EDS 230 Assignment 2

Elmera Azadpour, Mia Forsline, Alex Vand

2022-04-12

Introduction

The goal of this assignment is to build a simple model to predict annual almond yield based on the paper by Lobell et al. (2006).

Set up: Load necessary libraries and scripts

Function description

Key

- `variable_name` = description [units]

Inputs

- `temp` = average minimum temperature in February [degrees C]
- `precip` = average precipitation in January [mm]

Parameters

- `temp_coeff1`
- `temp_coeff2`
- `precip_coeff1`
- `precip_coeff2`

Note that all parameter values are based on the original paper.

Outputs

- `almond_yield_anomaly` = almond yield anomaly [ton/acre]

Read in climate data

```r
# read in the data from .txt file
clim_raw <- read.csv(file = "clim.txt",
                     header = T,
                     sep = "")

# clean data
clim_data <- clim_raw %>%
             janitor::clean_names() %>%
             mutate(d = lubridate::as_date(d), #convert d column to Date format rather than a character
                    year = lubridate::year(d))
```

Use the `almond_yield_anomaly_annual()` function using coefficients from the paper

- Note that this function averages minimum February temperatures (ºC) and January precipitation (mm)

```r
# set coefficient parameter values
temp_coeff1 <- -0.015
temp_coeff2 <- -0.0046
precip_coeff1 <- -0.07
precip_coeff2 <- 0.0043
constant <- 0.28

# calculate annual almond yield for each year
annual_almond_yield <- cbind(unique(clim_data$year), almond_yield_anomaly_annual(clim_data)) %>%  #crea
  as.data.frame() %>% #transform into a data frame
  rename(year = V1,
         yield = V2)
```

```
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
```

```
## Warning in cbind(unique(clim_data$year),
## almond_yield_anomaly_annual(clim_data)): number of rows of result is not a
## multiple of vector length (arg 2)
```

Check the model outputs

```r
yield_1999 <- subset(annual_almond_yield, year == 1999)[[2]]
yield_1999
```

```
## [1] 9.599988
```

```r
yield_2000 <- subset(annual_almond_yield, year == 2000)[[2]]
yield_2000
```

```
## [1] 159.512
```

```
yield_2001 <- subset(annual_almond_yield, year == 2001)[[2]]
yield_2001
```

```
## [1] 0.2450914
```

Plot the annual trend in almond yield over time

```
ggplot(annual_almond_yield,
       aes(year, yield)) +
  geom_line()+
  geom_point() +
  labs(x = "Year", y = "Almond Yield Anomaly (tons/acre)") +
  theme_classic()
```

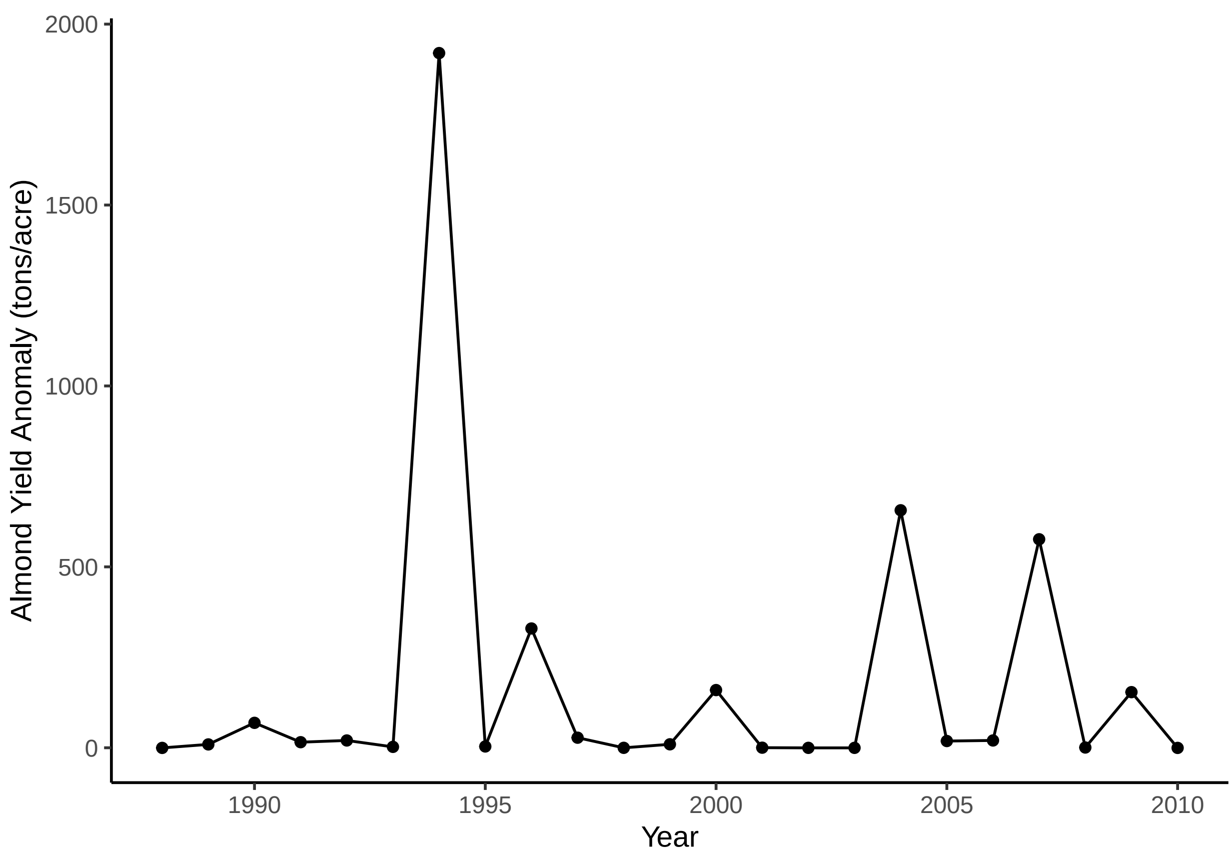

Figure 1: Line graph of almond yield anomolies (tons/acres) in California from 1988 - 2010. During this time, the average almond yield anomaly was approximately 173.5 tons/acre per year. Average almond yield anomaly peaked at approximately 1920 tons/acre in 1994 while some years reported negative average almond yield anomalies. This model is based on Lobell et al. (2006).

Results and conclusions

```
max_yield <- max(annual_almond_yield$yield)
min_yield <- min(annual_almond_yield$yield)
avg_yield <- mean(annual_almond_yield$yield)
```

We built a simpler replica of the almond yield model built by Lobell et al. (2006). From 1988 to 2010, the average almond yield anomaly was 173.5408846 tons/acre per year. During this time period, average almond yield anomaly peaked at 1919.9811511 in 1994 while some years reported negative average almond yield anomalies. Over time, there is no apparent visual positive or negative trend in average almond yield anomaly. There are multiple and sporadic peaks starting after 1994.

Moving forward, it would be interesting to further investigate outlier years like 1994 to discern if that high of an average almond yield is feasible or if the model can be improved.