UNIVERSITY OF PUERTO RICO
RIO PIEDRAS CAMPUS
FACULTY OF NATURAL SCIENCES
DEPARTMENT OF MATHEMATICS

MODELING INFANTILE ASTHMA IN PUERTO RICO

By
Jaime A. Roura Monllor

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN MATHEMATICS
AT THE UNIVERSITY OF PUERTO RICO

JUL 2011

UNIVERSITY OF PUERTO RICO
DEPARTMENT OF
DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the Faculty of Faculty of Natural Sciences for acceptance a thesis entitled "**Modeling Infantile Asthma in Puerto Rico**" by **Jaime A. Roura Monllor** in partial fulfillment of the requirements for the degree of **Master of Science in Mathematics**.

Dated: Jul 2011

Supervisor: _____
L. R. Pericchi Guerra, Ph.D.

Readers: _____
M.E. Pérez, Ph.D.

_____
D. Reetz, Ph.D.

UNIVERSITY OF PUERTO RICO

Date: **Jul 2011**

Author: **Jaime A. Roura Monllor**

Title: **Modeling Infantile Asthma in Puerto Rico**

Department: **Department of Mathematics**

Degree: **M.Sc.** Convocation: **Jul** Year: **2011**

Permission is herewith granted to University of Puerto Rico to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

I dedicate this work to all the families that struggle with asthma. I hope that you may find a way of life that diminishes the impact of this disease on you.

Above all I dedicate this research to God.

Table of Contents

List of Tables

List of Figures

Abstract

In response to the high incidence of infantile asthma in Puerto Rico (PR), this study aimed to predict infantile asthma based on fungal spore, pollen, pollutant concentrations, and/or meteorological factors. A predictive model would allow for the creation of an alert to inform the general public about the risk of infantile asthma on a daily basis. Simulating dynamic linear models with discount factors using *OpenBUGS*, *R*, and *R2OpenBUGS* we constructed models which explained pediatric asthma cases at San Jorge Children's Hospital and the University of PR's Carolina Hospital by atmospheric ozone concentration or by a combination of total airborne fungal spore and ozone concentrations with and without interaction. High autocorrelation of residuals led us deseasonalize using a Fourier model for San Jorge Children's Hospital. By minimizing the deviance information criterion (DIC), and analyzing model coefficients and residuals we chose model $y_t = \beta_{0,t} + \beta_{1,t} totalSpores_t + \beta_{,t} ozone_t + \epsilon_t$ where ϵ_t represents errors. Residuals seemed to follow right-skewed distributions and we did not manage to approximate normality for any model. This result undoubtedly serves as a dynamically predictive model for the monthly data and might serve as starting points for future research with more complete, hopefully daily, data. Further analysis should attempt to better clarify outliers and to fit a regressive model explaining seasonality at San Jorge. Furthermore we would like to compare our model with a non-regressive one that combines linear growth and seasonality. Future research should also work on establishing a network of daily and island-wide asthma case registry from hospitals and physicians. This type of information would greatly assist in creating a model with much better predictive capacity than the present one.

Acknowledgments

We would like to thank Dr Yiamira Oquendo (University of Puerto Rico Pediatric Hospital), Dr Jeannette Lube (University of Puerto Rico Hospital at Carolina), Elizabeth Quintero (UPR-Medical Sciences Campus, Microbiology Department) and Dr Benjamín Bolaños (National Allergy Bureau (NAB) American Academy of Asthma, Allergy, and Immunology (AAAAI)), the Puertorican Environmental Quality Board (JCA by its Spanish name) and NOAA for providing the original data and initiative for this project. We also thank David Torres at the Center for Biostatistics and Bioinformatics (CBB) at the University of Puerto Rico at Río Piedras. This research was sponsored in part by the National Science Foundation "STEM: Alan Turing Fellowship" (DUE:0630927), Merck, and the CBB.

Río Piedras, Jaime Roura
April 22, 2011

Introduction

Infantile asthma in Puerto Rico poses a serious public health issue that affects 28.9% of children during their lifetimes which amounts to approximately $300,080$ individuals (95% CI $266,073 - 334,086$) compared to 13.3% for the United States [16]. Furthermore 33.3% of schoolchildren suffer from respiratory conditions out of which 51.3% receive treatment (Perez-Perdomo et al. in [19]. In general asthma accounts for about 37 deaths per million inhabitants in Puerto Rico which ranks 5^{th} internationally [15]. This study aims to predict infantile asthma based on fungal spore, pollen, pollutant concentrations, and/or meteorological factors and to create an alert to inform the general public about the risk of infantile asthma on a daily basis.

Studies in other countries have attempted to model the relationship between asthma and environmental factors including meteorological, air quality, and aeroallergenic information. In general they show that a positive relationship exists between atmospheric fungal spore concentration and hospitalizations and/or emergency room visits for pediatric asthma. Some results disagree with respect to the type of fungal spore that shows association.

Recently Agarwal et al. [1] published a review of the current evidence linking severe asthma and fungi. They affirmed that uncontrolled severe asthma may hinder normal activity and, in severe cases, lead to death. References within noted that

"2/3 of asthmatics are atopic to common allergens" with environmental fungi implicated as one of the most common along with dust mites and pet dander. Despite general association, the review confirms that past results disagree on the type of fungal sensitization and severe asthma: Aspergillus or other (Alternaria, Cladosporium, Epicoccum, Helminthosporium, Penicillum, Candida, Aurebasidium, Trichophyton). In another review Denning et al. [11] acknowledged that severe asthma ranked 8^{th} among reasons to visit a physician and that "the role of fungi as a primary exogenous driver of asthma has been incompletely explored, possibly because exposure is universal but highly variable in time and intensity and hard to measure".

Dales et al. [7] analyzed visits to Ontario's Regional Children's Hospital with a primary diagnosis of asthma for the years 1993-1997 (months March to October). They analyzed a time series model which smoothed seasonal trends using locally weighted nonparametric regression and smoothing scatter plots (LOESS), filtered for day of the week effects noticing an increase in ER visits over weekends, and minimized Akaike's Information Criterion (AIC). Their results indicated that each relative humidity and ozone on its own on the date of visit (lag = 0) predicted day-to-day changes in emergency visits using a forward inclusion (AIC) stepwise regression analysis. They also detected deuteromycetes and basidiomycetes (including ganoderma) with lags 2 and 3 days, respectively. Considering an increase equal in magnitude to its mean in fungal spore families independent from each other they measured a significant ($p < 0.05$) increase in emergency department visits of 1.9% (SE 0.9%) for deuteromycetes, 4.1% (1.6%) for basidiomycetes, 2.8% (1.0%) for ascomycetes, and 8.8% for these combined. When they considered the three groups of spores the effect of total spores disappeared from the model which suggested apparent independent effects.

In a follow up study regarding thunderstorm related asthma during six years, Dales et al. [8] also found a positive association between fungal spores and emergency department visits at the same hospital. This time they measured an increase in emergency department visit of 2.2% (0.9% SE) related to an increase in total spore count equal in magnitude to its mean.

In yet another study, Dales et al. [9] expanded their time series analysis to include 7 years (1993-2000) of daily rates of hospitalizations for asthma in the 10 largest cities across Canada which comprised about half of the population of the entire country. They adjusted time series per city for long term trends, day of week, climate, and air pollution considering lags 0 to 5 days. They considered a generalized additive model with a rigorous convergence criteria ($\epsilon < 10^{-14}$); fully parametric natural cubic splines with 3 knots per year to account for long term confounding factors (viral or other). They determined the degrees of freedom and number of knots per year using AIC. They then selected the model with either minimum AIC or maximum evidence of no structure in residuals. They adjusted each time series for temporal trends with natural cubic spline functions with 5 knots per year including serial correlation by Bartlett's test and for day of week and daily changes in variables using backwards stepwise regression ($p < 0.05$ with AIC). There existed a significant variability of allergens and interaction terms between cities. Only the interaction between ozone and tree pollen was significant ($p < 0.05$) for all cities and it accounted for a 0.22% further increase in admissions in the model ($p < 0.05$). Associated to a daily increase equivalent to the mean value of each allergen this model showed "the following percentage increase in asthma hospitalizations: 3.3% (95% CI, 2.3 to 4.1) for basidiomycetes, 3.1% (95% CI, 2.8 to 5.7) for ascomycetes, 3.2% (95% CI, 1.6 to 4.8) for deuteromycetes, 3.0%

(95% CI, 1.1 to 4.9) for weeds, 2.9% (95% CI, 0.9 to 5.0) for trees, and 2.0% (95% CI, 1.1 to 2.8) for grasses. After accounting for the independent effects of trees and ozone, the combination of the two was associated with an additional 0.22% increase in admissions averaged across cities ($p < .05$)".

Atkinson et al. [2] studied asthma in London. They gathered counts of asthma cases as general practitioner consults, emergency rooms visits, and hospital admissions and analyzed aeroallergens from 1992 to 1993 including fungal spores (30 taxa divided into 3 mayor families: deuteromycetes, basidiomycetes, and ascomycetes) and pollens (trees, grass, and nettles). They also collected meteorological (daily average temperature and relative humidity) and air quality information. Considering a Poisson regression model that adjusted for seasonal trends and calendar effects (day of week, etc.) they accounted for unmeasured confounders by including a non-linear function of time, and for known and measured variables with linear and non-linear models. They performed time series analysis using penalized splines with weekly knots.

They treated spores counts as a whole and by family or taxa. Total spore count had a 4 level factor according to quartile permitting the assessment of non-linear association. Spore counts by family or taxa had one of 3 different analyses: 4 level (quartile), 3 level, or 2 level (present, absent) depending on measured quantities. The model considered lags 0, 1, 2, and 3 days. Furthermore they compared models constructed using the generalized cross validation method and the partial auto correlation function (PACF). Patterns of associations under both models were similar which "suggested that associations were unlikely to be confounded by inappropriate seasonal control". Results show seasonal behavior of asthma with an increase during

the start of autumn (October). Children up to the age of 14 showed weak evidence of a positive association between total spores and either hospital admissions or ER visits. "The relative risks for increases in the number of hospital accident and emergency visits and hospital admissions associated with changes in fungal spore concentrations from the lower to upper quartiles were 1.06 (95% CI 0.94 to 1.18) and 1.07 (0.97 to 1.19) respectively. The addition of pollen or air pollutants had little impact on the observed associations. A number of individual spore taxa, in particular Alternaria, Epicoccum, Agrocybe, Mildews, and both coloured and colourless Basidiospores and Ascospores, were associated with increases in the number of emergency visits and hospital admissions for asthma, although the precision of these estimates were low".

They showed an association between deuteromycetes and basidiomycetes with hospital admissions and between basidiomycetes and ascomycetes with emergency room visits. The pediatric model used lag 0. They argued that "one possible explanation for this pattern of responses was that the more severe asthma exacerbations, resulting in an emergency hospital admission, were provoked by the higher spore concentrations".

This study had the advantage of counting with a large population. Its main limitation came from the fact that ER visits made self diagnoses which may introduce bias. It also used only 2 years of data with no prior knowledge. A lesser limitation came from having a single spore trap which they estimated to measure a radius of 50km. Atkinson et al. also restrained their analysis for months ranging from June until mid October due to the absence or low levels of fungal spores during the colder months. In PR, however, we have fungus year round.

Rosas et al [22] analyzed the relationships of aeroallergens, air pollution, and weather with daily asthma emergency admissions to a hospital in Mexico City. They

performed a time-series analysis by building a generalized linear model (GLM) under a Poisson distribution and a logarithmic link function. They considered lag up to 7 days. For children (< 15 years old) the GLM associated grass pollen for both wet and dry seasons. The wet model season positively associated ascospore concentration and negatively associated maximum temperature and rainfall. This ascospore influence proved to be almost as strong as that for grass pollen. The dry season model positively associated deuteromycetes. They found no strong association with air pollutants although they note that other studies show a link. Additionally their results suggest that delay or accumulation of environmental factors takes 1-2 days. This study had the limitation of analyzing only one year of data.

In Auckland, New Zealand, Black et al. [3] found that 20 of 37 patients (54%) admitted to an intensive care unit (ICU) had a positive skin test for one or more fungal allergens compared with 15 of 50 patients (30%) in each of the other groups ($p = 0.005$). The ICU patients were no more likely to have positive skin tests for the grass mix, cat dander, or house-dust mite than the other patients. They concluded that "a positive skin test for fungal allergens is a risk factor for admission to an intensive care unit with an acute attack of asthma".

In Ankara, Turkey, Inal et al. [13] studied a group of 19 children mono-sensitized to only molds for 1 year. Since they lacked controls each child acted as her own. They found a significant correlation of outdoor fungi concentrations with average monthly rhinitis score ($r = 0.877, p < 0.001$), asthma score ($r = 0.831, p = 0.001$), morning peak expiratory flow (PEF) ($r = -0.741, p = 0.006$), evening PEF ($r = -0.720, p = 0.008$), and climatic conditions. Despite these results they could not correlate symptom scores with concentrations of specific fungal taxa even though 14

(73%) of the patients exhibited Cladosporium and/or Alternaria sensitization in skin prick tests. They used a Spearman correlation test. They did not consider trends and seasonality which may confound their analysis. They also seemed to have overlooked other confounding factors such as viruses.

In a panel study analyzing longitudinal data in Southern California, Delfino et al. [10] followed 22 individuals for 8 weeks. They tested the relation between symptom score, PEF, and as-needed inhaler use explained by fungal spore concentrations, pollen counts, air quality and meteorological factors. Using random effects longitudinal regression models controlled for autocorrelation and weather they found an association between fungal spores and worsening symptom scores, PEF, and inhaler use. They also noted that high temperatures improved symptoms and PEF. They concluded that "exposure to fungal spores can adversely effect the daily respiratory status of some asthmatics".

Raphoz et al. [21] performed a time series analysis of emergency department visits for pediatric asthma in Montreal. Performing a time series analysis for daily asthma visits as Poisson random variables they found "lagged positive effects between Cladosporium and Basidiospores and initial emergency department visits". Some past studies however have shown no association between fungal spores and asthma. Furthermore they commented that "a study conducted by Lierl and Hornung found no association between concentrations of total spores and emergency department visits and hospitalizations for asthma. It is difficult to interpret these negative findings".

Many studies focus in adults and have temperate settings which differs greatly from our issue. We want to explain pediatric asthma in the tropical archipelago of Puerto Rico. The tropical setting accounts for year-round fungal spores at levels

that often surpass main land US levels. Furthermore for our purposes we note that the tropical climate in Puerto Rico may differ greatly from that of all other studies exposed here since they took place in temperate climates.

Chapter 1

Methods

1.1 Data

We obtained monthly emergency room visit and hospitalization counts due to pediatric asthma counts at the University of Puerto Rico Hospital at Carolina(UPR-HC) from Dr Jeannette Lube and at the University of Puerto Rico Pediatric Hospital (UPR-PH) from Dr Yiamira Oquendo. Additionally Dr Lourdes Pedraza provided hospitalization counts from San Jorge Children's Hospital (SJCH) . Cases reported varied in age by hospital. UPR-HC counted asthma from 0 to 17 years, UPR-PH from 0 to 18 years, and SJCH from 0 to 21 years. The diagnostic corresponded to ICD-9 codes 493.90, 493.91, or 493.92. These differed from human respiratory syncytial virus (RSV) bronchiolitis (code 466.11) and non-RSV bronchiolitis (code 466.19).

Dr Benjamín Bolaños (National Allergy Bureau (NAB) of the American Academy of Asthma, Allergy, and Immunology (AAAAI)) supplied monthly airborne fungal spore concentrations and pollen counts. Spore counts included total spores, Ascospores, Aspergillus/Penicillium, Basidiospores, Cladosporium, Coprinus, Diatripellas, and Ganoderma. Total spore concentrations dated from June 2006 up to December 2008 while all other spores dated from January 2006 up to December 2008. Dr

Bolaños's team used a Burkard air sampler located at the roof (30m above ground level) of the University of Puerto Rico's Medical Sciences Campus main building in San Juan. The National Allergy Bureau of the American Academy of Asthma Allergy and Immunology certified this station for the report 24/7 the pollen and mold spores levels in San Juan. The station was established in 2005.

Additionally, the Puertorican Environmental Quality Board (JCA by its Spanish name) specified carbon monoxide (CO), ozone (O_3), and 2.5 and 10 micron particulate matter levels ($PM_{2.5}$ and PM_{10}, respectively) from 2006 through 2008. CO data got measured at Baldorioty de Castro Ave. and Fernández Juncos Ave. in San Juan; O_3 at Rte. 165 in Cataño and Rte. 183 in Juncos (Juncos not used due to missing values); PM_{10} at Baldorioty de Castro Ave., William Jones St. (Río Piedras), and Covadonga Terminal in San Juan, and $PM_{2.5}$ at Baldorioty de Castro Ave in San Juan.

Finally the National Oceanic and Atmospheric Administration (NOAA) facilitated meteorological information. Measures corresponded to monthly averages of daily dew point, temperature, precipitation, wind and gust speed and direction, and maximum and minimum relative humidity. We ultimately excluded pollen and PM_{10} from our analyses due to missing values. We also excluded minimum relative humidity due to high variability.

Please refer to table C.1 in appendix C for a list of the variables used, their source and how we used them.

1.2 Exploratory Statistical Analysis

We used statistical software *R* [20] complemented with *Microsoft Excel* [6], *OpenBUGS* [14], and *R* packages *dlm* [18] and *R2OpenBUGS* [24] for our analyses.

1.2.1 Preliminary Data Sets

This project started with a database of infantile asthma cases provided by a local insurance company reporting island wide infantile asthma cases. We attempted to analyze some generalized linear models (GLMs) distributed as a Poisson family, but found no suitable model. Since this data came from a private insurance company and it did not include a population at risk we supposed that a strong bias for the data and continued to look alternate sources.

While waiting for data from hospitals to arrive we attempted to reproduce results from [7]. We approximated the summary of hospital visits and fungal spore counts from a graph in their paper and obtained historical meteorological information from [25]. Using generalized linear models distributed as a Poisson family we found that temperature explained hospital visits best (results not shown). Our analysis had strong limitations such as lacking both original data and time-series analysis.

1.2.2 Visual Inspection and Correlations

Once UPR-HC, UPR-PH, and SJCH provided data we realized that quality greatly improved despite the obvious limitation of having monthly counts. We visually inspected graphs of the ordered hospital, meteorologic, aero-allergen, and air quality measures to note any apparent relations. We observed that UPR-PH had very few cases per month including many months without any cases (6 out of 31) and decided

best to disregard this hospital for further analysis. UPR-HC and SJCH exhibited similar annually seasonal patterns with SJCH showing it more clearly and with more cases. Furthermore SJCH had an increasing trend in pediatric asthma cases. In contrast UPR-HC and UPR-PH did not have any clear increasing or decreasing trend.

Through visual comparison of SJCH cases with meteorological, aero-allergen, and air quality data we noted some apparent correlations. Asthma cases seemed to show inverse relation with relative humidity, dew point, and temperature. They also seemed to have a direct relation with ozone. Other variables did not show any clear relations. Other visual comparisons did not reveal any consistent clear trends: SJCH and UPR-HC with total spore counts, and meteorological and air quality data with spore counts.

Correlation analysis did not yield any particularly insightful information between spore counts and environmental variables. We considered Pearson, Spearman, and Kendall correlation tests for pairwise complete observations. The Spearman coefficients seemed to have the tendency of lying between the other two. Different spore counts showed the highest correlations with total spores (minimum Spearman correlation coefficient 0.63 for Diatripellas and ranging up to 0.89 for Basidiospores) except for Cladosporium (0.27). Apart from these fairly strong correlations, only the y component of wind gusts showed some correlation with total spores (0.50).

Asthma at SJCH showed highest Pearson correlation coefficient with Ozone (0.53). At UPR-HC the maximum occurred with the x-component of wind gusts (0.58) and with Diatripellas (0.61) when considering total spores and by family, respectively.

1.2.3 Generalized Linear Models

Next UPR-HC, UPR-PH, and SJCH provided data which allowed us to consider Poisson distributed GLM for each hospital. We also transformed this Poisson data

under a square root yielding normal data with mean equal to the square root of the original underlying Poisson distribution and fixing the variance at 0.25 (refer to appendix A for a proof). Attempting to minimize either Akaike's information criterion (AIC) or the Bayesian information criterion (BIC) [23] we still obtained inadequate models.

As a preliminary indicator of interaction between explanatory variables we used Poisson distributed GLMs for each hospital's cases explained by a spore count (either total or by family), a meteorologic or air quality variable, and the interaction term between the two. We chose significant terms as those with coefficients having p-values smaller than 0.05.

1.2.4 Time Series Analysis

Since we obtained monthly pediatric asthma counts at three local hospitals we started performing time series analysis of the data. Under the square root transform we considered dynamic linear regression models with normally distributed errors and time varying variances for each hospital according to the conventions used by Petris et al. [18]:

$$\begin{aligned} y_t &= F_t\theta_t + \nu_t, & \nu_t &\sim N_m(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + \omega_t, & \omega_t &\sim N_p(0, W_t), \end{aligned}$$

for time $t \geq 1$ where y_t represents pediatric asthma counts, θ_t explanatory variables, ν_t error in the observations, ω_t system errors, and F_t and G_t known coefficient matrices ($m \times p$ and $p \times p$ respectively). This type of model, also called Gaussian linear state space models, have a Normal prior distribution for the p-dimensional state vector at

time $t = 0$ [18]:

$$\theta_0 \sim N_p(m_0, C_0).$$

In our case we modeled hospitals independently so we worked with three separate models with $m = 1$. The value p varied and equaled the number of explanatory variables each model considered.

We used library *dlm* in *R* to generate models using function *dlmModReg*. We also considered seasonal factor (*dlmModSeas*) and Fourier models (*dlmModTrig*). We chose to analyze each of these models separately. Additionally we considered models that included both regressive and seasonal components (for example *dlmModSeas + dlmModReg*). We calculated the Kalman filter and smoother for each model using functions *dlmFilter* and *dlmSmooth*, respectively. Unfortunately the singular value decompositions (SVDs) related to the Kalman smoothing of these models contained negative values which prevented us from calculating variance and probability intervals. The *dlm* package failed to calculate non-negative SVDs whether or not we used the variance stabilizing square root transformation.

Noticing an increasing trend at SJCH we decided to eliminate the tendency and model the residuals. We justified this technique since only SJCH exhibited a growing tendency which may result from marketing of the private hospital. We used successive differences and residuals from a Poisson distributed GLM to take out the tendency where the former method proved more useful. Unfortunately the models still produced negative SVDs.

Despite the SVD limitation the dynamic linear models helped with model selection by calculating mean absolute deviance (MAD), mean square error (MSE), and Theil's

U (U) criteria. We define these as:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}e_t^2$$

$$U = \sqrt{\frac{\sum_{t=1}^{n}e_t^2}{\sum_{t=1}^{n}(y_t - y_{t-1})^2}}$$

where $e_t = y_t - f_t$, and the f_t's represent the Kalman filtered forecasts. Theil's U "compares the MSE of the model with the MSE of the trivial 'no-change' model that predicts the next observation to be the same as the current one" [18].

1.3 Simulations: Markov Chain Monte Carlo Methods

Simulations using Markov Chain Monte Carlo (MCMC) methods based on the Gibbs sampler allowed us to get around the negative SVD problem and produce models with credibility intervals. We used the open-source software *OpenBUGS* (**B**ayesian inference **U**sing **G**ibbs **S**ampling) via *R* using package *R2OpenBUGS* . We continued using the square root transform for Poisson data to normality. Since SJCH exhibited an increasing trend we used the residuals of a linear model explained by time:

$$SJCH = 6.4496 + 0.1343t$$

where $t = 1, 2, 3, \ldots, 31$. The residuals of this model effectively took out the tendency of the original data. Subsequent models attempted to explain variances in these residuals. Furthermore, for multi-variate models we subtracted the medians from explanatory variables (before calculating an interaction term if applicable) in order to eliminate auto-correlation of model parameters between iterations.

We adjusted a model similar to Peter Congdon's market share, promotion and prices dynamic example (Example 8.9 in [5]). We applied the varying coefficient model:

$$y_t = \beta_{0,t} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t} + \beta_{3,t}x_{1,t}x_{2,t} + \epsilon_t. \tag{1.3.1}$$

The dynamic linear model with discounting assigns normal (random walk) priors and a vague gamma prior for time varying regressors and observation precisions, respectively. We chose regression priors

$$\beta_{j,t} \sim N(\beta_{j,t-1}, W_{j,t})$$

where we specified the time varying precisions, $P_{j,t} = 1/W_{j,t}$, using discount factors

$$P_{j,t} = (\delta_j)^{t-1} P_{j,1}$$

with $\delta_j = 0.90$ for $j = 0, 1, 2, 3$. $P_{j,1}$ represents the initial precision which we distributed as $P_{j,1} \sim Ga(0.1, 0.1)$. Additionally the model discounted a vague gamma prior for time varying precisions of observations:

$$P_{\epsilon,t} = 1/V_{\epsilon,t} = (\delta^*)^{t-1} P_{\epsilon,1}, \qquad t = 2, \ldots 31, \text{ with initial}$$
$$P_{\epsilon,1} = 1/V_{\epsilon,1} \sim Ga(0.01, 0.01)$$

where $\delta^* = 0.99$. In each time period our choice of $\delta_j = 0.90$, $j = 0, 1, 2, 3$, and $\delta^* = 0.99$ represent approximately a 10% and a 1% increase in uncertainty for regressor and observation precisions, respectively.

Please note that for the rest of this thesis we will use the following notation to refer to the dynamic model above (1.3.1). We take the liberty to write a variable

identifier instead of $x_{i,t}, i = 1, 2$:

$$y_t \sim x_{1,t} + x_{2,t} \text{ for a two variable model and} \tag{1.3.2}$$

$$y_t \sim x_{1,t} : x_{2,t} \text{ for the three variable model with interaction.} \tag{1.3.3}$$

In order to eliminate autocorrelation between model parameters (coefficients, precisions, etc.) we ran 1 chain with $10,000$ iterations, an initial burn of 500 and a thinning of $1,000$. After the first 500 iterations got discarded *OpenBUGS* stored 1 out of every 1,000 iterations (discarded 999) until it reached 9,500. Through R (using *R2OpenBUGS*) we generated initial values in order to initialize the model where we assumed that regressors had no effect on outcomes: $P_{\epsilon,t} \sim Ga(0.5, 0.5)$, $\beta_{0,t} \sim N(\bar{y}, W = 100)$, and $\beta_{j,t} = 0$ for all values of j and t.

OpenBUGS introduced the ease of working with the deviance information criterion (DIC) for model selection. According to Spiegelhalter et al. "differences in DIC are estimates of differences in expected loss in prediction" where lower DIC imply better overall quality-of-fit [23]. Consequently we aimed to minimize DIC for our models. Furthermore, models which differ by 1-2 are highly comparable. Those which differ by 3-7 have "considerably less support" [23]. Consequently we opted to choose models with lower DIC which correspond to those greater predictive potential. We combined DIC with the other model selection criteria previously discussed: MAD, MSE and Theil's U.

We generated one-step ahead forecast, posterior mean, coefficient, normal quantile-quantile, autocorrelation function (ACF), and Ljung box-statistic plots, among other to analyze residuals. Due to high ACF for SJCH we re-adjusted the dynamic model after taking the residuals of a seasonal (1 cycle per year) Fourier model with observational variance equal to the variance of historical data ($V = mean(y)$), system

variance $W = 0.01$, initial mean vector equal to monthly averages (months 1-11; eg.: $m0[l] = mean(y_{l'})$ where the $y_{l'}$'s are all the y_t that fall on the same calendar month) and initial variance matrix identity. Afterwards we dropped the first value and considered one-step ahead forecast residuals for modeling. Note that we adjusted this Fourier model after calculating the square root transform and linear model residuals.

Lastly, due to right-skewed residuals we decided to take natural logarithmic transform on observational and/or explanatory variables. If we took the natural logarithm of the observations (y_t) then we did not apply the square root transform and vice-versa. Furthermore for the SJCH models we considered the natural logarithm with and without applying the Fourier seasonal model residuals.

We include the *R* code for simulations along with the model passed to *OpenBUGS* using *R2OpenBUGS* in appendix B.

Chapter 2

Results

2.1 MCMC Simulations using *OpenBUGS*

Models for pediatric asthma cases simulated using total spore counts performed as good or better (smaller or equal MAD, MSE, Theil's U, and DIC) than models using individual spore families. For both San Jorge Children's Hospital (SJCH) and the University of Puerto Rico's Hospital at Carolina (UPR-HC) we chose the one with total spores, average monthly temperature and interaction as explanatory variables. We present results of model selection criteria, one-step ahead forecasts, time varying posterior means, and regressor coefficients, among others including 95% credibity intervals (CIs).

By aiming to minimize the aforementioned selection criteria we narrowed down dynamic linear regression models to the following single- and bi-variate alternatives[1]:

$$y_t \sim \beta_{0,t} + \beta_{1,t}x_{1,t},$$

$$y_t \sim \beta_{0,t} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t}, \text{ and}$$

$$y_t \sim \beta_{0,t} + \beta_{1,t}x_{1,t} + \beta_{2,t}x_{2,t} + \beta_{3,t}x_{1,t}x_{2,t}$$

[1] refer to (1.3.2) for details on notation.

where t represents time in months ($t = 1 \ldots 31$ with $t = 1$ corresponding to June 2006), y_t asthma cases at the hospital, and the $x_{i,t}$'s either total spore concentration, ozone concentration, average temperature, average maximum relative humidity, and/or average dew point for $i \in 1, 2$. In the single variate case we chose either total spore or ozone concentration or average temperature for $x_{i,t}$. In the bivariate case without interaction we narrowed down choices to having total spore concentration as $x_{1,t}$ and either ozone concentration, average temperature, average maximum relative humidity, or average dew point as $x_{2,t}$. Finally in the bivariate case with interaction we ended up with total spore concentration as $x_{1,t}$ and either ozone concentration, average temperature, or average dew point as $x_{2,t}$. These represent the general considerations across hospitals but vary for individual ones.

2.2 SJCH

Originally the DIC served as our selection criterion, yet later (page 36) regressor coefficients helped further refine our selection. Model DIC's ranged from 39.8 (for ozone concentration on its own) up to 51.6 (for total spores + dew point with interaction). Total spores and ozone combined fared second best with DIC equal to 43.3 and 43.4 for models with and without interaction, respectively.

The other selection criteria did not vary as much. MAD stayed constant at 1. MSE ranged from 1.4 up to 1.6. Finally Theil's U remained at 1.1. All of MAD, MSE, and Theil's U had rather small variations in their corresponding 95% CIs (mostly overlapped).

Table 2.1 shows the results for model selection criteria and selected models.

Upon analyzing autocorrelation function (ACF) and Ljung box-statistic plots for

model residuals explaining pediatric asthma cases at SJCH by ozone concentration alone we observed high ACF at lag one and significant Ljung box-statistic p-values (< 0.05) for lags 1, 4, 5, 7, 8, and 12. Figure 2.1 shows relevant plots. Consequently we proceeded to take Fourier seasonal model residuals for modeling.

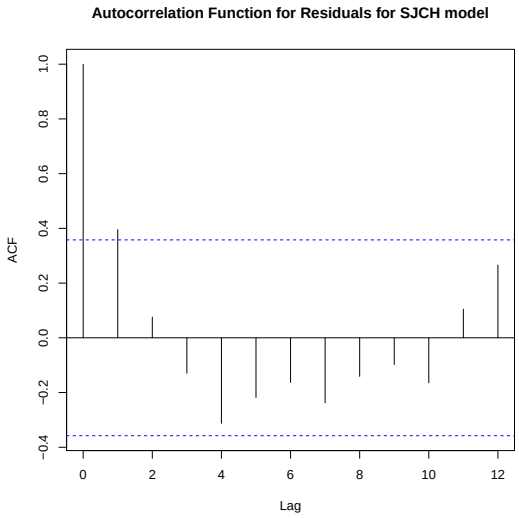
Model for SJCH	**DIC**	**MAD**	**MSE**	**Theil's U**
Total Spores	45.4	1 (0.8, 1.3)	1.4 (1, 2.5)	1.1 (0.9, 1.4)
O3	39.8	1 (0.8, 1.2)	1.4 (1, 2.2)	1.1 (0.9, 1.3)
totSpores + O3	43.4	1 (0.8, 1.3)	1.4 (1, 2.4)	1.1 (0.9, 1.4)
totSpores + Temp	46.4	1 (0.8, 1.4)	1.5 (1.1, 2.9)	1.1 (0.9, 1.5)
totSpores + Dew	47.5	1 (0.8, 1.4)	1.5 (1.1, 3)	1.1 (0.9, 1.6)
totSpores : O3	43.3	1 (0.8, 1.3)	1.4 (1, 2.5)	1.1 (0.9, 1.4)
totSpores : Temp	50.9	1 (0.8, 1.4)	1.5 (1.1, 2.9)	1.1 (0.9, 1.5)
totSpores : Dew	51.6	1 (0.9, 1.5)	1.6 (1.1, 3.6)	1.1 (0.9, 1.7)

Table 2.1: Results for model selection criteria at SJCH included mean absolute deviance (MAD), mean square error (MSE), Theil's U, and deviance information criterion (DIC) with 95% credibity intervals. Regressors were total spore concentration (Total Spores, totSpores), ozone concentration (O_3), average monthly temperature (Temp), and average monthly dew point (Dew). Models showing a colon (:) have both regressors and an interaction term.

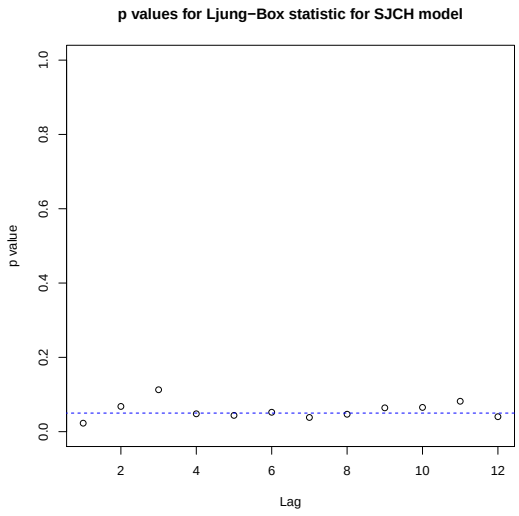
2.2.1 SJCH - Fourier

Just as before, the DIC served as our first indicator, yet later (page 36) regressor coefficients helped further refine our selection. Model DIC's ranged from 11.4 (for ozone concentration on its own) up to 42.5 (for total spores + average temperature + interaction). Total spores and ozone combined fared second best with DIC equal to 20.1 and 20 for models with and without interaction, respectively.

The other selection criteria did not vary as much. MAD took values from 1.9 up to 2.1. MSE ranged from 6.5 up to 7.2. Finally Theil's U stayed between 1.1 and 1.2. All of MAD, MSE, and Theil's U had rather small variations in their corresponding



(a) Autocorrelation Function for SJCH ∼ Ozone



(b) Ljung Box-Statistic for SJCH ∼ Ozone

Figure 2.1: Residual autocorrelation for SJCH ∼ Ozone. Pediatric asthma cases at SJCH explained by ozone concentration.

95% CIs (mostly overlapped).

Table 2.2 shows the results for Fourier residual model selection criteria and selected models. Figure 2.2 shows one step ahead forecasts and the posterior mean calculated by the model explaining pediatric asthma cases at SJCH with atmospheric O_3 concentration. We show residual analysis of ACF and Ljung box statistic for this model in figure 2.3.

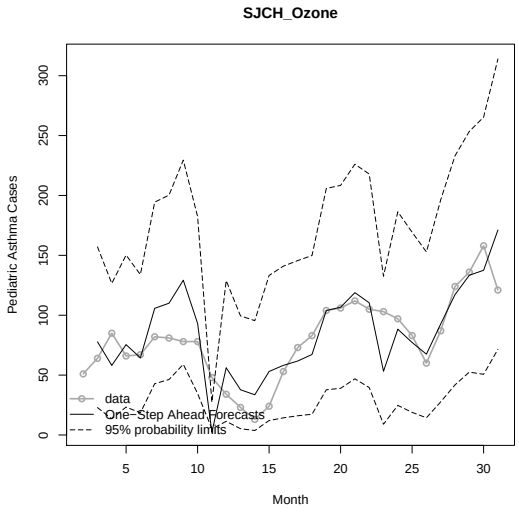
Additionally we include the second best models which interest us since they include total fungal spore concentrations: SJCH $\sim$ Total Spores + O_3 with and without interaction. Plots for forecasts, and residual correlation analysis appear in figures 2.4, 2.5, 2.6, and 2.7. Analysis of residual distribution (along with plots) get discussed on page 40.

Model for SJCH	**DIC**	**MAD**	**MSE**	**Theil's U**
Total Spores	25.3	2 (1.5, 2.8)	6.8 (3.8, 12.3)	1.1 (0.9, 1.6)
O_3	11.4	2 (1.5, 2.8)	6.8 (3.6, 12.2)	1.1 (0.8, 1.6)
totSpores + O_3	20	2.1 (1.5, 2.9)	7.2 (3.8, 13.3)	1.2 (0.9, 1.6)
totSpores : O_3	20.1	2.1 (1.5, 2.9)	7.2 (3.8, 13.2)	1.2 (0.9, 1.6)
totSpores : temp	42.5	1.9 (1.4, 2.7)	6.5 (4.1, 11.7)	1.1 (0.9, 1.5)

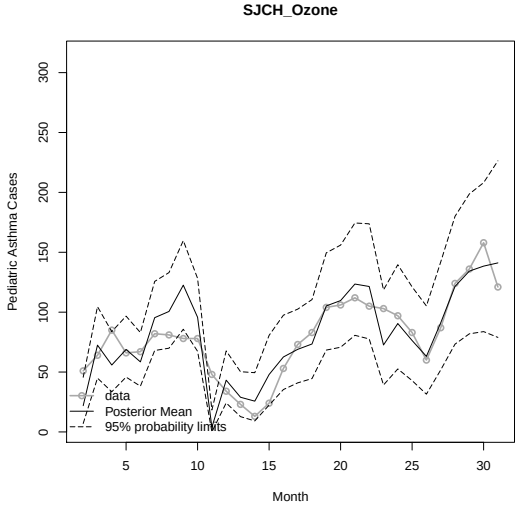
Table 2.2: Results for model selection criteria at SJCH after taking Fourier seasonal model residuals included mean absolute deviance (MAD), mean square error (MSE), Theil's U, and deviance information criterion (DIC) with 95% credibity intervals. Regressors were total spore concentration (Total Spores, totSpores), ozone concentration (O_3), and average monthly temperature (Temp). Models showing a colon (:) have both regressors and an interaction term.

2.3 UPR-HC

The DIC served as our first indicator as it did for SJCH model selection and later (page 36) regressor coefficients helped further refine our selection. Model DIC's ranged

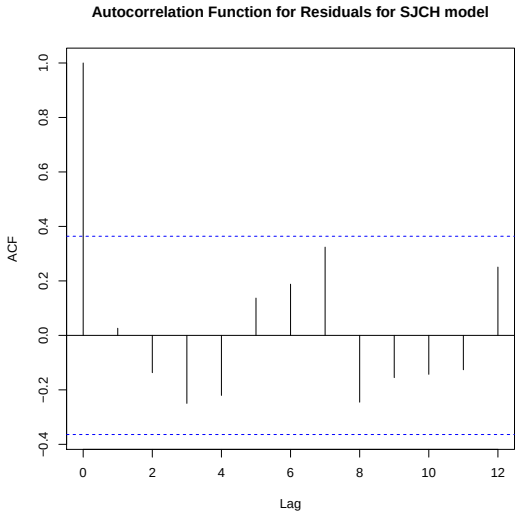


(a) One-Step Ahead Forecasts for SJCH $\sim$ Ozone after Fourier

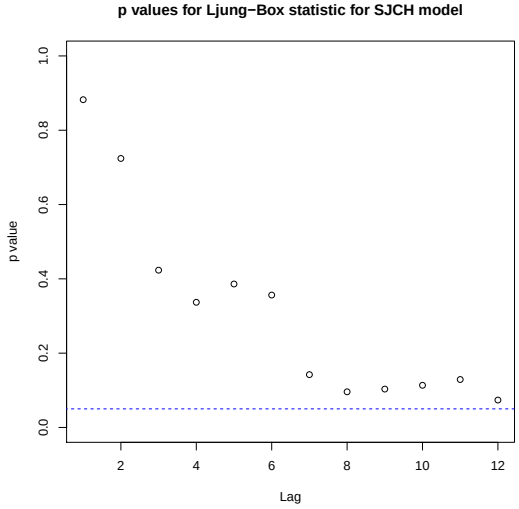


(b) Posterior Means for SJCH $\sim$ Ozone after Fourier

Figure 2.2: Forecasts for SJCH $\sim$ Ozone after Fourier. Pediatric asthma cases at SJCH explained by ozone concentration after considering Fourier seasonal model residuals.

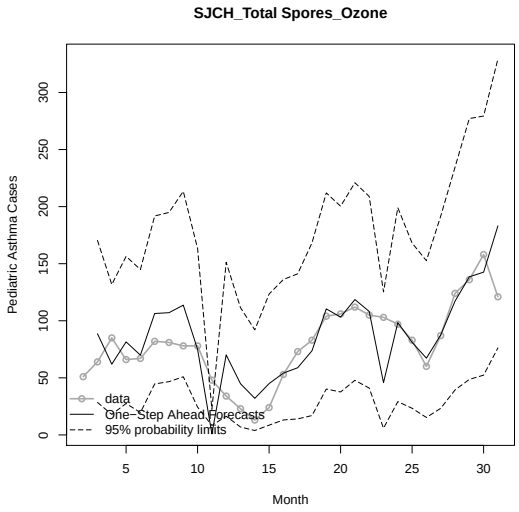


(a) Autocorrelation Function for SJCH ∼ Ozone after Fourier

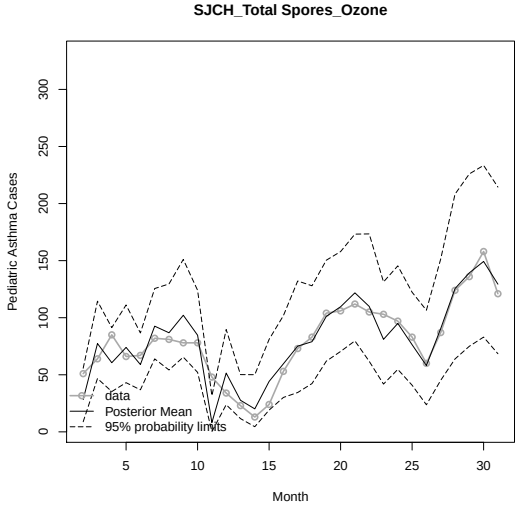


(b) Ljung Box-Statistic for SJCH ∼ Ozone after Fourier

Figure 2.3: Residual autocorrelation for SJCH ∼ Ozone after Fourier. Pediatric asthma cases at SJCH explained by ozone concentration after considering Fourier seasonal model residuals.

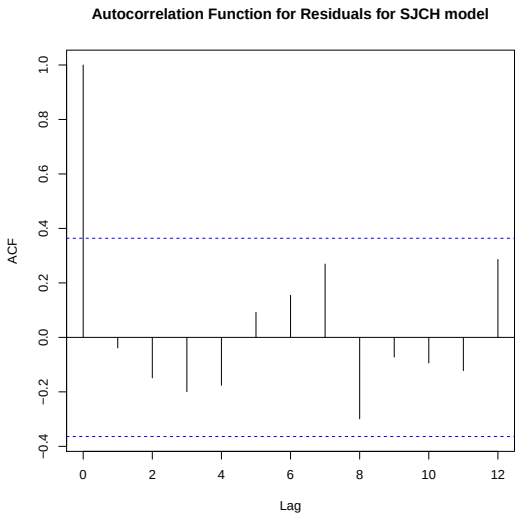


(a) One-Step Ahead Forecasts for SJCH $\sim$ Total Spores + O_3 after Fourier

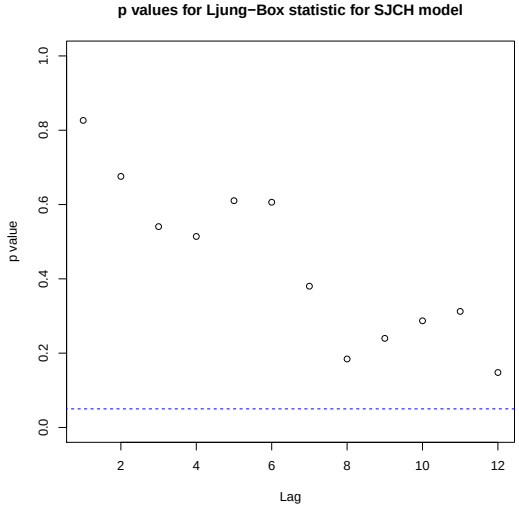


(b) Posterior Means for SJCH $\sim$ Total Spores + O_3 after Fourier

Figure 2.4: Forecasts for SJCH $\sim$ Total Spores + O_3 after Fourier. Pediatric asthma cases at SJCH explained by total spore and ozone concentration after considering Fourier seasonal model residuals.

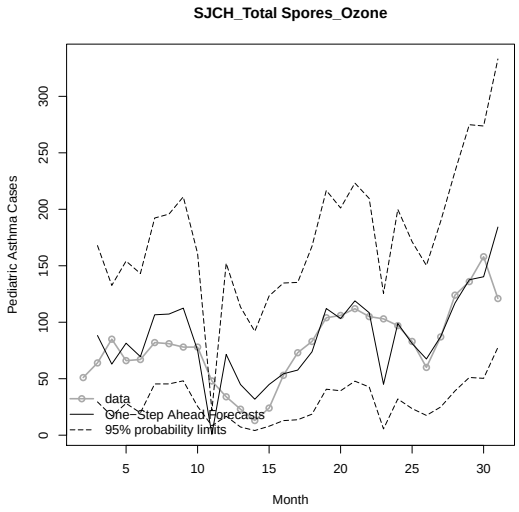


(a) Autocorrelation Function for SJCH $\sim$ Total Spores + O_3 after Fourier

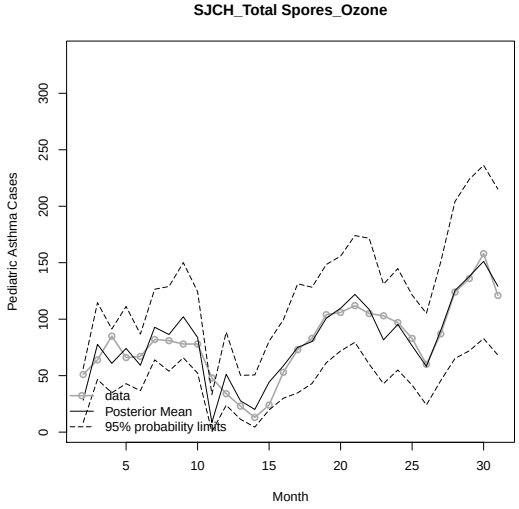


(b) Ljung Box-Statistic for SJCH $\sim$ Total Spores + O_3 after Fourier

Figure 2.5: Residual autocorrelation for SJCH $\sim$ Total Spores + O_3 after taking Fourier seasonal model residuals. Pediatric asthma cases at SJCH explained by total spore and ozone concentration after considering Fourier seasonal model residuals.

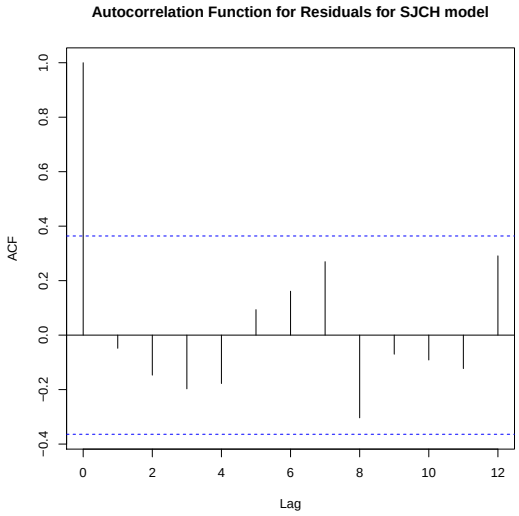


(a) One-Step Ahead Forecasts for SJCH $\sim$ Total Spores : O_3 after Fourier

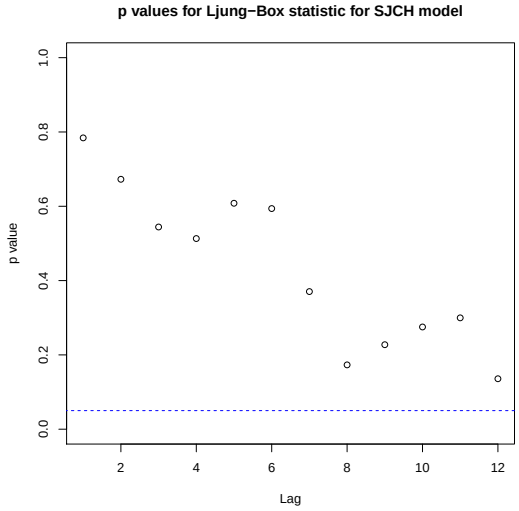


(b) Posterior Means for SJCH $\sim$ Total Spores : O_3 after Fourier

Figure 2.6: Forecasts for SJCH $\sim$ Total Spores : O_3 after taking Fourier seasonal model residuals. Pediatric asthma cases at SJCH explained by total spore and ozone concentration with interaction after considering Fourier seasonal model residuals.



(a) Autocorrelation Function for SJCH $\sim$ Total Spores : O_3 after Fourier



(b) Ljung Box-Statistic for SJCH $\sim$ Total Spores : O_3 after Fourier

Figure 2.7: Residual autocorrelation for SJCH $\sim$ Total Spores : O_3 after taking Fourier seasonal model residuals. Pediatric asthma cases at SJCH explained by total spore and ozone concentration with interaction after considering Fourier seasonal model residuals.

Model UPR-HC	**DIC**	**MAD**	**MSE**	**Theil's U**
Total Spores	39.6	0.9 (0.7, 1.2)	1.3 (0.9, 2.3)	1.1 (0.9, 1.5)
O_3	12.7	1 (0.7, 1.3)	1.5 (0.8, 2.7)	1.2 (0.9, 1.6)
temp	76	1.3 (0.7, 3.4)	3.4 (0.9, 18.4)	1.6 (1, 4.3)
totSpores + O_3	25.7	1 (0.7, 1.3)	1.5 (0.9, 2.8)	1.2 (0.9, 1.7)
totSpores + temp	44.8	0.9 (0.7, 1.2)	1.2 (0.9, 2.2)	1.1 (0.9, 1.5)
totSpores : O_3	25.7	1 (0.7, 1.3)	1.5 (0.9, 2.8)	1.2 (0.9, 1.7)
totSpores : temp	49.2	0.9 (0.7, 1.2)	1.2 (0.9, 2.3)	1.1 (0.9, 1.5)

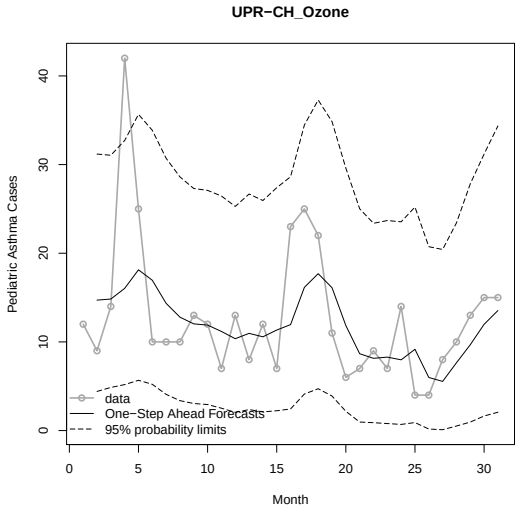
Table 2.3: Results for model selection criteria at UPR-HC included mean absolute deviance (MAD), mean square error (MSE), Theil's U, and deviance information criterion (DIC) with 95% credibity intervals. Regressors were total spore (Total Spores, totSpores), and ozone concentrations (O_3). Models showing a colon (:) have both regressors and an interaction term.

from 12.7 (for ozone) up to 76 (for temperature). Total spores plus ozone with and without interaction fared second best with both DICs equal to 25.7.

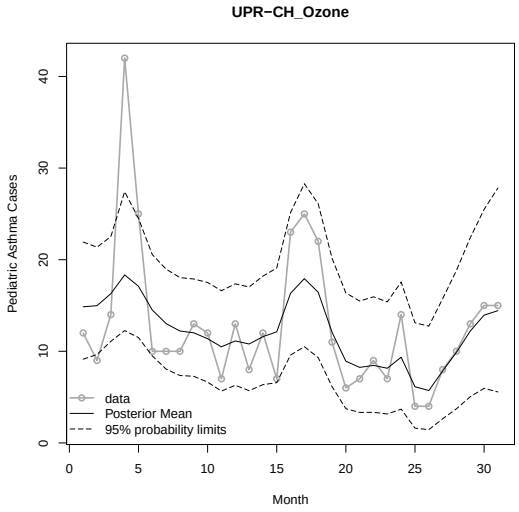
The other selection criteria did not vary as much. MAD ranged from 0.9 to 1.3. MSE mostly ranged from 1.3 up to 1.5, yet UPR-HC $\sim$ temperature had MSE equal to 3.4 (95% CI 0.9 to 18.4). Finally Theil's U stayed at 1.1 and 1.2 except for UPR-HC $\sim$ temperature which had U = 1.6 (95% CI 1 to 4.3). Disregarding the model with only temperature as an explanatory variable which showed greater intervals, all of MAD, MSE, and Theil's U had rather small variations in their corresponding 95% CIs.

Table 2.3 shows the results for model selection criteria and selected models. Figure 2.8 shows one step ahead forecasts and the posterior mean calculated by the model.

In contrast to models for SJCH, residual ACF and Ljung box-statistics did not show any autocorrelation between residuals so we did not analyze models with a seasonal Fourier component. Figure 2.9 shows relevant graphs. Analysis of residual distribution (along with plots) get discussed on page 40.

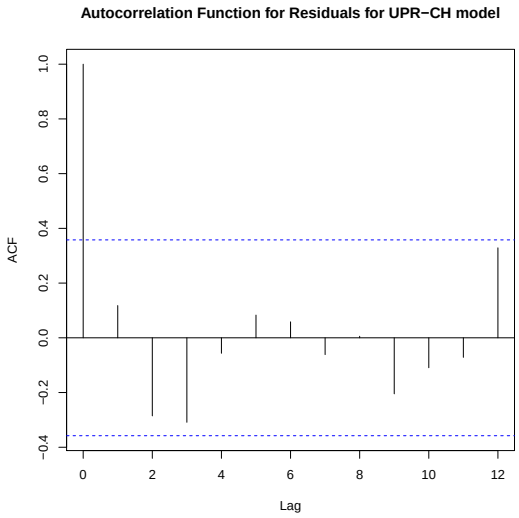


(a) One-Step Ahead Forecasts for UPR-HC $\sim$ O_3

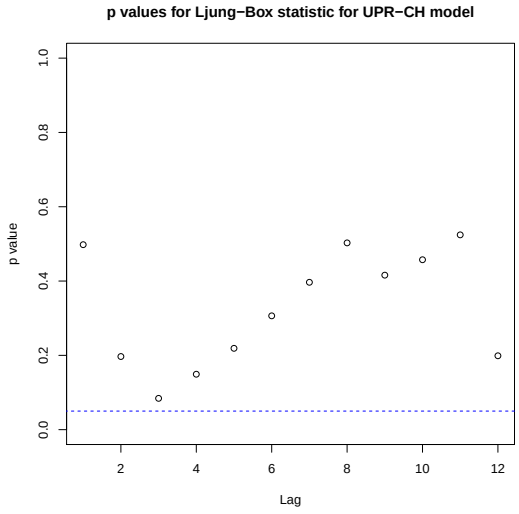


(b) Posterior Means for UPR-HC $\sim$ O_3

Figure 2.8: Forecasts for UPR-HC $\sim$ O_3. Pediatric asthma cases at UPR-HC explained by ozone concentration.



(a) Autocorrelation Function for UPR-HC $\sim$ O_3



(b) Ljung Box-Statistic for UPR-HC $\sim$ O_3

Figure 2.9: Residual autocorrelation for UPR-HC $\sim$ O_3. Pediatric asthma cases at UPR-HC explained by ozone concentration.

Chapter 3

Discussions and Conclusions

3.1 Data

3.1.1 Disadvantages and Limitations of the Data

Even though we managed to narrow down the selection of our models, we realize that this research has the main limitation of counting with biased and filtered data. Bias gets introduced when we only consider hospitals in Puerto Rico -San Jorge Children's Hospital (SJCH) and the University of Puerto Rico's Hospital at Carolina (UPR-HC)- and we do not have a target population for them. We consider the establishment of such a target population for these hospitals an impossibility since anybody can walk into a hospital and receive treatment. This applies to at least SJCH due to its private ownership.

The private nature of SJCH introduces additional biases. One relates to the measurement of a group that has the funds to pay for such the service. This would most likely exclude the lower classes or people without health insurance. Additionally we observed an increasing trend for SJCH (and did not show at UPR-HC) which we speculated might have resulted from marketing (including but not limited to billboards, T.V. commercials, newspaper ads, and a website).

Apart from biases introduced by region and private ownership, we also experienced the one produced by monthly totaling of asthma cases. Monthly data prevented us from analyzing the time series in its entirety since we could not compare daily aeroallergen, meteorologic, and air quality measures and compare them to the quantity of pediatric asthma cases on the same day (lag=0 days) or subsequent days (lag$\geq$ 1). Additionally, monthly conglomeration of asthma counts introduced the question of which statistics (such as totals, means, minimums, maximums, etc.) to consider for other variables in the study. Even though we considered monthly averages of explanatory variables we recognize that perhaps different statistics could serve better. Examples include maximum fungal spore concentrations, change in allergen concentration, presence of ozone (or other variables) above some threshold, and number of dry days per month.

We also had the technical limitation of receiving not a number (NaN) errors from *R* package *dlm* when we used our data. We still do not know why this happened but Petris et al. [17] faced a similar problem when working with package *KFAS*. They too got stuck with NaNs when dealing with Poisson data.

3.1.2 Advantages of the Data

Reporting of asthma by diagnostic codes in hospitals gives the data used for this research an advantage by eliminating possible confounders such as viral infections and bronchiolitis (due to respiratory syncytial virus or other causes) which has played a role in other research [7–9, 13]. This may also help resolve the issue of subjective diagnostic by the patient as seen in [2].

3.2 Models

Despite the limitations discussed above we selected a few models. Due to the limitations we recognize that these models serve as possible ones to predict infantile asthma in Puerto Rico. These should not get considered as definitive predictive models. They undoubtedly serve as dynamic predictive models for the monthly data and might serve as starting points for future research with, hopefully, complete daily data. Furthermore, we do not know if explanatory variables exhibit a causal relation to asthma cases. They simply serve as predictors.

At present the strongest model we have explains infantile asthma through just atmospheric ozone concentration(3.2.1). The same model resulted best for both SJCH and UPR-HC despite having considered Fourier residuals only at SJCH. This fact makes the model more credible. However, the consideration of Fourier residuals at SJCH to explain seasonal patterns simply serves as a proxy for other or unknown variables that could take part of the model. More analysis regarding this seasonality would benefit future research. Furthermore, second best models at both hospitals also matched up despite the same difference that only SJCH considered Fourier seasonal residuals. Pediatric asthma cases explained by total airborne fungal spore and ozone concentrations without (3.2.2) and with (3.2.3) interaction fared second best. The inclusion of ozone as an explanatory variable may reflect the increased sensitivity to allergens that people experience when they get exposed to the air pollutant [12].

Best Model According to DIC:

$$y_t \sim Ozone_t \tag{3.2.1}$$

Second Best Models According to DIC:

$$y_t \sim totalSpores_t + O3_t \text{ and} \tag{3.2.2}$$

$$y_t \sim totalSpores_t + O3_t + (totalSpores_t)(O3_t) \tag{3.2.3}$$

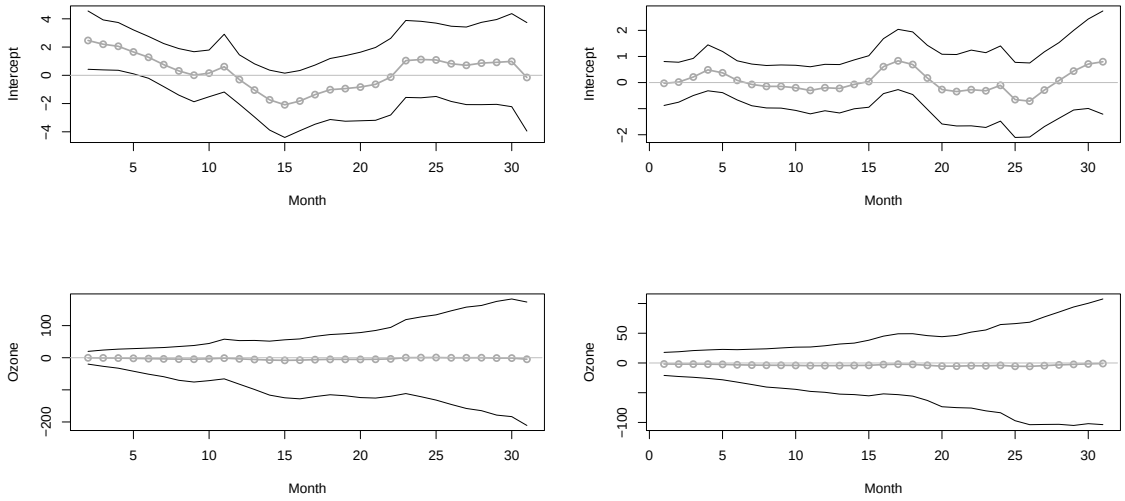
Please refer to equation (1.3.2) for details on notation.

Since Theil's U criterion never dropped below 1, we know that our models did not fare better, on average, with respect to Theil's model of "no-change". U values of 1.1 for model (3.2.1) and 1.2 for models (3.2.2) and (3.2.3) mean that in general the model that predicts today's asthma cases equal to yesterday's on average will have 10% and 20% less error than our models, respectively. Our models still have the innate advantage of having explanatory capacity bestowed by regressors which Theil's model lacks.

3.2.1 Coefficient Analysis

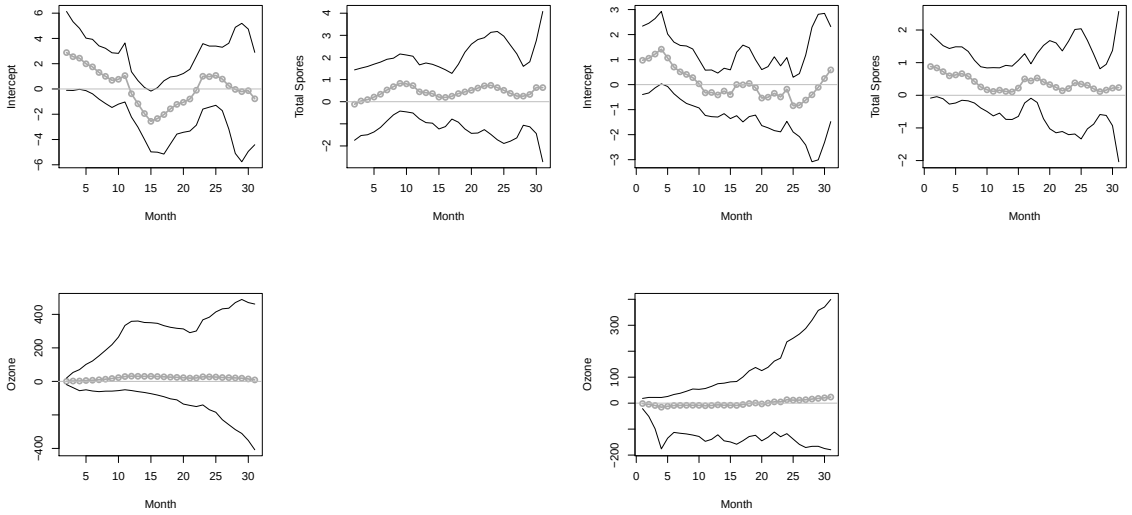
After analyzing model selection criteria we proceeded to look at individual model coefficients. The ozone coefficient for the best model (3.2.1) at either hospital stayed very close to 0 while the intercept varied more closely to model predictions. The 95% CI and variance of the ozone coefficient increased with time (see figure 3.1) which lead us to question this model and consider "second best" options.

Even though models with both fungal spore and ozone concentrations still showed a widening 95% probability band about ozone's coefficient they showed steadier CIs for total spore concentrations (figures 3.2 and 3.3). Total airborne fungal spore concentration had a positive relation with pediatric asthma on both models for both



(a) SJCH $\sim O_3$ (b) UPR-HC $\sim O_3$

Figure 3.1: Coefficient plots for pediatric asthma cases at SJCH and UPR-HC explained by atmospheric ozone concentration. SJCH model considered Fourier seasonal model residuals.

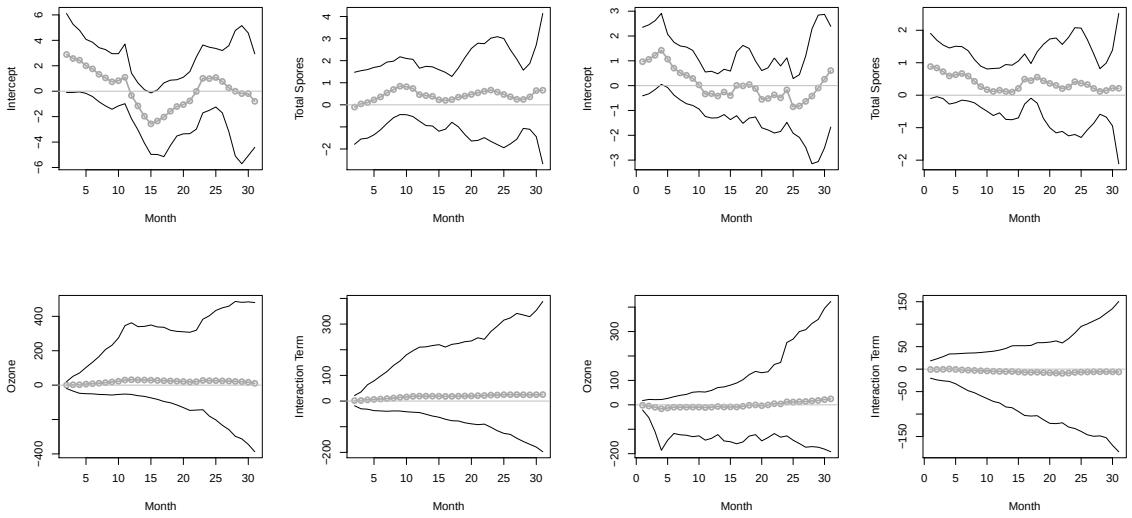


(a) SJCH $\sim$ totalSpores + O_3 (b) UPR-HC $\sim$ totalSpores + O_3

Figure 3.2: Coefficient plots for pediatric asthma cases at SJCH and UPR-HC explained by atmospheric fungal spore and ozone concentrations. SJCH model considered Fourier seasonal model residuals.

hospitals. Ozone also did except that it started negative and eventually went positive in the interaction model for UPR-HC. The interaction term varied between hospitals: it stayed positive for SJCH and negative for UPR-HC. This difference between interaction term coefficients led us to question if the model really needed it.

Widening 95% probability band for ozone coefficients puts in doubt the quality and credibility of the data as it suggests high variability. Furthermore, due to the nature of an air pollutant, the measurement may not necessarily represent the concentration across a region due to dispersion factors like wind speed and direction.



(a) SJCH $\sim$ totalSpores : O_3 (b) UPR-HC $\sim$ totalSpores : O_3

Figure 3.3: Coefficient plots for pediatric asthma cases at SJCH and UPR-HC explained by atmospheric fungal spore and ozone concentrations with interaction. SJCH model considered Fourier seasonal model residuals.

3.2.2 Residual Analysis

The shapes of Normal Q-Q plots suggested right-skewed residual distributions which we verified by drawing residual histograms. Figure 3.4 show the Normal Q-Q plots and corresponding residual histograms for models explaining asthma by just ozone concentration. Figures 3.5a & 3.5b, and 3.6a & 3.6b show those explaining SJCH with both total spore and ozone concentrations (with or without interaction) and suggested the positive skewed distributions. The two outliers at the top right-hand corners of the Normal Q-Q plots (labeled 9 & 21 for months $t = 11 and 23$) correspond to the month of April for years 2007 and 2008. This pattern seems to match the most drastic increases in precipitation experienced during the study: from March (average monthly precipitation 0.07in and 0.02in for 2007 and 2008, respectively) to April (average monthly precipitation 0.33in and 0.16in for 2007 and 2008, respectively). It seems that our model has difficulty adjusting for the effect that big changes in precipitation has on hospital admissions in SJCH. Table 3.1 shows precipitation for these months.

Date	Precipitation	Presence of Outlier
March 2007	0.07	No
April 2007	0.33	Yes
March 2008	0.02	No
April 2008	0.16	Yes

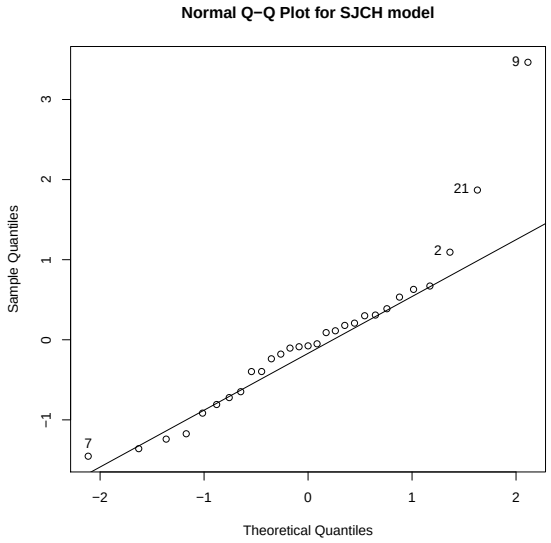
Table 3.1: Precipitation and Outliers. An increase in precipitation coincided with outliers for the models at SJCH.

Residuals for models with both total spore and ozone concentrations (with or without interaction) at UPR-HC also had two outliers at the top right-hand corner of the plots (figures 3.5c and 3.6c). These two points also suggested a distribution

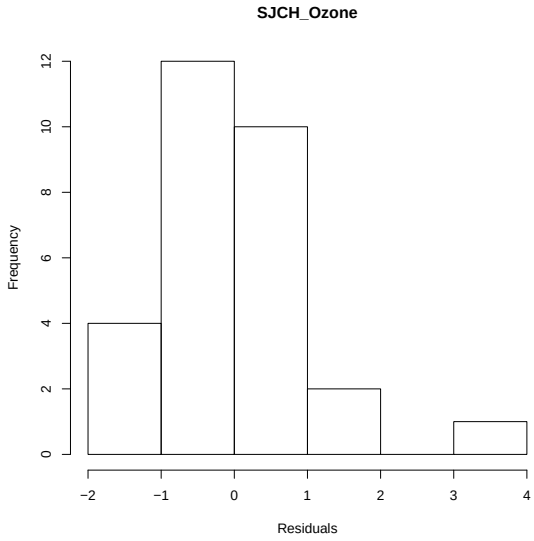
somewhat skewed to the right which we saw in the corresponding histograms (figures 3.5d and 3.6d). In this case points labeled 3 & 15 correspond to $t = 4, 16$ (month of September 2007 and 2008, respectively). These outliers (as do the ones for the SJCH models) merit further study as we did not identify any relation to precipitation or other variables for these outliers.

So given the deviations of residuals from Normal Q-Q lines and the distribution of histograms we suspect that residuals do not follow normal distributions. They exhibited somewhat right-skewed distributions. We proceeded to take a natural logarithmic transform to test if it would result in normally distributed residuals. We considered taking the natural logarithm of observations (with and without Fourier for SJCH) and/or explanatory variables. The resulting models at SJCH had high ACF and significant Ljung box-statistics, or residuals still showed right-skewness. At UPR-HC models still showed residuals with positive skew. For simplicity we just show the ACF and Normal Q-Q plots for the two variable model (total spores and ozone) using the logarithm of both response and explanatory variables at SJCH with Fourier (figures 3.7a, 3.6a) and UPR-HC (figures 3.7c, 3.6c). Due to high ACF and skewed residuals we decided to discard these models.

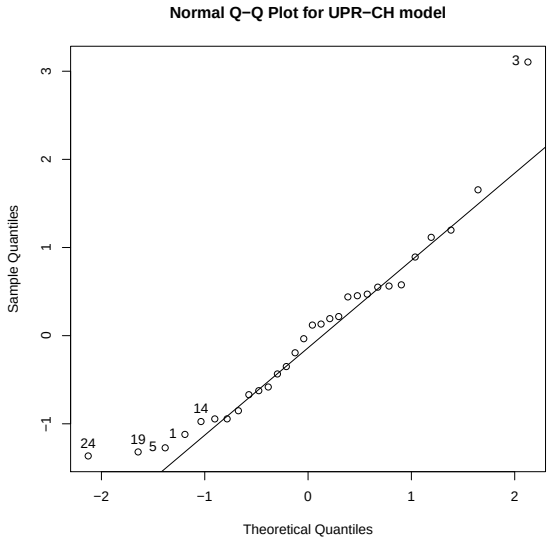
Using the square root transformation of the response variables, the model with just ozone had widening probability band on the regressor's coefficient. Additionally coefficients for the interaction term varied in sign between both hospitals so we discarded these models and had our final "best" model left, despite non-normal residuals and other limitations.



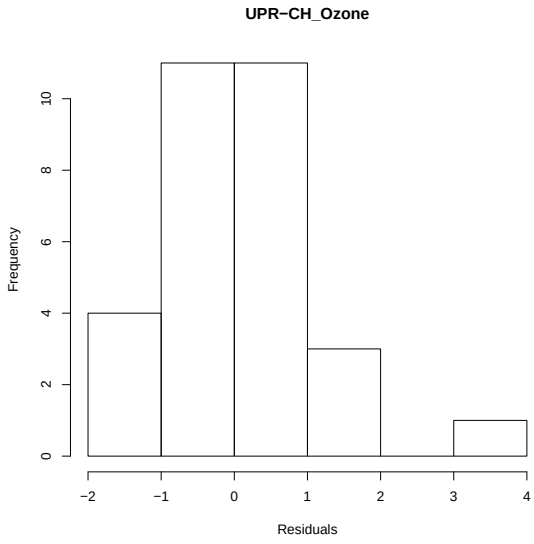
(a) SJCH∼O_3



(b) SJCH∼O_3

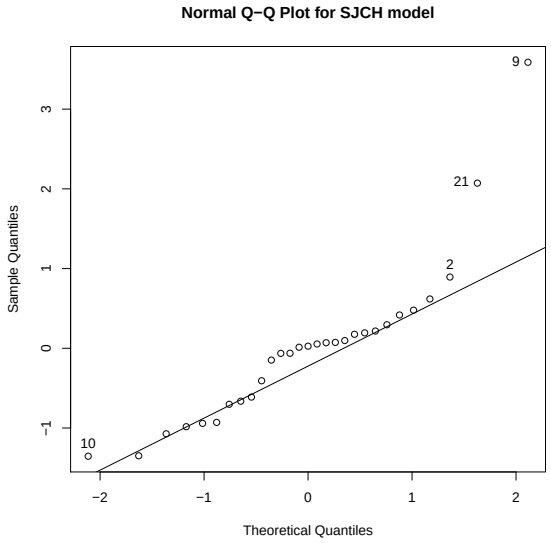


(c) UPR-HC∼O_3

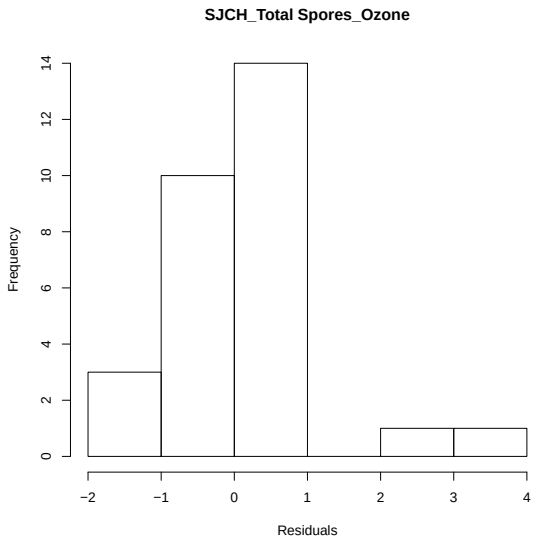


(d) UPR-HC∼O_3

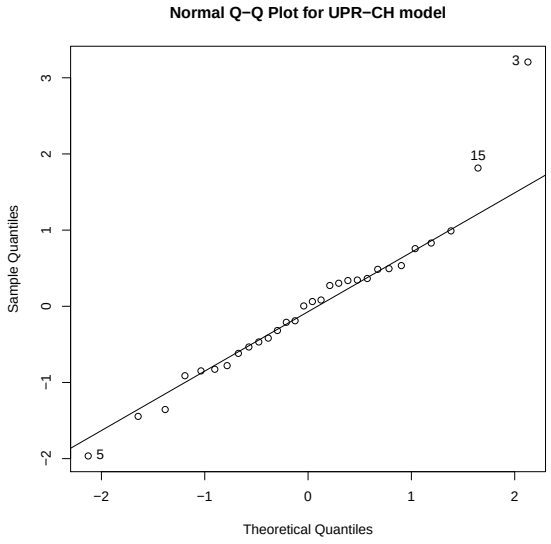
Figure 3.4: Residual diagnostic plots for pediatric asthma cases at SJCH and UPR-HC explained cases explained by atmospheric ozone concentrations. The model at SJCH considered Fourier seasonal residuals. Graphs include normal quantile-quantile (Q-Q) plots and residual histograms. Note that residuals seem to follow a right-skewed distribution.



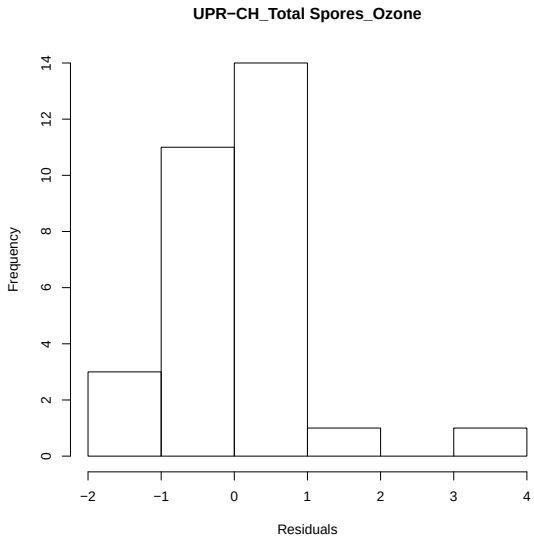
(a) SJCH∼totSpores+O_3



(b) SJCH∼totSpores+O_3

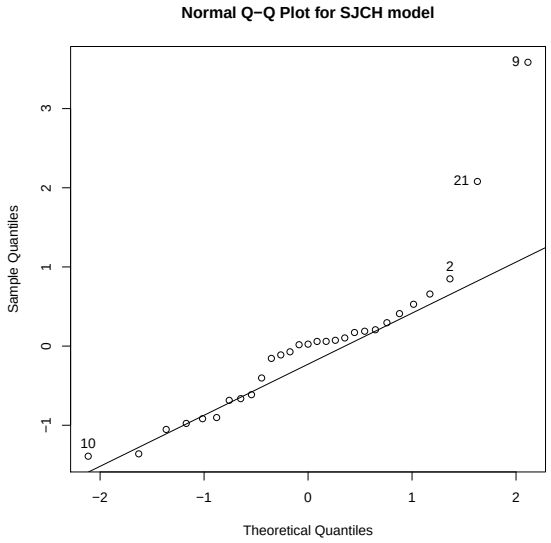


(c) UPR-HC∼totSpores+O_3

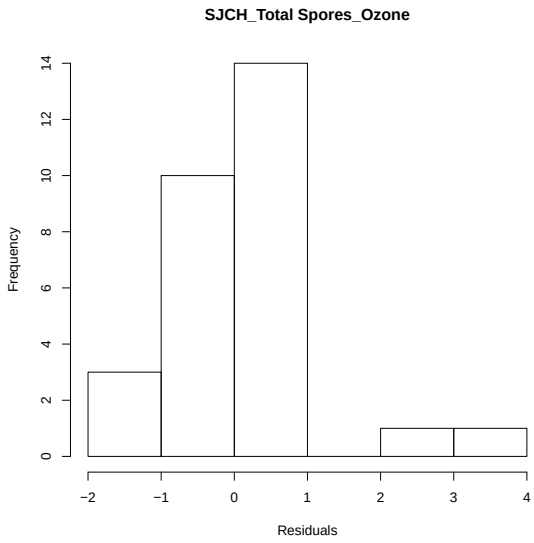


(d) UPR-HC∼totSpores+O_3

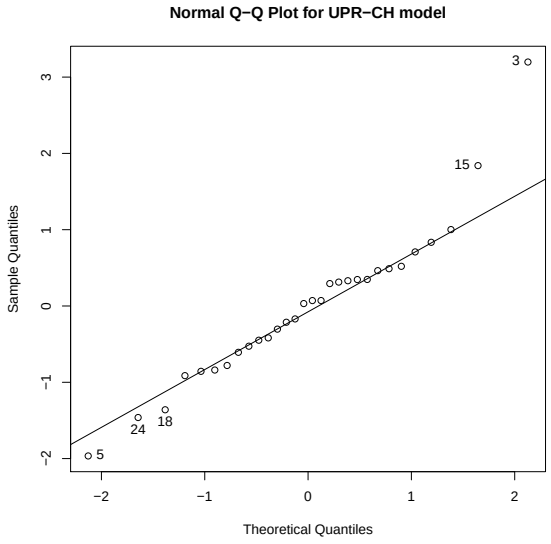
Figure 3.5: Residual diagnostic plots for pediatric asthma cases at SJCH and UPR-HC explained by total airborne fungal spore and ozone concentrations. The model at SJCH considered Fourier seasonal residuals. Graphs include normal quantile-quantile (Q-Q) plots and residual histograms. Note that residuals seem to follow a right-skewed distribution.



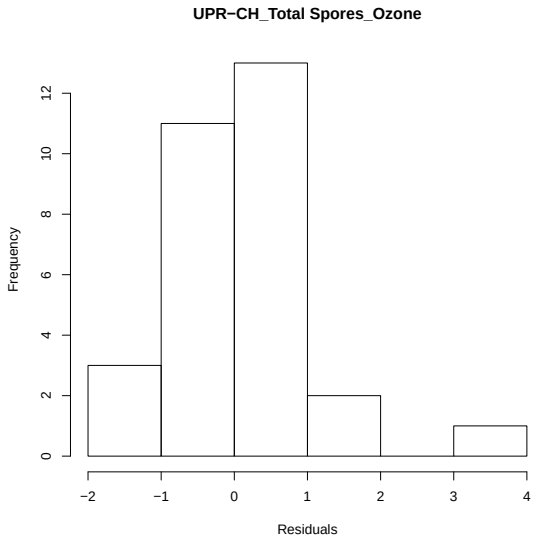
(a) SJCH∼totSpores:O_3



(b) SJCH∼totSpores:O_3

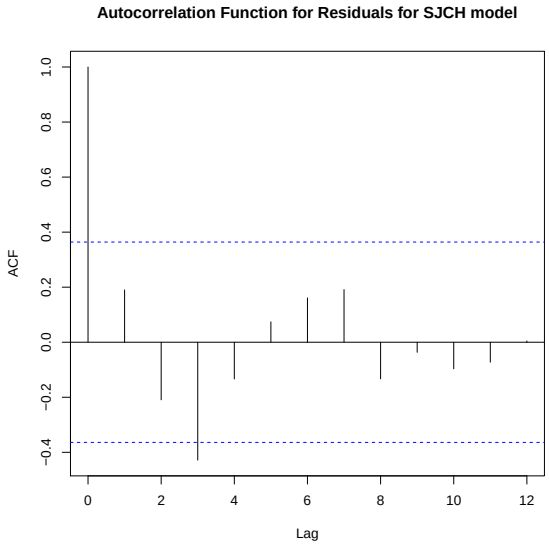


(c) UPR-HC∼totSpores:O_3

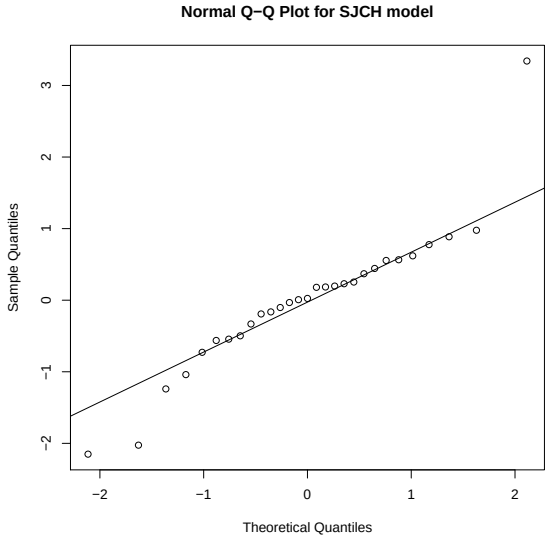


(d) UPR-HC∼totSpores:O_3

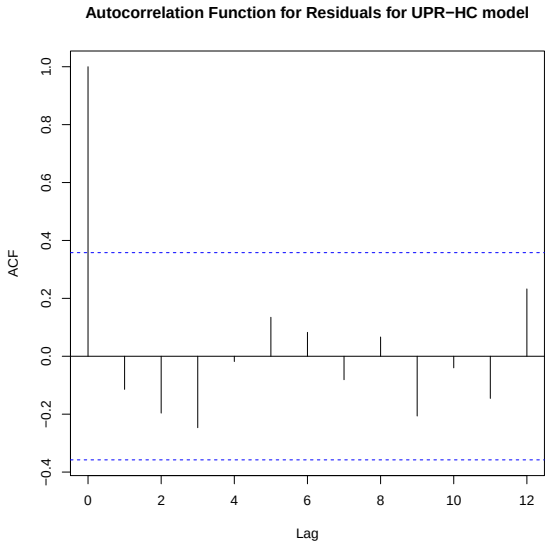
Figure 3.6: Residual diagnostic plots for pediatric asthma cases at SJCH and UPR-HC explained by total airborne fungal spore and ozone concentrations with interaction. The model at SJCH considered Fourier seasonal residuals. Graphs include normal quantile-quantile (Q-Q) plots and residual histograms. Note that residuals seem to follow a right-skewed distribution.



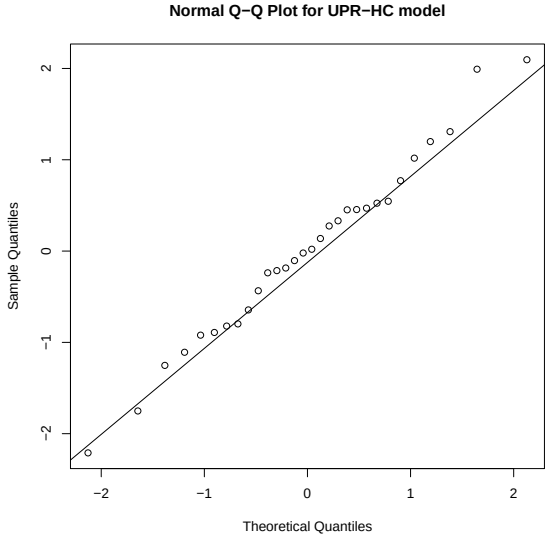
(a) log(SJCH)∼log(totSpores)+log(O_3)



(b) log(SJCH)∼log(totSpores)+log(O_3)



(c) log(UPR-HC)∼log(totSpores)+log(O_3)



(d) log(UPR-HC)∼log(totSpores)+log(O_3)

Figure 3.7: Residual diagnostic plots for pediatric asthma cases at SJCH and UPR-HC explained by total airborne fungal spore and ozone concentrations with interaction on natural logarithmic transformation for all variables. The model at SJCH considered Fourier seasonal residuals. Graphs include autocorrelation function (ACF) and normal quantile-quantile (Q-Q) plots. Residuals seem to follow a right-skewed distribution as with previous models.

"Best" Model:

$$y_t \sim totalSpores_t + O3_t$$

3.3 Conclusions

In response to the high incidence of infantile asthma in Puerto Rico (PR), this study aimed to predict infantile asthma based on fungal spore, pollen, pollutant concentrations, and/or meteorological factors. A predictive model would allow for the creation of an alert to inform the general public about the risk of infantile asthma on a daily basis. Simulating dynamic linear models with discount factors using *OpenBUGS*, *R*, and *R2OpenBUGS* we constructed models which explained pediatric asthma cases at two local hospitals (San Jorge Children's Hospital and the University of PR's Carolina Hospital) by atmospheric ozone concentration or by a combination of total airborne fungal spore and ozone concentrations with and without interaction. High autocorrelation of residuals led us deseasonalize using a Fourier model for San Jorge Children's Hospital. By minimizing the deviance information criterion (DIC), and analyzing model coefficients and residuals we chose model $y_t \sim totalSpores_t + O3_t$. Residuals seemed to follow right-skewed distributions and we did not manage to approximate normality for any model. This result undoubtedly serves as a dynamically predictive model for the monthly data and might serve as starting points for future research with more complete, hopefully daily, data.

The inclusion of both ozone and fungal spores as explanatory variables in the resulting model held particular importance as it may reflect the increased sensitivity to allergens (such as fungal spores) that people experience from exposure to ozone [12].

The location of Ozone observation (Route 165 in Cataño) may or may not represent levels throughout the metropolitan area. It likely does not represent the entire island. This limitation related to the spot of measurement may impact on the increasing variability of ozone's coefficient, yet we cannot ascertain. A similar argument applies to fungal spore counts, yet we did not observe increasing coefficient variability in this case. Additionally, having fungal spores compares to similar research in Canada [7, 9, 21].

Further analysis should attempt to better clarify outliers and to fit a regressive model explaining seasonality at San Jorge. Furthermore we would like to compare our model with a non-regressive one that combines linear growth and seasonality. Future research should also work on establishing a network of daily and island-wide asthma case registry from hospitals and physicians. This type of information would greatly assist in creating a model with much better predictive capacity than the present one. Perhaps the federal regulation that will require hospitals to keep electronic records will help...

Appendix A

Variance Stabilizing Square Root Transformation of Poisson Data to Normality

Theorem A.0.1. *(Delta Method) Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

$$\sqrt{n}[g(Y_n) - g(\theta)] \to N(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.} \tag{A.0.1}$$

Consider a sequence of k independent and identically distributed (*iid*) Poisson random variables Y_i, $i = 1, 2, \ldots, k$, such that $Y_1 + Y_2 + \ldots + Y_k = Y \sim \text{Poisson}(\lambda)$ in distribution. Now $E(Y) = Var(Y) = \lambda$ [4]. We choose g such that $g(Y) \sim N(\lambda, 1/4), 0 \leq \lambda < \infty$.

Claim A.0.2. $g(Y) = \sqrt{Y}$

Proof. $g(Y) = \sqrt{Y} \Rightarrow g'(Y) = 1/(2\sqrt{Y})$. Without loss of generality suppose that $Y > 0$ since we cannot divide by 0. Now since $Y \sim \text{Poisson}(\lambda)$ we get $[g(Y) - g(\lambda)] = [\sqrt{Y} - \sqrt{\lambda}] \to N(0, \lambda[g'(\lambda)]^2)$ in distribution according to the Delta Method. Now

$$\lambda[g'(\lambda)]^2 = \lambda[1/(2\sqrt{\lambda})]^2 = \frac{\lambda}{[2\sqrt{\lambda}]^2} = \frac{1}{4} \text{ so}$$

$$[\sqrt{Y} - \sqrt{\lambda}] \to N(0, 1/4) \text{ in distribution}$$

which implies that

$$\sqrt{Y} \to N(\sqrt{\lambda}, 1/4) \text{ in distribution.}$$

□

Appendix B

R Code Specifications and *OpenBUGS* Model

Below we include *R* code along with the model passed to *OpenBUGS* using *R2OpenBUGS*. Note that for 1 and 2 variable models we simply modified the following instructions.

B.1 *R* Code using package *R2OpenBUGS*

Square root transform and linear model (lm) residuals.

```
#Variance stabilizing square root transform
y <- sqrt(y)

#Take out tendency with a linear model (lm)
t = 1:length(y)
ylm <- lm(y~t)
y <- ylm$residuals
```

Seasonal Fourier model fitting.

```
N = N-1
freq=12
m0=1:(freq-1)
C0=diag(nrow=freq-1)
firstObs = 1
```

```
    for(l in 1:(freq-1))
    {
      m0[l]= mean(y[(time(y) %% freq)== l]) #Monthly Averages
    }
    modTrig = dlmModTrig(s=freq, dV=var(y), dW=0.01, m0=m0 ,C0=C0)
    modFilt = dlmFilter(y, modTrig)
    modSmooth = dlmSmooth(modFilt) #dlmSmooth(y, modTrig)
```

R2OpenBUGS variables, data, initial value, and parameter declarations along with call to run model.

```
x1 = as.vector(asma.ts[,sporeIDs[2,j]])
x2 = as.vector(asma.ts[,meteoIDs[2,k]])
x1 = x1 - median(x1); x2 = x2 - median(x2)
x3 = x1*x2
data <- list ("N", "y", "x1", "x2", "x3", "elim",
               "DFStar", "DF0", "DF1", "DF2", "DF3")
inits <- function(){list(P = rgamma(N,shape=0.5,scale=0.5),
                         b0 = rnorm(N, mean(y), 100), b1 = rep(0,N),
                         b2 = rep(0,N), b3 = rep(0,N))}
parameters <- c("b0", "b1", "b2", "b3", "mu",
                 "y.one", "Pobs", "Pb", "MAD", "U", "MSE")

asma.sim <- bugs(data, inits, parameters, model.file,
     n.chains=1, n.iter=10000, n.burnin=500, n.thin=1000)
```

B.2 *OpenBUGS* Model

```
model
{     # observational variance with discounting
                    P ~ dgamma(0.01,0.01)
                    Pobs[1] <- P
   for (t in 2:N) { Pobs[t] <- P*pow(DFStar,t-1)}
# one step ahead forecasts
   for (t in 2:N) { y.one[t]   ~ dnorm(mu[t-1],Pobs[t-1])
                        resid.step[t] <- y.one[t] - y[t]
                        absdev[t] <- abs(resid.step[t])
                        squaredev[t] <- pow(resid.step[t],2)
```

```
                        y.step[t] <- y[t] - y[t-1]
                        y.step.sq[t] <- pow(y.step[t],2)}
                        MAD <- sum(absdev[(elim+1):N])/(N-elim)
                        MSE <- sum(squaredev[(elim+1):N])/(N-elim)
                        U <- sqrt(sum(squaredev[(elim+1):N])/
                              sum(y.step.sq[(elim+1):N]))
# regression priors with discounting
      for (j in 1:4) {Pbeta[j] ~ dgamma(0.1,0.1)}
      for (t in 2:N) {b0[t] ~ dnorm(b0[t-1],Pb[t,1])
                      b1[t] ~ dnorm(b1[t-1],Pb[t,2])
                      b2[t] ~ dnorm(b2[t-1],Pb[t,3])
                      b3[t] ~ dnorm(b3[t-1],Pb[t,4])
# discount factors for level and predictors
                      DF[1,t] <- pow(DF0,t-2)
                      DF[2,t] <- pow(DF1,t-2)
                      DF[3,t] <- pow(DF2,t-2)
                      DF[4,t] <- pow(DF3,t-2)
      for (j in 1:4) {Pb[t,j] <- DF[j,t]*Pbeta[j]}   }
# Initialising regression priors
      b0[1] ~ dnorm(0,0.01)
      b1[1] ~ dnorm(0,0.01)
      b2[1] ~ dnorm(0,0.01)
      b3[1] ~ dnorm(0,0.01)
# on-line mean
      for (t in 1:N) {  y[t]   ~ dnorm(mu[t],Pobs[t])
               mu[t] <- b0[t] + b1[t]*x1[t]  + b2[t]*x2[t] + b3[t]*x3[t]}
}
```

We used discount factors DFStar $= 0.99$ and DF$j = 0.90$ for $j = 0, 1, 2, 3, 4$.

Appendix C

Table of Variables Used

The following table details the nature, source, and usage of variables that we had for the study.

Table C.1: Legend: Emergency Room (ER), Hospitalizations (Hosp), University of Puerto Rico Hospital at Carolina (UPR-HC), University of Puerto Rico Pediatric Hospital (UPR-PH), San Jorge Children's Hospital (SJCH), and Respiratory Syncytial Virus (RSV).

Variable	Source	How did we process it?
ER + Hosp Asthma at UPR-HC	Dr Jeannette Lube	Used raw and under square root or natural logarithm for modeling.
ER + Hosp Asthma at UPR-PH	Dr Yiamira Oquendo	Used raw and under square root or natural logarithm for modeling.
Hosp Asthma at SJCH	Dr Lourdes Pedraza	Used raw and under square root or natural logarithm for modeling.

Variable	Source	How did we process it?
ER, Hosp RSV at UPR-HC	Dr Jeannette Lube	Did not use (outside scope of study).
ER, Hosp Bronchiolitis at UPR-HC	Dr Jeannette Lube	Did not use (outside scope of study).
Hosp RSV at UPR-PH	Dr Yiamira Oquendo	Did not use (outside scope of study).
ER, Hosp Bronchiolitis at UPR-PH	Dr Yiamira Oquendo	Did not use (outside scope of study).
Hosp RSV at SJCH	Dr Lourdes Pedraza	Did not use (outside scope of study).
Hosp Bronchiolitis at SJCH	Dr Lourdes Pedraza	Did not use (outside scope of study).
Total Fungal Spores	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Ganoderma	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Ascospores	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Basidiospores	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Diatripellas	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Aspergillus/Penicillum	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Coprinus	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.
Cladosporium	Dr Benjamín Bolaños	Used raw and under natural logarithm for modeling.

Variable	Source	How did we process it?
Daily Tree Pollen Count	Dr Benjamín Bolaños	Averaged monthly. Did not use due to missing or null values.
Daily Grass Pollen Count	Dr Benjamín Bolaños	Averaged monthly. Did not use due to missing or null values.
Daily Weed Pollen Count	Dr Benjamín Bolaños	Averaged monthly. Did not use due to missing or null values.
Carbon Monoxide Concentration at Ave Baldorioty de Castro	JCA	Used for modeling.
Carbon Monoxide Concentration at Ave Fernández Juncos	JCA	Used for modeling.
Ozone Concentration at Cataño (Rte 165)	JCA	Used for modeling.
Ozone Concentration at Juncos (Rte 183)	JCA	Did not use due to missing values.
10 micron Particulate Matter at Ave Baldorioty de Castro	JCA	Did not use due to missing values.
10 micron Particulate Matter at Río Piedras (William Jones St)	JCA	Did not use due to missing values.
10 micron Particulate Matter at Old San Juan (Covadonga Terminal)	JCA	Did not use due to missing values.
2.5 micron Particulate Matter at Ave Baldorioty de Castro	JCA	Used for modeling.

Variable	Source	How did we process it?
Daily Dew Point	NOAA	Averaged monthly. Used for modeling.
Maximum Daily Relative Humidity	NOAA	Averaged monthly. Used for modeling.
Minimum Daily Relative Humidity	NOAA	Averaged monthly. Used for modeling.
Average Monthly Temperature	NOAA	Used for modeling.
Average Monthly Precipitation	NOAA	Used for modeling.
Wind speed and direction	NOAA	Calculated x and y components then used these for modeling.
Gust speed and direction	NOAA	Calculated x and y components then used these for modeling.

Bibliography

[1] Ritesh Agarwal and Dheeraj Gupta, *Severe asthma and fungi: current evidence*, Medical Mycology **49** (2011), no. S1, S150–S157.

[2] R W Atkinson, D P Strachan, H R Anderson, S Hajat, and J Emberlin, *Temporal associations between daily counts of fungal spores and asthma exacerbations*, Occupational and Environmental Medicine **63** (2006), no. 9, 580–590.

[3] PN Black, AA Udy, and SM Brodie, *Sensitivity to fungal allergens is a risk factor for life-threatening asthma*, Allergy **55** (2000), no. 5, 501–504.

[4] George Casella and Roger L. Berger, *Statistical inference*, 2 ed., Duxbury Press, 2001.

[5] P. Congdon, *Bayesian statistical modelling*, Wiley, 2006.

[6] Microsoft Corporation, *Microsoft office excel 2007*.

[7] Robert E. Dales, Sabit Cakmak, Richard T. Burnett, Stan Judek, Frances Coates, and Jeffrey R. Brook, *Influence of ambient fungal spores on emergency visits for asthma to a regional children's hospital*, Am. J. Respir. Crit. Care Med. **162** (2000), no. 6, 2087–2090.

[8] Robert E. Dales, Sabit Cakmak, Stan Judek, Tom Dann, Frances Coates, Jeffrey R. Brook, and Richard T. Burnett, *The role of fungal spores in thunderstorm asthma**, Chest **123** (2003), no. 3, 745–750.

[9] Robert E. Dales, Sabit Cakmak, Stan Judek, Tom Dann, Frances Coates, Jeffrey R. Brook, and Richard T. Burnett, *Influence of outdoor aeroallergens on hospitalization for asthma in canada*, Journal of Allergy and Clinical Immunology **113** (2004), no. 2, 303 – 306.

[10] R.J. Delfino, R.S. Zeiger, J.M. Seltzer, D.H. Street, R.M. Matteucci, P.R. Anderson, and P. Koutrakis, *The effect of outdoor fungal spore concentrations on daily asthma severity.*, Environmental health perspectives **105** (1997), no. 6, 622.

[11] D. W. Denning, B. R. O'Driscoll, C. M. Hogaboam, P. Bowyer, and R. M. Niven, *The link between fungi and severe asthma: a summary of the evidence*, European Respiratory Journal **27** (2006), no. 3, 615–626.

[12] Maria A. Fierro, Mary Kay O'Rourke, and Jefferey L. Burgess, *What is ozone?*, http://www.airinfonow.org/html/ed_ozone.html, Accessed July 2011.

[13] A. Inal, G.B. Karakoc, D.U. Altintas, M. Pinar, T. Ceter, M. Yilmaz, and S.G. Kendirli, *Effect of Outdoor Fungus Concentrations on Symptom Severity of Children with Asthma and/or Rhinitis Monosensitized to Molds*, Asian Pacific Journal of Allergy and Immunology **26** (2008), 11–17.

[14] David Lunn, David Spiegelhalter, Andrew Thomas, and Nicky Best, *The bugs project: Evolution, critique and future directions*, Statistics in Medicine **28** (2009), no. 25, 3049–3067.

[15] Nation Master, *Mortality statistics > asthma (per capita) (most recent) by country*, http://www.nationmaster.com/graph/mor_ast_percap-mortality-asthma-per-capita, 2004; Accessed July 2011.

[16] Centers for Disease Control National Center for Environmental Health, Air Pollution Prevention, and Respiratory Health Branch, *Child Lifetime Asthma Prevalence Rate (Percent) and Prevalence (Number) by State or Territory:*

BRFSS2008, http://www.cdc.gov/asthma/brfss/08/child/lifetime/tableL1.htm, 2008; Accessed July 2011.

[17] G. Petris and S. Petrone, *State space models in r*, Journal of Statistical Software **41** (2011), no. 4, 1–25.

[18] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic linear models*, Dynamic Linear Models with R (2009), 31–84.

[19] Elizabeth Quintero, Felix Rivera-Mariani, and Benjamín Bolaños Rosero, *Analysis of environmental factors and their effects on fungal spores in the atmosphere of a tropical urban area (san juan, puerto rico)*, Aerobiologia **26** (2010), 113–124, 10.1007/s10453-009-9148-0.

[20] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0, http://www.R-project.org.

[21] M. Raphoz, M.S. Goldberg, M. Garneau, L. Héguy, M.F. Valois, and F. Guay, *Associations between atmospheric concentrations of spores and emergency department visits for asthma among children living in montreal*, Archives of Environmental and Occupational Health **65** (2010), no. 4, 201–210.

[22] I. Rosas, H. A. McCartney, R. W. Payne, C. Calderón, J. Lacey, R. Chapela, and S. Ruiz-Velazco, *Analysis of the relationships between environmental factors (aeroallergens, air pollution, and weather) and asthma emergency admissions to a hospital in mexico city*, Allergy **53** (1998), no. 4, 394–401.

[23] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde, *Bayesian measures of model complexity and fit*, Journal of the Royal Statistical Society. Series B (Statistical Methodology) **64** (2002), no. 4, pp. 583–639 (English).

[24] Sibylle Sturtz, Uwe Ligges, and Andrew Gelman, *R2winbugs: A package for running winbugs from r*, Journal of Statistical Software **12** (2005), no. 3, 1–16.

[25] Weather Underground, *Ottawa, ontario weather station history*, http://www.wunderground.com, 2011.