

# Rain

Matthew Wiens, Kelly Kung

December 20, 2017

## Abstract

Abstract Here

## 1 Introduction

Residents and visitors to the Pacific Northwest (Oregon and Washington) notice a distinct pattern year over year: most days during the winter are rainy, and then they feel that there a distinct point during the spring or early summer when the weather becomes dry most days. Similarly, during early fall the nice weather switches back to rain. However, the timing of the switch varies throughout the area, with the prominent Cascade Mountains being a key factor. There are many resorts on the east side of the Cascade Range that cater to residents of the west side looking to escape the rain.

Widmann and Bretherton developed a methodology to control for local variation in topography and precipitation data and generated a dataset of estimated precipitation data over 46 years in a 50km by 50km grid. In comparison, general models of atmospheric weather operate on the order of hundreds of kilometers, so there significant room to improve the spatial resolution of weather models. In particular, local topographic features are not captured by general model, which is a limitation in regions like the Pacific Northwest. The temporal and spatial correlation is also highly dependent on the topography and difficult to model.

## 2 Background

### 2.1 Dataset

To analyze the precipitation in the Pacific Northwest, we use the same dataset as Widmann and Bretherton. This data contains precipitation logs (measured in millimeters) from 1949 - 1994 that was collected from the National Weather Service, which was then corrected by the National Climatic Data Center<sup>12</sup>. The dataset is a 3-dimensional array with a 2-dimensional grid-cell that consists of 16 longitude subsections and 17 latitude subsections, where each cross-section is approximately  $0.48^\circ$  (latitude) by  $0.62^\circ$  (longitude). The third dimension contains temporal data (16,801 days), which means we have approximately 4 million data points. However, note that there are approximately 32,000 missing values which correspond to areas with few functional weather stations, such as the ocean and some areas in the southeast. Using this dataset, we then transformed and created several variables in order to obtain a model to analyze the precipitation.

---

<sup>1</sup>The data was corrected during the 'Validated Historical Daily Data' project, which aimed to verify the data collected.

<sup>2</sup>T. Reek, S. Doty, and T. Owen. A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network. In Bulletin of the American Meteorological Society 73, no. 6 (1992): 753-62.

## 2.2 Grouping of Areas

The Pacific Northwest has a varied climate, from temperate rainforests on the Pacific coast to desert in Southeastern Oregon, which has a major impact on the seasons. Therefore, each grid cell is classified into one of nine climatic zones as follows:

1. Washington Coastline (Temperate Rainforest)
2. Oregon Coastline
3. Western Washington / Seattle Metropolitan Area
4. Portland Metropolitan Area / Willamette Valley / Southwestern Oregon
5. Cascade Range
6. Columbia Plateau
7. Selkirk Mountains
8. Blue Mountains
9. Eastern Oregon / High Desert / Great Basin

These zones were decided by considering average precipitation over the time period of the data and the topographic features of the region. The rainfall for the region was aggregated by considering if there was measurable rainfall in each grid by day, and then for each day if the majority of grid cells reported rain, then a value of *rain* was assigned to the area, else it was *dry*

## 2.3 Indicators for Dry and Wet Season

Using the grouping of areas as mentioned in Section 2.2, we then proceeded to create indicators of when the dry season and wet season starts. To do this, we first created indicators of precipitation for each day in each area, where a 1 indicates that more than 0.1 mm of precipitation occurred. We then aggregated the indicators in each of the 9 areas for each day and determined that there was precipitation in the area for that day if a majority of the cells in the area had more than 0.1 mm of precipitation. The areas with precipitation were deemed to be *wet* and the areas without precipitation were deemed to be *dry*. Afterwards, we created an algorithm that chooses indicators of when the dry season starts ( $X_{min}$ ) and when the wet season starts ( $X_{max}$ ) such that it maximizes the number of *dry* days in the dry season and maximizes the number of *wet* days in the wet season. For example, we expected the  $X_{min}$  days to be approximately 150 and the  $X_{max}$  days to be approximately 300 because these dates correspond to the beginning of May and the beginning of November. Using this algorithm, we found the start dates for the *dry* and *wet* seasons for the 46 years.<sup>3</sup>

## 3 Model

We further developed our model by considering the probability of rain on each day for the summer (dry) and winter (rainy) seasons. This is motivated by the idea of trying to plan outdoor activity in each area of the Pacific Northwest and deciding if there is value in traveling to a different area to escape rain. Therefore, the following model is proposed:

---

<sup>3</sup>Note that when creating the indicators for years, we did not take into account of leap years, which resulted in data points that spilled over to the next year. We just take the years in which we have a year's worth of data for.

- The data is binary data representing if it rained or not for each day of the year,
- $\pi_w$  and  $\pi_d$  are the probability of rain during the wet season and dry season, respectively.
- $\theta_1$  and  $\theta_2$  are the cutoff points for the wet and dry seasons. The wet season runs from January 1st until the day before  $\theta_1$  and from  $\theta_2$  to December 31st, and the dry season runs from  $\theta_1$  until the day before  $\theta_2$ , and the natural restriction is imposed that  $\theta_1 < \theta_2$
- Let  $y_j$  be the data on day  $j$ . Then  $p(y_j) = \pi_w 1_{j < \theta_1} + \pi_d 1_{\theta_1 \leq j < \theta_2} + \pi_w 1_{j \geq \theta_2}$

## 4 Analysis

## 5 Discussion

## 6 Conclusion