. 1, pp. 3–36, 2011.

[22] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," *Journal of the Royal Statistical Society (B)*, vol. 82, no. 4, pp. 1059–1086, 2020.

[23] P. Breheny, Y. Zeng, and R. Kurth, "Regularization paths for regression models with grouped covariates," *R package version*, pp. 3–1, 2018.

[24] J. Huang, P. Breheny, and S. Ma, "A selective review of group selection in high-dimensional models," *Statistical Science*, vol. 27, no. 4, pp. 481–499, 2012.

[25] Y. Min, M. Singh, G. De Roche, S. Poudel, P. Hidalgo-Gonzalez, R. Deshmukh, M. Mildenberger, and G. Wu, "Survey-based residential electrification adoption pathways and equity implications," *in preparation*, 2026.

I. Review #49A

A. Paper summary

The authors describe a model for hourly (fractional hour) residential load forecasting considering impacts of climate change. They use a bottom approach based on the residential housing stock data and appliance models.

B. Comments for authors

It is certainly challenging and important to understand how load profiles and, in particular, peak loads, may change in the future. The authors develop a bottom up approach using housing and residental appliance component models. This is certainly not new and many such tools exist today. Their particular statistical approach appears valid and may provide improvement over such models. These models tend to be useful for specific studies but do not provide much in the way of reliable forecasting. There are too many variables (climate change, technology changes, economic development in particlar areas, and so on). Nothing in the results struck this reviewer as particular surprising for a 10 year horizon, i.e., the growth in mean hourly load, daily peak load, or seasonal shifts were well within expectations. The presented work did not explore possible impacts on the grid, although this was mentioned as on-going efforts.

C. Summarize your recommendation in one to two sentences

The others are analyzing an important problem. The reviewer finds the results reasonable but the contribution modest.

— Kevin Tomsovic

II. Review #49B

A. Paper summary

This paper develops a statistical pipeline to project how climate change will affect residential electricity load at the substation level across the California grid. Generalized Additive Models are trained on outputs from the ResStock building simulation tool and then evaluated under 2035 weather conditions drawn from four downscaled global climate models, with attention to extrapolation beyond the training weather range and correction of spatial weather mismatches between substations and PUMA-level inputs. The findings provide spatially resolved climate impact projections at the substation level for the CAISO grid.

B. Comments for authors

The paper is a well-executed applied contribution. The methodological choices are well-motivated and the spatially resolved findings are useful for grid planning. The main weakness is that the methodological contribution is limited: the paper applies existing statistical tools thoughtfully but does not propose anything new. This should be framed as an applied rather than methods paper. Validation seems limited since there is no held-out predictive test against observed loads or a separate ResStock year and the spatial disaggregation at the substation level depends on another methodology that is deferred to a companion publication. The spread in projections across GCMs deserves more analysis given its direct implications for grid planning under uncertainty.

C. Summarize your recommendation in one to two sentences

An applied paper that delivers useful, spatially resolved climate impact projections for the CAISO grid. Recommend acceptance, noting that the paper should be framed explicitly as an applied contribution and that predictive validation against held-out data would strengthen the results.

— Anonymous reviewer

III. Review #49C

A. Paper summary

This paper presents an approach for climate-driven forecasting of substation-level residential heating and cooling loads for CAISO in 2035 using downscaled GCM climate data, ResStock data, and other building footprint data. The proposed approach uses generalized additive models (GAMs) regressed on daily metrological weather variables to capture load behavior and develops 220,000 individual load models covering more than 1,800 substations, five building types, two load types (heating and cooling), and 12 hourly load models.

B. *Comments for authors*

This is an interesting paper with strong motivation and a meaningful contribution to the power and energy systems community. The use case at the scale of the CAISO grid is comprehensive and useful for informing grid planning studies.

The results are comprehensive and the discussion is detailed and clear. The important modeling implications and caveats are stated very clearly, providing a good understanding of what the model implies and where improvements may be needed.

The paper could be improved by highlighting how the results of this study can be validated. For example, it would help to demonstrate a benchmark case for the model and show how well it reproduces past observations. It would also be useful to include some discussion of how such long-term forecasting models are typically validated.

C. *Summarize your recommendation in one to two sentences*

This is a strong and well-motivated paper with a comprehensive CAISO-scale use case and clear discussion of modeling implications and caveats. The main area for improvement is validation, and the paper would benefit from a clearer discussion of benchmark cases, reproduction of past observations, and standard approaches for validating long-term forecasting models.

— Anamika Dubey

IV. Brief response to reviewers (optional)

Thank you for the insightful comments. Minor changes were done to the body of the paper and Appendix Section C was added, see summary of changes.

V. Brief summary of changes made to the final paper

I greatly appreciate the reviewers' constructive feedback across the 3 anonymous reports.

In general, there were no major changes made in the final version. In response to reviews, it was clarified that this is first and foremost an applied work that summarizes the overall pipeline and the main findings from running it on the CAISO region and respective housing stock. Further methodological details and validation will appear in a dedicated statistical publication, and further discussion and implication for grid impact will appear in a forthcoming large-team publication. Besides clarifying several passages, updating and de-anonymizing the references, and fixing minor typos, the biggest change is to add a new section to the Appendix outlining the validation of the GAM models, and illustrating model goodness-of-fit via Fig A.10.