

Arizona climate signal heat deaths study

Robbie M Parks

The Earth Institute, Columbia University

robbie.parks@columbia.edu

This work was conducted in collaboration with Columbia Journalism Investigations and the Center for Public Integrity for the story available here [LINK]. Elisabeth Gawthrop of CJI contributed to shaping the research questions.

Data

Death and population data

We used data on all deaths in Arizona from 1980 to 2017 through the National Center for Health Statistics (NCHS) (https://www.cdc.gov/nchs/nvss/dvs_data_release.htm) and on the population from the NCHS bridged-race dataset for 1990–2017 (https://www.cdc.gov/nchs/nvss/bridged_race.htm) and from the US Census Bureau before 1990 (<https://www.census.gov/data/tables/time-series/demo/popest/1980s-county.html>).

We calculated monthly population counts through linear interpolation, assigning each yearly count to July. We filtered for deaths which took place in the summer months (May to September). We calculated death rates by dividing the deaths counts in each month by the population estimate (Figure 1).

Temperature data

We obtained data on temperature from ERA5 (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>), which uses data from global in situ and satellite measurements to generate a worldwide meteorological dataset, with full space and time coverage over our analysis period. We used gridded estimates measured four times daily at a resolution of 30km to generate monthly temperatures by county.

Temperature anomaly

We defined a measure of anomalous temperature for each county and month, which represents the deviation from the average temperature of the county in that month over the entire analysis period. We adapted and described the anomaly metric described in further detail **here**, and briefly below.

To calculate the magnitude of temperature anomaly, we first calculated average temperatures for each summer month in each county over the entire 38 years of analysis (Figure 2). We subtracted these

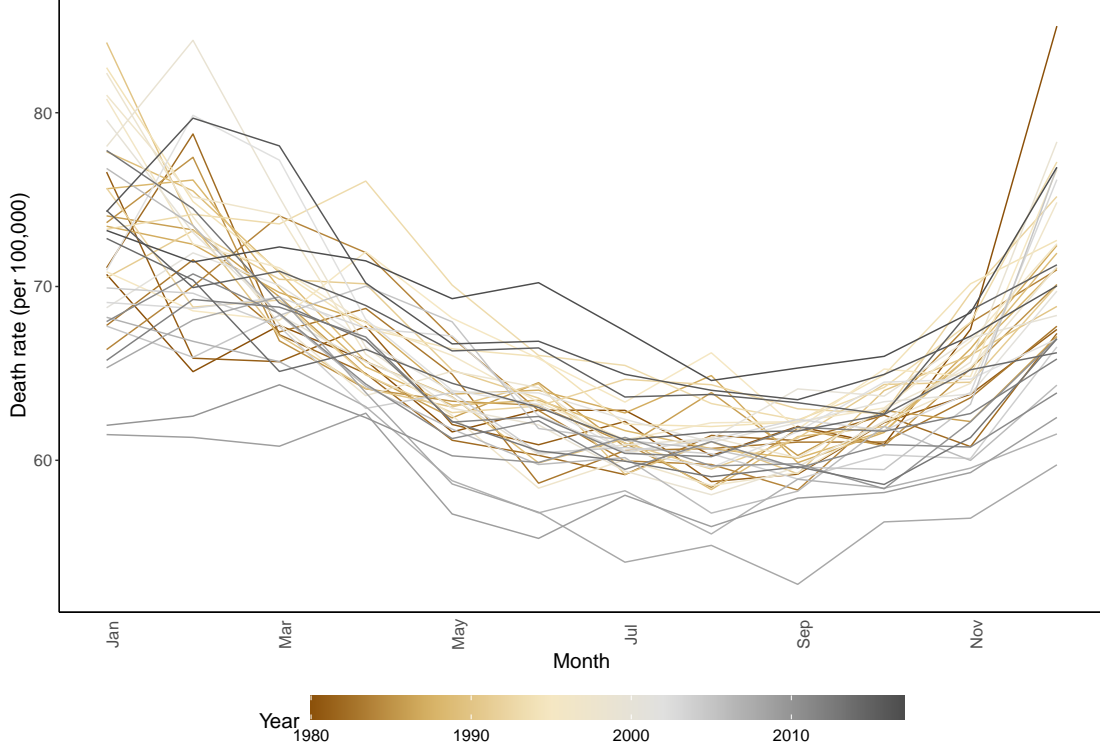


Figure 1: Arizona all-cause death rates from 1980 to 2017, by month.

long-term average temperatures from respective monthly temperature values to generate a temperature anomaly time series for each month and year in each county. County-level temperature anomalies were aggregated to the state level with the use of population weights for analyzing their associations with mortality.

Methods

Statistical model

We analyzed the association of monthly all-cause death rates for Arizona with anomalous temperature using a Bayesian temporal model, as adapted and described in further detail [here](#), and briefly below.

We modeled the number of deaths in each year as following a Poisson distribution:

$$\text{deaths}_{time} \sim \text{Poisson}(\text{rate}_{time} \cdot \text{population}_{time}) \quad (1)$$

The variation in death rates across month of the year and through time are due to several factors, which we capture with the model below, and finds log-transformed death rates modelled as a sum of components that depend on month of year, overall time (*time* in months) and temperature anomaly

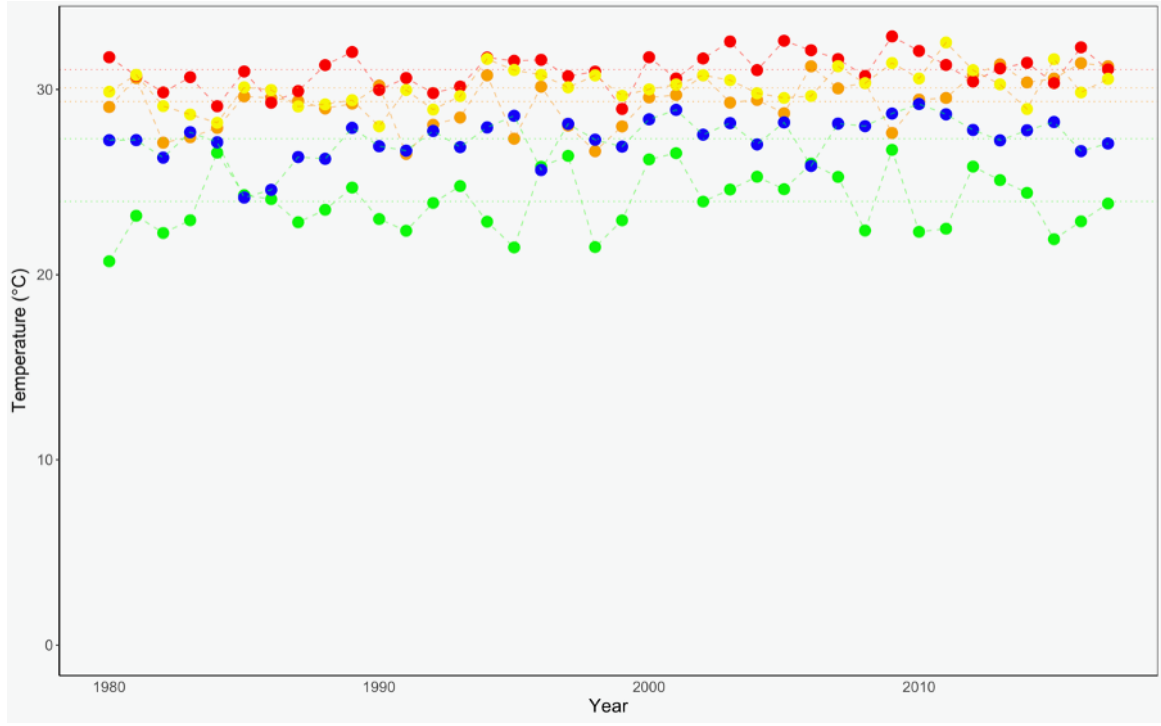


Figure 2: Arizona summer monthly temperatures. Each color represents a particular summer month, with the corresponding dotted lines the long-term average temperatures.

(*Anomaly*):

$$\log(rate_{time}) = \alpha_0 + \beta_0 \cdot (time) + \alpha_{month} + \beta_{month} \cdot (time) + \gamma_{month} \cdot (Anomaly_{time}) + \nu_{time} + \epsilon_{time} \quad (2)$$

The model contained terms that represent the overall level and trend in mortality, with α_0 as the common intercept and β_0 the common slope with overall time. Death rates also vary by month, which may be partly related to temperature and partly due to other monthly factors; monthly variations tend to be smooth across adjacent months. Therefore, we allowed each month of the year to systematically have a different mortality level and trend, with α_{month} the month-specific intercept and β_{month} the month-specific slope with overall time. We used a first-order random walk prior for the monthly random intercepts and slopes, widely used to characterize smoothly varying trends. The random walk had a cyclic structure, so that December was adjacent to January.

We included a term that relates log-transformed death rate to the above-defined state temperature anomaly, $\gamma_{month} \cdot (Anomaly_{time})$. The coefficients γ_{month} represent the logarithm of the monthly death rate ratio per 1°C increase in anomaly. There was a separate coefficient for each month, which means that an anomaly of the same magnitude could have different associations with mortality in different months. The coefficients that we try to infer using statistical analysis are using the additional variation in death rates over and above the other accounted-for factors. As with the month-specific intercepts and trends, we used a cyclic first-order random walk to smooth the coefficient of the temperature anomaly across months.

Non-linear change over overall time (in months) was captured by a first-order random walk, ν_{time} . To ensure identifiability, each set of random walk terms or random effects was constrained to sum to zero. An over-dispersion term (ϵ_{time}) captured the variation unaccounted for by other terms in the model, modeled as $N(0, \sigma_\epsilon^2)$. We used weakly informative priors so that parameter estimation was driven by the data. Hyper-priors were defined on the logarithm of the precisions of the random effects, in other words on $\log(1/\sigma^2)$. These were modeled as $\log\text{Gamma}(\theta, \delta)$ distributions with shape $\theta=1$ and rate $\delta=0.001$. The same hyper-priors were used for all precision parameters of the random effects in the model. For the common slope, we used $N(0, 1,000)$ and for the common intercept a flat prior.

We fitted the models using integrated nested Laplace approximation (INLA), and the R-INLA software, which is computationally more efficient than traditional Markov Chain Monte Carlo for Bayesian inference.

Climate signal deaths

We used the resultant risk estimates from the *Anomaly* term, and the all-cause death rates in Arizona in 2017, to calculate deaths due to monthly temperatures above the long-term norm temperature for each month. For these calculations, where temperatures were above average, we multiplied the actual death counts for each month in each year from 1980 to 2017 by the corresponding excess relative risk, which was calculated as the exponential of the coefficient of the temperature anomaly term from the above analysis. To sum results across the summer months, we obtained 1,000 draws from the posterior marginal of each month's excess relative risk, and took the mean of each 1,000 draw combinations across months.

In other words, this method estimates how temperatures on the monthly (climate) scale contribute to heat-related deaths. Warmer-than-average monthly temperatures come from a combination of natural fluctuations and climate change, and we cannot say which of these 'climate' heat deaths come from climate variability versus change. But by understanding the climate signal in heat deaths now, we can also better understand how mortality could increase as climate change leads to warmer monthly temperatures.

Code for main model

```
#####

## year.month, year.month2, year.month3 - month and year over time, values 1-456

## month1, month2, month3, month4, month5, month6, month7 - indexes for month, values 1-12

## anomaly - temperature anomaly

# hyperparameter value
hyper_value = 0.001

#INLA formula

fml <- deaths ~
# global terms
1 + # global intercept
year.month + # global slope
# month specific terms
f(month, model='rw1',cyclic = TRUE, hyper = list(prec = list(prior = "loggamma",
  param = c(1, hyper_value)))) + # month specific intercept
f(month2, year.month2, model='rw1', cyclic= TRUE, hyper = list(prec = list(prior =
  "loggamma", param = c(1, hyper_value)))) + # month specific slope
# climate specific term
f(month3, anomaly, model="rw1", cyclic=TRUE, hyper = list(prec = list(prior =
  "loggamma", param = c(1, hyper_value)))) + # month specific temperature slope
# random walk across time
f(year.month3, model="rw1", hyper = list(prec = list(prior = "loggamma", param =
  c(1, hyper_value)))) + # rw1
# overdispersion term
f(e, model = "iid", hyper = list(prec = list(prior = "loggamma", param = c(1,
  hyper_value)))) # overdispersion term
```

```

# INLA model rough
system.time(mod.rough <-
inla(formula = fml,
family = "poisson",
data = dat.input,
E = population,
control.compute = list(dic=TRUE),
control.predictor = list(link = 1),
control.inla = list(diagonal=10000, int.strategy='eb',strategy='gaussian'),
verbose=TRUE
))

```

```

# INLA model proper
system.time(mod <-
inla(formula = fml,
family = "poisson",
data = dat.input,
E = population,
control.compute = list(config=TRUE, dic=TRUE),
control.predictor = list(link = 1),
control.inla=list(diagonal=0),
control.mode = list(result = mod.rough, restart = TRUE),
verbose=TRUE
))

```

```
#####
```
