Analysis of data services performance in UMTS networks using data analytics

Carlos Jorge Almeida Martins

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Examination Committee

Chairperson: Prof. José Eduardo Charters Ribeiro da Cunha Sanguino

Supervisor: Prof. Luís Manuel de Jesus Sousa Correia

Member of Committee: Prof. Paulo Luís Serras Lobato Correia

Eng. Paulino Correa

November 2015

To the ones I love

Acknowledgements

My first acknowledgement goes to Prof. Luís M. Correia, for giving me the opportunity to write this thesis under his guidance, support and professionalism. I am very thankful for the weekly meetings and for the good advices he gave me, not only academically but also for my future professional life.

Special acknowledgements are also given to Eng. Paulino Correa and Eng. Ricardo Correia, who followed my work from Vodafone's point of view and helped me with their constant assistance, guidance, and technical support, which were essential during this journey.

To all GROW's members, a special thanks for all the support, especially to my master thesis' colleagues and friends Ana Cláudia Castro, Andrea Marotta, João Pires, Miguel Sá and Ricardo Gameiro for their friendship, advices and mutual sharing of information.

To my family, a big thanks for their support, affection and for making me what I am today, especially to my Mother, Maria de Lurdes Martins, to my Father, Carlos Martins, and to my sister Carla Martins, for always being terrific role models and for their endless love and encouragement.

Last, but not least, to the love of my life, Francisca, for all the support, encouragement and unconditional love, and for giving me the strength and motivation, without which, the completion of this work would have been a more difficult task.

Abstract

The main objective of this thesis was to develop a model that enables network optimisation, based on data analysis from UMTS networks, since these carry most of the data services. An analysis of the data traffic and the number of users per operating system and type of device registered during the last 3 years on 4 specific RNCs located in the south of Portugal is done, in order to have a perception of the growing trend of those data along those years and a better understanding of the human behaviour and needs in terms of the usage of mobile communications. For that, one resorts to different statistical distributions that fit to the average of the different data related to the 3 years for each RNC, through MATLAB *software*, to better represent the progress of the analysed parameters, thus contributing to predict their behaviour throughout time. For the different scenarios, it is concluded that is possible to have models with a proximity above 95% in relation to the average of the original data, which can then be implemented on the prediction models of the following 2 years. In order to be aware of a possible network saturation, it is predicted that the users' traffic in 2016 will be 1.74 times larger than the one obtained in 2014, and that the data traffic in 2016 will be 1.57 times greater than the one obtained in 2014 too. Smartphones and Androids' usage is also expected to continue to grow apace.

Keywords

UMTS, Data Services, Data Analytics, Fitting Models, Prediction.

Resumo

O principal objetivo desta dissertação é desenvolver um modelo que permita a otimização de rede, com base na análise de dados provenientes das redes UMTS, uma vez que estas transportam a maior parte dos serviços de dados. É feita uma análise do tráfego de dados e do número de utilizadores por sistema operativo e por tipo de dispositivo registados durante os últimos 3 anos em 4 RNCs específicos localizados a sul de Portugal, de modo a ter uma perceção da tendência de crescimento desses dados ao longo desses anos e a compreender melhor o comportamento humano e as suas necessidades no que toca às comunicações móveis. Para isso, recorre-se a diferentes distribuições estatísticas que melhor se ajustam à média dos diferentes dados relativos aos 3 anos de cada RNC, através do *software* MATLAB, para melhor representar o andamento dos parâmetros em análise, contribuindo desta forma para prever os seus comportamentos ao longo do tempo. Para os diferentes cenários, conclui-se que é possível obter modelos com um grau de aproximação superior a 95% em relação à média dos dados originais, que depois são então implementados nos modelos de previsão dos 2 anos seguintes. De forma a prevenir uma possível saturação da rede, prevê-se que o tráfego de utilizadores e o tráfego de dados em 2016 sejam 1.74 e 1.57 vezes maiores que os mesmos obtidos em 2014, respetivamente. Também é previsto que o uso de Smartphones, Androids e outros dispositivos e plataformas continue a crescer rapidamente.

Palavras-chave

UMTS, Serviço de Dados, Análise de Dados, Modelos de Aproximação, Previsão.

Table of Contents

List of Figures

List of Tables

List of Acronyms

1G	First generation mobile communications system
2G	Second generation mobile communications system
3G	Third generation mobile communications system
3GPP	3rd Generation Partner Project
4G	Fourth Generation mobile communications system
AN	Access Network
AMC	Adaptive Modulation and Coding
ATM	Android Traffic Model
BS	Base Station
BSS	Business Support System
CAGR	Compound Annual Growth Rate
CD	Coefficient of Determination
CN	Core Network
CoMP	Co-ordinated Multi-Point
CRM	Customer Relationship Management
CS	Circuit Switch
CSP	Communication Service Provider
DA	Data Analytics
DC-HSDPA	Dual Cell High-Speed Downlink Packet Access
DL	Downlink
DRNC	RNC Data Traffic Model
DS-CDMA	Direct-Sequence Code Division Multiple Access
FDD	Frequency Division Duplex
FTM	Facebook Traffic Model
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GMSC	Gateway Mobile Location Register
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HDMI	High-Definition Multimedia Interface
HLR	Home Location Register
HSDPA	High-Speed Downlink Packet Access
HSPA	High-Speed Packet Access
HSPA+	HSPA evolution
HSUPA	High-Speed Uplink Packet Access
HTML5	Hypertext Markup Language 5
IMT-Advanced	International Mobile Telecommunications-Advanced
iOS	iPhone OS
iOSTM	iOS Traffic Model
ITU	International Telecommunication Union

LTE	Long-Term Evolution
M2M	Machine-to-Machine
MBB	Mobile Broadband
ME	Mobile Equipment
MIMO	Multiple Input Multiple Output
MLTM	Multi Linear Traffic Model
MSC	Mobile Switching Centre
MT	Mobile Terminal
MTM	MBB Traffic Model
OD	Organisational Development
OS	Operating System
OSN	Online Social Network
OSS	Operations Support System
OTM	Others Traffic Model
OTT	Over-The-Top
PC	Personal Computer
PLMN	Public Land Mobile Network
PPS	Packet Per Second
PS	Packet Switch
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quaternary Phase Shift Keying
RMSE	Root Mean Squared Error
RNC	Radio Network Controller
RRM	Radio Resource Management
RTM	Room Traffic Model
RTU	RNC Total Users Traffic Model
RTT	Round Trip Time
SD	Secure Digital
SGSN	Serving GPRS Support Node
SMS	Short Message Service
STM	Smartphone Traffic Model
SWIFT	Society of Worldwide Interbank Financial Telecommunication
TDTM	Total Data Traffic Model
TTI	Transmission Time Interval
TTM	Tablet Traffic Model
TwTM	Twitter Traffic Model
UE	User Equipment
UHD	Ultra-High Definition
UL	Uplink
UMTS	Universal Mobile Telecommunications System
URNC	RNC User Traffic Model
USB	Universal Serial Bus
USD	United States Dollar
USIM	UMTS Subscriber Identity Module
UTM	Unclassified Traffic Model
UTRA	Universal Terrestrial Radio Access

UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visitor Location Register
VoIP	Voice over IP
WATM	WhatsApp Traffic Model
WCDMA	Wideband Code Division Multiple Access
WiTM	Windows Traffic Model
WLAN	Wireless Local Area Network
WTM	Water Traffic Model
WWW	World Wide Web

List of Symbols

$\sqrt{\overline{\varepsilon^2}}$	Root Mean Squared Error
μ	Average obtained from the samples
μ_n	*n*th sample
σ	Standard deviation obtained from the samples
σ_{avg}	Average of the standard deviations
σ_n	*n*th sample
a_{atm}	Power initial value
a_{drnc1}	Exponential initial value
a_{drnc2}	Exponential initial value
a_{drnc3}	First exponential initial value
a_{drnc4}	Power initial value
a_{iostm}	Power initial value
a_{mtm1}	First exponential initial value
a_{mtm2}	Second exponential initial value
a_{otm}	Exponential initial value
a_{rtu}	Power initial value
a_{stm}	Power initial value
a_{tdtm}	Exponential initial value
a_{ttm}	Exponential initial value
a_{urnc1}	Gaussian amplitude
a_{urnc2}	Exponential initial value
a_{urnc3}	Exponential initial value
a_{urnc4}	First exponential initial value
a_{utm}	Power initial value
a_{witm}	Power initial value
a_{wtm}	Power initial value
b_{atm}	Power decay factor
b_{drnc1}	Exponential decay factor
b_{drnc2}	Exponential decay factor
b_{drnc3}	First exponential decay factor
b_{drnc4}	Power decay factor
b_{iostm}	Power decay factor
b_{mtm1}	First exponential decay factor

b_{mtm2}	Second exponential decay factor
b_{otm}	Exponential decay factor
b_{rtu}	Power decay factor
b_{stm}	Power decay factor
b_{tdtm}	Exponential decay factor
b_{ttm}	Exponential decay factor
b_{urnc1}	Month shifted peak value
b_{urnc2}	Exponential decay factor
b_{urnc3}	Exponential decay factor
b_{urnc4}	First exponential decay factor
b_{utm}	Power decay factor
b_{witm}	Power decay factor
b_{wtm}	Power decay factor
c_{atm}	Linear function offset
c_{drnc1}	Exponential offset
c_{drnc2}	Exponential offset
c_{drnc3}	First exponential offset
c_{drnc4}	Linear function offset
c_{ftm}	Linear function constant value
c_{iostm}	Linear function offset
c_{mltm1}	First linear function constant value
c_{mltm2}	Second linear function constant value
c_{mltm3}	Third linear function constant value
c_{mtm1}	First exponential offset
c_{mtm2}	Second exponential offset
c_{otm}	Exponential offset
c_{rtm1}	First linear function offset
c_{rtm2}	Second linear function offset
c_{rtu}	Linear function offset
c_{stm}	Linear function offset
c_{tdtm}	Exponential offset
c_{ttm}	Exponential offset
c_{twtm}	Linear function constant value
c_{urnc1}	Gaussian deviation
c_{urnc2}	Exponential offset
c_{urnc3}	Exponential offset
c_{urnc4}	First exponential offset
c_{utm}	Linear function offset
c_{watm}	Linear function constant value

c_{witm}	Linear function offset
c_{wtm1}	First linear function offset
c_{wtm2}	Second linear function offset
d_{atm}	Polynomial offset
d_{drnc1}	Polynomial offset
d_{drnc2}	Polynomial offset
d_{drnc3}	Second exponential initial value
d_{drnc4}	Exponential initial value
d_{iostm}	Polynomial offset
d_{otm}	Polynomial offset
d_{rtu}	Polynomial offset
d_{stm}	Polynomial offset
d_{tdtm}	Polynomial offset
d_{ttm}	Polynomial offset
d_{urnc1}	Gaussian offset
d_{urnc2}	Polynomial offset
d_{urnc3}	Polynomial offset
d_{urnc4}	Second exponential initial value
d_{utm}	Polynomial offset
d_{witm}	Polynomial offset
d_{wtm}	Exponential initial value
f_{atm}	Android Traffic Model function
f_{drnc1}	RNC1 Data Traffic Model function
f_{drnc2}	RNC2 Data Traffic Model function
f_{drnc3}	RNC3 Data Traffic Model function
f_{drnc4}	RNC4 Data Traffic Model function
f_{ftm}	Facebook Traffic Model function
f_{iostm}	iOS Traffic Model function
f_{mltm}	Multi Linear Traffic Model function
f_{mtm}	MBB Traffic Model function
f_{otm}	Others Traffic Model function
f_{rtm}	Room Traffic Model function
f_{rtu}	RNC Total Users Traffic Model function
f_{stm}	Smartphone Traffic Model function
f_{tdtm}	Total Data Traffic Model function
f_{ttm}	Tablet Traffic Model function
f_{twtm}	Twitter Traffic Model function
f_{urnc1}	RNC1 User Traffic Model function
f_{urnc2}	RNC2 User Traffic Model function

f_{urnc3}	RNC3 User Traffic Model function
f_{urnc4}	RNC4 User Traffic Model function
f_{utm}	Unclassified Traffic Model function
f_{watm}	WhatsApp Traffic Model function
f_{witm}	Windows Traffic Model function
f_{wtm}	Water Traffic Model function
g_{drnc3}	Second exponential decay factor
g_{drnc4}	Exponential decay factor
g_{urnc1}	Polynomial offset
g_{urnc4}	Second exponential decay factor
g_{wtm}	Exponential decay factor
h_{drnc3}	Second exponential offset
h_{drnc4}	Exponential offset
h_{urnc4}	Second exponential offset
m_{ftm}	Linear function slope
m_{mltm1}	First linear function slope
m_{mltm2}	Second linear function slope
m_{mltm3}	Third linear function slope
m_{rtm1}	First linear function slope
m_{rtm2}	Second linear function slope
m_{twtm}	Linear function slope
m_{watm}	Linear function slope
N	Number of samples
p_{atm}	Polynomial scale factor
p_{drnc1}	Polynomial scale factor
p_{drnc2}	Polynomial scale factor
p_{iostm}	Polynomial scale factor
p_{otm}	Polynomial scale factor
p_{rtu}	Polynomial scale factor
p_{stm}	Polynomial scale factor
p_{tdtm}	Polynomial scale factor
p_{ttm}	Polynomial scale factor
p_{urnc1}	Polynomial scale factor
p_{urnc2}	Polynomial scale factor
p_{urnc3}	Polynomial scale factor
p_{utm}	Polynomial scale factor
p_{witm}	Polynomial scale factor
q_{atm}	Polynomial shift
q_{drnc1}	Polynomial shift

q_{drnc2}	Polynomial shift
q_{iostm}	Polynomial shift
q_{otm}	Polynomial shift
q_{rtu}	Polynomial shift
q_{stm}	Polynomial shift
q_{tdtm}	Polynomial shift
q_{ttm}	Polynomial shift
q_{urnc1}	Polynomial shift
q_{urnc2}	Polynomial shift
q_{urnc3}	Polynomial shift
q_{utm}	Polynomial shift
q_{witm}	Polynomial shift
t_{atm}	Breakpoint shifted month
t_{drnc1}	Breakpoint shifted month
t_{drnc2}	Breakpoint shifted month
t_{drnc3}	Breakpoint shifted month
t_{drnc4}	Breakpoint shifted month
t_{iostm}	Breakpoint shifted month
t_{ftm}	Duration time (months)
t_{mltm1}	First breakpoint shifted month
t_{mltm2}	Second breakpoint shifted month
t_{mtm}	Breakpoint shifted month
t_{otm}	Breakpoint shifted month
t_{rtm}	Breakpoint shifted month
t_{rtu}	Breakpoint shifted month
t_{stm}	Breakpoint shifted month
t_{tdtm}	Breakpoint shifted month
t_{ttm}	Breakpoint shifted month
t_{twtm}	Duration time (months)
t_{urnc1}	Breakpoint shifted month
t_{urnc2}	Breakpoint shifted month
t_{urnc3}	Breakpoint shifted month
t_{urnc4}	Breakpoint shifted month
t_{utm}	Breakpoint shifted month
t_{watm}	Duration time (months)
t_{witm}	Breakpoint shifted month
t_{wtm}	Breakpoint shifted month
R^2	Coefficient of determination
x_i	Mean value or the standard deviation for the ith month

$(x_i)_N$	Mean value or the standard deviation for the *i*th month already normalised
y	Expression of each branch
$\bar{y}$	Average of the observations of the variable
$y_{1st\ Aug}$	Value that the parameter takes on the 1st August available (where the first big peak occurs)
y_i	*i*th observation
$\hat{y_i}$	Estimated or predicted value of y_i
Y_i	Observed (true) values
$\hat{Y_i}$	Model (fitted) values
y_j	Value that the parameter takes in the *j*th month
$(y_j)_N$	Value of the parameter in the *j*th month already normalised
$(y)_N$	New normalised expression of the branch
z_{max}	Maximum value of the approximation
z_{min}	Minimum value of the approximation

List of Software

Adobe Photoshop	Raster graphics editor
MATLAB	Numerical computing environment
Microsoft Excel	Spreadsheet application
Microsoft Visio	Diagramming and vector graphics application
Microsoft Word	Word processor
Wolfram Mathematica 9	Computational software

Chapter 1

Introduction

The present chapter introduces the theme of this dissertation, in particular, over a contextual and motivational perspective, simultaneously providing an overview of the assumptions established for the work development. Furthermore, it establishes the scope of the work performed together with its main contributions, followed by the detailed presentation of the work's structure.

1.1 Overview

Mobile communications are a technology that achieves the communication any time and place between terminal objects. Mobile devices and connections are not only getting smarter in their computing capabilities, but are also evolving from a lower-generation network connectivity to a higher-generation one. When device capabilities are combined with faster, higher bandwidth and more intelligent networks, it leads to a wide adoption of advanced multimedia applications that contribute to increase mobile and Wi-Fi traffic.

Second generation (2G) mobile communications systems, which included Global System for Mobile Communications (GSM), are using digital modulations techniques, very much different from the first generation (1G) ones, in order to support low-speed data services. GSM was originally designed to carry voice traffic, but later on, data capability was added. Data usage has increased, but the traffic volume in GSM is clearly dominated by voice.

Currently, the number of mobile phones is exceeding the number of landline ones, and the mobile phone penetration is higher in several markets. As it is known, the data-handling capabilities of 2G systems are limited. So, in order to have higher bit-rate services that enable high-quality images and video to be transmitted, and to provide access to Internet with higher data rates, third generation (3G) mobile communications systems were developed.

A 3G standard is Universal Mobile Telecommunications System (UMTS) which boosts considerably data usage. UMTS data growth is driven by high-speed radio capability, flat-rate pricing schemes, and simple device installation, and its introduction has marked the transition of mobile communications from voice-dominated networks to packet-data ones.

4G, the fourth generation mobile communication system, corresponding to Long-Term Evolution (LTE), allows to transmit high quality video images similar to high-definition television. If 3G can provide a high-speed transmission in wireless communications environment, then 4G communication is an ultra-high-speed mobile network, being faced as a super Internet highway without cables. This new generation brings fast data access, unified messaging and broadband multimedia, in the form of new interactive services [Pent11].

Figure 1.1 shows the forecast of the global mobile devices and connections over the coming years. Globally, the relative share of 3G-capable devices and connections will surpass 2G-capable devices and connections by 2016 (48% and 44% relative share). By 2018, 15% of all global devices and connections will be 4G capable.

In order to fulfil consumer demand needs, the most significant change is the transition from a clear dominance of voice traffic to a data one, as it can be seen in Figure 1.2.

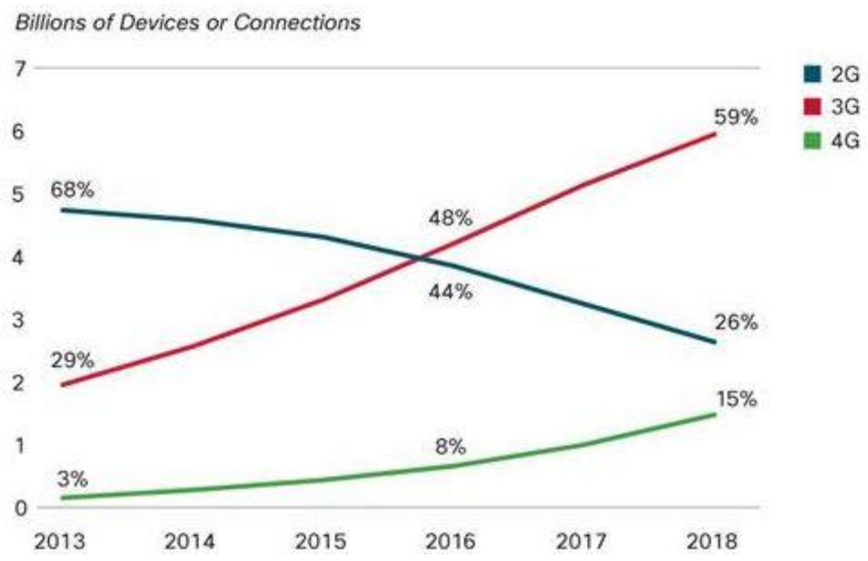

Figure 1.1. Global Mobile Devices and Connections (extracted from [Cisc13]).

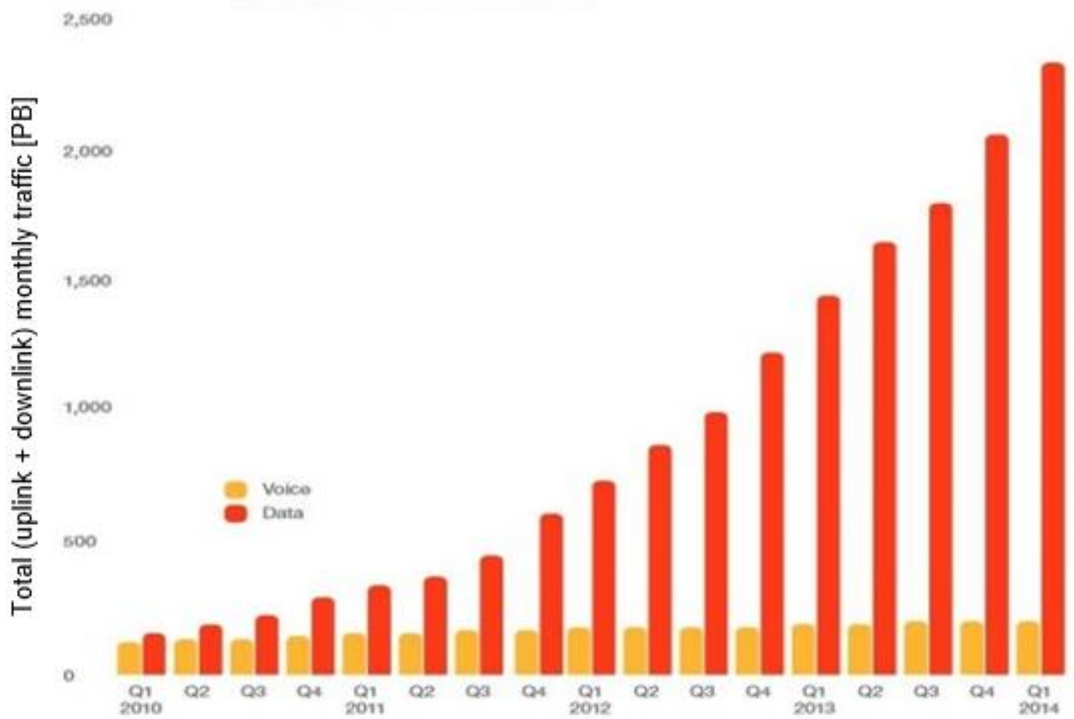

Figure 1.2. Global total traffic in mobile networks (adapted from [Eric14]).

In 1999, the 3rd Generation Partner Project (3GPP) launched UMTS, also called Release 99, which uses Wideband Code Division Multiple Access (WCDMA) in its air interface, featuring data rates up to 384 kbps for the Downlink (DL) and Uplink (UL), and a theoretical maximum for DL of 2 Mbps [Moli11].

The major focus for all 3GPP Releases is to make the system backwards and forwards compatible wherever possible, in order to ensure that the operation of user equipment is uninterrupted. All of these advances provide a high degree of continuity in the evolving systems, allowing the existing equipment to be prepared for future features and functionalities, delivering higher data rates, quality of service and cost efficiencies. Figure 1.3 shows the evolution of UMTS (3GPP releases) over the years.

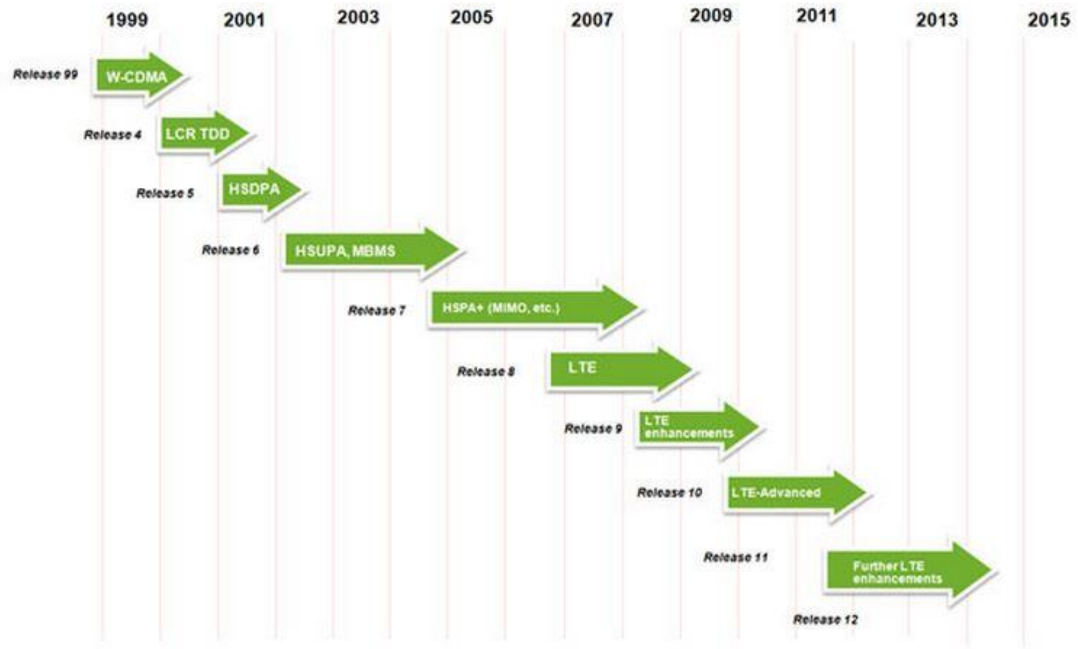

Figure 1.3. Evolution of 3GPP releases (adapted from [3GPP14]).

An evolution of the first release was the High-Speed Downlink Packet Access (HSDPA) in Release 5, which included major improvements in DL data rates and capacity. Similar technical solutions were applied to UL as part of Release 6 with High-Speed Uplink Packet Access (HSUPA). Release 7 brought a number of further substantial enhancements to end-user performance, network capacity and network architecture.

Release 7 solutions also brought High-Speed Packet Access (HSPA) capabilities closer to LTE targets. LTE was specified as part of Release 8 and a further push to higher radio capabilities. Releases 7 and 8 solutions for HSPA evolution were worked in parallel with LTE development, and some aspects of LTE were also expected to reflect on HSPA evolution. The HSPA evolution is also known as HSPA+ and was specified in Release 7 with the introduction of the Multiple Input Multiple Output (MIMO) feature, which allows users to achieve data rates up to 42 Mbps [HoTo07].

Release 9 provides further enhancements to HSPA+ and to the initial Release 8. Release 9 was the support of simultaneous MIMO and Dual Cell High-Speed Downlink Packet Access (DC-HSDPA) operation, as well as enhancements to the transmit diversity modes to improve performance with non-MIMO capable devices. With the completion of Release 9, focus in 3GPP has turned to Releases 10 and 11.

Release 10 adds further functionalities and performance enhances to both HSPA and LTE. For HSPA, additional multi-carrier and MIMO options were introduced in Release 10. For LTE, Release 10 marks the introduction of LTE-Advanced, which has been certified as complying with all of the International Telecommunication Union (ITU) International Mobile Telecommunications-Advanced (IMT-Advanced) requirements.

Release 11 focuses on continued HSPA+ and LTE/LTE-Advanced enhancements, including Co-ordinated Multi-Point (CoMP), Carrier Aggregation and ICIC enhancements. The proposed enhancements to HSPA+ in Release 11 include 8-carrier HSDPA, UL dual antenna beamforming and MIMO, and DL multi-point transmission.

1.2 Motivation and Contents

Currently, UMTS is still the system that carries most of the data services, since GSM is preferential for voice and LTE has not yet achieved a general penetration. By analysing the performance of data services in UMTS, one can gain a view that can be used in LTE, enabling network optimisation when penetration will reach higher levels. By taking real data from networks, on user sessions, data volume, profile, location, among other parameters, correlating with network metrics, and establishing models, one can improve the efficiency and the effectiveness of network planning and development, as well as user experience by proper network parameters tuning. The ultimate goal of this approach is network optimisation, via the usage of developed models, and then to extrapolate to LTE. This thesis addresses this problem, aiming at establishing the models assessed by data analytics, and guidelines for implementation.

This thesis was done in collaboration with Vodafone Portugal, which belongs to a multinational network operator. Conclusions of the developed work are intended to give some guidelines to the operator, essentially on network optimisation. The collected data was supplied by Vodafone and consist of the data traffic and the number of users per operating system and type of device registered during the last 3 years (2012/2013/2014) on 4 specific Radio Network Controllers (RNCs) located in the south of Portugal.

From this, one resorts to different statistical distributions that fit to the data (fitting models), in order to have a perception of the growing trend of those data along the years, and a better understanding of human behaviour and needs in terms of mobile communications usage. Finally, these distributions are used in the prediction of data behaviour (prediction models) throughout time (for 2015 and 2016), which in turn will allow to draw important conclusions.

The present thesis is composed of 5 different chapters and 3 annexes, as shown in Figure 1.4; it provides the general understanding of the work performed, used methods and obtained results.

The present chapter provides a general understanding of mobile communications systems and their adoption throughout the world, which justifies the need to be constantly developing new systems and new techniques that improve the performance of existing ones.

Chapter 2 provides a brief description of UMTS network architecture and radio interface, presenting their main elements. Services and applications considerations are explored, including Quality of Service (QoS) requirements and priorities. A research on mobile terminals and traffic evolution is done in order to understand the impact they have in the network. Finally, the state of the art is presented, containing the latest work developed related to this thesis.

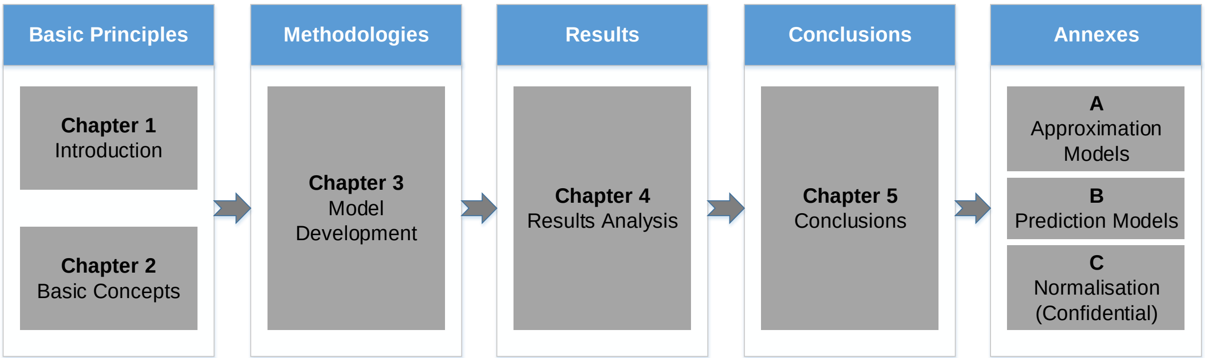

Figure 1.4. Thesis Structure.

Chapter 3 concerns the model developed for the purpose of this thesis. A detailed description of the simulator conceived to implement the models, including algorithms and important considerations and choices made, is presented. The process to obtain the fitting and the prediction models of the RNCs traffic (users, and data traffic) is also explained. Finally, a study to understand the tendency of user's behaviour as well as the most used mobile applications in these days is presented.

Chapter 4 presents the description of the scenarios considered in this thesis and the analysis of results. Some of the fitting and prediction models obtained are presented along with some important values that are useful for evolving towards network optimisation.

Chapter 5 summarises this thesis work, pointing out the main results and conclusions, followed by suggestions for future work.

At the end, a group of annexes is presented containing additional information. Annex A addresses some fitting models done to the collected data. Annex B presents the prediction models considered for the related scenarios. Annex C identifies the RNCs and presents the values used for the normalisation of the models.

.

Chapter 2

Basic Concepts

This chapter provides an overview of the fundamental concepts to understand this thesis. Section 2.1 presents the network architecture and radio interface of UMTS. Section 2.2 addresses services and applications for UMTS. Section 2.3 gives a detailed view of mobile terminals. Section 2.4 gives a brief overview about traffic evolution. Section 2.5 concludes the chapter with the state of the art.

2.1 UMTS Network

First, an overview of the fundamental concepts regarding the UMTS is done, based on [HoTo04]. The network architecture is presented together with its main elements, as well as a brief description of the radio interface.

2.1.1 Network Architecture

The UMTS architecture refers to the interconnection of the Access Network (AN), the UMTS Terrestrial Radio Access Network (UTRAN) to the adapted pre-Release 99 GSM/General Packet Radio Service (GPRS) Core Network (CN) infrastructure.

Functionally, the network elements are grouped into:

- UTRAN, which handles all radio-related functionalities.
- Core Network, which is the responsible for switching and routing calls and data connections to external networks.

It also includes the User Equipment (UE) that interfaces with the user and the radio interface.

Figure 2.1 shows the high-level network architecture divided into sub-networks.

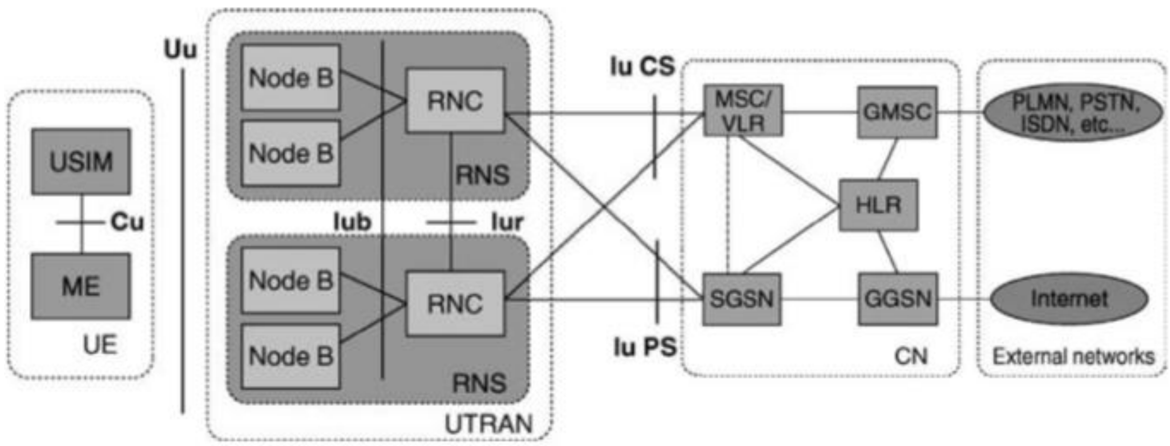

Figure 2.1. UMTS network architecture (extracted from [HoTo07]).

From a specification and standardisation point of view, the UE and UTRAN consist of completely new protocols, as they are based on the requirements of WCDMA, while the CN is adopted from GSM.

The UE sub-network consists of two parts:

- UMTS Subscriber Identity Module (USIM): a smartcard that holds the subscriber identity, performs authentication algorithms, and stores authentication, encryption keys, and some subscription information that is needed at the Mobile Terminal (MT).
- Mobile Equipment (ME): the radio terminal used for radio communications over the Uu

interface.

The UTRAN sub-network consists of two distinct elements:

- Node B: it corresponds to the more generic term Base Station (BS) and converts the data flow between the Iub and Uu interfaces. It also participates in Radio Resource Management (RRM).
- RNC: it owns and controls the radio resources in its domain (the Node B is connected to it by the lub interface), because it is the network element responsible for the control of the radio resources. The RNC is the service access point for all services UTRAN provides to the CN (e.g., management of connections to the UE, and handover control from BS to BS).

The CN is a migration of what was established in GSM, but with further overlaid elements to enable additional functionalities demanded by UMTS, being responsible for switching and routing calls and data to external networks. These external networks are:

- Circuit Switch (CS), which provides circuit-switched connections (telephony service).
- Packet Switch (PS), which provides connections for packet data services (Internet).

The main elements of the CN are the following:

- Home Location Register (HLR): a database that contains all the administrative information about each subscriber, such as allowed services, user location for routing calls, and preferences.
- Mobile Switching Centre/Visitor Location Register (MSC/VLR): the switch (MSC) that manages CS calls and the database (VLR) that serves the UE in its location.
- Gateway MSC (GMSC): it is effectively the interface to the external networks, the switch at the point where UMTS Public Land Mobile Network (PLMN) is connected to external CS networks. All incoming and outgoing CS connections go through GMSC.
- Serving GPRS Support Node (SGSN): it provides a number of functions (Mobility management, Session management, Interaction with other-areas-of-the-network, and Billing), and has similar functionalities to the MSC, but for PS.
- Gateway GPRS Support Node (GGSN): the central element between the PS network and external ones, and is analogous to that of GMSC, for PS.

There are some interfaces connecting the UTRAN internally or externally to other functional entities:

- Cu: electrical interface between the USIM smartcard and the ME.
- Uu: interface connecting the Node B with the UE.
- Iub: interface connecting a Node B and a RNC.
- Iur: interface connecting two RNCs with each other.
- Iu: external interface that connects the RNC to the CN.

2.1.2 Radio Interface

The MT and the BS communicate using WCDMA as the standard radio transmission, which is a wideband Direct-Sequence Code Division Multiple Access (DS-CDMA) system, whose main parameters are summarised in Table 2.1 [HoTo04].

With a chip rate of 3.84 Mcps, leading to a carrier bandwidth of 5 MHz, WCDMA supports highly variable data rates. Each user is allocated with 10 ms frames, during which the data rate is kept constant, being able to change only from frame to frame. Using the 5 MHz channel bandwidth, it has the capacity to carry around 100 simultaneous voice calls, or data at speeds up to 2 Mbps in its original version.

However, with the later enhancements of HSDPA and HSUPA included in later releases of the standard, the data transmission speeds have been increased to 14.4 Mbps.

Table 2.1. Main WCDMA parameters (extracted from [HoTo04]).

Multiple access method	DS-CDMA
Duplexing method	Frequency division duplex/time division duplex
Base station synchronisation	Asynchronous operation
Chip rate	3.84 Mcps
Frame length	10 ms
Service multiplexing	Multiple services with different quality of service requirements multiplexed on one connection
Multirate concept	Variable spreading factor and multicode
Detection	Coherent using pilot symbols or common pilot
Multiuser detection, smart antennas	Supported by the standard, optional in the implementation

Frequency Division Duplex (FDD) is the relevant mode of operation in UMTS, using a pair of 5 MHz band carrier frequencies, one for the UL and another for the DL. This operation mode is suited for applications in macro- and micro-cell environments, and high mobility.

Universal Terrestrial Radio Access (UTRA) usually occupies the band [1920, 1980] MHz for UL and [2110, 2170] MHz for DL [Moli11], but there are also implementations in the GSM 900 MHz band. Table 2.2 shows the values of DL and UL Transmission and Round Trip Time (RTT), and which release is used in 3GPP evolution.

After Release 99, HSDPA and HSUPA have emerged (patented as Release 5 and Release 6, respectively), both known as HSPA. HSDPA was designed to be deployed together with Release 99 and works as an enhancement of UMTS initial version, increasing DL packet data throughput by means of fast physical layer (L1) retransmission and combining transmission, and promoting fast link adaptation controlled by the Node B. In order to allow higher data rates, this new technology supports the 16 Quadrature Amplitude Modulation (16-QAM) with 4 bits per symbol, which can only be used under good radio signal quality due to additional decision boundaries [Frad11].

Table 2.2. 3GPP evolution and peak data rate (extracted from [Stuh12]).

	UMTS\HSPA\HSPA+			LTE	
Release	Rel 99/4	Rel 5/6	Rel 7	Rel 8/9	Rel 10/11
DL Transmission	384 kbps	14 Mbps	28 Mbps	150 Mbps	1 Gbps
UL Transmission	128 kbps	5.7 Mbps	11 Mbps	75 Mbps	500 Mbps
Round Trip Time	150 ms	100 ms	50 ms	10 ms	10 ms

Unlike Release 99, where the scheduling control was based on the RNC and the Node B only had power control functionalities, in HSDPA scheduling and fast link adaptation based on physical layer retransmissions were moved to the Node B, reducing latency and providing a whole change in RRM, improving in this way capacity and spectral efficiency.

In Release 99, radio transmissions are structured in frames of 10 ms, and transport data blocks are transmitted over an integer number of frames, with the Transmission Time Interval (TTI) being usually between 10 and 80 ms. Unlike this, HSPA supports a frame length of 2 ms, which leads to the reduction of latency and a fast scheduling among different users.

HSDPA does not support soft handover or fast power control. Quaternary Phase Shift Keying (QPSK) is mainly used to maximise coverage and robustness. HSDPA introduces Adaptive Modulation and Coding (AMC), which adjusts the modulation and coding scheme to the radio channel conditions, and, together with 16-QAM, allows higher data rates.

2.2 Services and Applications

3G systems are characterised by supplying the user with services beyond voice, or simple data transmission. A service is a set of capabilities that work in a complementary or cooperative way, in order to allow the user to establish applications, while an application is a task that needs communication among two or more points, being characterised by parameters associated with services, communications, and traffic. A service can be divided into three basic components: audio, video, and data, which are grouped into classes, according to their characteristics [Corr10]. In order to characterise traffic in UMTS, 3GPP specified four different QoS classes, which are summarised in Table 2.3.

The main distinguishing factor between these four classes is how delay sensitive the traffic is. While the Conversational class is meant for traffic which is very delay sensitive, the Background class is the most delay insensitive of them all, as it is described in [3GPP11].

The Conversational and Streaming classes are mainly intended to be carry real-time traffic flows. The

Conversational class is the one that raises the strongest and most stringent QoS requirements, as it is the only one where the required characteristics are strictly given by human perception, examples of video conferencing and Voice over IP (VoIP). The Streaming class preserves time relation between information entities of the stream and require bandwidth to be maintained like Conversational, but tolerate some delay variations that are hidden by dejitter buffer in the receiver. Interactive and Background classes are mainly meant to be used by traditional Internet applications like World Wide Web (WWW), e-mail, and File Transfer Protocol (FTP).

Table 2.3. Service classes and requirements for UMTS QoS (extracted from [3GPP11]).

Service Class	**Conversational**	**Streaming**	**Interactive**	**Background**
Real Time	Yes	Yes	No	No
Symmetric	Yes	No	No	No
Switching	CS	CS	PS	PS
Guaranteed Bit Rate	Yes	Yes	No	No
Affordable Delay	Minimum (Fixed)	Minimum (Variable)	Moderate (Variable)	High (Variable)
Buffer	No	Yes	Yes	Yes
Bursty	No	No	Yes	Yes
Example	Voice	Video-Clip	Web Browsing	E-mail

Due to looser delay requirements, compared to Conversational and Streaming classes, Interactive and Background provide better error rate by means of channel coding and retransmission. The main difference between Interactive and Background is that the former is mainly used by interactive applications, e.g., Web browsing, while the latter is meant for background traffic, e.g., e-mails or background file downloading. The responsiveness of interactive applications is ensured by separating Interactive and Background applications. Traffic in the Interactive class has higher priority in scheduling than Background one, so Background applications use transmission resources only when Interactive applications do not need them. This is very important in a wireless environment where the bandwidth is low compared to fixed networks [ETSI04].

It is also important to know the usual bit rates and typical file dimensions associated with the different services, as well as their QoS priorities. Table 2.4 shows some examples. The first services to be reduced, if reduction strategies of data rates are used, are those with the lowest QoS priority (a value corresponding to the highest priority), according to the traffic classes shown in Table 2.3.

Mobile Internet is the combination of mobile communications and Internet. Mobile communications and Internet have gained their own great achievements. However, their terminal modes, network architectures, application categories, and user behaviours obviously differ [HUAW12].

If the Internet, mainly providing data service, is integrated into mobile communications, which provide voice service, great impacts are inflicted on network resource efficiency, capacity, and signalling. The use of cloud computing services (cloud computing) overcomes the limitations of memory and computing power of mobile devices that prevented them from acting as major consumers of content and consequently mobile data. Services like Netflix, YouTube, Pandora, Facebook, Spotify, Dropbox, SkyDrive, among others, have overcome these limitations. Currently a Smartphone that uses these services generates twice the traffic volume in relation to another that only uses web applications and e-mail. Machine-to-Machine (M2M) communications are being used by many industries.

Table 2.4. Bit rates and applications of different services (adapted from [Lope08], [3GPP05] and [Khat14]).

Service		Bit rate [kbps]		QoS priority	Characteristics		
		DL	UL		Average volume/duration	DL	UL
Voice		12.2		1	Call duration [s]	120	
Web		[512, 1536]	[128, 512]	2	Page size [kB]	300	20
Streaming		[512, 1024]	[64, 384]	3	Video size [MB]	9.6	0.02
Email		[384, 1536]	[128, 512]	4	File size [kB]	100	
FTP		[384, 2048]	[128, 512]	5	File size [MB]	10	
Chat		[64, 384]		6	MSN message size [B]	50	
Peer-to-peer		[128, 1024]	[64, 384]	7	File size [MB]	12.5	
M2M	Smart Meters	200		5	File size [kB]	2.5	
	e-Health	200		1	File size [MB]	56.	
	ITS	200		2	File size [kB]	0.06	
	Surveillance	[64, 384]		2	File size [kB]	5.5	

In addition to the increasing number of devices, new created services will grow the volume of data transmitted by mobile networks. Likewise, the bandwidth is increasingly becoming a concern. It is expected that such equipment uses 3G and 4G instead of 2G. With the development of mobile Internet in recent years, its service categories and characteristics are different from the traditional Internet. Table 2.5 describes the categories of current mobile Internet and their main characteristics.

The preceding features are defined as follows:

- If packet per second (PPS) is greater than 20, the data is transmitted continuously.
- If PPS is less than 10, the data is transmitted less frequently.
- A data packet larger than 1000 bytes is defined as a big packet.
- A data packet less than 600 bytes is defined as a small packet.

The main traffic volume for mobile Internet is used for web browsing, and the rest is used for streaming media and file transfer. Mobile Internet is widely deployed and the traffic rate increases.

Smartphones and Tablets are equipped with more functions. Mobile streaming media services are widely used and the main traffic volume is occupied by video service. Instant communications with text, voice, and video are more preferable, and network access becomes more frequent. Meanwhile, the Hypertext Markup Language 5 (HTML5) technique becomes increasingly mature. Cloud service replaces traditional web browsing and file transfer as the dominant player. More and more smart machine terminals and M2M services, such as smart electrical household appliances, auto meter reading, and mobile payment come into play [HUAW12].

Table 2.5. Mainstream mobile Internet categories and characteristics ([adapted from HUAW12]).

Category	**Description**	**Typical Application**	**Characteristic**	**Impact**
IM	Sending or receiving instant messaging	WhatsApp, WeChat, iMessage	Small packets are sent occasionally	Increasing signalling for calling and called parties and reduced resource efficiency
VoIP	Internet telephone service, including voice and video calls	Viber, Skype, Tango, Face Time	Small packets are sent continuously	Reduced resource efficiency
Streaming	Streaming media such as HTTP audios and videos, Peer-to-peer videos	YouTube, Spotify, Pandora, PPStream	Big packets are sent continuously	Large amount of downlink data
SNS	Social networking websites	Facebook, Twitter, Sina Weibo	Small packets are sent less frequently	Increasing signalling for calling and called parties and increasing uplink and downlink data
Web Browsing	Web page browsing including wireless access protocol	Typical web browsers are Safari and UC Browser	Big packets are sent less frequently	Increasing signalling and downlink data
Cloud	Applications, including cloud computing and online cloud applications	Siri, Evernote, iCloud	Big packets	Increasing signalling and uplink data
Email	Emails, including Web mail, Post Office Protocol 3 and Simple Mail Transfer Protocol	Gmail	Big packets are sent less frequently	Increasing signalling and uplink and downlink data
File Transfer	File transfer including Peer-to-peer file sharing, file storage, application download and update	Mobile Thunder, App Store	Big packets are sent continuously	Increasing signalling and uplink and downlink data

Table 2.5 (contd.). Mainstream mobile Internet categories and characteristics ([adapted from HUAW12]).

Gaming	Mobile gaming such as social gaming and bridges	Angry Birds, Draw Something, Words with Friends	Big packets are sent less frequently	Increasing signalling and uplink and downlink traffic data
M2M	Machine Type, communication	Auto meter reading, mobile payment	Small packets	Increasing signalling for calling and called parties and reduced resource efficiency

2.3 Mobile Terminals

In recent years, there has been a strong growth in data transfer between terminals due to the increasing capacity and functionality thereof. Nowadays, Smartphones and Tablets are the most used terminal devices, which have immense resources. In these devices, mobile calls have been losing interest compared to data transfer. The mass access to the global network, the Internet, allows an exchange of information easily and quickly dropping physical location barriers, being desirable to be accessible as long as possible. Thus, mobile systems must have the ability to meet all requests and increased data.

In order to better frame this theme in this section, it is important to define what a mobile terminal is. This is a broad concept that in a generalist view encompasses two meanings. The first relates to the possibility of physical mobility of the device. This definition can include multiple terminals. If it is easy to classify a Smartphone and a Laptop as a mobile terminal, associated with the context of physical mobility, the same cannot be said for a mainframe server. The second meaning relates to the access to information from the device. This refers to the communication characteristics of the device when it comes to Internet connection. The terminal can only display links when one is physically stationary, although it can move losing connectivity in those moments, or it can guarantee connectivity during its displacement, usually dubbed as mobile Internet terminal.

Within this thesis, the focus lies at the terminals that comprise the second meaning, enabling communication during their physical movement. These are now part of everyday life in the form of phones with access to mobile data communication, such as Smartphones, and more recently Tablets, that are used in countless ways in which the interaction without barriers to mobility is desirable.

Nowadays, it is easy to find access points for wireless networks spread in any urban centre - inside and outside of buildings. In recent years, there has been the deployment of successfully better mobile data technologies (UMTS, HSPA, HSPA+ and LTE) that allow sufficient bandwidth for the increasing flow of information. The increasing number of wireless devices that are accessing mobile networks worldwide is one of the primary contributions to traffic growth.

Each year, several new devices in different form factors and increased capabilities and intelligence are being introduced in the market. According to Figure 2.2, in 2017 there will be 8.6 billion mobile or personal ready mobile devices. Tablets have been gaining market day by day: at first they were very limited in their functionality, resembling a large Smartphone screen, but over the years they won numerous functions, such as creating and editing files, cloud connection, Secure Digital (SD) card inputs, Universal Serial Bus (USB), High-Definition Multimedia Interface (HDMI), excellent processing and storage, camera, and long term batteries. Figure 2.3 and Figure 2.4 show the most used screen resolutions in Tablets and Mobiles in 2014 in Europe, i.e., 1024x768 for the former and in 568x320 for the latter.

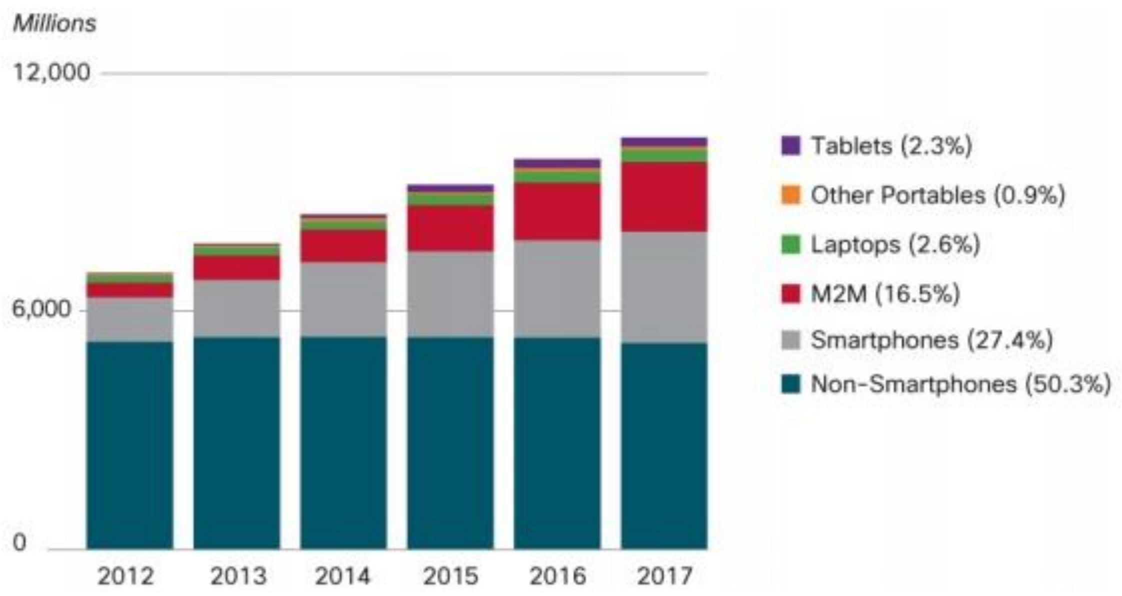

Figure 2.2. Global mobile devices (extracted from [Cisc13]).

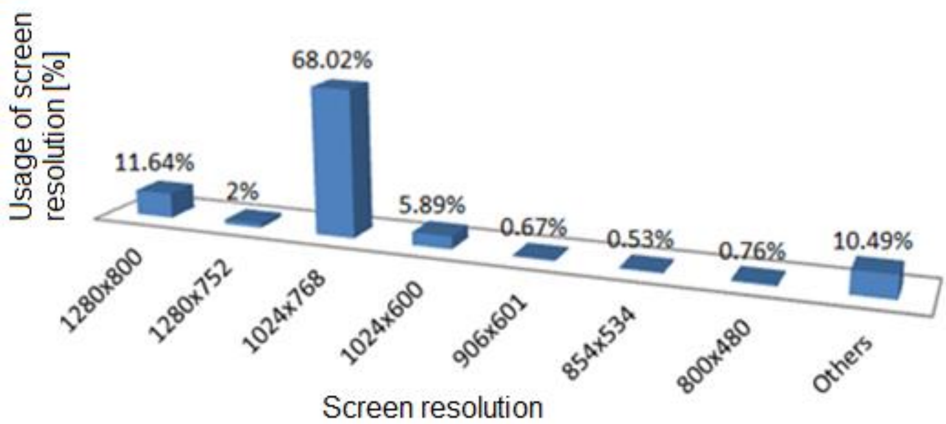

Figure 2.3. Utilisation ratio of different Tablet screen resolutions (adapted from [GISt14]).

The signalling data of cellular phones can be used as valuable information for state-of-the-art traffic applications, especially in urban areas. If a mobile phone is moving, different types of signalling events are generated, which may provide suitably accurate location data of terminals and thus can be exploited to realise location based services [Küpp05]. Most of the Smartphones these days run under iOS (iPhone OS) or Android operating systems. Based on mature iOS and software on protocol stack, Apple devices provide services of fast dormancy, being online permanently, and push notifications.

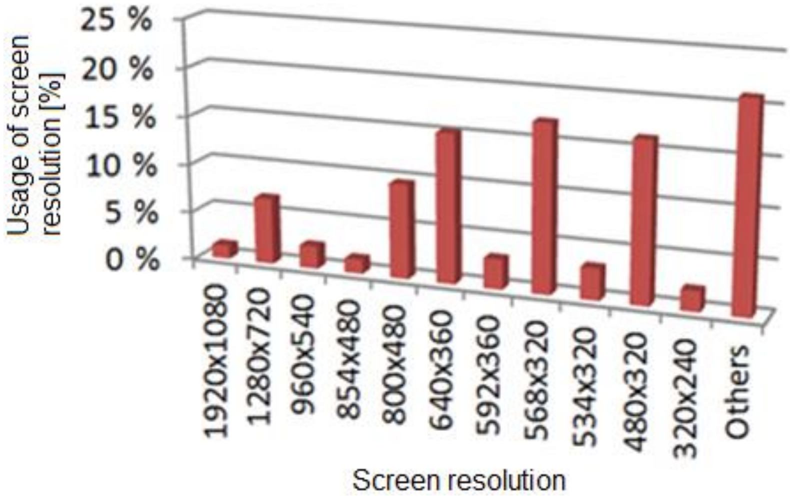

Figure 2.4. Utilisation ratio of different Mobile screen resolutions (adapted from [GlSt14]).

The network resource utilisation and user experience of push services due to permanent online requirement are different for iOS and Android devices. For iOS, background applications do not generate cellular data flows: the heartbeats of background services are regarded as those for Apple push server, these services being in the deactivated status. For Android, most background services have a single heartbeat. The unified heartbeat mechanism in iOS reduces the frequent network connection requests and disconnection signalling during screen off [HUAW12]. Table 2.6 describes the comparison of background behaviours for screen off between iOS and Android devices.

According to Table 2.6, network connection requests in one hour for iOS and Android are 2 and 30 respectively. When the terminal is in the connected status, but without push messages, the number of connections for devices with Android is 15 times more than those for devices using iOS. Frequent connection requests from devices with Android bring congestion for network.

Table 2.6. Background behaviours for screen off between iOS and Android devices ([adapted from HUAW12]).

Background Behaviour	Android	iOS
QQ	Heartbeat cycle: 540s	No heartbeat
WhatsApp	Double heartbeats: one with cycle of 285s, and the other with a cycle of 900s	No interaction if heartbeat stops in 15 minutes of screen off
Facebook	Heartbeat cycle: 3600s	No heartbeat
Twitter	Heartbeat: cycle 900s	No heartbeat
Sina microblog	No heartbeat	No heartbeat
OS heartbeat	Gtalk cycle: 28 minutes	Heartbeat cycle adaptive to firewall aging time: 30 minutes
Number of interactions per hour	30	2

As one can see in Figure 2.5, the most widely used operating systems for Smartphones and Tablets are Apple iOS, Google Android, and Microsoft Windows Phone. While iOS can only be found in iPhones, iPads and other unique models of Apple, Android is generally used by many different brands.

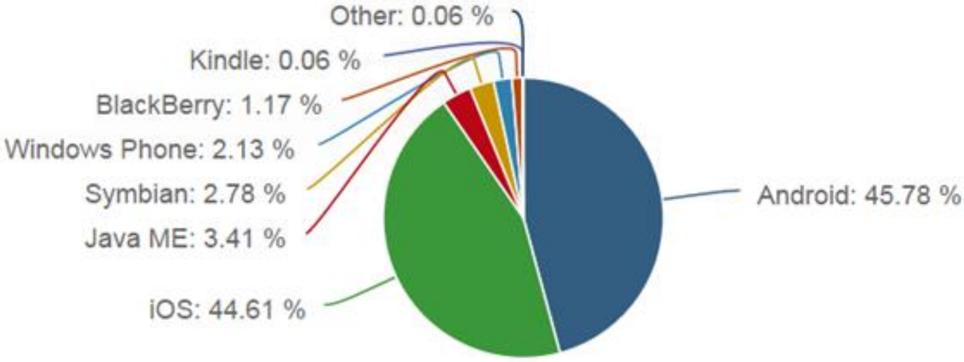

Figure 2.5. Mobile/Tablet OS Market Share in 2014 (extracted from [NMSh14]).

Figure 2.6 shows the share of in mobile data traffic growth. Today, laptops generate a relevant amount of traffic, but Smartphones, Tablets and M2M nodes will begin to account for a more significant portion of traffic by 2017. It is possible to see that Smartphones lead the traffic growth.

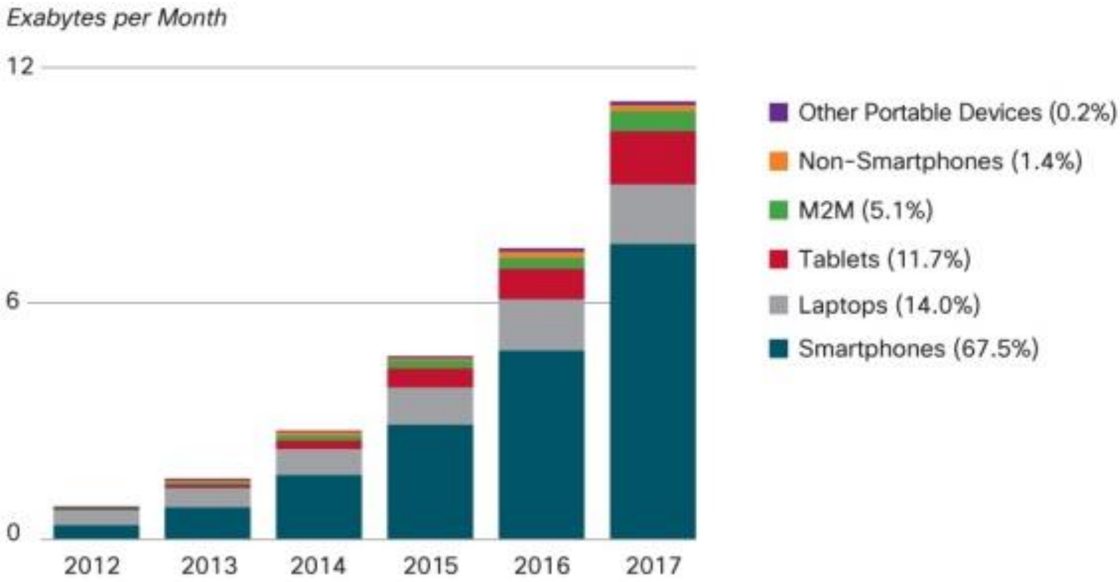

Figure 2.6. Mobile data traffic growth (extracted from [Cisc13]).

Video is the largest and fastest growing segment of mobile data traffic. It is expected to grow by approximately 45% annually through to 2020, as it is shown in Figure 2.7. YouTube still dominates video traffic in most mobile networks and accounts for 40% a 60% of total video traffic volume in many mobile networks.

Audio traffic, still expected to increase at an annual rate of around 35%, is in line with total mobile traffic growth. Today, social networking constitutes around 15% of total mobile data traffic. Its overall market share will remain at the same level in 2020, even though social networking will increasingly include data-rich content [Eric14]. The relative share of traffic generated by web browsing will decline by 2020 as a result of stronger growth in categories such as video and social networking. Consumer preferences are shifting towards more video and app-based use relative to web browsing.

The emergence of new applications can shift the relative volumes of different types of traffic, but the proliferation of specific devices can also affect the traffic mix – for example, Tablets are associated with a much higher share of online video traffic than Smartphones.

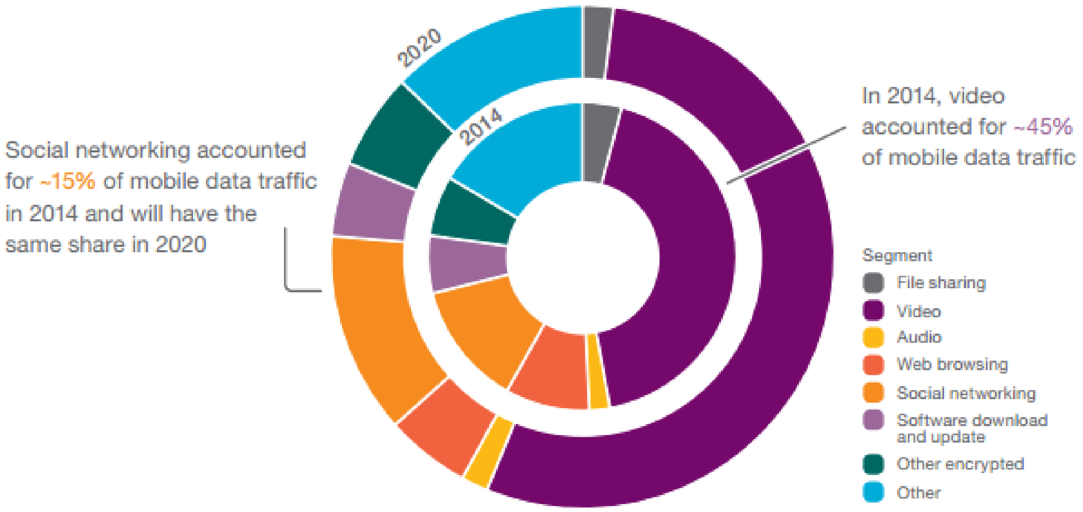

Figure 2.7. Share of mobile data traffic by application type (extracted from [Eric14]).

Several factors are contributing to a growth in mobile video traffic. Prominent among them is the number of video-capable devices that are being used by consumers. The devices themselves are also evolving: many now have larger screens, enabling higher picture quality for streamed video. Video content is increasingly appearing as part of other online applications, e.g., news, advertisements, and social media. Streaming video is growing strongly, primarily driven by Over-The-Top (OTT) providers, e.g. YouTube and Netflix [Eric14]. There are several variables that affect the performance of a mobile connection: the number of devices sharing a BS, signal strength availability, or the type of application; depending on the type of application, DL and UL speeds and latency characteristics vary widely. Figure 2.8 shows that the speed of Tablets is foreseen to be higher than the one of Smartphones.

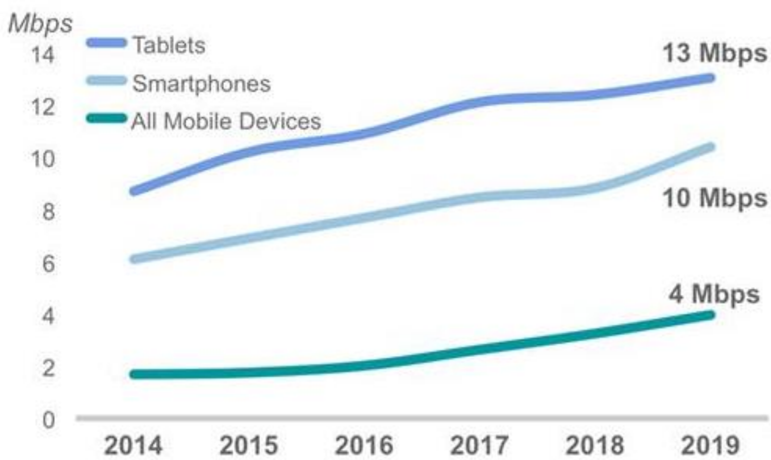

Figure 2.8. Mobile Speeds by Device from 2014 to 2019 (extracted from [Cisc15]).

2.4 Traffic Evolution

There are several reasons for the growth of mobile data. Technological developments, the diversity of new equipment and services, and consumer habits have changed over the past few years. Currently, there are little more than 6.8 billion mobile subscriptions worldwide, and this number is predicted to reach 9.2 billion by 2019, while global Mobile Broadband (MBB) subscriptions should reach 7.6 billion by the same year [Eric14]; by the end of 2019, MBB subscriptions are expected to account for more than 80% of all mobile ones. The mobile voice service is already considered a necessity by most people, and mobile data, video, and TV services are becoming an essential part of consumers' life. Used extensively by consumer as well as enterprise segments, with impressive uptakes in both developed and emerging markets, mobility has proven to be transformational.

Mobile subscribers are growing rapidly, and bandwidth demand due to data and video is increasing. Mobile M2M connections continue to increase. The next years are projected to provide a fast mobile video adoption. The proliferation of mobile and portable devices, such as Smartphones, Tablets and Laptops connected to mobile networks, is great at generating traffic, since these devices are able to offer the consumer content and applications that were not supported by mobile devices of the previous generation [Cisc15]; there is an imminent need for networks to allow all these devices to be connected transparently, with the network providing high-performance computing and delivering enhanced real-time video and multimedia. This openness will broaden the range of applications and services that can be shared, creating a highly enhanced MBB experience.

The expansion of wireless presence will increase the number of consumers who access and rely on mobile networks, creating a need for greater economies of scale and lower cost per bit [Cisc15]. In general, traffic occurs in bursts and clumps that are only partly predictable. Many intranet sites have activity peaks at the beginning and end of the day, and around lunchtime; however, the exact size of these peaks will vary from day to day and, of course, the actual traffic load changes from moment to moment. There is a direct relationship between the amount of traffic and the network bandwidth needed. As the number of subscribers increases, more capacity is required by the server [Micr14].

From Figure 2.9, it is possible to conclude that as time goes by, the total traffic generated in the world begins to be distributed by more people, therefore not being so concentrated. It appears that there is an increase in the traffic generated by the main subscribers in the months of February and March.

By analysing Figure 2.10, the months of April, June, July and August are the ones in which a larger amount of traffic from the main subscribers is realised, mostly because people are on holidays.

The tendency to carry traffic has risen gradually over time and it has a significant increase in July, August and September, as shown in Figure 2.11.

The percentage of mobile data users generating more than 2 GB per month was nearly 25% by September 2013, and the percentage of users generating more than 200 MB per month reached 75% in September 2013.

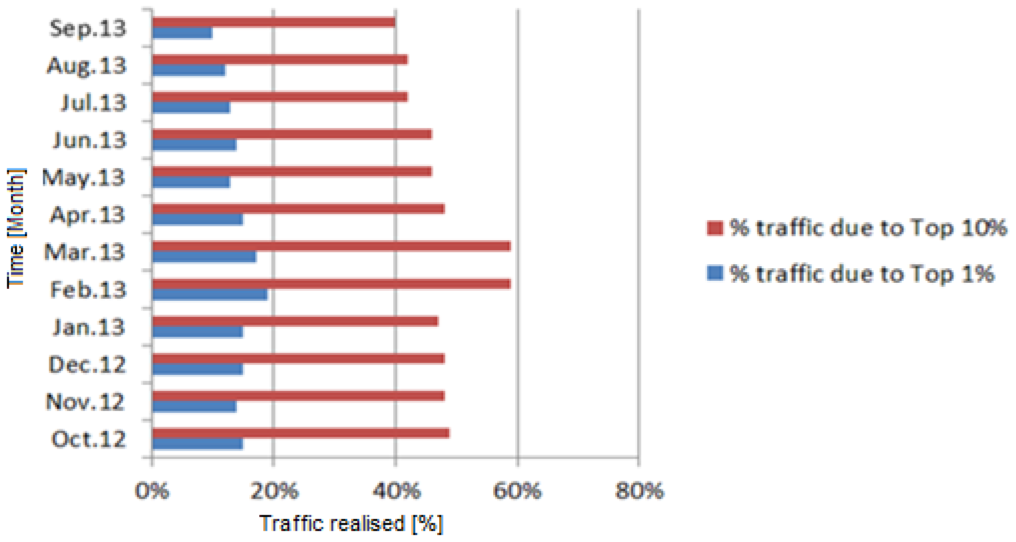

Figure 2.9. Percentage of Traffic by User Tier (adapted from [Cisc14]).

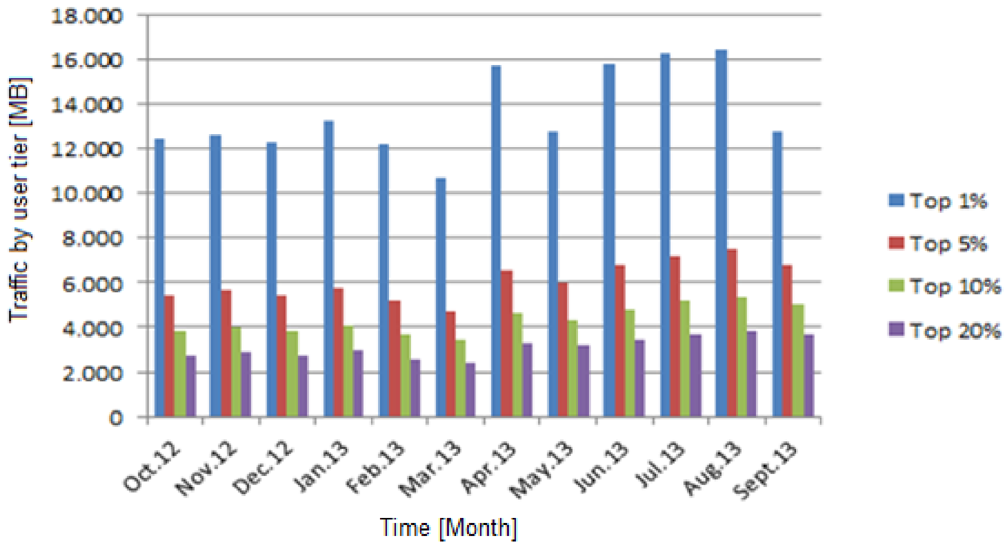

Figure 2.10. Average Traffic by User Tier in MB per Month (adapted from [Cisc14]).

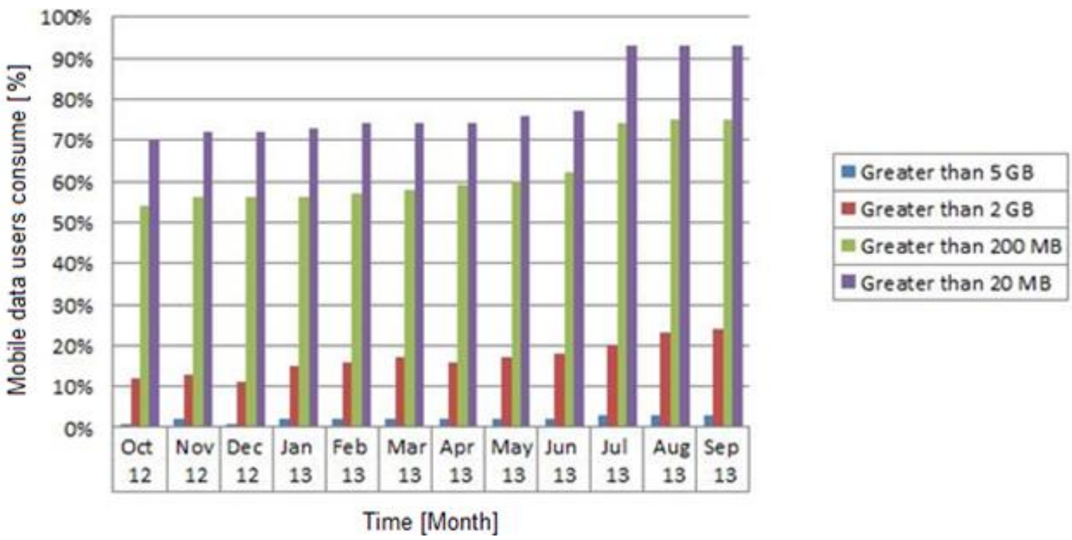

Figure 2.11. One percentage of Mobile Data Users Consume 5 GB per Month (adapted from [Cisc14]).

The tiered pricing plans have a lower MB per month consumption compared to unlimited plans. The rapid increase in data usage presents a challenge to operators that have implemented tiers defined solely in terms of usage limits, and the average consumption per connection continues to increase for both tiered and unlimited plans [Cisc14]. It is possible to see in Figure 2.12 that Android devices led in average MB per month usage in unlimited plans.

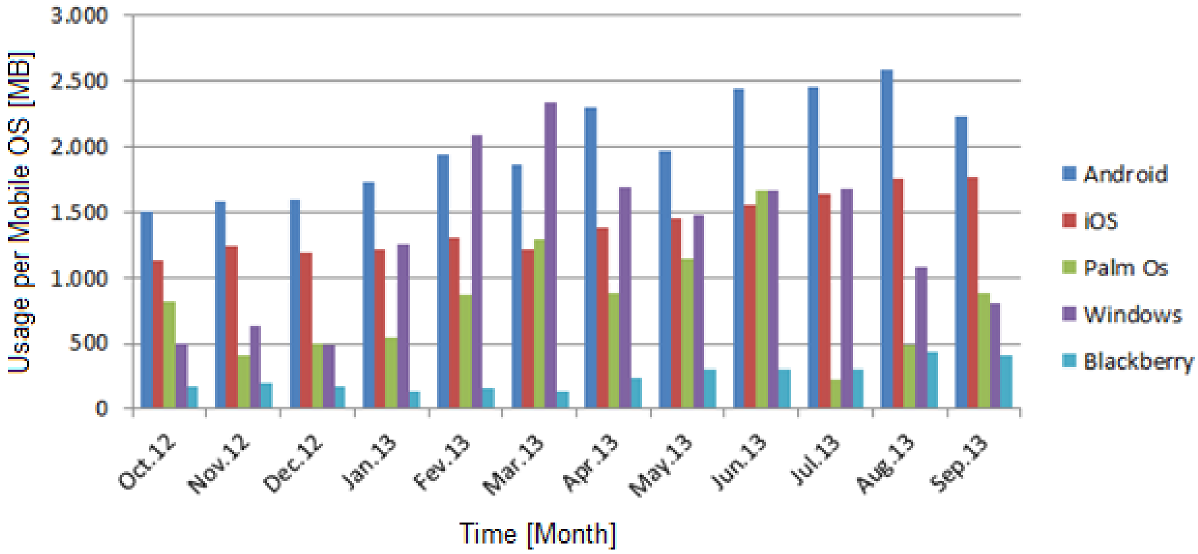

Figure 2.12. MB per Month Usage per Mobile OS in Unlimited Plans (adapted from [Cisc14]).

Otherwise, when the tiered pricing plan is used, iOS devices lead in average MB per month as shown in Figure 2.13.

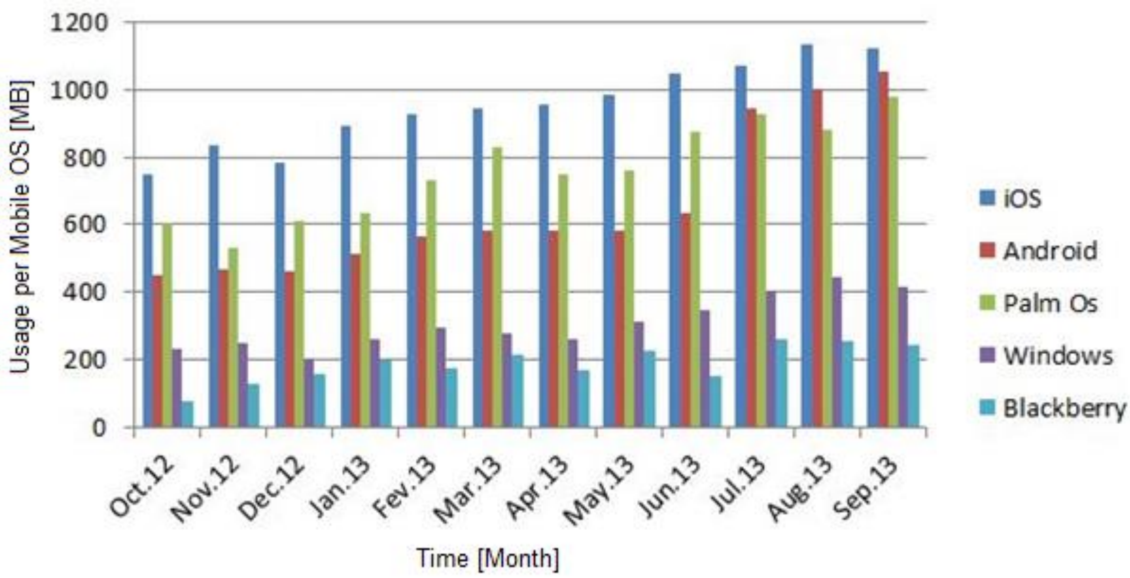

Figure 2.13. MB per Month Usage per Mobile OS in Tiered Pricing Plans (adapted from [Cisc14]).

Cellular networks got faster, and Smartphone screens got bigger. As a result, people's consumption of mobile data increased as well. The growth in usage per device outpaces the growth in the number of devices. As it is shown in Table 2.7, the growth rate of mobile data traffic from new devices is two to five times greater than the growth rate of users.

Table 2.7. Comparison of Global Device Unit Growth and Global Mobile Data Traffic Growth (extracted from [Cisc15]).

Device Type	Growth in Devices, 2014-2019	Growth in Mobile Data Traffic, 2014-2019
Smartphone	16.7%	60.1%
Tablet	32.0%	83.4%
Laptop	5.3%	22.3%
M2M Module	45.5%	102.7%

It is possible to see the evolution of data traffic in Figure 2.14, which is expected to grow to 24.3 EB per month by 2019, nearly a tenfold increase over 2014. According to [Cisc15], the mobile data traffic will grow at 57% from 2014 to 2019.

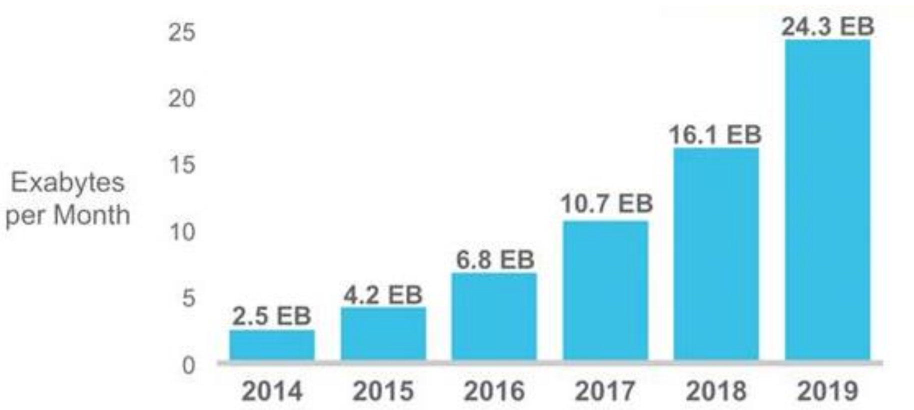

Figure 2.14. Global Mobile Data Traffic from 2014 to 2019 (extracted from [Cisc15]).

2.5 State of the Art

A brief overview of the state of the art is presented in this section, in order to show what has been done in this field up to now, thus, emphasising the importance of this work.

Analytics is the discovery and communication of meaningful patterns in data (corporate, product, channel, and customer), and is changing expectations and business strategies. How organisations instrument, capture, create and use data is fundamentally changing the dynamics of work, life and leisure [Pred14]. Data analytics (DA) examines raw data with the purpose of drawing conclusions about that information, and is used to allow companies and organisations to make better business decisions, as well as to verify or disprove existing models or theories. Data miners sort through huge data sets using sophisticated software in order to identify undiscovered patterns and establish hidden relationships.

People and devices are constantly generating data. While streaming a video, playing the latest game with friends, or making in-app purchases, user activity generates data about their needs and preferences, as well as the quality of their experiences. Even when they put their devices in their pockets, the network is generating location and other data that keep services running and ready to use.

In 2014, estimates put worldwide data generation at a staggering 7 ZB [ViOE11], and by 2018 each Smartphone is expected to generate 2 GB of data every month [Eric13]. At the same time, the big data technology and services market is expected to grow at a 40% Compound Annual Growth Rate (CAGR), with revenues expected to reach United States Dollar (USD) 16.9 billion in 2015 [IDCA15].

Clearly, the age of big data has begun. Operators can make use of these big data in order to drive a wide range of important decisions and activities, such as: designing more competitive offers, prices and packages; recommending the most attractive offers to subscribers during the shopping and ordering process; communicating with users about their usage, spending and purchase options; configuring the network to deliver more reliable services; and monitoring Quality of Experience (QoE) to proactively correct any potential problems [Eric13].

The profound impact that increased broadband networking has on society also creates business opportunities in new areas for operators. With improved real-time connectivity and data management comes the possibility to create tailored data sets, readily available for analysis and machine learning.

For an operator, big data not only means a fundamental shift in the way data is stored and managed – it also entails deploying powerful real-time analytics and visualisation tools, collaboration platforms, and the ability to automatically create links with existing applications vital to the operator's business, such as operations and business support systems (OSSs/BSSs) and customer relationship management (CRM) [Eric13].

Combining what people write in social media with network data can play a vital role in fully understanding the way users experience a service or knowing whether they are likely to churn. In practical terms, there are several important factors to be considered to manage the expansion of data, devices and network complexity:

- The growth of MBB.
- Enhanced connectivity and the adoption of M2M.
- The power of cloud computing.
- Service expansion and the shift of scale and context with relation to such services.
- Easy and cheaper access to big data technologies.

The complexity of handling this expanded universe of data sources is compounded by the need to link, match and transform data across business heterogeneous entities and systems, while managing scale and timeliness. Organisations need to understand key data relationships, such as complex hierarchies and links between data types and sources [Eric13].

There are several recent measurement studies that cover different aspects of cellular traffic [FaLy10]. However, these studies either consider overall cellular traffic (agnostic to device types), or focus specifically on Smartphone usage. None of them have considered the differences between Smartphones, feature phone, and air cards. In [FaMa10], a comprehensive study of the Smartphone usage is presented from four perspectives:

- User interactions with the phone.
- Application usage.
- Network traffic.
- Energy drain.

[ShWa11] is one of the previous works that studied the differences among various types of cellular devices. Their study showed that the accuracy of the models describing cellular Internet traffic volume can be improved by customisation based on device differences. In particular, cellular devices were grouped according to their trademarks and version numbers. On the other hand, [LiGa12] conducts a comparative study of the usage characteristics of three different device types: Smartphones, feature phones, and air cards. Moreover, one considers overall usage of the cellular device, including not only data traffic, but also voice and short messages.

Mobile phone data is massive, having a wide scope of coverage, and real-time acquisition. The accuracy positioning can meet the requirement, and time interpolation processing has greatly improved Data Availability. In [DoCh14], on the basis of vast mobile phone data from Beijing, combined with urban road network data, the residents' traveling Organisational Development (OD) analysis and the feature analysis of traffic zone are performed. The results show that using mobile phone data can be a convenient and rapid way to access the important traffic characteristics, such as the OD information, for traffic planning and management. In future studies, much more deep and extending researches may be taken, such as analysing travel trajectory characteristics based on special contextual features and the identification of transportation modes based on travel features.

[PaKi12] investigates the effect of a new mobile communication service, LTE, on the number of subscribers and traffic volume of the traditional service, 3G, in Korea. The methods of forecasting 3G traffic volume are divided into the number of service subscribers and the traffic volume per user. Using these methods, it was found that the use of LTE in July was higher than the use of 3G in the same month. Due to the traffic consuming characteristics of LTE, induced traffic rate was shown to be greater than the substitution rate for 3G, considering the rate of LTE subscribers. Further study is planned to investigate the effect of LTE adoption with a more sophisticated forecasting model.

Chapter 3

Model Development

In this chapter, the database search algorithm is described, followed by its application and some test results presentation. The way how the problem of this thesis is solved and how the models are implemented is also described. Finally, some examples of users' behaviour are presented.

3.1 Temporal Modelling of User Behaviour

Users' online behaviour and interests will play an increasing role in future mobile networks. Mobile networks are evolving and integrating with every aspect of people's lives. Laptops, Tablets and Smartphones are becoming ubiquitous, providing almost continuous Internet access. This creates a tight coupling between users and mobile networks, where various characteristics of users' online behaviour significantly affect network performance. Much of the previous mobility or web usage modelling focused on individual behaviour. However, while individual behaviour is important, investigating group behaviours and trends is more challenging and involved.

Consumers have clear preferences when it comes to the selection of screens to perform particular activities, whether for work or entertainment. Figure 3.1 shows the usage of different devices by people to perform various activities in 23 countries [Eric14], where it can be concluded that browsing the Internet and social networking are common activities across all devices. However, when it comes to communication, Smartphones are the preferred devices, whereas shopping and work/study related activities are more commonly performed on Personal Computers (PCs), with Tablets being used for more video-centric activities, such as gaming and watching videos.

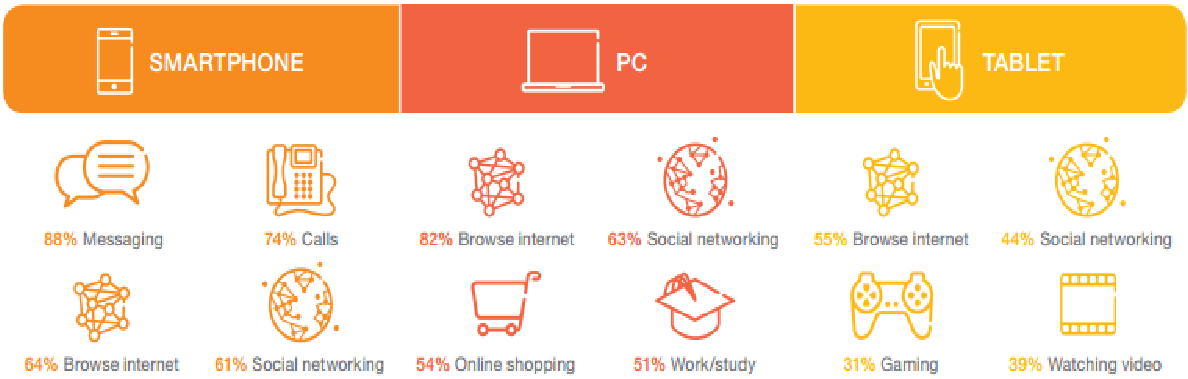

Figure 3.1. Top four services and the proportion of people performing each activity on a particular screen several times a week (extracted from [Eric14]).

The data traffic variation profile along the day is achieved by gathering daily data traffic. The profile evolution can be analysed as the general tendency of daily data traffic generation at the mobile level. Figure 3.2 presents the data traffic variation profile along the day, showing that the usage volume on Smartphones, Tablets, and Desktop Computers varies along the day.

Tablets are increasing in popularity to become our "third screen", and our browsing habits on Smartphones and Desktops shift to accommodate the new medium [Goog15]. Desktops/Laptops are used predominantly during business hours. Usage rises at 9am and falls at 6pm, with a small spike around 8pm. Smartphone usage increases throughout the day, spiking during the morning and in the evening, and Tablets get a rest during the day, but are used intensively in the evening.

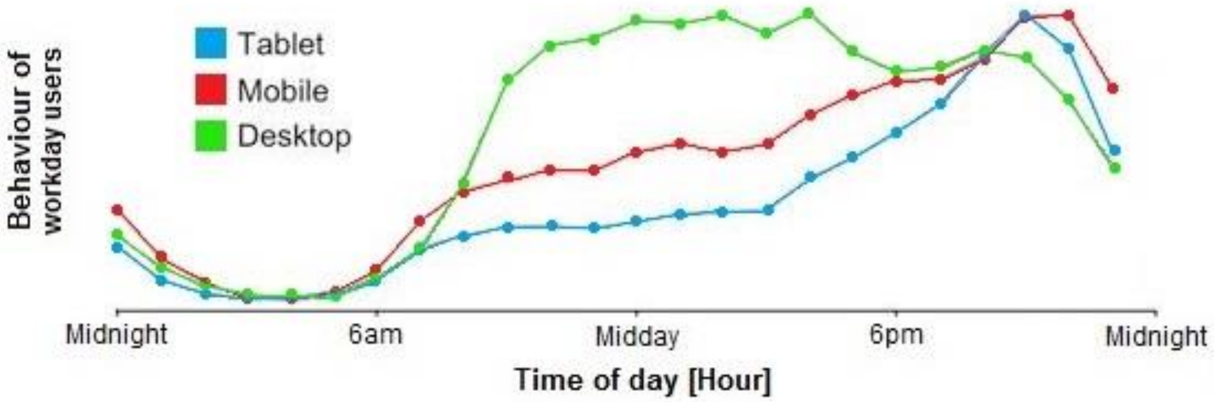

Figure 3.2. Behaviour of workday users per device during the day (adapted by [Goog15]).

Users' behaviour is changing, resulting in video being consumed on all types of devices and in higher quantities. Higher video resolutions, such as Ultra High Definition (UHD), are also emerging, although its impact on mobile devices has yet to be seen. All of these factors drive mobile video traffic volumes. The volume of encrypted video is also increasing rapidly, and technological improvements, such as new video compression techniques, will lead to more effective usage of data throughput, and will help mobile network operators to accommodate the increasing demand.

Nowadays, the access to a mobile application is done in a simple touch. Figure 3.3 shows that a good time for news publishers to reach consumers on their devices is during the morning commute. The readerships of the News application hit a peak at 7am, dropping off at 9am; for other categories, usage shows a steady improvement from 6am until about 9pm.

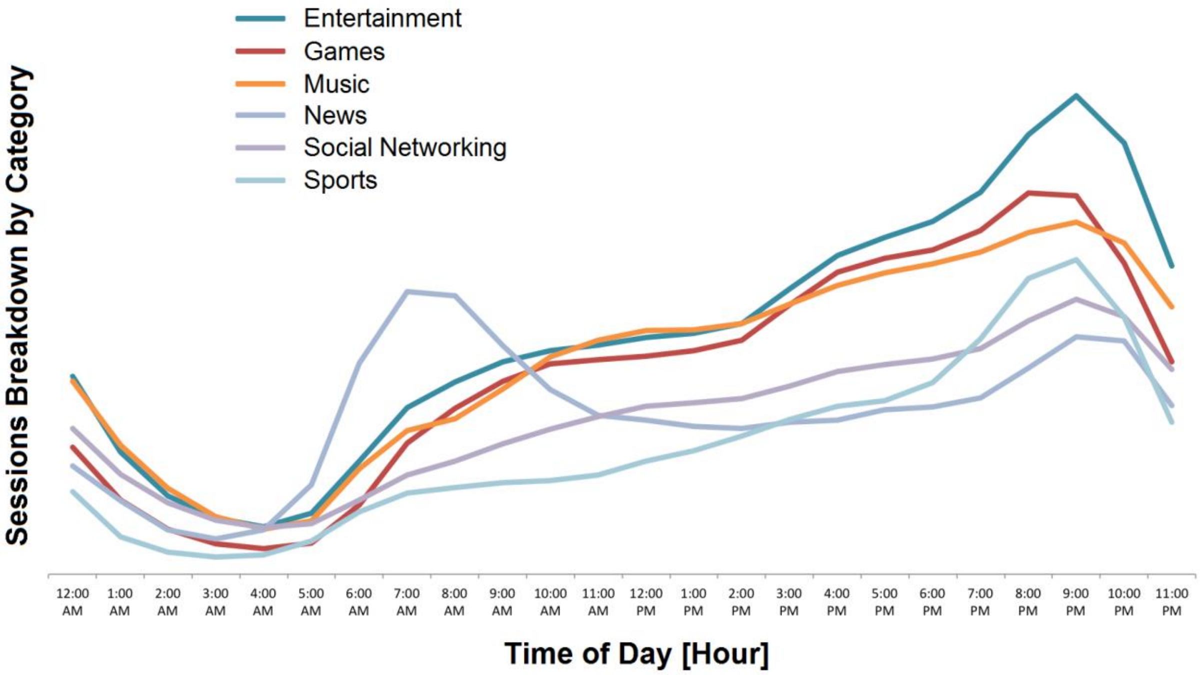

Figure 3.3. Behaviour of the queries of applications during the day (extracted from [Loca15]).

3.2 Model Overview and Implementation

3.2.1 Model Overview

In order to implement the models and evaluate the data, the MATLAB *software* was used and a programme was developed in the Wolfram Mathematica *software*, in order to aid in the data processing and calculations. A flowchart of the programme is illustrated in Figure 3.4. The input data, *.csv files, was taken in the time period January 2012 and December 2014. Besides inputs and outputs, one can also see the three main blocks that constitute the model: Initial Data Analysis, Fitting Models, and Prediction Models. The flowchart of each block is presented in the following sections. Data are properly analysed, and different fitting models are created, obtaining approximations of the annual behaviour of these parameters, in order to successfully build the predictive models for the behaviour for the trend in 2015 and 2016, which is the output.

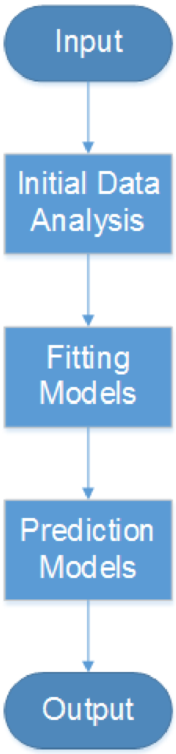

Figure 3.4. Model Overview.

A lot of effort was put into these modelling approaches and implementations, in order to get a more realistic approach to the network's behaviour; all models have been implemented from scratch in the thesis. In order to get a more efficient approach, a careful analysis off all files has been done.

3.2.2 Initial Data Analysis

After receiving the input, the 4 different RNCs are identified, as well as the respective parameters to analyse, which are the number of users and the corresponding data traffic according to the brand, device type and operating system. The main structure of this step is shown in Figure 3.5.

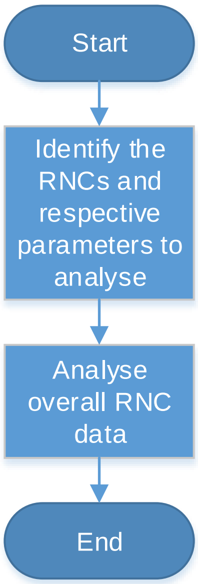

Figure 3.5. Initial Data Analysis of the model.

After this step, data are properly organised into several tables in order to study the evolution over the years of:

- Number of users of each individual RNC (4 predictions).
- Data traffic of each individual RNC (4 predictions).
- Total number of users of the 4 RNCs (1 prediction).
- Total data traffic of the 4 RNCs (1 prediction).
- Number of users depending on the type of operating system of the 4 RNCs (5 predictions):
    - Android.
    - iOS.
    - Windows.
    - Others.
    - Unclassified.
- Number of users depending on the type of device of the 4 RNCs (3 predictions):
    - MBB.
    - Smartphone.
    - Tablet.

There is a total of 18 different predictions to be obtained as a results of the model.

In certain cases, due to lack of data from some months, it was necessary to perform a linear interpolation, allowing to build a new set of data from a discrete set of previously known data points.

3.2.3 Fitting Models

For each of the data sets for which it is intended to obtain a prediction, it is necessary to obtain a fitting model that fits the data. The main structure of this step is provided in Figure 3.6.

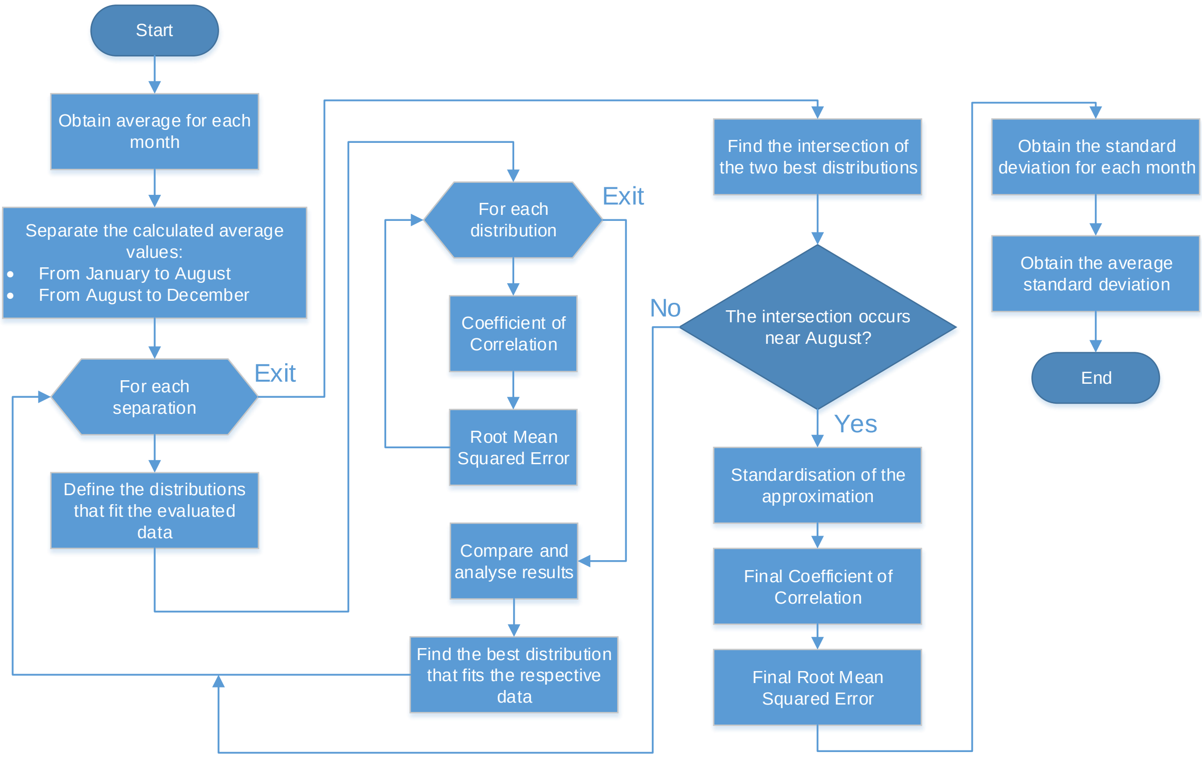

Figure 3.6. Fitting Model process.

The average data values are obtained for each month in accordance with the amount of available information. Then, these mean values for each month are separated into two parts: in the first, one has the values related to the months between January and August, and in the second to the months between August and December. This separation is done because the initial data analysis led to the conclusion that the studied parameters always have a peak in August. Therefore, the fitting curves are better if they are divided into two branches, one for the first part of the year and the other for the second one, i.e., up to August and after August, when the highest peak occurs.

For each branch, it is necessary to find the distributions that fit the data and determine which one fits best; MATLAB was used, since it contains an algorithm (the Curve Fitting Tool) to process and evaluate the data and to ascertain, through a set of distributions, which ones fits the investigated parameters best. Since this tool provides a number of different curve fit types (linear, polynomial, exponential, logarithmic, power, rational, among others), choosing a model for a particular data set was not a straightforward task. In order to decide on a curve fit model, one needs to analyse the underlying properties associated with the data, to make a good decision on which model to use, being required to examine both graphical and numerical fit results.

In order to find the best distribution that fits the evaluated data, one can use the corresponding Root Mean Squared Error (RMSE) and Coefficient of Determination (CD). The RMSE is frequently used for calculating the error between a predicted value and the actual one, giving the standard deviation of the model prediction error, therefore, a smaller value indicates better model performance; when the data are normalised from 0 to 1, the RMSE can only take values in between 0 and 1, so a score of 0 indicates a perfect fit, while a score of 1 indicates a very poor fit. The expression for RMSE is [EvsK10]:

$$\sqrt{\overline{\varepsilon^2}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{Y_i} - Y_i\right)^2} \quad (3.1)$$

where:

- $\sqrt{\overline{\varepsilon^2}}$: Root Mean Squared Error.
- N: Number of samples.
- $\widehat{Y_i}$: Model (fitted) values.
- Y_i: Observed (true) values.

The Coefficient of Determination measures how successful the fit is, taking the variation of the data, where a result of 1 means total correlation and 0 represents no correlation between the variables. The expression is given by [EvsK10]:

$$R^2 = \frac{\sum_{i=1}^{N}(\widehat{y_i} - \bar{y})}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (3.2)$$

where:

- R^2: Coefficient of Determination.
- y_i: ith observation.
- $\bar{y}$: Average of the observations of the variable.
- $\widehat{y_i}$: Estimated or predicted value of y_i.
- N: Number of samples.

When the best distributions that are suited to the two parts of the year are obtained, the instant when the intersection of the two distributions occurs is determined, being confirmed that it always occurs in August. Consequently, it is possible to check the quality of the approximation, since if this intersection does not occur near August, the approximation is not adequate and one has to look again at the different distributions for each branch.

Finally, a normalisation of the approximation for a whole year, defined by the two selected branches, is made in order to ensure the confidentiality of data, and the CD and RMSE values for the normalised approximation are calculated. This normalisation consists of placing the approximation between 0 and 1, using the following expression [Mato07]:

$$(y)_N = \frac{y - z_{min}}{z_{max} - z_{min}} \quad (3.3)$$

where:

- z_{min}: Minimum value of the approximation.
- z_{max}: Maximum value of the approximation.
- y: Expression of each branch.
- $(y)_N$: New normalised expression of the branch.

In this way, it is possible to get the best fitting model that suits the data. In order to display the graphics of these approximations along with the mean values and a confidence interval, it is necessary to first proceed with the normalisation of the initial data and of the mean values initially calculated from the actual data according to the normalisation of the approximation, using:

$$(x_i)_N = \frac{x_i - z_{min}}{z_{max} - z_{min}} \tag{3.4}$$

where:

- z_{min}: Minimum value of the approximation.
- z_{max}: Maximum value of the approximation.
- x_i: Mean value or one of the given values for the *i*th month.
- $(x_i)_N$: Mean value or one of the given values for the *i*th month already normalised.

Furthermore, in order to get the confidence interval, the standard deviations of the initial data already normalised are obtained for each month. Then, the calculation of the average of these standard deviations (σ_{avg}) is done to obtain the value that is used for the confidence interval (average ± average of the standard deviation). A confidence interval provides a range of values for the parameter of interest, indicating the accuracy of the approximations.

Average and standard deviation metrics are crucial elements that support subsequent analyses, because they are essential to calculate a confidence interval. In order to ensure the statistical relevance of results for each analysis, the average and the standard deviation for each scenario and for each parameter are obtained using:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} \mu_n \tag{3.5}$$

where:

- N: Number of samples.
- μ_n: *n*th sample.
- μ: Average obtained from the samples.

and

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \sigma_n^2} \tag{3.6}$$

where:

- N: Number of samples.
- σ_n: *n*th sample.
- σ: Standard deviation obtained from the samples.

This model is then used in the next step where it is intended to predict the evolution parameters.

3.2.4 Prediction Models

After determining the best fitting model for a parameter, one can then proceed to its forecast/prediction. The main structure of this step is shown in Figure 3.7.

Initially, a normalisation of the data for the whole period duration (two or three years) is done, by:

$$(y_j)_N = \frac{y_j}{y_{1st\ Aug}} \tag{3.7}$$

where:

- y_j: Value that the parameter takes in the jth month.
- $y_{1st\ Aug}$: Value that the parameter takes on the 1st August available (where the first big peak occurs).
- $(y_j)_N$: Value of the parameter in the jth month already normalised.

Then a linear regression of the data over the years under analysis is carried out, leading to the growth line, and a linear regression only for the peaks occurring in August is also carried out, in order to obtain the peaks growth line. In some cases, there can be two or three peaks (according to the number of years): in cases considering all RNCs there are three peaks, except those that study each particular operating system, which only has two peaks; in cases considering each individual RNC there are also 3 peaks, except those that consider the RNC2 and RNC4, with only two peaks.

Initially, to obtain a forecast/prediction for the following two years (2015/2016), a pure and simple extension of the model is made, using the fitting model obtained earlier. For this, the average value of the approximation is determined by collecting approximation curve values at various time points. From this value the two occurrences of that value in the approximation are determined, thus obtaining two instants of time.

With this, the model extension initially consists of extending the two lines of growth, being able to maintain, increase or decrease this growth in each year. Then, it is imposed, in each year, that the approximation curves intercept the growth line of data at the instants calculated, and coincide, in the instant already determined (near August), with the growth line of peaks.

In this way it is possible to obtain the blind application of the fitting model, where two discontinuities appear in the transition to the model each year (from December to January) that no one is expecting it to happen in reality. Thus, the method based on the average value of the approach is not meant to predict the evolution of the endpoint being investigated, and therefore it is necessary to do a rectification of this prediction model. This amendment consists of changing the approximation curve relative to the first part of each year, proceeding to the change of the impositions previously made.

In the first year of the prediction, it is imposed that this curve has to start in the data corresponding to the last month of the previous year (2014) and that in the end it has to intersect the line of peaks, and in the second year, it is established that the curve has to start in the value that the model takes at the end of 2015 and has to end intersecting the line of peaks again.

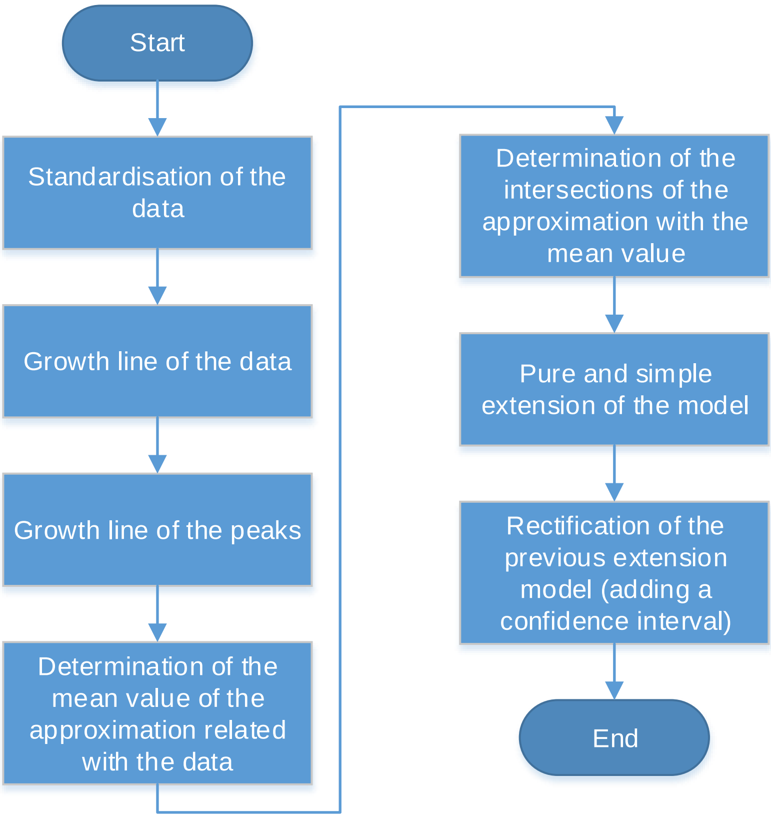

Figure 3.7. Prediction Models process.

In conclusion, the most important in this study is to have a model with a versatility to introduce new variables or to change some, because they allow studying the various scenarios. The approximation models are always the same in terms of the model output, but the equations are different and the value of the parameters vary.

3.3 Data Traffic Models of Users' Behaviour

For the data traffic models, some of certain human behaviours in some monthly traffic variation tendencies are studied in order to developed models for these variations. The chosen scenarios are the tendency of the traffic access to the Internet and the sending Short Message Service (SMS) in relation to banking. The occupation of rooms at the hotels in Algarve is also studied, as well as the water consumption in that region.

Internet access from mobile devices in Portugal continues to grow apace, and it is shown in Figure 3.8 that Internet data traffic has increased over the years, although in April 2013 there was a greater restraint in the use of mobile data.

August 2014 was the moment when the highest peak occurred compared to 2013. This traffic growth trend is associated with an increased number of users and with the number and growth of connections and devices. Consequently, mobile operators are working to continuously expand their product portfolio by delivering new products and services to the consumer, and also by expanding their network. Mobile data continues to drive rapid traffic growth for mobile operators with all regions showing impressive data volume growth rates as more and more people connect to the Internet via mobile.

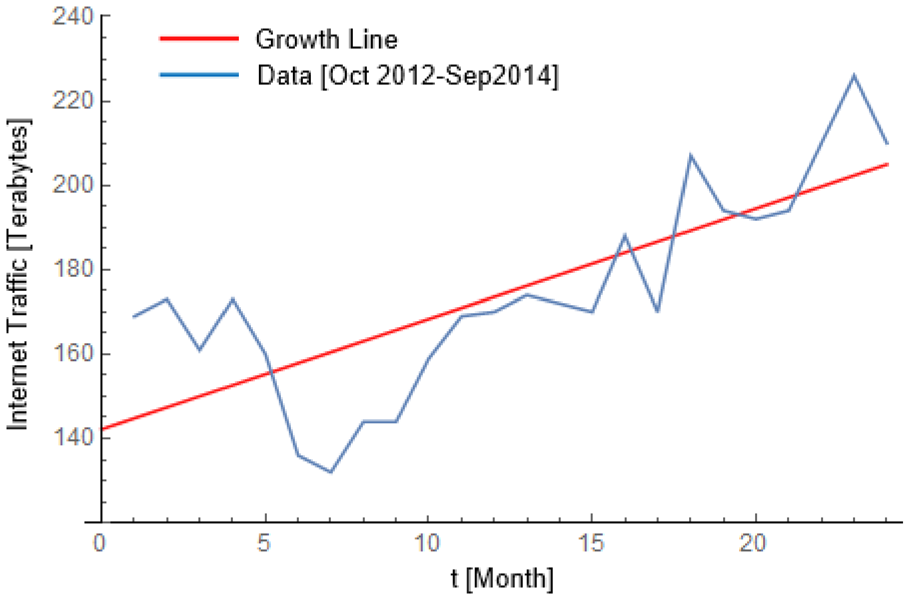

Figure 3.8. Mobile Operator Internet Traffic per month (adapted from [ANAC14]).

It is not possible to develop a model for the annual growth, in a consistent way, because years have distinct behaviours, making it difficult to predict what the behaviour for the next year will be. But it can be assumed that growth will continue to happen, and one can establish trends under certain assumptions.

The next chosen scenario, shown in Figure 3.9, is the Society of Worldwide Interbank Financial Telecommunication (SWIFT) FIN traffic evolution, which only has data from 2 years (2013 and 2014).

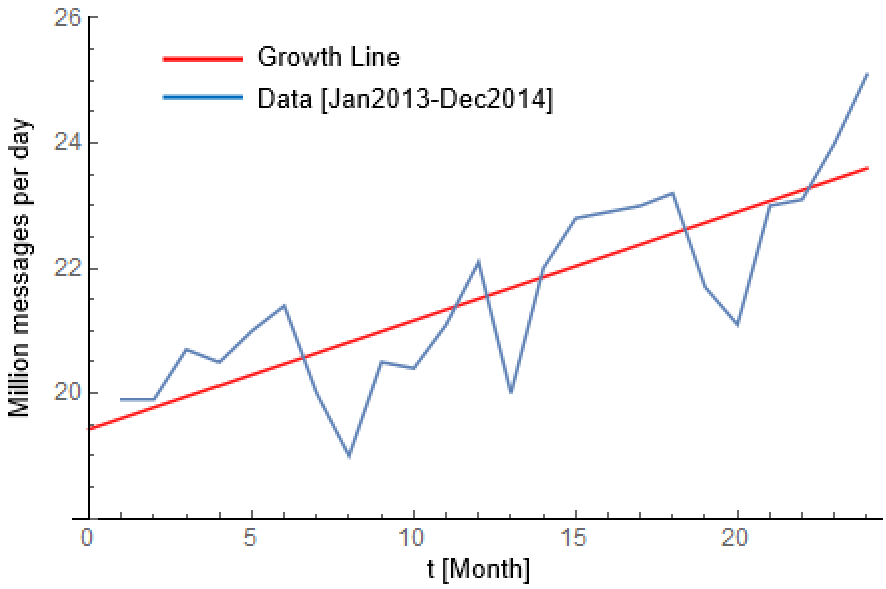

Figure 3.9. Tendency of the growth of messages per month (extracted from [SWIF15]).

As one can see, December 2014 stands as the month with most sending messages (25.1 million). As the data from 2013 is very similar to the one from 2014, the average of the two years is calculated, and then the best fit is found. This study is done to see certain trends that occur during the year, in a cyclical way, in order to then be able to develop a model.

The model that gives a good approximation to the data is the Multi Linear Traffic Model (MLTM), which consists of linear functions, as seen in the following expression, the values of the parameters being in Table 3.1:

$$f_{mltm}(t_{shift}) = \begin{cases} m_{mltm1} \times t_{shift} + c_{mltm1} & 0 \leq t_{shift} \leq t_{mltm1} \\ m_{mltm2} \times t_{shift} + c_{mltm2} & t_{mltm1} \leq t_{shift} \leq t_{mltm2} \\ m_{mltm3} \times t_{shift} + c_{mltm3} & t_{mltm2} \leq t_{shift} \leq 12 \end{cases} \tag{3.8}$$

with

$$f_{mltm}(t_{mltm1}) = m_{mltm1} \times t_{mltm1} + c_{mltm1} = m_{mltm2} \times t_{mltm1} + c_{mltm2} \tag{3.9}$$

and

$$f_{mltm}(t_{mltm2}) = m_{mltm2} \times t_{mltm2} + c_{mltm2} = m_{mltm3} \times t_{mltm2} + c_{mltm3} \tag{3.10}$$

where:

- f_{mltm}: Multi Linear Traffic Model function.
- m_{mltm1}: First linear function slope.
- c_{mltm1}: First linear function constant value.
- t_{mltm1}: First breakpoint shifted month.
- m_{mltm2}: Second linear function slope.
- c_{mltm2}: Second linear function constant value.
- t_{mltm2}: Second breakpoint shifted month.
- m_{mltm3}: Third linear function slope.
- c_{mltm3}: Third linear function constant value.

Table 3.1. Multi Linear Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**		
Left	Linear	[0; 5.79]	m_{mltm1}	c_{mltm1}	t_{mltm1} = 5.79
			0.120	0	
Middle	Linear	[5.79; 7.78]	m_{mltm2}	c_{mltm2}	
			-0.313	2.51	t_{mltm2} = 7.78
Right	Linear	[7.78; 12]	m_{mltm3}	c_{mltm3}	
			0.220	-1.64	
$R^2 = 0.953$		$\sqrt{\varepsilon^2} = 0.0848$		$\sigma_{avg} = 0.416$	

The duration/extension chosen to the interval of the model is explained in Chapter 4. One can see in Figure 3.10 that there is a good approximation between the data and the model. A confidence interval is also calculated to indicate the reliability of the estimation.

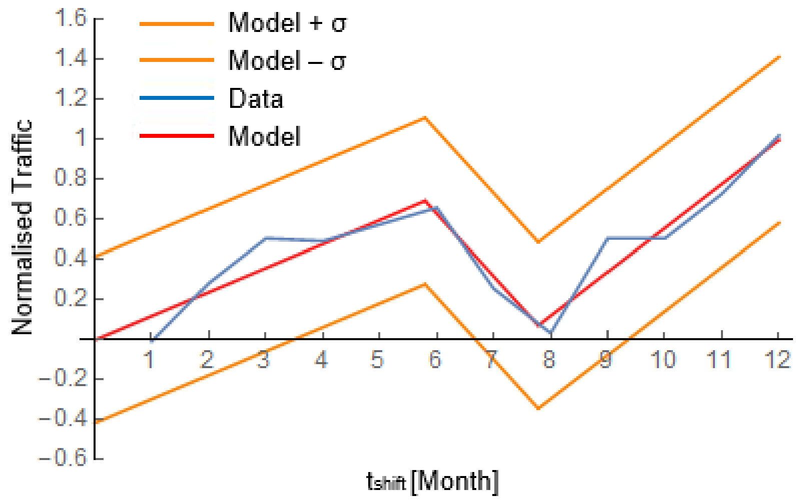

Figure 3.10. Multi Linear Model, with the respective data and a confidence interval.

The case of the scenario of the Room occupancy rate in Algarve is shown in Figure 3.11.

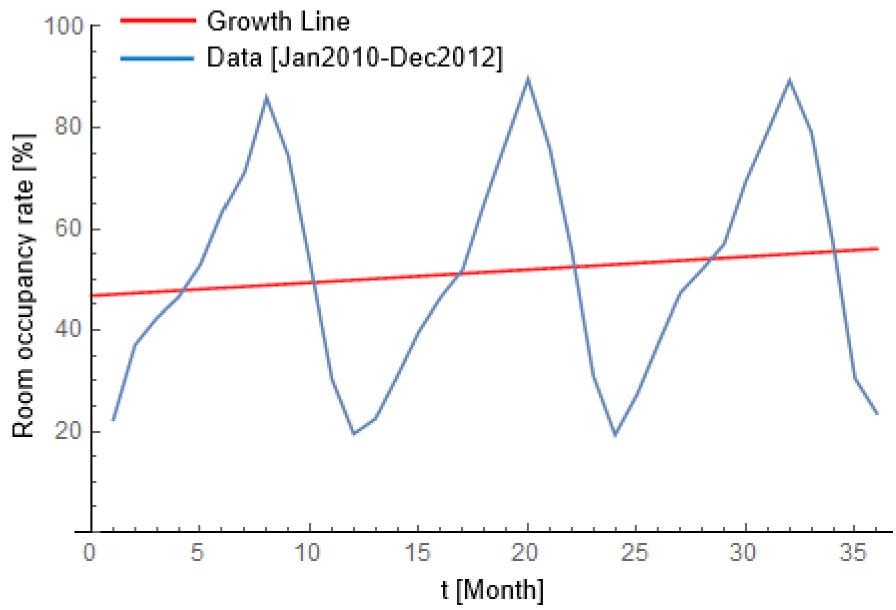

Figure 3.11. Room occupancy rate of National + Foreign in Algarve per month (extracted from [TuPo15]).

It is important to refer some key aspects of the analysis to room occupancy rates of ART Algarve. For 2010, the figure presents a strong seasonality, with room occupancy rates varying between a minimum of 19.5% in December and a maximum of 86% in August. It is also worthwhile noting the peak demand from June to September, the time of the year when room occupancy rates are above 63%. For 2011, a similar seasonality is also presented, with room occupancy rates varying between a minimum of 19.3% in December and a maximum of 89.6% in August. The peak demand occurs also from June to September, the time of the year when the room occupancy rates are above 65%. For 2012, the same trend is observed, with room occupancy rates varying between a minimum of 23.6% in December and a maximum of 89.4% in August, and peak demand occurring from June to September, with room occupancy rates above 69%.

Once again, the average of the 3 years was calculated in order to give a better understanding to develop the model. The developed model is the Room Traffic Model (RTM), and uses two linear functions, with the following expression:

$$f_{rtm}(t_{shift}) = \begin{cases} m_{rtm1} \times t_{shift} + c_{rtm1} & 0 \leq t_{shift} \leq t_{rtm} \\ m_{rtm2} \times t_{shift} + c_{rtm2} & t_{rtm} < t_{shift} \leq 12 \end{cases} \quad (3.11)$$

with

$$f_{rtm}(t_{rtm}) = m_{rtm1} \times t_{rtm} + c_{rtm1} = m_{rtm2} \times t_{rtm} + c_{rtm2} \quad (3.12)$$

where:

- f_{rtm}: Room Traffic Model function.
- m_{rtm1}: First linear function slope.
- c_{rtm1}: First linear function offset.
- t_{rtm}: Breakpoint shifted month.
- m_{rtm2}: Second linear function slope.
- c_{rtm2}: Second linear function offset.

The parameters for the RTM are also calculated as shown in Table 3.2.

Table 3.2. Room Traffic Model parameters.

Section	Function	Interval [Month]	Parameters		
Left	Linear	[0; 8.21]	m_{rtm1}	c_{rtm1}	t_{rtm} = 8.21
			0.122	0	
Right	Linear	[8.21; 12]	m_{rtm2}	c_{rtm2}	
			-0.254	3.08	
R^2 = 0.990		$\sqrt{\overline{\varepsilon^2}}$ = 0.0416		σ_{avg} = 0.0382	

The best approximation is shown in Figure 3.12, from which one can conclude the existence of a good behaviour for the model.

In the last scenario, shown in Figure 3.13, the evolution of the volume of water supplied in 2010-2012 by the municipality of Lagoa and by the company Águas do Algarve is studied.

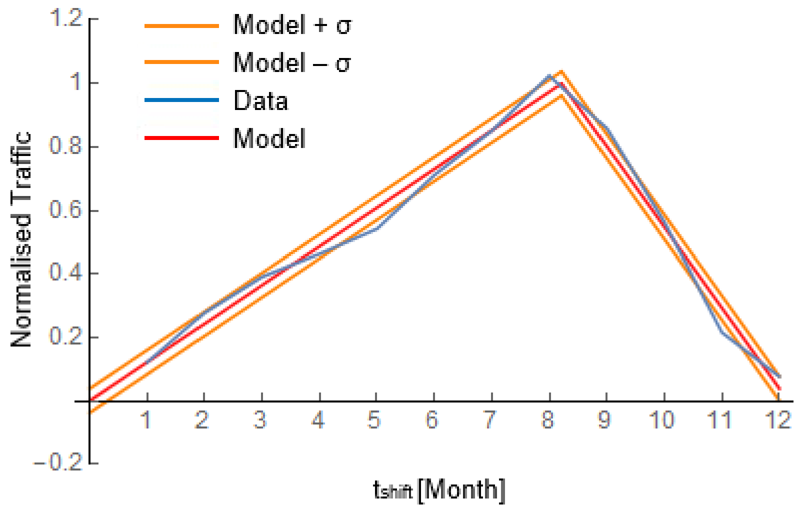

Figure 3.12. Room Traffic Model, with the respective data and a confidence interval.

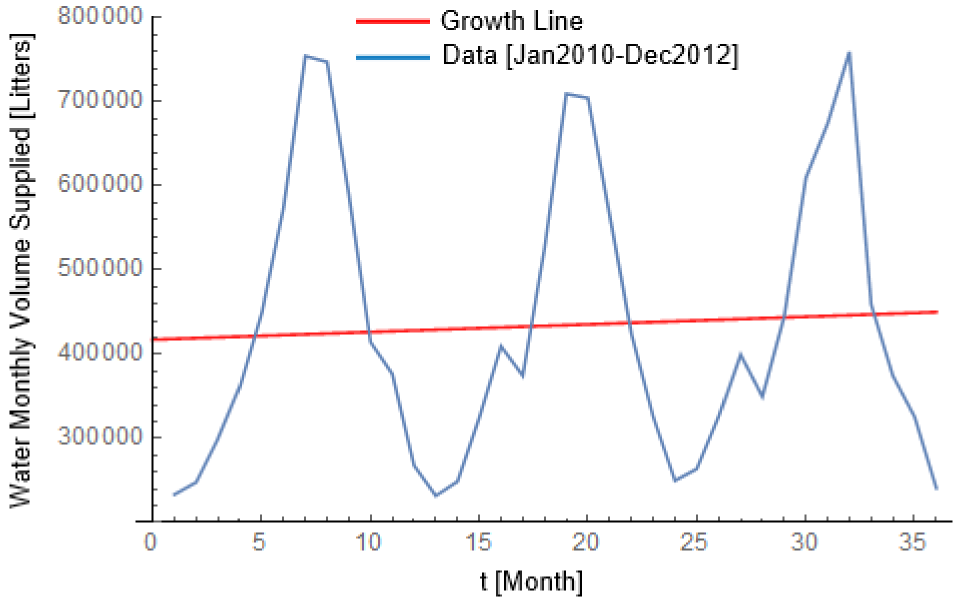

Figure 3.13. Evolution of volume of water supplied by Águas do Algarve per month (adapted from [CMLa15]).

For each year, the largest consumption of water occurs in the summer season, due to tourism. For 2010 the consumption varies with a minimum of 233 500 litters in January and a maximum in July with 755 000 litters. For 2011 the rates of consumption of water vary with a minimum in January of 232 000 litters, being the lowest value of the 3 years, and with a maximum in July of 710 000 litters, which means that there was lower water consumption in this summer. 2012 presents a different scenario comparing to the previous years, because water consumption was minimum in December with a consumption of 241 000 litters, while the maximum was no longer in July but in August, with 760 000 litters.

The developed model is the Water Traffic Model (WTM), with power and exponential functions being used for each section. This study is important because the water consumption in Algarve is increasing year by year and this variation in water consumption reflects on most tourist towns of Algarve, such as Albufeira, Portimão and Loulé, whose monthly bills of water can reach one million euros during the summer.

The expression for the WTM is given by:

$$f_{wtm}(t_{shift}) = \begin{cases} a_{wtm} \times t_{shift}^{b_{wtm}} + c_{wtm1} & 0 \leq t_{shift} \leq t_{wtm} \\ d_{wtm} \times e^{t_{shift} \times g_{wtm}} + c_{wtm2} & t_{wtm} < t_{shift} \leq 12 \end{cases} \tag{3.13}$$

with

$$f_{wtm}(t_{wtm}) = a_{wtm} \times t_{wtm}^{b_{wtm}} + c_{wtm1} = d_{wtm} \times e^{t_{wtm} \times g_{wtm}} + c_{wtm2} \tag{3.14}$$

where:

- f_{wtm}: Water Traffic Model function.
- a_{wtm}: Power initial value.
- b_{wtm}: Power decay factor.
- c_{wtm1}: First linear function offset.
- t_{wtm}: Breakpoint shifted month.
- d_{wtm}: Exponential initial value.
- g_{wtm}: Exponential decay factor.
- c_{wtm2}: Second linear function offset.

For the Water Traffic Model, the same parameterisation is made, and the values obtained for this model are shown in Table 3.3.

Table 3.3. Water Traffic Model parameters.

<table>
<tr><th>Section</th><th>Function</th><th>Interval [Month]</th><th colspan="4">Parameters</th></tr>
<tr><td rowspan="2">Left</td><td rowspan="2">Power</td><td rowspan="2">[0; 7.93]</td><td>a_{wtm}</td><td>b_{wtm}</td><td>c_{wtm1}</td><td rowspan="4">t_{wtm} = 7.93</td></tr>
<tr><td>0.0202</td><td>1.88</td><td>0</td></tr>
<tr><td rowspan="2">Right</td><td rowspan="2">Exponential</td><td rowspan="2">[7.93; 12]</td><td>d_{wtm}</td><td>g_{wtm}</td><td>c_{wtm2}</td></tr>
<tr><td>97.1</td><td>-0.577</td><td>0</td></tr>
<tr><td colspan="2">R^2 = 0.990</td><td colspan="3">$\sqrt{\overline{\varepsilon^2}}$ = 0.0483</td><td colspan="2">σ_{avg} = 0.0700</td></tr>
</table>

As it has been referred, for the left section the best fitting curve is the power function due to the variation of the recorded values, and for the right section the best fit is the exponential one, as shown in Figure 3.14. In general, the proximity between model and data is good, which is clearly seen with the obtained values for the fitting requirements.

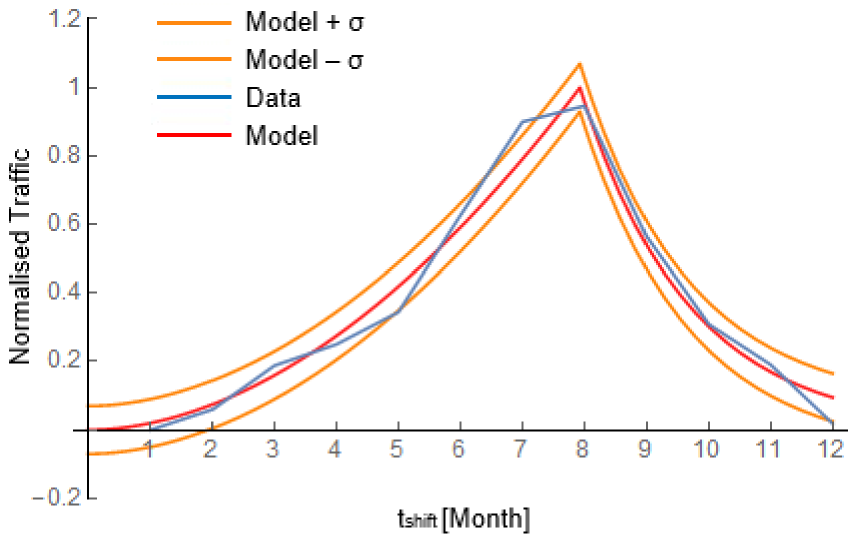

Figure 3.14. Water Traffic Model, with the respective data and a confidence interval.

3.4 Online Social Networks Model

Online Social Networks (OSNs) are becoming extremely popular. More than two-thirds of the global online population visit and participate in social networks and blogs [FTMV12]. In fact, social networking and blogging account for nearly 41% of all time spent on the Internet [Glob15]. These statistics suggest that OSNs have become a fundamental part of the global online experience. Through OSNs, users connect with each other, share and find content, and disseminate information. Numerous sites provide social links, for example, networks of professionals and contacts and networks for sharing contents. Understanding how users behave when they connect to these sites is important for a number of reasons. The studies of users' behaviours allow the performance of existing systems to be evaluated, and lead to better site design and advertisement placement policies. Accurate models of user behaviour in OSNs are crucial in social studies as well as in viral marketing.

Understanding how users behave when they connect to social networking sites creates opportunities for better interface design, richer studies of social interactions, and improved design of content distribution systems [FTMV12].

In mobile networks, social networks are experiencing an emergence of popularity similar to the one seen in fixed networks, being currently among the fastest growing applications. This growth can be explained by the easy access to these services through Smartphones, and also by the ability to access these services at any time. In the future, it is expected that these applications will continue to boost the consumption of mobile data, and will promote a development in mobile applications that provide these services. For instance, viral marketers might want to exploit models of user interaction in order to spread their content or promotions quickly and widely. Understanding how the workload of social networks is re-shaping the Internet traffic is valuable in designing the next-generation networks infrastructure.

The impact of applications in society has been exponential and varied, following the technological evolution of mobile devices. Nowadays, written messages are as used as phone calls; some people say it depends on the segment, but the paradigm has changed, and now one can send and receive text messages, voice and video without spending money. With the Internet present in Smartphones and Tablets, it is enough to have an application that provides all of these contacts at zero cost.

OTT applications include messaging services to third-party liability that do not use the means of delivering SMS of operators. These services resort to delivering messages via data connection, and in this way there is no charge for the service. There is no specific number, but it is considered that each user with these services sends an average of 32.6 messages a day, while traditional SMS users send an average of 5 messages per day [INFO15].

There are many applications, but in this thesis only a few of them are studied to understand the trend of recent years: Facebook, Twitter, WhatsApp, Skype, KIK Messenger and Viber. Facebook, as it is well-known, lets users sending messages to other users for free. It provides a very quick way to communicate, also enabling video calls using Smartphones and Tablets. A data traffic model called Facebook Traffic Model (FTM) was developed:

$$f_{ftm}(t_{shift}) = m_{ftm} \times t_{shift} + c_{ftm} \qquad 0 < t_{shift} \leq t_{ftm} \tag{3.15}$$

with

$$f_{ftm}(t_{ftm}) = m_{ftm} \times t_{ftm} + c_{ftm} \tag{3.16}$$

where:

- f_{ftm}: Facebook Traffic Model function.
- m_{ftm}: Linear function slope.
- c_{ftm}: Linear function constant value.
- t_{ftm}: Duration time (months).

A parameterisation was made for FTM and the values obtained are shown in Table 3.4.

Table 3.4. Facebook Traffic Model parameters.

Section	Function	Interval [Month]	Parameters		
1	Linear	[3; 66]	m_{ftm}	c_{ftm}	t_{ftm}
			18.2	163	66
$R^2 = 0.997$			$\sqrt{\overline{\varepsilon^2}} = 0.0311$		

As shown in Figure 3.15, it is possible to see the tendency (Jan2009 – Jun2014) of the FTM, the proximity between model and data being very good. A confidence interval is not created because the growth has been steady and the trend is the continuation of this steady growth.

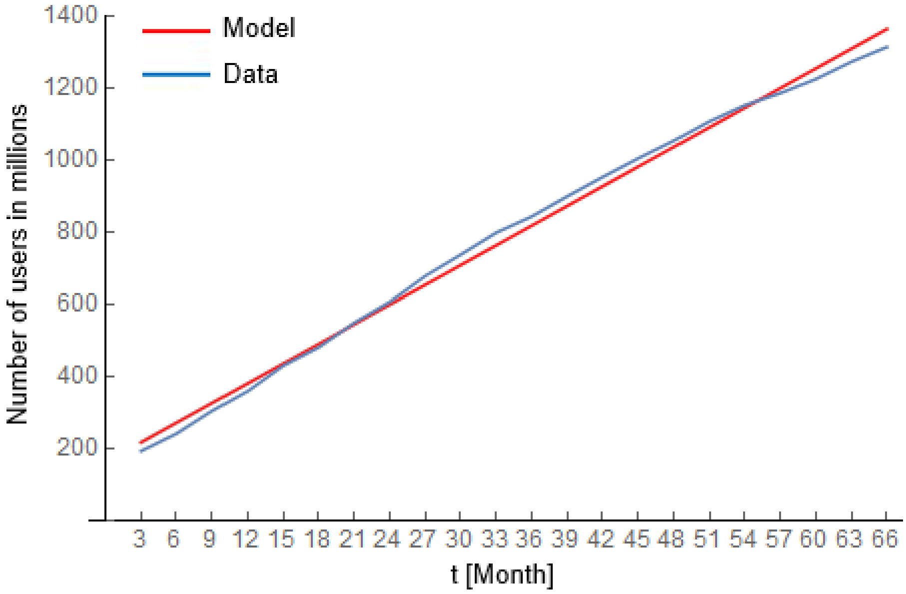

Figure 3.15. Facebook Traffic Model with the average profile (adapted from [Stat15]).

The second application studied is Twitter, which allows users to send and read personal updates from other contacts (texts up to 140 characters, known as "tweets") via the own Website. For this, a data traffic model called Twitter Traffic Model (TwTM) was developed:

$$f_{twtm}(t_{shift}) = m_{twtm} \times t_{shift} + c_{twtm} \qquad 0 < t_{shift} \leq t_{twtm} \tag{3.17}$$

with

$$f_{ttm}(t_{twtm}) = m_{twtm} \times t_{twtm} + c_{twtm} \tag{3.18}$$

where:

- f_{twtm}: Twitter Traffic Model function.
- m_{twtm}: Linear function slope.
- c_{twtm}: Linear function constant value.
- t_{twtm}: Duration time (months).

The TwTM is characterised by the parameters shown in Table 3.5.

Table 3.5. Twitter Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**		
1	Linear	[3; 60]	m_{twtm}	c_{twtm}	t_{twtm}
			4.92	3.93	60
$R^2 = 0.997$			$\sqrt{\overline{\varepsilon^2}} = 0.0392$		

In Figure 3.16, the tendency (Jan2010 – Mar2014) of Twitter usage is shown. As one can see, there is a good behaviour of data and model in terms of proximity.

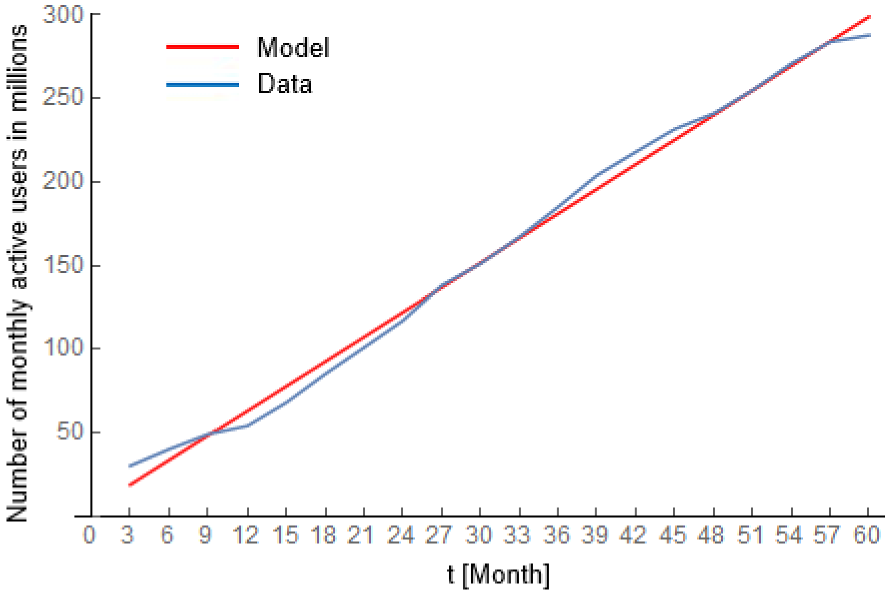

Figure 3.16. Twitter Traffic Model with the average profile (adapted from [Stat15]).

Another application studied was WhatsApp, which has a huge success and allows the user to exchange messages, pictures and videos with other users without paying the cost of an SMS. The WhatsApp Traffic Model (WATM) as developed, having the following expression:

$$f_{watm}(t_{shift}) = m_{watm} \times t_{shift} + c_{watm} \qquad 0 < t_{shift} \leq t_{watm} \tag{3.19}$$

with

$$f_{watm}(t_{watm}) = m_{watm} \times t_{watm} + c_{watm} \tag{3.20}$$

where:

- f_{watm}: WhatsApp Traffic Model function.
- m_{watm}: Linear function slope.
- c_{watm}: Linear function constant value.
- t_{watm}: Duration time (months).

The parameters for the WhatsApp Traffic Model are in Table 3.6.

Table 3.6. WhatsApp Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**		
1	Linear	[1; 24]	m_{watm}	c_{watm}	t_{watm}
			24.4	108	24
R^2 = 0.999			$\sqrt{\overline{\varepsilon^2}}$ = 9.72×10^{-3}		

Figure 3.17 shows the impact that this application had during 2 years (Jan2013 – Dec2014). A good approximation is clearly visible between model and data.

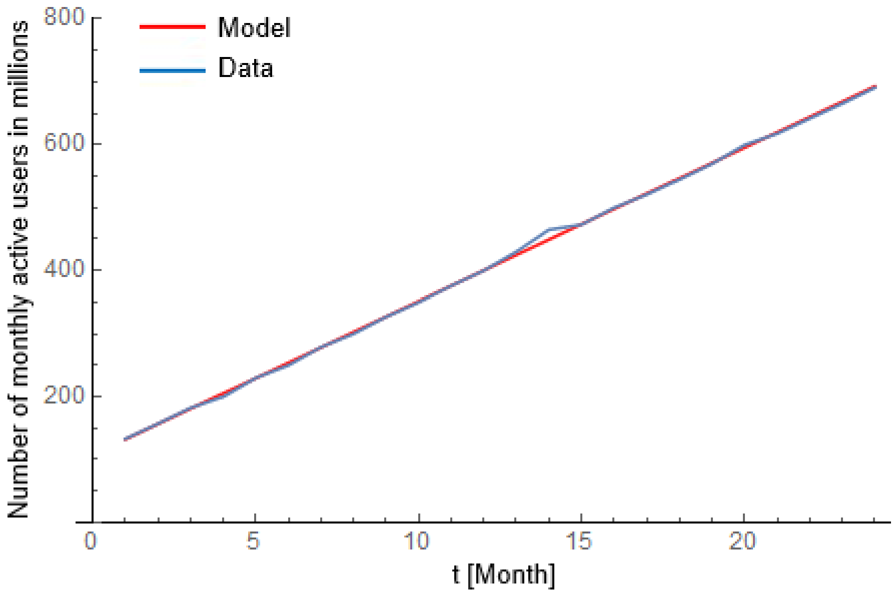

Figure 3.17. WhatsApp Traffic Model with the average profile (adapted from [Stat15]).

For the next applications being studied, such as Skype, KIK Messenger and Viber, it is not possible to develop models due to the huge discontinuities and the highs and lows that these applications have on the level of usage. Figure 3.18 shows the behaviour (Sep2013 – Sep2014) of these applications and a perception of their usage.

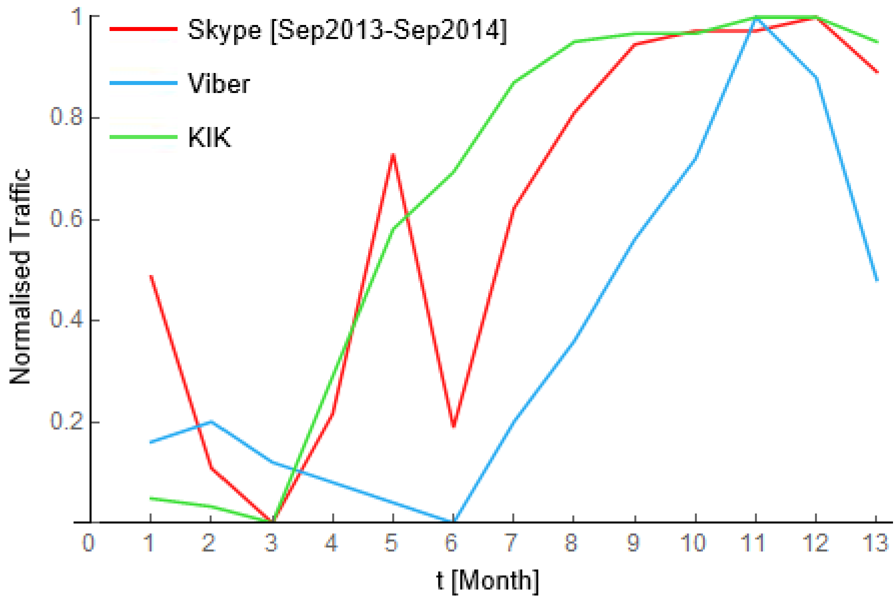

Figure 3.18. Utilisation of the Skype, Viber and KIK Messenger applications during the year (adapted from [Coms15]).

As it is known, Skype is one of the most widely spoken tools of communication nowadays. Skype was pioneer in VoIP service and brought many ideas to the actual communications market. One can talk via video, messages and even leave a videotaped message for those who are offline. The KIK Messenger application is a new generation of instant messaging for Smartphones and Tablets and is faster and more reliable in this market segment application. The Viber application allows the user to make calls and send messages free of charge to users who have the application installed. This application was pioneer when it comes to allowing the use of infrastructures to send text and voice messages without spending money. The sound quality and the ability to be used in multiplatform gave Viber recognition to be one of the most popular applications worldwide.

KIK Messenger shows a continuous evolution, allowing to conclude that the users are joining more and more to this application.

The establishment of a model for these applications would require more data over more years. However, given that they are recent ones, it is not possible yet to get such a model.

Chapter 4

Results analysis

This chapter provides an overview of the considered scenario along with the associated results and respective analysis. Firstly, a brief description of the parameters and environments of the fitting models used is done, and afterwards, the prediction results are presented for the network usage evolution.

4.1 Scenario

The national mobile communications market has been able to maintain a certain degree of competitive dynamism, due to innovative solutions in terms of promotions and campaigns that allow price decrease and increase in the number of the available devices, the main beneficiary being the consumer. With new technological developments, mobile penetration in Portugal has increased, consumers are more demanding, and the result is that mobile operators are constantly offering new products and services.

The scenario used for this work is the access network of UMTS, installed in south of Portugal, considering the sites that are connected to 4 specific RNCs (designated as RNC1, RNC2, RNC3 and RNC4). The total number of users is less than 950 000, and the collected data consists of monthly traffic values from January 2012 to December 2014 for the RNC1 and RNC3, and in some cases only from April/June 2013 to December 2014 for the RNC2 and RNC4. It is important to mention that in some cases a linear interpolation was taken, particularly in the early months of the year, due of lack of data. These collected data were used to study the behaviour as well as the effect of the migratory season of population during holidays in those regions.

In the data files, the RNCs, the categories, such as the number of users and the data traffic according to the brand, the device type (MBB, Smartphone, Tablet) and the operating system (Android, iOS, Windows, Others, Unclassified) were identified for each month of the several years.

The study of each category was done separately, and organised in order to have an idea of what has been the trend in the respective years. For each RNC, the number of users and data traffic were checked, as well as the total of users and data traffic for the 4 RNCs. An individual verification for the operating systems and the different types of devices was also done for the total users of the 4 RNCs.

The next step was to calculate the average of the samples depending on the RNC and the category, in order to obtain a single curve after the study to find the best possible approximation based on CD and RMSE (explained in Section 3.2.3). After getting this best approximation, the average of the standard deviation was calculated in order to ensure a confidence interval, to enable operators to get a sense of how much the data traffic can vary.

It should be noted that with Internet access, mobile phones are going beyond communication and entertainment, entering the realm of everyday activities. This growth is especially driven by the apps that are emerging and available to users, which are perceived as less time-consuming than browsers and less complex than applications on PCs; simply put, they provide direct access and right functionality.

Normally the data traffic used by users in the various devices, in particular Tablets, Smartphones and Computers, is normal in the daily file, with not very pronounced peaks. However, the highest peak of data traffic usually takes place in August, because it is when the culmination of several effects occurs.

At first, there is the commercial effect of the campaigns of Smartphones and Tablets, which generate an interest to buy among consumers. Then, there are two times of the year when people are more consumerist: Christmas (in December) and the vacations (in August), because that is when usually users leave the comfort zone and go to different areas of the country. It is also worth noting that people tend to use more Tablets and Smartphones instead of using Television or Computer when they are in vacations. Otherwise, this devices are used differently to be more convenient.

With the rollout of 3G networks, few people could have predicted how rapidly and eagerly users would embrace MBB and how it would lead to unprecedented growth in data traffic. The rapid uptake of Smartphones, Tablets and MBB has significantly increased the complexity of networks, making them more difficult to manage. Knowing this and having calculated, through data analysis, the best approaches for each category, these models are then extrapolated to do the prediction for the years 2015 and 2016, allowing one to know what the trend is, therefore being able to help Vodafone to avoid problems from the high load of the network in these regions.

4.2 Fitting Models for the RNCs' Usage

At first, the usage of the 4 RNCs is studied in order to understand the migration in south of Portugal during the year.

4.2.1 User Traffic per RNC

Figure 4.1 shows the collected data on the number of users per each RNC. These data were normalised in relation to the 1st August peak, using (3.7).

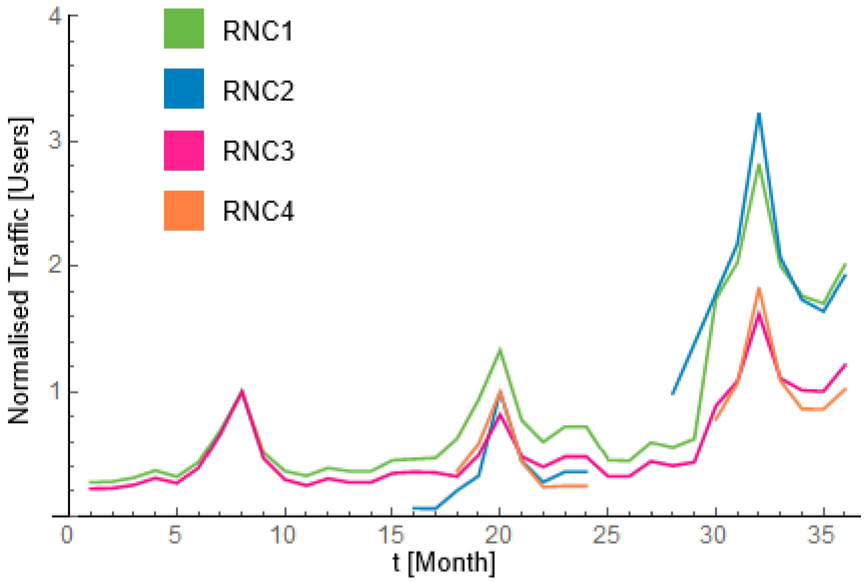

Figure 4.1. Behaviour of the number of users for each RNC from 2012 to 2014.

For RNC1 and RNC3, the first peak of August is in 2012 and for RNC2 and RNC4 it is in 2013. Due to this normalisation, it is not possible to show which RNC has a higher or lower number of users.

The next step is the calculation of the average, i.e., the average over the two or three years, in order to find the best approximation that suits the data. In order to obtain the traffic models, one uses simple analytical functions and, in order to get a better approximation, the models are composed of piecewise sections of these functions, with the guarantee of continuity in between the distinct sections. This process is explained in Chapter 3 in the implementation of the model (Section 3.2) and is used in the remaining fitting models presented in the current chapter.

For a brief explanation of this process, RNC1 is chosen, because it is one of the RNCs for which data is complete for 3 years (2012/2013/2014) and it also has a better approximation for R^2 compared with RNC3 within the requirements.

The normalisation process preserves the year traffic load variation shape despite of the size of the RNC, while the shift minimises the error between profiles and models, as shown in Figure 4.2. Each sector profile is analysed in order to perceive tendencies on the annual evolution of the traffic load.

In Figure 4.2, the blue curve represents the average obtained for an annual variation of the data collected for the three years, shown in Figure 4.1. From this curve, two curves, shown in red, are obtained, representing the two sections of the fitting model, the left and the right one, and corresponding to the best functions that fit the data. As mentioned before, the parameters used for this choice are CD and RMSE. As one can see, these two red curved are joined into one, whose junction point occurs at the 8th month of the year, i.e., in August, which is the moment when the highest peak is supposed to occur (as explained in Chapter 3). Finally, due to the variations in the data for the three years, a confidence interval (average of data value from the collected data ± average of standard deviation from the collected data) is obtained, which corresponds to the orange curves, giving a certain probability that data will be within that range.

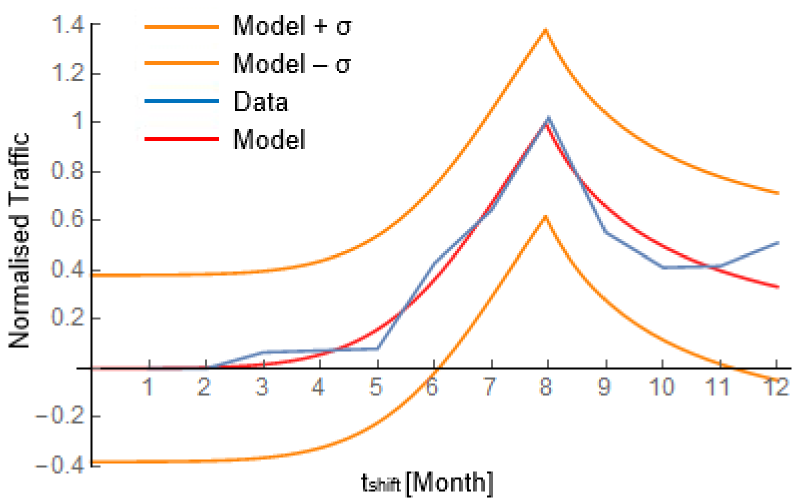

Figure 4.2. RNC1 User Traffic Model, with the respective data and a confidence interval.

The developed traffic model for the chosen RNC is designated by RNC1 User Traffic Model (URNC1); the parameters are in Table 4.1, and the expression is the following:

$$f_{urnc1}(t_{shift}) = \begin{cases} a_{urnc1} \times e^{(-((t_{shift}-b_{urnc1})/c_{urnc1})^2)} + d_{urnc1} & 0 \leq t_{shift} \leq \ t_{urnc1} \\ \dfrac{p_{urnc1}}{t_{shift} + q_{urnc1}} + g_{urnc1} & t_{urnc1} < t_{shift} \leq \ 12 \end{cases} \quad (4.1)$$

with

$$f_{urnc1}(t_{urnc1}) = a_{urnc1} \times e^{(-((t_{urnc1}-b_{urnc1})/c_{urnc1})^2)} + d_{urnc1} = \frac{p_{urnc1}}{t_{urnc1} + q_{urnc1}} + g_{urnc1} \quad (4.2)$$

where:

- f_{urnc1}: RNC1 User Traffic Model function.
- a_{urnc1}: Gaussian amplitude.
- b_{urnc1}: Month shifted peak value.
- c_{urnc1}: Gaussian deviation.
- d_{urnc1}: Gaussian offset.
- t_{urnc1}: Breakpoint shifted month.
- p_{urnc1}: Polynomial scale factor.
- q_{urnc1}: Polynomial shift.
- g_{urnc1}: Polynomial offset.

It is important to mention that an extension is made in the traffic model (from Month 0 to Month 1) in order to have an accuracy of 13 months, because in the prediction model, which is addressed in Section 4.6, it is necessary to connect the month of December from the last year with the month of January of the following one when making the transition in between years from the actual data to the prediction.

Table 4.1. RNC1 User Traffic Model parameters.

Section	Function	Interval [Month]	Parameters				
Left	Gaussian	[0; 7.95]	a_{urnc1}	b_{urnc1}	c_{urnc1}	d_{urnc1}	$t_{urnc1} = 7.95$
			1.29	9.53	3.14	-1.25×10^-4^	
Right	Rational	[7.95; 12]	p_{urnc1}	q_{urnc1}	g_{urnc1}		
			2.03	-5.92	-1.25×10^-4^		
$R^2 = 0.970$			$\sqrt{\varepsilon^2} = 0.0744$			$\sigma_{avg} = 0.381$	

Figure 4.3 shows the obtained models that explain the behaviour of the number of users of each RNC, and Table 4.2 presents the parameters of the models that show the quality of the approximations. In Annex A, the approximation models for the others RNCs are shown, where it is possible to see the functions used for each section and their expressions, as well as the values obtained.

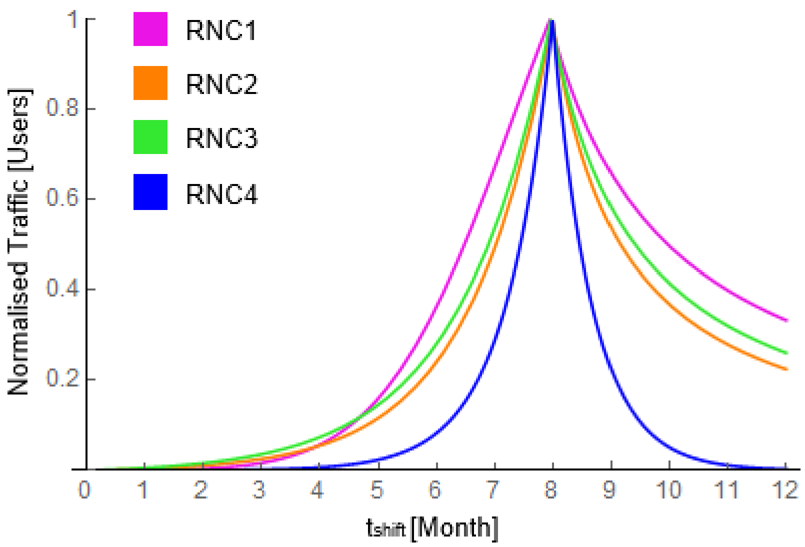

Figure 4.3. User Traffic Model of each RNC.

Table 4.2. User Traffic Model parameters for each RNC.

RNC	RNC1	RNC2	RNC3	RNC4
CD	0.970	0.965	0.967	0.990
RMSE	0.0744	0.0709	0.0726	0.0473

For the results regarding the usage of each RNC, it is observed that the left section is better fitted by the exponential function in 75% of the cases, and that the right one by the rational function in 75% of the cases too, since they offer the best CD and RMSE.

From the comparative results among the traffic models that are shown in Table 4.2, one can conclude that RNC4 is the one with the best approximation model (99% of proximity) due to its low value of RMSE and high CD, but it should be noted that this RNC has less data, which causes the aforementioned parameters to be the best in this case.

However, the CD and RMSE values are high and low enough in all cases to validate all of the developed models, as it would be expected from the graphical representation of these models and their respective average sector profiles shown in Annex A.

4.2.2 Total Users of the 4 RNCs

The procedure from the last subsection is used again, but in this case for the sum of the users of the 4 RNCs, which is shown in Figure 4.4.

Figure 4.5 presents the average values of these 3 years and the traffic model with the respective confidence interval. The parameters of the developed model are shown in Table 4.3.

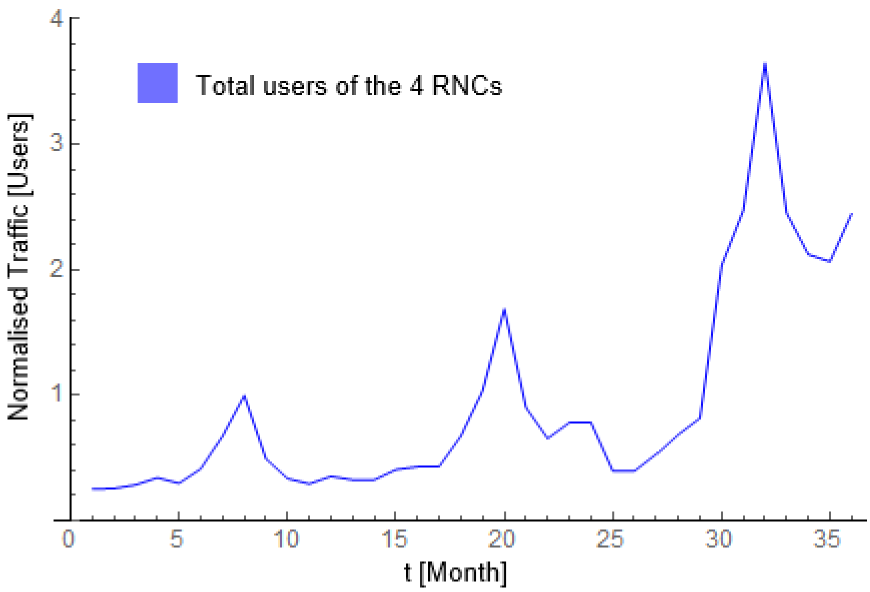

Figure 4.4. Trend of the total users of the 4 RNCs from 2012 to 2014.

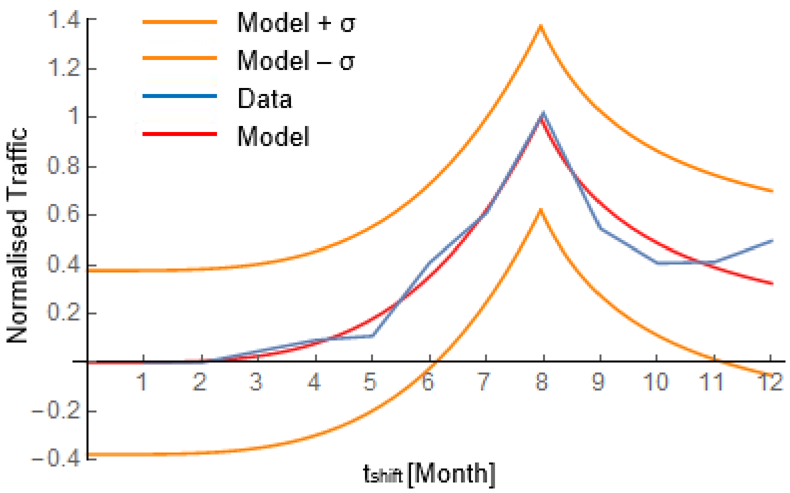

Figure 4.5. RNC Total Users Traffic Model, with the respective data and a confidence interval.

This model is called the RNC Total Users Traffic Model (RTU), and is defined by two sections, the left one with a power function adjusted to the first 8 months, and the right one with a rational function adjusted to the last 4 months. The expression is the following:

$$f_{rtu}(t_{shift}) = \begin{cases} a_{rtu} \times t_{shift}^{b_{rtu}} + c_{rtu} & 0 \leq t_{shift} \leq t_{rtu} \\ \dfrac{p_{rtu}}{t_{shift} + q_{rtu}} + d_{rtu} & t_{rtu} < t_{shift} \leq 12 \end{cases} \tag{4.3}$$

with

$$f_{rtu}(t_{rtu}) = a_{rtu} \times t_{rtu}^{b_{rtu}} + c_{rtu} = \frac{p_{rtu}}{t_{rtu} + q_{rtu}} + d_{rtu} \tag{4.4}$$

where:

- f_{rtu}: RNC Total Users Traffic Model function.
- a_{rtu}: Power initial value.
- b_{rtu}: Power decay factor.
- c_{rtu}: Linear function offset.
- t_{rtu}: Breakpoint shifted month.
- p_{rtu}: Polynomial scale factor.
- q_{rtu}: Polynomial shift.
- d_{rtu}: Polynomial offset.

Table 4.3. RNC Total Users Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**			
Left	Power	[0; 7.95]	a_{rtu}	b_{rtu}	c_{rtu}	$t_{rtu} = 7.95$
			4.64×10^{-4}	3.70	0	
Right	Rational	[7.95; 12]	p_{rtu}	q_{rtu}	d_{rtu}	
			1.94	-6.01	0	
$R^2 = 0.972$			$\sqrt{\overline{\varepsilon^2}} = 0.0694$		$\sigma_{avg} = 0.376$	

By analysing Figure 4.5, it is visible that there is a substantial growth up to August, because in the last years Portuguese people tend to travel less abroad, so there is an increased migration to the southern region of Portugal, being reflected in the huge increase of the number of users. Another growing trend occurs in the months close to December since it is known that southern Portugal has a strong tradition of celebration of New Year's Eve, receiving many visitors from the most varied origins. For this reason, the right section fails to have a better approximation, due to the fact that it has fewer samples relative to the left section and because the available distributions for this section are monotone decreasing not explaining the sudden growth that happens in October. In order to explain this growth, another separation of the piecewise functions should be done around this month. However, in a general way, there is a good proximity between the data and the approximation model.

4.3 Fitting Models for the Usage of Different Types of Devices

Prior to the Smartphone, consumers tended to confine their online Internet activities to when they were close to a Computer. Currently people are using the Internet constantly and Smartphones are not the only devices being used in this new era of constant Internet access, Tablets taking an important role in this growth too.

The constant connection to the cloud is weaving itself ever deeper into the daily routines of consumers, which are increasingly interacting with their Smartphones and Tablets and the Internet before they even get out of bed. For many people, the alarm clock is now used is in their phone, so when they wake up, it is the first thing they touch. Furthermore, if they are turning off an alarm in a Smartphone/Tablet, they also have a quick connection to the Internet in their hands [Eric11].

Figure 4.6 shows the trends for the different types of devices, normalised to the 1st August peak, which in this case corresponds to 2012. Then again, the annual average for each different type of device was calculated, followed the calculation of the best fitting model.

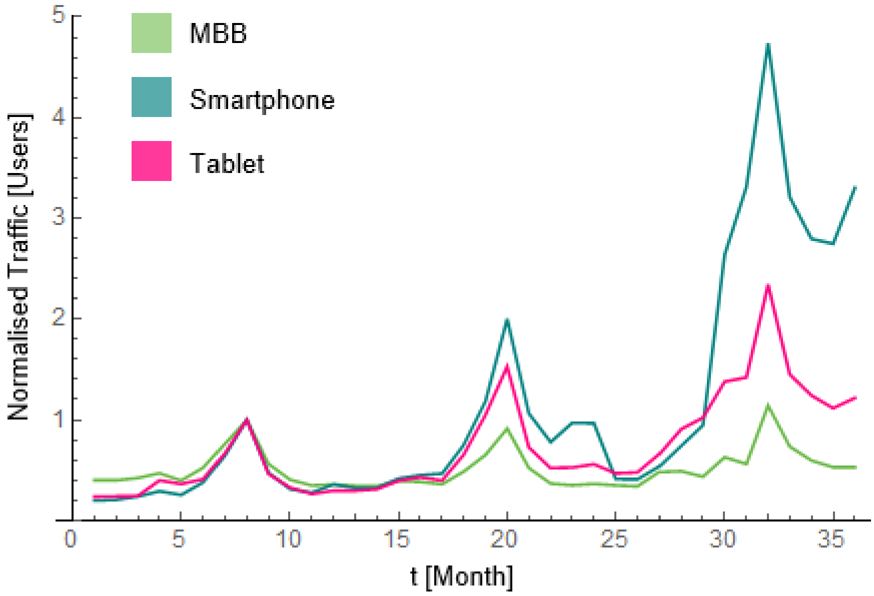

Figure 4.6. Behaviour of the number of users per type of device from 2012 to 2014.

As one can see in Figure 4.7, the largest market lies in Smartphones compared to other devices (Tablets and MBBs). This figure represents the annual average of the usage of the types of devices in the 4 RNCs.

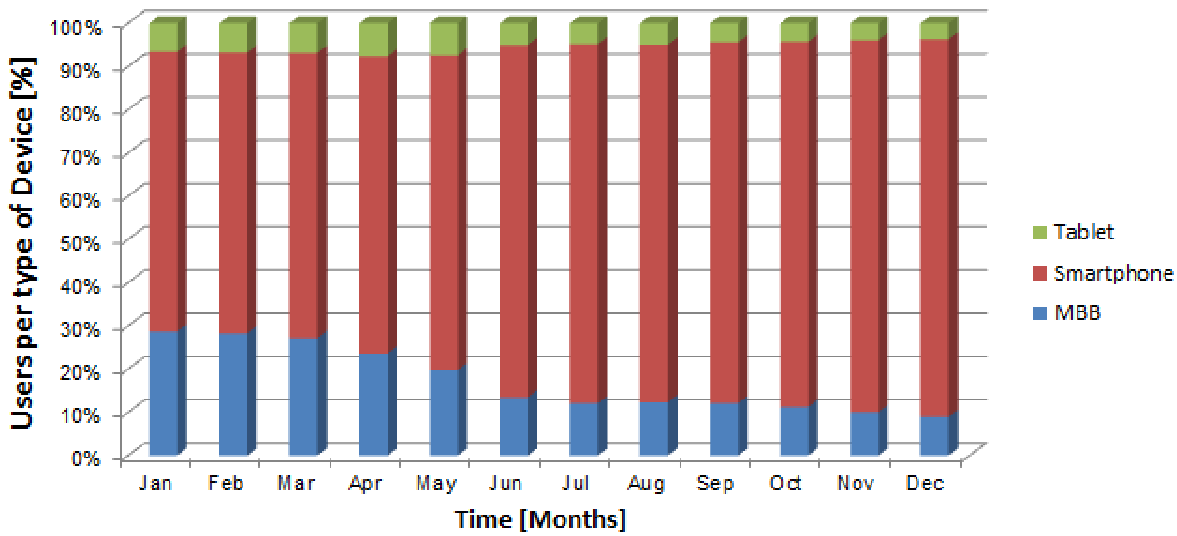

Figure 4.7. Usage of the types of devices in the 4 RNCs during the year.

It is seen that the percentage of Smartphone users among all the types of devices usually has a steady increase throughout the year, in contrast with MBB which has a decline.

A traffic model called Smartphone Traffic Model (STM) was developed for the Smartphones' usage in order to explain the behaviour of the calculated average of the 3 years. The best fit chosen for the left section is a power function, and for the right one is a rational function:

$$f_{stm}(t_{shift}) = \begin{cases} a_{stm} \times t_{shift}^{b_{stm}} + c_{stm} & 0 \leq t_{shift} \leq t_{stm} \\ \dfrac{p_{stm}}{t_{shift} + q_{stm}} + d_{stm} & t_{stm} < t_{shift} \leq 12 \end{cases} \tag{4.5}$$

with

$$f_{stm}(t_{stm}) = a_{stm} \times t_{stm}^{b_{stm}} + c_{stm} = \frac{p_{stm}}{t_{stm} + q_{stm}} + d_{stm} \tag{4.6}$$

where:

- f_{stm}: Smartphone Traffic Model function.
- a_{stm}: Power initial value.
- b_{stm}: Power decay factor.
- c_{stm}: Linear function offset.
- t_{stm}: Breakpoint shifted month.
- p_{stm}: Polynomial scale factor.
- q_{stm}: Polynomial shift.
- d_{stm}: Polynomial offset.

The parameters are also calculated for the STM, as shown in Table 4.4, in order to get the best approximation to the variation of the average sectors profile.

Table 4.4. Smartphone Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 7.92]	a_{stm}	b_{stm}	c_{stm}	t_{stm} = 7.92
			5.39×10^{-4}	3.64	0	
Right	Rational	[7.92; 12]	p_{stm}	q_{stm}	d_{stm}	
			2.46	-5.46	0	
R^2 = 0.969			$\sqrt{\varepsilon^2}$ = 0.0765		σ_{avg} = 0.418	

Figure 4.8 shows that there is a good proximity between the model and the data.

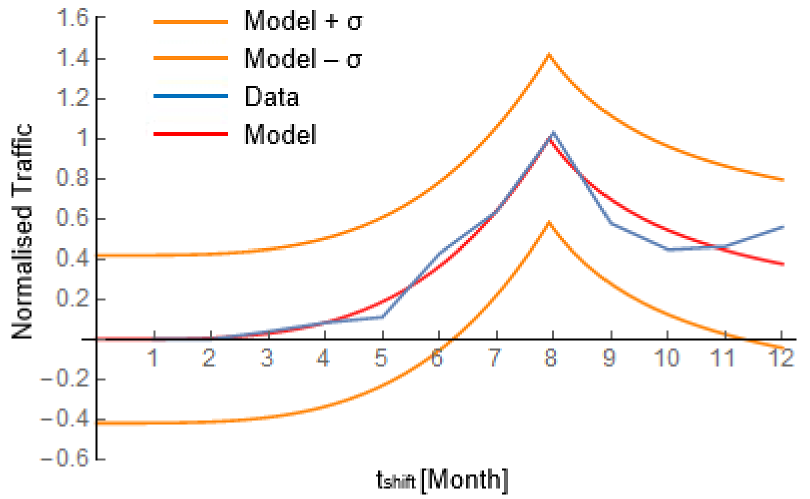

Figure 4.8. Smartphone Traffic Model, with the respective data and a confidence interval.

In Annex A, the approximation models for the other types of devices (MBB and Tablets) are shown, where it is possible to see the functions used for each section and their expressions, such as the values obtained. The Smartphones' usage was chosen for this section since this is the most used device, as it has been shown previously. Regarding the approximations relative to the use of each type of device, one can see again that the exponential function fits best the left section in 66.6% of the cases, and that the rational function does for the right one in 66.6% of the cases too.

The comparison of devices is shown in Figure 4.9 and Table 4.5. After analysing each of the best approaches for each category of data, it may be concluded that the best approach model corresponds to the Tablet Traffic Model one, due to its CD and RMSE values.

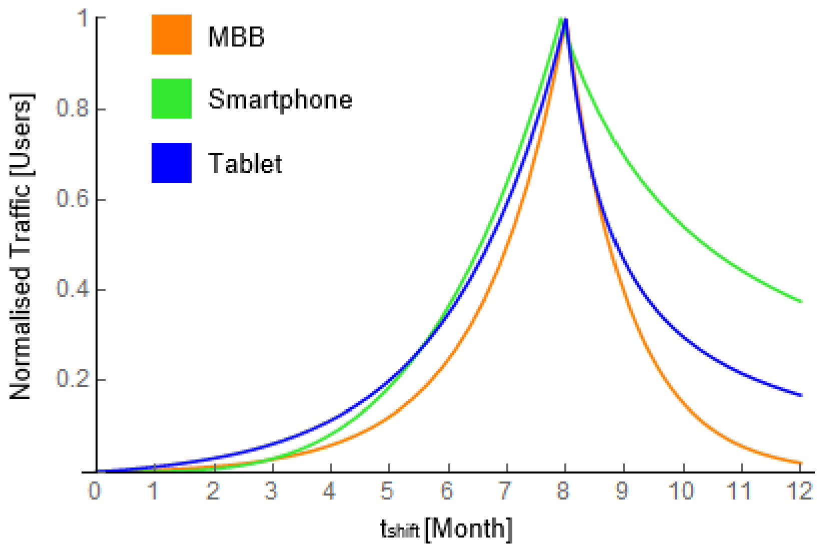

Figure 4.9. Traffic Models for the usage of the different types of devices.

Table 4.5. Traffic Model parameters for each type of device.

	MBB	Smartphone	Tablet
CD	0.988	0.969	0.988
RMSE	0.0451	0.0765	0.0426

4.4 Fitting Models for the Usage of Different Operating Systems

In the national and world markets, there is a huge competition for the creation of rich applications for mobile devices and to offer a better experience and satisfaction to users. Some development platforms for Smartphones and Tablets have attracted the attention of many consumers and emerges in this market, as Google's Android, iOS (iPhone) Apple, and Windows from Microsoft.

The market for mobile devices is now a great engine for technological development in this area. Almost everything is designed and developed in order to follow market trends and consumer preferences.

Data was collected only from 2 years (2013/2014) for the OSs, and once again the normalisation to the 1st August peak is done, as shown in Figure 4.10. In the first year, it is visible that the behaviour of all OSs is very similar, but in the second year there are some slight differences. These operating systems (Android, iOS and Windows) are currently transforming themselves in the main OSs for Smartphones and Tablets, being competitors with each other, not only in their architecture but also in their functionality. One of the interesting points of this competition is the growth and relevance of these systems. By definition, a mobile operating system is a set of programs with the function of managing the hardware and software features for mobile devices, and providing an end-user interface [SiGG04].

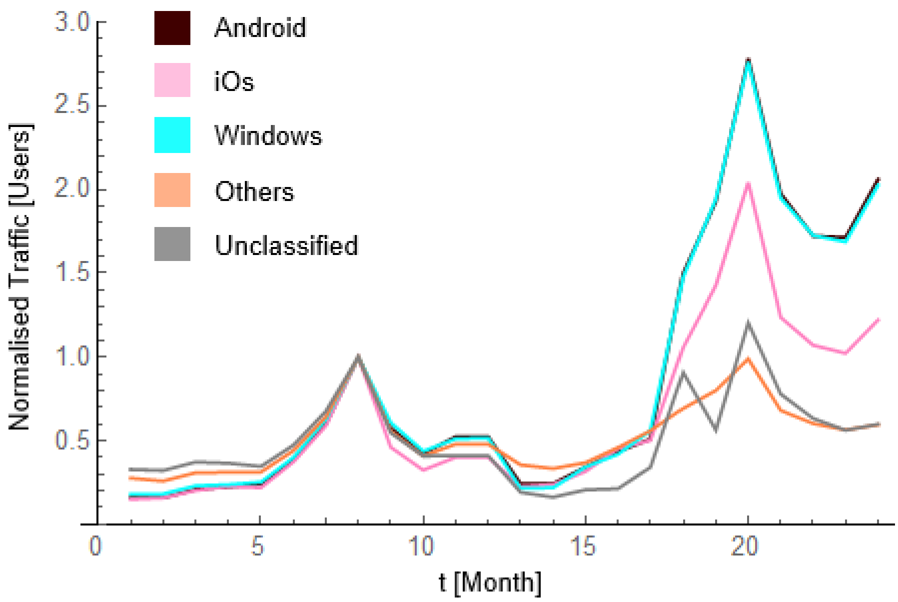

Figure 4.10. Behaviour of the number of users per OS over 2013 and 2014.

The average of the data was done for each OS, and the choice for mobile operating systems were Android, iOS and Windows, because of their market relevance and of wide use by manufacturers. The three analysed mobile operating systems have characteristics that make them very interesting choices. They present a set of features that fulfil the steps and expectations considered very important in mobile operating systems for Smartphones and Tablets. Figure 4.11 shows the percentage of users of each operating system in each month, and one can see that Android has consolidated its position as the most used operating system, but that there is a great deal of people who choose iOS.

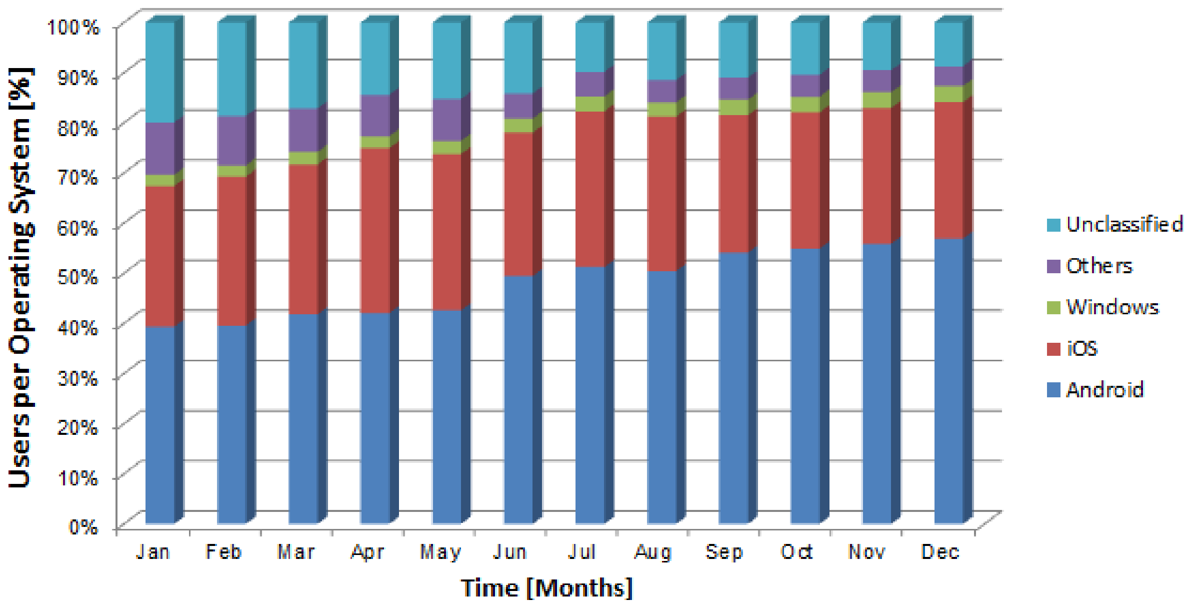

Figure 4.11. Usage of the OSs in the 4 RNCs during the year.

By analysing this figure, one can verify that the share of Android in the market shows a significant increase, that iOS' share remains virtually constant during the year, and that Microsoft is the third major OS designed for mobile devices, but still with a very small expression in the segment and a long way to face up the competition.

It is worth mentioning that there are other operating systems, such as Symbian, Blackberry, Bada, among others, but their numbers are far less significant today. There is also the Unclassified share, which represents the equipment from which it was not possible to determine the operating system. However, due to the advancement of technology and equipment, it will become possible to identify all types of equipment and operating systems on the network, thereby the Unclassified category will become negligible over the years.

The Android platform is based on the Linux kernel and has dethroned, little by little, Apple and Microsoft, according to the latest market research. The developed traffic model that shows the best fit for the average profile of the Android's usage is called Android Traffic Model (ATM), and is shown in Figure 4.12: for the left section the best fit is a power function, and for the right one a rational function:

$$f_{atm}(t_{shift}) = \begin{cases} a_{atm} \times t_{shift}^{b_{atm}} + c_{atm} & 0 \leq t_{shift} \leq t_{atm} \\ \dfrac{p_{atm}}{t_{shift} + q_{atm}} + d_{atm} & t_{atm} < t_{shift} \leq 12 \end{cases} \tag{4.7}$$

with

$$f_{atm}(t_{atm}) = a_{atm} \times t_{atm}^{b_{atm}} + c_{atm} = \frac{p_{atm}}{t_{atm} + q_{atm}} + d_{atm} \tag{4.8}$$

where:

- f_{atm}: Android Traffic Model function.
- a_{atm}: Power initial value.
- b_{atm}: Power decay factor.
- c_{atm}: Linear function offset.
- t_{atm}: Breakpoint shifted month.
- p_{atm}: Polynomial scale factor.
- q_{atm}: Polynomial shift.
- d_{atm}: Polynomial offset.

As one can see in Figure 4.12, there is a fairly sharp increase over the months from May to July. These growths are always dependent on the number of devices, and it is well known that in summer, especially June, July and August, there is a huge increase in this number compared with the other months. In November and December there is an increase again, which has to do with the New Year's Eve. In Table 4.6 the parameters corresponding to the ATM are shown, providing a decent approximation.

Table 4.6. Android Traffic Model parameters.

<table>
<tr><th>Section</th><th>Function</th><th>Interval [Month]</th><th colspan="4">Parameters</th></tr>
<tr><td rowspan="2">Left</td><td rowspan="2">Power</td><td rowspan="2">[0; 7.86]</td><td>a_{atm}</td><td>b_{atm}</td><td>c_{atm}</td><td rowspan="4">t_{atm} = 7.86</td></tr>
<tr><td>6.52×10^{-4}</td><td>3.56</td><td>0</td></tr>
<tr><td rowspan="2">Right</td><td rowspan="2">Rational</td><td rowspan="2">[7.86; 12]</td><td>p_{atm}</td><td>q_{atm}</td><td>d_{atm}</td></tr>
<tr><td>4.21</td><td>-3.65</td><td>0</td></tr>
<tr><td colspan="3">R^2 = 0.969</td><td colspan="2">$\sqrt{\overline{\varepsilon^2}}$ = 0.0823</td><td colspan="2">σ_{avg} = 0.356</td></tr>
</table>

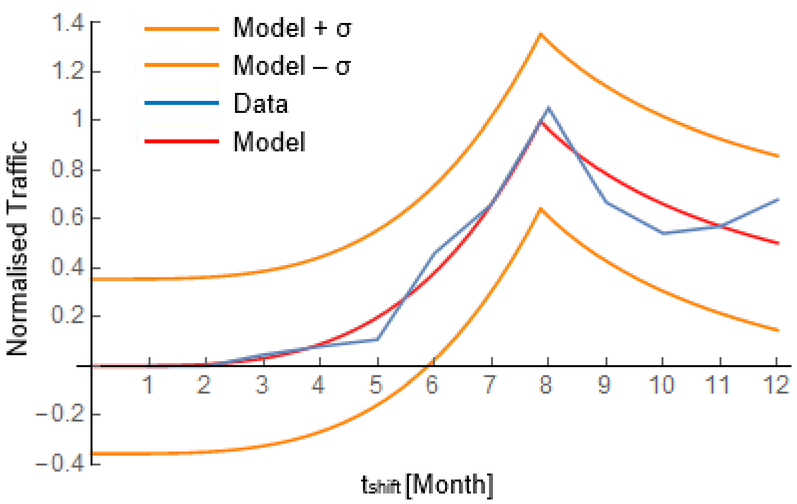

Figure 4.12. Android Traffic Model, with the respective data and a confidence interval.

In Annex A, the approximation models for the other operating systems (iOS, Windows, Others and

Unclassified) are shown, where it is possible to see the functions used for each section and their expressions, such as the values obtained. Android's usage was chosen for this section since it is the most used OS. Regarding the OSs, the most used function for the left section is the power one, which fits in 80% of the cases, and for the right section is the rational one, which is used in all cases. A comparative analysis of the results obtained from the models is presented in Table 4.7 and Figure 4.13. Considering CD and RMSE values, the Others Traffic Model has the best approximation. It can be observed that the left section has a similar behaviour for all OSs, whereas in the right section the behaviour is different because some OSs have a more marked decline.

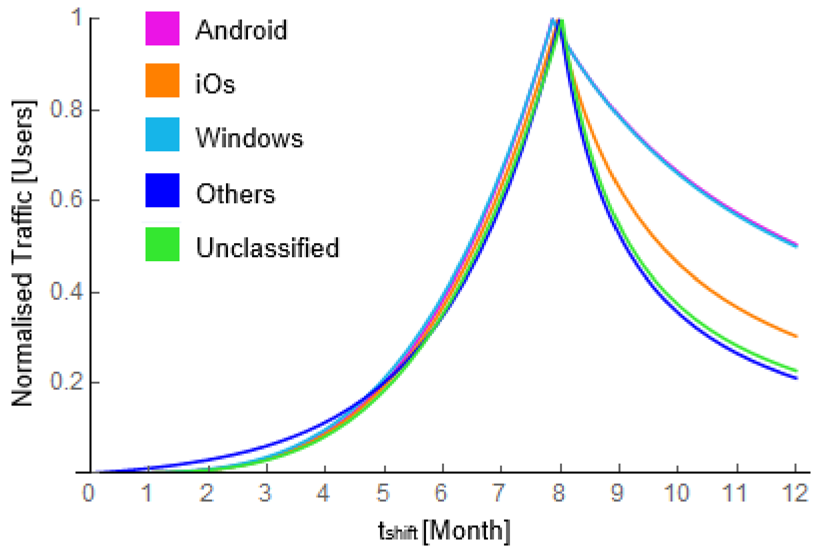

Figure 4.13. Traffic Models for the usage of the different OSs.

Table 4.7. Traffic Model parameters for each operating system.

	Android	**iOS**	**Windows**	**Others**	**Unclassified**
CD	0.969	0.970	0.972	0.983	0.960
RMSE	0.0823	0.0707	0.0769	0.0506	0.0806

4.5 Fitting Models for the RNCs' Data Traffic

A study on the number of existing users during the studied period has been already done. In what follows, the study is conducted for the data traffic made by those users.

4.5.1 Data Traffic per RNC

In Figure 4.14, the specific data traffic for each RNC is represented. One should remember that RNC1 and RNC3 have information from the full 3 years (2012/2013/2014), while RNC2 and RNC4 have only some information from 2 years (2013/2014). Despite this lack of information, the essential information for this study is present. Once again, the annual average is done for each RNC, and its behaviour is studied in order to find the best approximation and the corresponding confidence interval.

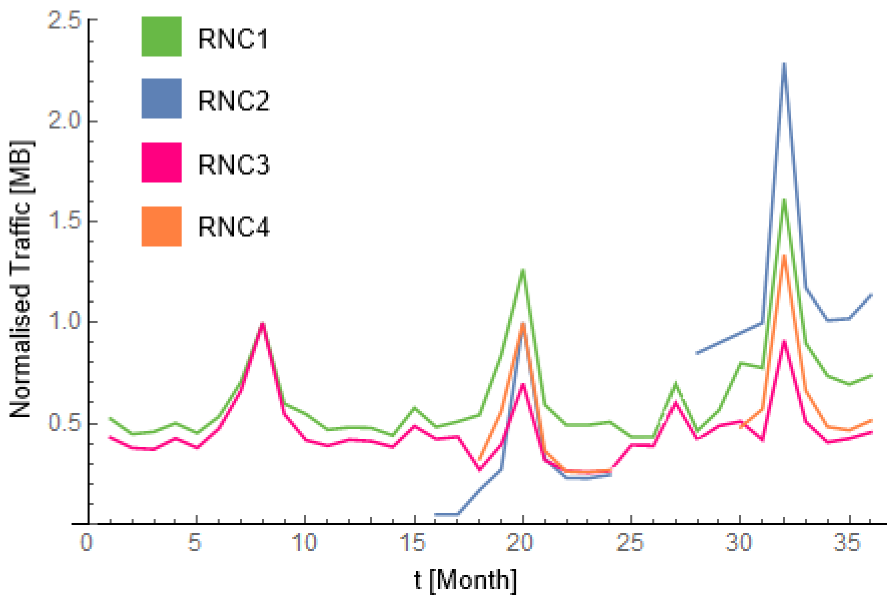

Figure 4.14. Behaviour of the data traffic for each RNC from 2012 to 2014.

Previously, a detailed explanation of RNC1's user traffic has been presented, and now it is possible to complement that information with its data traffic information. For this, a model called RNC1 Data Traffic Model (DRNC1) was developed, which consists of an exponential function for the left section and a rational function to the right one. The parameters of this model are shown in Table 4.8, and the model is represented in Figure 4.15, being given by:

$$f_{drnc1}(t_{shift}) = \begin{cases} a_{drnc1} \times e^{t_{shift} \times b_{drnc1}} + c_{drnc1} & 0 \leq t_{shift} \leq t_{drnc1} \\ \dfrac{p_{drnc1}}{t_{shift} + q_{drnc1}} + d_{drnc1} & t_{drnc1} < t_{shift} \leq 12 \end{cases} \tag{4.9}$$

with

$$f_{drnc1}(t_{drnc1}) = a_{drnc1} \times e^{t_{drnc1} \times b_{drnc1}} + c_{drnc1} = \frac{p_{drnc1}}{t_{drnc1} + q_{drnc1}} + d_{drnc1} \tag{4.10}$$

where:

- f_{drnc1}: RNC1 Data Traffic Model function.
- a_{drnc1}: Exponential initial value.
- b_{drnc1}: Exponential decay factor.
- c_{drnc1}: Exponential offset.
- t_{drnc1}: Breakpoint shifted month.
- p_{drnc1}: Polynomial scale factor.
- q_{drnc1}: Polynomial shift.
- d_{drnc1}: Polynomial offset.

As one can see in Figure 4.15, a good approximation of the average values is achieved. In March, a higher value than the nearby months is shown, perhaps due to some seasonal effect in this region.

Table 4.8. RNC1 Data Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 8.00]	a_{drnc1}	b_{drnc1}	c_{drnc1}	$t_{drnc1} = 8.00$
			1.59×10^{-3}	0.806	-1.59×10^{-3}	
Right	Rational	[8.00; 12]	p_{drnc1}	q_{drnc1}	d_{drnc1}	
			0.467	-7.54	-1.59×10^{-3}	
$R^2 = 0.985$		$\sqrt{\overline{\varepsilon^2}} = 0.0496$			$\sigma_{avg} = 0.133$	

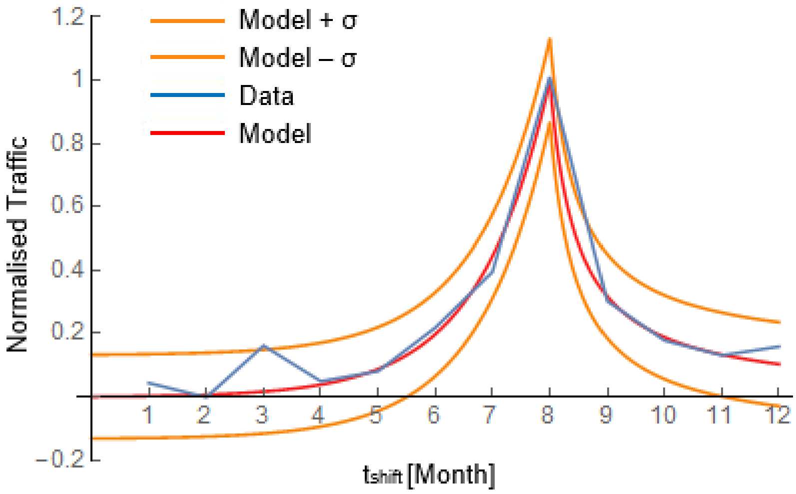

Figure 4.15. RNC1 Data Traffic Model, with the respective data and a confidence interval.

Figure 4.16 shows the best approximations of the data traffic of each RNC and Table 4.9 presents the parameters of the developed models. For the data traffic of each RNCs, the exponential function is used for the left section in 75% of the cases and the rational and the exponential function fit best the right section in 50% of the cases each.

Table 4.9. Data Traffic Model parameters for each RNC.

RNC	RNC1	RNC2	RNC3	RNC4
CD	0.985	0.989	0.952	0.999
RMSE	0.0496	0.0463	0.0953	0.0156

One can easily reach the conclusion that RNC4 is again the one that has a better approximation model, taking CD (better if closer to 1) and RMSE (better if closer to 0) into account, but it also happens due to the lack of data samples.

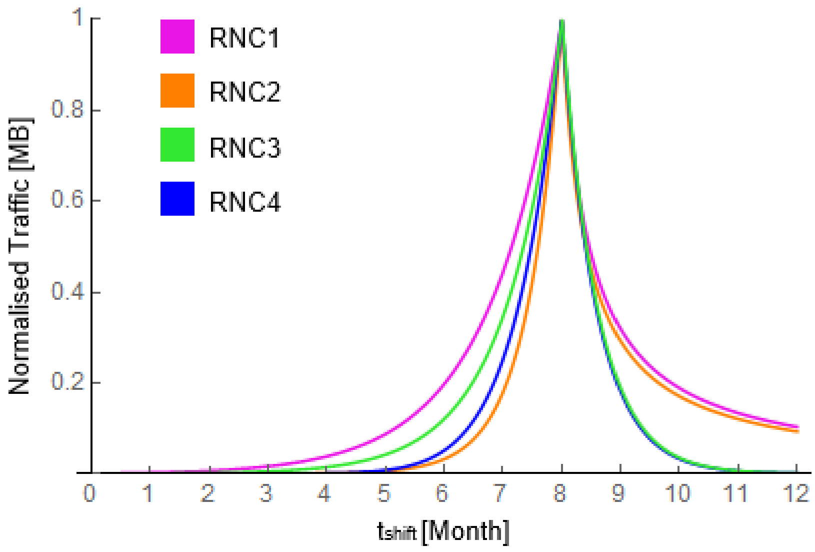

Figure 4.16. Data Traffic Model of each RNC.

Annex A shows the parameters and the graphical representation of the data traffic models of the other RNCs and their respective average sector profiles.

4.5.2 Total Data Traffic of the 4 RNCs

The behaviour of the sum of the data traffic of the 4 RNCs is shown in Figure 4.17.

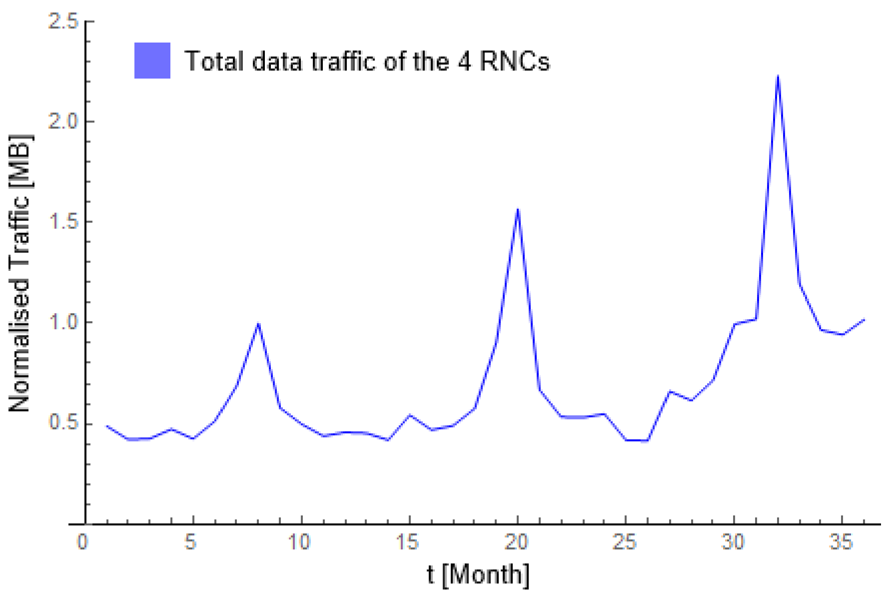

Figure 4.17. Trend of the total data traffic of the 4 RNCs from 2012 to 2014.

The model called Total Data Traffic Model (TDTM) is developed for the total of data traffic of the 4 RNCs in order to describe the behaviour of the average values. The proximity between the model and the average profile is visible in Figure 4.18, which allows to conclude that the model is good. The parameters' values obtained for the traffic model are shown in Table 4.10.

The best fitting approximations for the model's sections are the exponential and the rational functions:

$$f_{tdtm}(t_{shift}) = \begin{cases} a_{tdtm} \times e^{t_{shift} \times b_{tdtm}} + c_{tdtm} & 0 \leq t_{shift} \leq t_{tdtm} \\ \dfrac{p_{tdtm}}{t_{shift} + q_{tdtm}} + d_{tdtm} & t_{tdtm} < t_{shift} \leq 12 \end{cases} \tag{4.11}$$

with

$$f_{tdtm}(t_{tdtm}) = a_{tdtm} \times e^{t_{tdtm} \times b_{tdtm}} + c_{tdtm} = \frac{p_{tdtm}}{t_{tdtm} + q_{tdtm}} + d_{tdtm} \tag{4.12}$$

where:

- f_{tdtm}: Total Data Traffic Model function.
- a_{tdtm}: Exponential initial value.
- b_{tdtm}: Exponential decay factor.
- c_{tdtm}: Exponential offset.
- t_{tdtm}: Breakpoint shifted month.
- p_{tdtm}: Polynomial scale factor.
- q_{tdtm}: Polynomial shift.
- d_{tdtm}: Polynomial offset.

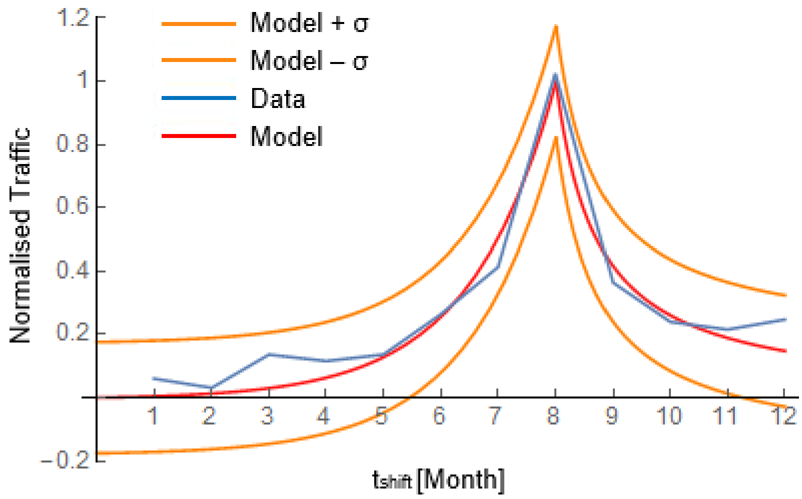

Figure 4.18. Total Data Traffic Model, with the respective data and a confidence interval.

Table 4.10. Parameters of the Total Data Traffic Model.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 8.01]	a_{tdtm}	b_{tdtm}	c_{tdtm}	$t_{tdtm} = 8.01$
			4.62×10^{-3}	0.672	-4.62×10^{-3}	
Right	Rational	[8.01; 12]	p_{tdtm}	q_{tdtm}	d_{tdtm}	
			0.717	-7.30	-4.62×10^{-3}	
$R^2 = 0.980$		$\sqrt{\overline{\varepsilon^2}} = 0.0584$			$\sigma_{avg} = 0.176$	

It is important to refer that when the data tends down in the right section the approach is always better because the right function never approximates the rise that happens by the end of the year

4.6 Prediction Models

A major requirement for any network operator is to ensure that the network is operating to its maximum efficiency. This prediction section demonstrates the behaviour that the network will probably have in 2015 and 2016, based on data analysis from the collected data in order to have a network optimisation.

4.6.1 Discontinuities of the Model

As it is explained in Section 3.2.4, initially a normalisation of the specific set of data from the 2 or 3 years being analysed is done in relation to the August peak of the first year.

Having found the approximation models of the annual behaviour of the data being studied, the next step is to calculate their average values in order to proceed to the extension of the model for the next two years and have a good enough prediction. Having done almost all of the steps introduced in Section 3.2.4 (except the final rectification one) and applying a 10% higher growth in the evolution for the predicted years (represented by the green lines), one can obtain for the evolution of the total number of users of the 4 RNCs the brown curves that are shown in Figure 4.19.

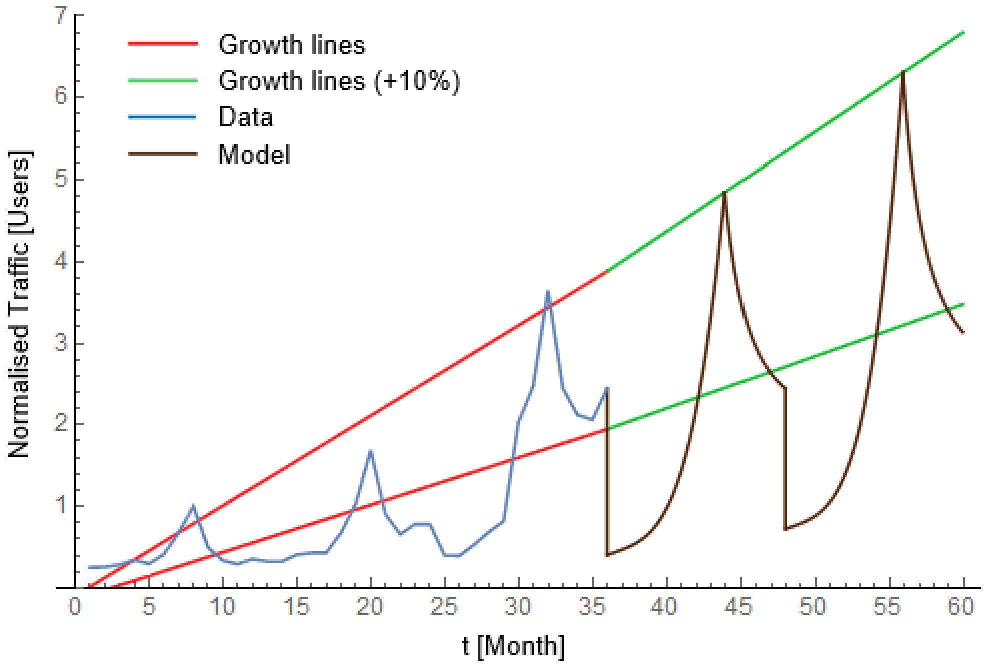

Figure 4.19. Prediction of the number of total users of the 4 RNCs for 2015 and 2016 with a pure and simple extension of the model.

By analysing this figure, one can reach the conclusion that this pure and simple extension does not work well, because it originates discontinuities in December 2014 and 2015 that obviously no one is expecting them to happen.

After this, one can also conclude that the left and the right curves of an annual extension must be made independently, because the left one has to ensure continuity.

Hereupon, the next step is to rectify the model as described in the end of Section 3.2.4, stretching either up or down the curves and correcting the gaps.

In the first year of prediction, it is imposed that the left curve has to start in the point corresponding to the last month of the previous year (2014) and that in the end it has to intersect the line of peaks. In the second year it is established that the left curve has to start in the value that the model takes at the end of 2015 and has to end intersecting the line of peaks again. This way, the right curves do not need to be corrected.

4.6.2 Prediction of the Total Users and the Total Data Traffic of the 4 RNCs

The technological development that has been made in recent decades in some areas has had a strong impact on society, not only from an economic point of view but also regarding the behaviour of people. The continued growth of mobile phone users obliges the operators to continuously increase the capacity of their networks, in order to fulfil customers' desired quality of service. One of the aims of this thesis is to understand the behaviour of the network over the years in order to help the mobile operator in its optimisation.

If the network is not performing to its maximum, this can directly result in poor performance and a slow response to data downloads. As a result, network performance, and hence the service quality seen by users, is a key differentiator. A good level of perceived quality will help retain users, whereas poor service will lead to high levels of churn for the network operator.

Considering the huge amounts of investment that mobile operators require to set up their systems, it is necessary to ensure that they operate to their greatest efficiency. As a result, mobile network optimisation tools are an essential requirement for any network these days. They enable a faster response to performance issues, a better service quality across the network, lower operating costs for the cellular network, and a much more efficient use of the invested capital.

In this study, a prediction for 2 years (2015/2016) is made, so in order to differentiate the collected data from the prediction, a use of different colours is done. In Figure 4.20 and Figure 4.21, the blue colour represents the collected data, and the two red lines are growth lines: the red top line is the average of the three higher peaks that always occur in August and the red bottom line is the mean curve, i.e., it is the average trend of growth of the 3 years. The green line represents the start of the prediction, and a 10% higher growth compared to the previous years, obviously, instead of 10%, any other value (positive or negative) can be chosen; the developed program allows the user to insert different values of growth in each one, and then both years being predicted. The value of 10% growth has been chosen because on the one hand one can expect a continued growth, therefore imposing its linearity, and on the other hand, given the expansion of mobile communications that occurred in the past years, one can also expect a slight extra boost; anyway, this growth may stabilise, which is why such a "low value" is chosen. Finally, the brown curves represent the approximation model developed for each case and the orange curves define the confidence interval for each model.

The evolution of the total number of users of the 4 RNCs is presented in Figure 4.20. Annex B shows other prediction models, particularly for each RNC.

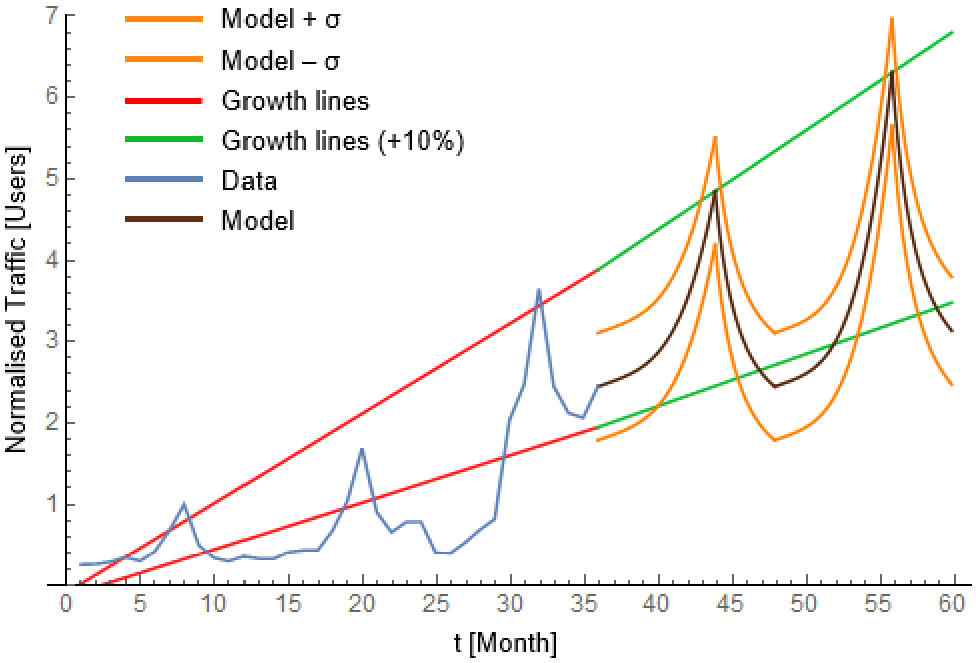

Figure 4.20. Prediction of the number of total users of the 4 RNCs for 2015 and 2016.

By observing Figure 4.20, one can conclude that there has been an annual growth, especially in the summer months due to increased migration, and that the number of users had a much larger increase in 2014 than 2012 and 2013, mainly due to new facilities in purchasing a new mobile equipment, and also because nowadays mobile operators invest in more plans and promotions, which increase access to mobile Internet. Although 2014's jump is great, it is believed that in 2015 the increase will no longer be so abrupt even taking a 10% increase in the annual growth line into account.

In order to avoid unpleasant surprises, like slower data traffic or even a saturation of services, some aspects need to be taken into account in this study. After a comparison of the user traffic in the summer in different years, as this is the moment of the year that the traffic usually has a big peak, one can infer through Figure 4.20 that in 2015 the traffic will be 1.32 times bigger than the one observed in 2014, and that in 2016 it will be 1.74 times greater than the traffic registered in 2014, which means that the network must be prepared to deal with this growth. Although these values should provide a good enough prediction of the user traffic growth, one should not forget that a 10% higher growth is being imposed, and that there is a confidence interval that covers possible deviations of the model helping the operator to have an estimate of how far this model can go.

Nowadays, it is still common to find operators that offer data plans for each equipment, but with the increasing number of devices per user, overall packages begin to emerge accommodating various devices and boosting the increase of consumption data. Operators are obliged to follow this development with the aim of attracting more customers and retain those that they already have. Figure 4.21 shows that there has been a steady growth of data traffic over the first three years and specially in the summer, which is the time of the year in the region being studied that has a higher propensity for an increased use of mobile devices.

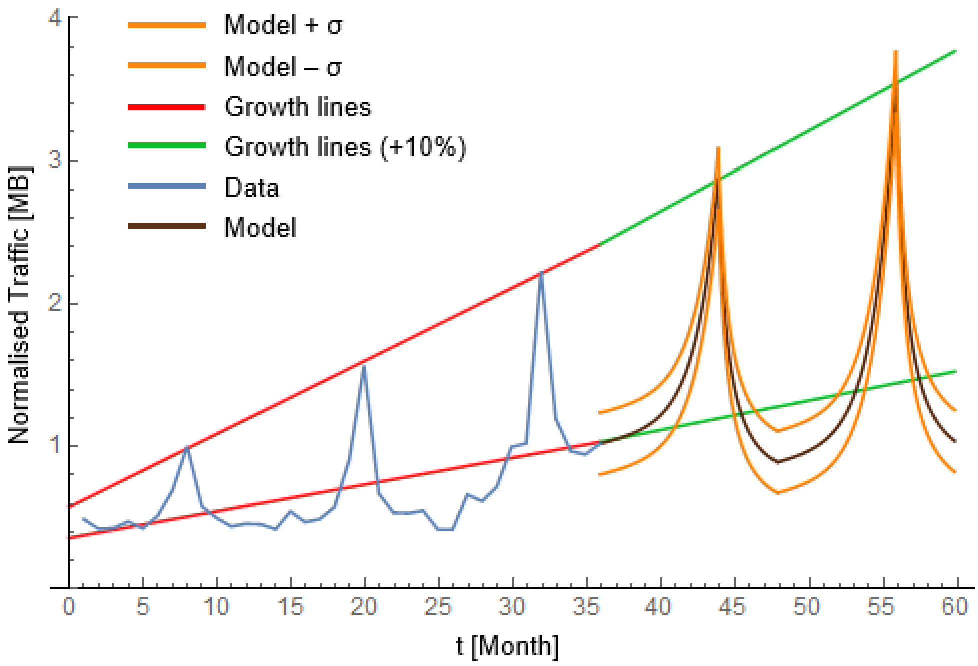

Figure 4.21. Prediction of the data traffic of the 4 RNCs for 2015 and 2016.

The global mobile data traffic is also driven by the growing number of mobile users. By analysing Figure 4.21, it can be concluded that in order to be possible to maintain the desired performance of the network and to ensure that there is enough capacity to for the generated traffic, the network in August 2015 will need to be prepared to cope with 1.29 times more than in August 2014, while in August 2016 data traffic will be 1.57 times larger than in August 2014.

4.6.3 Prediction of the Usage of Different Types of Devices

The proliferation of new devices such as Smartphones, Tablets and Computers (using the MBB) connected to mobile networks is great at generating traffic, since these devices are able to offer the consumer content and applications that are not supported by mobile devices of the previous generation.

The global adoption of Smartphones is accelerating rapidly, and the way people use these devices in mobile networks is changing too. Voice no longer dominates: today's Smartphone users are most interested in Internet browsing, social media, instant messaging and mobile app usage [Eric15b]. Figure 4.22 shows the growth of Smartphones' usage during 2012, 2013 and 2014 and its prediction for 2015 and 2016 considering a 10% higher growth. It can be seen that there has been a gradual increase along the years, and that in 2014 there was a big rise in the usage of Smartphones reaching a value that is 4.74 greater than the one registered in 2012. Regarding network optimisation, one concludes that the network must be prepared in August 2015 for Smartphone traffic 1.33 times bigger than the one observed in August 2014. In August 2016 the network will need to be prepared to a Smartphone usage 1.78 times greater than the use registered in August 2014. Again, if one only considers the 2013 and 2014 data, Smartphone usage will reach even higher levels than the ones predicted with the model, because there is an abrupt growth between these two years.

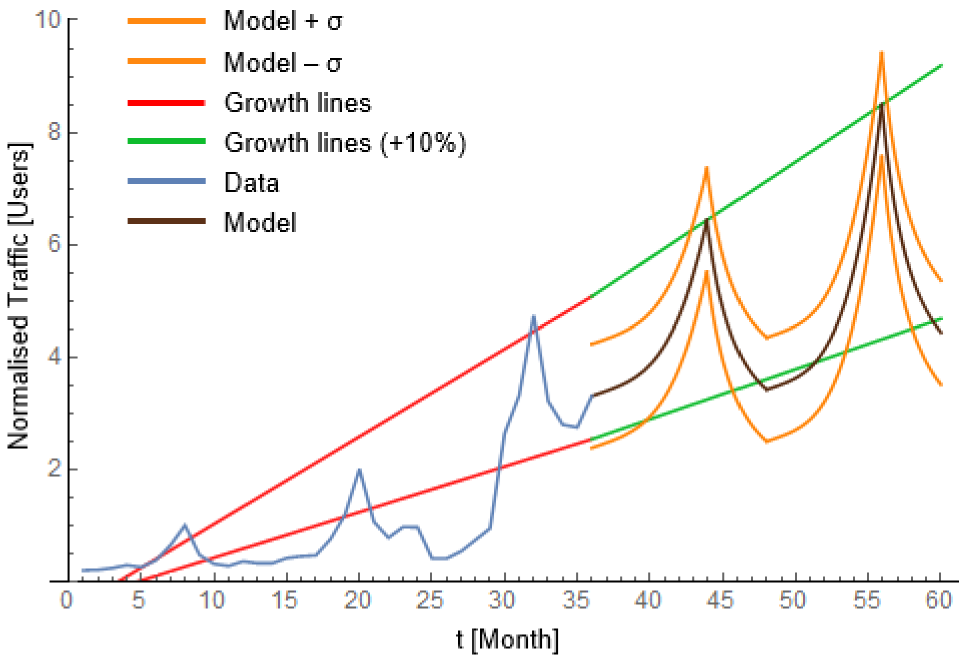

Figure 4.22. Prediction of the usage of Smartphones for 2015 and 2016.

The prediction models regarding the other devices (MBB and Tablet) are shown in Annex B. Table 4.11 shows that Smartphones should have a greater relative rise of usage in 2015 and 2016 compared to 2014, but a rise that is not necessarily bigger than the one of the other devices (due to the normalisation). As one can see, the use of all of them will probably increase in the next years.

Table 4.11. Increase in devices' traffic in 2015 and 2016 compared to 2014.

	MBB	Smartphone	Tablet
2015	1.01	1.33	1.24
2016	1.06	1.78	1.58

4.6.4 Prediction of the Usage of Different Operating Systems

As mentioned before, the market is now divided into three major operating systems for Smartphones and Tablets: Android, iOS and Windows. The other systems are failing to impose themselves as alternatives and are at risk of disappearing or leaving the European market soon. Among the big three, Google's Android, although remaining as the market leader, is being challenged by Apple's iOS in high-end devices.

Android is currently the operating system deployed in the largest number of mobile devices in the world and has the largest market share and sales, so it makes sense that there is a growing trend that occurs in the last two years (2013 and 2014), as it can be seen in Figure 4.23. The most important thing to analyse are the peaks that occur in August in order to predict what will happen in the coming years. As one can see, Android's usage in 2014 was 2.78 times bigger than the one registered in 2013. Applying a 10% higher growth again in the prediction model, Figure 4.23 shows that the usage of Tablets will probably be 1.68 times greater in 2015 than the usage in 2014, and 2.37 times greater in 2016 than in 2014 too.

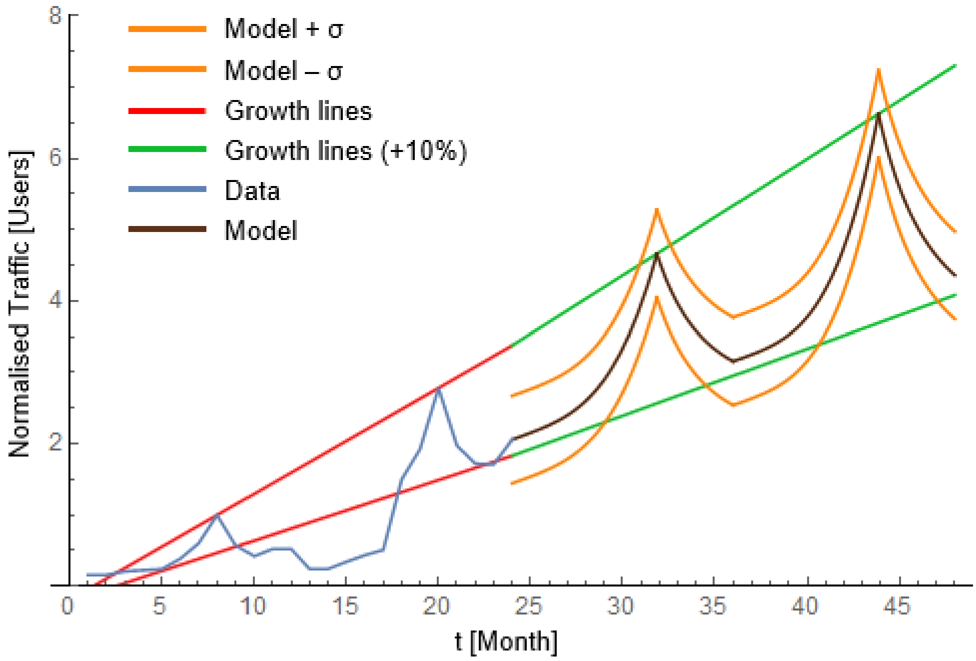

Figure 4.23. Prediction of the usage of the Android for 2015 and 2016.

The prediction models regarding the other OSs (iOS, Windows, Others and Unclassified) are shown in Annex B. Table 4.12 shows the increase of OSs' usage in the network.

Table 4.12. Increase in OSs' traffic in 2015 and 2016 compared to 2014.

	Android	iOS	Windows	Unclassified
2015	1.68	1.53	1.66	1.18
2016	2.37	2.10	2.38	1.36

As one can see, it is predicted that the main OSs will continue to grow. The biggest relative rise for

2015 belongs to Android, while in the second year of the prediction (2016) the system with a larger relative increase in the network is the Windows, but a rise that is not necessarily bigger than the one of the other devices (due to the normalisation).

Finally, the prediction models regarding the number of users and the data traffic of each RNC are also shown in Annex B.

Chapter 5

Conclusions

This chapter provides a summary of all the work carried out under the scope of this thesis. Then some important aspects, like useful guidelines for a network optimisation, are given as well as some hints for future work.

Currently, Telecommunications play an increasingly important role not only in society but also in the economy of countries, being the market segment with the highest growth at national and international levels. This thesis intends to develop a model for data analysis from UMTS networks, for different user data service profiles, leading to network optimisation. With this traffic profiling, one intends to do a prediction for 2015 and 2016. For that, statistical methods, regression analysis and mathematical trend projections were used. These models are based on the assumption that the future is a function of the past. In other words, these models look at what has happened over a period of time and use a number of previous data in order to make a prediction.

In Chapter 1, a brief description of the mobile communications systems evolution over time and of the growing consumer and traffic demand is presented, followed by an explanation on the motivation and contents of this thesis.

Chapter 2 is focused on the UMTS basic concepts, presenting the network architecture, radio interface, services and applications, mobile terminals, traffic evolution, and the state of the art. Section 2.1.1 addresses the network architecture, describing the most important network elements, such as the UE, UTRAN, CN, External networks, among others. Section 2.1.2 addresses the radio interface, focusing on the multiple access techniques, modulation, FDD, 3GPP evolution and peak data rate for DL and UL. In Section 2.2, services and applications are described and characterised concerning data rates. Section 2.3 is dedicated to mobile terminals, focusing on the growth of the different mobile devices, the usage of different types of devices and OSs, the behaviour of the most used OSs, the share of traffic by application type, and the mobile speed by device. Section 2.4 discusses traffic evolution, such as the mobile data consumption, the usage of OSs with unlimited plans or with tiered pricing plans, the growth in devices, and about the global mobile data traffic in the next years. Finally, Section 2.5 describes the state of the art, which contemplates relevant works important for this thesis, focusing on different approaches.

One of the main conclusions that can be drawn is that the number of mobile-connected devices is growing fast, which has exceeded the world’s population in 2014. Moreover, it is estimated that monthly global mobile data traffic will surpass 24.3 Exabytes by 2019. Around the world, the Smartphone, along with the Tablet, and a fast-expanding of “wearables” and other “smart” devices, are transforming the way people live, work, play, connect, and interact. In the process, they are converting the digital revolution into an increasingly mobile phenomenon. Smartphones will reach three-quarters of mobile data traffic by 2019, and Tablets will exceed more than 10% of global mobile data traffic by 2017. In the same year (2017), Smartphones, Laptops and Tablets will account for 93% of global mobile data traffic. Regarding operating systems, the market for the different types of devices is essentially dominated by two platforms, Android and iOS, which together are dominating the world.

In Section 3.1, the temporal modelling of users’ behaviour is described, which shows that the Smartphone remains the device of choice throughout the day. Section 3.2 gives an overview and implementation details of the developed model, which was entirely developed from scratch. The core of the model is a distribution fitting algorithm that allows an extrapolation to the prediction of trends. In the model overview section, a systemic view of the model developed for the purpose of this thesis is depicted, in a way that one can easily identify which are the inputs and outputs. Moreover, the various blocks that make up the model are presented: Initial Data Analysis, Fitting Models, and Prediction Models. Section 3.3 addresses the human behaviour throughout the year, in relation to data traffic and messages sent, and also water consumption and hotel reservations in south of Portugal in order to understand what happens and with that to obtain study models that certify the developed model. Section 3.4 shows the behaviour and the evolution of mobile applications applying the developed model. One can conclude that some mobile applications (like Facebook, Twitter and WhatsApp) are growing fast, very different from other applications (like Skype, Viber and KIK messenger) that are also growing, but that do not allow the use of the developed model due to huge discontinuities.

Chapter 4 shows the distinct scenarios descriptions and the results obtained. It is important to understand what effect the strategic actions taken by the mobile network operators cause in the behaviours and perceptions of mobile equipment holders. The obtained predictions give an idea about the way how each user uses the mobile terminal, which is important because it allows one to have an idea about how much the services provided by the operator, the tariffs they practice and the types of terminal they have affect people's behaviours and desires. The better the service and the price, the more people will request it. From this point of view, it is important to help the mobile operator to get an idea of what network behaviour will be in times of a larger movement of people to a region.

The high quality of the approximation models obtained (proximity above 95%) shows that users' behaviour is not random throughout the year. There is quite a growth in the first half of the year and then the usage becomes progressively lower until its end. Moreover, it appears that the exponential function is the one that fits best the growth in the first part of the year in 55.6% of cases, followed by the power function that is the best fit in 38.9% of them. Regarding the second half of the year, the function that better explains the decrease in traffic is the rational in 77.8% of cases. This shows that the traffic in this region has an approximately exponential growth in the first part of the year and then gradually decreases until the beginning of next one.

In order to give a better support to a mobile operator concerning the evolution of the network for the future, and in order to beware of a possible network saturation, one should note that it is predicted that users' traffic in 2016 will be 1.74 bigger than the one obtained in 2014, and that data traffic in 2016 will be 1.57 greater than the one obtained in 2014 too. Smartphones and Androids' usage is expected to continue to grow apace. In 2016, it is expected that the Smartphone users will be 1.78 times more than the ones that existed in 2014, and that the Android users will be 2.37 times more than the ones that were registered in 2014 too.

Although the objectives of this thesis are accomplished and the model works good enough to get useful conclusions, the analysis done in this work has some limitations, namely in the amount of samples processed. In order to make a good prediction a fairly large source of data is needed, but in this study it was not possible to obtain a large sample of past data. In addition to the fact that some studies are based on data from only 2 years (regarding the RNC2 and the RNC4), the remaining are not provided with much information too. If it would be possible to have more data samples, a more precise and accurate analysis would have been possible.

Finally, with this study it is possible to create approximation models that are added in the prediction models that explain the temporal growth of the data provided by the mobile operator. One has imposed a linear growth of the traffic throughout the year. However, this should not always correspond to reality, since the 3G growth will eventually stabilise over the years, not least because the 4G technology is growing, which brings a large increase in bandwidth and allows the use of voice, Internet and multimedia services via streaming.

For future work, it is suggested that measurements campaign in the network should be considered, as it provides a better understanding of how a real network behaves, giving a general perspective of the influence of certain parameters on system performance. Furthermore, a more extensive process of data collecting is also important. Additionally, what was previously done in UMTS can now be done for LTE, since this is a technology that is growing fast nowadays. 4G is increasing in consumers' lives, since it provides fast data traffic. It is important for operators to know more about this growth, because sooner or later the 2G and 3G markets will tend to disappear and they need to be prepared for the traffic that this new technology will bring to the network. For that, some frequencies can be reused, which is already happening.

With 3G, the goal changed from voice-centric to data-centric. Moreover total mobility became an objective to pursuit. In this generation it is possible to combine voice, multimedia applications and mobility in a never experienced manner. However, global mobility, while an important objective, was never really reached by this generation. At the same time new applications demand more bandwidth and lower costs. The newcomer 4G tries to address this problem by integrating all different wireless technologies. In spite of all the evolving technologies, the final success of new mobile generations will be dictated by the new services and content made available to users. These new applications must meet user expectations, and give added value over existing offers. So, it is important to have a model for the characterisation of different user data service profiles data applications in the temporal domain, leading to network optimisation, based on real data analysis from cellular networks.

Annex A

Fitting Models

In this annex, the approximation models for the users and data traffic of the RNCs used in the present work are shown.

A.1 User Traffic per RNC

RNC1

$$f_{urnc1}(t_{shift}) = \begin{cases} a_{urnc1} \times e^{(-((t_{shift}-b_{urnc1})/c_{urnc1})^2)} + d_{urnc1} & 0 \leq t_{shift} \leq t_{urnc1} \\ \dfrac{p_{urnc1}}{t_{shift} + q_{urnc1}} + g_{urnc1} & t_{urnc1} < t_{shift} \leq 12 \end{cases} \quad \text{(A.1)}$$

with

$$f_{urnc1}(t_{urnc1}) = a_{urnc1} \times e^{(-((t_{urnc1}-b_{urnc1})/c_{urnc1})^2)} + d_{urnc1} = \frac{p_{urnc1}}{t_{urnc1} + q_{urnc1}} + g_{urnc1} \quad \text{(A.2)}$$

where:

- f_{urnc1}: RNC1 User Traffic Model function.
- a_{urnc1}: Gaussian amplitude.
- b_{urnc1}: Month shifted peak value.
- c_{urnc1}: Gaussian deviation.
- d_{urnc1}: Gaussian offset.
- t_{urnc1}: Breakpoint shifted month.
- p_{urnc1}: Polynomial scale factor.
- q_{urnc1}: Polynomial shift.
- g_{urnc1}: Polynomial offset.

Table A.1. RNC1 User Traffic Model parameters.

Section	Function	Interval [Month]	Parameters				
Left	Gaussian	[0; 7.95]	a_{urnc1}	b_{urnc1}	c_{urnc1}	d_{urnc1}	$t_{urnc1} = 7.95$
			1.29	9.53	3.14	-1.25×10^{-4}	
Right	Rational	[7.95; 12]	p_{urnc1}	q_{urnc1}	g_{urnc1}		
			2.03	-5.92	-1.25×10^{-4}		
$R^2 = 0.970$			$\sqrt{\overline{\varepsilon^2}} = 0.0744$		$\sigma_{avg} = 0.381$		

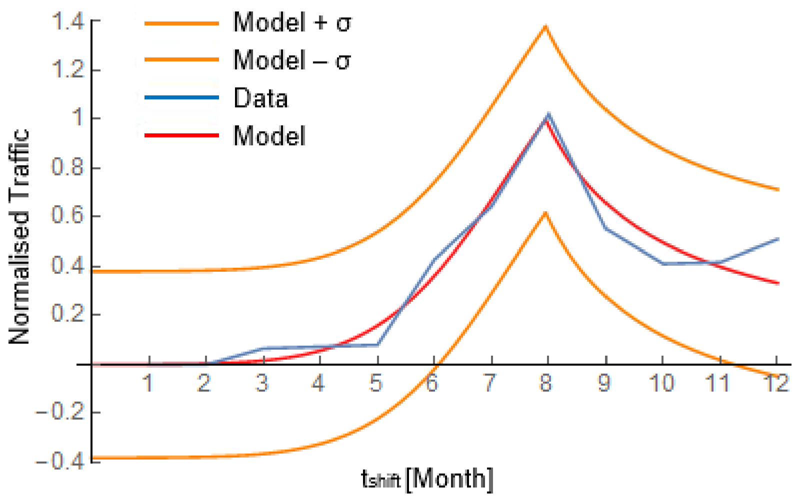

Figure A.1. RNC1 User Traffic Model, with the respective data and a confidence interval.

RNC2

$$f_{urnc2}(t_{shift}) = \begin{cases} a_{urnc2} \times e^{t_{shift} \times b_{urnc2}} + c_{urnc2} & 0 \leq t_{shift} \leq t_{urnc2} \\ \dfrac{p_{urnc2}}{t_{shift} + q_{urnc2}} + d_{urnc2} & t_{urnc2} < t_{shift} \leq 12 \end{cases} \tag{A.3}$$

with

$$f_{urnc2}(t_{urnc2}) = a_{urnc2} \times e^{t_{urnc2} \times b_{urnc2}} + c_{urnc2} = \frac{p_{urnc2}}{t_{urnc2} + q_{urnc2}} + d_{urnc2} \tag{A.4}$$

where:

- f_{urnc2}: RNC2 User Traffic Model (URNC2) function.
- a_{urnc2}: Exponential initial value.
- b_{urnc2}: Exponential decay factor.
- c_{urnc2}: Exponential offset.
- t_{urnc2}: Breakpoint shifted month.
- p_{urnc2}: Polynomial scale factor.
- q_{urnc2}: Polynomial shift.
- d_{urnc2}: Polynomial offset.

Table A.2. RNC2 User Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 7.99]	a_{urnc2}	b_{urnc2}	c_{urnc2}	t_{urnc2} = 7.99
			3.42×10^{-3}	0.711	-3.42×10^{-3}	
Right	Rational	[7.99; 12]	p_{urnc2}	q_{urnc2}	d_{urnc2}	
			1.19	-6.81	-3.42×10^{-3}	
R^2 = 0.965			$\sqrt{\overline{\varepsilon^2}}$ = 0.0709		σ_{avg} = 0.694	

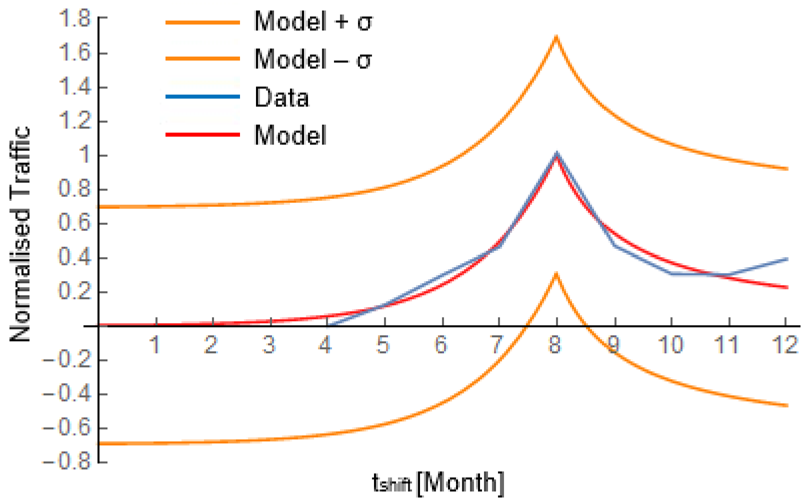

Figure A.2. RNC2 User Traffic Model, with the respective data and a confidence interval.

RNC3

$$f_{urnc3}(t_{shift}) = \begin{cases} a_{urnc3} \times e^{t_{shift} \times b_{urnc3}} + c_{urnc3} & 0 \leq t_{shift} \leq t_{urnc3} \\ \dfrac{p_{urnc3}}{t_{shift} + q_{urnc3}} + d_{urnc3} & t_{urnc3} < t_{shift} \leq 12 \end{cases} \quad \text{(A.5)}$$

with

$$f_{urnc3}(t_{urnc3}) = a_{urnc3} \times e^{t_{urnc3} \times b_{urnc3}} + c_{urnc3} = \frac{p_{urnc3}}{t_{urnc3} + q_{urnc3}} + d_{urnc3} \quad \text{(A.6)}$$

where:

- f_{urnc3}: RNC3 User Traffic Model (URNC3) function.
- a_{urnc3}: Exponential initial value.
- b_{urnc3}: Exponential decay factor.
- c_{urnc3}: Exponential offset.
- t_{urnc3}: Breakpoint shifted month.
- p_{urnc3}: Polynomial scale factor.

- q_{urnc3}: Polynomial shift.
- d_{urnc3}: Polynomial offset.

Table A.3. RNC3 User Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 7.97]	a_{urnc3}	b_{urnc3}	c_{urnc3}	$t_{urnc3} = 7.97$
			6.28×10^{-3}	0.637	-6.28×10^{-3}	
Right	Rational	[7.97; 12]	p_{urnc3}	q_{urnc3}	d_{urnc3}	
			1.47	-6.51	-6.28×10^{-3}	
$R^2 = 0.967$			$\sqrt{\varepsilon^2} = 0.0726$		$\sigma_{avg} = 0.291$	

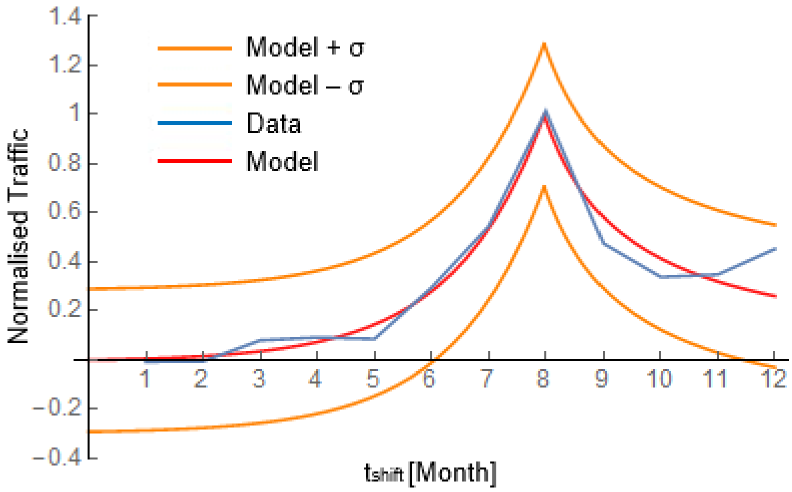

Figure A.3. RNC3 User Traffic Model, with the respective data and a confidence interval.

RNC4

$$f_{urnc4}(t_{shift}) = \begin{cases} a_{urnc4} \times e^{t_{shift}\times b_{urnc4}} + c_{urnc4} & 0 \leq t_{shift} \leq t_{urnc4} \\ d_{urnc4} \times e^{t_{shift}\times g_{urnc4}} + h_{urnc4} & t_{urnc4} < t_{shift} \leq 12 \end{cases} \quad \text{(A.7)}$$

with

$$f_{urnc4}(t_{urnc4}) = a_{urnc4} \times e^{t_{urnc4}\times b_{urnc4}} + c_{urnc4} = d_{urnc4} \times e^{t_{urnc4}\times g_{urnc4}} + h_{urnc4} \quad \text{(A.8)}$$

where:

- f_{urnc4}: RNC4 User Traffic Model (URNC4) function.
- a_{urnc4}: First exponential initial value.
- b_{urnc4}: First exponential decay factor.

- c_{urnc4}: First exponential offset.
- t_{urnc4}: Breakpoint shifted month.
- d_{urnc4}: Second exponential initial value.
- g_{urnc4}: Second exponential decay factor.
- h_{urnc4}: Second exponential offset.

Table A.4. RNC4 User Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**			
Left	Exponential	[0; 8.00]	a_{urnc4}	b_{urnc4}	c_{urnc4}	t_{urnc4} = 8.00
			4.63×10^{-5}	1.25	-4.63×10^{-5}	
Right	Exponential	[8.00; 12]	d_{urnc4}	g_{urnc4}	h_{urnc4}	
			1.45×10^{5}	-1.49	-4.63×10^{-5}	
R^2 = 0.990		$\sqrt{\overline{\varepsilon^2}}$ = 0.0473			σ_{avg} = 0.509	

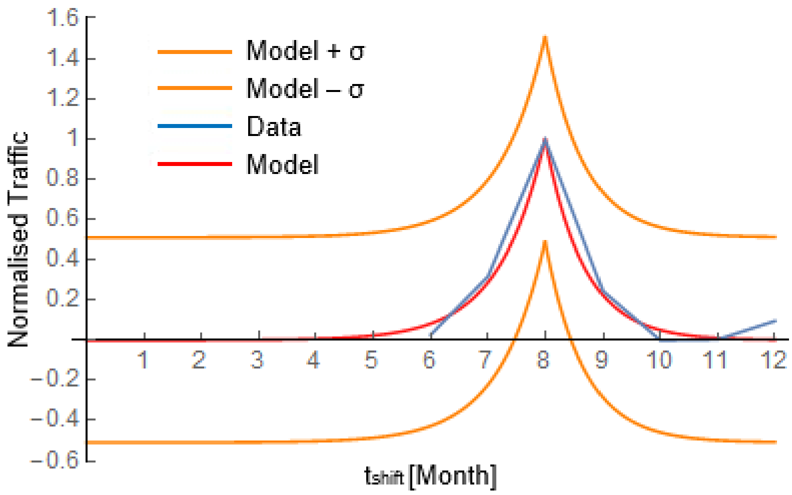

Figure A.4. RNC4 User Traffic Model, with the respective data and a confidence interval.

A.2 User Traffic per Type of Device

MBB

A traffic model called MBB Traffic Model (MTM) is developed to the average of the usage of MBB, which is composed of two exponentials, one for each section. The expression for MTM is given by:

$$f_{mtm}(t_{shift}) = \begin{cases} a_{mtm1} \times e^{t_{shift} \times b_{mtm1}} + c_{mtm1} & 0 \leq t_{shift} \leq t_{mtm} \\ a_{mtm2} \times e^{t_{shift} \times b_{mtm2}} + c_{mtm2} & t_{mtm} < t_{shift} \leq 12 \end{cases} \quad \text{(A.9)}$$

with

$$f_{mtm}(t_{mtm}) = a_{mtm1} \times e^{t_{mtm} \times b_{mtm1}} + c_{mtm1} = a_{mtm2} \times e^{t_{mtm} \times b_{mtm2}} + c_{mtm2} \quad \text{(A.10)}$$

where:

- f_{mtm}: MBB Traffic Model function.
- a_{mtm1}: First exponential initial value.
- b_{mtm1}: First exponential decay factor.
- c_{mtm1}: First exponential offset.
- t_{mtm}: Breakpoint shifted month.
- a_{mtm2}: Second exponential initial value.
- b_{mtm2}: Second exponential decay factor.
- c_{mtm2}: Second exponential offset.

The MTM is characterised by the parameters shown in Table A.5.

Table A.5. MBB Traffic Model parameters.

Section	**Function**	**Interval [Month]**	**Parameters**			
Left	Exponential	[0; 8.00]	a_{mtm1}	b_{mtm1}	c_{mtm1}	$t_{mtm} = 8.00$
			4.01×10^{-3}	0.690	-4.01×10^{-3}	
Right	Exponential	[8.00; 12]	a_{mtm2}	b_{mtm2}	c_{mtm2}	
			1.80×10^{3}	-0.936	-4.01×10^{-3}	
$R^2 = 0.988$		$\sqrt{\overline{\varepsilon^2}} = 0.0451$			$\sigma_{avg} = 0.119$	

It is visible in Figure A.5 that there is a good approximation of the developed model to the averaged data even though apparently it has been a greater compliance to MBB in the months of March and April that is not explained by the model.

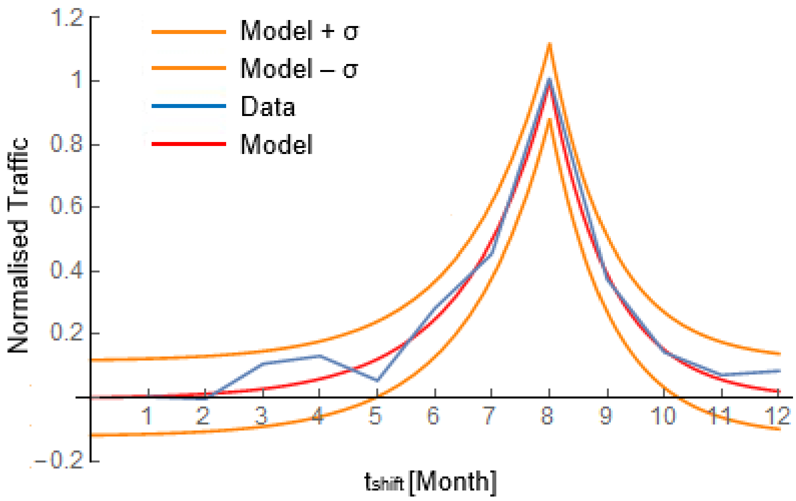

Figure A.5. MBB Traffic Model, with the respective data and a confidence interval.

Smartphone

A traffic model is also developed for this category in order to explain the behaviour of the calculated average of the 3 years (2012/2013/2014), the Smartphone Traffic Model (STM). The best fit chosen for the left section is a power function, and for the right section a rational function, as it is observed in the following expression:

$$f_{stm}(t_{shift}) = \begin{cases} a_{stm} \times t_{shift}^{b_{stm}} + c_{stm} & 0 \leq t_{shift} \leq t_{stm} \\ \dfrac{p_{stm}}{t_{shift} + q_{stm}} + d_{stm} & t_{stm} < t_{shift} \leq 12 \end{cases} \tag{A.11}$$

with

$$f_{stm}(t_{stm}) = a_{stm} \times t_{stm}^{b_{stm}} + c_{stm} = \frac{p_{stm}}{t_{stm} + q_{stm}} + d_{stm} \tag{A.12}$$

where:

- f_{stm}: Smartphone Traffic Model function.
- a_{stm}: Power initial value.
- b_{stm}: Power decay factor.
- c_{stm}: Linear function offset.
- t_{stm}: Breakpoint shifted month.
- p_{stm}: Polynomial scale factor.
- q_{stm}: Polynomial shift.
- d_{stm}: Polynomial offset.

The parameters are also calculated for the STM, as shown in Table A.6, in order to get the best approximation to the variation of the average sectors profile. Figure A.6 shows that there is a good proximity between them.

Table A.6. Smartphone Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 7.92]	a_{stm}	b_{stm}	c_{stm}	t_{stm} = 7.92
			5.39×10^{-4}	3.64	0	
Right	Rational	[7.92; 12]	p_{stm}	q_{stm}	d_{stm}	
			2.46	-5.46	0	
R^2 = 0.969		$\sqrt{\overline{\varepsilon^2}}$ = 0.0765		σ_{avg} = 0.418		

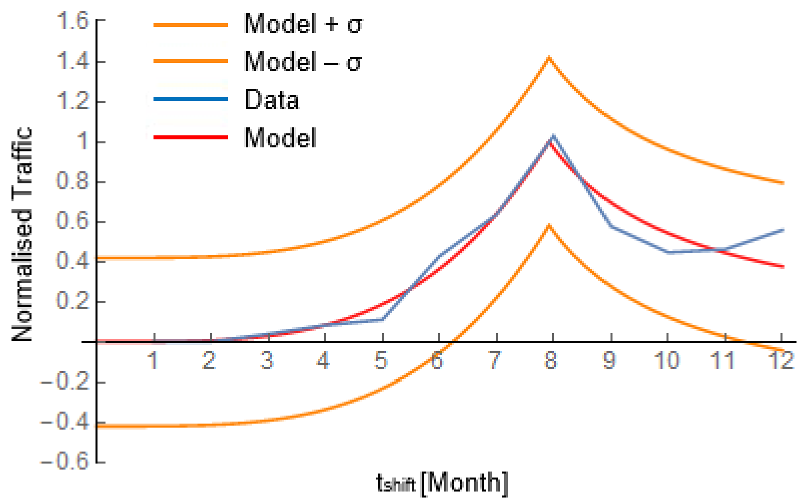

Figure A.6. Smartphone Traffic Model, with the respective data and a confidence interval.

Tablet

The traffic model for this case is called Tablet Traffic Model (TTM). It is defined by an exponential function for the left section and a rational function for the right one, because these functions are the ones that have the best fitting to the average data, as shown in the following expression:

$$f_{ttm}(t_{shift}) = \begin{cases} a_{ttm} \times e^{t_{shift} \times b_{ttm}} + c_{ttm} & 0 \leq t_{shift} \leq t_{ttm} \\ \dfrac{p_{ttm}}{t_{shift} + q_{ttm}} + d_{ttm} & t_{ttm} < t_{shift} \leq 12 \end{cases} \tag{A.13}$$

with

$$f_{ttm}(t_{ttm}) = a_{ttm} \times e^{t_{ttm} \times b_{ttm}} + c_{ttm} = \frac{p_{ttm}}{t_{ttm} + q_{ttm}} + d_{ttm} \tag{A.14}$$

where:

- f_{ttm}: Tablet Traffic Model function.
- a_{ttm}: Exponential initial value.
- b_{ttm}: Exponential decay factor.
- c_{ttm}: Exponential offset.
- t_{ttm}: Breakpoint shifted month.
- p_{ttm}: Polynomial scale factor.
- q_{ttm}: Polynomial shift.
- d_{ttm}: Polynomial offset.

The proximity between the model and the average profile is visible for the both sections in Figure A.7, and the respective parameters are shown in Table A.7.

Table A.7. Tablet Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 8.00]	a_{ttm}	b_{ttm}	c_{ttm}	$t_{ttm} = 8.00$
			0.0170	0.511	-0.0170	
Right	Rational	[8.00; 12]	p_{ttm}	q_{ttm}	d_{ttm}	
			0.908	-7.11	-0.0170	
$R^2 = 0.988$			$\sqrt{\overline{\varepsilon^2}} = 0.0426$		$\sigma_{avg} = 0.304$	

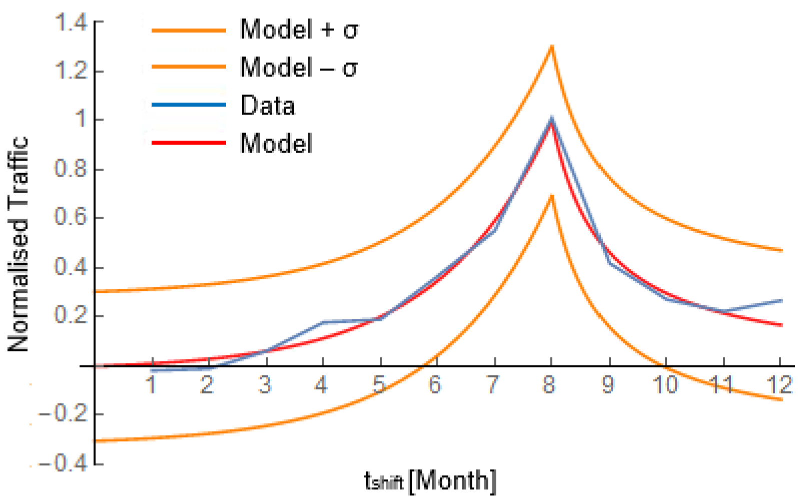

Figure A.7. Tablet Traffic Model, with the respective data and a confidence interval.

A.3 User Traffic per Operating System

Android

The developed traffic model that shows the best fit for the average profile is called Android Traffic Model (ATM), and is shown in Figure A.8: for the left section the best fit is a power function, and for the right one a rational function. The expression that is used to develop this model is:

$$f_{atm}(t_{shift}) = \begin{cases} a_{atm} \times t_{shift}^{b_{atm}} + c_{atm} & 0 \leq t_{shift} \leq t_{atm} \\ \dfrac{p_{atm}}{t_{shift} + q_{atm}} + d_{atm} & t_{atm} < t_{shift} \leq 12 \end{cases} \quad \text{(A.15)}$$

with

$$f_{atm}(t_{atm}) = a_{atm} \times t_{atm}^{b_{atm}} + c_{atm} = \frac{p_{atm}}{t_{atm} + q_{atm}} + d_{atm} \quad \text{(A.16)}$$

where:

- f_{atm}: Android Traffic Model function.
- a_{atm}: Power initial value.
- b_{atm}: Power decay factor.
- c_{atm}: Linear function offset.
- t_{atm}: Breakpoint shifted month.
- p_{atm}: Polynomial scale factor.
- q_{atm}: Polynomial shift.
- d_{atm}: Polynomial offset.

As one can see in Figure A.8, there is a fairly sharp increase over the months from May to July. These growths are always dependences from the number of devices, and it is formerly known that in the summer months, especially June, July and August, there is a huge increase in this number compared with the other months of the year. In November and December there is an increase again, which has to do with the New Year's Eve. In Table A.8 the parameters corresponding to the ATM are shown, providing a decent approximation.

Table A.8. Android Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 7.86]	a_{atm}	b_{atm}	c_{atm}	t_{atm} = 7.86
			6.52×10^{-4}	3.56	0	
Right	Rational	[7.86; 12]	p_{atm}	q_{atm}	d_{atm}	
			4.21	-3.65	0	
R^2 = 0.969		$\sqrt{\overline{\varepsilon^2}}$ = 0.0823		σ_{avg} = 0.356		

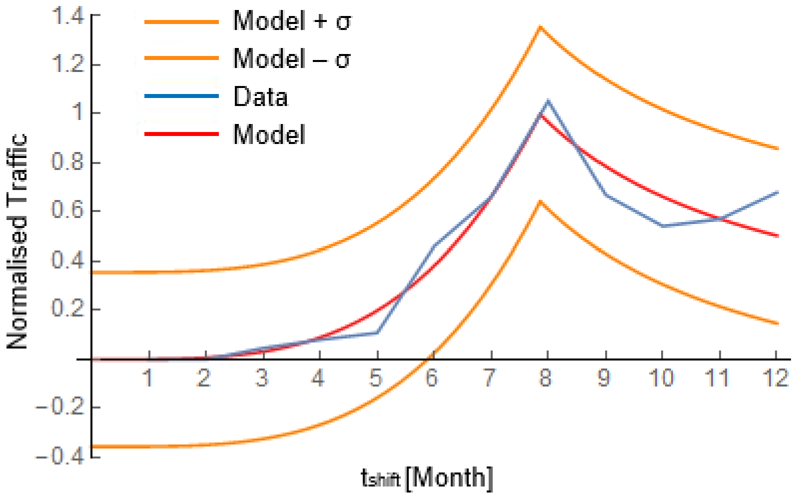

Figure A.8. Android Traffic Model, with the respective data and a confidence interval.

iOS

For the iOS usage, a model called iOS Traffic Model (iOSTM) is created, which uses one power function for the left section, until the maximum value of traffic load, and then a rational function in the right section to represent the decrease of traffic until December, as shown in Figure A.9. The expression of this model is given by:

$$f_{iostm}(t_{shift}) = \begin{cases} a_{iostm} \times t_{shift}^{b_{iostm}} + c_{iostm} & 0 \leq t_{shift} \leq t_{iostm} \\ \dfrac{p_{iostm}}{t_{shift} + q_{iostm}} + d_{iostm} & t_{iostm} < t_{shift} \leq 12 \end{cases} \quad \text{(A.17)}$$

with

$$f_{iostm}(t_{iostm}) = a_{iostm} \times t_{iostm}^{b_{iostm}} + c_{iostm} = \frac{p_{iostm}}{t_{iostm} + q_{iostm}} + d_{iostm} \quad \text{(A.18)}$$

where:

- f_{iostm}: iOS Traffic Model function.
- a_{iostm}: Power initial value.
- b_{iostm}: Power decay factor.
- c_{iostm}: Linear function offset.
- t_{iostm}: Breakpoint shifted month.
- p_{iostm}: Polynomial scale factor.
- q_{iostm}: Polynomial shift.
- d_{iostm}: Polynomial offset.

The parameters obtained for the iOSTM are shown in Table A.9.

Table A.9. iOS Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 7.96]	a_{iostm}	b_{iostm}	c_{iostm}	$t_{iostm} = 7.96$
			6.47×10^{-4}	3.54	0	
Right	Rational	[7.96; 12]	p_{iostm}	q_{iostm}	d_{iostm}	
			1.75	-6.20	0	
$R^2 = 0.970$		$\sqrt{\varepsilon^2} = 0.0707$		$\sigma_{avg} = 0.270$		

As one can see in Figure A.9, the model provides a good approximation to the respective data.

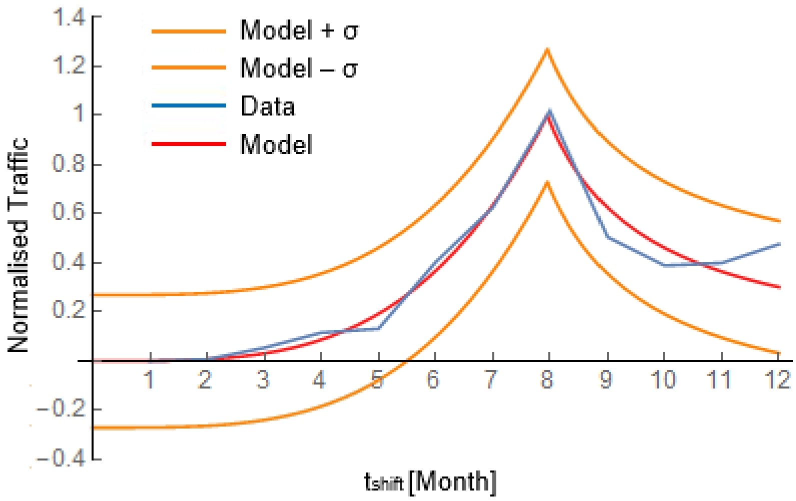

Figure A.9. iOS Traffic Model, with the respective data and a confidence interval.

Windows

The developed traffic model for the Windows usage is the Windows Traffic Model (WiTM), very similar to the ATM. A power and a rational function are used for each section, as one can see in Figure A.10, and are given by the following expression:

$$f_{witm}(t_{shift}) = \begin{cases} a_{witm} \times t_{shift}^{b_{witm}} + c_{witm} & 0 \leq t_{shift} \leq t_{witm} \\ \dfrac{p_{witm}}{t_{shift} + q_{witm}} + d_{witm} & t_{witm} < t_{shift} \leq 12 \end{cases} \tag{A.19}$$

with

$$f_{witm}(t_{witm}) = a_{witm} \times t_{witm}^{b_{witm}} + c_{witm} = \frac{p_{witm}}{t_{witm} + q_{witm}} + d_{witm} \tag{A.20}$$

where:

- f_{witm}: Windows Traffic Model function.

- a_{witm}: Power initial value.
- b_{witm}: Power decay factor.
- c_{witm}: Linear function offset.
- t_{witm}: Breakpoint shifted month.
- p_{witm}: Polynomial scale factor.
- q_{witm}: Polynomial shift.
- d_{witm}: Polynomial offset.

The values obtained for the parameters of the WiTM are shown in Table A.10.

Table A.10. Windows Traffic Model parameters.

<table>
<tr><th>Section</th><th>Function</th><th>Interval [Month]</th><th colspan="4">Parameters</th></tr>
<tr><td rowspan="2">Left</td><td rowspan="2">Power</td><td rowspan="2">[0; 7.86]</td><td>a_{witm}</td><td>b_{witm}</td><td>c_{witm}</td><td rowspan="4">$t_{witm} = 7.86$</td></tr>
<tr><td>8.20×10^{-4}</td><td>3.45</td><td>0</td></tr>
<tr><td rowspan="2">Right</td><td rowspan="2">Rational</td><td rowspan="2">[7.86; 12]</td><td>p_{witm}</td><td>q_{witm}</td><td>d_{witm}</td></tr>
<tr><td>4.13</td><td>-3.74</td><td>0</td></tr>
<tr><td colspan="3">$R^2 = 0.972$</td><td colspan="2">$\sqrt{\overline{\varepsilon^2}} = 0.0769$</td><td colspan="2">$\sigma_{avg} = 0.372$</td></tr>
</table>

The behaviour of this model is very similar to the one seen in the ATM and gives a decent approximation to the average of the data.

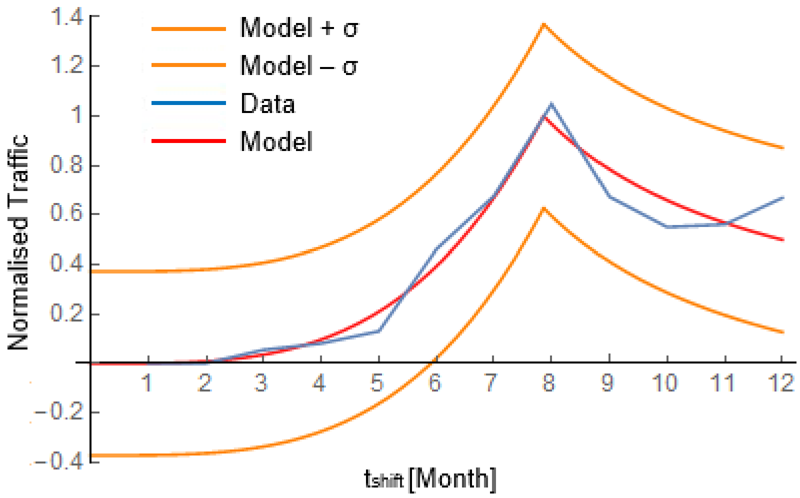

Figure A.10. Windows Traffic Model, with the respective data and a confidence interval.

Others

The data for the next model are collected from multiple operating systems still used by certain users, although increasingly scarce. The Others Traffic Model (OTM) is defined by one exponential and one rational function that are presented in Figure A.11. The use of these two functions is explained by the distinct behaviours of the traffic variation. At first there is an increase from January to August, hence the use of the exponential function, followed by a decrease of the traffic load from August to December, whose best fit is made by the rational function. The expression for the OTM is given by:

$$f_{otm}(t_{shift}) = \begin{cases} a_{otm} \times e^{t_{shift} \times b_{otm}} + c_{otm} & 0 \leq t_{shift} \leq t_{otm} \\ \dfrac{p_{otm}}{t_{shift} + q_{otm}} + d_{otm} & t_{otm} < t_{shift} \leq 12 \end{cases} \tag{A.21}$$

with

$$f_{otm}(t_{otm}) = a_{otm} \times e^{t_{otm} \times b_{otm}} + c_{otm} = \frac{p_{otm}}{t_{otm} + q_{otm}} + d_{otm} \tag{A.22}$$

where:

- f_{otm}: Others Traffic Model function.
- a_{otm}: Exponential initial value.
- b_{otm}: Exponential decay factor.
- c_{otm}: Exponential offset.
- t_{otm}: Breakpoint shifted month.
- p_{otm}: Polynomial scale factor.
- q_{otm}: Polynomial shift.
- d_{otm}: Polynomial offset.

The parameters for the Others Traffic Model are shown in Table A.11. There is an excellent approximation between the data and the model developed in relation to the whole year, as seen in Figure A.11, except in the last two months, due to variations such as the Christmas holidays among other phenomena.

Table A.11. Others Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 7.98]	a_{otm}	b_{otm}	c_{otm}	t_{otm} = 7.98
			0.0163	0.518	-0.0163	
Right	Rational	[7.98; 12]	p_{otm}	q_{otm}	d_{otm}	
			1.17	-6.83	-0.0163	
R^2 = 0.983		$\sqrt{\varepsilon^2}$ = 0.0506		σ_{avg} = 0.133		

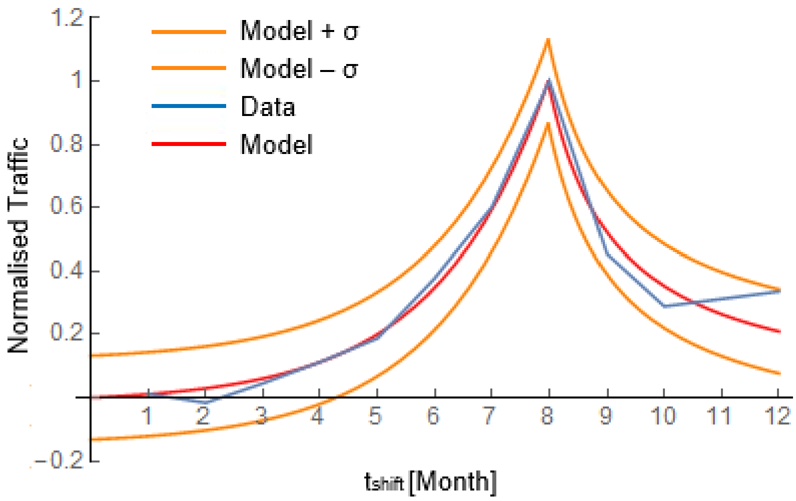

Figure A.11. Others Traffic Model, with the respective data and a confidence interval.

Unclassified

The developed data traffic model for the Unclassified usage is called Unclassified Traffic Model (UTM) and uses a power function for the left section, until the maximum value of the data traffic load, and a rational function for the right section to represent the finish of the summer and the entry for the remaining months of the year. The expression of the UTM is:

$$f_{utm}(t_{shift}) = \begin{cases} a_{utm} \times t_{shift}^{b_{utm}} + c_{utm} & 0 \leq t_{shift} \leq t_{utm} \\ \dfrac{p_{utm}}{t_{shift} + q_{utm}} + d_{utm} & t_{utm} < t_{shift} \leq 12 \end{cases} \tag{A.23}$$

with

$$f_{utm}(t_{utm}) = a_{utm} \times t_{utm}^{b_{utm}} + c_{utm} = \frac{p_{utm}}{t_{utm} + q_{utm}} + d_{utm} \tag{A.24}$$

where:

- f_{utm}: Unclassified Traffic Model function.
- a_{utm}: Power initial value.
- b_{utm}: Power decay factor.
- c_{utm}: Linear function offset.
- t_{utm}: Breakpoint shifted month.
- p_{utm}: Polynomial scale factor.
- q_{utm}: Polynomial shift.
- d_{utm}: Polynomial offset.

The Unclassified Traffic Model is characterised by the parameters shown in Table A.12. Analysing Figure A.12 one can conclude that there is a good approximation between the model and the data, although a peak occurs in June that is not visible in the approximation.

Table A.12. Unclassified Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 8.02]	a_{utm}	b_{utm}	c_{utm}	t_{utm} = 8.02
			5.65×10^{-4}	3.59	0	
Right	Rational	[8.02; 12]	p_{utm}	q_{utm}	d_{utm}	
			1.16	-6.86	0	
R^2 = 0.960			$\sqrt{\overline{\varepsilon^2}}$ = 0.0806		σ_{avg} = 0.153	

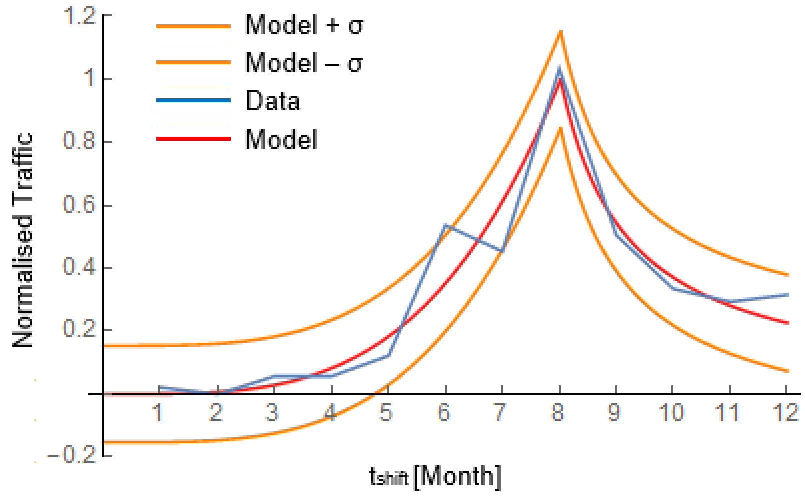

Figure A.12. Unclassified Traffic Model, with the respective data and a confidence interval.

A.4 Data Traffic per RNC

RNC1

$$f_{drnc1}(t_{shift}) = \begin{cases} a_{drnc1} \times e^{t_{shift} \times b_{drnc1}} + c_{fdrnc1} & 0 \leq t_{shift} \leq t_{drnc1} \\ \dfrac{p_{drnc1}}{t_{shift} + q_{drnc1}} + d_{drnc1} & t_{drnc1} < t_{shift} \leq 12 \end{cases} \quad \text{(A.25)}$$

with

$$f_{drnc1}(t_{drnc1}) = a_{drnc1} \times e^{t_{drnc1} \times b_{drnc1}} + c_{drnc1} = \frac{p_{drnc1}}{t_{drnc1} + q_{drnc1}} + d_{drnc1} \quad \text{(A.26)}$$

where:

- f_{drnc1}: RNC1 Data Traffic Model function.

- a_{drnc1}: Exponential initial value.
- b_{drnc1}: Exponential decay factor.
- c_{drnc1}: Exponential offset.
- t_{drnc1}: Breakpoint shifted month.
- p_{drnc1}: Polynomial scale factor.
- q_{drnc1}: Polynomial shift.
- d_{drnc1}: Polynomial offset.

Table A.13. RNC1 Data Traffic Model parameters.

<table>
<tr><th>Section</th><th>Function</th><th>Interval [Month]</th><th colspan="4">Parameters</th></tr>
<tr><td rowspan="2">Left</td><td rowspan="2">Exponential</td><td rowspan="2">[0; 8.00]</td><td>a_{drnc1}</td><td>b_{drnc1}</td><td>c_{drnc1}</td><td rowspan="4">$t_{drnc1} = 8.00$</td></tr>
<tr><td>1.59×10^{-3}</td><td>0.806</td><td>-1.59×10^{-3}</td></tr>
<tr><td rowspan="2">Right</td><td rowspan="2">Rational</td><td rowspan="2">[8.00; 12]</td><td>p_{drnc1}</td><td>q_{drnc1}</td><td>d_{drnc1}</td></tr>
<tr><td>0.467</td><td>-7.54</td><td>-1.59×10^{-3}</td></tr>
<tr><td colspan="3">$R^2 = 0.985$</td><td colspan="2">$\sqrt{\overline{\varepsilon^2}} = 0.0496$</td><td colspan="2">$\sigma_{avg} = 0.133$</td></tr>
</table>

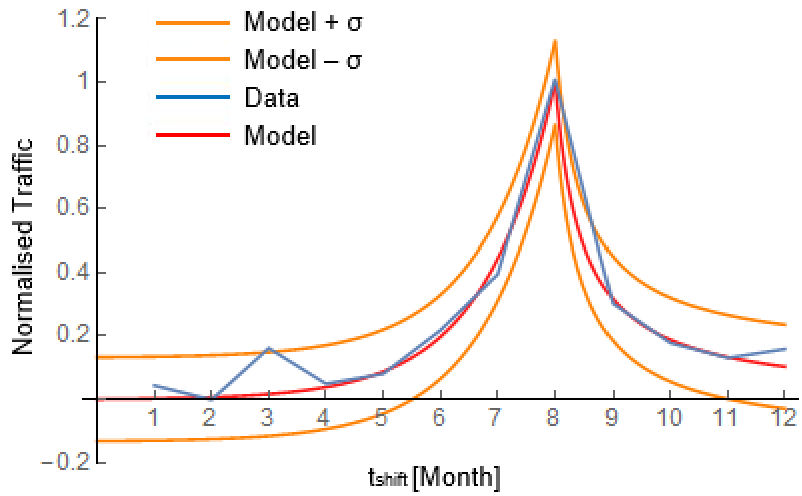

Figure A.13. RNC1 Data Traffic Model, with the respective data and a confidence interval.

RNC2

$$f_{drnc2}(t_{shift}) = \begin{cases} a_{drnc2} \times e^{t_{shift} \times b_{drnc2}} + c_{drnc2} & 0 \leq t_{shift} \leq t_{drnc2} \\ \dfrac{p_{drnc2}}{t_{shift} + q_{drnc2}} + d_{drnc2} & t_{drnc2} < t_{shift} \leq 12 \end{cases} \tag{A.27}$$

with

$$f_{drnc2}(t_{drnc2}) = a_{drnc2} \times e^{t_{drnc2} \times b_{drnc2}} + c_{drnc2} = \frac{p_{drnc2}}{t_{drnc2} + q_{drnc2}} + d_{drnc2} \tag{A.28}$$

where:

- f_{drnc2}: RNC2 Data Traffic Model (DRNC2) function.
- a_{drnc2}: Exponential initial value.
- b_{drnc2}: Exponential decay factor.
- c_{drnc2}: Exponential offset.
- t_{drnc2}: Breakpoint shifted month.
- p_{drnc2}: Polynomial scale factor.
- q_{drnc2}: Polynomial shift.
- d_{drnc2}: Polynomial offset.

Table A.14. RNC2 Data Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 8.00]	a_{drnc2}	b_{drnc2}	c_{drnc2}	$t_{drnc2} = 8.00$
			1.09×10^{-6}	1.72	-1.09×10^{-6}	
Right	Rational	[8.00; 12]	p_{drnc2}	q_{drnc2}	d_{drnc2}	
			0.410	-7.59	-1.09×10^{-6}	
$R^2 = 0.989$			$\sqrt{\overline{\varepsilon^2}} = 0.0463$		$\sigma_{avg} = 0.512$	

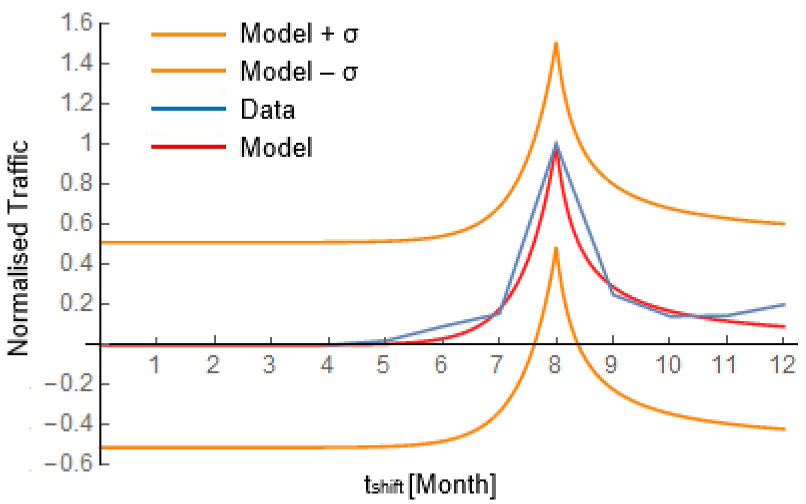

Figure A.14. RNC2 Data Traffic Model, with the respective data and a confidence interval.

RNC3

$$f_{drnc3}(t_{shift}) = \begin{cases} a_{drnc3} \times e^{t_{shift} \times b_{drnc3}} + c_{drnc3} & 0 \leq t_{shift} \leq t_{drnc3} \\ d_{drnc3} \times e^{t_{shift} \times g_{drnc3}} + h_{drnc3} & t_{drnc3} < t_{shift} \leq 12 \end{cases} \quad \text{(A.29)}$$

with

$$f_{drnc3}(t_{drnc3}) = a_{drnc3} \times e^{t_{drnc3} \times b_{drnc3}} + c_{drnc3} = d_{drnc3} \times e^{t_{drnc3} \times g_{drnc3}} + h_{drnc3} \quad \text{(A.30)}$$

where:

- f_{drnc3}: RNC3 Data Traffic Model (DRNC3) function.
- a_{drnc3}: First exponential initial value.
- b_{drnc3}: First exponential decay factor.
- c_{drnc3}: First exponential offset.
- t_{drnc3}: Breakpoint shifted month.
- d_{drnc3}: Second exponential initial value.
- g_{drnc3}: Second exponential decay factor.
- h_{drnc3}: Second exponential offset.

Table A.15. RNC3 Data Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Exponential	[0; 8.01]	a_{drnc3}	b_{drnc3}	c_{drnc3}	$t_{drnc3} = 8.01$
			2.20×10^{-4}	1.05	-2.20×10^{-4}	
Right	Exponential	[8.01; 12]	d_{drnc3}	g_{drnc3}	h_{drnc3}	
			7.07×10^{5}	-1.68	-2.20×10^{-4}	
$R^2 = 0.952$		$\sqrt{\overline{\varepsilon^2}} = 0.0953$			$\sigma_{avg} = 0.169$	

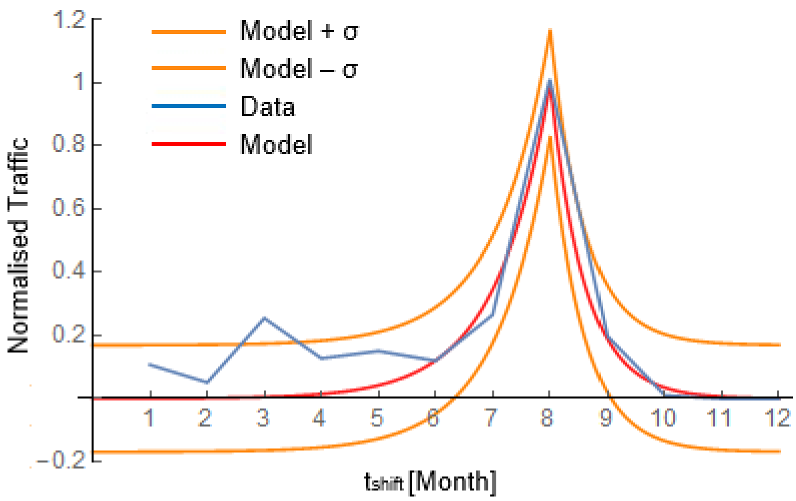

Figure A.15. RNC3 Data Traffic Model, with the respective data and a confidence interval.

RNC4

$$f_{drnc4}(t_{shift}) = \begin{cases} a_{drnc4} \times t_{shift}^{b_{drnc4}} + c_{drnc4} & 0 \leq t_{shift} \leq t_{drnc4} \\ d_{drnc4} \times e^{t_{shift} \times g_{drnc4}} + h_{drnc4} & t_{drnc4} < t_{shift} \leq 12 \end{cases} \tag{A.31}$$

with

$$f_{drnc4}(t_{drnc4}) = a_{drnc4} \times t_{drnc4}^{b_{drnc4}} + c_{drnc4} = d_{drnc4} \times e^{t_{drnc4} \times g_{drnc4}} + h_{drnc4} \tag{A.32}$$

where:

- f_{drnc4}: RNC4 Data Traffic Model (DRNC4) function.
- a_{drnc4}: Power initial value.
- b_{drnc4}: Power decay factor.
- c_{drnc4}: Linear function offset.
- t_{drnc4}: Breakpoint shifted month.
- d_{drnc4}: Exponential initial value.
- g_{drnc4}: Exponential decay factor.
- h_{drnc4}: Exponential offset.

Table A.16. RNC4 Data Traffic Model parameters.

Section	Function	Interval [Month]	Parameters			
Left	Power	[0; 8.00]	a_{drnc4}	b_{drnc4}	c_{drnc4}	$t_{drnc4} = 8.00$
			4.79×10^{-10}	10.3	0	
Right	Exponential	[8.00; 12]	d_{drnc4}	g_{drnc4}	h_{drnc4}	
			9.09×10^{5}	-1.72	0	
$R^2 = 0.999$			$\sqrt{\overline{\varepsilon^2}} = 0.0156$		$\sigma_{avg} = 0.186$	

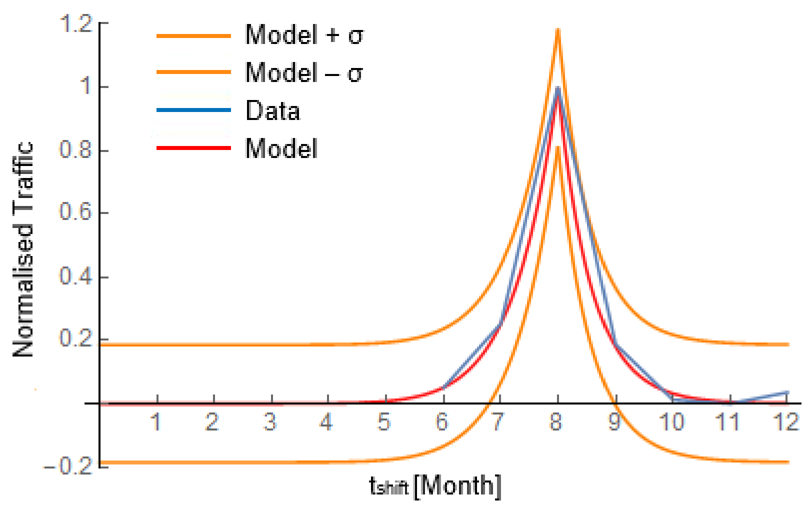

Figure A.16. RNC4 Data Traffic Model, with the respective data and a confidence interval.

Annex B

Prediction Models

In this annex, the prediction models of the RNCs are presented.

B.1 User Traffic per RNC

RNC1

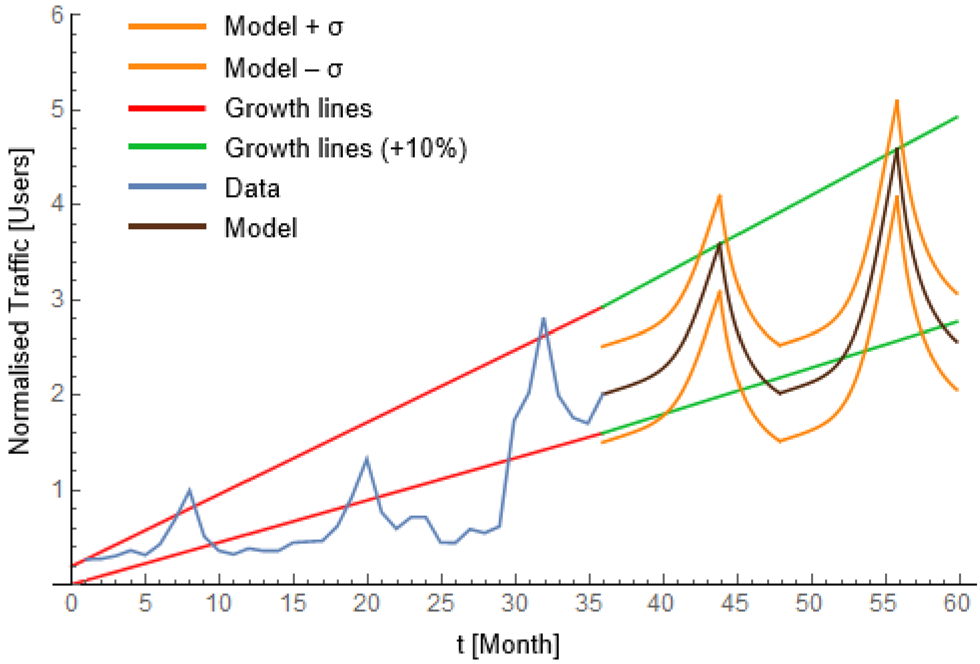

Figure B.1. Prediction of the number of users of the RNC1 for 2015 and 2016.

It is predicted that the number of users of this RNC in 2015 will be 1.28 times bigger than the one registered in 2014, and that in 2016 it will be 1.63 times greater than in 2014 too.

RNC2

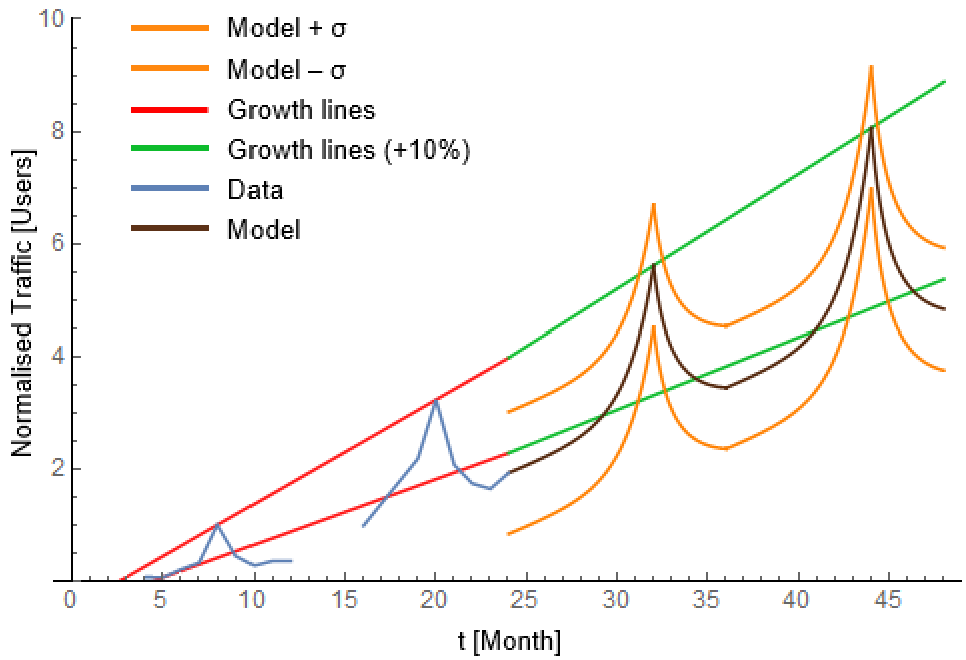

Figure B.2. Prediction of the number of users of the RNC2 for 2015 and 2016.

It is predicted that the number of users of this RNC in 2015 will be 1.74 times bigger than the one registered in 2014, and that in 2016 it will be 2.49 times greater than in 2014 too.

RNC3

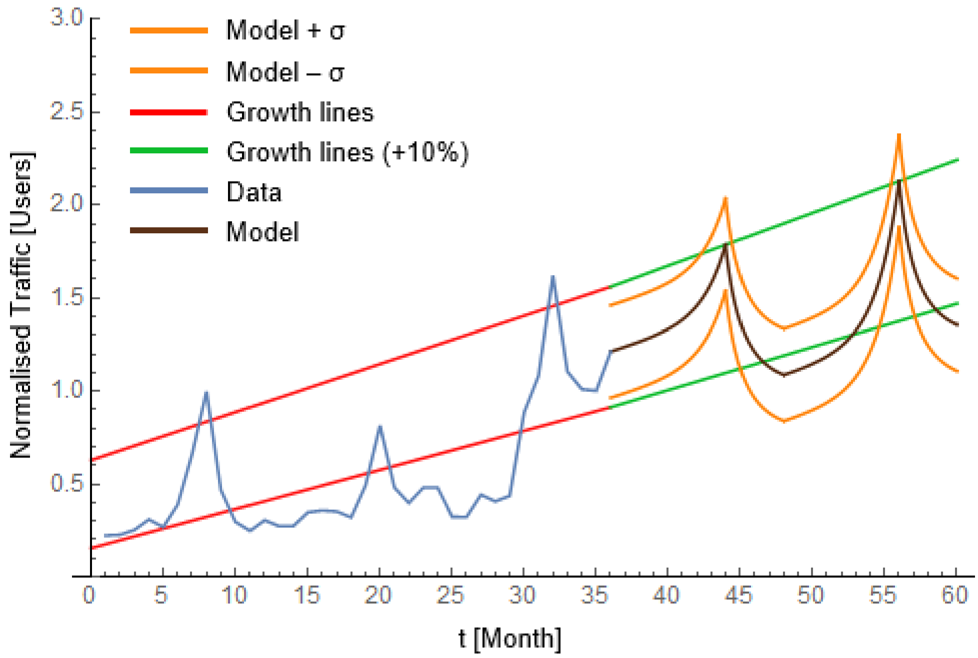

Figure B.3. Prediction of the number of users of the RNC3 for 2015 and 2016.

It is predicted that the number of users of this RNC in 2015 will be 1.10 times bigger than the one registered in 2014, and that in 2016 it will be 1.32 times greater than in 2014 too.

RNC4

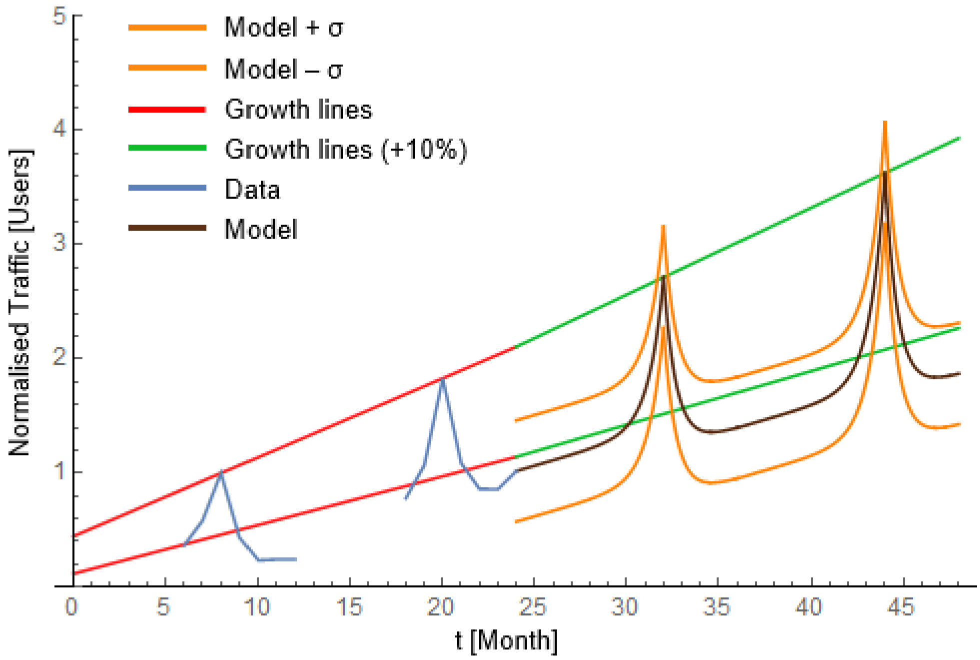

Figure B.4. Prediction of the number of users of the RNC4 for 2015 and 2016.

It is predicted that the number of users of this RNC in 2015 will be 1.48 times bigger than the one registered in 2014, and that in 2016 it will be 1.98 times greater than in 2014 too.

B.2 Usage of Different Types of Devices

MBB

Figure B.5 shows the prediction of the MBB usage for 2015 and 2016. It should be noted that MBB is a prerequisite for collaborating, socialising and innovating in the Networked Society [Eric15a]. Operators with the best performing networks are the most successful in terms of growth, revenue and valuation.

It is possible to observe that 2013 is the year in which there was a lower usage of the MBB, but in 2014 it returned to gain more strength, including having the highest peak in relation to the previous years. For August 2015 it is expected that the MBB usage will be 1.01 times greater than in August 2014, and 1.06 times greater in August 2016 compared to August 2014 too. Thus, the increase will not be significant to worry about, except if one despises the 2012 data. If one only considers the 2013 and 2014 data, it is possible to notice that the MBB usage growth from one summer to another was very high, phenomenon that can occur again in the next two years and that consequently needs some attention in this thesis.

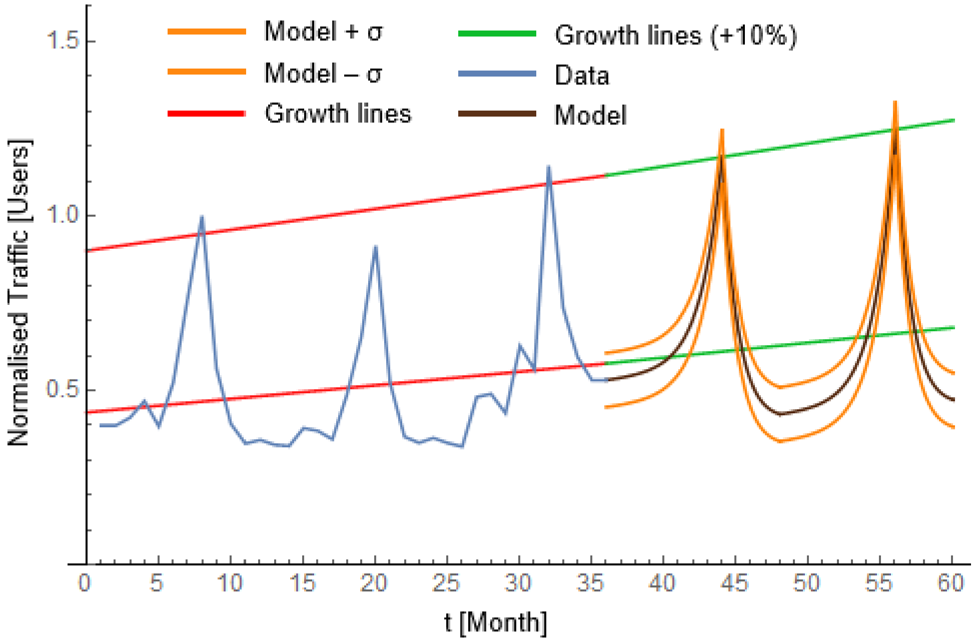

Figure B.5. Prediction of the usage of MBB for 2015 and 2016.

Tablet

Tablet usage over the years can be seen in Figure B.6.

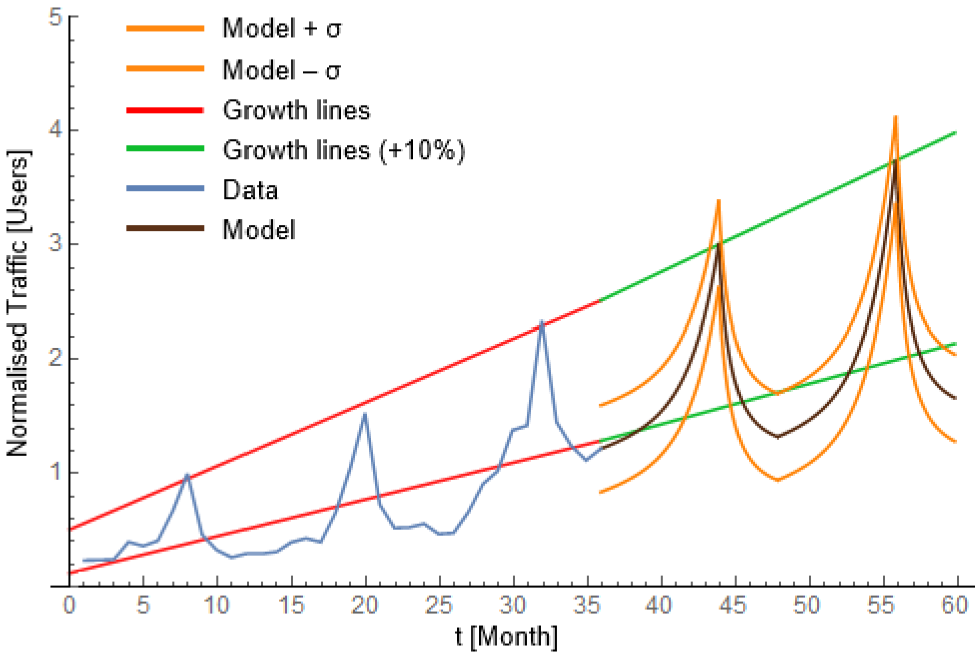

Figure B.6. Prediction of the usage of Tablets for 2015 and 2016.

As one can see, there is a continuous growth along the years which means that the market of Tablets continues in high growth, promoting a strong competition between the major manufacturers to rapidly produce and develop devices with high-quality hardware and software. In order to anticipate possible network saturation, it is concluded that the network in 2015 and 2016 must be prepared for a Tablet traffic 1.24 and 1.58 times greater than the one observed in 2014, respectively.

B.3 Usage of Different Operating Systems

iOS

Figure B.7 shows the tendency of growth of the iOS operating system, which is the OS deployed in the second largest number of mobile devices. In order to understand the possible annual behaviour of this OS for 2015 and 2016, the prediction is made with a 10% higher growth again. Analysing the figure, one can see that the iOS usage in 2014 was 2.04 times bigger than the one registered in 2013, and that it is predicted that it will be 1.53 and 2.10 times greater in 2015 and 2016, respectively, compared to 2014.

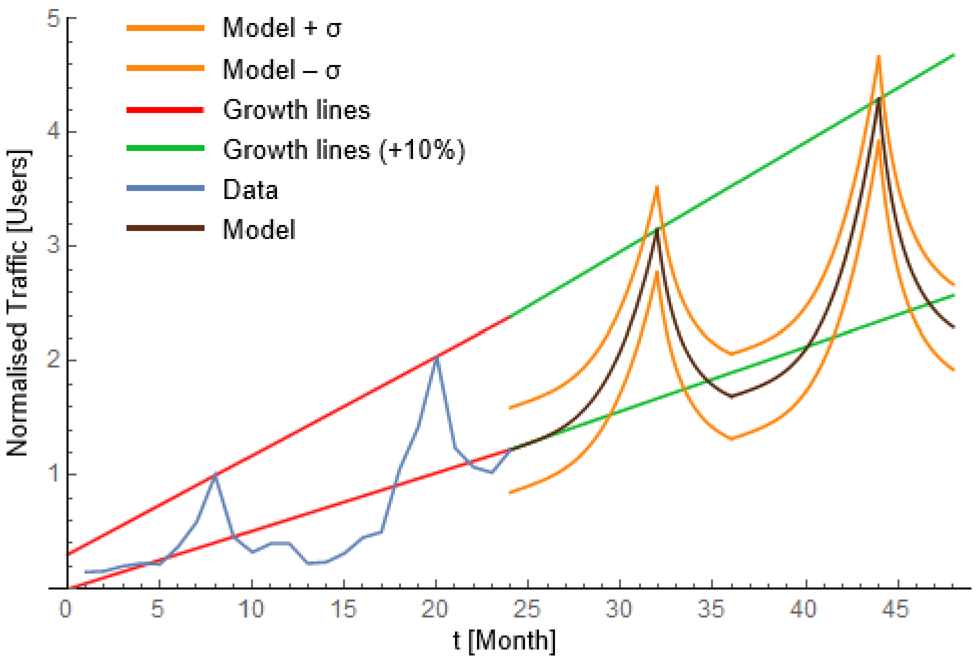

Figure B.7. Prediction of the usage of the iOS for 2015 and 2016.

Windows

The Windows is the operating system deployed in the third largest number of mobile devices, even though it modestly begins to gradually gain more market. Figure B.8 shows that there was a significant increase in 2014 over the previous year. In order to predict the users' behaviour for this operating system, a 10% higher growth was imposed again. It is found that the Windows' usage in August 2015, the time of the year when a greater number of users occurs, will probably be 1.66 times greater than in 2014, and that this value will be 2.38 times greater in 2016 than 2014 too.

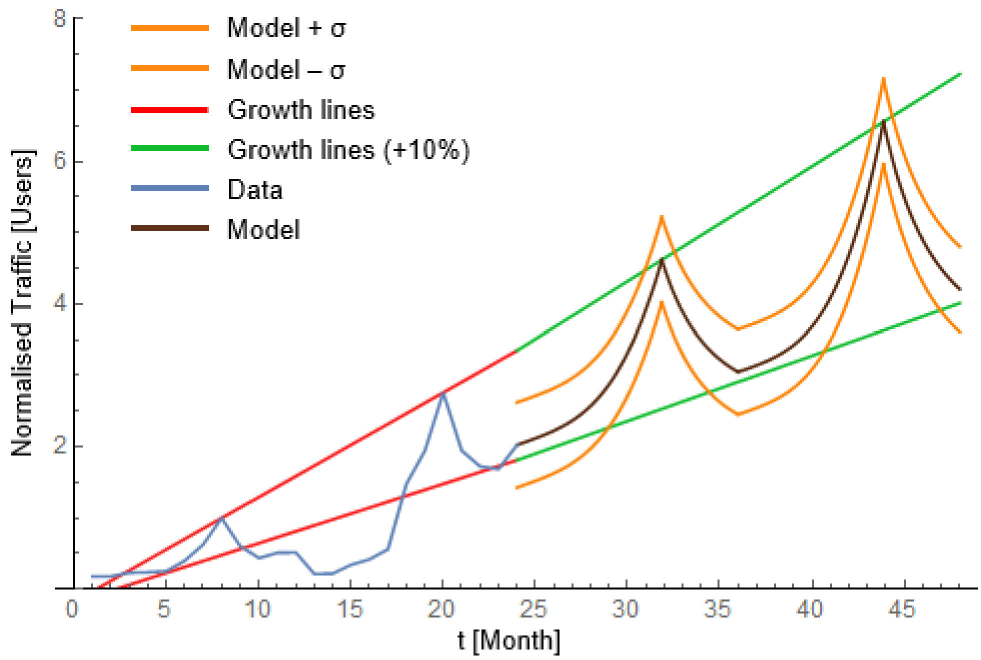

Figure B.8. Prediction of the usage of the Windows for 2015 and 2016.

Others

Figure B.9 shows the usage of the other operating systems that are expressionless in the market. The model developed for this case does not work well, because the two rising trend lines (green ones) intersect on May 2016, causing the model to fail. Despite this, a small decrease in the usage of the rest of the operating systems is expected in 2015 (the model does not fail in this year; the user traffic in 2015 will correspond to 98.3% of the traffic in 2014) and then it will probably tend to a constant value or even decrease more. In other words, it is predicted that these OSs will no longer have commercial speech in few years due to the growth of the major ones.

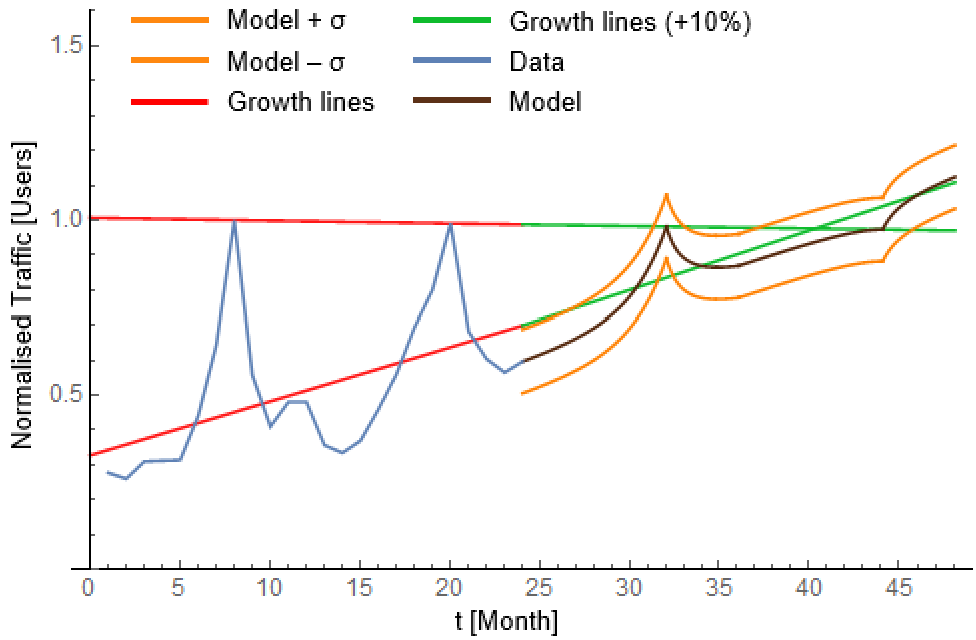

Figure B.9. Prediction of the usage of the other operating systems for 2015 and 2016.

Unclassified

In relation to the not defined operating systems, Figure B.10 shows the prediction of this evolution for 2015 and 2016 too. There is a slight upward trend between 2014 and 2015 (traffic 1.18 times bigger in 2015 in relation to 2014) and between 2015 and 2016 (traffic 1.36 times bigger in 2016 compared to 2014).

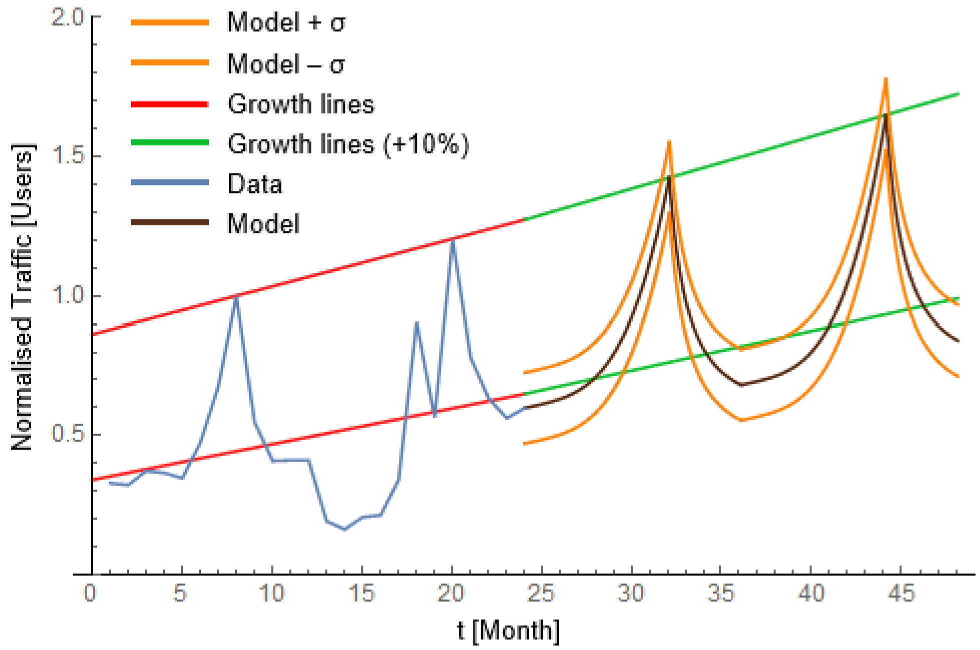

Figure B.10. Prediction of the usage of the Unclassified for 2015 and 2016.

B.4 Data Traffic per RNC

RNC1

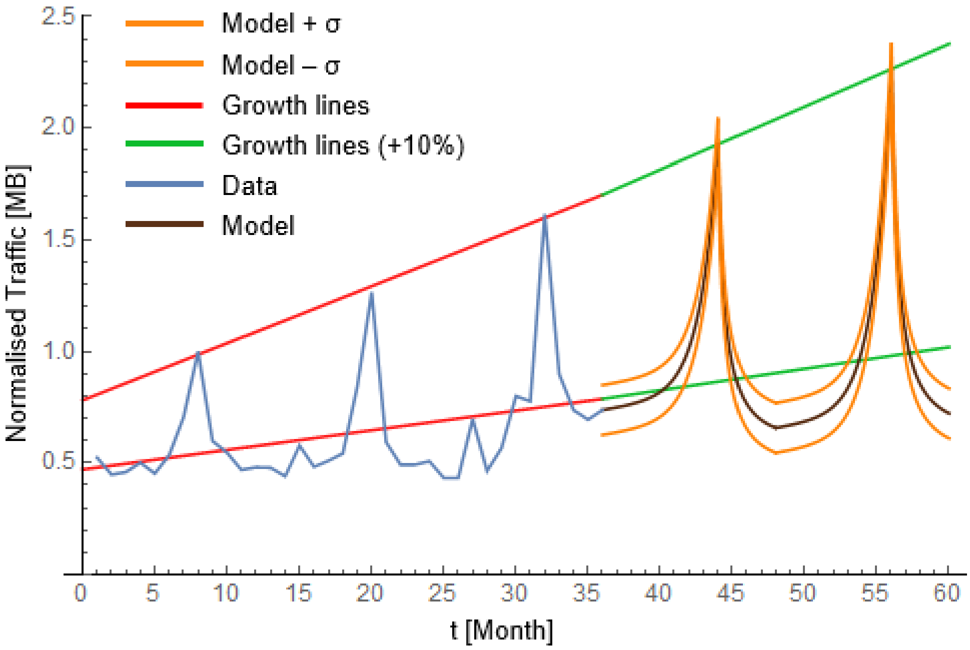

Figure B.11. Prediction of the data traffic of the RNC1 for 2015 and 2016.

It is predicted that the data traffic of this RNC in 2015 will be 1.19 times bigger than the one registered in 2014, and that in 2016 it will be 1.40 times greater than in 2014 too.

RNC2

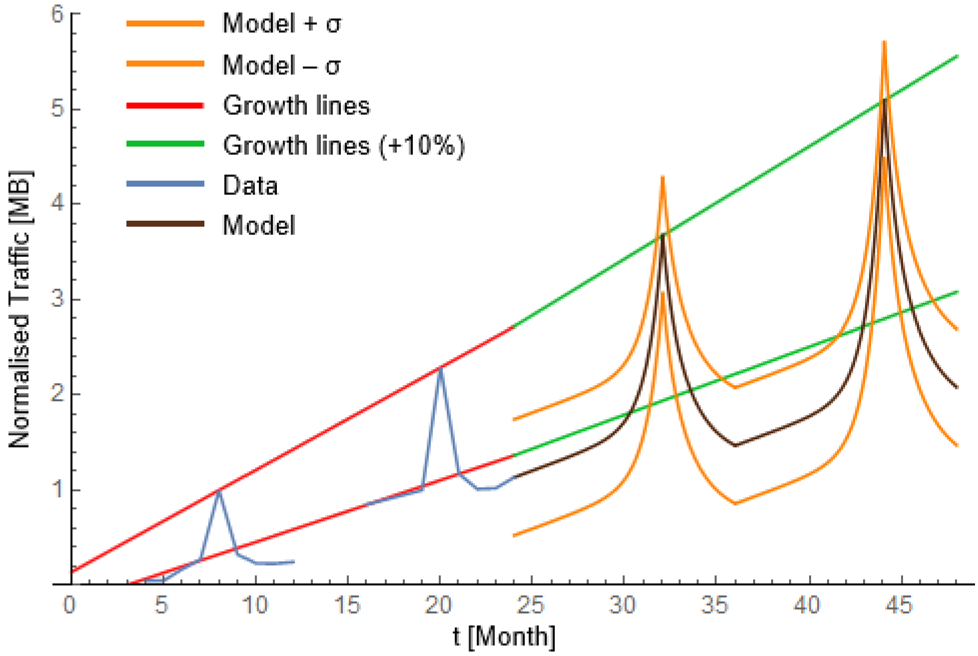

Figure B.12. Prediction of the data traffic of the RNC2 for 2015 and 2016.

It is predicted that the data traffic of this RNC in 2015 will be 1.61 times bigger than the one registered in 2014, and that in 2016 it will be 2.24 times greater than in 2014 too.

RNC3

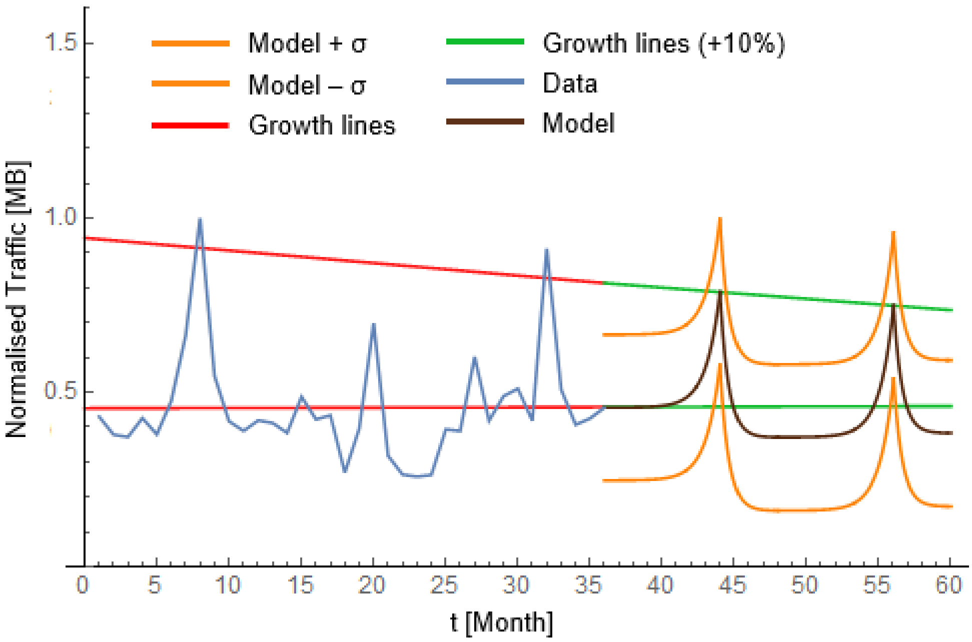

Figure B.13. Prediction of the data traffic of the RNC3 for 2015 and 2016.

It is predicted that the data traffic of this RNC in 2015 will correspond to 86.7% of the traffic registered in 2014, and that in 2016 it will correspond to 83.6% of the traffic in 2014 too.

RNC4

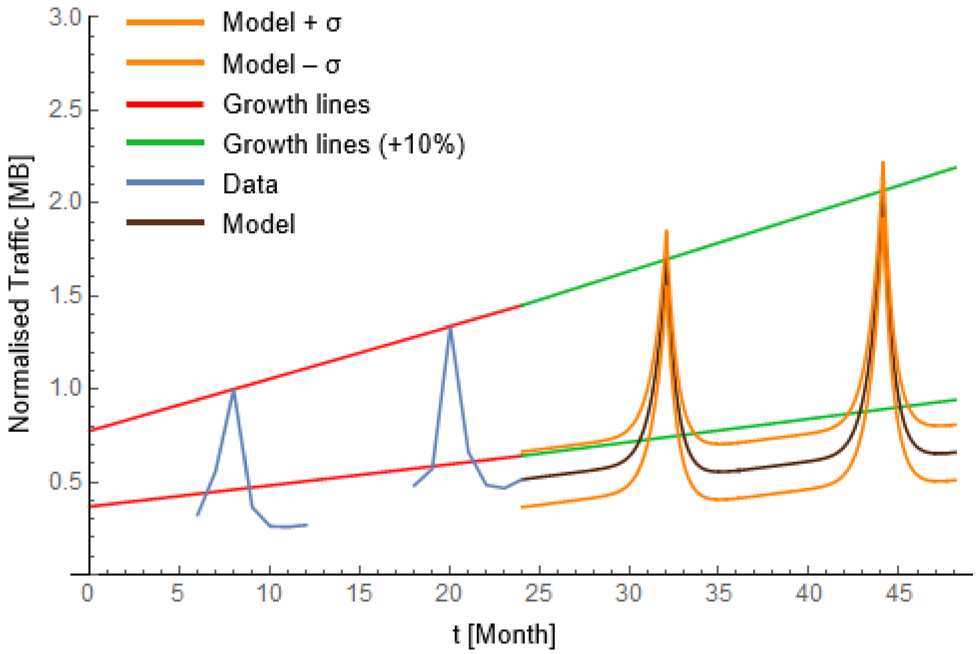

Figure B.14. Prediction of the data traffic of the RNC4 for 2015 and 2016.

It is predicted that the data traffic of this RNC in 2015 will be 1.27 times bigger than the one registered in 2014, and that in 2016 it will be 1.63 times greater than in 2014 too.

References

[3GPP05] 3GPP, Technical Specification Group - Services and System Aspects, *Services and Service Capabilities (Release 6),* Report TS 22.105, V6.4.0, Sep. 2005 (http://www.3gpp.org/ftp/Specs/html-info/22105.htm).

[3GPP11] 3GPP, *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Quality of Service (QoS) concept and architecture*, ETSI TS, No. 23.107, Ver. 10.1.0, June 2011 (http://www.3gpp.org).

[3GPP14] http://www.3gpp.org/about-3gpp, Nov. 2014.

[ANAC14] ANACOM, *Information Internet Access Service of Statistics* (in Portuguese), Public Consultation, Portugal, Oct. 2014 (www.anacom.pt/streaming/SAI_1trimestre2014.pdf?contentId=1236152&field=ATTACHED_FILE).

[Cisc13] Cisco VNI Forecast, *Cisco Visual Networking Index Global Mobile Data Traffic Forecast Update, 2012-2017*, Feb. 2013 (http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf).

[Cisc14] Cisco VNI Forecast, *Cisco Visual Networking Index Global Mobile Data Traffic Forecast Update, 2013-2018*, Feb. 2014 (http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf).

[Cisc15] Cisco VNI Forecast, *Cisco Visual Networking Index Global Mobile Data Traffic Forecast Update, 2014-2019*, Feb. 2015 (http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf).

[CMLa15] http://www.cm-lagoa.pt/menu/853/sala-de-desporto--escola-eb-23-de-est%C3%B4mbar.aspx, Feb. 2015.

[Coms15] http://www.comscore.com/, Mar. 2015.

[Corr10] Correia,L.M., *Mobile Communication System – Course Notes*, Instituto Superior Técnico, Lisbon, Portugal, 2010.

[DoCh14] Dong,H. and Chu,L., “Urban Traffic Commuting Analysis Based on Mobile Phone Data”, in *Proc. of ITSC - 17th IEEE International Conference on Intelligent Transportation Systems*, Qingdao, China, Oct. 2014.

[Eric11] Ericsson ConsumerLab, *From Apps to Everyday Situations*, Public Consultation, Stockholm, Sweden, 2011 (http://www.ericsson.com/res/docs/2011/silicon_valley_brochure_letter.pdf).

[Eric13] Ericsson, *Ericsson White Paper, Big Data Analytics 13,* Public Consultation, Stockholm, Sweden, Aug. 2013 (http://www.ericsson.com/res/docs/whitepapers/wp-big-data-v7.pdf).

[Eric14] Ericsson, *Ericsson Mobility Report, On the Pulse of the Networked Society*, Public Consultation, Stockholm, Sweden, June 2014 (http://www.ericsson.com/res/docs/2014 /ericsson-mobility-report-november-2014 .pdf).

[Eric15a] Ericsson, *Ericsson Mobility Report, On the Pulse of the Networked Society*, Public Consultation, Stockholm, Sweden, June 2015 (http://www.ericsson.com/res/docs/2015 /ericsson-mobility-report-june-2015.pdf).

[Eric15b] http://www.ericsson.com/news/150217-3g-smartphones-optimizing-user-experience-and-network-efficiency_244069647_c, June 2015.

[ETSI04] ETSI, *Technical Report, Universal Mobile Telecommunications System (UMTS); QoS Concept*, Public Consultation, Valbonne, France, June 2004 (ftp://www.3gpp.org /tsg_sa/WG2_Arch /TSGS2_04/tempdoc/S2-99223.doc).

[EvSk10] Everitt,B.S. and Skrondal,A., *The Cambridge Dictionary of Statistics*, Cambridge University Press, Cambridge, UK, 2010.

[FaLy10] Falaki,H. and Lymberopoulos,D., "A first look at traffic on Smartphones," in *Proc. of IMC – 10th ACM Conference on Internet measurement*, New York, USA, Nov. 2010.

[FaMa10] Falaki,H. and Mahajan,R., "Diversity in Smartphone usage", in *Proc. of MobiSys – 8th ACM International Conference on Mobile systems, applications, and services*, New York, USA, June 2010.

[Frad11] Fradinho,V., *Advanced Performance Analysis of GSM and UMTS Radio Networks*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2011.

[FTMV12] Benevenuto,F., Rodrigues,T., Cha,M. and Almeida,V., "Characterising user navigation and interactions in online social networks", *Information Sciences*, Vol. 195, July 2012, pp.1-24.

[GISt14] http://gs.statcounter.com/, Nov. 2014.

[Glob15] http://globalwebindex.net, Mar. 2015.

[Goog15] http://googlemobileads.blogspot.pt/2011/09/new-search-data-reveal-that-when-sun.html, Apr. 2015.

[HoTo04] Holma,H. and Toskala,A., *WCDMA for UMTS – Radio Access For Third Generation Mobile Communications*, Wiley & Sons, Chichester, UK, 2004.

[HoTo07] Holma,H. and Toskala,A., *HSDPA/HSUPA for UMTS – High Speed Radio Access for Mobile Communications*, Wiley & Sons, Chichester, UK, 2007.

[HUAW12] Huawei Technologies, *Smartphone Solutions White Paper*, Shenzhen, China, 2012 (www.huawei.com/ilink/en/download/HW_19303).

[IDCA15] http://www.idc.com/getdoc.jsp?containerId=prUS23355112, Jan. 2015.

[INFO15] http://info.abril.com.br/noticias/tecnologia-pessoal/uso-de-apps-para-mensagens-supera-sms-pela-primeira-vez-29042013-8.shl, Mar. 2015.

[Khat14] Khatibi,S., *Radio Resource Management Strategies in Virtual Networks*, Ph.D. Thesis draft, Instituto Superior Técnico, Lisbon, Portugal, 2014.

[Küpp05] Küpper,A., *Location-based Services: Fundamentals and Operation*, John Wiley & Sons, Chichester, UK, 2005.

[LiGa12] Liu,H. and Gatmir-motahari,S., "Evolving Landscape of Cellular Network Traffic", in *Proc. of ICCCN – 21st IEEE International Conference on Computer Communications and Networks*, Munich, Germany, July 2012.

[Loca15] www.pando.com, Mar. 2015.

[Lope08] Lopes,J., *Performance of UMTS/HSDPA/HSUPA at the Cellular Level*, M.Sc. Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2008.

[Mato07] Matos,M., *Standardisation and metrics* (in Portuguese), Public Consultation, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, 2014 (https://web.fe.up.pt/~mam/normetr.pdf).

[Micr14] https://msdn.microsoft.com/en-us/library/bb742409.aspx, Dec. 2014.

[Moli11] Molisch,A.F., *Wireless Communications*, John Wiley & Sons, Chichester, UK, 2011.

[NMSh14] http://marketshare.hitslink.com/, Nov. 2014.

[PaKi12] Park,S. and Kim,J., "Effect of LTE Service Adoption on Mobile Data Traffic in Korea", in *Proc. of ICCCT – 7th IEEE International Conference on Computing and Convergence Technology*, Seoul, South Korea, Dec. 2012.

[Pent11] Penttinen,J., *The LTE/SAE Deployment Handbook*, John Wiley & Sons, Chichester, UK, 2011.

[Pred14] https://practicalanalytics.wordpress.com/predictive-analytics-101/, Dec. 2014.

[ShWa11] Shafiq,M. and Wang,J., "Characterising and modelling internet traffic dynamics of cellular devices", in *Proc. of SIGMETRICS – ACM International Conference on Measurement and modelling of computer systems*, New York, USA, June 2011.

[SiGG04] Silberschatz,A., Galvin,P.B. and Gagne,G., *Operating Systems with Java* (in Portuguese), Editora Campus-Elsevier, Rio de Janeiro, Brazil, 2004.

[Stat15] http://www.statista.com/, Mar. 2015.

[Stuh12] Stuhlfauth,R., *UMTS Long Term Evolution (LTE),* Leipzig, Germany, 2012 (http://www.slideshare.net/RohdeSchwarzNA/lte-eutran-rsanov2012day1).

[SWIF15] https://www.swift.com/about_swift/company_information/fin_traffic_new, Feb. 2015.

[TuPo15] http://www.turismodeportugal.pt/Portugu%C3%AAs/ProTurismo/estat%C3%ADsticas/quadrosestatisticos/Pages/Quadrosestat%C3%ADsticos.aspx, Feb. 2015.

[ViOE11] Villars,R.L., Olofson,C.W. and Eastwood,M., *Big Data: What It is and Why you should care*, Public Consultation, USA, June 2011 (http://sites.amd.com/us/Documents/IDC_AMD_Big_Data_Whitepaper.pdf).