BAYESIAN MODELLING OF NON-GAUSSIAN TIME SERIES OF SEVERE ACUTE RESPIRATORY ILLNESS

by

RAYMOND N. MUSYOKA

A thesis submitted for the degree of

Doctor of Philosophy

in Biostatistics

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

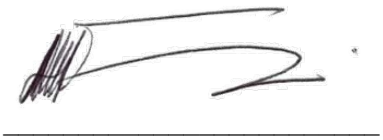
South Africa

Date: 14th January 2019

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Biostatistics, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by the University and Centers for Disease Control and Prevention (CDC)

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.



Signed: Prof. Henry Mwambi

Date: 11 March 2019

Signed: Prof. Thomas N.O. Achia

Date: 14th January 2019

Signed: Dr. Antony Gichangi

Date: 14th January 2019

DECLARATION 1: PLAGIARISM

Note that two declaration sections are required if there are papers emanating from the dissertation/thesis. The first (obligatory) declaration concerns plagiarism and the second declaration specifies your role in the published papers.

I, Raymond Nyoka Musyoka, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Raymond Nyoka Musyoka

Date: 14th January 2019

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated. The * indicates corresponding author.

1. Nyoka R*, Omony J, Mwalili SM, Achia TNO, Gichangi A, Mwambi H (2017), Effect of climate on incidence of respiratory syncytial virus infections in a refugee camp in Kenya: A non-Gaussian time-series analysis. PLoS ONE 12(6): e0178323. https://doi.org/10.1371/journal.pone.0178323
2. Nyoka R*, Achia TNO, Omony J, Mwalili SM, Gichangi A, Mwambi H, Time Series Non-Gaussian Bayesian Bivariate Model Applied to Data on HMPV and RSV: A Case of Dadaab in Kenya. (Submitted to BioMed Central)
3. Nyoka R*, Mwalili SM, Achia TNO, Gichangi A, Mwambi H, A Non-Gaussian Bayesian Model of Multiple Time Series Epidemics of Acute Respiratory Illness: Case of Dabaab in Kenya. (To be submitted)

The research reported on is based on the data collected from a disease surveillance system designed by Kenya Medical Research Institute (KEMRI) and the Centers for Disease Control and Prevention (CDC). I analysed the data and wrote the papers.

Signed: Raymond Nyoka Musyoka

Date: 14th January 2019

ABSTRACT

Respiratory syncytial virus (RSV), Human metapneumovirus (HMPV) and Influenza are some of the major causes of acute lower respiratory tract infections (ALRTI) in children. Children younger than 1 year are the most susceptible to these infections. RSV and influenza infections occur seasonally in temperate climate regions. We developed statistical models that were assessed and compared to predict the relationship between weather and RSV incidence in chapter 2.

Human metapneumovirus (HMPV) have similar symptoms to those caused by respiratory syncytial virus (RSV). Currently, only a few models satisfactorily capture the dynamics of time series data of these two viruses. In chapter 3, we used a negative binomial model to investigate the relationship between RSV and HMPV while adjusting for climatic factors. In chapter 4, we considered multiple viruses incorporating the time varying effects of these components. The occurrence of different diseases in time contributes to multivariate time series data. In this chapter, we describe an approach to analyze multivariate time series of disease counts and model the contemporaneous relationship between pathogens namely, RSV, HMPV and Flu. The use of the models described in this study, could help public health officials predict increases in each pathogen infection incidence among children and help them prepare and respond more swiftly to increasing incidence in low-resource regions or communities. We conclude that, preventing and controlling RSV infection subsequently reduces the incidence of HMPV.

EXTENDED ABSTRACT

Respiratory syncytial virus (RSV) is one of the major causes of acute lower respiratory tract infections (ALRTI) in children. Children younger than 1 year are the most susceptible to RSV infection. RSV infections occur seasonally in temperate climate regions. Based on RSV surveillance and climatic data, we developed statistical models that were assessed and compared to predict the relationship between weather and RSV incidence among refugee children younger than 5 years in Dadaab refugee camp in Kenya. Most time-series analyses rely on the assumption of Gaussian-distributed data. However, surveillance data often do not have a Gaussian distribution. We used a generalised linear model (GLM) with a sinusoidal component over time to account for seasonal variation and extended it to a generalised additive model (GAM) with smoothing cubic splines. Climatic factors were included as covariates in the models before and after timescale decompositions, and the results were compared. Models with decomposed covariates fit RSV incidence data better than those without. The Poisson GAM with decomposed covariates of climatic factors fit the data well and had a higher explanatory and predictive power than GLM. The best model predicted the relationship between atmospheric conditions and RSV infection incidence among children younger than 5 years.

Human metapneumovirus (HMPV) have similar symptoms to those caused by respiratory syncytial virus (RSV). The modes of transmission and dynamics of these epidemics still remain poorly understood. Climatic factors have long been suspected to be implicated in impacting on the number of cases for these epidemics. Currently, only a few models satisfactorily capture the dynamics of time series data of these two viruses. In this study, we used a negative binomial model to investigate the relationship between RSV and HMPV while adjusting for climatic factors. We specifically aimed at establishing the

heterogeneity in the autoregressive effect to account for the influence between these viruses. Our findings showed that RSV contributed to the severity of HMPV. This was achieved through comparison of 12 models of various structures, including those with and without interaction between climatic cofactors.

Most models do not consider multiple viruses nor incorporate the time varying effects of these components. Common ARIs etiologies identified in developing countries include respiratory syncytial virus (RSV), human metapneumovirus (HMPV), influenza viruses (Flu), parainfluenza viruses (PIV) and rhinoviruses with mixed co-infections in the respiratory tracts which make the etiology of Acute Respiratory Illness (ARI) complex. The occurrence of different diseases in time contributes to multivariate time series data. In this work, the surveillance data are aggregated by month and are not available at an individual level. This may lead to over-dispersion; hence the use of the negative binomial distribution. In this paper, we describe an approach to analyze multivariate time series of disease counts. A previously used model in the literature to address dependence between two different disease pathogens is extended. We model the contemporaneous relationship between pathogens, namely; RSV, HMPV and Flu from surveillance data in a refugee camp (Dadaab) for children under 5 years to investigate for serial correlation. The models evaluate for the presence of heterogeneity in the autoregressive effect for the different pathogens and whether after adjusting for seasonality, an epidemic component could be isolated within or between the pathogens. The model helps in distinguishing between an endemic and epidemic component of the time series that would allow the separation of the regular pattern from irregularities and outbreaks. The use of the models described in this study, can help public health officials predict increases in each pathogen infection incidence among children and help them prepare and respond more swiftly to increasing incidence in low-resource regions or communities. This knowledge helps public health officials to prepare for, and respond more effectively to increasing RSV

incidence in low-resource regions or communities. The study has improved our understanding of the dynamics of RSV and HMPV in relation to climatic cofactors; thereby, setting a platform to devise better intervention measures to combat the epidemics. We conclude that, preventing and controlling RSV infection subsequently reduces the incidence of HMPV.

ACKNOWLEDGMENTS

The work presented in this thesis was made possible through the help and generous support of several people without whom this would not have been easy for me to accomplish. In particular I am very grateful to the University of KwaZulu-Natal for their scholarship in my first three years of this journey. I extend the same to my employer the US Centers for Disease Control and Prevention (CDC). The US CDC made available resources which included funding for the data collection, scientific and ethical review of study protocols and funding for my travels to meetings and conferences that I attended in South Africa.

I would like to express my sincere gratitude to my supervisor Prof. Henry Mwambi and co-supervisors Prof. Achia and Prof. Mwalili for their inspiration, guidance and mentorship over the course of my doctoral research work. They shared their immense knowledge and experience, and provided insights that really helped to mould my research ideas. They always provided well thoughtful and very useful feedback and suggestions to my queries that I raised.

I extend the same gratitude to my wife Lucy who did very much understand me when I could not fully stand for the family chores at home. She provided a very conducive environment for my studies and not forgetting my children Luke, Cynthia and Dan who were very accommodative. My mum Marygorret (a teacher) for instilling my mathematical background when I was still very young and my dad Dominic for being there with me at all times supporting my financial needs throughout my education.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research (nature and scope)

The burden on mobility estimates for respiratory syncytial virus (RSV), Human metapneumovirus (HMPV) and influenza associated disease in the refugee camps in Kenya are limited. Although RSV has been shown to circulate throughout the year with biennial peaks, the timing and the meteorological determinants associations has never been described. The correlation between RSV and HMPV in the refugee settings and even in the tropical region has not been studied. Influenza is also known to circulate throughout the year but how its outbreaks relate with those of RSV has not been known. The modelling of the time series events of these viruses will not only help in the prediction of their outbreaks but also in estimating which outbreaks precede each other. The results could also be used by other countries in the tropical zone in Africa with similar settings to inform vaccination timings as control measures to prevent outbreaks.

1.2 Justification

Worldwide, as of 2005, 99% of deaths from RSV were reported by the World Health Organization (WHO) to occur in developing countries. The highest mortalities in 2015 caused by Acute Respiratory Infections (ARI) among children less than five years of age were in Sub-Saharan Africa. Epidemiological knowledge of the respiratory system has been mostly related to developed countries, though the burden of respiratory virus infections (RVIs) is more manifested in developing countries with very high hospitalization and mortality rates. Higher mortality is associated with increased displacement into overcrowded refugee camps. The RSV adversely impacts the health of adults and immunocompromised patients, and is associated with significant mortality and morbidity, particularly in young children and vulnerable infants. Most time-series analyses rely on the assumption of Gaussian-distributed data. However, surveillance data often do not have a Gaussian distribution. The wide range of

statistical methods used to explore the link between RSV outbreaks and climate makes it difficult to elucidate a definitive relationship. It is, therefore, crucial to establish good RSV surveillance systems in developing countries to help understand the dynamics of the disease. The Human metapneumovirus (HMPV) has similar symptoms to those caused by the respiratory syncytial virus (RSV). The modes of transmission and dynamics of these epidemics still remain poorly understood. Currently, only a few models satisfactorily capture the dynamics of time series data of these three viruses.

1.3 Aims

The Main aim of this thesis was to make use of available surveillance data to come up with models that could help explore the dynamics of acute respiratory infections.

1.4 Objectives

This thesis evaluated data for equatorial climatic region to aid accurate predictions of RSV, HMPV and Influenza outbreaks. A better understanding of the relationship between climate and RSV helps in making reliable predictions of its incidence and to establish good surveillance systems in developing countries to help understand the dynamics of the disease. Specifically our objectives were;

i. To explore the best model that predicts the relationship between RSV incidence and climatic factors along spatio-temporal scales to determine whether a seasonal pattern of RSV infection exists.

ii. To investigate the relationship between RSV and HMPV while adjusting for climatic factors.

iii. To establish the heterogeneity in the autoregressive effect to account for the influence

between RSV, HMPV and Influenza viruses.

iv. To assess the presence of influence of high incidences between these viruses and whether higher incidences of one virus are influenced by another and to investigate for serial correlation between them.

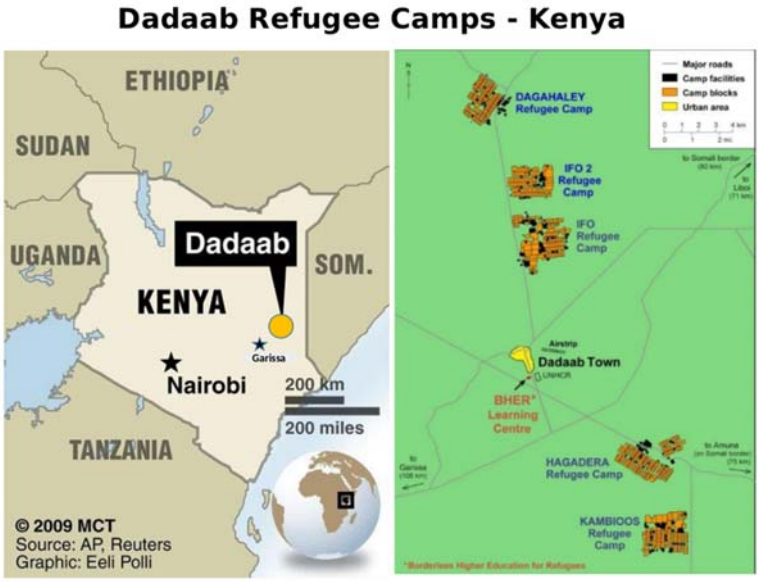


Figure 1.1 A map showing the location of Dadaab refugee camps

1.5 Outline of thesis structure

The objectives listed above were addressed through a surveillance study conducted in the Dadaab refugee camps in Kenya as shown in Figure 1.1 above. Dadaab is located in the east

of the country (Latitude 0ºN, Longitude 40ºE) and consists of five camps namely, Dagahaley, Ifo, Ifo2, Hagadera and Kambios. The neighboring weather station in Garissa, is about 100 kilometers away from the Dadaab camp. The objectives were presented in the chapters described below.

In chapter 2, we model RSV data that also include climatic data and discuss the effects of the climatic conditions on the RSV incidence. Based on RSV surveillance and climatic data, we develop statistical models that we use to assess, compare and predict the relationship between weather and RSV incidence among refugee children younger than 5 years in Dadaab refugee camps in Kenya. Surveillance data often do not have a Gaussian distribution. We therefore, use a generalised linear model (GLM) with a sinusoidal component over time to account for seasonal variation and extend it to a generalised additive model (GAM) with smoothing cubic splines. Climatic factors are included as covariates in the models before and after timescale decompositions, and the results are compared.

In chapter 3, we use a bivariate non-Gaussian time series model to describe RSV. We use a negative binomial model to investigate the relationship between RSV and HMPV while adjusting for climatic factors. We specifically aim at establishing the heterogeneity in the autoregressive effect to account for the influence between these viruses. Our findings show that RSV contributes to the severity of HMPV. This is achieved through comparisons of 12 models of various structures, including those with and those without interactions between the climatic factors.

In chapter 4, we model the multivariate associations of three time series of RSV, HMPV and Influenza and predict their outbreak detections. We describe an approach to analyse multivariate time series of disease counts. A previously used model in the literature to address dependence between two different disease pathogens is extended to investigate for serial correlation. The models evaluate for the presence of heterogeneity in the autoregressive effect

for the different pathogens and whether after adjusting for seasonality, an epidemic component could be isolated within or between the pathogens.

Finally, we provide the overall key findings from the studies presented in this thesis in a discussion in chapter 5. We also discuss the implication of the findings and conclude with the recommendations for further studies.

CHAPTER 2: Effects of Climate on Incidence of Respiratory Syncytial Virus Infections in a Refugee Camp in Kenya: A non-Gaussian Time-series Analysis

2.1 Abstract

Respiratory syncytial virus (RSV) is one of the major causes of acute lower respiratory tract infections (ALRTI) in children. Children younger than 1 year are the most susceptible to RSV infection. RSV infections occur seasonally in temperate climate regions. Based on RSV surveillance and climatic data, we developed statistical models that were assessed and compared to predict the relationship between weather and RSV incidence among refugee children younger than 5 years in Dadaab refugee camp in Kenya. Most time-series analyses rely on the assumption of Gaussian-distributed data. However, surveillance data often do not have a Gaussian distribution. We used a generalised linear model (GLM) with a sinusoidal component over time to account for seasonal variation and extended it to a generalised additive model (GAM) with smoothing cubic splines. Climatic factors were included as covariates in the models before and after timescale decompositions, and the results were compared. Models with decomposed covariates fit RSV incidence data better than those without. The Poisson GAM with decomposed covariates of climatic factors fit the data well and had a higher explanatory and predictive power than GLM. The best model predicted the relationship between atmospheric conditions and RSV infection incidence among children younger than 5 years. This knowledge helps public health officials to prepare for, and respond more effectively to increasing RSV incidence in low-resource regions or communities.

Keywords: *Respiratory syncytial virus, time series, seasonal, climate, modeling.*

2.2 Introduction

Respiratory syncytial virus (RSV) is one of the major causes of acute lower respiratory tract infections (ALTRI) in infants and young children (1)(2). The RSV infections occur seasonally in temperate climate regions (3). The RSV adversely impacts the health of adults and immunocompromised patients, and is associated with significant mortality and morbidity, particularly in young children and vulnerable infants (4). Children younger than 1 year are most susceptible to RSV infection; often 60-70% of children in this age group have been infected at least once, and re-infection can occur throughout their lifetime (4)(5)(6).

The RSV is shed in saliva and nasopharyngeal secretions (7). Infected hosts shed higher quantities of viral particles upon exposure to higher-ambient temperatures (8). Low humidity during winter enhances RSV viability, and enables its survival for up to 12 hours on nonporous surfaces (9). In dry air conditions, large droplets evaporate and remain air-borne for longer periods of time. Some studies have shown that airborne transmission appears to be sensitive to ambient humidity and temperature in temperate regions (8)(10). The RSV outbreaks show some seasonality that suggests a connection with atmospheric and environmental conditions (11)(12). Most RSV infections in temperate locations occur between November and April (13). The RSV infection has been associated with winter in these regions because people spend more time indoors, potentially in crowded conditions (14). Such climatic regions are different from those of Kenya, which is located on the equator and experiences bimodal seasonal rainfall due to the interaction of the Northern and Southern Hemisphere monsoon systems (15). However, variations in climatic factors, such as humidity, temperature, wind speed, rainfall etc., can have a significant impact on disease dynamics.

Therefore, it is essential that the RSV incidence be evaluated for equatorial climatic regions to aid accurate predictions of RSV outbreaks. (16)(17).

The wide range of statistical methods used to explore the link between RSV outbreaks and climate makes it difficult to elucidate a definitive relationship. Pearson correlation analysis was previously used to explain the associations of RSV-positive cases with meteorological variables (11). The univariate analysis of variance (ANOVA), multiple regression analysis, and Spearman's rank correlation were used to assess the association between RSV incidence and meteorological parameters (18). A better understanding of the relationship between climate and RSV helps in making reliable predictions of its incidence.

Worldwide, as of 2005, 99% of deaths from RSV were reported by the World Health Organization (WHO) to occur in developing countries (19). It is, therefore, crucial to establish good RSV surveillance systems in developing countries to help understand the dynamics of the disease. In 2006, the U.S. Centers for Disease Control and Prevention (CDC) and the Kenya Medical Research Institute (KEMRI) established a respiratory illness surveillance system to detect disease outbreaks in Kenyan refugee camps (20). We used RSV incidence data from this system to explore the best model that predicts the relationship between RSV incidence and climatic factors along spatio-temporal scales to determine whether a seasonal pattern of RSV infection exists. A generalised linear model (GLM) with a sinusoidal component over time was used to account for seasonal variation and compared with a generalised additive model (GAM) with smoothing cubic splines. Climatic factors were included as covariates in the models before and after timescale decompositions.

2.3 Methods

2.3.1 Data

Surveillance for viral respiratory illnesses, including adenovirus, human metapneumovirus, influenza virus, parainfluenza viruses 1, 2 and 3, and RSV was implemented in Dadaab refugee camp in north eastern Kenya in 2007. Paediatric and adult patients who presented at a camp medical unit, and met the case definition for influenza-like illness (ILI) or severe acute respiratory infection (SARI), were enrolled into the laboratory-enhanced respiratory surveillance system and tested for all of the above diseases after an informed consent form was completed by adults, older minors, and guardians of all minors <15 years (20). The number of laboratory-confirmed cases was recorded on a daily basis from September 2007 to August 2011. The monthly counts of all RSV cases among children younger than 5 years were included in the present analysis; the main outcome of interest being monthly RSV incidence rate in this age group. The RSV incidence rate per 1,000 children younger than 5 years was calculated by dividing monthly RSV counts by the monthly population of children younger than age 5 years in the camp. Local weather and climatic data, including: the mean temperature and mean dew point for the day (both in ^{0}F); mean sea level pressure for the day in millibars; mean visibility for the day in miles; mean wind speed for the day in knots; minimum and maximum temperature (^{0}F) reported during the day; and the total precipitation (in inches) reported during the day were obtained from the World Meteorological Organization's (WMO's), World Weather Watch Program, according to WMO Resolution 40 (Cg-XII) (available at http://www7.ncdc.noaa.gov/CDO/cdo). The meteorological dataset consisted of measurements recorded at successive, equally spaced time points (covariates

used in the present study are provided in Table 2.1). Data used in the analysis are available upon request from the authors.

Table 2.1 Covariates and their description. Non-decomposed and decomposed covariates into the seasonal (S), trend (T), and random (R) components.

Covariate	Description
x_{t1}	Wind speed
x_{t2}	amount of Rainfall
x_{t3}	Temperatures
x_{t4}	mean Dew point
x_{t5}	Visibility
x_{t1S}	Seasonal, wind
x_{t1T}	Trend, wind
ξ_{t1R}	Random, wind
x_{t2S}	Seasonal, rainfall
x_{t2T}	Trend, rainfall
ξ_{t2R}	Random, rainfall
x_{t3S}	Seasonal, temperature
x_{t3T}	Trend, temperature
ξ_{t3R}	Random, temperature
x_{t4S}	Seasonal, dew
x_{t4T}	Trend, dew
ξ_{t4R}	Random, dew

Covariate	Description
x_{t5S}	Seasonal, visibility
x_{t5T}	Trend, visibility
ξ_{t5R}	Random, visibility

234

235 ***2.3.2 Generalised Linear Models and Generalised Additive Models***
236

237 A Poisson distribution model was used in this analysis, as the outcome of interest (incident
238 RSV cases) was non-Gaussian count data. Some authors have used Gaussian vector
239 autoregressive models on multivariate counts that are serially correlated. Brandt and others
240 used vector autoregressive methods that were based on Gaussian error process (21). However,
241 such an assumption is not applicable to event count data because it produces biased estimates
242 (22). So, as many of those methods apply for count series that approximate normality, they
243 may not hold to dynamic events like the ones applied here. In the first model, seasonal effects
244 on RSV incidence were analysed by using a generalised linear model (GLM) with a
245 sinusoidal component to account for seasonal variation. The second model extended the GLM
246 model to a generalised additive model (GAM) by applying smoothing cubic splines. The
247 GAM is an extension of the GLM and is adaptable to non-normally distributed variables (23).
248 GLM uses linear predictors specified as the expected value of a response variable (Y_j), which
249 is expressed as $\eta = \Sigma_j \beta_j(X_j)$. Here, β_j is a coefficient parameter and X_j represents the j-th
250 explanatory variable. The GAMs extend these by replacing them with $\eta = \Sigma_j f_j(X_j)$, where
251 $f_j(X_j)$ are unspecified nonparametric functions estimated by including smoothing splines
252 (24). GAMs allow for adjustments of the nonparametric, nonlinear, confounding effects of
253 seasonality, trends, and weather variables, which have been previously used in modeling time-

254 series data (25). In the present analysis, climatic time-series covariates were included in the
255 GLM and GAM models and implemented in R language v3.1.0 (26). Both models were
256 optimized for predictive accuracy and precision.

257 Data were decomposed into three components. namely, trend, seasonal and random
258 components, in order to independently evaluate the existence and strength of associations
259 between RSV incidence and covariates on each time scale. Data decomposition was
260 accomplished using Loess smoothing, a regression method that assigns a weighted
261 polynomial to each component (27). The idea is that the time series is decomposed into trend,
262 seasonal and remainder components. $Y_\upsilon = T_\upsilon + S_\upsilon + R_\upsilon$ for $\upsilon = 1$ to N data points. The
263 seasonal-trend decomposition approach uses the Loess (LOcal regrESSion) smoothing. For y_i
264 and x_i measurements, a smooth estimate g(x) is provided by Loess for y at all values of x. A
265 positive integer q is chosen to calculate g where a larger q yields greater smoothing. Closer q
266 values of x_i to x are selected and each is weighted by how far it is from x. The weight given as
267 $\nu_i = W(|x_i - x|)$ where W is the tricube weight function and $\lambda_g(x)$ is the distance from the
268 q[th] farthest point (for $q < N$. if $q \geq N$, additional scale terms must be used). For selected
269 (x_i, y_i) and with weights ν_i, a polynomial of degree d is fit. Some data points are considered
270 more heavily in the regression depending on the weights allowed. We introduced a GLM for
271 time-series data, with a sinusoidal component over time to account for seasonal variations.
272 The GLM was extended to include a smoothing function using the GAM approach to the
273 Poisson distribution (28) In each model, a data-driven smoothing function of time was fitted,
274 and compared with those fitted, using sine and cosine functions in the Fourier basis.

275 The observed number of RSV counts, Y_t at a given month $t = 1, \cdots, n$ from the population at
276 risk is assumed to follow a Poisson random variable: $Y_t \sim \text{Poisson}(\mu_t)$. We let n_t be the
277 population of children age 5 years and younger who are at risk of RSV in the camp. The

expected value (mean) of Y_t is $E(Y_t) = \mu_t = n_t\Theta_t$ where the dependence of covariates on Θ_t is modeled by $\Theta_t = e^{x_t^T\beta}$. Therefore, a Poisson GLM of the form $E(Y_t) = \mu_t = n_t e^{x_t^T\beta}$ is used. More explicitly, to model the incidence of RSV, we use:

$$\log \mu_t = \beta_0 + \alpha y_{t-1} + \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q}\beta_{ksl}x_{(t-l)ks} + \eta_1 cos\left[\frac{2\pi t}{T}\right] + \eta_2 sin\left[\frac{2\pi t}{T}\right] + \log n_t\,, \tag{2.1}$$

where μ_t is the infection rate for the month, t. β_0 is the intercept, α is the coefficient of the lagged RSV counts by one month, which is represented by y_{t-1}, $x_{(t-l)ks}$, is the decomposed measured covariate, β_{ksl} their corresponding coefficients with $k = 1, \cdots, m$ covariates and $s = 1, \cdots, r$ corresponding to r-th decomposition of the k-th covariate. This additive time-scale decomposition of the k-th covariate into the seasonal (S), trend (T), and random (R) components is;

$$\beta_{ksl}x_{(t-l)ks} = \beta_{ksl}x_{(t-l)kS} + \beta_{kTl}x_{(t-l)kT} + \xi_{(t-l)kR,} \tag{2.2}$$

for every k in $\{1, \cdots, m\}$. In the above case, s takes on three levels S, T, and R. This decomposition helps in assessing for the significance of the seasonal and trend components of the covariates in explaining the RSV incidence. The combination of the seasonal and trend components makes up the patterns in the covariates. The $l = 0,1, \cdots, q$ distributed lags where q is the maximum lag and $t = 1, \cdots, n$ are the time points. The terms η_1 and η_2 are coefficients of the sine and cosine function, respectively. Here, T is the number of time periods described by one cosine function over the interval $[0,2\pi]$.

Using a cosine function, we specified two periods: one that defines the measure of RSV infection (month) and the other that is described by one cosine cycle. After fitting all covariates in the GLM model, the most parsimonious model was identified. The maximum

lag for each covariate was obtained by comparing different lagged models using Akaike information criterion (AIC). The maximum lag for each covariate was used to run "crossbasis" in the "dlnm" package for time-series models (29)(30). The same covariates were used to fit the GAM model.

The corresponding GAM for the Poisson model is:

$$\log \mu_t = \beta_0 + \alpha y_{t-1} + \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q} \beta_{ksl}\Psi_k\left(x_{(t-l)ks}, \lambda_{(t-l)ks}\right) + \Psi_{k+1}(t, \lambda_{k+1}) + \log n_t, \quad (2.3)$$

where $\lambda_{(t-l)ks}$ is the smoothing parameter or the degrees of freedom for covariates, λ_{k+1} is a smoothing parameter for time and $\Psi.$ is the smoothing function. Larger values of $\lambda.$ are indicative of a less-smooth function.

The trend cycles represent long-term changes in the levels or values of the covariate, while the periodic changes are the fluctuations of constant length. The GLM model (2.1) has the Logit link function. The residual deviance for these models takes on the form $D = -2\log(L_{test}/L_{sat})$, where L_{test} and L_{sat} are the maximized likelihoods under the test and saturated models, respectively. The model selection and fitting was done using the "glmulti" package (31) and "gam" (32) in "mgcv" package (24) in the R language v3.1.0.

2.3.3 Ethical Considerations

Ethical approval for the surveillance activities was obtained from the KEMRI Ethical Review Committee (SSC Protocol Number 1161). Institutional review was waived by CDC because

the study was considered to be a non-research public health activity. Informed written consent was obtained from all participants and from the guardians of minors.

2.4 Results and discussions

2.4.1 Data exploration

A peak in RSV incidence occurred every 11-12 months, particularly from October to January (Figure 2.1). Other than these peaks, there was relatively low RSV incidence (≤ 20 cases per 1000 person months), the RSV incidence rate per 1,000 children younger than 5 years.

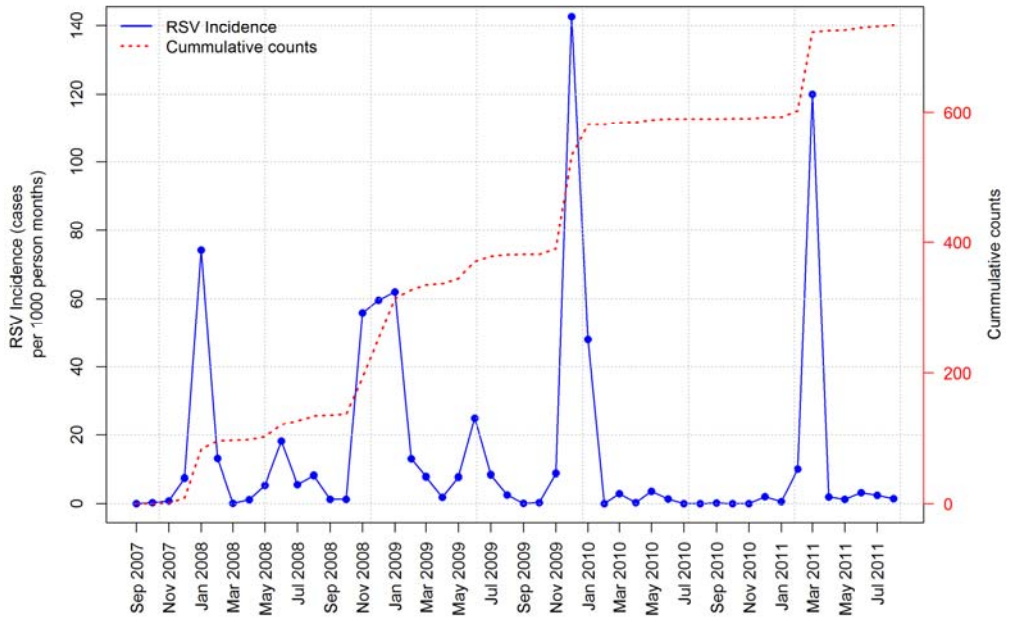


Figure 2.1 Plot of RSV incidence in Dadaab. Fluctuations in the data are roughly constant over time, indicating that the RSV time series could likely be described using an additive model.

The decomposed data, seasonal pattern, trend line, and random component of the RSV, wind, rainfall, and temperature time series are shown in Figure 2.6 - 2.9. The seasonal pattern of RSV incidence regularly repeated itself, with two distinct peaks annually (Figure 2.6). The magnitude of the seasonal components of the decomposed covariates did not vary annually (Figure 2.6 – 2.11). This justifies the use of additive, rather than multiplicative decomposition. There was a positive correlation between temperature and RSV incidence (Figure 2.2B). There was a significant moderate correlation between RSV incidence and wind speed ($\rho = -0.4651, p = 0.001$) (Figure 2.2A); an insignificant weak correlation between RSV incidence and temperature ($\rho = 0.1850, p = 0.224$) (Figure 2.2B); an insignificant weak correlation between RSV incidence and dew point ($\rho = 0.230, p = 0.128$) (Figure 2.2C); and for temperature and wind speed (Figure 2.2D). We fitted a parabolic curve using: $x_3 = \mathfrak{r}_0 + \mathfrak{r}_1(x_1 - \mathfrak{r}_2)^2$ where $\mathfrak{r}_{0,1,2}$ are constants, and the regression fit was significant ($p = 0.002$). Here, x_1 and x_3 represent wind speed and temperature, respectively. Since the relationship between the two is quadratic, there is no problem of multi-collinearity between them.

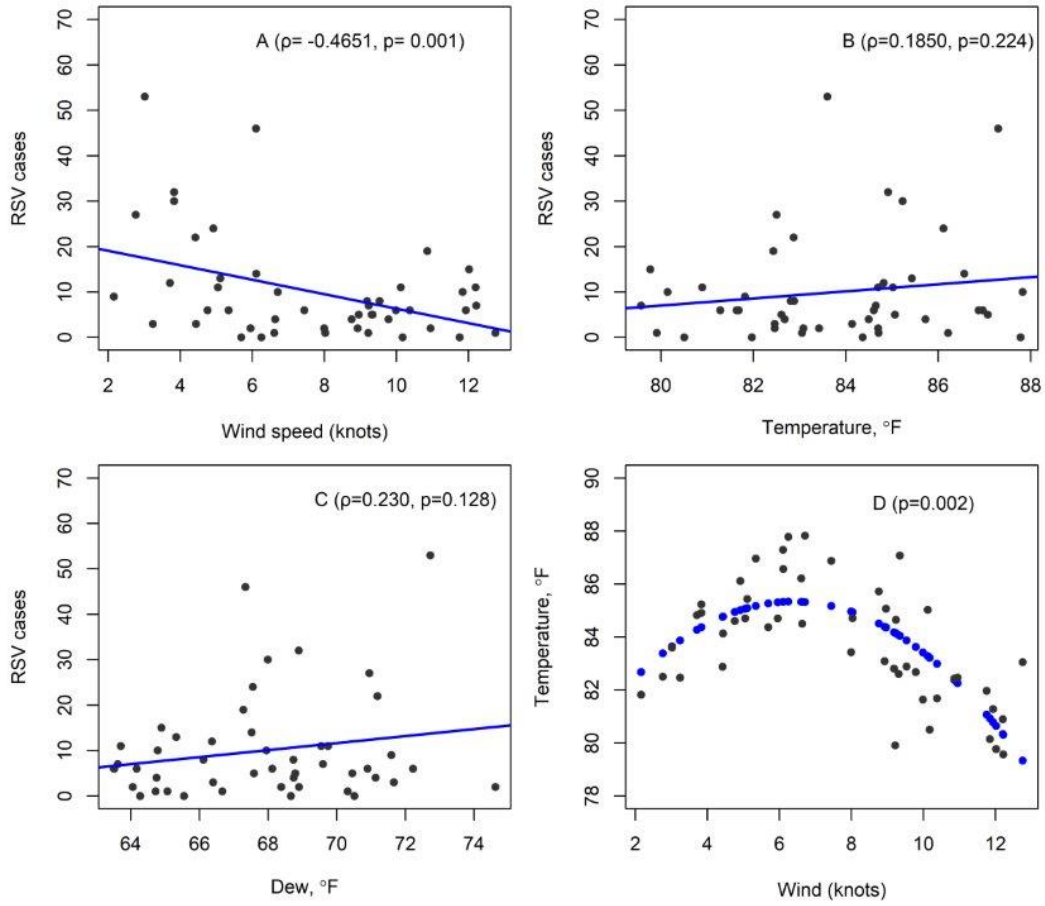


Figure 2.2 Correlation-regression analysis. A: Correlation between RSV incidence and wind speed; B: Correlation between RSV incidence and temperature; C: Correlation between incidence and dew point; and D: Correlation between temperature and wind speed. In these plots, the regression lines of best fit are indicated by bold blue lines between RSV incidence and dew point; and D: Correlation between temperature and wind speed. In these plots, the regression lines of best fit are indicated by bold blue lines.

2.4.2 Model assessment and comparison

The trend component of the wind decomposition model decreased slightly immediately after 2008, then increased steadily to a peak in early 2009, followed by a decrease to a minimum value in late 2010 (Figure 2.7). To determine the best predictive model, we compared the performance of the four models described in the methods chapter. The best GLMs and GAMs from the Poisson were compared using the AIC and residual deviances (Table 2.2).

Table 2.2 Model diagnostic and performance results

Model	**Deviance explained (%)**	**AIC**
Poisson GLM[a] $\log \mu_t = \beta_0 + \alpha y_{t-1} + \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q}\beta_{ksl}x_{(t-l)ks} + \eta_1 cos\left[\frac{2\pi t}{T}\right] + \eta_2 sin\left[\frac{2\pi t}{T}\right] + \log n_t$	34.3	446.86
Poisson GLM[b] $\log \mu_t = \beta_0 + \alpha y_{t-1} + \eta_1 cos\left[\frac{2\pi t}{T}\right] + \eta_2 sin\left[\frac{2\pi t}{T}\right] + \log n_t$	29.4	477.25
Poisson GAM[a] $\log \mu_t = \beta_0 + \alpha y_{t-1} + \sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{s=1}^{r}\sum_{l=0}^{q}\beta_{ksl}\Psi_k\,(x_{(t-l)ks}, \lambda_{(t-l)ks}) + \Psi_{k+1}(t, \lambda_{k+1}) + \log n_t$	65.3	317.17
Poisson GAM[b] $\log \mu_t = \beta_0 + \alpha y_{t-1} + \Psi_{k+1}(t, \lambda_{k+1}) + \log n_t$	59.5	346.44

The superscripts a,b indicate models with and without covariate decomposition, respectively.

The AIC was used to judge the best model from the set of models that had a good fit. The best models all had covariates with $p < 0.05$. This was the case for models with and without decomposed covariates. Of all the models that were evaluated, the Poisson GAM with decomposed covariates had the best fit to the data (AIC = 317.17 and a Deviance explained = 65.3%, Table 2.2). Figure 2.3 shows the best model fit to the RSV incidence data with decomposed covariates comparing the Poisson GLM and the Poisson GAM, where the Poisson GAM fits the data well.

Table 2.3 ANOVA model for the best performing model, the Poisson GAM with covariate decomposition.

Variable	lag	df	F-value	p-value
Seasonal, wind speed	1	4	28.81	<0.0001
Trend, wind speed	3	4	17.99	0.0012
Seasonal, rainfall	0	4	27.70	<0.0001
Trend, mean dew point	2	4	45.59	<0.0001
Trend, visibility	2	4	68.18	<0.0001
Month		3	2.48	0.4784

The best model in its reduced form is the Poisson GAM[a] (Table 2.4). Table 2.3 contains the corresponding ANOVA results for the Poisson GAM[a]. From this table, the wind with both the trend and seasonal effects (seasonal effect of rainfall, trend mean dew point, and the trend effect of visibility) significantly explained RSV incidence. We note that time in months did not significantly explain RSV incidence, further demonstrating the importance of using climactic factors to explain the seasonality of RSV.

366 The direction of effects demonstrated nonlinear relationships with RSV incidence, except in
367 the case of seasonal wind speed, which had a linear relationship (Figure 2.4). High wind
368 speed within the same month had a significant negative effect on the RSV incidence. The
369 trend component of the wind speed in the two months preceding incident RSV cases had a
370 nonlinear relationship with RSV incidence. As the wind speed increased, incidence fluctuated
371 from low to high, returning to low incidence when the speeds were highest. An increase in the
372 seasonal component of rainfall in the four months preceding RSV cases was associated with
373 an increase in RSV incidence. When rainfall was at its lowest, RSV incidence increased then
374 returned to baseline when rainfall reached its maximum. The trend effect of the mean dew
375 point 1 month preceding incident cases was associated with an increase in RSV incidence
376 until dew point reached its maximum. The increase in visibility trend component 2 months
377 preceding incident RSV cases demonstrated a constant effect on RSV incidence, which
378 peaked when the visibility was 19.5 miles and troughed when the visibility was at its highest.

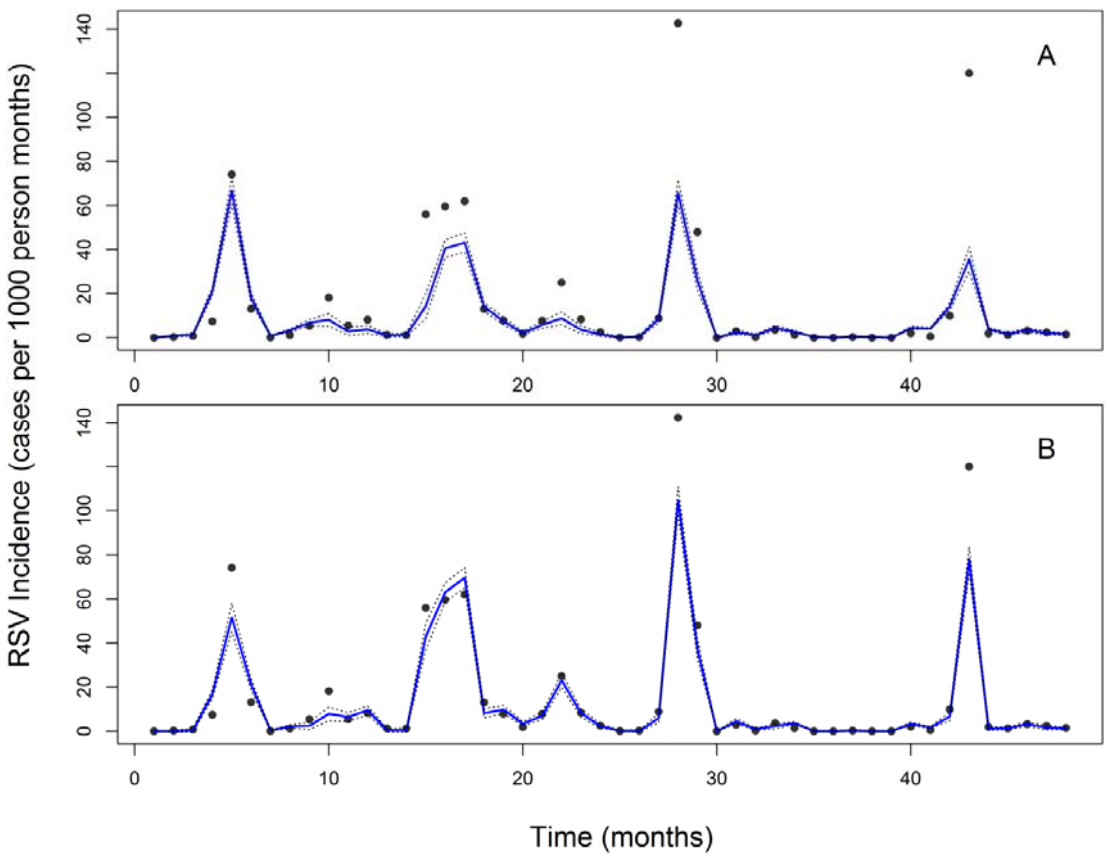


Figure 2.3 Best model fit to the RSV incidence data (bold lines) with decomposed covariates. A: Poisson, GLM. B: Poisson, GAM. The standard error bars to the model fit are indicated by the dotted lines (95% confidence bounds). The base year in all these plots was September 2009.

379

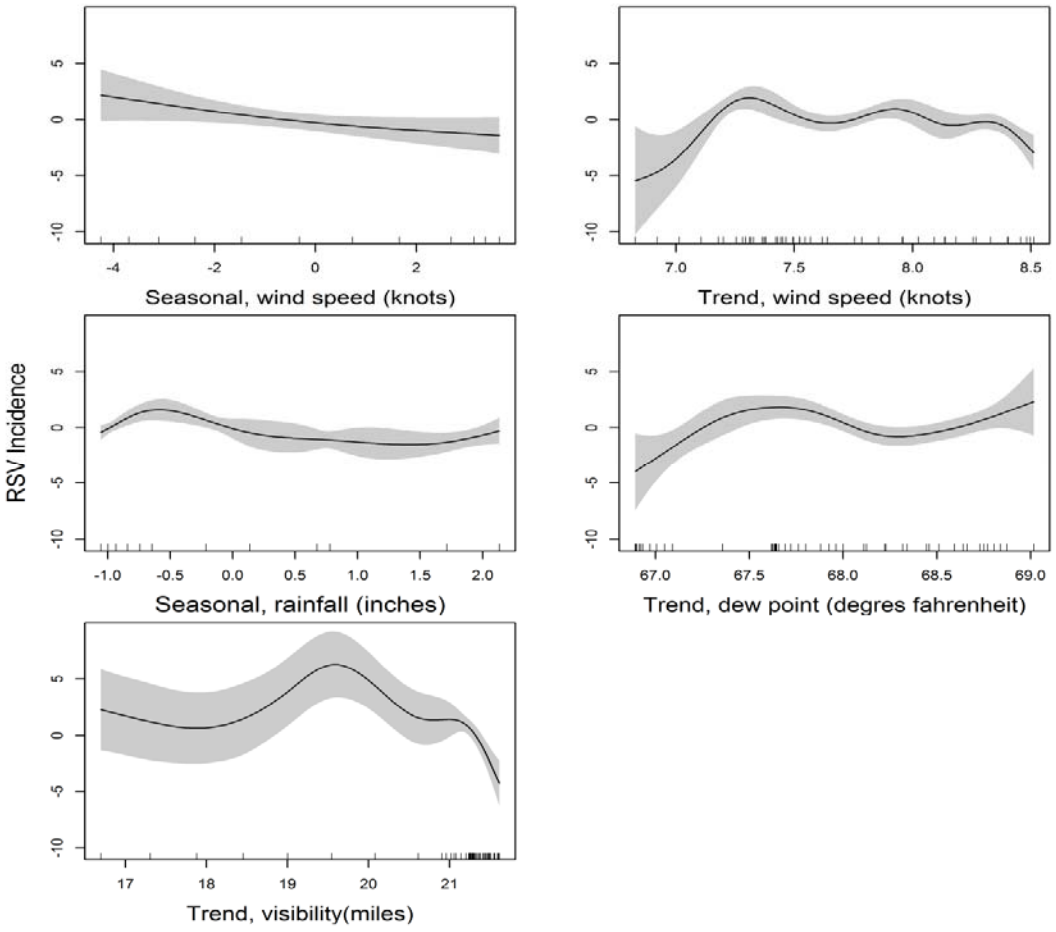


Figure 2.4 Best model fit (Poisson, GAM) to the RSV incidence data with the signifficant decomposed covariates. Seasonal wind speed, Trend wind speed, Seasonal rainfall, Trend rainfall and Trend visibility. The standard error bars to the model fit are indicated by the grey shade (95% confidence bounds). RSV incidence units as cases per 1,000 person months.

380

381

The best model residual analysis (Figure 2.5) reveals that there was a slight improvement to the fit when the decomposed covariates were included into the model. Due to having less data points and many parameters there was an over-fit to the data in general.

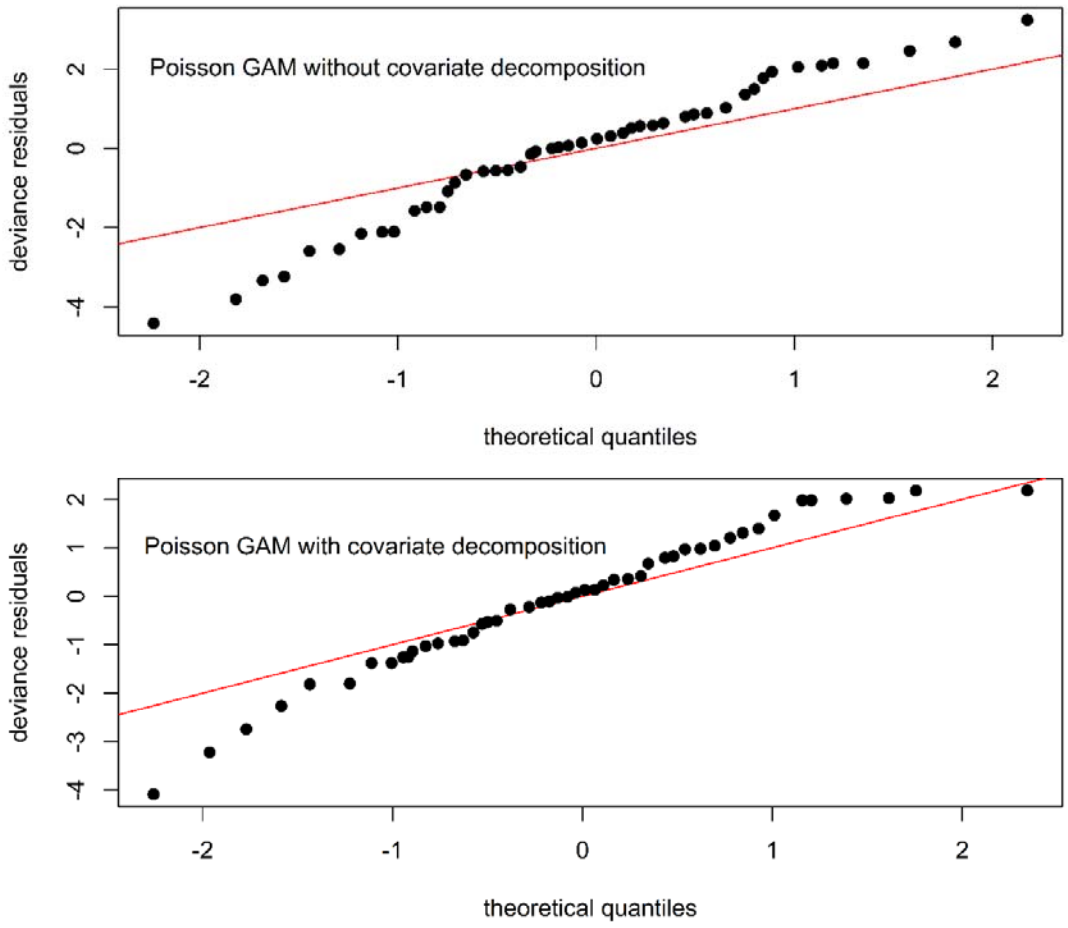


Figure 2.5 Residual plots to the best GAM models (Poisson, GAM without covariate decomposition and Poisson, GAM with covariate decomposition) to the RSV incidence data.

2.4.3 Implication of results and comparison to related studies

Our data showed seasonal variations for RSV incidence (Figure 2.6). The Poisson GAM with decomposed covariates out-performed the GLM variant, thereby relaxing its linearity.

Generally, the role of climatic factors in determining disease dynamics is rather complex to decipher (33). In the literature, there is strong evidence that the relationship between climatic factors and RSV incidence varies widely between geographical regions (18). Previous studies have shown that climatic factors might be associated with RSV, although it remains unclear what these factors are or exactly how they impact RSV incidence. We performed a correlation analysis for each covariate with RSV by fitting regression lines to test the level of significance between the climatic variables (Figures 2.2A-2.2C). A recent study by Agoti *et al* (34) on RSV strains using the same RSV surveillance data showed that there were six epidemic peaks within the three year study period: two peaks each year; the first and the last peaks were composed of group B strains and the other four peaks were composed of group A strains. Agoti's study, in conjunction with our findings, show that onset of RSV infections in Kenya can be reliably predicted. Our findings, in comparison with other studies, also suggest that the relationship between RSV incidence and climatic factors varies widely; for instance, from 2004 to 2012 in tropical and sub-tropical zones such as Hong Kong, China, Singapore, Kuala Lumpur, Malaysia, Medellin and Colombia outbreaks occurred primarily during the hot and rainy seasons (14).

The ability to predict increases in RSV incidence, based on prevailing meteorological conditions, could potentially inform the application of public health interventions and provisions of healthcare in Kenya, and perhaps, in other regions with a similar climate and equatorial location. Currently, there is no RSV vaccine available; however, in developed countries, infants at risk of severe outcomes can be administered monthly doses of the anti-RSV antibody, palivizumab, during outbreaks of RSV (3)(8). Because predicting the incidence of RSV could optimize the cost-effectiveness of immunoprophylaxis; our model might be useful to apply in a cost-benefit analysis of this approach in Kenya. In most

417 temperate climate regions, RSV occurs as an annual epidemic. For instance, Noyola and
418 Mandeville found that temperature was the predominant atmospheric condition explaining the
419 annual spread and variability of RSV incidence in San Luis Potosi, Mexico (35). Using
420 correlation and regression analysis, Noyola and colleagues (35) observed that the weekly
421 number of RSV incidence between October 2002 and May 2006 was correlated to ambient
422 temperature, barometric pressure, relative humidity, vapor tension, dew point, precipitation,
423 and hours. Our findings corroborate what they observed for the same climatic factors. The
424 modeling has aided identification of factors influencing RSV incidence and provided
425 indicators for devising measures to prevent the spread of the disease.

426 Our analysis showed that other climatic factors affecting RSV seasonality can improve the
427 performance of a predictive model. Khor *et al* (18) demonstrated that, in Malaysia, ambient
428 temperature was inversely associated with RSV activity, even though the highest number of
429 cases may not always coincide with the lowest temperature. A negative correlation between
430 the mean minimum temperatures and RSV incidence was recently reported in Italy (11). The
431 RSV transmission that occurs during cold weather is facilitated by its stability in secretions,
432 since inhalation of cold air slows down the mucociliary escalator. This reduces phagocytic
433 activity of leukocytes, increasing the host's vulnerability to infection. There is evidence of
434 RSV epidemics occurring in tropical areas with high temperatures during rainy seasons, a
435 phenomenon that our data are exhibiting (36)(37). However, the exact mechanisms of how
436 climatic factors affect RSV incidence requires further investigations, especially across
437 geographically diverse regions. The relationship between the dynamics in wind speed and
438 direction, and how these dynamics influence the climate of geographical regions like Dadaab,
439 remains unclear. Understanding such complex relationships between the co-factors explaining
440 the spread of RSV is essential to predict its incidence.

A foreseeable limitation of our models is that with log- or logit-links; the mean value zero corresponds to an infinite range on a linear predictor scale. For count data with a relatively large number of zeros clustered closely within the covariate space, GAMs might suffer from identifiability problems, especially the Poisson family. For the over-dispersion parameter, the assumption of equal mean and variance inherent in the Poisson GAM might be violated; hence, it has to be replaced by variances that exceed the mean. Our data show a cyclic and seasonal behaviour for RSV incidence among children (Figure 2.1). The Poisson GAM from this analysis demonstrated that climatic factors, including wind speed, rainfall, dew point and visibility, significantly affected RSV incidence. The use of atmospheric condition data can help public health officials predict increases in RSV infection incidence among children and help them prepare and respond more swiftly to increasing RSV incidence in low-resource regions or communities. While specific vaccines, antiviral medications and immunoglobulins are not available to control RSV in these settings, agencies responsible for managing healthcare in crisis-affected populations can increase preparedness for RSV outbreaks by establishing additional patient-isolation areas and bed space, ensuring that all healthcare workers are provided with adequate personal protective equipment (e.g., facial masks and gloves) and appropriate amounts of hand sanitizers and adequate hand-washing facilities for healthcare workers are readily available.

Health education is important; crisis-affected populations should be made aware of the symptoms and signs of RSV, how it spreads, and how to protect themselves and their loved ones. Health education should focus on how to cover coughs, keep appropriate social distancing (e.g., not being too close to others, not shaking hands), and the importance of washing hands with soap. In particular, our model indicates that when the wind speed in knots change from high to low, these interventions should be enhanced to prevent spread of RSV

infections in Kenya. In the future, these models could be validated with new RSV surveillance data to see how well they perform to predict increases in RSV incidence particularly for geographical regions with similar climatic attributes to Dadaab.

2.5 Acknowledgements

The authors wish to acknowledge the CDC Kenya Refugee Health Program for their tireless work in the surveillance in the refugee camp and their assistance with the data collection and management. We also thank Nina Marano and Rachael Joseph for their assistance with the study design and data collection.

489 **2.6 Supporting Information**

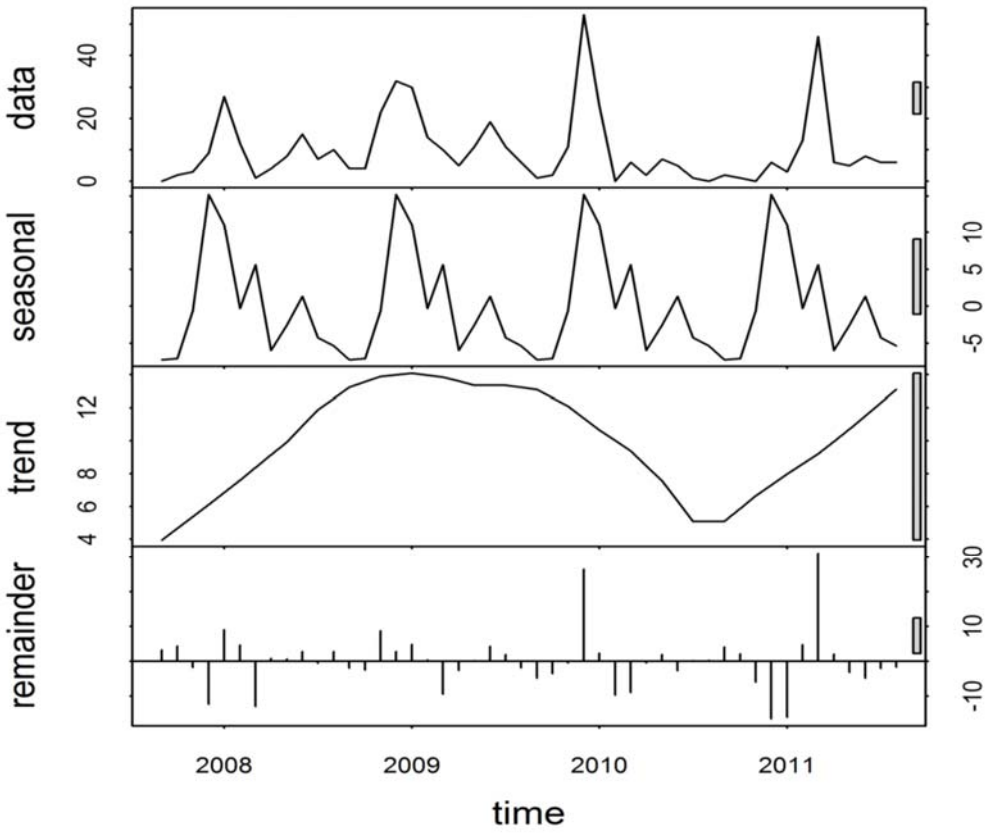


Figure 2.6 Decomposition of RSV time-series data. The variation in the remainder component is approximately the same as the variation in the data. The variations in the seasonal and trend components are about 3 to 4 times smaller than that observed in the data. The long-term trend components appear to be generally increasing. The random (remainder), the bottom plot, show the residual variation in the data after the long-term trend and seasonality are removed.

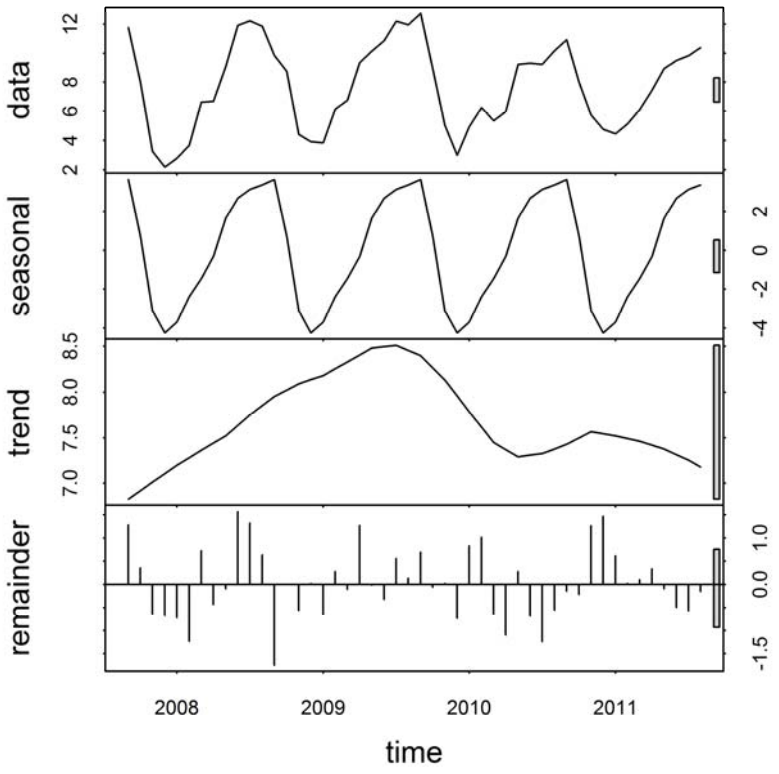


Figure 2.7 Decomposition of wind time-series data. The variation in the trend is much smaller than that in the data. The variations in the seasonal and remainder components are marginally smaller than the variation in the data (grey bars on the right).

490

491

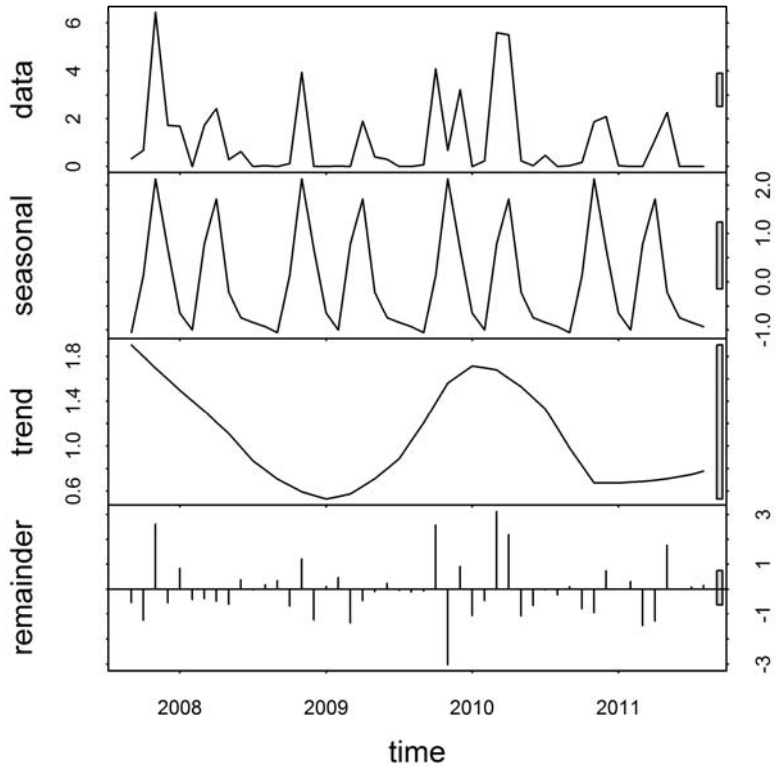


Figure 2.8 Decomposition of rainfall time-series data. The variations in the seasonal and remainder components do not deviate much from that in the data. The variation in the trend component is roughly 4 times less than the variation in the data.

492

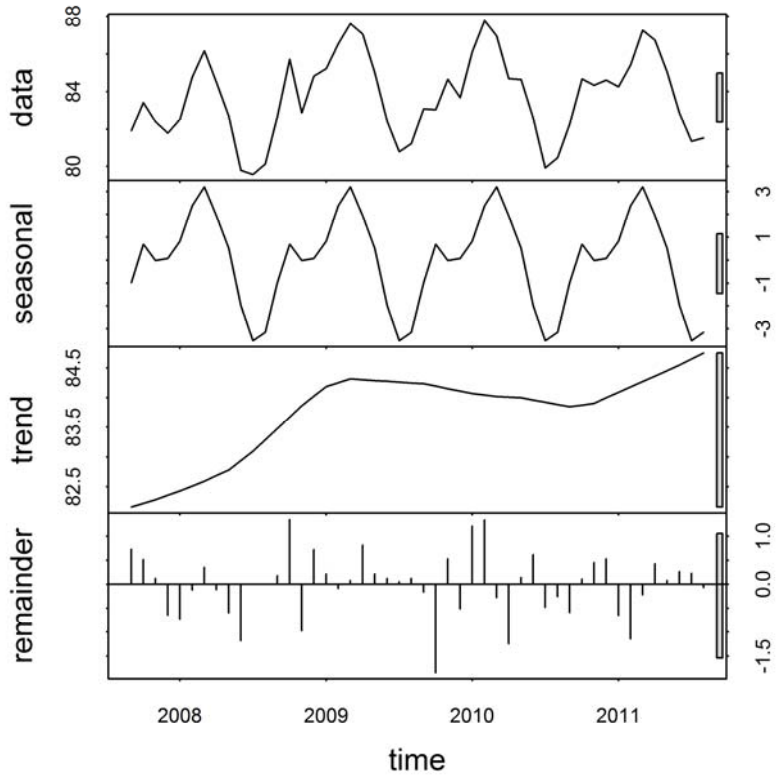


Figure 2.9 Decomposition of temperature time-series data. The trend component has a much smaller variation than that in the data. The seasonal and remainder components show marginally smaller variation than that observed in the data. The long-term trend components appear to be generally increasing.

493

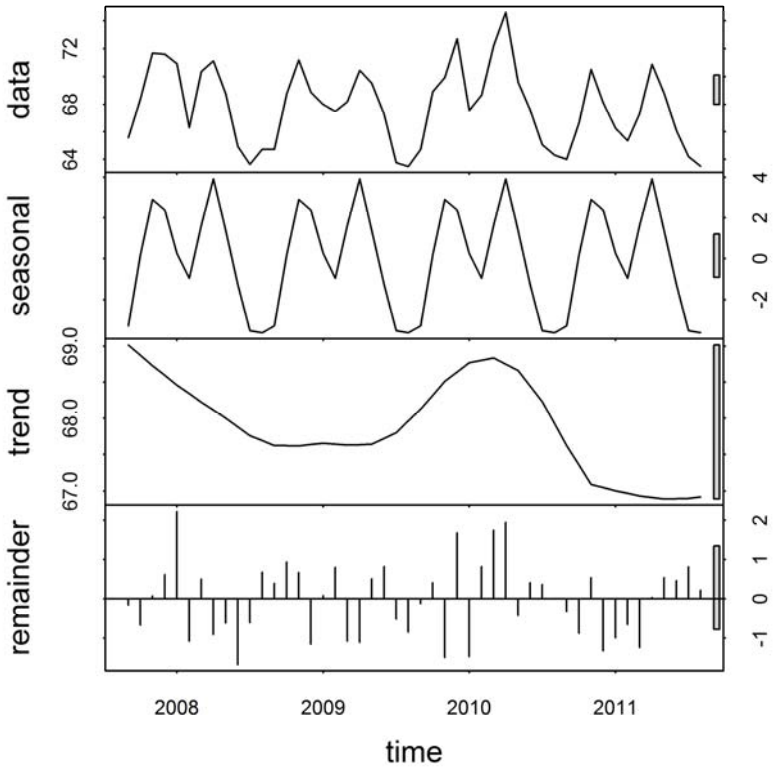


Figure 2.10 Decomposition of Dew time-series data. The trend exhibits approximately 3 times the overall variation in the Dew data (large grey bar relative to the grey bar on the right-hand of the data plot). The long-term trend components appear to be generally increasing.

494

495

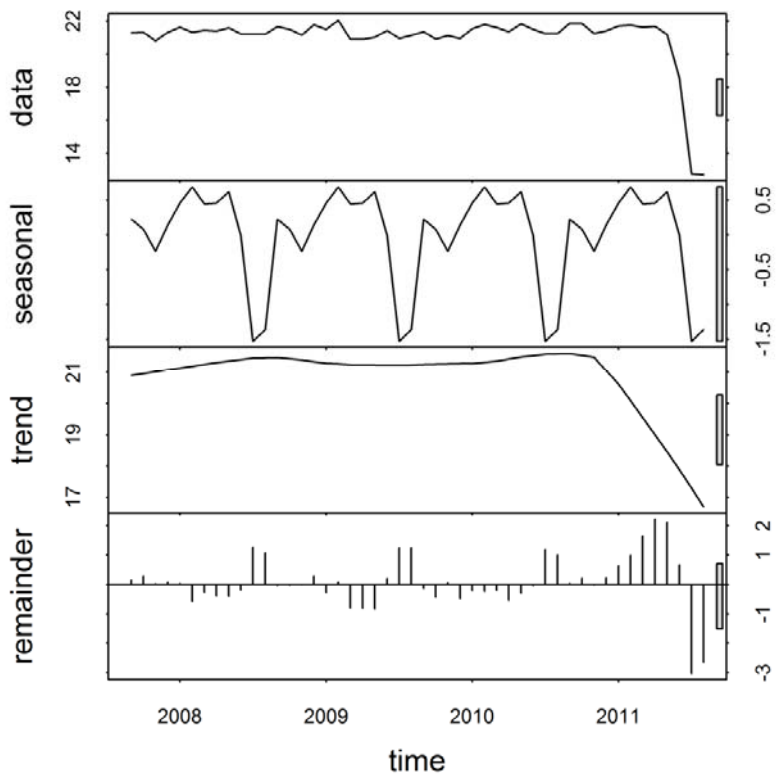


Figure 2.11 Decomposition of visibility time-series data. The season accounts for a very small portion of the overall variation in the visibility value (large grey bar relative to the grey bar on the right-hand of the data plot). The long-term trend components seem to be generally decreasing.

507 **Table 2.1 Selected Poisson candidate models.**

Covariates	**GLM[a]**	**GLM[b]**	**GAM[a]**	**GAM[b]**
	Est(se)	**Est(se)**	**Est(se)**	**Est(se)**
Intercept	-0.69(0.074)***	-0.71(0.070)***	-2.71(4.111)	-2.09(0.637)**
y_{t-1}	0.004(0.004)	0.02(0.004) ***	-0.0006(0.007)	0.01 (0.006).
x_{t1}		-0.82(0.441).		
x_{t2}		-1.21(0.427) **		
x_{t3}		-2.43(0.408) ***		
x_{t4}		1.33(0.454) **		
x_{t1S}	-6.45(0.917) ***			
x_{t1T}	1.69(0.286) ***			
x_{t2S}	-1.12(0.246) ***			
x_{t4T}	0.630(0.254) *			
x_{t5T}	-1.22(0.423) **			
Cos(2πt/12)	1.27(0. 277)***	-0.70(0.182)***		
Sin(2πt/12)	-1.20(0.210) ***			
ns(x_{t1},4)1				-5.45(0.792)***
ns(x_{t1},4)2				-1.67 (0.784)*
ns(x_{t1},4)3				-0.32 (1.287)
ns(x_{t1},4)4				-1.78(0.723)*
ns(x_{t2},4)1				-1.03(0.337)**
ns(x_{t2},4)2				1.51(0.399)***
ns(x_{t2},4)3				0.00(0.000)
ns(x_{t2},4)4				-3.37(0.614)***

Covariates	GLM[a]	GLM[b]	GAM[a]	GAM[b]
	Est(se)	Est(se)	Est(se)	Est(se)
ns(x_{t3},4)1				-1.41(0.484)**
ns(x_{t3},4)2				-1.34(0.602)*
ns(x_{t3},4)3				-1.25 (1.043)
ns(x_{t3},4)4				-0.14(0.687)
ns(x_{t4},4)1				2.80(0.448)***
ns(x_{t4},4)2				2.43(0.490)***
ns(x_{t4},4)3				4.72(0.997)***
ns(x_{t4},4)4				3.21(0.732)***
ns(x_{t1S}, 4)1			-10.97(6.153).	
ns(x_{t1S}, 4)2			-3.71(2.241).	
ns(x_{t1S}, 4)3			-13.25(10.110)	
ns(x_{t1S}, 4)4			-7.19(4.857)	
ns(x_{t1T}, 4)1			2.86(1.406)*	
ns(x_{t1T}, 4)2			-0.07(1.076)	
ns(x_{t1T}, 4)3			1.01(2.711)	
ns(x_{t1T}, 4)4			2.05(0.700)**	
ns(x_{t2S}, 4)1			-2.31(3.019)	
ns(x_{t2S}, 4)2			3.87(2.245).	
ns(x_{t2S}, 4)3			-0.88(2.411)	
ns(x_{t2S}, 4)4			0.46(0.781)	
ns(x_{t4T}, 4)1			4.12(1.307)**	
ns(x_{t4T}, 4)2			7.33(1.125)***	

Covariates	GLM[a]	GLM[b]	GAM[a]	GAM[b]
	Est(se)	Est(se)	Est(se)	Est(se)
ns(x_{t4T}, 4)3			13.36(2.437)***	
ns(x_{t4T}, 4)4			-2.032(0.948)*	
ns(x_{t5T}, 4)1			-6.31(1.646)***	
ns(x_{t5T}, 4)2			-5.53(1.563)***	
ns(x_{t5T}, 4)3			-3.17(1.835).	
ns(x_{t5T}, 4)4			-10.03(1.334)***	
ns(t, 4)1			5.65(4.180)	2.08(1.156) .
ns(t, 4)2			5.87(4.794)	1.01(0.786)
ns(t, 4)3			0.00(0.000	2.49(0.666) ***
ns(t, 4)4			-5.57 (4.299)	-1.43(0.326) ***

Signif. codes: $p < 0$ '***'; $p < 0.001$ '**'; $p < 0.01$ '*'; $p < 0.05$ '.'; $p < 0.1$ ' '; $p < 1$

ns stands for natural splines, and the numbers inside and outside the brackets represent the degrees of freedom for the splines.

CHAPTER 3: Time Series Non-Gaussian Bayesian Bivariate Model Applied to Data on HMPV and RSV: A Case of Dadaab in Kenya

3.1 Abstract

Human metapneumovirus (HMPV) have similar symptoms to those caused by respiratory syncytial virus (RSV). The modes of transmission and dynamics of these epidemics still remain poorly understood. Climatic factors have long been suspected to be implicated in impacting on the number of cases for these epidemics. Currently, only a few models satisfactorily capture the dynamics of time series data of these two viruses. In this study, we used a negative binomial model to investigate the relationship between RSV and HMPV while adjusting for climatic factors. We specifically aimed at establishing the heterogeneity in the autoregressive effect to account for the influence between these viruses. Our objective was to assess the presence of influence of high incidences between the viruses and whether higher incidences of one virus are influenced by the other. Our findings showed that RSV contributed to the severity of HMPV. This was achieved through comparison of 12 models of various structures, including those with and without interaction between climatic cofactors. The study has improved our understanding of the dynamics of RSV and HMPV in relation to climatic cofactors there by setting a platform to devise better intervention measures to combat the epidemics. We conclude that preventing and controlling RSV infection subsequently reduces the incidence of HMPV.

Keywords: *Non-Gaussian Bivariate Bayesian model, RSV, HMPV, epidemic, time series, climatic factors.*

3.2 Introduction

Epidemiological knowledge of the respiratory system has been mostly related to developed countries, though the burden of respiratory virus infections (RVIs) is more manifested in developing countries with very high hospitalization and mortality rates (38). Higher mortality is associated with increased displacement into overcrowded refugee camps (39). The burden of RVIs is considerably high during crises times (40) and is more severe in infants (41). Recently, Pastula *et al*. (42) highlighted that hospitalization for respiratory syncytial virus (RSV) is not limited to infants but also to adults. In 2001, HMPV was identified as a potential etiologic agent for respiratory infections (43). A study at Queen Mary Hospital in Hong Kong showed that the peaks of HMPV and that of RSV activity occurred in spring and the early months of summer and viral diagnoses during the study period showed that RSV and HMPV had similar seasonality (44). Guerrero *et al*. (45) indicate that RSV but not HMPV induces a productive infection in human monocyte-derived dendritic cells. Reinfection by RSV has a great impact on human health and may cause long-term effects on the host immune response (46). Greensill *et al*. (47) detected HMPV in 21 out of 30 infants infected with severe RSV and were hospitalized requiring intensive-care unit ventilator support. Konig *et al*. (48) found out that 60% cases with HMPV had RSV. They also found that HMPV contributed to the severity of Lower respiratory tract infections (LRTIs) at a lower rate than RSV and coinfection was considered a cause of severe lower respiratory tract disease. The HMPV infections have similar symptoms to those caused by RSV (49)(50). The HMPV and RSV share similar risk factors (51) and simultaneous detection times (52). The HMPV and RSV may cross-react directly or indirectly because they are both co-viruses to each other (53).

In this paper, we used surveillance data aggregated by month in a time series model and the negative binomial distribution to address the issue of over-dispersion. We model the relationship between two viruses, namely, RSV and HMPV. Meteorological variables were included in the model to help assess for serial correlation. Held *et al*. (54) suggested that environmental factors can be incorporated in these models to improve model fit to data and predictions. These models help to assess the presence of influence of high incidences between the viruses and whether higher incidences of one virus are influenced by another. They also aid in evaluating if an epidemic component can be isolated within or between the viruses and how the autoregressive component captures the residual temporal dependence in the time-series, after adjusting for seasonal effects. In section 3.3, we show the statistical model fitting with and without climatic covariates to a bivariate time-series. In section 3.4 we show the applicability of the models illustrated with a real world example and discuss the results obtained and finally conclude in section 3.5.

3.3 Methods

3.3.1 Statistical modelling

Modeling count data is faced with many challenges since count outcomes do not meet the usual normality assumption required of many standard statistical tests. Typical log-transformation to induce normality does not often work, or categorization of the outcome may lead in loss of information as described by O'Hara and Kotze (55). The most commonly used models to study the dynamics of epidemics and predict future outbreaks using count data are the Poisson (56) and the negative binomial distributions (57). In this work we model the time-evolution of two epidemics using a bivariate approach suggested by Held *et al*. (54). We

assume that we have $i = 1, \dots, m$ 'viruses' and denote with y_{it} the number of cases in virus *i* at time *t*. The general model for the multivariate time series of count events $\{y_{it}, i = 1, \dots, m;\ t = 1, \dots, T\}$ for different virus type i at time t assumes a Poisson distribution with conditional mean μ_{it} given by

$$\log(\mu_{it}) = \lambda_{i,t-1} y_{i,t-1} + \phi_{i,t-1} \sum_{j \neq i} \omega_{ij} y_{j,t-1} + \eta_{i,t} \nu_{it}\,. \qquad (3.1)$$

It holds $\mathrm{VAR}(y_{i,t} | y_{i,t-1}) = \mathrm{E}(y_{i,t} | y_{i,t-1}) = \mu_{it}$. Hence, in the case of a conditional Poisson response model the conditional mean μ_{it},is identical to the conditional variance δ of the observed process.

In model 3.1, $\lambda_{i,t-1}$ is the autoregressive parameter representing the proportion of epidemic cases from the total number of cases for virus type *i* at time *t.* When $\lambda_{i,t-1} \geq 1$ (an outbreak occurs) there is an influx of the endemic cases, and $\lambda_{i,t-1} < 1$ means the process is stable (no outbreak occurs). The $\phi_{i,t-1}$ quantifies the influence of all other virus types j on i; $\eta_{i,t}$ corresponds to an offset term in the model (the monthly varying population counts at time t on virus type i) and ν_{it} is the endemic component as subsequently shown in equation (3.5). The variable $y_{j,t-1}$ denotes the number of cases observed in virus type j at time $t-1$. $\omega_{ij} =$ 1 if pathogens j and i have an autoregressive effect on each other and 0 otherwise,

This model is aggregation consistent where the aggregated counts $y_t = \sum_{i=1}^{m} y_{it}$ have the mean,

$$log(\boldsymbol{\mu}_t) = \boldsymbol{\lambda} \boldsymbol{y}_{t-1} + \boldsymbol{\phi}_{t-1} \boldsymbol{Z}_{t-1} + \boldsymbol{\eta}_t \boldsymbol{\nu}_t,$$

where, $\boldsymbol{Z}_{t-1} = \sum_{j \neq i} \omega_{ij} y_{j,t-1}$, $\boldsymbol{\eta}_t = \sum_{i=1}^{m} \eta_{i,t}$, $\boldsymbol{\phi}_t = \sum_{i=1}^{m} \phi_{i,t}$, $\boldsymbol{\nu}_t = \sum_{i=1}^{m} \nu_{i,t}$.So, the parameter $\boldsymbol{\lambda}$ has the same interpretation for the aggregated counts similar to the counts y_{it}. In

the presence of over-dispersion, the Poisson model is replaced by a negative binomial model where the conditional mean remains unchanged but the variance δ is modified to $\boldsymbol{\mu}_t(1 + \boldsymbol{\mu}_t\boldsymbol{\psi})$ with over-dispersion parameter $\boldsymbol{\psi} > 0$. The extent of over-dispersion is captured by how far the term $\boldsymbol{\psi}$ deviates from zero. An extensive discussion on handling over-dispersion can be found in the work of Ver Hoef and Boveng (58). We are interested in two different types of viruses transmitted through the same route, i.e. respiratory illness. Let $x_{k,t-1}$ denote climatic covariates with $\boldsymbol{\tau}_k$ coefficients in the model and $k = 1, \dots, K$ covariates. In the model, it is assumed that the cases follow a negative binomial distribution, $\boldsymbol{y}_t|\boldsymbol{y}_{t-1} \sim \text{NegBin}(\boldsymbol{\mu}_t, \boldsymbol{\psi})$, with conditional mean

$$log(\boldsymbol{\mu}_t) = \boldsymbol{\lambda}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{\tau}_k x_{k,t-1} + \boldsymbol{\phi}_{t-1}\boldsymbol{Z}_{t-1} + \exp(\boldsymbol{\eta}_t) \quad (3.2)$$

and conditional variance

$$\boldsymbol{\mu}_t(1 + \boldsymbol{\mu}_t\boldsymbol{\psi}). \quad (3.3)$$

The incidence of the disease $\boldsymbol{\mu}_t$was additively decomposed into two parts. The first part,

$$\boldsymbol{\xi}_t = \boldsymbol{\lambda}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{\phi}_{t-1}\boldsymbol{z}_{t-1} + \boldsymbol{\tau}_k x_{k,t-1} \quad (3.4)$$

is the epidemic component explaining the outbreaks or irregularities in the data including the interaction between viruses. The second part is $\nu_{i,t} = \exp(\eta_{i,t})$, which is expressed in log-scale as

$$\log(\nu_{it}) = \alpha_i + \sum_{s=1}^{S}\{\boldsymbol{\gamma}_s \sin(\omega_s t) + \boldsymbol{\delta}_s \cos(\omega_s t)\} \; . \quad (3.5)$$

This is the endemic component that explains the baseline incidence rate of cases. The endemic and epidemic components of the time series were explored and studied allowing for the separation of the regular pattern from irregular ones in estimating the epidemic peaks. The parameter α_i allows for different incidence levels of the viruses, and S is the virus specific number of harmonic waves. The term in curly brackets captures seasonal variations. $\boldsymbol{\gamma}_s$ and $\boldsymbol{\delta}_s$ are the seasonal parameters, while $\omega_s = 2\pi s/12$ for monthly data are the Fourier frequencies.

3.3.2 Likelihood and posterior distribution

The counts $\boldsymbol{y}_t$, conditional on the previous observation $\boldsymbol{y}_{t-1}$ (Only lag one was applied in our case because more than one lag did not fit the data well) are assumed to follow a Negative binomial distribution with mean

$$\boldsymbol{\mu}_t\boldsymbol{\theta} \equiv \boldsymbol{\mu}_t = \boldsymbol{\xi} + \boldsymbol{\nu}\,, \tag{3.6}$$

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{m,}\psi_1, \ldots, \psi_m)^T$ The log-likelihood of the observation $\boldsymbol{y}_t$ is given as

$$l(\boldsymbol{\theta}) = \sum_t l_t(\boldsymbol{\theta}, \boldsymbol{\psi}) \tag{3.7}$$

and the likelihood as,

$$f(\boldsymbol{y}_t|\boldsymbol{\theta}) = exp\left\{\sum_t l_t(\boldsymbol{\theta}, \boldsymbol{\psi})\right\}, \tag{3.8}$$

where,

$$l_t(\boldsymbol{\theta}, \boldsymbol{\psi}) \propto log\ \Gamma\left(\boldsymbol{y}_t + \frac{1}{\boldsymbol{\psi}}\right) - log\,\Gamma\left(\frac{1}{\boldsymbol{\psi}}\right) + \frac{1}{\boldsymbol{\psi}} log\left(\frac{1}{1+\boldsymbol{\psi}\boldsymbol{\mu}_t(\boldsymbol{\theta})}\right) +$$

$$\boldsymbol{y}_t log\left(\frac{\boldsymbol{\psi}\boldsymbol{\mu}_t(\boldsymbol{\theta})}{1+\boldsymbol{\psi}\boldsymbol{\mu}_t(\boldsymbol{\theta})}\right), \tag{3.9}$$

and $\Gamma(.)$ is the gamma function and $\boldsymbol{\psi}$ and $\boldsymbol{\tau}$ are the dispersion parameters. The gamma priors are assumed for $\boldsymbol{\psi}$ and $\boldsymbol{\tau}$,

$$\boldsymbol{\psi} \sim Ga(\alpha_{\boldsymbol{\psi}}, \beta_{\boldsymbol{\psi}}),$$

$$\boldsymbol{\tau} \sim Ga(\alpha_{\boldsymbol{\tau}}, \beta_{\boldsymbol{\tau}}).$$

The virus dependent effects α_i are assumed to be independent and normally distributed with a large variance,

$$\alpha = (\alpha_1, \ldots, \alpha_I) \sim N(0, \sigma_\alpha^2 I),\ \sigma_\alpha^2 = 10^6,$$

where I is an identity matrix. All model parameters are non-negative and therefore we propose gamma prior distributions for them. The rate parameters $\boldsymbol{\lambda}_t$, assumes independent gamma priors with gamma hyperpriors on the second parameter,

$\boldsymbol{\lambda}_t \sim Ga(\alpha_{\boldsymbol{\lambda}}, \beta_{\boldsymbol{\lambda}})$ and $\beta_{\boldsymbol{\lambda}} \sim Ga(a, b)$.

Where we use $\alpha_{\boldsymbol{\lambda}} = 1, a = 10$ and $b = 10$, with values for $\alpha_{\boldsymbol{\lambda}}$, a and b chosen arbitrarily.

Independent normal priors are assumed for $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$,

$$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_I) \sim N(0, \sigma_\gamma^2 I),\ \sigma_\gamma^2 = 10^6,$$

$$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_I) \sim N(0, \sigma_\delta^2 I),\ \sigma_\delta^2 = 10^6.$$

The parameter $\boldsymbol{\phi}_t$ assumes gamma priors, $\boldsymbol{\phi}_t \sim Ga(\alpha_{\boldsymbol{\phi}}, \beta_{\boldsymbol{\phi}})$.

The posterior distribution is therefore given as,

$$f(\boldsymbol{\theta}|\boldsymbol{y}_t) \propto f(\boldsymbol{y}_t|\boldsymbol{\theta})f(\boldsymbol{\theta}),$$

which can be expressed as,

$$f(\boldsymbol{\theta}|\boldsymbol{y}_t) \propto exp\left\{\sum_t l_t(\boldsymbol{\theta},\boldsymbol{\psi})\right\} \times \prod_{s=1}^{S} e^{-\frac{1}{2c}\sigma_\gamma^2} \times \prod_{s=1}^{S} e^{-\frac{1}{2c}\sigma_\delta^2} \times \prod_{i=1}^{m} e^{-\frac{1}{2c}\sigma_{\alpha_i}^2}$$

$$\times \prod_{i=1}^{m} {\lambda_i}^{\alpha_{\lambda_i}-1} e^{-\beta_{\lambda_i}^{\lambda_i}} {\lambda_i}^{a-1} e^{-b\lambda_i} \times \prod_{i=1}^{m} {\psi_i}^{\alpha_{\psi_i}-1} e^{-\beta_{\psi_i}^{\psi_i}} \times \prod_{i=1}^{m} {\phi_i}^{\alpha_{\phi_i}-1} e^{-\beta_{\phi_i}^{\phi_i}}$$

$$\times \prod_{i=1}^{m} {\tau_i}^{\alpha_{\tau_i}-1} e^{-\beta_{\tau_i}^{\tau_i}}. \qquad (3.10)$$

3.3.3 Simulations

We investigated the proposed model performance on simulated data. We simulated bivariate data using a frequentist approach in R software using the package "Surveillance" previously used by Held *et al*.(59)(60). We used the function "hhh4" with the class "disprog" to simulate two disease pathogen counts replicated 10000 times. We then applied the Bayesian approach to compare different models based on varied scenarios. We considered a situation with the presence of overdispersion where parameter $\psi_i \neq 0$ assumed the negative binomial distribution and where $\psi_i = 0$ assumed the Poisson distribution. We also considered the presence and absence of parameter λ_i (the 'epidemic' component) to evaluate temporal dependence. In this simulation we disregarded the linear trend. It is evident from Table 3.1, that the simulation results show that $\psi_i = 0$ and therefore the best performing model is the Poisson (model 3.2) with the presence of the epidemic component having the least AIC = 1626.58.

Table 3.1 Simulation results including Parameter estimates, Standard errors and measure of model Goodness of Fit.

Parameter	Model1 (ψ=0 λ=0)	Model2 (ψ=0 $\lambda\neq$0)	Model3 ($\psi\neq$0 λ=0)	Model4 ($\psi\neq$0 $\lambda\neq$0)
ψ_1	-	-	0.0000 (0.0000)	0.0000 (0.0000)
ψ_2	-	-	0.0000 (0.0000)	0.0000 (0.0001)
λ_1	-	0.1730 (0.3135)	-	0.1743 (0.3072)
λ_2	-	0.4337 (0.2010)	-	0.4482 (0.2115)
ϕ_1	0.4727 (0.2262)	0.4586 (0.2300)	0.4726 (0.3092)	0.4585 (0.2300)
ϕ_2	0.8123 (0.0420)	0.0963 (0.2204)	0.3034 (0.2204)	0.1485 (0.2424)
AIC	1644.14	1626.58	1636.92	1630.33

This method of analysis failed to detect over-dispersion in the simulated data and there was temporal dependence.

3.3.4 Application on data

Let $\{y_{it}, i = 1, 2;\ t = 1, \ldots, 48\}$ be the time series of virus counts for RSV (y_{1t}) and HMPV (y_{2t}) over the 48 months study time-frame. The bivariate model for the two time series would therefore be;

$$\log\begin{pmatrix}\mu_{1,t}\\ \mu_{2,t}\end{pmatrix} = \begin{pmatrix}\lambda_{1,t-1} & \phi_{1,t-1}\\ \phi_{2,t-1} & \lambda_{2,t-1}\end{pmatrix}\begin{pmatrix}y_{1,t-1}\\ y_{2,t-1}\end{pmatrix} + \begin{pmatrix}\tau_{1,1} & \tau_{1,2} & \tau_{1,3} & \tau_{1,4}\\ \tau_{2,1} & \tau_{2,2} & \tau_{2,3} & \tau_{2,4}\end{pmatrix}\begin{pmatrix}x_{1,t-1}\\ x_{2,t-1}\\ x_{3,t-1}\\ x_{4,t-1}\end{pmatrix} + \eta_t\begin{pmatrix}\nu_{1,t}\\ \nu_{2,t}\end{pmatrix},$$

where

$$\nu_{1,t} = \alpha_1 + \gamma_{1,1}\sin(\omega_1 t) + \delta_{1,1}\cos(\omega_1 t) + \gamma_{1,2}\sin(\omega_2 t) + \delta_{1,2}\cos(\omega_2 t),$$

$$\nu_{2,t} = \alpha_2 + \gamma_{2,1}\sin(\omega_1 t) + \delta_{2,1}\cos(\omega_1 t) + \gamma_{2,2}\sin(\omega_2 t) + \delta_{2,2}\cos(\omega_2 t)$$

and $x_{1,t-1}, x_{2,t-1}, x_{3,t-1}$ and $x_{4,t-1}$ are the climatic factors representing rainfall, wind speed, mean dew point and visibility, respectively. The term η_t corresponds to an offset term in the model (the monthly varying population counts at time *t*).

The models were compared for their fit to the epidemic data. Naturally, models are compared for their performance based on the ability to fit well on data and their reliability in predicting future epidemic outbreaks. Fundamentally, in our model fitting to data we searched for the model that provided the best trade-off between the fit to data and the model structure complexity. Often, approaches such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are sufficient for ranking and selecting the best performing models. However, when the data is non-Gaussian and the model is Bayesian, like in our case, then the deviance information criterion (DIC) is more appropriate. For the comparison of our models, we used the DIC proposed by Spiegelhalter *et al*. (61), specifically for Bayesian based models and it is a Bayesian generalisation of the AIC and BIC. The model with the smallest DIC value gives the better trade-off between model fit and complexity; therefore, it is considered as the model that best predicts a replication of a data set with a similar structure as that which was observed currently (62).

To further assess the model performance with regards to the parameters, sensitivity analysis to alternative prior assumptions was performed because there are no true priors in the Bayesian analysis. In order to ensure reliable and robust results from our best model, it was crucial to verify how sensitive the resulting posteriors were for each prior input for the epidemic parameter λ_{it} and ϕ_{it}, the parameter that quantifies the influence of one virus on the other.

Therefore, we assumed independent Gamma priors with uniform hyper-priors on the second parameter, $\lambda_{it} \sim Ga(\alpha_{\lambda,}\beta_{\lambda})$ and $\beta_{\lambda} \sim Beta(a, b)$ using $\alpha_{\lambda} = 1, a = 0.5$ and $b = 0.5$. Similarly for the influential parameter we used the uniform prior, $\phi_{it} \sim Beta(\alpha_{\phi,}\beta_{\phi})$. To our understanding this comparison of models has not yet been done using RSV and HMPV time series data. All the models in our work were run and tested in the statistical software WinBUGS 14. The models differed on the epidemic part $\xi_{i,t}$ by the assumptions made on the interactions between the viruses. We used 6 models depending on the assumptions applied as explained below with each model with a corresponding inclusion of climatic factors giving rise to a total of 12 models. (Table 3.2).

Table 3.2 Models of the epidemic part $\xi_{(i,t)}$ with assumptions made on interactions between the viruses with and without the climatic factors.

Model	$\xi_{i,t}$ **(with climatic factors)**	$\xi_{i,t}$ **(without climatic factors)**
1	$\lambda y_{i,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda y_{i,t-1}$
2	$\lambda y_{i,t-1} + \phi \sum_{j \neq i} w_{ji}\, y_{j,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda y_{i,t-1} + \phi \sum_{j \neq i} w_{ji}\, y_{j,t-1}$
3	$\lambda_i y_{i,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda_i y_{i,t-1}$
4	$\lambda_i y_{i,t-1} + \sum_{j \neq i} w_{ji}\, \phi_i y_{j,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda_i y_{i,t-1} + \sum_{j \neq i} w_{ji}\, \phi_i y_{j,t-1}$
5	$\lambda_{i,t-1} y_{i,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda_{i,t-1} y_{i,t-1}$
6	$\lambda_{i,t-1} y_{i,t-1} + \sum_{j \neq i} w_{ji}\, \phi_{i,t-1} y_{j,t-1} + \tau_{i,k} x_{k,t-1}$	$\lambda_{i,t-1} y_{i,t-1} + \sum_{j \neq i} w_{ji}\, \phi_{i,t-1} y_{j,t-1}$

In model 1 it is assumed that the incidence rate is the same in every virus; hence, no interactions between the viruses. Model 2 assumes that there is the interaction between viruses where the sum of related viruses at the same time point has an equal rate. In Table 3.2, Models 3 and 4 are generalisations of models 1 and 2, respectively; with a different rate for each virus. Models 5 and 6 generalise model 3 and 4, respectively; with a different rate for each virus per time point. The best model was then evaluated on whether; there were interactions between cases of RSV and HMPV (alternatively stated as $\phi_{\mathrm{RSV}} \neq \phi_{\mathrm{HMPV}} \neq 0$), the existence of the influence of RSV on HMPV ($\phi_{\mathrm{RSV}} = 0, \phi_{\mathrm{HMPV}} \neq 0$), the existence of the influence of HMPV on RSV ($\phi_{\mathrm{HMPV}} = 0, \phi_{\mathrm{RSV}} \neq 0$) or there were no interactions at all ($\phi_{\mathrm{RSV}} = \phi_{\mathrm{HMPV}} = 0$).

3.4 Results and Discussions

3.4.1 Data

The monthly observed number of RSV and HMPV cases in Dadaab from September 2007 to August 2011 that were collected in the surveillance system was plotted (Figure 3.1). Similar trends were also observed by Agoti *et al* (63)and Nyoka *et al* (64). Wilkesmann *et al*. (65) showed that HMPV and RSV causes similar symptoms and clinical severity with similar seasonality. A similar finding was reached by Kim *et al*. (66) who investigated the clinical and epidemiological assessment of HMPV and RSV in Seoul, Korea, 2003-2008.

3.4.2 Exploratory Data Analysis (EDA)

Figure 3.1 shows the monthly counts of RSV and HMPV epidemics plotted against time. The plot shows cumulative counts of HMPV cases that were approximately 2.5 times less than the RSV counts for the same timeframe.

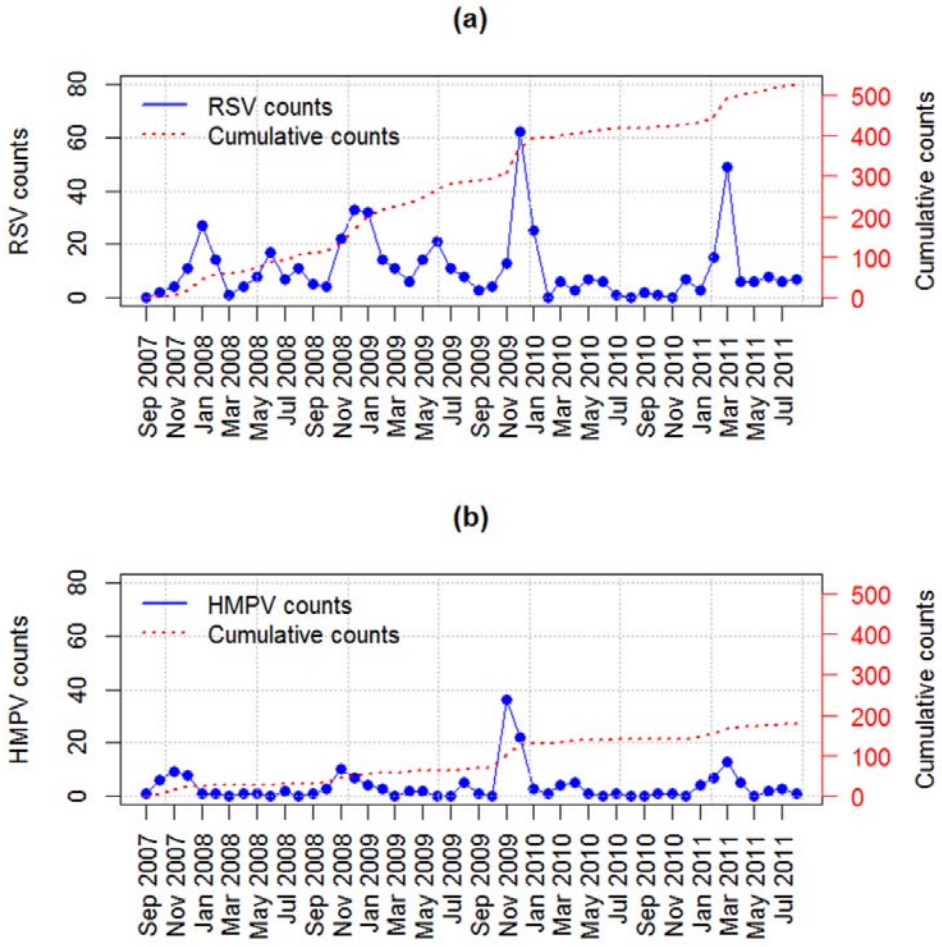


Figure 3.1 The monthly counts of epidemics (a) RSV and (b) HMPV plotted against time. The cumulative counts of HMPV cases were approximately 2.5 times less than the RSV counts for the same time-frame

The HMPV data shows a strong seasonality pattern as indicated by the four peaks during November of the years 2007, 2008 and 2009 while a fourth peak appears in March 2011 (Figure 3.1(b)). Figure 3.2 show that the epidemics coincide in timing of their occurrence peaks, especially in March 2011.

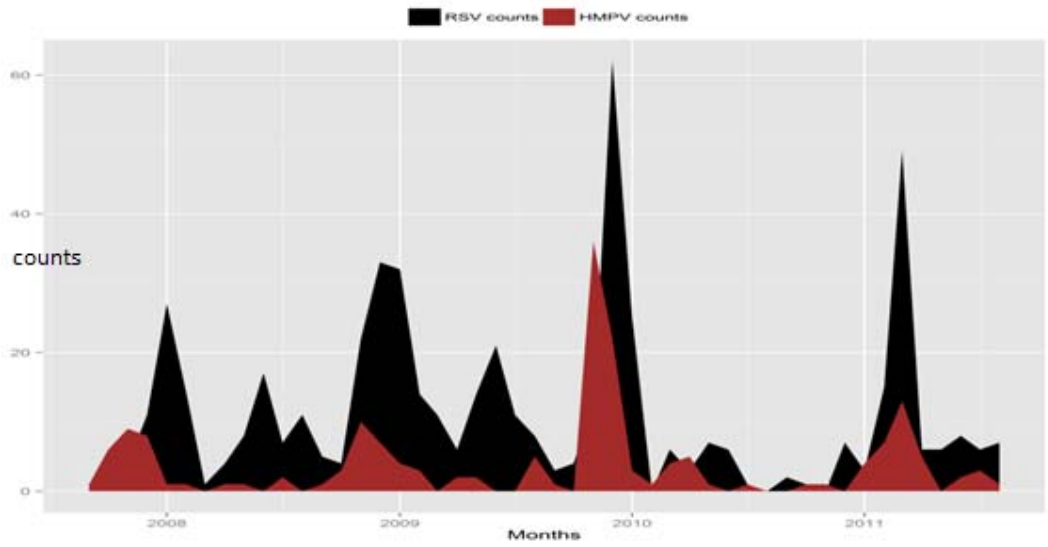


Figure 3.2 The monthly counts of RSV and HMPV plotted against time. Overall, the epidemics coincide in timing of their occurrence peaks, especially in March 2011

This plot shows that the HMPV peaks coincide with the RSV peaks. In their paper, Cuevas *et al*. (43) observed that HMPV incidence had increased with increases in RSV incidence. Another study in Yemeni children younger than 2 years identified co-infections of RSV and HMPV, and also showed that there were seasonal variations of RSV and HMPV with a peak of RSV in December and January and a peak of HMPV in February and March (67).

Figure 3.3 shows a plot of time versus the disease counts along with a smoothing spline fit to the data. Both plots accentuate the need to include the nonlinearity effect in our models. The histograms indicate the disease counts are skewed to the left and therefore are not normally

distributed. In this case, we know that we need to use non-Gaussian techniques to model this data. These insights are used to develop the mean function for the models.

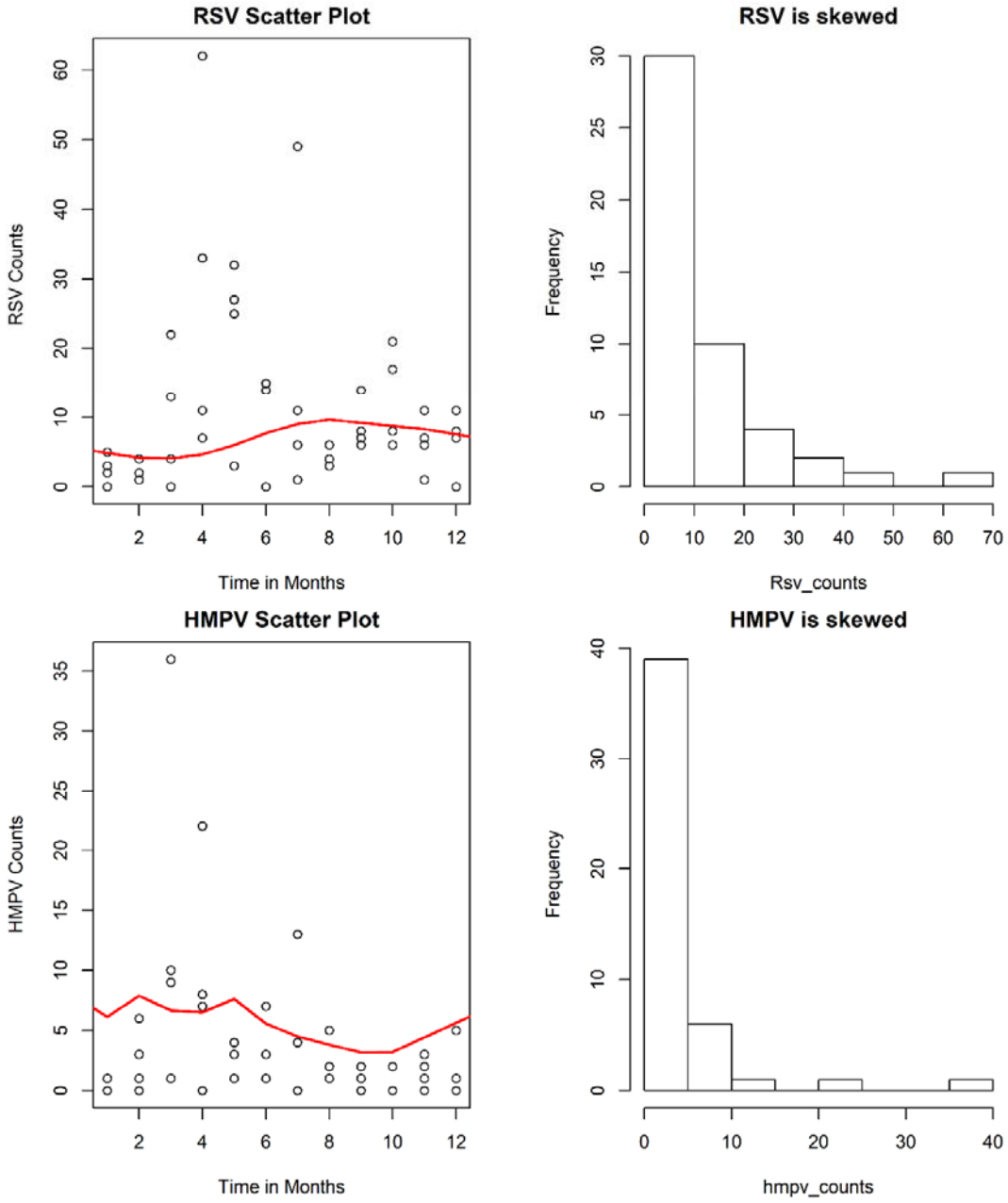


Figure 3.3 Scatter plots and histograms for RSV and HMPV counts. Both are skewed to the right (the red solid lines denote the fitted curves using smoothing splines)

Changes of dispersion and dependence in time are accounted for in the covariance structure of the model. As time progresses, the change in dispersion is evident from Figure 3.1 and Figure 3.3. The correlation plot in Figure 3.4 captures the strength of dependence.

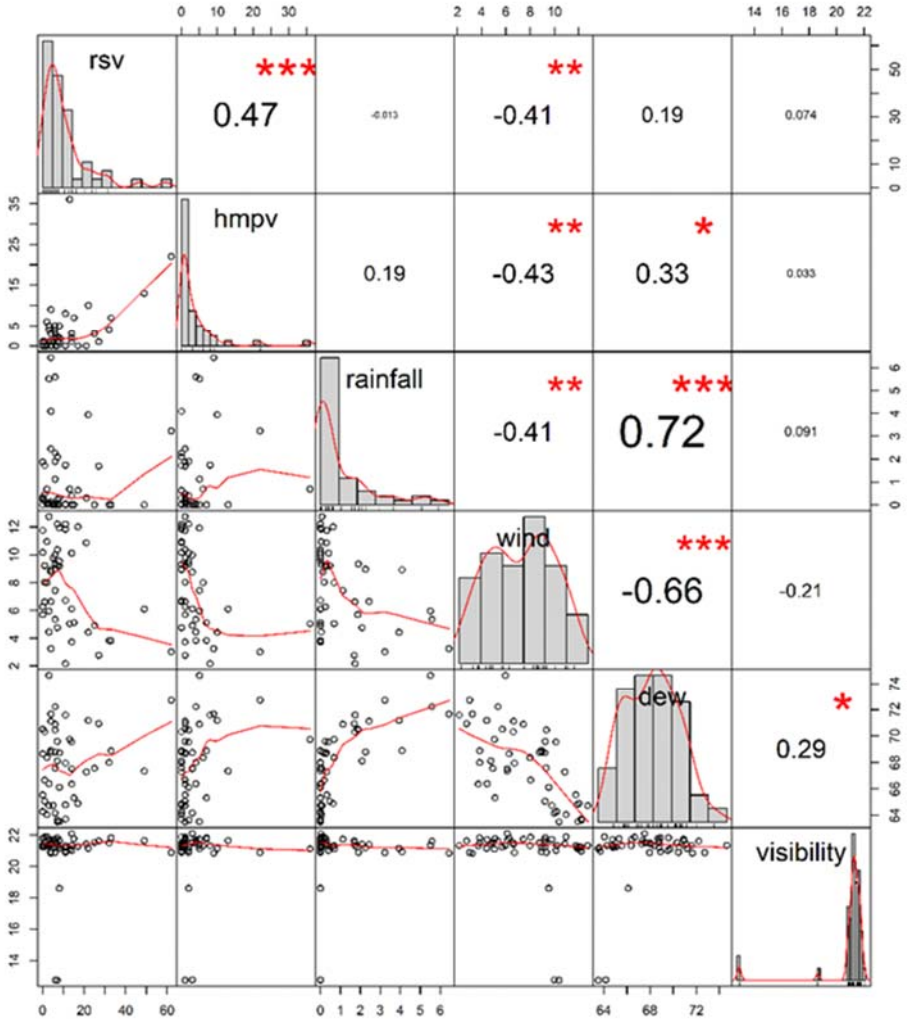


Figure 3.4 Correlation matrix and marginal distribution of the disease counts and the climatic factors. Signif. codes for correlations: $pvalue < 0$ '*'; $pvalue < 0.001$ '**'; $pvalue < 0.01$ '*'**

There is a significant moderate positive dependence with disease pathogens and a significant moderate negative correlation of both the disease pathogens with the speed of wind. The

marginal distribution for each of the disease pathogen and each of the climatic variables is examined in Figure 3.4. The distributions of the disease pathogens as previously highlighted in Figure 3.3 are right-skewed and have outliers, as shown in Figure 3.4.

3.4.3 Model Results

We compared 12 models with various structures (Table 3.2) and the results for the DIC values are given in Table 3.3.

Table 3.3 Comparison DIC values for different models.

Model	1	2	3	4	5	6
DIC (with climatic factors)	490.43	558.30	559.46	558.45	502.17	173.52
DIC (without climatic factors)	549.82	541.11	548.44	536.09	571.72	744.22

Model 6 and 1 with climatic factors clearly out-perform the other models since, overall, they have lower DIC values. Model 6 with climatic factors had the least DIC value (173.52) and provided the best fit and explanation for the variation observed in the data. This is probably due to the seasonality nature of the climatic factors therefore by including them in the model supports the seasonality of RSV and HMPV thereby explaining the data better. The models showed that the inclusion of climatic factors play an important role in the estimation of the number of cases for the two epidemics (RSV and HMPV). From our previous work using the same dataset we noted a similar conclusion that the use of climatic factors explained the seasonality of RSV (64). This implies that having considered the different rate for each virus at every time point, the models with the best fit to data were those with climatic factors. We

840 further considered different scenarios on the best model with four sub-models (results are
841 shown in Table 3.4).

842 **Table 3.4 Four sub-models from the best model. The symbols "–" and "√" mean the**
843 **absence and presence of interactions, respectively. Model 6 (i) no interactions between**
844 **HMPV and RSV (ϕ_HMPV=ϕ_RSV=0); Model 6 (ii) influence of HMPV on RSV**
845 **(ϕ_RSV≠0,ϕ_HMPV=0), Model 6 (iii) influence of RSV on HMPV**
846 **(ϕ_RSV=0,ϕ_HMPV≠0) and Model 6 (iv) interactions between HMPV and RSV**
847 **(ϕ_HMPV≠ϕ_RSV≠0).**

Model	HMPV→RSV	RSV→HMPV	DIC
6(i)	–	–	543.68
6(ii)	√	–	457.61
6(iii)	–	√	112.14
6(iv)	√	√	173.52

848

849 Model 6(i) in Table 3.4 does not allow for interactions between HMPV and RSV ($\phi_{\mathrm{HMPV}} =$
850 $\phi_{\mathrm{RSV}} = 0$) and its DIC value is 543.68. Model 6(ii) includes the influence of HMPV on RSV
851 with influence of RSV on HMPV equal to zero. This model yielded a DIC value of 457.61.
852 Model 6(iii) includes the influence of RSV on HMPV where the influence of HMPV on RSV
853 is zero. Compared to the others, this model yielded the smallest DIC value of 112.14 (Table
854 3.4). The results from sensitivity analysis shown in Figure 3.5, indicates that this model is
855 robust and insensitive to the prior distribution since its posterior distribution did not
856 dramatically change upon altering the base prior parameter values. Model 6(iv) has both the

influence of RSV on HMPV and the influence of HMPV on RSV which is the full model with a DIC value of 173.52 (Table 3.4).

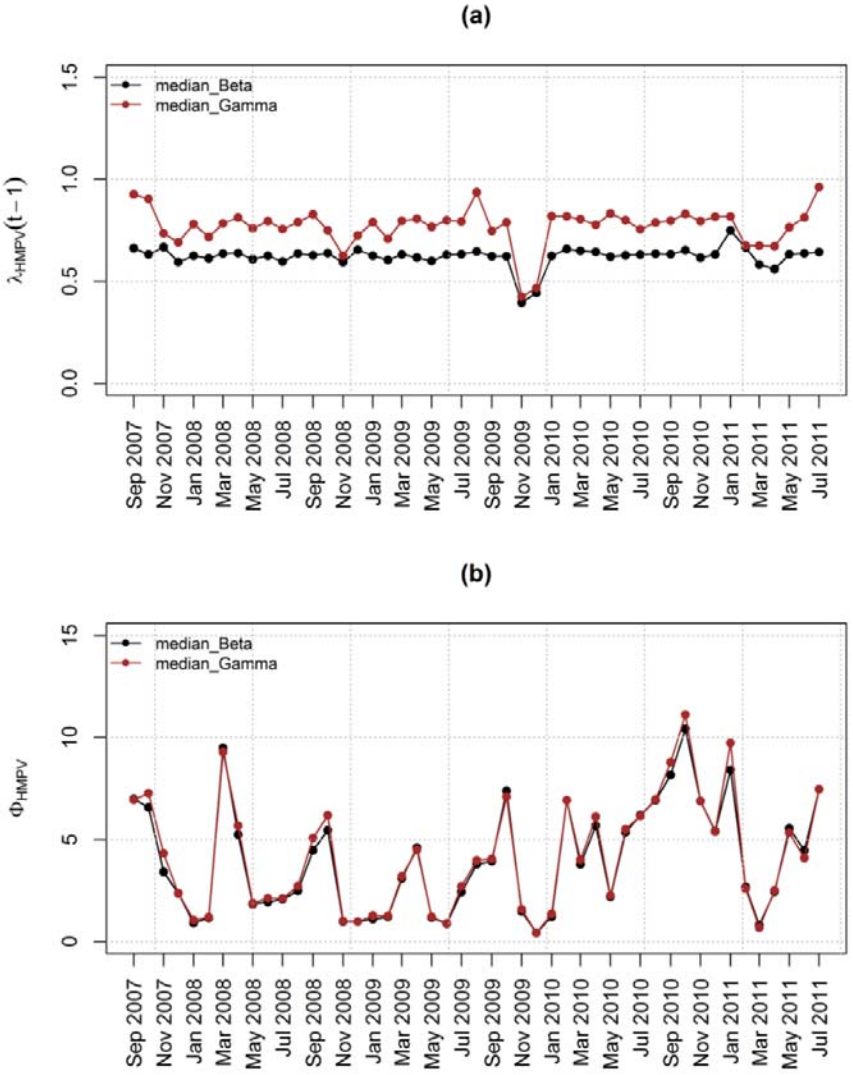


Figure 3.4 Posterior median values for the priors with Gamma and Beta distributions for the best model. Plots showing the Posterior median values of (a) λ_HMPV and (b) ϕ_HMPV for model 6(iii). Median_Beta and median_Gamma are the posterior medians from the Beta distribution and the Gamma distribution priors respectively.

This indicates that the additional parameter (i.e., influence of HMPV on RSV) into model 6(iii) does not significantly improve the model fit to data. A similar observation was made by Lazar *et al*. who noted that HMPV did not contribute to the severity of RSV (68). In our study we have shown that incidence of RSV influenced that of HMPV from the best model fit. This is corroborated in findings from a similar investigation of the influence of RSV on HMPV by Greensill *et al*. (47) in which 70% of children infected with RSV were co-infected with HMPV.

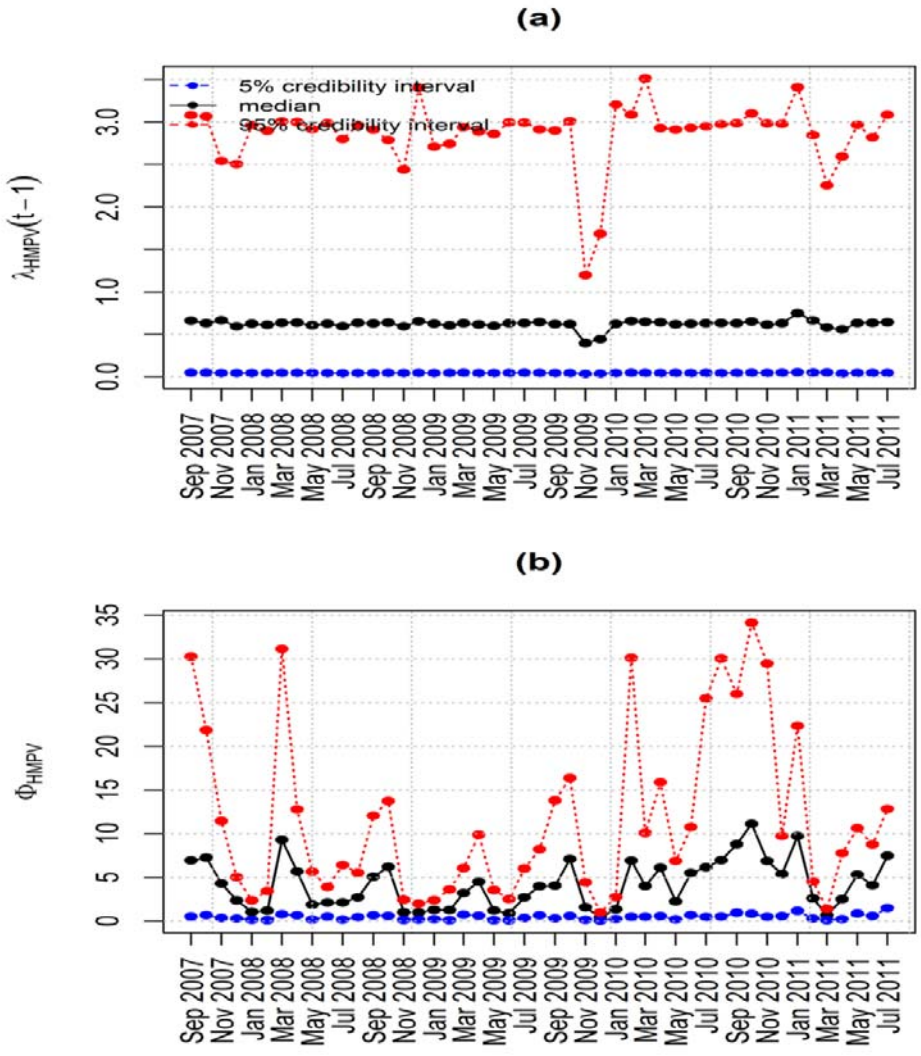
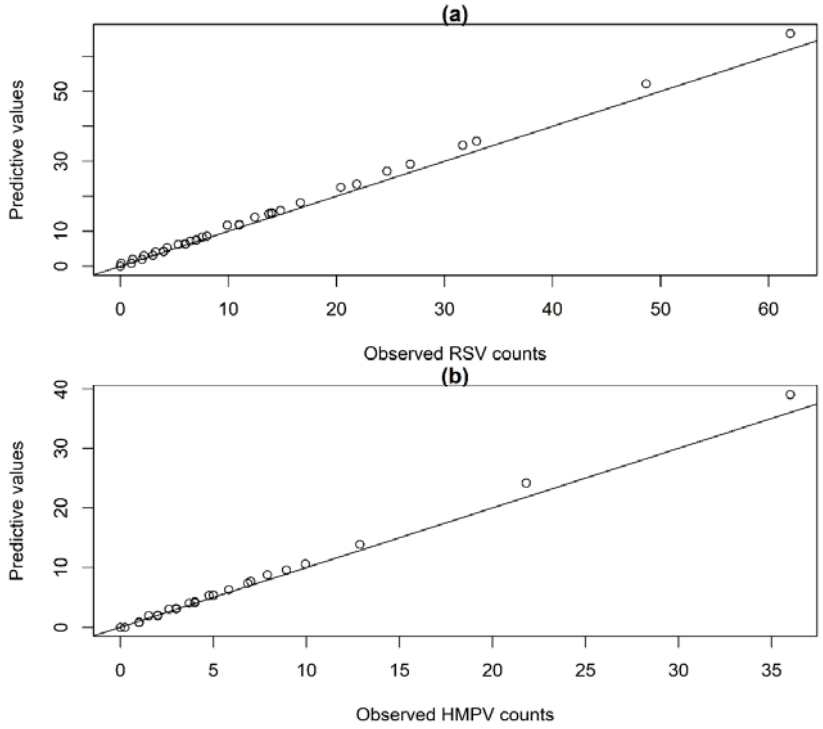


Figure 3.6 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (a) λ_HMPV and (b) ϕ_HMPV for model 6(iii).

Elsewhere, Cuevas *et al*. (43) observed that HMPV incidence increased with increasing number of RSV cases, suggesting the presence of a strong association between the dynamics of the two epidemics.

The epidemic parameter λ_{HMPV} for model 6(iii) in Figure 3.6(a) does not exceed the value 1. This implies that the time series is stable without a detection of an outbreak of HMPV due to the influence of RSV. Figure 3.6(b) shows the influence of RSV on HMPV with biannual peaks noted over the study period. The other parameters estimated in this model are shown in the supplementary materials (Table 3.5) that includes the posterior median and point-wise 95% credibility intervals. In particular from this table (Table 3.5), the posterior median and the point-wise 95% credibility intervals for the overdispersion parameters ψ_{HMPV} and ψ_{RSV} were 7.762(0.238, 116.1) and 4.688(0.090, 97.33) respectively. This indicates the existence of overdispersion which relaxes our adoption of the negative-binomial modelling. Figure 3.8 - 3.11 in supplementary materials show the posterior median and point-wise 95% credibility intervals for the climatic factors. Figure 3.7 shows the scatter plots of realized vs. posterior predictive values for RSV and HMPV from the best model fit measuring the discrepancies between observed and predictive values. As can be seen, there was some systematic difference between the realized and posterior predictive values but this was the best fit among the models fitted.



894

895 **Figure 3.7 Realised vs. Posterior Predictive Values of RSV and HMPV Disease Counts for**
896 **the best model**

897

898 Some of the limitations of this study were that the available time series data for the viruses
899 was only for a four year time-frame which is short for time series analysis and that the
900 climatic factors were from the neighboring weather station in Garissa, which is about 100
901 kilometers away from the Dadaab camp. Nevertheless, the weather measurements are a good
902 representation of the actual weather around Dadaab. There was no establishment of whether
903 patients were co-infected during virus testing. We used the DIC which is an approximation to
904 a penalized loss function based on the deviance to evaluate the models. DIC under-penalized
905 the more complex models and therefore its application is valid only when the number of
906 parameters is much smaller than the number of independent observations(69). Classical model

907 selection was used that assumes that there is at least a best model for deducing inferences
908 from the data. The criterion used to select the best model did not allow for the computation of
909 weights of each fitted model to quantify for uncertainty, that is the model averaging
910 techniques were not used(70).

911 **3.5 Conclusion**

912

913 We provided a comprehensive comparison of RSV and HMPV in a refugee camp setting by
914 using a bivariate non-Gaussian model to jointly model the epidemics. By comparing various
915 model structures, we identified a model that could better explain the variations although it did
916 not satisfactorily fit the epidemic data. The models and estimated parameters also provided
917 clues into the dynamics and stability of the two epidemics. Our results demonstrated the
918 influence of RSV on HMPV while adjusting for climatic factors. The climatic factors played a
919 significant role in explaining the influence of RSV incidence on HMPV incidence. These
920 models are important to the public health implication since controlling the incidence of RSV
921 would consequently reduce the incidence of HMPV.

922 **3.6 Acknowledgments**

923

924 The authors wish to acknowledge the CDC Kenya Refugee Health Program for their tireless
925 work in the surveillance in the refugee camp and their assistance with the data collection and
926 management.

927 **3.7 Data files**

928

929 The data files and supplementary materials used for this study can be found at

930 https://figshare.com/s/e8a735c22f554d8372e3 DOI: 10.6084/m9.figshare.5340724 .

931 **3.8 Conflict of interest**

932 The authors declare that there is no conflict of interest.

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

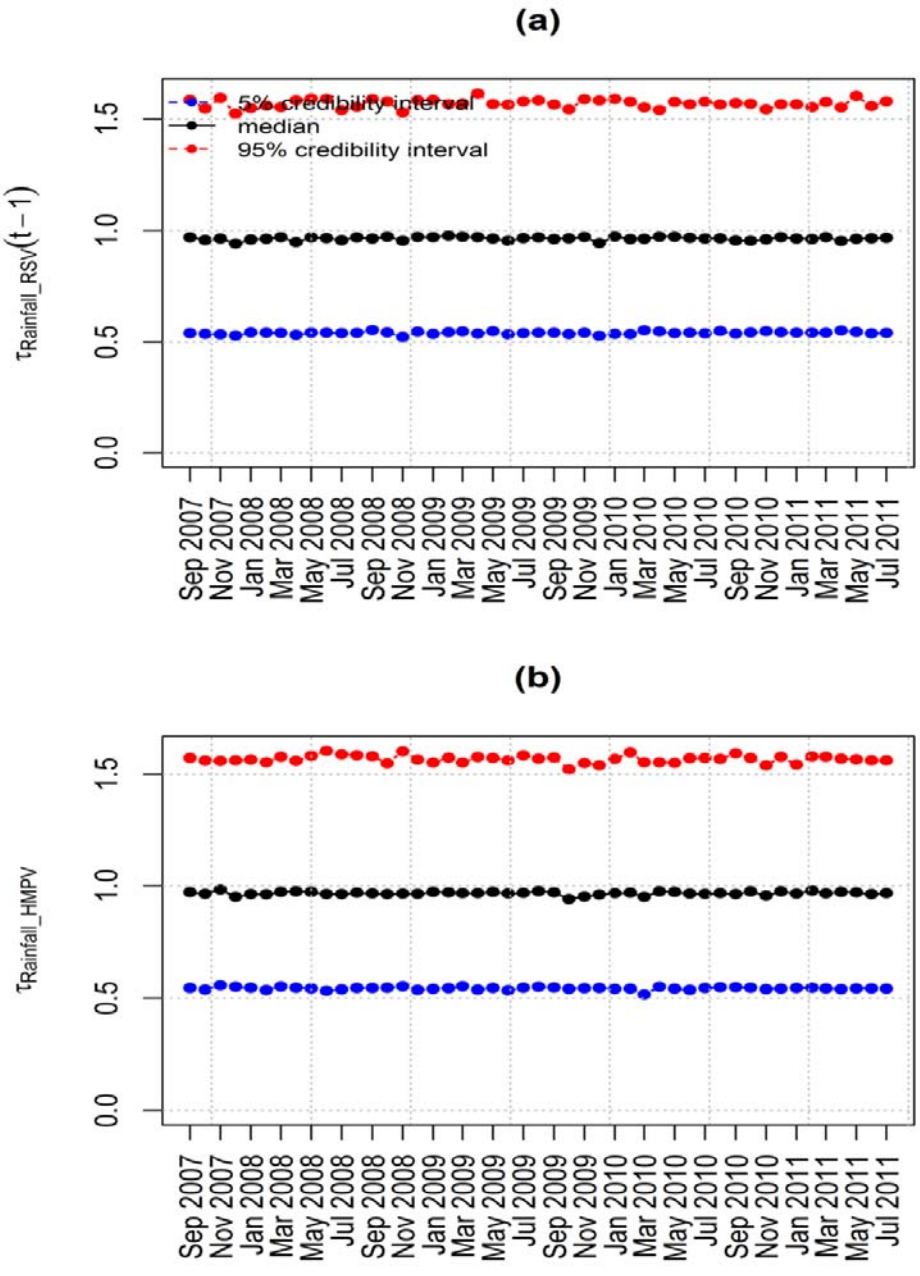
Supplementary materials

Table 3.5 Posterior median and point-wise 95% credibility intervals for the best model.

Parameter	5.0%	Median	95%
alpha1	-4.283	-3.998	-3.683
alpha2	-3.765	-3.765	-3.481
delta11	-2.564	-2.564	-1.979
delta21	-4.783	-4.783	-4.023
gamma11	-6.303	-5.653	-4.812
gamma21	-9.209	-7.965	-6.934
psi1	0.238	7.762	116.1
psi2	0.090	4.688	97.33

965

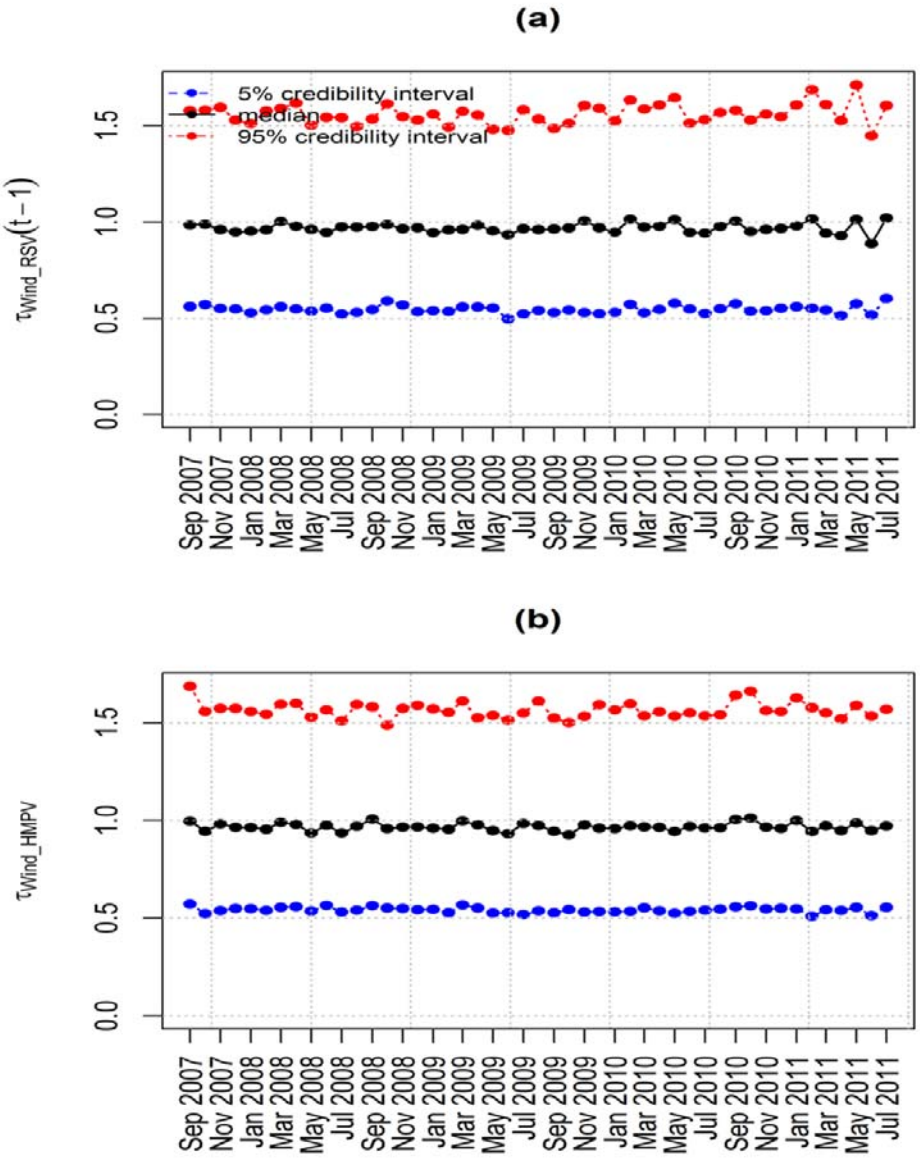
966 **Figure Legends**



967

968 **Figure 3.8 Posterior median and point-wise 95% credibility intervals for the best model.**
969 **Plots showing the Posterior median and point-wise 95% credibility interval of (a)**
970 **τ_(Rainfall_RSV) and (b) τ_(Rainfall_HMPV) for model 6(iii).**

971



972

973 **Figure 3.9 Posterior median and point-wise 95% credibility intervals for the best model.**
974 **Plots showing the Posterior median and point-wise 95% credibility interval of (a)**
975 **τ_(Wind_RSV) and (b) τ_(Wind_HMPV) for model 6(iii).**

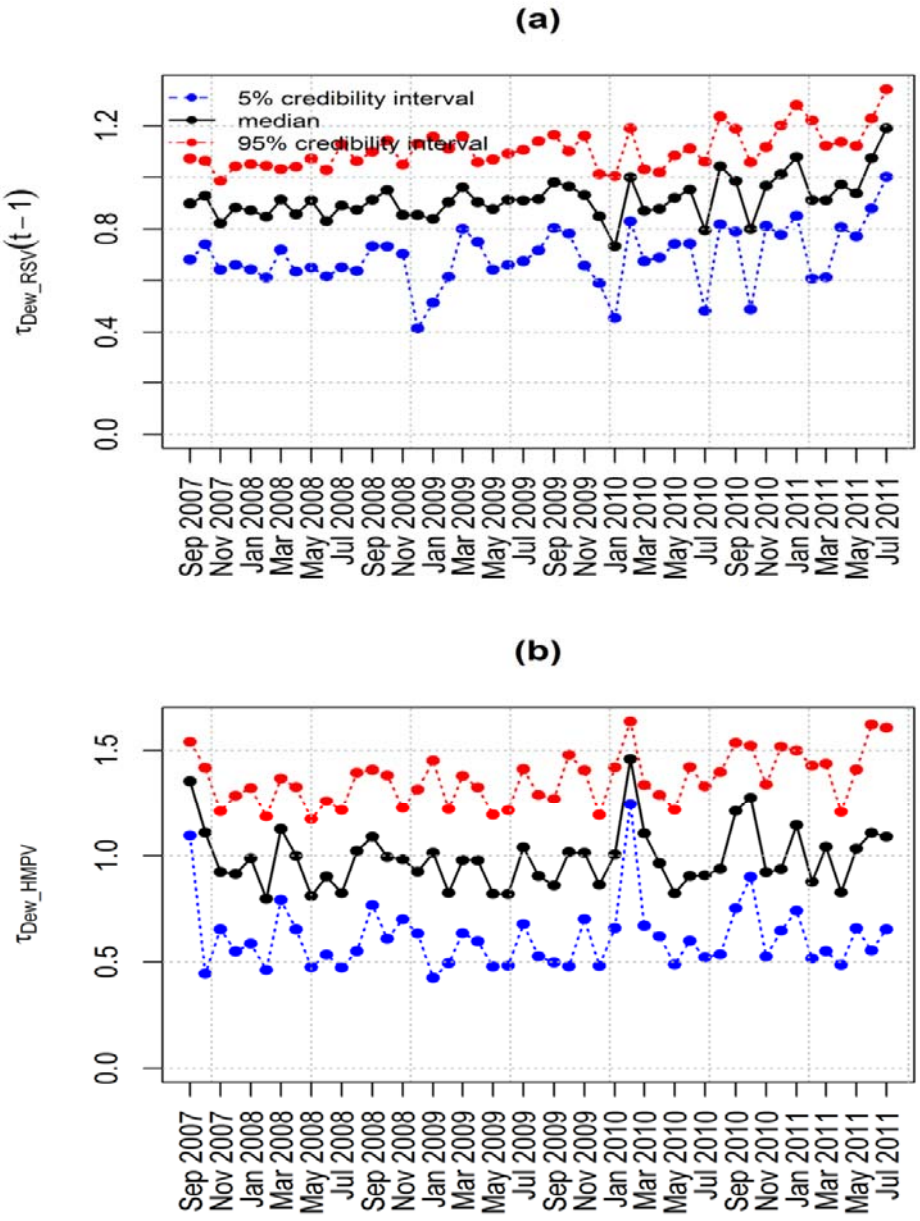


Figure 3.10 Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (a) τ_(Dew_RSV) and (b) τ_(Dew_HMPV) for model 6(iii).

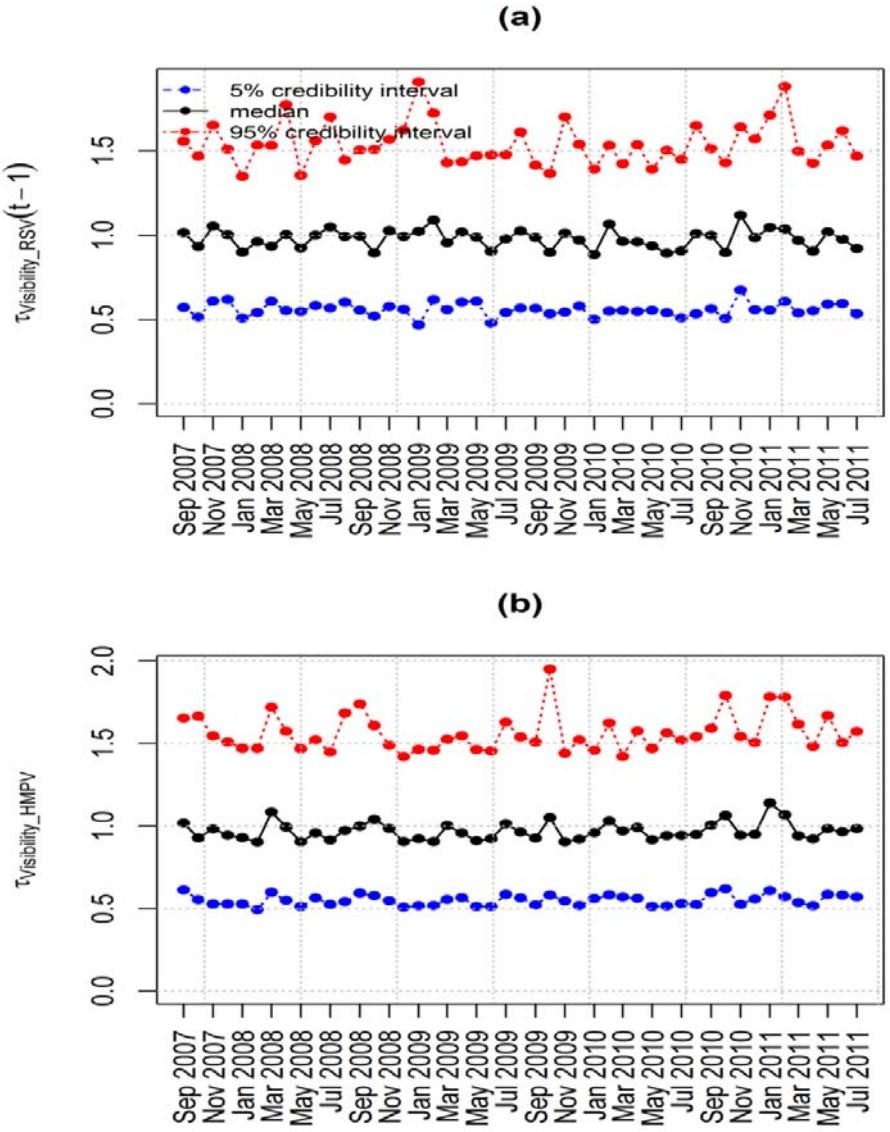


Figure 3.11 **Posterior median and point-wise 95% credibility intervals for the best model. Plots showing the Posterior median and point-wise 95% credibility interval of (a) τ_(Visibility_RSV) and (b) τ_(Visibility_HMPV) for model 6(iii).**

CHAPTER 4: A Non-Gaussian Bayesian Model of Multiple Time Series Epidemics of Acute Respiratory Illness: Case of Dabaab in Kenya

4.1 Abstract

Most models do not consider multiple viruses nor incorporate the time varying effects of these components. Common ARIs etiologies identified in developing countries include respiratory syncytial virus (RSV), human metapneumovirus (HMPV), influenza viruses (Flu), parainfluenza viruses (PIV) and rhinoviruses with mixed co-infections in the respiratory tracts which make the etiology of Acute Respiratory Illness (ARI) complex. The occurrence of different diseases in time contributes to multivariate time series data. In this work, the surveillance data are aggregated by month and are not available at an individual level. This may lead to over-dispersion; hence the use of the negative binomial distribution. In this paper, we describe an approach to analyze multivariate time series of disease counts. A previously used model in the literature to address dependence between two different disease pathogens is extended. We model the contemporaneous relationship between pathogens namely, RSV, HMPV and Flu from surveillance data in a refugee camp (Dadaab) for children under 5 years to investigate for serial correlation. The models evaluate for the presence of heterogeneity in the autoregressive effect for the different pathogens and whether after adjusting for seasonality, an epidemic component could be isolated within or between the pathogens. The model helps in distinguishing between an endemic and epidemic component of the time series that would allow the separation of the regular pattern from irregularities and outbreaks. The use of the models described in this study, could help public health officials predict increases in each pathogen infection incidence among children and help them prepare and respond more swiftly to increasing incidence in low-resource regions or communities.

Keywords: *Respiratory syncytial virus, human metapneumovirus, influenza, time series, seasonal, modeling.*

4.2 Introduction

Common ARIs etiologies identified in developing countries include respiratory syncytial virus (RSV), human metapneumovirus (HMPV), influenza viruses (Flu), parainfluenza viruses (PIV) and rhinoviruses with mixed co-infections in the respiratory tracts which make the etiology of Acute Respiratory Illness (ARI) complex (71). The highest mortalities in 2015 caused by ARI among children less than five years of age were in Sub-Saharan Africa (72). Co-infections of multiple viral etiologies of lower respiratory tract was detected among Egyptian children under 5 years of age (73). In their paper, Ivana et. al (74) assessed and explored the proportional contribution of mixed viral infections and their separate contributions of flu, PIV, and adeno viruses to severe acute lower respiratory infections in children less than 5 years. In the past decade, progressive availability of vaccines against influenza has led to reduced mobility and mortality. Due to lack of useful vaccines to prevent the infection of respiratory viruses, knowledge of etiology of viruses is vital to successful implementation of prevention, control and treatment strategies (75). There is therefore urgency to study the characteristics and the epidemiology of respiratory tract pathogen infections in developing countries with limited data (76). Causal link between viruses and association between mixed infections and increase in disease severity is also challenging (77).

Most models do not consider multiple viruses nor incorporate the time varying effects of these components (78). The extent of illness in children caused by relative contributions of different pathogens is not available(48). The log-linear, Poisson, binomial and logistic regression are widely used to analyze event count data in univariate models (71)(78)(79)(80). Currently, only a few methods address dynamic multiple time series of count data. Vector autoregressive

models are used to identify the lead, lag and contemporaneous relationships within and between time series. It is challenging to model these relationships using likelihood based methods. The lead and lag relationships of the within and between time series are specified using the spatial correlation structure (81). Jung *et al*. (2011) proposed a dynamic factor model for multivariate count time series that allows for serial correlation and idiosyncratic factors. This represents a non-trivial contemporaneous and temporal interaction across the series (82). The occurrence of different diseases in time contributes to multivariate time series data. For this, an integer-valued autoregressive model of order 1 for count data on bivariate time series was used (83).

Infectious surveillance data has been used to model age groups and geographical regions as different time series data for the multivariate case. The spatio-temporal dependence is considered in the later, in which case the minimum likelihood estimation was used in the model formulation. However, the covariance information is not obtained from their model (54). Models with a tendency of pollution causing respiratory disease were done for three disease categories. Available independent series of observations and covariates made it possible to model the fixed and random effects for between-series variation. The covariates included weather conditions and seasonal effects that depicted modeling factors with an acute effect on subjects (79). Jorgensen *et al* (80) analyzed daily visit counts for respiratory diseases where counts were categorized as asthma, bronchitis and ear infection, with the covariates such as temperature, maximum and minimum relative humidity.

In this paper, the surveillance data are aggregated by month and are not available at an individual level. This may lead to over-dispersion; hence the use of the negative binomial distribution. We model the contemporaneous relationship between pathogens, namely, RSV, ADENO, HMPV, Flu and PIV and the meteorological variables to investigate for serial

correlation. Held *et al*. suggested that environmental factors could be incorporated in the model if they do exist (54). The meteorological and seasonal variables are assumed to have an immediate effect on disease incidence. The models evaluate for the presence of heterogeneity in the autoregressive effect for the different pathogens and whether after adjusting for seasonality, an epidemic component could be isolated within or between the pathogens. The model helps in distinguishing between an endemic and epidemic component of the time series that would allow the separation of the regular pattern from irregularities and outbreaks.

4.3 Methods

4.3.1 Model formulation

We denote $\{y_{i,t}; i = 1, \ldots, I, t = 1, \ldots, T\}$ the multivariate time series of disease counts for the specific disease pathogens. Here *T* denotes the length of the time series and *I* denotes the number of pathogens monitored. The methods in this study are motivated by a branching process with immigration by Paul *et al* (84) where the model below for the multivariate time series of infectious disease counts is suggested ,

$$\boldsymbol{\mu}_t = \boldsymbol{\Lambda} \boldsymbol{y}_{t-1} + \boldsymbol{\nu}_t \ , \tag{4.1}$$

as the mean incidence which comprises of two additive components namely: an epidemic or autoregressive component $\boldsymbol{\Lambda} \boldsymbol{y}_{t-1}$, and an endemic component $\boldsymbol{\nu}_t$. The vectors $\boldsymbol{\mu}_t, \boldsymbol{\Lambda} \boldsymbol{y}_{t-1}$ and $\boldsymbol{\nu}_t$ are of length *m* and $\boldsymbol{\Lambda}$ is a *m* x *m* matrix with λ_i on the diagonal and elements $(\Lambda)_{ij} = \phi_i \omega_{ij}$ for $i \neq j$. Taking the variation of different pathogens into account, the inclusion of $\phi_i \sum \omega_{ij} \, y_{j,t-1}$ in the epidemic component leads to the model,

$$\boldsymbol{\mu}_t = \boldsymbol{\lambda}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{\phi}_{t-1}\boldsymbol{Z}_{t-1} + \boldsymbol{\eta}_t\boldsymbol{\upsilon}_t\,, \tag{4.2}$$

where $\boldsymbol{\phi}_{t-1}$ is the autoregressive effect of pathogen *j* on pathogen *i* and $\boldsymbol{\eta}_t$ corresponds to an offset and $\boldsymbol{Z}_{t-1} = \sum_{j\neq i}\omega_{ij}y_{j,t-1}$where, $\omega_{ij} = 1$ if pathogens *j* and *i* have an autoregressive effect on each other and 0 otherwise. The endemic component $\upsilon_{i,t}$ can be expressed as,

$$log\left(\upsilon_{i,t}\right) = \alpha_i + \left\{\sum_{s=1}^{S}\boldsymbol{\gamma}_s\, sin(\omega_s t) + \boldsymbol{\delta}_s\, cos(\omega_s t)\right\}, \tag{4.3}$$

where α_i is an intercept and the terms in curly brackets are the seasonal variation.

Letting $x_{k,t-1}$ denote climatic covariates with $\boldsymbol{\tau}_k$ coefficients in the model and $k = 1, \ldots, K$ covariates, then the conditional mean becomes

$$log(\boldsymbol{\mu}_t) = \boldsymbol{\lambda}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{\tau}_k x_{k,t-1} + \boldsymbol{\phi}_{t-1}\boldsymbol{Z}_{t-1} + \exp(\boldsymbol{\eta}_t)$$

4.3.2 Simulations study

The proposed model performance is investigated in this chapter on simulated data. We use both the frequentist and Bayesian approaches to compare the models with varying parameter estimates. The frequentist approach simulated multivariate data for five time series using package 'Surveillance' in R software as used earlier on by Held *et al*. in 2005(54) applied to model (4.2) above. The package uses the retrospective analysis of epidemic spread providing tools for visualization and simulation. Multivariate count time series models are estimated by 'hhh4' function as applied by Meyer and Held in 2016(84) and by Paul and Held in 2011(59).

The function uses the object of class 'disProg' that simulates the disease pathogen counts 10000 times. The absence of the over-dispersion parameter ψ_i indicates that Poisson distribution was assumed while its presence assumes the negative binomial distribution. In this study, both distributions are used and evaluated. The autoregressive parameter λ_i (the 'epidemic' component) was varied in the models to allow for the evaluation of whether the inclusion or exclusion of previous cases allowed for temporal dependence. The other autoregressive parameter ϕ_i was included in all the models for the adjacent pathogens where we assumed that all of them were correlated to each other and therefore the observations $y_{j,t-1}$ at previous time points were used for the autoregression. The linear trend was not used in the simulation in this study. The simulation returned a list with the following elements: *Data*, which is a 'disProObj' of simulated data, *mean* which is a matrix with mean $\mu_{i,t}$ used to simulate the data, *endemic* which is a matrix with the endemic part $\upsilon_{i,t}$ and *coefs* which is a list of all the parameters of the model. The simulated data was then used to fit the models for purposes of comparison. The simulated data is assumed to follow a Poisson distribution with the conditional mean shown in model 4.1 above. In the presence of over-dispersion, the Poisson model is replaced by a negative binomial model where the conditional mean remains unchanged but the variance δ is modified to $\mu_{i,t}\left(1+\ \mu_{i,t}\psi_i\right)$ with over-dispersion parameter $\psi_i > 0$ for every *i* –th virus.

The results of the simulation study as presented in Table 4.1 show the standard errors for the model parameter estimates and the measure of model goodness of fit. The model in Table 4.1 with the autoregressive parameter λ_i (model 3) and with $\psi_i \neq 0$ was the best with an AIC of 5037.81. This implies that the inclusion of previous cases $y_{i,t-1}$ allowed for temporal dependence. Model 1 with fixed $\psi_i = 0$, had its AIC value greater than that of model 3 and

that the presence of the parameter ψ_i in the best model indicates that there was over-dispersion.

Table 4.1 Simulation results including Parameter estimates, Standard errors and measure of model Goodness of Fit.

Parameter	Model1 ψ=0 $\lambda\neq0$	Model2 ψ=0 λ=0	Model3 $\psi\neq0$ $\lambda\neq0$	Model4 $\psi\neq0$ λ=0
ψ_1	-	-	15.7978 (0.0162)	9.8814 (0.0253)
ψ_2	-	-	21.0527 (0.0146)	12.1951 (0.0185)
ψ_3	-	-	62.8931 (0.0148)	29.4117 (0.0177)
ψ_4	-	-	7.8802 (0.0283)	4.2355 (0.0363)
ψ_5	-	-	13.6426 (0.0208)	2.6144 (0.0475)
λ_1	0.6174 (0.0341)	-	0.6187 (0.0476)	-
λ_2	0.4090 (0.0487)	-	0.4115 (0.0621)	-
λ_3	0.3562 (0.0624)	-	0.3875 (0.0674)	-
λ_4	0.5362 (0.0466)	-	0.5914 (0.0706)	-
λ_5	0.8634 (0.0336)	-	0.8273 (0.0428)	-
ϕ_1	0.1060 (0.0086)	0.1639 (0.0080)	0.0945 (0.0108)	0.2042 (0.0132)
ϕ_2	0.1491 (0.0125)	0.2519 (0.0073)	0.1409 (0.0151)	0.2373 (0.0102)

Parameter	Model1 ψ=0 $\lambda \neq 0$	Model2 ψ=0 λ=0	Model3 $\psi \neq 0$ $\lambda \neq 0$	Model4 $\psi \neq 0$ λ=0
$\boldsymbol{\phi_3}$	0.0842 (0.0130)	0.1626 (0.0075)	0.0870 (0.0138)	0.1678 (0.0088)
$\boldsymbol{\phi_4}$	0.0449 (0.0072)	0.1431 (0.0062)	0.0573 (0.0110)	0.1731 (0.0102)
$\boldsymbol{\phi_5}$	0.0000 (0.0000)	0.1268 (0.0060)	0.0000 (0.0000)	0.2183 (0.0153)
$\boldsymbol{AIC}$	5449.11	5984.87	5037.81	5463.39

4.3.3 Cointegration analysis

We explored for cointegration of the five time series using the Johansen procedure which allowed us to test whether the time series formed a cointegrating relationship. This test is due to Johansen (85) and summarized by QuantStart Team(86). This procedure assumes a vector autoregressive model of the form

$$\mathbf{X_t} = \boldsymbol{\mu} + \boldsymbol{A}_1\mathbf{X_{t-1}} + \cdots + \boldsymbol{A}_p\mathbf{X_{t-p}} + \mathbf{W_t} \quad , \tag{4.4}$$

where $\boldsymbol{\mu}$ is the vector-valued mean of the series, $\boldsymbol{A}_i$ are the coefficient matrices for each lag and $\mathbf{W_t}$ is the multivariate Gaussian noise.

A Vector Error Correction Model (VECM) can be formed by differencing the series in Eqn 4.4 as,

$$\Delta \boldsymbol{X}_t = \boldsymbol{\mu} + \boldsymbol{A}\mathbf{X_{t-1}} + \boldsymbol{\Gamma}_1\Delta \boldsymbol{X}_{t-1} + \cdots + \boldsymbol{\Gamma}_p\Delta \boldsymbol{X}_{t-p} + \mathbf{W_t}, \tag{4.5}$$

where $\Delta \boldsymbol{X}_t \coloneqq \mathbf{X_t} - \boldsymbol{X}_{t-1}$ is the differencing operator, $\boldsymbol{A}$ is the coefficient matrices for the first lag and $\boldsymbol{\Gamma}_i$ are the matrices for each differencing lag.

The test checks for the situation of no cointegration, which occurs when **A** = **0.** The Johansen test can check for multiple linear combinations of time series for forming stationary portifolios by carrying out an eigenvalue decomposition of **A**. The rank of the matrix **A** is given by *r* and the Johansen test sequentially tests if this rank is equal to zero, one, through $r = n - 1$, where *n* is the number of time series under test. The null hypothesis is $r = 0$ which means that there is no cointegration and a rank $r > 0$ indicates that a cointegrating relationship between two or more time series exist.

4.3.4 Bayesian analysis

The set of parameters, $\theta_i = \left(\lambda_i, \phi_i, \alpha_i, \gamma_{i,1}, \dots, \gamma_{i,s_i}, \delta_{i,1}, \dots, \delta_{i,s_i}\right)^T$, $\theta = (\theta_1, \dots, \theta_m, \psi_1, \dots, \psi_m)^T$ from the log-likelihood of the observation $y_{i,t}$ in model 4.2 and model 4.3 are given as

$$l(\theta) = \sum_{i,t} l_{i,t}(\theta_i, \psi_i), \tag{4.5}$$

and the likelihood as,

$$f\left(y_{i,t} \middle| \theta_i\right) = exp\left\{\sum_{i,t} l_{i,t}(\theta_i, \psi_i)\right\}, \tag{4.6}$$

where,

$$l_{i,t}(\theta_i, \psi_i) \propto log\ \Gamma\left(y_{i,t} + \frac{1}{\psi_i}\right) - log\,\Gamma\left(\frac{1}{\psi_i}\right) + \frac{1}{\psi_i} log\left(\frac{1}{1+\psi_i\,\mu_{it}(\theta_i)}\right) + y_{i,t} log\left(\frac{\psi_i\,\mu_{it}(\theta_i)}{1+\psi_i\,\mu_{it}(\theta_i)}\right), \quad (4.7)$$

and $\Gamma(.)$ is the gamma function and the dispersion parameters ψ_i for $i = 1, \ldots, m$. The gamma priors are assumed,

$$\psi_i \sim Ga(\alpha_{\psi,}\beta_\psi).$$

The pathogen dependent effects α_i are assumed to be independent and normally distributed with a large variance since our data is counts where n is large,

$$\alpha = (\alpha_1, \ldots, \alpha_I) \sim N(0,\sigma_\alpha^2 I),\ \sigma_\alpha^2 = 10^6,$$

where I is an identity matrix. Since all model parameters are non-negative we propose gamma prior distributions for them. The rate parameters λ_i, assumes independent gamma priors with gamma hyperpriors on the second parameter,

$$\lambda_i \sim Ga(\alpha_{\lambda,}\beta_\lambda) \text{ and } \beta_\lambda \sim Ga(a, b)$$

$$\text{using } \alpha_\lambda = 1, a = 10 \text{ and } b = 10.$$

Independent normal priors are assumed for γ_i and δ_i,

$$\gamma = (\gamma_1, \ldots, \gamma_I) \sim N(0,\sigma_\gamma^2 I),\ \sigma_\gamma^2 = 10^6,$$

$$\delta = (\delta_1, \ldots, \delta_I) \sim N(0,\sigma_\delta^2 I),\ \sigma_\delta^2 = 10^6.$$

The parameter ϕ_i assumes gamma priors, $\phi_i \sim Ga(\alpha_{\phi,}\beta_\phi)$.

The posterior distribution is therefore given as,

$$f(\theta_i|y_{i,t}) \propto f(y_{i,t}|\theta_i) f(\theta_i)\,,$$

which can be expressed as,

$$f(\theta_i|y_{i,t}) \propto exp\left\{\sum_{i,t} l_{i,t}(\theta_i, \psi_i)\right\} \times \prod_{s=1}^{s_i} e^{-\frac{1}{2c}\sigma^2_{\gamma_i}} \times \prod_{s=1}^{s_i} e^{-\frac{1}{2c}\sigma^2_{\delta_i}} \times \prod_{i=1}^{m} e^{-\frac{1}{2c}\sigma^2_{\alpha_i}}.$$

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were sufficiently used for ranking and selecting the best performing models. For the models done using Bayesian, the deviance information criterion (DIC) is more appropriate and was used.

4.4 Results

4.4.1 Exploratory Data Analysis (EDA)

The data consisted of monthly counts of health facility visits in Dadaab for respiratory viruses, namely: RSV, ADENO, HMPV, Flu and PIV as monthly cases obtained from September 2007 to August 2011. A combined plot of the five epidemics time profiles was analysed for similarities in timing and overall dynamics of the epidemics. We evaluate for the occurrence in the peaks of the epidemics. In our analysis, a time-evolution analysis of the epidemics is provided (Figure 4.1). In this figure we see that the incidence for some of the epidemic like Flu is lower compared to the other epidemics. Such low levels of incidence lead to sparse data, which can be problematic for accurate and reliable predictive modeling. Figure 4.2 shows a plot of time versus the disease counts along with a smoothing spline fit to the data. All the plots accentuate the need to include the nonlinearity effect in our models. The histograms indicate the disease counts are skewed to the left and therefore are not normally distributed. In this case, we know that we need to use non-Gaussian techniques to model this data. These insights are used to develop the mean function for these models. The correlation

plot in Figure 4.3 captures the strength of dependence.

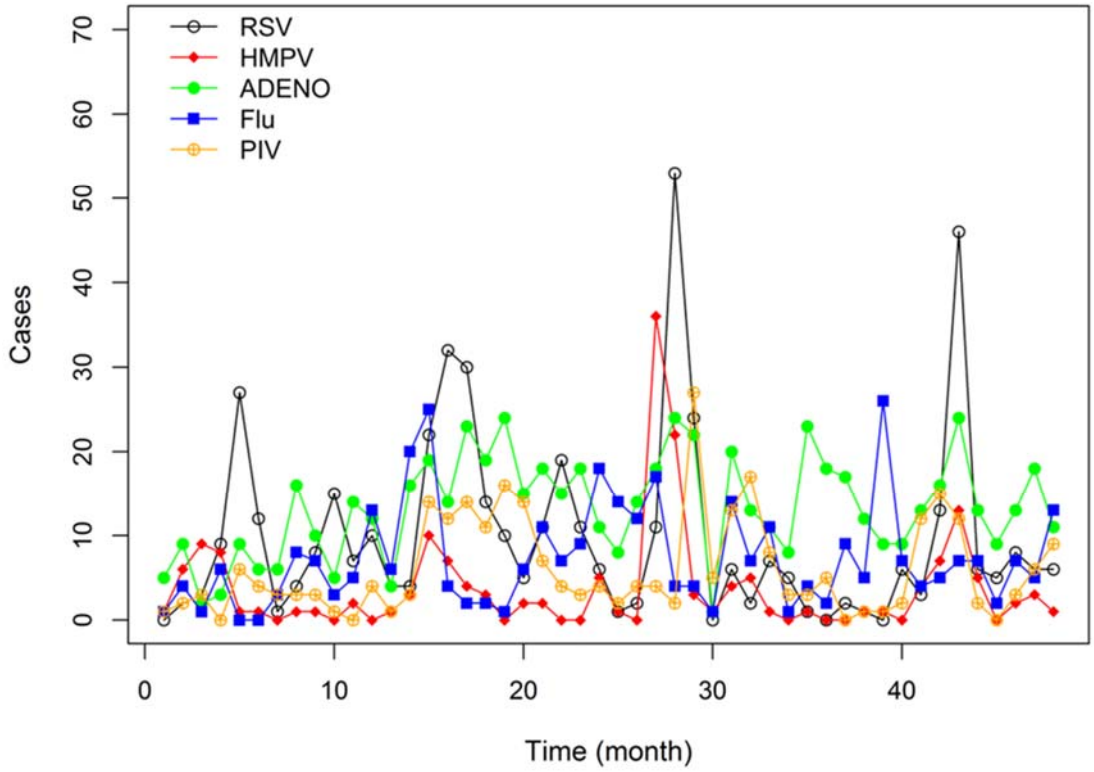
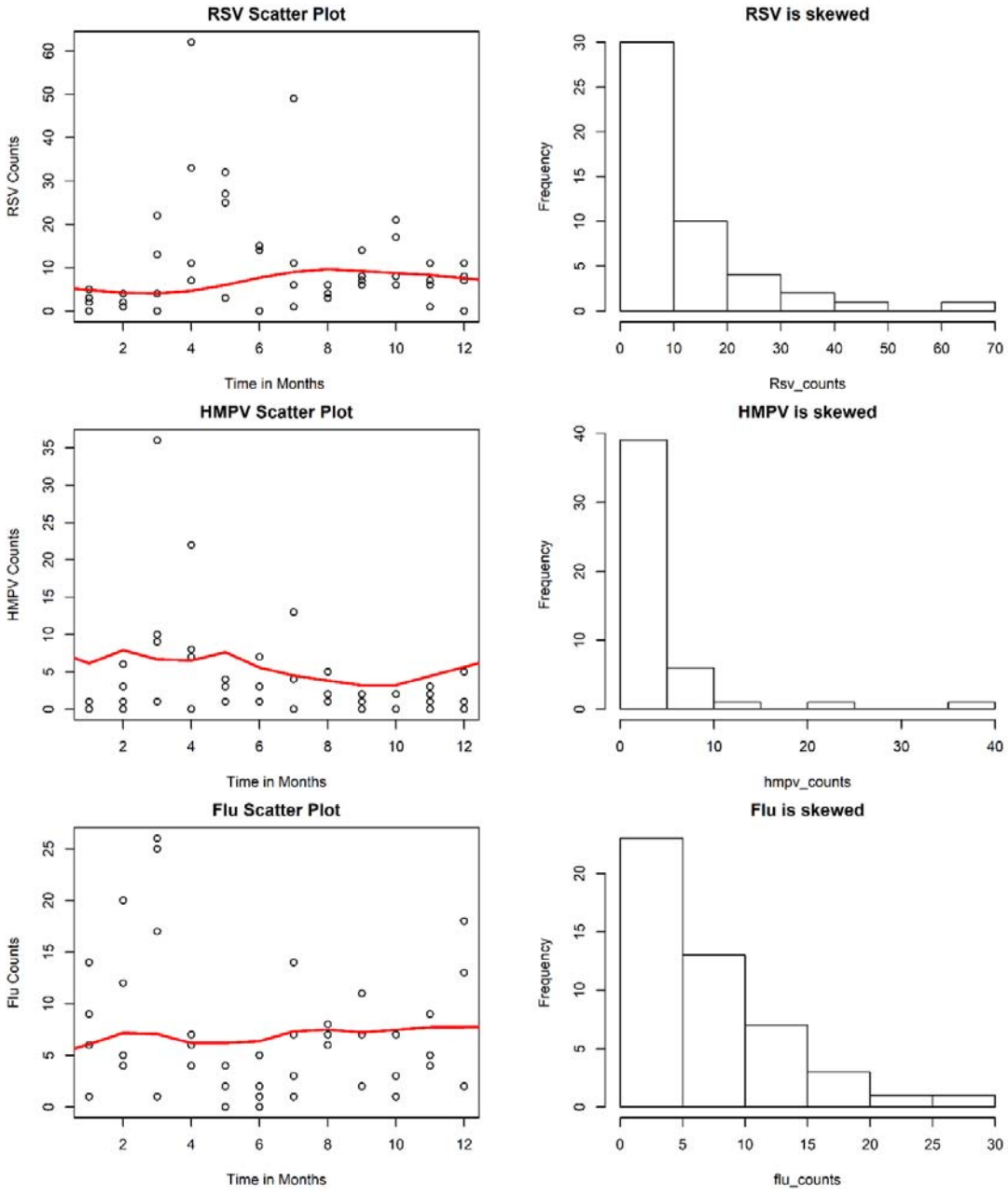


Figure 4.1 Time series plot of the epidemics. Overall, the number of HMPV cases were the lowest compared to all the other epidemics. There is no synchrony in the time of occurrence of the epidemic peaks. The intensity of the few peaks shows variations in both amplitude and frequency.



1234

1235 **Figure 4.2 Scatter plots and histograms for RSV, HMPV and Flu counts. All are skewed**
1236 **to the right (the red solid lines denote the fitted curves using smoothing splines)**

1237

1238 There is a significant positive dependence between RSV and HMPV disease pathogens.

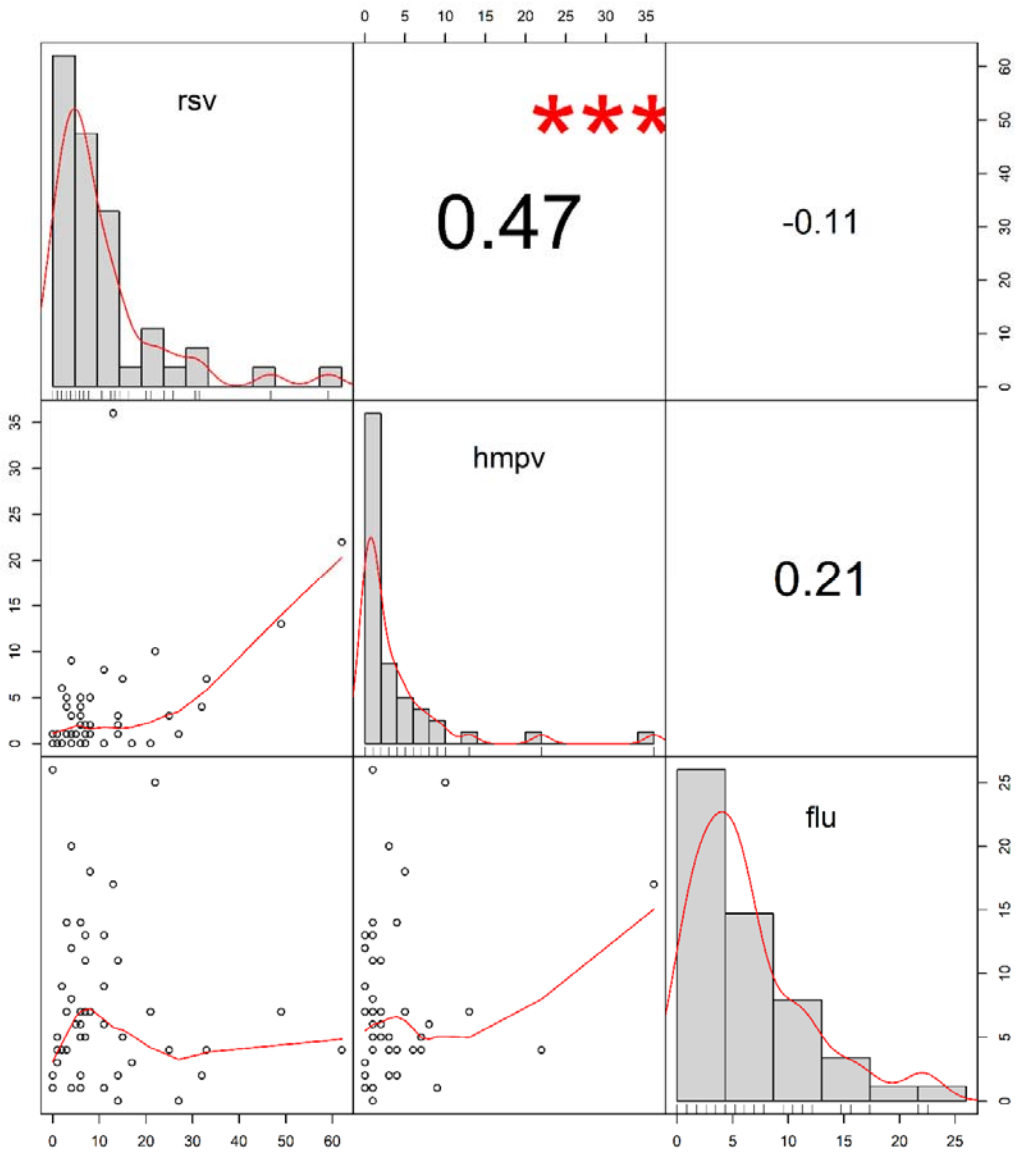


Figure 4.3 Correlation matrix and marginal distribution of the disease counts. Signif. codes: $p < 0$ '*'**

In order to avoid spurious correlation in fitting the time series models, we evaluated for cointegration between them as seen in Table 4.2.

Table 4.2 Values of test statistic and critical values of cointegration tests.

(a)	Test Statistic	Significance levels			(b)	Test Statistic	Significance levels		
		10%	5%	1%			10%	5%	1%
r<= 4	7.72	6.5	8.18	11.65					
r<= 3	20.47	15.66	17.95	23.52					
r<= 2	39.13	28.71	31.52	37.22	r <= 2	14.11	6.5	8.18	11.65
r<= 1	66.25	45.23	48.28	55.43	r <= 1	34.04	15.66	17.95	23.52
r= 0	98.2	66.49	70.6	78.87	r= 0	56.61	28.71	31.52	37.22

This was done by sequentially carrying out the hypothesis tests beginning with the null hypoothesis of $r = 0$ versus the alternative hypothesis of $r > 0$. From Table 4.2(a), there was clear evidence to reject the null hypothesis at the 1% level (98.20> 78.87) and we could likely conclude that $r > 0$. Similarly when we carried out the $r \leq 1$ null hypothesis versus the $r > 1$ alternative hypothesis we had sufficient evidence to reject the null hypothesis at the 1% level (66.25> 55.43) and could conclude $r > 1$. Similarly when we carried out the $r \leq 2$ null hypothesis versus the $r > 2$ alternative hypothesis we had sufficient evidence to reject the null hypothesis at the 1% level (39.13 > 37.22) and could conclude $r > 2$. However, for the $r \leq 3$ hypothesis we could only reject the null hypothesis at the 5% level (20.47 > 17.95). This was a weaker evidence than the previous hypotheses and, although it suggested we could reject the null at this level, *r* might equal three, rather than exceeded three. What this meant, was that it may be possible to form a linear combination with only three pathogens rather than requiring all five to form a cointegrating portfolio. Upon testing the various combinations of pathogens, a set consisting of RSV, HMPV and Flu was not cointegrated. Table 4.2 (b) shows

that when we carried out the $r \leq 2$ null hypothesis versus the $r > 2$ alternative hypothesis we had sufficient evidence to reject the null hypothesis at the 1% level (14.11 > 11.65).

4.4.2 Model Results

We extended the analysis of the monthly incidences adjusting for seasonality and checked whether an epidemic component could be isolated within or between this set of three pathogens that were not cointegrated using a Bayesian approach. After fitting the possible combinations of the pathogen interactions as is seen in Table 4.3, model 4 that includes the influence of HMPV on RSV and the influence of Flu on RSV with the rest of the interactions equal to zero yielded a DIC value of 817.969.

Table 4.3 Models for different interaction combinations of the three pathogens, RSV, HMPV and Flu. The symbols "–" and "√" mean the absence and presence of interactions, respectively and measure of model Goodness of Fit (DIC).

Model	RSV→HMPV	HMPV→RSV	RSV→Flu	Flu→ RSV	HMPV →Flu	Flu→HMPV	DIC
1	–	–	–	–	–	–	829.102
2	√	√	–	–	–	–	822.606
3	√	–	√	–	–	–	827.406
4	√	–	–	√	–	–	817.969
5	√	–	–	–	√	–	831.144
6	√	–	–	–	–	√	821.404

Model	RSV→HMPV	HMPV→RSV	RSV→Flu	Flu→ RSV	HMPV →Flu	Flu→HMPV	DIC
7	–	√	√	–	–	–	830.879
8	–	√	–	√	–	–	832.900
9	–	√	–	–	√	–	832.917
10	–	√	–	–	–	√	833.097
11	–	–	√	√	–	–	830.361
12	–	–	√	–	√	–	825.633
13	–	–	√	–	–	√	830.895
14	–	–	–	√	√	–	827.758
15	–	–	–	√	–	√	830.973
16	√	√	√	–	–	–	819.647
17	√	–	√	√	–	–	819.511
18	√	–	–	√	√	–	821.346
19	√	–	–	–	√	√	821.651
20	–	√	√	√	–	–	832.529
21	–	√	–	√	√	–	832.807
22	–	√	–	–	√	√	832.961

Model	RSV→HMPV	HMPV→RSV	RSV→Flu	Flu→ RSV	HMPV →Flu	Flu→HMPV	DIC
23	−	−	√	√	√	−	827.288
24	−	−	√	−	√	√	827.647
25	−	−	−	√	√	√	827.942
26	√	√	√	√	−	−	821.193
27	√	−	√	√	√	−	819.503
28	√	−	−	√	√	√	821.418
29	−	√	√	√	√	−	832.333
30	−	√	−	√	√	√	832.702
31	−	−	√	√	√	√	827.602
32	√	√	√	√	√	√	821.475

Compared to the others, this model yielded the smallest value and therefore resulted to be the best. In our study we have shown that the incidence of RSV peaks influence those of HMPV and consequently, the incidence of Flu peaks influence those of RSV as shown from the best model fit. This intuitively implies that incidence of Flu peaks influence those of HMPV. The model shows that the interactions that best describes the best model are those of RSV with HMPV and Flu with RSV. We further fitted the best model with including the climatic factors. The model with the climatic factors yielded smaller DIC value of 759.219. This

implied the need to include the climatic factors to improve the fit. The model parameter estimates for the best model are shown in Table 4.4.

Table 4.4 Posterior median and point-wise 95% credibility intervals for the best model.

Parameter	5%	median	95%
alpha$_{(RSV)}$	-6.079	-3.902	-3.047
alpha$_{(HMPV)}$	-8.226	-5.106	-4.194
alpha$_{(Flu)}$	-5.257	-4.689	-4.129
beta$_{(RSV)}$	0.9619	1.401	2.038
beta$_{(HMPV)}$	0.6685	1.063	1.649
beta$_{(Flu)}$	0.8314	1.353	1.947
delta$_{(RSV)}$	-6.302	-3.577	-2.344
delta$_{(HMPV)}$	-6.556	-3.737	-2.282
delta$_{(Flu)}$	-4.699	-2.555	-1.56
gamma$_{(RSV)}$	-5.437	-3.883	-1.677
gamma$_{(HMPV)}$	-11.49	-2.38	-0.1207
gamma$_{(Flu)}$	-5.073	-2.31	-0.7587
Psi$_{(RSV)}$	0.1564	5.257	108.1
psi$_{(HMPV)}$	0.1525	4.448	102.0
psi$_{(Flu)}$	0.4901	7.574	113.3

They are the posterior median with their corresponding 95% credibility intervals. It is shown that the over-dispersion parameter ψ_i for every *i –th* virus from Table 4.1 is greater than zero for the best model. This guarantees the use of the negative binomial distribution rather than

the Poisson distribution that assume $\psi_i = 0$. Figure 4.4 shows the scatter plots of realized vs. posterior predictive values for RSV, HMPV and flu from the best model fit measuring the discrepancies between observed and predictive values.

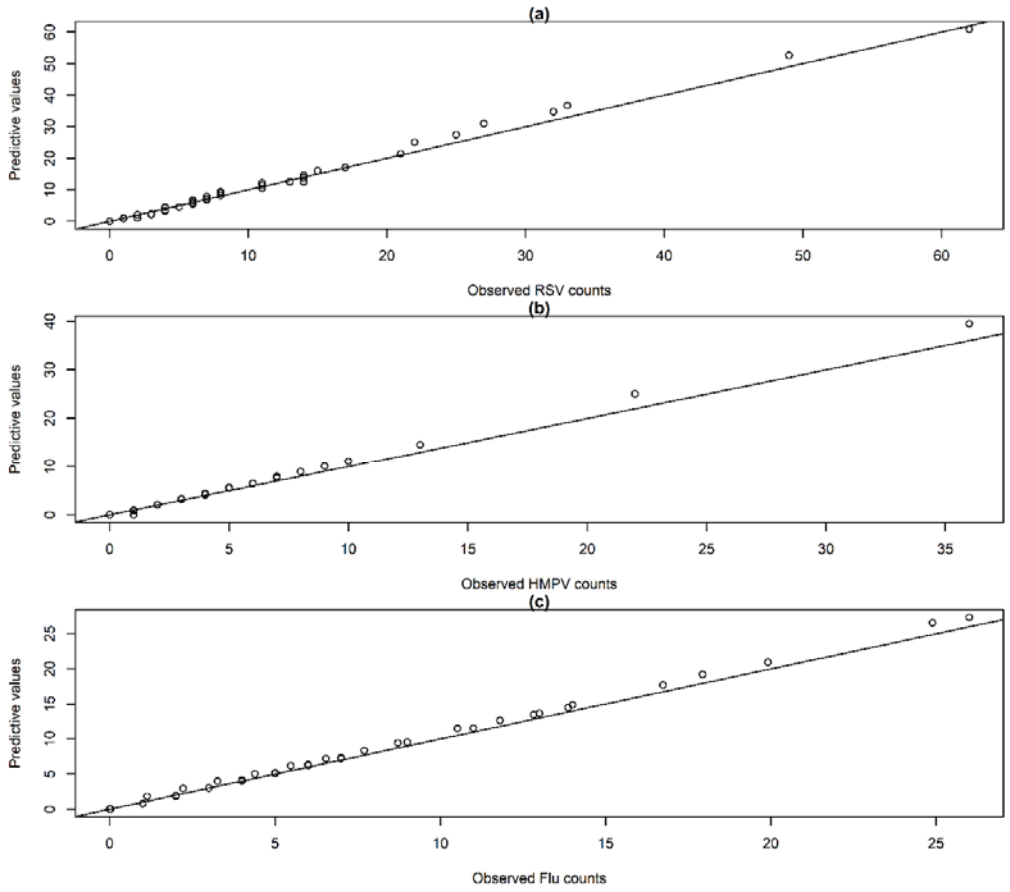


Figure 4.4 Realised vs. Posterior Predictive Values of RSV, HMPV and Flu Disease Counts for the best model

As can be seen, there was some systematic difference between the realized and posterior predictive values but this was the best fit among the models fitted.

4.5 Discussion

Common causes of ARI are respiratory viruses. The RSV and influenza are associated with an increased number of hospitalizations. In the tropical regions, there is an expected increase during the rainy seasons. Comprehensive insight into the recent state of Flu-A and Flu-B viruses was discussed by Matheka *et al*. (87). In their work, they explored the then existing epidemic patterns, discussed challenges associated to combating the epidemic and suggested intervention and control measures particularly in East Africa. The RSV has been associated with climatic factors and has been found to have biennial peaks(64). Some previous work has been done to show that there is interaction between RSV and influenza(88)(89). In their work they were able to show that the first peak of ARI was explained by one pathogen while the next peak was associated with the other pathogen. Alternating patterns of RSV and influenza have been reported in San Luis Potosi(88). These variations in patterns at the same time could be due to the presence of influenza and RSV occurrence(89). Poisson regression models were used for the analysis of time series data for RSV and influenza(90)(91).The two pathogens were found to be seasonally related and it was difficult to disentangle one from the other. Analysis of RSV from the effect of season and influenza infection was mentioned to have begun(90). A recent study in the United States that used seasonal variation in the analysis of time trends of deaths showed that RSV contribution to mortality was less than that by influenza(91). Viral interference has been argued to be there between influenza and RSV and this interference has been seen among viruses by many studies(92). Arguedas-Flatts et. al (92) in their paper, used a two pathogen epidemic model to study the interaction of influenza and RSV. In our study, we extended this bivariate modeling to include HMPV pathogen and aimed at understanding the multiple interactions among the three pathogens.

The HMPV and RSV has shown to cause similar clinical severity and symptoms with similar seasonality(65). Kim *et al*. (66) who investigated the epidemiological and clinical assessment of HMPV and RSV in Seoul, Korea, 2003-2008 had similar findings. Cuevas *et al*. (43) observed the increase in HMPV incidence as RSV incidence increased. Co-infections of RSV and HMPV were identified in a study in Yemeni among children younger than 2 years which showed seasonal variations of HMPV and RSV with peaks of RSV in December and January and for HMPV in February and March (67).

The models used in our study have been able to capture the serial correlation between the three pathogens namely, RSV, HMPV and influenza. Our results show that there is presence of interference between the three viruses which lead to occurrence of sequential peaks supporting for superinfection.

Some limitations to this study include the short time series of four years monthly data points of infected counts. The model used did not consider all the possible three pathogen combinations for serial correlations among those five available in the surveillance data. We did not adjust for the climatic variables as covariates in our models because this would have made them complex to evaluate. We would recommend the use of more time series data and adjustments with climatic variables as covariates to be used in the future research to help in the understanding of these interactions.

4.6 Conclusion

Whereas seasonal influenza, RSV and HMPV have long been recognised as causes of mobility and mortality in countries with temperate climates, recent studies have shown that only influenza has a vaccine to prevent its advances and control outbreaks. This study has

used data and models that could be useful to detect outbreaks of these viruses and for developments of effective vaccines. The model could help in managing the occurrence of outbreaks by use of flu vaccines. Due to the interactions of the three viruses studied in the model, vaccinating against influenza subsequently reduces HMPV and RSV infections.

4.7 Acknowledgments

The authors wish to acknowledge the CDC Kenya Refugee Health Program for their tireless work in the surveillance in the refugee camp and their assistance with the data collection and management.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

5.1 Introduction

Globally, a large number of human respiratory tract infections are associated with viruses including RSV, HMPV and influenza. In this final chapter, we present the synthesis and overview of the key findings of the studies included in this thesis. In this chapter we highlight the objectives, discuss the challenges of the studies, make recommendations for further research that is needed, and finalise with the general conclusion.

5.2 Aims and objectives

The aim of our models was to study any relationships of influenza with other pathogens so that vaccinating against it would help reduce the spread of the others. The main objective of this thesis was to better understand the relationship between climate and RSV, HMPV and influenza in making reliable predictions of their incidence and to establish good surveillance systems in developing countries to help understand the dynamics of the disease. Specifically our objectives were;

i. To explore the best model that predicts the relationship between RSV incidence and climatic factors along spatio-temporal scales to determine whether a seasonal pattern of RSV infection exists.
ii. To investigate the relationship between RSV and HMPV while adjusting for climatic factors.
iii. To establish the heterogeneity in the autoregressive effect to account for the influence between RSV, HMPV and Influenza viruses.
iv. To assess the presence of influence of high incidences between these viruses and

whether higher incidences of one virus are influenced by another and to investigate for serial correlation between them.

These objectives were addressed in chapter 2 to chapter 4 of this thesis.

5.3 Challenges

Lack of adequate surveillance data remains a challenge to this study. Collection of more data would help in the future to model outbreak detection. Most models have not considered multiple viruses nor incorporated the time varying effects. The extent of illness in children caused by relative contributions of different pathogens is not available. The log-linear, Poisson, binomial and logistic regression are widely used to analyze event count data in univariate models and currently, only a few methods address dynamic multiple time series of count data. Furthermore, variations in climatic factors, such as humidity, temperature, wind speed and rainfall can have a significant impact on disease dynamics. These climatic factors are essential in evaluating ARI for equatorial climatic regions to aid accurate predictions of their outbreaks.

5.4 Future possibilities

Modeling of these viruses requires more data to enable the estimation of predictions to detect outbreaks so that these could be controlled in a timely manner. We would recommend the use of more time series data and adjustments with climatic variables as covariates to be used in the future research to help in the understanding of these interactions. Exploring zero inflated models with the Poisson GAM models that we fitted would be recommended.

5.5 Final comments and summary conclusions

We provided a comprehensive comparison of RSV and HMPV in a refugee camp setting by using a bivariate non-Gaussian model to jointly model the epidemics. By comparing various model structures, we identified a model that satisfactorily fits the epidemic data, thereby explaining most of the observed variation therein. The models and estimated parameters also provided clues into the dynamics and stability of the three epidemics. The modelling of the time series events of these viruses helped in the prediction of their outbreaks but also in estimating which outbreaks preceded each other. Due to lack of useful vaccines to prevent the infection of respiratory viruses, knowledge of etiology of viruses is vital to successful implementation of prevention, control and treatment strategies.The results could also be used by other countries in the tropical zone in Africa with similar settings to inform vaccination timings as control measures to prevent outbreaks.

Our results demonstrated the influence of RSV on HMPV while adjusting for climatic factors. The climatic factors played a significant role in explaining the influence of RSV incidence on HMPV incidence. These models are important to the public health implication since controlling the incidence of RSV would consequently reduce the incidence of HMPV. The models could help in managing the occurrence of outbreaks by use of flu vaccines. Due to the interactions of the three viruses studied in the model, vaccinating against influenza subsequently could reduce HMPV and RSV infections.

Analysis codes

```
library(tsModel);
library(gam);
library(rJava);
```

```
library(glmulti);
library(lattice);
library(TSA);
library(DAAG);
library(mgcv);
library(forecast);
library(splines)
library(modEvA);
install.packages("modEvA", repos="http://R-Forge.R-project.org")
##---------------------------------------------------------------
## load survey data
windtimeseries <- ts(RSV$wind,frequency=12, start=c(2007,9)) # wind tmeseries
t = 1:length(windtimeseries); # timevector
rainfalltimeseries <- ts(RSV$rainfall,frequency=12, start=c(2007,9)) # rainfall timeseries
Temptimeseries  <- ts(RSV$Temp,frequency=12, start=c(2007,9))    # Temp timeseries
dewtimeseries   <- ts(RSV$dew,frequency=12, start=c(2007,9))     # dew  timeseries
VISIBtimeseries <- ts(RSV$VISIB,frequency=12, start=c(2007,9))   # VISIB timeseries
RSV_proptimeseries <- ts(RSV$prop,frequency=12, start=c(2007,9)) # RSV incidence
timeseries
RSV_proptimeseriescomponents <- decompose(RSV_proptimeseries)
RSV_postimeseries <- ts(RSV$rsv_pos,frequency=12, start=c(2007,9)) #RSV events
timeseries
setwd("C:\\Data\\Course\\PhD\\Analysis\\Doc\\Papers\\Paper 1")
#------------------------- STL decomposition-----------------------------------
 RSV.stl   = stl(RSV_postimeseries, s.window="periodic")  # RSV
```

```
1477    wind.stl  = stl(windtimeseries, s.window="periodic")     # wind
1478    rain.stl  = stl(rainfalltimeseries, s.window="periodic") # rainfall
1479    temp.stl  = stl(Temptimeseries, s.window="periodic")     # temperature
1480    dew.stl   = stl(dewtimeseries, s.window="periodic")      # dew
1481    visib.stl = stl(VISIBtimeseries, s.window="periodic")    # precipitation
1482
1483  tiff("S1_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1484    plot(RSV.stl); dev.off() # S1_Fig
1485  tiff("S2_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1486    plot(wind.stl); dev.off()  # S2_Fig
1487  tiff("S3_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1488    plot(rain.stl); dev.off()  # S3_Fig
1489  tiff("S4_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1490    plot(temp.stl); dev.off()  # S4_Fig
1491  tiff("S5_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1492    plot(dew.stl); dev.off()  # S5_Fig
1493  tiff("S6_Fig.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
1494    plot(visib.stl); dev.off() # S6_Fig
1495
1496  rsv_events <- as.vector(RSV_postimeseries) # RSV incidence
1497  wnd  <- as.vector(windtimeseries) # wind
1498  rain <- as.vector(rainfalltimeseries) # rainfall
1499  tmp  <- as.vector(Temptimeseries) # temperature
1500  dew  <- as.vector(dewtimeseries) # dew
1501  precipitation <- as.vector(VISIBtimeseries) # precipitation
```

```
##----------------plot Figure 2.A to F below------------------------------
tiff("Figure2.tiff", res=600, compression = "lzw", height=7, width=8, units="in")
op02 <- par(mfrow = c(2,2), oma = c(2,2,0,0) + 0.17, mar = c(4,4,1,1) + 0)
plot(wnd,rsv_events,pch=16,xlab="Wind speed (knots)",ylab="RSV cases",col="gray20",
    ylim=c(0,70)) # wnd and RSv events
  model_rsv_wnd <- lm(rsv_events ~ wnd); abline(model_rsv_wnd,col="blue",lwd=2)
     text(9,65,"A (??= -0.4651, p= 0.001)")
plot(tmp,rsv_events,pch=16,xlab= expression(paste("Temperature, ",degree,"F")),
    ylab="RSV cases",col="gray20", ylim=c(0,70)) # temp and RSv events
  model_rsv_tmp <- lm(rsv_events ~ tmp); abline(model_rsv_tmp,col="blue",lwd=2)
     text(86,65,"B (??=0.1850, p=0.224)")
plot(dew,rsv_events,pch=16,xlab=expression(paste("Dew, ",degree,"F")),
    ylab="RSV cases",col="gray20",
    ylim=c(0,70)) # dew and RSv events
  model_rsv_dew <- lm(rsv_events ~ dew); abline(model_rsv_dew,col="blue",lwd=2)
     text(72,65,"C (??=0.230, p=0.128)")
  h1 <- mean(wnd); k1 <- max(tmp); a1 <- -1
  yfit<- nls(tmp ~ a*(wnd - h)^2 + k, start = list(a = a1, h = h1, k = k1))
  yfitParm <- summary(yfit)$para[,1]; cbind(h1,k1,a1)
  ymod <- yfitParm[1]*(wnd - yfitParm[2])^2 + yfitParm[3];
  tmp.fitted <- (summary(yfit)$coef[1])*(wnd - (summary(yfit)$coef[2]))^2 +
summary(yfit)$coef[3];
plot(tmp.fitted ~ wnd,xlab="Wind (knots)",ylab=expression(paste("Temperature,
",degree,"F")),
```

```
1527        col="blue",lty=1,pch=16, ylim=c(78,90)) #
1528    plot(wnd,predict(yfit),col="black",type="l",pch=1)
1529      lines(wnd,tmp,col="gray20",lwd=2,pch=16,type="p") # wind and temperature
1530      text(10,89,"D (p=0.002)")
1531    par(op02); dev.off()
1532
1533    ##------------------------set variables for model fit ----------------------------
1534    ###  RSV:
1535      RSV.s <- ts.union(RSVts=RSV.stl$time.series[,1])
1536      RSV.t <- ts.union(RSVts=RSV.stl$time.series[,2])
1537      RSV.r <- ts.union(RSVts=RSV.stl$time.series[,3])
1538      # par(mfrow=c(3,1)); plot(RSV.s, main="Seasonal"); plot(RSV.t, main="Trend");
1539    plot(RSV.r, main="Remainder")
1540    ##  rainfall:
1541      rain.s <- ts.union(raints=rain.stl$time.series[,1])
1542      rain.t <- ts.union(raints=rain.stl$time.series[,2])
1543      rain.r <- ts.union(raints=rain.stl$time.series[,3])
1544      # par(mfrow=c(3,1)); plot(rain.s, main="Seasonal"); plot(rain.t, main="Trend"); plot(rain.r,
1545    main="Remainder")
1546    ##  wind:
1547      wind.s <- ts.union(windts=wind.stl$time.series[,1])
1548      wind.t <- ts.union(windts=wind.stl$time.series[,2])
1549      wind.r <- ts.union(windts=wind.stl$time.series[,3])
1550      # par(mfrow=c(3,1)); plot(wind.s, main="Seasonal"); plot(wind.t, main="Trend");
1551    plot(wind.r, main="Remainder")
```

```
1552  ##  temperature:
1553    temp.s <- ts.union(tempts=temp.stl$time.series[,1])
1554    temp.t <- ts.union(tempts=temp.stl$time.series[,2])
1555    temp.r <- ts.union(tempts=temp.stl$time.series[,3])
1556    # par(mfrow=c(3,1)); plot(temp.s, main="Seasonal"); plot(temp.t, main="Trend");
1557  plot(temp.r, main="Remainder")
1558  ##  dew:
1559    dew.s <- ts.union(dewts=dew.stl$time.series[,1])
1560    dew.t <- ts.union(dewts=dew.stl$time.series[,2])
1561    dew.r <- ts.union(dewts=dew.stl$time.series[,3])
1562    # par(mfrow=c(3,1)); plot(dew.s, main="Seasonal"); plot(dew.t, main="Trend"); plot(dew.r,
1563  main="Remainder")
1564  ##  visib: precipitation
1565    visib.s <- ts.union(visibts=visib.stl$time.series[,1])
1566    visib.t <- ts.union(visibts=visib.stl$time.series[,2])
1567    visib.r <- ts.union(visibts=visib.stl$time.series[,3])
1568    # par(mfrow=c(3,1)); plot(visib.s, main="Seasonal"); plot(visib.t, main="Trend");
1569  plot(visib.r, main="Remainder")
1570
1571
1572  ## ---------- Correlation analsys: model fits----------
1573  # plot(wnd,rsv_events,pch=16,col="gray20",xlab="Wind speed", ylab="RSV
1574  incidence",main="A") # wnd and RSv events
1575  model_rsv_wnd <- lm(rsv_events ~ wnd); summary(model_rsv_wnd)
1576  cor.test(rsv_events,wnd)
```

```
1577  # abline(model_rsv_wnd,col="black"); model_rsv_wnd
1578  model_rsv_tmp <- lm(rsv_events ~ tmp); summary(model_rsv_wnd)
1579  cor.test(rsv_events,tmp)
1580  model_rsv_dew <- lm(rsv_events ~ dew ); summary(model_rsv_dew )
1581  cor.test(rsv_events,dew )
1582  model_tmp_wnd <- lm(tmp ~ wnd); summary(model_tmp_wnd)
1583  cor.test(tmp ,wnd)
1584  ## ------------------------------------------------------------------------
1585  ## Dadaab RSV Poisson GLM model
1586  probRSVti <- RSV$prop ; # incidence
1587  #RSV$rsv_pos <- cbind(RSV$rsv_pos,(RSV$pop - RSV$rsv_pos));
1588  # plot(t,probRSVti,pch=16,col="gray20",xlab="Time",ylab="incidence")
1589
1590  ##-------- no-decomposed covariates----------------------------------------------
1591  x1 <- as.vector(windtimeseries)     # wind (x1)
1592  x2 <- as.vector(rainfalltimeseries) # rainfall (x2)
1593  x3 <- as.vector(Temptimeseries)     # temp (x3)
1594  x4 <- as.vector(dewtimeseries)      # dew (x4)
1595  x5 <- as.vector(VISIBtimeseries)    # visib/precipitation (x5)
1596  ##-------- decomposed covariates------------------------------------------------
1597  x1S <- as.vector(wind.s); # windtimeseriescomponents$seasonal; (x1S)
1598  x1T <- as.vector(wind.t); # windtimeseriescomponents$trend; (x1T)
1599  x1R <- as.vector(wind.r); # windtimeseriescomponents$random; (x1R)
1600  x2S <- as.vector(rain.s); # rainfalltimeseriescomponents$seasonal; (x2S)
1601  x2T <- as.vector(rain.t); # rainfalltimeseriescomponents$trend; (x2T)
```

```
x2R <- as.vector(rain.r); # rainfalltimeseriescomponents$random; (x2R)
x3S <- as.vector(temp.s); # Temptimeseriescomponents$seasonal; (x3S)
x3T <- as.vector(temp.t); # Temptimeseriescomponents$trend; (x3T)
x3R <- as.vector(temp.r); # Temptimeseriescomponents$random; (x3R)
x4S <- as.vector(dew.s); # dewtimeseriescomponents$seasonal; (x4S)
x4T <- as.vector(dew.t); # dewtimeseriescomponents$trend; (x4T)
x4R <- as.vector(dew.r); # dewtimeseriescomponents$random; (x4R)
x5S <- as.vector(visib.s); # VISIBtimeseriescomponents$seasonal; (x5S)
x5T <- as.vector(visib.t); # VISIBtimeseriescomponents$trend; (x5T)
x5R <- as.vector(visib.r); # VISIBtimeseriescomponents$random; (x5R)
RSVcases.s <- as.vector(RSV$prop); # proportion of RSV cases, response.
cosA <- as.numeric(cos(2*pi*(1/12)*t)); sinA <- as.numeric(sin(2*pi*(1/12)*t));
##-------------------------------------------------------------------------------------
# par(mfrow=c(1,1)); plot(cosA,type="l",col="black",xlab="Time index",ylab="Amplitude")
# lines(sinA,type="l",col="red")

Dsquared <- function(binglm.1a, adjust = FALSE) {
# calculates the explained deviance of a GLM
# model: a model object of class "glm"
# adjust: logical, whether or not to use the adjusted deviance taking into acount the nr of
observations
# and parameters (Weisberg 1980; Guisan & Zimmermann 2000)
  d2 <- (binglm.1a$null.deviance - binglm.1a$deviance) / binglm.1a$null.deviance
  if (adjust) {
    n <- length(binglm.1a$fitted.values);
```

```
1627      p <- length(binglm.1a$coefficients)
1628      d2 <- 1 - ((n - 1) / (n - p)) * (1 - d2)}
1629    return(d2)} # end Dsquared function
1630
1631  ## Data exploratory Analysis
1632  data2<-data.frame(RSV$prop,RSV$wind,RSV$Temp,RSV$rainfall,RSV$dew,RSV$VISIB)
1633    attach(data2)
1634    data2.ts <- ts(data2,frequency=12, start=c(2007,9)); plot(data2.ts)
1635  par(mfrow=c(2,2))
1636    cc1=ccf(RSV.wind,RSV.prop);cc2=ccf(RSV.rainfall,RSV.prop)
1637    cc3=ccf(RSV.Temp,RSV.prop);cc4=ccf(RSV.dew,RSV.prop)
1638    cc5=ccf(RSV.VISIB,RSV.prop)
1639    cc1; cc2; cc3; cc4; cc5
1640  library(astsa); library("dlnm")
1641    lag2.plot(RSV.rainfall,RSV.prop,11)
1642    lag2.plot(RSV.wind,RSV.prop,11)
1643    lag2.plot(RSV.Temp,RSV.prop,11)
1644    lag2.plot(RSV.dew,RSV.prop,11)
1645    lag2.plot(RSV.VISIB,RSV.prop,11)
1646  ## Define the function LAG
1647  LAG <- function(x,k){n = length(x)
1648    xx=x
1649    xx[1:(n-k)]=x[(k+1):n]
1650    xx[(n-k+1):n]=NA
1651    xx}
```

```
z=1:20
LAG(z,2)
lagpad <- function(x, k) {
    c(rep(NA, k), x)[1 : length(x)]
}
z=1:20
lagpad(z,2)

RSV$lag_yt_1<-lagpad(RSV$rsv_pos,1)
RSV$lag_yt_2<-lagpad(RSV$rsv_pos,2)
library(MASS)

##Choosing the best glm Poisson models to be compared using the "glmulti" package
library(glmulti)

posglmulti.1a <- glmulti(rsv_pos ~
lag_yt_1 + x1 + x2 + x3 + x4 + x5 + cosA + sinA + offset(log(RSV$pop/1000)),# best model
without decomposition of the incidence rate per 1000 children
data=RSV, family = poisson,level = 1, method = "h", crit = "aic",
confsetsize = 5,plotty = F, report = F)
summary(posglmulti.1a)
posglmulti.1a@formulas
summary(posglmulti.1a@objects[[1]])


```

```
posglmulti.1b <- glmulti(rsv_pos ~
lag_yt_1 + x1S + x1T + x2S + x2T + x3S + x4S + x4T + x5S + x5T + cosA + sinA  +
offset(log(RSV$pop/1000)), # best model with decomposition
data=RSV, family = poisson,level = 1, method = "h", crit = "aic", #rsv_pos ~ 1 + x1S + x1T +
x2S + x4T + x5T + cosA + sinA
confsetsize = 5,plotty = F, report = F)
summary(posglmulti.1b)
posglmulti.1b@formulas
summary(posglmulti.1b@objects[[1]])


## Dadaab RSV Poisson GLM model ====== MODEL 3.a
=========================================
## Testing for max lag for each of the covariates in the GLM models without decomposition
selected above
LAGG=4
posglm.x1<-glm(rsv_pos ~ LAG(x1,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x1)

LAGG=3
posglm.x2<-glm(rsv_pos ~ LAG(x2,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x2)

```

```
LAGG=2
posglm.x3<-glm(rsv_pos ~ LAG(x3,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x3)

LAGG=1
posglm.x4<-glm(rsv_pos ~ LAG(x4,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x4)


LAGG=3
posglm.x1S<-glm(rsv_pos ~ LAG(x1S,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x1S)

LAGG=5
posglm.x1T<-glm(rsv_pos ~ LAG(x1T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x1T)

LAGG=4
posglm.x2S<-glm(rsv_pos ~ LAG(x2S,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
family = poisson)
summary(posglm.x2S)
```

```
1727
1728  LAGG=5
1729  posglm.x4T<-glm(rsv_pos ~ LAG(x4T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
1730  family = poisson)
1731  summary(posglm.x4T)
1732
1733  LAGG=10
1734  posglm.x5T<-glm(rsv_pos ~ LAG(x5T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
1735  family = poisson)
1736  summary(posglm.x5T)
1737
1738
1739
1740  library(dlnm)
1741  ## specify the two cross-bases for each of the variables for the GLM model
1742  basisposglm.x1 <- crossbasis(x1, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 3, maxlag
1743  = 4)
1744    summary(basisposglm.x1 )
1745  basisposglm.x2 <- crossbasis(x2, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 2, maxlag
1746  = 3)
1747    summary(basisposglm.x2 )
1748  basisposglm.x3 <- crossbasis(x3, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 1, maxlag
1749  = 2)
1750    summary(basisposglm.x3 )
```

```
basisposglm.x4 <- crossbasis(x4, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 3, maxlag
= 4)
  summary(basisposglm.x4 )

basisposglm.x1S <- crossbasis(x1S , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 0,
maxlag = 1)
  summary(basisposglm.x1S )
basisposglm.x1T <- crossbasis(x1T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 4,
maxlag = 5)
  summary(basisposglm.x1T )
basisposglm.x2S <- crossbasis(x2S , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 3,
maxlag = 4)
  summary(basisposglm.x2S )
basisposglm.x4T <- crossbasis(x4T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 4,
maxlag = 5)
  summary(basisposglm.x4T )
basisposglm.x5T <- crossbasis(x5T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 9,
maxlag = 10)
  summary(basisposglm.x5T )

## fitting the models
## Dadaab RSV Poisson GLM model ====== MODEL 3.a
=========================================

posglm.1a <- glm(rsv_pos ~
```

```
# lag_yt_1 + x1 + x2 + x3 + x4 + cosA + offset(log(RSV$pop/1000)), # model without
decomposition
lag_yt_1 + basisposglm.x1 + basisposglm.x2 + basisposglm.x3 + basisposglm.x4 + cosA  +
offset(log(RSV$pop/1000)), # best model without decomposition of the Poisson GLM model
  data=RSV, family = poisson)
  summary(posglm.1a)
  Dsquared(posglm.1a, adjust = TRUE)
#posglm.1a@formulas
#posglm.1a@objects
#summary(posglm.1a@objects[[1]])  ## plot(residuals(posglm.1a@objects[[1]]))
#weightable(posglm.1a)
# plot(posglm.1a)
# plot(posglm.1a,type="r")
# par(mfrow=c(1,2)); plot(posglm.1a,type="s"); plot(posglm.1a,type="w")

## Dadaab RSV Poisson GLM model ====== MODEL 3.b
==========================================

posglm.1b <- glm(rsv_pos ~
# x1S + x1T + x2S + x4T + x5T
# + cosA + sinA + offset(log(RSV$pop/1000)), # model with decomposition
basisposglm.x1S + basisposglm.x1T + basisposglm.x2S + basisposglm.x4T +
basisposglm.x5T +
cosA + sinA + offset(log(RSV$pop/1000)), # best model with decomposition
data=RSV, family = poisson)
```

```
summary(posglm.1b)
Dsquared(posglm.1b, adjust = TRUE)
#posglm.1b@formulas
#posglm.1b@objects
#summary(posglm.1b@objects[[1]])  ## plot(residuals(posglm.1b@objects[[1]]))
#weightable(posglm.1b)
# plot(posglm.1b)
# plot(posglm.1b,type="r")
# par(mfrow=c(1,2)); plot(posglm.1b,type="s"); plot(posglm.1b,type="w")


## Testing for max lag for each of the covariates in the GAM models for the same covariates
fitted in the GLM model above
library(mgcv)

LAGG=2
posgam.x1<-gam(rsv_pos ~ LAG(x1,LAGG) + offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x1)

LAGG=1
posgam.x2<-gam(rsv_pos ~ LAG(x2,LAGG) + offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x2)
```

```
LAGG=2
posgam.x3<-gam(rsv_pos ~ LAG(x3,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x3)

LAGG=1
posgam.x4<-gam(rsv_pos ~ LAG(x4,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x4)

LAGG=4
posgam.x1S<-gam(rsv_pos ~ LAG(x1S,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x1S)

LAGG=6
posgam.x1T<-gam(rsv_pos ~ LAG(x1T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x1T)

LAGG=8
posgam.x2S<-gam(rsv_pos ~ LAG(x2S,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x2S)

```

```
LAGG=3
posgam.x4T<-gam(rsv_pos ~ LAG(x4T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x4T)

LAGG=6
posgam.x5T<-gam(rsv_pos ~ LAG(x5T,LAGG)+ offset(log(RSV$pop/1000)), data=RSV,
 family = poisson)
aic(posgam.x5T)

## specify the two cross-bases for each of the variables for the GAM model
basisposgam.x1 <- crossbasis(x1, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 1,
maxlag = 2)
  summary(basisposgam.x1 )
basisposgam.x2 <- crossbasis(x2, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 0,
maxlag = 1)
  summary(basisposgam.x2 )
basisposgam.x3 <- crossbasis(x3, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 1,
maxlag = 2)
  summary(basisposgam.x3 )
basisposgam.x4 <- crossbasis(x4, vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 0,
maxlag = 1)
  summary(basisposgam.x4 )

```

```
basisposgam.x1S <- crossbasis(x1S , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 3,
maxlag = 4)
  summary(basisposgam.x1S )
basisposgam.x1T <- crossbasis(x1T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 5,
maxlag = 6)
  summary(basisposgam.x1T )
basisposgam.x2S <- crossbasis(x2S , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 7,
maxlag = 8)
  summary(basisposgam.x2S )
basisposgam.x4T <- crossbasis(x4T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 2,
maxlag = 3)
  summary(basisposgam.x4T )
basisposgam.x5T <- crossbasis(x5T , vartype = "bs", vardegree = 44, vardf = 4,  lagdf = 5,
maxlag = 6)
  summary(basisposgam.x5T )

## Dadaab RSV Poisson GAM model  ====== MODEL 4.a
========================================


posgam.1a <- gam(rsv_pos ~ lag_yt_1 + ns(basisposgam.x1,4) + ns(basisposgam.x2,4) +
ns(basisposgam.x3,4)
   + ns(basisposgam.x4,4)+ ns(mont,4)+ offset(log(RSV$pop/1000)), # best, without
decomposition
data=RSV, family = poisson)
```

```
summary(posgam.1a)
posgam.1a$coeff
summary.gam(posgam.1a)$sp.criterion # measure of performance
summary.gam(posgam.1a)$r.sq # adj.R-sq, measure of performance
AIC(posgam.1a)
plot(posgam.1a,pages=1)
gam.check(posgam.1a)


## Dadaab RSV Poisson GAM model  ====== MODEL 4.b
=========================================

posgam.1b <- gam(rsv_pos ~
ns(basisposgam.x1S,2*2) + ns(basisposgam.x1T,2*2) + ns(basisposgam.x2S,2*2) +
ns(basisposgam.x4T,2*2) + ns(basisposgam.x5T,2*2) + ns(mont,2*2)+
offset(log(RSV$pop/1000)),
data=RSV, family = poisson)# with decomposition
summary(posgam.1b)
posgam.1b$coeff
summary.gam(posgam.1b)$sp.criterion # measure of performance
summary.gam(posgam.1b)$r.sq # adj.R-sq, measure of performance
AIC(posgam.1b)
plot(posgam.1b,pages=1)
gam.check(posgam.1b)

```

```
# gam.check(posgam.1b) # overview on model performance
# vis.gam(posgam.1b,view=c("x1T","x2T"))


posgam.1bb <- gam(rsv_pos ~
s(x1S) + s(x1T) + s(x2S) + s(x4T) + s(x5T) + ns(mont,2)+ offset(log(RSV$pop/1000)),
data=RSV, family = poisson)

tiff("Figure4.tiff", res=600, compression = "lzw", height=8, width=8, units="in")
op05 <- par(mfrow = c(3,2), mar=c(4,5,0.5,0.5), oma=c(1.5,3,1,1))
##oma = c(5,4,0,0) + 0.1, mar = c(0,0,1,1) + 0.7)
# par(mfrow = c(3,2))
plot(posgam.1bb, select=1, shade=TRUE, xlab="Seasonal, wind speed (knots)",
    cex.lab=1.5, ylab="")
text(3,40,"A", cex=1.6)
# screen(2)
plot(posgam.1bb, select=2, shade=TRUE, xlab="Trend, wind speed (knots)",
    cex.lab=1.5, ylab="")
text(8.3,40,"B", cex=1.6)
plot(posgam.1bb, select=3, shade=TRUE, xlab="Seasonal, rainfall (inches)",
    ylab="RSV Incidence", cex.lab=1.6)
text(1.8,40,"C", cex=1.6)
# screen(4)
plot(posgam.1bb, select=4, shade=TRUE, xlab="Trend, dew point (degres fahrenheit)",
    cex.lab=1.5, ylab="")
```

```
text(1.8,40,"D", cex=1.6)
# screen(5)
plot(posgam.1bb, select=5, shade=TRUE, xlab="Trend, visibility(miles)",
    cex.lab=1.5, ylab="")
text(21,40,"E", cex=1.6)

# title(ylab="RSV Incidence (cases per 1000 person months)",
#       cex.lab=1.5, cex.axis=1.5, outer = TRUE, line = 3)
# close.screen(all=TRUE)
par(op05);
dev.off()


##-------------------------COMPARE MODEL PERFORMANCES--------------------------------
-

print(AIC(posgam.1a)) # Poisson, GAM, without decomposition
print(AIC(posgam.1b)) # Poisson, GAM, with decomposition

anova(posgam.1a,posgam.1b, test="Chisq")
anova(posgam.1b)

##----------------plot Residual-----------------------------
tiff("Residual.tiff", res=600, compression = "lzw", height=7, width=8, units="in")
op03 <- par(mfrow = c(2,1), oma = c(2,2,0,0) + 0.17, mar = c(4,4,1,1) + 0)
```

```
gam.check(posgam.1a)
text(-1,2,"Poisson GAM without covariate decomposition")
gam.check(posgam.1b)
text(-1.09,1,"Poisson GAM with covariate decomposition")
par(op03); dev.off()



csN <- RSV$rsv_pos
popN <- 1000
fracP <- csN/popN
##----------------plot Figure 1.A and B, below-------------------------------
tiff("Figure1.tiff", res=600, compression = "lzw", height=9, width=8, units="in")
op01 <- par(mfrow = c(1,1), mar=c(4, 6, 2, 1) + 0.1)
# op01 <- par(mfrow = c(2,1), oma = c(2,2,0,0) + 0.17, mar = c(4,4,1,1) + 0.4)
plot((RSV_proptimeseriescomponents$x)*popN*csN, xlab="Time (month/year)",
    ylab="RSV Incidence (cases \nper 1000 person
months)",lwd=1,pch=16,col="gray20",type="o", #ylim=c(0,0.8),
    main=expression(A))

grid (NULL,NULL,col="grey") # Figure_1.a
periodogram(RSV_proptimeseriescomponents$x,log='no',plot=TRUE,
        ylab="Periodogram", xlab="Frequency",lwd=2,main=expression(B))  # Figure_1.b
par(op01)
dev.off()
```

```
2000
2001  ## =================PLOTS MODEL FITS WITHOUT DECOMPOSED
2002  COVARIATES===============
2003
2004  # best GLM without decomposed covariates, Poisson model-1 ------------------------
2005    p.posglm.1a <- predict(posglm.1a,RSV,type = "link",se.fit = TRUE) # @objects[[1]]
2006    prob.posglm.1a <- as.vector(exp(p.posglm.1a$fit)/(RSV$pop + exp(p.posglm.1a$fit))); #
2007  RSV$
2008    sd.e_posglm.1a <- as.vector(exp(p.posglm.1a$se.fit)/(RSV$pop + exp(p.posglm.1a$se.fit)));
2009    upr.pglm.1a <- prob.posglm.1a + (1.96*sd.e_posglm.1a);
2010    lwr.pglm.1a <- prob.posglm.1a - (1.96*sd.e_posglm.1a);
2011  # best GAM without decomposed covariates, Poisson model-2 ------------------------
2012    p.posgam.1a <- predict(posgam.1a,se.fit = TRUE)
2013    prob.posgam.1a <- as.vector(exp(p.posgam.1a$fit)/(RSV$pop + exp(p.posgam.1a$fit)));
2014  #RSV$
2015    sd.e_posgam.1a <- as.vector(exp(p.posgam.1a$se.fit)/(RSV$pop +
2016  exp(p.posgam.1a$se.fit))); #RSV$
2017    upr.pgam.1a <- prob.posgam.1a + (1.96*sd.e_posgam.1a);
2018    lwr.pgam.1a <- prob.posgam.1a - (1.96*sd.e_posgam.1a);
2019
2020  ##----plot best model fits to data: plot Figure 10.A to C, below-------------
2021  tiff("Figure3.tiff", res=600, compression = "lzw", height=8, width=10, units="in")
2022  op03 <- par(mfrow = c(2,1), oma = c(5,4,0,0) + 0.2, mar = c(0,0,1,1) + 0.7)
2023  plot(t, csN*(probRSVti*popN),col="gray20",pch=16,xlab="Time") #ylim=c(0,0.8))
2024    lines(t, csN*popN*prob.posglm.1a,col="blue",lwd=2)
```

```
2025    lines(t, csN*popN*upr.pglm.1a,col="black",type="l",lty=3,lwd=1);
2026    lines(t, csN*popN*lwr.pglm.1a,col="black",type="l",lty=3,lwd=1)
2027    text(47,0.9*max(csN*probRSVti*popN),"A", cex=1.5)
2028  plot(t, csN*(probRSVti*popN),col="gray20",pch=16,xlab="Time") #ylim=c(0,0.8))
2029    lines(t, csN*popN*prob.posgam.1a,col="blue",lwd=2)
2030    lines(t, csN*popN*upr.pgam.1a,col="black",type="l",lty=3,lwd=1);
2031    lines(t, csN*popN*lwr.pgam.1a,col="black",type="l",lty=3,lwd=1)
2032    text(47,0.9*max(csN*probRSVti*popN),"B", cex=1.5)
2033  title(xlab = "Time (months)",ylab = "RSV Incidence (cases per 1000 person months)",
2034        cex.lab=1.5, cex.axis=1.5, outer = TRUE, line = 3)
2035  par(op03);
2036  dev.off()
2037
2038  ## ==================PLOTS MODEL FITS WITH DECOMPOSED
2039  COVARIATES===================
2040  # best GLM with decomposed covariates, Poisson model-3 ----------------------------
2041    p.posglm.1b <- predict(posglm.1b,RSV,type = "link",se.fit = TRUE)
2042    prob.posglm.1b <- as.vector(exp(p.posglm.1b$fit)/(RSV$pop + exp(p.posglm.1b$fit)));
2043    sd.e_posglm.1b <- as.vector(exp(p.posglm.1b$se.fit)/(RSV$pop + exp(p.posglm.1b$se.fit)));
2044    upr.pglm.1b <- prob.posglm.1b + (1.96*sd.e_posglm.1b);
2045    lwr.pglm.1b <- prob.posglm.1b - (1.96*sd.e_posglm.1b);
2046  # best GAM with decomposed covariates, Poisson model-4 -------------------------
2047    p.posgam.1b <- predict(posgam.1b,se.fit = TRUE)
2048    prob.posgam.1b <- as.vector(exp(p.posgam.1b$fit)/(RSV$pop + exp(p.posgam.1b$fit)));
```

```
  sd.e_posgam.1b <- as.vector(exp(p.posgam.1b$se.fit)/(RSV$pop +
exp(p.posgam.1b$se.fit)));
  upr.pgam.1b <- prob.posgam.1b + (1.96*sd.e_posgam.1b);
  lwr.pgam.1b <- prob.posgam.1b - (1.96*sd.e_posgam.1b);

##-----plot best model fits to data: plot Figure 3.A to C, below----------------------------------
tiff("FigureS1.tiff", res=600, compression = "lzw", height=5, width=5, units="in")
incide<- probRSVti*1000
op04 <- par(mfrow = c(2,2), oma = c(5,4,0,0) + 0.1, mar = c(0,0,1,1) + 0.7)
plot(t,incide,col="gray20",pch=16,xlab="Time",ylab="RSV Incidence",ylim=c(0,2))
  lines(t,prob.posglm.1b,col="orangered",lwd=2)
  lines(t,upr.pglm.1b,col="black",type="l",lty=3,lwd=1);
  lines(t,lwr.pglm.1b,col="black",type="l",lty=3,lwd=1); text(45,1.5,"C")
plot(t,incide,col="gray20",pch=16,xlab="Time",ylab="RSV Incidence",ylim=c(0,2))
  lines(t,prob.posgam.1b,col="orangered",lwd=2)
  lines(t,upr.pgam.1b,col="black",type="l",lty=3,lwd=1);
  lines(t,lwr.pgam.1b,col="black",type="l",lty=3,lwd=1); text(45,1.5,"D")
title(xlab = "Time (months)",ylab = "RSV Incidence",outer = TRUE, line = 3)
par(op04); dev.off()

# Bivariate Negative Binomial Model: RSV and HMPV

model{
#Likelihood
for (t in 2:T)
```

```
2074            {
2075            rsv[t] ~ dnegbin(p1[t],psi1)
2076            p1[t]<-psi1/(psi1+mu1[t])
2077            hmpv[t] ~ dnegbin(p2[t],psi2)
2078            p2[t]<-psi2/(psi2+mu2[t])
2079            nu1[t] <-alpha1+gamma11*sin(2*3.14/12)+delta11*cos(2*3.14/12)
2080            nu2[t] <-alpha2+gamma21*sin(2*3.14/12)+delta21*cos(2*3.14/12)
2081
2082            log(mu1[t]) <-lambda1[t-1]*rsv[t-1]+tao11[t-1]*
2083            rainfall[t-1] + tao12[t-1]*wind[t-1] + tao13[t-1]*dew[t-1] + tao14[t-1]*visibility[t-1] +
2084    log(pop[t])*nu1[t]
2085            log(mu2[t]) <-lambda2[t-1]*hmpv[t-1]+phi2[t-1]*rsv[t-1]+tao21[t-1]*rainfall[t-1] +
2086    tao22[t-1]*wind[t-1] + tao23[t-1]*dew[t-1] + tao24[t-1]*visibility[t-1] +log(pop[t])*nu2[t]
2087            }
2088
2089    # log-likelihood ll[t] <-
2090
2091
2092    # Priors
2093
2094            psi1~dgamma(0.1,0.01)
2095            psi2~dgamma(0.1,0.01)
2096            gamma11~dnorm(0.0, 0.001)
2097            gamma21~dnorm(0.0, 0.001)
2098            delta11~dnorm(0.0, 0.001)
```

```
2099        delta21~dnorm(0.0, 0.001)
2100        alpha1~dnorm(0.0, 0.001)
2101        alpha2~dnorm(0.0, 0.001)
2102
2103        beta1~dgamma(10,10)
2104        beta2~dgamma(10,10)
2105
2106
2107
2108        for (t in 1:(T-1))
2109        {
2110                phi2[t]~dgamma(1,0.5)
2111                lambda1[t]~dgamma(1,beta1)
2112                lambda2[t]~dgamma(1,beta2)
2113                tao11[t]~dgamma(10,10)
2114                tao12[t]~dgamma(10,10)
2115                tao13[t]~dgamma(10,10)
2116                tao14[t]~dgamma(10,10)
2117          tao21[t]~dgamma(10,10)
2118                tao22[t]~dgamma(10,10)
2119                tao23[t]~dgamma(10,10)
2120                tao24[t]~dgamma(10,10)
2121        }
2122        }
2123
```

```
2124
2125  # Multivariate Poisson Model: RSV,HMPV,FLU
2126
2127  model{
2128  #Likelihood
2129  for (t in 2:T)
2130          {
2131          rsv[t] ~  dnegbin(p1[t],psi1)
2132          p1[t]<-psi1/(psi1+mu1[t])
2133          hmpv[t] ~ dnegbin(p2[t],psi2)
2134          p2[t]<-psi2/(psi2+mu2[t])
2135          flu[t] ~ dnegbin(p3[t],psi3)
2136          p3[t]<-psi3/(psi3+mu3[t])
2137
2138          nu1[t] <-alpha1+gamma11*sin(2*3.14/12)+delta11*cos(2*3.14/12)
2139          nu2[t] <-alpha2+gamma21*sin(2*3.14/12)+delta21*cos(2*3.14/12)
2140          nu3[t] <-alpha3+gamma31*sin(2*3.14/12)+delta31*cos(2*3.14/12)
2141
2142  log(mu1[t]) <-lambda1[t-1]*rsv[t-1]+phi1[t-1]*hmpv[t-1] +tao11[t-1]*
2143          rainfall[t-1] + tao12[t-1]*wind[t-1] + tao13[t-1]*dew[t-1] + tao14[t-1]*visibility[t-1] +
2144  log(pop[t])*nu1[t]
2145  log(mu2[t]) <-lambda2[t-1]*hmpv[t-1]+phi3[t-1]*rsv[t-1] +tao21[t-1]*rainfall[t-1] + tao22[t-
2146  1]*wind[t-1] + tao23[t-1]*dew[t-1] + tao24[t-1]*visibility[t-1] +log(pop[t])*nu2[t]
2147  log(mu3[t]) <-lambda3[t-1]*flu[t-1]+tao31[t-1]*rainfall[t-1] + tao32[t-1]*wind[t-1] + tao33[t-
2148  1]*dew[t-1] + tao34[t-1]*visibility[t-1] +log(pop[t])*nu3[t]
```

```
2149
2150            }
2151
2152    # log-likelihood ll[t] <-
2153
2154
2155    # Priors
2156
2157            psi1~dgamma(0.1,0.01)
2158            psi2~dgamma(0.1,0.01)
2159            psi3~dgamma(0.1,0.01)
2160
2161            gamma11~dnorm(0.0, 0.001)
2162            gamma21~dnorm(0.0, 0.001)
2163            gamma31~dnorm(0.0, 0.001)
2164
2165            delta11~dnorm(0.0, 0.001)
2166            delta21~dnorm(0.0, 0.001)
2167            delta31~dnorm(0.0, 0.001)
2168
2169            alpha1~dnorm(0.0, 0.001)
2170            alpha2~dnorm(0.0, 0.001)
2171            alpha3~dnorm(0.0, 0.001)
2172
2173            beta1~dgamma(10,10)
```

```
2174            beta2~dgamma(10,10)
2175            beta3~dgamma(10,10)
2176
2177            for (t in 1:(T-1))
2178            {
2179                    phi1[t]~dgamma(1,0.1)
2180                    phi3[t]~dgamma(1,0.1)
2181                    lambda1[t]~dgamma(1,beta1)
2182                    lambda2[t]~dgamma(1,beta2)
2183                    lambda3[t]~dgamma(1,beta3)
2184                    tao11[t]~dgamma(10,10)
2185                    tao12[t]~dgamma(10,10)
2186                    tao13[t]~dgamma(10,10)
2187                    tao14[t]~dgamma(10,10)
2188              tao21[t]~dgamma(10,10)
2189                    tao22[t]~dgamma(10,10)
2190                    tao23[t]~dgamma(10,10)
2191                    tao24[t]~dgamma(10,10)
2192                    tao31[t]~dgamma(10,10)
2193                    tao32[t]~dgamma(10,10)
2194                    tao33[t]~dgamma(10,10)
2195                    tao34[t]~dgamma(10,10)
2196            }
2197    }
2198
```

References

1. Iwane MK, Edwards KM, Szilagyi PG, Walker FJ, Griffin MR, Weinberg GA, et al. Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. Pediatrics [Internet]. 2004;113(6):1758–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15173503
2. Stensballe LG, Devasundaram JK, Simoes EAF. Respiratory syncytial virus epidemics: the ups and downs of a seasonal virus. Pediatr Infect Dis J. 2003;22(2):S21–32.
3. Leecaster M, Gesteland P, Greene T, Walton N, Gundlapalli A V, Rolfs RT, et al. Modeling the variations in pediatric respiratory syncytial virus seasonal epidemics. BMC Infect Dis [Internet]. 2011;11(1):105. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3094225&tool=pmcentrez&rendertype=abstract
4. Simões EAF, DeVincenzo JP, Boeckh M, Bont L, Crowe JE, Griffiths P, et al. Challenges and opportunities in developing respiratory syncytial virus therapeutics. J Infect Dis. 2015;211:S1–20.
5. Mangtani P, Hajat S, Kovats S, Wilkinson P, Armstrong B. The association of respiratory syncytial virus infection and influenza with emergency admissions for respiratory disease in London: an analysis of routine surveillance data. Clin Infect Dis. 2006;42(5):640–6.
6. Mwambi H, Ramroop S, White LJ, Okiro EA, Nokes DJ, Shkedy Z, et al. A frequentist approach to estimating the force of infection and the recovery rate for a respiratory disease among infants in coastal Kenya. Stat Methods Med Res [Internet]. 2011;20(5):573. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19221168

7. Loeffelholz M, Chonmaitree T. Advances in diagnosis of respiratory virus infections. International Journal of Microbiology. 2010.
8. Piedimonte G, Perez MK. Respiratory syncytial virus infection and bronchiolitis. Pediatr Rev [Internet]. 2014;35(12):519–30. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84913533040&partnerID=tZOtx3y1
9. Hall CB. Nosocomial respiratory syncytial virus infections: the "Cold War" has not ended. Clin Infect Dis [Internet]. 2000;31(2):590–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10987726
10. Aintablian N, Walpita P, Sawyer MH. Detection of Bordetella pertussis and respiratory synctial virus in air samples from hospital rooms. Infect Control Hosp Epidemiol [Internet]. 1998;19(12):918–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9872529
11. Vandini S, Corvaglia L, Alessandroni R, Aquilano G, Marsico C, Spinelli M, et al. Respiratory syncytial virus infection in infants and correlation with meteorological factors and air pollutants. Ital J Pediatr [Internet]. 2013;39(1):1. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3553040&tool=pmcentrez&rendertype=abstract
12. Vitor P, Simas M, Gustavo L, Gardinassi A, Nogueira FC, Measso C, et al. Virus Reviews and Research, Analysis of climatic factors impact on RSV infection distribution in children attending childcare at northwest region of Sao Paulo, Brazil. Virus Rev Res. 2012;23–8.
13. Ampofo K, Bender J, Sheng X, Korgenski K, Daly J, Pavia AT, et al. Seasonal invasive pneumococcal disease in children: role of preceding respiratory viral infection. Pediatrics [Internet]. 2008;122(2):229–37. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/18676537

14. Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller M a., et al. Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. PLoS One. 2013;8(2):3–4.

15. Wolff C, Kristen-Jenny I, Schettler G, Plessen B, Meyer H, Dulski P, et al. Modern seasonality in Lake Challa (Kenya/Tanzania) and its sedimentary documentation in recent lake sediments. Limnol Oceanogr [Internet]. 2014;59(5):1621–36. Available from: http://doi.wiley.com/10.4319/lo.2014.59.5.1621

16. Hall CB, Simőes EAF, Anderson LJ. Clinical and epidemiologic features of respiratory syncytial virus. Curr Top Microbiol Immunol [Internet]. 2013;372:39–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24362683

17. DeVincenzo JP, de Swart RL, Osterhaus ADME. 2007 International Congress on Respiratory Viruses, Colorado, USA, 20-22 July 2007. [Internet]. Vol. 27, Pediatric Infectious Disease Journal. 2008. p. S1–145. Available from: http://www.pidj.com/

18. Khor C-S, Sam I-C, Hooi P-S, Quek K-F, Chan Y-F. Epidemiology and seasonality of respiratory viral infections in hospitalized children in Kuala Lumpur, Malaysia: a retrospective study of 27 years. BMC Pediatr [Internet]. 2012;12(1):32. Available from: http://www.biomedcentral.com/1471-2431/12/32

19. Dnb N, Theodoratou E, Rudan I, Campbell H, Nokes DJ, Hnd N, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet. 2010;375:1545–55.

20. Kim C, Nyoka R, Ahmed JA, Winchell JM, Mitchell SL, Kariuki Njenga M, et al. Epidemiology of respiratory infections caused by atypical bacteria in Two Kenyan refugee camps. J Immigr Minor Heal. 2012;14(1):140–5.

21. Patrick T. Brandt and John T. Williams. Multiple time series models. No: 148. Sage;

2007.

22. Patrick T. Brandt and John T. Williams. A Linear Poisson Autoregressive Model: The Poisson AR(p) Model “Political Analysis, vol. 9, no. 2” [Internet]. 2001. 164-184 p. Available from: www.jstor.org/stable/25791638

23. Hastie TJ, Tibshirani R. Generalized additive models [Internet]. Vol. 1, Statistical Science. 1990. p. 297–318. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Generalized+additive+models#0

24. Wood SN. Generalized additive models : an introduction with R. Texts Stat Sci [Internet]. 2006;xvii, 392 p. Available from: http://catdir.loc.gov/catdir/enhancements/fy0702/2006040209-d.html%5Cnhttp://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&doc_number=014829960&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

25. Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol. 2002;156(3):193–203.

26. Development Core Team R. R: A Language and Environment for Statistical Computing [Internet]. Vol. 0, R Foundation for Statistical Computing Vienna Austria. 2011. {ISBN} 3-900051-07-0. Available from: http://www.r-project.org

27. Cleveland B. Robert, Cleveland S. William, McRae E. Jean TI. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. J Off Stat [Internet]. 1990;6(No. 1):3–73. Available from: https://www.wessa.net/download/stl.pdf

28. Peng RD, Dominici F. Statistical Methods for Environmental Epidemiology with R [Internet]. Media. 2008. 1-151 p. Available from: http://link.springer.com/10.1007/978-

0-387-78167-9%5Cnhttps://books.google.com/books?id=QbDBxSSXIjsC&pgis=1

29. Gasparrini A. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. J Stat Softw [Internet]. 2011;43(8):1–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3191524&tool=pmcentrez&rendertype=abstract

30. Gasparrini A. Distributed lag linear and non-linear models for time series data. 14310721250 [Internet]. 2014;1–12. Available from: http://143.107.212.50/web/packages/dlnm/vignettes/dlnmTS.pdf

31. Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J Am Stat Assoc [Internet]. 1988;83(403):596–610. Available from: http://dx.doi.org/10.2307/2289282

32. Calcagno V, Mazancourt C De. glmulti : An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J Stat Softw. 2010;34(12):1–29.

33. Jónsson G. The environmental factor in migration dynamics – a review of African case studies. Int Migr Inst Work Pap [Internet]. 2010;1–34. Available from: http://www.imi.ox.ac.uk/pdfs/wp-21-gjonsson

34. Agoti CN, Mayieka LM, Otieno JR, Ahmed J a, Fields BS, Waiboci LW, et al. Examining strain diversity and phylogeography in relation to an unusual epidemic pattern of respiratory syncytial virus (RSV) in a long-term refugee camp in Kenya. BMC Infect Dis [Internet]. 2014;14(1):178. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021307&tool=pmcentrez&rendertype=abstract

35. Noyola DE, Mandeville PB. Effect of climatological factors on respiratory syncytial virus epidemics. Epidemiol Infect [Internet]. 2008;136(10):1328–32. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2870732&tool=pmcentrez

&rendertype=abstract

36. Haynes AK, Manangan AP, Iwane MK, Sturm-Ramirez K, Homaira N, Brooks WA, et al. Respiratory syncytial virus circulation in seven countries with global disease detection regional centers. J Infect Dis. 2013;208(SUPPL. 3).

37. Donaldson GC. Climate change and the end of the respiratory syncytial virus season. Clin Infect Dis [Internet]. 2006;42(5):677–9. Available from: http://discovery.ucl.ac.uk/61843/

38. Simoes E a. F, Cherian T, Chow J, Shahid-Salles S a., Laxminarayan R, John TJ. Acute Respiratory Infections in Children. Dis Control Priorities Dev Ctries [Internet]. 2006;1–24. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11786/

39. Checchi F, Gayer M, Grais RF, Mills EJ. Public Health in Crisis-Affected Populations: A Practical Guide for Decision-Makers [Internet]. Vol. 61, Humanitarian Practice Network. 2007. Available from: http://www.atha.se/sites/default/files/public_health_in_crisis-affected_populations-_a_practical_guide_for_decision-makers.pdf

40. Bellos A, Mulholland K, O'Brien KL, Qazi SA, Gayer M, Checchi F. The burden of acute respiratory infections in crisis-affected populations: a systematic review. Confl Health [Internet]. 2010;4(1):3. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-77954739853&partnerID=tZOtx3y1

41. Tregoning JS, Schwarze J. Respiratory viral infections in infants: Causes, clinical symptoms, virology, and immunology. Clin Microbiol Rev. 2010;23(1):74–98.

42. Pastula ST, Hackett J, Coalson J, Jiang X, Villafana T, Ambrose C, et al. Hospitalizations for respiratory syncytial virus (RSV) among adults in the United States, 1997 - 2012. Open Forum Infect Dis [Internet]. 2017;48105:ofw270. Available from:

https://academic.oup.com/ofid/article-lookup/doi/10.1093/ofid/ofw270

43. Cuevas LE, Nasser AM Ben, Dove W, Gurgel RQ, Greensill J, Hart CA. Human metapneumovirus and respiratory syncytial virus, Brazil. Emerg Infect Dis. 2009;9(12):1626–8.

44. Peiris JSM, Tang WH, Chan KH, Khong PL, Guan Y, Lau YL, et al. Children with respiratory disease associated with metapneumovirus in Hong Kong. Emerg Infect Dis. 2003;9(6):628–33.

45. Guerrero-Plata A, Casola A, Suarez G, Yu X, Spetch L, Peeples ME, et al. Differential response of dendritic cells to human metapneumovirus and respiratory syncytial virus. Am J Respir Cell Mol Biol. 2006;34(3):320–9.

46. Le Nouen C, Munir S, Losq S, Winter CC, McCarty T, Stephany DA, et al. Infection and maturation of monocyte-derived human dendritic cells by human respiratory syncytial virus, human metapneumovirus, and human parainfluenza virus type 3. Virology. 2009;385(1):169–82.

47. Greensill J, McNamara PS, Dove W, Flanagan B, Smyth RL, Hart CA. Human metapneumovirus in severe respiratory syncytial virus bronchiolitis. Emerg Infect Dis. 2003;9(3):372–5.

48. Konig B, Konig W, Arnold R, Werchau H, Ihorst G, Forster J. Prospective Study of Human Metapneumovirus Infection in Children Less Than 3 Years of Age. Society. 2004;42(10):4632–5.

49. Tang RS, Schickli JH, Macphail M, Fernandes F, Bicha L, Spaete J, et al. Effects of Human Metapneumovirus and Respiratory Syncytial Virus Antigen Insertion in Two 3J Proximal Genome Positions of Bovine / Human Parainfluenza Virus Type 3 on Virus Replication and Immunogenicity. 2003;77(20):10819–28.

50. Akhras N, Weinberg JB, Newton D. Human metapneumovirus and respiratory syncytial

virus: Subtle differences but comparable severity. Infect Dis Rep. 2010;2(2):35–9.

51. Moe N, Krokstad S, Stenseng IH, Christensen A, Risnes KR, Arne S, et al. Comparing Human Metapneumovirus and Respiratory Syncytial Virus : Viral Co- Detections , Genotypes and Risk Factors for Severe Disease. PLoS One [Internet]. 2017;12(1):1–19. Available from: https://doi.org/10.1371/journal.pone.0170200

52. You H-L, Chang S-J, Yu H-R, Li C-C, Chen C-H, Liao W-T. Simultaneous detection of respiratory syncytial virus and human metapneumovirus by one-step multiplex real-time RT-PCR in patients with respiratory symptoms. BMC Pediatr [Internet]. 2017;17(1):89. Available from: http://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-017-0843-7

53. Ditt V, Lüsebrink J, Tillmann RL, Schildgen V, Schildgen O. Respiratory infections by HMPV and RSV are clinically indistinguishable but induce different host response in aged individuals. PLoS One. 2011;6(1):1–9.

54. Held L, Höhle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. Stat Modelling. 2005;5(3):187–99.

55. O'Hara RB, Kotze DJ. Do not log-transform count data. Methods Ecol Evol [Internet]. 2010;1(2):118–22. Available from: http://doi.wiley.com/10.1111/j.2041-210X.2010.00021.x

56. Nishiura H. Early Detection of Nosocomial Outbreaks Caused by Rare Pathogens: A Case Study Employing Score Prediction Interval. Osong Public Heal Res Perspect [Internet]. 2012;3(3):121–7. Available from: http://dx.doi.org/10.1016/j.phrp.2012.07.010

57. Buckeridge DL, Okhmatovskaia A, Tu S, O'Connor M, Nyulas C, Musen M a. Predicting outbreak detection in public health surveillance: quantitative analysis to enable evidence-based method selection. AMIA Annu Symp Proc. 2008;76–80.

58. Ver Hoef JM, Boveng PL. Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? Ecology. 2007;88(11):2766–72.

59. Paul M, Held L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Stat Med. 2011;30(10):1118–36.

60. Paul M, Meyer S. The function hhh4 in the R -package surveillance. 2014;1–15.

61. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity anf Fit. J R Stat Soc Ser B (Statistical Methodol. 2002;64(4):583–639.

62. Adrion C, Mansmann U. Bayesian model selection techniques as decision support for shaping a statistical analysis plan of a clinical trial: an example from a vertigo phase III study with longitudinal count data as primary endpoint. BMC Med Res Methodol [Internet]. 2012;12:137. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3554595&tool=pmcentrez&rendertype=abstract

63. Agoti CN, Mayieka LM, Otieno JR, Ahmed JA, Fields BS, Waiboci LW, et al. Examining strain diversity and phylogeography in relation to an unusual epidemic pattern of respiratory syncytial virus (RSV) in a long-term refugee camp in Kenya. BMC Infect Dis [Internet]. 2014;14(1):178. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021307&tool=pmcentrez&rendertype=abstract

64. Nyoka R, Omony J, Mwalili SM, Achia TNO, Gichangi A, Mwambi H. Effect of climate on incidence of respiratory syncytial virus infections in a refugee camp in Kenya : A non-Gaussian time-series analysis. PLoS One. 2017;12(6):1–14.

65. Wilkesmann A, Schildgen O, Eis-Hübinger AM, Geikowski T, Glatzel T, Lentze MJ, et al. Human metapneumovirus infections cause similar symptoms and clinical severity as respiratory syncytial virus infections. Eur J Pediatr. 2006;165(7):467–75.

66. Kim CK, Choi J, Callaway Z, Kim H Bin, Chung JY, Koh YY, et al. Clinical and epidemiological comparison of human metapneumovirus and respiratory syncytial virus in Seoul, Korea, 2003-2008. J Korean Med Sci. 2010;25(3):342–7.

67. Al-Sonboli N, Hart CA, Al-Aeryani A, Banajeh SM, Al-Aghbari N, Dove W, et al. Respiratory syncytial virus and human metapneumovirus in children with acute respiratory infections in Yemen. Pediatr Infect Dis J. 2005;24(8):734–6.

68. Lazar I, Weibel C, Dziura J, Ferguson D, Landry ML, Kahn JS. Metapneumovirus and Severity of Respiratory Disease. Emerg Infect Dis [Internet]. 2004;10(7):7–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16607540

69. Plummer M. Penalized loss functions for Bayesian model comparison. Biostatistics. 2008;9(3):523–39.

70. Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. Vol. 33, Sociological Methods and Research. 2004. p. 261–304.

71. Ouedraogo S, Traore B, Nene Bi ZA, Yonli FT, Kima D, Bonane P, et al. Viral etiology of respiratory tract infections in children at the pediatric hospital in ouagadougou (burkina faso). PLoS One. 2014;9(10):e110435.

72. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Infect Dis [Internet]. 2017;3099(17):1–29. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1473309917302761

73. Shafik CF, Mohareb EW, Yassin AS, Amin MA, El Kholy A, El-Karaksy H, et al. Viral etiologies of lower respiratory tract infections among Egyptian children under five years of age. BMC Infect Dis [Internet]. 2012;12(1):350. Available from: http://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-12-350

74. Hatipoglu N, Somer A, Badur S, Unuvar E, Akcay-Ciblak M, Yekeler E, et al. Viral etiology in hospitalized children with acute lower respiratory tract infection. Turk J Pediatr [Internet]. 2011;53(5):508–16. Available from: http://www.turkishjournalpediatrics.org/uploads/pdf_TJP_966.pdf

75. Liu T, Li Z, Zhang S, Song S, Julong W, Lin Y, et al. Viral Etiology of acute respiratory tract infections in hospitalized children and adults in Shandong Province, China. Virol J [Internet]. 2015;12(1):168. Available from: http://www.virologyj.com/content/12/1/168

76. Liu WK, Liu Q, Chen DH, Liang HX, Chen XK, Chen MX, et al. Epidemiology of Acute Respiratory Infections in Children in Guangzhou: A Three-Year Study. PLoS One [Internet]. 2014;9(5):e96674. Available from: http://dx.plos.org/10.1371/journal.pone.0096674

77. Debiaggi M, Canducci F, Ceresola ER, Clementi M. The role of infections and coinfections with newly identified and emerging respiratory viruses in children. Virol J [Internet]. 2012;9(1):247. Available from: http://virologyj.biomedcentral.com/articles/10.1186/1743-422X-9-247

78. Van Asten L, Van Den Wijngaard C, Van Pelt W, Van De Kassteele J, Meijer A, Van Der Hoek W, et al. Mortality attributable to 9 common infections: Significant effect of influenza A, respiratory syncytial virus, influenza B, norovirus, and parainfluenza in elderly persons. J Infect Dis. 2012;206(5):628–39.

79. Jørgensen B, Lundbye-Christensen S, Song P-K, Sun L. A state space model for multivariate longitudinal count data. Biometrika. 1999;86(1):169–81.

80. Jørgensen B, Lundbye-Christensen S, Song XK, Sun L. A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. Stat Med [Internet]. 1996;15(7–9):823–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9132908

81. Ngueyep R, Serban N. Large Vector Auto Regression for Multi-Layer Spatially Correlated Time Series. Technometrics [Internet]. 2014;(January):00–00. Available from: http://www.tandfonline.com/doi/full/10.1080/00401706.2014.902775

82. Jung RC, Liesenfeld R, Richard J-F. Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity. J Bus Econ Stat. 2011;29(1):73–85.

83. Pedeli X, Karlis D. A bivariate INAR(1) process with application. Stat Model [Internet]. 2011 Aug 1;11(4):325–49. Available from: http://smj.sagepub.com/content/11/4/325.abstract

84. Paul M, Meyer S. hhh4 : An endemic-epidemic modelling framework for infectious disease counts. 2016;(2010):1–17.

85. Johansen S. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica. 1991;59(No. 6 (November, 1991)):1551–80.

86. QuantStart Team. Johansen Test for Cointegrating Time Series Analysis in R [Internet]. Available from: https://www.quantstart.com/articles/Johansen-Test-for-Cointegrating-Time-Series-Analysis-in-R

87. Matheka DM, Mokaya J, Maritim M. Overview of influenza virus infections in Kenya: past, present and future. Pan Afr Med J. 2013;14:138.

88. Comas-García A, García-Sepúlveda CA, Méndez-De Lira JJ, Aranda-Romo S, Hernández-Salinas AE, Noyola DE. Mortality attributable to pandemic influenza A (H1N1) 2009 in San Luis Potosí, Mexico. Influenza Other Respi Viruses. 2011;5(2):76–82.

89. Alexander M, Kobes R. Effects of vaccination and population structure on influenza epidemic spread in the presence of two circulating strains. BMC Public Health [Internet].

2011;11(Suppl 1):S8. Available from: http://www.biomedcentral.com/1471-2458/11/S1/S8

90. Mangtani P, Hajat S, Kovats S, Wilkinson P, Armstrong B. The Association of Respiratory Syncytial Virus Infection and Influenza with Emergency Admissions for Respiratory Disease in London : An Analysis of Routine Surveillance Data. Clin Infect Dis. 2017;42(December):640–6.

91. Nicholson KG. Impact of influenza and respiratory syncytial virus on mortality in England and Wales from January 1975 to December 1990. Epidemiol Infect [Internet]. 1996;116(1):51–63. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2271234&tool=pmcentrez&rendertype=abstract

92. Arguedas-Flatts YN, Capistran MA, Christen JA, Noyola DE, Velasco-Hernandez JX. An analysis of the interaction between influenza and respiratory syncytial virus based on acute respiratory infection records. arXiv:13120594. 2013;1–14.