U.PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

THE POISSON PROCESS AND ITS APPLICATION TO WILDFIRES

Paula Sofia Sousa Resende

Dissertation

Master in Modeling, Data Analysis and Decision Support System

---

Supervised by

Ana Cristina Gomes Monteiro Moreira de Freitas
Jorge Miguel Milhazes de Freitas

2019

Acknowledgments

I would like to thank to my supervisors PhD Ana Cristina Gomes Monteiro Moreira de Freitas and PhD Jorge Miguel Milhazes de Freitas for all support, guidance and patience. Their help was essential to solve problems of dissertation and their opinion was crucial in all steps.

To my parents, they deserve all the thanks. They wouldn't let me give up. They heard me when I needed and they always had a right word to say.

To my boyfriend, he had an important role. He never left me alone. He did not understand the subject but he always said I could do it.

To my little cousin Beatriz, because of she I chose the theme about wildfires. Beatriz is afraid of wildfires and she said this project was very important to her because it would help the stakeholders to prevent wildfires.

I would also like to thank to my colleagues of Transnautica, in specially, my manager Cláudia for all support and encourage.

Lastly, I would also like to thank to my all friends and colleagues of MADSAD for all advices and for helping me to keep my mind off of the problems when I need it.

Abstract

Portugal has been affected by wildfires for many years. However, the wildfires were specially severe in 2017, causing a big number of deaths and, consequently, panic in population.

Our main goal is try to give more information to understand their occurrences. Therefore, we will estimate the time occurrences of wildfires in Continental Portugal.

The Poisson Process is used to model the occurrence of rare events like storm water overflows and understand how their occurrences appear scattered on the time line.

The model has two types: homogeneous Poisson Process and inhomogeneous Poisson Process. Both types are related because a deterministic time change can transform a homogeneous Poisson process into an inhomogeneous Poisson Process, and vice versa. We are going to study their properties and we are going to apply the theoretical results of the Poisson Process to a real database, which includes the records of wildfires between 2009 and 2015 in Continental Portugal. The analysis revealed that the adequacy of the model is strongly sensitive to factors such as seasonality and interdependence between occurrences.

Keywords: Poisson Process, Homogeneous Poisson Process, Inhomogeneous Poisson Process, Stochastic Processes

Resumo

Portugal foi afetado por incêndios florestais durante muitos anos. Contudo, os incêndios florestais foram muito severos no ano 2017 provocando algumas mortes e, consequentemente, pânico na população.

O nosso principal objetivo é tentar fornecer mais informação para prevenção dos incêndios florestais. Assim sendo, nós iremos estimar o tempo das ocorrências dos incêndios florestais em Portugal Continental.

O processo de Poisson é utilizando para modelar a occorrência de eventos raros como inundações provocadas por tempestades e a disposição temporal dos mesmos.

O modelo tem duas vertentes: processo de Poisson homogéneo e o processo de Poisson não homogéneo. Ambos estão relacionados pois uma mudança deterministica no tempo pode tranformar um processo de Poisson homogéneo num processo de Poisson não homogéneo e vice-versa. Nós iremos estudar as suas propriedades e iremos aplicar os resultados téoricos do processo de Poisson a uma base dados real que contém as ocorrências dos incêndios florestais em Portugal Continental entre 2009 e 2015. Durante a análise, nós encontramos fatores que influenciam o modelo tais como: a sazonalidade e a interdependência entre as occorrências.

Palavras-chave: Processo de Poisson, Processo de Poisson Homogéneo, Processo de Poisson não homogéneo, Processos Estocásticos

List of Figures

List of Tables

Contents

Chapter 1

Introduction

The dissertation is based on events that have particularly affected Portugal during the last years, which are wildfires. There have been severe wildfires in Portugal that caused panic in the population. It is necessary to help stakeholders with relevant information to prevent it.

We decided to consider the applicability of the Poisson Process as a model for the time occurrences of wildfires. The Poisson Process is used to model the occurrence of rare events like storm water overflows see Gallager (2012) and the time points at which events happen in a given time interval (Tse (2014)).

In the Literature Review, we present the model and its properties. The Poisson Process has two types: homogeneous Poisson Process and inhomogeneous Poisson Process. The Poisson Process is homogeneous when its expected value is a linear function. Otherwise, it is considered an inhomogeneous Poisson Process. Nevertheless, a deterministic time change can transform a homogeneous Poisson process into an inhomogeneous Poisson Process, and vice versa. We take the opportunity to explain some important properties: the first one is a Markov property which says that the past and future are conditionally independent; then we will identify the Poisson Process as a Renewal process, which means that the inter-arrival times are an iid sequence of exponential random variables; and lastly we study thoroughly the order statistics property, which says that for a Poisson Process with intensity λ, the distribution of the occurrence times on the interval $(0, t)$ is uniform between 0 and t.

In the next chapter we try to fit the model to a real database which has the occurrences of wildfires in Continental Portugal in the period from January 1, 2009, until December 31, 2015. The database is available on the website of the *Instituto da Conservação da Natureza e das Florestas* (ICNF). It has 131199 records and several information. We used the alert hour and date of occurrence of each wildfire in order to obtain the arrivals and inter-arrival times and for categorizing data we constructed a new variable that divided the country into 3 regions: the North, Center and South. After that, we analyzed if the model is appropriate for the whole period and all regions. We found the presence of some factors such as seasonality (since wildfires occur in specific season) and the interdependence between occurrences (we cannot guarantee the events are mutually independent because wildfires can cause new ones). Then, we decided to study if the model is suitable in a region which has less events and in a

shorter period of time. We started to study the South Region which has less records, during the period from the 1st May until the 30th September of each year. The model provided a good fitting. We turned to the Center Region and in the last sub-chapter we presented the results for the North Region and compared with the previous ones.

Lastly, we present the conclusions of our analysis.

Chapter 2

Motivation

Wildfires are considered one of the most dangerous natural disasters around the world. Every year several countries are affected by this phenomenon and populations say that nothing was done by public institutions to prevent it.

Governments said that it is urgent to take action to prevent the wildfires and they applied some measures, such as, buying airplanes, hiring more fire fighters and forcing populations to clean their forests.

Portugal has been affected by wildfires for years. However in 2017 the forest fires had severe consequences: more than 440 thousand hectares of forest burnt and approximately 150 people died. The Portuguese government said that it was urgent to take action to prevent the wildfires, however, in 2018 the forest fires reached other populations and although the consequences were not as dramatic as in the previous year, the forest fires go on being one of the biggest problems in our country.

Wildfires have several consequences like causing the decrease of oxygen, the increase of carbon dioxide, the increase of water pollution and soil pollution, the acceleration of the recycling time of important minerals in the ashes and the change of the species composition and hence the landscape. The consequences lead to a decrease in the quality of life of the populations (Møller and DÍAZ-AVALOS (2010)).

I hope the results of our study help stakeholders and decisions makers to make better decisions to prevent it.

Chapter 3

Literature Review

In this chapter will be presented the homogeneous Poisson Process, the inhomogeneous Poisson process and their properties. This chapter is based on the references Mikosch (2009) and Mingola (2013).

The Poisson process has been applied to Risk Theory since Filip Lundberg exploited it as model for the claim number processes. Filip Lundberg introduced a simple model that is capable of describing a homogeneous insurance portfolio.

1. Claims happen at the times T_i satisfying $0 \leq T_1 \leq T_2 \leq T_{...} \leq T_n$. We call them claim arrivals or claim times or claim arrival times or simply, arrivals;

2. The ith claim arrival at time T_i causes the claim size or claim severity X_i. The sequence (X_i) constitutes an independent and identically distributed (iid) sequence of non-negative random variables;

3. The claim size process (X_i) and the claim arrival process (T_i) are mutually independent.

The iid property of the claim sizes, X_i, reflects the fact that there is a homogeneous probabilistic structure in the portfolio. Now we can define the claim number process:

$$N_t = N(t) = \#\{i \geq 1 : T_i \leq t, t \geq 0\} \tag{3.1}$$

i is the occurrence, $i = 1$ corresponds to the first occurrence, N is a counting process on $[0, \infty)$, so $N(t)$ is the number of occurrences that occurred up to time t and let us define $N(t_{i-1}, t) = N(t) - N(t_{i-1})$ as the number of arrivals in the interval $(t_{i-1}, t]$.

A stochastic process $N = (N(t))_{t \geq 0}$ is a Poisson process if the following conditions are satisfied:

1. The process starts at zero: $N(0) = 0$;

2. The process has independent increments: for any t_i, i=0,...,n, and $n \geq 1$ such that $0 = t_0 < t_1 < ... < t_n$, the increments $N(t_{i-1}, t_i)$, $i = 1, ..., n$, are mutually independent;

3. There is a non-decreasing right-continuous function $\mu : [0, \infty) \rightarrow [0, \infty)$ with $\mu\{0\} = 0$ such that the increments $N(s,t)$ for $0 < s < t < \infty$ have a Poisson distribution $Po(\mu(s,t))$. We call μ the mean value function of N;

4. With probability 1, the sample paths $(N_t(\omega))_{t \geq 0}$ of the process N are right-continuous for $t \geq 0$ and have limits from the left for $t > 0$.

N_t can be thought of as the number of arrivals up to time t or the number of occurrences up to time t (Tse (2014)).

The Poisson process will be homogeneous if the following condition is satisfied:

$$\mu(t) = \lambda t, t \geq 0 \tag{3.2}$$

The quantity λ is the intensity or rate of the homogeneous Poisson process and is equal to the mean number of events occurring during the unit interval (Dobrow (2016)). If $\lambda = 1$, N is called a standard homogeneous Poisson process. When $\mu(t)$ is not a linear function, the Poisson process is said to be inhomogeneous because the intensity varies in time $\lambda(t)$.

In general, the literature says that N has an intensity function or rate function λ if μ is absolutely continuous, with respect to Lebesgue measure, i.e., for any $s < t$ the increment $\mu(s,t]$ has the following representation:

$$\int_s^t \lambda(y)dy, s < t, \tag{3.3}$$

for some non-negative measurable function λ.

If N is homogeneous, time evolves linearly: $\mu(s,t] = \mu(s+h, t+h]$ for any h>0 and $0 \leq s < t < \infty$. So, this suggests that claims arrive approximately uniformly over time.

If N has non-constant intensity function λ time *slows down* or *speeds up* taking into consideration the magnitude of $\lambda(t)$. When the event tends to occur in a specific season, then λ is non-constant.

A homogeneous Poisson process with intensity λ has:

1. With probabibilty 1, the sample paths $(N_t(\omega))_{t \geq 0}$ of the process N are right-continuous for $t \geq 0$ and have limits from the left for $t > 0$;

2. starts at zero, with probability 1;

3. has independent and stationary increments, if the intervals have the same length, the probability of occurring n events is the same in both intervals;

4. Then, $N(t)$ is a Poisson random variable with parameter λt, i.e.,

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{3.4}$$

An example of a homogeneous Poisson process is the Cramér-Lundberg model which is one of the most used models in non-life insurance mathematics.

After defining homogeneous Poisson process, we will see some of its properties. The first one is the Markov Property.

The Poisson process is a particular case of a Markov process on $[0, \infty)$ with state space $N_0 = \{0, 1, \dots\}$.

A stochastic process is said to be a Markov process if and only if the following condition is satisfied:

$$P\big(N(t_n) = k_n | N(t_1) = k_1, \dots, N(t_{n-1}) = k_{n-1}\big) = P\big(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}\big) \tag{3.5}$$

The Markov property says the past and the future are conditionally independent (Marais et al. (2010)).

This is a consequence of the independent increment property. In fact, if we have a Poisson Process with $N = (N(t))_{t \geq 0}$ which has a continuous intensity function λ on $[0, \infty)$, we have

$$\lambda_{k,k+h}(t) = \begin{cases} \lambda(t) & \text{if} \quad h = 1, \\ 0 & \text{if} \quad h > 1 \end{cases}$$

for $k \geq 0$.

In other words, we say that the intensity function $\lambda(t)$ of the Poisson Process N is the intensity of the Markov process N for the transition from state k to state $k+1$. The probability of at least one event occurring in a time interval length h is:

$$P\big(N((t, t+h]) \geq 1\big) = \lambda h + o(h) \text{ as } h \downarrow 0$$

$o(h)$ is a general and unspecified remainder term of smaller order than h, that is $o(h)/h \to 0$. So, we can say that λ in a Poisson Process is like the proportionality constant in the probability of an event occurring during an arbitrary small interval h (Pkj (2007)). Furthermore in an interval of length h the probability of happening two or more events is:

$$\big(P(N((t, t+h]) \geq 2) = o(h), \quad h \downarrow 0 \tag{3.6}$$

And the probability of occurring one event is:

$$\big(P(N(t, t+h] = 1) = \lambda(t)h(1 + o(1)) \tag{3.7}$$

The equations 3.6 and 3.7 make sure that a Poisson Process N with continuous intensity function λ is unlikely to have jumps larger than 1. It is intuitively a Markov process because of the memoryless feature of exponential variables.

3.1 Inhomogeneous Poisson Process

In the following to understand more properties we need to explain the relation between the homogeneous and inhomogeneous Poisson process. They are very closely related because a

deterministic time change can transform a homogeneous Poisson process into an inhomogeneous Poisson process, and vice-versa.

Let μ be the mean value function of a Poisson process N and $\widetilde{N}$ be a standard homogeneous Poisson process. Then the following conditions are satisfied:

1. The process $(\widetilde{N}(\mu(t)))_{t \geq 0}$ is Poisson with mean value function μ;

2. If μ is continuous, increasing and $\lim_{t \to \infty} \mu(t) = \infty$ then $(N(\mu^{-1}(t)))_{t \geq 0}$ is a standard homogeneous Poisson process.

If λ is piecewise constant, which means that μ is piecewise linear, then, in order to transform an inhomogeneous Poisson Process into a standard homogeneous Poisson process, we have to transform the arrivals according to the equation:

$$\mu(T_i) = (T_i - T_{i_{lastperiod}})\lambda_{period} + \mu(T_{i_{lastperiod}}) \tag{3.8}$$

Therefore, $T_{i_{lastperiod}}$ is the last arrival that happened in the previous period.

These conditions suggest a simple way of simulating paths of an inhomogeneous Poisson process $\widetilde{N}$ from the paths of a homogeneous Poisson process.

One of the goals of this dissertation is to try to model the number of occurrences and their time occurrences. So we are going to study the sequence of the arrival times $0 \leq T_1 \leq T_2 \leq \ldots$ of a homogeneous Poisson process. To study the arrival times we need to see the Poisson process as a renewal process.

A Renewal process is an arrival process in which the inter-arrival intervals are positive, independent and iid random variables (Gallager (2012)).

$$N(t) = \#\{i \geq 1 : T_i \leq t\}, t \geq 0, \tag{3.9}$$

where,

$$T_n = W_1 + \cdots + W_n, n \geq 1, \tag{3.10}$$

and W_i is an iid sequence of exponential random variables, with parameter λ.

Theorem 1. *The homogeneous Poisson process is a renewal process, if and only if:*

1. The process N given by 3.9 and 3.10 with an iid exponential with parameter λ sequence (W_i) constitutes a homogeneous Poisson process with intensity λ.

2. N is a homogeneous Poisson process with intensity λ and arrival times $0 \leq T_1 \leq T_2 \leq \ldots$. Then N has a representation given by equation (3.9), and (T_i) has a representation given by equation (3.10) for an iid sequence of exponential random variables with parameter λ.

We provide here a simple argument that gives the Poisson property of N given by 3.9 and 3.10, which is part of the proof of statement 1. See Mikosch (2009) for the full proof of the Theorem.

The following relationship is crucial:

$$\{N(t) = n\} = \{T_n \leq t < T_{n+1}\}, n \geq 0 \qquad (3.11)$$

If we consider the sum of n iid exponentially distributed with parameter λ random variables ($T_n = W_1 + \cdots + W_n$), then a classical result says that T_n has a gamma distribution with parameters n and λ for $n \geq 1$, i.e,

$$P(T_n \leq x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} \qquad (3.12)$$

Hence,

$$P(N(t) = n) = P(T_n \leq t) - P(T_{n+1} \leq t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \qquad (3.13)$$

This proves the Poisson property of $N(t)$.

If we consider a homogeneous Poisson process as a renewal process then the inter-arrival times are equal to:

$$W_i = T_i - T_{i-1}, i \geq 1, \qquad (3.14)$$

According to the law of the large numbers, we have that $\frac{T_n}{n} \overset{\text{a.s.}}{\to} EW_1 = \lambda^{-1} > 0$, which means that T_n grows approximately like $\frac{n}{\lambda}$ and there are no limit points in the sequence (T_n) at any finite instant of time. This means that the values $N(t)$ of a homogeneous Poisson Process are finite a.s. on any finite time interval [0,t].

An inhomogeneous Poisson process N with mean value function μ can be understood as a deterministic time changed standard homogeneous Poisson Process $\tilde{N}$:

$$(N(t))_{t \geq 0} \overset{\text{d}}{=} (\tilde{N}(\mu(t)))_{t \geq 0} \qquad (3.15)$$

Let $(\tilde{T}_i)_i$ be the arrival sequence of $\tilde{N}$ and μ be increasing and continuous. Therefore there is an inverse μ^{-1} and

$$\tilde{N}(t) = \#\{i \geq 1 : \tilde{T}_i \leq \mu(t)\} = \#\{i \geq 1 : \mu^{-1}(\tilde{T}_i) \leq t\}, t \geq 0 \qquad (3.16)$$

Therefore, the arrival times of an inhomogeneous Poisson Process with mean value function μ have the representation:

$$T_n = \mu^{-1}(\tilde{T}_n), \quad \tilde{T}_n = \widetilde{W}_1 + \cdots + \widetilde{W}_n, \quad n \geq 1, \quad \widetilde{W}_i \quad iid \quad Exp(1). \qquad (3.17)$$

Let N be a Poisson process on $[0, \infty)$ with a continuous a.e. positive intensity function λ. Then the following conditions are satisfied:

1. The vector of the arrival times $(T_1, \ldots, T_n)$ has density:

$$f_{T_1,\ldots,T_n}(x_1, \ldots, x_n) = e^{-\mu(x_n)} \prod_{i=1}^{n} \lambda(x_i) I_{\{0 < x_1 < \cdots < x_n\}} \qquad (3.18)$$

2. The vector of inter-arrival times $(W_1, \ldots, W_n) = (T_1, T_2 - T_1, \ldots, T_n - T_{n-1})$ has density:

$$f_{W_1,\ldots,W_n}(x_1, \ldots, x_n) = e^{-\mu(x_1+\cdots+x_n)} \prod_{i=1}^{n} \lambda(x_1 + \cdots + x_i), x_i \geq 0. \qquad (3.19)$$

This means that the inter-arrival times $W_1, \ldots, W_n$ are independent (and identically distributed) only in the case of a homogeneous Poisson Process. This property distinguishes the homogeneous Poisson process within the class of all Poisson processes on $[0, \infty)$.

3.2   Order Statistics Property

The next feature that will be present is one of the most important properties of the Poisson Process. It is the order statistics property. This property says the locations of events are uniformly distributed when conditioning on a fixed total number of events in a certain interval.

Consider the next example proposed by Pinsky and Karlin (2010): we start with a line segment t units long and a fixed number n of darts. We throw the darts along the line segment in such a way that their position is uniformly distributed along the segment, independently of the other darts and their location. Let U_1 be the position of the first dart, U_2 the position of the second, and so on up to U_n. The probability density function is the uniform density:

$$f_U(\mu) = \begin{cases} 1/t & \text{for} \qquad 0 \leq \mu \leq t, \\ 0 & \text{elsewhere.} \end{cases}$$

Now let $T_1 \leq T_2 \leq \cdots \leq T_n$ denote these same positions, not in the order in which the darts were thrown, but instead in the order in which they are placed along the line. The next figure 3.1 shows a typical relation between $U_1, U_2, \ldots, U_n$ and $T_1, T_2, \ldots, T_n$. The joint probability density function for $T_1, T_2, \ldots, T_n$ is:

$$f_{T_1,\ldots,T_n}(x_1, \ldots, x_n) = n! t^{-n}, 0 < x_1 < \cdots < x_n < t \qquad (3.20)$$

For example, if one consider n=2 we have:
$$f_{T_1,T_2}(x_1, x_2) dx_1 dx_2 = 2t^{-2} dw_1 dw_2$$

Dividing by $dx_1 dx_2$ and passing to the limit gives 3.20. When n=2, there are two ways that U_1 and U_2 can be ordered; either U_1 is higher than U_2 or U_1 is less than U_2. There are usually $n!$ arrangements of $U_1, \ldots, U_n$ that lead to the same ordered values $T_1 \leq \cdots \leq T_n$.
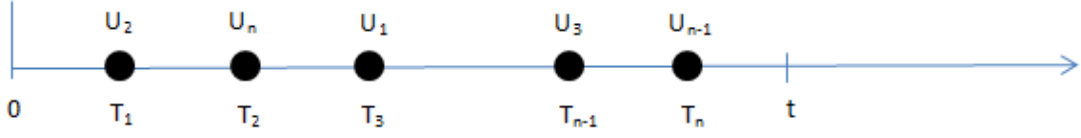
Figure 3.1: $T_1, T_2, \ldots, T_n$ can be seen as the values of $U_1, U_2, \ldots U_n$ displayed in increasing order.

Theorem 2. *Consider the Poisson process $N = (N(t))_{t \geq 0}$ with continuous a.e. positive intensity function λ and arrival times $0 < T_1 < T_2 < \ldots$ a.s. Then the conditional distribution of $(T_1, \ldots, T_n)$ given $\{N(t) = n\}$ is the distribution of ordered sample $(X_{(1)}, \ldots, X_{(n)})$ of an iid sample $X_1, \ldots, X_n$ with common density $\frac{\lambda(x)}{\mu(t)}, 0 < x \leq t$,*

$$(T_1, \ldots, T_n | N(t) = n) \overset{d}{=} (X_{(1)}, \ldots, X_{(n)}) \tag{3.21}$$

This means that the left-hand vector has conditional density:

$$f_{T_1, \ldots, T_n}(x_1, \ldots, x_n | N(t) = n) = \frac{n!}{(\mu(t))^n} \prod_{i=1}^{n} \lambda(x_i), \quad 0 < x_1 < \cdots < x_n < t \tag{3.22}$$

In the case of a homogeneous Poisson process, we have:

$$f_{T_1, \ldots, T_n}(x_1, \ldots, x_n | N(t) = n) = n! t^{-n}, 0 < x_1 < \cdots < x_n < t \tag{3.23}$$

Therefore, if there are n arrivals of a homogeneous Poisson process in the interval $(0, t)$, these arrivals constitute the points of an ordered sample of n uniformly distributed random variables on $(0, t)$. This conclusion is independent of the intensity λ.

Proof. To prove this property, we start by giving a well-known result on the distribution of the order statistics:

$$X_{(1)} \leq \cdots \leq X_{(n)},$$

of an iid sample $X_1, \ldots, X_n$.

If the iid $X_i's$ have density f then the density of the vector $(X_{(1)}, \ldots, X_{(n)})$ is given by:

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i) I_{\{x_1 < \cdots < x_n\}} \tag{3.24}$$

The support of the vector $(X_{(1)}, \ldots, X_{(n)})$ is the set:

$$C_n = \{(x_1, \ldots, x_n) : x_1 \leq \cdots \leq x_n\} \subset \mathbb{R}^n. \tag{3.25}$$

The existence of a density $f_{X_i, \ldots, X_n}$ implies that all elements of the iid sample $X_1, \ldots, X_n$ are different a.s., which means that in definition of C_n the $\leq$ can be replaced by $<$.

10

In what follows, we will give the proof of (3.21) and (3.22). The proof starts with a limit:

$$lim_{h_i \downarrow 0, i=1,\ldots,n} \frac{P(T_1 \in (x_1, x_1+h_1], \ldots, T_n \in (x_n, x_n+h_n] | N(t)=n)}{h_1, \ldots, h_n}$$

The limit can be interpreted like a density for the conditional probability distribution of $(T_1, \ldots, T_n)$, given $\{N(t) = n\}$.

One can choose the $h_i's$ so small that the intervals $(x_i, x_i + h_i] \subset [0, t], i = 1, \ldots, n$, become disjoint with $0 < x_1 < \cdots < x_n < t$. Then the following expression is obtained:

$$
\begin{aligned}
&\left\{ T_1 \in (x_1, x_1 + h_1], \ldots, T_n \in (x_n, x_n + h_n], N(t) = n \right\} \\
&= \left\{ N(0, x_1] = 0, N(x_1, x_1 + h_1] = 1, N(x_1 + h_1, x_2] = 0, \right. \\
&\quad N(x_2, x_2 + h_2] = 1, \ldots, N(x_{n-1} + h_{n-1}, x_n] = 0, \\
&\quad \left. N(x_n, x_n + h_n] = 1, N(x_n + h_n, t] = 0 \right\}
\end{aligned}
\tag{3.26}
$$

Taking probabilities on both sides and using the property of independent increments of the Poisson Process N, we obtain:

$$
\begin{aligned}
&P\big(T_1 \in (x_1, x_1 + h_1], \ldots, T_n \in (x_n, x_n + h_n], N(t) = n\big) \\
&= P(N(0, x_1] = 0)\, P(N(x_1, x_1 + h_1] = 1)\, P(N(x_1 + h_1, x_2] = 0) \\
&\quad P(N(x_2, x_2 + h_2] = 1) \cdots P(N(x_{n-1} + h_{n-1}, x_n] = 0) \\
&\quad P(N(x_n, x_n + h_n] = 1)\, P(N(x_n + h_n, t] = 0) \\
&= e^{-\mu(x_1}[\mu(x_1, x_1 + h_1]\, e^{-\mu(x_1, x_1 + h_1]}]\, e^{-\mu(x_1 + h_1, x_2]}[\mu(x_2, x_2 + h_2] \\
&\quad e^{-\mu(x_2, x_2 + h_2]}] \cdots e^{-\mu(x_{n-1} + h_{n-1}, x_n]}[\mu(x_n, x_n + h_n] \\
&\quad e^{-\mu(x_n, x_n + h_n]}]\, e^{\mu(x_n + h_n, t]} \\
&= e^{-\mu(t)}\mu(x_1, x_1 + h_1] \cdots \mu(x_n, x_n + h_n]
\end{aligned}
\tag{3.27}
$$

Dividing by $P(N(t) = n) = e^{-\mu(t)}\frac{(\mu(t))^n}{n!}$, we get the scaled conditional probability:

$$
\begin{aligned}
&\frac{P(T_1 \in (x_1, x_1+h_1], \ldots, T_n \in (x_n, x_n+h_n] | N(t)=n)}{\prod\limits_{i=1}^{n} h_i} \\
&= \frac{n!}{(\mu(t))^n} \frac{\mu(x_1, x_1+h_1]}{h_1} \cdots \frac{\mu(x_n, x_n+h_n]}{h_n} \\
&\to \frac{n!}{(\mu(t))^n}\ \lambda_{(x_1)}, \ldots, \lambda_{(x_n)}, \quad as \quad h_i \downarrow 0, \quad i = 1, \ldots, n
\end{aligned}
$$

In the last step, we apply the continuity of λ to show that $\mu'(x_i) = \lambda(x_i)$.

Then we obtained equation 3.23. □

Chapter 4

Pre-processing Data and Solution Approach

In this chapter, we are going to apply the theoretical results of the Poisson Process to a real database which has the occurrences of wildfires in Continental Portugal in the period from January 1, 2009, until December 31, 2015.

The database is available on the ICNF website and it has 131199 records. We would like to explore data about 2017 but public institutions have not analyzed it, yet.

The database has several information, but essentially we used the alert hour and date of occurrence of each wildfire in order to create a variable T_i, which keeps record of the elapsed time in hours since the first day (1st January of 2009) until the occurrence of that particular wildfire. Based on this variable T_i we defined the waiting times, W_i according to equation 3.14.

We also created a variable in order to categorize data according to the region where the wildfire took place. We split the country into 3 regions by grouping the following districts:

1. North Region: Viana do Castelo, Porto, Braga, Vila Real, Bragança, Guarda, Viseu;

2. Center Region: Aveiro, Coimbra, Santárem, Castelo Branco, Portalegre, Leiria;

3. South Region: Lisboa, Setúbal, Beja, Faro, Évora.

4.1   Continental Portugal

In this section we are going study if the theoretical results adjust to a real database in several years. We used SPSS, Excel and R to analyze the data.

We plotted the arrivals (Figure 4.1a) with a straight line that has a slope 0.47. According to the law of large numbers, the slope represents $\hat{\lambda}^{-1} = T_n/n$. The $\hat{\lambda}$ is the maximum likelihood estimator of λ under the hypothesis that the inter-arrival times W_i are iid Exp(λ). We assessed the validity of this statement by applying a Kolmogorov Smirnov test. This method

is implemented when one wants test if a given sample follows a certain distribution. In this case we want to know if W_i follows an exponential distribution with $\lambda = 1/0.47 = 2.13$, i. e., we consider

$$H_0 = W_i \sim Exp(2.13)$$
$$H_1 = W_i \nsim Exp(2.13)$$

and obtained a $p - value = 0.2976 > 0.01$, we do not reject null hypothesis and for a significance level 1% and since the p-value is quite large we have some indication that the inter-arrival times may follow an exponential distribution. Then we plotted the corresponding inter-arrival times (Figure 4.1b).

(a) Arrivals   (b) Inter-arrival times
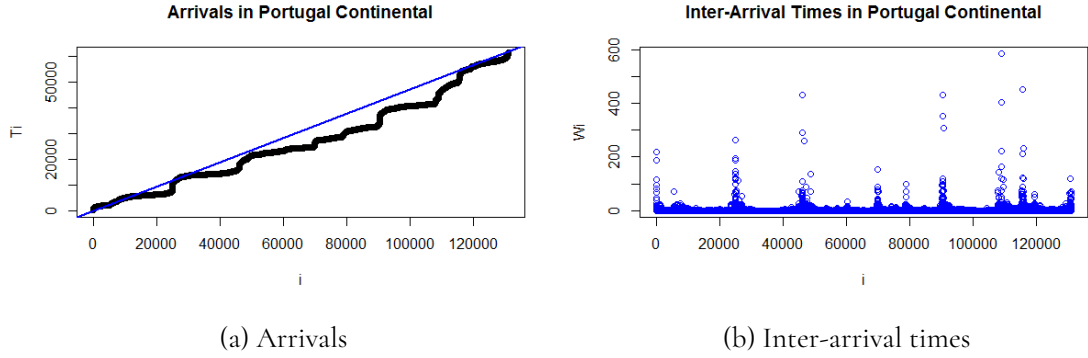
Figure 4.1: Arrivals and corresponding inter-arrival times in Continental Portugal

We also built a table (Table 4.1) with the main statistics of the inter-arrival times for each year, for the whole period and all regions.

As mentioned earlier, the inverse of the mean corresponds to an estimator of the intensity function. The table suggests that during the years there is a tendency for an intensity decrease. We can prove this statement taking into consideration time series methods that analyze trends and seasonal components. The moving average is a non-parametric approach that estimates a deterministic trend which influenced a stationary time series based on the past values (Brockwell et al. (2002)). The estimate of the mean inter-arrival time is defined by:

$$(\hat{\lambda}(i))^{-1} = (2m+1)^{-1} \sum_{j=max(1,i-m)}^{min(n,i+m)} W_j \quad for \quad m = 3000. \tag{4.1}$$

The corresponding estimates for $(\hat{\lambda}(i))^{-1}$ can be interpreted as the reciprocals of the values of the Poisson intensity. We plotted the annual estimates supported by moving average estimate $(\hat{\lambda}(i))^{-1}$ and the corresponding reciprocals and there is a tendency for the intensity to decrease when time goes by (Figure 4.2). Moreover we presented the boxplots of each year

where one can see that the distribution of inter-arrival times of the claims is more spread towards the end of 2015 and concentrated around the value 0.15 in contrast to 0.08 at the beginning (Figure 4.3). Furthermore, the annual claim number decreased over the years.

Year	2009	2010	2011	2012	2013	2014	2015	Total
sample size	24816	21241	23977	20374	18633	6800	15358	131199
minimum	0	0	0	0	0	0	0	0
1st quartile	0.03	0.03	0.03	0.03	0.03	0.10	0.05	0.03
median	0.08	0.07	0.08	0.10	0.10	0.27	0.15	0.10
mean	0.35	0.41	0.37	0.43	0.47	1.29	0.57	0.47
$\hat{\lambda}$=1/mean	2.83	2.43	2.74	2.32	2.13	0.77	1.76	2.14
3rd quartile	0.20	0.20	0.22	0.25	0.23	0.67	0.37	0.25
maximum	218.65	291.60	432.00	352.82	429.55	584.43	230.67	584.43

Table 4.1: Descriptive Statistics of Continental Portugal
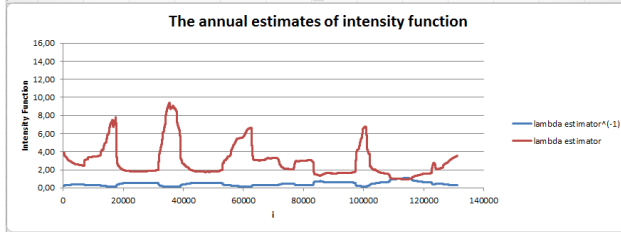
Figure 4.2: The annual estimates by moving average estimate $(\hat{\lambda}(i))^{-1}$ and their reciprocals
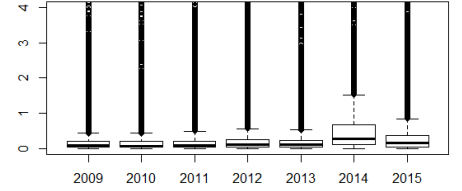
Figure 4.3: Boxplots of the inter-arrival times for each year and all regions together

According to the literature review, a homogeneous Poisson Process has a constant λ in the time interval of analysis. Then, we have obtained statistical evidence that the arrival times of wildfires in Continental Portugal are possibly modeled by an inhomogeneous Poisson Process.

However, there is a relation between an inhomogeneous and homogeneous Poisson Process. As mentioned above, if one transforms the arrivals T_i into $\mu(T_i)$, the inhomogeneous Poisson Process will become a standard homogeneous Poisson Process. The difficulty is then the estimation of μ. We calculated $\mu(T_i)$ using the equation 3.8, assuming that $\mu(t)$ is piecewise linear, which means that we considered that λ is piecewise constant. We used the usual estimator of λ during the intervals of constancy. In our first approach to the global analysis, we assumed these periods corresponded to each year. To analyze the performance of our estimates we assessed the quality of the standard homogeneous fitting by verifying the following criteria.

1. The homogeneous Poisson Process has a linear function; To prove this statement we plotted the transformed arrivals $\mu(T_i)$ (Figure 4.4).

2. According to previous chapter in the homogeneous Poisson Process the claims arrive approximately uniformly over time. Then, we did the histogram of the values $\mu(T_i)$ (Figure 4.5).

3. For the homogeneous Poisson Process the inter-arrival times follow an exponential distribution. Therefore, we draw a QQ-plot of the values $\mu(T_i) - \mu(T_{i-1})$ against the standard exponential distribution (Figure 4.6).

4. The homogeneous Poisson Process is standard if λ is equal to 1. We plotted the estimates of the intensity function using the approach of time series with horizontal line equal to 1 (Figure 4.7).

We were expected to observe a closely straight line in the Figure 4.4 however there are several jumps between arrivals. As we can see in the Figure 4.5 there is statistical evidence that transformed arrivals do not follow an uniform distribution. We can observe in the Figure 4.6 there is a clear indication of a right tail that is heavier than exponential. Finally in the Figure 4.7 the estimates vary wildly around 1. Indeed, there is evidence that the process may not be Poisson and other models probably could be more appropriate.

Possibly reasons for such poor agreement are: seasonality, as one knows wildfires tend to occur in days that temperatures are high and very windy, therefore during winter and autumn there are several periods of heavy rain that cause the huge differences between the arrivals as one can observe in Figure 4.1b; the second reason is related with the number observations. As we referred above the Poisson Process is a suitable approach for rare events which have a small probability of occurring. In six years we have 131199 records, therefore it is hard to say that wildfires are rare events; the last reason has to do with interdependence which is also related with the number of observations. If there are more wildfires there is a tendency to increase the number of events because the temperature increases and the residuals that were freed by wildfires may fall on soil and cause a new event, so we cannot conclude that the inter-arrivals are completely mutually independent, which is one of the conditions for a Poisson Process.

Consolidating our analysis, we decided to divide the country into three regions and we chose to study more thoroughly the South Region because it is the region that has less events and where there are less longer periods of heavy rainfall, which means less interdependence and ultimately a better fitting.

4.2   South Region

We are going to apply the theoretical results of the Poisson Process to the South Region between January 1, 2009, and December 31, 2015.

Firstly, we plotted the arrivals with a linear function $T_i = 4.25i$. If we observe the figure 4.8 we can see the jumps in arrivals, which may indicate that model is not suitable. We believe there is a strong seasonality in the data.
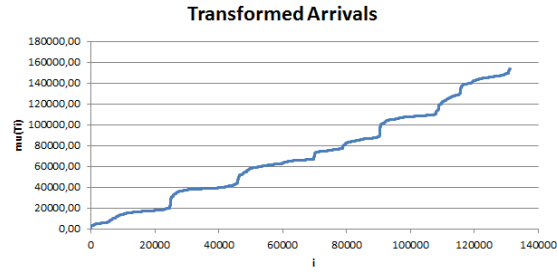
Figure 4.4: The transformed arrivals $\mu(T_i)$
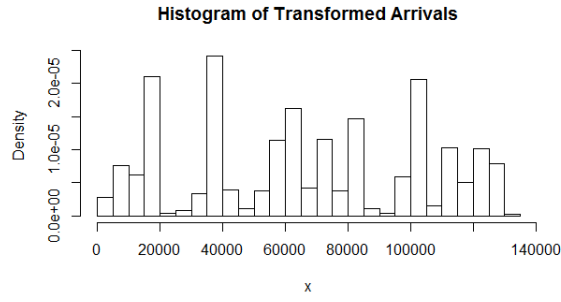
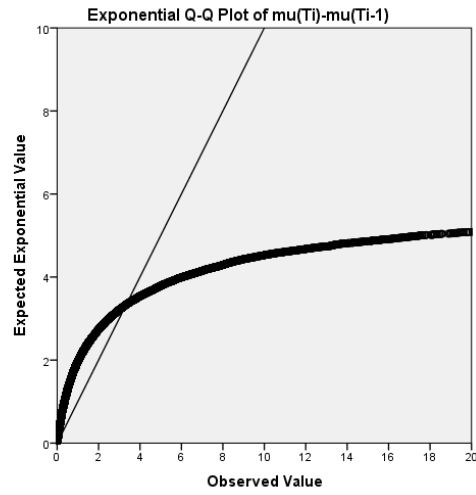Figure 4.5: The histogram of the values $\mu(T_i)$

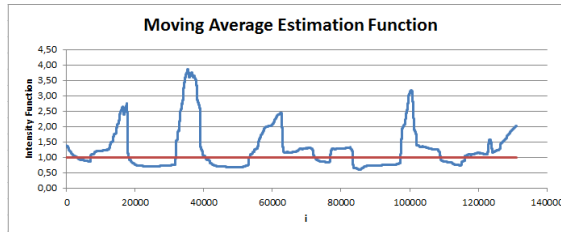Figure 4.6: QQ-plot of the values $\mu(T_i) - \mu(T_{i-1})$

Figure 4.7: Estimation of the intensity function corresponding to the transformed sequence $\mu(T_i)$

The seasonality is shown in the figures 4.9 and 4.10 where it is clear that most arrivals occurred between May and September. Therefore, we decided restrict the study to these periods ever year.
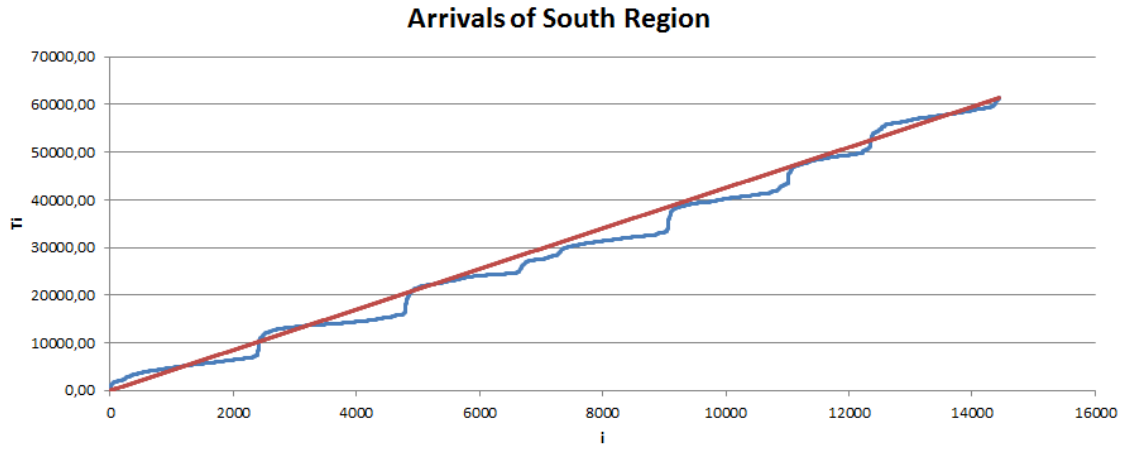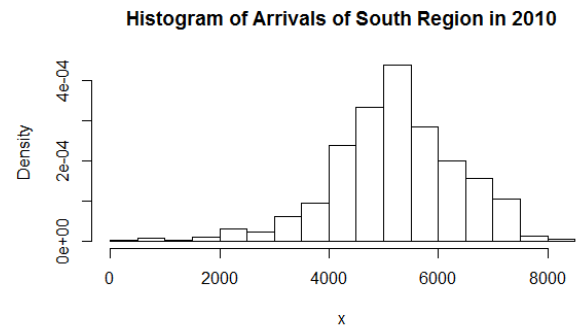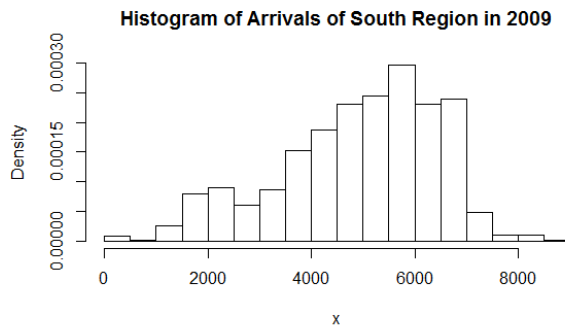

Figure 4.8: Arrivals of South Region

Figure 4.9: Histogram of arrivals of South in 2009   Figure 4.10: Histogram of arrivals of South in 2010

The first period between 1st May 2009 to 30th September we have 1811 occurrences which is more than 50% of all events recorded during this year. We divide it into 22 short periods which correspond to one week. Therefore, the first is 1st May 2009 until 7th May 2009, the second is 8th May 2009 until 15st May 2009, and so on.

Firstly we built a table with relevant information about each week of inter-arrival time (Table 4.2). As we can see in the table the intensity is not constant and varies a lot, showing a tendency to increase. Therefore, we calculated the intensity according to the moving average formula 4.1 with $m = 50$ because we decreased the number of events from 131199 to 1811 and we can observe that there is a clear tendency for the intensity to increase (Figure 4.11). Next, we draw the boxplots per week and we can observe in the first period that the distribution

Weeks	Sample Size	Mean	$\hat{\lambda} = 1/mean$
1	50	3.249	0.308
2	19	8.942	0.112
3	51	3.311	0.302
4	52	3.279	0.305
5	67	2.484	0.403
6	38	4.464	0.224
7	71	2.337	0.428
8	103	1.598	0.626
9	55	3.137	0.319
10	95	1.761	0.568
11	83	2.040	0.490
12	88	1.904	0.525
13	102	1.655	0.604
14	82	2.012	0.497
15	106	1.603	0.624
16	96	1.761	0.568
17	121	1.370	0.730
18	127	1.333	0.750
19	118	1.415	0.707
20	76	2.233	0.448
21	101	1.651	0.606
22	110	1.298	0.770
Total	1811	2.026	0.494

Table 4.2: Sample size and Mean about South Region between 1st May 2009 and 30th September 2009

of inter-arrival times is around the value 1.4 which is greater than in last week that is about the value 0.6 (Figure 4.12). Furthermore, the number of events increases over the weeks.

So, we have gained statistical evidence that the arrivals are modeled by an inhomogeneous Poisson Process. As mentioned above to transform an inhomogeneous Poisson Process into a homogeneous Poisson Process we need to calculate $\mu(T_i)$. Once $\mu(T_i)$ is computed taking into consideration equation 3.8 and using a constant λ per week, we will check the fitting with respect to a standard homogeneous Poisson Process after transforming the inhomogeneous process using the condition 2 of the beginning of Section 3.1.

Firstly we observed if the transformed process looks like a linear function. We plotted the transformed arrivals and compared with estimated mean value function of insurance arrivals $\mu(t)$. In the left graph we have the estimated mean value function $\mu(t)$. As one can see the slopes correspond to the intensities from Table 4.2. On the right we draw the transformed arrivals. We can observe that the transformed data is very much aligned with a straight line,
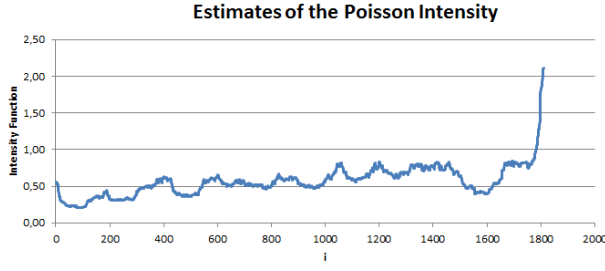
Figure 4.11: Estimates of intensity function by moving average approach
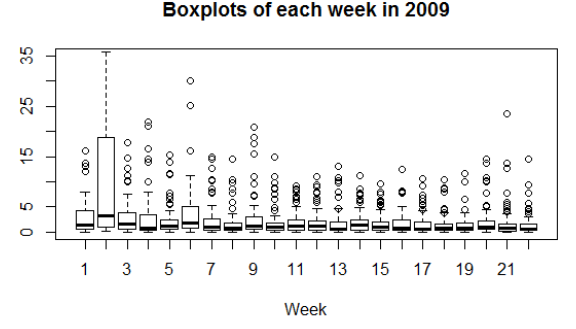
Figure 4.12: Boxplots for each periof of 2009

while the left graph shows several deviations of the arrivals from a straight line (Figure 4.13).
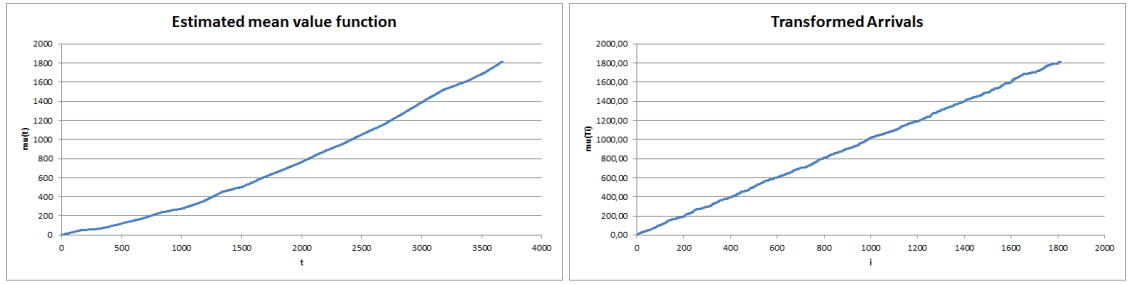
(a) Estimated mean value function $\mu(t)$

(b) Transformed arrivals $\mu(T_i)$

Figure 4.13: Comparison between estimated mean value function $\mu(t)$ and transformed arrivals $\mu(T_i)$

According to the literature the arrivals of a homogeneous Poisson Process follows an uniform distribution on any fixed interval, conditionally on the number of claims in this interval. To check it we plotted the histogram of the transformed arrivals and the histogram of arrivals. We can observe that the histogram of the transformed arrivals resembles a histogram of an uniform sample (Figure: 4.14).

Next, we focus on the distribution of inter-arrival times, which should be exponentially distributed. To check this statement we draw a QQ-plot of the distribution of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$ against the standard exponential distribution (Figure 4.15b). We also draw the QQ-plot of the distribution of W_i against the standard exponential distribution (Figure 4.15a). In both graphs there is a clear indication of a right tail of the distribution which is heavier than the tail of the exponential distribution. However there is a clear approximation of the straight line in the right graph, which means that assuming an exponential distribution for the inter-arrival times is not absurd.

Next, we focus on the constant intensity λ. According to literature if we change a inhomogeneous Poisson Process into homogeneous it will become a standard homogeneous

(a) Histogram of T_i
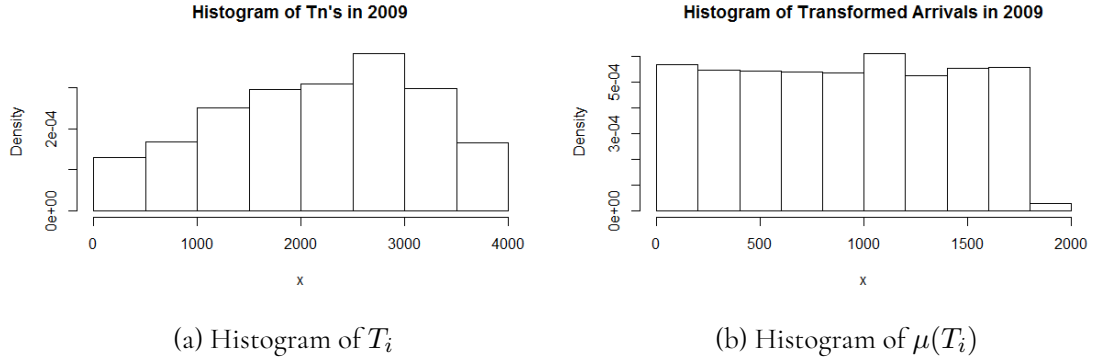
(b) Histogram of $\mu(T_i)$

Figure 4.14: Histogram of arrivals and histogram of transformed arrivals

(a) QQ-plot of the inter-arrival times W_i

(b) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$
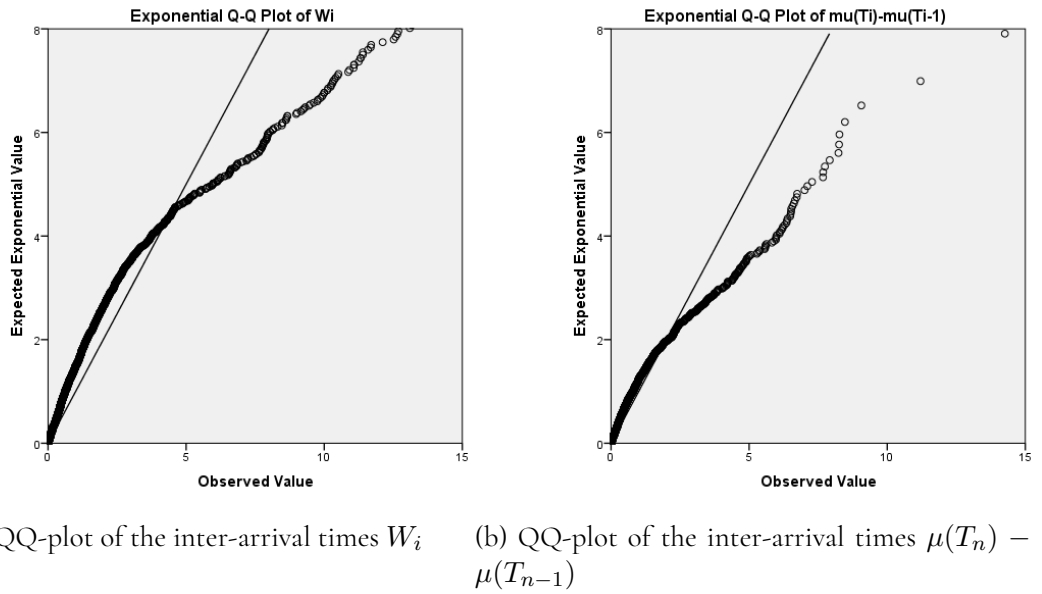
Figure 4.15: Comparison between QQ-plot of the inter-arrival times W_i and QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$

Poisson Process, which means that λ is equal to 1. We estimated λ by moving average approach using W_i (Figure 4.16a) and $\mu(T_n) - \mu(T_{n-1})$ (Figure 4.16b). We can observe in the right graph that the estimates of intensity function around the value 1 in contrast in the left graph the estimates of intensity function are increasing.
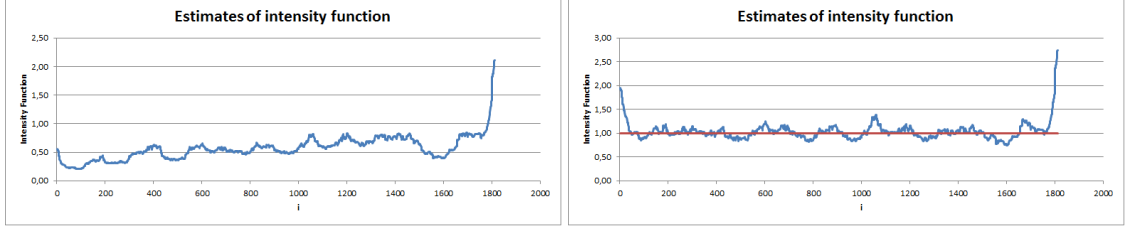
(a) Estimates of intensity function using W_i
(b) Estimates of intensity function using $\mu(T_n) - \mu(T_{n-1})$

Figure 4.16: Comparison between estimates of intensity function using W_i and using $\mu(T_n) - \mu(T_{n-1})$

We may conclude that an inhomogeneous Poisson Process provides an appropriate model for the period between the 1st May of 2009 and the 30th September of 2009.

In the following, we studied if the model is suitable for whole years in the same period, we followed the same steps that we did for 2009. Firstly we did a table with important information about the years (Table 4.3). The table shows us that there is not any year that the estimates of intensity function $\hat{\lambda} = 1/mean$ is constant, therefore the homogeneous Poisson Process is not appropriate in any year.

| Years	2010			2011			2012			2013			2014			2015		
Weeks	i	Mean	1/Mean	i	Mean	1/Mean	i	Mean	1/Mean	i	Mean	1/Mean	i	Mean	1/Mean	i	Mean	1/Mean
1	24	6.640	0.150	9	18.052	0.055	1	115.633	0.009	29	5.770	0.173	26	6.247	0.160	14	11.500	0.087
2	4	29.446	0.034	16	10.658	0.094	23	9.370	0.107	36	4.635	0.216	48	3.523	0.284	39	4.451	0.225
3	34	6.485	0.154	10	16.280	0.061	18	9.504	0.105	23	7.372	0.136	31	5.311	0.188	78	2.110	0.474
4	40	4.265	0.234	19	9.075	0.110	28	6.055	0.165	50	3.276	0.305	23	7.608	0.131	73	2.330	0.429
5	31	5.387	0.186	16	10.676	0.094	34	4.790	0.209	84	2.035	0.491	66	2.526	0.396	93	1.829	0.547
6	22	6.619	0.151	22	7.579	0.132	51	3.376	0.296	23	7.084	0.141	34	4.841	0.207	72	2.335	0.428
7	78	2.491	0.401	39	4.305	0.232	61	2.724	0.367	69	2.494	0.401	55	3.118	0.321	40	4.185	0.239
8	77	2.169	0.461	36	4.521	0.221	62	2.757	0.363	85	1.997	0.501	20	8.368	0.119	65	2.590	0.386
9	80	2.119	0.472	49	3.484	0.287	82	2.035	0.491	105	1.584	0.631	39	4.370	0.229	61	2.765	0.362
10	125	1.309	0.764	70	2.445	0.409	53	3.119	0.321	143	1.180	0.847	59	2.842	0.352	80	2.100	0.476
11	102	1.676	0.597	69	2.434	0.411	76	2.252	0.444	52	3.154	0.317	89	1.889	0.529	111	1.512	0.661
12	140	1.211	0.826	73	2.313	0.432	143	1.178	0.849	61	2.757	0.362	56	2.994	0.334	100	1.642	0.609
13	216	0.777	1.287	73	2.287	0.437	70	2.369	0.422	76	2.261	0.442	91	1.828	0.547	86	1.990	0.502
14	152	1.103	0.907	52	3.144	0.318	100	1.696	0.590	79	2.109	0.474	86	1.956	0.511	70	2.396	0.417
15	203	0.814	1.228	88	1.981	0.505	122	1.370	0.730	102	1.672	0.598	103	1.648	0.607	125	1.335	0.749
16	133	1.286	0.777	72	2.282	0.438	70	2.416	0.414	98	1.677	0.596	108	1.558	0.642	81	2.082	0.480
17	109	1.531	0.652	66	2.569	0.389	87	1.924	0.520	105	1.620	0.617	82	1.995	0.501	69	2.399	0.417
18	138	1.219	0.821	36	4.613	0.217	166	1.012	0.988	114	1.491	0.671	89	1.931	0.518	63	2.689	0.372
19	77	2.196	0.455	51	3.363	0.297	108	1.539	0.650	73	2.201	0.454	40	3.475	0.288	93	1.835	0.545
20	89	1.878	0.533	59	2.856	0.350	102	1.670	0.599	81	2.146	0.466	10	14.727	0.068	42	3.887	0.257
21	53	3.114	0.321	99	1.690	0.592	88	1.837	0.544	100	1.654	0.604	9	20.691	0.048	87	1.984	0.504
22	97	1.506	0.664	111	1.275	0.784	24	6.188	0.162	21	3.437	0.291	21	7.880	0.127	67	2.124	0.471
Total	2024	1.813	0.551	1135	3.232	0.309	1569	2.339	0.428	1609	2.235	0.447	1185	3.092	0.323	1609	2.281	0.438

Table 4.3: Main Statistics of whole years during the period 1st May until 30th September

Then, we are going to study if we transform the T_i into $\mu(T_i)$, whether the process is well

fitted by a standard homogeneous Poisson Process.

For each year, we draw the same graphs as we did for 2009 (Figures 4.17, 4.18, 4.19, 4.20, 4.21 and 4.22) to check if the fitting is good or not. In the left we put the inhomogeneous plots and the homogeneous graphs on the right.

We observed that, in all years, from May until September the Poisson fitting is reasonable. So, we grouped all years between May and September and obtained $\hat{\lambda} = 0.62$ which means that in each 0.62 hours a wildfire occur, i.e., every 37 minutes a new wildfire occur in the South Region.

However, the Poisson Process does not fit perfectly in some years. For example, in 2012 the estimates of intensity function using moving average approach in the beginning of period is very close to 1 in contrast to the end there is a widely variation of intensity function. This phenomenon happened because in the weeks 18 to 20, there were more than 100 wildfires in each week in contrast to the weeks 16 and 17 and weeks 21 and 22 when there were 157 and 112 wildfires occurred, respectively, in each 2 week period (Figure 4.19h). If we consider only the period between May and August we will obtain better fitting. According to a climate report from IPMA in the last week of September rained copiously.

Another year that the Poisson Process does not fit perfectly is 2015. As observed in figure 4.22c, we see the arrivals are more approximately uniformly distributed in the end of period in contrast to the beginning. If we analyzed the statistics of inter-arrival times we can verify the maximum of inter-arrival times are in the first week, if we analyze the mean of inter-arrivals times per month we can see on May the mean is 3.068 higher than other months that are respectively 2.518, 1.911, 1.965 and 2.303. If we exclude May the Poisson Process fitting will be better. According to a climate report from IPMA, in May, a heat wave started on the 9th May. In fact, the mean of inter-arrival times the 1st May to 8th May is 10.54 hours while in the remaining days is 2.47 hours.

Therefore we can conclude that the Poisson Process is appropriate to model rare events. However, certain weather phenomena may perturb its fitting. Seasonality affects the fitting and long raining periods will introduce huge differences between arrivals. Dependence between occurrences plays also an important role since a wildfire may generate favorable conditions for new wildfires in close locations.

4.3   Center Region

In this section, we are going to study if the Poisson Process is a suitable model for the Center Region. The Center Region has more occurrences than the South Region. It is important to understand if we increase the number of events, the Poisson Process will still be able to model the number of occurrences of wildfires due to interdependence of events.

Firstly, we plotted the graphs of the two first years to understand if seasonality was also presented in the Center Region. In Figures 4.23 and 4.24, we detected seasonality. So, to perform our analysis we intend to study a shorter period, namely, between May to September because we would like to compare with the South Region.
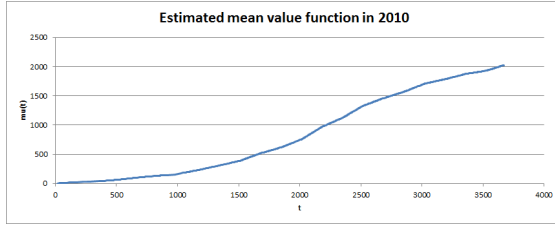
(a) The estimated mean value function $\mu(t)$
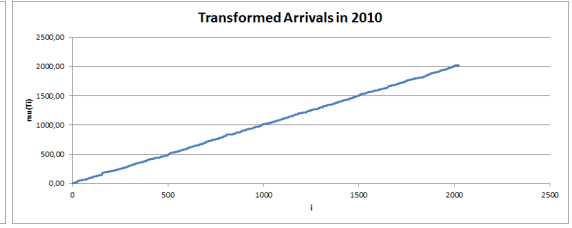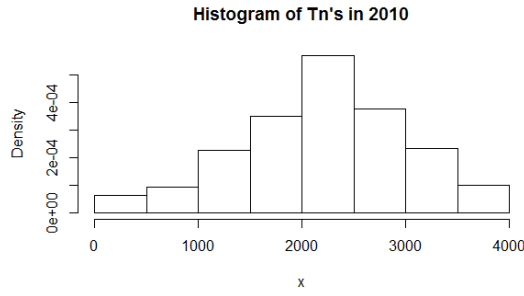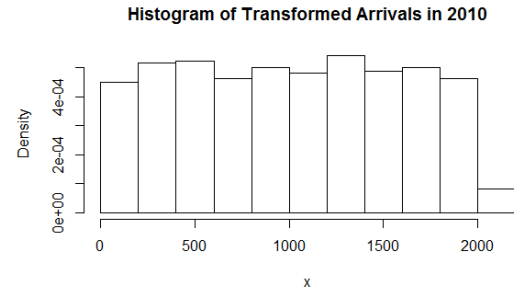
(b) The transformed arrivals $\mu(T_i)$

(c) Histogram of T_i

(d) Histogram of the $\mu(T_i)$
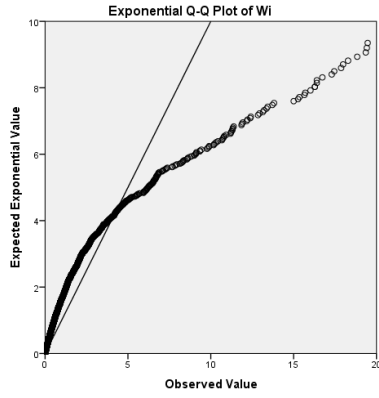
(e) QQ-plot of the inter-arrival times W_i
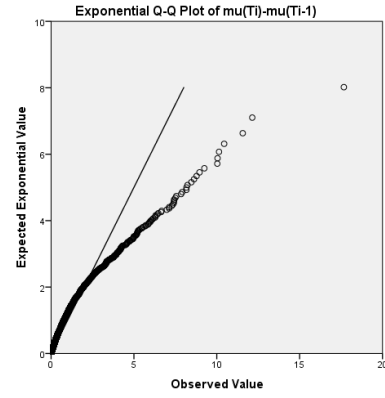
(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$
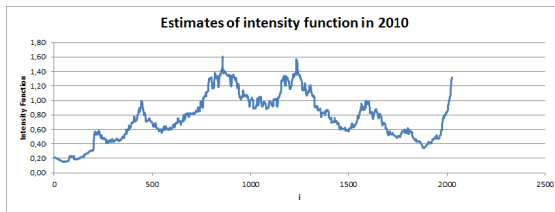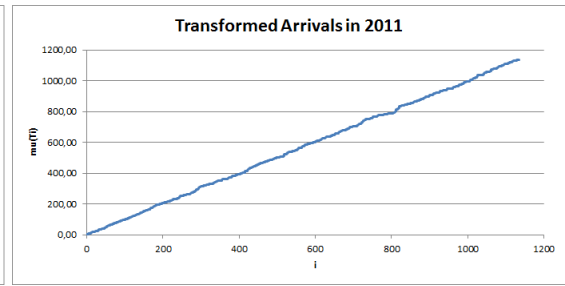

(g) Estimates of intensity function using W_i
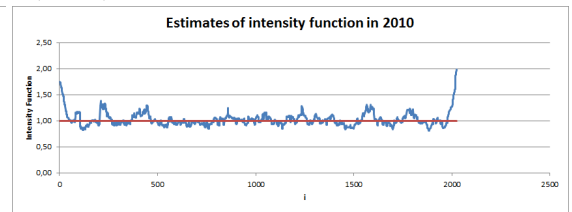
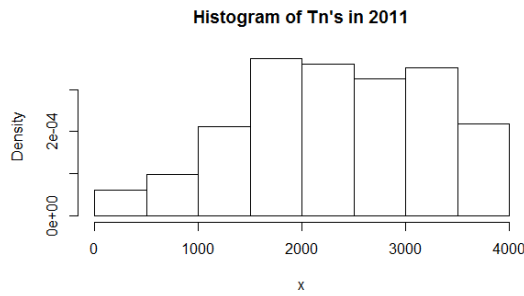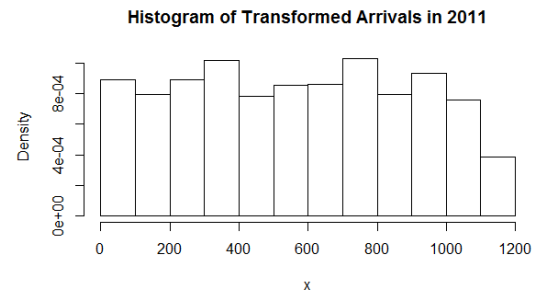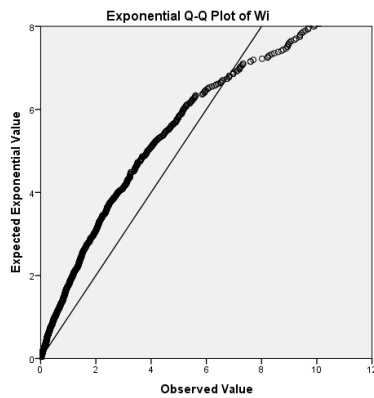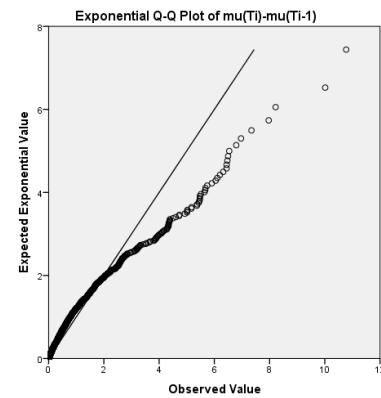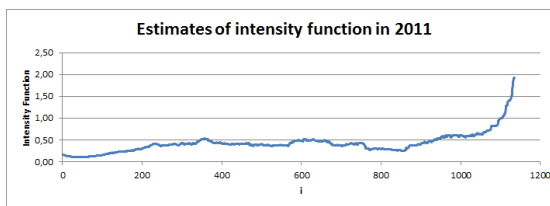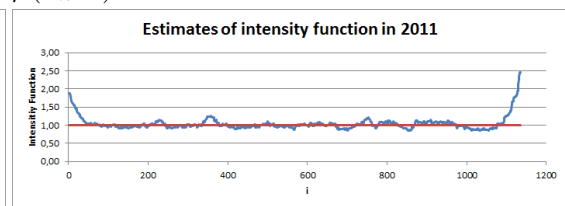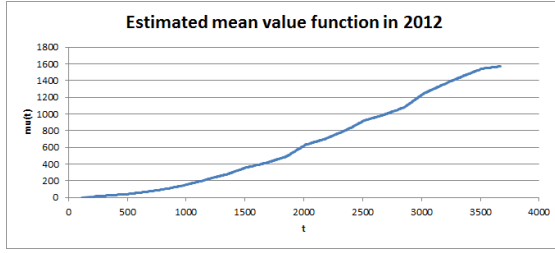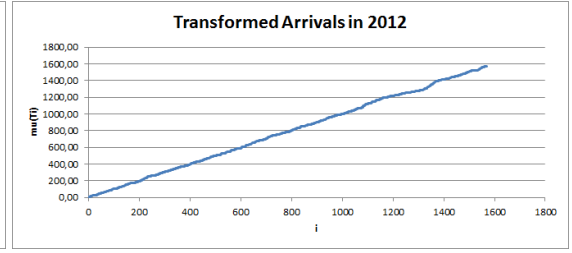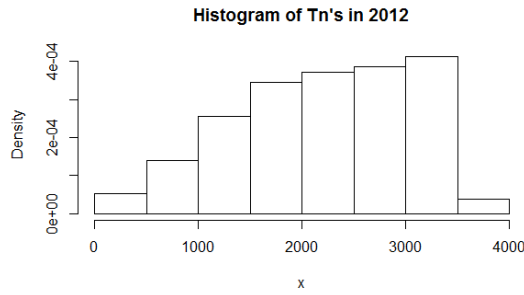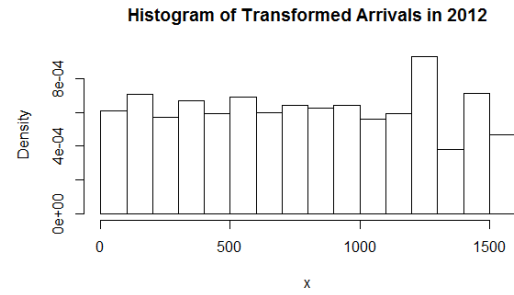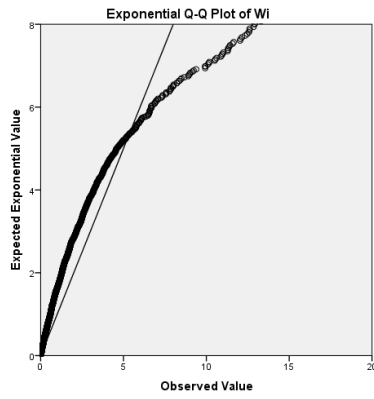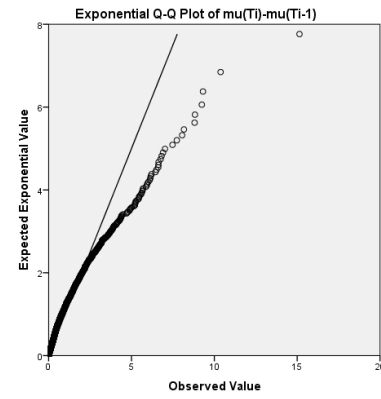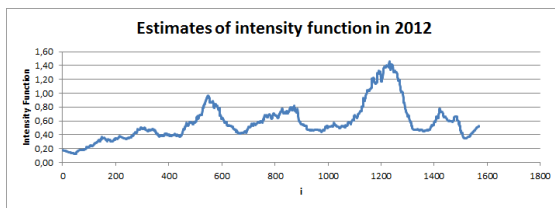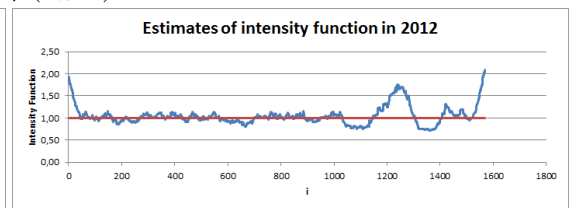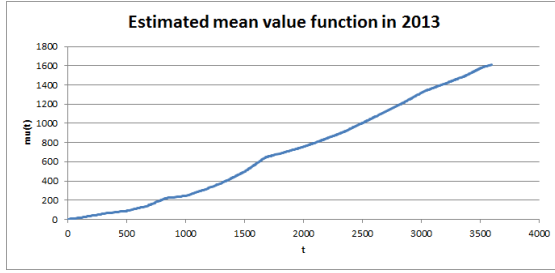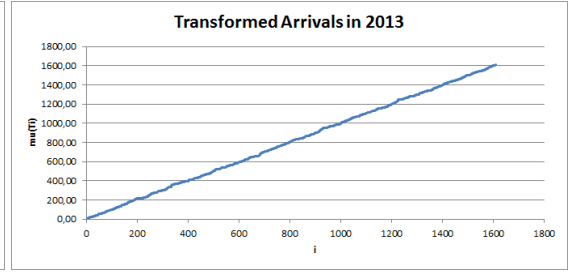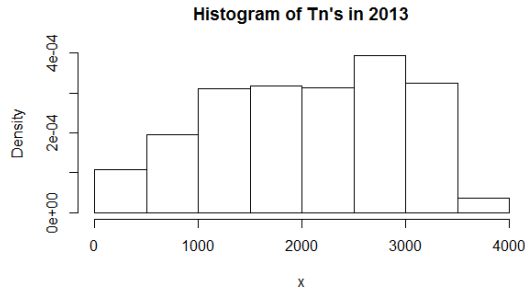(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$
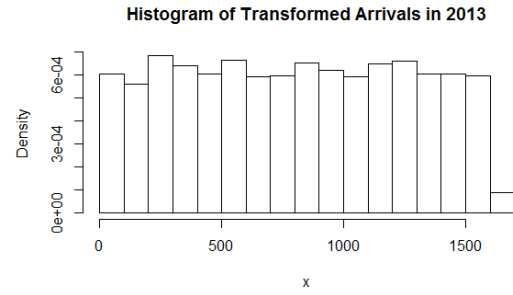
Figure 4.17: Graphs for 2010
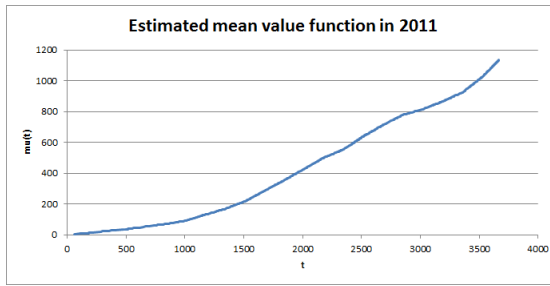
23

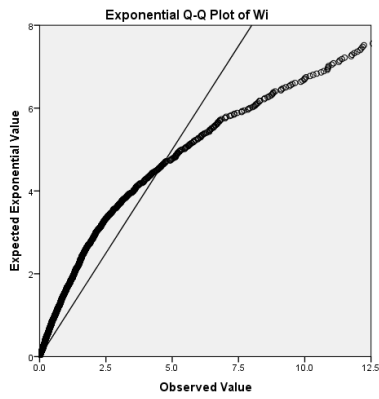(a) The estimated mean value function $\mu(t)$

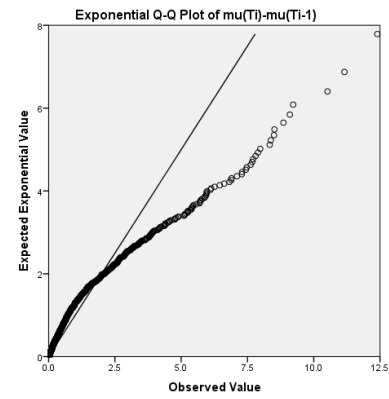(b) The transformed arrivals $\mu(T_i)$

(c) Histogram of T_i

(d) Histogram of the $\mu(T_i)$

(e) QQ-plot of the inter-arrival times W_i

(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$

(g) Estimates of intensity function using W_i

(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$

Figure 4.18: Graphs for 2011

(a) The estimated mean value function $\mu(t)$

(b) The transformed arrivals $\mu(T_i)$
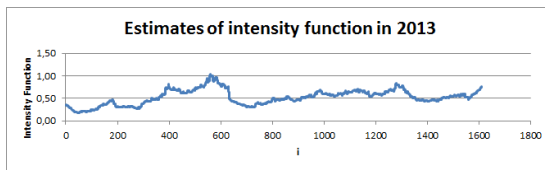

(c) Histogram of T_i
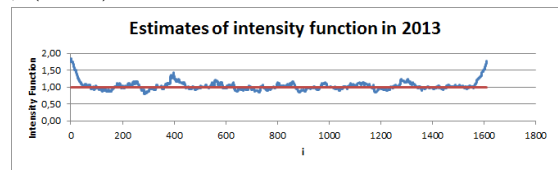

(d) Histogram of the $\mu(T_i)$

(e) QQ-plot of the inter-arrival times W_i

(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$

(g) Estimates of intensity function using W_i

(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$

Figure 4.19: Graphs for 2012

(a) The estimated mean value function $\mu(t)$

(b) The transformed arrivals $\mu(T_i)$

(c) Histogram of T_i

(d) Histogram of the $\mu(T_i)$

(e) QQ-plot of the inter-arrival times W_i

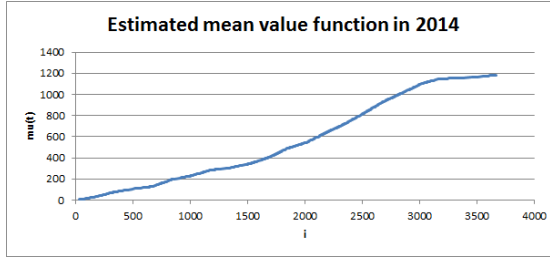(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$

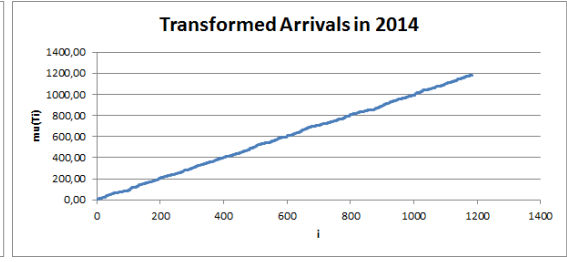(g) Estimates of intensity function using W_i

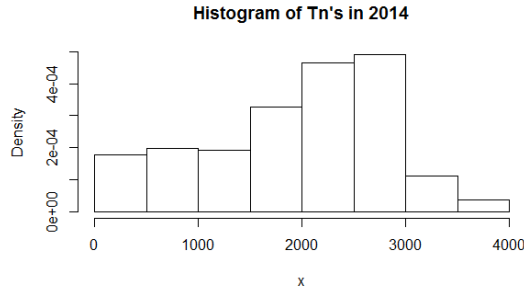(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$
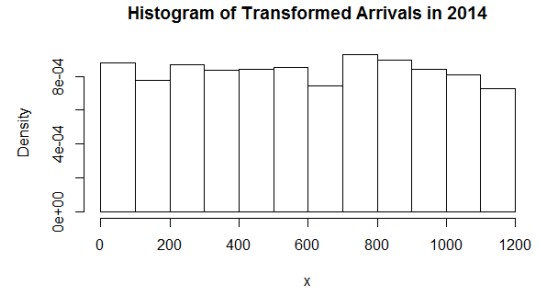
Figure 4.20: Graphs for 2013

(a) The estimated mean value function $\mu(t)$

(b) The transformed arrivals $\mu(T_i)$

(c) Histogram of T_i

(d) Histogram of the $\mu(T_i)$

(e) QQ-plot of the inter-arrival times W_i

(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$

(g) Estimates of intensity function using W_i

(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$

Figure 4.21: Graphs for 2014

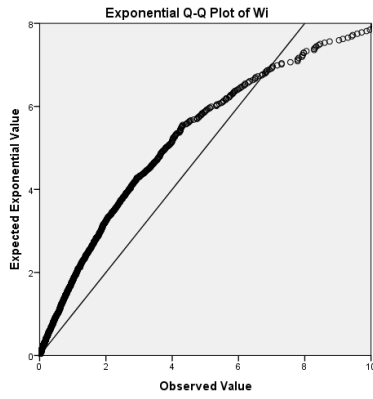(a) The estimated mean value function $\mu(t)$

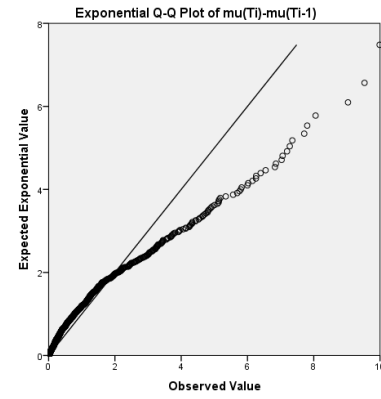(b) The transformed arrivals $\mu(T_i)$
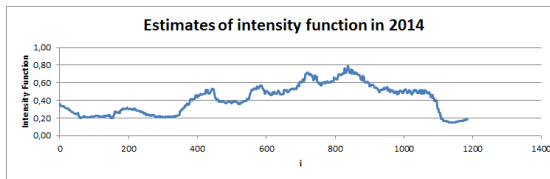

(c) Histogram of T_i
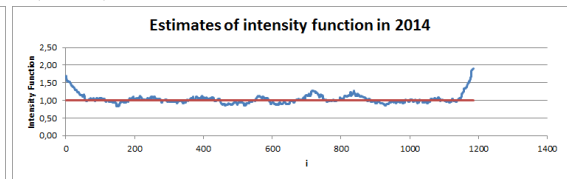

(d) Histogram of the $\mu(T_i)$

(e) QQ-plot of the inter-arrival times W_i

(f) QQ-plot of the inter-arrival times $\mu(T_n) - \mu(T_{n-1})$
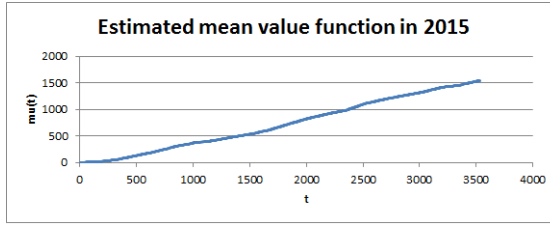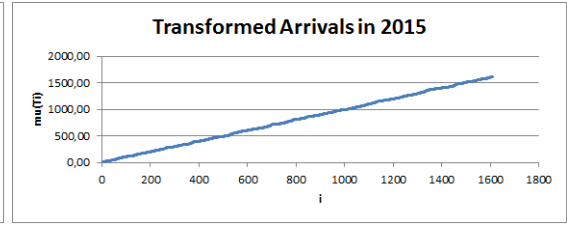

(g) Estimates of intensity function using W_i

(h) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$
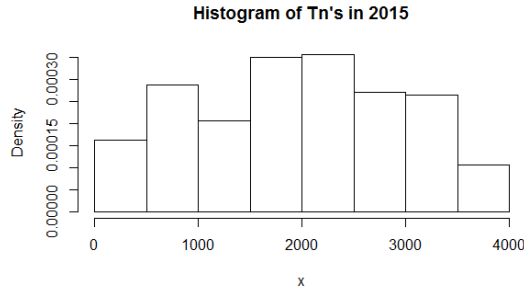
Figure 4.22: Graphs for 2015
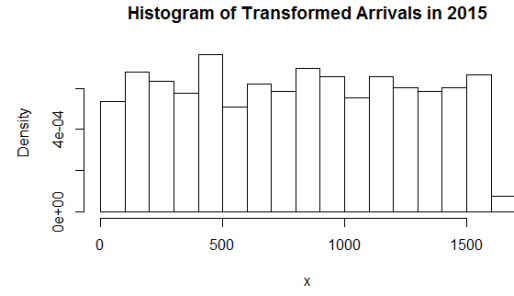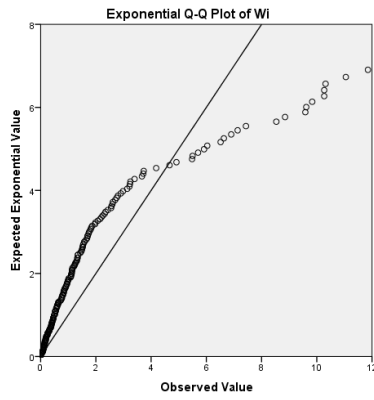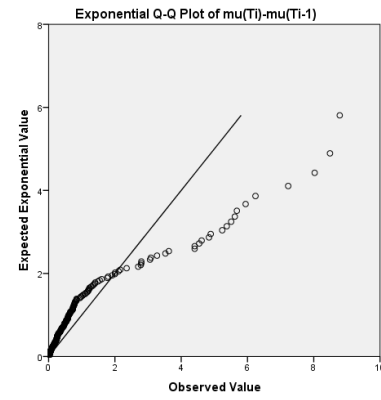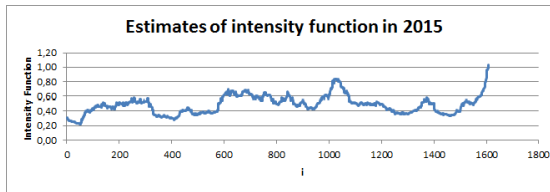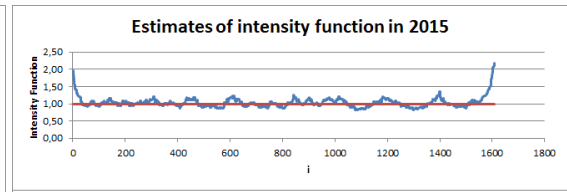
28
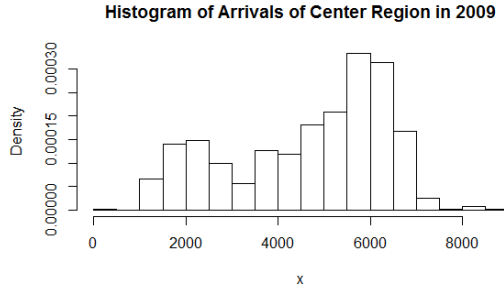
Figure 4.23: Histogram of arrivals of Center in 2009
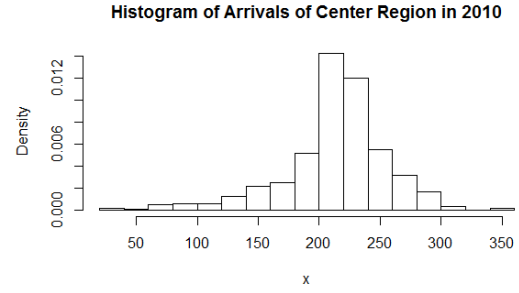
Figure 4.24: Histogram of arrivals of Center in 2010

We built a table with descriptive information about each week. Table 4.4 shows us that λ increases along the weeks. According to the literature review, a homogeneous Poisson Process has a constant λ, so we have gained statistical evidence the arrivals are modeled by an inhomogeneous Poisson Process.

An inhomogeneous Poisson Process will become a standard homogeneous Poisson process if we transform the arrivals T_i into $\mu(T_i)$. After that, we verified if the conditions to be a standard homogeneous Poisson Process are respected in the Center Region.

The first assumption is related to $\mu(t)$. As one can observe in Figure 4.25a, $\mu(t)$ is not linear (as it should be if the process was homogeneous). Therefore, we assumed that $\mu(t)$ was piecewise linear with constant slopes in each week and estimated this slopes with usual estimator for λ. Equipped with this approximation for μ, we transformed the data and plotted $\mu(T_i)$, which corresponds to Figure 4.25b. As expected by the theory, as long as, we have a good approximation of $\mu(t)$ the transformed data should look as those from a standard homogeneous Poisson process, which means that one should observe approximately a line of slope 1, as Figure 4.25b indeed shows.
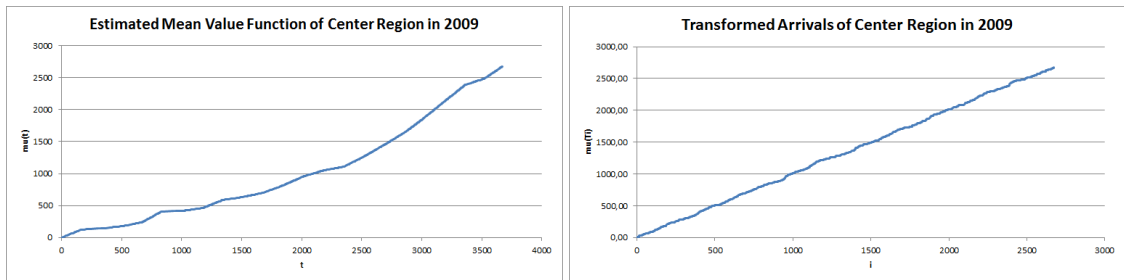
(a) The estimated mean value function $\mu(t)$ of Center Region in 2009

(b) The transformed arrivals $\mu(T_i)$ of Center Region in 2009

Figure 4.25: Comparison between estimated mean value function $\mu(t)$ and the transformed arrivals of Center Region in 2009

Weeks	Sample Size	Mean	$\hat{\lambda} = 1/mean$
1	124	1.336	0.748
2	20	8.192	0.122
3	33	5.284	0.189
4	63	2.657	0.376
5	168	0.973	1.028
6	8	21.146	0.047
7	45	3.802	0.263
8	127	1.320	0.758
9	45	3.703	0.270
10	66	2.568	0.389
11	118	1.432	0.699
12	137	1.208	0.828
13	95	1.772	0.564
14	58	2.920	0.342
15	160	1.053	0.950
16	179	0.929	1.077
17	191	0.882	1.134
18	238	0.710	1.409
19	257	0.657	1.523
20	251	0.669	1.496
21	107	1.573	0.636
22	182	0.778	1.285
Total	2672	1.373	0.728

Table 4.4: Sample size and Mean about Center Region between 1$^{\text{st}}$ May 2009 and 30$^{\text{th}}$ September 2009

The next condition is about arrivals T_i. According to the property of order statistics the arrivals of a standard homogeneous Poisson Process are approximately uniformly distributed. To check this we plotted the histogram of T_i against the histogram of $\mu(T_i)$. In the Figure 4.26, we see that the right graph resembles the histogram of uniform sample in contrast to the left one.

Next, we are going to check if the inter-arrival times are exponentially distributed. We built two QQ-plots, the left side has a QQ-plot using W_i and the right side has a QQ-plot taking into consideration $\mu(T_i) - \mu(T_{i-1})$. In the figure 4.27 we visualize that both graphs have a right tail which is heavier than the tail of the exponential distribution, however the right graph is more close to the straight line than left graph, then it is not unreasonable to assume that $\mu(T_i) - \mu(T_{i-1})$ are approximately exponentially distributed.

Lastly, if a homogeneous Poisson Process is standard, λ is equal to 1. To verify this statement we computed the estimates of the intensity function for W_i and $\mu(T_i) - \mu(T_{i-1})$ using

Histogram of Ti's of Center in 2009

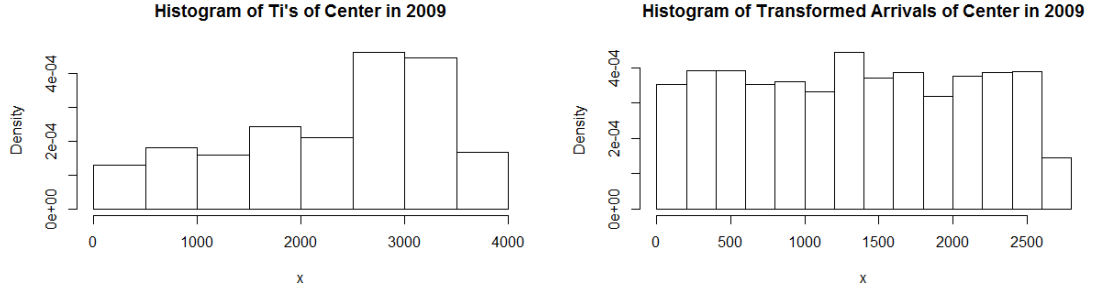**Histogram of Transformed Arrivals of Center in 2009**

(a) Histogram of T_i of Center Region in 2009

(b) Histogram of transformed arrivals $\mu(T_i)$ of Center Region in 2009

Figure 4.26: Comparison between histogram of T_i and histogram of $\mu(T_i)$ of Center Region in 2009

Exponential Q-Q Plot of Wi

Exponential Q-Q Plot of mu(Ti)-mu(Ti-1)
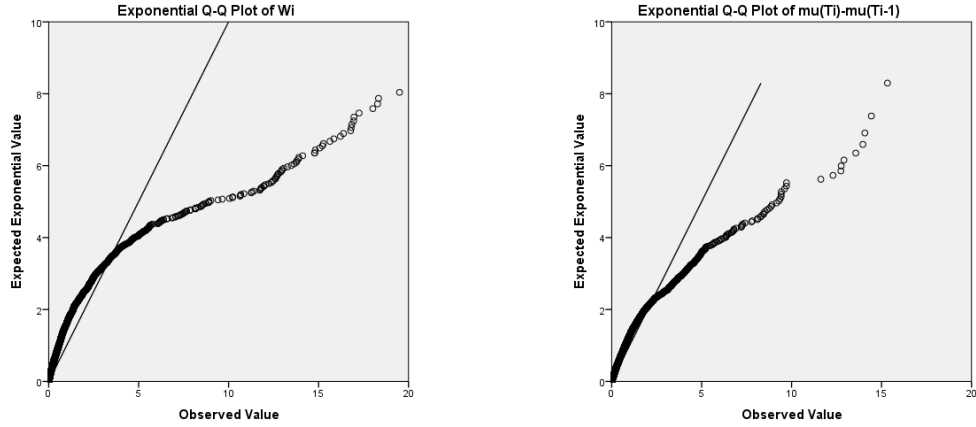
(a) QQ-plot using W_i of Center Region in 2009

(b) QQ-plot using $\mu(T_i) - \mu(T_{i-1})$ of Center Region in 2009

Figure 4.27: Comparison between QQ-plot using W_i and QQ-plot using $\mu(T_i) - \mu(T_{i-1})$ of Center Region in 2009

the moving average approach with $m = 75$ because the number of events is higher. Figure 4.28 shows that the intensity function estimates vary wildly on the left, while, on the right, the graph is close to 1.

(a) Estimates of intensity function using W_i
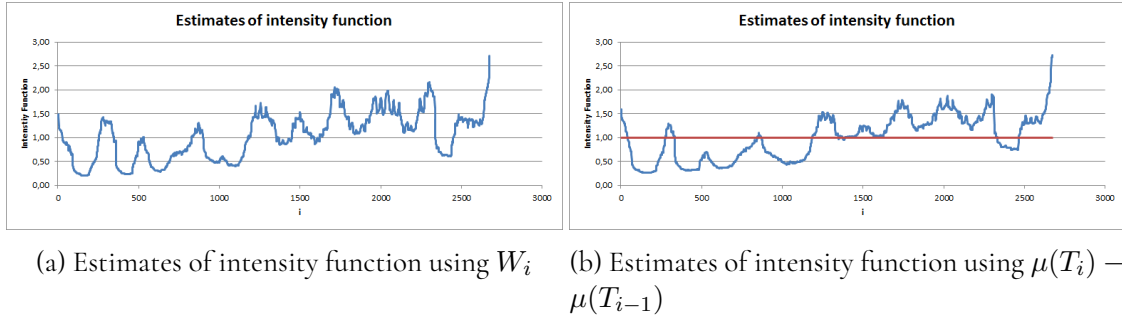
(b) Estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$

Figure 4.28: Comparison between estimates of intensity function using W_i and estimates of intensity function using $\mu(T_i) - \mu(T_{i-1})$ of Center Region in 2009

If we compare this period with the same period in the South Region we can conclude the Poisson Process is more appropriate for the South Region than for the Center Region. We observe that the sample size in the South Region counts 1811 wildfires and in the Center Region counts 2672 events from May to September. Comparing the left graph of Figures 4.14 with 4.26b, we see that the histogram of transformed arrivals in the South Region is better approximated by the uniform distribution than the histogram of transformed arrivals in the Center Region. This may be explained by the number of occurrences. If we increase the number of events, it will be more likely that the events are dependent of each other and the fitting is worse.

4.4   North Region

In the last section of this chapter we are going to compare the same periods analyzed earlier for the North Region. So, we are going to study the period between 1$^{\text{st}}$ May and 30$^{\text{th}}$ September of 2009 to compare with the South Region and the Center Region.

Firstly we compiled a table with the main information about the three regions. We can observe on table 4.5 that the number of events in North Region is seven times greater than in the South Region and five times greater than in the Center Region. We will see that the Poisson Process does not provide a good fitting for this region and given the weight of the contribution of the number of occurrences in this region for the global number, this may explain why there is not a good adjustment for the whole country. To check this we plotted all graphs for the North Region to compare with other regions. Note that, to estimate the intensity function we took $m = 1500$.

In Figure 4.29, we can observe that the North Region is not modeled by the Poisson Process. The transformed arrivals are not linear and one can see different slopes in the graph 4.29a. The histogram 4.29d does not resemble one from an uniform sample, the QQ-plot

| Regions	South			Center			North		
Weeks	i	Mean	1/Mean	i	Mean	1/Mean	i	Mean	1/Mean
1	50	3.249	0.308	124	1.336	0.748	284	0.587	1.703
2	19	8.942	0.112	20	8.192	0.122	107	1.514	0.661
3	51	3.311	0.302	33	5.284	0.189	58	2.981	0.335
4	52	3.279	0.305	63	2.657	0.376	141	1.207	0.828
5	67	2.484	0.403	168	0.973	1.028	381	0.439	2.277
6	38	4.464	0.224	8	21.146	0.047	13	12.835	0.078
7	71	2.337	0.428	45	3.802	0.263	98	1.725	0.580
8	103	1.598	0.626	127	1.320	0.758	273	0.615	1.627
9	55	3.137	0.319	45	3.703	0.270	79	2.127	0.470
10	95	1.761	0.568	66	2.568	0.389	169	0.997	1.003
11	83	2.040	0.490	118	1.432	0.699	293	0.574	1.741
12	88	1.904	0.525	137	1.208	0.828	418	0.386	2.592
13	102	1.655	0.604	95	1.772	0.564	194	0.898	1.114
14	82	2.012	0.497	58	2.920	0.342	202	0.835	1.198
15	106	1.603	0.624	160	1.053	0.950	636	0.264	3.784
16	96	1.761	0.568	179	0.929	1.077	829	0.203	4.934
17	121	1.370	0.730	191	0.882	1.134	956	0.176	5.690
18	127	1.333	0.750	238	0.710	1.409	1595	0.105	9.496
19	118	1.415	0.707	257	0.657	1.523	1585	0.106	9.439
20	76	2.233	0.448	251	0.669	1.496	1789	0.094	10.645
21	101	1.651	0.606	107	1.573	0.636	883	0.190	5.253
22	110	1.298	0.770	182	0.778	1.285	1435	0.100	9.980
Total	1811	2.026	0.494	2672	1.373	0.728	12418	0.296	3.382

Table 4.5: Main Statistics of all regions between 1st May and 30th September
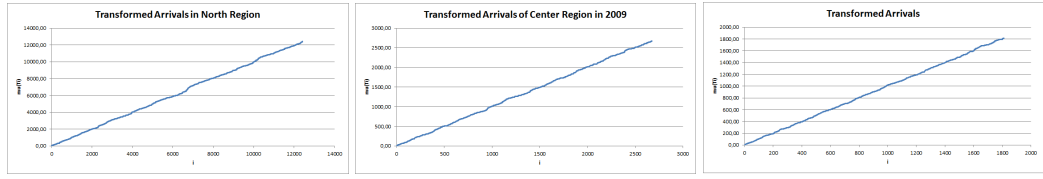
(a) Transformed arrivals in North Region

(b) Transformed arrivals in Center Region

(c) Transformed arrivals in South Region
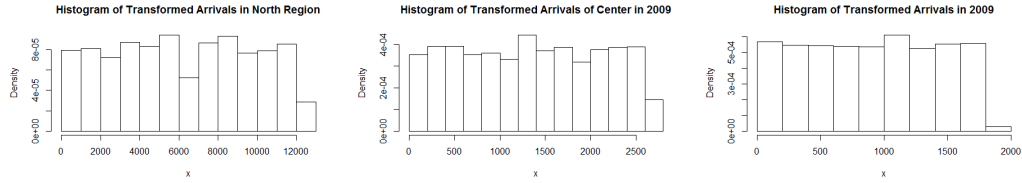
(d) Histogram of transformed arrivals in North Region

(e) Histogram of transformed arrivals in Center Region

(f) Histogram of transformed arrivals in South Region
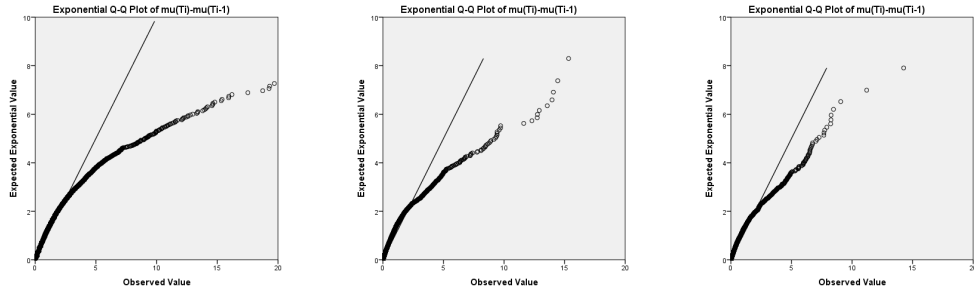
(g) QQ-plot of inter-arrival times $\mu(T_n) - mu(T_{n-1})$ in North Region

(h) QQ-plot of inter-arrival times $\mu(T_n) - mu(T_{n-1})$ in Center Region

(i) QQ-plot of inter-arrival times $\mu(T_n) - mu(T_{n-1})$ in South Region
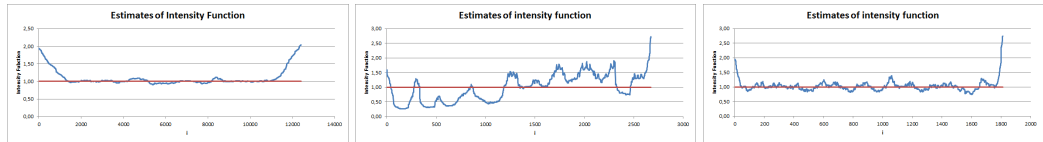
(j) Estimates of intensity function in North Region

(k) Estimates of intensity function in Center Region

(l) Estimates of intensity function in South Region

Figure 4.29: Comparison of Regions during the period between 1st May and 30th September of 2009

has a right tail higher than the tail of the exponential distribution and finally the estimates function is round to 1. We have gained statistical evidence that North Region is not modeled by Poisson Process.

If we analyze all regions we can see the model suits the region that has less wildfires which means if the number of events increase there is more probability the occurrences are dependent. If we look at the graphs from left to right we can see the effect of interdependence. The best plots are on right and the worst in the left.

Chapter 5

Conclusion

The Poisson Process is an appropriate approach to model the number of occurrences and their arrival times in rare events which have small probability of occurring.

In our case the model does not work for the whole period and all regions because of two factors: the seasonality which influences the arrivals because there are huge values of inter-arrival times which are caused by rains periods and the interdependence. As we mentioned above, for a Poisson Process, inter-arrival times must be independent and identically distributed and, in this case, there is a high likelihood that some of the wildfires occur because of other ones and, furthermore, if the climate conditions are equally extreme for all regions there is a huge correlation between occurrences. If we study this model for wildfires in Norway, Sweden or Denmark (see Mikosch (2009)) the model is more adequate.

However we think the model can be suitable if we divide the country in regions. So, we grouped districts into regions and we formed three regions. After that, we chose to study the model for all regions after excluding seasonality, by selecting a period from the 1st May until 30th September.

We started with the South Region because it has less events and we believe the interdependence was smaller. We concluded that in the South Region, the events were reasonably modeled by an inhomogeneous Poisson Process, which can be transformed into a standard homogeneous Poisson Process if we change the arrivals T_i into transformed arrivals $\mu(T_i)$. After applying this method, we conclude that the South Region is well modeled by an inhomogeneous Poisson Process in all years. We estimated λ for all years and we obtained $\hat{\lambda} = 0.62$, which means that in, every 0.62 hours, a new wildfire occurs in the South Region.

After that, we studied the Center Region for the year that produced better results, which was 2009. The Center Region showed more occurrences than the South Region. If we increase the number of events there is an higher risk that the occurrences are dependent. We applied the model for the same period that was studied in the South Region and the model does not suit as well as in the South Region. However, it still is reasonable in the Center Region.

Finally, in our analysis we considered the results for the North Region. If we see Figure 4.29 which is the comparison between all regions, the model does not seem to fit well, in the North Region, for the period between the 1st May and the 30th September, we believe this

is due to interdependence of increments and consequently of arrivals. The worse fitting in North Region may explain why there is not a good adjustment for the whole country.

Bibliography

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to time series and forecasting*, volume 2. Springer.

Dobrow, R. P. (2016). *Introduction to stochastic processes with R*. John Wiley & Sons.

Gallager, R. G. (2012). *Discrete stochastic processes*, volume 321. Springer Science & Business Media.

Marais, C. A. et al. (2010). *Markov processes in disease modelling: estimation and implementation*. PhD thesis, University of Pretoria.

Mikosch, T. (2009). *Non-life insurance mathematics: an introduction with the Poisson process*. Springer Science & Business Media.

Mingola, P. (2013). A study of poisson and related processes with applications.

Møller, J. and DÍAZ-AVALOS, C. (2010). Structured spatio-temporal shot-noise cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, 37(1):2–25.

Pinsky, M. and Karlin, S. (2010). *An introduction to stochastic modeling*. Academic press.

Pkj, P. (2007). Lecture 3: Continuous times markov chains. poisson process. birth and death process.

Tse, K.-K. (2014). Some applications of the poisson process. *Applied Mathematics*, 5(19):3011.