Spatial-temporal Poisson Regression Model with Spatial Components using INLA

October 29, 2024-provided with Heat-Health Algorithms deliverable due 11/1/24

John Molitor, Diane Garcia-Gonzales, David P. Eisenman, Michael Jerrett, Katherine McNamara, Sydney Monte-Sano, Larry Kalkstein, Marinelle Villanueva.

Four Definitions for Heat-Health Metric (see Deliverable 1)

Definition 2 used for current analyses

Health Outcome	ICD-9	ICD-10	Definition 1	Definition 2	Definition 3	Definition 4
1. Diseases of the circulatory system	390-459, 785	I00-I99; R00 R00.2, R57.9, R57.0, R6.21, R57.8, R57.1	X			
1.1 Ischemic heart disease	410-414	I20-I25	X		X	X
1.2 Acute myocardial infarction	410	I21	X		X	X
1.3 Cardiac dysrhythmias	427	I47-I49	X		X	X
1.4 Heart failure	428	I50	X		X	
1.5 Hypotension	458	I95	X		X	X
1.6 Essential hypertension	401	I16	X			
1.7 Hemorrhagic stroke	430-432	I60-I62	X			
1.8 Aneurysm	441-442	I71-I72	X			
1.9 Ischemic stroke	433-436	I63-I65, I67	X		X	X
2. Diseases of the respiratory system	460-519	J00-J94, J96-99	X		X	X
2.1 Pneumonia	480-486	J12-J18	X		X	X
2.2 Asthma	493	J45	X		X	X
2.3 Chronic bronchitis/ emphysema	491-492	J41-J44	X		X	X
3. Endocrine, nutritional, and metabolic disorders	249, 250, 253.5	E08-E13, E23.2	X		X	X
3.1 Diabetes	250	E08-E11, E13	X		X	X
4. Dehydration	276	E86	X	X	X	X
5. Diseases of the genitourinary system	584-586	N17-N19	X	X	X	X
5.1 Acute renal failure	584, 586	N17, N19	X	X	X	X
6. Heat illness and heat stroke	992	T67	X	X	X	X

Model Components

A **random walk** describes a process where something like ER visits changes gradually over time, with each new value being based on the previous value plus a bit of randomness. We apply a random walk to track seasonal trends to capture smooth, natural changes in ER visits across the year (increases during the flu season). Using a random walk ensures that the model reflects long-term seasonal patterns without being distracted by sudden, day-to-day fluctuations.

Besag-York Mollie (BYM2 model) smooths out random noise in the data while keeping meaningful patterns. It gives more reliable estimates, especially for places with small populations or sparse data. This way, we avoid overreacting to areas with unusually high or low rates due to chance, and instead focus on real patterns.

Random effects allow a model to handle natural differences between groups or areas without overfitting to each one. They help us get more accurate, balanced results by recognizing that not everything needs to be explained in detail—some differences are just part of the variability in the real world. On the other hand, fixed effects are like constants in a model – they capture specific characteristics of a group or factor that do not change over time or between conditions.

Model Formula w/ Components

$$\log(\lambda_i) = \boxed{\beta_{0z_i}} + \boxed{\mathrm{f}(\mathrm{doy_i})} + \boxed{\mathrm{dow.fct_i}} + \boxed{\beta_{1,z_i}} heat_i + \boxed{f(heat_i)}$$

Intercept with spatial smoothing

Random Walk for DoY

Fixed Effects

Slope coefficient with spatial smoothing

Interaction term

The model has been used on historic weather data but can also be used on future weather data, real or forecasted.

Model Formula

$$\log(\lambda_i) = \beta_{0z_i} + \mathrm{f}(\mathrm{doy_i}) + \mathrm{dow.fct_i} + \beta_{1,z_i} heat_i + f(heat_i)$$

λ_i: Expected number of ER visits for observation *i*

B_{0zi}: Intercept (baseline ER visit rate) can vary with smoothing

$f(\mathrm{doy_i})$: Random walk component modeling the effect of the day of the year (captures seasonal trends).

$\mathrm{dow.fct_i}$: Categorical fixed effect for the day of the week.

B_{1zi}: Effect of heat exposure with spatial smoothing.

$f(\mathrm{heat}_i)$: Random walk component on heat.

Model Formula

$$\log(\lambda_i) = \beta_{0z_i} + \mathrm{f}(\mathrm{doy_i}) + \mathrm{dow.fct_i} + \beta_{1,z_i} heat_i + f(heat_i)$$

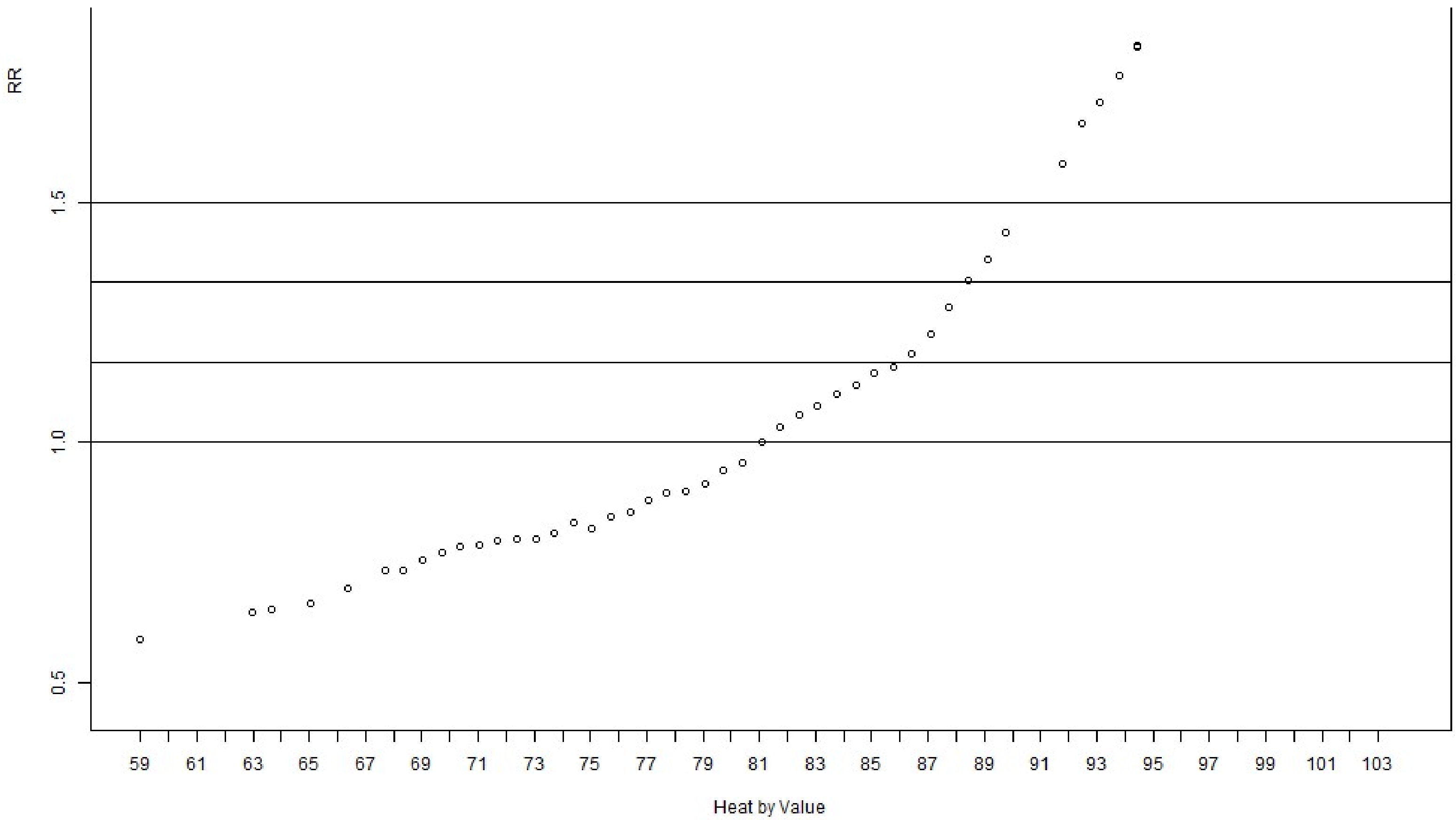
λ_i: Expected number of ER visits for observation *i*

B_{0zi}: Intercept (baseline ER visit rate) can vary with smoothing

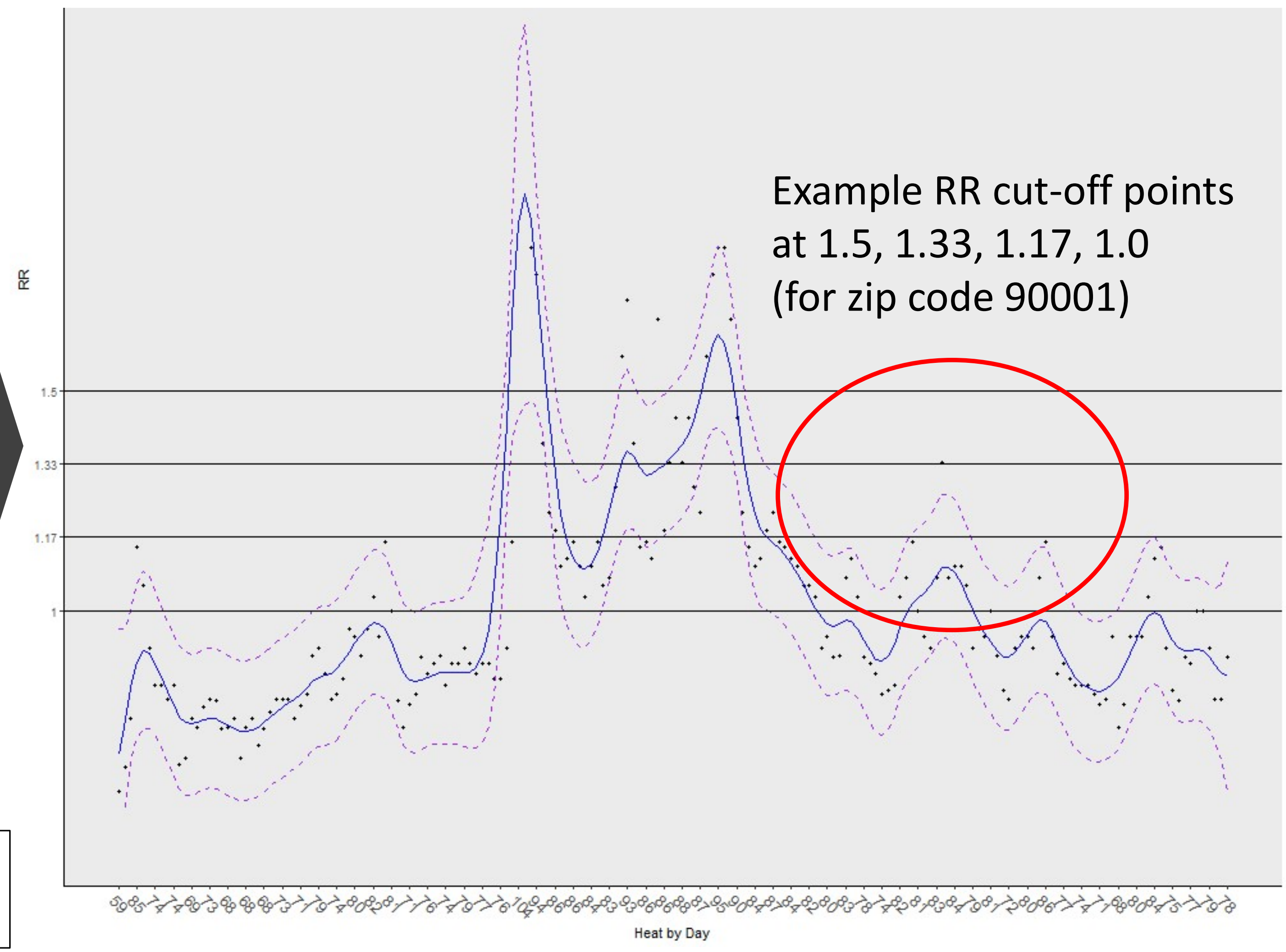
To get Relative Risk, we take the observed (modelled) ER visits data (β_{0Zi}) and divide by the expected ER visits (λ_i) for an average temperature for that day in that zip code over the total data timeframe.

We can adjust apparent temperature for the expected ER visits in the RR estimate.

Relative Risk Outputs for one Zip Code (90001) ordered by increasing daily apparent temperatures



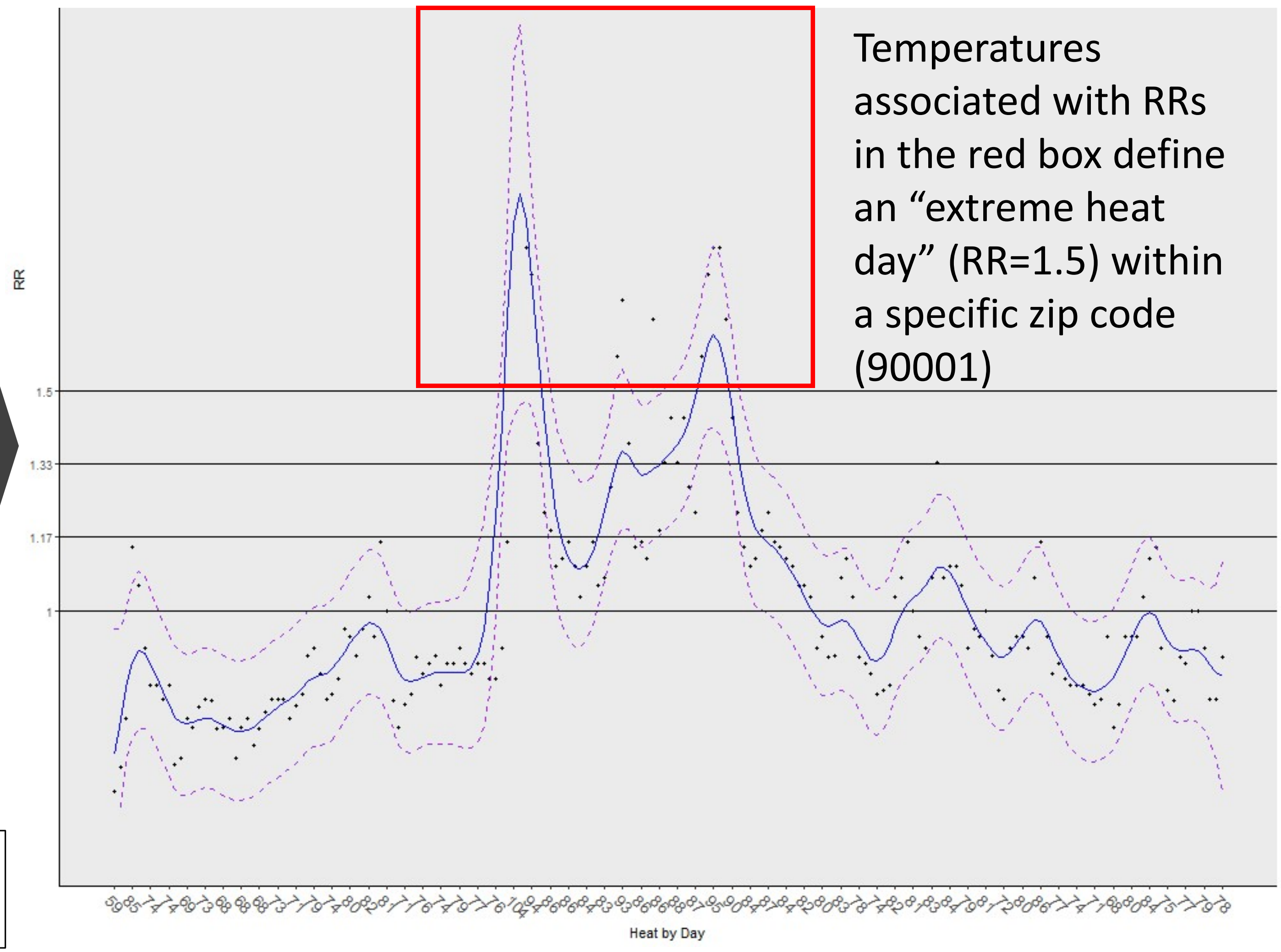
Relative Risk Outputs (by day, zip)



Time series of one heat season (May to October)

Dark Blue Line = Central Estimate of Relative Risk
Top Purple Line = Upper Confidence Limit
Bottom Purple Line = Lower Confidence Limit

Relative Risk Outputs (by day, zip)



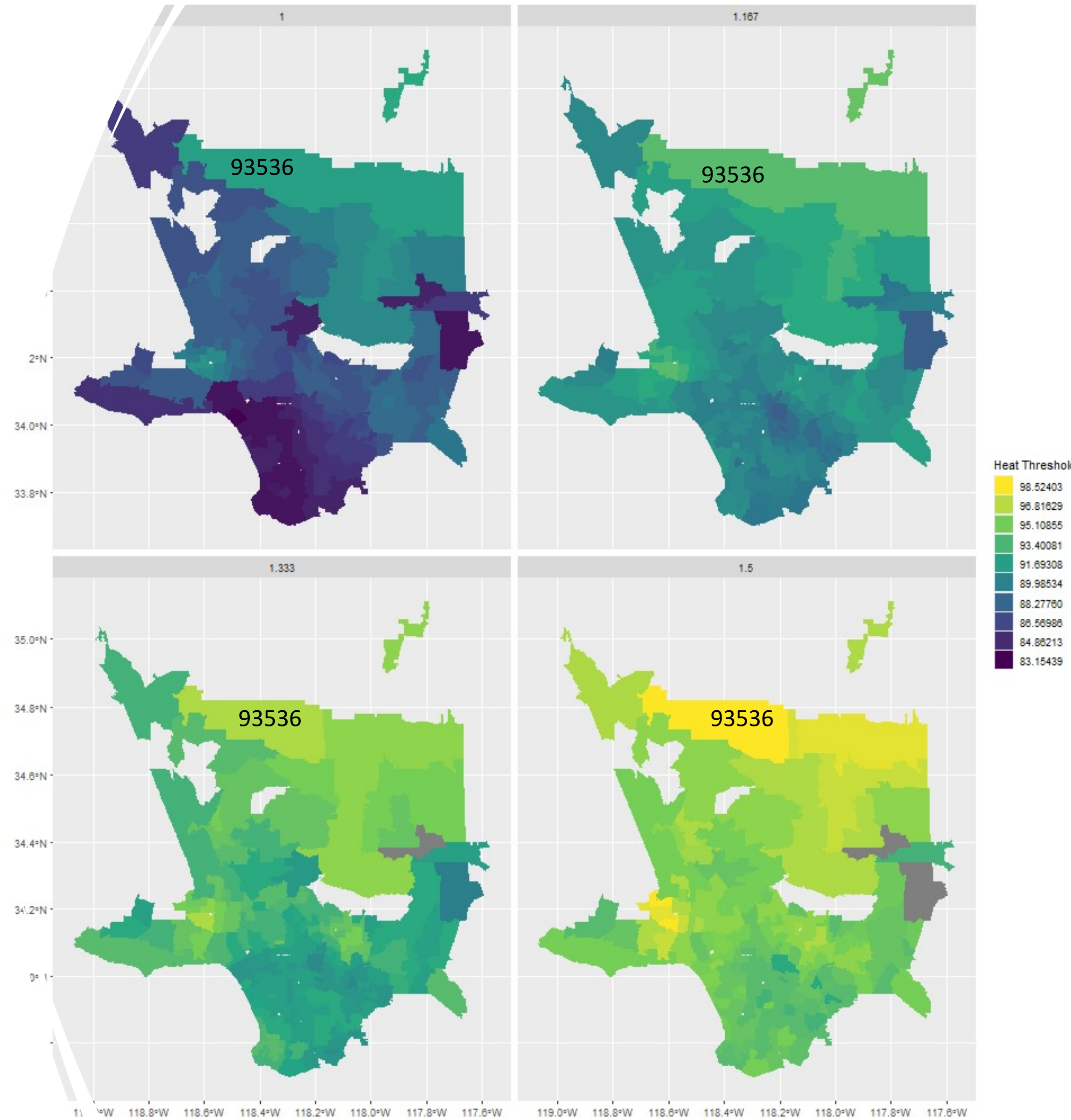
Temperatures associated with RRs in the red box define an “extreme heat day” (RR=1.5) within a specific zip code (90001)

Time series of one heat season (May to October)

Dark Blue Line = Central Estimate of Relative Risk
Top Purple Line = Upper Confidence Limit
Bottom Purple Line = Lower Confidence Limit

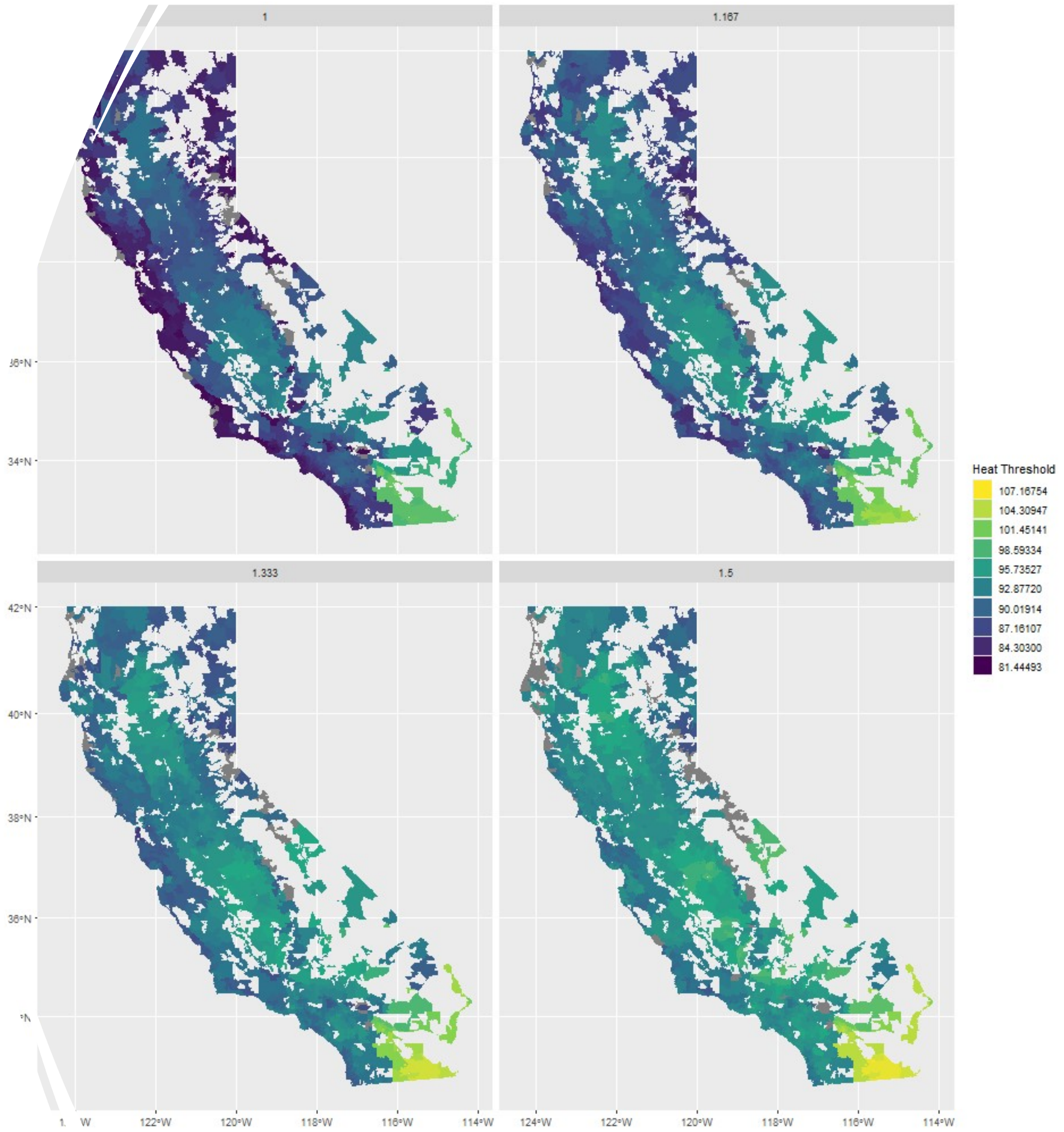
Map of Temperature Warning Thresholds for each Zipcode, by RR

- Using our model, CalHeat Score system can assign apparent temperatures for the different heat warning levels by selecting RRs for the model, at zip code level.
  - For zipcode 93536: an AT of 96 leads to a RR=1.33 and AT=98 leads to a RR=1.5
- Uses 2016 – 2018 data for definition 2



Map of Temperature Warning Thresholds for each Zipcode, by RR

- Using our model, CalHeat Score system can assign apparent temperatures for the different heat warning levels by selecting RRs for the model, at zip code level.
- Uses 2016 – 2018 data for definition 2.
- Can provide output as a data table.



Covariates and Metadata

Variable	Covariate Description	Column Name	Source (years)	R Prep Code (or source)	Function Description (if available)
DoY	Includes day of year ranging from 1 – 366	DoY (model component used to capture seasonal trends)	HCAI (May – Oct 2008 – 2018)	doy = yday(ymd(date))	Derived from date of service column using HCAI individual level ER dataset for each year using yday() function
DoW	Includes the day of the week ranging from 1-7 (Monday = 1)	DoW	HCAI (May – Oct 2008 – 2018)	dow = wday(ymd(date))	Derived from date of service column using HCAI individual level ER dataset for each week in each year using wday() function
week	Week of the year ranging from 1-52	week	HCAI (May – Oct 2008 – 2018)	week = week(ymd(date))	Represents the week of the year derived from the date of service column using HCAI individual level ER dataset and week() function
month	Includes months ranging from 1–12	month	HCAI (May – Oct 2008 – 2018)	month = month(ymd(date))	Represents the week of the year derived from the date of service column using HCAI individual level ER dataset and month() function
Apparent Temperature - High	temp + humidity index, 1-day max value by zip	heat	GridMet (2008 – 2018)	GridMet (/Users/DGGonzales/DataFiles/GRIDMet/HeatDays_MaxHI_MayOct_2008_2019_Per95_1day.csv)	NA
Apparent Temperature Data - Min	temp + humidity index, 1-day min value by zip	Heat_min	GridMet (2008 – 2018)	GridMet (/Users/DGGonzales/DataFiles/GRIDMet/HeatDays_MaxHI_MayOct_2008_2019_Per95_1day.csv /Users/DGGonzales/DataFiles/GRIDMet/MinHI_oct_2008_2019.csv)	Two datasets merged together to include October for all years
Population (2020)	2016-2020 5-year ACS (B01003_001E)	Population	ACS (2016-2020 -5-year)	ACS (/Users/dianegarcia-gonzales/Dropbox/C_Solutions/Projects/OEHHA/R_Code/20240930_HeatMapping_PopData.Rmd)	See RMD file for full data prep details
ZCTA	ZCTA data crosswalk	zcta	GitHub (2017 zctas)	GitHub (/Users/DGGonzales/DataFiles/Geography/zip_zcta_xref.csv)	Used to merge ACS data to HCAI. Data from https://github.com/censusreporter/acs-aggregate/blob/master/crosswalks/zip_to_zcta/ZIP_ZCTA_README.md

Model Components & R Code

$$\log(\lambda_i) = \beta_{0z_i} + \mathrm{f}(\mathrm{doy_i}) + \mathrm{dow.fct_i} + \beta_{1,z_i} heat_i + f(heat_i)$$

Component	Description	R Code & Parameters	Formula
Intercept	Baseline level	er ~ 1	β_{0Zi}
Year Effect	yearly variation	year.fct (fct = factor)	*Only included if data includes multiple years
Random Effect on Climate Zones	Unstructured random effect with independent identically distributed (iid) model.	cz model='iid', constr=TRUE	*climate zone
Day of Year	Random walk of order 1 (rw1) to capture seasonal trends throughout the year.	doy, model='rw1', hyper=prior.list$dunif, scale.model=TRUE	f(doy_i) (f = function)
Day of Week	Categorical effect for day of the week to account for differences that may occur if event occurred Mon - Sun.	dow.fct	$dow.fct_i$
Interaction on Zip Code & Heat	Effect of heat exposure with spatial smoothing.	zip.num2 x heat.grp.scale, model='besag', graph=H, scale.model=TRUE, adjust.for.con.comp=TRUE	$\beta_{1,Zi}$ x $heat_i$
Random Walk on Heat Scale	Random walk (rw1) on the heat exposure scale to capture temporal trends in heat effects.	f(heat.grp.scale2), model='rw1', hyper=prior.list$dunif	f($heat_i$)

Next Steps

- Which RR do we want to use to identify extreme heat categories?
  - Provide corresponding data table.
- Review models with external colleagues.
- Investigate lags, minimum temperature, SES clusters.
- Investigate impacts using other definitions of heat related outcomes.