Exploratory analysis of Los Angeles meteorological data

Henry Toulmin

May 2020

Table of Contents

[Introduction 2](#_Toc39839622)

[Methods 3](#_Toc39839623)

[Residual Analysis 3](#_Toc39839624)

[Data Normalization 3](#_Toc39839625)

[Results 3](#_Toc39839626)

[Average Monthly Temperature 4](#_Toc39839627)

[Average Monthly Humidity 5](#_Toc39839628)

[Total Yearly Rainfall 7](#_Toc39839629)

[Discussion 9](#_Toc39839630)

[Conclusion 10](#_Toc39839631)

[References 10](#_Toc39839632)

knitr::opts\_chunk$set(message=FALSE, warning=FALSE, echo = TRUE)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

library(grid)  
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.6.3

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

LA\_airport <- readRDS("../Data/LA\_airport.rds")  
dat <- na.omit(LA\_airport) %>%   
 filter(aa1\_1 != 99, aa1\_2 !=999.9, aa1\_4 != 3) %>%   
 mutate(year = year(time),  
 month = month(time),  
 monthyear = decimal\_date(ymd(paste(year,month,1, sep = "-"))),  
 aa1\_2 = aa1\_2/aa1\_1) %>%   
 filter(year != 1972, year != 2020)

## Introduction

The scientific consensus surrounding the question of climate change is that it is not only occurring, but it is a result of human activity. The 2018 IPCC Special Report on the impacts of global warming claims that anthropogenic climate change has resulted in an increase in global temperatures by approximately 1.0°C above pre-industrial levels and expects that this figure will jump to 1.5°C between 2030 and 2052. They point out the consequences of anthropogenic climate change have included but are not limited to rising global temperatures, rising sea levels, and an increased frequency in natural disasters such as droughts, floods, and storms. The report claims that the risk of these consequences is significantly higher if global warming reaches 2°C than if it is limited to 1.5°C, and the IPCC stresses that the world needs to act now if we are going to succeed in keeping temperature rise to below 1.5°C (IPCC).

There are some places on Earth that are most susceptible to climate change consequences than others. One of these places is Los Angeles, California. In recent years, Los Angeles county, as well as a large share of California, has been experiencing an ongoing drought. The longest duration of drought began in December 2011 and persisted until March 29, however 29% of California is still in drought. (drought.gov) Additionally, being located on the pacific coast sets Los Angeles up to be one of the places most affected by sea level rise. The Los Angeles area is expected to match global sea level rise estimates of 0.1 - 0.6 m from 2000 to 2050 and 0.4 - 1.7 m from 2000 to 2100, with financial losses from building stock ranging from 410 million to 714 million US dollars (Griffman et al).

Because of the high-risk climate change poses to Los Angeles and its nearly 19 million residents, it is vitally important for Los Angeles to look into mitigation techniques. However, before these techniques can be accurately developed, it is essential to understand the trends of local climate variables over the past several decades to help predict how they will behave in the future. Using meteorological data from Los Angeles Airport, this study explores the trends of temperature, relative humidity, and depth of liquid precipitation from 1972 to 2019 to extrapolate how these variables will behave in the future.

## Methods

This study utilizes bivariate analysis to estimate trends of Los Angeles’ temperature, relative humidity, and total liquid precipitation with respect to time. The dataset used in this study to analyze the aforementioned trends was pulled from <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/> and represents hourly meteorological data for Los Angeles International Airport from 1973 to 2019. All data analysis was performed in R studio Version 1.1.456 and formatted in rstudio.cloud. After removing erroneous and incomplete data from the original dataset, I mutated the dataset, which was originally expressed as hourly data, to create year, month, and combined year and month date objects using **lubridate** and averaged the data of each meteorological variable by each of these date objects. Each summarized variable was then plotted against the

### Residual Analysis

Each plot showing a meteorological variable over time is accompanied with a residual dependency plot, a residual quantile-quantile plot, and a spread-location plot of the relevant data. The residual dependency plot is used to evaluate how well the fit assigned to the meteorological variable’s relationship to time. A loess fit was used for the residual dependency plot, as it will account for shifts in trends throughout the data rather than the overall trend of the data. If this fit is flat, the fit assigned to the fitted values is appropriate. The quantile-quantile plot is used to show how normalized the spread of residuals is. The residuals will be normally distributed if the sample distribution equals the theoretical distribution, meaning that the points line up perfectly along x=y. The spread-location plot shows whether the function behaves differently across different locations of the plot.

### Data Normalization

As this study is simply exploratory in nature and does not employ any hypothesis tests, I opted not to reexpress the data to normalize residuals. I wanted to represent the data as it is was originally measured, as I wanted the trends of the data to be as visually clear as possible. The appropriate transformations needed to normalize data for hypothesis testing is outlined for each figure in the results section.

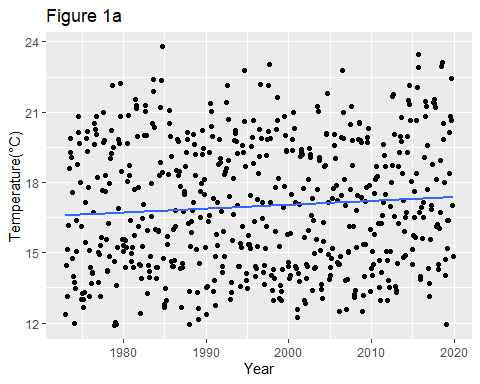
## Results

The data analysis yielded that over the 48-year span of data collection, the average temperature increased by 0.0166°C per year, the relative humidity decreased by 0.0653% every year, and total precipitation decreased exponentially at a rate of 0.08214x^2 - 0.03236x.

### Average Monthly Temperature

The trend of average temperature per year, as shown in figure 1a, is described by the equation y = 0.0166x - 16.17, meaning over the 47 year span of the data, the average temperature in Los Angeles has risen 0.7802°C. Figure 1b is a residual dependency plot for the temperature data that shows a relatively flat loess fit, showing the residuals of figure 1a to be distributed relatively evenly across a linear fit, meaning a parametric fit of order 1 is appropriate for this data. Figure 1c is a quantile-quantile plot of the temperature data which shows the distribution of the residuals of figure 1a to be abnormal, as the first and fourth quartile of the data deviate upward and downward respectively, compared to the normal distribution shown by the trendline. This data would need to be transformed through raising the y (avg.temp) variable to the 3rd power if a hypothesis test were to be utilized. Figure 1d is a spread location plot, which shows a flat loess fit, meaning there is very little change in function as a product of location.

dat1 <- dat %>%  
 group\_by(monthyear) %>%   
 summarise(avg.temp = mean(temp),  
 avg.rh = mean(rh),  
 total.rf = sum(aa1\_2))  
  
# Plot Temp vs. Time  
ggplot(dat1, aes(x=monthyear, y=avg.temp)) + geom\_point() +  
 geom\_smooth(method = MASS::rlm, formula = y~x, se=FALSE) +  
 ggtitle("Figure 1a") +  
 xlab("Year") + ylab("Temperature(°C)")



# Check the residual dependance plot  
M1 <- MASS::rlm(avg.temp~monthyear, dat1)  
dat1$res1 <- M1$residuals  
  
p1a <- ggplot(dat1) + aes(x=monthyear, y=res1) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1), size=0.5) +  
 ggtitle("Figure 1b") +  
 xlab("Year") + ylab("Residuals")  
  
# Check the residual for abnormality  
p1b <- ggplot(dat1, aes(sample=res1)) +  
 geom\_qq(distribution = qnorm, size=0.5) +  
 geom\_qq\_line(distribution = qnorm, size=0.5) +  
 ggtitle("Figure 1c") +  
 xlab("Theoretical Distribution") + ylab("Sample Distribution")  
  
# Check the spread-location plot  
s1 <- data.frame(std.res = sqrt(abs(residuals(M1))),  
 fit = fitted.values(M1))  
p1c <- ggplot(s1) + aes(x=fit, y=std.res) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1), size=0.5) +  
 ggtitle("Figure 1d") +  
 xlab("Temperature(°C)") + ylab("Standard Residuals")  
  
grid.arrange(p1a, p1b, p1c, nrow=1)

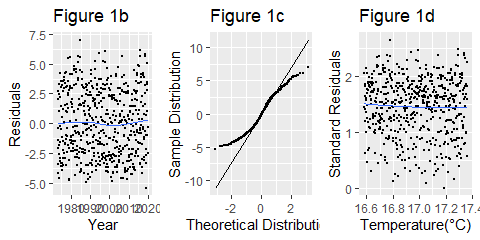
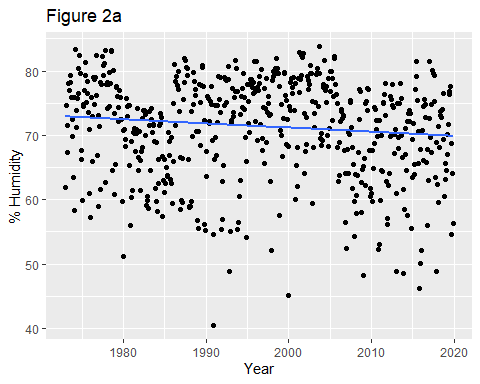


Figure 1. Average temperature (°C) as a function of time (months).

### Average Monthly Humidity

The trend of average relative humidity per year is described by the equation y = -0.0653x + 201.7, meaning over the 47-year span of the data, the average relative humidity in Los Angeles has dropped 3.069%. Figure 2b is a residual dependency plot for the humidity data that shows a relatively flat loess fit, showing the residuals of figure 2a to be distributed relatively evenly across a linear fit, meaning a parametric fit of order 1 is appropriate for this data. Figure 2c is a quantile-quantile plot of the humidity data which shows the distribution of the residuals of figure 2a to be abnormal, as the fourth quartile of the data deviates downward with respect to the normal distribution shown by the trendline. This data would need to be transformed through raising the y (avg.rf) variable to the 2nd power if a hypothesis test were to be utilized. Figure 2d is a spread location plot, which shows a relatively flat loess fit, meaning there is very little change in function as a product of location. There appears to be an abnormality in the data from 1980 to 1986, where there are no months with an average humidity of 75% or greater. The data does show many hourly measurements of above 75% within this span, so this abnormality is likely not due to holes in the data. Perhaps there was a natural phenomenon during this period which made the climate of Los Angeles relatively drier than usual.

ggplot(dat1, aes(x=monthyear, y=avg.rh)) + geom\_point() +  
 geom\_smooth(method = MASS::rlm, formula = y~x, se=F) +  
 ggtitle("Figure 2a") +  
 xlab("Year") + ylab("% Humidity")



# Check the residual dependance plot  
M2 <- MASS::rlm(avg.rh~monthyear, dat1)  
dat1$res2 <- M2$residuals  
  
p2a <- ggplot(dat1) + aes(x=monthyear, y=res2) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1), size=0.5) +  
 ggtitle("Figure 2b") +  
 xlab("Year") + ylab("Residuals")  
  
# Check the residual for abnormality  
p2b <- ggplot(dat1, aes(sample=res2)) + geom\_qq(distribution = qnorm, size=0.5) +  
 geom\_qq\_line(distribution = qnorm, size=0.5) +  
 ggtitle("Figure 2c") +  
 xlab("Theoretical Distribution") + ylab("Sample Distribution")  
  
# Check the spread-location plot  
s2 <- data.frame(std.res = sqrt(abs(residuals(M2))),  
 fit = fitted.values(M2))  
p2c <- ggplot(s2) + aes(x=fit, y=std.res) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1), size=0.5) +  
 ggtitle("Figure 2d") +  
 xlab("% Humidity") + ylab("Standard Residuals")  
  
grid.arrange(p2a, p2b, p2c, nrow=1)

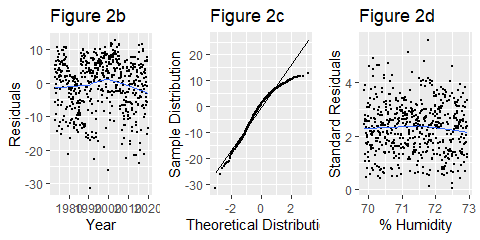
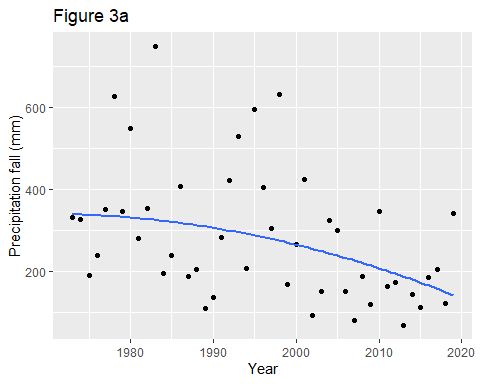


Figure 2. Average relative humidity (%) as a function of time (months).

### Total Yearly Rainfall

The trend of total rainfall per year is described by the equation y = -0.08214x^2 + 0.03236x - 3.184, meaning over the 47 year span of the data, the total yearly rainfall in Los Angeles has dropped 179.9mm. Figure 3b is a residual dependency plot for the precipitation data that shows a relatively flat loess fit, showing the residuals of figure 3a to be distributed relatively evenly across a parabolic fit, meaning a parametric fit of order 2 is appropriate for this data. Figure 3c is a quantile-quantile plot of the precipitation data which shows the distribution of the residuals of figure 3a to be abnormal, as the first and first and fourth quartiles of the data both deviate upward with respect to the normal distribution shown by the trendline. This data would need to be transformed through taking using the log of the y (total.rf) variable if a hypothesis test were to be utilized. Figure 3d is a spread location plot, which shows an upward loess fit, meaning as precipitation increases, the location of the residuals increases until around 275 mm, then decreases. However, this fit should become flatter when the rediduals are normalized as stated above. Because the data yielded many months with little to no rainfall, I decided to represent this data per year rather than month, as it gave a clearer picture of the trend of rainfall.

dat2 <- dat %>%  
 group\_by(year) %>%   
 summarise(avg.temp = mean(temp),  
 avg.rh = mean(rh),  
 total.rf = sum(aa1\_2))  
  
ggplot(dat2, aes(x=year, y=total.rf)) + geom\_point() +  
 geom\_smooth(method = MASS::rlm, formula = y~x + I(x^2), se=F) +  
 ggtitle("Figure 3a") +  
 xlab("Year") + ylab("Precipitation fall (mm)")



# Check the residual dependance plot  
M3 <- MASS::rlm(total.rf~year + I(year^2), dat2)  
dat2$res3 <- M3$residuals  
  
p3a <- ggplot(dat2) + aes(x=year, y=res3) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1), size=0.5) +  
 ggtitle("Figure 3b") +  
 xlab("Year") + ylab("Residuals")  
  
# Check the residual for abnormality  
p3b <- ggplot(dat2, aes(sample=res3)) + geom\_qq(distribution = qnorm, size=0.5) +  
 geom\_qq\_line(distribution = qnorm, size=0.5) +  
 ggtitle("Figure 3c") +  
 xlab("Theoretical Distribution") + ylab("Sample Distribution")  
  
# Check the spread-location plot  
s3 <- data.frame(std.res = sqrt(abs(residuals(M3))),  
 fit = fitted.values(M3))  
p3c <- ggplot(s3) + aes(x=fit, y=std.res) + geom\_point(size=0.5) +  
 geom\_smooth(method = "loess", se=F, method.args = list(degree = 1, family="symmetric"), size=0.5) +  
 ggtitle("Figure 3d") +  
 xlab("Precipitation (mm)") + ylab("Standard Residuals")  
   
grid.arrange(p3a, p3b, p3c, nrow=1)

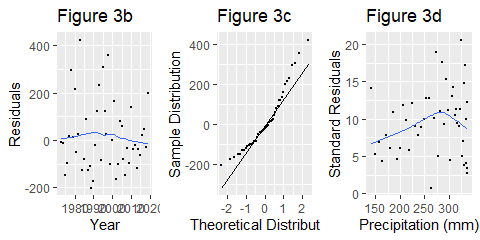


Figure 3. Average liquid precipitation depth (mm) as a function of time (years).

## Discussion

The analysis of this data supports the claim that average yearly temperatures are rising, which is consistent with the climate consensus that global temperatures are rising. According to [climate.gov](https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature), the average global increase in temperature has been 0.18°C per decade, which is not far off from the rate of 0.166°C per decade calculated in this study (percent difference = 8.1%). This would suggest that Los Angeles is warming relatively slightly slower than the rest of the world. An explanation for this could be that Los Angeles was already a relatively hot location, so it won’t have as larger of a temperature change compared to locations at higher latitudes. Because the temperature rise rate from this study is so similar to the global temperature rise, I expect Los Angeles to hit the 1.5°C within the span given in the IPCC report. Ideally, this study would have estimated reaching this threshold based off of preindustrial temperature averages in Los Angeles, but I could not find reliable temperature data for that long ago.

Additionally, the data yielded an exponential decrease in total yearly precipitation depth. I was unexpecting to find this to be the case, especially considering the decrease in recent drought as shown by [drought.gov](https://www.drought.gov/drought/states/california). This significant decrease in rainfall signifies an increased risk for further drought in the future. The sustained decrease in humidity also shows Los Angeles to becoming a drier climate, another contributor to increased droughts.

A decrease in temperature also shows potential contribution to California’s drought. There have been studies that link rising temperatures to increased melting of snowcapped mountains in the Sierra Nevada mountain range, which provides 60% of California’s drinking water. Because snow is melting faster, it delivers more meltwater than the aquifers can handle at once, so much of this water is lost. For the rest of the summer, there is no consistent meltwater flow, which results in drought (Barnett et al). In this way, the increase in temperature seen in Los Angeles could be contributing to California’s drought. However, to confirm this, temperature data from the mountain range would need to be collected and compared to the data from this study to ensure the range is facing the same temperature increase as Los Angeles. Additionally, the increase in temperature shown in this study does suggest a response in sea level rise, but an appropriate next step in this research would be to analyze sea level rise data and compare it with temperature rise trends to see if there is a cause and effect relationship that can be gleaned.

## Conclusion

Ultimately, this study shows that given the risk of climate change consequences, it is imperative that Los Angeles invest in climate change mitigation strategies. The increases in temperature and the decreases in humidity and precipitation show that Los Angeles is trending towards a warmer, drier climate. This data does not show a direct linkage to anthropogenic climate change, but further research utilizing trends in variables such as carbon emissions and deforestation may yield cause and effect relationships with the trends analyzed in this study. Luckily for Los Angeles, California is one of the most forward thinking and innovative states when it comes to climate action, so as long as climate variables such as those in this study are measured, monitored, and projected, we can keep a close watch on the way in which our climate is changing.

## References

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). *Potential impacts of a warming climate on water availability in snow-dominated regions*. Nature, 438(7066), 303-309. <https://www.nature.com/articles/nature04141>. 04 May 2020

Dahlman, L., Lindsey, R. (2018). *Climate Change: Global Temperature*. NOAA. <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>. 04 May 2020

Drought.gov. *Drought in California.* National Integrated Drought Information System. <https://www.drought.gov/drought/states/california>. 04 May 2020

Gimond, Manuel. *Exploratory Data Analysis in R.* <http://mgimond.github.io/ES218/index.html>. 04 May 2020.

Grifman, P., Hart, J., Ladwig, J., Newton, A., & Schulhof, M. (2013). *“Sea level rise vulnerability study for the City of Los Angeles”*. University of Southern California. <https://dornsife.usc.edu/assets/sites/291/docs/pdfs/SeaLevelRiseDocs/City_of_LA_SLR_Vulnerability_Study_FINAL_Online_w_appen_sm.pdf>. 04 May 2020

IPCC, 2018: Global Warming of 1.5°C.An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)].. <https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Full_Report_Low_Res.pdf>. 04 May 2020

NOAA. <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/>. 04 May 2020