Bayesian probabilistic heat-health warnings across all California neighborhoods for equitable resource allocation

Supplementary Information

Supplementary Methods: Estimation procedure

Here we describe some of the mathematical details behind the modeling that forms the basis of the heat-warning system.

Methodology overview

The sociodemographic covariates are standardized to have mean zero and a standard deviation of one across ZIP codes. Heat is standardized with a single statewide daily mean and standard deviation, so the chronic term $\bar{H}_i^{\text{scale}}$ has mean zero but a smaller across-ZIP spread (SD ≈ 0.62). This places covariates on a common scale, which aids in prior specification for the integrated nested Laplace approximation (INLA) models fit below. The main pipeline is carried out in three steps, as described below (Supplementary Figure 1).

- The first step deals with data preparation. We let i denote ZIP codes and d denote days, and write H_{id}^{scale} for the standardized daily maximum heat index. We break up heat into a between/within decomposition, noting

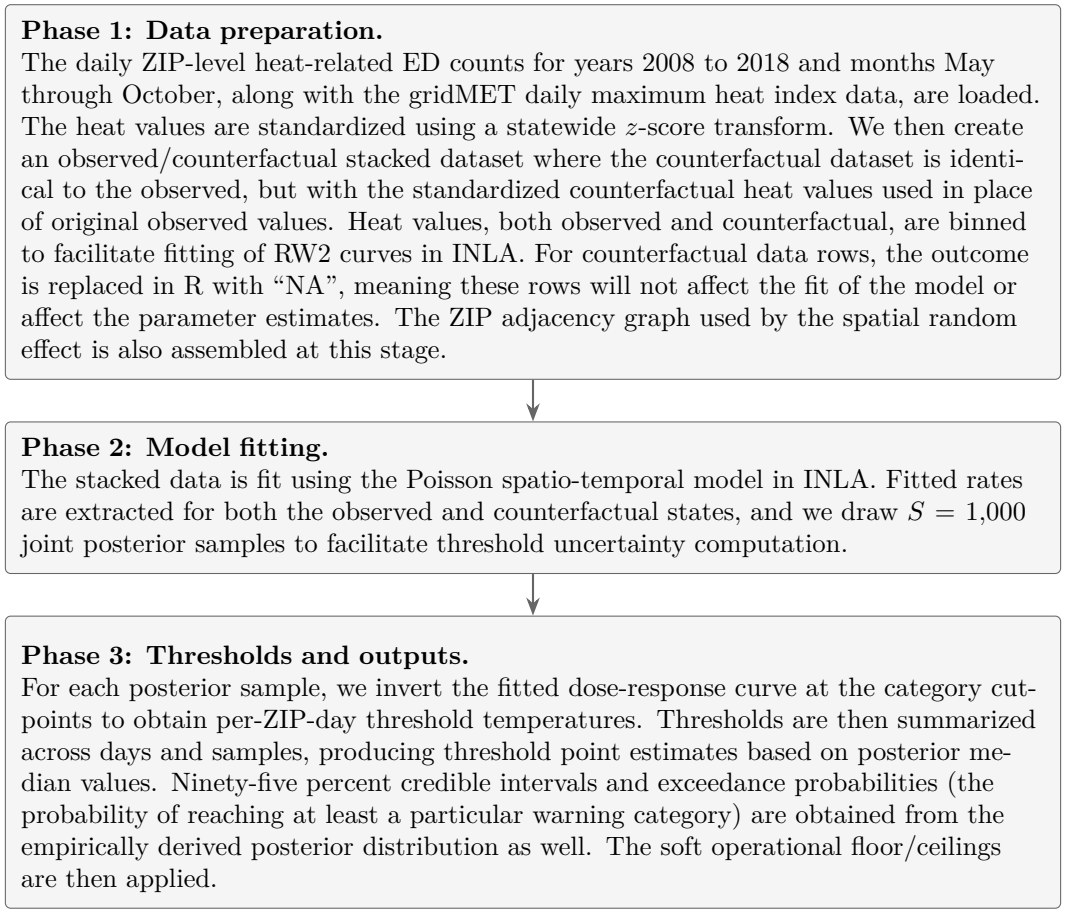
$$H_{id}^{\text{scale}} = \underbrace{\bar{H}_i^{\text{scale}}}_{\text{chronic}} + \underbrace{\left(H_{id}^{\text{scale}} - \bar{H}_i^{\text{scale}}\right)}_{\text{within-ZIP (acute)}}, \tag{1}$$

  where $\bar{H}_i^{\text{scale}}$ is the ZIP's study-period mean. This decomposition breaks up heat into chronic long-term exposure and within-ZIP exposure. We also construct a 31-year per-(ZIP, month) counterfactual exposure to be used for counterfactual comparison, and a ZIP-level spatial adjacency matrix used for the spatially smoothed random effects.

- We model emergency department (ED) counts via a Poisson log-link spatio-temporal model, which is fit using the computationally efficient INLA package[1]. We fit the chronic heat $\gamma_1 \bar{H}_i^{\text{scale}}$ as a fixed-effect covariate which modifies the baseline (intercept) log ED rate at the ZIP level. The acute exposure is a ZIP-level demeaned exposure, namely the daily maximum heat for a particular ZIP minus the 11-year average max heat value. The acute exposure enters the model through a nonlinear second-order random-walk (RW2) dose-response model that is shared across all ZIPs and allowed to deviate at the climate-zone level. The chronic exposure enters the intercept, or baseline log-ED rate, which also carries a spatially smoothed Besag–York–Mollié (BYM2) random effect. This fitted model produces posterior samples corresponding to both the observed and counterfactual exposures. The attributable-risk (AR) values are computed, with uncertainty, from the sampled parameters.

- In the last step, we invert the fitted model to get a temperature threshold for each ZIP-day combination. The daily thresholds are obtained for each set of sampled parameters described above. From these sampled values we tabulate exceedance probabilities by ZIP, warning

category, month, and temperature, and then apply the operational floor/ceiling anchoring to obtain the operational thresholds.



Supplementary Figure 1: Overview of estimation process. The first phase prepares the inputs for the model. The second phase fits the Bayesian spatio-temporal model to the stacked observed and counterfactual data. The third phase converts the fitted dose-response into per-ZIP threshold temperatures, exceedance probabilities, and operational warning levels.

Spatial confounding

The within-between split described above also aids in assessing concerns related to spatial confounding, a situation that occurs when a spatially informed random effect can alias the effect of the main exposure in question. Here we incorporate the chronic, long-term heat exposure $\bar{H}_i^{\text{scale}}$ into the "baseline" or intercept as $\gamma_1 \bar{H}_i^{\text{scale}}$, and then express the acute exposure as $H_{id}^{\text{scale}} - \bar{H}_i^{\text{scale}}$, which represents the "slope" term, or effect of heat, in the dose-response function. This construct creates a situation where the deviation between heat and the group mean is orthogonal to any ZIP-constant quantity, and thus helps deal with model assumption violations that occur when random effects and exposures correlate. See Bafumi and Gelman[2] and Bell et al.[3] for examples of this within-between modeling approach. Though the approach is generally formulated with i.i.d. random effects, we find that it is effective for spatially varying random effects as well. In particular, the correlation between the joint BYM2 intercept term (combined i.i.d. and spatial effects)

and the centered exposure is essentially zero across all 3,412,464 fitted observations. Across the 1,686 ZIP codes we find the BYM2 random intercept is likewise essentially uncorrelated with mean heat exposure $\bar{H}_i$: Pearson $r = -0.018$. We thus see little sign that chronic heat is being aliased by the random effect term, indicating virtually no evidence of spatial confounding. Finally, the attributable-risk computation behind our thresholds depends on the chronic-heat term $\gamma_1 \bar{H}_i^{\text{scale}}$ and the BYM2 spatial intercept s_i only through their sum $\alpha_i^{\text{H}} = \gamma_1 \bar{H}_i^{\text{scale}} + s_i$, and never on the two separately. Our thresholds therefore depend on the identified total.

Counterfactual construction

The counterfactual exposure represents the temperature that individuals in a particular ZIP code during a particular month would experience under "normal" conditions, where normal is defined as the 31-year 1995–2025 average of daily maximum heat index values for that ZIP code and calendar month, denoted $H_{im}^{\text{cf,raw}}$. Using this counterfactual exposure, we estimate, on each day in each ZIP code, the modeled expected ED count under observed and counterfactual exposure. This can easily be achieved by "stacking" the data. We first take the original dataset, and create the counterfactual dataset, which is identical to the original but, for each row, we replace the exposure H_{im} with its counterfactual value, which, incidentally, is standardized in the same manner as original exposures. The resulting dataset is of size $2N$, where N is the size of the original data. Rows $i = 1, \ldots, N$ represent original data while rows $i = (N+1), \ldots, 2N$ represent counterfactual data, with outcomes for rows $(N+1)$ to $2N$ replaced by "NA", so they do not contribute to the model fit. The difference between estimated counts for rows 1 versus $N+1$, 2 versus $N+2$, and so on represents the modeled excess counts used to compute the attributable risk. Note that each ZIP code's study-period mean exposure is left as is, so the chronic term is untouched and only the acute exposure differs between the two halves of the stack.

Multilevel model and RW2 binning

The non-linear RW2 curve fitted to establish the dose-response heat/health relationship cannot, for matters of computational tractability, be fit on all exposure/outcome values. INLA's `inla.group` function provides a "cut" method for creating the binning needed to fit the dose-response curve accurately but in a computationally feasible manner. Here we follow that same cut method, implemented directly in our pipeline and performed upstream of the model fit, using 50 equal-width bins spanning the observed range of the demeaned heat values, with the median of values within each bin used to fit the curve. These bin values are shared across all ZIPs, so the shared curve and each climate-zone deviation sit on the same lattice.

We express our model in multilevel form. Let i denote ZIP codes, d days, $\text{cz}[i]$ the climate zone to which ZIP i belongs, n_i the population offset, H_{id} and $\bar{H}_i$ the *standardized* exposure quantities, and Y_{id} the aggregate count of heat-related ED visits in ZIP i on day d. The model is,

$$Y_{id} \sim \text{Poisson}\big(n_i \, \mu_{id}\big), \tag{2}$$

$$\begin{aligned} \log \mu_{id} = {} & \alpha_i + \beta_{y(d)} + \beta_{w(d)} + \beta_{m(d)} + f_{\text{doy}}(d) \\ & + f_{\text{heat},\,\text{cz}[i]}\big(H_{id} - \bar{H}_i\big), \end{aligned} \tag{3}$$

$$\alpha_i = \gamma_0 + \gamma_1 \bar{H}_i + \mathbf{x}_i' \boldsymbol{\beta}_{\text{sociodem}} + s_i, \tag{4}$$

$$f_{\text{heat},\,\text{cz}[i]}(\cdot) = f_{\text{shared}}(\cdot) + f_{\text{dev},\,\text{cz}[i]}(\cdot). \tag{5}$$

The ZIP-level intercept α_i combines the chronic-heat fixed effect $\gamma_1 \bar{H}_i$, the sociodemographic vector $\mathbf{x}_i$ denoting seven scaled American Community Survey (ACS) covariates consisting of low-income, English-limited, Hispanic, non-Hispanic Black, outdoor-work, no-insurance, and over-65 percentages, along with the spatial random effect s_i consisting of the i.i.d. plus spatial BYM2 random effects. The likelihood is offset by temporal terms denoting year, day-of-week, and month fixed effects, along with a day-of-year RW2 smoother, f_{doy}. The effect of heat, $f_{\text{heat, cz}[i]}$, consists of a shared RW2 smooth term f_{shared} common to all ZIPs, and a climate-zone-specific deviation $f_{\text{dev, cz}[i]}$ shared by all ZIPs in the same California climate zone (30 zones statewide). The climate-zone heat-effect curves are not subject to spatial smoothing due to their relatively large geographies; as such, spatial borrowing-of-strength constructs are not particularly beneficial and would impose unnecessary modeling assumptions with little benefit. We instead give the deviations an exchangeable prior across zones, and for identifiability we impose sum-to-zero constraints on the shared curve, each deviation, the seasonal smoother, and the BYM2 intercept.

Model adequacy: Poisson goodness-of-fit

We chose to fit a Poisson likelihood, which is common for small-area estimation, as opposed to a negative-binomial alternative, which contains an extra random effect term to model extra-Poisson variability. The heat-ED counts are sparse, and the raw marginal variance-to-mean ratio of 1.155 goes to near unity once we condition on the rich set of random effects in our model, as evidenced by the residual overdispersion of 2.6%, with a Pearson dispersion statistic of 1.07. The observed distribution of ED counts is also well estimated. (See Supplementary Table 1.) We thus judged that the extra random term the negative-binomial likelihood provides is unnecessary here, given the existing random effects already in the model, and its inclusion could introduce identifiability concerns and degrade interpretation of model parameters.

Count	Observed	Poisson-predicted
$Y = 0$	0.9198	0.9186
$Y = 1$	0.0719	0.0737
$Y = 2$	0.0073	0.0069
$Y \geq 3$	0.00101	0.00077

Supplementary Table 1: Observed count distribution versus the masses implied by the fitted per-row Poisson($\hat{\lambda}_{id}$). There is strong correspondence between fitted and observed ED counts, with the exception of the rare high counts represented by the upper tail, where the Poisson fit slightly under-predicts these spikes.

Threshold inversion per day

A heat warning level is triggered when the temperature crosses a threshold defined by attributable risk (AR) at the ZIP-day level, with level k defined by a fixed AR cutpoint a_k. In particular, for ZIP i, day d and temperature T, the AR can be broken into,

$$\text{AR}_{id}(T) \;=\; \mu_{id}^{\text{cf}} \big(\text{RR}_i(T) - 1\big), \tag{6}$$

with μ_{id}^{cf} denoting the daily fitted counterfactual rate, which varies with day-of-year via the seasonal smoother $f_{\text{doy}}(d)$, and $\text{RR}_i(T)$ the fitted rate ratio. The multiplicative nature of the model means the baseline ED count estimate cancels in the rate ratio, so the RR is fixed across all days of

the month for a given ZIP code. While all ZIP codes within a climate zone share the same dose-response curve formulation, the curve is evaluated at each ZIP code's own centered exposure values. Solving $\mathrm{AR}_{id}(T) = a_k$ for the threshold $T_{id}^{(k)}$ and dividing through by the baseline, we obtain the day-specific RR value,

$$\mathrm{RR}_i(T_{id}^{(k)}) \;=\; 1 + \frac{a_k}{\mu_{id}^{\mathrm{cf}}}. \tag{7}$$

This RR value is inverted to obtain $T_{id}^{(k)}$ in °F. We invert on each day as μ_{id}^{cf} is larger in summer than in spring, which results in lower thresholds for the same excess target. The analytic pipeline produces summaries of $T_{id}^{(k)}$ across days within ZIP over the May–October study season. Medians of the posterior threshold distributions are used as the threshold estimates, prior to the floor/ceiling anchoring described below.

Unit invariance of thresholds

While the AR values are expressed in terms of per-10,000 population units, the thresholds derived are invariant to this population scaling. In particular $\mathrm{AR}_i(T) = \mathrm{rd\_rate}_i(T) \cdot (c/s)$, where c denotes the chosen population unit, here $c = 10{,}000$, s denotes the offset scale used in the Poisson model, $p_k(\cdot)$ denotes the preset 20th, 40th, 60th or 80th percentile of its argument, forming the basis of the thresholds for each warning level $k = 1, \ldots, 4$, and $\mathrm{rd\_rate}_i$ denotes the difference between the modeled visit rate under the observed temperature and under the counterfactual, on that offset scale. Since $\mathrm{AR}_i(T) = \mathrm{rd\_rate}_i(T) \cdot (c/s)$, $\mathrm{rd\_rate}_i(T) = \mathrm{AR}_i(T)/(c/s)$. The cutpoint a_k is the corresponding percentile of the pooled positive AR distribution, $a_k = p_k(\mathrm{rd\_rate}_{\mathrm{pooled}} \cdot (c/s))$. Since a percentile p_k scales with its argument, $a_k = p_k(\mathrm{rd\_rate}_{\mathrm{pooled}}) \cdot (c/s)$. Further since, for cutpoint a_k at level k, $\mathrm{rd\_rate}_i(T_i^{(k)}) = a_k/(c/s)$, the c/s term cancels, meaning that the threshold is thus invariant to scaling. The argument holds separately on every day, so population scaling cancels for each $T_{id}^{(k)}$. Fixed AR cutpoints set by policy consideration are thus likewise invariant, as a cutpoint based on AR ≥ 1 per 10,000 is identical to AR ≥ 0.1 per 1,000.

Posterior sampling chain

One of the main features of our warning system is the uncertainty estimated for each threshold produced at each warning level. Such uncertainty is obtained by drawing $S = 1{,}000$ samples from INLA's joint estimated posterior distribution using its built-in sampling routine. From this we extract samples of the joint latent field and samples for all modeled parameters. For each s, $s = 1, \ldots, S$, we recompute the RW2 curves to obtain per-day $\mathrm{RR}_i^{(s)}$, then per-day $T_{id}^{(k)(s)}$. From these we obtain point and interval estimates as percentiles (median, 2.5% and 97.5%) of the sampled posterior distributions. All threshold estimates come from these sampled distributional values.

Exceedance probabilities

For each ZIP code, warning category and month we calculate the exceedance probabilities using the posterior samples, on a grid of temperatures running from 60°F to 120°F in increments of one degree Fahrenheit. The exceedance probability represents the probability that a per-day threshold falls below temperature T, and we estimate it as the fraction of (sample × day) pairs whose threshold falls below T. Pooling over both dimensions means these probabilities carry the posterior dose-response uncertainty and the seasonal day-to-day movement in the baseline rate at once. The tabulated probabilities are obtained directly from the sampled thresholds and no post-hoc parametric form is fitted on the sampled output.

Operational floor/ceiling anchoring

For operational deployment of the heat-warning system, we "softly" adjust per-category heat thresholds toward two bounds drawn from the heat-warning literature, letting each threshold's uncertainty govern how far it moves. Consider the lower end and the "Moderate" warning level. By default, a ZIP goes into this category if the observed temperature exceeds the posterior median of the threshold distribution. Effectively, this means a day is designated as Moderate if the exceedance probability of reaching that category is greater than or equal to 50%. However, if this median threshold falls below the Moderate floor of 75°F (say it sits at 70°F), then we instead consider the temperature corresponding to a 90% probability of the Moderate level, which may be, say, 78°F. If this 90%-based value is above the floor, as here, it is lowered down to floor level. If it is below the floor, the threshold is raised only that far. This approach ensures that no threshold is lifted above the floor, but also that not all thresholds below the floor are uniformly raised to it. Some, due to climate and various aspects of model uncertainty, will be raised more towards the floor than others. On the upper end, a ceiling and a 10% exceedance probability are imposed in mirror fashion. This asymmetry is on purpose. At the upper end, in Extreme conditions, where the failure of, say, air conditioning can be catastrophic, we effectively lower overly high thresholds toward the ceiling. At the lower end we raise thresholds that would otherwise be too sensitive, as these are based on daily maximum heat values, which are spotty and less persistent on the coast.

This uncertainty-based floor/ceiling approach leaves most thresholds alone. Across all 6,744 thresholds (1,686 ZIPs × four levels), the rule adjusts 15.8% of the thresholds. Approximately 15.4% are lifted toward the floor and 0.4% pulled toward the ceiling, leaving the remaining 84.2% at the posterior median. At Level 4 (Extreme), the rule modifies 204 thresholds, or about 12%. Of those, 144 ZIPs (largely coastal) are raised to the literature floor, 42 are given a partial lift, and 18 (largely desert) are pulled toward the ceiling.

Supplementary References

[1] Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* **71**, 319–392 (2009).

[2] Bafumi, J. & Gelman, A. Fitting multilevel models when predictors and group effects correlate. Working paper, prepared for the 2006 Annual Meeting of the Midwest Political Science Association, Chicago, IL (2006). URL https://ssrn.com/abstract=1010095.

[3] Bell, A., Fairbrother, M. & Jones, K. Fixed and random effects models: Making an informed choice. *Quality & Quantity* **53**, 1051–1074 (2019).